Resilient Packet Ring Alliance

An Introduction to Resilient Packet Ring Technology

A White Paper by the
Resilient Packet Ring Alliance
October 2001

An Introduction to Resilient Packet Ring Technology

Table of Contents

An Introduction to Resilient Packet Ring Technology

1.0 Introduction

An important trend in networking is the migration of packet-based technologies from Local Area Networks to Metropolitan Area Networks (MANs). 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 is considered by many to be the only alternative for scaling metro networks to meet the demand.

2.0 Ethernet in the Metro

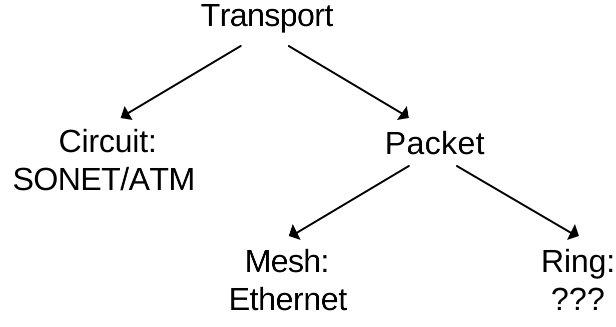
Defined simply, an Ethernet service is any data service offered via an Ethernet interface (10 Mbps, 100 Mbps, 1 Gbps Ethernet port). A key difference between Ethernet services and legacy data services such as leased lines, Frame Relay or ATM is the scalability of the service interface.

With legacy data services, physical interface requirements vary with the speed of the service. Thus hardware required for a T1 service is completely different from that required for DS-3 or OC-3 services.

With Ethernet service, on the other hand, a service provider can drop a Fast Ethernet (100 Mbps capacity) or Gigabit Ethernet (1000 Mbps capacity) port to a subscriber once and upgrade many times, without additional truck rolls beyond the initial installation. Bandwidth and other service changes can be administered remotely, simplifying and accelerating service provisioning.

Ethernet services are widely viewed as an offering that holds promise for rapid acceptance in the marketplace. The question remains as to what infrastructure can cost effectively scale to meet this demand.

Figure 1: Packet Rings: The Next Step in Packet-Based Transport



Ethernet has evolved over the past 25 years from 10 Mbps to 100 Mbps to 1 Gbps and now to 10 Gbps. These and other changes adopted by the IEEE make Optical Gigabit Ethernet technology, capable of supporting fiber spans of more than 50 miles, now emerge as a viable alternative for data transport in public networks. As nearly all data packets begin and end their trip across the Internet as Ethernet frames, carrying data in a consistent packet format from start to finish throughout the entire transport path eliminates the need for additional layers of protocol and synchronization that result in extra costs and complexities. In addition to efficient handling of IP packets, Ethernet has the advantages of familiarity, simplicity, and low cost.

Gigabit Ethernet, however, is only the first step in the evolution of packet-based transport in the MAN. Though well suited for point-to-point and mesh network topologies, it is difficult to deploy Ethernet in ring configurations and as a shared media. Rings act as a shared media and need media access control (MAC) mechanisms to manage access across multiple users. Ethernet has evolved to support full duplex switched infrastructures and lacks this MAC. Yet, most of the existing fiber plant in metro areas is in ring form, because the incumbent transport technology, SONET, is typically deployed over fiber rings.

Ring topologies also enable SONET to implement a fast (sub 50ms) protection mechanism that can restore connectivity using an alternate path around the ring in case of fiber cuts or equipment failure. Unlike SONET, Ethernet does not have a built-in fast protection mechanism. There are, therefore, great benefits in a new technology that can fully exploit fiber rings (in particular, ring resiliency) while retaining all the inherent advantages of a packet-based transport mechanism like Ethernet.

The emerging solution for metro data transport applications is Resilient Packet Ring (RPR) 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. In addition, RPR solves problems such as fairness and congestion control that have not been addressed heretofore by incumbent technologies. Several vendors are already developing and introducing RPR technologies to address this emerging market.

