

Overlay Protection Against Link Failures Using Network Coding

Ahmed E. Kamal, Aditya Ramamoorthy, Long Long, Shizheng Li

Abstract—This paper introduces a network coding-based protection scheme against single and multiple link failures. The proposed strategy ensures that in a connection, each node receives two copies of the same data unit: one copy on the working circuit, and a second copy that can be extracted from linear combinations of data units transmitted on a shared protection path. This guarantees instantaneous recovery of data units upon the failure of a working circuit. The strategy can be implemented at an overlay layer, which makes its deployment simple and scalable. While the proposed strategy is similar in spirit to the work of Kamal '07 & '10, there are significant differences. In particular, it provides protection against multiple link failures. The new scheme is simpler, less expensive, and does not require the synchronization required by the original scheme. The sharing of the protection circuit by a number of connections is the key to the reduction of the cost of protection. The paper also conducts a comparison of the cost of the proposed scheme to the 1+1 and shared backup path protection (SBPP) strategies, and establishes the benefits of our strategy.

I. INTRODUCTION

Research on techniques for providing protection to networks against link and node failures has received significant attention [1]. Protection, which is a proactive technique, refers to reserving backup resources in anticipation of failures, such that when a failure takes place, the pre-provisioned backup circuits are used to reroute the traffic affected by the failure. Several protection techniques are well known, e.g., in 1+1 protection, the connection traffic is simultaneously transmitted on two link disjoint paths. The receiver, picks the path with the stronger signal. On the other hand in 1:1 protection, transmission on the backup path only takes place in the case of failure. Clearly, 1+1 protection provides instantaneous recovery from failure, at increased cost. However, the cost of protection circuits is at least equal to the cost of the working circuits, and typically exceeds it. To reduce the cost of protection circuits, 1:1 protection has been extended to 1:N protection, in which one backup circuit is used to protect N working circuits. However, failure detection and data rerouting are still needed, which may slow down the recovery process. In order to reduce the cost of protection, while still providing instantaneous recovery, references [13], [15] proposed the sharing of one set of protection circuits by a number of working circuits, such that each receiver in a connection is able to receive two copies of the same data unit: one on the working circuit, and another one

from the protection circuit. Therefore, when a working circuit fails, another copy is readily available from the protection circuit. The sharing of the protection circuit was implemented by transmitting data units such that they are linearly combined inside the network, using the technique of network coding [16]. Two linear combinations are formed and transmitted in two opposite directions on a p-Cycle [4]. We refer to this technique as 1+N protection, since one set of protection circuits is used to simultaneously protect a number of working circuits. The technique was generalized for protection against multiple failures in [14].

In this paper, we propose a new method for protection against multiple failures that is related to the techniques of [15], [14]. Our overall objective is still the same; however, the proposed scheme improves upon the previous techniques in several aspects. First, instead of cycles, we use paths to carry the linear combinations. This reduces the cost of implementation even further, since in the worst case the path can be implemented using the cycle less one segment (that may consist of several links). Moreover, a path may be feasible, while a cycle may not. Second, each linear combination includes data units transmitted from the same round, as opposed to transmitting data units from different rounds as proposed in [15]. This simplifies the implementation and synchronization between nodes. This aspect is especially important when considering a large number of protection paths, since synchronization becomes a critical issue in this case. The protocol implementation is therefore self-clocked since data units at the heads of the local buffers in each node are combined provided that they belong to the same round. Overall, these improvements result in a simple and scalable protocol that can be implemented at the overlay layer. The paper also includes details about implementing the proposed strategy. Recently, a network coding scheme to protect against a combination of adversarial errors and failures under a similar model was proposed in [2]. In this case more protection resources are required.

This paper is organized as follows. In Section II we introduce our network model and assumptions. In Section III we introduce the modified technique for protection against single failures. Implementation issues are discussed in Section IV. In Section V we present a generalization of this technique for protecting against multiple failures. The encoding coefficient assignment is discussed in Section VI. In Section VII we present an integer linear programming formulation to provision paths to protect against single failures. Section VIII provides some results on the cost of implementing the proposed tech-

The authors are with the Dept. of Electrical and Computer Engineering at Iowa State University, Ames, IA 50011 (email: {kamal, adityar, longlong, szli}@iastate.edu). The material in this paper has appeared in part at the 42nd Annual Conf. on Information Sciences and Systems (CISS), 2008. This work was funded in part by grants CNS-0626741 and CNS-0721453 from NSF, and a gift from Cisco Systems.

nique, and compares it to 1+1 protection and SBPP. Section IX concludes this paper with a few remarks.

II. MODEL AND ASSUMPTIONS

In this section we introduce our network model and the operational assumptions. We also define a number of variables and parameters which will be used throughout the paper.

A. Network Model

We assume that the network is represented by an undirected graph, $G(V, E)$, where V is the set of nodes and E is the set of edges. Each node corresponds to a switching node, e.g., a router, a switch or a crossconnect. Network users access the network by connecting to input ports of such nodes, possibly through multiplexing devices. Each undirected edge corresponds to two transmission links, e.g., fibers, which carry data in two opposite directions. The capacity of each link is a multiple of a basic transmission unit, which can be wavelengths, or smaller tributaries, such as DS-3, or OC-3. In this paper, we do not impose an upper limit on the capacity of a link, and we assume that it carries a sufficiently large number of basic tributaries, i.e., we consider the uncapacitated case.

In order to protect against single link failures, the network graph needs to be at least 2-connected. That is, between each pair of nodes, there needs to be at least two link disjoint paths. The number of protection paths, and the connections protected by each of these paths depends on the connections and their end points, as well as the network graph. An example of connection protection in NSFNET will be given in Section III. In general, for protection against M link failures, the graph needs to be $(M + 1)$ -connected.

Since providing protection to connections will require the use of finite field arithmetic, these functions are better implemented in the electronic domain. Therefore, we assume that protection is provided at a layer that is above the optical layer, and this is why we refer to this type of protection as *overlay protection*.

B. Operational Assumptions

We make the following operational assumptions:

- 1) The protection is at the connection level, and it is assumed that all connections that are protected together will have the same transport capacity, which is the maximum bit rate that has to be handled by the connection. We refer to this transport capacity as B^1 .
- 2) All connections are bidirectional.
- 3) Paths used by connections that are jointly protected are link disjoint.
- 4) A set of connections will be protected together by a protection path. The protection path is bidirectional, and it passes through all end nodes of the protected connections. The protection path is also link disjoint from the paths used by the protected connections.

¹Throughout this paper we assume that all connections that are protected together have the same transport capacity. The case of unequal transport capacities can also be handled, but will not be addressed in this paper.

- 5) Links of the protection path protecting a set of connections have the same capacity of these connections, i.e., B .
- 6) Segments of the protection path are terminated at each connection end node on the path. The data received on the protection path segment is processed, and retransmitted on the outgoing port, except for the two extreme nodes on the protection path.
- 7) Data units are fixed and equal in size.
- 8) Nodes are equipped with sufficiently large buffers. The upper bound on buffer sizes will be derived in Section IV.
- 9) When a link carrying active (working) circuits fails, the receiving end of the link receives empty data units. We regard this to be a data unit containing all zeroes.
- 10) The system works in time slots. In each time slot a new data unit is transmitted by each end node of a connection on its primary path². In addition, this end node also transmits a data unit in each direction on the protection path. The exact specification of the protocol, and the data unit is given later.
- 11) The amount of time consumed in solving a system of equations is negligible in comparison to the length of a time slot. This ensures that the buffers are stable³.

The symbols used in this paper are listed in Table I, and will be further explained within the text. The upper half of the table defines symbols which relate to the working, or primary connections, and the lower half introduces the symbols used in the protection circuits. All operations in this paper are over the finite field $GF(2^m)$ where m is the length of the data unit in bits. It should be noted that all addition operations (+) over $GF(2^m)$ can be simply performed by bitwise XOR's. In fact, for protection against single-link failures we only require addition operations, which justifies the last assumption above.

III. 1+N PROTECTION AGAINST SINGLE LINK FAILURES

In this section we introduce our strategy for implementing network coding-based protection against single link failures.

Consider a set of \mathbb{N} bidirectional, unicast connections, where the number of connections is given by $N = |\mathbb{N}|$. Connection $i \leftrightarrow j$ is between nodes S_i and T_j . Nodes S_i and T_j belong to the two ordered sets \mathcal{S} and \mathcal{T} , respectively. Data units are transmitted by nodes in \mathcal{S} and \mathcal{T} in rounds, such that the data unit transmitted from S_i to T_j in round n is denoted by $d_i(n)$, and the data unit transmitted from T_j to S_i in the same round is denoted by $u_j(n)$ ⁴. The data units received by nodes S_i and T_j are denoted by \hat{u}_j and \hat{d}_i , respectively, and can be zero in the case of a failure on the primary circuit between S_i and T_j .

²The terms primary and working circuits, or paths, will be used interchangeably.

³Typically, a single connection will have a bit rate on the order of 10's or 100's of Mbps that is much lower than the capacity of a fiber or a wavelength. Therefore, we assume that the processing elements of a switching node will be able to process the data units within the transmission time of one data unit.

⁴For simplicity, the round number, n , may be dropped when it is obvious.

TABLE I

LIST OF SYMBOLS: UPPER HALF ARE SYMBOLS USED FOR WORKING PATHS, AND LOWER HALF ARE SYMBOLS FOR PROTECTION PATHS.

