Optimal Capacity Placement for Path Restoration in STM or ATM Mesh-Survivable Networks

Rainer R. Iraschko, M. H. MacGregor, Member, IEEE, and Wayne D. Grover, Senior Member, IEEE

Abstract—The total transmission capacity required by a transport network to satisfy demand and protect it from failures contributes significantly to its cost, especially in long-haul networks. Previously, the spare capacity of a network with a given set of working span sizes has been optimized to facilitate span restoration [11], [12]. Path restorable networks can, however, be even more efficient by defining the restoration problem from an end to end rerouting viewpoint. We provide a method for capacity optimization of path-restorable networks which is applicable to both synchronous transfer mode (STM) and asynchronous transfer mode (ATM) virtual path (VP)-based restoration. Lower bounds on spare capacity requirements in span- and path-restorable networks are first compared, followed by an integer program formulation based on flow constraints which solves the spare and/or working capacity placement problem in either span- or path-restorable networks. The benefits of path and span restoration, and of jointly optimizing working path routing and spare capacity placement, are then analyzed.

Index Terms—Capacity placement, mesh restoration, survivable networks.

I. INTRODUCTION

Mesh-based survivable network architectures exploit the intelligence of the transport network to minimize the amount of spare capacity required to protect working demands [1]–[8]. In mesh-restorable networks the spare capacity on one span contributes to the survivability of many other spans. These networks are called “mesh restorable” to reflect the ability of the restoration mechanism to exploit a meshlike topology through highly diverse and efficient rerouting of failed signal units.

Mesh-restorable networks either reroute failed signal units over a set of replacement paths through the spare capacity of a network between the two nodes terminating a span cut, referred to as span restoration, or over a set of replacement paths between all source and destination nodes affected by a failure, referred to as path restoration. Fig. 1 illustrates the difference. When span $S-T$ fails, span restoration finds replacement path segments directly between nodes $S$ and $T$, whereas path restoration finds end-to-end replacement paths between demand pairs $A-C$ and $B-D$. Path restoration is expected to be more capacity efficient than span restoration because it avoids “backhauls,” as shown in Fig. 1, and distributes replacement paths over a wider region of the network. Path restoration also provides for restoration of transit paths after node loss. However, path restoration is more complex than span restoration because it requires finding a set of replacement paths between numerous source and destination pairs, as opposed to rerouting the same total number of failed signal units between a single pair of nodes.

At the time of this work, two open problems in the design of survivable mesh networks were: 1) optimal spare capacity placement for path restoration and 2) combined optimization of working path routing and spare capacity placement for both span and path restoration. Our interest is in the use of integer programming (IP) based on flow constraints to determine the minimum capacity requirements of a span- or path-restorable network against all individual span failures. This paper compares the different capacity requirements of generic span- and path-restorable networks, independent of the restoration mechanism. Issues related to the restoration mechanisms themselves are addressed in related works [1], [2], [4], [21], [27].

Previous work used an IP approach based on max-flow min-cut considerations to solve the spare capacity placement problem in a span-restorable network [4], [11], [20]. This approach uses cutsets in the IP formulation and populates the constraint sets with cutsets iteratively until the solution provides a spare capacity placement which is 100% span restorable and minimizes the total spare capacity of the network.

A more recent and advantageous approach has been to specify flow constraints based on a suitable set of predefined routes over which restoration pathsets may be implemented [12], [19]. This is referred to as the flow-based approach. An IP formulation using this approach for a 100% span-restorable network was previously reported [12]. In this formulation the working capacity design is given and the IP is based on eligible restoration routes between each pair of nodes terminating a span. An IP which jointly optimizes working and spare capacity using flow constraints in the context of a restorable asynchronous transfer mode (ATM) network has been reported elsewhere [19].

The flow-based approach is preferred over the cutset approach because only a single IP execution is needed to obtain the spare capacity assignment. The same process also yields the exact routing used to restore each span failure, which is not the case in the cutset-based approach; the feasibility of the
required restoration pathsets is guaranteed, but they remain to be found by separate means in the latter. Obtaining the routing that accompanies the capacity placement is helpful when evaluating the performance of a distributed restoration mechanism operating in the target capacity design [2], or in a centralized restoration mechanism that accesses a database of optimal-capacity restoration paths [8].

The capacity placement technique presented here uses the flow-based constraints approach. However, unlike the IP’s presented previously [12], [19], our IP formulation optimizes working and spare capacity, or only spare capacity, in either span- or path-restorable networks. In our approach eligible working and restoration routes are segregated into two sets such that the various capacity placement options are clearly identified. Furthermore, the basic formulation given exactly specifies the capacity requirements of a network which does not assume splitting of individual transport signals or VP’s. The IP formulation presented here is therefore directly applicable to synchronous transfer mode (STM) networks, and also to ATM networks based on replacement of each working VP with a single end-to-end alternate backup VP.

