On Providing ATM Multipoint Services Survivability

William Yurcik * David Tipper †

Department of Information Science and Telecommunications
University of Pittsburgh
Pittsburgh, PA. 15260 USA
{*yurcik,tipper*}*@tele.pitt.edu*

Deepankar Medhi ‡

Department of Computer Networking
University of Missouri - Kansas City
Kansas City, MO 64110 USA
dmedhi@cstp.umkc.edu

July 16, 1998

Abstract

In this paper we examine the potential benefits of allocating a dedicated backup connection to provide survivability for ATM multipoint communications. Specifically we examine the feasibility of providing survivability using a working multipoint connection with a disjoint dedicated backup multipoint connection where the multipoint connections are setup using either Virtual Rings, shared multicast trees, or VC Mesh groups. We introduce a novel optimization formulation which identifies a "Disjoint Steiner Ring" within a general graph in order to construct Virtual Rings. Numerical results show that disjoint backup shared multicast trees and disjoint backup VC Mesh groups are not always available while backup Virtual Rings exist in all cases tested. In addition, experiments comparing the cost of providing survivability between Virtual Rings, shared multicast trees, and VC Mesh groups show that Virtual Rings are lower in cost for the topologies considered.

*author for correspondence, additional contact information: telephone (412) 624-9411, FAX (412) 624-2788; supported in part by NASA Earth Systems Science grant # NGT-30019, Defense Advanced Research Projects Agency grant # F30602-97-1-0257, and SAE International - The Engineering Society for Advancing Mobility Land Sea Air and Space

†supported in part by Defense Advanced Research Projects Agency grant # F30602-97-1-0257 and National Science Foundation grant # NCR-95-06652

‡supported in part by Defense Advanced Research Projects Agency grant # F30602-97-1-0257

1 Introduction

The importance of multipoint communications[1] and the need to efficiently and reliably support it within a B-ISDN network based on ATM has attracted a great deal of attention. Multicast mechanisms provide multipoint communications by reducing the amount of duplicate traffic in the network to conserve bandwidth and switch resources. When a sender transmits data to multiple receivers, some delivery paths may have common links. These common links can be shared such that the data is sent only once over the common links and replicated at branch points. Thus multicast is efficient because it minimizes aggregate network load by decreasing the number of links and nodes utilized in a conversation between group members. This efficiency increases with the size of the network and the number of group members. The network resources expended to support a multipoint communication using an underlying multicast mechanism grows less than linearly as the number of group members increases. While a succession of point-to-point unicast connections could provide multipoint communications, this approach is inherently inefficient and is unlikely to support new multipoint services with increasing resource requirements (e.g., video teleconferencing, distributed interactive simulation).

Supporting native ATM multipoint communications is difficult because ATM is inherently connection-oriented: (1) Steiner Tree heuristics developed for bidirected networks do not distinguish sending nodes from receiving nodes and (2) ATM requires VP/VC set-up establishment which is not compatible with group dynamics of participants joining and leaving [12]. The search for an ATM multipoint solution began with the introduction of ATM multicast capability with point-to-multipoint VCs in the ATM Forum's User Network Interface (UNI) specification 3.1. The ATM Forum MPOA (Multi-Protocol Over ATM) subworking group proposes two current approaches for ATM intracluster multipoint: (1) the VC Mesh model and (2) the Multicast Server model (MCS).[2] As a result of the perceived complexity and inefficiency of these two approaches, other techniques have been proposed such as shared multicast trees and Virtual Ring multicasting.

ATM network survivability has focused on point-to-point communications at the Virtual Path (VP) or virtual circuit/channel level (VC) and there has been very little work on the survivability of ATM multipoint communications [29]. The same types of vulnerabilities present in ATM unicast are also present in ATM multicast but the potential risk of these vulnerabilities is considerably greater in multipoint. For instance, ATM multicast traffic traverses potentially many more physical links than does a single unicast communication, thereby creating more of an opportunity to encounter link or node failures. While an ATM unicast path is a collection of links and nodes between one source and one destination, ATM multicast can have multiple such paths with the number of paths bounded by the factorial of the number of group members. To exacerbate this vulnerability, it is often more complex to determine and control which links are in use for a given multicast routing scheme as compared to the equivalent point-to-point communication. Additionally, users may require an all or nothing restoration where if all group members can not be restored simultaneously then the multipoint communication is abandoned.

Preplanned, dedicated bandwidth, disjoint backup route techniques have been widely used to provide

[1]a note on terminology: Multipoint communications will be used to describe any type of communication service within a group of multiple participants (it is equivalent to group communication). Multicast will be used to describe the underlying point-to-multipoint directional transmission mechanism (i.e., delivery routing tree, multipoint circuit) which is used to create a multipoint service. Thus, multipoint has a more general connotation and will refer to the group communication network service provided to the user.

[2]A cluster is defined as a set of ATM interfaces choosing to participate in native ATM connections. Traffic between ATM interfaces belonging to different clusters pass through an intercluster device.

survivability in circuit switched networks and have been proposed for the restoration of ATM point-to-point VPs and VCs. This paper investigates the extension of these point-to-point techniques to the restoration of ATM group communications. The basic concept proposed is each multipoint group communication requiring survivability will establish a dedicated backup set of connections with reserved bandwidth for restoration. When a single link or node fault occurs within a "working" group, traffic flow is rerouted to the corresponding "backup" group. Our results show that using this backup approach leads one to favor Virtual Ring multicasting for survivable ATM multipoint communication.

The remainder of this paper is organized as follows: Section 2 briefly describes previously proposed schemes for ATM group communications. Section 3 discusses survivability issues and tradeoffs of these proposed schemes. Section 4 introduces a novel optimization formulation which provides a theoretical basis upon which ATM Virtual Rings can be implemented. Section 5 presents numerical results comparing the feasibility and cost of restoration via Virtual Ring multicasting, shared multicast trees, and VC Mesh groups. The paper closes with a summary of conclusions and future research directions in Section 6.

2 Proposed ATM Group Communication Schemes

In this section, we briefly overview some of the techniques proposed for ATM multipoint communication beginning with the two current models proposed for intracluster multicast by the ATM Forum.

2.1 VC Mesh Model

In the VC Mesh Model, a unidirectional VC originates from each sender to all members of a group. Each group member must terminate one VC for each active source in the cluster. Each sender is essentially the root of its own source-based tree. If a member joins or leaves the group, each point-to-multipoint VC for each sender in the group must be modified (a signaling load that scales with the number of senders in a group). The ATM Forum UNI 4.0 specification allows a new member to join a point-to-multipoint VC without the intervention of the sender by using a Leaf-Initiated Join (LIJ) capability [5]. Receivers discover the existence of point-to-multipoint VCs through a directory and are able to signal a request to join a particular point-to-multipoint VC. LIJ does not specify where the new member should be joined to the existing point-to-multipoint VC. For example, the new member (leaf) may join at the nearest node which is already part of the point-to-multipoint VC or it may join via the shortest path to the sender (root).

Figure 1A shows an example of a group of three nodes using the VC Mesh Model. For a single group containing multiple senders, a unidirectional point-to-multipoint VC for each sender is established on the network with no coordination between VCs except for capacity allocation on shared links between senders. For multiple groups on the same network, a point-to-multipoint VC for each sender is handled independently except for capacity allocation on shared links between groups. This criss-crossing of VCs across a network gives rise to the name "VC Mesh" [4]. Using the VC Mesh model to form groups does not assume groups contain only symmetrical circuits; rather each source in turn identifies circuit(s) to connect with the remainder of the group. For example, a 3-switch group may consist of the following combinations of circuits: 6 point-to-point circuits; 3 point-to-multipoint circuits; 2 point-to-point/2 point-to-multipoint circuits; or 4 point-to-point/1 point-to-multipoint circuits (in 3 different permutations). In the asymmetric circuit case, traffic flow between node pairs may travel along different paths. In the symmetric circuit case, traffic flow between node pairs share the same path but in different directions.

2.2 Multicast Server Model

In the MCS Model, a server is chosen within each cluster to serve as a proxy for all senders in order to relieve senders from direct VC connection setup and release operations resulting from group dynamics. Conceptually the MCS serves as a centralized connection manager mediating and disassociating senders and receivers within a group [32]. All senders establish a unidirectional point-to-point VC to the MCS. The MCS establishes a unidirectional point-to-multipoint VC to the rest of the multicast group. During transmission, the MCS reassembles cells into packets arriving on all incoming VCs (avoiding interleaving cells from different packets) and queues cells belonging to particular packets for transmission on an outgoing point-to-multipoint VC.[3] Figure 1B shows an example of a group of three nodes using the MCS Model. The most basic MCS architecture only supports one group at a time, reflecting all incoming cells to the outbound point-to-multipoint VC. A more complex MCS architecture can support multiple groups simultaneously with the addition of minimal layer-3-specific-processing which allows the MCS to look inside received cells to determine which group and thus which point-to-multipoint VC to send the cell. All requests to create a group, delete a group, to add a member to a group, or delete a member from a group are sent to the MCS which maintains state information for all existing groups.

2.3 Shared ATM Multicast Tree Schemes

Many different schemes have been proposed for ATM group communications based on a shared ATM multicast tree approach. Here we summarize three typical schemes: SMART, SEAM and SPAM [18, 19, 21]. These schemes are similar in that they try to provide a general-purpose control architecture by modifying in-band control mechanisms of ATM switches [28]. Resources are reserved in both directions on all of the VC links of a shared multicast tree until the connections are released[18]. Figure 1C shows an example of a shared multicast tree supporting a group of three nodes communicating. Each of these schemes individually address two inherent problems of a shared tree architecture: (1) cell interleaving because cells from different sources may arrive interleaved at one destination and (2) resource management because resources allocated to a connection are shared between a number of different sources.