Symbol	Meaning
\mathbb{N}	set of connections to be protected
N	number of connections = $ \mathbb{N} $
\mathcal{S}, \mathcal{T}	two disjoint ordered sets of communicating nodes, such that a node in \mathcal{S} communicates with a node in \mathcal{T}
S_k, T_k	sets of connection end nodes protected by \mathbf{P}_k
S_i, T_j	nodes in \mathcal{S} and \mathcal{T} , respectively
d_i, u_j	data units sent by nodes S_i and T_j , respectively
\hat{d}_i, \hat{u}_j	data units sent by nodes S_i and T_j , respectively, on the primary paths, which are received by their respective receiver nodes
$T(S_i)$	node in \mathcal{T} transmitting to and receiving from S_i
$S(T_j)$	node in \mathcal{S} transmitting to and receiving from T_j
B	the capacity protected by the protection path
n	round number
M	total number of failures to be protected against ($M = 1$ in Section III).
\mathbf{P} (or \mathbf{P}_k)	bidirectional path used for protection
\mathbb{P}	set of protection paths
\mathbf{S}, \mathbf{T}	unidirectional paths of \mathbf{P} started by S_1 and T_1 , respectively
$\sigma(S_i)(\sigma(T_j))$	the next node downstream from S_i (respectively T_j) on \mathbf{S}
$\sigma^{-1}(S_i)(\sigma^{-1}(T_j))$	the next node upstream from S_i (respectively T_j) on \mathbf{S}
$\tau(S_i)(\tau(T_j))$	the next node downstream from S_i (respectively T_j) on \mathbf{T}
$\tau^{-1}(S_i)(\tau^{-1}(T_j))$	the next node upstream from S_i (respectively T_j) on \mathbf{T}
$\chi_w(\chi_{\mathbf{P}})$	delay over working (protection) path
$F_{\mathbf{S}}(S_i)(F_{\mathbf{T}}(S_i))$	buffers at node S_i used for transmission on the \mathbf{S} (\mathbf{T}) paths
$\alpha_{i \leftrightarrow j, k}$	scaling coefficient used for connection between S_i and T_j on \mathbf{P}_k
$y_e(z_e)$	The data unit transmitted on link $e \in \mathbf{S}$ ($e \in \mathbf{T}$ respectively)
K	The total number of protection paths, i.e., $ \mathbb{P} $

The two ordered sets, $\mathcal{S} = (S_1, S_2, \dots, S_N)$ and $\mathcal{T} = (T_1, T_2, \dots, T_N)$ are of equal lengths, N , which is the number of connections that are jointly protected. If two nodes communicate, then they must be in different ordered sets. These two ordered sets define the order in which the protection path, \mathbf{P} , traverses the connections' end nodes. The ordered set of nodes in \mathcal{S} is enumerated in one direction, and the ordered set of nodes in \mathcal{T} is enumerated in the opposite direction on the path. The nodes are enumerated such that one of the two end nodes of \mathbf{P} is labeled S_1 . Proceeding on \mathbf{P} and inspecting the next node, if the node does not communicate with a node that has already been enumerated, it will be the next node in \mathcal{S} , using ascending indices for S_i . Otherwise, it will be in \mathcal{T} , using descending indices for T_i . Therefore, node T_1 will always be the other end node on \mathbf{P} . The example in Figure 1 shows how ten nodes, in five connections are assigned to \mathcal{S} and \mathcal{T} . The bidirectional protection path is shown as a dashed line.

Under normal working conditions the working circuit will be used to deliver d_i and u_j data units from S_i to T_j and from T_j to S_i , respectively. The basic idea for receiving a second copy of data u_j by node S_i , for example, is to receive on two

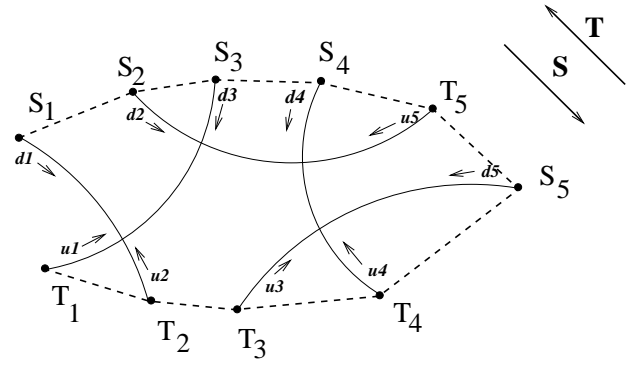


Fig. 1. An example of enumerating the nodes in five connections. Node T_5 is the first node to be encountered while traversing \mathbf{S} , which communicates with a node in \mathcal{S} that has already been enumerated (S_2).

opposite directions on the protection path, \mathbf{P} , the signals given by the following two equations, where all data units belong to the same round, n :

$$\sum_{k, S_k \in \mathbf{A}} d_k + \sum_{k, T_k \in \mathbf{B}} \hat{u}_k \quad (1)$$

$$u_j + \sum_{k, T_k \in \mathbf{B}} u_k + \sum_{k, S_k \in \mathbf{A}} \hat{d}_k \quad (2)$$

where \mathbf{A} and \mathbf{B} are disjoint subsets of nodes in the ordered set of nodes \mathcal{S} and \mathcal{T} , respectively, such that a node in \mathbf{A} communicates with a node in \mathbf{B} , and vice versa. If the link between S_i and T_j fails, then u_j can be recovered by S_i by simply adding equations (1) and (2).

We now outline the steps involved in the construction of the primary/protection paths and the encoding/decoding operations at the individual nodes.

A. Protection Path Construction and Node Enumeration

- 1) Find a bidirectional path⁵, \mathbf{P} , that goes through all the end nodes of the connections in \mathbb{N} . \mathbf{P} consists of two unidirectional paths in opposite directions. These two unidirectional paths do not have to traverse the same links, but must traverse the nodes in the opposite order. One of these paths will be referred to as \mathbf{S} and the other one as \mathbf{T} .
- 2) Given the set of nodes in all N connections which are to be protected together, construct the ordered sets of nodes, \mathcal{S} and \mathcal{T} , as explained above
- 3) A node S_i in \mathcal{S} (T_j in \mathcal{T}) transmits d_i (u_j) data units to a node in \mathcal{T} (\mathcal{S}) on the primary path, which is received as \hat{d}_i (\hat{u}_j).
- 4) Transmissions on the two unidirectional paths \mathbf{S} and \mathbf{T} are in rounds, and are started by nodes S_1 and T_1 , respectively. All the processing of data units occurs between data units belonging to the same round.

⁵The path is not necessarily a simple path, i.e., vertices and links may be repeated. We make this assumption in order to allow the implementation of our proposed scheme in networks where some nodes have a nodal degree of two. Although the graph theoretic name for this type of paths is a *walk*, we continue to use the term *path* for ease of notation and description.

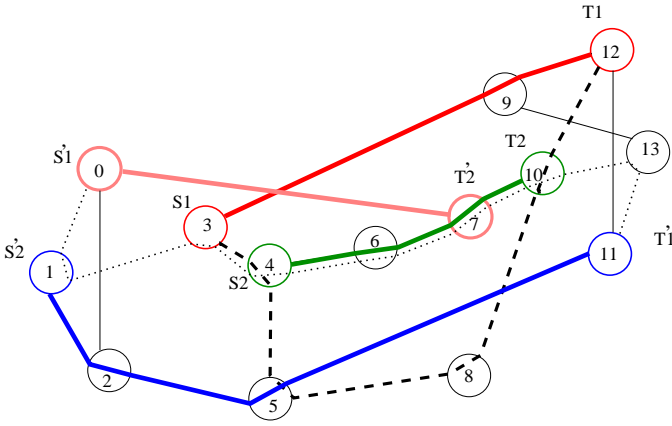


Fig. 2. An example of provisioning and protecting four connections on NSFNET.

It is to be noted that it may not be possible to protect all connections together, and therefore it would be necessary to partition the set of connections, and protect connections in each partition together. We illustrate this point using the example shown in Figure 2, where there are four connections (shown using bold lines) that are provisioned on NSFNET: $C_1 = (3, 12)$, $C_2 = (4, 10)$, $C_3 = (0, 7)$ and $C_4 = (1, 11)$. It is not possible to protect all four connections together using one protection path that is link disjoint from all four connections. Therefore, in this example, we use two protection paths: one protection path (3,4,5,8,10,12) protecting C_1 and C_2 , and is shown in dashed lines; and another protection path (0,1,3,4,6,7,10,13,11) protecting C_3 and C_4 , and is shown in dotted lines. Notice that all connections that are protected together, and their protection path are link disjoint. The end nodes in C_1 and C_2 are labeled S_1 , S_2 , T_1 and T_2 , while the end nodes in C_3 and C_4 are labeled S'_1 , S'_2 , T'_1 and T'_2 , respectively. In the above example, it is assumed that each connection is established at an electronic layer, i.e., an overlay layer above the physical layer. For example, the working path of a connection can be routed and established as an MPLS Label Switched Path (LSP), which can be explicitly routed in the network, as shown in the figure, and therefore the paths of the connections which are jointly protected, e.g., C_1 and C_2 in the above example, can be made link disjoint. However, when it comes to the protection path, since the data units transmitted on this path need to be processed, the protection path can be provisioned as segments, where each segment is an MPLS LSP which is explicitly routed. For the example of Figure 2, the protection path protecting connections C_1 and C_2 can be provisioned as three MPLS LSPs, namely, (3,4), (4,5,8,10) and (10,12).

B. Encoding Operations on \mathbf{S} and \mathbf{T}

The network encoding operation is executed by each node in \mathbf{S} and \mathbf{T} . To facilitate the specification of the encoding protocol we first define the following.

- $T(S_i)$: node in \mathbf{T} transmitting to and receiving from S_i , e.g. in Fig.1, $T(S_1) = T_2$.
- $S(T_j)$: node in \mathbf{S} transmitting to and receiving from T_j .
- $\sigma(S_i)/\sigma(T_j)$: the next node downstream from S_i (respectively T_j) on \mathbf{S} , e.g., in Fig.1, $\sigma(S_2) = S_3$.
- $\sigma^{-1}(S_i)/\sigma^{-1}(T_j)$: the next node upstream from S_i (respectively T_j) on \mathbf{S} , e.g., in Fig.1, $\sigma^{-1}(T_5) = S_4$.
- $\tau(S_i)/\tau(T_j)$: the next node downstream from S_i (respectively T_j) on \mathbf{T} , e.g., in Fig. 1, $\tau(T_4) = S_5$.
- $\tau^{-1}(S_i)/\tau^{-1}(T_j)$: the next node upstream from S_i (respectively T_j) on \mathbf{T} , e.g., in Fig.1, $\tau^{-1}(S_5) = T_4$.

We denote the data unit transmitted on link $e \in \mathbf{S}$ by y_e and the data unit transmitted on link $e \in \mathbf{T}$ by z_e . Assume that nodes S_i and T_j are in the same connection. The encoding operations work as follows, where all data units belong to the same round.

- 1) *Encoding operations at S_i .* The node S_i has access to data units d_i (that it generated) and data unit \hat{u}_j received on the primary path from T_j .

- a) It computes $y_{\sigma^{-1}(S_i) \rightarrow S_i} + (d_i + \hat{u}_j)$ and sends it on the link $S_i \rightarrow \sigma(S_i)$; i.e.

$$y_{S_i \rightarrow \sigma(S_i)} = y_{\sigma^{-1}(S_i) \rightarrow S_i} + (d_i + \hat{u}_j).$$

- b) It computes $z_{\tau^{-1}(S_i) \rightarrow S_i} + (d_i + \hat{u}_j)$ and sends it on the link $S_i \rightarrow \tau(S_i)$; i.e.

$$z_{S_i \rightarrow \tau(S_i)} = z_{\tau^{-1}(S_i) \rightarrow S_i} + (d_i + \hat{u}_j).$$

- 2) *Encoding operations at T_j .* The node T_j has access to data units u_j (that it generated) and data unit \hat{d}_i received on the primary path from S_i .

