Reprint to appear in Proceedings of OptiComm 2002, Boston, MA, USA, July 29 – Aug. 2, 2002.

Mesh-restorable networks with complete dual failure restorability and with selectively enhanced dual-failure restorability properties

Matthieu Clouqueur, Wayne D. Grover

TR*Labs*, #800 10611-98th Avenue, Edmonton, Alberta, Canada T5K 2P7
and Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Alberta
Contact: {clouqueur, grover}@trlabs.ca

ABSTRACT

We consider extensions of the most common mesh-restorable network capacity design formulation that enhance the dual-failure restorability of the designs. A significant finding is that while design for complete dual-failure restorability can require triple the spare capacity, dual failure restorability can be provided for a fairly large set of *priority* paths with little or no more spare capacity than required for single-failure restorability. As a reference case we first study the capacity needs under complete dual-failure restorability. This shows extremely high capacity penalties to support 100% dual-failure restorability. A second design model allows a user to specify a total capacity budget limit and obtain the highest average dual-failure restorability possible for that investment limit. This formulation can also be used to trace-out the capacity-versus-availability trade-off curve for a mesh network. A third design strategy supports multiple-restorability service class definitions at minimum total cost. Restorability can range from best-efforts-only on any failure to an assurance of complete single *and dual*-failure restorability, on a per-demand basis. Prior work has only considered multiple service classes ranging from assured single-failure restorability and downwards in quality. This work shows how to economically support an added service class in the upward quality direction: assured dual failure survivability. This lets a network operator tailor the investment in capacity to provide ultra-high availability on a selective basis, while avoiding the very high investment required for complete dual-failure restorability for all.

Keywords: optical networks, transport network design, restoration and protection, dual-failure restorability, capacity planning, restorability, availability, network optimization

1 INTRODUCTION

Dual-failure situations, or the equivalent, are not as rare as might be thought in an extensive national or regional optical network. Certainly the bulk of transport requirements will be well served by designing for them to withstand any one failure at a time. But we have learned from industry colleagues that there is a significant number of other applications and customers for whom there really is a desire to support dual-failure restorability. Considering rates of fibre cuts in some networks, span maintenance operations which create situations similar to dual-failures [1], and shared risk link groups, we are motivated to study and understand issues and phenomena surrounding dual failure restorability in mesh networks and to see what steps can be taken to enhance or even design for certain specific abilities to withstand dual failure scenarios.

1.1 Prior Work and Current Goals

The present work was preceded by a related study [2] that analyzed mesh-restorable networks from a service path availability standpoint considering all possible dual-failures as the primary contributor to unavailability. The aim in [2] (and a more preliminary effort in [3]) was to develop a means of determining the availability of service paths through mesh-restorable networks that are designed for full restorability to single failures and may involve an active, state-adaptive, recovery mechanism, but are then presented with dual-failure situations. The prior work ([2][3]) was thus essentially one of *analysis* whereas the present effort is to be one of design *synthesis* to specific new goals. One of the most striking findings of [2][3] was, however, that a span-restorable mesh network that is designed for full restorability to any single span failure is rather naturally able to protect a high average proportion of working capacity against *dual* span-failures. This was found to be especially true when the restoration mechanism is adaptive to changes in the spare capacity layer following any prior failure when the second failure arises. This basic finding gives considerable motivation to the idea of trying to specifically harness or otherwise exploit this basic property by design.

However, an important consideration about the high average-case dual-failure restorability (denoted R_2) found in [2] and [3] is that while this is of obvious benefit to the *overall* network availability as a whole, there is no way at present to target or otherwise structure the dual-failure restorability so that specific paths, such as those of premium customers, would be the ones assured to benefit. Therefore, we are motivated to set the next two goals:

(1) *How can we take advantage of the naturally high restorability of mesh networks to dual failures to specifically serve a new category of extreme high availability service paths that would enjoy full dual span-failure restorability?* And,

(2) *How can we specifically design the network capacity to serve any given mixture of single failure and dual failure restorability requirements, designated on a path-by-path basis, at minimum total cost?*

The present study combines concepts and observations from [2] and [3] with Integer Linear Programming (ILP) techniques to address these questions and to use the resulting capability to observe the overall trade-off between availability and network capacity. It also provides new design methods for support of restorability-related service level agreements.

There has been prior work both on methods to design for 100% dual failure restorability, and, separately, on providing multiple classes of service, including multiple classes of restoration or protection, referred to as "multi QoP." As early as 1992,. Sakauchi et al [4] gave a formulation for spare capacity design of a span-restorable mesh network to withstand all possible dual span failure scenarios, although they did not produce results with the model. As far as we know the first quantitative results for how much extra spare capacity would be required for complete dual failure restorability was provided in [3]. The present paper includes a confirmation of the basic finding in [3] that design for complete dual-failure restorability would as much as triple the amount of spare capacity needed (relative to the single failure restorable design case). In this regard what is novel in the present work is that we go on to consider dual-failure restorability design, not for the network as a whole, but specifically for only a subset of premium-service paths and we also introduce a way to maximize dual failure restorability given a specific budget limit. We think this moves the study of dual-failure restorability design from the rather impractical and limited scope of simply designing the whole network to dual-failure restorability, with an enormous capacity penalty. In contrast, applications of the design models given here could actually be considered in practise either to support a selective subset of high priority paths with dual-failure restorability, or, to assign a limited budget allowance for generalized enhancement of the networks dual-failure restorability. The latter includes the option to set the budget for extra capacity at zero, in which case the method provided effects a redistributing optimization of the spare capacity needed for withstanding single failures to enhance the dual-failure restorability as much as possible.