SMART (Shared Many-to-many ATM ReservaTions) proposes a shared tree of one to several associated many-to-many VCs where the number of VCs is independent of the number of group members. The same service class and traffic parameters are given to all connections of the tree at call setup. By granting access to a shared tree through the use of resource management (RM) cells, SMART ensures that only one source is active on a tree at any given instant. A shortcoming of SMART is whether a single access control mechanism will be sufficient for all group communication requirements [28].

SEAM (Scalable and Efficient ATM Multicast) is another attempt to support multipoint-to-multipoint VCs in a shared tree approach. By requiring ATM switches to detect the first and last cells in a packet, SEAM relies on per-VC queueing and per-packet forwarding to avoid interleaving of cells from different packets. SEAM introduces two concepts: (1) "cut-through forwarding" which is a mechanism to forward all cells from a packet together while buffering cells from later packets until the last cell is forwarded and (2) "short-cutting" which is a signaling mechanism allowing cells to follow the shortest path along a shared tree instead of having to go all the way to the "root" and be forwarded back to receivers. The two major shortcomings of SEAM are: (1) a slow stream can potentially block fast input ports and (2) there is potential for a large penalty if the cell containing the End-Of-Packet (EOP) bit flag is lost.

[3] Reassembly of incoming AAL_SDUs is required at the multicast server as AAL5 does not support cell level multiplexing of different AAL_SDUs on a single outgoing VC [4].

SPAM (a Simple Protocol for ATM Multicast) is a third attempt to support multipoint-to-multipoint VCs in a shared tree approach by incrementally improving SEAM. In contrast to the forwarding model of SEAM where there is no multiplexing of cells from different packets, SPAM multiplexes cells from different packets through the use of a proposed AAL-SPAM and "native cut-through forwarding". As a compromise between AAL5 and AAL3/4, AAL-SPAM is AAL5 plus an added 16 bit MID (Multiplexing IDentifier) field. Unique MIDs are assigned to senders within a multicast group such that receivers can distinguish cells from different senders. "Native cut-through forwarding" refers to cells from different packets being switched immediately and multiplexed over the shared many-to-many VC connecting the group. Results comparing SPAM with SEAM report the following: (1) SPAM buffer requirements are independent of packet size while SEAM buffer requirements increase significantly with increase in packet size and (2) SPAM delay is the expected propagation and transmission delay while certain SEAM switches in a network will experience maximum delay due to increased blocking [21]. The major shortcomings of SPAM are: (1) the practicality of modifying AAL5 to AAL-SPAM and (2) the global allocation of unique MIDs.

2.4 Virtual Ring Multicasting

Ofek and Yener have proposed to the Virtual Ring [27, 33, 7] for multipoint group communications. The Virtual Ring is a unidirectional circular overlay of connections that join all desired members for a group communication. Using Virtual Rings is straightforward since a message sent from one member of the group travels around the ring and back to the sender [7]. Figure 1D shows an example of a Virtual Ring connecting a group of three nodes. The main advantage of the Virtual Ring is that it extends multicast to large groups. To overcome the "ACK implosion effect" while retaining a high degree of reliability, the Virtual Ring employs a combined optimistic/pessimistic strategy of Implicit ACKs and Explicit NACKs. The optimistic strategy is to avoid sending positive feedback messages (ACKs) when there is a high probability the network layer is normal and instead rely on Implicit ACKs with the circulation of cells to be received back at the originating source. The Implicit ACK procedure considers each cell that circulates the Virtual Ring and returns to its source as an implicit positive acknowledgment. The cells that do not return to the source are detected by sequence number or a group time-out mechanism. In less probable pessimistic situations when a cell is missing at an upper layer within a group member receiver, the error control load is shifted from the source to the group of receivers with the use of NACKs. An Explicit NACK is initiated by the receiver (above the transport layer) when ordered cells are missing. The Explicit NACK is a unicast retransmission request sent from a receiver directly to the source requesting retransmission of only the missing cell(s). These combined techniques also minimize source/receiver processing requirements for state information.

In the application of Virtual Ring multicasting to ATM networks, each unicast connection between adjacent nodes on the Virtual Ring can be implemented as a separate point-to-point VC. The Virtual Ring provides a global sense of direction for concatenated VCs from switch to switch around an enclosed ring. Possible drawbacks of the Virtual Ring for ATM multicasting are the increased average end-to-end delay and the lack of mechanisms to accommodate group dynamics.

The first criticism is that a ring topology is generally longer (more hops end-to-end) than its corresponding tree topology. The size of a ring increases faster than a tree with an increasing number of participants. A larger topology will increase the average end-to-end delay between distant nodes on a unidirectional ring. Ofek and Yener claim that although the routing length in a Virtual Ring may be longer than a least cost tree, the ultimate delay experienced in a Virtual Ring can be shorter. They base this upon the claim that Virtual Ring minimizes retransmissions, congestion, and buffering within the network while tree networks under high load have increasing losses and retransmissions which will increase end-to-end delay.

The second criticism of Virtual Ring multicasting is that a ring topology may be more difficult to maintain with group dynamics (join/leave requests) than other topologies (i.e., star, tree). When original members leave a Virtual Ring, there are no mechanisms for the ring to shrink to include only the remaining group members. Shrinking a ring due to leave requests is less of a problem than expanding a ring due to new members joining, but leave requests create an opportunity for inefficient Virtual Rings after many group members leave a larger ring (inefficient in terms of bandwidth utilization and node processing). One solution for a Virtual Ring to handle group dynamics is to establish the initial ring for the maximum number of identified participants (a closed set) and then allow group members to leave without changing the Virtual Ring. This is analogous to the operation of the Internet MBONE in that sites must arrange for a feed from the MBONE distribution tree and be multicast-capable (routers, local area network, and host) but can choose to participate or not participate in individual MBONE broadcast sessions.

3 Survivability Issues and Tradeoffs

The two most relevant survivability issues that translate from unicast restoration to multipoint restoration are: (1) the location to initiate restoration and (2) the method of identifying a backup route for restoration. Considering the tradeoffs we briefly outline, we have decided to pursue end-to-end preplanned restoration since it can simultaneously provide optimal and guaranteed restoration.

There are two locations for initiating restoration: end-to-end restoration initiated between sender and receivers and local restoration initiated between the upstream and downstream nodes next to the detected fault. While local restoration offers greater speed and ease of implementation, it may also result in blocking and "backhaul" situations where the backup route is not disjoint from the remaining working route. In general, the further restoration is initiated away from a fault results in slower restoration but a globally optimal solution, less blocking, and avoidance of backhaul situations. Results in [10] verify that end-to-end restoration methods are always more capacity efficient than local restoration but the size of this advantage depends upon network load. In the ATM context, local restoration is available only if VPs/VCs are concatenated together to form an end-to-end route. A single VP/VC that extends over multiple hops will perform end-to-end restoration unless the fault is handled entirely at a lower layer transparent to ATM (i.e., SONET ring rerouting at the transmission layer).

There are two methods of identifying a backup route for restoration: a preplanned backup route can be computed in advance or a dynamic search for a backup route can be computed in real-time. In the preplanned method, each working route has a prearranged backup route when initially assigned. In the dynamic search method, senders and receivers broadcast signaling messages after being notified of a network failure and backup routes are selected via a specified algorithm. In the preplanned method, the backup route has reserved bandwidth that remains dedicated for restoration purposes. In dynamic search, no bandwidth is reserved for backup routes so restoration is not guaranteed (i.e., residual bandwidth may not be available). In the ATM context, the preplanned method is faster and dependent only on the number of VPs/VCs to be rerouted while the dynamic search method is slower and dependent on network conditions (traffic loads, network topology, fault location, ATM signaling) in addition to the number of VPs/VCs to be rerouted. Considering group dynamics, the preplanned method incurs a significant setup delay for reassignment of working and backup VPs/VCs as members join and leave a group while dynamic search does not need updates since the only group membership to be considered is when a fault occurs. Leaving these general survivability issues, we will now develop survivability issues specific to the ATM multipoint schemes we have introduced.

The VC Mesh Model has a number of survivability issues that need to be addressed. In the VC Mesh Model, group dynamics drive the establishment of a new point-to-multipoint VC for each sender in a group in order to reflect updated membership after each join or leave operation.[4] Also in the VC Mesh Model, multiple senders transmitting to the same group members use separate point-to-multipoint VCs while in the MCS Model multiple senders each establish point-to-point VCs to the MCS and then the MCS has the capability to integrate point-to-multipoint transmissions to the same group members. For both of these reasons, the number of VCs utilized in the VC Mesh Model increases with the number of senders at a faster rate than the MCS Model. If all group members are active senders at some point over the lifetime of a particular group, then the resulting paths will be a fully-connected mesh. As a result, there is potentially a larger number of routes over which a failure may occur for the VC Mesh which means an increased exposure to link and node failures.

Other survivability issues specific to the VC Mesh Model include increased signaling traffic due to more VCs and increased termination requirements (end hosts need one VC termination per source per each ATM interface). These issues both represent performance limitations due to lack of model scalability. As group sizes grow, providing survivability becomes more difficult due to the increased capacity required for signaling and the exhaustion of end host ports due to previous sender transmissions.