- a) It computes $y_{\sigma^{-1}(T_j) \rightarrow T_j} + (\hat{d}_i + u_j)$ and sends it on the link $T_j \rightarrow \sigma(T_j)$; i.e.

$$y_{T_j \rightarrow \sigma(T_j)} = y_{\sigma^{-1}(T_j) \rightarrow T_j} + (\hat{d}_i + u_j)$$

- b) It computes $z_{\tau^{-1}(T_j) \rightarrow T_j} + (\hat{d}_i + u_j)$ and sends it on the link $T_j \rightarrow \tau(T_j)$; i.e.

$$z_{T_j \rightarrow \tau(T_j)} = z_{\tau^{-1}(T_j) \rightarrow T_j} + (\hat{d}_i + u_j)$$

An example in which three nodes perform this procedure in the absence of failures is shown in Figure 3.

Consider $S' \subseteq \mathbf{S}$ and let $\mathcal{N}(S')$ represent the subset of nodes in \mathbf{T} that have a primary path connection to the nodes in S' (similar notation shall be used for a subset $T' \subseteq \mathbf{T}$). Let $D_{\mathbf{S}}(S_i)$ and $U_{\mathbf{S}}(S_i)$ represent the set of downstream and upstream nodes of S_i on the protection path \mathbf{S} (similar notation shall be used for the protection path \mathbf{T}). When all nodes in \mathbf{S} and \mathbf{T} have performed their encoding operations, the signals received at a node S_i on the \mathbf{S} and \mathbf{T} paths, respectively, are

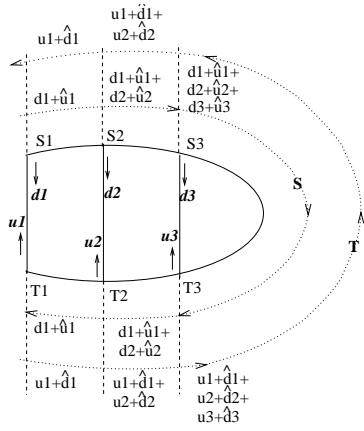


Fig. 3. Example of three nodes performing the encoding procedure. Note that the addition (bitwise XOR) of two copies of the same data unit, e.g., d_i and \hat{d}_i , removes both of them.

as follows

$$\begin{aligned}
& y_{\sigma^{-1}(S_i) \rightarrow S_i} \\
&= \underbrace{\sum_{\{k: S_k \in U_{\mathbf{S}}(S_i) \cap \mathbf{S}\}} d_k + \sum_{\{k: T_k \in \mathcal{N}(U_{\mathbf{S}}(S_i) \cap \mathbf{S})\}} \hat{u}_k}_{\text{From nodes upstream of } S_i \text{ on } \mathbf{S} \text{ in } \mathcal{S}} \\
&+ \underbrace{\sum_{\{k: T_k \in U_{\mathbf{S}}(S_i) \cap \mathbf{T}\}} u_k + \sum_{\{k: S_k \in \mathcal{N}(U_{\mathbf{S}}(S_i) \cap \mathbf{T})\}} \hat{d}_k}_{\text{From nodes upstream of } S_i \text{ on } \mathbf{S} \text{ in } \mathcal{T}}, \text{ and} \quad (3)
\end{aligned}$$

$$\begin{aligned}
& z_{\tau^{-1}(S_i) \rightarrow S_i} \\
&= \underbrace{\sum_{\{k: S_k \in U_{\mathbf{T}}(S_i) \cap \mathbf{S}\}} d_k + \sum_{\{k: T_k \in \mathcal{N}(U_{\mathbf{T}}(S_i) \cap \mathbf{S})\}} \hat{u}_k}_{\text{From nodes upstream of } S_i \text{ on } \mathbf{T} \text{ in } \mathcal{S}} \\
&+ \underbrace{\sum_{\{k: T_k \in U_{\mathbf{T}}(S_i) \cap \mathbf{T}\}} u_k + \sum_{\{k: S_k \in \mathcal{N}(U_{\mathbf{T}}(S_i) \cap \mathbf{T})\}} \hat{d}_k}_{\text{From nodes upstream of } S_i \text{ on } \mathbf{T} \text{ in } \mathcal{T}} \quad (4)
\end{aligned}$$

Similar equations can be derived for node T_j .

C. Recovery from failures

The encoding operations described in Subsection III-B allow the recovery of a second copy of the same data unit transmitted on the working circuit, hence protecting against single link failures. To illustrate this, suppose that the primary path between nodes S_i and T_j fails. In this case, S_i does not receive u_j on the primary path, and it receives $\hat{u}_j = 0$ instead. Moreover, $\hat{d}_i = 0$. However, S_i can recover u_j by adding equations (3) and (4). In particular node S_i computes

$$\begin{aligned}
y_{\sigma^{-1}(S_i) \rightarrow S_i} + z_{\tau^{-1}(S_i) \rightarrow S_i} &= \sum_{\{k: S_k \in \mathcal{S} \setminus \{S_i\}\}} d_k + \sum_{\{k: T_k \in \mathcal{T}\}} u_k \\
&+ \sum_{\{k: T_k \in \mathcal{T} \setminus \{T_j\}\}} \hat{u}_k + \sum_{\{k: S_k \in \mathcal{S}\}} \hat{d}_k \\
&= \hat{d}_i + u_j \\
&= u_j \quad (\text{since } \hat{d}_i = 0.) \quad (5)
\end{aligned}$$

Similarly, T_j can recover d_i by adding the values it obtains over \mathbf{S} and \mathbf{T} . For example, if the working path between S_2 and T_2 in Figure 3 fails, then at node S_2 adding the signal received on \mathbf{S} to the signal received on \mathbf{T} , then u_2 can be recovered, since T_2 generated u_2 . Also, node T_2 adds the signals on \mathbf{S} and \mathbf{T} to recover d_2 .

Notice that the reception of a second copy of u_2 and d_2 at S_2 and T_2 , respectively, when there are no failures, requires the addition of the d_2 and u_2 signals generated by the same nodes, respectively.

As a more general example, consider the case in Figure 1. Node S_5 , for example, will receive the following signal on \mathbf{S} :

$$(d_1 + \hat{u}_2) + (d_2 + \hat{u}_5) + (d_3 + \hat{u}_1) + (d_4 + \hat{u}_4) + (u_5 + \hat{d}_2), \quad (6)$$

and will receive the following on \mathbf{T} :

$$(u_1 + \hat{d}_3) + (u_2 + \hat{d}_1) + (u_3 + \hat{d}_5) + (u_4 + \hat{d}_4). \quad (7)$$

If the link between S_5 and T_3 fails, then $\hat{d}_5 = 0$, and adding equations (6) and (7) will recover u_3 at S_5 .

IV. IMPLEMENTATION ISSUES

In this subsection we address a number of practical implementation issues.

A. Round Numbers

Since linear combinations include packets belonging to the same round number, the packet header should include a round number field. The field is initially reset to zero, and is updated independently by each node when it generates and sends a new packet on the working circuit. Note that there will be a delay before the linear combination propagating on \mathbf{S} and \mathbf{T} reaches a given node. For example, in Figure 3 assuming that all nodes started transmission at time 0, node S_3 shall receive the combination corresponding to round 0 over \mathbf{S} , $d_1(0) + \hat{u}_1(0) + d_2(0) + \hat{u}_2(0)$ after a delay corresponding to the propagation delay between nodes S_1 and S_3 , in addition to the processing and transmission times at nodes S_1 and S_2 . However since the received data unit shall contain the round number 0, it shall be combined with the data unit generated by S_3 at time slot 0.

The size of the round number field depends on the delay of the protection path, including processing and transmission times, as well as propagation time, and the working circuit delay. It is reasonable to assume that the delay of any working circuit is shorter than that of the protection circuit; otherwise, the protection path could have been used as a working path. Thus, when a data unit on the protection path corresponding to a particular round number reaches a given node, the data unit of that round number would have already been received on the primary path of the node.

In this case, it is straightforward to see that once a data unit is transmitted on the working circuit, then it will take no more than twice the delay of the protection path to recover the backup copy of this data unit by the receiver. Therefore, round numbers can then be reused. Based on this argument,

the size of the set of required unique round numbers is upper bounded by $2a$, where

$$a = \lceil \frac{\chi_{\mathbf{P}}}{(\text{Protection data unit size in bits})/B} \rceil. \quad (8)$$

$\chi_{\mathbf{P}}$ in the above equation is the delay over the protection circuit, and B is the transport capacity of the protection circuit, which, as stated in Section II-B, is taken as the maximum over all the transport capacities of the protected connections. A sufficiently long round number field will require no more than $\log_2(2a)$ bits.

B. Synchronization

An important issue is node synchronization to rounds. This can be achieved using a number of strategies. A simple strategy for initialization and synchronization is the following:

- In addition to buffers used to store transmitted and received data units, each node $S_i \in \mathcal{S}$ has two buffers, $F_{\mathbf{S}}(S_i)$ and $F_{\mathbf{T}}(S_i)$, which are used for transmissions on the \mathbf{S} and \mathbf{T} paths, respectively. Node $T_j \in \mathcal{T}$ also has similar buffers, $F_{\mathbf{S}}(T_j)$ and $F_{\mathbf{T}}(T_j)$.
- Node S_1 starts the transmission of $d_1(0)$ on the working circuit to $T(S_1)$. When S_1 receives $\hat{u}_{T(S_1)}(0)$, it forms $d_1(0) + \hat{u}_{T(S_1)}(0)$ and transmits it on the outgoing link in \mathbf{S} . Similarly, node T_1 will transmit $u_1(0)$ on the working circuit, and $u_1(0) + \hat{d}_{S(T_1)}(0)$ on the outgoing link in \mathbf{T} .
- Node S_i , for $i > 0$, will buffer the combinations received on \mathbf{S} in $F_{\mathbf{S}}(S_i)$. Assume that the combination with the smallest round number buffered in $F_{\mathbf{S}}(S_i)$ (i.e., head of buffer) corresponds to round number n . When S_i transmits $d_i(n)$ and receives $\hat{u}_{T(S_i)}(n)$, then it adds those data units to the combination with the smallest round number in $F_{\mathbf{S}}(S_i)$ and transmits the combination on \mathbf{S} . The combination with round number n is then purged from $F_{\mathbf{S}}(S_i)$. Similar operations are performed on $F_{\mathbf{T}}(S_i)$, $F_{\mathbf{S}}(T_j)$ and $F_{\mathbf{T}}(T_j)$. Note that purging of the data unit from the buffer only implies that the combination corresponding to round n has been sent and should not be sent again. However node S_i needs to ensure that it saves the value of the data unit received on \mathbf{S} as long as needed for it to be able to decode $u_{T(S_i)}(n)$ if needed. An illustration of the use of those buffers is shown in Figure 4.

