An Alternative Approach for Enhanced Availability Analysis and Design Methods in p-Cycle-Based Networks

Dev Shankar Mukherjee, *Student Member, IEEE*, Chadi Assi, *Member, IEEE*, and Anjali Agarwal, *Senior Member, IEEE*

***Abstract*— We study the unavailability of end-to-end traffic in p-cycle based mesh networks, which are designed to protect against single link failures. It has been shown earlier by Grover and Clouqueur [6], [7] that the p-cycle length as well as its topology play a vital role in determining the availability of span(s) which are protected by the p-cycle. Similarly, we derive the relationship between the unavailability of a span(s) and the topology of the p-cycle(s) which is allocated for the restoration of the span(s). Based on these insights and on the fact that the end-to-end unavailability of a working path depends not only on the length of the restoration path but also on the number of spans along the working path, we try to design a method for allocating p-cycles such that the end-to-end unavailability is bounded by an upper limit and the upper limit can be varied as desired. As expected, results show that more capacity is required to guarantee a lower end-to-end unavailability. Our results also show that shorter service paths tend to use longer p-cycles than longer service paths, to obtain the same level of availability; this is expected since the path length, apart from the p-cycle length, also plays a role in determining the availability of the service path. We compare this formulation with a formulation which rather limits the hop count of candidate p-cycles to provide a lower end-to-end unavailability. We notice that directly limiting the end-to-end unavailability, as proposed by this paper, gives better results in terms of spare capacity redundancy than limiting the hop count of p-cycles. That is because the former allows shorter working paths to use p-cycles with higher hop count and therefore a better utilization of the allocated spare capacity.**

***Index Terms*— Mesh networks, optical transport network, p-cycles, availability analysis, network survivability, restoration, protection, ILP.**

I. INTRODUCTION

IN high-capacity optical wavelength-division multiplexing (WDM) mesh networks, a failure of a network component can lead to severe disruption of traffic. Hence, fast recovery mechanism against network component failures is critically important in future optical mesh networks. Recovery mechanism can be classified into two main categories: protection and restoration techniques [8], [18]. The former allocates and reserves back-up resources in advance, providing a fast recovery on pre-planned paths, whereas the latter makes use of real time availability of resources to provide back-up paths on the occurrence of a failure in the network. A variety of protection and restoration methods exist for optical transport networks [8]. These include automatic protection switching (APS), p-cycles, shared backup path protection (SBPP), mesh span restoration and mesh path restoration among others. These methods vary in their spare capacity requirements and restoration speed. In p-cycle based design, it has been shown that cycle length plays an important role in determining the unavailability of the portion of the working path which is protected by the p-cycle [6], [7]. In this paper we show that since the length of the working path along which a demand is routed also determines the unavailability of that demand, a smaller working path could be allocated longer p-cycles to protect its spans and still limit the unavailability of the working path (and hence the demand) to a certain upper bound. The interest in allocating longer p-cycles to a shorter working path arises from the fact that longer p-cycles generally have more straddling spans and hence are more efficient from the capacity redundancy point of view [1], [2], [3].

Manuscript received January 16, 2006; revised October 4, 2006.

D. S. Mukherjee was with Concordia University, and is currently with Juniper Networks, San Jose, California (email: (ds_mukhe@encs.concordia.ca).

C. Assi is with the CIISE Department, Concordia University, Montréal, Canada.

A. Agarwal is with the ECE Department, Concordia University, Montréal, Canada.

Digital Object Identifier 10.1109/JSAC-OCN.2006.025006.

A. On Availability in General

Availability is defined as the probability of the system being found in the operating state at some time t in the future given that the system started in the operating state at time t = 0. For repairable systems, an equilibrium is reached between the failure arrival processes and the repair processes, both characterized by the respective rates and resulting in the fraction of the total time that is “up” time [18]. The most widely used expression for availability in repairable systems is given [10], [18] by:

$$A = \frac{MTTF}{MTTF + MTTR} \tag{1}$$

where $MTTF$ (Mean Time to Failure) is the mean time till the failure occurs and $MTTR$ is the Mean Time to Repair. The probabilistic complement of availability is the unavailability U where:

$$U = 1 - A \tag{2}$$

Now, the availability of a system comprising of N elements in a series structure as shown in Figure 1 is equal to the product of the availability of the individual elements; that is

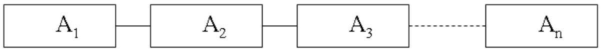


Fig. 1. Series structure of n elements.

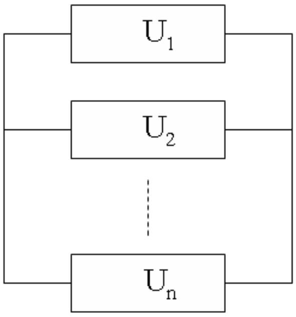


Fig. 2. Parallel structure of n elements.

all the elements in the system needs to be available for the system to be available [10], [18].

$$A_s = A_1 A_2 A_3 .... A_n = \prod_{i=1}^{n} A_i \tag{3}$$

$$U_s = 1 - A_s = 1 - \prod_{i=1}^{n} A_i = 1 - \prod_{i=1}^{n} (1 - U_i) \tag{4}$$

$$U_s \approx \sum_{i=1}^{n} U_i \tag{5}$$

An approximation in equation (5) is done by ignoring the higher power of U_i, as it is assumed that $U_i << 1$.

Alternatively, the unavailability of a system comprising of N elements in parallel as shown in Figure 2 is given by the product of the unavailability of the individual elements; that is all the system would be unavailable if all the elements in the system are unavailable [10], [18].

$$U_p = \prod_{i=1}^{n} U_i \tag{6}$$

To summarize, for systems with elements in series the unavailability of each element needs to be added up, which is an approximation considering that $Ui << 1$ and for systems with elements in parallel the unavailability of each element is multiplied to obtain the total unavailability.

B. Overview of p-Cycle Protection

The method of pre-configured protection cycles (p-cycles) proposed by W. Grover's research group [1], [3] can achieve ring-like high speed protection with mesh-like high efficiency in the use of spare capacity. This is because the p-cycle method makes use of the ring protection function (originally used in SDH/SONET rings) to perform fast protection. Additionally it can protect not only the on-cycle spans but also the straddling spans in the network.

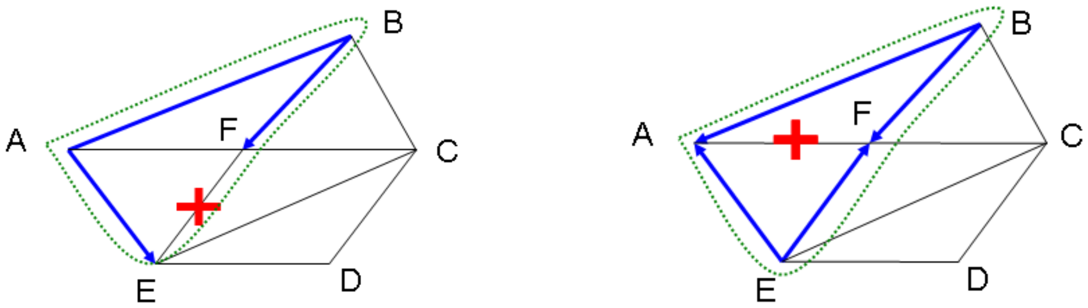
There are two types of p-cycles. Link p-cycles protect the individual channels within a link. On the other hand, node encircling p-cycles [4] are routed through all neighbor nodes of a specific node and protect all the connections traversing through that node. Figure 3 illustrates the way in which link p-cycles may be used for restoration, in ways that are not possible with rings. In Figure 3(a), a span on the cycle breaks and the surviving arc of the cycle is used for restoration using a method similar to the ring based BLSR protection mechanism. Figure 3(b) shows how the p-cycle can also be accessed for restoration of working paths (e.g. AF) that are not on the cycle. In fact, two restoration paths are possible for every straddling link.

