

Network Availability based Service Differentiation

M. Durvy *, C. Diot †, N. Taft † and P. Thiran *

* Inst. of Communication Systems (LCA, I&C) at EPFL

† Sprint Advanced Technology Labs

{mathilde.durvy, patrick.thiran}@epfl.ch, {nina, cdiot}@sprintlabs.com

Abstract

Numerous approaches have been proposed to manage Quality of Service in the Internet. However, up to now none of them was successfully deployed in a commercial IP backbone, mostly because of their complexity. In this paper, we take advantage of the excess network bandwidth to offer a degraded class of traffic. We identify and analyze the impact of link failures on such a service and show that under certain circumstances they provide the main vector to service differentiation. We simulate our QoS scheme on a real IP backbone network and derive Service Level Agreements for the new degraded service. We find that by adding a degraded class of traffic in the network, we can at least double the link utilization with no impact on the current backbone traffic.

I. INTRODUCTION

Although link failures occur everyday in backbone networks [1], providers are able to guarantee impressive traffic performance at the price of significant over-dimensioning. In other words, enough spare bandwidth must be made available to reroute the traffic without degradation of performance in case of link failures. In the absence of failure, the excess bandwidth is unused and the average resource utilization is low. This waste is a major concern for providers who are always looking for new ways to maximize the return on investment from their network infrastructure. The target of this work is to provide a mean to increase the network traffic load without any penalty on the current backbone traffic (denoted by legacy traffic). The problem is non trivial since simply scaling the amount of the legacy traffic would result in performance degradation.

We propose to take advantage of the unused bandwidth to offer a degraded class of service. The legacy traffic remains served with an absolute priority. Degraded traffic is added if bandwidth is available, in such a way that it does not affect the performance of the legacy traffic. Its performance will thus be very sensitive to the total link load and to link failures. More specifically, if there is no link failure and if the links are not overloaded (which translates in a link utilization below 80% [2]), the performance of the two traffic classes should be very similar. Instead, in the event of a link failure, the degraded service is dropped, if necessary, to accommodate the legacy traffic that cannot suffer from network outages. As a consequence the degraded service can endure severe disruption to protect the legacy traffic.

The contribution of this work are threefold:

- 1) This is the first time the impact of link failures is taken into consideration in the definition and evaluation of a QoS scheme in backbone networks.
- 2) We analyze by simulation the performance of the degraded service for a real IP backbone network topology.
- 3) We identify and define a new SLA metric, service availability, in order to capture the service uptime as perceived by the users.

In today's backbone networks, failures are a potential cause of congestion and packet losses and are thus a main problem in maintaining high quality services. Therefore, failure events should be included in any study relative to the deployment of a QoS scheme in IP networks. We define the *network availability* to be the percentage of the total network bandwidth available to route the traffic. When averaged over time, network availability effectively reflects the frequency and impact of failures in a given network. We perform a thorough analysis of the relative performance of the two traffic classes based on realistic failure patterns. We observe that if the total traffic remains under 80% of the link capacity in the no failure state, the difference of performance between the two classes of service is mainly explained in terms of network availability.

Despite the multitude of existing QoS models, none was successfully deployed in the Internet. It is too easy to blame the conservatism of network engineers. In fact most of the proposed schemes involve a significant increase in complexity while providing no explicit, or difficult to enforce, end-to-end guarantees. In response to these issues, our two-class scheme is solely based on local, and very simple decisions. We run an extensive set of *ns-2* simulations using the Sprint domestic backbone topology and its observed failure patterns. We study the range of Service Level Agreements that can be offered to the degraded class of traffic. We introduce a new SLA metric named *service availability*. We define service availability as the fraction of time the service is available to a customer. In failure prone environment, packet loss and delay may reach level where most applications are unable to function properly. In such cases we consider the service to be unavailable to the customer. We believe that service availability is an important parameter in the quality of a service perceived by the user, and should thus be included in the SLAs.

The degraded class of traffic proposed in this work is in line with the strategy of some ISPs that recently started to provide a "no SLA" service (even though this service is not proposed in conjunction with a high quality class of traffic). We also believe that our solution is of special interest for networks with very volatile bandwidth such as wireless networks. In such environments, it might be difficult to predict what resource will be available at the time it is needed and a very simple QoS mechanism that provides absolute priority to a subset of the traffic may ultimately be the only feasible approach. However, we choose to validate our approach on the Sprint IP backbone network that represents a more complex and demanding environment in terms of traffic performance and network availability.

The remainder of the paper is organized as follows. Section II defines the two classes of service and discuss implementation related issues. In Section III, we introduce important characteristics of IP backbones including current network engineering practices and recent data on failure patterns. Section IV explains the role of Service Level Agreements in commercial networks and provide an enhanced definition. In Section V and VI we present experimental settings and results. We discuss related work and conclude in Section VII and VIII respectively.

II. SERVICE DEFINITION AND IMPLEMENTATION

We now give an overview of the two service classes and discuss implementation related issues.

A. *Legacy Service: Fully Available (FA)*

The Fully Available traffic corresponds to the existing traffic in backbone networks. Its performance is defined by traffic engineering rules. Only unlikely events could visibly affect its performance. The network is designed in such a way that the FA service is available 99.999% of the time, end-to-end delays are close to the propagation delays and loss are below 0.3%. FA traffic is currently the default service available on major Tier-1 backbone networks.

B. *Degraded Service: Partially Available (PA)*

The Partially Available (or degraded) traffic is designed to have *no* impact on the existing traffic. PA is a low priority traffic that runs exclusively on the bandwidth unused by the FA traffic. PA traffic does not affect network engineering rules. The performance of the PA traffic will thus depend on the network traffic load and on the occurrence of failures. As a consequence, the PA service can occasionally become unavailable to its users. Introducing the PA traffic in backbone networks will allow carriers to reduce the amount of unused bandwidth while providing a cheaper service to their customers.

C. *Implementation Issues*

One of the main concerns is that the QoS scheme proposed must be easy to deploy in a backbone network. Our goal is not to derive a new scheduling policy but instead to introduce a simple (possibly already available) scheduler in network routers, and to evaluate the resulting traffic performance.

To satisfy the FA traffic requirements we isolate the two classes of traffic in two separate queues. The FA queue has an absolute priority over the PA queue. Note that giving a strict priority to the FA traffic may lead to starvation of the PA traffic. This is part of the PA service definition. In practice however, the link overprovisioning is such that we will seldom observe a complete starvation of the PA traffic.

This scheme conforms to strict priority scheduling with two drop tail queues. It can be implemented in most of the current routers and does not imply modification in the routing infrastructure. The only additional requirement is that the two classes of traffic needs to be identified by a single bit marking in the IP header. In particular, this QoS scheme does not require signaling, known to account for most of the complexity in QoS architectures.