Other authors have investigated the implementation of optical digital cross connect (DCS) [25], and the capacity placement problem from a trunking perspective [26], which unlike this work, allows for service denial. Though the relative efficiency of path and span restoration has been investigated [29], an exact solution to the capacity placement problem in a path-restorable network has not.

This paper proceeds as follows. Section II derives some simple bounding considerations on the intrinsic efficiency of path versus span restoration. This motivates the work on the capacity design of path-restorable networks and also provides some insight about the nature of the effectiveness of path restoration. These hypotheses are revisited during inspection of the final results. Section III contains the main development of the integer program formulation and explains its variations to represent span or path restoration and, for path restoration, the use or omission of stub release. In closing, Section III makes clear the relationship and applicability of these formulations to ATM virtual path (VP)-based restorable networks, showing that IP’s formulated for STM have the same structure as IP’s formulated explicitly for ATM, a point apparent in [28]. This is an area where this paper has been extended from its related conference version [22]. Section IV presents the test networks and results for the various design cases. Section V is devoted to diagnostic inspection and discussion of the results, relating the findings back to the initial insights from Section II, and comparing the effects of stub release and joint working and spare planning on total network capacity. Section VI summarizes this paper.

II. LOWER BOUNDS ON REDUNDANCY

As a preliminary exercise to motivate the design of path-restorable networks, a lower bound on the redundancy of a path-restorable network is compared to that of a span-restorable network next. Redundancy is defined as the ratio of total spare to working capacity and may, in general, be distance weighted. In the following let:

- \( S \) number of spans in a network;
- \( N \) number of nodes in a network;
- \( d \) 2 \( S/N \), i.e., average degree of nodes in the network;
- \( w_i \) working capacity of span \( i \);
- \( D_i \) total number of bidirectional demand pairs affected by span cut \( i \);
- \( \text{Cap}_S \) total spare capacity in a network;
- \( \text{Cap}_{w} \) total working capacity in a network.

The average demand per relation severed by a span cut is then

\[
W_{avg} = \frac{1}{S} \sum_{i=1}^{S} D_i
\]


and the average working capacity per span is then

\[
W_{avg} = \frac{1}{S} \sum_{i=1}^{S} w_i.
\]

If we assume, for bounding considerations only, that 1) network restoration is limited by the number of spares incident with the end-nodes of a failure; 2) each span carries the same capacity \( \text{Cap}_{w} \); and 3) each relation has the same demand \( w_{avg} \) affected by a span cut; then immediate egress for rerouting of failed capacity for the end-nodes of a failed span is possible if the total spare capacity on the unaffected spans terminated at these nodes is enough to accommodate the lost capacity \( W_{avg} \).
A theoretical lower bound on the redundancy of 100% span-restorable networks is [1]

\[ R_{\text{span}} = \frac{\text{Cap}_S}{\text{Cap}_{PW}} = \frac{W_{\text{avg}} \cdot S}{d-1 \cdot W_{\text{avg}} \cdot S} = \frac{1}{d-1}. \] (2.1)

Equation (2.1) states that in a 100% span-restorable network the combined spare capacity of the \( d - 1 \) surviving spans terminated at a node must exceed the working capacity of the cut \( d \)th span.

In path restoration, as in span restoration, the egress capacity of the local end-nodes for rerouting a particular demand pair is the sum of the spare capacity on the \((d-1)\) other unaffected spans. However, unlike span restoration, many end-nodes are usually affected by a failure in a path-restorable network, and each pair is only concerned with the restoration of that portion of \( W_{\text{avg}} \) that pertains to them \(\frac{w_p}{W_{\text{avg}}} \). In this limiting case each span must have \( \frac{w_p}{W_{\text{avg}}/(d-1)} \) spare capacity to fully restore failures. The total spare capacity in the network is then

\[ \text{Cap}_S = \frac{w_p}{d-1} \cdot S. \]

Given that the total working capacity is

\[ \text{Cap}_{PW} = W_{\text{avg}} \cdot S \]

a lower limit on the redundancy of a path-restorable network is

\[ R_{\text{path}} = \frac{\text{Cap}_S}{\text{Cap}_{PW}} = \frac{\frac{w_p}{d-1} \cdot S}{d-1 \cdot W_{\text{avg}} \cdot S} = \frac{1}{d-1} \cdot \frac{w_p}{W_{\text{avg}}} \] (2.2)

Equation (2.2) states that in a 100% path-restorable network the combined spare capacity of the \( d - 1 \) surviving spans terminated at a node must exceed the fraction of working capacity lost by this node.

This result indicates that the redundancy of a path-restorable network improves if demands for each node pair are kept to a minimum on each span of the network, and implies the notion of dispersion of demand routing, a concept which we return to in Section V.