The MCS Model has a number of survivability issues which make it an unattractive solution if not addressed. The MCS constrains the number of potential paths over which a failure may occur by concentrating traffic on a server which is both a performance bottleneck and single point-of-failure for a group's communication. If an MCS is disabled, all group sessions using that MCS terminate.

Survivability issues common to both the VC Mesh and MCS Models: (1) due to group dynamics, it can be complex to identify and thus protect links/paths that will be used for multipoint communications; (2) since it is complex to identify utilized links/paths in advance, this makes provisioning capacity for multipoint restoration difficult especially given a network of multiple groups with different bandwidth requirements; and (3) group access control such that only members of a group can send-to or receive-from the group, both the VC Mesh and MCS Models require external mechanisms to identify group members. The failure of individual senders (or receivers) within a group will degrade but not terminate group communications if there are other senders (or receivers) still operating. We will not be addressing strategies to recover from individual sender or receiver failures but rather focus on the survivability of a group's communication.

For the shared ATM multicast tree, the dominant survivability issue is handling link or node failures in the "trunk" of the shared tree. By "trunk" we mean the portion of the tree where traffic from all group members traverse. A link or node failure in the "trunk"' of the multipoint-to-multipoint circuit(s) connecting all group members will interrupt all group communications. On the other hand, sharing a common tree makes the identification and protection of links and authentication of traffic a straightforward task. Signaling to handle group dynamics will require LIJ capabilities to join either to the nearest node, nearest capable node, or a specified access node of the shared tree. Without LIJ capabilities, the Shared Multicast Tree is restricted to static group membership or the establishment of a new multipoint-to-multipoint VC with each join or leave operation.

[4] applies to ATM Forum UNI 3.1 specification without LIJ capability; the intention of the LIJ capability is to avoid all parties being aware of group dynamics but LIJ still may have a similar property depending on implementation

A general technique to provide survivability to tree-based approaches (VC MESH, shared multicast tree) is to restore faults on the "working" tree by rerouting to a preplanned disjoint "backup" shared multicast tree. Figure 2A depicts a "working" shared multicast tree and its corresponding "backup" disjoint shared multicast tree.

3.1 Survivability via the Virtual Ring

The formulation of the Virtual Ring as proposed by Ofek and Yener does not result in a least cost link and node disjoint ring.[5] Thus a single node or link fault may disconnect a Virtual Ring in multiple places. Ofek and Yener propose to restore interrupted Virtual Ring multicast connections from link and node faults by a tree-based approach they refer to as a Tree Embedded Ring (TER) [33]. This approach computes a new spanning tree of the remaining links and nodes available after a fault and then rebuilds the Virtual Ring over the new spanning tree. The rebuilt Virtual Ring resulting from a TER may not include all available group members. Ofek and Yener also do not take into account restoration timing thresholds or the feasibility of processing a rebuilt Virtual Ring in real-time. In general fault recovery should take place as quickly as possible in order to maintain service continuity and minimize retransmissions.

Here we propose an alternate approach using concepts similar to self-healing SONET ring systems [3, 1, 30, 31, 20] and studies on Virtual Path (VP) -based ATM ring architectures (differentiated from our proposed VC-based ATM Virtual Ring architecture).

We propose to extend Virtual Ring multicasting to provide survivability to ATM multipoint communications with two contributions: (1) a novel formulation of a *least cost* Virtual Ring connecting all desired group members, that is *link and node disjoint*, such that a single link or node fault will only disconnect a ring in one place and (2) a real-time ring reconfiguration mechanism to provide restoration of single link or single node faults while maintaining the established Virtual Ring architecture (no need to compute a reformulation of the Virtual Ring in real-time).

A real-time ring reconfiguration mechanism to provide restoration of single link or single node faults can be achieved by bandwidth provisioning of circuits for counter-rotating Virtual Rings (a working Virtual Ring of VCs and a backup Virtual Ring of VCs in the reverse direction). Figure 2B shows a real-time Virtual Ring reconfiguration based on a link failure between node E and node H. A real-time ring reconfiguration mechanism must have the following five operations:

1. The switches in the Virtual Ring adjacent to a ring fault are notified of the failure via ATM operations and maintenance (OAM) signaling.

2. The switch directly upstream to the fault will redirect all unicast ring transmissions from the incoming working Virtual Ring interface to the outgoing backup Virtual Ring interface (node E in Figure 2B).

3. The switch directly downstream to the fault will redirect all unicast ring transmissions from the incoming backup Virtual Ring interface to the outgoing working Virtual Ring interface (node H in Figure 2B).

[5] A link and node disjoint ring R is a set of nodes and paths that form a closed circle (i.e., a set whose nodes and links form a cycle from node y back again to node y passing through no nodes twice and using no links twice).

4. All switches, except the one switch directly downstream from the fault, will continue to unicast transmit on the outgoing working Virtual Ring interface and continue to receive unicast transmissions from the incoming working Virtual Ring interface (while continuing to apply Implicit ACK/Explicit NACK reliability).

5. The one switch directly downstream from the fault will continue to unicast transmit on the outgoing working Virtual Ring interface but will reconfigure to receive unicast transmissions from the incoming backup Virtual Ring interface (while continuing to apply Implicit ACK/Explicit NACK reliability).

The real-time ring reconfiguration mechanism to provide restoration of single link or single node faults as described requires no new ATM signaling capabilities. The major ramification of the ring reconfiguration is an increase in end-to-end delay for multicast transmissions on the Virtual Ring. The number of links a cell traverses in the reconfigured Virtual Ring will increase from L to at most $2(L-2)$ for a single node failure and from L to at most $2(L-1)$ for a single link failure. Thus the size of a Virtual Ring which incorporates real-time restoration of faults is limited by the maximum end-to-end delay that can be tolerated in a reconfigured Virtual Ring. This may be a major limiting factor since a single ring topology is generally longer than a tree topology and a reconfigured Virtual Ring (two counter-rotating rings coordinated for restoration) will be longer than a single ring.

4 Disjoint Steiner Ring Formulation

Implementation of working and backup Virtual Rings to provide ATM multipoint survivability will require the identification of a least cost link and node disjoint ring connecting all desired group members within an arbitrary network (i.e., general graph). This problem is a generalization of providing the least cost pair of disjoint paths for two nodes as formulated in [25]. We refer to this as the Disjoint Steiner Ring (DSR) Problem because it is a variant of the Steiner Tree Problem in Graphs (find the disjoint least cost ring in a general graph G - a Steiner Ring - joining a subset of specified nodes). Our DSR formulation is similar to the minimum spanning tree (MST) approximation to the Steiner Tree Problem in Graphs, using abstraction to only consider paths between desired group members. While the MST approximation to the Steiner Tree Problem in Graphs is suboptimal because all trees containing Steiner Points are not considered, our DSR formulation provides an exact solution because all rings containing Steiner Points are considered. Abstraction in terms of paths results in a DSR formulation directly analogous to a generalization of the Traveling Salesman Problem (TSP) due to the least cost ring solution with added constraints to ensure disjoint links and nodes. The TSP can likewise be thought of a relaxation of the DSR; find the least cost ring among all nodes (as opposed to some subset of nodes) in a fully-connected graph (as opposed to a general graph with Steiner Points).

To our knowledge, the DSR problem has never been formulated before. The closest work is from researchers at Cidade University in Portugal who have investigated bounds for a complete graph problem in terms of maximizing revenue costs minus link costs and from researchers at Complutense University in Madrid, Spain who propose an approximation algorithm based on network reduction and exhaustive search [16, 26]. Bellcore researchers Cosares, et. al., state the following: "Since the ring-selection problem has not been formulated or solved before, we cannot compare our solutions to alternatives or even to good lower bounds, since none exist. ..."[8]

Very few formulations have been proposed for Steiner Tree variant problems in unidirected networks and all of them operate under special conditions (e.g., acyclic networks) and thus cannot be applied to practical networks [17]. This DSR formulation can assume either unidirected or bidirected links as specified in a given general graph G. Paths between node pairs are formed based on the unidirected or bidirected links available from graph G. The result of the DSR formulation is a ring of unidirected or bidirected links corresponding to original link assumptions. A DSR result of bidirected links can be routed in two directions. A DSR result of unidirected links can be routed only in the direction of the resulting solution. A unidirected $link_{u,v}$ between $node_u$ and $node_v$ is considered not disjoint from unidirected $link_{v,u}$ (i.e., opposite unidirected links in reverse directions) due to the practical survivability assumption that these links share common facilities and thus have dependent fault probabilities.

Let graph $G = (N, L)$ consist of a set of nodes N and a set of unidirected or bidirected links L. Each node pair is connected by at most one link in each direction (i.e., no multiple links between node pairs in a given direction). From N we select m multicast nodes to form set M where $M \subset N$. All simple paths[6] (i.e., no cycles) between multicast node pairs are considered. The ring solution will be denoted as an ordered set $R(M)$ containing all nodes in M arranged corresponding to relative node adjacency in the minimum cost ring and an ordered set $R(P)$ containing indices for the *disjoint* paths selected between adjacent node pairs. A reference to a node n in the ring is to be interpreted as a reference to the node labeled (n MOD m) where m ($m \geq 2$) is the number of multicast nodes within M. The node in the first position in the solution $R(M)$ is labeled as node 0 and the node in the last position in the solution set $R(M)$ is labeled as node $(m-1)$. The order of the nodes in set $R(M)$ corresponds directly to the order of the paths in set $R(P)$. A reference to a path q in the first position of set $R(P)$ is to be interpreted that the qth path between node 0 and node 1 was selected in the ring solution. No loops exist in the ring solution (a loop is a single link whose tail node is the same as its head node).[7] $R(M)$ and $R(P)$ can be combined into a single solution set $R(M)^*$ which is an ordered set of all nodes in the disjoint minimum cost ring solution, including m multicast nodes from set M and Steiner Point non-multicast nodes from set $(N - M)$. The Steiner Point non-multicast nodes are specified by the selected disjoint paths indicated by the partial ring solution $R(P)$. Since there are no multiple links between node pairs and no loops, the single solution set $R(M)^*$ contains all node and path information to completely specify an unambiguous disjoint minimum cost ring solution.

