

A General Framework for Service Availability for Bandwidth-Efficient Connection-Oriented Networks

Ori Gerstel, *Fellow, IEEE*, and G. Sasaki

Abstract—Availability in connection-oriented service in networks has traditionally been “all-or-nothing,” i.e., when a failure occurs, a connection either is unprotected or fully protected. The differences in availability and cost between these two extremes can be quite high. A general framework for service availability will be presented that fills the gap. It is shown how network resources and cost are related to service parameters of the framework for networks that are bandwidth-efficient. In addition, a simple revenue model is presented and characterized, revealing when nontraditional service agreements may be attractive.

Index Terms—Protection switching, service availability, service level agreements, survivable networks.

I. INTRODUCTION

MOST protection schemes in a network attempt to achieve the availability specified in a customer’s service level agreement¹ (SLA) by “all-or-nothing” switching, i.e., whenever a fault occurs, a connection on the fault is completely protected or not for the duration of the fault. There are great differences in availability and cost between these two extremes. For example, a 1 + 1 protected connection may have a high availability of 99.999%, while an unprotected connection could have a much lower availability of 99.9% or even lower. In addition, the 1 + 1 connection may use more than twice the network resources as an unprotected connection since the disjoint working and protection paths of a 1 + 1 connection are together at least twice a shortest path of an unprotected connection. While shared protection schemes reduce resource usage, they still require significant protection resources—especially for sparse topologies such as rings—and do not provide availability guarantees for connections that are not 99.999% protected. So among the limited selection of classical protection services, there is a high tradeoff between availability and network cost, which ultimately affects customer prices. What is needed is to bridge the gap between

these classical protection services so that a customer will be able to find the right SLA at the right price.

The purpose of this paper is to propose a general framework for the SLA toward this goal while keeping in mind the underlying technologies for implementation. The framework has the following practical advantages. It can be implemented with current and foreseeable technologies at both the optical layer and electronic packet switched layer (e.g., Ethernet, MPLS, IP). Its parameters can be measured to verify SLA compliance. It leads to bandwidth efficiency in the network, which in this paper means that there is minimal or possibly no additional bandwidth beyond the necessary working bandwidth. This is important to keep connection prices low.

The framework allows protection bandwidth to be a fraction of the working bandwidth such as in [1] and [2]. It also allows connections that are not directly on a fault to be interrupted to free surviving bandwidth for protection such as in [3] and [4]. This is a departure from the usual telecommunication practice of not disturbing established connections, but it allows greater flexibility in optimizing surviving bandwidth, leading to a more holistic network protection:

Definition: Network protection is a means to redistribute the limited bandwidth that survived a network failure, among all the services supported by the network with the single goal of ensuring they all meet their SLAs.

In addition, the framework will introduce features that address the following weakness of conventional availability specifications. The availability of a connection is typically specified by a percentage, e.g., 99.9%. For an operating period of say a year, the connection will be unavailable for at most 8 hours and 46 min. Note that the connection can be continually down for 8 hours and 46 min and still meet its SLA. This may be too long for a customer, who may prefer to limit any continuous downtime to a couple of hours, and spread out the downtimes. The SLA framework of this paper addresses this by ensuring availability over short periods.

The paper is organized as follows. Related work is discussed in Section II, and the SLA framework is presented in Section III. Section IV describes how the SLA framework affects network cost, and in particular the required link bandwidths. In the section, and throughout this paper, the system is assumed to be composed of two network nodes connected by a pair of connections that pass through a network as shown in Fig. 1. The connections basically serve as a pair of links between the two nodes, and they will be referred to as “links” 1 and 2. The “links” are assumed to have the same bandwidth and have disjoint physical paths, and that they do not fail together. It will be assumed throughout the paper that the total time that link k ($= 1$ or 2) is down is at most F_k . In addition, the time to repair a link is

Manuscript received March 09, 2008; revised November 06, 2008; February 21, 2009; and August 13, 2009; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor A. Somani. First published April 19, 2010; current version published June 16, 2010.

O. Gerstel is with Carrier Routing Business Unit, Cisco Systems, Natanya 42504, Israel (e-mail: ori@ieee.org).

G. H. Sasaki is with the University of Hawaii, Honolulu, HI 96822 USA (e-mail: galens@hawaii.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TNET.2010.2046746

¹For sake of simplicity, we refer to the availability defined in the SLA as “the SLA.” The SLA contains many other aspects, such as maximum latency, jitter, as well as support guarantees, but they are outside the scope of this paper.

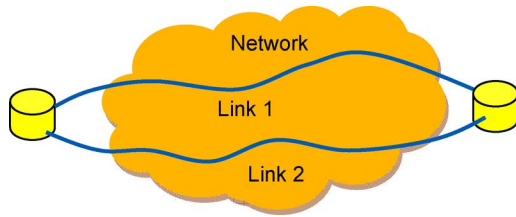


Fig. 1. Point-to-point system through a network.

at most f . The values F_1 , F_2 , and f are assumed to be known *a priori*. Presumably, at least conservative estimates are known by service providers. Throughout the paper, the following notation will be used: $\Sigma F = F_1 + F_2$. Note that ΣF is an upper bound on the total amount of time that there is some failure.

It will also be assumed that the links carry a total of K connections, and each connection operates over the time interval $[0, T]$. The duration T will be referred to as the *lifetime of the network*. In Section IV, a lower bound on link bandwidth requirements is given. It will be shown that the lower bound is nearly achievable for simple but important special cases.

Section V presents scenarios when nontraditional protection schemes may be economically attractive. Simple economic models are used to illustrate the tradeoffs. Implementation is discussed in Section VI. It is worth noting that the network must now keep track of state information per connection. Section VII has final remarks. It includes directions for future research and a discussion of generalizations of the assumption shown in Fig. 1 from two links to multiple links.

II. RELATED WORK

There have been a number of proposals to bridge the gap between full protection and unprotected service. Several papers—such as [5]—propose that connections be given priorities based on their SLA, and survivability of a connection depends on its relative priority among the other connections. Such “best effort” approaches that depend on the SLA of other connections may be unsuitable for applications that require service guarantees.

The Quality of Protection (QoP) framework in [1] is an SLA framework with survivability guarantees that are independent of other connections. Here, the availability of connections is accommodated in one of two ways: 1) connections are given the amount of bandwidth they were promised under a failure condition—even if a fraction of the bandwidth under normal conditions; 2) a probabilistic scheme in which connections get their full bandwidth according to a probability that is based on their priority. Note that the QoP framework the first way is also considered in [2]. Also note that the QoP framework the second way is an “all-or-nothing” protection switching. Another probabilistic approach is presented in [6], which describes a mathematical optimization framework. In [7], routing under a probabilistic framework is considered.

In [3] and [4], connections that are not on a fault can be interrupted or “victimized,” allowing greater flexibility to meet SLAs of all connections. In the case of [3], the connections protect a fraction of their working bandwidth. This approach

can be viewed as a generalization of the “extra traffic” concept in SONET. Our approach also blurs the boundary between working and protection bandwidth, as ultimately the system can use both to protect connections. It should be noted that the selection of which connections to victimize must be done carefully to ensure SLAs are satisfied.

In [4], the accumulated downtime of connections are kept track of and used to determine the connections to victimize. It is shown in [4] that naive greedy approaches can fail. Accumulated downtimes of connections are also used in [8]. In [9], it is discussed that the guaranteed maximum accumulated downtime may need a safety margin to take into account the time to repair.

III. SLA FRAMEWORK

The SLA framework for a connection k will be described. The framework has components referred to as SLA1–SLA4.

SLA1: The connection has two states, *working* and *protection*, corresponding to two bandwidths, the working state bandwidth B_k and the protection state bandwidth b_k , respectively.² The working state is the normal state for the connection, and the protection state occurs only when there is a fault somewhere in the network. So when there are no faults in the network, all connections are in their working state, and if there is a fault, then some connections may be in the protection state, and the rest of the connections are in the working state. The set of connections that are in the protection state may change over time during a fault. The protection state bandwidth b_k is a fraction of B_k and can be zero. The following simplifying assumption will be used in the subsequent sections. For all connections k , the working state bandwidth is $B_k = B$, and the protection state bandwidth is $b_k = b$.

