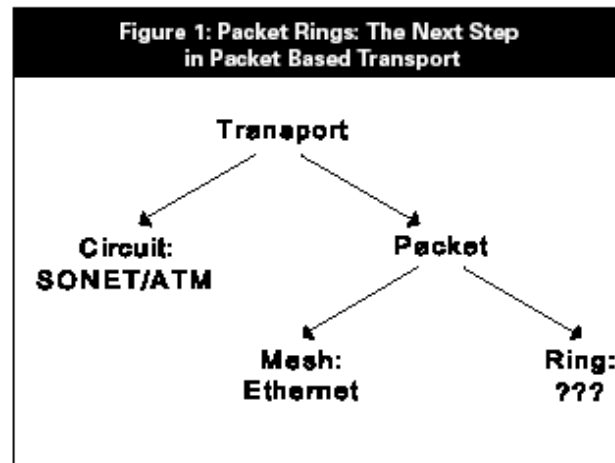


## INTRODUCTION

An important trend in networking is the migration of packet-based technologies from local Area Networks to Metropolitan Area Networks. The rapidly increasing volume of data traffic in metro networks is challenging the capacity limits of existing transport infrastructures based on circuit-oriented technologies like SONET and ATM. Inefficiencies associated with carrying increasing quantities of data traffic over voice-optimized circuit-switched networks makes it difficult to provision new services, and increases the cost of building additional capacity beyond the limits of most carriers' capital expense budgets. Packet-based transport technology, a natural fit with the now ubiquitous IP protocol, is considered by many to be the only alternative for scaling metro networks to meet the demand.



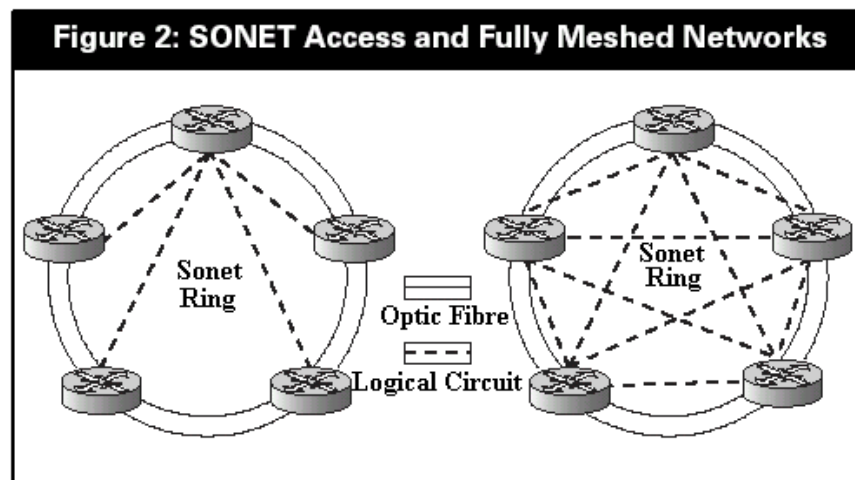
The emerging solution for metro data transport applications is Packet Ring technology. It offers two key features that have heretofore been exclusive to SONET: efficient support for ring topology and fast recovery from fiber cuts and link failures. At the same time, Packet Ring technology can provide data efficiency, simplicity, and cost advantages that are typical to Ethernet. Even though there is currently no standard for Packet Rings operating at Gigabit speeds and higher, many vendors are developing and introducing Packet Ring technologies to address this emerging market.

To be a viable contender for data transport in the MAN, Packet Ring technology should provide support for multi-Gigabit data speeds and integrate seamlessly with existing Ethernet and SONET networks. Packet Ring solutions should be available in various form factors and link speeds, and at prices that are competitive with Ethernet. Finally, an industry standard that defines the link layer for Packet Rings needs to be developed to achieve vendor interoperability and customer acceptance.

## LIMITATIONS OF SONET AND ETHERNET

### SONET

Most metro area fiber is in ring form. Ring topology is a natural match for SONET-based TDM networks that constitute the bulk of existing metro network infrastructure. However, there are well-known disadvantages to using SONET for transporting data traffic (or point-to-point SONET data solutions, like Packet over SONET [POS]). SONET was designed for point-to-point, circuit-switched applications (e.g. voice traffic), and most of limitations stem from these origins. Here are some of the disadvantages of using SONET Rings for data transport:



- Fixed Circuits.

SONET provisions point-to-point circuits between ring nodes. Each circuit is allocated a fixed amount of bandwidth that is wasted when not used. For the SONET network that is used for access in Figure 2 (left), each node on the ring is allocated only one quarter of the ring's total bandwidth (say, OC-3 each on an OC-12 ring). That fixed allocation puts a limit on the maximum burst traffic data transfer rate between endpoints. This is a disadvantage for data traffic, which is inherently bursty.

- Waste of Bandwidth for Meshing.

If the network design calls for a logical mesh, (right), the network designer must divide the OC-12 of ring bandwidth into 10 provisioned circuits. Provisioning the circuits necessary to create a logical mesh over a SONET Ring is not only difficult but also results in extremely inefficient use of ring bandwidth. As the amount of data traffic that stays within metro networks is increasing, a fully meshed network that is easy to deploy, maintain and upgrade is becoming an important requirement.

- Multicast Traffic.

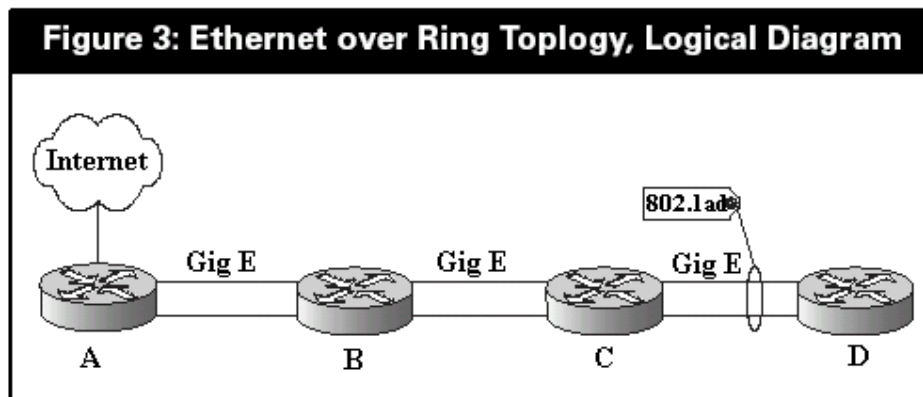
On a SONET Ring, multicast traffic requires each source to allocate a separate circuit for each destination. A separate copy of the packet is sent to each destination. The result is multiple copies of multicast packets traveling around the ring, wasting bandwidth.

- Wasted Protection Bandwidth.

Typically, 50 percent of ring bandwidth is reserved for protection. While protection is obviously important, SONET does not achieve this goal in an efficient manner that gives the provider the choice of how much bandwidth to reserve for protection.

## **Ethernet**

How about Ethernet over a ring? Ethernet does make efficient use of available bandwidth for data traffic, and does offer a far simpler and inexpensive solution for data traffic. However, because Ethernet is optimized for point-to-point or meshed topologies, it does not make the most of the ring topology.

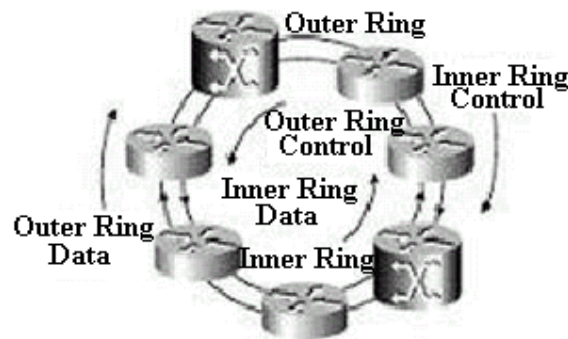


Unlike SONET, Ethernet does not take advantage of a ring topology to implement a fast protection mechanism. Ethernet generally relies on the spanning tree protocol to eliminate all loops from a switched network. Even though spanning tree protocol can be utilized to achieve path redundancy, it recovers comparatively slowly from a fiber cut since the recovery mechanism requires the failure condition to be propagated serially to each upstream node. Link aggregation (802.1ad) can provide a link level resiliency solution, but it is comparatively slow (~500ms vs. ~50ms) and not appropriate for providing path level protection. Ethernet is also not good at implementing global “fairness” policies for sharing ring bandwidth. Ethernet switches can provide link-level fairness, but this does not necessarily or easily translate into global fairness.

## RPR OPERATION

RPR technology uses a dual counter rotating fiber ring topology. Both rings (inner and outer) are used to transport working traffic between nodes. By utilizing both fibers, instead of keeping a spare fiber for protection, RPR utilizes the total available ring bandwidth. These fibers or ringlets are also used to carry control (topology updates, protection and bandwidth control) messages. Control messages flow in the opposite direction of the traffic that they represent. RPR has the ability to differentiate between low and high priority packets. In addition, RPR nodes also have a transit path, through which packets destined to downstream nodes on the ring flow.

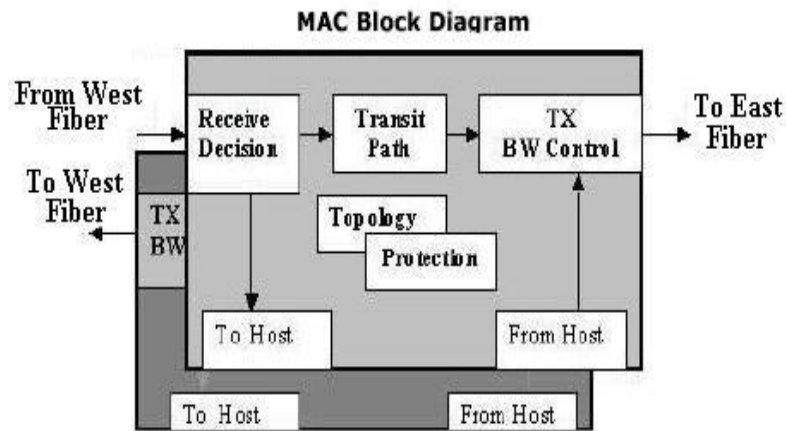
### RPR Terminology



With a transit buffer capable of holding multiple packets, RPR nodes have the ability to transmit high priority packets while temporarily holding other low priority packets in the transit buffer.

### The RPR MAC

As a Layer-2 network protocol, the MAC layer contains much of the functionality for the RPR network. The RPR MAC is responsible for providing access to the fiber media. The RPR MAC can receive, transit, and transmit packets.



## Receive Decision

Every station has a 48-bit MAC address. The MAC will receive any packets with a matching destination address. The MAC can receive both unicast and multicast packets. Multicast packets are copied to the host and allowed to continue through the transit path. Matching unicast packets are stripped from the ring and do not consume bandwidth on downstream spans. There are also control packets that are meant for the neighboring node; these packets do not need a destination or source address.

## Transit Path

Nodes with a non matching address are allowed to continue circulating around the ring. Unlike point-to-point protocols such as Ethernet, RPR packets undergo minimal processing per hop on a ring. RPR packets are only inspected for a matching address and header errors.

## Transmit And Bandwidth Control

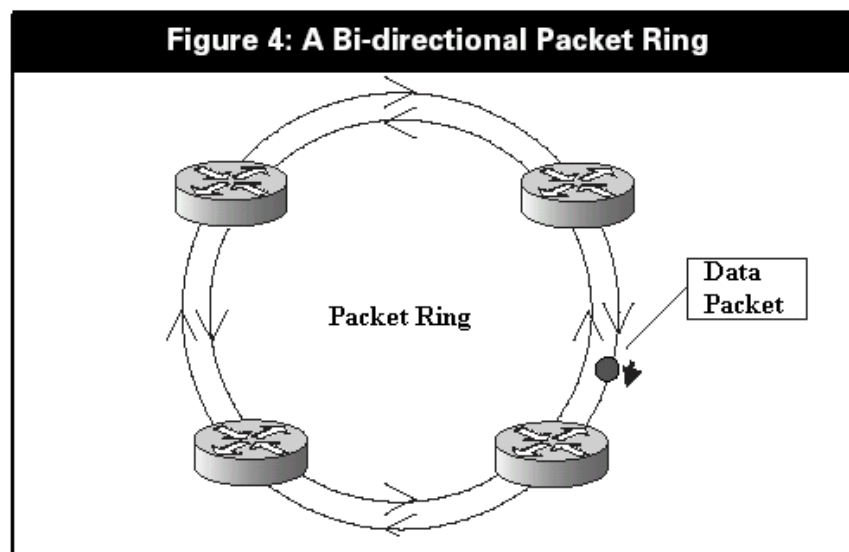
The RPR MAC can transmit both high and low priority packets. The bandwidth algorithm controls whether a node is within its negotiated bandwidth allotment for low priority packets. High priority packets are not subjected to the bandwidth control algorithm.

## **Topology Discovery**

RPR has a topology discovery mechanism that allows nodes on the ring to be inserted/removed without manual management intervention. After a node joins a ring, it will circulate a topology discovery message to learn the MAC addresses of the other stations. Nodes also send these messages periodically (1 to 10 seconds). Each node that receives a topology message appends its MAC address and passes it to its neighbor. Eventually, the packet returns to its source with a topology map (list of addresses) of the ring.

## THE ADVANTAGES OF PACKET RING NETWORKING

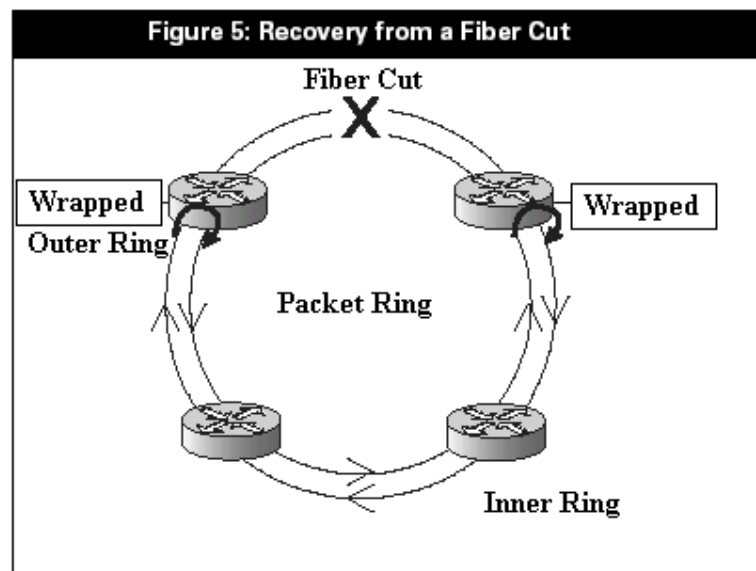
As we've seen, neither SONET nor Ethernet is ideal for handling data traffic on a ring network. SONET does take advantage of the ring topology, but does not handle data traffic efficiently, wasting ring bandwidth. Ethernet, while a natural fit for data traffic, is in fact difficult to implement on a ring, and does not make the most of the ring's capabilities. Packet Ring protocols, on the other hand, promise the best of both worlds. Packet Ring protocols create a full, packet-based networking solution that avoids the provisioning complications and inflexibility of SONET and provides the ring protection and global fairness features missing from Ethernet. Here are the general and specific advantages of Packet Ring networking.



The basic advantage of a Packet Ring is that each node can assume that a packet sent on the ring will eventually reach its destination node regardless of which path around the ring has taken. Since the nodes "know" they are on a ring, only three basic packet-handling actions are needed: insertion (adding a packet into the ring), forwarding (sending the packet onward), and stripping (taking the packet off the ring). This reduces the amount of work individual nodes have to do to communicate with each other, especially as compared with mesh networking where each node has to make a forwarding decision about which exit port to use for each packet.

## Resiliency

Packet Rings have a natural resiliency advantage. A ring that is built using switches needs to distribute failure information across an entire network to recover fully from a fiber cut. In the Ethernet case, this can be accomplished using a spanning tree protocol. On the other hand, a Packet Ring protocol can use a “ring wrap” at the nodes surrounding the cut (see Figure 5).



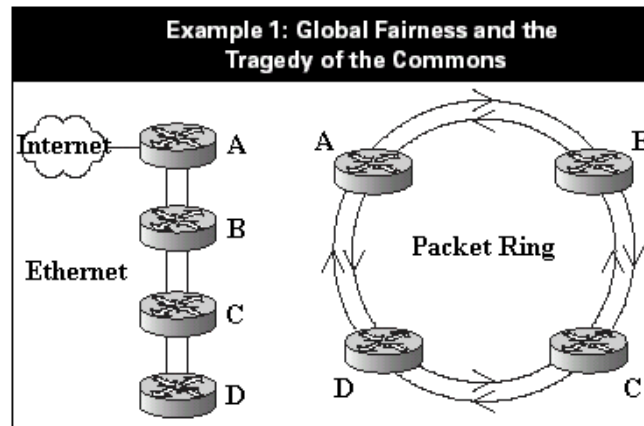
In this case, only nodes that are adjacent to the failure need to take any action. Wrapped traffic can reach the original destination by going around the ring in the opposite direction. Ring fail-over is often described as “self-healing” or “automatic recovery.” In practice, ring-based transport systems have reliably achieved <50ms fail-over periods.

## Bandwidth Fairness

Packet Rings also have an inherent advantage for implementing fairness algorithms to regulate bandwidth usage. Ring bandwidth is a shared resource, and is vulnerable to exploitation by individual users or nodes, in the network version of the “the tragedy of the commons.” A fairness algorithm is a mechanism that gives every node on the ring a predetermined “fair” share of the ring bandwidth, ideally without

the straitjacket of a provisioned circuit. A ring-level fairness algorithm can and should allocate ring bandwidth as one global resource. Bandwidth policies that can allow maximum ring bandwidth to be utilized between any two nodes when there is no congestion can be implemented without the inflexibility of a fixed circuit based system like SONET, but with greater effectiveness than point-to-point Ethernet.

For example, a Packet Ring can simply control the rate of packet forwarding relative to packet sourcing. This is an easy method of preventing neighboring nodes from acting as “bandwidth hogs.” SONET also implements point-to-point circuits that allocate and reserve a fixed amount of bandwidth for each connection, but lack of flexibility is the problem. Adding or subtracting bandwidth requires manual configuration of new circuits, and the reservation of such circuits wastes bandwidth.



In Example 1, the chain of switches on the left, we can see that Node D is vulnerable to the bandwidth usage patterns of Nodes A, B, and C. Ethernet switches typically allocate output port bandwidth fairly among all input ports. If each node, for example, is trying to send 2Gbs traffic to the Internet between the hours of 8pm to 12pm, Node A will be able to send 1Gbs, Node B will be able to send 500Mbs and nodes C and D will only be able to send 250Mbs each. As the number of nodes in the chain grows, the unfairness of Ethernet switches to nodes further upstream becomes even more significant.

One solution is to implement link-level rate limits on each node. For example, the ingress traffic at each switch might be rate limited to 500Mbs. But link-

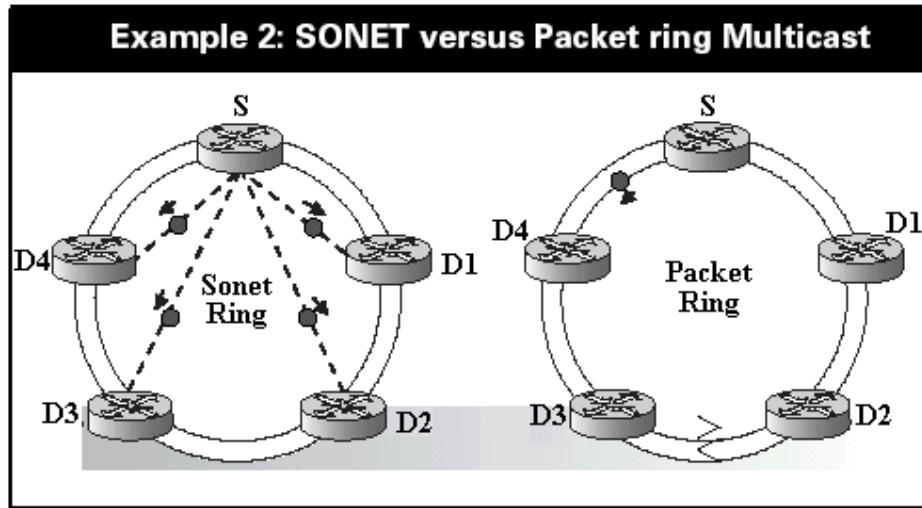
level rate limiting policies don't translate into a global fairness policy and create a provisioning and efficiency problem for best effort data traffic similar to SONET circuits. If at 4am no one else is using any bandwidth, traffic from D need not be limited to 500Mbps.

On the Packet Ring on the right, it is much easier to implement a global fairness policy, because the Packet Ring implements fairness policies at the level of the entire ring, not individual links. The service provider can set rules that govern the rate at which packets are forwarded from upstream or downstream nodes in relation to packets sourced by the node. That way, if none of the ring bandwidth is being used, Node D is free to source as many packets as needed. On the other hand, if Nodes A-C are each using 100Mbps, a Packet Ring can then automatically limit the amount of packets D is allowed to put on the ring, by controlling the source/forward relationship.

### **Broadcast or Multicast Traffic**

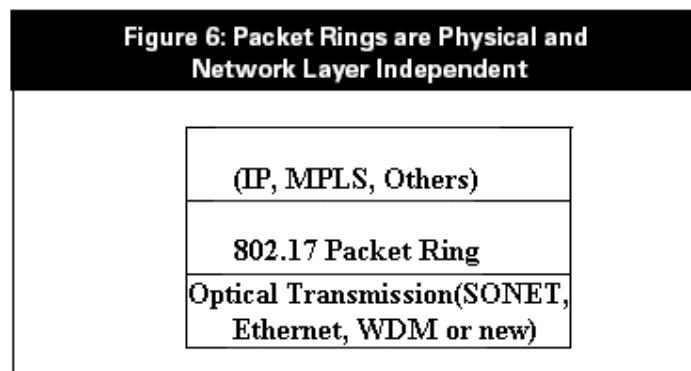
Packet Rings are a natural fit for broadcast and multicast traffic. As detailed above, for unicast traffic, nodes on a Packet Ring generally have the choice of stripping packets from the ring or forwarding them. However, for a multicast, the nodes can simply receive the packet and forward it, until the source node strips the packet. This makes it possible to multicast or broadcast a packet by sending only one copy around the ring.

In Example 2, Source Node S wants to broadcast a packet to destination Nodes D1-D4. Using a POS network, S must replicate the packet and send a separate copy to each provisioned circuit. On a Packet Ring, Source Node S simply sends a single packet onto the ring that is received in turn by each D Node, and forwarded. The Packet Ring, in this example, uses one quarter the bandwidth as the SONET Ring for the same multicast.



**Physical Layer Versatility**

The Packet Ring standards now in development create only a new Media Access Control (MAC) addressing scheme designed for ring-based topologies. This has the advantage of leaving Layer 1 open. Hence, Packet Ring technologies will be compatible with Ethernet, SONET, and DWDM physical layer standards.

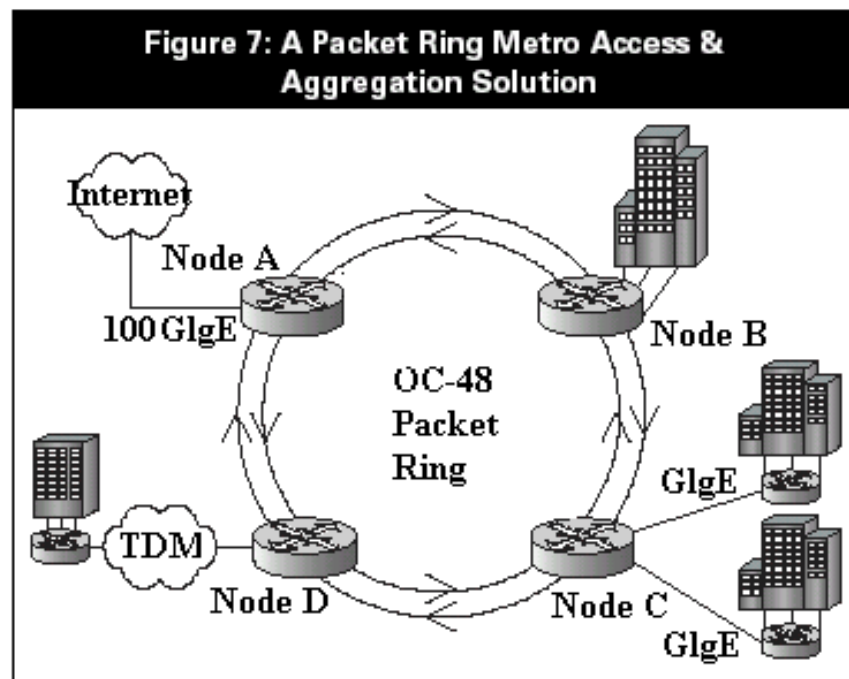


This means that Packet Rings can operate over an existing SONET configuration, creating better data capabilities without the need to replace SONET equipment or abandon SONET for TDM voice traffic. Similarly, a DWDM configuration can make use of Packet Ring at Layer 2 for some or all wavelengths. Finally, where dark fiber is available, a Packet Ring can operate without the need to purchase expensive Layer 1 gear.

## Packet Rings in Application

Packet Rings suit the needs of many types of service providers in the metro area. A Packet Ring is useful anywhere that data transport over a fiber ring is needed. Resiliency and bandwidth sharing are important, but the cost or provisioning complexity of SONET is undesirable.

Let's focus first on a metro access/aggregation solution that might be used by a Metro Service Provider (i.e., an ILEC, CLEC, or BLEC) or (in modified form) a Broadband cable/DSL carrier. These carriers are typically looking to provide data transport services, preferably without the SONET overhead. Packet Ring technology is a perfect fit for these needs.



In this solution, a single Packet Ring serves the needs of multiple buildings in a metro area, over dark fiber (only three buildings are pictured, but in practice a single ring will be capable of supporting dozens of nodes/buildings). This solution has several notable features:

1. The Packet Ring at the center of this solution will allow the service provider to sell true protected bandwidth to the end-users, with a promise of <50 ms fail-over on all services. Each access node provides bandwidth to be added/dropped as desired. Higher layer protocols such as IP or MPLS can be used for service creation, while the resilient transport serves to guarantee service levels.
2. The fairness algorithms built into the Packet Ring will prevent any single node of the ring from “starving” other nodes, or imposing excessive delays.
3. Small form-factor Packet Ring routers will mean node equipment that can easily fit into existing configured spaces, such as a building basement (Node B), or a vault (Node C and D). In general, a small form factor will also mean a less expensive Packet Ring solution.
4. The presented solution demonstrates a healthy integration of Ethernet and Packet Ring technologies. Ethernet is used here for access and uplink, either in 10/100 (Node B), Gigabit (Node C), or 10 Gigabit (Node A) forms. In addition, for customers reachable only by TDM technologies (Node D), Riverstone’s integration of WAN and Packet Ring technologies in a single platform means access for customers in every environment.

## **STANDARDIZATION**

Efforts are underway to create a Resilient Packet Ring (RPR) standard. The IEEE 802.17 RPR working group is developing a standard that will define a Resilient Packet Ring Access Protocol for use in Local, Metropolitan, and Wide Area Networks. The 802.17 working group has been approved and had its first meeting in January 16-17, 2001. An 802.17 plenary meeting was held March 12-16, 2001 with 166 attendees from 89 organizations. At the end of this meeting, 21 objectives and requirements for the 802.17 RPR standard was accepted; acceptance of technical issues requires a 75 percent vote in IEEE standards groups.

Some of the goals of the 802.17 working group are:

- (1) Support for dual counter rotating ring topology;
- (2) Full compatibility with IEEE's 802 architecture as well as 802.1D, 802.1Q and 802.1f;
- (3) Protection mechanism with sub 50ms fail-over;
- (4) Destination stripping of packets;
- (5) Adoption of existing physical layer medium to avoid technical risk.

The 802.17 working group plans to achieve a standard by March 2003. This preliminary schedule, and the clear benefits of RPR technologies, suggests that pre-standard products will be in widespread use before the standard is finished.

## **CONCLUSION**

Fiber Rings are extremely common in the metro and other networking environment. The time is ripe for a transport technology that both fully exploits the potential of ring networking, and is also easy to integrate with existing Ethernet and SONET technology. RPRs provide a reliable, efficient, and service-aware transport for both enterprise and service-provider networks. Combining the best features of legacy SONET/SDH and Ethernet into one layer, RPR maximizes profitability while delivering carrier-class service. RPR will enable the convergence of voice, video and data services transport.

## **GLOSSARY**

**10 Gigabit Ethernet (10GbE)** — The emerging IEEE standard for Ethernet operation at 10 Gbps.

**Add/Drop Multiplexer (ADM)** — A multiplexer capable of extracting or inserting lower-rate signals from a higher-rate multiplexed signal without completely demultiplexing the signal.

**Asynchronous Transfer Mode (ATM)** — A cell-based, fast packet technology that provides a protocol for transmitting voice and data over high-speed networks. ATM is a connection oriented technology used in both LAN and WAN environments. It is asynchronous in that the recurrence of cells depends on the required or instantaneous bit rate.

**Border Gateway Protocol (BGP)** — Protocol for communications between a router in one autonomous system and routers in another.

**Carrier Sense Multiple Access/Collision Detection (CSMA/CD)** — A channel access mechanism wherein devices wishing to transmit first check the channel for a carrier. If no carrier is sensed for some period of time, devices can transmit. If two devices transmit simultaneously, a collision occurs and is detected by all colliding devices, which subsequently delays their retransmissions for some random length of time. CSMA/CD access is used by Ethernet and IEEE 802.3.

**Data Link Layer**—Layer 2 of the OSI reference model. This layer takes a raw transmission facility and transforms it into a channel that appears, to the network layer, to be free of transmission errors. Its main services are addressing, error detection, and flow control.

**Ethernet** — (1) A baseband LAN specification invented by Xerox Corporation and developed jointly by Xerox, Intel, and Digital Equipment Corporation. Ethernet networks operate at 10 Mbps using CSMA/CD to run over coaxial cable. Ethernet is similar to a series of standards produced by IEEE referred to as IEEE 802.3.

(2) A very common method of networking computers in a local area network (LAN). Ethernet will handle about 10,000,000 bps and can be used with almost any kind of computer.

**Fast Ethernet**— Term given to IEEE 802.3u (called Fast Ethernet) for Ethernet operating at 100 Mbps over Cat-3 or 5 UTP.

**Fiber Distributed Data Interface (FDDI)**—An emerging high speed networking standard. The underlying medium is fiber optics, and the topology is a dual-attached, counter-rotating Token Ring. FDDI networks can often be spotted by the orange fiber “cable.” The FDDI protocol has also been adapted to run over traditional copper wires. An ANSI-defined standard specifying a 100 Mbps token-passing network using fiber-optic cable. Uses a dual-ring architecture to provide redundancy.

**Fiber Optic Cable**—A transmission medium that uses glass or plastic fibers, rather than copper wire, to transport data or voice signals. The signal is imposed on the fiber via pulses (modulation) of light from a laser or a light-emitting diode (LED). Because of its high bandwidth and lack of susceptibility to interference, fiber-optic cable is used in long-haul or noisy applications.

**Fiber Optics**—A method for the transmission of information (sound, pictures, data). Light is modulated and transmitted over high purity, hair-thin fibers of glass. The bandwidth capacity of fiber optic cable is much greater than that of conventional cable or copper wire.

**Gigabit Ethernet**—A 1Gbps standard for Ethernet.

**Gigabits Per Second (Gbp/s)**—Billion bits per second. A measure of transmission speed.

**IEEE 802.1p**—An IEEE draft standard that extends the 802.1D Filtering Services concept to provide both prioritized traffic capabilities and support for dynamic multicast group establishment.

**IEEE 802.2**—IEEE LAN protocol that specifies an implementation of the logical link control sub layer of the link layer. IEEE 802.2 handles errors, framing, flow control, and the Layer 3 service interface.

**IEEE 802.3u**—IEEE LAN protocol that specifies an implementation of the physical layer and MAC sub layer of the link layer. IEEE 802.3 uses CSMA/CD access at a variety of speeds over a variety of physical media. One physical variation of IEEE 802.3 (10Base5) is very similar to Ethernet.

**IEEE 802.5**—The Token Ring protocol. IEEE 802.5 uses token passing access at 4 or 16 Mbps over shielded twisted pair wiring and is very similar to IBM Token Ring.

**IEEE 802.17** — IEEE standard to create a resilient packet ring technology.

**Institute of Electrical and Electronic Engineers (IEEE)**— Professional organization that defines network standards. IEEE LAN standards are the predominant LAN standards today, including protocols similar or virtually equivalent to Ethernet and Token Ring.

**Media Access Control (MAC)**—IEEE specifications for the lower half of the data link layer (layer 2) that defines topology dependent access control protocols for IEEE LAN specifications.

**Media Access Control Sub Layer (MAC Sub layer)** — As defined by the IEEE, the lower portion of the OSI reference model data link layer. The MAC sub layer is concerned with media access issues, such as whether token passing or contention will be used.

**Media Independent Interface (MII)**—The standard in Ethernet devices to transparently interconnect the MAC sublayer and the PHY physical layer, regardless of media.

**Media Interface Connector (MIC)**—FDDI de facto standard connector.

**Megabit (Mb/s)** — One million bits per second.

**Megabits per Second (Mbps)**—A digital transmission speed of millions of bits per second.

**Metropolitan Area Network (MAN)**—A data communication network covering the geographic area of a city (generally, larger than a LAN but smaller than a WAN).

**Multiple Protocol Label Switching (MPLS)**—A set of IETF standards that are designed to allow packet flows to be switched on the basis of labels instead of the full destination addresses, thereby promoting higher performance and allowing traffic engineering.

**Multicast** — General term referring to a message intended for receipt by multiple parties.

**Open Shortest Path First (OSPF)** — A routing protocol used in IP networks.

**Operations Administration Maintenance and Provisioning (OAM&P)**—Tasks performed by the management and administrative systems in a network, especially with reference to public networks.

**Optical Add Drop Multiplexer (OADM)**—An ADM used with fiber optics.

**Optical Cable Level 3 (OC-3)**—Defined standard for the optical equivalent of Synchronous Transport Signal 3 (STS 3) transmission rate or STS 3c Synchronous Optical Network Transport Systems (SONET) transmission rate. The signal rate for these standards is 155.52 Mbps.

**Optical Carrier 1 (OC-1)**—ITU-ISS physical standard for optical fiber used in transmission systems operating at 51.84 Mbps.

**Optical Carrier 3 (OC-3)**—Optical Carrier level 3, SONET rate of 155.52 Mbit/s, matches STS-3.

**Optical Carrier- N (OC-N)**—Higher SONET level, N times 51.84 Mbit/s.

**Physical Coding Sublayer (PCS)** — One of the sublayers defined for the Ethernet protocol stack.

**Physical Layer (PHY)** — The bottom layer of the OSI and ATM protocol stack, which defines the interface between ATM traffic and the physical media. The PHY consists of two sublayers: the transmission convergence (TC) sublayer and the physical medium-dependent (PMD) sublayer.

**Points of Presence (POP)** — A term used by Internet service providers to indicate the number of geographical locations from which they provide access to the Internet.

**Protocol Data Unit (PDU)**—A discrete piece of information like a frame or a packet in the appropriate format for encapsulation and segmentation in the payload of a cell.

**Packet Ring** — Refers generally to technologies which transmit information in packet form but are optimized for a ring topologies. Examples include the emerging RPR / 802.17 standard, and Token Ring technology.

**Resilient Packet Ring (RPR)** — A term that refers to the specific efforts of the IEEE 802.17 working group to generate a resilient packet ring protocol for Wide and Metro Area networks.

**Quality of Service (QoS)** — Term for the set of parameters and their values which determine the performance of a given virtual circuit.

**Shared Ethernet**—Conventional CSMA/CD Ethernet configuration to which all stations are attached by a hub and share 10 or 100 Mbps of bandwidth. Only one session can transmit at a time. This is the most popular network type today

**Spatial Reuse Protocol (SRP)** — A proprietary Cisco packet ring technology.

**Synchronous Digital Hierarchy (SDH)** — ITU-TSS international standard for transmission over optical fiber.

**Synchronous Optical Network (SONET)** — A set of standards for transmitting digital information over optical networks. “Synchronous” indicates that all pieces of the SONET signal can be tied to a single clock. A CCITT standard for synchronous transmission up to multigigabit speeds

**Time Division Multiplexing (TMD)**—A form of transmission in which different flows are combined on the basis of time slots.

**Token Ring** — Refers to the packet ring technology defined in IEEE 802.5 and elsewhere, intended for use in local area networks. On a token ring, transmission on the ring is mediated by nodes sequentially passing a token which allows the possessor to broadcast on the ring.

**Transport Control Protocol/Internet Protocol (TCP/IP)**—A protocol (set of rules) that provides reliable transmission of packet data over networks.

**Wide Area Network (WAN)**—A network which encompasses interconnectivity between devices over a wide geographic area. Such networks would require public rights-of-way and operate over long distances.

**Wave Division Multiplexing (WDM)**—A technology that allows multiple wavelengths to be multiplexed over a single strand of fiber. Comes in various forms including Dense and Wide depending on the number of wavelengths involved.