The other sense in which this work is original is regarding the "multi QoP" aspects of prior work. What distinguishes the present work from prior considerations of multi QoP in both industry standards forums and in the research literature is that we enable a *new* QoP class which stands above the previously considered hierarchies of gold, silver, bronze, etc. For instance gold usually means assured single-failure restorability, silver is best efforts, bronze is non-protected, and there may be another lower class of pre-emptible economy class services as well. But here we find an economical way to design-in support for what could be called a platinum service class that stacks *above* the existing QoP paradigms, in the sense illustrated in Figure 1. The particular contribution made by the end of the paper is thus of showing how a new super-high availability service class can be added to the range of possible QoP classes and that it can be supported with remarkably little additional cost within span restorable networks that are already efficiently designed only to support single failure restorability, i.e., the typical gold QoP class. [5][6]

1.2 Outline

The rest of the Introduction section covers the main method for basic capacity design of span-restorable networks leading up to how we define and measure the dual-failure restorability of such networks. Section 2 then proposes three new design formulations having to do with dual-failure restorability design considerations. It is through these new design models that we can study, and also manipulate, the capacity-availability trade-off within a span-restorable mesh network. Section 3 discusses experimental results obtained with these formulations. Section 4 summarizes and discusses the impact and significance of these advances in design methods and the network strategies they suggest.

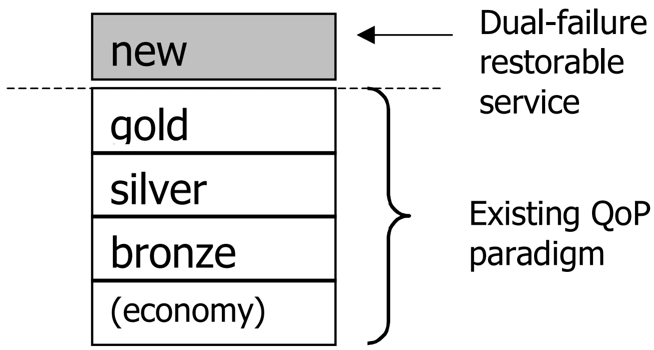

Figure 1. Quality of Protection Classes

1.3 Common approach to mesh network capacity design

The basic capacity design approach and optimization model introduced by Herzberg and Bye [7] is covered in this section and serves as the 'baseline' for the following work. First, we will look at the basic formulation for the spare capacity assignment (SCA). In this approach demands are first routed on shortest-path. This results in an accumulation of working demands crossing each span, called the w_i values, where i denotes the span number. The problem is then posed as minimizing the total cost of a spare capacity allocation s_i to all spans so that the required number of restoration paths are feasible through surviving spare capacity for each single span cut. The problem, formulated as an LP, is in fact that of finding an assignment of restoration flows to the eligible routes on which restoration for each failure could be considered so as to minimize the total amount of required spare capacity. These so-called 'distinct eligible routes' are obtained before running the formulation. The graph topology is first processed to find all the distinct logical routes that are "eligible" for use in the restoration routing for each failure scenario. More details on this follow in section 1.4. The problem as described, (where the working demands are first routed and then the spare capacity optimized,) is called the *non-joint* mesh capacity design problem. In a *joint* formulation, which will be an aspect of this work, the routing of demands is simultaneously optimized with the placement of spare capacity so as to minimize *total* capacity.

For this and subsequent formulations, we use the following notation:

Parameters (inputs):

C_j	Cost of each unit of capacity on span j
$\boldsymbol{S}$	Set of spans in the network
$\boldsymbol{P}_i$	Set of eligible routes for restoration of span i
w_j	Number of working links (capacity units) on span j
$\delta_{i,j}^{p}$	Equal to one if p^{th} eligible route for span i uses span j, zero otherwise

Variables:

f_i^p	Restoration flow assigned to p^{th} route for span i
s_j	Number of spare capacity units placed on span j

SCA (Herzberg):

$$\text{Minimize} \quad \sum_{j \in \boldsymbol{S}} C_j \cdot s_j \qquad (1)$$

Subject to:

$$\sum_{p \in \boldsymbol{P}_i} f_i^{\,p} = w_i \qquad \forall i \in \boldsymbol{S} \qquad (2)$$

$$s_j \geq \sum_{p \in \boldsymbol{P}_i} \delta_{i,j}^{p} \cdot f_i^{\,p} \qquad \forall (i,j) \in \boldsymbol{S}^2, i \neq j \tag{3}$$

The constraint set in (2) ensures that restoration for span failure *i* meets the target level (100% restoration will be assumed). The set of constraints in (3) forces sufficient spare capacity on each span *j* such that the sum of the restoration paths routed over span *j* is met, for every failure span *i*. The largest simultaneously imposed set of restoration paths on a span effectively sets the minimum feasible s_i value on each span under a given assignment of restoration flows. Hence the formulation works by finding the assignment of flows that pushes up these span-wise minimums to a global minimum on total spare capacity. For the joint working and spare capacity assignment problem (JCA) the model changes as follows:

Additional Parameters:

$\boldsymbol{D}$	Set of O-D pairs with non-zero demand
d^r	Number of demand units for O-D pair *r*
Q^r	Set of eligible working routes available for demand pair *r*
$\zeta_j^{r,q}$	Equal to 1 if q^{th} eligible route for working demands between O-D pair *r* uses span *j*