Note that (2.1) can be derived from (2.2). In span restoration only the end-nodes of a span cut are directly involved in restoring a failure, in which case \( w_p = W_{\text{avg}} \) and (2.1) and (2.2) become equivalent, as they should. This also implies that the redundancy of a path-restorable network should be less than or equal to the redundancy of a span-restorable network. Equation (2.2) also implies that as \( w_p \) decreases, the redundancy of a path-restorable network may also decrease. This suggests that in an optimal network design the total working capacity between each node pair may be spread over many diverse routes such that a small amount of demand is lost by many relations after a failure, rather than a large amount being lost by a few relations.

In addition, it is apparent that increasing the connectivity of a network can help decrease the value of \( W_{\text{avg}} \) by reducing the capacity on each span. The number of diverse paths over which demands can be spread will also increase, tending to decrease \( w_p \). However, the extent to which we can decrease \( w_p \) and increase \( W_{\text{avg}} \) to minimize the redundancy of a path-restorable network will also depend on the counteracting desire to minimize the sum of the working links required. The IP-based results presented in Section IV are inspected later to see the extent to which this “dispersion” hypothesis is manifested.

III. IP FORMULATION

A flexible IP formulation is now presented which can be used to optimize the placement of only the spare capacity when working demands are already routed, or jointly optimize both working and spare capacity, of either a span- or path-restorable network. The approach uses flow constraints based on a set of eligible predefined routes over which pathsets may be implemented. The solution to the IP tableau specifies the optimal capacity placement per span as well as the actual paths used to restore each possible span failure.

The IP formulation uses the following notation, illustrated in Fig. 2. Those variables which must be specified, rather than solved for, when the IP is used to optimize the placement of only the spare capacity of a network, are identified:

- \( C_j \): cost of a link (working or spare) assigned to span \( j \);
- \( I_{j+i}^r \): restoration level required for demand pair \( r \) upon the failure of span \( i \). \( 0 \leq I_{j+i}^r \leq 1 \) (for 100% network restorability \( I_{j+i}^r = 1 \) for all \( r \) and all \( i \));
- \( D \): total number of nonzero demand pairs in the demand matrix;
- \( d_r^p \): number of demand units between end-node pair \( r \);
- \( X_r^p \): number of demand units lost by demand pair \( r \) upon the failure of span \( i \). (Must be specified when the IP is used to optimize the placement of spare capacity);
- \( P_r^p \): total number of eligible restoration routes for demand pair \( r \) upon the failure of span \( i \);
- \( Q_r^p \): total number of working routes available to satisfy the demand between node pair \( r \);
- \( f_{r+i,j}^{p,P} \): restoration flow through the \( j \)th restoration route for demand pair \( r \) upon the failure of span \( i \);
- \( y_{r+i,j}^{p,q} \): working capacity required on the \( q \)th working route to satisfy the demand between node pair \( r \). (Must be specified when the IP is used to optimize the placement of spare capacity);
- \( \delta_{r+i,j}^{p,P} \): takes the value of one if the \( p \)th restoration route for demand pair \( r \) after the failure of span \( i \) uses span \( j \), and zero, otherwise;
- \( \zeta_{r+i,j}^{p,q} \): takes the value of one if the \( q \)th working route for demand pair \( r \) uses span \( i \);
- \( u_{r+i,j} \): number of working links (i.e., transport capacity units) on span \( j \);
- \( s_j \): number of spare links on span \( j \).

A. IP for Spare Capacity Placement in Networks with Predefined Demand Routing

The following formulation optimizes the spare capacity placement of a path-restorable network given a fixed working capacity design. The objective function is

\[ \min \left\{ \sum_{j=1}^{S} C_j(s_j) \right\}. \]
Fig. 2. Integer program notation.

The structure of this objective function assumes that the cost of adding a link to a particular span is the same for all links added to that span, and that the cost of adding a link to a particular span doesn’t change while solving the IP formulation. One could interpret $C_j$ as the length of link $j$ and $C_j(S_j)$ as the total physical length of all the spare links on span $j$.

The constraints to be satisfied are:

1) restoration flow meets target restoration levels for each demand pair $r$

$$\sum_{j=1}^{R} f_i^{rP} \geq [X_i^r \cdot L^r_i]$$

\forall r = 1, 2, \cdots, D_r, \forall i = 1, 2, \cdots, S

2) span $j$’s spare capacity is sufficient to meet the simultaneous demands of all restoration routes that use it to restore a single span failure

$$(s_j) - \left( \sum_{i=1}^{D_i} \sum_{j=1}^{R} \delta_i^{rP} \cdot f_i^{rP} \right) \geq 0$$

\forall (i, j) = 1, 2, \cdots, S; \quad i \neq j

3) total demand lost by relation $r$ after the failure of span $i$ is the sum of the flows over relation’s $r$’ working routes traversing span $i$

$$\sum_{q=1}^{Q} g_i^{rQ} \cdot f_i^{rQ} = X_i^r$$

\forall r = 1, 2, \cdots, D_r, \forall i = 1, 2, \cdots, S

4) restoration flows $f_i^{rP}$ and working flows $g_i^{rQ}$ are non-negative integers;

5) spare capacities $s_j$ and working capacities $w_j$ are non-negative integers.