Most of the previous work on protection has focused on 100% restorability[1] against single span failure. Single span failure restoration methods have been suggested using both heuristic [5] as well as ILP methods [2]. Clearly single failure restorability is of great overall benefit to network integrity, but that qualitative recognition does not guarantee that the availability of service path will be 100%. To users of a WDM network, the end-to-end availability is of utmost concern. Hence, being able to analyze the factors that affect and determine the end-to-end availability is of great interest not only academically but also for network operators as it assists them in offering their customers some guarantees for high service availability. Such guarantees are commonly specified in service level agreements (SLA) which has to be honored by the network operator and typically a failure to provide service as specified per the SLA may force the network operators to pay certain penalties. Given this, it is vital for the network operators to have quantitative measures of end-to-end availability in order to offer competitive SLAs.

C. Related Work

A number of papers have been published on p-cycles schemes, most of which are related to the issue of optimizing spare capacity placement required to support single failure restoration [1]- [5]. Multiple and Dual failure restoration have recently been studied by Schupke et al in [15], [16], [17]. In [16] it is shown that the dual failure restorability and the protection capacity can vary significantly for cycle-configurations with different numbers of deployed p-cycles. In [11], the existing p-cycle network design theory is extended to include the capability of direct restriction on protection path lengths, rather than indirect restriction through circumference limits. On the other hand, the authors of [12] developed mathematical models for path availability and provisioning resources required in various strategies for realizing high availability service paths in bidirectional line-switched rings or shared protection WDM rings. More related work by Clouqueur and Grover can be found in [6], [7], [13], [14]; in [6] and [13], the authors explain the relation between the path availability and the restorability of a network to dual

[1]Restorability has been defined as a fraction of demands that are affected by a failure scenario, but which survive by virtue of recovery mechanisms using the spare capacity of the network [15].



(a) (b)

Fig. 3. A p-cycle example

failures. The availability analysis of a span restorable network is based on the computational analysis of the restorability of a network to all possible dual-failure scenarios. In [7] the concept of availability analysis in span restorable network is extended to p-cycles. The paper computes the availability of spans protected by p-cycles, by considering all dual failure sequences that result in an outage of a path protected by p-cycles. The authors consider the working path to consist of one or more domains where each domain is one or more spans in the working path, which is protected by the p-cycle and they compute the unavailability of a domain. The authors showed the importance that cycle size plays in terms of availability and suggested strategies for achieving high availability of paths in a network protected by p-cycles.

D. Outline

In section II we show, with the help of an example, that in a span restorable network the length of the restoration path of a particular demand as well as the length of its corresponding working path, together determine the unavailability of the demand. In section III we present a computational analysis of the unavailability of the portion of the working path which is protected by a single p-cycle. In section IV we use the computational analysis developed in section III to develop an integer linear program (ILP) model, which optimizes the p-cycle allocation to a pre-routed network and allocates p-cycles such that the end-to-end unavailability of every working route in the network is less than a certain upper limit. This upper limit is an input parameter to the ILP model. The results obtained from the ILP are provided in section V. Finally, section VI concludes our finding.

II. Unavailability of Span Restorable Service Path

In any network, the most common aim in designing for survivability is to achieve restorability against all single failures with an objective to minimize the required spare capacity. This leaves dual failure situation as the main factor which can contribute to the unavailability of the service. It is reasonable to neglect the probability of occurrence of triple failures in networks as it is significantly less than the probability of occurrence of dual failures. Moreover, in networks where pre-connected protection schemes have been used to achieve

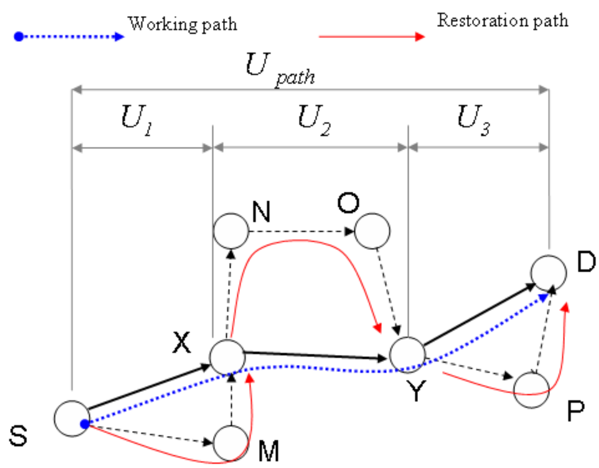


Fig. 4. An example of span restorable end-to-end path.

restorability (as in the case of p-cycles) against single failures, the recovery speed is expected to be very fast, e.g., in the order of 50ms to a maximum of 1 or 2 sec. In such networks the unavailability of service during recovery time is also very insignificant and can be neglected. Numerical examples in [6] have shown that the effect of dual span-failures are in fact more important in determining the expected service path unavailability in long haul networks, and considering only dual failures is sufficient to obtain a good estimate on the availability of the service. Here, we will analyze the end-to-end unavailability of span restorable service path. We will do that with the help of an example, as shown in Figure 4.

In this example, the total unavailability of the service path S-X-Y-D is the sum of the unavailability of traffic between the nodes S and X, X and Y and Y and D. The network between the nodes S and X, X and Y and Y and D is represented as three domains and the un-availabilities of these three domains are represented as U_1, U_2, U_3. These elements are in a series structure and as per equation (5):

$$U_{sxyd} = U_1 + U_2 + U_3 \tag{7}$$

Further, each of the domains is composed of a restoration path and a working span in a parallel structure. In this example it is assumed that the restoration paths are dedicated to the working path S-X-Y-D. As per equation (6) the unavailability for each domain would be calculated as follows:

$$U_1 = U_{sx} \times U_{smx} \tag{8}$$

$$U_2 = U_{xy} \times U_{xnoy} \tag{9}$$

$$U_3 = U_{yd} \times U_{ypd} \tag{10}$$

where U_{smx}, U_{xnoy}, U_{ypd} are the unavailabilities of the restoration paths of each of the spans S-X, X-Y and Y-D respectively. Each of the restoration paths is a series structure of spans. The unavailability of the restoration path would hence be given by the following equations:

$$U_{smx} = U_{sm} + U_{mx} = 2U \tag{11}$$

$$U_{xnoy} = U_{xn} + U_{no} + U_{oy} = 3U \tag{12}$$

$$U_{ypd} = U_{yp} + U_{pd} = 2U \tag{13}$$

To simplify the equations, we have considered the unavailability of all spans to be equal (e.g., U in this case), though it is understood that the unavailability of spans depends on various characteristics like length, the terrain that the link is through, etc. Hence, the unavailability of each domain and the total unavailability are given by:

$$\begin{aligned} U_1 &= 2U^2 \\ U_2 &= 3U^2 \\ U_3 &= 2U^2 \end{aligned} \tag{14}$$

$$U_{path} = (2 + 3 + 2)U^2 \tag{15}$$

Now, the unavailability of each domain can also be calculated by finding the number of possible dual failures which cause an outage. For example, in domain 2 there are three possible dual failures which can cause an outage. These are X-Y and X-N, X-Y and N-O, X-Y and O-Y. The possibility of each dual failure is U^2, as the failure in each span is an independent event and hence the unavailability of two spans would be the product of the unavailability of each of the spans. Given this, the unavailability of domain 2 is therefore the probability that any of these three pair of spans is unavailable, which is $3U^2$.

Through observation we can deduce the following:

1) The unavailability of each domain in the working path tends to worsen as longer restoration paths are allocated. In the case of a p-cycle based design, the length of the restoration path has a relation with the size of the cycle allocated to restore the corresponding span.
2) The end-to-end unavailability is likely to be more for a service path with more number of domains than for a shorter service path. For p-cycle based design, it would imply that a shorter service path can be protected by longer p-cycles (or p-cycles with larger hop count) than longer service path to achieve the same level of end-to-end availability.