C. Buffer Size

Assuming that all nodes start transmitting simultaneously, then all nodes would have decoded the data units corresponding to a given round number in a time that does not exceed

$$\chi_{\mathbf{P}} + \max_{1 \leq w \leq N} \chi_w$$

where χ_w is the delay over working path w .

Based on this, the following upper bounds on buffer sizes can be established:

- The transmit buffer, as well as the $F_{\mathbf{S}}$ and $F_{\mathbf{T}}$ buffers are upper bounded by

$$\lceil \frac{\chi_{\mathbf{P}} + \max_{1 \leq w \leq N} \chi_w}{\text{Data unit size in bits}/B} \rceil.$$

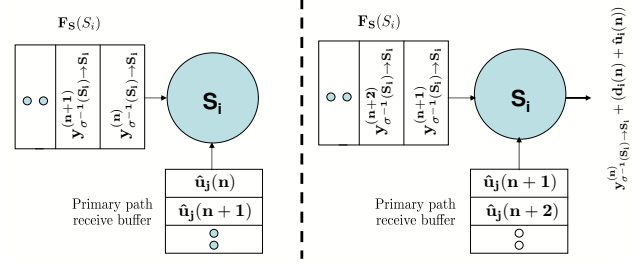


Fig. 4. An illustration of the use of node buffer $F_{\mathbf{S}}(S_i)$. (a) Shows the status of the buffers before data unit at round n has been processed. (b) Shows the status of the buffers after the data unit at round n has been processed. Note that the data units corresponding to round n have been purged from both $F_{\mathbf{S}}(S_i)$ and the primary path receive buffer. The operation of other buffers is similar.

This is because it will take χ_w units of time over the path w used by the connection $S(T_1) \leftrightarrow T_1$ to receive $\hat{d}_{S(T_1)}$, and then start transmission on the \mathbf{T} path. An additional $\chi_{\mathbf{P}}$ units of time is required for the first combination to reach S_1 . The numerator in the above equation is the maximum of this delay.

- The receive buffer is upper bounded by

$$\lceil \frac{\chi_{\mathbf{P}} + \max_{1 \leq w \leq N} \chi_w - \min_{1 \leq w \leq N} \chi_w}{\text{Data unit size in bits}/B} \rceil.$$

The numerator in the above equation is derived using arguments similar to the transmit buffer, except that for the first data unit to be received, it will have to encounter the delay over the working circuit; hence, the subtraction of the minimum such delay.

V. PROTECTION AGAINST MULTIPLE FAULTS

We now consider the situation when protection against multiple (more than one) link failures is required. In this case it is intuitively clear that a given primary path connection needs to be protected by multiple bi-directional protection paths. To see this we first analyze the sum of the signals received on \mathbf{S} and \mathbf{T} for a node S_i that has a connection to node T_j when the primary paths $S_i \leftrightarrow T_j$ and $S_{i'} \leftrightarrow T_{j'}$ protected by the same protection path are in failure. In this case we have $\hat{d}_i = \hat{d}_{i'} = \hat{u}_j = \hat{u}_{j'} = 0$. Therefore, at node S_i we have,

$$\begin{aligned} y_{\sigma^{-1}(S_i) \rightarrow S_i} + z_{\tau^{-1}(S_i) \rightarrow S_i} &= \sum_{\{k: S_k \in \mathcal{S} \setminus \{S_i\}\}} d_k + \sum_{\{k: T_k \in \mathcal{T}\}} u_k \\ &+ \sum_{\{k: T_k \in \mathcal{T} \setminus \{T_j\}\}} \hat{u}_k + \sum_{\{k: S_k \in \mathcal{S}\}} \hat{d}_k \\ &= (d_{i'} + u_{j'}) + u_j. \end{aligned}$$

Note that node S_i is only interested in the data unit u_j but it can only recover the sum of u_j and the term $(d_{i'} + u_{j'})$, in which it is not interested.

We now demonstrate that if a given connection is protected by multiple protection paths, a modification of the protocol presented in Section III-B can enable the nodes to recover from multiple failures. In the modified protocol a node multiplies the sum of its own data unit and the data unit received over

its primary path by an appropriately chosen scaling coefficient before adding it to the signals on the protection path. The scheme in Section III-B can be considered to be a special case of this protocol when the scaling coefficient is 1 (i.e., the identity element over $GF(2^m)$).

It is important to note that in contrast to the approach presented in [14], this protocol does not require any synchronization between the operation of the different protection paths.

As before, suppose that there are N bi-directional unicast connections that are to be protected against the failure of any M links, for $M \leq N$. These connections are now protected by K protection paths $\mathbf{P}_k, k = 1, \dots, K$. Protection path \mathbf{P}_k passes through all nodes $\mathcal{S}_k \subseteq \mathcal{S}$ and $\mathcal{T}_k \subseteq \mathcal{T}$ where the nodes in \mathcal{S}_k communicate bi-directionally with the nodes in \mathcal{T}_k . Note that $\cup_{k=1}^K \mathcal{S}_k = \mathcal{S}$ and $\cup_{k=1}^K \mathcal{T}_k = \mathcal{T}$. The ordered sets \mathcal{S}_k and \mathcal{S}_l are not necessarily disjoint for $l \neq k$, i.e., a primary path can be protected by different protection paths. However, if two protection paths are used to protect the same working connection, then they must be link disjoint.

A. Modified Encoding Operation

Assume that nodes S_i and T_j are protected by the protection path \mathbf{P}_k . The encoding operations performed by S_i and T_j for path \mathbf{P}_k are explained below (the operations for other protection paths are similar). In the presentation below we shall use the notation $\sigma(S_i), \sigma^{-1}(S_i), \tau(S_i), \tau^{-1}(S_i)$ to be defined implicitly over the protection path \mathbf{P}_k . Similar notation is used for T_j .

The nodes S_i and T_j initially agree on a value of the scaling coefficient denoted $\alpha_{i \leftrightarrow j, k} \in GF(2^m)$. The subscript $i \leftrightarrow j, k$ denotes that the scaling coefficient is used for connection S_i to T_j over protection path \mathbf{P}_k .

- 1) *Encoding operations at S_i .* The node S_i has access to data units d_i (that it generated) and data unit \hat{u}_j received on the primary path from T_j .

- a) It computes $y_{\sigma^{-1}(S_i) \rightarrow S_i} + \alpha_{i \leftrightarrow j, k}(d_i + \hat{u}_j)$ and sends it on the link $S_i \rightarrow \sigma(S_i)$; i.e.

$$y_{S_i \rightarrow \sigma(S_i)} = y_{\sigma^{-1}(S_i) \rightarrow S_i} + \alpha_{i \leftrightarrow j, k}(d_i + \hat{u}_j).$$

- b) It computes $z_{\tau^{-1}(S_i) \rightarrow S_i} + \alpha_{i \leftrightarrow j, k}(d_i + \hat{u}_j)$ and sends it on the link $S_i \rightarrow \tau(S_i)$; i.e.

$$z_{S_i \rightarrow \tau(S_i)} = z_{\tau^{-1}(S_i) \rightarrow S_i} + \alpha_{i \leftrightarrow j, k}(d_i + \hat{u}_j).$$

- 2) *Encoding operations at T_j .* The node T_j has access to data units u_j (that it generated) and data unit \hat{d}_i received on the primary path from S_i .

- a) It computes $y_{\sigma^{-1}(T_j) \rightarrow T_j} + \alpha_{i \leftrightarrow j, k}(\hat{d}_i + u_j)$ and sends it on the link $T_j \rightarrow \sigma(T_j)$; i.e.

$$y_{T_j \rightarrow \sigma(T_j)} = y_{\sigma^{-1}(T_j) \rightarrow T_j} + \alpha_{i \leftrightarrow j, k}(\hat{d}_i + u_j)$$

- b) It computes $z_{\tau^{-1}(T_j) \rightarrow T_j} + \alpha_{i \leftrightarrow j, k}(\hat{d}_i + u_j)$ and sends it on the link $T_j \rightarrow \tau(T_j)$; i.e.

$$z_{T_j \rightarrow \tau(T_j)} = z_{\tau^{-1}(T_j) \rightarrow T_j} + \alpha_{i \leftrightarrow j, k}(\hat{d}_i + u_j)$$

It should be clear that we can find expressions similar to the ones in (3) and (4) in this case as well.

B. Recovery from failures

Suppose that the primary paths $S_i \leftrightarrow T_j$ and $S_{i'} \leftrightarrow T_{j'}$ fail, and they are both protected by \mathbf{P}_k . Consider the sum of the signals received by node S_i over \mathbf{S}_k and \mathbf{T}_k . Similar to our discussion in III-C, we can observe that

$$y_{\sigma^{-1}(S_i) \rightarrow S_i} + z_{\tau^{-1}(S_i) \rightarrow S_i} = \alpha_{i' \leftrightarrow j', k}(d_{i'} + u_{j'}) + \alpha_{i \leftrightarrow j, k}u_j$$

Note that the structure of the equation allows the node S_i to treat $(d_{i'} + u_{j'})$ as a single unknown. Thus from protection path \mathbf{P}_k , node S_i obtains one equation in two variables. Now, if there exists another protection path \mathbf{P}_l that also protects the connections $S_i \leftrightarrow T_j$ and $S_{i'} \leftrightarrow T_{j'}$, then we can obtain the following system of equations in two variables

$$\begin{bmatrix} \alpha_{i' \leftrightarrow j', k} & \alpha_{i \leftrightarrow j, k} \\ \alpha_{i' \leftrightarrow j', l} & \alpha_{i \leftrightarrow j, l} \end{bmatrix} \begin{bmatrix} (d_{i'} + u_{j'}) \\ u_j \end{bmatrix} = \begin{bmatrix} x_{S_i}^k \\ x_{S_i}^l \end{bmatrix}, \quad (9)$$

where $x_{S_i}^k$ and $x_{S_i}^l$ represent values that can be obtained at S_i and therefore u_j can be recovered by solving the system of equations. The choice of the scaling coefficients needs to be such that the associated 2×2 matrix in (9) is invertible. This can be guaranteed by a careful assignment of the scaling coefficients. More generally we shall need to ensure that a large number of such matrices need to be full-rank. By choosing the operating field size $GF(2^m)$ to be large enough, i.e., m to be large enough we can ensure that such an assignment of scaling coefficients always exists [24]. The detailed discussion of coefficient assignment can be found in Section VI.

C. Conditions for Data Recovery:

We shall first discuss the conditions for data recovery under a certain failure pattern. To facilitate the discussion on determining which failures can be recovered from, we represent the failed connections, and the protection paths using a bipartite graph, $G_{DR}(V, E)$, where the set of vertices $V = \mathbb{N} \cup \mathbb{P}$, and the set of edges $E \subseteq \mathbb{N} \times \mathbb{P}$ where \mathbb{N} is the set of connections to be protected, and \mathbb{P} is the set of protection paths. There is an edge from connection $N_i \in \mathbb{N}$ to protection path $\mathbf{P}_k \in \mathbb{P}$ if \mathbf{P}_k protects connection N_i . In addition, each edge has a label that is assigned as follows. Suppose that there exists an edge between N_i (between nodes $S_{i'}$ and $T_{j'}$) and \mathbf{P}_k . The label on the edge is given by the scaling coefficient $\alpha_{i' \leftrightarrow j', k}$.