III. BACKBONES: DESIGN AND ENGINEERING PRINCIPLES

To users, backbone networks appear as a black box delivering high quality service. However, to understand the rationales behind the proposed QoS scheme, and to measure the impact of such a scheme, it is necessary to understand some backbone design and engineering practices.

A. Network Topology

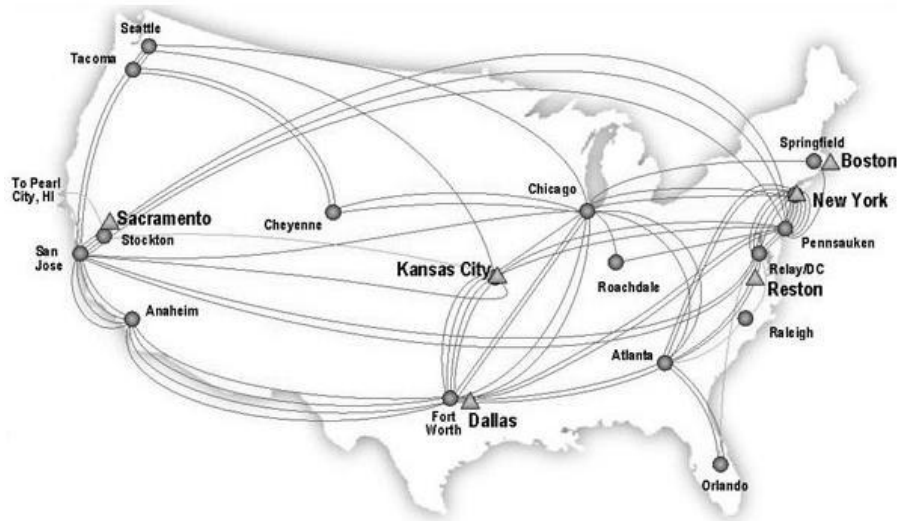


Fig. 1. Simplified view of the Sprint U.S IP domestic backbone (circles represent PoPs)

At the logical level, a backbone network can be represented as a graph whose vertices are *Point-of-Presence (PoPs)* and edges are *inter-PoP links* (Figure 1).

A PoP is a collection of access and core routers collocated in the same site. Clients connect to the network via access routers, which are in turn connected to at least two core routers. The number of core routers per PoP can vary. However, core routers are typically highly meshed with each other.

A pair of neighboring PoPs is connected by multiple, high capacity links (OC-48 and OC-192); each of these parallel links initiates and terminates on different backbone routers (Figure 2(a)). Having numerous parallel inter-PoP links between two given PoPs increases the robustness of the network. It also increases the opportunities for load balancing.

B. Network Engineering

Today, the majority of large backbones use IP over DWDM technology. SONET protection has been removed because of its high cost, although SONET framing is kept for failure detection purpose. Protection and restoration are thus provided at the IP layer only.

Link state protocols such as OSPF or IS-IS are used for intra-domain routing. When a link fails, the traffic is rerouted on the path with the smallest weight sum. In networks with multiple parallel inter-PoP links we differentiate between two types of rerouting events:

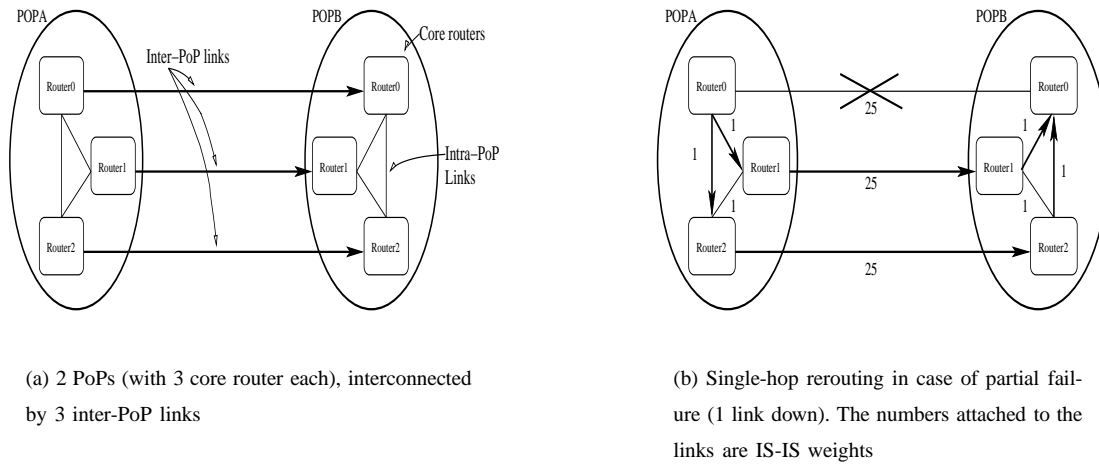


Fig. 2. 2-PoP topology

- *Single-hop rerouting.* The traffic of the failed link is load balanced on the n remaining links (Figure 2(b)). The result is an implicit link protection scheme similar in concept to a $1 : n$ protection scheme at the optical level. Single-hop rerouting, is possible if the parallel inter-PoP links have equal weights (in Figure 2(b) the three inter-PoP links have an IS-IS weight of 25 while the intra-PoP links have an IS-IS weight of 1) and enough excess bandwidth to support the rerouted traffic. The main advantage of single-hop rerouting is to limit the geographical impact of failures and thus the additional propagation delay incurred by the rerouted traffic. Note however that single-hop rerouting is not significantly faster than multi-hop rerouting [1].
- *Multi-hop rerouting.* The traffic is rerouted through an additional PoP. Multi-hop rerouting only happens when all links between two PoPs fail at the same time (eg: if all the links between San Jose and Anaheim fail, the traffic is rerouted through RelayDC, Atlanta and Fort Worth).

We include both types of rerouting events in our simulations.

C. Failures in Backbone Networks

To the best of our knowledge a complete characterization of failures in backbone networks is not available. However some preliminary results, based on monitoring data from the Sprint network, were presented in [1]. We summarize their findings below.

- *Failure duration and causes.*
 - 20% of the inter-PoP link failures last longer than 10 minutes. Possible causes are fiber cuts, equipment failures and/or link upgrades.
 - 30% of the failures last between 1 and 10 minutes. Likely causes include router reboots and software problems.
 - 50% of the failures last less than 1 minute. These failures could be the result of oscillatory effects when a router mistakenly considers the adjacency to be down or could be due to optical equipment.

- The *mean time between successive failure events* is of the order of 30 minutes.
- The *failure distribution across the links* is far from uniform. Some links hardly fail while three links account for 25% of the failures.
- *Failures can be strongly correlated*. Depending on the mapping of the logical topology on the optical topology, a single fiber cut may bring down several logical links. Failures of logical links mapped on disjoint fiber path are, on the other hand, close to being independent.

The knowledge of failure characteristics is an important step toward a model of failure in backbone networks. It is also mandatory to provide realistic evaluations of QoS models.