As stated, this IP can be adapted to optimize the spare capacity placement for either a span- or path-restorable network. If a span-restorable design is desired, the set of all node pairs affected by a failure is restricted to just the single pair of nodes terminating the severed span, i.e., $D_i = 1$, and $X_i^r = w_i$.

In a path-restorable network it is advantageous to release the surviving upstream and downstream portions of a cut working path and make those links available to the restoration process. This option is called stub release. Stub release is an option in a path-restorable network but does not arise in a span-restorable network because span restoration only replaces the cut portion of a transport signal. Stub release could be functionally implemented by detecting the alarm indication signal (AIS) alarm at a DCS and issuing a release command. However, whether it is practical to implement stub release, given that it makes reverting to a network’s prefailure state more difficult, requires further investigation. To represent stub release, constraint 2) can be augmented as follows:

2a) span $j$’s spare dimensioning is sufficient to meet the simultaneous demands of all restoration routes that use it to restore a single span failure (first double sum) after releasing the surviving portions of cut paths (second double sum)

$$(s_j) - \left( \sum_{i=1}^{D_i} \sum_{j=1}^{R} \delta_i^{rP} \cdot f_i^{rP} \right) + \left( \sum_{i=1}^{D_i} \sum_{q=1}^{Q} \epsilon_i^{rQ} \cdot f_i^{rQ} \right) \geq 0$$

\forall (i, j) = 1, 2, \cdots, S; \quad i \neq j

Restoration route information is specified in $\delta_i^{rP}$. These restoration routes are not predetermined; rather, they form a set of routes that may be used by the IP to optimize the capacity placement. There will, in general, be more eligible routes than actually used to support restoration paths in the solution.

Working route information is specified in $\epsilon_i^{rQ}$. The working routes are specified to the IP when optimizing only the spare capacity of a network. For the jointly optimized formulation, presented in the next section, the working routes, like the restoration routes, are not predetermined and are part of the set of eligible routes used by the IP to optimize the capacity placement.

All distinct routes between demand pair end-nodes (or span end-nodes, depending on the case) are ideally represented in the constraint system. However, because the number of all distinct routes in a network of $S$ spans is $O(S^2)$, the number typically has to be restricted in practice. The route sets used here were restricted by limiting the length of eligible
routes in a manner that is in part similar to the “hop-limited” approach described in [12]. However, rather than an explicit hop limit, eligible routes were limited to be no longer than the length of the respective shortest route between nodes plus a limited number of additional hops and a limited additional geographical distance. Furthermore, the hop- and distance-limited set of distinct routes was supplemented with the \( k \) successively shortest span-disjoint routes between all pairs of nodes without distance or hop limits. This set of supplemental \( k \) shortest disjoint routes is relatively small, but ensures full representation of the network’s topological connectivity between all nodes, which a simple hop limit may fail to do in some networks.

B. Jointly Optimized Spare and Working Capacity Placement

The constraint system in Section III-A is applicable for span- and path-restorable designs where the working routes are given and only sparing is to be optimized. The IP is here extended to simultaneously optimize the working path routing and spare capacity placement of a path-restorable network. By adding the following constraints \( \{6\) and \(7\)\] to the previous IP formulation, the solution will include the values of \( u_{ij} \) and \( g^{r,q} \), which minimize the total capacity cost in a path-restorable network.

The objective function becomes

\[
\min \left\{ \sum_{j=1}^{S} C_j(s_j + w_{ij}) \right\}
\]

subject to constraints 1)–5), defined previously, and

6) total capacity on the working routes allocated to node pair \( r \) can carry all the demand of relation \( r \)

\[
\sum_{q=1}^{Q} g^{r,q} = d^r \quad \forall r = 1, 2, \ldots, D
\]

7) span \( j \)'s working capacity is sufficient to meet the prefailure demands of all relations which cross it

\[
(w_{ij}) - \left( \sum_{r=1}^{D} \sum_{q=1}^{Q} c_{ij}^{r,q} \cdot g^{r,q} \right) = 0 \quad \forall j = 1, 2, \ldots, S.
\]

As before, the joint formulation can be adapted to a span-restorable network by designating the source and destination of all working paths cut by a span failure as the immediate end-nodes of the severed span (i.e., \( D_0 = 1 \), and \( X_{ij}^r = w_{ij} \)) and by eliminating stub release [i.e., using constraint 2), not 2a)].