To formulate the optimization problem we define the following variables:

Let

i = index for each path considered

j = index for multicast nodes considered

$P(j, (j+1)MOD\ m)$ = set of paths from node j to node $((j+1)MOD\ m)$

$$P(j,(j+1)MOD\ m)_i \quad \begin{cases} = 0 & \text{if path } P(j,(j+1)MOD\ m)_i \text{ is not selected} \\ = 1 & \text{if path } P(j,(j+1)MOD\ m)_i \text{ is selected} \end{cases}$$

Given

$Q(j)$ = number of paths from node j to node $((j+1)MOD\ m)$

$C(j, (j+1)MOD\ m)_i$ = cost of path $P(j, (j+1)MOD\ m)_i$

[6] A path is a sequence of links and nodes starting at a head node and ending at a tail node. A simple path is a path without repetition of links and nodes. A cycle is a simple path where the head node and tail node are the same.

[7] terminology found in [2] p. 25

k = number of unidirected links within general graph G

n = number of nodes within set N

$\mathcal{L}\ (\ P(j,(j+1)MOD\ m)_i\ )\ =\ [\ L_k\ ]$ vector indicating the unidirected/bidirected links contained within path $P(j,(j+1)MOD\ m)_i$

$$L_k \quad \begin{cases} = 0 & \text{if } L_k \text{ is not contained within path } P(j,(j+1)MOD\ m)_i \\ = 1 & \text{if } L_k \text{ is contained within path } P(j,(j+1)MOD\ m)_i \end{cases}$$

$\mathcal{N}\ (\ P(j,(j+1)MOD\ m)_i\ )\ =\ [\ N_n\ ]$ vector indicating interior nodes (not end nodes) contained within path $P(j,(j+1)MOD\ m)_i$

$$N_n \quad \begin{cases} = 0 & \text{if } N_n \text{ is not contained within path } P(j,(j+1)MOD\ m)_i \text{ (or an end node)} \\ = 1 & \text{if } N_n \text{ is an interior node within path } P(j,(j+1)MOD\ m)_i \end{cases}$$

DSR optimization problem

$$\min_{P(j,(j+1)MOD\ m)_i} \left(\sum_{j=0}^{m-1} \sum_{i=1}^{Q(j)} C(j,(j+1)MOD\ m)_i \;\; P(j,(j+1)MOD\ m)_i \right)$$

subject to

(1): $\sum_{i=1}^{Q(j)} P(j,(j+1)MOD\ m)_i = 1 \;\; \forall j$

(2): $\sum_{j=0}^{m-1} \sum_{i=0}^{Q(j)} \mathcal{L}\ (\ P(j,(j+1)MOD\ m)_i\ )\ \ P(j,(j+1)MOD\ m)_i \ \le 1 \quad \forall k$

(3): $\sum_{j=0}^{m-1} \sum_{i=0}^{Q(j)} \mathcal{N}\ (\ P(j,(j+1)MOD\ m)_i\ )\ \ P(j,(j+1)MOD\ m)_i \ \le 1 \quad \forall n$

(4): $P(j,(j+1)MOD\ m)_i\ = 0 \text{ or } 1 \quad \forall i,j$

The DSR decision variables to be determined are the paths selected between multicast nodes for all valid non-isomorphic ring permutations ($P(j,(j+1)MOD\ m)_i$). The DSR objective function dictates that the solution is the combination of paths within a particular ring permutation that minimizes total cost over all possible paths over all possible rings. All paths between multicast nodes are accessible via preprocessing given the network topology and ring permutations for a given group size m are accessible via preprocessing. The structure of the DSR objective function forms a ring of paths between all multicast nodes in set M and the ring preprocessing makes available different permutations of set M over which to search for the solution. The following discussion will focus on the preprocessing of ring permutations since the set of all possible ring permutations is an intractable search space that contains redundant and unnecessary information. DSR constraint 1 restricts the solution to selecting a single path between each multicast node pair in a ring. DSR constraint 2 restricts the solution to link disjoint paths and should be interpreted that each bidirectional link in graph G should only be used once in the ring solution (i.e., different paths cannot use the same link). A consequence of DSR constraints 1 and 2 in combination is that each multicast node must have a degree of at least two for a ring solution to exist (given that all paths are simple paths and no loops exist). DSR constraint 3 restricts the solution to node disjoint paths and should be interpreted that the partial ring solution set $R(M)$ must contain each multicast node exactly once and each node in the complete ring solution set $R(M)^*$ should be listed exactly once with no nodes repeated (i.e., in the ring solution each node - multicast node or Steiner Point node - has exactly two incident links).

Our DSR formulation is part of a well-known class of algorithms based on the straightforward idea of enumerating all the permutations of the nodes in M and then checking each of them for inducing a Hamiltonian cycle [22]. There are, however, several ways to significantly reduce the search space. First, consider only valid rings (Hamiltonian cycles). Exhaustive search includes many disconnected and otherwise malformed rings that are not Hamiltonian cycles. Second, consider only the $\frac{(m-1)!}{2}$ non-isomorphic[8] Hamiltonian cycles. Third, as the permutations are generated in lexicographic order then many hopeless candidates can be bypassed immediately and not enumerated (i.e., branch-and-bound). For instance, adjacent nodes in the ring solution should be "nearby neighbors" such that nodes in the ring solution proposed to be adjacent should be close in terms of cost or in terms of a hop-limit heuristic. Fourth, as potential Hamiltonian cycles are being generated, inclusion of certain paths may force the exclusion of other paths which may reduce the total number of non-isomorphic Hamiltonian cycles.

Enumerating all non-isomorphic Hamiltonian cycles in a graph is an infamous graph theory problem referred to as the "necklace-problem" or the "keys-on-a-ring problem".[9] The solution enumerates only the Hamiltonian cycle permutations whose first term is some prescribed node (i.e., node 0). All other cycles which start at other labeled nodes are isomorphic to a cycle whose first term is node 0 as can be visualized by the ring being "rotated". This reduces the search space from $m!$ to $(m-1)!$. Then enumerating only the permutations whose labels are in clockwise (or counterclockwise) order around the cycle (referred to as "flipping") reduces the search space from $(m-1)!$ to $\frac{(m-1)!}{2}$. Flipping is possible because we have retained the symmetric TSP problem for a bidirected link assumption (i.e., costs of reverse paths are equal). For an unidirected link assumption, the non-isomorphic search space remains $(m-1)!$ without flipping.

The importance of restricting the search space is due to the phenomenal rate of growth of the computation time as m increases. For m of quite moderate values, $m!$ is too large to make the computation feasible. In fact, determining whether or not a ring even exists shorter than a specified length is an NP-complete problem (an important problem for delay-constrained multicasting).[10] There is no known way to verify "yes" to the existence of a ring less than a specified length without examining all possible rings (or a large proportion of them). In other words, no polynomial time nondeterministic algorithm for the minimum cost ring problem is known.

[8] Two graphs $G_a = (N_a, L_a)$ and $G_b = (N_b, L_b)$ are said to be isomorphic if there exists a one-to-one correspondence between their node sets $N_a \leftrightarrow N_b$ and a one-to-one correspondence between their link sets $L_a \leftrightarrow L_b$ such that the corresponding links in G_a and G_b are incident on the same corresponding nodes in G_a and G_b.

[9] see *Graphical Enumeration* by Frank Harary and Edgar Palmer, Academic Press, 1973, pages 44-46 or *Introductory Combinatorics* 2nd edition by Kenneth Bogart, HBJ Publishers, problem 65 on page 13

[10] see reference [14] p. 234, reference [2] pp. 790-793, reference [13] pp. 116-117

5 Comparative Evaluation

Here we report the results of experimentation to compare the usefulness of preplanned dedicated backup route restoration techniques to provide survivability for ATM group communications implemented via the Virtual Ring (VR), the Shared Multicast Tree (SMT), and the VC MESH Model (VCMESH). The MCS model was not included because of the obvious single-point-of-failure vulnerability which requires a different set of restoration techniques. To enable comparisons of feasibility and cost we make the following assumptions: (1) only single faults (single link/node) in a single group are considered; (2) the bandwidth of any traffic flow is less than or equal to link capacity; (3) all traffic flow within a group is symmetric with equal bandwidth and QoS requirements; and (4) all groups have equal traffic flow. We simulated two different cases of the cost function C_i in our simulations. In the first case, experiments one and two assign each link equal cost such that the cost function C_i defaults to a link count metric per path. Assuming all links have equal cost implies that the number of links involved in a group is linearly related to cost. In the second case we assign random link costs. The experiments were implemented in MATLAB on a Sun Enterprise 4000 with ten 250 MHz ULTRASparc-II CPUs and 2.5 GB of RAM.

We consider two different network configurations used in [23, 9, 24, 15] as shown in Figure 3. Both configurations include switching nodes that can originate and terminate traffic, as well as cross-connect nodes that relay traffic. A 10 switching-node network with 18 cross-connect nodes and 27 bi-directional transmission links will be referred to as NET-1. A 15 switching-node network with 23 cross-connect nodes and 33 bi-directional transmission links will be referred to as NET-2.