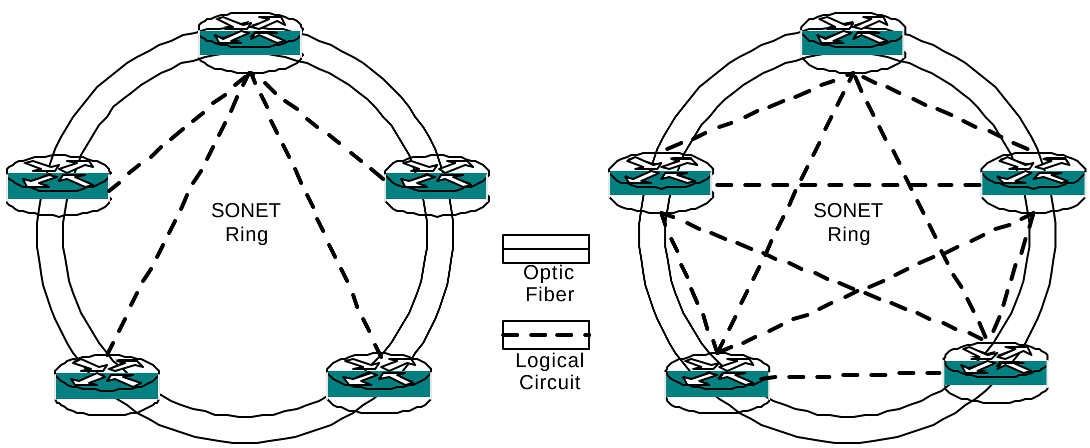
This paper introduces RPR networking, explains its advantages in the metro environment, and gives some examples that illustrate applications that can make the best use of Packet Ring technology.

3.0 The Limitations of SONET and Ethernet in Metro Rings

3.1 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. Below are some of the disadvantages of using SONET Rings for data transport.

Figure 2: SONET Access and Fully Meshed Networks



- *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 1 (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.

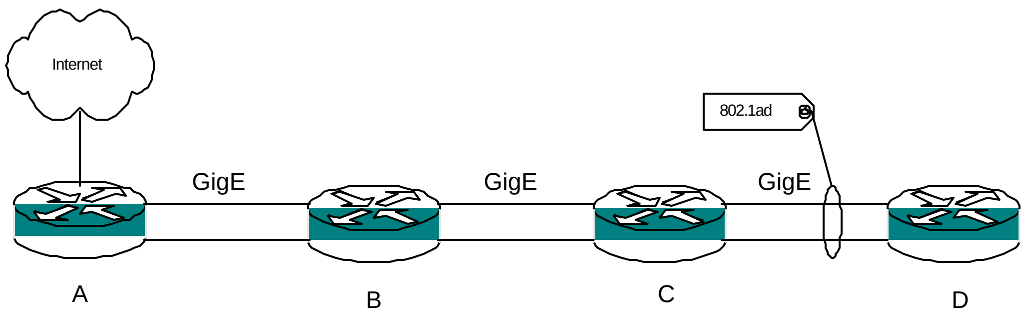
- *Multicast Traffic*. On a SONET ring, multicast traffic requires each source to allocate a separate circuit for each destination. A separate copy of the packet is sent to each destination. The result is multiple copies of multicast packets traveling around the ring, wasting bandwidth.

- *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.

3.2 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.

Figure 3: Ethernet over Ring Topology, Logical Diagram



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 as 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. As described in detail below (see Example 1), a simpler and more efficient method comes from taking advantage of the ring topology to create a global fairness policy.

4.0 Resilient Packet Ring – Emerging Metro Network Architecture

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.

Resilient Packet Ring is an emerging network architecture and technology designed to meet the requirements of a packet-based metropolitan area network. Unlike incumbent architectures based on Ethernet switches or SONET ADMs (Add-drop muxes), RPR approaches the metro bottleneck problem with a clean slate.

In the past few years there have been fiber ring deployments in most metro areas. The challenge for service providers is to tap into the latent capacity available on these fiber rings and carve out as many profitable, revenue-generating services as possible.