Link failures occur everyday in backbone network. However their impact on traffic performance can greatly vary. In the remaining of this work we make a distinction between *partial* and *complete* failures. A failure is complete when all the links between two PoPs fail simultaneously. A failure is considered as partial if only a subset of the parallel inter-PoP links fail at the same time. We expect complete failures to induce more severe traffic perturbation, since they imply rerouting of the traffic through one or more additional PoPs. Complete failures are also more costly in terms of resource usage. Note that the Sprint backbone is designed in such a way that the probability of a complete failure is very low [3]. We use these observations to model the failures we consider in our evaluation of the degraded traffic performance.

IV. SERVICE LEVEL AGREEMENTS

A Service Level Agreement (SLA) specifies a contractual service target between a provider and its customer and spells out penalties for non compliance. The tradeoff between the SLA and the service pricing is often a significant factor in the success of an offered service.

A. SLAs in commercial networks

The SLAs offered by Tier-1 provider typically include packet loss, packet delay and port availability. The first two metrics are computed network wide and usually averaged over a one month period. The loss metric reports the average percentage of packet lost in a transmission while the delay metric (or latency) reports the round-trip transmission time averaged over all PoP pairs in the ISP backbone network. Contrary to the other SLA metrics, port availability does not capture the performance of the traffic inside the backbone. Instead, port availability measures the fraction of time a customer's physical connectivity to the ISP's network is up. Note that the notion of port availability may differ between providers.

Table I reports SLA values inside the continental USA for some Tier-1 providers. For comparison purposes, Table II presents the actual measured traffic performance for the Sprint U.S domestic backbone. Port availability cannot be included as it is measured per customer. We observe no SLA violation during the observation period (the second half of year 2002).

B. SLA: enhanced definition

For the purpose of this work we slightly modify the SLA metric definition. Our aim is to better capture the local behavior of each class of traffic and to provide a network wide counterpart to port availability. To make

	AT&T	C&W	Genuity	Sprint	UUNET
Latency	60ms	50ms	55ms	55ms	55ms
Loss	0.7%	0.5%	0.5%	0.3%	0.5%
Port Availability	95%	99.97	99.97%	99.9%	100.0%

TABLE I

INTRA-US SLAS¹: PACKET LOSS AND LATENCY METRIC (DECEMBER 2002)

	July	Aug	Sept	Oct	Nov	Dec
Latency	45.68ms	46.36ms	46.76ms	46.76ms	47.08ms	47.68ms
Loss	0.01%	0.06%	0.06%	0.01%	0.00%	0.00%

TABLE II

SPRINT MEASURED PERFORMANCE FOR THE LAST SIX MONTHS OF YEAR 2002

the number of simulations manageable, we reduce the SLA computation period to 10 days rather than the conventional one month. However, we use a measurement granularity of one minute which is much lower than the one used in commercial networks. As a result the number of samples averaged to compute our SLA is greater than for a commercial SLA despite the shorter SLA computation period. We now provide a high-level definition of our three SLA metrics:

- *Packet loss*: the loss rate averaged over all inter-PoP links.
- *Packet delay*: the packet delays averaged over all inter-PoP links.
- *Service availability*: the fraction of time the service is available. The service is available if the following conditions are verified:
 - All possible source-destination pairs are connected by at least one route.
 - No traffic transmission experiences persistent (10 consecutive minutes or more) high (above 50%) loss rates.

At this point of the paper it is not possible to provide a more detailed description of the SLA metrics, since their computation is closely related to the simulation environment. The exact methodology used to compute the value of each SLA metric will be explained in Section VI.B.

We believe that these three metrics capture accurately the performance of a service while being easier to compute than their commercial equivalent. In particular, we consider service availability to be a good measure of the service quality as perceived by the user and thus a valuable addition to the SLA metrics.

¹<http://ipnetwork.bgtmo.ip.att.net/averages.html>, <http://sla.cw.net/sla/Help.jsp>, <http://netperformance.genuity.com/ourdata.htm>, <http://www.sprintbiz.com/business/network/slas.html>, <http://www.worldcom.com/global/about/network/latency/>

V. CAPTURING CARRIER BACKBONES

The target of this work is to derive quantitative SLAs for the degraded service. To do so we perform *ns-2* simulations on the Sprint U.S. backbone topology. Performing realistic simulations on a backbone network is a challenge. The generation of traffic matrices and failure patterns, are two research topics by themselves. Our approach was thus to use known models or available monitoring data. In addition, we had to make several assumptions to make the simulations computationally tractable. We provide hereafter a pragmatic and experimental justification of these assumptions and show that they do not impact our observations.

A. Topology

To analyze the degraded service SLA, we use the exact topology of the IP Sprint domestic backbone (not shown here for confidentiality issues). For each of the 91 inter-PoP links, we specify the propagation delay, the bandwidth and the IS-IS weight. The intra-PoP topology is fully meshed in all the PoPs.

The typical bandwidth of inter-PoP links in backbone networks is 2.5 or 10 Gb/s. To reduce the simulation complexity we set the bandwidth of inter-PoP links to 10Mb/s. Such a huge reduction of the link capacities might appear to be a rather severe simplification. However, we expect the two classes of traffic to react in similar ways. Section V.D shows that the relative performance of FA and PA traffic are indeed maintained.

B. Traffic

It is not possible today to obtain an exact PoP-to-PoP traffic matrix for an ISP backbone via direct measurement, and techniques for inferring traffic matrices are still under development. Some recent studies [4] have shown the gravity models can capture reasonably well properties of PoP-to-PoP flow volumes. The basic idea behind the gravity models is that the flow between a pair of PoPs is proportional to the product of two factors, one which is a metric of the ingress node and one a metric of the egress node. These metrics should capture key features of the total volume flowing from a PoP into (or out of) the backbone of the ISP. We generated an FA traffic matrix according to these principals and checked that the resulting link loads matched actual Sprint backbone link loads closely on most links. Using the FA traffic matrix we then create approximately 900 router-to-router data flows.

We made an attempt to run the simulations on the Sprint network with TCP traffic but the requirements in terms of memory and time exceeded the capacity of our simulation platform. Therefore, the simulations presented in this paper use UDP constant rate traffic with exponentially distributed On and Off periods. The results for the 2-PoP topology (Section V.D) show that the UDP performance provide an appropriate lower bound for the performance of TCP traffic. In addition, UDP offers a simple and convenient way to look at service availability as it does not adapt its sending rate in case of network congestion.