Generating all possible paths between every node pair up to the inclusion of all network nodes results in 16,178 paths for NET-1 and 63,880 paths for NET-2. No path length or "hop-limit" constraint below the number of "hops" (nodes) in each network was utilized so that cost of each technique could be evaluated independently of unavailable backups due to long backup path lengths. For NET-1, the average number of eligible paths between node pairs is 179.76 and the average length of each of these eligible paths is 11.36 links. For NET-2, the average number of eligible paths between node pairs is 304.19 and the average length of each of these eligible paths is 14.17 links.

For NET-1 and NET-2, ATM multipoint groups of specified sizes are formed for minimum cost using the VR, SMT, and VCMESH techniques. First a minimum cost "working" ring (VR), tree (SMT), or group (VCMESH) is found for each technique respectively. The term "working" refers to an individual ring, tree, or group initially established for communication. The term "backup" refers to an individual ring, tree, or group that is reserved for restoration and not used otherwise (i.e., dedicated bandwidth). When a fault occurs with a link or node within a "working" group, traffic flow is automatically rerouted to the corresponding "backup" group. For the case of the VC Mesh and the shared multicast tree, the "backup" group is link and node disjoint such that a single fault in the "working" group will not affect the "backup" group.

Minimum cost reconfigurable Virtual Rings (working and backup rings over the same links and nodes) are calculated by an implementation of the DSR formulation we have presented. The SMT approaches leave the issue of how to build the shared tree to an external routing protocol so we have assumed the best case scenario by identifying the minimum cost (Steiner) tree [19]. The VCMESH approach specifies that each sender establishes circuit(s) to connect with all other group members but does not specify the mechanism so we have assumed the best case scenario such that each sender optimally selects the set of point-to-multipoint/point-to-point circuits that minimizes cost. Minimum cost working and disjoint backup shared multicast trees and VC Mesh groups are calculated using a implementation variant of a Steiner Tree procedure due to Lawler known as the spanning tree enumeration algorithm [6].

Table 1A shows the first set of experiments we performed on NET-1 and NET-2. In the first set of experiments we found all the possible VRs on two different networks for varied group sizes in order to show consistency of VR results for different network topologies. Group sizes selected (three and four out of a possible ten nodes) are consistent with the range of multipoint group sizes relative to network size suggested in the literature (i.e., 10-60%). A given group size for different networks will result in a different number of simulated groups. For example, for a group size of 3 there are 120 simulated groups for NET-1 (i.e., combination of 10 switches taken 3 at a time) and 455 simulated groups for NET-2.

Table 1B shows the second set of experiments we performed. In the second set of experiments we compared the feasibility and cost of the different restoration techniques on NET-1 for group sizes of three and four. Using the VCMESH technique each member of a group is an eligible source which can optimize circuits to the other members of the group. Thus for a group size of 3, there are 3 possible sources and thus 3 possible source-receivers circuits.

Table 1C shows the third set of experiments we performed. In the third set of experiments we modified the cost function to assign random weights to each link within NET-1 and NET-2 in order to show consistency of feasibility and cost comparison results given different link cost assignments. Actual link costs may be preprocessed or computed from network information. The path cost, represented by the cost function C_i, may consist of multiple links with different costs reflecting possible distance, bandwidth, and economic actual costs. Figure 3 shows two different random link cost assignments ("A" and "B") for NET-1 and NET-2 respectively. Link cost assignments "A" and "B" contain different random permutations of the same random weights for each network.

Table 2 shows results from the first set of experiments comparing the feasibility and cost of VRs of different group sizes (3,4) on two different networks (NET-1, NET-2). We assign all links equal cost for this set of experiments. We found VRs of both sizes were 100% feasible on both networks (no cases where a VR was unavailable). The average and extreme values of all reconfigurable (working and backup) VRs show consistency from NET-1 to NET-2. The average cost of a Virtual Ring for group size = 3 and group size = 4 on NET-1 are below the corresponding average cost of a Virtual Ring on NET-2 by a constant value (approximately 4 cost units). Since NET-2 is a larger network (more nodes, larger diameter), it is expected that the VRs on NET-2 are higher in cost than the VRs on NET-1.

Tables 3 - 5 show the results from the second set of experiments comparing the feasibility and cost of the different restoration techniques (VR, SMT, VCMESH) on NET-1 for different group sizes. We assign all links equal cost for this set of experiments.

Table 3 shows results comparing the feasibility of using VR, SMT, and VCMESH techniques to provide survivability on NET-1 for group sizes = 3, 4. Note that VRs are 100% feasible on NET-1 but this is not the case with the SMT and VCMESH techniques. The feasibility of providing survivability begins for a group size of 3 at approximately 88% and 58% via the SMT and VCMESH techniques respectively. As group size increases, the percentage of available disjoint backup SMTs and VCMESH groups decreases commensurately. Table 3B shows the feasibility of restoration for the different techniques for all groups combined.

The explanation for the poor feasibility of the SMT and VCMESH techniques in NET-1 is the existence of "multipoint traps". A "multipoint trap" is a topology where a corresponding backup

SMT/VCMESH is not available due to the disjointness constraint. For example, Figure 4 shows a minimum cost "working" SMT for two groups on NET-1. The removal of the links and nodes for these "working" SMTs leaves no potential disjoint backup SMTs.

"Multipoint traps" become more prevalent as the number of group members increases because more links must be used for the "working" SMT/VCMESH leaving less links for forming a disjoint "backup" SMT/VCMESH. "Multipoint traps" also explain why the VCMESH technique has a dramatically lower feasibility than the SMT technique. For VCMESH restoration to be available, all group members must form a "working" and corresponding "backup" set of circuits (point-to-point/point-to-multipoint) to all other members of the group. For example, a group size of 3 requires 3 sets of corresponding "working"/"backup" circuits. If a "multipoint trap" exists for just one of the group members then VCMESH restoration will not be available.

"Multipoint traps" arise as a function of how the "working" and "backup" SMTs and VCMESH groups are sequentially derived. A "working" SMT/VCMESH is computed independent of the availability of a corresponding "backup" SMT/VCMESH. A minimum cost "working" SMT/VCMESH is first computed for each group of member nodes and then second a disjoint "backup" SMT/VCMESH is computed based on the identified "working" SMT/VCMESH group. Algorithmically, the disjoint "backup" SMT/VCMESH is computed by a Steiner Tree procedure considering the remaining network after the removal of the links and nodes of the "working" SMT/VCMESH. We found many instances where choosing the minimum cost "working" SMT/VCMESH results in a "multipoint trap" topology while selecting a higher cost "working" SMT/VCMESH would result in a feasible disjoint "backup" SMT/VCMESH. Elimination of "multipoint traps" and a least cost restoration solution can be obtained if both "working"/"backup" SMT/VCMESH were to be computed jointly instead of sequentially but such a solution includes multiple NP-complete problems and is thus computationally infeasible. Point-to-point traps are discussed in more detail in [11].

Table 4 shows results comparing the costs of using VR, SMT, and VCMESH techniques to provide survivability on NET-1 for group sizes = 3, 4. The working VR/SMT/VCMESH is computed independent of backup availability. The backup and working plus backup cost statistics are only from groups where a corresponding backup is available. General observations include:

- costs monotonically increase as the group size increases
- costs are significantly different at a 95% confidence level (one exception noted)
- cost differences between the techniques diverge as group size increases
- the relative ranking order of each technique by cost is consistent across working, backup, and working plus backup measurements: VR (lowest cost), SMT (middle cost), and VCMESH (highest cost)
- the cost of the working VR and backup VR is equal
- the cost of the <u>backup</u> SMT/VCMESH is dramatically higher than the cost of its corresponding <u>working</u> SMT/VCMESH

Table 5 shows results of total cost and individual cost comparisons between the VR, SMT, and VCMESH techniques to provide survivability on NET-1 for group sizes = 3, 4. The network manager and the individual user both share a primary goal in the feasibility of restoration techniques. If a given restoration technique is unavailable for a given group due to a "multipoint trap" then cost is irrelevant. From a network manager's viewpoint, however, the secondary goal is to minimize the total cost of forming all possible groups on a given network. From an individual user's viewpoint, however, the secondary goal is which technique has the highest probability that it will deliver with the least cost for a particular instance.

Table 5A and 5B investigate the secondary goal from the network manager's viewpoint. Table 5A shows a three-way total cost comparison between the different techniques with statistics only from groups where all three corresponding backups are available (VR/SMT/VCMESH). For both group sizes (3 and 4), the VR technique has a lower total cost than the SMT technique (20.40%) and a lower total cost than the VCMESH technique (56.50%). Table 5B shows a two-way total cost comparison between the VR and SMT techniques with statistics only from groups where both corresponding backups are available.[11] For both group sizes (3 and 4), the VR technique has a lower total cost than the SMT technique (29.03%).

Table 5C investigates the secondary goal from the individual user's viewpoint by showing a one-to-one individual cost comparison between the VR and SMT techniques.[12] For an individual group, which technique has the highest probability that it will be preferable in terms of availability and cost? The results show the percentage of groups where VR restoration is both available *and* least cost. A VR is considered preferable regardless of cost when a corresponding SMT backup is unavailable. For both group sizes (3 and 4), the VR technique is both feasible and least cost in 98.39% of the cases (328/330). Figure 5 shows the only two cases where the SMT technique is less costly than the VR technique.