III. Unavailability Analysis in p-Cycle Based Network

In [7], Clouqueur and Grover used the concept of "protection domains" to analyze the unavailability of a set of spans which are protected by a p-cycle. We use a similar concept in this section to find the unavailability of any path. Any working path could be protected by a set of p-cycles, where each p-cycle could be protecting one or more spans of the working path. We define a "protection domain" as the set of spans which are protected by the same p-cycle. In short, each path is divided into sub-sections based on the p-cycle allocated to protect each sub-section. Here, according to this definition we are modelling the path as a series structure of elements where the elements are the protection domains. Please note that in [7] if a span on a path was protected by a p-cycle as an on-cycle span and another span on the same path was protected as a straddling span, then these two spans were counted as two different domains. According to our definition of protection domains all spans in a path protected by the same p-cycle belong to the same protection domain and hence the two spans in the above case also belong to the same domain.

A. Unavailability Analysis of a Protection Domain

Here we will analyze the unavailability of a section of the path, which belongs to the same protection domain. We perform this by finding all possible combinations of dual failures within the protection domain that can result in an outage on the corresponding service path. The ability to analyze the unavailability of a protection domain will enable us to analyze the end-to-end unavailability as protection domains are serially connected to form the end-to-end path. In our analysis we assume the physical unavailability of each span to be equal to U. Extension of the unavailability expression developed in this paper to consider span specific unavailability should be fairly straightforward. We have partitioned the p-cycle (x) that is protecting the span/spans along a path p into sets using the following notations:

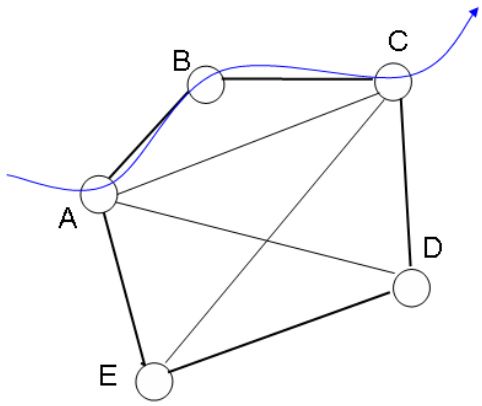
- O_x^p: The set of spans in p-cycle (x) that are on-cycle spans and are also in the working path (p).
- $O_x^{\overline{p}}$: The set of spans in p-cycle (x) that are on-cycle spans but are not in the working path (p).
- S_x^p: The set of spans in p-cycle (x) that are straddling spans and are also in the working path (p).
- $S_x^{\overline{p}}$: The set of spans in p-cycle (x) that are straddling spans but are not in the working path (p).

On further analysis we can categorize any outage causing dual failure scenarios within a p-cycle corresponding to a protection domain into one of the following categories (we will be referring to these categories throughout the rest of the paper):

Category-1: Dual failure scenario in which one of the spans belongs to O_x^p and the other span belongs to $O_x^{\overline{p}}$.

Category-2: Dual failure scenario in which one of the spans belongs to O_x^p and the other span belongs to $S_x^{\overline{p}}$.

Category-3: Dual failure scenario in which one of the spans belongs to O_x^p and the other span belongs to S_x^p.



As per the notations followed:

$O_x^p = \{A-B, B-C\}$, $O_x^{\overline{p}} = \{C-D, D-E, E-A\}$

S_x^p = NULL, $S_x^{\overline{p}} = \{A-C, C-E, A-D\}$

Fig. 5. Working path protected as on-cycle spans.

Category-4: Dual failure scenario in which one of the spans belongs to S_x^p and the other span belongs to $O_x^{\overline{p}}$.

Category-5: Dual failure scenario in which both spans belong to the S_x^p.

Category-6: Dual failure scenario in which one of the spans belongs to S_x^p and the other span belongs to $S_x^{\overline{p}}$.

We will analyze the probability of an outage caused in each such scenario with the help of a few examples.

B. *Example 1*

The first example presented in Figure 5 shows a part of a network in which a working path is traversing a p-cycle (A-B-C-D-E-A). Two spans (A-B and B-C) along the working path are protected by the p-cycle as on-cycle spans. These two spans together form a "protection domain" of the end-to-end working path. Our aim is to find the unavailability of this protection domain. In other words, what is the probability of finding that the traffic entering the protection domain at A suffers an outage before it leaves the protection domain at C given that there was no outage between the Source and A?

In Figure 6 we show the case where the first failure occurs on one of the spans AB or BC and the second failure occurs on one of the on-cycle spans that are not on the working path. Such a scenario is certain to cause an outage. Obviously the case when the two failures occur in the reverse order is also guaranteed to cause outage for the path. In this particular protection domain, there are 2 categories of dual failures, which can cause an outage. In the first category (Category-1) one of the spans belongs to O_x^p (i.e., either of the spans A-B and B-C) and the other span belongs to $O_x^{\overline{p}}$ (i.e., one among the spans A-E, E-D and D-C). The example shown in Figure 6 belongs to this category of dual-failures. Obviously, the number of possible combinations of dual failures in this category is $|O_x^p|.|O_x^{\overline{p}}|$. Hence, the contribution of these kinds of dual failures to the unavailability of the protection domain is $|O_x^p|.|O_x^{\overline{p}}|.U^2$, where U is the physical unavailability of any span in the network. Formally, we have:

$$U_{Category-1} = |O_x^p|.|O_x^{\overline{p}}|.U^2 \tag{16}$$

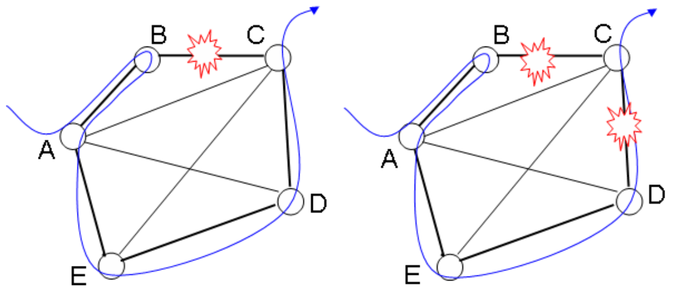


Fig. 6. Dual failure case : One span belongs to the set O_x^p and the other span belongs to the set $O_x^{\overline{p}}$.

The second category of dual failures that may result in an outage in this example is Category-2 in which one of the spans belongs to O_x^p (i.e., either of the spans A-B or B-C) and the other span belongs to $S_x^{\overline{p}}$ (i.e., any one of the spans A-C, A-D or C-E). In this category, the order in which the failures occur is important. If the first failure occurs in a span belonging to the set O_x^p (i.e., on-cycle span) then the path would be switched to the back-up path. The second failure occurring on a straddling link will not affect the restoration path of the span, which failed first and hence there would be no service outage.