Variables:

s_j	Number of spare capacity units placed on span *j*
$g^{r,q}$	Working capacity assigned to the q^{th} eligible working route for demand pair *r*
w_j	Number of working capacity units on span *j*

JCA:

$$\text{Minimize} \quad \sum_{j \in \boldsymbol{S}} C_j \left(w_j + s_j \right) \tag{4}$$

Subject to (additional constraints):

$$\sum_{q \in \boldsymbol{Q}^r} g^{r,q} = d^r \qquad \forall r \in \boldsymbol{D} \tag{5}$$

$$\sum_{r \in \boldsymbol{D}} \sum_{q \in \boldsymbol{Q}^r} \zeta_j^{r,q} \cdot g^{r,q} = w_j \qquad \forall j \in \boldsymbol{S} \tag{6}$$

The new constraint set (5) ensures that all working demands are routed. Constraint set (6) generates the working capacity required on each span j to satisfy the sum of all pre-failure demands routed over it.

1.4 Concept of "Eligible Routes"

A technical aspect of implementing either of the joint or non-joint design models above is creating the eligible route sets to populate the formulations. By eligible routes we are referring back to the "arc-path" nature of the basic Herzberg approach in [7]. That is to say, of posing the problem as assignment of restoration flows to "eligible" distinct routes over the network graph. In practice this approach is desirable so that restoration route properties can be under engineering control for length, loss, or any other eligibility criteria. In this work we consider both jointly and non-jointly optimized designs. In the joint case, we need two families of eligible routes: one to represent the routings possible for restoration of each failure, and another to represent the ways that each working demand might be routed so as to provide choices for joint optimization with the restoration strategy. Different considerations apply to the two types of eligible route representations.

For restoration, the idea is to represent all distinct routes between the end nodes of each span failure, excluding the failed span, up to the "threshold" hop limit, *H*. A practical problem comes when the network contains long chains or otherwise requires a high hop limit to reach the limiting efficiency in sharing of spare capacity. With *H* much over 6 or 7 the number of distinct eligible routes can quickly be above memory limits in medium or large size networks. For the test cases that follow it has been possible to represent all distinct routes up to the threshold hop limits for *restoration* routes.

For the representation of eligible *working* routes in the joint formulation we use the following strategy. The idea is not to presume a specific hop limit and attempt to generate all distinct working routes up to the limit. Rather, we set a minimum number of distinct route options that we want to have available for the working path routing of any demand pair. A route-finding program that varies the hop limit adaptively for each O-D pair so that *at least* the target number of eligible routes are found. At this target hop limit, all distinct routes at or below that number of hops are included.

1.5 Dual-failure restorability

In the network design methods just surveyed, the common aim is to minimize capacity requirements with an assurance of a target level of restorability on each *single* span failure. Reasonably enough, a side-effect of designing for this goal can be that certain amounts of working capacity may also survive certain dual failure scenarios. This is a characteristic that can be measured through the combination of computer-experimental re-routing trials and analysis as described in [2]. The "dual span-failure restorability" $R_2(i,j)$ of a given pair (i,j) of span failures is defined as the fraction of the total failed working capacity of spans i and j that can be restored in the case of the overlapping failure states on spans i and j. In other words if $N(i,j)$ is the number of non restorable working links in the case of failure of spans i and j bearing w_i and w_j working links respectively then:

$$R_2(i,j) = 1 - \frac{N(i,j)}{w_i + w_j} \tag{7}$$

The network wide average of $R_2(i,j)$ over all ordered (i,j) dual-failure combinations is also defined and denoted simply R_2.

Although our ongoing and direct focus in this paper is on design to give certain assurances in terms of R_2 it is important to appreciate how R_2 relates to the end-to-end availability of service paths. This involves the concept of equivalent link unavailability. If we first imagine a mesh network over which a path d is provisioned over n spans, but with no restoration mechanism, we would fairly accurately estimate that path's availability to be:

$$A_{path}(d) \simeq 1 - \sum_{i=1}^{n} U_{\text{link}}^{p}(i) \tag{8}$$

where $U_{\text{link}}^{p}(i)$ is the *physical unavailability* of the i^{th} link in the path, in other words the probability of finding link i in the failed and non-restored state at any point in time. Therefore one way of thinking about the action of an active span restoration mechanism is that it is a transformer of *physical* span unavailability to a lower *equivalent* unavailability of links on the span. This viewpoint argues that from the standpoint of an end-to-end path, there are two equally acceptable ways in which a link along the path can be in "up" state: either it is physically working, or, it is physically "down" but transparently replaced by a restoration path between its end nodes. Thus if we define the *equivalent unavailability* of a link $U_{\text{link}}^{*}(i) = p(\text{link } i \text{ down } \cap \text{ link } i \text{ not restored})$, then the path availability has the same form as Eq. (8) but is based on U_{link}^{*} not U_{link}^{p}. This line of reasoning reduces the problem of calculating path availability to determining the *equivalent unavailability* of links U_{link}^{*} in a span-restorable network based on the capacity in the network and the particulars of the restoration mechanism. Further recognizing that single span failures do not really contribute to unavailability of links in a network designed for 100% restorability to single failures, and if we consider the contribution of triple span failures as negligible compared to dual span-failures, it can be shown that under standard assumptions of independent failures, the equivalent link unavailability can be expressed as follows:

$$U_{\text{link}}^{*} = U_s^2 \left(S-1\right)\left(1-R_2\right) \tag{9}$$

where U_s represents the physical span unavailability and S the number of spans in the network. Thus, certainly as we would expect intuitively, the dual span-failure restorability R_2, has a direct influence on end-to-end service availability as shown by Equation (9). In a slight refinement from Equations (7) and (9), which together make only an average-case statement about a path of a certain length through the network, we can think of specific path over specific links and using specific channels which only have either $R_2(i,j) = 1$ or $R_2(i,j) = 0$ for any other particular (i,j) failure combination. In the latter orientation the overall R_2 on a network need not manifest itself only as an aggregate average integrity measure of the network as a whole, but we can instead take steps to specifically structure the capacity investment and assign the individual channels on links for which $R_2(i,j) = 1$ coherently to premium class paths end-to-end. We now begin the series of steps that will lead us to the ability to design a minimum-cost restorable network in which a specific subset of end-to-end paths enjoy $R_2 = 1$ on each link they traverse.

2 OPTIMIZING THE CAPACITY DESIGN FOR HIGH AVAILABILITY

Let us now develop three specific optimization models through which we can explore and understand the trade-offs and opportunities to design for dual failures and to specifically structure the dual failure restorability so that it is targeted onto desired paths. As mentioned, previous work [2] showed that surprisingly high levels of dual-failure restorability tend to arise simply as a side effect of placing spare capacity to assure the restorability of single failures, a condition we can denote as R_1=1. In other words R_1 tends to beget high average R_2 without further special effort of any type. This prompts a natural first exercise which is to simply see how much it would cost to directly design for restorability to all dual-failures. This we will refer to as the problem of dual-failure minimum capacity design (DFMC). The results of DFMC are somewhat surprising: although the average $R_2(i,j)$ may be relatively high, it turns out that achieving strictly $R_2(i,j)$=1 for all (i,j) is very expensive. The next question is therefore to see just how high an R_2 can be achieved for a given set limit on additional capacity expenditure over the R_1=1 condition. This is called the dual-failure maximum restorability (DFMR) problem. These preliminary investigations lead to the final goal of deliberately placing capacity at minimum cost to serve certain paths at R_2 =1 and others at only R_1=1. This is called multi-restorability capacity placement design (MRCP). To summarize we will now consider each of the following problems:

(i) <u>Dual-failure minimum capacity design (DFMC)</u>: What is the minimum spare capacity assignment that guarantees R_2 =100%?

(ii) <u>Dual-failure maximum restorability (DFMR)</u>: Given a finite budget of total spare capacity, find the spare capacity assignment that maximizes R_2 in the network, given the available spare budget.

(iii) <u>Multi-restorability capacity placement design (MRCP)</u>: Find the minimum total (spare plus working) capacity assignment that satisfies a mixture of end-to-end path restorability objectives (R_0, R_1 or R_2) for each demand.

In the first two problems DFMC, DFMR, the working demands are first shortest-path routed, generating the w_i values which are inputs to the problem. MRCP is a type of joint optimization problem.

2.1 Dual Failure Minimum Capacity (DFMC)

This DFMC formulation finds a minimum total spare capacity assignment that guarantees full restorability of all dual span-failure scenarios. This formulation will tell us what the minimum "price" is for reaching R_2 = 1 restorability. For obvious reasons, a feasible solution to this problem cannot be found for a network with degree-2 nodes or 2-edge cuts of the network graph [2]. Therefore our tests of this formulation are limited to those having a graph topology that qualifies. In practice few transport networks have cuts of only 2 edges not involving degree-2 nodes but many have degree-2 nodes. In practice DFMC can therefore be considered a formulation primarily for the assurance of R_2 = 100% on the "mesh backbone" component of the overall transport network. In this context the formulation can easily be applied by logical removal of a degree-2 node between spans i, j and assertion that $w_i' = \max(w_i, w_j)$ where w_i' applies to the single logical edge arising from the degree-2 node removal.

There are no new parameters or variables to introduce at this stage, other than:

$f_{i,j}^p$ Restoration flow assigned to p^{th} restoration route of span i when span j has failed simultaneously (integer)

DFMC:

$$\text{Minimize} \quad \sum_{j \in S} C_j \cdot s_j \tag{10}$$

Subject to:

$$\sum_{p \in \boldsymbol{P}_i} f_{i,j}^p = w_i \qquad \forall (i,j) \in \boldsymbol{S}^2, i \neq j \tag{11}$$

$$f_{i,j}^p \leq C_\infty \left(1 - \delta_{i,j}^p\right) \qquad \forall (i,j) \in \boldsymbol{S}^2, i \neq j, \forall p \in \boldsymbol{P}_i \tag{12}$$

$$\sum_{p \in \boldsymbol{P}_i} f_{i,j}^p \cdot \delta_{i,k}^p + \sum_{p \in \boldsymbol{P}_j} f_{j,i}^p \cdot \delta_{j,k}^p \leq s_k \qquad \forall (i,j,k) \in \boldsymbol{S}^3,\ i \neq j,\ i \neq k,\ j \neq k \tag{13}$$

The constraint set (11) ensures full restoration in all dual span-failures. Constraints (12) use an arbitrarily high "capacity constant", C_∞, to ensure that span j can support restoration flow when it is not part of the failure scenario, but that it will not be used for restoration of span i in the (i,j) scenario. Constraints (13) ensure that there is enough spare capacity on each span of the network to support all the restoration flows assigned for the restoration of any dual span failure. No explicit statement of the single-failure restorability requirement is needed. This is implicitly provided for in specifying the restorability of all *dual* failures.

