

# Ethernet DLSw+ Redundancy

This appendix discusses network design issues in a DLSw+ environment with Ethernet-attached end systems. It first reviews SRB. Then it describes how DLSw+ provides redundancy and load balancing when end systems are connected over any media that supports SRB—that is, Token Ring, FDDI, Token Ring LANE, and Token Ring ISL. Finally, it explains the limitations in providing redundancy in Ethernet environments and recommends design techniques for addressing these limitations.

## SRB Redundancy and Load Balancing

SRB allows multiple concurrently active paths in a bridged network. To prevent loops, SRB packets are sent along a route that is specified in the RIF in each packet. The RIF is placed in the packet by the source end system (hence the term source routing). The RIF lists each ring number and bridge along the path.<sup>1</sup>

The source end system discovers the route by use of an SRB explorer. When an LLC2 connection is first established (or for connectionless traffic, when the cache has expired), the source end system sends a TEST or XID frame in an SRB explorer.<sup>2</sup> Every source-route bridge that sees the frame copies the packet to each of its locally attached rings and concurrently updates the RIF to include its bridge number and the next ring number.

If a bridge notices that the next ring matches a ring already in the RIF, the frame is not copied to that ring. This prevents loops in the network. When the explorer reaches the destination, the destination flips a bit to indicate that the RIF should be read in the opposite direction, and it returns the frame to the source. If there are multiple paths through the network, the end system receives multiple responses. The source end system typically selects the first response back (presumed to be the best path at that moment) and uses its RIF for the connection or the cache. In SNA, each SNA PU has its own LLC2 connection, so each PU can potentially find a new path through an SRB network. This enables rudimentary load balancing of SNA traffic across a source-route bridged network.

SRB also allows duplicate, concurrently active MAC addresses in the network, a feature that has made SRB a key component in many SNA network designs. A common technique used in SNA network design is to give multiple FEPs or CIPs the same MAC address. Using this technique, if one of the channel gateways fails, sessions are dropped but are automatically reconnected over the alternate channel gateway. In addition, multiple gateways can be used concurrently to more effectively utilize resources and minimize the impact of any single

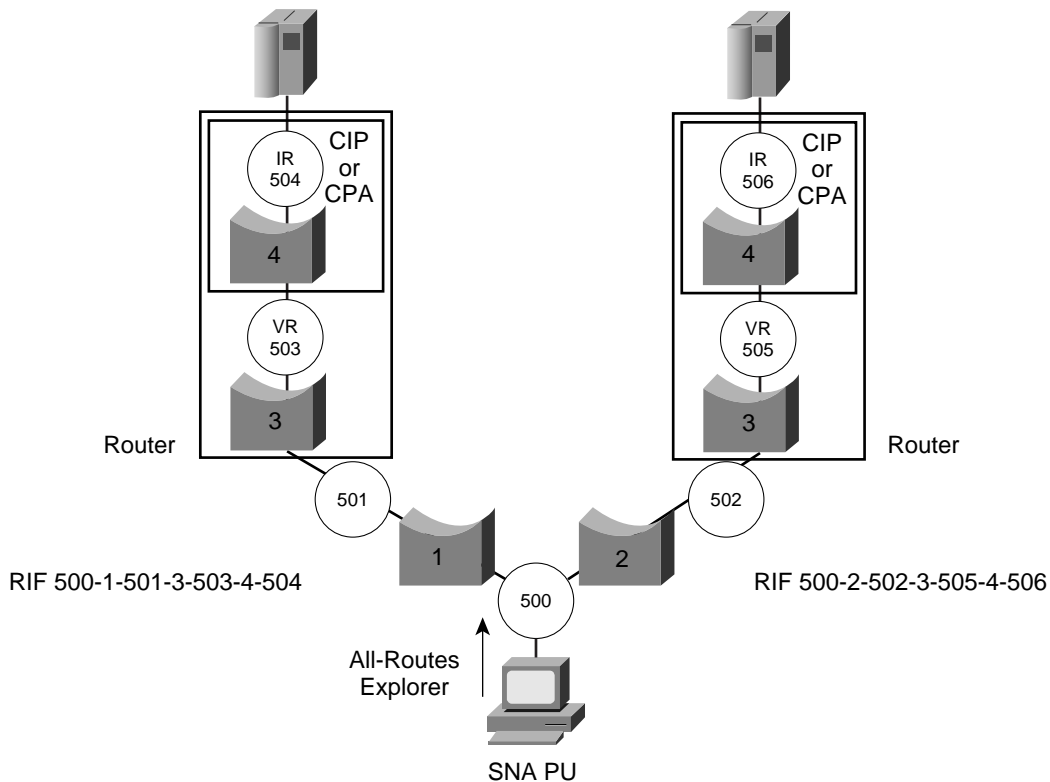
1. Each Token Ring in an SRB network must be assigned a unique Token Ring number, and each bridge connecting the same pair of rings must have a unique bridge number.

2. There are two types of explorers: all-routes explorers and spanning-tree explorers. Source-route bridges copy all-routes explorers onto every possible ring. Source-route bridges copy spanning-tree explorers to only one ring, relying on the SRB spanning-tree algorithm to determine which ring to use. For simplicity, this paper describes only all-routes explorers and describes the typical operation of SNA end systems.

failure. Load balancing in SRB environments is not very deterministic (basically, the first adapter to respond to a TEST frame is the adapter that is used for the duration of the PU connection), but it helps in periods of high traffic where one adapter may get congested. Assigning duplicate MAC addresses to both a FEP and a CIP provides a safe means to migrate from a FEP to a CIP, allowing one to back up the other automatically in the case of a failure.

Figure C-1 shows an example of an SRB network. In this figure, there are two CIPs with the same MAC address. When the SNA PU sends out a TEST frame in an SRB explorer, it gets two responses, each with a different RIF. It uses the first response it receives for the LLC2 connection.

Figure C-1 SRB-Enabled Load Balancing and Redundancy



## DLSw+ Redundancy and Load Balancing

DLSw+ enables and improves upon the capabilities of SRB by using a more deterministic method for load balancing, or alternatively, provides customizable selection of the preferred path. DLSw+ also eliminates session outages in the event of certain failures or error conditions (using IP to reroute around link failures and TCP to retransmit dropped frames or frames in error). Finally, DLSw+ extends the load balancing capability to end systems that connect over serial media or Ethernet (with some limitations).

With DLSw+, remote branch routers can peer to multiple central site routers. If multiple central site routers can reach a given MAC address, there are two ways to control which central site router is used. These methods, or modes, are known as fault tolerant and load balance.

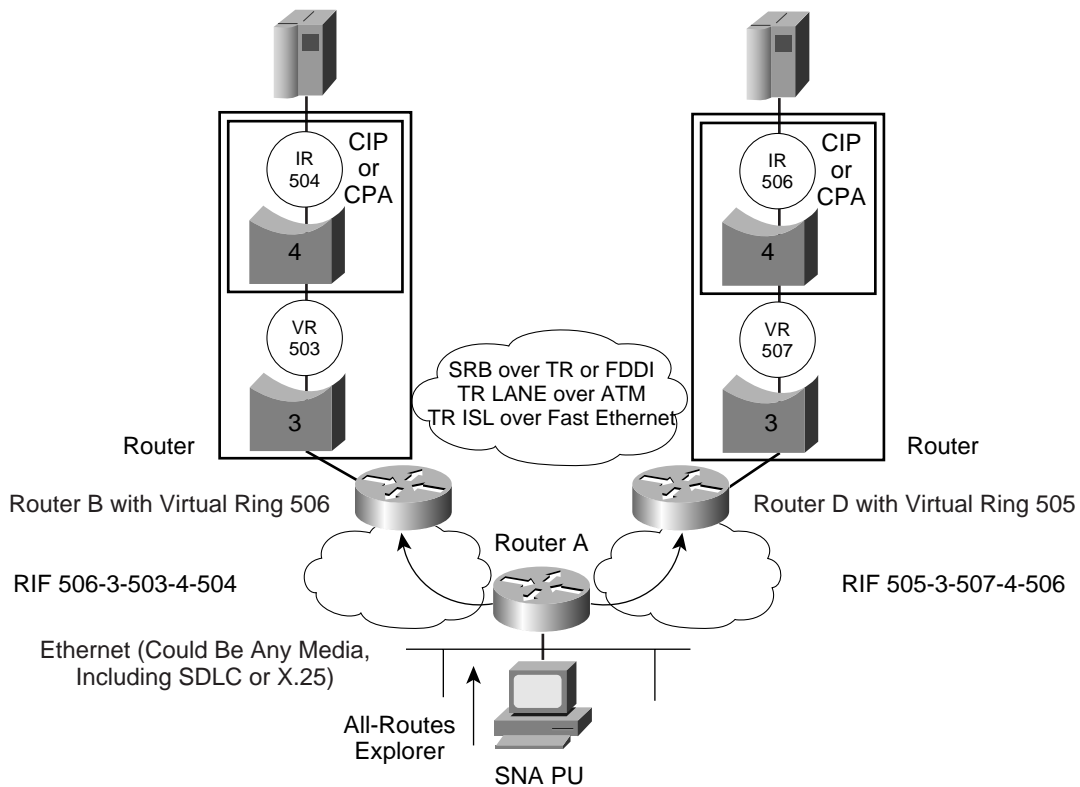
The default is fault tolerant. In this mode, remote routers peer to two or more central site routers. During circuit establishment, the remote router sends a CANUREACH frame to each of its central site peers. The first router to respond is used for the circuit. Alternatively, you can configure the DLSw+ cost parameter to control which router is the preferred peer and which ones are alternates. In either case, if the preferred peer is not available, an alternate one is automatically selected.

If `dls w load-balance` is specified in the remote routers, the load balance mode is used. DLSw+ sets up each new circuit with a different DLSw+ peer, alternating through the list of capable peers in either a round-robin or enhanced load balancing fashion. (See the “Advanced Features” chapter for details.) Each SNA PU uses a unique circuit.

In addition to allowing remote routers to load balance across central site routers, DLSw+ also allows central site routers to load balance across alternate RIFs or Token Ring ports. Simply specify `dls w load-balance` in the central site routers. Although round robin does not provide perfect load balancing, it is vastly better than traditional SRB.

DLSw+ allows branch end systems to attach to DLSw+ over Ethernet (or serial protocols) and still benefit from duplicate Token Ring adapters at a central site. Figure C-2 shows a network where the remote end system is attached to the DLSw+ router over Ethernet, but the upstream SNA device is attached over an SRB-capable medium. Note that the central site DLSw+ routers could attach upstream over either Token Ring, FDDI, Asynchronous Transfer Mode (ATM), or even Fast Ethernet. As long as SRB is supported on the medium (for example, by using Token Ring LANE on ATM or Token Ring ISL on Fast Ethernet), load balancing across duplicate RIFs or SRB ports is supported.

Figure C-2 Load Balancing Using DLSw+



## Transparent Bridging Redundancy

SRB is not supported on Ethernet. Instead, transparent bridging is used. As the name implies, transparent bridges are not visible to the end systems. For that matter, they are not visible to each other. When a frame arrives at a transparent bridge, the bridge has no way to determine where the frame has been. Hence, the only way to prevent loops in this environment is to not have any loops. At any given point in time, there can be only one path between any two MAC addresses on an Ethernet LAN. In addition, duplicate, concurrently active MAC addresses are not supported. These two characteristics of transparent bridging prevent load balancing, but redundancy is still possible. The IEEE Spanning-Tree Protocol allows multiple paths between two points in a transparent bridge environment by ensuring that only one path is active at any given point in time.

A spanning tree is a subset of a transparently bridged network in which exactly one path exists between any pair of nodes. The Spanning-Tree Protocol defines a means for transparent bridges to communicate with each other and determine which ones will be forwarding and which ones will be listening. The forwarding bridges forward frames. The forwarding bridges and Ethernet LANs comprise the spanning tree. The listening bridges simply listen to determine if they need to take over for a forwarding bridge.

The DLSw standard does not support Ethernet IEEE spanning tree.

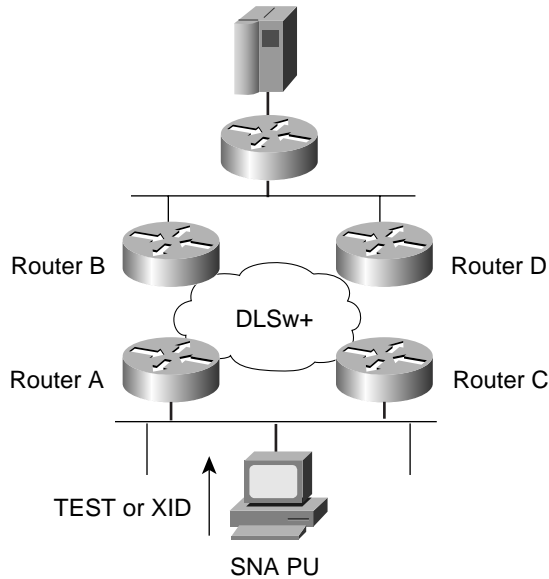
## DLSw+ Redundancy in an Ethernet Environment

As long as the Ethernet end system that initiates a connection has only one way to get to a DLSw+ network, DLSw+ can provide load balancing and redundancy at the central site, as illustrated in Figure C-2. This configuration addresses the requirements of most environments, because most environments only have a single router at a branch. However, if an end system has multiple paths into a DLSw+ network, loops are possible.

## Ethernet-to-Ethernet

Figure C-3 illustrates an invalid configuration with multiple active paths between two Ethernet LANs. In this example, we will assume Router A peers to Router B and Router C peers to Router D. Because all four DLSw+ routers are “forwarding” bridges (from the perspective of the Spanning-Tree Protocol), a packet destined for an unknown MAC address would loop endlessly (if these were standard DLSw+ routers without the plus features).

Figure C-3 Invalid DLSw+ Ethernet Configuration



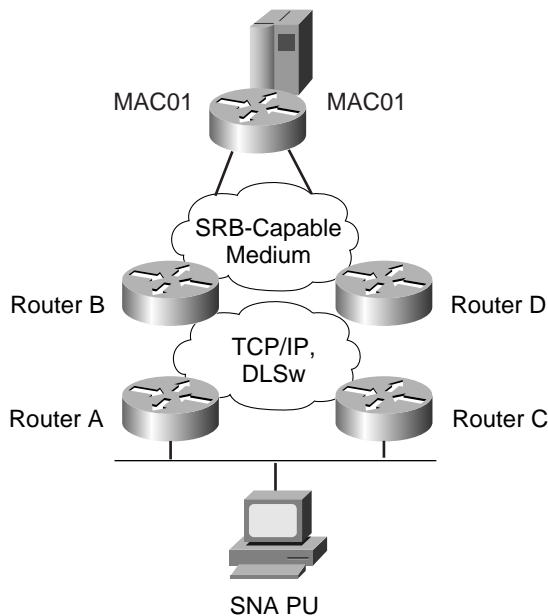
DLSw+ has a number of features designed to minimize unnecessary broadcasts. For example, in the situation described in Figure C-3, DLSw+ prevents endless looping with its explorer firewalling feature. If DLSw+ notes that it is already searching for a particular resource, it blocks subsequent explorers for that resource. Another DLSw+ feature is local resource learning. When DLSw+ determines that a resource is local, it blocks remote searches for that resource. This feature, which is usually quite useful, can create problems in an environment such as the one illustrated in Figure C-3. When the SNA PU issues a TEST frame, it is picked up by both Router A and Router C. For simplicity, we will ignore Router C for now. Router A forwards a CANUREACH frame to Router B. Router B drops the TEST frame on the Ethernet LAN where it is seen not only by the channel gateway, but also by Router D. Router D records the SNA PU as a local resource, even though it is in fact remote. When Router B gets a positive TEST response from the CIP, it forwards an ICANREACH frame to Router A. Router A drops a positive TEST response on the LAN. Router C sees that TEST response from the channel gateway (a CIP in this example) and inappropriately learns that the CIP is local (because the source address of the CIP is picked up in a TEST frame on the local LAN port). Likewise, if Router C gets a positive TEST response, it drops it on the LAN and Router A marks the CIP as local. As long as these entries are in the caches, neither DLSw+ peer forwards CANUREACH frames across the WAN and hence new circuits cannot be established.

Another problem with this configuration is that duplicate circuits can be created for the same SNA session. When the SNA PU sends a TEST frame, both DLSw+ routers see the frame and forward it in a CANUREACH frame to their respective peers. Both Router B and Router D copy the TEST frame onto the Ethernet LAN at the central site, specifying the source address of the SNA PU. The CIP sees two frames and responds to both of them. Both Router B and Router D see the first response and send an ICANREACH frame back to its respective peer and establish a circuit. (Both ignore the second response.) The next frame from the SNA PU is an XID. It is picked up by both Router A and Router C, sent to both Router B and Router D, and then forwarded to the CIP. The CIP, upon receiving two LLC2 packets with the same sequence number, terminates the LLC2 connection and, hence, the SNA connection is never established.

## Using DLSw+ to Connect Ethernet LANs to an SRB-Capable Medium

If one end of the connection supports SRB, the situation is improved, but you can still have problems. In Figure C-4, assume that Router B peers to Router A and Router D peers to Router B. Both sets of DLSw+ peers see the traffic between the client workstation and the CIP. Neither pair of DLSw+ peers is aware of the other, and both pairs set up a circuit to transport the traffic between the client and the CIP. Duplicate packets arrive on the SNA session, resulting in the session being terminated. Also, if the virtual rings in Router B and Router D are not the same, packets dropped on the SRB-capable medium by Router B are picked up by Router D and forwarded back across the DLSw+ network causing a loop. With proper network design and some advanced DLSw+ features, however, you can avoid these problems.

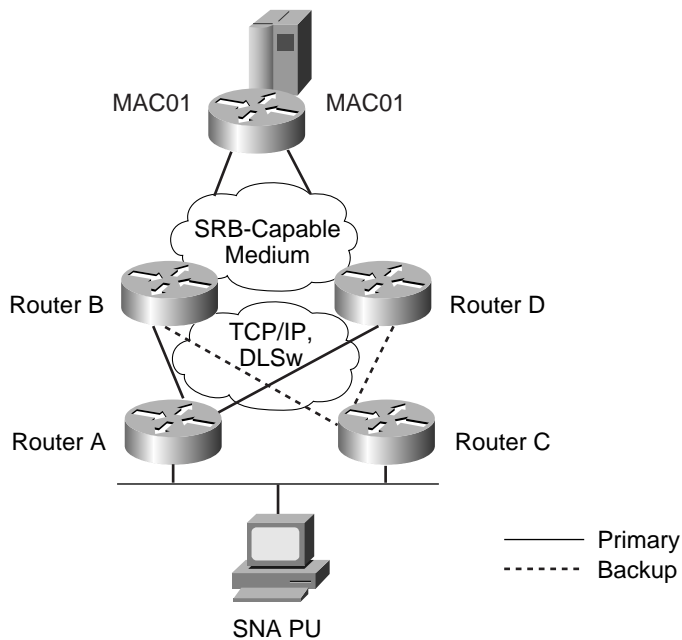
Figure C-4 Ethernet-to-SRB Configuration



To prevent these issues, you can use the network design shown in Figure C-5. Either TCP or FST<sup>3</sup> encapsulation can be used to transport the SNA traffic across an IP network. Router A and Router C are passive (meaning that they do not initialize their peer connections) and promiscuous (meaning that they establish a peer connection with a remote DLSw+ router that was not predefined). Neither router defines any remote peers. Router B and Router D define remote peers and initialize the DLSw+ peer connections. Both Router B and Router D are configured with the same virtual ring number to prevent loops. In both of the central site routers, Router A is configured as a peer, and Router C is configured as a backup peer to Router A, as shown in Figure C-5. The `linger` parameter on the backup peer statement is set to 0. With this configuration, as long as Router A is available, all DLSw+ circuits are between Router A and either Router B or Router D.

3. FST encapsulation with media conversion requires Cisco IOS Release 11.3 or higher.

Figure C-5 Designing SR/TLB for Ethernet Redundancy



To load balance, specify `dls load-balance`. By configuring load balancing in the remote routers (Router A and Router C), they evenly distribute circuits across the central site routers, Router B and Router D. By configuring load balancing in Router B and Router D, they load balance across duplicate RIFs<sup>4</sup> or ports. This also allows you to evenly distribute traffic across the pair of CIPs or FEPs, which appear to SRB to simply be two RIFs to the same adapter or MAC address. By load balancing traffic across a pair of central site routers and a pair of CIPs, any single failure disrupts only half of the network. Hence, recovery is faster.

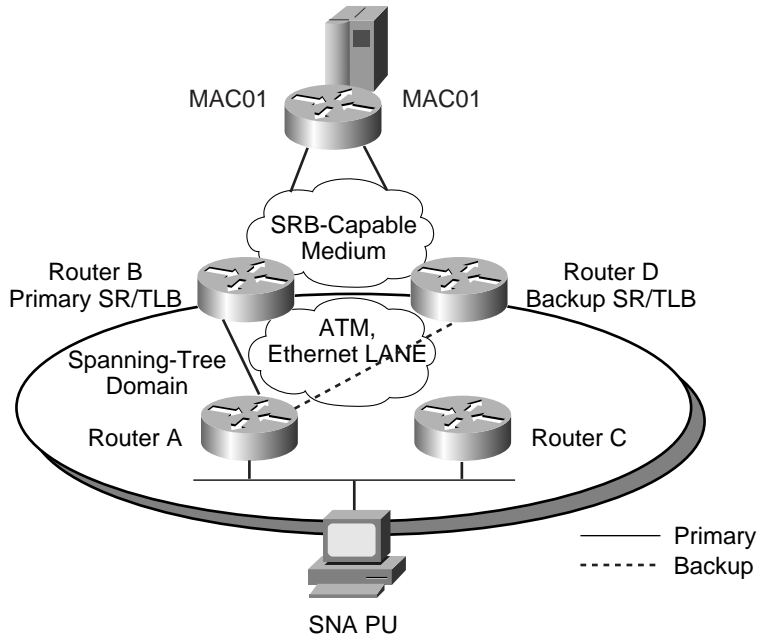
Router A is a single point of failure (that is, sessions are disrupted if you lose Router A). However, if Router A fails, Router C dynamically takes over. If Router A then becomes operational, it is imperative that Router C be disconnected immediately (hence, `linger = 0`) to prevent looping. Unfortunately, this means SNA sessions are disrupted again when the fallback occurs. All recovery is dynamic, requiring no intervention.

## Using ST/TLB to Connect Ethernet LANs to an SRB-Capable Medium

An alternative means to connect Token Ring to Ethernet is SR/TLB. In general, bridging is never a better alternative than DLSw+ across a WAN, but in an ATM environment (for example, across a metropolitan area network), SR/TLB merits investigation. Some ATM environments might use Ethernet LANE to transfer SNA across the ATM backbone and use SR/TLB for media conversion, as illustrated in Figure C-6. This design assumes that both SR/TLB routers are at the central site, the channel-attached router has two CIPs, and each SR/TLB router is connected to both of the CIPs in the central site router.

4. Load balancing across duplicate RIFs requires Cisco IOS Release 11.3(4.1) or higher

Figure C-6 Designing SR/TLB for Ethernet Redundancy



At the central site, only one translational bridge can be active at a time. Manual intervention is required to cause Router D to take over for Router B. At the remote site, only Router A or Router C is forwarding at any one time, relying on spanning tree to determine which is forwarding and which is listening. No manual intervention is required for Router C to take over for Router A.

SR/TLB allows duplicate concurrently active MACs. When a unicast frame is bridged from the transparent bridge domain to the source-route bridged domain, SR/TLB first checks to see if it has an entry in its RIF cache matching the DMAC address. If it does, SR/TLB uses that RIF. If it does not, SR/TLB sends a spanning tree explorer, which finds both of the CIPs. Each CIP responds to the spanning-tree explorer with a directed response. The SR/TLB router caches the first response it receives. All sessions use the RIF in that response and, hence, the same CIP.

If one of the CIPs fails, all sessions using that CIP are disrupted. The RIF cache times out, another explorer is sent out, and the alternative CIP is found. All subsequent sessions are established over that CIP.

Table C-1 compares the trade-offs from each of these solutions.

Table C-1 DLSw+ and SR/TLB Feature Comparison

Feature	DLSw+	SR/TLB
Dynamic Recovery from Failure of Central Site DLSw+ or SR/TLB Router	Yes	No, manual intervention required
Dynamic Recovery from Failure of a CIP/Channel	Yes	Yes (if each SR/TLB router is connected to both CIPs as described previously)
Fallback	Dynamic and immediate but disruptive	Manual intervention required; disruptive, but can be done when convenient
Load Balancing across Central Site Routers	Yes	No
Load Balancing across Duplicate MACs	Yes	No

Table C-1 DLSw+ and SR/TLB Feature Comparison (Continued)

Feature	DLSw+	SR/TLB
Performance	Approximate; assumes TCP encapsulation; FST encapsulation is faster. Cisco 7500-RSP4: 3600 pps at 70% CPU Cisco 7200: 300–3000 pps at 70% CPU Cisco 7200: 150–1500 pps at 70% CPU Cisco 4700: 1400 pps at 70% CPU <sup>1</sup>	<ul style="list-style-type: none"> <li>Fast switched SR/TLB</li> <li>Cisco 7500-RSP2: 14,000 pps (100% CPU utilization), assuming Token Ring to 10-Mbps Ethernet; LANE impact not factored in</li> </ul>
Future	Redundancy to be supported in a future release	No enhancements planned

1. 1400 pps, when sending data from one mainframe to another, can represent significant throughput. DLSw+ is not impacted by packet size, and mainframe-to-mainframe traffic can use packets up to the 1500-byte limit imposed by Ethernet. Hence, this could be 2.1 MB of traffic, depending on packet size.

## Remote Ethernet Switches

If switches are installed at the remote site, there are additional design considerations. Figure C-7 shows an example of a network with remote Ethernet switches. The best design in this case is the design recommended in Figure C-5.

Figure C-7 Ethernet Switches at Remote Site

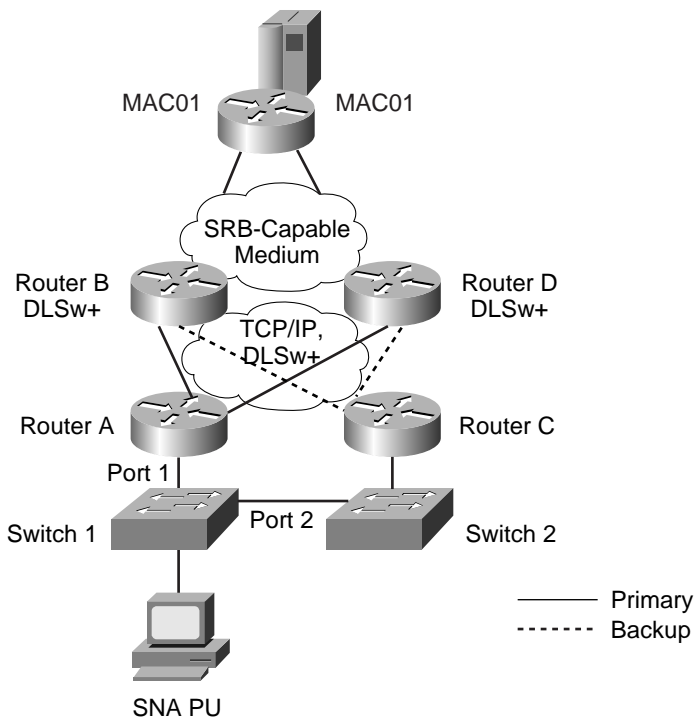


Figure C-7 shows the same configuration as Figure C-5 except that in this diagram, two Ethernet switches have been added. During normal operation, Router B and Router D each establish a peer connection with Router A. When the SNA PU sends the first TEST frame, Switch 1 sends it over both Port 1 and Port 2. Router C has no active peer connections and, hence, merely updates its local reachability table and discards the TEST frame. Router A forwards the TEST frame to Router B and Router D and establishes a circuit with one of them. As part of the circuit establishment, Router A sends a positive response to the SNA PU. When Switch 1 sees the response, it updates its forwarding table (also known as content addressable memory [CAM]) to indicate that the MAC address of the CIP or IBM 3745 (for simplicity, referred to as MAC01) is reachable through Port 1.

If Router A fails, Switch 1 clears its entry for MAC01. (This scenario assumes that the entire router fails.) The next time the SNA PU sends a TEST frame, it is again sent out both ports, but this time, Router C establishes the circuit (let's assume that it uses Router B) and sends the TEST response. Hence, Switch 1 updates its CAM showing that MAC01 is reachable via Port 2. Because Router A is the primary peer and Router C is only the backup peer, Router B continually tries to reestablish a peer connection with Router A. Because linger was not set to 0, if Router A succeeds, it immediately terminates its connection to Router C (and terminates any SNA sessions using that peer connection).

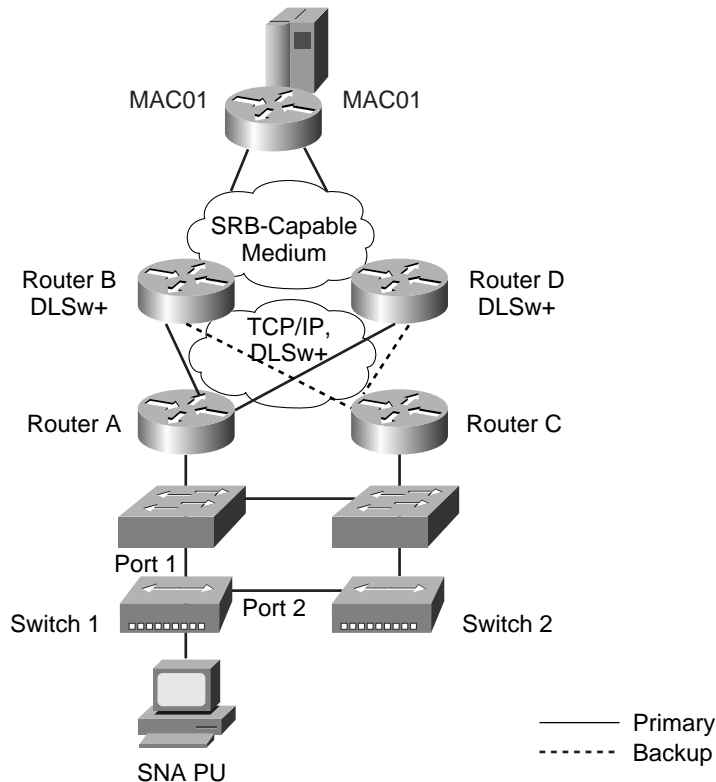
Because the SNA session is terminated, the SNA PU sends out another TEST frame. Unfortunately, it only goes to Router C because the CAM in Switch 1 still points to Port 2. Router C is active, but the DLSw+ peer connection is not active. Hence, until the CAM is cleared, Router A cannot see the TEST frame and the circuit cannot be established.

To minimize the impact of this problem, you can set the CAM aging timers lower. However, note that when the CAM timer for an entry expires, frames destined for that MAC address are flooded to all the ports on a switch, so there is a trade-off that must be considered. Another alternative is shown in the next example.

## Remote Ethernet Switches with Hubs

Figure C-8 is almost the same diagram as Figure C-7, with the addition of two hubs. The hubs allow Switch 1 to reach both DLSw+ routers through either Port 1 or Port 2. Because the hubs introduce a loop in the network, the switches use the Ethernet Spanning-Tree Protocol to ensure that only one of these ports is active at a time. Regardless of which port is active, both Router A and Router C are reachable. Hence, the CAM problem described in the previous scenario is eliminated. As soon as Router A resumes its peer connection to Router B, Router A sees the TEST frames from the SNA PU and establishes the circuit.

Figure C-8 Remote Ethernet Switches Connected by One or More Hubs





## Conclusion

Providing true router redundancy in an Ethernet environment is much more complex and difficult than in an equivalent Token Ring (or to be more precise, SRB) environment. This is because SRB allows multiple active paths and uses the RIF to prevent loops. Transparent bridging, used in Ethernet environments, relies on spanning tree to prevent loops and does not allow duplicate active paths. With careful design, redundancy is possible, although perhaps not optimal.

If you are running Cisco IOS Release 12.0(5)T and later, you can configure the DLSw+ Ethernet redundancy feature. This feature provides redundancy and load balancing when end systems are connected over Ethernet. See Chapter 12 “DLSw+ Ethernet Redundancy Feature” for more details.