Alternatively, if the first failure occurs on a straddling span and assuming that the p-cycle is fully loaded [2], the straddling span would be restored using both arcs of the p-cycle. In this situation a failure of a span in O_x^p will result in an outage. We formally denote the unavailability due to a dual failure in this category as:

$$U_{Category-2} = \frac{1}{2}|O_x^p|.|S_x^{\overline{p}}|.U^2 \tag{17}$$

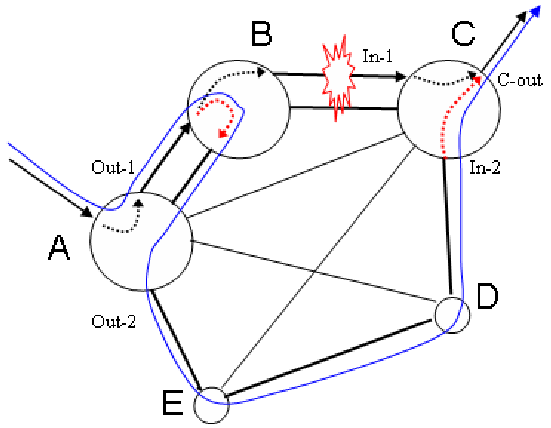
Here, the factor 0.5 denotes the probability that a straddling span fails first. The unavailability of the protection domain (A-B-C) can therefore be expressed as the sum of the unavailability obtained in equations (16) and (17):

$$U_{total} = U_{Category-1} + U_{Category-2} \tag{18}$$

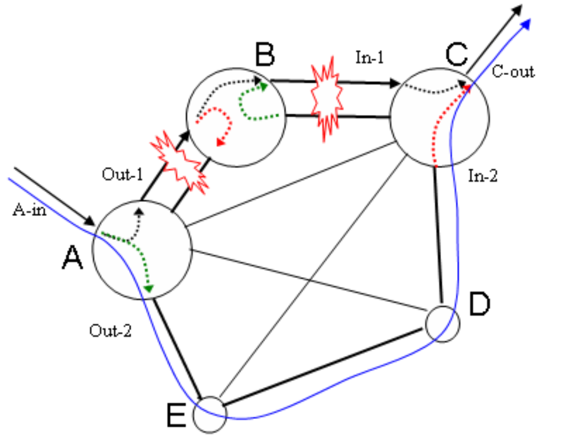
It can be shown that if spans A-B and B-C fail there would be no outage. The manner in which the cross-connects at nodes A, B and C are pre-configured to handle span failures on spans A-B and B-C would force the demand (originally traversing A-B-C) to be restored using the arc A-E-D-C in the event that both A-B and B-C fail. This is shown graphically in Figure 7.

In this example (Figure 7) the changes in the switch configuration which were effective after the occurrence of the second failure (failure of span A-B) are not biased by the fact that a prior failure had occurred on span B-C, but rather are premeditated. Although the loop back performed at node B after the failure of span A-B is not useful, the change in configuration at node A enables the working path to use the arc A-E-D-C as the restoration path. In the above example, the on-cycle spans present on the working path were adjacent spans and a dual failure involving both spans clearly did not

[2]Fully loaded p-cycle is one which provides restoration to two units of working capacity in all straddling spans and one unit of working capacity to all on-cycle spans.

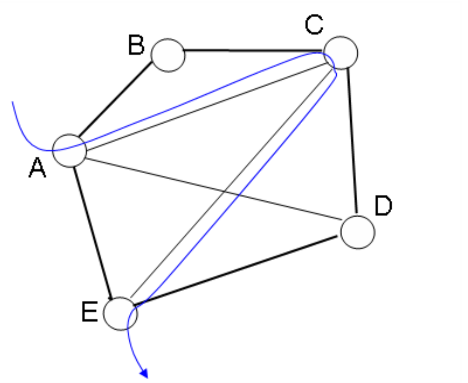


(a) The first failure in span B-C is restored by switching to a new, but pre-configured XC configuration at node B and C. Node B performs a loop-back to divert the working path to the protection cycle. Node C changes the XC configuration from (In-1→ C-out) to (In-2 → C-out).



(b) A failure in span A-B is restored by switching to a new, but pre-configured XC configuration at node A and B. Node A changes the XC configuration from (A-in→ Out-1) to (A-in→ Out-2). Node B performs a loop-back to divert the traffic in the protection path to the working path.

Fig. 7. Dual failure case: Both spans belong to the set O_x^p



As per the notations followed:

O_x^p = NULL, $O_x^{\overline{p}}$= $\{A-B, B-C, C-D, D-E, E-A\}$

S_x^p = $\{A-C, C-E\}$, $S_x^{\overline{p}}$ = $\{A-D\}$

Fig. 8. Working path protected as straddling spans.

result in any outage. Using a similar analysis, it can also be shown that the same holds true for on-cycle spans, which are not adjacent and are traversed by the same working path.

C. Example 2

The second example (shown in Figure 8) shows part of a network in which a working path crosses a p-cycle (A-B-C-D-E-A) at two spans (A-C and C-E), which are straddling to the p-cycle. The portion of the working path crossing the p-cycle forms a protection domain. In such a protection domain, there are 3 categories of dual failures (please refer to Figure 9) which can result in an outage to the working path as per our classification of dual-failure categories.

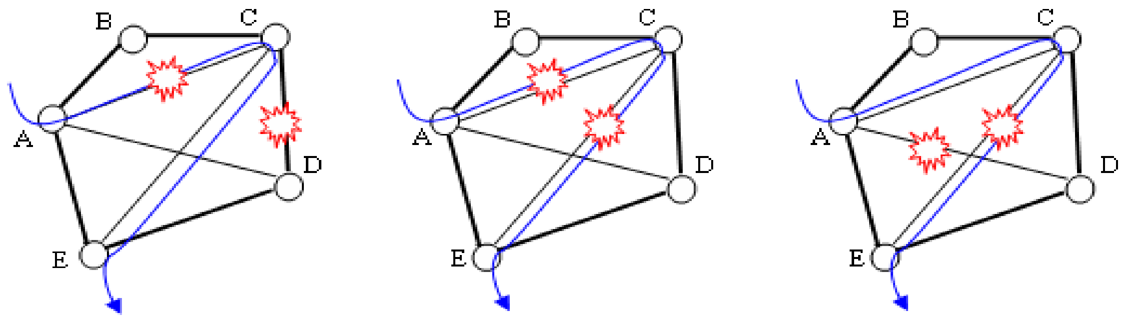
- Category-4, where one span among the straddling spans crossed by the working path fails and one span among the on-cycle spans not crossed by the working path fails.
- Category-5, where 2 straddling spans crossed by the working path fail.
- Category-6, where one span among the straddling spans crossed by the working path fails and one span among the straddling span not crossed by the working path fails.

For each of these three categories we find the number of combinations of dual failures that can result in an outage and accordingly we calculate the contribution that each category of dual failures makes towards the unavailability of the protection domain.

Category-4: There is a 50% chance that the on-cycle span fails first which is then followed by a second failure occurring on the straddling span (which is crossed by the concerned working path); this permutation of dual failures will definitely lead to service outage along the path. On the other hand, a span failure could first affect the straddling span and the second span failure could occur on an on-cycle span. In this case there is 50% chance that the on-cycle span would affect the back-up path used by the straddling span because the straddling span could, without discrimination, use any of the two halves of the p-cycle (see Figure 9-a). Given this (un-ordered) combination of failures, the probability of an outage for the concerned path is 75%. The number of combination of dual-failures in this category is $|S_x^p|.|O_x^{\overline{p}}|$. We formally denote the unavailability due to a dual failure in this category by:

$$U_{Category-4} = \frac{3}{4}|S_x^p|.|O_x^{\overline{p}}|.U^2 \qquad (19)$$

Category-5: The occurrence of the failure on a straddling span will activate both halves of the p-cycle; therefore, a following second failure on another straddling span (which is on the working path) will certainly result in an outage for the working path. Here we have assumed that the p-cycle is fully loaded or utilized, hence both halves of the p-cycle would be utilized, one half by the concerned working path and the other half by a working path (which is not of concern) flowing through the failed span but in the reverse direction. By assuming a fully loaded p-cycle, the unavailability calculations are pessimistic. In reality the unavailability due to dual failures



(a) Category 4 (b) Category 5 (c) Category 6

Fig. 9. Illustrations showing one example in each of the three categories of dual failure that can effect the working path.

will be somewhat less than our calculations. The number of combination of dual-failures in this category is given by $(|S_x^p|.|S_x^p - 1|)/2$; hence the unavailability contribution by this type of dual failures can formally be written as:

$$U_{Category-5} = \frac{1}{2}|S_x^p|.|S_x^p - 1|.U^2 \qquad (20)$$

Category-6: Here we are dealing with a dual-failure occurring on two straddling spans, one which is traversed by the working path (e.g., C-E) and the other is not (e.g., A-D). In this combination, there is a 50% chance that the straddling span which is traversed by the working path fails first; here a second failure occurring on another straddling span will not affect the restoration path of the span which failed first and hence there would be no outage for the working path. If the two failures occurred in the reverse order, a disruption is guaranteed to occur and hence there is a 50% chance that any failure in this category will result in an outage. Obviously, the possible number of such dual-failure combinations in any protection domain is $|S_x^p|.|S_x^{\overline{p}}|$, which leads the unavailability contribution due to failures in this category to be:

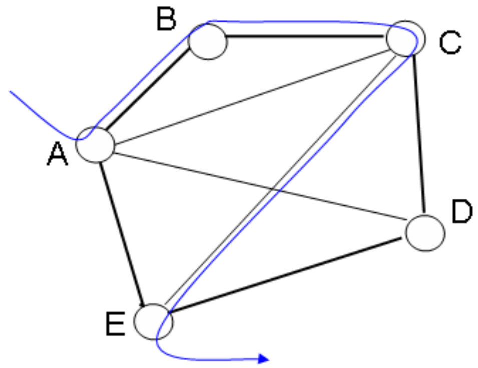
$$U_{Category-6} = \frac{1}{2}|S_x^p|.|S_x^{\overline{p}}|.U^2 \qquad (21)$$

The unavailability of the protection domain in this example is the sum of the unavailability calculated in equations (19), (20) and (21).

D. Example 3

Finally we take a broader example of a protection domain in which none of the sets O_x^p, $O_x^{\overline{p}}$, S_x^p and $S_x^{\overline{p}}$ in the protection domain is a NULL set. Figure 10 shows a working path traversing a p-cycle through on-cycle spans as well as straddling spans. The portion of the working path, which crosses the p-cycle forms the protection domain that we are interested in analyzing.

Dual-failures related to all six categories can take place in such a protection domain, as shown in Figure 10. It should be appreciated that the unavailability as shown in equations (16), (17), (19), (20) and (21) would also apply to this example. The additional type of dual-failures that could affect this protection domain is when an on-cycle span and a straddling span both traversed by the working path fail; e.g., the dual-failure combination of A-B and C-E or B-C and C-E. This



As per the notations followed:

$O_x^p = \{A-B, B-C\}$, $O_x^{\overline{p}} = \{C-D, D-E, E-A\}$

$S_x^p = \{C-E\}$, $S_x^{\overline{p}} = \{A-D, A-D\}$

Fig. 10. Working path protected as on-cycle span as well as straddling span.

category of dual-failure is Category-3 as per our previous classification. Irrespective of the order in which the failures occur, the working path outage is guaranteed. The first failure occurring on the straddling span would result in the spare capacity in both halves of the p-cycle to be used to protect traffic on the straddling span and the second failure on the on-cycle span will result in an outage as the p-cycle would already be pre-occupied. If the failures happen in the reverse order, the failed on-cycle span would be protected by the remaining arc of the p-cycle and the second failure on the straddling span would not be restorable. It should be noted that the working capacity in both directions of the straddling span is restored by the p-cycle (here we are assuming fully loaded/utilized p-cycles).

The number of combinations of such failures is clearly $|O_x^p|.|S_x^p|$ and hence the contribution to the total unavailability as a result of dual-failures in this category is given by:

$$U_{Category-3} = |O_x^p|.|S_x^p|.U^2 \qquad (22)$$

Without much elaboration it can be stated that the total unavailability of a protection domain can the given by the sum of the contribution towards unavailability by each category. Formally, we have:

$$U_{domain} = |O_x^p|.|O_x^{\overline{p}}|.U^2 + \frac{1}{2}.|O_x^p|.|S_x^{\overline{p}}|.U^2 + |O_x^p|.|S_x^p|.U^2 + \frac{3}{4}.|S_x^p|.|O_x^{\overline{p}}| + \frac{1}{2}.|S_x^p|.|(S_x^p - 1)|.U^2 + \frac{1}{2}.|S_x^p|.|S_x^{\overline{p}}|.U^2 \quad (23)$$

Finally, since the working path is a series of domains, the total unavailability of the working path would the summation of the unavailability of each individual domain:

$$U_{path} = \sum_{domains} U_{domain} \quad (24)$$

IV. Formulation to Limit the End-to-End Unavailability

In this section we provide a formulation to optimize the allocation of spare capacity to find the minimum cost capacity placement that enables us to guarantee that every working path is protected against single span failures. The model will also ensure that the unavailability of all end-to-end working paths is less than a certain user set upper bound. The demands are pre-routed and all the working paths are provided as input parameters to the Integer Linear Program (ILP). The routing of the demands is done before hand and could be done using any load balancing routing algorithm or any other suitable algorithm. The optimization is therefore a non-joint optimization problem. We restrict our discussions to bi-directional p-cycles. We further assume a WDM network with full wavelength conversion, or networks with similar characteristics. For this formulation we use the following notations.

Input parameters:

- S Set of spans.
- C_k Cost of a span k.
- P Set of simple p-cycles available for allocation.
- R Set of working paths.
- π_k^p On-cycle relation between a p-cycle (p) and a span (k). π_k^p is equal to '1' if span k is an on-cycle span in p-cycle p , otherwise it is '0'.
- δ_k^p Protection relation between a p-cycle (p) and a span (k). δ_k^p is equal to '1' if span k is an on-cycle span in p-cycle p, '2' if the span k is a straddling link in p-cycle p and '0' if the span is not protected by the p-cycle p.
- ϕ_k^r ϕ_k^r is equal to '1' if span k is part of the working path r, '0' otherwise.
- MU Maximum unavailability of any working path after the allocation of p-cycle.

Output parameters:

- S_k Number of spare units allocated in span k.
- N_r^p N_r^p is is '1' if p-cycle p is allocated for the protection of working path r, '0' otherwise.
- N^p Number of instances of p-cycle p allocated.
- O_r^p Number of on-cycle spans in p-cycle p which are also part of the working path r. Note that in the previous section O_r^p was defined as the set of such on-cycle spans and not the number of such spans.
- $O_r^{\overline{p}}$ Number of on-cycle spans in p-cycle p which are not part of the working path r.
- S_r^p Number of straddling spans in p-cycle p which are also part of the working path r.
- $S_r^{\overline{p}}$ Number of straddling spans in p-cycle p which are not part of the working path r.
- U_r^p Unavailability of a protection domain in working path r protected by p-cycle p.
- U_r Total end-to-end unavailability of working path r.

The objective is to minimize the total protection capacity cost:

$$Minimize \sum_{k \in S} C_k S_k \quad (25)$$

The constraint (26) given below finds the set of p-cycles that have to be allocated to ensure that every working path r is protected against single failures on every span along the working path. A p-cycle could either be allocated for a certain route or not, and N_r^p contains this information:

$$\sum_{p \in P; \delta_k^p > 0} N_r^p \geq \phi_k^r, \forall r \in R; k \in S \quad (26)$$

For a given p-cycle that is allocated for a working path, the spans on the working path, which are protected by the p-cycle belong to a "protection domain". From the intermediate output N_r^p we can determine the protection domains for each working path. Equations (27) to (30) calculate the variables of interest: O_r^p, $O_r^{\overline{p}}$, S_r^p and $S_r^{\overline{p}}$.

$$O_r^p = \sum_{k \in S} N_r^p \phi_k^r \pi_k^p, \forall r \in R; p \in P \quad (27)$$

$$O_r^{\overline{p}} = \sum_{k \in S} N_r^p (1 - \phi_k^r) \pi_k^p, \forall r \in R; p \in P \quad (28)$$