This problem of effectively managing a shared resource (in this case the fiber ring is the resource that needs to be shared across thousands of subscribers in a metro area) is most efficiently solved at the MAC layer of the protocol stack. RPR (IEEE 802.17) will be a new MAC protocol designed for metro fiber ring networks.

By creating a MAC protocol for ring networks, RPR attempts to find a fundamental solution to the metro bottleneck problem. Other solutions attempt to make incremental changes to existing products but do not address the basic problem and hence are inefficient. Neither SONET ADMs nor Ethernet switches address the need for a MAC layer designed for the metropolitan area environment. SONET employs Layer 1 techniques (point-to-point connections) to manage capacity on a ring. Ethernet switches rely on Ethernet bridging or IP routing for bandwidth management. Consequently, the network is either underutilized in the case of SONET or non-deterministic in the case of Ethernet switches.

5.0 RPR Characteristics

RPR has several unique characteristics that make it an ideal platform for delivery of data services in metro networks.

5.1 Packet ADM Architecture

It is interesting to compare the packet ADM architecture of RPR devices with the packet switch architecture of Ethernet switches. As shown in Figure 1, a metro network built with Ethernet switches consists of nodes connected by point-to-point links. Network traffic gets queued and scheduled at every intermediate node between the source and the destination. This creates scalability issues. Each node now has to process traffic coming in from the network at line rate. Packet processing technology at each node may be able to handle the lower rates of 1 Gbps or 2.5 Gbps. But when the network scales to 10 Gbps and beyond, this approach breaks down.

5.2 Physical Layer Versatility

The RPR standard now in development create only a new MAC protocol 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.

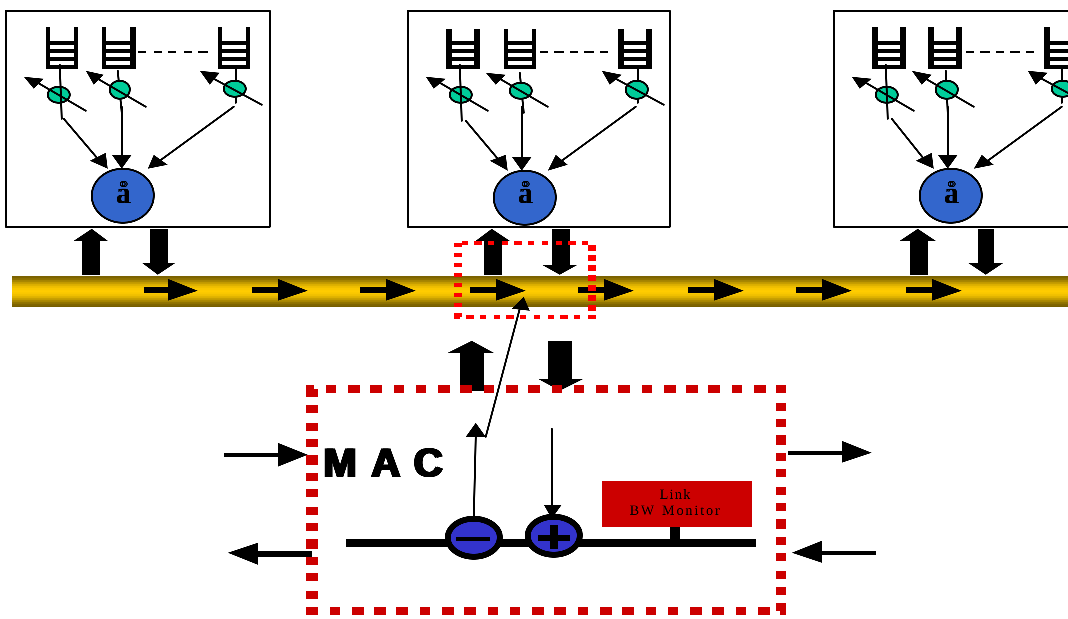


Figure 4: RPR devices act as packet ADMs connected to a shared medium.