C. Relation to ATM

The IP formulation presented so far is directly applicable to, and was developed specifically for, STM networks. However, the formulation for path restoration with stub release also applies directly to two special cases of ATM networks. The first is a network where all VP’s have the same bandwidth allocation. In this case each restoration or working path in the formulation is equivalent to one VP of the common size in an ATM network in which each backup VP provides an equal bandwidth replacement for failed working VP’s. The

fully general ATM problem will admit real-valued bandwidth allocations per VP and multiple VP’s per demand pair. The cases considered here are special in the sense that either each demand pair has only one VP of \( d^r \) bandwidth units or \( d^r \) VP’s each of one bandwidth unit, and all units are integers.

The second type of ATM network modeled by this formulation is the case of an ATM network where VP’s are of differing size but then these VP’s are allowed to be split into a basic constituent size for restoration. In this case one restoration path or working path corresponds to the unit of bandwidth that is the least common multiple of all VP’s, e.g., a single synchronous transport signal (STS)-1. This type of operation may or may not be technically feasible or desired, but the formulation is a direct model for such an ATM system without any extensions. In any case it is useful as a lower bound on the capacity requirements of an ATM network.

The IP can also be extended to exactly represent the case of an ATM network where VP’s are of varying size and are not split for restoration. When VP’s are of differing size, and restoration must occur over a backup VP which is of the same size as the cut working VP, the IP formulation is modified so that the flow of a single working VP \( g^{r,q} \) is constrained to follow a single restoration route. In this case the variable \( X_{ij}^r \)

Fig. 3. Topology of network 1.

Fig. 4. Topology of network 2.
Fig. 5. Topology of network 3.

is replaced with the new variable $q^p_q$, defined subsequently, and constraints 1) and 3) are modified as follows:

1a) backup VP’s are sufficient to meet the target restoration for all working VP’s

$$f^p_q = \sum_{q=1}^{Q^p_q} g^q_q \cdot t^p_q \cdot L^p_q \quad \forall p = 1, 2, \ldots, P^p_q,$$

$$\forall r = 1, 2, \ldots, D; \quad \forall i = 1, 2, \ldots, S$$

where $r^p_q \in (0, 1) \forall (r, q, p)$ and is 1 if restoration route $p$ from relation $r$ is chosen for the backup of VP $q$ and 0 otherwise;

3a) only one backup VP can be used for each working VP, i.e., VP flows are not split

$$\sum_{p=1}^{P^q_p} r^p_q = 1$$

$$\forall q = 1, 2, \ldots, Q^p_q; \quad \forall r = 1, 2, \ldots, D.$$

When the capacity of an ATM network is designed using the IP formulation presented above, each nonzero route is designated a VP. The extension of the STM IP formulations to the ATM case, as shown here, illustrates that IP’s formulated for STM have a structure similar to those formulated explicitly for ATM. Results for the ATM case are presented and analyzed separately [24]. The results for the STM case are presented next.

IV. RESULTS

A. Networks Investigated

Five networks and demand matrices were used to test the IP formulations presented in Section III. Some characteristics of each network are presented in Table I and the topology of each network is shown in Figs. 3–7. Network 1 has a uniform point-to-point demand matrix with two demand units between all node pairs. This is “SmallNet” [16]. Network 2 is a metropolitan area model which was published with a demand matrix, used here, based on industry data [15]. Network 3 is another metropolitan area model and the demand data set is based on a Canadian city. Networks 4 and 5 and their point-to-point demands are representative of long-haul networks.

B. Test Cases and Methodology

The capacity placement for each network presented in Section IV-A was optimized to restore all single-span failures. The first three cases optimized the placement of spare capacity in span- and path-restorable networks given a working capacity design. The working capacity design in Cases 1, 2, and 3 split the routing of demand between a node pair as evenly as possible over the node pair’s equally logically shortest disjoint routes. The latter three cases, i.e., Cases 4, 5, and 6, repeat the first three but with jointly optimized placement of spare and working capacity. The six capacity design cases are as follows.

Case 1: Optimize the placement of spare capacity in a span-restorable network.

Case 2: Optimize the placement of spare capacity in a path-restorable network without stub release.

Case 3: Optimize the placement of spare capacity in a path-restorable network with stub release.

Case 4: Optimize the placement of working and spare capacity in a span-restorable network.

Case 5: Optimize the placement of working and spare capacity in a path-restorable network without stub release.