2.2 Dual Failure Maximum Restorability (DFMR)

The second formulation is for the converse problem of maximizing the achievable R_2 level with a *given* total spare capacity investment. The same parameters and variables apply as so far defined with the addition of:

New Parameter: B, total budget available for spare capacity

New Variable: $N(i,j)$, number of non-restored working units under dual failure of spans (i, j) (integer)

DFMR:

$$\text{Minimize} \sum_{(i,j) \in \boldsymbol{S}^2, i \neq j} N(i,j) \tag{14}$$

Subject to:

$$\sum_{p \in \boldsymbol{P}_i} f_i^p = w_i \qquad \forall i \in \boldsymbol{S} \tag{15}$$

$$N(i,j) = w_i + w_j - \left(\sum_{p \in \boldsymbol{P}_i} f_{i,j}^p + \sum_{p \in \boldsymbol{P}_j} f_{j,i}^p \right) \qquad \forall (i,j) \in \boldsymbol{S}^2, i \neq j \tag{16}$$

$$\sum_{p \in \mathbf{P_i}} f_{i,j}^p \leq w_i \qquad \forall (i,j) \in \boldsymbol{S}^2, i \neq j \tag{17}$$

$$f_{i,j}^p \leq C_\infty \left(1 - \delta_{i,j}^p\right) \qquad \forall (i,j) \in \boldsymbol{S}^2, i \neq j, \forall p \in \boldsymbol{P}_i \tag{18}$$

$$\sum_{p \in \boldsymbol{P}_i} f_i^p \cdot \delta_{i,k}^p \leq s_k \qquad \forall (i,k) \in \boldsymbol{S}^2, i \neq k \tag{19}$$

$$\sum_{p \in \boldsymbol{P}_i} f_{i,j}^p \cdot \delta_{i,k}^p + \sum_{p \in \boldsymbol{P}_j} f_{j,i}^p \cdot \delta_{j,k}^p \leq s_k \qquad \forall (i,j,k) \in \boldsymbol{S}^3,\ i \neq j,\ i \neq k,\ j \neq k \tag{20}$$

$$\sum_{k \in \boldsymbol{S}} C_k \cdot s_k \leq B \tag{21}$$

Constraints (15) ensure that $R_1 = 1$ for every single span-failure case. Eq. (16) defines $N(i,j)$, which is the number of working capacity units that are *not* restored in case of a dual failure on spans i, j. By minimizing the sum of $N(i,j)$ over all dual failure scenarios, one is maximizing R_2 (through Equation (7) and thereby also maximizing the availability through Equation (9)). Constraints (17) ensure that the number of paths assigned to the restoration of a span in a dual span-failure scenario is at most equal to the number of working links to be restored. (18) is identical to (12), forcing the exclusion of restoration routes for span *i* from using span *j* and vice-versa during the (*i*,*j*) scenario, while allowing use of span *j* for all other scenarios. Eq. (19) ensures adequate spare capacity for every single failure case. (20) is identical form to (13) but here it serves to assure adequate spare capacity only for the dual failure scenarios that the budget-limited formulation *chooses to cover*. The aspect of selectivity, not present in DFMC, is effected through (17) which, (in contrast to (15)), is not required to provide fully adequate restoration flows for all dual-failure scenarios. This is also why an explicit assurance of single failure restorability is present (15) whereas it is not in DFMC. Constraint (21) imposes the budget limit on the total cost of spare capacity.

2.3 Multi-service Restorability Capacity Placement Design (MRCP)

The last model is the most computationally challenging but perhaps also the most practically interesting and directly useful in the business of network operators. The idea is to explicitly design the allocation of spare capacity and the routing of demands to support *a multi-service restorability classification of the demands served*. The previous formulation maximizes the network R_2 as an average over all spans of the network. All demands served in the network will share in the increased availability in a general way. The MRCP formulation will, however, allow us to target and structure the high availability investment *specifically to the intended services or customers*. Every demand will receive a specific class of restorability guarantee on every span end-to-end over its route. In the first two formulations a given demand might cross a mixture of spans with various R_2 levels, whereas here we will be able to stipulate that a priority service will exclusively travel over spans where its restorability to dual span failures is guaranteed, thus deriving an overall R_2 = 1 path guarantee for the service class. More generally, to each demand unit we assign one of the following restorability service class designations:

(i) 'R_0 restorability': no assured restorability – best efforts in both single and dual failures

(ii) 'R_1 restorability': assured restorability to any single span-failure, best efforts for dual failures

(iii) 'R_2 restorability': assured restorability to any single or dual span-failure.

The formulation finds the minimum total cost of capacity (working plus spare) and the routing of each working demand so as to satisfy the restorability class-of-service of each demand. For MRCP a demand *group* is now defined as one or more demand units on the same O-D pair and in the same service class. Demand groups must be defined now to distinguish between demands of different service classes on each O-D pair. The sum of all demand groups on an O-D pair here equals the prior single-service demand quantities on each O-D pair. *r* now indexes not just over all O-D pairs but over all demand groups.