Note that in general one could have link failures on primary paths as well as protection paths. Suppose that a failure pattern is specified as a set $F = \{N_{i_1}, \dots, N_{i_n}\} \cup \{\mathbf{P}_{j_1}, \dots, \mathbf{P}_{j_{n'}}\}$ where $\{N_{i_1}, \dots, N_{i_n}\}$ denotes the set of primary paths that have failed and $\{\mathbf{P}_{j_1}, \dots, \mathbf{P}_{j_{n'}}\}$ denotes the set of protection paths that have failed. The determination of whether a given node can recover from the failures in F can be performed in the following manner.

- 1) *Initialization.* Form the graph $G_{DR}(V, E)$ as explained above.
- 2) *Edge pruning.*

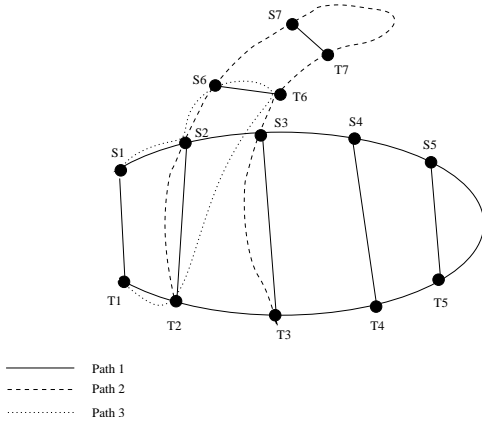


Fig. 5. An example of a network protected against multiple faults.

- a) For all connections $N_i \in \mathbb{N} \setminus F$ remove N_i and all edges in which it participates from G_{DR} .
- b) For all protection paths $\mathbf{P}_i \in F$ remove \mathbf{P}_i and all edges in which it participates from G_{DR} .
- 3) *Checking the system of equations.* Let the residual graph be denoted $G'_{DR} = (\mathbb{N}' \cup \mathbb{P}', \mathbb{E}')$. For each connection $N_i \in \mathbb{N}'$, do the following steps.
 - a) Let the subset of nodes in \mathbb{P}' that have a connection to N_i be denoted $\mathcal{N}(N_i)$. Each node in $\mathcal{N}(N_i)$ corresponds to a linear equation that is available to the nodes participating in N_i . The linear combination coefficients are determined by the labels of the edges. Identify this system of equations.
 - b) Check to see whether a node in N_i can solve this system of equations to obtain the data unit it is interested in.

In Figure 6 we show an example that applies to the network in Figure 5. Figure 6.(a) shows the bipartite graph for the entire network, while Figures 6.(b) and 6.(c) show the graph corresponding to the following two failing patterns, respectively:

- (S_2, T_2) , (S_6, T_6) and (S_5, T_5)
- \mathbf{P}_2 , (S_2, T_2) and (S_6, T_6)

Let us assume that the encoding coefficients are chosen to make sure the equation obtained by each node has unique solution. From Figure 6.(b), the failures of connections (S_2, T_2) and (S_6, T_6) can be recovered from because each node obtains two equations in two unknowns. More specifically, at node S_2 we obtain the following system of equations (the equation from \mathbf{P}_1 is not used).

$$\begin{bmatrix} \alpha_{2 \leftrightarrow 2,2} & \alpha_{6 \leftrightarrow 6,2} \\ \alpha_{2 \leftrightarrow 2,3} & \alpha_{6 \leftrightarrow 6,3} \end{bmatrix} \begin{bmatrix} u_2 \\ (d_6 + u_6) \end{bmatrix} = \begin{bmatrix} x_{S_2}^2 \\ x_{S_2}^3 \end{bmatrix},$$

which has a unique solution if $(\alpha_{2 \leftrightarrow 2,2}\alpha_{6 \leftrightarrow 6,3} - \alpha_{2 \leftrightarrow 2,3}\alpha_{6 \leftrightarrow 6,2}) \neq 0$. As pointed out in Section V-B, the choice of the scaling coefficients can be made so that all possible matrices involved have full rank by working over a large enough field size. Thus in this case S_2 and T_2 can recover from the failures. By a similar argument we can

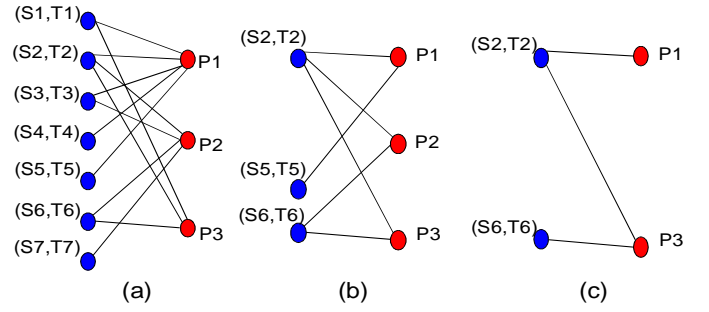


Fig. 6. Applying the bipartite graph representation verify if failures will be recovered.

observe that S_6 and T_6 can also recover from the failures by using the equations from \mathbf{P}_2 and \mathbf{P}_3 . However, S_5 and T_5 cannot recover from the failure since they can only obtain one equation from \mathbf{P}_1 in two variables that corresponds to failures on (S_2, T_2) and (S_5, T_5) . In Figure 6.(c), path \mathbf{P}_2 does not exist, and (S_6, T_6) is protected only by path \mathbf{P}_3 , which protects two failed connections. Therefore, it cannot recover from the failure. However, (S_2, T_2) can still recover its data units by using path \mathbf{P}_1 .

In general, this procedure needs to be performed for every possible failure pattern that needs to be protected against, for checking whether all nodes can still recover the data unit that they are interested in. However, usually the set of failure patterns to be protected against is the set of all single link failures or more generally the set of all possible $M \geq 1$ link failures. Those M link failures can happen anywhere, on primary paths or protection paths.

Next, we consider general conditions for data recovery. First, we describe the general model for multiple failures. In order to make expressions simple, we assume that the data unit obtained by a node of a failed connection, say S_i , from protection path \mathbf{P}_k is the sum of the data units from S_k , T_k . Adding up with $\alpha_{i \leftrightarrow j,k} d_i$, which is the data units generated at node S_i , we denote this sum by p_k where $p_k = y_{\sigma^{-1}(S_i) \rightarrow S_i} + z_{\tau^{-1}(S_i) \rightarrow S_i} + \alpha_{i \leftrightarrow j,k} d_i$. Note that d_i is the local data units, which is always available. In this case, each node on one protection path \mathbf{P}_k obtains the same equation in terms of the same variables. By denoting the set of failed primary connections protected by \mathbf{P}_k as $F(\mathbf{P}_k)$, the equation for this protection path \mathbf{P}_k is

$$\sum_{(S_i \leftrightarrow T_j) \in F(\mathbf{P}_k)} \alpha_{i \leftrightarrow j,k} (d_i + u_j) = p_k. \quad (10)$$

In equation (10), each $d_i + u_j$ is considered as one variable and the coefficients assigned to d_i and u_j are the same. Each node of a failed connection will obtain one equation from each intact protection path that protects it and consequently forms a system of linear equations. The number of equations that node S_i obtains is the number of intact protection paths that protect S_i . The number of variables is the total number of failed connections protected by the protection paths that also provide protection to the failed connection between S_i and T_j .

S_i needs to solve the system of equations and obtain $d_i + u_j$. By subtracting d_i , it can get u_j , which is the data unit S_i wants to receive while T_j can retrieve the data d_i by subtracting u_j from $d_i + u_j$.

Each protection path maps to an equation in terms of a number of variables representing the combination of the data units generated at two end nodes of the failed connections protected by this path. We can form a system of equations that consists of at most K equations like equation (10) where K is the total number of protection paths. Each failure of a primary path introduces a variable whereas each failure occurring on a protection path erases the corresponding equation from the matrix. In general, the system of equations that a node obtains also depends on the topology. If all of the connections are not protected by the same protection paths, there are zeros in the coefficient matrix because a failed connection is not protected by all protection paths, implying that some variables will not appear in all equations.

In order to recover from any failure pattern of M failures, we require the following necessary conditions.

Theorem 1: In order for the network to be guaranteed protection against any M link failures, the following necessary conditions should be satisfied.

- 1) Each node should be protected by at least M link-disjoint protection paths.
- 2) Under any failure pattern with M failures, a subset of equations that each node obtains should have a unique solution.

proof: The first condition can be shown by contradiction. If a node is protected by $M - 1$ protection paths, the failure could happen on these $M - 1$ protection paths and on the primary path in which this node participates. Then, this node does not have any protection path to recover from its primary path failure.

The second condition is to ensure that each node can recover the data unit under any failure pattern with M failures. Note that for necessary condition, we don't require that the whole system of equations each node obtains has unique solution because one node is only interested in recovering the data unit sent to it. As long as it can solve a subset of the equations, it recovers from its failure. ■

We emphasize that the structure of the equations depends heavily on the network topology, the connections provisioned and the protection paths. Therefore it is hard to state a more specific result about the conditions under which protection is guaranteed. However, under certain structured topologies it may be possible to provide a characterization of the conditions that can be checked without having to verify each possible system of equations.

For example, if all connections are protected by M protection paths, it is easy to see the sufficient condition for data recovery from any M failures is that the coefficient matrix of the system of equations each node obtains under any failure pattern with M failures has full rank. As will be shown next, our coefficient assignment methods are such that the sufficient conditions above hold.

Next we construct a $K \times N$ matrix to facilitate the discussion of coefficient assignment. According to the encoding protocol, each connection $S_i - T_j$ has coefficient $\alpha_{i \leftrightarrow j, k}$ for encoding on \mathbf{P}_k . In general, there are at most $K \times N$ coefficients for a network with N primary paths $S_{i_1} \leftrightarrow T_{j_1}, S_{i_2} \leftrightarrow T_{j_2}, \dots, S_{i_l} \leftrightarrow T_{j_l}, \dots, S_{i_N} \leftrightarrow T_{j_N}$ and K protection paths $\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_K$. We form a $K \times N$ matrix \mathcal{A} where $\mathcal{A}_{kl} = \alpha_{i_l \leftrightarrow j_l, k}$ if $S_{i_l} \leftrightarrow T_{j_l}$ is protected by \mathbf{P}_k , $\mathcal{A}_{kl} = 0$ otherwise. Here, l is the index for primary paths and each column of \mathcal{A} corresponds to a primary path. Each row of \mathcal{A} corresponds to a protection path. This matrix contains all encoding coefficients and some zeros induced by the topology in general. It is easy to see that under any failure pattern, the coefficient matrix of the system of equations at any node of any failed connection is a submatrix of matrix \mathcal{A} . We require these submatrices of \mathcal{A} to have full rank. We shall discuss the construction of \mathcal{A} , i.e., assign proper coefficients in Section VI.