SLA2: The connection has a maximum accumulated time D_k that the connection is in the protection state. Note that the service availability of the connection must be at least $1 - D_k/T$, where T is the lifetime of the network. For example, if the availability is 99.999% and T is a year, then D_k is 5.26 min. D_k will be referred to as the *maximum accumulated protection state time* for connection k .

SLA1 and SLA2 cover classical connection services: unprotected, fully protected, and low-priority preemptible (extra traffic in SONET nomenclature) services. For unprotected connections k , $D_k = F_{n(k)}$, where $n(k)$ is the link that connection k is normally carried. For fully protected connections (i.e., availability is 100%), $D_k = 0$. For low-priority preemptible connections, $D_k = \Sigma F$, since any failure will impact these connections due to their preemption—even if they were not directly impacted. SLA1 and SLA2 also cover protection schemes of [1]–[3] and [4].

The next two SLA components, SLA3 and SLA4, specify the availability of a connection over short time periods.

SLA3: Whenever the connection goes into the working state, it must remain in the state for at least a minimum amount of time μ_k before going to the protection state. This ensures that services that require the working state bandwidth have sufficient time to be completed. For example, video streams for movies

²Such an SLA is feasible if connection interfaces can transmit at two bandwidth rates. Section VI presents more implementation details.

require bandwidth for 60 to 90 min. The parameter μ_k will be referred to as the *minimum working state duration*.

The following assumption will be used in the next section to ensure SLA3 with minimal link bandwidth. Let the time before the first failure and the times between consecutive faults be referred to as *fault-free periods*. Assume that the *minimum fault-free duration* is at least $\max_i \mu_i$, where μ_i is the minimum working state duration for connection i . Note that this assumption should be reasonable if the minimum working state durations are much smaller than the mean time to failure for the links. Without this assumption, a link could go down, come back up, and then immediately go back down again, possibly leading to a violation of SLA3. For example, suppose each link is 30 Gb/s and carries three connections, where the working state bandwidth is $B = 10$ Gb/s and the protection state bandwidth is $b = 5$ Gb/s. Note that the links have minimum bandwidth for the connections in their working state, and if a link goes down, then the surviving link has just enough bandwidth for all connections in their protection state. Now when one of the links goes down, all six connections go into their protection states, each with 5 Gb/s on the surviving link. When the link comes back up, all connections are required to be in their working state. However, when the link immediately goes back down again, there is not enough surviving bandwidth to ensure all connections can be in their working state for their minimum working state durations, violating SLA3. To ensure SLA3, more link bandwidth is needed, but then the links are less bandwidth-efficient.

SLA4: Whenever the connection goes into the protection state, it must transition to the working state after at most δ_k amount of time. The parameter δ_k will be referred to as the *maximum protection state duration*. In addition, over any time period $[x, y]$, the amount of time that the connection is in the working state is at least $\max\{0, \alpha_k \cdot (y - x - \delta_k)\}$, where α_k is a parameter satisfying $0 \leq \alpha_k \leq 1$ and referred to as the *short-term availability rate*. Note that this has the same form as the quality-of-service guarantee definition in [10].

This ensures that working state bandwidth will resume within a tolerable prescribed delay δ_k . An example application is rescheduling a video conference meeting within a couple of hours. It also ensures that the connection has access to working state bandwidth for a fraction of time that is approximately α_k during faults, and α_k can be chosen high enough to provide good average throughput. An example application that needs good throughput is offline backup.

The next lemma shows how the maximum protection state duration δ_k and minimum working state duration μ_k imply an access rate to working state bandwidth.

Lemma 1: Suppose a connection k has a minimum working state duration μ_k and a maximum protection state duration δ_k . Then, during any interval $[x, y]$, the amount of time the connection is in the working state is at least $\frac{\mu_k}{\mu_k + \delta_k} \cdot \max\{y - x - \delta_k, 0\}$.

The proof of the lemma is given in Appendix A. From the lemma, it can be assumed without loss of generality that $\alpha_k \geq \frac{\mu_k}{\mu_k + \delta_k}$. To achieve bandwidth efficiency, in many cases, the value of α_k must be strictly larger than $\frac{\mu_k}{\mu_k + \delta_k}$. For example, suppose in Fig. 1 there are two connections, 1 and 2, that are on links 1 and 2, respectively, when there

are no faults. Suppose they have working state bandwidth B and protection state bandwidth 0. Suppose connection 1 has $\mu_1 = \delta_1 = 1$ hour, and connection 2 has $\mu_2 = \delta_2 = 4$ hours. Suppose $\alpha_k = \frac{\mu_k}{\mu_k + \delta_k} = 0.5$, so whenever link 2 has a fault, the connections are in the working state for approximately 50% of the time. However, with these values of μ_k and δ_k , if there is a fault on link 2 for say 8 hours, then connection 2 will be in the working state for at least 4 hours, during which connection 1 will also be in the working state at some time. Since both connections will be in the working state on link 1 at the same time, link 1 must have bandwidth $2B$. So even though the two connections each only require average bandwidth $B/2$ on link 1 during the fault, the link must have bandwidth $2B$, and the link will be utilized at only about 50%.

Note that subsets of the SLA components can be disabled by choosing appropriate parameter values. For example, to disable SLA1, SLA2, SLA3, or SLA4, the parameters can be chosen so that $b_k = 0$, $D_k = \Sigma F$, $\mu_k = 0$, or $\delta_k = D_k$, respectively.

The following are mild assumptions on the SLA parameters to simplify results in Section IV.

Assumption 1: Without loss of generality, each connection k is assumed to satisfy: $D_k \leq \Sigma F$ (for SLA2); $\delta_k \leq D_k$ (for SLA4); and $\alpha_k \geq \frac{\mu_k}{\mu_k + \delta_k}$ (for SLA4).

A. Examples of Service Mixes

To add some intuition to the variety of options that our general SLA framework provides, this example presents possible service mixes for Fig. 1 when each link is 30 Gb/s, and each link has three 10-Gb/s connections, so there are six connections altogether. The following are services that could be supported by the system.

Service Mix 1: All six connections are unprotected.

Service Mix 2: Three of the connections are fully protected, and the other three are low-priority. This is a classical SONET scenario with extra traffic.

Service Mix 3: Whenever a fault occurs, all six connections have protection state bandwidth $b_k = 5$ Gb/s. This scenario would apply to a real-time application that can fall back to a degraded performance at a certain reduced data rate. An example is given in the Section VI of connections carrying high-definition TV (HDTV) video, but when there is a fault, the connections reduce their bandwidth to carry standard-definition TV (SDTV) video.

Service Mix 4: Whenever a fault occurs, half of the connections are at 10 Gb/s, and the other half have no bandwidth. The connections take turns at having the working state bandwidth of 10 Gb/s and switch every 2 hours. This is a “rolling blackout” strategy to share the surviving bandwidth. This corresponds to Proposition 4 in Section IV, where $\mu_k = \delta_k = 2$ hours, $b_k = 0$, and a parameter h of the proposition is equal to 2. This scenario corresponds to a nonreal-time application, such as data center backup during off hours.

Service Mix 5: Whenever a fault occurs, four connections have protection state bandwidth of 2.5 Gb/s, and two of the connections have working state bandwidth of 10 Gb/s. The connections take turns at having the working state bandwidth of 10 Gb/s every 2 hours. This corresponds to Proposition 4 in Section IV,

where $\mu_k = 2$ hours, $\delta_k = 4$ hours, $b_k = 2.5$ Gb/s, and parameter h of the proposition is equal to 3. It corresponds to a mixture of services—some real-time applications with a degraded fallback option and some offline backup services.

IV. NETWORK COST

In this section, the link bandwidth requirements will be discussed for the system in Fig. 1 given the SLA framework of Section III. We will present upper and lower bounds on the link bandwidth that depend on SLA parameter values. These bounds are presented in four propositions numbered 1–4. The propositions also have constraints on the values of the parameters that show how the parameters are related to each other. For example, the choice of the minimum working state durations μ_k and the short-term availability rates α_k will determine how small the maximum protection state durations δ_k can be.

Proposition 1 has simple lower bounds. Proposition 2 presents a simple upper bound, but where the short-term availability conditions SLA3 and SLA4 are essentially ignored. Propositions 3 and 4 have upper bounds when SLA3 and SLA4 are included. Proposition 4 has additional restrictions on the values of minimum working state durations, maximum protection state durations, and short-term availability rates that lead to lower protection state durations. Though Proposition 4 has restricted SLA parameter values, it can still be applied to useful services such as Service Mixes 4 and 5 in Section III.