New Parameters:

$\boldsymbol{D}$	Set of demand groups, index *r*
$\boldsymbol{Q}^r$	Set of eligible working routes for demand group *r*
d^r	Size of the r^{th} demand group (number of individual demand units)
ψ_1^r	Equal to 1 if r^{th} demand group is in the R_1 or R_2 service class, 0 otherwise
ψ_2^r	Equal to 1 if r^{th} demand group is in the R_2 service class, 0 otherwise

MRCP:

$$\text{Minimize} \quad \sum_{k\in \boldsymbol{S}} C_k \left(w_k + s_k \right) \qquad (22)$$

Subject to:

$$\sum_{q\in \boldsymbol{Q}^r} g^{r,q} = d^r \qquad \forall r \in \boldsymbol{D} \qquad (23)$$

$$\sum_{r\in \boldsymbol{D}} \sum_{q\in \boldsymbol{Q}^r} \zeta_i^{r,q} \cdot g^{r,q} = w_i \qquad \forall i \in \boldsymbol{S} \tag{24}$$

$$\sum_{p\in \boldsymbol{P}_i} f_i^{\,p} = \sum_{r\in \boldsymbol{D}} \sum_{q\in \boldsymbol{Q}^r} \zeta_i^{r,q} \cdot \psi_1^r \cdot g^{r,q} \qquad \forall i \in \boldsymbol{S} \tag{25}$$

$$\sum_{p\in \boldsymbol{P}_i} f_{i,j}^{\,p} = \sum_{r\in \boldsymbol{D}} \sum_{q\in \boldsymbol{Q}^r} \zeta_i^{r,q} \cdot \psi_2^r \cdot g^{r,q} \qquad \forall (i,j) \in \boldsymbol{S}^2, i \neq j \tag{26}$$

$$f_{i,j}^{\,p} \le C_\infty \left(1 - \delta_{i,j}^p\right) \qquad \forall (i,j) \in \boldsymbol{S}^2, i \neq j, \forall p \in \boldsymbol{P}_i \tag{27}$$

$$\sum_{p\in \boldsymbol{P}_i} f_i^{\,p} \cdot \delta_{i,k}^p \le s_k \qquad \forall (i,k) \in \boldsymbol{S}^2, i \neq k \tag{28}$$

$$\sum_{p\in \boldsymbol{P}_i} f_{i,j}^{\,p} \cdot \delta_{i,k}^p + \sum_{p\in \boldsymbol{P}_j} f_{j,i}^{\,p} \cdot \delta_{j,k}^p \le s_k \qquad \forall (i,j,k) \in \boldsymbol{S}^3,\ i \neq j,\ i \neq k,\ j \neq k \tag{29}$$

Constraint set (23) ensures that each demand group is fully served. Constraint set (24) ensures that there is enough working capacity on each span to support the routing of demands. Constraint set (25) ensures that adequate restoration flow exists for each single-span failure affecting demands that require R_1 restorability. Constraint set (26) ensures that adequate restoration flow exists for each dual-failure scenario affecting demand groups that require R_2 restorability. The rest of the constraints are as previously explained.

3 EXPERIMENTAL RESULTS

The three formulations were solved on the test networks described in Table 1. Each formulation was implemented as an AMPL model and solved (with all capacity and flow variables integer) using the CPLEX 6 MIP solver on a four × 250 MHz Sun Enterprise processor running the Sun Solaris Operating System 2.6 with 892 MB of RAM. All DFMC and DFMR problems solved in less than 30 min, most of them in seconds. Good feasible solutions to the MRCP problems were typically found in a few minutes, but complete solution could take several hours [1]. All results are based on a full CPLEX termination or a MIPGAP of under 0.001 (i.e. solutions are provably within 0.1% of optimal.) In all cases problems were formulated with the complete set of distinct routes up to a hop limit of five for restoration flows. In MRCP problems the policy for routing of working demands (which is jointly solved with the spare capacity in MRCP only) was as follows: R_1 demand groups were routed based on the complete route set at the lowest hop limit that provided a minimum of three distinct routes for each demand. That is, the working path hop limit would be increased until a minimum of three routes exists, then all routes at that hop limit are represented. For R_2 demands the set of eligible routes was based on the hop limit determined for R_1 demands on the same O-D pair, plus one. This route-enumeration policy typically provided 10 to 15 eligible routes for R_1 demands and at least twice that number for R_2 demands. R_0 demand groups, when present, would be always explicitly routed over shortest paths because this is optimal when no corresponding investment in spare capacity is made for them.

[1] While we comment in general about the runtimes, all were manageable. Beyond this, tabulations of individual run times are not included, partly for space reasons, but more so because the main interest is in the networking concepts, not the computational aspects of the formulations. Heuristics and relaxations can be developed in the future for faster solving of these problems if the basic networking ideas are of sufficient value.

Table 1. Test network characteristics

Network	Nodes	Spans	Cap. Redundancy Non modular	Cap. Redundancy Modular	min cut < 3
Bellcore	11	23	55.9 %	92.1 %	3
EuroNet	19	37	50.6 %	78.3 %	5
6n14s	6	14	44.1 %	59.4 %	0
11n20s1	11	20	91.6 %	105.8 %	0
11n20s2	11	20	47.5 %	73.2 %	0
12n18s1	12	18	99.8 %	112.1 %	0
12n20s1	12	20	104.4 %	130.1 %	3
12n24s1	12	24	48.4 %	89.6 %	0
15n28s1	15	28	58.1 %	84.7 %	2
16n26s1	16	26	78.8 %	83.0 %	0
22n41s1	22	41	53.5 %	73.0 %	2

Table 2 shows results with the DFMC formulation. The column 'R_2 Redundancy' indicates the minimum spare capacity redundancy required to obtain R_2 = 1. The third column indicates the increase in total capacity required in going from R_1 to R_2 restorability. Combined with the findings of Section IV these results show that although high R_2 levels arise just from R_1 design, the price strictly to assure R_2 = 1 is very high. Achieving the "last few percents" in R_2 is extremely expensive.