VI. ENCODING COEFFICIENT ASSIGNMENT

In this section, we shall discuss encoding coefficient assignment strategies for the proposed network coding schemes, i.e., construct \mathcal{A} properly. Under certain assumptions on the topology, two special matrix based assignments can provide tight field size bound and efficient decoding algorithms. We shall also introduce matrix completion method for general topologies.

Note that the coefficient assignment is done before the actual transmission. Once the coefficients have been determined, during data transmission they need not be changed. Thus, for the schemes that guarantee successful recovery with high probability, we can keep generating the matrix \mathcal{A} until the full rank condition discussed at the end of the previous section satisfies. This only needs to be done once. After that, during the actual transmission, the recovery is successful for sure.

A. Special matrix based assignment

In this and the next subsection, we assume that all primary paths are protected by the same protection paths. This implies that matrix \mathcal{A} only consists of encoding coefficients. It does not contain zeros induced by the topology. Thus, we can let \mathcal{A} to be a matrix with some special structures such that any submatrix of \mathcal{A} has full rank. The network will be able to recover from any failure pattern with M (or less) failures. Without loss of generality, we shall focus on the case when $M = K$, where K is the number of protection paths. If M failures happen, in which t_1 failures happen on primary paths, each node will get $M - (M - t_1) = t_1$ equations with t_1 unknowns corresponding to t_1 primary path failures. The $t_1 \times t_1$ coefficient matrix is a square submatrix of \mathcal{A} and they are the same for each node under one failure pattern.

First, we shall show a Vandermonde matrix-based coefficient assignment. It requires the field size to be $q \geq N$. If all failures happen on primary paths, the recovery at each node is guaranteed. In this assignment strategy, we pick up N distinct elements from $GF(q)$: $\lambda_1, \dots, \lambda_N$ and assign them to each primary paths. At nodes S_{i_l} and T_{j_l} , λ_l^{k-1}

is used as encoding coefficient on protection path \mathbf{P}_k , i.e., $\mathcal{A}_{kl} = \alpha_{i_l \leftrightarrow j_l, k} = \lambda_l^{k-1}$. In other words, \mathcal{A} is a Vandermonde matrix [26, Section 6.1]:

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_N \\ \lambda_1^2 & \lambda_2^2 & \cdots & \lambda_N^2 \\ \cdots & \cdots & \cdots & \cdots \\ \lambda_1^{K-1} & \lambda_2^{K-1} & \cdots & \lambda_N^{K-1} \end{bmatrix}.$$

Suppose M failures happen on primary paths, the indices of failed connections are e_1, \dots, e_M , every node gets a system of linear equations with coefficient matrix having this form:

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \lambda_{e_1} & \lambda_{e_2} & \cdots & \lambda_{e_M} \\ \lambda_{e_1}^2 & \lambda_{e_2}^2 & \cdots & \lambda_{e_M}^2 \\ \cdots & \cdots & \cdots & \cdots \\ \lambda_{e_1}^{M-1} & \lambda_{e_2}^{M-1} & \cdots & \lambda_{e_M}^{M-1} \end{bmatrix}.$$

This matrix is a $M \times M$ Vandermonde matrix. As long as $\lambda_{e_1}, \lambda_{e_2}, \dots, \lambda_{e_M}$ are distinct, this matrix is invertible and $S_{i_{e_1}}$ can recover $u_{j_{e_1}}$. We choose $\lambda_1, \dots, \lambda_N$ to be distinct so that the submatrix formed by any M columns of \mathcal{A} has full rank. The smallest field size we need is the number of connections we want to protect, i.e., $q \geq N$. Moreover, the complexity of solving linear equation with Vandermonde coefficient matrix is $O(M^2)$ [19]. Thus, we have a more efficient decoding because if the coefficients are arbitrarily chosen, even if it is solvable, the complexity of Gaussian elimination is $O(M^3)$.

If $M - t_1$ failures happen on protection paths, we require that any $t_1 \times t_1$ square submatrix formed by choosing any t_1 columns and t_1 rows from \mathcal{A} has full rank. Although the chance is large, the Vandermonde matrix can not guarantee this for sure [20, p.323, problem 7],[22],[23]. We shall propose another special matrix to guarantee that for combined failures, the recovery is successful at the expense of a slightly larger field size compared to Vandermonde matrix assignment.

In order to achieve this goal, we resort to Cauchy matrix [20], of which any square submatrix has full rank if the entries are chosen carefully.

Definition 2: Let $\{x_1, \dots, x_{m_1}\}, \{y_1, \dots, y_{m_2}\}$ be two sets of elements in a field F such that

- (i) $x_i + y_j \neq 0, \forall i \in \{1, \dots, m_1\} \forall j \in \{1, \dots, m_2\}$;
- (ii) $\forall i, j \in \{1, \dots, m_1\}, i \neq j : x_i \neq x_j$ and $\forall i, j \in \{1, \dots, m_2\}, i \neq j : y_i \neq y_j$.

The matrix $C = (c_{ij})$ where $c_{ij} = 1/(x_i + y_j)$ is called a Cauchy matrix.

If $m_1 = m_2$, the Cauchy matrix becomes square and its determinant is [20]:

$$\det(C) = \frac{\prod_{1 \leq i < j \leq m_1} (x_j - x_i)(y_j - y_i)}{\prod_{1 \leq i, j \leq m_1} (x_i + y_j)}$$

Note that in $GF(q)$ where q is some power of 2, the addition and subtraction are equivalent. Therefore, as long as $x_1, \dots, x_{m_1}, y_1, \dots, y_{m_1}$ are distinct, Cauchy matrix has full rank and its any square submatrix is also a Cauchy matrix (by definition) with full rank. For our protection

problem, we let matrix \mathcal{A} to be a $K \times N$ Cauchy matrix. $\{x_1, \dots, x_K\}, \{y_1, \dots, y_N\}$ are chosen to be distinct. Thus, the smallest field size we need is $K + N$. Suppose there are t_1 failures on primary paths and $M - t_1$ failures on protection paths, the coefficient matrix of the system of equations obtained by a node is a $t_1 \times t_1$ submatrix of \mathcal{A} . It is still a Cauchy matrix by definition and invertible. Thus, the network can be recovered from any M failures. Moreover, the inversion can be done in $O(t_1^2)$ [21], which provides an efficient decoding algorithm.

B. Random assignment

We could also choose the coefficients from a large finite field. More specifically, we have the following claim [27].

Claim 3: When all coefficients are randomly, independently and uniformly chosen from $GF(q)$, the probability that a t_1 -by- t_1 matrix has full rank is $p(t_1) = \prod_{i=1}^{t_1} (1 - 1/q^i)$, $1 \leq t_1 \leq M$.

Under one failure pattern with t_1 failures on the primary paths and $M - t_1$ failures on the protection paths, every failed connection obtains the equations that have the same t_1 -by- t_1 coefficient matrix. The probability that it is full rank is $p(t_1)$ and it goes to 1 when q is large. Note that there are $\sum_{t_1=1}^M \binom{N}{t_1} \binom{M}{M-t_1}$ possible failure patterns when the total number of failures is M . Thus, by union bound, the probability of successful recovery under any failure pattern with M failures is $1 - \sum_{t_1=1}^M \binom{N}{t_1} \binom{M}{M-t_1} (1 - p(t_1))$, and it approaches 1 as q increases.

C. Matrix completion for general topology

If the primary paths are protected by different protection paths, like in Figure 5, there are some zeros in \mathcal{A} induced by the topology. We want to choose encoding coefficients so that under every failure pattern with M or less failures, the coefficient matrix of the system of equations obtained by every node is invertible. We can view the encoding coefficients in \mathcal{A} as indeterminates to be decided. The matrices we require to have full rank are a collection $\mathcal{C}_{\mathcal{A}}$ of submatrices of \mathcal{A} , where $\mathcal{C}_{\mathcal{A}}$ depends on the failure patterns and the network topology. Each matrix in $\mathcal{C}_{\mathcal{A}}$ consists of some indeterminates and some zeros. The problem of choosing encoding coefficients can be solved by matrix completion [24]. A simultaneous max-rank completion of $\mathcal{C}_{\mathcal{A}}$ is an assignment of values from $GF(q)$ to the indeterminates that preserves the rank of all matrices in $\mathcal{C}_{\mathcal{A}}$. After completion, each matrix will have the maximum possible rank. Matrix completion can be done by deterministic algorithms [24]. Moreover, simply choosing a completion at random from a sufficiently large field can achieve the maximum rank with high probability [25]. Hence, we can choose encoding coefficients randomly from a large field.

VII. ILP FORMULATION FOR SINGLE-LINK FAILURE

The problem of provisioning the working paths and their protection paths in a random graph is a hard problem. This is due to the fact that the problem of finding link disjoint paths between multiple pairs of nodes in a graph is known to

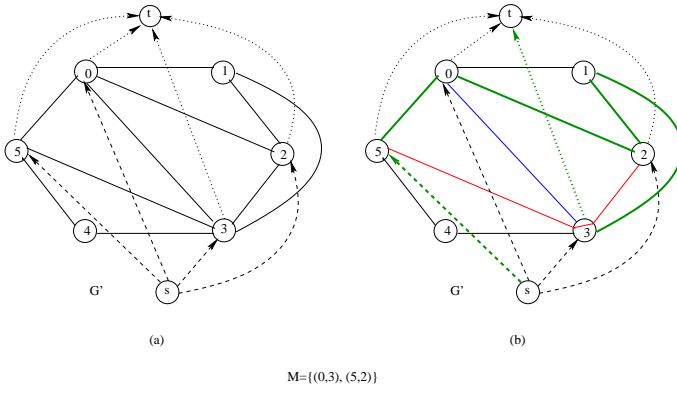


Fig. 7. An example to show: (a) the graph G in solid line and its modified graph G' ; (b) the provisioning of the connections $\{(0-3), (5-3-2)\}$ and their protection path $(s-5-0-2-1-3-t)$, where the two links $(s-5)$ and $(3-t)$ are not included in the cost of the protection circuit.

be NP-complete [17]. Therefore, in this section we formulate an integer linear program that optimally provisions a set of unicast connections, and their protection paths against single-link failure. The optimality criterion is the minimization of the sum of the working and protection resources.