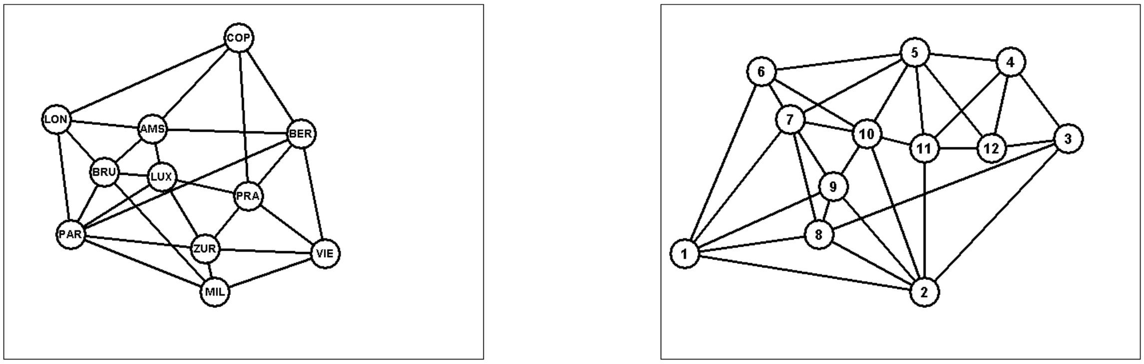
$$S_r^p = \sum_{k \in S; \delta_k^p = 2} N_r^p \phi_k^r, \forall r \in R; p \in P \quad (29)$$

$$S_r^{\overline{p}} = \sum_{k \in S; \delta_k^p = 2} N_r^p (1 - \phi_k^r), \forall r \in R; p \in P \quad (30)$$

From the output of the equations (27) to (30) and using equation (23) we compute U_r^p , which is the unavailability of the protection domain in working path r, which is protected by p-cycle p:

$$U_r^p = \{O_r^p.O_r^{\overline{p}} + \frac{1}{2}.O_r^p.S_r^{\overline{p}} + O_r^p.S_r^p + \frac{3}{4}.S_r^p.O_r^{\overline{p}} + \frac{1}{2}.S_r^p(S_r^p - 1) + \frac{1}{2}.S_r^p.S_r^{\overline{p}}\}.U^2, \quad \forall r \in R; p \in P \quad (31)$$

The end-to-end unavailability is calculated by summing U_r^p over all domains of a working path r:



(a) COST239 network (b) 12n30s network

Fig. 11. Test Networks used.

$$U_r = \sum_{p \in P} U_r^p, \forall r \in R; p \in P \tag{32}$$

We now constrain the unavailability of the working path to a certain upper limit - which is our main objective. The upper limit is an input parameter to our ILP. The p-cycles will be allocated such that this constraint is satisfied. That is, for a lower value of the constraint, p-cycles with less hop-count will be allocated and for a more relaxed value of the constraint the optimizer will tend to allocated longer p-cycles. Obviously if the desired value for unavailability is too low, a solution may not exist.

$$U_r \leq MU; \forall r \in R \tag{33}$$

The equation (34) given below finds the number on unit protection capacity allocated on the p-cycle (p). The R.H.S. of the equation is the number of instances of p-cycle (p) required by any span (k) so as to protect all the working paths, which traverse k. The R.H.S side is not necessarily an integer value. For example if a p-cycle protects a certain span as a straddling span for 7 units of working capacity, then the number of instances of the p-cycles required by that span would be calculated as 3.5, as each instance of the p-cycle can provide protection to 2 units of straddling span. But N^p is an integer value and is more than or equal to the R.H.S side for all spans k.

$$N^p \geq \sum_{r \in R; \delta_k^p = 2} \frac{1}{2} N_r^p \phi_k^r + \sum_{r \in R; \delta_k^p = 1} N_r^p \phi_k^r, \forall k \in S \tag{34}$$

The total spare capacity in each span is given by the equation below:

$$S_k = \sum_{p \in P} N^p \pi_k^p, \forall k \in S, \forall p \in P \tag{35}$$

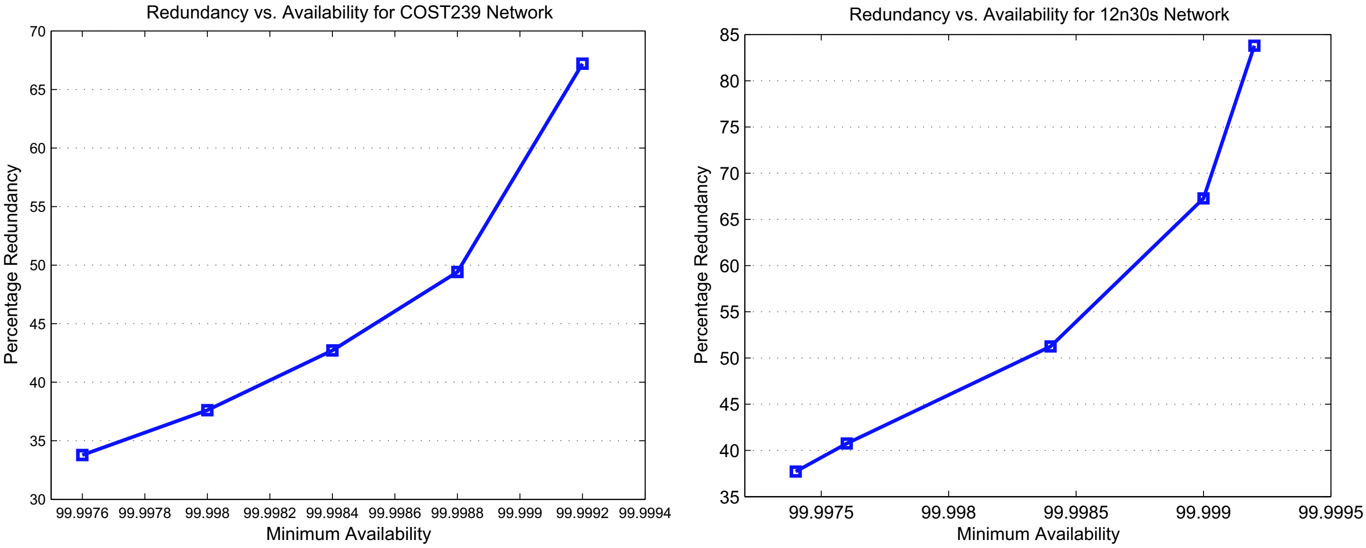
The spare capacity given by equation (35) is optimized as per the objective (25).

It is worth noting here that as the number of nodes in the network increases, the number of candidate p-cycles increases exponentially especially in dense networks. Finding an optimal set of p-cycle using the ILP formulation presented previously can be shown to be an NP-hard problem [21] though limiting the number of candidate p-cycles can reduce the computation time by compromising the optimality.

V. Numerical Results

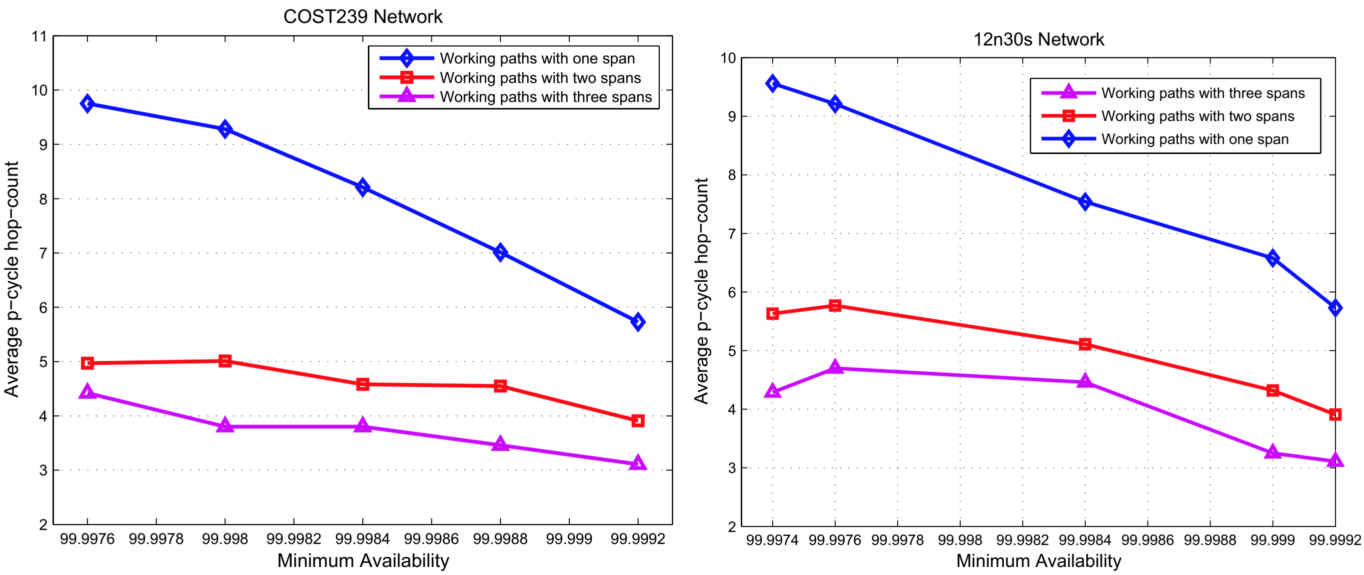
We have chosen the COST239 and the 12n30s networks to implement the ILP formulation we provided in the previous section, and these networks are shown in Figure 11. Without loss of generality we assume that each node can perform full wavelength conversion. We also assume that each span has enough capacity to support the protection capacity required by the optimal solution. The formulation is implemented as an AMPL model and solved using the solver CPLEX 9.1.3 [20]. A separate C model is used to find the input parameters for the AMPL model. In our implementation, the demands between any two pair of nodes are randomly assigned to a certain number of lightpaths ranging from 0 to 10. The demands are symmetric; i.e., every source-destination pair has equal number of lightpaths in either direction. The demands are routed using a Dijkistra shortest path algorithm with metrics reciprocal to the free capacity of the span [2]. Lightpaths between a pair of nodes are routed individually and two lightpaths between the same pair of nodes can possibly have different working routes. Each lightpath demand is associated with a working route; all working routes are provided as inputs to the AMPL model. Candidate p-cycles are found using a depth-first search and are pre-selected using AE metrics [19] and provided as input to the AMPL model.

Figure 12.a (Figure 12.b) compares the redundancy obtained for the desired level of availability for COST239 (12n30s respectively). The redundancy here is defined as the ratio of the sum of the spare capacity in each span to the sum of the working capacity in each span. For computational simplicity, we assume the unavailability of each span to be equal. In our case, we assumed the unavailability of each span to be equal to 10^{-3}. As expected, the results for both networks show that the availability increases at the cost of additional spare capacity requirement. The same set of demands is used to find the redundancy for each of the desired availability requirement.



(a) (b)

Fig. 12. Redundancy versus minimum availability for (a) COST239 network, (b) 12n30s network.



(a) (b)

Fig. 13. Average p-cycle hop count for different working path length for (a) COST239 network, (b) 12n30s network

Another interesting observation is with respect to the average length of the p-cycles (in terms of hop-count), which are allocated for the restoration of working paths. Here, we are interested in analyzing the average cycle length because longer cycles tend to be more efficient from spare capacity redundancy point of view. We segregate the working paths in our network into three groups: working paths with only one span, working paths with two spans and working paths with three spans. In our simulation, none of the working paths covered more than 3 spans. The average cycle length of the p-cycles allocated to paths in a particular group is calculated, and the same calculation is done for all the three groups and compared. Figure 13.a shows the average cycle length for the three groups for different levels of availability for COST239 network. It can be observed that:

1) Longer p-cycles are allocated to working paths covering smaller number of spans. For example to achieve an availability level of 99.9976%, the working paths with one span are allocated p-cycles with an average cycle hop-count of 9.75, whereas working paths containing two spans are allocated p-cycles with an average hop-count of 4.97 and the average hop-count of p-cycles allocated to working paths containing three spans is 4.42.
2) In general the average cycle length decreases as the level of availability is increased.

The same results can also be observed for the 12n30s network, as shown in Figure 13.b.

Next we simulate our ILP without constraining the unavailability to an upper bound, given by equation (33), but rather by limiting the length of candidate p-cycles (i.e., limiting their hop-count). The results found by this method would be the same as those found by the conventional single failure restoration ILP provided by Grover et al [1], wherein the ILP model finds a set of allocated p-cycles with an objective to minimize the redundancy. We find out the working path with the worst case (i.e., minimum) availability, when the candidate p-cycles are pre-selected based on the hop-count. Figure 14 compares the worst case availability of any working path in the entire network when the length of candidate p-cycles is restricted (based on the hop-count), with that of the worst case availability of any working path when the unavailability is constrained by equation (33) and the p-cycle hop-count is not restricted. Figure 14 shows results from the COST239 network. For example, on pre-selecting candidate cycles with hop-count not exceeding 5 hops, we obtain a set of p-cycles through the ILP (in which the unavailability is not constrained), such that there exists a demand which has an availability of 99.998% and the overall redundancy in the network is 56.10%. We compare the availability of the demand which has the worst availability rather than comparing the average unavailability of the demands in order to find a lower bound on the availability of all the demands. Clearly, directly limiting the unavailability proves to be a better design option than limiting the hop-count of the candidate p-cycles. As the figure shows, our method achieves a better (i.e., lower) spare capacity redundancy for the same availability (or unavailability) or better availability for almost the same redundancy.

Finally, we show the average hop count (vs. the availability) of the p-cycles allocated for (1) the unavailability constrained design and (2) the design in which the hop count of the p-cycle was limited; the results are shown in Figure 15. It can be seen that as the availability requirement increases, the hop count decreases for our design. The same is also true for the hop count limited design except that in the hop count limited design, the availability is obtained as a post design calculation and is not a design constraint. We can also see that the unavailability constrained design tends to have a higher average hop count for the same availability level as compared to the hop count limited design which therefore explains the lower redundancy obtained by our design.

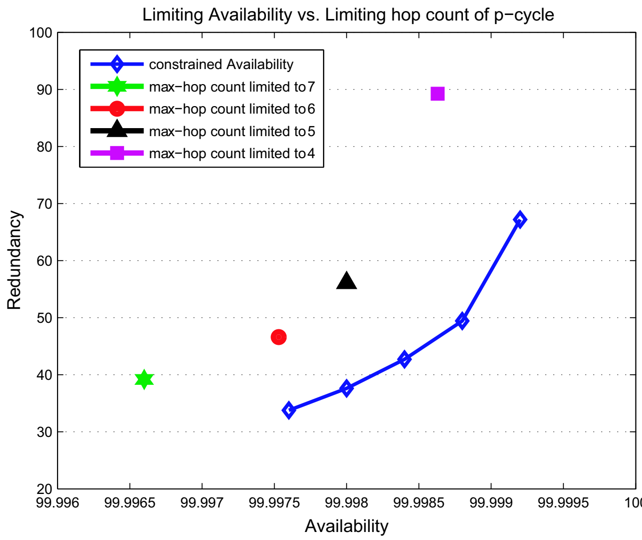


Fig. 14. Directly limiting unavailability vs. limiting hop-count of candidate p-cycles for COST239 network.

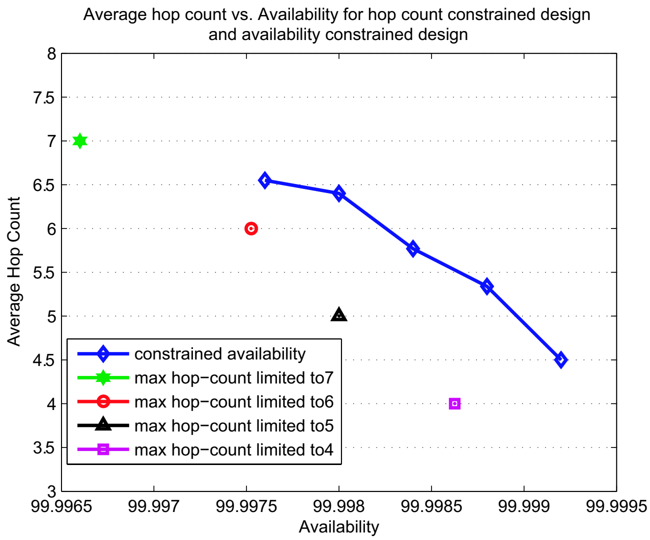


Fig. 15. Average hop count vs Availability for both the design methods - unavailability constrained design and the hop count limited design. Network used is COST239.