Table 2. Test results for DFMC formulation

Network	R_2 redundancy	Total Cap. increase/R_1
6n14s	116.8 %	50.5 %
11n20s1	258.9 %	87.3 %
11n20s2	161.3 %	77.1 %
12n18s1	268.6 %	84.5 %
12n24s1	145.9 %	65.7 %
16n26s1	248.2 %	94.7 %

Fig. 2 shows results for the "budget-oriented" DFMR formulation on three of the test networks. Each curve shows the improvement in R_2 as the spare capacity budget is increased and its allocation optimized under DFMR. In each case there is an initially better-than-linear growth of R_2 as the spare capacity allowance is increased but this slows greatly as R_2 nears unity and merges with the very high capacity results of the limiting case of R_2 = 1 in DFMC. Note the special interpretation of "pure re-distribution" of spare capacity under the DFMR formulation when B is no more than that required for basic R_1. With no "additional" budget for spare capacity (beyond that essential for R_1 design), the DFMR formulation is performing a redistribution of the spare capacity for the single-failure restorability to enhance the achievable R_2. In results so far, this "zero-cost" redistribution benefit to R_2 is relatively small, however, under 5% in all cases.

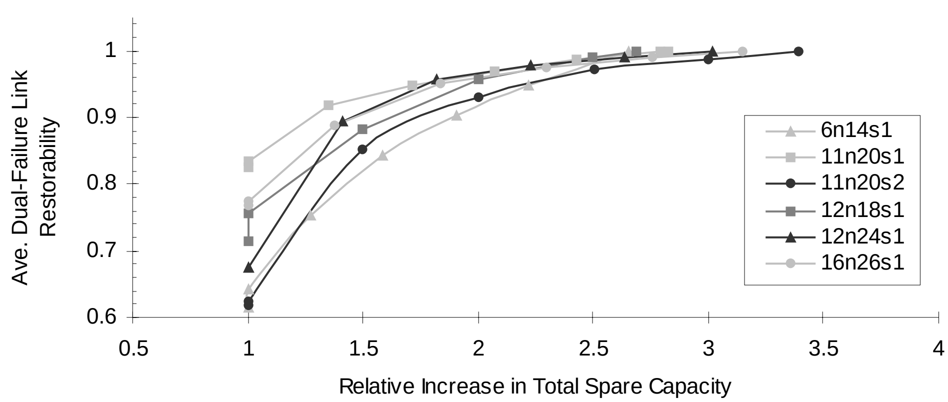

Figure 2. Achievable R_2 restorability vs. total spare capacity

Table 3 gives test results for the multi-service MRCP formulation. For these tests the basic demand matrices are identical to those in the DFMC and DFMR test and results in [2] but for MRCP we overlay a status classification of individual demands on each O-D pair with respect to the multi-service restorability categories. Although the general formulation allows R_0, R_1, R_2 categories, it was thought more meaningful and stringent for test purposes to retain R_1 as a minimum requirement for all demands in the tests here. This is conservative in terms of assessing the potential R_2-serving ability and it reduces the dimensionality of the results to be presented. It also makes results more directly comparable to the prior "conventional" spare capacity design (for R_1) (although for the "jointly optimized" case [8][9]). For tests of MRCP the fraction of demands on each 0-D pair that were given R_2 status was varied from 0 to 40%, as detailed in Table 3. Table 3 shows the total (working plus spare) capacity requirements to support the given mix of "R_1-assured" and "R_2-assured" services (R_2 class implies an R_1 assurance too). Fig. 3 shows the increase in total capacity relative to that for the single-class R_1 restorability requirements.

Table 3. Test results for MRCP formulation

Network	R_1 Total Capacity Req.	R_1 + 20% R_2 Capacity Req.	R_1 + 30% R_2 Capacity Req.	R_1 + 40% R_2 Capacity Req.
6n14s	203	203	206	211
11n20s1	944	948	969	1074
11n20s2	405	405	425	449
12n18s1	894	1030	1342	1541
12n24s1	569	570	587	611
16n26s1	1497	1572	1970	2218

The results suggest that through this formulation *a significant class of R_2 customers could be served with relatively small increases in total capacity requirement*. For instance in 11n20s1 the total capacity to serve all demands with the basic requirement of R_1 = 1 is 944 units. But under MRCP design 20% of demands *on every O-D pair* can also be given R_2-class service with a total capacity of 948 units. For a larger network, 16n26s1, 30 % of all demands could be given R_2-class service for a 1970 / 1497 = 30% increase in total capacity. Fig. 3 shows that as the fraction of R_2 class demands increases, the increase in required capacity is relatively slow up to 20 - 30% R_2 depending on the network. Overall this suggests a tendency of these networks to be capable of supporting a select class of "ultra high availability" R_2 = 1 customers without the huge capacity requirement for the network to have R_2 = 1 as a whole. Inspection of the resulting designs show that the satisfaction of the R_2 service class arises not only through re-distribution of spare capacity to target high R_2 on the key spans, but also from changes in the routing of working paths that bring them through regions of the network that are more easily made R_2 restorable. These results are probably also fairly conservative estimates of the operational potential for multi-service class design because the present MRCP formulation is non-modular and because we present results with no R_0 service class members. As seen in [2] modularity only increases the inherent R_2 level. And R_0 class services require no spare capacity so can only increase the fraction of R_2-servable customers at the same total capacity investment.