Next, the four propositions will be presented, followed by a comparison of the bounds.

Proposition 1: Suppose there are K connections. Suppose B is the working state bandwidth and b is the protection state bandwidth from SLA1. For each connection k , let D_k be the maximum accumulated protection state time from SLA2, μ_k be the minimum working state duration from SLA3, and δ_k be the maximum protection state duration and α_k be the short-term availability rate from SLA4. Suppose Assumption 1 is true. Then, the bandwidth on a link is at least:

- $\frac{1}{2}KB$;
- $Kb + (B - b) \sum_k \rho_k$, where $\rho_k = 1 - \frac{D_k}{\Sigma F}$; and
- $Kb + (B - b) \sum_k \tilde{\alpha}_k$, where $\tilde{\alpha}_k = \alpha_k(1 - \frac{\delta_k}{f})$.

Proof: Part (a) is the link bandwidth required for K connections in their working states when there are no faults. For part (b), suppose the accumulated times when there is a fault is ΣF . SLA2 implies that, during faults, the fraction of time that a connection k is in the protection state is at most $D_k/\Sigma F$, or equivalently the fraction of time that the connection is in the working state is at least ρ_k . Thus, during the fault, the average bandwidth for connection k is at least $b + \rho_k \cdot (B - b)$. This implies part (b). For part (c), suppose there is a single fault with duration f . SLA4 implies that during the fault, connection k will be in the working state for at least $\alpha_k \cdot (f - \delta_k)$ amount of time, or equivalently the fraction of time that the connection is in the working state is at least $\tilde{\alpha}_k$. Thus, during the fault, the average bandwidth required by connection k is at least $b + \tilde{\alpha}_k \cdot (B - b)$. This implies part (c). **QED**

Proposition 2: Suppose there are K connections. Suppose B is the working state bandwidth and b is the protection state bandwidth from SLA1. For each connection k , let D_k be the maximum accumulated protection state time from SLA2. Suppose

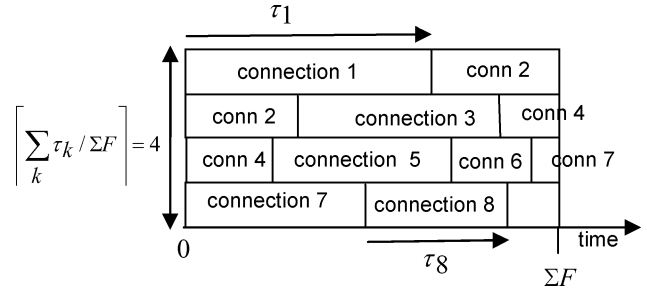


Fig. 2. A surviving bandwidth schedule S .

the short-term availability constraints SLA3 and SLA4 are ignored, i.e., for each connection k , minimum working state duration $\mu_k = 0$, and maximum protection state duration $\delta_k = D_k$. Suppose Assumption 1 is true. To satisfy the SLA, the following link bandwidth is sufficient:

$$\max \left\{ \left\lceil \frac{K}{2} \right\rceil B, K \cdot b + \left[\sum_k \rho_k \right] \cdot (B - b) \right\},$$

where ρ_k is defined in Proposition 1.

Proof: A bandwidth schedule for the connections will be presented that assumes that there is one down link and one surviving link during the time interval $[0, \Sigma F]$. The schedule determines when the connections are in their working and protection states and is defined over $[0, \Sigma F]$. This will be referred to as the *surviving bandwidth schedule* S and is illustrated in Fig. 2. In the figure, the rows correspond to connections that are in their working states over time. For example, at time 0, connections 1, 2, 4, and 7 are in their working states; then after some time, connection 4 goes into its protection state, and connection 5 goes into its working state. Finally, at time ΣF , connections 2, 4, and 7 are in their working states. Note that a valid schedule requires that a connection cannot appear in two or more rows at the same time.

The connections are arranged in the figure as follows. Each connection k is scheduled to be in the working state for an accumulated time $\tau_k = \Sigma F - D_k$, so it is in the protection state for an accumulated time at most D_k , and SLA2 is satisfied. Connections are scheduled in order, starting with connection 1, filling one row at a time, with wraparound at the end of the row (at time ΣF) to the beginning of the next. Since each connection k has $\tau_k \leq \Sigma F$, it appears in at most one row at any time. Connections 2, 4, and 7 are examples of connections that have their schedules wrapped around from one row to the next.

The surviving bandwidth schedule S will be used as follows. Whenever a fault occurs, the connections follow S , where on the first fault, the connections begin following S from its starting time 0. When there is no fault, all connections go to their working states. When the next fault occurs, the connections resume following S from where they left off. For example, suppose schedule S is Fig. 2, and there are three faults of durations L_1, L_2 , and L_3 . Then, Fig. 3 shows how the schedule is followed during faults. For example, fault 2 begins with connections 1, 3, 5, and 7 in their working states and ends with connections 1, 3, 5, and 8 in their working states.

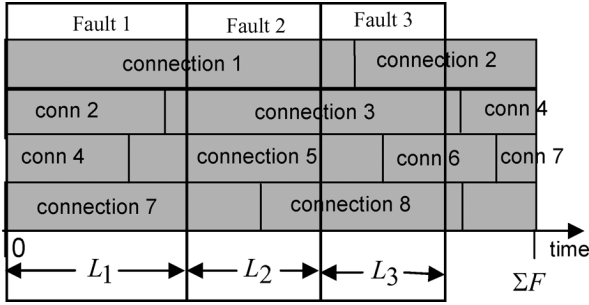


Fig. 3. An example for three faults.

Note that the connections k satisfy SLA2 since they are in their protection states only during faults, during which schedule S has them in their protection states for accumulated time at most D_k .

The link bandwidth of the proposition will be shown to be sufficient. When there are no faults, then $\lceil K/2 \rceil \cdot B$ per link is sufficient to carry K connections. When there is a fault, the connections follow schedule S . Then, the number of connections in their working states is at most $\lceil \sum_k \tau_k / \Sigma F \rceil$, which is the number of rows in Fig. 2. Since $\lceil \sum_k \tau_k / \Sigma F \rceil = \lceil \sum_k \rho_k \rceil$, the required bandwidth for the connections on the surviving link is $K \cdot b + \lceil \sum_k \rho_k \rceil (B - b)$. Therefore, the link bandwidth of the proposition is sufficient. **QED**

Proposition 3: Suppose there are K connections. Let B be the working state bandwidth and b be the protection state bandwidth from SLA1. Suppose each connection k has maximum accumulated protection state time D_k for SLA2, minimum working state duration μ_k for SLA3, and maximum protection state duration δ_k and short-term availability rate α_k for SLA4. Suppose Assumption 1 is true. Suppose the minimum fault-free duration is at least $\max_i \mu_i$. Suppose each connection k satisfies

$$\delta_k \geq \frac{2\mu_k}{\max\{\alpha_k, \tilde{\rho}_k\}} + \max_i \mu_i - \mu_k \cdot \left(1 + 1 / \left[\sum_i \max\{\alpha_i, \tilde{\rho}_i\} \right] \right) \quad (1)$$

where $\tilde{\rho}_i = \frac{\rho_i + \omega_i}{1 + \omega_i - \max_j \omega_j}$, $\omega_i = \mu_i / \Sigma F$, and ρ_i is from Proposition 1. To satisfy the SLA, the following link bandwidth is sufficient:

$$\max \left\{ \left\lceil \frac{K}{2} \right\rceil B, K \cdot b + \left\lceil \sum_k \max\{\tilde{\rho}_k, \alpha_k\} \right\rceil \cdot (B - b) \right\}. \quad (2)$$

The proof of the proposition is presented in Appendix B. The proposition shows how the SLA parameters are related. For example, (1) implies that the maximum protection state duration δ_k is proportional to the minimum working state duration μ_k and inversely proportional to the short-term availability rate α_k .