The problem can be stated as follows: *Given an bidirectional graph $G = (V, E)$ and a traffic demand matrix of unicast connections, \mathbb{N} , establish a connection for each bidirectional traffic request $j \in \mathbb{N}$, and a number of protection paths that travel all the end nodes of the connections in \mathbb{N} , defined by set \mathbb{C} , such that:*

- A path protecting a connection must pass through the end nodes of the connection.
- The connections jointly protected by the same path must be mutually link disjoint, and also link disjoint from the protection path.
- The total number of edges used for both working and protection paths is minimum.

We also assume that the network is uncapacitated.

In order to formulate this problem, we modify the graph G to obtain the graph G' by adding a hypothetical source s and a hypothetical sink t . We also add a directed edge from s to each node v , where $v \in \mathbb{C}$, as well as a directed edge from each such node v to t . An example is shown in Figure 7. Figure 7.(a) shows a graph G with six nodes and ten bidirectional edges and the corresponding modification to the graph G' given two traffic requests $\mathbb{N} = \{(0, 3), (5, 2)\}$. Figure 7.(b) shows the provisioning of the two connections in \mathbb{N} and their protection path from s to t . Therefore, the problem of finding the protection paths turns out to be establishing connections from node s to t that traverse all the nodes $v \in \mathbb{C}$. For each subset of connections that are protected together, the two ends nodes of these traffic requests have to be traversed by the same protection path.

This disjoint paths routing problem can be formulated with ILP as follow: (Note that $G = (V, E)$ and $G' = (V', E')$ denote the original and modified graph in the formulation). It

is to be noted that the number of protection paths must satisfy:

$$1 \leq \text{number of protection paths} \leq N.$$

We may have more than one protection path because it is possible that the primary connections are partitioned into several sets and each set of primary connections share the protection of path. However, the worst case is that each primary path requires a unique protection path (the case of 1+1 protection), which results in a total of N protection paths. In the formulation, therefore, we have a maximum of $2N$ paths:

- Connections indexed from 1 to N are the ones given by the set \mathbb{N} , and these should be provisioned in the network.
- Connections indexed from $N+1$ to $2N$ are hypothetical connections, which correspond to protection connections, and at least one of them should be provisioned.

The ILP is formulated as a network flow problem, where there is a flow of one unit between each pair of end nodes of a connection, and there is also a flow of one unit from s to t for each protection path.

We define the following parameters, which are input to the ILP:

$G(V, E)$:	the original network graph
$G'(V', E')$:	the modified graph
\mathbb{N} :	the set of unicast connections
c_{mn} :	a constant, the cost of link $(m, n) \in E$
v_j :	set of end nodes of connection j in \mathbb{N} , $v_j = \{s_j, t_j\}$, which are different notations from the previous definition of a connection, denoted by S_i, T_j where i, j are the indices for the nodes.

We also define the following binary variables which are computed by the ILP:

f_{mn}^i	binary, equals 1 if the protection path i traverses link (m, n) in G
Zf_m^i	integer, the number of times that the node $m \in V$ is traversed by path i
U_j^i	binary, equals 1 if connection j is protected by path i
p_{mn}^j	binary, equals 1 if the working flow of j traverses link $(m, n) \in G$
q_{mn}^j	binary, equals 1 if the protection flow of j traverses link $(m, n) \in G$
Zp_m^j	integer, the number of times that node $m \in V$ is traversed by the working flow of j
Zq_m^j	integer, the number of times that node $m \in V$ is traversed by the protection flow of j

The objective function is:

$$\text{Minimize: } \sum_{(m,n) \in E} \left(\sum_{1 \leq j \leq N} p_{mn}^j c_{mn} + \sum_{N < i \leq 2N} f_{mn}^i c_{mn} \right)$$

The objective function minimizes the total cost of links used by the working paths (first term) and by the protection paths (second term). Note that a protection path at s and end at t in the modified graph, G' , but we only consider the cost of links in the original graph G .

The constraints are such that:

1) **Working Flow Conservation:**

$$\sum_{\{n:(s_j,n) \in E\}} p_{s_j n}^j = 1, \quad j \leq N; \quad (11)$$

$$\sum_{\{n:(m,n) \in E\}} p_{mn}^j = 2Zp_m^j, \quad \forall m \in V \setminus c_j. \quad (12)$$

The constraints (11) and (12) are standard flow conservation for working traffic which ensures that a bidirectional path is established between end nodes s_j and t_j of connection j .

2) **Protection Flow Conservation:**

For $\forall j \leq N, N < i \leq 2N$:

$$\sum_{\{n:(s_j,n) \in E\}} q_{s_j n}^j = 1; \quad (13)$$

$$\sum_{\{n:(m,n) \in E\}} q_{mn}^j = 2Zq_m^j, \quad \forall m \in V \setminus v_j; \quad (14)$$

Constraints (13) and (14) make sure that each connection j has a protection flow.

$$\sum_{\{n:(s,n) \in E'\}} f_{sn}^i \leq 1; \quad (15)$$

$$\sum_{\{n:(m,n) \in E\}} f_{mn}^i = 2Zf_m^i, \quad \forall m \in V; \quad (16)$$

The flow conservation of protection paths is ensured by constraints (15) and (16). It is worth noting that not every protection path i ($N < i \leq 2N$) is required unless it is used for protection.

$$\sum_{N < i \leq 2N} U_j^i = 1; \quad (17)$$

$$\frac{1}{N} \sum_{j \leq N} U_j^i \leq \sum_{\{n:(s,n) \in E'\}} f_{sn}^i; \quad (18)$$

$$f_{mn}^i \geq q_{mn}^j + U_j^i - 1, \quad \forall (m, n) \in E; \quad (19)$$

Each working flow should be protected by exactly one protection path, guaranteed by constraint (17). Meanwhile, any protection path i is provisioned only if it is used to protect any working path j . Otherwise, we do not need to provision it. Therefore, equation (18) ensures this constraint. Furthermore, constraint (19) ensures that if a protection path i protects connection j , it should traverse the same links used by the protection flow q_{mn}^j .

3) **Protection Path Sharing:**

For $\forall (m, n) \in E, N < i \leq 2N$:

$$p_{mn}^j + q_{mn}^j \leq 1, \quad \forall j \leq N; \quad (20)$$

$$p_{mn}^j + f_{mn}^i + U_j^i \leq 2, \quad \forall j \leq N; \quad (21)$$

$$p_{mn}^j + p_{mn}^k + U_j^i + U_k^i \leq 3, \quad \forall j < k \leq N. \quad (22)$$

The working flow and protection flow of each connection j should be link disjoint, reflected by constraint (20). Each protection path may protect multiple connections so that it needs to traverse multiple corresponding protection flows. Thus, each protection path should also be link disjoint to all the working flow it protects. This constraint is ensured by equation (21). Meanwhile, if two connections are protected by the same path f , their working flow should also be link disjoint such that codewords can be decoded at each end nodes through the protection path. The last constraint is guaranteed by equation (22).

The total number of variables used in the ILP is $(3N|V| + 3N|E| + N^2)$ and the total number of constraints is $(6N|V| + 2N + 2N^2|E| + N|E| + N^2(N-1)|E|)$, which is dominated by $O(N^3|E|)$.

VIII. NUMERICAL RESULTS

This section presents numerical results of the cost of our proposed protection scheme and compares it to 1+1 protection and Shared Backup Path Protection (SBPP) in terms of total resource requirements for protection against single-link failure. SBPP has been proven to be the most capacity efficient protection scheme and can achieve optimal solutions [12]. However, it is also a reactive protection mechanism and takes time to detect, localize and recover from failures. We consider two realistic network topologies, NSFNET and COST239, as shown in Fig. 8 and 9, respectively. Both networks are bidirectional and each bidirectional span e has a cost c_e , which equals the actual distance in kilometers between two end nodes.

We first compare three schemes in terms of the total connection and protection provisioning cost in both networks as shown in Fig. 10 and 11, respectively. We obtained the results by formulating the problems as ILPs using three different approaches. The x-axis denotes the number of connections in the static traffic matrix and y-axis denotes the total network design cost. Each value is the average cost over ten independent cases and all approaches used identical traffic requests for each case.

Since SBPP is the most capacity efficient scheme, it achieves the minimum cost. 1+N approach uses much lower cost than 1+1, but is higher than SBPP in both networks. We express the extra cost ratio of a scheme over SBPP by: $(Cost_{scheme} - Cost_{SBPP})/Cost_{SBPP}$. The extra cost ratio of 1+N in NSFNET increases from 5.2% to 23% as the number of connections increases from 2 to 7. Meanwhile, the extra cost ratio of 1+1 over SBPP increases from 12% to 45%, which is almost twice that of 1+N at each case. The advantage of 1+N over 1+1 in COST239 is even more significant than NSFNET

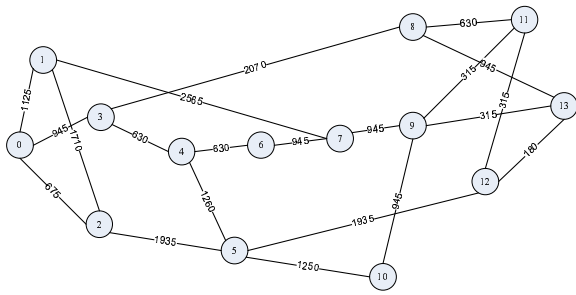


Fig. 8. NSFNET (N=14, E=21)

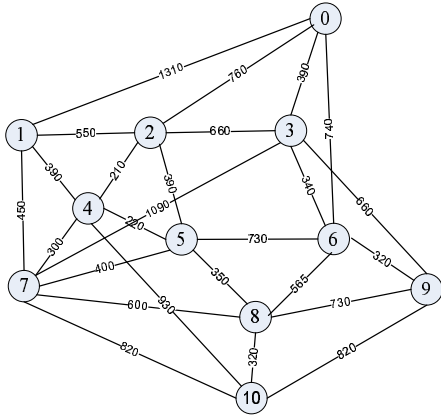


Fig. 9. COST239 (N=11, E=26)

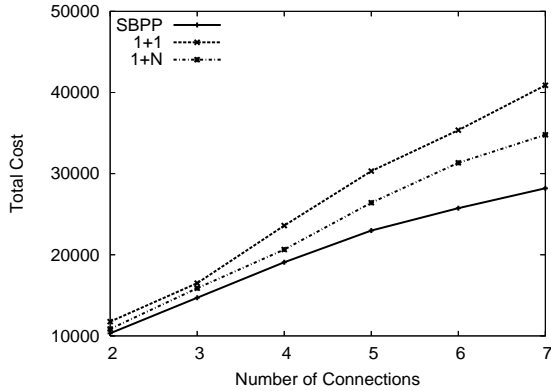


Fig. 10. Comparison of total cost in NSFNET

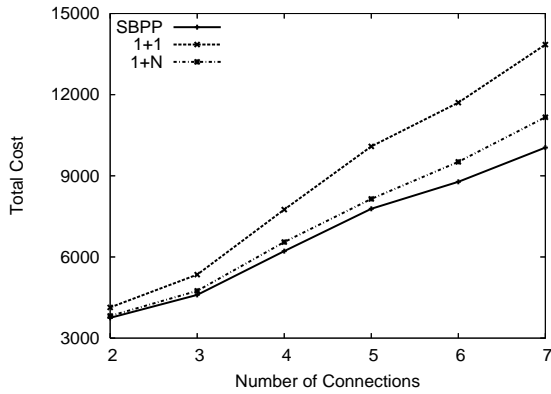


Fig. 11. Comparison of total cost in COST239 network

due to the larger average nodal degree, 4.6, compared to, 3, in NSFNET. Hence, there is a higher chance for multiple primary paths to share the same protection path, which results in lower overall cost. Based on the results, we can observe that the extra cost ratio of 1+N over SBPP in COST239 increases from 1.8% to 11.1% whereas the ratio of 1+1 over SBPP increases from 10.2% to 38%, as the number of connections increases from 2 to 7. Actually, the cost of using 1+N is very close to the optimal in COST239 network. The extra cost required by 1+N over the optimal solution is less than 27% of that achieved by 1+1 scheme.