Case 6: Optimize the placement of working and spare capacity in a path-restorable network with stub release.
The IP formulation for each of these cases was solved using a commercially available software package known as CPLEX [23]. The number of variables (columns) and constraints (rows) in the IP tableau for networks 1 and 4 are presented in Table II. Table II also shows the number of eligible routes available to the IP and the time it took CPLEX to solve the IP formulations on a DEC Alpha workstation with 69 MB of RAM. The set of eligible working and spare routes available to the IP formulation were found using software developed by the authors. This software found all distinct routes between demand pairs up to a given hop and distance limit, as well as the set of $k$ shortest span-disjoint routes, as explained at the end of Section III-A.

Figs. 8–12 summarize the 30 designs (6 design cases times 5 networks) in terms of total network capacities. The functional correctness of each design was verified using independent tools which tested that the working capacity satisfied all entries in the demand matrix, and that all failed working paths were indeed 100% restorable using independent programs to implement span or path restoration, as appropriate within the sparing provided by the IP designs.

V. DISCUSSION OF CAPACITY PLACEMENT RESULTS

A. Path Restoration in Networks with Predefined Demand Routing

To quantify the capacity benefits of path restoration in networks with established working routes, the total capacity required in Cases 2 and 3 was normalized to that of the comparable span-restorable design (Case 1). Table III shows that when the placement of spare capacity is optimized, path-restorable designs without stub release require between 4%–15% less total capacity than the span-restorable designs. The corresponding reductions in spare capacity were from 5%–19%. (The spare/working/total capacity of a network is the

---

**TABLE I**

<table>
<thead>
<tr>
<th>Network</th>
<th>No. of nodes</th>
<th>No. of spans</th>
<th>Avg. network degree</th>
<th>No. of pt-to-pt demands</th>
<th>Total amount of demand</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>22</td>
<td>4.40</td>
<td>45</td>
<td>90</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>28</td>
<td>3.73</td>
<td>67</td>
<td>824</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>31</td>
<td>3.10</td>
<td>153</td>
<td>2 152</td>
</tr>
<tr>
<td>4</td>
<td>53</td>
<td>79</td>
<td>2.98</td>
<td>347</td>
<td>858</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>59</td>
<td>3.93</td>
<td>263</td>
<td>8 312</td>
</tr>
</tbody>
</table>

Fig. 7. Topology of network 5.

Fig. 8. Network 1 designs.

Fig. 9. Network 2 designs.
distance-weighted sum of all of the spare/working/(working + spare) links in the network.

Results also show that stub release can further reduce the total capacity required in a path-restorable network between 1%–8%, as seen in Table III. Whether this further savings is worthwhile in practice may depend on the extent to which stub release complicates reversion to the original prefailure state. In large networks the economic benefits of stub release may be substantial considering that stub release decreased the total number of links (which may each be optical carrier (OC)-η’s

**TABLE II**

<table>
<thead>
<tr>
<th>Network</th>
<th>Case</th>
<th>No. of eligible routes</th>
<th>No. of constraints</th>
<th>No. of variables</th>
<th>CPLEX run time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>6360</td>
<td>485</td>
<td>6382</td>
<td>34 sec.</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>7526</td>
<td>570</td>
<td>7549</td>
<td>24 min</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>7526</td>
<td>570</td>
<td>7549</td>
<td>2.7 min</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>21097</td>
<td>552</td>
<td>21142</td>
<td>9.3 hrs</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>36260</td>
<td>956</td>
<td>19253</td>
<td>2.7 days</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>36260</td>
<td>956</td>
<td>19253</td>
<td>9.2 hrs</td>
</tr>
</tbody>
</table>

**TABLE III**

**SPARE CAPACITY DESIGNS**

<table>
<thead>
<tr>
<th>Network</th>
<th>Normalized spare capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>1.00</td>
</tr>
<tr>
<td>Case 2</td>
<td>0.94</td>
</tr>
<tr>
<td>Case 3</td>
<td>0.93</td>
</tr>
<tr>
<td>2</td>
<td>1.00</td>
</tr>
<tr>
<td>3</td>
<td>0.91</td>
</tr>
<tr>
<td>4</td>
<td>0.85</td>
</tr>
<tr>
<td>5</td>
<td>0.96</td>
</tr>
</tbody>
</table>

Fig. 10. Network 3 designs.

Fig. 11. Network 4 designs.

Fig. 12. Network 5 designs.
of some level) required in network 5 by 2695 (Fig. 12, Case 2 versus Case 3).

B. Capacity Savings from Joint Working and Spare Optimization

The total capacity savings gained when working path routing and spare capacity placement are jointly optimized is seen in Figs. 8–12 when Cases 1–3 are compared to Cases 4–6. The benefit of combined optimization ranges from a total capacity reduction of 4%–27% in the span-restorable designs (comparing Case 1 and Case 4 results). It averages 8% in the path-restorable designs without stub release (comparing Case 2 and Case 5 results) and is about 7% in the path-restorable designs with stub release (comparing Case 3 and Case 6 results).