Note that $\tilde{\rho}_k$ in the link bandwidth formula (2) is approximately equal to ρ_k , which is part of previous link bandwidth formulas in Propositions 1 and 2. It is a close approximation if

TABLE I
COMPARISON OF ρ_k AND $\tilde{\rho}_k$ WHEN AVAILABILITY IS 99.9%

| | ΣF | | | |
|--|------------|--------|--------|--------|
| | 12 hrs | 18 hrs | 24 hrs | 30 hrs |
| ρ_k | 0.27 | 0.51 | 0.64 | 0.71 |
| $\tilde{\rho}_k (\mu = 0.25 \text{ hr})$ | 0.29 | 0.53 | 0.65 | 0.72 |
| $\tilde{\rho}_k (\mu = 0.50 \text{ hr})$ | 0.31 | 0.54 | 0.66 | 0.72 |
| $\tilde{\rho}_k (\mu = 1.00 \text{ hr})$ | 0.35 | 0.57 | 0.68 | 0.74 |
| $\tilde{\rho}_k (\mu = 2.00 \text{ hr})$ | 0.44 | 0.62 | 0.72 | 0.77 |

the minimum working state durations $\max_i \mu_i$ are much smaller than the accumulated time of all failures ΣF . Table I compares the values of $\tilde{\rho}_k$ and ρ_k for different values of μ_k and ΣF when the network lifetime T is a year, the long-term availability $1 - D_k/T$ is 99.9% (i.e., $D_k = 8.76$ hours), and assuming all connections have the same minimum working state duration $\mu_k = \mu$. Note that the values are about the same if the minimum working state durations are at most an hour.

In the next Proposition 4, there are restrictions on the values of the minimum working state duration, maximum protection state duration, and short-term availability rate. By constraining the values, the connections can be scheduled more efficiently during faults, and this can shorten the maximum protection state duration for a given working state duration and short-term availability rate.

Proposition 4: Suppose there are K connections. Let B be the working state bandwidth and b be the protection state bandwidth from SLA1. For each connection k , let D_k be the maximum accumulated protection state time for SLA2. Suppose all connections have the same minimum working state duration μ for SLA3. For SLA4, suppose there is an integer $h \geq 2$, and for each connection k , there is a nonnegative integer $m(k)$ such that the maximum protection state duration is $\delta_k = \mu \cdot (h^{m(k)} - 1)$ and the short-term availability rate $\alpha_k = \frac{\mu_k}{\mu_k + \delta_k} = h^{-m(k)}$. In addition, suppose the maximum accumulated protection state time D_k is sufficiently large so that $D_k \geq (1 - \alpha_k) \cdot (\Sigma F + \mu)$. Suppose Assumption 1 is true. Suppose the minimum fault-free duration is at least μ . To satisfy the SLA, the following link bandwidth is sufficient:

$$\max \left\{ \left\lceil \frac{K}{2} \right\rceil B, K \cdot b + \left\lceil \sum_k \alpha_k \right\rceil (B - b) \right\}.$$

The proof of the proposition is left in Appendix C. The proof relies on a bandwidth schedule when failures occur. Also, the constraint $D_k \geq (1 - \alpha_k) \cdot (\Sigma F + \mu)$ ensures SLA2 is satisfied under the schedule. This constraint implies that the short-term availability rate α_k (which affects the link bandwidth) must satisfy $\alpha_k \geq 1 - \frac{D_k}{\Sigma F + \mu}$ under the assumptions of the proposition.

To illustrate that Proposition 4 leads to smaller values of maximum protection state duration δ_k , suppose that the assumptions of Proposition 4 are true and for all connections k , $\tilde{\rho}_k = \alpha_k$. Then, Propositions 3 and 4 have the same link bandwidth value. Note that $\delta_k \geq \mu \cdot \left(\frac{2}{\alpha_k} - 1 / \lceil \sum_i \alpha_i \rceil \right)$ for Proposition 3, whereas

TABLE II
COMPARISON OF LINK BANDWIDTH FORMULAS

| | |
|---------------------|---|
| <i>Prop 2:</i> | $\max\left\{\left\lceil\frac{K}{2}\right\rceil B, K \cdot b + \left\lceil\sum_k \rho_k\right\rceil \cdot (B-b)\right\}$ |
| <i>Lower bound:</i> | $\max\left\{\frac{K}{2} B, K \cdot b + \sum_k \rho_k \cdot (B-b)\right\}$ |
| <i>Prop 3:</i> | $\max\left\{\left\lceil\frac{K}{2}\right\rceil B, K \cdot b + \left\lceil\sum_k \max\{\tilde{\rho}_k, \alpha_k\}\right\rceil \cdot (B-b)\right\}$ |
| <i>Lower Bound:</i> | $\max\left\{\frac{K}{2} B, K \cdot b + \max\left\{\sum_k \rho_k, \sum_k \tilde{\alpha}_k\right\} \cdot (B-b)\right\}$ |
| <i>Prop 4:</i> | $\max\left\{\left\lceil\frac{K}{2}\right\rceil B, K \cdot b + \left\lceil\sum_k \alpha_k\right\rceil \cdot (B-b)\right\}$ |
| <i>Lower Bound:</i> | $\max\left\{\frac{K}{2} B, K \cdot b + \sum_k \tilde{\alpha}_k \cdot (B-b)\right\}$ |

$\delta_k = \mu \cdot (h^{-m(k)} - 1) = \mu \cdot \left(\frac{1}{\alpha_k} - 1\right)$ for Proposition 4. Therefore, the constraints of Proposition 4 could reduce the maximum protection state duration δ_k by more than a half for the same link bandwidth value.

To check that the link bandwidths of Propositions 2, 3, and 4 are minimal, Table II compares them to the lower bounds from Proposition 1. Note that the link bandwidth of the propositions are close to the lower bounds if $\alpha_k \approx \tilde{\alpha}_k$ and $\rho_k \approx \tilde{\rho}_k$. A case when $\alpha_k \approx \tilde{\alpha}_k$ is where the maximum protection state duration δ_k is much smaller than the maximum time f to repair a link. Cases when $\rho_k \approx \tilde{\rho}_k$ are shown in Table I.

Table III presents an example set of values from the formulas in Table II for ΣF ranging from 12 to 30 hours. For example, ‘‘Prop. 2’’ and ‘‘Lower Bound 2’’ in Table III correspond to ‘‘Prop. 2’’ and its lower bound in Table II. The values in Table III are for the parameter values $T = 1$ year, $f = 6$ hours, $K = 32$ connections, and for all connections k , the long-term availability $1 - D_k/T$ is 99.9% (i.e., $D_k = 8.76$ hours), the minimum working state duration μ_k is 1 hour, the working state bandwidth $B_k = 1$, the protection state bandwidth $b_k = 0.25$, and the short-term availability rate $\alpha_k = 0.5$. The maximum protection state duration δ_k for the link bandwidth of Proposition 3 was chosen to be the minimum possible under constraint (1). The values were 4.0, 3.6, 3.0, and 2.7 for ΣF values of 12, 18, 24, and 30 hours, respectively. The value of δ_k is 1 hour for the link bandwidth of Proposition 4. Note that in the table, no values are given for ‘‘Prop. 4’’ and its lower bound when $\Sigma F = 18, 24$, and 30 hours because then the constraint $D_k \geq (1 - \alpha_k) \cdot (\Sigma F + \mu)$ of Proposition 4 is violated.

Table III shows that the link bandwidths of the propositions can be close to the lower bounds. Also, note that a link bandwidth of 16 is minimum since it is required for the 32 connections when there are no failures. Therefore, the link bandwidth of Proposition 2 for $\Sigma F = 12$ hours does not require additional protection bandwidth.

TABLE III
COMPARISON OF LINK BANDWIDTH VALUES

| | ΣF | | | |
|---------------|------------|----------|----------|----------|
| | 12 hours | 18 hours | 24 hours | 30 hours |
| Prop. 2 | 16.00 | 20.75 | 23.75 | 25.25 |
| Lower Bound 2 | 16.00 | 20.32 | 23.24 | 24.99 |
| Prop. 3 | 20.00 | 22.25 | 24.50 | 26.00 |
| Lower Bound 3 | 16.48 | 21.65 | 24.24 | 25.79 |
| Prop. 4 | 20.00 | | | |
| Lower Bound 4 | 18.00 | | | |

V. ECONOMIC CONSIDERATIONS

While the theoretical merit of the general SLA framework in Section III is from its ability to express all types of protection guarantees under one formal umbrella, this approach will only be adopted in real networks if it provides economic advantages over current approaches.