RPR devices on the other hand implement the notion of a transit path. At each node, traffic that is not destined for the node simply passes through; it does not get queued and scheduled. As shown in Figure 2, the MAC entity on each node performs three functions: "add" for insertion of subscriber traffic from the node, "drop" or removal of traffic destined for a subscriber on the node, and "pass" or direct transfer of transit traffic from one network link to another. The transit path effectively becomes a part of the transmission medium and makes the RPR ring behave as one continuous medium shared by all the RPR nodes. Because a packet ADM node does not process transit traffic, the packet ADM architecture can scale more easily to higher data rates.

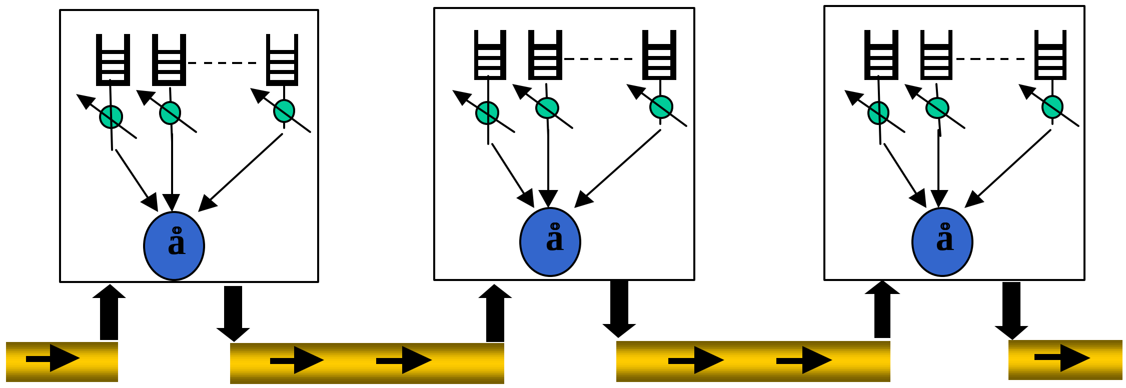


Figure 5: Ethernet switches connected by point-to-point links queue and schedule traffic at each node.

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.

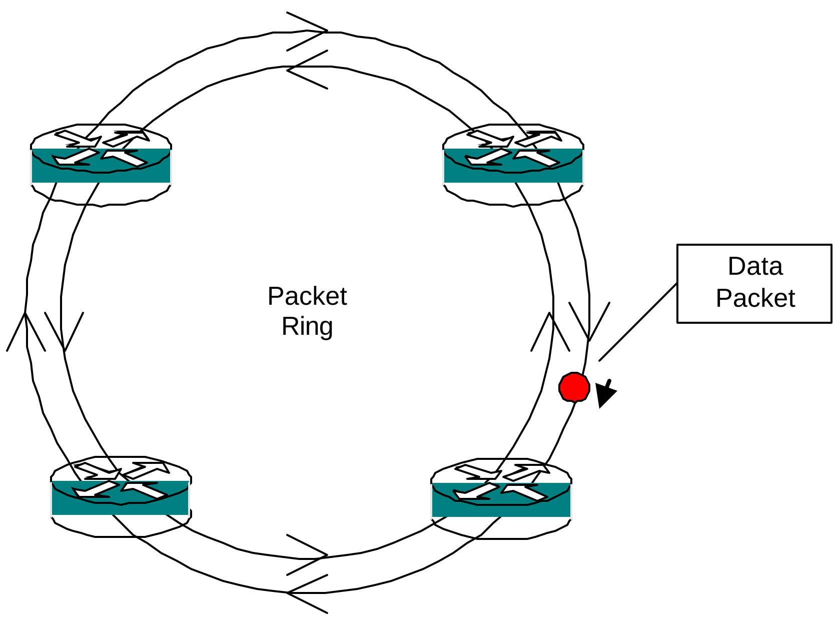
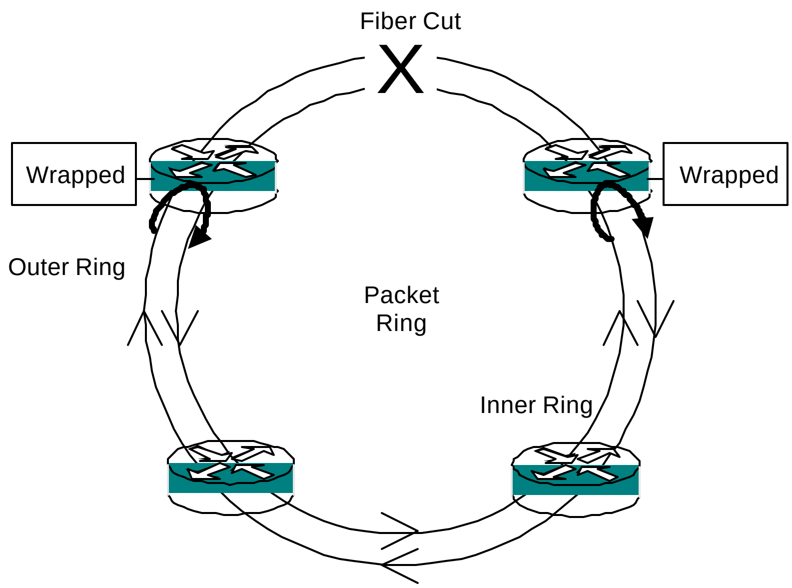


Figure 6: A Bi-Directional Packet Ring

5.3 Resiliency

Packet Rings have a natural resiliency advantage. In the Ethernet case, this can be accomplished using a spanning tree protocol. This restoration mechanism is relatively slow. 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. A Packet Ring protocol can initiate a "ring wrap" at the nodes surrounding the cut (see Figure 2) or packet "steering" by causing the sending node to redirect packets. In either case, traffic can reach the original destination by going around the ring in the opposite direction in the event of a fiber cut.

Figure 7: Recovery from a Fiber Cut



5.4 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 customer on the ring a "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.

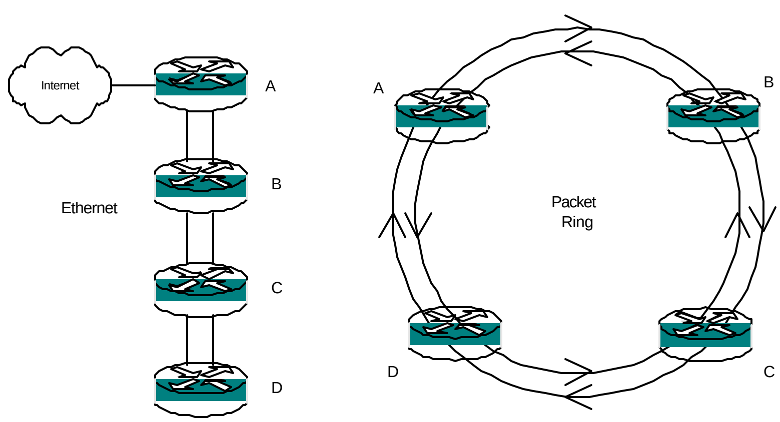
SONET also implements point-to-point circuits that allocate and reserve a fixed amount of bandwidth for each connection but lack of flexibility is the problem. Adding or subtracting bandwidth requires manual configuration of new circuits, and the reservation of such circuits wastes bandwidth.

In a network with dynamically changing traffic patterns (which is typical in any packet network), the only way to optimize network utilization without discarding traffic is to have a feedback mechanism built into the network. The feedback mechanism informs the traffic sources of the capacity available on the network so that the sources can adjust the rate at which they inject traffic into the network.

The MAC entity on each node monitors the utilization on its immediate links and makes that information available to all the nodes on the ring. Each node can then either send in more data or throttle back. This efficient use of bandwidth enables RPR rings to scale beyond 95% of their total capacity.

Ethernet switches or SONET ADMs have no bandwidth management capabilities and hence cannot maximize network utilization.

Figure 8: Global Fairness and the Tragedy of the Commons



In 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 8 am to 12 pm, Node A will be able to send 1 Gbps, Node B will be able to send 500 Mbps and nodes C and D will only be able to send 250 Mbps each. As the number of nodes in the chain grows, the unfairness of Ethernet switches to nodes further upstream becomes even more significant.

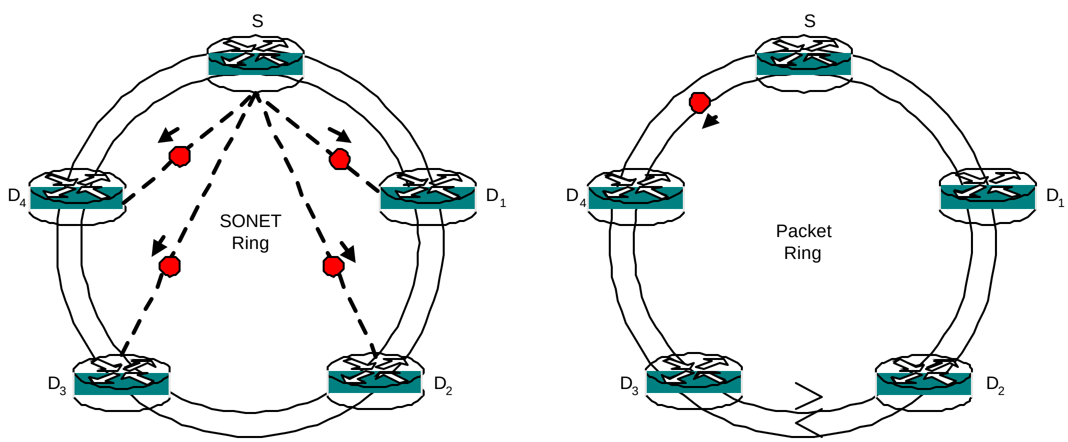
One solution is to implement link-level rate limits on each node. For example, the ingress traffic at each switch might be rate limited to 500 Mbps. But link-level rate limiting policies don't translate into a global fairness policy and create a provisioning and efficiency problem for best effort data traffic similar to SONET circuits. If at 4 am no one else is using any bandwidth, traffic from D need not be limited to 500 Mbps.

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 100 Mbps, 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.

5.5 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.

Figure 9: SONET versus Packet Ring Multicast



In this example, 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.

5.6 Simplified Service Provisioning

A common complaint from data service customers is that it takes too long for carriers to provision services. Activation times on the order of six weeks to six months for DS1 and DS3 services are quite common, with services at OC-3 rates and higher taking even longer.

A significant portion of this delay in service activation can be attributed to the underlying SONET infrastructure and its circuit-based provisioning model. Creating an end-to-end circuit takes many steps. First, the network operator identifies the circuit's physical endpoints to the management system. The operator must then configure each node within the ring for all the required pass-through and add-drop connections. This provisioning operation is time and labor-intensive.

Newer SONET systems automate some of the provisioning steps. But the network operator still needs to perform traffic engineering manually to optimize bandwidth utilization on the ring. The operator must be aware of the network topology, the traffic distribution on the ring, and the available bandwidth on every link traversed by the circuit.

Service provisioning on a network of Ethernet switches is slightly better because they do not require provisioning of circuits through each node. Provisioning, however, still happens a node at a time. In addition, if providers wish to deliver SLAs over the network, the operator still needs to manually traffic engineer the network.

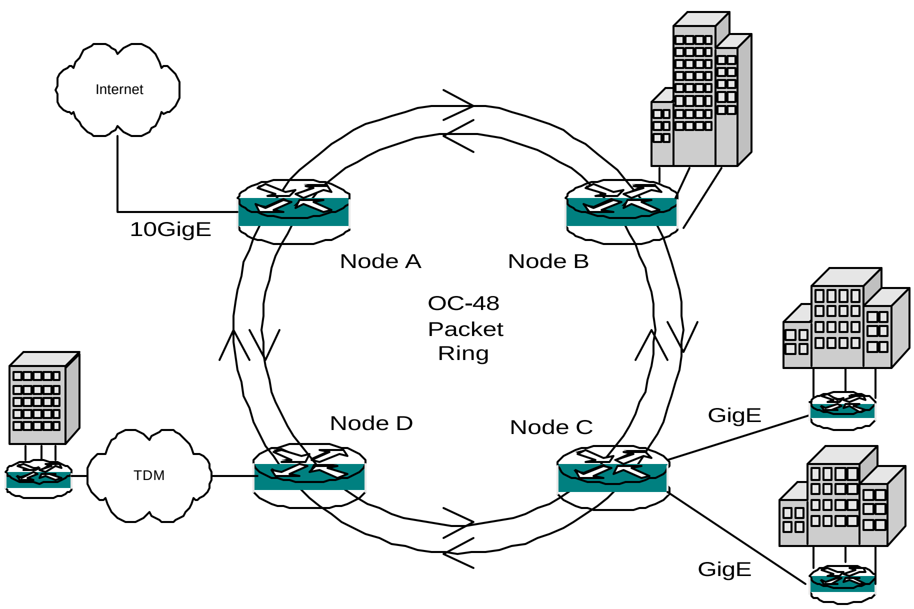
An RPR system, by contrast, offers a very simplified service model. In an RPR system, the ring functions as a shared medium. All the nodes on the ring share bandwidth on the Packet Ring. Each node has visibility into the capacity available on the ring. Provisioning a new service is therefore far simpler. There is no need for a node-by-node and link-by-link capacity planning, engineering, and provisioning exercise. The network operator simply identifies a traffic flow and specifies the quality of service it should get as it traverses the ring.

6.0 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.

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.

Figure 10: A Packet Ring Metro Access and Aggregation Solution



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 Resilient 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

2. The fairness algorithms built into the Resilient Packet Ring will prevent any single node of the ring from "starving" other nodes or imposing excessive delays.

3. Small form-factor Resilient Packet Ring devices will mean node equipment that can easily fit into existing configured spaces, such as a building basement (Node 1), or a vault (Node 3 and 4). 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 2), Gigabit (Node 3), or 10 Gigabit (Node 1) forms.

7.0 Packet Ring Standard Development in IEEE 802.17

Efforts are underway to create an 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 was approved in December 2000. Over 200 individuals from more than 90 companies regularly participate in the working group.

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; and (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 complete. A working draft of the standard is expected by early 2002.

Supplementing and supporting the 802.17 effort is the RPR Alliance, a group of companies that are cooperating to promote the Resilient Packet Ring standard. The mission of the RPRA is to “nurture and help develop a broad market by promoting the proliferation of Resilient Packet Ring into the broadly-defined networking market, including LANs, MANs, and WANs. It is also committed to promoting multi-vendor interoperability.” The RPR Alliance will play an important role in communicating the RPR message to the world, building consensus for the RPR standard, and ensuring interoperability among RPR vendors.

8.0 Conclusion

With all the comparisons with Ethernet switches and SONET ADMs, the reader might be led to believe that RPR intends to compete with two well-entrenched, time-tested technologies. Nothing could be further from the truth.

In reality RPR is complementary to both SONET and Ethernet. Both SONET and Ethernet are excellent Layer 1 technologies. Whereas SONET was designed as a Layer 1 technology, Ethernet has evolved into one. Through its various incarnations, Ethernet has transformed from the CSMA/CD shared media network architecture to a full duplex, point-to-point switched network architecture. Most of the innovation in Ethernet has been focused on its physical (PHY) layer, increasing the speed at which it operates. The MAC layer has been largely left untouched and is practically irrelevant. The portion of the MAC layer that continues to thrive is the MAC frame format.

RPR is a MAC protocol and operates at Layer 2 of the OSI protocol stack. By design, RPR is Layer 1 agnostic, which means that RPR can run over either SONET or Ethernet. RPR enables service providers to build more scalable and efficient metro networks using SONET or Ethernet as physical layers.

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 technologies. The RPR MAC in conjunction with Ethernet or SONET PHYs offers a standards-based approach to building highly efficient metro networks.

Note

Authors: Gunes Aybay, Mannix O'Connor, Kanaiya Vasani, and Tim Wu
The authors are all active participants in the Resilient Packet Ring Alliance. Gunes Aybay and Tim Wu are with Riverstone Networks; Mannix O'Connor and Kanaiya Vasani are with Lantern Communications.