Combined working and spare capacity optimization requires a larger number of routes to define its constraint system relative to those required for spare capacity optimization alone. This is because eligible working routes need to be specified in addition to eligible restoration routes. Given that the set of eligible routes is much larger when determining a combined capacity design, the constraint systems for Cases 4, 5, and 6 include a smaller proportion of all possible routes than the constraint systems used to optimize the placement of spare capacity alone. Similarly, because path restoration requires eligible restoration routes between all node pairs while span restoration requires them only for every pair of adjacent nodes, the path-restorable designs also include a smaller fraction of all possible routes when compared to the span-restorable designs. For example, as shown in Table II, the number of eligible restoration routes for span restoration in network number 1, Case 1, is 6360, but more than 7526 eligible restoration routes are required for the path restoration case in that network, i.e., Case 3. In contrast, the number of distinct restoration and working routes in network 1 for Case 4 (joint optimization) is over 21,097. As a practical matter, it was only possible to include the complete set of all possible distinct routes in the constraint set for networks 1, 2, and 3 in the nonjoint span-restorable design (Case 1), and the other results are based on strictly incomplete eligible route sets.

Obviously, a design is, in principle, going to be nearer the global optimum when the IP is equipped with a larger number of routes. Therefore, the path-restorable and combined capacity designs here may not be strictly as close to their minimal capacity requirement as are the span-restorable and nonjoint designs. This implies that the theoretical benefits for path restoration (and especially jointly planned path restoration) may be even greater than the practically achievable benefits found here. However, the difference is not expected to be very significant because of the large size of the route set that is used in each case in view of the relatively small number of routes from that set that are actually used by the IP. For example, of 6360 or more restoration routes available for network 1, only 105, 125, and 187 routes, respectively, are actually used by the IP in Cases 1, 3, and 4. As a practical matter, therefore, we have been finding that as long as several dozen or more eligible routes exist for each restoration path or working path choice to be made, the IP does essentially as well as if the entire route set was present, although the latter might contain hundreds of route options for each requirement. This is especially true if the incomplete route set contains the \( k \) shortest span-disjoint routes for enrichment of the route set as described at the end of Section III-A.

C. Routing Effects of Joint Working and Spare Optimization

The jointly optimized designs have some interesting properties which, upon diagnosis, reveal characteristics associated with demand routing dispersion, a principle first introduced as part of the considerations presented in Section II. When the IP is allowed to jointly optimize the placement of working and spare capacity in a network, it will tend to choose working paths which are coordinated with the network restoration process. This means that demands may sometimes be routed via paths longer than the shortest path. Table IV indicates that an outcome of joint optimization is relatively more even \( u_i \) quantities on spans compared to the \( u_i \) levels generated by shortest-path demand routing, which was used in the nonjointly planned cases. This arises because a more even distribution of span working quantities tends to prevent a single span from significantly dominating the spare capacity requirements of other spans in a span- or path-restorable network. With the latitude to seek longer working routes, the IP does so to gain this benefit in the jointly optimized design cases.

From the considerations in Section II, we expect that it should also be advantageous in path-restorable networks to disperse demands over working paths such that a span failure affects many relations, even if such paths are slightly longer than the shortest route. This hypothesis is confirmed in Table

<table>
<thead>
<tr>
<th>Network</th>
<th>Avg. standard deviation of ( w_i ) in case 1, 2, 3 designs</th>
<th>Avg. standard deviation of ( w_i ) in case 4 designs</th>
<th>Avg. standard deviation of ( w_i ) in case 5 designs</th>
<th>Avg. standard deviation of ( w_i ) in case 6 designs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.59</td>
<td>0.891</td>
<td>0.940</td>
<td>0.656</td>
</tr>
<tr>
<td>2</td>
<td>68.2</td>
<td>34.6</td>
<td>33.4</td>
<td>34.6</td>
</tr>
<tr>
<td>3</td>
<td>115</td>
<td>97.5</td>
<td>94.5</td>
<td>106</td>
</tr>
<tr>
<td>4</td>
<td>27.3</td>
<td>22.2</td>
<td>22.3</td>
<td>25.7</td>
</tr>
<tr>
<td>5</td>
<td>617</td>
<td>536</td>
<td>590</td>
<td>593</td>
</tr>
</tbody>
</table>
V. Table V shows that the number of demand pairs to be restored after a failure generally increases when the IP jointly optimizes spare and working capacity. In the event that the number of demand pairs to be restored after a failure decreases rather than increases, as shown in Table V for network 1, it is likely due to the fact that it is impossible to always evenly distribute span working quantities and simultaneously spread the impact of a failure over a large region of the network.