A. Comparing Protection Services Given Fixed Link Bandwidth

Service Mixes 1–4 in Section III will be used to illustrate when there are economic advantages of nontraditional protection service. Note that they all have six connections with working state bandwidth of 10 Gb/s, and so the only differences are their protection. In addition, the link bandwidths are the same at 30 Gb/s per link. Thus, network bandwidth costs are the same. Switching and management costs will be ignored in this section, but discussed in Section VI.

Service Mixes 1 and 2 are classical protection services, where Service Mix 1 is classical unprotected service, while Service Mix 2 is a mix of 50% of 1:1 fully protected service and 50% of low-priority, preemptible service. We now provide formulas for the revenue per connection. Let $R(0)$ denote the revenue generated by an unprotected connection, and $R(1)$ denote the revenue by fully protected service. Here, the notation $R(r)$ denotes the revenue of a connection when a fraction r of its working bandwidth is protected. Note that $R(r)$ is increasing with r because more bandwidth is protected. Let $\tilde{R}(0)$ denote the revenue by a low-priority, preemptible connection. Note that $R(0) \geq \tilde{R}(0)$ because the availability of working state bandwidth for unprotected service is higher than low-priority preemptible service. Note that the revenue per connection for Service Mix 1 is $R(0)$, and the revenue connection for Service Mix 2 is $\frac{\tilde{R}(0)+R(1)}{2}$.

Service Mix 3 is a nontraditional protection scenario such that whenever a fault occurs, all six connections go to their protection state bandwidth of 5 Gb/s. It has the same availability of working state bandwidth as low-priority preemptible service, but it has the advantage of having protection state bandwidth while low-priority preemptible service has none. Let $\tilde{R}(0.5)$ denote its revenue per connection. A possible application of this service mix is low-cost protection for HDTV. During faults, the HDTV connection is downgraded to SDTV, which requires 20%–25% of the HDTV bandwidth. Therefore, it maintains connectivity during faults, which would be important for HDTV of live events or video conferencing.

Then, Service Mix 3 will have better revenue per connection than classical protection methods if

$$\tilde{R}(0.5) > \max \left\{ \frac{\tilde{R}(0) + R(1)}{2}, R(0) \right\}.$$

Also, Service Mix 3 compares even more favorably with 1+1 protection since a 1 + 1 connection uses 10 Gb/s on both links, and so its revenue per connection is $R(1)/2$.

Service Mix 4 is a nontraditional service that has the six connections equally sharing the bandwidth on the surviving link when a fault occurs, similar to a “rolling blackout” strategy. The minimum working state duration and the maximum protection state duration are 2 hours. This is an improvement over classical unprotected service, where a fault-prone link can leave connections without any service for long periods, e.g., 8 hours or more. Thus, presumably Service Mix 4 will lead to better revenues than classical unprotected service.

B. Link Bandwidth Grows With Protection Level

In the previous discussion, the link bandwidth was fixed. We now consider a case when the amount of link bandwidth will depend on the protection level. It will refer to the following.

Scenario 1. There are K connections. The connections have the same working state bandwidth B and protection state bandwidth $r \cdot B$, where r is a fraction. The connections adhere to the standard telecommunication practice of not disturbing established connections. Therefore, in the case of Fig. 1, the amount of link bandwidth is $(K/2) \cdot B \cdot (1 + r)$, and the network bandwidth cost is $Q(r) = K \cdot B \cdot (1 + r)$. Thus, the network bandwidth cost grows linearly with the protection level r . Note that the cases of $r = 0$ and 1 correspond to classical protection, where $r = 0$ is unprotected service, and $r = 1$ is fully protected service, e.g., 1+1 or 1:1. The cases of nontraditional protection are when $0 < r < 1$. The next propositions present conditions when nontraditional protection is attractive or unattractive. Let $R(r)$ denote the revenue per connection, so the total revenue is $K \cdot R(r)$. The profit is $P(r) = K \cdot R(r) - Q(r)$.

Proposition 5: For Scenario 1, nontraditional protection service ($0 < r < 1$) will be attractive if the profit P satisfies $P(r) > \max\{P(0), P(1)\}$. (Note that classical fully protected service is $P(1)$ and classical unprotected service is $P(0)$.)

The following is an observation about how the “shape” of the function $R(r)$ can imply that classical protection will be preferred. It applies to more general network topologies and connections.

Proposition 6: Consider Scenario 1, but extended from two links between a pair of nodes to a network with arbitrary topology, and the connections are between arbitrary pairs of nodes. Suppose for each routing x of the working and protection paths of the connections, the network bandwidth cost is a linear function of r , which we denote by $Q_x(r)$. Assume that the routing is optimal, and so the network cost is $Q(r) = \min_x Q_x(r)$. Suppose the revenue per connection $R(r)$ is a convex function of r . Then, the profit is maximized at one of the extreme values of $r = 0$ or 1, which corresponds to classical protection service.

Proof: Since Q is the minimum of linear functions of r , it is a concave function of r . Note that profit $P(r)$ is a linear combination of $R(r)$ and $-Q(r)$, which is a convex function of r . If $R(r)$ is a convex function of r , then $P(r)$ is convex and maximized at one of the extreme values of $r = 0$ or 1. **QED**

VI. IMPLEMENTATION NOTES

This section discusses implementation issues of the SLA framework. First, we note that the framework can be implemented for both circuit-switched networks (e.g., SONET/SDH, OTN, or wavelength connections) and packet switched networks (e.g., IP, MPLS, or Ethernet).

To support SLA1, mechanisms are required to support the switching between the working state bandwidth B_k and the protection state bandwidth b_k . In the case of circuit switching, this may be implemented by inverse multiplexing lower speed connections, e.g., SONET or OTN Virtual Concatenation (VCAT), to achieve variable bandwidth. In order to switch between these modes in a graceful way, without losing traffic, the system can use the LCAS protocol.

In the case of packet switching, this may be implemented with input packet regulators, e.g., leaky bucket regulators with their rates adjustable to B_k or b_k . An alternative method is to prioritize packets and drop the packets of low priority. Most packet networks can be very easily adapted to support the protection approach by exploiting the existing priority field in packets. This can be done in one of two ways. The first way is dynamic priority based on the current priority of the service compared to other services at the same endpoints. For example, if a particular connection should be victimized, its priority is decreased until it is victimized.

The second way is a mix of high-priority and low-priority packets to support two bit rates as defined in SLA1. For example, consider an HDTV–SDTV service where HDTV signals are transported under normal operation, but lower bandwidth SDTV signals may be transported if there is a fault. Thus, if the SDTV bandwidth is 20% of the HDTV bandwidth, then b_k is equal to 20% of B_k . Encoding methods such as hierarchical or layered coding [11] can help realize this approach. The encoding has two layers of information, where the first layer is the encoded SDTV signal and the second layer is used to build on the first to get the HDTV signal. The first layer is sent in high-priority packets, while the second layer is sent on low-priority packets. When a fault occurs, all packets are switched to the surviving link. Then, if the congestion becomes high, the low-priority packets can be dropped. However, the end-user will still receive the high-priority packets and the SDTV video stream. (Another example of using hierarchical encoding for video is presented in [12], though in this case, feedback is used for the server to adapt its transmission to the available bandwidth.) In both ways, the network must be modified at the edge to set the priorities in a proper way, but the core of the network need not be aware of this.

Supporting SLA2, SLA3, and SLA4 requires that the network keep state information per connection, e.g., to keep track of how much time connections have been in working and protection states. It may require scheduling algorithms to determine when connections are in the working and protection states. The proofs

of Propositions 2, 3, and 4 suggest that this schedule can be computed and stored in advance and followed when faults actually occur. While adding such state is nontrivial, it should be noted that edge routers and similar devices already contain per-user state, although the current state is not as dynamic as needed for the proposed scheme.

Finally, the scheme requires coordination between the two nodes to decide which connections are in their working state and which are in their protect state and to switch between working and protection states. This can be implemented with extensions to the SONET Automatic Protection Switching (APS) protocol.

VII. FINAL REMARKS

A general framework for availability was presented that leads to bandwidth efficiency in point-to-point systems across a network. The framework allows two levels of bandwidth, the working state and protection state bandwidth levels. It addresses short-term availability, which is lacking in conventional availability guarantees in SLAs.

It is straightforward to extend the results from the two links between two nodes in Fig. 1 to a multiple of J links between the two nodes assuming at most one link can fail at any time. The bandwidth formulas in Propositions 1–4 have two components: bandwidth under normal operation and bandwidth when there is a failure. Under normal operation, all K connections are in their working states, so each link must have at least $K \cdot B/J$ bandwidth in Proposition 1(a) and $\lceil K \cdot B/J \rceil$ bandwidth in Propositions 2–4. During failures, the lower bound required in (b) and (c) of Proposition 1 is divided among the $J - 1$ surviving links. During failures, the bandwidth of Propositions 2–4 is due to schedules such as in Fig. 2. For example, in Proposition 2, the schedule has $\lceil \sum_k \rho_k \rceil$ connections in the working state and $K - \lceil \sum_k \rho_k \rceil$ connections in the protection state at any time, and recall that connections in the working state need B bandwidth and those in the protection state need b bandwidth. (Note that ρ_k is a function of ΣF , which is now the sum of F_k over all J links.) A simple bandwidth allocation to ensure that the surviving $J - 1$ links have sufficient bandwidth for the connections is for each link to have $\lceil \lceil \sum_k \rho_k \rceil / (J - 1) \rceil \cdot B + \lceil (K - \lceil \sum_k \rho_k \rceil) / (J - 1) \rceil \cdot b$ bandwidth. Similar link bandwidth allocations apply to Propositions 3 and 4.

A future direction of research is to extend the framework beyond the configuration of two connection terminating nodes to a more general network of connections. If only SLA1 is considered, then determining the required bandwidth appears to be straightforward. Here, connections have working and protection paths, and link bandwidths are determined by considering all possible faults. On the other hand, the signaling and management may be nontrivial because when a fault occurs, connections not on the fault may have to switch to their protection path and bandwidth, and it may be difficult to coordinate this in a distributed fashion.

Another direction of future research is to generalize the SLA framework to multiple protection states per connection, each state with their own bandwidth and duration. This may be of interest if the service represents an aggregation of multiple applications with their distinct service needs. Also, one could consider other SLA constraints where connections are guaranteed

bandwidth at specific times. For example, a connection k may require its working state bandwidth B_k from 1–3 p.m. every day. However, such constraints can create traffic demand “peaks” at popular time periods, which may lead to significant underutilization during traffic demand “valleys.”

APPENDIX A PROOF OF LEMMA 1

First, it will be shown that during any time interval of duration $L \leq \mu_k + \delta_k$, the accumulated time a connection k is in its working state is at least

$$\max\{L - \delta_k, 0\}. \quad (\text{A.1})$$

If connection k goes into the protection state more than once, then between protection states, the connection is in the working state for duration at least μ_k , which is at least (A.1). Otherwise, the connection is in the protection state at most once and, for accumulated time, at most δ_k . In either case, the connection is in the working state for accumulated time at least (A.1).

Now consider an arbitrary interval $[x, y]$. It can be divided into subintervals, where the first $\lfloor \frac{y-x}{\mu_k + \delta_k} \rfloor$ subintervals have duration $\mu_k + \delta_k$ and the last subinterval has duration L , where L is a value that satisfies $0 \leq L < \mu_k + \delta_k$, i.e.,

$$y - x = (\mu_k + \delta_k) \cdot \left\lfloor \frac{y - x}{\mu_k + \delta_k} \right\rfloor + L. \quad (\text{A.2})$$

Applying (A.1) to each of the subintervals, implies that the amount of time connection k is in the working state during time interval $[x, y]$ is at least

$$\begin{aligned} & \mu_k \cdot \left\lfloor \frac{y - x}{\mu_k + \delta_k} \right\rfloor + \max\{L - \delta_k, 0\} \\ & \geq \mu_k \cdot \left\lfloor \frac{y - x}{\mu_k + \delta_k} \right\rfloor + \frac{\mu_k}{\mu_k + \delta_k} \max\{L - \delta_k, 0\} \\ & = \frac{\mu_k}{\mu_k + \delta_k} \left((\mu_k + \delta_k) \cdot \left\lfloor \frac{y - x}{\mu_k + \delta_k} \right\rfloor + \max\{L - \delta_k, 0\} \right) \\ & \geq \frac{\mu_k}{\mu_k + \delta_k} \max \left\{ (\mu_k + \delta_k) \cdot \left\lfloor \frac{y - x}{\mu_k + \delta_k} \right\rfloor + L - \delta_k, 0 \right\} \\ & = \frac{\mu_k}{\mu_k + \delta_k} \max\{y - x - \delta_k, 0\} \end{aligned}$$

where the last equality is due to (A.2). The lemma is implied.

APPENDIX B PROOF OF PROPOSITION 3

Note that it is sufficient to prove SLA2, SLA3, and SLA4 are true for the given link bandwidth. As in the Proof of Proposition 2, a *surviving bandwidth schedule* S for the connections will be presented that assumes that during the time interval $[0, \Sigma F]$, there is one down link and one surviving link. The schedule determines when the connections are in their working and protection states over the interval.

Surviving Bandwidth Schedule S : Consider the packet switched system in [13], which is a queueing system with parallel transmission links with unit capacity. It has flows of packets, which are transmitted according to a scheduling

algorithm referred to as MSFQ in [13]. The algorithm is based on Generalized Processor Sharing (GPS) [14] and uses a GPS server model as a reference. Each packet flow k has a parameter ϕ_k that determines the rate of service to the flow. We will assume that there are $\lceil \sum_k \phi_k \rceil$ parallel links in the system. Then, each packet flow k will have a guaranteed service rate ϕ_k from the GPS server reference model. We will also assume that for each packet flow k : 1) packets use exactly μ_k amount of time to transmit on a link; 2) the first packet arrives at time 0; and 3) packet interarrival times are constant and equal to μ_k/ϕ_k . Note that under the GPS server reference model, a packet will commence service when it arrives and use at most μ_k/ϕ_k time to complete service, which is before the next packet arrives for flow k . The upper bounds in [14] imply that a packet will complete transmission in time at most $\mu_k/\phi_k + \beta_k$ after it arrives, where $\beta_k = \mu_k + \max_i \mu_i - \mu_k/\lceil \sum_i \phi_i \rceil$.

These results are applied to define schedule S as follows. Each connection k corresponds to a “flow of packets,” where value ϕ_k is equal to $\max\{\alpha_k, \tilde{\rho}_k\}$. A “packet” being “transmitted” on a “link” corresponds to when connection k is in the working state. The value μ_k is the minimum working state duration of the connection.

As in the Proof of Proposition 2, the connections follow schedule S whenever there is a fault. When there are no faults, then all connections are in their working states.

Lemma B.1: Assume that if the connections k follow schedule S , then they have maximum accumulated protection state time D_k , minimum working state duration μ_k , maximum protection state duration δ_k , and short-term availability rate α_k . Now, suppose connections follow schedule S only during faults (as in the Proof of Proposition 2) and are in their working states when there are no faults. In addition, suppose the minimum fault-free period is at least $\max_i \mu_i$. Then, the connections k will still satisfy SLA2 with parameter D_k , SLA3 with parameter μ_k , and SLA4 with parameters δ_k and α_k .

Proof: To check SLA2, note that connection k is in the protection state only during faults, during which it follows schedule S . By assumption, schedule S will keep the connection’s accumulated time in the protection state to at most D_k , so SLA2 is true.

To check SLA3, consider an arbitrary working state period π of connection k . Consider the following two cases. In the first case, suppose π overlaps a fault-free period. Then, its duration is at least the duration of the fault-free period, which is at least $\max_i \mu_i \geq \mu_k$. In the second case, suppose π does not overlap a fault-free period. Then, it is within a faulty period, during which the connection follows schedule S . Therefore, by the assumption on schedule S , π has duration at least μ_k . Thus, in either case, the minimum working state duration is at least μ_k .

To check SLA4, it will be shown that the amount of time a connection k is in the working state during an arbitrary interval $[x, y]$ is at least $\alpha_k \cdot (y - x - \delta_k)$. First, consider the accumulated amount of time during $[x, y]$ when there is a fault, and denote the amount by z . Note that during these times, the connection follows the surviving bandwidth schedule S over some time interval $[x', y']$ of the schedule S . Therefore, $y' - x' = z$. From the assumption in the lemma, during these times, the amount of time the connection is in the working state is at least

$\alpha_k \cdot (z - \delta_k)$. The rest of the times in $[x, y]$ are fault-free, and the connection is in the working state. Thus, the amount of time a connection k is in the working state during $[x, y]$ is at least $\alpha_k \cdot (z - \delta_k) + 1 \cdot (y - x - z) \geq \alpha_k \cdot (y - x - \delta_k)$.

Finally, it will be shown that the maximum protection state duration for a connection k is at most δ_k by contradiction. Suppose there is a protection state duration $L > \delta_k$. Then, during that period, the amount of time that the connection is in the working state is at least $\alpha_k \cdot (L - \delta_k) > 0$, so the connection spends some time in the working state. This contradicts the assumption that the entire period is a protection state period. Therefore, the maximum protection state duration is at most δ_k . **QED**

Given Lemma B.1, the following lemma implies that the connections satisfy SLA2, SLA3, and SLA4.

Lemma B.2: Suppose the connections follow schedule S over the time interval $[0, \Sigma F]$. Then, for the connections k : (a) the maximum accumulated protection state time is D_k ; (b) the minimum working state duration is μ_k ; (c) the maximum protection state duration is δ_k ; and (d) the short-term availability rate is α_k .

The proof of the lemma will be presented after the following result.

Lemma B.3: Suppose a connection k follows surviving bandwidth schedule S . Then, for any subinterval $[x', y']$ of time interval $[0, \Sigma F]$, the accumulated amount of time the connection is in the working state during $[x', y']$ is at least $\phi_k \cdot (y' - x' - \hat{\delta}_k)$, where $\hat{\delta}_k = 2\mu_k/\phi_k + \max_i \mu_i - \mu_k \cdot (1 + 1/\lceil \sum_i \phi_i \rceil)$.

Proof: Consider the following “transmission” schedule S^* , which will be used as a reference to compare to schedule S : For connection k , suppose its “packets” begin “transmission” as soon as they arrive. Since “packets” arrive every μ_k/ϕ_k time units and are “transmitted” in μ_k amount of time, it corresponds to protection state durations $\mu_k \cdot (1/\phi_k - 1)$ and working state durations μ_k .

Let a_n denote the arrival time of “packet” n for connection k . Then, under schedule S^* , “packet” n completes its “transmission” at time $a_n + \mu_k$. On the other hand, under schedule S , “packet” n is scheduled by the MSFQ algorithm [13] and may be “transmitted” later than $a_n + \mu_k$. From upper bounds in [13], a “packet” n completes its transmission at time no later than $a_n + \mu_k/\phi_k + \beta_k$, where β_k has been defined earlier in the description of schedule S . Therefore, the “transmission” of “packet” n by S compared to S^* may be delayed, but by at most $a_n + \mu_k/\phi_k + \beta_k - (a_n + \mu_k) = \mu_k \cdot (1/\phi_k - 1) + \beta_k$.

This implies that all portions of “packets” that are “transmitted” under schedule S^* in the time interval $[x', y' - (\mu_k \cdot (1/\phi_k - 1) + \beta_k)]$ will be “transmitted” in time interval $[x', y']$ under schedule S . Thus, to complete the proof, it will be shown that at least $\phi_k \cdot (y' - x' - \hat{\delta}_k)$ amount of “packets” are “transmitted” in time interval $[x', y' - (\mu_k \cdot (1/\phi_k - 1) + \beta_k)]$ under schedule S^* .

Schedule S^* has working state durations μ_k and protection state durations $\mu_k \cdot (1/\phi_k - 1)$. Therefore, Lemma 1 can be applied, and “transmissions” of schedule S^* during the time interval $[x', y' - (\mu_k \cdot (1/\phi_k - 1) + \beta_k)]$ are at least $\phi_k \cdot \max\{y' - x' - (2\mu_k \cdot (1/\phi_k - 1) + \beta_k), 0\}$. Lemma B.3 is implied because $2\mu_k \cdot (1/\phi_k - 1) + \beta_k = \hat{\delta}_k$. **QED**

Proof of Lemma B.2: To prove part (a) for a connection k , note that Lemma B.3 implies that under schedule S , the amount of time that the connection is in the working state over the duration of the schedule is at least

$$\begin{aligned}
\phi_k \cdot (\Sigma F - \hat{\delta}_k) &= \phi_k \cdot \Sigma F - \phi_k \cdot \hat{\delta}_k \\
&= \phi_k \cdot \Sigma F - 2\mu_k - \phi_k \cdot \max_i \mu_i \\
&\quad + \phi_k \cdot \mu_k \cdot \left(1 + 1/\left[\sum_i \phi_i\right]\right) \\
&\geq \phi_k \cdot \Sigma F - 2\mu_k - \phi_k \cdot \max_i \mu_i + \phi_k \cdot \mu_k \\
&= \phi_k \cdot (\Sigma F + \mu_k - \max_i \mu_i) - 2\mu_k \\
&\geq \tilde{\rho}_k \cdot (\Sigma F + \mu_k - \max_i \mu_i) - 2\mu_k \\
&= \Sigma F - D_k
\end{aligned}$$

where the last equality is due to $\tilde{\rho}_k = \frac{\Sigma F - D_k + 2\mu_k}{\Sigma F + \mu_k - \max_i \mu_i}$. Thus, connection k is in the protection state for at most time D_k , and part (a) is true.

For part (b), note that the minimum working state duration for a connection k is μ_k because a working state duration corresponds to a ‘‘packet transmission’’ of schedule S , which has duration μ_k .

To show parts (c) and (d), consider an arbitrary time interval $[x, y]$, and the amount of time a connection k is in the working state. Let z denote the accumulated amount of time in $[x, y]$ when there is some fault. Note that during these times, the connection follows the surviving bandwidth schedule S over some time interval $[x', y']$ of the schedule S . Thus, $y' - x' = z$. Then, from Lemma B.3, the amount of time the connection is in the working state in $[x', y']$ by following S is at least $\phi_k \cdot (z - \hat{\delta}_k)$. When there are no faults, the connection is in the working state. Therefore, the amount of time the connection is in the working state during $[x, y]$ is at least $\phi_k \cdot (z - \hat{\delta}_k) + 1 \cdot (y - x - z) \geq \alpha_k \cdot (y - x - \hat{\delta}_k) \geq \alpha_k \cdot (y - x - \delta_k)$, where the last inequality is due to $\delta_k \geq \hat{\delta}_k$ from (1). Thus, α_k is the short-term availability rate, and that proves (d).

We have just proven that for time interval $[x, y]$, the amount of time a connection k is in the working state is at least $\alpha_k \cdot (y - x - \delta_k)$. Therefore, we can use the argument at the end of the Proof of Lemma B.1 to show that δ_k is the maximum protection state duration. That verifies (c). **QED**

To complete the proof, it will be shown that the link bandwidth of the proposition is sufficient. When there are no faults, the link bandwidth of $\lceil K/2 \rceil \cdot B$ is sufficient for all K connections in their working states. When there is a fault, the connections follow the surviving bandwidth schedule S , which has at most $\lceil \sum_k \phi_k \rceil = \lceil \sum_k \max\{\alpha_k, \tilde{\rho}_k\} \rceil$ connections in the working state. Thus, it is sufficient that each link have bandwidth $K \cdot b + \lceil \sum_k \max\{\alpha_k, \tilde{\rho}_k\} \rceil \cdot (B - b)$.

APPENDIX C PROOF OF PROPOSITION 4

Note that it is sufficient to prove SLA2, SLA3, and SLA4 are true for the given link bandwidth. As in the Proof of Proposition 2, a *surviving bandwidth schedule* S for the connections

will be presented that assumes that during the time interval $[0, \Sigma F]$, there is one down link and one surviving link. The schedule determines when the connections are in their working and protection states over the interval.

Surviving Bandwidth Schedule S : Partition the connections into $\lceil \sum_k \alpha_k \rceil$ groups, where each group G satisfies $\sum_{i \in G} \alpha_i = 1$ except possibly the last group, that satisfies $\sum_{i \in G} \alpha_i \leq 1$. This is possible because the rates α_k have values that are powers of h^{-1} .

The connections in a group G take turns being in the working state by using the following schedule. Time is slotted from time 0 with time slots of duration μ . Connections are assigned time slots, during which they are in their working states. The connections are assigned time slots as follows. For connection k , let $m(k)$ be referred to as its ‘‘level.’’ We will use ‘‘tree scheduling’’ from [15]. We create a tree, where it is initially just a single root node at level 0. If there is a connection at level 0, then assign it to this node. If not, then create h children nodes from it and define them to be at level 1. Any connections at level 1 are assigned to distinct nodes at level 1. For each node at level 1 that does not have a connection, create h children nodes from it and define them to be at level 2. Any connections at level 2 are assigned to distinct nodes at level 2, and so on. This process of creating children nodes at different levels and assigning connections to them at the same level continues until all connections are assigned to some tree node. At this point, any leaf nodes without a connection are assigned ‘‘dummy’’ connections. In [15], the tree defines a schedule for the connections to the time slots, where exactly one connection (real or dummy) is assigned to any slot, and each connection k has a periodic schedule with period equal to $h^{m(k)}$. Then, a connection k has one slot per $h^{m(k)}$ time slots. Thus, its working state durations are μ and protection state durations are $\delta_k = \mu \cdot (h^{m(k)} - 1)$.

Note that for this schedule S , the number of connections in their working states is at most the number of groups $\lceil \sum_k \alpha_k \rceil$.

We can use Lemma B.1 from Appendix B. Then, the following lemma implies that the connections satisfy SLA2, SLA3, and SLA4.

Lemma C.1: Suppose the connections follow schedule S over the time interval $[0, \Sigma F]$. Then, for the connections k : (a) the maximum accumulated protection state time is D_k ; (b) the minimum working state duration is μ ; (c) the maximum protection state duration is $\delta_k = \mu \cdot (h^{m(k)} - 1)$; and (d) the short-term availability rate is $\alpha_k = h^{-m(k)}$.

Proof: To prove parts (b) and (c), note that schedule S has connection k in the working state for durations of μ and in the protection state for durations of δ_k . To prove part (d), note that parts (b) and (c) and Lemma 1 imply that for any interval $[x, y]$, the connection is in the working state for at least $\alpha_k \cdot (y - x - \delta_k)$ amount of time. Therefore, it has a short-term availability rate of α_k , and so part (d) is true.

Next, note that from (c) and (d), the amount of time that connection k is in the working state is at least $\alpha_k \cdot (\Sigma F - 0 - \delta_k) = \alpha_k \cdot (\Sigma F - \delta_k) = \alpha_k \cdot \Sigma F - \alpha_k \cdot \delta_k = \alpha_k \cdot \Sigma F - h^{-m(k)} \cdot \mu \cdot (h^{m(k)} - 1) = \alpha_k \cdot \Sigma F - \mu \cdot (1 - \alpha_k) = \Sigma F - (1 - \alpha_k) \cdot (\Sigma F + \mu) \geq \Sigma F - D_k$, where the last inequality is due to the assumption $D_k \geq (1 - \alpha_k) \cdot (\Sigma F + \mu)$ in the proposition. Thus,

the connection is in the protection state for time at most time D_k , and part (a) is true. **QED**

To complete the proof, it will be shown that the link bandwidth of the proposition is sufficient. When there are no faults, the K connections have working bandwidth B , so link bandwidth of $\lceil K/2 \rceil \cdot B$ is sufficient. When there is a fault, the connections follow the surviving bandwidth schedule, which has at most $\lceil \sum_k \alpha_k \rceil$ connections in their working states. Therefore, they require link bandwidth of $Kb + \lceil \sum_k \alpha_k \rceil \cdot (B - b)$. Thus, the link bandwidth of the proposition is sufficient.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their insightful comments.

REFERENCES

- [1] O. Gerstel and G. Sasaki, "Quality of protection (QoP): A quantitative unifying paradigm to protection service grades," *Opt. Netw. Mag.*, vol. 3, no. 3, pp. 40–49, May/June 2002.
- [2] J. Fang, M. Sivakumar, A. Somani, and K. Sivalingam, "On partial protection on groomed optical WDM networks," in *Proc. DSN*, Yokohama, 2005, pp. 228–237.
- [3] O. Gerstel and G. Sasaki, "A new protection paradigm for digital video distribution networks," in *Proc. ICC*, Istanbul, Turkey, Jun. 2006, vol. 6, pp. 2518–2523.
- [4] O. Gerstel and G. Sasaki, "Meeting SLAs by design: A protection scheme with memory," in *Proc. OFC*, Anaheim, CA, 2007, pp. 1–3.
- [5] N. Golmie, T. D. Ndousse, and D. H. Su, "A differentiated optical service for WDM networks," *IEEE Commun. Mag.*, vol. 38, no. 2, pp. 68–73, Feb. 2000.
- [6] N. Bambos, S. Gitzenis, A. Miura, O. Gerstel, and L. Paraschis, "A service risk-management approach to capacity protection in optical networks," in *Proc. IEEE LANMAN*, 2004, pp. 69–74.
- [7] R. Banner and A. Orda, "The power of tuning: A novel approach for the efficient design of survivable networks," *IEEE/ACM Trans. Netw.*, vol. 15, no. 4, pp. 737–748, Aug. 2007.
- [8] L. Song and B. Mukherjee, "Accumulated-downtime-aware restoration approach approach for dynamic SLA-differentiated services in survivable mesh networks," in *Proc. OFC*, San Diego, CA, 2008, pp. 1–3.
- [9] L. Zhou and W. Grover, "A theory for setting the "safety margin" on availability guarantees in an SLA," in *Proc. 5th DRCN*, Ischia (Naples), Italy, Oct. 16–19, 2005.
- [10] R. L. Cruz, "Quality of service guarantees in virtual circuit switched networks," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 6, pp. 1048–1056, Aug. 1995, Special Issue on "Advances in the Fundamentals of Networking".
- [11] T. Chiang and D. Anastassiou, "Hierarchical coding of digital television," *IEEE Commun. Mag.*, vol. 32, no. 5, pp. 38–45, May 1994.
- [12] R. Rejaie, M. Handley, and D. Estrin, "Layered quality adaption for internet video streaming," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 12, pp. 2530–2543, Dec. 2000.
- [13] J. Blanquer and B. Ozden, "Fair queueing for aggregated multiple links," in *Proc. ACM SIGCOMM*, 2001, pp. 189–197.
- [14] A. K. Parekh and R. G. Gallager, "A generalized processor sharing approach to flow control in integrated services networks—The single node case," *IEEE/ACM Trans. Netw.*, pp. 344–357, Jun. 1993.
- [15] A. Bar-Noy, V. Dreizen, and B. Patt-Shamir, "Efficient periodic scheduling by trees," in *Proc. IEEE INFOCOM*, New York, Jun. 2002, vol. 2, pp. 791–800.



Ori Gerstel (F'08) received the Ph.D. degree in computer science from the Technion, Haifa, Israel, in 1995.

He is a Senior Technical Leader in the Core Routing Business Unit at Cisco, Natanya, Israel, and is responsible for the architecture of IP over Transport. Prior to this role, he was in charge of Cisco's Optical Advanced Technology team and was the key inventor behind advanced capabilities of Cisco's DWDM product. Before joining Cisco in 2002, he was a Senior Systems Architect for Nortel Networks', Santa Clara, CA, MEMS-based photonic crossconnect product. Until 2000, he was the Systems and Software Architect for the Optical Networking Group at Tellabs, Hawthorne, NY, where he designed the first commercial mesh DWDM system (TITAN 7100). Prior to that, he performed early optical networking research at IBM Research, Hawthorne, NY. He authored over 50 papers in international conferences and journals and over 20 patents on optical networks.

Dr. Gerstel served on conference committees such as OFC and IEEE INFOCOM and has been the Technical Co-Chair of Broadnets and IPoP. He also serves as an Editor for several international journals such as the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS and has been teaching short courses at OFC.



Galen Sasaki received the B.S. degree in electrical engineering from the University of Hawaii, Honolulu, in 1981, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Illinois, Urbana-Champaign, in 1984 and 1987, respectively.

He has been an Associate Professor with the Department of Electrical Engineering, University of Hawaii, since 1992. He was an Assistant Professor with the Department of Electrical and Computer Engineering, University of Texas at Austin, from 1987 to 1992. He held visiting positions at the Naval Research Laboratory in Maryland; IBM Research, Hawthorne, NY; Tellabs, Hawthorne, NY; Xros, Sunnyvale, CA; and Nortel Networks, Santa Clara, CA. His research interests are in communication networks, optimization algorithms, and performance evaluation.