Table 6 shows the results from the third set of experiments comparing the feasibility of the different restoration techniques on NET-1 and NET-2 for different link cost assignments. We consider only group size equal to three for this set of experiments. Note VRs of either link cost assignment were 100% feasible on NET-1 and NET-2 but again this is not the case with the SMT and VCMESH techniques. For NET-1, the feasibility of providing survivability via the SMT and VCMESH techniques begins at approximately 82% and 57% respectively. For NET-2, the feasibility of providing survivability via the SMT and VCMESH techniques begins at approximately 71% and 56% respectively. Consistent with results from experiment two, experiment three results show that the relative ranking order of each technique by feasibility remains the same for both random link cost assignments tested: VR (100% feasibility), SMT (middle feasibility), VCMESH (lowest feasibility). It is observed that changing the link cost assignments does not change the relative feasibility of the different techniques but rather changes the relative number and location of multipoint traps. Further results show that the average, extreme, and relative ranking order of costs for the different techniques for the tested networks (NET-1, NET-2) are the same independent of link cost assignment. A one-to-one individual cost comparison between the VR and SMT techniques for random link cost assignments results in the VR technique being both feasible and least cost in 95% of the cases for NET-1 and 98% of the cases for NET-2.

[11]Note there are more groups measured in Table 5B than Table 5A since Table 5A excludes many groups lacking a VCMESH backup.

[12]A one-to-one individual cost comparison between the VR and VCMESH techniques or the SMT and VCMESH techniques is not needed because the VCMESH will be dominated in either scenario since the VCMESH is always unavailable if an SMT backup is unavailable and the VCMESH is the highest cost in all cases measured.

6 Summary

Preplanned, dedicated bandwidth, disjoint backup route techniques have proven to be very effective for the restoration of ATM point-to-point Virtual Paths and Virtual Circuits. This research has investigated the extension of these techniques for the restoration of ATM group communications implemented via Virtual Ring multicasting, shared multicast trees, and the VC Mesh Model. When a single link or single node fault occurs within a working group, traffic flow is rerouted to a corresponding backup group that is link and node disjoint with dedicated bandwidth reserved for restoration.

Our results show that for the two networks studied the use of disjoint dedicated backup, while 100% feasible for the Virtual Ring technique, is often not possible for SMT and VCMESH techniques. Guaranteed survivability begins for a group size of the three at only about 88% for SMT and 58% for VCMESH and decreases rapidly for larger group sizes due to "multipoint trap" topologies. These results are consistent across different networks, group sizes, link cost assignments, and also intuitive because as group size increases, the SMT/VCMESH techniques find it harder to identify *disjoint* paths linking all group participants.

Given the use of disjoint dedicated backup SMT/VCMESH techniques to provide ATM multipoint survivability is not feasible for groups on arbitrary networks, there still exist densely connected networks where these techniques can provide guaranteed survivability. By being identified in advance, some groups may have the option of being protected using either of the VR/SMT/VCMESH techniques. For these groups only, the question of restoration cost is relevant. Results show that for the two networks studied the relative ranking order of each technique by cost is consistent across different networks, group sizes, and link cost assignments: VR (lowest cost), SMT (middle cost), and VCMESH (highest cost).

We conclude that providing survivability to ATM multipoint communications via Virtual Ring multicasting is both feasible and the least cost preplanned backup restoration technique. The underlying DSR formulation we have introduced provides a theoretical basis upon which VRs can be implemented. In its essence, the DSR formulation solves for both working and backup jointly via a reconfigurable dual ring. The "multipoint traps" inherent in the SMT and VCMESH restoration techniques are due to the independent identification of working and backup groups sequentially.

Lastly, in this paper we presented exact theoretical comparisons between techniques to provide ATM multipoint survivability. Future research is needed to develop less complex but accurate VR heuristics that can be implemented on large networks, bandwidth provisioning procedures to allocate capacity for dual reconfigurable VRs on actual networks, and approaches to integrate these results for delay-constrained and capacity-constrained dual reconfigurable VRs.

References

[1] S. Ahn, "A Fast VP Restoration Scheme Using Ring-Shaped Sharable Backup VPs," in *Proceedings of IEEE Global Telecommunications Conference* (GlobeCom'97), Phoenix AZ USA, 1997, pp. 1383-1387.

[2] R.K. Ahuja, T. L. Magnanti, and J.B. Orlin, *Network Flows: Theory, Algorithms, and Applications*, Prentice Hall, NJ, 1993.

[3] J. Anderson, "Virtual Path Group Switching - A Method for Fast ATM Network Survivability," *Bell Labs Technical Journal*, Spring 1997, pp. 213-233.

[4] G. Armitage, "Support for Multicast Over UNI 3.0/3.1 Based ATM Networks," *IETF RFC*, rfc2022.txt, November 1996.

[5] ATM Forum Technical Committee, *ATM User-Network Interface (UNI) Signaling Specification Version 4.0*, ATM Forum Document af-sig-0061.000, July 1996.

[6] V. K. Balakrishnan, *Network Optimization*, Chapman & Hall, London U.K., 1995.

[7] M. Baldi, T. Ofek, and B. Yener, "Adaptive Real-Time Multicast," in *Proceedings of IEEE INFOCOM'97* Kobe, Japan, 1997.

[8] S. Cosares, D. Deutsch, I. Saniee, and O. Wasem, "SONET Toolkit: A Decision Support System for Designing Robust and Cost-Effective Fiber Optic Networks," *Interfaces*, Vol. 25, No. 1, January/February 1995, pp. 20-40.

[9] R. Cotter, D. Medhi, and D. Tipper, "Traffic Backlog and Impact on Network Dimensioning for Survivability for Wide-Area VP-Based ATM Networks," in *Proceedings of the 15th International Teletraffic Conference* (ITC 15), Washington D.C., 1997, pp. 691-700.

[10] R. Doverspike and B. Wilson, "Comparison of Capacity Efficiency of DCS Network Restoration Routing Techniques," *Journal of Network and Systems Management*, Vol. 2, No. 2, 1994, pp. 95-123.

[11] D. Dunn, W. Grover, and M. MacGregor, "Comparison of k-Shortest Paths and Maximum Flow Routing for Network Facility Restoration," *IEEE Journal on Selected Areas in Communications*, Vol. 12, No. 1, January 1994, pp. 88-99.

[12] Z. Dziong, *ATM Network Resource Management*, McGraw-Hill, New York, 1997.

[13] M.R. Garey and D.S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W.H. Freeman and Company, New York, 1979.

[14] A. Gibbons, *Algorithmic Graph Theory*, Cambridge University Press, Cambridge U.K., 1985.

[15] A. Girard and B. Sanso, "Multicommodity Flow Models, Failure Propagation, and Reliable Loss Network Design," *IEEE/ACM Transactions on Networking*, Vol. 6, No. 1, February 1998, pp. 82-93.

[16] L. Goiuveia and J. M. Pires, "Bounds for a Steiner Ring Network Design Problem with Revenue Costs," in *Proceedings of the 6th International Conference on Telecommunication Systems*, Nashville TN USA, March 1998, pp. 126-130.

[17] F.K. Hwang and D.S. Richards, "Steiner Tree Problems," *Networks*, Vol. 22, 1992, pp. 55-89.

[18] E. Gauthier, J-Y. Le Boudec, and P. Oechslin, "SMART: A Many-To-Many Multicast Protocol for ATM," *IEEE Journal on Selected Areas in Communications*, Vol. 15, No. 3, April 1997, pp. 458-472.

[19] M. Grossglauser and K.K. Ramakrishnan, "SEAM: Scalable and Efficient ATM Multicast," in *Proceedings of IEEE INFOCOM'97*, Kobe Japan, 1997.

[20] Y. Kajiyama, N. Tokura, and K. Kikuchi, "ATM Self-Healing Rings," in *Proceedings of IEEE Global Telecommunications Conference* (GlobeCom'92), Orlando FL USA, 1992, pp. 639-643.

[21] S. Komandur and D. Mossé, "SPAM: A Data Forwarding Model for Multipoint-to-Multipoint Connection Support in ATM Networks," *Sixth International Conference on Computer Communications and Networks* (IC3N), IEEE Computer Society, Las Vegas, NV, 1997.

[22] E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan, and D.B. Shmoys (editors), *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization*, John Wiley and Sons, New York 1985.

[23] D. Medhi, "Models for Network Design, Servicing, and Monitoring of ATM Networks Based on the Virtual Path Concept," *Computer Networks and ISDN Systems*, Vol. 29, 1997, pp. 373-386.

[24] D. Medhi, "A Unified Approach to Network Survivability for Teletraffic Networks: Models, Algorithms and Analysis," *IEEE Transactions on Communications*, Vol. 42 No. 2/3/4, March/April 1994, pp. 534-548.

[25] D. Medhi, "Diverse Routing for Survivability in a Fiber-Based Sparse Network," *IEEE International Conference on Communications* (ICC'91), Denver CO USA, 1991, session 22.4, pp. 672-676.

[26] A. Nitchiporenko, A. Mata-Rodriguez, and J. Yanez, "Efficient Ring Creation Procedure For Design of SDH/SONET Transmission Networks," in *Proceedings of the 4th International Conference on Telecommunication Systems*, Nashville TN USA, March 1996, pp. 473-476.

[27] Y. Ofek and B. Yener, "Reliable Concurrent Multicast From Bursty Sources," *IEEE Journal on Selected Areas in Communications*, Vol. 15, No. 3, April 1997, pp. 434-444.

[28] J.E. van der Merwe and I.M. Leslie, "Service-Specific Control Architectures for ATM," *IEEE Journal on Selected Areas in Communications*, Vol. 16, No. 3, April 98, pp. 424-436.

[29] C.S. Wu, S.W. Lee and Y.T. Hou, "Backup VP Preplanning Strategies for Survivable Multicast ATM Networks," in *Proceedings of IEEE International Conference on Communications 1997* (ICC'97), Montreal, Canada, pp. 267-271.

[30] T-H. Wu, D.T. Kong, R.C. Lau, "An Economic Feasibility Study for a Broadband Virtual Path SONET/ATM Self-Healing Ring Architecture," *IEEE Journal on Selected Areas in Communications*, Vol. 10, No. 9, December 1992, pp. 1459-1473.

[31] T-H. Wu, D.T. Kong, R.C. Lau, "A Broadband Virtual Path SONET/ATM Self-Healing Ring Architecture and Its Economic Feasibility Study," in *Proceedings of IEEE Global Telecommunications Conference* (GlobeCom'92), Orlando FL USA, 1992, pp. 834-840.

[32] Y. Xie, M.J. Lee, and T.N. Saadawi, "Multicasting Over ATM Using Connection Server," in *Proceedings of IEEE International Conference on Communications 1997* (ICC'97), Montreal, Canada, pp. 1376-1380.

[33] B. Yener, Y. Ofek, and M. Yung, "Combinatorial Design of Congestion-Free Networks," *IEEE/ACM Transactions on Networking*, Vol. 5, No. 6, December 1997, pp. 989-1000.

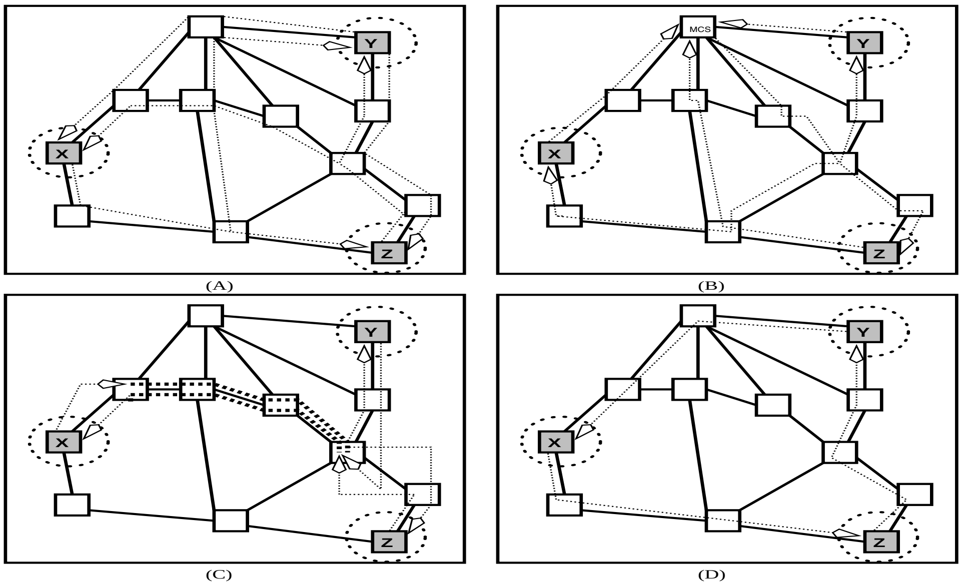


Figure 1: *ATM Group Communication Schemes: (A) VC Mesh; (B) Multicast Server; (C) Shared ATM Multicast Tree; (D) The Virtual Ring*

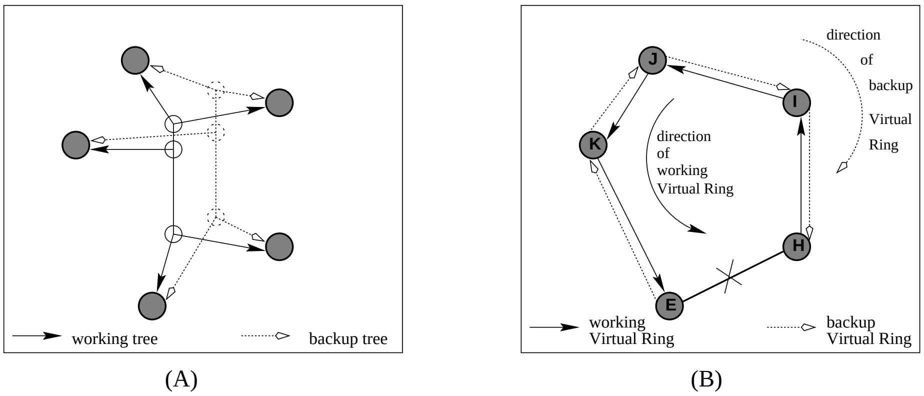
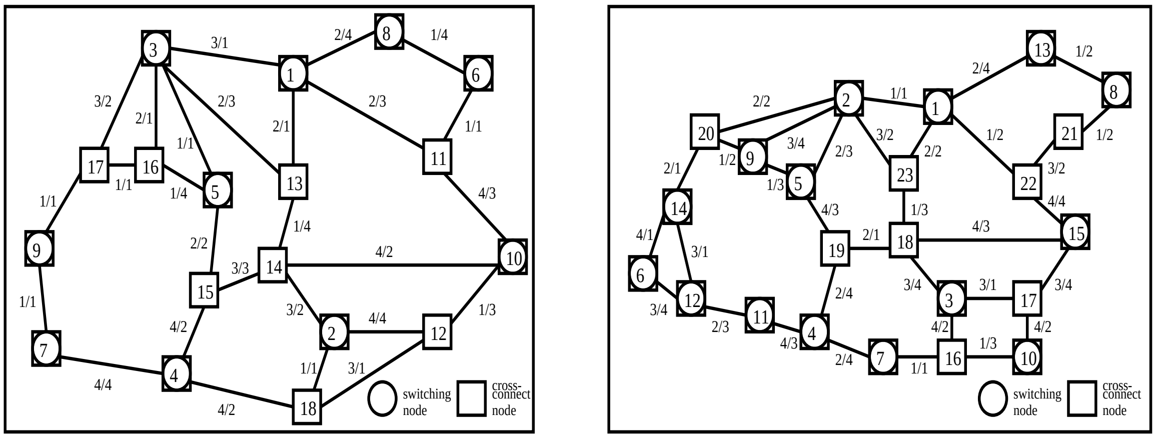


Figure 2: *Restoration Techniques to Provide ATM Multipoint Survivability: (A) a fault in working tree is restored via disjoint backup tree; (B) a Virtual Ring single link fault restoration showing a coordinated ring reconfiguration between a working Virtual Ring and backup Virtual Ring sharing bidirected links (in reverse directions)*



(A) (B)

Figure 3: *Network Topologies: (A) Topology of NET-1 (10 switching nodes, 18 cross-connect nodes) with random link cost assignments "A"/"B"; (B) Topology of NET-2 (15 switching nodes, 23 cross-connect nodes) with random link cost assignments "A"/"B"; (networks used in the following references [23, 9, 24, 15])*

(A)	*GROUP SIZE*	*NET-1 10 switches/18 nodes*	*NET-2 15 switches/23 nodes*	-
	3	120 groups	455 groups	-
	4	210 groups	1,365 groups**	-
(B)	*GROUP SIZE*	*Virtual Ring*	*Shared Multicast Tree*	*VC Mesh Groups*
	3	120 groups	120 groups	120 groups ⇒ 360 ckts.
	4	210 groups	210 groups	210 groups ⇒ 840 ckts.
(C)	*LINK COST ASSIGNMENT*	*NET-1 group size = 3*	*NET-2 group size = 3*	-
	assignment A	120 groups	455 groups	-
	assignment B	120 groups	455 groups	-

** random sample of groups selected to compute a confidence interval with a relative precision $\leq$ 5%

Table 1: *Simulation Experiments: (A) First Set: Virtual Rings on NET-1 and NET-2 for different group sizes; (B) Second Set: comparing restoration techniques on NET-1; (C) Third Set: comparing restoration techniques on NET-1 and NET-2 for different random link cost assignments*

NETWORK / GROUP SIZE	*MEAN COST*	*95% CONF. INT.*	*MIN COST*	*MAX COST*
(A) NET-1 / group size = 3	19.52	(18.79 - 20.24)	8	28
NET-2 / group size = 3	23.56	(23.08 - 24.03)	6	34
(B) NET-1 / group size = 4	22.34	(21.85 - 22.83)	14	30
NET-2 / group size = 4	26.71	(25.41 - 28.00)	14	34

(all values in cost units = number of links utilized)

Table 2: *Experiment One Results - Total Cost to Provide Virtual Ring Survivability: (A) for group size = 3, the cost of all dual reconfigurable working and backup Virtual Rings on NET-1 and NET-2 (all links equal cost); (B) for group size =4, the cost of dual reconfigurable working and backup Virtual Rings on NET-1 and NET-2 (all links equal cost)*

NETWORK / GROUP SIZE	*VIRTUAL RING*	*SHARED MULTICAST TREE*	*VC MESH*
(A) NET-1 / group size = 3	100.00%	88.33%	58.33%
NET-1 / group size = 4	100.00%	78.10%	25.24%
(B) NET-1 / group size = 3 + 4	100.00%	81.81%	32.27%

Table 3: *Experiment Two Results - Feasibility of Providing Virtual Ring Survivability on NET-1 For Varied Group Sizes: (A) availability of disjoint backup restoration using different techniques for group sizes 3 and 4; (B) availability of disjoint backup restoration using different techniques for all group sizes combined*

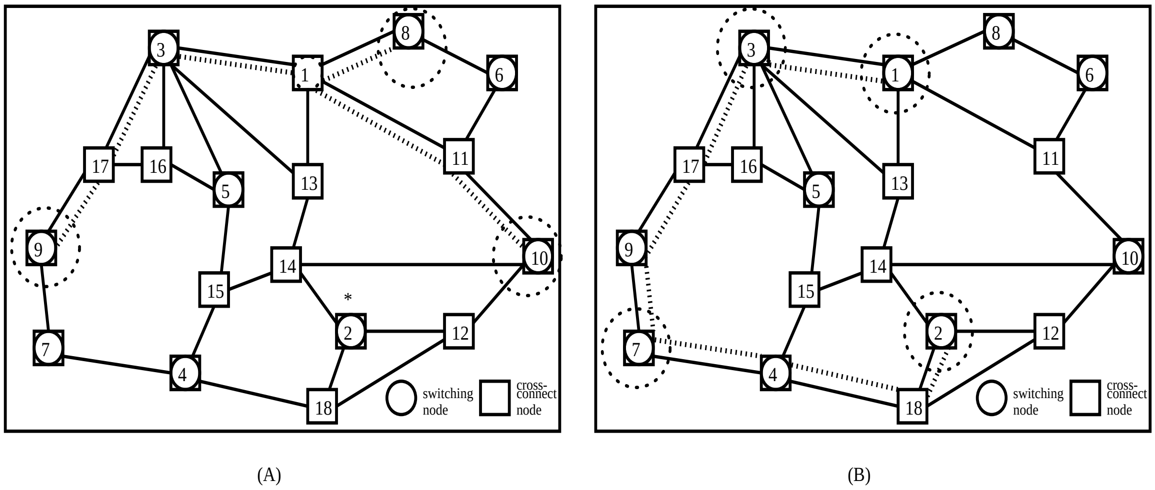


Figure 4: *Experiment Two Results - "Multipoint Trap" Topologies in NET-1 (where a disjoint "backup" SMT does not exist): (A) a "multipoint trap" topology resulting from a "working" SMT for the group of member nodes {8,9,10}; (B) a "multipoint trap" topology resulting from a "working" SMT for the group of member nodes {1,2,3,7}*

RESTORATION TECHNIQUE	*MEAN COST*	*95% CONF. INT.*	*MIN COST*	*MAX COST*
NET-1 / GROUP SIZE = 3				
(A) working VR	9.76	(9.39 - 10.12)	4	14
working SMT	9.55	(9.11 - 9.99)	4	16
working VCMESH	14.33	(13.66 - 14.99)	6	24
(B) backup VR	9.76	(9.39 - 10.12)	4	14
backup SMT	15.32	(14.48 - 16.16)	8	30
backup VCMESH	21.57	(20.47 - 22.67)	12	33
(C) working + backup VR	19.52	(18.79 - 20.24)	8	28
working + backup SMT	24.68	(23.63 - 25.73)	12	40
working + backup VCMESH	35.03	(33.47 - 36.59)	18	48
NET-1 / GROUP SIZE = 4				
(D) working VR	11.17	(10.93 - 11.42)	7	15
working SMT	12.27	(11.96 - 12.58)	6	18
working VCMESH	24.49	(23.91 - 25.07)	12	36
(E) backup VR	11.17	(10.93 - 11.42)	7	15
backup SMT	20.55	(19.63 - 21.47)	12	42
backup VCMESH	37.64	(36.06 - 39.22)	24	50
(F) working + backup VR	22.34	(21.85 - 22.83)	14	30
working + backup SMT	32.52	(31.53 - 33.52)	20	54
working + backup VCMESH	60.70	(58.49 - 62.91)	42	77

Table 4: *Experiment Two Results - Cost Comparisons Between Techniques for Varied Group Sizes on NET-1 (cost/group size) : (A) working/3; (B) backup/3; (C) working + backup/3; (D) working/4; (E) backup/4; (F) working + backup/4; (all values in cost units = number of links utilized)*

GROUP SIZE	*VR*	*SMT*	*VCMESH*	*VR < SMT*	*VR < VCMESH*
(A) group size = 3	1318	1588	2452	17.00%	46.25%
group size = 4	1148	1510	3217	23.97%	64.31%
group size = 3 + 4	2466	3098	5669	20.40%	56.50%
(B) group size = 3	2042	2616	-	21.94%	-
group size = 4	3600	5334	-	32.51%	-
group size = 3 + 4	5642	7950	-	29.03%	-
GROUPSIZE	*# OF GROUPS*	*VR > SMT*	*VR > SMT*	*VR ≤ SMT*	*VR ≤ SMT*
(C) group size = 3	120	2/120	1.66%	118/120	98.33%
group size = 4	210	0/210	0.00%	210/210	100.00%
group size = 3 + 4	330	2/330	1.61%	328/330	98.39%

((A) and (B) total cost values in cost units = number of links utilized)

Table 5: *Experiment Two Results - Cost Comparisons Between Techniques on NET-1 for Varied Group Sizes: (A) Three-Way Total Cost Comparison; (B) Two-Way Total Cost Comparison; (C)One-to-One Individual Cost Comparison*

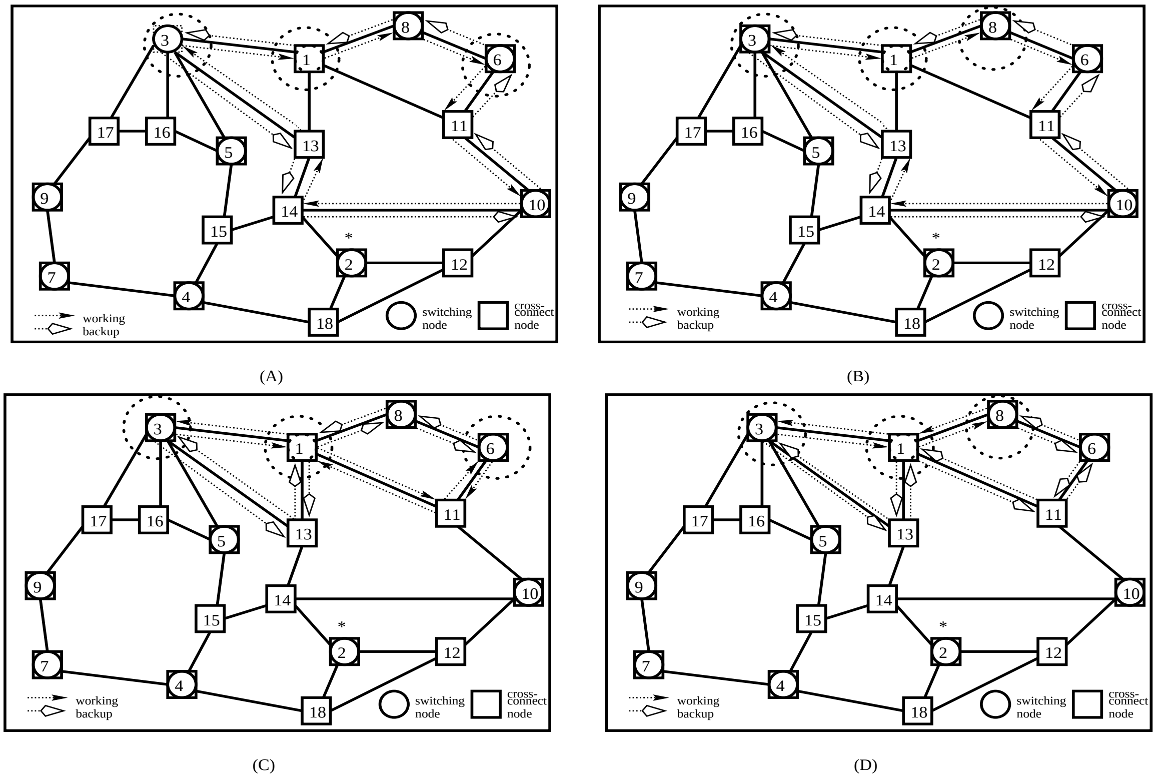


Figure 5: *Experiment Two Results - The Only Two Cases (Out of 330 Cases) Where SMT < VR on NET-1 for group size = 3,4 : (A) working/backup VR for group members {1,3,6}, cost = 16; (B) working/backup VR for group members {1,3,8}, cost = 16; (C) working/backup SMT for group members {1,3,6}, cost = 14; (D) working/backup SMT for group members {1,3,8}, cost = 14;*

	LINK COST ASSIGNMENTS								
NETWORK	*VIRTUAL RING*			*SHARED MULTICAST TREE*			*VC MESH*		
	EQUAL	*"A"*	*"B"*	*EQUAL*	*"A"*	*"B"*	*EQUAL*	*"A"*	*"B"*
(*A*) NET-1	100%	100%	100%	88.33%	85.83%	72.50%	58.33%	60.83%	52.50%
(combined)	*100% (360/360)*			*82.22% (296/360)*			*57.22% (206/360)*		
(*B*) NET-2	100%	100%	100%	78.46%	69.67%	65.05%	59.78%	59.78%	49.67%
(combined)	*100% (1365/1365)*			*71.06% (970/1365)*			*56.41% (770/1365)*		

Table 6: *Experiment Three Results - Feasibility of Providing Survivability on NET-1/NET-2 For Different Link Cost Assignments: (A) availability of disjoint backup restoration on NET-1 using different techniques for link cost assignments = EQUAL, "A", "B"; (B) availability of disjoint backup restoration on NET-2 using different techniques for link cost assignments = EQUAL, "A", "B";*