C. Failure scenario

As we explained earlier, a probabilistic model of failure in backbone networks is not yet available. Therefore, we decided to replay failure sequences as they appear in the Sprint backbone in an attempt to reproduce the failure characteristics observed in Section III.C. Figure 3 shows the distribution of link failures in the Sprint

network between December 2001 and March 2002. We use a 10-day period from February 1st to February

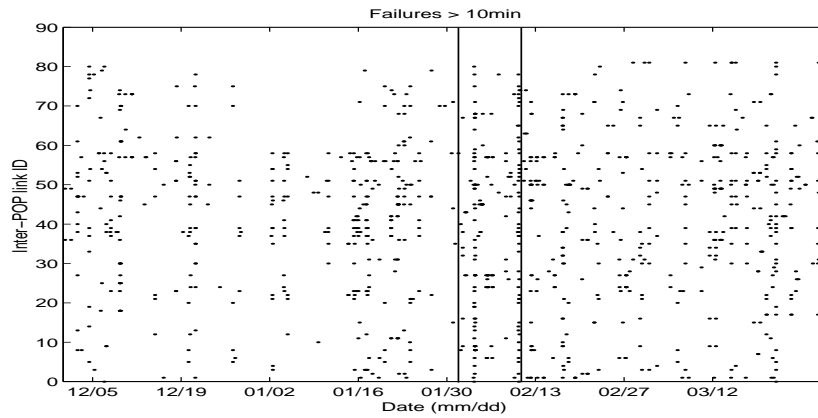


Fig. 3. Failure distribution in the Sprint network. Each dot corresponds to a link failure lasting at least 10 minutes. The two vertical lines indicate the 10-day period selected to run our simulations.

11th to run our simulations. This time interval was chosen because it is representative of an heavily perturbed period.

We isolate each failure event by grouping simultaneous, equal length, failures together, this leads to 15 multi-link failure events. Figure 4 shows the length of the failure and the number of links involved in each of the failure event. Up to 12 links can fail simultaneously and the longest failure event lasted for 8 hours. Half of the failure events include at least one PoP pair that experiences a complete failure (and as a consequence multi-hop rerouting of the traffic). Based on the failure events we construct our failure scenario. We define a

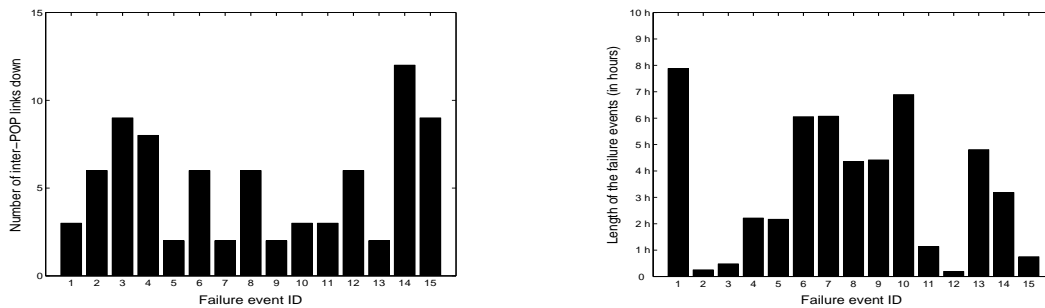


Fig. 4. Individual failure event included in the SLA computation

failure scenario as a sequence of failure events separated by time intervals where the network is not in a failure state. The failure scenario used to run our simulation corresponds to a 10-day snapshot of the network, with a total cumulative failure event time of approximately 2 days.

We made two minor assumptions relative to the failure patterns. First we do not consider intra-PoP failures. The main reason is that intra-PoP link failures have a much smaller impact on traffic since core routers are fully meshed. Second, we only replicate failure events that last more than 10 minutes. Even if those failure events represent only a minority of the total number of failures, they are the ones that significantly affect the

traffic performances, and as a consequence the SLA of the degraded class of traffic.

D. Proof of concept and validation of assumptions

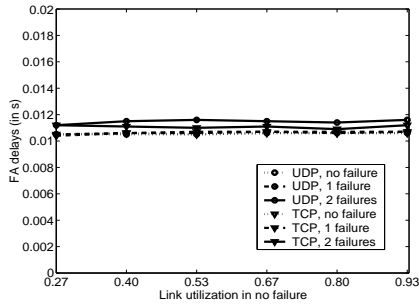
We use the 2-PoP topology shown in Figure 2 to validate our assumptions and provide an initial intuition of the PA performance. The three inter-PoP links have a 10 ms propagation delay and a bandwidth of either 10Mb/s (the link bandwidth chosen for our simulations on the Sprint network) or 2Gb/s (\simeq OC-48). All the inter-PoP links have the same IS-IS weight. When a link fails, the traffic is thus load balanced on the remaining links. The queue size is set equal to the bandwidth-delay product of the link. Each router in PoPA (see Figure 2) generates the same total amount of traffic. The FA traffic load is fixed and occupies approximately 27% of the link bandwidth in the no-failure case. The amount of FA traffic generated was chosen to yield a 80% FA link utilization in the most severe failure event (i.e., when two inter-PoP links are down). The PA traffic is added progressively until we reach a link utilization close to 100%. The packet size is set to 500 bytes for both UDP and TCP traffic.

1) General observations: The goal of the 2-PoP topology is not to provide quantitative traffic performance, as it could be very different on a large network, but to verify that the two classes of traffic behave as expected. First, we notice that FA traffic is not affected by the addition of the PA traffic (as shown by the flat curves of Figure 5(a)). As a direct consequence, the FA service will have the same SLA as the current backbone traffic. Second, we observe that after a short and sharp increase the PA delay stabilizes (see one-failure curves in Figure 5(b)). At this point the PA traffic occupies all the bandwidth unused by the FA traffic, i.e., the link utilization is 100%. Its performance are then dictated only by the amount of spare bandwidth unused by FA traffic (60% in the one-failure case and 20% in the two-failure case). In general, the PA performance will thus depend on the total traffic load (FA load and PA load), and on the quantity of failures. Indeed, Figure 5(d) clearly shows that, as long as there is no failure and the link utilization remains under 80%, the two classes of traffic have very similar performance.

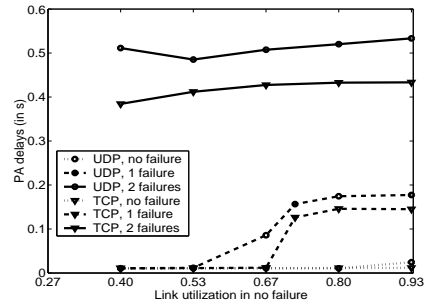
2) Experimental justification of our assumptions:

UDP performance provides a lower bound to TCP performance. Figures 5(a), 5(b) and 5(c) compare packet delay and packet loss performance for UDP and TCP traffic. The link bandwidth is set to 10Mb/s. We did not report the packet loss metric for the FA traffic since its value is null by design. The curves of Figure 5(a) and 5(b) show that delays for UDP traffic are only slightly higher than delays for TCP traffic and follow the same trend. UDP delays thus provide a good approximation of TCP delay performance. However, the PA packet losses for UDP traffic are substantially larger than the corresponding TCP loss rates. Contrary to TCP, UDP does not adapt its sending rate to the amount of congestion in the network. A high level of UDP loss thus reflects a poor availability of the service. Although our decision to use UDP traffic for our simulations on the Sprint network was mainly motivated by complexity considerations for large-scale network simulations, UDP loss rate happens to be a convenient metric to measure network availability.

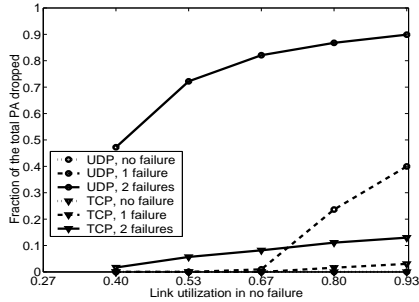
Simulated link bandwidth has a limited impact. Figure 5(d) shows the delay performance of the PA and FA service on links of 10 Mb/s and 2Gb/s capacity. The traffic type is UDP and the three inter-PoP links are up. We observe that the traffic performance on the different bandwidth links are very close to each other and that



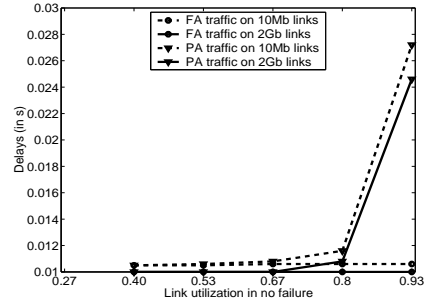
(a) UDP and TCP packet delays for the FA traffic vs. the total link utilization.



(b) UDP and TCP packet delays for the PA traffic vs. the total link utilization.



(c) UDP and TCP packet loss rate for PA traffic vs. the total link utilization.



(d) UDP packet delays for FA and PA traffic in the no-failure case. The link bandwidth is either 10Mb/s or 2Gb/s.

Fig. 5. Experimental results on 2-PoP topology

the relative performance of the FA and PA service are indeed maintained. Note however, that queuing delays tend to be slightly lower on high bandwidth links. This is due essentially to transmission delay and will be neglected in the remaining of the study.

VI. EXPERIMENTAL DESIGN AND RESULTS

Experimental design is an area of statistics used to maximize the information gain obtained from a finite set of simulations and to provide an accurate analysis of the simulation results. In statistical design of experiments, the outcome of an experiment is called the *response variable*, the parameters that affect the response variable are called *factors* and the value that the factor can take *levels*. In this section we first present the different factors studied. We briefly explain how the response variable is computed from the simulation output. Finally, we analyze the effect of factors and combination of factors on the response variable and discuss SLA for the degraded class of traffic.

A. Factors evaluated

We identify three important factors which could potentially affect the performance of the degraded traffic.

1) *The PA traffic generation strategy*: In order to consider a variety of demand scenarios for the degraded traffic class we have chosen two different strategies to generate the PA traffic:

- *PA-fraction*. The PA traffic matrix is a fraction of the FA traffic matrix. The justification for this approach is simply that the demand for the new traffic class is likely to be proportional to the existing demand (this is also in agreement with the gravity model [4]). The PA-fraction generation strategy may, however, yield poor PA traffic performance since links with an already high utilization receive the largest PA traffic load. Non-uniformity in the traffic distribution across the links will thus be amplified rather than attenuated.
- *PA-optimal*. This strategy is meant to reproduce an 'optimal' placement of the PA traffic. We compute the PA traffic matrix to yield an equal link utilization on all inter-PoP links. If the utilization of FA traffic alone on a link is already larger than the target average utilization, we do not add any PA traffic on the link. Although such a distribution of the PA traffic is unlikely to arise in reality, it represents the best case and thus allows us to assess the best possible SLA this traffic class could receive.

2) *The PA load*: The main goal of this study is to determine how much PA traffic can be added to the network before too much degradation in the SLA occurs. The FA load in the network is constant and is determined by the FA traffic matrix. The average FA link utilization (i.e the total FA load divided by the sum of the link capacities) is approximately 16%. The PA traffic load is variable and is added as a multiple of the FA traffic load in the network. In the context of this work we run simulation for PA load equals to up to four times the FA load. This corresponds to an average link utilization between 16% and 80%.

3) *Network availability*: To take into account the variation of failure rates in backbone network, we derive the SLAs for different levels of *network availability*, α . We define network availability as the percentage of the total bandwidth available to carry the backbone traffic. Network availability is averaged over time. $\alpha = 0$ corresponds to a network where all the links are always down and $\alpha = 1$ corresponds to a network which is never subject to any failure. For the 10-day failure scenario described in Section V.C we can compute the network availability as follows:

$$\alpha = \frac{1}{\sum_{i=0}^{15} t_i} (t_0 + t_1 \alpha_1 + t_2 \alpha_2 + \dots + t_{15} \alpha_{15}) = 0.99$$

where t_i the length of the i^{th} failure event (as reported by Figure 4), t_0 is the time spent in the no-failure state and α_i the percentage of links up during the i^{th} failure event. The actual level of network availability in the Sprint network is 0.99. We vary the network availability by scaling the time spent in failure events (t_i , $0 < i \leq 15$), versus the time spent in the no failure-state (t_0) and were thus able to run simulations for $\alpha \in [0.95, 1.00]$ ($\alpha=0.95$ corresponds to $t_0 = 0$, i.e., the Sprint network is permanently in one of the failure state).

B. Computation of the response variable

We are interested in multiple response variables, namely the packet loss (p_l), the packet delay (p_d) and the availability (s_{avail}) of a service. As we have seen previously we can group those three metrics together under the notion of SLA. To compute the response variables we simulate each of the 15 multi-link failure events and record the average performance of the two classes of traffic. The simulation time is 1min per failure event. It was not necessary to run the simulations longer since the performance become stable after a couple of seconds. The simulation output for the i^{th} failure event can be described as the tuple $(p_{l_i}, p_{d_i}, s_{avail_i})$. We calculate a response variable by performing a weighted average of the metric of interest observed during each failure event:

$$\begin{cases} p_l &= \frac{1}{\sum_{i=0}^{15} t_i} (t_0 p_{l_0} + t_1 p_{l_1} + t_2 p_{l_2} + \dots + t_{15} p_{l_{15}}) \\ p_d &= \frac{1}{\sum_{i=0}^{15} t_i} (t_0 p_{d_0} + t_1 p_{d_1} + t_2 p_{d_2} + \dots + t_{15} p_{d_{15}}) \\ s_{avail} &= \frac{1}{\sum_{i=0}^{15} t_i} (t_0 s_{avail_0} + t_1 s_{avail_1} + t_2 s_{avail_2} + \dots + t_{15} s_{avail_{15}}) \end{cases}$$

For each possible combination of factor levels we thus run a set of 16 simulations (15 multi-link failure events + the no failure case) and compute the resulting SLA as the tuple (p_l, p_d, s_{avail}) rounded up with a suitable granularity.

C. Impact of the different factors on the response variable

We now apply the experimental design methodology to study the impact of each factor, and each combination of factors, on the response variables, and to identify factors with the highest influence on the traffic SLAs. The importance of each factor, is measured by the proportion of the total variation in the response that is explained by the factor. Several types of experimental design are available, we use the very popular 2^k factorial design [5], [6]. A 2^k factorial design is used to determine the effect of k factors each of which have two alternatives or levels. In our experiment $k = 3$ and the two level selected are the maximum and the minimum values for each of the factor (except in the case of the PA generation strategy where there is only two possible values). To make our analysis more precise we perform two distinct 2^3 factorial designs. In the first factorial design, we assume that the amount of PA traffic added to the network is lower than the current FA traffic load (i.e PA load level ≤ 1). In the second one, we relax this assumption and move to PA load > 1 . We will see later that for PA load > 1 we are no longer able to guarantee similar performance to FA and PA traffic even in the no failure case.

1) *Network availability based service differentiation:* Table III summarizes the impact of each factor and combination of factors on the different SLA metrics (values under 1% are omitted). For example, the first cell of the table tells us that the PA load is responsible for 4.3% of the variation in the PA loss performance. Table III only considers PA load ≤ 1 . Results in terms of FA service availability are undefined since there is no variation in the response (i.e. the FA service availability is uniformly equal to 100% across all factor levels). The result of the 2^k factorial design shows that at low PA traffic load the network availability accounts for the major variation in the traffic performance. Since the two classes of service have initially the same performance we can thus conclude that, for PA load ≤ 1 , the difference of performance between the FA and PA service will

	PA loss	FA loss	PA delay	FA delay	PA service availability	FA service availability
PA load (Ld)	4.3%		17.4%			NaN
Network Availability (NAv)	71.1%	98.4%	17.6%	99.5%	100.0%	NaN
PA generation Strategy (St)	5.9%		12.2%			NaN
Ld,NAv	4.3%		14.7%			NaN
Ld,St	4.2%		13.0%			NaN
NAv,St	5.9%		13.5%			NaN
Ld,NAv,St	4.2%		11.7%			NaN
Ld+NAv+St	81.3%	99.2%	47.1831%	99.8%	100.0%	NaN

TABLE III

IMPACT OF EACH FACTOR FOR PA LOAD UNDER 1 (VALUES UNDER 1% ARE OMITTED, NaN STANDS FOR NOT A NUMBER).

be essentially a function of the frequency of failures in the network. Notice though that PA delay are sensitive to all factors and combination of factors.

2) *Link load as a main factor in service differentiation:* Table IV reports the impact of each factor and combination of factors on the different SLA metrics for PA load > 1. At high traffic load, the PA load becomes

	PA loss	FA loss	PA delay	FA delay	PA service availability	FA service availability
PA load (Ld)	52.6%		63.7%			NaN
Network Availability (NAv)	22.6%	98.9%	18.9%	99.7%	68.2%	NaN
PA generation Strategy (St)	18.3%		16.5%		15.8%	NaN
Ld,NAv	1.8%					NaN
Ld,St	4.0%					NaN
NAv,St					15.8%	NaN
Ld,NAv,St						NaN
Ld+NAv+St	93.4%	99.4%	99.0%	99.9%	84.1%	NaN

TABLE IV

IMPACT OF EACH FACTOR FOR PA LOAD ABOVE 1.

the dominant factor affecting the delay and loss of the PA traffic. However, the FA traffic remains only influenced by the level of network availability, the PA generation strategy and the amount of PA traffic in the network have not impact on its performance. This confirms the immunity of the legacy traffic SLA to the adjunction of the degraded traffic in the Sprint network. Table IV also shows that, contrary to other PA performance metrics, the PA service availability is mainly affected by the network availability. We conclude that even for

large PA loads the service availability could remain high as long as there are few failures in the network (i.e. the network availability is high). Yet, in the event of failures, PA customers should expect a severe reduction in the availability of their service. Finally, we observe that contrary to low PA loads the interactions between the different factors are small and can thus be neglected.

3) *Discussion - Implication of the results:* Figure 6 illustrates the results of the 2^k factorial designs in terms of network availability. For PA delay and PA losses (Figure 6(a) and 6(b)) we observe essentially a vertical

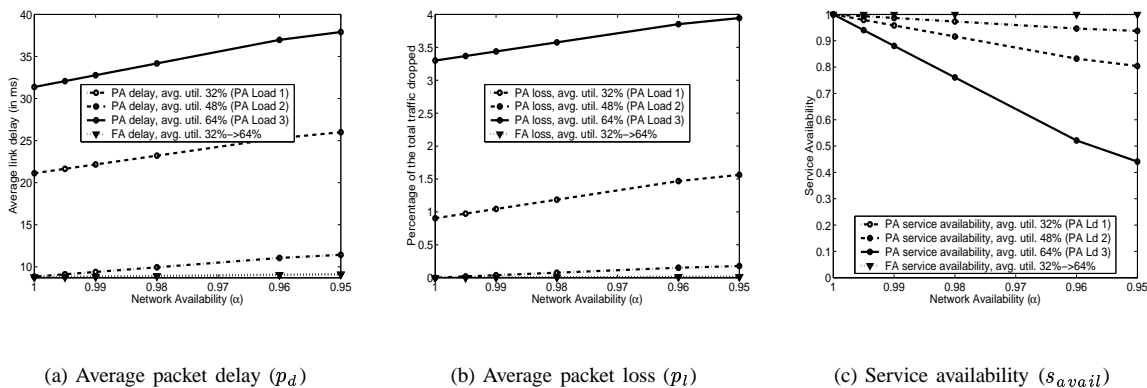


Fig. 6. Impact of the network availability on the different SLA parameters. Three sets of curves, for low, medium and high PA load. The PA-fraction generation strategy was used and FA traffic performance are shown for reference.

translation of the performance curves as additional PA load is added to the network. On the other hand, for the PA service availability (Figure 6(c)) we mainly notice an increased inclination of the curves when the average link utilization becomes higher. The “translation” phenomenon reflects the impact of the PA load on the PA performance while the “inclination” or the slope of the curves shows the influence of the network availability factor. As expected, the impact of the network availability on the delay and loss rate is significant, but still acceptable by most non interactive applications. However, service availability seems to suffer severely from a reduced network availability. A PA service availability of 100% can drop to less than 50% when we move from a network availability of 100% to 95%. This result is surprising and limits the applicability of the PA service. Lets illustrate these results targeting a service available 75% of the time (that would correspond to the service being unavailable during peak hours and in case of major failure events). An Internet Service Provider (ISP) could simply double its resource utilization (PA load equal to FA load) and systematically match above 90% service availability with very little delay and loss degradation. If the resource utilization is tripled (i.e., PA load of two) the service availability remains above the 75% threshold at the expense of higher packet losses and delays. On the other hand, a PA traffic load of three would only allow a service availability of 75% for a network availability above 98% (which is below the current network availability in the Sprint backbone).

Therefore, our degraded class of service seems to allow a visible service differentiation at the cost of marginal additional complexity in backbone routers. Furthermore, service availability appears to provide a natural distinction between the FA and PA services quality and is likely to limit the introduction of the degraded traffic in a failure prone environment.

D. Analysis of the SLA for the degraded class of service

To conclude our analysis, we exhibit upper and lower “bounds” on the degraded traffic performance, and present SLAs for the current level of network availability in the Sprint network. Please note that we do not intend to use the term “bound” in its mathematical sense. The bounds provided in this section were found by simulations and represent a favorable respectively an unfavorable setting for the introduction of degraded traffic in the network.

Figures 7(a), 7(b) and 7(c) show performance intervals for packet delay, packet loss and service availability, as a function of the PA load added to the network. To compute the lower bound we measured the traffic performance under the PA-fraction generation strategy and a network availability of 95%. The upper bound corresponds to the traffic performance observed when we set the traffic generation strategy to PA-optimal and the network availability to the level of 100%.

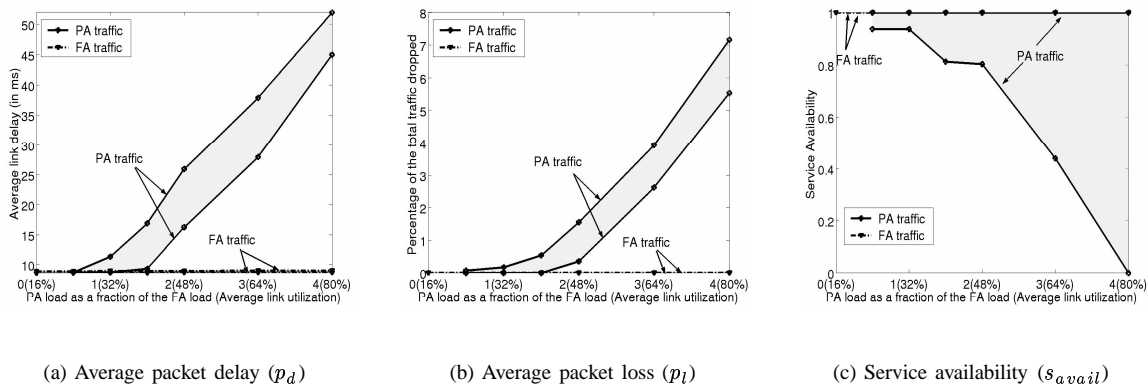


Fig. 7. Performance interval for the FA and PA class of traffic. Two extreme cases are shown: 1) The PA-fraction generation strategy combined with a network availability of 95%; 2) The PA-optimal generation strategy with a network availability of 100%

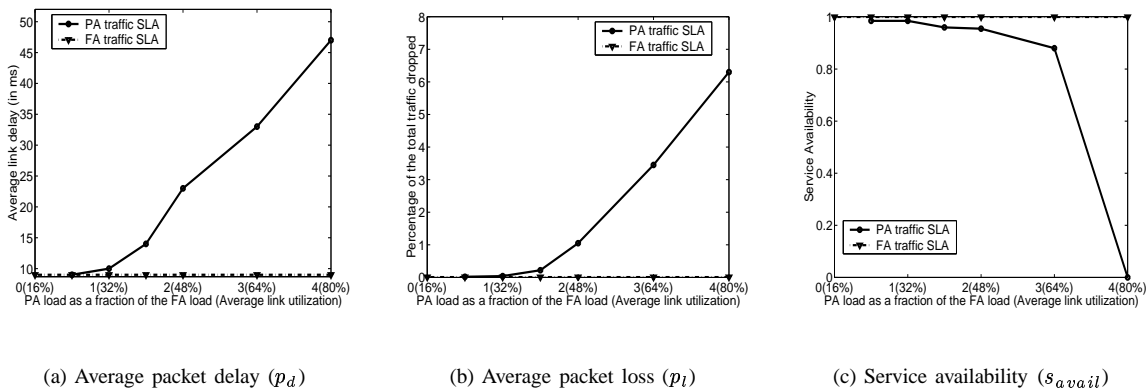


Fig. 8. FA and PA SLAs for the current level of network availability in the Sprint network

The tight bounds on the FA traffic performance are a good indication that the Sprint network is carefully provisioned to support the current backbone traffic. They demonstrate that even at high levels of failures the performance of the FA traffic remains stable. They also show that the addition of the PA traffic has no impact on the FA traffic as demonstrated earlier by our factorial design analysis.

In the case of PA traffic delays and losses, we observe a small performance interval at low PA load which increases until the PA load in the network is twice the FA load. It is interesting to notice though that for high PA loads the PA performance interval stops expanding. The PA service availability exhibits different properties. Its performance interval keeps increasing as more traffic is added to the network while in the best case the service availability stays equal to 100%. It reflects the fact that even for large traffic load the service availability can remain high as long as the network availability is high too and the PA traffic distribution is homogeneous.

Figure 8 depicts SLAs for the current level of network availability ($\alpha = 0.99$) in the Sprint backbone. The reported SLAs are based on measured traffic performance rounded up to the nearest ms for delay, the nearest 0.01% for loss and the nearest 0.5% for service availability. For each PA load, we present the worst performance observed among the two PA traffic generation strategies. Based on the result presented in Figure 8 we can extract SLAs for different PA loads (examples are given in TableV). We observe that if the amount of added PA traffic

PA/FA=1.0	Drop	Delay	Service Availability	PA/FA=2.0	Drop	Delay	Service Availability
PA SLA	0.04%	10ms	98.5%	PA SLA	1.05%	23ms	95.5%
FA SLA	0.01%	9ms	100%	FA SLA	0.01%	9ms	100%

TABLE V

SLAs FOR A PA LOAD EQUAL TO THE FA LOAD (LEFT), AND TO TWICE THE FA LOAD (RIGHT)

is 50% of the FA traffic, then the PA will experience the same SLA as FA traffic. Moreover, we can add an amount of PA traffic equal to the amount of FA traffic, and still leaves PA's SLA only marginally degraded. If more PA is added beyond this amount, then the PA SLA begins to degrade monotonically as the load increases. In particular, the sharp decrease in service availability between PA loads of 3 and 4 indicates that we could at most quadruple the traffic load carried by the Sprint network. Nevertheless, these results are encouraging for carriers. Depending on the quality of the degraded service they want to sell, they can at least double the total traffic load in their networks. The corresponding increase in revenue should then be determined through marketing studies and is outside the scope of this work.