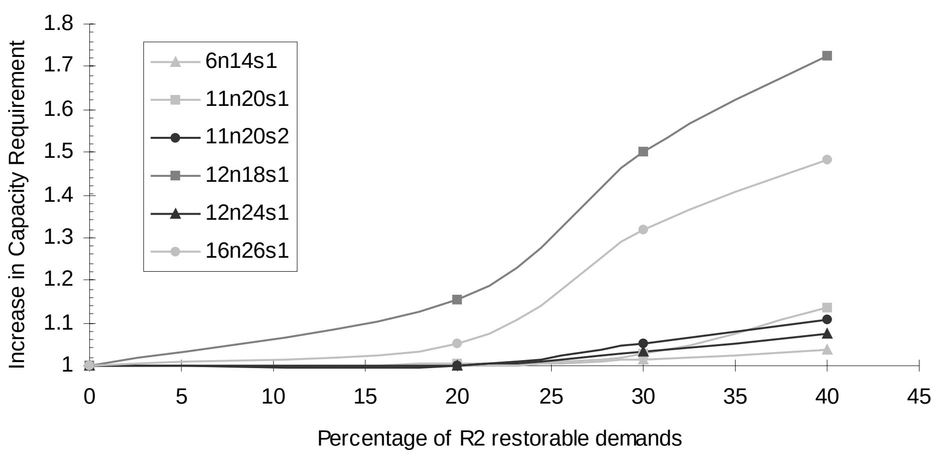

Figure 3. Plot of Results obtained with the MRCP formulation

4 CONCLUSION

This work has treated three ways in which the basic formulation for span-restorable mesh network design can be extended to design for higher availability through strategies for control of R_2. To start with we have again confirmed that designing to support complete dual-failure restorability requires inordinately high additions of spare capacity relative to the single-failure design case. We have, however, gone past that limited design paradigm to provide two much more practical options for enhancing dual-failure survivability. With the DFMR formulation one can maximize the network-wide R_2 subject to any desired budgetary allocation for this purpose. Results showed that the overall average R_2 could be significantly increased for a small increase of capacity. In addition as a check on results, the budget requirement to reach R_2 = 1 is consistent with that predicted by DFMC. The findings with the MRCP method may, however, be of the most practical importance because they directly support the desired business model of multiple restoration service classes and show how to support an ultra-high availability service class which withstands all dual failures with little or no additional capacity. The key to how MRCP does this is that the inherently high average R_2 levels of ordinary single-failure designs are in effect given internal structuring so that this property is coherently targeted on the priority paths rather than being a network-average property only. This makes it possible to guarantee extremely high availability levels to selected demands, with minimal added capacity. As such MRCP directly addresses a long-held issue for network operators: "How to earn new revenue from investing in survivability."

5 REFERENCES

[1] W. D. Grover, M. Clouqueur, T. Bach, "Quantifying and Managing the Influence of Maintenance Actions on the Survivability of Mesh-restorable networks," *Proceedings 17th Annual National Fiber Optic Engineers Conference (NFOEC 2001)*, July 8-12, Baltimore, 2001

[2] W. D. Grover and M. Clouqueur, "Availability analysis of span-restorable mesh networks," *IEEE JSAC Special Issue on Recent Advances in Fundamentals of Network Management*, vol. 20, no. 4, May 2002.

[3] M. Clouqueur, W. D. Grover, "Computational and Design Studies on the Unavailability of Mesh-restorable Networks", in Proc. *IEEE/VDE Design of Reliable Communication Networks (DRCN 2000*), Munich, Germany, April 2000, pp. 181-186.

[4] H. Sakauchi, Y.Okanoue, S. Hasegawa, "Spare-channel design schemes for self-healing networks," IEICE Trans. Comm., vol. E75-B, no.7, July 1992, pp. 624-633.

[5] O. Gerstel and G. Sasaki, "Quality of Protection (QoP): A Quantitative Unifying Paradigm to Protection Service Grades," Gigabit Networking Workshop, Anchorage Alaska, April 23, 2001

[6] Murari Sridharan, Arun K. Somani and Murti V. Salapaka, "Approaches for capacity and revenue optimization in survivable WDM networks," Journal of High Speed Networks vol. 10, no.2, pp. 109-125, Aug. 2001.

[7] M. Herzberg and S. Bye, "An optimal spare-capacity assignment model for survivable networks with hop limits," in *Proc. IEEE Globecom'94*, 1994, pp.1601-1607.

[8] J. Doucette and W. D. Grover "Influence of modularity and economy-of-scale effects on design of mesh-restorable DWDM networks," *IEEE J. on Sel. Areas in Comm.*, vol. 18, no. 10, pp. 1912-1923, Oct. 2000.

[9] R. R. Iraschko, M. H. MacGregor, and W. D. Grover, "Optimal capacity placement for path restoration in STM or ATM mesh-survivable networks," *IEEE/ACM Trans. Networking*, vol. 6, pp. 325-336, June 1998.