VI. Conclusions

In this paper we studied the relationship between the unavailability of a span and the topology of the p-cycle, which is allocated for the restoration of the spans in mesh networks with p-cycle based protection method. We then provided a method for allocating p-cycles to restore single link failures such that the unavailability of all the demands in the network is bounded by an upper limit. Results have provided us with an insight on the relationship between the length of the working path and the length of the p-cycles, which are allocated to protect the spans along the working path. Allocating p-cycles based on the desired availability (or unavailability) rather than limiting the hop-count of candidate p-cycles appears to be a promising method. That is because longer cycles can be allocated to protect shorter working paths and achieve the same availability; with longer p-cycles, better sharing of the protection capacity among multiple spans (on cycle and straddling spans) can be obtained. As a topic for future study, demands in a network can be further classified into different priority levels, where each priority level would have a different upper bound for the end-to-end unavailability, thus providing different classes of protection and availability to various classes of service.

Acknowledgment

We thank the Natural Sciences and Engineering Research Council of Canada (NSERC), the FQRNT, and Concordia University for their generous support.

References

[1] W. Grover and D. Stamatelakis, "Bridging the ring-mesh dichotomy with p-cycles," in *Proc. Second International Workshop on the Design of Reliable Communication Networks (DRCN 2000)*, pp. 92-104.

[2] D. A. Schupke, C. G. Gruber, and A. Autenrieth, “Optimal configuration of p-cycles in WDM networks,” in *Proc. IEEE International Conference on Communications (ICC) 2002*, pp. 2761-2765.
[3] W. D Grover and D. Stamatelakis, “Cycle-oriented distributed preconfiguration: Ring-like speed with mesh-like capacity for self-planning network restoration,” in *Proc. IEEE International Conference on Communications (ICC) 1998*, pp. 537-543.
[4] D. Stamatelakis and W. D. Grover, “Rapid span or node restoration in IP networks using virtual protection cycles,” in *Proc. Third Canadian Conference on Broadband Research (CCBR ’99)*, pp. 33-44.
[5] Z.-R..Zhang, W.-D. Zhong, and B. Mukherjee, “A heuristic method for design of survivable WDM networks with p-cycles,” *IEEE Commun. Lett.*, vol. 8, no. 7, pp. 467-469, July 2004.
[6] M. Clouqueur and W. D. Grover, “Availability analysis of span-restorable mesh networks,” *IEEE J. Sel. Areas Commun.*, special issue on recent advances in fundamentals of network management, vol. 20, no. 4, pp. 810-821, May 2002.
[7] M. Clouqueur and W. D. Grover, “Availability analysis and enhanced availability design in p-cycle-based networks,” *Photonic Network Communications*, Springer Science, vol. 10, no. 1, pp. 55-71, 2005.
[8] S. Ramamurthy and B. Mukherjee, “Survivable WDM mesh networks– part I: protection,” in *Proc. IEEE Infocom ’99*, pp. 744-751.
[9] I. Jurdana and B. Mikac, “An availability analysis of optical cables,” in *Proceedings WAON’98*, pp. 153-160.
[10] I. Rados, P. Tarulija, and T. Sunaric, “Availability model of bidirectional line switched ring,” in *Proc. Transparent Optical Networks 2001*, pp. 312-316.
[11] A. Kodian, A. Sack, and W. D. Grover, “p-cycle network design with hop limits and circumference limits,” in *Proc. BroadNets 2004*, pp. 244-253.
[12] W. D. Grover, “High availability path design in ring-based optical networks,” *IEEE/ACM Trans. Networking*, vol. 7, no. 4, pp. 558-574, Aug. 1999.
[13] M. Clouqueur and 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)*, pp. 181-186
[14] M.Clouqueur and W. D. Grover, “Quantitative comparison of end-to-end availability of service paths in ring and mesh-restorable networks,” in *Proc. 19th Annual National Fiber Optics Engineers Conference (NFOEC 2003)*, pp. 317-326.
[15] D. A. Schupke, “Multiple failure survivability in WDM networks with p-cycles,” in *Proc. IEEE International Symposium on Circuits and Systems (ISCAS) 2003*, pp. 866-869.
[16] D. A. Schupke, “The tradeoff between the number of deployed p-cycles and the survivability to dual fiber duct failures,” in *Proc. IEEE International Conference on Communications (ICC) 2003*, pp. 1428-1432.
[17] D. A. Schupke, W. D. Grover, and M. Clouqueur, “Strategies for enhanced dual failure restorability with static or reconfigurable p-cycle networks,” in *Proc. IEEE International Conference on Communications (ICC) 2004*, pp. 1628-1633.
[18] W. D. Grover, *Mesh-Based Survivable Networks: Options and Strategies for Optical, MPLS, SONET and ATM Networking*. Upper Saddle River, NJ: Prentice Hall PTR, 2003.
[19] W. D. Grover and J. Doucette, “Advances in optical network design with p-cycles: joint optimization and pre-selection of candidate p-cycles,” in *Proc. IEEE LEOS Tropical Meetings*, pp. 49-50.
[20] CPLEX/AMPL: http://www.ilog.com/products/ampl/
[21] K. Lo, D. Habibi, Q. V. Phung, H. N. Nguyen, “Dynamic p-cycles selection in optical WDM mesh networks,” in *Proc. 13th International Conference on Networks (ICON2005)*, pp. 844-849.

Dev Shankar Mukherjee received a B.Tech (Hons) degree in Electronics and Electrical Communication Engineering from Indian Institute of Technology, Kharagpur in 1999 and a M.A.Sc in Electrical Engineering from Concordia University, Montreal in 2006.

Dev is currently a hardware development engineer in the core technology group in Juniper Networks. Prior to joining Juniper Networks, he held engineering positions in Transwitch Inc. where he worked on design and architecture of ASICS for SONET/SDH application. Dev’s research interests are in the area of DWDM, SONET/SDH and IP Networks.

Chadi Assi received the B.S. degree in engineering from the Lebanese University, Beirut, Lebanon, in 1997 and the Ph.D. degree from the Graduate Center, City University of New York, New York, in April 2003. He was a Visiting Researcher at Nokia Research Center, Boston, MA, from September 2002 to August 2003, working on quality-of-service in optical access networks. He joined the Concordia Institute for Information Systems Engineering (CIISE), Concordia University, Montreal, QC, Canada, in August 2003 as an Assistant Professor. Dr. Assi received the Mina Rees Dissertation Award from the City University of New York in August 2002 for his research on wavelength-division-multiplexing optical networks. His current research interests are in the areas of provisioning and restoration of optical networks, wireless and ad hoc networks, and security.

Anjali Agarwal (SM’03) received her Ph.D. (Electrical Engineering) in 1996 from Concordia University, Montreal, M.Sc. (Electrical Engineering) in 1986 from University of Calgary, Alberta, and B.E. (Electronics and Communication Engineering) in 1983 from Delhi College of Engineering, India.

She is currently an Associate Professor in the Department of Electrical and Computer Engineering at Concordia University, Montreal. Prior to joining faculty in Concordia, she has worked as a Protocol Design Engineer and a Software Engineer in industry, where she was involved in providing specifications and design issues for TCP/IP and Voice over IP support, and in the software development life cycle of real time embedded software. Her current research interests are in the various aspects of real-time and multimedia communication over Internet and over the access networks.