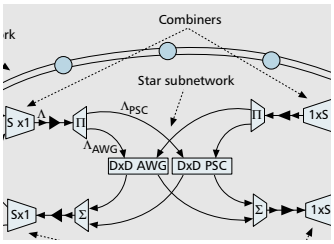

RINGOSTAR: AN EVOLUTIONARY PERFORMANCE-ENHANCING WDM UPGRADE OF IEEE 802.17 RESILIENT PACKET RING

Martin Herzog, Technical University Berlin
Martin Maier, Institut National de la Recherche Scientifique

ABSTRACT

To upgrade IEEE 802.17 Resilient Packet Ring networks by WDM, most previously reported approaches deploy WDM on the ring, requiring modifications of all nodes at the architecture and/or protocol level without significantly enhancing the performance of RPR apart from increased capacity and optical bypassing capability. In this article we aim at rethinking the box. Leveraging on the dark fiber infrastructure abundantly available in today's metropolitan areas, a subset of RPR nodes are interconnected by a single-hop star WDM subnetwork in a pay-as-you-grow manner. Building on RPR's destination stripping, wrapping, and steering techniques, we describe and examine two novel evolutionary performance-enhancing techniques: *proxy stripping* and *protectoration*. Our findings show that the resultant RINGOSTAR network not only dramatically increases spatial reuse, provides fairness as well as fast and efficient resilience against multiple failures but also supports both metro core's uniform and in particular metro edge's strongly hubbed traffic in a highly efficient way.

INTRODUCTION

The new standard, IEEE 802.17, Resilient Packet Ring (RPR), aims at combining synchronous optical network/synchronous digital hierarchy's (SONET/SDH's) carrier-class functionalities of high availability, reliability, and profitable time-division multiplexing (TDM) service support and Ethernet's high bandwidth utilization, low equipment cost, and simplicity [1–3]. RPR is a ring-based architecture consisting of two counterdirectional optical fiber rings with up to 256 nodes. Similar to SONET/SDH, RPR is able to provide fast recovery from a single link or node failure, and carry legacy TDM traffic with a high level of quality of service (QoS). Similar to Ethernet, RPR provides advantages of low equipment cost and simplicity, and exhibits an improved bandwidth utilization due to statistical multiplexing. The bandwidth utilization is further increased by means of spatial reuse. In RPR, packets are removed from the ring by the corresponding destination node. This so-called *destination stripping* enables nodes in different ring segments to transmit simultaneously, resulting in spatial reuse and an increased bandwidth utilization. Furthermore, RPR provides fairness, as opposed to today's Ethernet, and allows the full ring bandwidth to be utilized under normal (failure-free) operation conditions, as opposed to today's SONET/SDH rings where 50 percent of the available bandwidth is reserved for protection. Current RPR networks are single-channel systems (i.e., each fiber carries a single wavelength channel) and are expected to be primarily deployed in metro edge and metro core areas.

Today's metro networks present a significant bandwidth bottleneck between increasingly higher-speed access networks and the huge bandwidth pipes of backbone networks [4]. This bottleneck, often called the *metro gap*, prevents end users from tapping into the vast amount of backbone bandwidth. Next-generation metro networks have to bridge the metro gap in order to tap into the vast amount of backbone bandwidth, enable new emerging services, and stimulate revenue growth. To this end, RPR is likely to be upgraded from a single-channel system to a multichannel system by means of wavelength-division multiplexing (WDM). Clearly, one approach to upgrading RPR by WDM is to use multiple wavelength channels on the fiber rings. To date, a plethora of WDM-upgraded ring network architectures in conjunction with various access and fairness control protocols have been proposed. Previous WDM upgrade approaches of optical ring networks can be categorized into the design of all-optical (OOO) node structures, optical bypassing, traffic grooming, and so-called meshed rings. Due to space constraints we need to refer the interested reader to [5] for a comprehensive survey on and in-depth discussion of WDM rings, including access control, fairness, and QoS support. However, most of these WDM upgrades require modifications of RPR at the node architec-

This work was supported in part by the European Commission within the Network of Excellence e-Photon/ONe and the German research funding agency "Deutsche Forschungsgemeinschaft (DFG)" under the graduate program "Graduiertenkolleg 621 (MAGSI/Berlin)."

ture and/or protocol level, resulting in a revolutionary WDM upgrade. More important, deploying WDM on the fiber rings implies that all network nodes need to be WDM upgraded, be it by wavelength (de)multiplexers or transceiver arrays. Such WDM upgrades that affect the entire network are not well suited to meet today's operators' needs to provide cautious upgrades of existing networks and realize their survival strategy in a highly competitive environment [6].

In this article we report on a novel *evolutionary* WDM upgrade of RPR that builds on its node architecture and protocols. In our WDM upgrade, called RINGOSTAR henceforth, only a *subset* of ring nodes need to be WDM upgraded and interconnected by an arrayed waveguide grating (AWG)-based star WDM network in a pay-as-you-grow manner. By capitalizing on the spatial wavelength reuse capability of the AWG, star WDM networks with modular upgradability, transparency, flexibility, efficiency, reliability, and protection can be realized [7]. In our preliminary investigations we have analyzed RINGOSTAR in terms of mean hop distance, spatial reuse, and capacity and compared it with unidirectional, bidirectional, and meshed WDM ring networks. It was shown in [8] that by WDM upgrading and interconnecting only 64 nodes of a 256-node RINGOSTAR network, the mean hop distance is less than 5 percent of that of bidirectional WDM rings with destination stripping and shortest path routing. In terms of capacity, a 256-node RINGOSTAR network with a single additional (tunable) transceiver at only 64 nodes significantly outperforms unidirectional, bidirectional, and meshed WDM rings in which each of the 256 nodes needs to be WDM upgraded by using an array of 16 (fixed-tuned) transceivers. The contributions of this article are twofold. First, we provide a comprehensive yet comprehensible tutorial overview of RINGOSTAR and its two underlying performance-enhancing techniques, proxy stripping and protectoration. Second, by means of simulations we investigate how recently reported improved RPR fairness control protocols can be extended to RINGOSTAR, which is the major original contribution of this article.

The remainder of the article is organized as follows. In the following subsection we provide a brief overview of RPR and outline its major limitations. We introduce a novel packet stripping technique used in RINGOSTAR. The architecture and access protocol of RINGOSTAR are explained. We examine the hybrid protection-restoration mechanism of RINGOSTAR, and fairness is investigated. We then conclude the article.

Resilient Packet Ring: Overview and Limitations

In this section we briefly highlight the salient features and limitations of RPR. For a more detailed description of RPR, the interested reader is referred to [1–3].

Overview of RPR — RPR is an optical dual-fiber bidirectional ring network where each fiber ring carries a single wavelength channel. Destination stripping in conjunction with shortest path routing is deployed to improve the spatial reuse of bandwidth. Each node is equipped with two fixed-tuned transmitters and two fixed-tuned receivers, one for each fiber ring. Each node has separate (electrical) transit and station queues for either ring. Specifically, for each ring a node has one or two transit queues for in-transit traffic, one transmission queue for locally generated data packets, one reception queue for packets destined for the local node, and one add_MAC queue that stores locally generated control packets. In RPR in-transit ring traffic is given priority over station traffic so that in-transit packets are not lost due to buffer overflow. Thus, the transit path is lossless, and a packet put on the ring is not dropped at downstream nodes. On the downside, however, a backlogged node has to wait for the transit path to be empty before it can send data. As a consequence, upstream nodes can easily starve downstream nodes, giving rise to fairness problems.

To achieve fairness, a distributed fairness control algorithm is deployed in RPR according to the so-called Ring Ingress Aggregated with Spatial Reuse (RIAS) reference model. In RIAS, the level of traffic granularity for fairness determination at a link is defined as an ingress aggregated (IA) flow (i.e., the aggregate of all flows originating from a given ingress node). Moreover, in RIAS bandwidth can be reclaimed by IA flows when it is unused to ensure maximal spatial reuse. The fairness control in RPR is realized by enabling a backlogged node to send fairness control packets based on its local measurements to upstream nodes in order to throttle their ingress data rates and thus alleviate the congestion.

Finally, RPR provides resilience against any *single* link or node failure by means of *wrapping* and *steering* protection mechanisms. Wrapping occurs locally and requires both nodes adjacent to the failure to perform protection switching. Steering is achieved by modifying the routing tables of each node after learning that a failure has occurred.

Limitations of RPR — Due to its underlying ring topology and the applied fairness control algorithm, RPR suffers from the following limitations.

Spatial reuse: In RPR, packets generally have to traverse multiple intermediate nodes in order to reach their destinations, and thus consume a considerable amount of ring bandwidth, resulting in limited spatial reuse.

Oscillations under unbalanced traffic: Spatial reuse in RPR is further decreased due to severe and permanent oscillations under unbalanced and constant rate traffic inputs [2]. Recently, novel fairness algorithms have been proposed that are able to mitigate the oscillations and achieve nearly complete spatial reuse [9. 10]. In particular, the so-called Distributed Virtual-Time Scheduling in Rings (DVSR) fairness control algorithm has attracted considerable attention [11]. We examine an extended version of DVSR later when discussing fairness in RINGOSTAR.

Single-failure resilience: RPR is able to recover only from a single link or node failure in a rather inefficient manner by wrapping and steering incoming traffic away from the failure on the opposite fiber ring. For instance, it was shown in [3] that in the event of a failure, the loss of traffic in a 63-node RPR network may be as high as 94 percent due to the increased length of the backup path. Furthermore, RPR's resilience against a single failure is poorly suited to provide survivability in the presence of multiple failures, which is paramount in metro core networks [12].

Hot-spot traffic pattern: Metro edge rings exhibit strongly hubbed (hotspot) traffic where most traffic originating from a given access network is outbound toward metro core rings [4]. RPR with its underlying ring topology supports such hotspot traffic inefficiently since outbound packets have to traverse many intermediate nodes along the fiber rings on their way to the hub due to missing alternate shorter paths.

In the following sections we describe RINGOSTAR together with its performance-enhancing techniques and explain how they alleviate RPR's shortcomings.

Proxy Stripping

To improve the spatial reuse, resilience, and bandwidth efficiency of RPR, we propose to augment the bidirectional ring with a single-hop star subnetwork, as shown in Fig. 1a for $N = 12$ ring nodes. A subset of $P \leq N$ ring nodes are connected to the single-hop star subnetwork, preferably by bidirectional pairs of *dark fiber*. Note that recently most conventional carriers, a growing number of public utility companies, and new network operators make use of their rights of way, especially in metropolitan areas, to build and offer so-called dark fiber

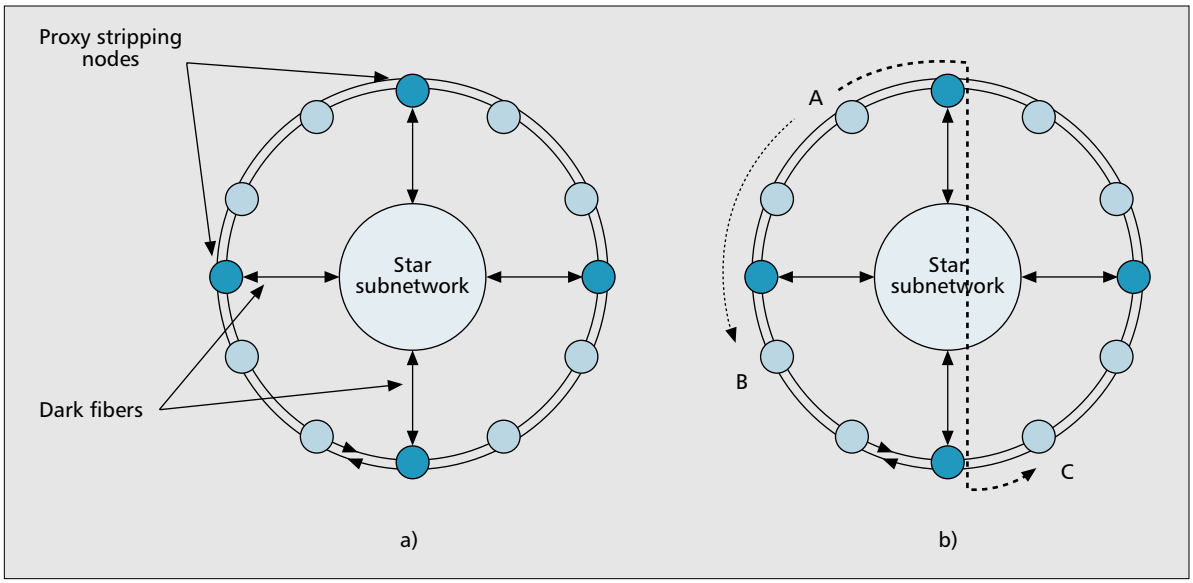

FIGURE 1. Proxy stripping technique: a) RPR with $N = 12$ nodes, where $P = 4$ of them are interconnected by a dark-fiber single-hop star subnetwork, b) proxy stripping in conjunction with destination stripping and shortest path routing for source node A and destination nodes B and C.

networks. These dark fiber providers have installed a fiber infrastructure that exceeds their current needs. The unlit fibers provide a cost-effective way to build very-high-capacity networks or upgrade the capacity of existing (ring) networks. Buying one's own dark fibers is a promising solution to reduce network costs as opposed to leasing bandwidth, which is an ongoing expense. Nodes can be attached to the single-hop star subnetwork one at a time in a pay-as-you-grow manner according to given traffic demands. The hub of the single-hop star network may be a passive star coupler (PSC), an AWG, or a combination of both. For more details on various AWG and PSC-based single-hop star network and node architectures together with medium access control (MAC) protocols, the interested reader is referred to [7]. Nodes attached to the star subnetwork perform *proxy stripping*, a novel packet stripping technique developed in RINGOSTAR.

Proxy stripping is illustrated in Fig. 1b. Recall that in RPR spatial reuse is achieved by means of shortest path routing and destination stripping, as shown in Fig. 1b for source node A and destination node B. Note that only source node A (shortest path routing) and destination node B (destination stripping) are involved, but the intermediate node attached to the star subnetwork performs simple forwarding on the ring. In this case the node attached to the star subnetwork does not pull packets destined for node B from the ring and does not send them across the star subnetwork since the path on the counterclockwise ring is the shortest path between nodes A and B in terms of hops. If, however, the shortcuts of the star subnetwork provide a shorter path than either peripheral fiber ring, intermediate nodes attached to the star subnetwork perform proxy stripping instead of simple forwarding. Proxy stripping makes use of RPR's built-in shortest path routing and destination stripping. As shown in Fig. 1b for source node A and destination node C, node A sends its packets destined for node C to its closest proxy-stripping node (shortest path routing). Now, instead of simply forwarding the packets on the clockwise peripheral ring, the proxy-stripping node pulls the packets from the ring and sends them across the single-hop star subnetwork to the proxy-stripping node closest to destination node C by using the MAC protocol of the given star subnetwork. The receiving proxy-stripping node forwards the packets on the shortest path along the counterclockwise ring toward node C, which finally takes the packet from the ring (destination stripping). Practically, proxy stripping can be done by monitoring an arriving packet's source and destination MAC addresses and making a lookup in each proxy-stripping node's topology database in order to decide whether a given packet has to be proxy stripped or not. The topology database is built and continuously updated using RPR's built-in topology discovery protocol [1].

By means of proxy stripping, the single-hop shortcuts of the star subnetwork are exploited to decrease the mean hop distance and diameter of the network. Thus, packet transmissions require fewer bandwidth resources on the ring, resulting in increased spatial reuse and improved throughput-delay performance. In [13] we have examined the dimensioning of the star subnetwork and the throughput-delay performance of RINGOSTAR by means of probabilistic analysis and simulation. To give the maximum achievable throughput-delay performance of proxy stripping and provide an upper bound that enables the performance comparison of the various proposed fairness control protocols, we did not take fairness control into account. In our investigations we considered both uniform and hotspot unicast traffic and a typical trimodal IP packet size distribution (50 percent 40-byte packets, 30 per-

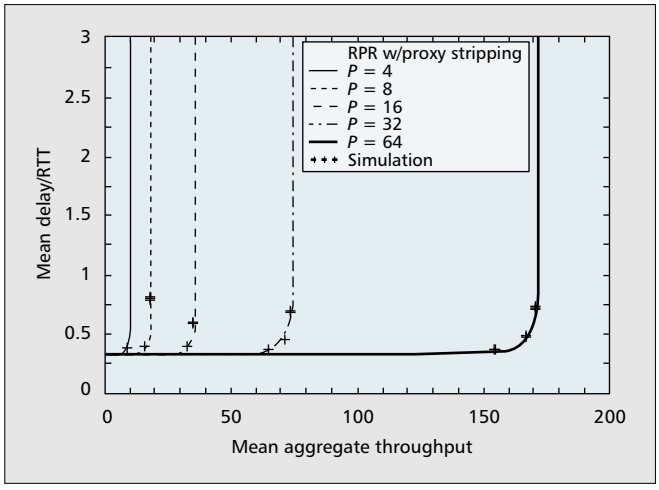

FIGURE 2. Mean delay (given in round-trip time (RTT) of ring) vs. mean aggregate throughput (given in number of simultaneously transmitting nodes in steady state) of RPR with $P \in \{4,8,16,32,64\}$ proxy stripping nodes for uniform traffic with $N = 256$.

cent 552-byte packets, and 20 percent 1500-byte packets). The star subnetwork was dimensioned such that it provides sufficient capacity to shortcut traffic. Figure 2 shows the throughput-delay performance of RINGOSTAR for different numbers of proxy stripping nodes $P \in \{4, 8, 16, 32, 64\}$ and $N = 256$ fixed (the maximum number of nodes in RPR) under uniform traffic, where the mean aggregate throughput denotes the mean number of simultaneously transmitting nodes, and the mean delay denotes the mean time period between packet generation at the source node and packet reception at the destination node, given in round-trip time (RTT) of the ring. Obviously, the throughput is dramatically improved by increasing P. For instance, by interconnecting $32/256 = 12.5$ percent of the nodes via a star subnetwork (i.e., $P = 32$), a maximum mean aggregate throughput of approximately 75 is achieved. Compared to the maximum mean aggregate throughput of 8 achievable in conventional RPR bidirectional rings without proxy stripping, this translates into a throughput improvement by a factor of almost 10. The throughput of RPR can be further improved by increasing P at the expense of more star transceivers and dark fibers. As shown in [13], for hotspot traffic the maximum mean aggregate throughput of RPR (without proxy stripping) drops to 4, which is half that obtained under uniform traffic. By deploying $P = 32$ proxy stripping nodes, the maximum mean aggregate throughput of the 256-node network is increased by a factor of more than 30. Thus, proxy stripping dramatically improves the throughput-delay performance not only under uniform traffic of metro core networks, but also in particular under hotspot traffic of metro edge networks. Finally, note that the star subnetwork in conjunction with proxy stripping can be used to protection-switch traffic around multiple link and/or node failures on the ring, as discussed in greater detail later.

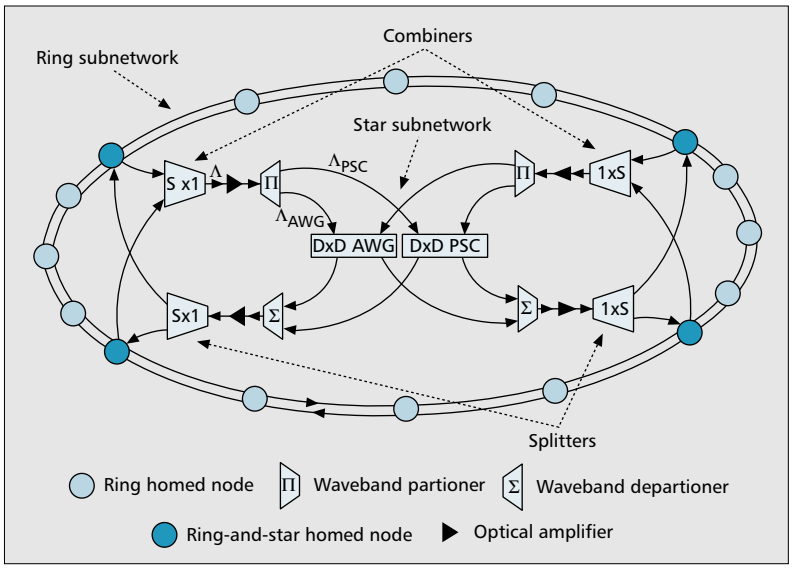

FIGURE 3. RINGOSTAR network architecture with $N = 16$ nodes, where $N_{rs} = D \cdot S = 2 \cdot 2 = 4$ are ring-and-star homed nodes and $N_r = N - N_{rs} = 12$ are ring homed nodes. There are $\Lambda_{PSC} = D \cdot S + 1 = 2 \cdot 2 + 1 = 5$ wavelengths on the PSC, $\Lambda_{AWG} = D \cdot R = 2 \cdot R$ wavelengths on the AWG, for a total of $\Lambda = \Lambda_{PSC} + \Lambda_{AWG} = 5 + 2 \cdot R$ wavelengths in the star subnetwork, where R denotes the number of used FSRs of the AWG.

ARCHITECTURE AND ACCESS PROTOCOL

In this section we describe the architecture and access protocol of RINGOSTAR in greater detail, paying particular attention to the star subnetwork.

ARCHITECTURE

As shown in Fig. 3, RINGOSTAR consists of the RPR bidirectional ring subnetwork and a star subnetwork.

Ring Subnetwork — The RPR ring subnetwork interconnects $N \geq 1$ nodes, which are subdivided into two subgroups of $N_{rs} = D \cdot S$ ring-and-star homed nodes, and $N_r = N - N_{rs}$ ring homed nodes, with $D \geq 1$ and $S \geq 1$. The N_{rs} ring-and-star homed nodes are equally spaced among the N_r ring homed nodes on the ring, as illustrated in Fig. 3 for $N = 16$ and $N_{rs} = D \cdot S = 2 \cdot 2 = 4$ (and $N_r = N - N_{rs} = 12$). Unlike the ring homed nodes, the ring-and-star homed nodes are also attached to the star subnetwork.

Star Subnetwork — The star subnetwork is based on a central hub that consists of a $D \times D$ AWG in parallel with a $D \times D$ PSC, where $D \geq 1$. The star subnetwork uniquely combines the merits of the wavelength-insensitive PSC (broadcast control) and the wavelength-routing AWG (spatial wavelength reuse). Furthermore, the PSC and AWG protect each other if either device fails, preventing the single point of failure of star networks [14]. Each ring-and-star homed node i, $i = 1, \ldots, N_{rs}$, has a home channel λ_i on the PSC (i.e., a unique wavelength channel λ_i on which node i receives data transmitted over the PSC). In addition, there is a control wavelength channel λ_c on the PSC. Consequently, there are $\Lambda_{PSC} = N_{rs} + 1 = D \cdot S + 1$ wavelength channels on the PSC, which make up the PSC waveband. The AWG waveband consists of $\Lambda_{AWG} = D \cdot R$ contiguous data wavelength channels, where $R \geq 1$ denotes the number of used free spectral ranges (FSRs) of the underlying $D \times D$ AWG. A total of $\Lambda = \Lambda_{AWG} + \Lambda_{PSC}$ contiguous wavelength channels are operated in the star subnetwork (as further detailed later).

The signals from S ring-and-star homed nodes on the Λ wavelength channels are transmitted on S distinct fibers to an $S \times 1$ combiner, which combines the signals onto the Λ wavelength channels of one fiber leading to a waveband partitioner. The waveband partitioner partitions the set of Λ wavelengths into the AWG and PSC wavebands, which are fed into an AWG and PSC input port, respectively. The signals from the opposite AWG and PSC output ports are collected by a waveband departitioner and then equally distributed to the S ring-and-star homed nodes by a $1 \times S$ splitter. If necessary, optical amplifiers are used between combiner and partitioner as well as splitter and departitioner to compensate for attenuation and insertion losses of the star subnetwork. A total of D of these arrangements, each consisting of combiner, amplifier, waveband partitioner, waveband departitioner, amplifier, and splitter, are used to connect all $N_{rs} = D \cdot S$ ring-and-star homed nodes to the central hub.

The architecture of ring homed nodes is identical to that of conventional RPR nodes described earlier. For the transmission and reception on the ring subnetwork each ring-and-star homed node deploys the same number and type of transceivers and queues as a ring homed node. In addition, for control transmission on the star subnetwork each ring-and-star homed node is equipped with a transmitter (FT) fixed tuned to the control wavelength channel λ_c of the PSC waveband. The remaining $D \cdot S$ wavelength channels of the PSC waveband and all $\Lambda_{AWG} = D \cdot R$ wavelength channels of the AWG waveband are accessed for data transmission by a tunable transmitter (TT) whose tuning range equals $D \cdot S + \Lambda_{AWG} = D(S + R)$. Similarly, for control reception on the star subnetwork each ring-and-star homed node is equipped with a receiver (FR) fixed tuned to the control wavelength channel λ_c of the PSC waveband. For data reception on the PSC ring-and-star homed node i has a separate fixed-tuned receiver

(FR) operating at its own dedicated *home channel* $\lambda_i \in \{1,2,\ldots, D \cdot S\}$. Each data wavelength channel of the PSC waveband is dedicated to a different ring-and-star homed node for reception. Thus, data packets transmitted on PSC data wavelength channels do not suffer from receiver collisions (a receiver collision occurs when the receiver of the intended destination node is not tuned to the wavelength channel on which the data packet was sent by the corresponding source node). Moreover, on the wavelength channels of the AWG waveband, data packets are received by a tunable receiver (TR) whose tuning range equals $\Lambda_{AWG} = D \cdot R$. Thus, for communication across the star subnetwork each ring-and-star homed node has an FT–FR^2–TT–TR structure (beside the two FTs and two FRs of the ring subnetwork). All transceivers of the star subnetwork of a given ring-and-star homed node are connected to its station queues. Note that the required tuning range of the tunable receiver (Λ_{AWG}) is smaller than that of the tunable transmitter ($D \cdot S + \Lambda_{AWG}$). These requirements take into account the current state of the art in tunable transceivers. While fast TTs with a relatively large tuning range have been shown to be feasible, TRs are less mature in terms of tuning time and/or tuning range.

Access Protocol

Ring homed nodes access the ring subnetwork channels like conventional RPR nodes, as described earlier. Ring-and-star homed nodes access the channels of the ring as well as star subnetworks. On the star subnetwork access to the wavelength channels is arbitrated by means of pretransmission coordination. Specifically, time is divided into frames that are repeated periodically. To prevent collisions of control traffic, in every frame each ring-and-star homed node is assigned a dedicated slot on the control channel λ_c for sending a reservation control packet. Each control packet consists of three fields:

- Destination address of the ring-and-star homed node that is either the destination itself or closest to the destination node
- Length of the corresponding data packet
- Priority of the corresponding data packet

After announcing the data packet in its assigned control slot, the ring-and-star homed node transmits the corresponding data packet on the home channel λ_i of the addressed ring-and-star homed node i in the subsequent L slots by using its TT, where L denotes the length of the data packet in number of slots. After an end-to-end propagation delay of the PSC of the star subnetwork all ring-and-star homed nodes receive the broadcast control packet by using their FRs fixed tuned to λ_c. The corresponding data packet is successfully received at the addressed ring-and-star homed node by using its FR fixed tuned to λ_i, unless one or more other ring-and-star homed nodes have transmitted data packets on λ_i in at least one of the aforementioned L slots. Since all control packets are sent collision-free, they are not retransmitted, and all ring-and-star homed nodes are able not only to detect data packet collisions but also to determine the corresponding source and destination nodes. Based on this information, the retransmission of a given collided data packet is scheduled by all ring-and-star homed nodes in a distributed collision-free fashion by using the appropriate wavelength channel on the AWG instead of PSC. In doing so, retransmissions do not interfere with data traffic on the PSC, resulting in collision-free retransmissions on the AWG and fewer collisions on the PSC, and thus improved throughput-delay performance of the star subnetwork. More precisely, the index j, $1 \le j \le D \cdot S$, of the used control slot and the destination and length fields of the control packet enable each ring-and-star homed node to determine whether the corresponding data packet has collided or not. Given the index j of the control slot, which uniquely identifies the input port of the AWG to which the source node is attached, together with the destination and length fields, the corresponding AWG wavelength channel can be chosen in a distributed fashion by all ring-and-star homed nodes such that data packets are retransmitted on the AWG without collision. If the receiving ring-and-star homed node is not the destination, it forwards the corresponding data packet toward the destination ring homed node on the shortest path similarly to conventional RPR nodes. Likewise, to send locally generated traffic on the ring subnetwork, each ring-and-star homed node deploys the same access rules as a conventional RPR node. To send locally generated traffic across the star subnetwork, each ring-and-star homed node deploys the aforementioned reservation protocol. For a more detailed description of the access protocol of the ring and star subnetworks, we refer the interested reader to [15].

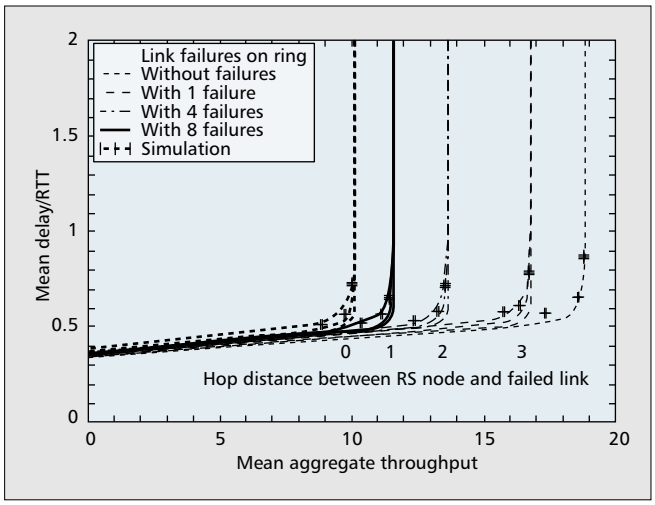

FIGURE 4. Mean delay (given in round-trip time (RTT) of ring) vs. mean aggregate throughput (given in number of simultaneously transmitting nodes in steady state) of RINGOSTAR for link failures with different locations on the ring subnetwork ($N = 64$, $D = 8$, $S = 1$).

Protectoration

In this section we explain how proxy stripping in conjunction with RPR's built-in wrapping and steering can be used to provide RPR with fast and efficient resilience capabilities against multiple failures. The proposed *protectoration* technique aims at combining the fast recovery time of protection (wrapping) and the bandwidth efficiency of restoration (steering together with proxy stripping). Moreover, since protectoration operates at the MAC layer, it does not require convergence of routing protocols at the (IP) network layer in response to failures, and avoids the complex interworking of protection and restoration schemes in layers 2 and 3. As a result, routing stability and network availability are improved.

Upon detection of a ring link or node failure, the two nodes adjacent to the ring failure wrap the incoming traffic away from the failure on the opposite fiber ring. By monitoring the *ring identifier bit*, which defines into which ring the packet was initially inserted, the ring-and-star homed node closest to the failure is able to detect wrapped data packets [1]. This ring-and-star homed node sends wrapped traffic across the star subnetwork to that ring-and-star homed node which is on the other side of the ring failure, using the single-hop short-cuts of the star subnetwork to bypass the ring failure efficiently. If a ring-and-star homed node goes down it is not further available for proxy stripping. In this case the two ring homed nodes adjacent to the failed proxy stripping node detect the failure and inform the remaining nodes by broadcasting topology messages. After learning about the failed proxy stripping node the remaining nodes do not send traffic to the failed ring-and-star homed node. Instead, the neighbor-

ing proxy stripping nodes of the failed proxy stripping node take over its role of proxy stripping. In case of link or device failures of the star subnetwork one or more ring-and-star homed nodes become disconnected from the star subnetwork. After detecting disconnection, the affected ring-and-star homed node informs all remaining nodes by broadcasting a control packet on either ring and acts subsequently as a conventional ring homed node. (For a detailed discussion of fault detection techniques in the star subnetwork, the interested reader is referred to [14].) After learning about the failure, the source node steers its traffic along the shortest path.

Note that the described protectoration technique is highly bandwidth-efficient since wrapped traffic neither makes a round-trip between source and wrapping nodes, nor does it take any long secondary path. More important, the star subnetwork divides the peripheral ring into several segments, each comprising the nodes between two adjacent ring-and-star homed nodes. Note that each segment is able to recover from a single link or node failure without losing full connectivity of the network. Thus, the number of fully recoverable link and/or node failures is identical to the number of ring-and-star homed nodes, provided that there is no more than one failure in each segment. As a result, RINGOSTAR is resilient against multiple link and/or node failures, as opposed to RPR, which would be divided into two or more disjoint subrings in the presence of multiple failures. In [15] we have examined the impact of multiple link failures on the throughput-delay performance of RINGOSTAR with $N = 64$, $D = 8$, and $S = 1$. The link failures are assumed to be 0, 1, 2, or 3 hops away from the next ring-and-star homed (RS) node. We observe from Fig. 4 that the location of link failures has a strong impact on network performance. For a given failure location, however, the protectoration technique is able to accommodate multiple link failures without significant performance loss. Similar results are obtained for multiple node failures.

Fairness Control

In this section we present an extended version of the DVSR fairness control protocol that incorporates proxy stripping.

Operation

Similar to DVSR, to establish RIAS fair transmission rates in RINGOSTAR, packets arriving at the transit queue(s) and station queues are first-in first-out (FIFO) queued at each node. One fairness control packet circulates upstream on each ring. Each fairness control packet consists of $(N + DS/2)$ fields. The first N fields contain the fair rates of all ring links and the remaining $DS/2$ fields contain the fair rates of the star links, where one control packet carries the rates of the even numbered and the other one the rates of the odd numbered star links. Each node monitors both fairness control packets and writes its locally computed fair rates in the corresponding fields of the fairness control packets. To calculate the fair link rates, each node measures the number of bytes l_k arriving from node k, including the station itself, during the time interval T between the previous and the actual arrival of the control packet. Each node performs separate measurements for either ring using two separate time windows. Proxy stripping nodes additionally count the number of bytes arriving from the star for each node and use the time window of the fairness control packet that carries the fair rate of the corresponding proxy stripping node. The fair rate F of a given link is equal to the max–min fair share among all measured link rates l_k/T with respect to the link capacity C currently available for fairness-eligible traffic.

Each node limits the data rate of its $(N - 1)$ ingress flows by using token buckets whose refill rates are set to the current fair rates of the corresponding destinations. Using the same two time windows as in the calculation of the link fair rates above, each node i counts the bytes ρ_{ij} sent to destination j during the time window. Thus, there are two sets of $(N - 1)$ byte counters, one for each time window. Each time a fairness control packet arrives a given node calculates the fair rate of each ingress flow as follows. According to RIAS, the total capacity available to a given node on a certain link equals the fair rate F which is shared among all its ingress flows crossing that link. Based on the measured ingress rates ρ_{ij}/T of these flows and the available capacity F, the max–min fair share f is calculated for each crossed link. The refill rate of each token bucket is set to the minimum fair share f of these links.

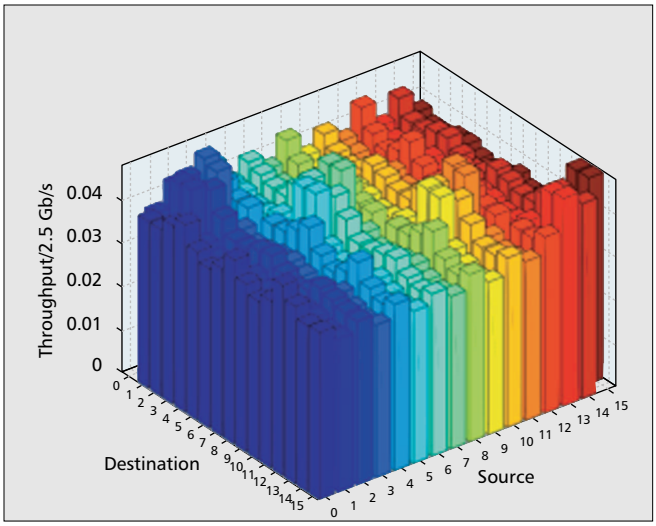

FIGURE 5. RIAS fair throughput (given in 2.5 Gb/s) between each pair of nodes for uniform self-similar traffic ($N = 16$, $D = 4$, $S = 1$).

Simulation Results

In the following we investigate the proposed fairness control protocol by means of simulation. We set $N = 16$, $D = 4$, and $S = 1$. We consider uniform self-similar traffic with Hurst parameter 0.75, where each node does not send any traffic to itself and sends a generated data packet to the remaining $(N - 1)$ nodes with equal probability $1/(N - 1)$. We consider best effort traffic class C, and assume that no bandwidth is reserved for traffic class A, and 10 percent of the ring bandwidth is left for traffic class B (i.e., class C traffic must not use more than 90 percent of the ring bandwidth. Each node is assumed to continuously have data to send on the ring, which operates at 2.5 Gb/s.

Figure 5 shows the RIAS fair throughput for all $16 \cdot 16 = 256$ source-destination node pairs, for each node pair given in 2.5 Gb/s. The throughput varies for different node pairs due to network symmetry. Specifically, there are three types of nodes: proxy stripping nodes (0, 4, 8, 12), nodes between two proxy stripping nodes (2, 6, 10, 12), and neighboring nodes of proxy stripping nodes (remaining nodes). All nodes of a given type achieve identical throughput to all destinations whose distance from the corresponding source node is the same. Proxy stripping nodes achieve higher than average throughput to all other proxy stripping nodes due to the single-hop links of the star subnetwork. Nodes within the same ring segment between two adjacent proxy stripping nodes achieve higher than average throughput if they communicate with each other. Traffic between nodes of different ring segments is bottlenecked by the ring links next to the intermediate proxy stripping node(s), resulting in lower than average throughput. Note that the aggregate throughput of all 256 source-destination nodes pairs is slightly smaller than 20 Gb/s, which is the maximum achievable aggregate throughput. To see this, note that due to shortest path routing and destination stripping, the mean hop distance equals $N/4 = 4$ on either fiber ring, resulting in a spatial reuse factor of 4 on each fiber ring. Given a line rate of 2.5

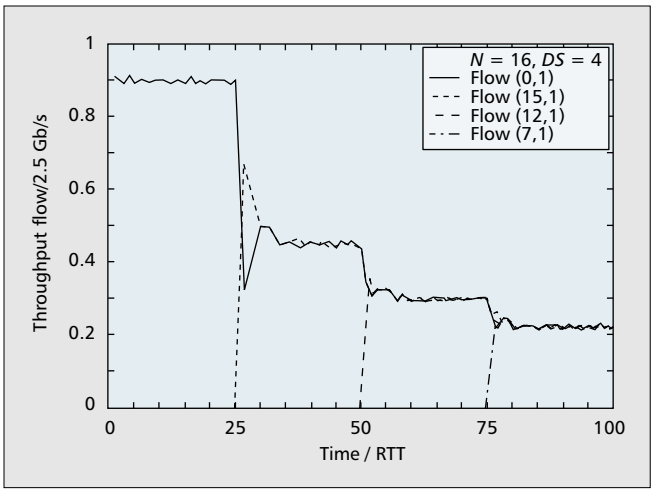

FIGURE 6. Convergence of transmission rates of flows between nodes (0,1), (15,1), (12,1), and (7,1) to their RIAS fair throughput rate (given in 2.5 Gb/s) vs. time given in ring RTT ($N = 16$, $D = 4$, $S = 1$).

Gb/s, this translates into a maximum aggregate throughput of $2 \cdot 4 \cdot 2.5 = 20$ Gb/s in the bidirectional RPR network.

The dynamics of the fairness control are illustrated in Fig. 6, which shows the throughput of four different flows vs. time, which is given in round-trip time (RTT) of the ring. All four flows cross the ring link (0,1) from node 0 to node 1, where node 0 is assumed to be a proxy stripping node. Initially, only flow (0,1) from source node 0 to destination node 1 is active, achieving a normalized throughput of 0.9. Next, flow (15,1) is activated at 25 RTTs. After a convergence time of approximately 10 RTTs both flows equally share the available bandwidth on link (0,1). Note that before the new fair rates are established, flow (15,1) fills up the transit queue of node 0, resulting in a throttled rate of flow (0,1) and a throughput peak of flow (15,1). After 50 RTTs flow (12,1) is activated. Flow (12,1) is first sent from node 12 to 0 via the star subnetwork and then uses link (0,1) to reach node 1. We observe that it takes about 10 RTTs to converge to the new fair rates after flow (12,1) has been activated. Finally, flow (7,1) is activated after 75 RTTs. The flow uses the star subnetwork as a shortcut from node 8 to node 0, similar to flow (12,1). Since the fair rate of link (0,1) is transmitted upstream, it takes longer for node 7 to receive changes of the fair rate of link (0,1) than for node 12. Note that some packets collide at node 0 and have to be retransmitted since now two flows use the star subnetwork as shortcut, resulting in increased delay. However, the convergence time remains approximately 10 RTTs. In summary, the sending rates adapt precisely to the theoretically expected rates in about 10 RTTs and do not suffer from severe oscillations after that.

CONCLUSIONS

To upgrade RPR via WDM, there are basically two complementary approaches, one saving on fiber requirements and the other saving on nodal upgrade requirements. Either WDM is deployed on the fiber ring without requiring an additional fiber infrastructure at the expense of WDM upgrading all ring nodes, or the ring is augmented by an additional (star) WDM network. Apart from increasing network connectivity and thus increasing network resilience, and decreasing both the diameter and mean hop distance of the network, RINGOSTAR provides several other advantages. As opposed to WDM rings, RINGOSTAR improves the spatial reuse and bandwidth efficiency of RPR dramatically. Furthermore, unlike WDM rings, where all ring nodes need to be WDM upgraded at the same time, RINGOSTAR allows for cautious pay-as-you-grow WDM upgrades of RPR by WDM upgrading and interconnecting only a subset of ring nodes according to given traffic demands and/or cost constraints. This incremental evolution of currently single-channel RPR to WDM RPR enables network operators to realize their survival strategy in a highly competitive environment.

The performance gain of RINGOSTAR comes at the expense of an additional star subnetwork. Note, however, that the star subnetwork can be built without costly construction work by using existing dark fibers, which are abundantly available in RPR's primary target metropolitan areas. Also, most parts of the star subnetwork are readily available off-the-shelf components. Furthermore, for a sufficiently small number of proxy-stripping nodes, the star subnetwork can be built completely passive, resulting in reduced maintenance costs.

ACKNOWLEDGMENT

We are grateful to Martin Reisslein for fruitful discussions and insightful comments.

REFERENCES

[1] F. Davik *et al.*, "IEEE 802.17 Resilient Packet Ring Tutorial," *IEEE Commun. Mag.*, vol. 42, no. 3, Mar. 2004, pp. 112–18.
[2] P. Yuan, V. Gambiroza, and E. Knightly, "The IEEE 802.17 Media Access Protocol for High-Speed Metropolitan-Area Resilient Packet Rings," *IEEE Network*, vol. 18, no. 3, May/June 2004, pp. 8–15.
[3] S. Spadaro *et al.*, " Positioning of the RPR Standard in Contemporary Operator Environments," *IEEE Network*, vol. 18, no. 2, Mar./Apr. 2004, pp. 35–40.
[4] N. Ghani, J.-Y. Pan, and X. Cheng, "Metropolitan Optical Networks," *Optical Fiber Telecommun.*, vol. IVB, 2002, pp. 329–403.
[5] M. Herzog, M. Maier, and M. Reisslein, "Metropolitan Area Packet-Switched WDM Networks: A Survey on Ring Systems," *IEEE Commun. Surveys and Tutorials*, vol. 6, no. 2, 2nd qtr. 2004, pp. 2–20.
[6] N. Ghani, " Metropolitan Networks: Trends, Technologies, and Evolutions," *Proc. IEEE ICC*, Apr.–May 2002.
[7] M. Maier and M. Reisslein, "AWG-Based Metro WDM Networking," *IEEE Commun. Mag.*, vol. 42, no. 11, Nov. 2004, pp. S19–S26.
[8] M. Herzog, M. Maier, and A. Wolisz, "RINGOSTAR: An Evolutionary AWG-Based WDM Upgrade of Optical Ring Networks," *IEEE/OSA J. Lightwave Tech.*, vol. 23, no. 4, Apr. 2005, pp. 1637–51.
[9] F. Alharbi and N. Ansari, "A Novel Fairness Algorithm for Resilient Packet Ring Networks with Low Computational and Hardware Complexity," *Proc. IEEE LANMAN*, Apr. 2004, pp. 11–14.
[10] Y. Robichaud *et al.*, "Access Delay Performance of Resilient Packet Ring under Bursty Periodic Class B Traffic Load," *Proc. IEEE ICC*, vol. 2, June 2004, pp. 1217–21.
[11] V. Gambiroza *et al.*, "Design, Analysis, and Implementation of DVSR: A Fair High-Performance Protocol for Packet Rings," *IEEE/ACM Trans. Net.*, vol. 12, no. 1, Feb. 2004, pp. 85–102.
[12] A. A. M. Saleh and J. M. Simmons, " Architectural Principles of Optical Regional and Metropolitan Access Networks," *IEEE/OSA J. Lightwave Tech.*, vol. 17, no. 12, Dec. 1999, pp. 2431–48.
[13] M. Herzog, S. Adams, and M. Maier, "PROXY STRIPPING: A Performance-Enhancing Technique for Optical Metropolitan Area Ring Networks," *OSA J. Optical Net.*, vol. 4, no. 7, July 2005, pp. 400–31.
[14] C. Fan, M. Maier, and M. Reisslein, " The AWG||PSC Network: A Performance-Enhanced Single-Hop WDM Network with Heterogeneous Protection," *IEEE/OSA J. Lightwave Tech.*, vol. 22, no. 5, May 2004, pp. 1242–62.
[15] M. Maier *et al.*, "PROTECTORATION: A Fast and Efficient Multiple-Failure Recovery Technique for Resilient Packet Ring (RPR) Using Dark Fiber," *IEEE/OSA J. Lightwave Tech.*, Special Issue on Optical Networks, vol. 23, no. 10, Oct. 2005, pp. 2816–38.

BIOGRAPHIES

MARTIN HERZOG (herzog@tkn.tu-berlin.de) received a Dipl.-Ing. degree (with distinctions) in computer engineering from the Technical University of Berlin, Germany, in 2002. He is currently a Ph.D. student at the Telecommunication Networks Group at the Technical University of Berlin and participates in a graduate interdisciplinary engineering research program. His research interests lie in the area of optical WDM networks with a focus on architectures and access protocols for metro networks.

MARTIN MAIER (maier@ieee.org) is an associate professor at Institut National de la Recherche Scientifique (INRS), Montréal, Canada. He was educated at the Technical University of Berlin, and received Dipl.-Ing. and Dr.-Ing. degrees (both with distinctions) in 1998 and 2003, respectively. Currently, his research activities focus on evolutionary WDM upgrades of optical access and metro networks. He is the author of the book *Metropolitan Area WDM Networks — An AWG Based Approach* (Springer, 2003).