In fact, if we only consider the cost of protection, i.e. exclude the cost of connection provisioning, 1+N protection uses much lower resources than 1+1 protection. For example, by examining one network scenario where there are seven connections in COST239 network, the average protection cost of using SBPP, 1+N and 1+1 protection schemes is 3586.0, 4313.5 and 6441.5, respectively. The saving ratio of 1+N to 1+1 is around 33%, which is higher than the saving ratio of joint capacity cost (19.3%). This example further illustrates the cost saving advantages of using 1+N protection over 1+1 protection.

In summary, 1+N protection has a traffic recovery speed which is comparable 1+1 protection. However, it performs significantly better than 1+1 scheme in terms of protection cost. Compared with the most capacity efficient protection scheme, SBPP, 1+N protection performs close to SBPP in terms of total capacity cost in dense networks. However, SBPP takes much longer to recover from failures due to the long switch reconfiguration time and traffic rerouting, which are not required in 1+N protection.

IX. CONCLUSIONS

This paper has introduced a resource efficient, and a fast method for providing protection for a group of connections such that a second copy of each data unit transmitted on the working circuits can be recovered without the detection of the failure, or rerouting data. This is done by linearly combining the data units using the technique of network coding, and transmitting these combinations on a shared set of protection circuits in two opposite directions. The reduced number of resources is due to the sharing of the protection circuit to transmit linear combinations of data units from multiple sources. The coding is the key to the instantaneous recovery of the information. This provides protection against any single link failure on any of the working circuits. The paper also generalized this technique to provide protection against multiple link failures.

The method introduced in this paper improves the technique introduced in [15] and [14]. In particular, (a) it requires fewer protection resources, and (b) it implements coding using a simpler synchronization strategy. A cost comparison study of providing protection against single link failures has shown that the proposed technique introduces a significant saving over typical protection schemes, such as 1+1 protection, while achieving a comparable speed of recovery. The numerical

results also show that the cost of our 1+N scheme is close to SBPP, the most capacity efficient protection scheme. However, the proposed scheme in our paper provides much faster recovery than SBPP.

REFERENCES

- [1] D. Zhou and S. Subramaniam, "Survivability in optical networks," *IEEE Network*, vol. 14, pp. 16–23, Nov./Dec. 2000.
- [2] S. Li and A. Ramamoorthy, "Protection against link errors and failures using network coding," *IEEE Transactions on Communications*, to appear. Available: <http://arxiv.org/abs/0905.2248>
- [3] D. Stamatelakis and W. D. Grover, "Ip layer restoration and network planning based on virtual protection cycles," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 10, pp. 1938–1949, 2000.
- [4] W. D. Grover, *Mesh-based survivable networks : options and strategies for optical, MPLS, SONET, and ATM Networking*. Upper Saddle River, NJ: Prentice-Hall, 2004.
- [5] C. Fragouli, J.-Y. LeBoudec, and J. Widmer, "Network coding: An instant primer," *ACM Computer Communication Review*, vol. 36, pp. 63–68, Jan. 2006.
- [6] D. Schupke and R. Prinz, "Performance of path protection and rerouting for wdm networks subject to dual failures," in *Optical Fiber Conference*, pp. 209–210, 2003.
- [7] S. Kim and S. Lumetta, "Evaluation of protection reconfiguration for multiple failures in WDM mesh networks," in *Optical Fiber Conference*, pp. 210–211, 2003.
- [8] J. Zhang, K. Zhu, and B. Mukherjee, "Backup provisioning to remedy the effect of multiple link failures in wdm mesh networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, pp. 57–67, Aug. 2006.
- [9] H. Choi, S. Subramaniam, and H.-A. Choi, "Loopback recovery from double-link failures in optical mesh networks," *IEEE/ACM Transactions on Networking*, vol. 12, pp. 1119–1130, Dec. 2004.
- [10] W. He, M. Sridharan, and A. K. Somani, "Capacity optimization for surviving double-link failures in mesh-restorable optical networks," *Photonic Network Communications*, vol. 9, pp. 99–111, Jan. 2005.
- [11] Y. Liu, D. Tipper, and P. Siripongwutikorn, "Approximating optimal spare capacity allocation by successive survivable routing," *IEEE/ACM Transactions on Networking*, vol. 13, pp. 198–211, Feb. 2003.
- [12] S. Ramamurthy and B. Mukherjee, "Survivable WDM mesh networks. part I-protection," in *Proceedings of IEEE INFOCOM, 1999*.
- [13] A. E. Kamal, "1+N protection in optical mesh networks using network coding on p-cycles," in *the proceedings of the IEEE Globecom, 2006*.
- [14] A. E. Kamal, "1+N protection against multiple faults in mesh networks," in *the proceedings of the IEEE International Conference on Communications (ICC), 2007*.
- [15] A. E. Kamal, "1+N Network Protection for Mesh Networks: Network Coding-Based Protection using p-Cycles," *IEEE/ACM Transactions on Networking*, Vol. 18, No. 1, Feb. 2010, pp. 67–80.
- [16] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, pp. 1204–1216, July 2000.
- [17] J. Vygen, "NP-completeness of some edge-disjoint paths problems," *Discrete Appl. Math.*, vol. 46, pp. 83–90, 1995.
- [18] R. Bhandari, *Survivable Networks: Algorithms for Diverse Routing*. Springer, 1999.
- [19] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge University Press, 1992.
- [20] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. North Holland, 1977.
- [21] J. Blomer, M. Kalfane, R. Karp, M. Karpinski, M. Luby, and D. Zuckerman, "An xor-based erasure-resilient coding scheme," *Int. Comput. Sci. Inst., Berkeley, CA, TR-95-048, 1995*. [Online]. Available: cseer.ist.psu.edu/blomer95xorbased.html
- [22] J. Lacan and J. Fimes, "Systematic MDS erasure codes based on Vandermonde matrices," *IEEE Communications Letters*, vol. 8, no. 9, Sep. 2004.
- [23] I. E. Shparlinski, "On singularity of generalized Vandermonde matrices over finite fields," *Finite Fields and Their Applications*, vol. 11, no. 2, pp. 193–199, 2005.
- [24] N. J. A. Harvey, D. R. Karger, and K. Murota, "Deterministic network coding by matrix completion," in *SODA '05: Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, 2005, pp. 489–498.
- [25] L. Lovasz, "On determinants, matchings and random algorithms," in *Fund. Comput. Theory 79, Berlin, 1979*.
- [26] S. Lin and D. J. Costello, *Error control coding: fundamentals and applications*. Prentice Hall, 2004.
- [27] C. Cooper, "On the distribution of rank of a random matrix over a finite field," in *Random Struct. Algorithms*, vol. 17, no. 3-4, pp. 197–212, 2000.



Ahmed E. Kamal Ahmed E. Kamal (S'82-M'87-SM'91) is a professor of Electrical and Computer Engineering at Iowa State University. His research interests include high-performance networks, optical networks, wireless and sensor networks and performance evaluation. He is a senior member of the IEEE, a senior member of the Association of Computing Machinery, and a registered professional engineer. He was the co-recipient of the 1993 IEEE Hartree Premium for papers published in Computers and Control in IEE Proceedings for his paper entitled

Study of the Behaviour of Hubnet, and the best paper award of the IEEE Globecom 2008 Symposium on Ad Hoc and Sensors Networks Symposium. He served on the technical program committees of numerous conferences and workshops, was the organizer and co-chair of the first and second Workshops on Traffic Grooming 2004 and 2005, respectively, and was the chair of co-chair of the Technical Program Committees of a number of conferences including the Communications Services Research (CNSR) conference 2006, the Optical Symposium of Broadnets 2006, and the Optical Networks and Systems Symposium of the IEEE Globecom 2007, the 2008 ACS/IEEE International Conference on Computer Systems and Applications (AICCSA-08), and the ACM International Conference on Information Science, Technology and Applications, 2009. He is also the Technical Program co-chair of the Optical Networks and Systems Symposium of the IEEE Globecom 2010. He is on the editorial boards of the Computer Networks journal, and the Journal of Communications.



Aditya Ramamoorthy Aditya Ramamoorthy received his B. Tech degree in Electrical Engineering from the Indian Institute of Technology, Delhi in 1999 and the M.S. and Ph.D. degrees from the University of California, Los Angeles (UCLA) in 2002 and 2005 respectively. He was a systems engineer at Biomorphic VLSI Inc. till 2001. From 2005 to 2006 he was with the data storage signal processing group at Marvell Semiconductor Inc. Since Fall 2006 he has been an assistant professor in the ECE department at Iowa State University. He has interned

at Microsoft Research in summer 2004 and has visited the ECE department at Georgia Tech, Atlanta in Spring 2005. His research interests are in the areas of network information theory and channel coding.



Long Long Long Long received his B.Eng degree in Electronic Information Engineering from Huazhong University of Science and Technology, Wuhan, China in 2002 and M.Sc degree in Software Engineering from Peking University, Beijing, China in 2005. Since fall 2006, he has been a Ph.D student in ECE department of Iowa State University, USA. His research interests are in the area of traffic grooming and survivability of optical networks.



Shizheng Li Shizheng Li received his B.Eng degree in information engineering from Southeast University (Chien-Shiung Wu Honors College), Nanjing, China, in 2007. He worked on error correction codes in National Mobile Communications Laboratory at Southeast University during 2006 and 2007. Since Fall 2007, he has been a Ph.D. student in the Department of Electrical and Computer Engineering, Iowa State University. His research interests include network coding, distributed source coding and network resource allocation. He received Microsoft Young

Fellowship from Microsoft Research Asia in 2006. He is a student member of IEEE.