VII. RELATED WORK

QoS is a very prolific area of research. A multitude of QoS models exist, but to date none was successfully deployed in the Internet. There is a common belief that QoS can help manage resources at no cost. This is wrong. QoS has a cost in terms of complexity, management and network robustness. This is especially true of stateful schemes such as Intserv [7]. Schemes which are almost stateless, such as Diffserv [8], seem a priori

better suited for backbone deployment. Nevertheless, most of those schemes greatly increases the complexity of edge routers, while offering no explicit guarantees.

Our service differentiation technique does not require complex admission control schemes or traffic profiles. A simple one bit marking at the ingress router and a strict priority scheduler with two drop tail queues are all we need to provide service differentiation in backbone networks.

Backbone networks have their own characteristics. Due to the large amount of overprovisioning all users experience a very high QoS. It is thus difficult to create a Class of Service (CoS) with improved performance. Instead we decided to provide a degraded class of service. The idea of a degraded service, though not new (Internet drafts [9], [10]), is especially appropriate for backbone networks. The first Internet draft [9], describes a Lower than Best Effort (LBE) Per-Hop Behavior (PHB). The primary goal is to separate the LBE traffic from best-effort traffic in congestion situations. LBE packets are discarded more aggressively than best-effort packets but nevertheless LBE traffic is guaranteed a minimal share of the bandwidth. Our proposal is, however, closer to the ideas presented in the second Internet draft. We believe that the creation of a new PHB is not required since existing PHB can be configured to forward packet of the degraded traffic only when the output link would otherwise be idle. Both of these Internet drafts are now expired.

QBone Scavenger Service (QBSS) and Alternative Best Effort (ABE) are two other examples of non-elevated services and are currently under investigation by QBone [11]. QBSS is similar to the PA service we investigate in this paper but it uses different queuing disciplines. Introducing the QBSS would not allow us to maintain the SLAs of the current backbone traffic.

Packet losses in backbone network are almost always the result of failures [1]. The notion of service differentiation as a function of failures was first proposed in [12]. However, in [12], the two classes of service were protected at two different layers, one at the WDM layer and one at the IP layer. In this paper we consider that both classes are protected at the IP layer. This reflects an important reality because many carriers find protection at the optical layer to be too costly. Moreover, we are the first to derive quantitative SLAs for a degraded service class in a commercial backbone. We are also the first to introduce real failure patterns in our analysis.

VIII. CONCLUSION

In this paper we introduce the idea of a new class of service intended to be a degraded service relative to the existing service offered by today's Internet. Our solution has three major advantages; first, it reduces the amount of unused bandwidth in the IP backbone thus improving resource utilization; second, it offers ISPs a way to increase their revenue; and third, it offers users a cheaper service which quality is sufficient for most non real-time applications. We show that, by introducing a very simple scheduling mechanism in backbone routers, it is possible to add a degraded class of traffic and still maintain the impressive performance of the legacy traffic.

The objective of this work was to evaluate the Service level Agreements which could be offered to the customers of the degraded traffic. To do so we carry out large-scale simulations that mimic the topology, traffic matrices and failure patterns of a commercial IP backbone (i.e., the Sprint U.S domestic backbone). Although, we had to make several simplifications to make our simulations tractable, we claim that our evaluation is realistic enough to prove the feasibility of our QoS scheme.

In particular, this work contains two major innovation in the area of QoS management:

- This is the first time network availability is used in the evaluation of a QoS scheme in a wired environment. Network availability effectively captures the impact of link failures on the network infrastructure.
- We introduce a new SLA metric, the service availability, in order to reflect the ability of a customer to successfully use its network connectivity at any time, independently of the destination end-point.

We demonstrate that when the amount of degraded traffic added to the network is lower than the load of the legacy traffic, the level of network availability is the primary factor affecting the degraded service quality. However, when higher loads of degraded traffic are added to the network, the average link utilization becomes the dominant factor influencing the delay and loss performance of the degraded traffic. Yet we noticed, that across all traffic load the service availability metric remains mainly affected by the level of network availability. The distribution of the degraded traffic across the links was shown to have a limited impact on its overall performance.

To conclude, our results demonstrate that it is possible to double the utilization of the network resources and still guarantee fairly high performance to the customers of the degraded service. In the context of a non-real time traffic such as web or peer-to-peer, carriers can even expect to triple or quadruple the existing traffic load depending on the level of network availability in their backbones.

The current evaluation environment has some limitation, mostly due to the simulation setting. A statistical model of resource failures in IP backbone is mandatory to provide a more accurate estimate of the degraded traffic performance and to refine the notion of network availability. In addition, the definition of service availability could be adjusted to target the needs of specific applications which are likely to generate a large share of the degraded traffic load. These subjects will form the basis of future work.

REFERENCES

- [1] G. Iannaccone, C.-N. Chuah, R. Mortier, S. Bhattacharyya, and C. Diot, "Analysis of link failures over an IP backbone," in *ACM SIGCOMM Internet Measurement Workshop*, Marseilles, France, Nov. 2002.
- [2] C. Fraleigh, F. Tobagi, and C. Diot, "Provisioning IP Backbone Networks to Support Latency Sensitive Traffic," in *IEEE Infocom*, San Francisco, Mar. 2003.
- [3] F. Giroire, A. Nucci, N. Taft, and C. Diot, "Increasing the Robustness of IP Backbones in the Absence of Optical Level Protection," in *IEEE Infocom*, San Francisco, Mar. 2003.
- [4] A. Medina, N. Taft, K. Salamatian, S. Bhattacharyya, and C. Diot, "Traffic Matrix Estimation: Existing Techniques and New Directions," in *ACM SIGCOMM*, Pittsburgh, USA, Aug. 2002.
- [5] R. Jain, *The art of computer systems performance analysis: techniques for experimental design, measurement, simulation, and modeling*. New York: Jon Wiley, 1991, no. 0-471-50336-3.
- [6] P. John, *Statistical Design and Analysis of Experiments*. 3600 University City Science Center, Philadelphia, PA 19104-2688: Society for Industrial and applied Mathematics, 0-89871-427-3, 1998, no. 0-89871-427-3.
- [7] R. Braden, D. Clark, and S. Schenker, *Integrated Services in the Internet Architecture: An Overview*, IETF RFC 1633, June 1994.
- [8] Y. e. a. Bernet, *A Framework for Differentiated Services*, IETF Internet Draft, February 1999.
- [9] R. Bless and K. Wehrle, *A lower Than Best-Effort Per-Hop Behavior*, Std., Expires 2000.
- [10] B. Carpenter and K. Nichols, "A bulk handling per-domain behavior for differentiated service," in *IETF Internet Draft*, Expires 2001.
- [11] "<http://qbone.internet2.edu/>."
- [12] A. Nucci, N. Taft, P. Thiran, H. Zang, and C. Diot, "Increasing the Link Utilization in IP-over-WDM networks," in *SPIE OPTICOMM*, Boston, Aug. 2002.