The most general aim when designing a restorable network is to minimize the sum of both the working and spare links required. Because longer working paths require more links, it is reasonable to expect that the IP design should have working paths which deviate only slightly from the shortest route. Table VI confirms that the average working path length in the combined capacity designs is longer than the average working path length in the spare capacity designs, but only slightly. This suggests that the effectiveness of path restoration can, in part, be attributed to the significant route dispersion of demands (Tables III and IV) without dramatically increasing route lengths (Table VI).

VI. CONCLUSIONS

This study has found that mesh-restorable networks using path restoration with stub release are the most capacity efficient. Spare capacity placement in such networks was seen to require up to 19% less total capacity than the corresponding span-restorable networks studied. The overall effectiveness of path restoration is also apparent when one considers that the average redundancy of all path-restorable designs which optimize the placement of spare capacity is 66%, as opposed to 87% for the span-restorable designs.

Results also showed that the capacity required in a 100% path-restorable network with stub release can further be minimized by an average of 7% when jointly optimizing the placement of spare and working capacity. The benefits of jointly optimizing working path routing and spare capacity placement are, however, far more pronounced in span-restorable designs. Comparing jointly optimized span-restorable designs to the baseline of spare-only optimized span-restorable designs, as much as 27% total network capacity may be saved, making jointly planned span-restorable networks almost competitive with path-restorable networks. As a practical matter, however, deploying and operating jointly optimized working routes and spare capacity allocation may be difficult because it implies that existing working routes may have to be rearranged in the face of changing overall demand to retain overall joint capacity optimality. It may therefore be a practical requirement to continue provisioning working capacity based on shortest-path routing and to separately optimize the placement of spare capacity. The results suggest that there would not be a major penalty for this continued mode of operation in a path-restorable network.
All of the results presented in this paper specify the capacity requirements of a network for individual span failures. However, given a path-restorable network, the methodology for determining the capacity requirements for multiple span cut scenarios is the same as for the individual span failure case because only the demand pairs affected by a failure scenario are entered into the IP formulation. Nonetheless, the capacity requirements to restore multiple span failure scenarios are not investigated here because economic realities usually dictate that only sufficient capacity be invested in a network to restore individual span failures.

The IP formulation presented here to solve the spare and combined capacity placement problem can accommodate span- or path-restorable networks, and stub release, if desired. All of the IP formulations presented may be used to design a transport network which adds the same basic unit of capacity to every span, and in which the cost of adding a unit of capacity to a span remains constant for that span while solving the IP. Further work is required to develop an IP formulation that takes into account economies of scale and the modularity of transmission systems.

Finally, a completed IP run specifies the exact spare and working capacity requirements of a mesh-restorable STM network, or an ATM network which doesn’t split VP’s, and the corresponding optimal routing of working and restoration paths. This routing information may be used directly in the case of a centralized restoration control strategy or used as the ideal reference solution in assessing the efficiency of a distributed real-time protocol for path restoration, as in recent studies of a real-time path restoration protocol [21].

REFERENCES


M. H. MacGregor (S’91–M’91) received the M.Sc. degree in chemical engineering (process control) in 1982 and the Ph.D. degree in computing science in 1991, both from the University of Alberta, Edmonton, Alta., Canada.

From 1990 to 1997 he was a Member of the Networks and Systems Group, TRLabs, Edmonton, Alta., Canada, where he worked on algorithms for survivable network design and operation. He is currently an Information Systems Specialist with TELUS PLANet, Edmonton, Alta., Canada, where he is responsible for the server and network architecture of a large residential ISP. He is also an Adjunct Professor with the Departments of Computing Science and Electrical Engineering, University of Alberta, Edmonton, Alta., Canada, where he currently teaches computer architecture.

Wayne D. Grover (S’74–M’76–SM’90) received the B.Eng degree from Carleton University, Ottawa, Ont., Canada, in 1976, the M.Sc. degree from the University of Essex, U.K., in 1981, and the Ph.D. degree from the University of Alberta, Edmonton, Alta., Canada, in 1989, all in electrical engineering.

From 1976 to 1986, he was with BNR (now Nortel Technology) as part of the Scientific Staff and Management, where he worked on digital networks, fiber optics, switching systems, digital radio, and other areas before joining TRLabs, Edmonton, Alta., Canada, as its founding Technical Vice President. With the startup of TRLabs he was responsible for the development of the TRLabs research program and contributing to the development of TRLabs’ sponsorship base. He is currently the Chief Scientist of Network Systems with TRLabs and a Professor of electrical and computer engineering with the University of Alberta, Edmonton, Alta., Canada. He holds U.S. and Canadian patents on 17 topics to date, over half of which have found use in industry. To date, he has published three book chapters and 40 journal papers, and has been involved in supervision of 21 graduate students.

Dr. Grover received the TRLabs Technology Commercialization Award in 1996 for the licensing of restoration-related technologies to industry. He also received the 1996/1997 McCalla Research Professorship in Engineering. He has served as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS.