Optical Metropolitan Networks: Packet Format, MAC Protocols and Quality of Service

Tülin Atmaca[1] and Viet-Hung Nguyen[2]

[1] Telecom SudParis, 9 rue Charles Fourier, 91011 Evry, France
tulin.atmaca@it-sudparis.eu
[2] Itron, 52 rue Camille Desmoulins, 92130 Issy-les-Moulineaux, France
viet-hung.nguyen@itron.com

Abstract. Optical Packet Switching (OPS) is among the most promising solutions for next generation metropolitan networks. The increase of packet-based services (video on demand, etc.) is pushing metropolitan networks providers to renew their infrastructures. Today, metropolitan networks are based on SONET/SDH circuit-switched networks, which are becoming inefficient and costly to support new requirements of quality of service and bandwidth of sporadic packet-based traffic. To solve this problem, many new network solutions are proposed recently, including Next Generation SONET/SDH, Resilient Packet Ring, etc. Among others, the optical networking technology appears a good choice thanks to its following benefits: huge transmission capacity, high reliability, and high availability. This paper is devoted to provide an overview of the metropolitan network infrastructure and particularly to its evolution towards OPS networks. It also highlights performance issues in terms of optical packet format, medium access control protocol and quality of service, as well as traffic engineering issues.

Keywords: Metropolitan Area Network, Optical Packet Switching, Optical Packet Format, Medium Access Control, Quality of Service, Dynamic Intelligent MAC, Circuit Emulation Service, Time Division Multiplexing, Fairness.

1 Introduction

Today's metropolitan area networks (MANs) are faced with a significant challenge; to maintain traditional circuit services (e.g. voice) while, at the same time, enabling new, value-added packet-based services (i.e. video and data) to be carried over the same packet-based network infrastructure. This challenge is the result of the unprecedented proliferation of packet-based services, which in turn has led to a rapid growth in demand in terms of bandwidth and sophisticated quality of service (QoS) requirements in metropolitan areas. MAN service providers must therefore renew their network infrastructures to adapt to these service requirements as well as deliver the bandwidth demanded.

Transformation of metro networks infrastructure can be accomplished by increasing capacity, but more importantly by introducing new technologies with high performance gains relative to infrastructure cost. Several solutions have been proposed to enable the deployment of next-generation of metro optical networks, which

D. Kouvatsos (Ed.): Next Generation Internet, LNCS 5233, pp. 808–834, 2011.

allow a gradual increasing of capacity in a scalable and cost effective manner. Some of them focus on enhancing and adapting existing Synchronous Digital Hierarchy / Synchronous Optical NETwork (SDH/SONET) technologies, while others are designed specifically to serve as a substitute for SDH/SONET. While the standardization in the area of SDH/SONET is not specific to metro networks, the importance of developing new network solutions for metropolitan area is reflected by the large number of recently initiated standardization activities and industry forums as Internet Engineering Task Force (IETF) working group for IP over RPR (IPoRPR), IEEE 802.17 Resilient Packet Ring working group (RPRWG), Metro Ethernet Forum (MEF) and Resilient Packet Ring alliance. That is, an important trend in networking in metropolitan area is the migration of packet-based technology from Local Area Networks (LANs) to MANs. Therefore, in the next generation of metro networks, transport functions will migrate from SDH/SONET architecture to packet-switching transport networks, and will complement service layer features to satisfy the full range of infrastructure and service-specific requirements.

The design of a substitute for actual SDH/SONET architecture is complex because the solutions are diverse and thus the election of the best approach remains a controversial issue. Over the past few years, many studies have been dedicated to the design of architecture for the next generation of metropolitan area networks. Unfortunately, their diverging approaches and viewpoints highlight the complexity of the problem. However, the major standardization works, as supported by Institute of Electrical and Electronics Engineers (IEEE), IETF and MEF have in common the following elements. Firstly, the bandwidth access should be dynamic, flexible and provide a large spectrum of granularity. The metro networks need to handle fine granularity traffic with highly variable characteristics. Moreover, a metro network must directly interoperate with a large range of protocol types (e.g., IP, ATM, Gigabit Ethernet, Frame Relay, etc.) and thus the aggregation is a more significant function than the transport. Secondly, the network should be capable to transport all types of traffic with specific QoS. Here, the TDM support (i.e. backward compatibility) is needed despite the huge growth in data traffic. Finally, the network resiliency will play an important role in electing the best approach. There are efforts to develop solutions capable to offer a sub-50 ms network resiliency as in SDH/SONET.

In this context, the optical networking technology appears a technology of choice for the next generation MAN. The main benefit of optical technology can be summarized in the following terms: *huge transmission capacity*, *high reliability*, and *high availability*. This tutorial is devoted to provide an overview of the MAN infrastructure and particularly to its evolution towards optical packet networks during the last decades. It also highlights performance issues in optical networking in metro area in terms of optical packet format, medium access control (MAC) protocol and quality of service (QoS), as well as traffic engineering issues. We begin with a brief state-of-the-art and perspective on optical metro network. Next, we provide a number of arguments for an answer to the problem of the choice of packet format to be adopted in future metropolitan optical packet switching networks. Here, we provide performance comparison between fixed length packet and variable length packet approaches. Then, we explore the performance issues at MAC layer and present some improvements for MAC protocol. This is followed by the discussion about how to guarantee QoS in multiservice optical packet switching (OPS) metro network, illustrated by some possible mechanisms allowing the transport of circuit-based TDM traffic on packet-based networks. Finally, we finish with some conclusions.

2 Overview and Perspective on Optical Metro Network

2.1 Evolution of Optical Metro Network

In an end-to-end connectivity perspective, a MAN is a network that interconnects many LANs and provides connections with other MANs through the backbone WAN (Fig. 1). A MAN typically provides access and services in a metro region (i.e., a large campus or a city).

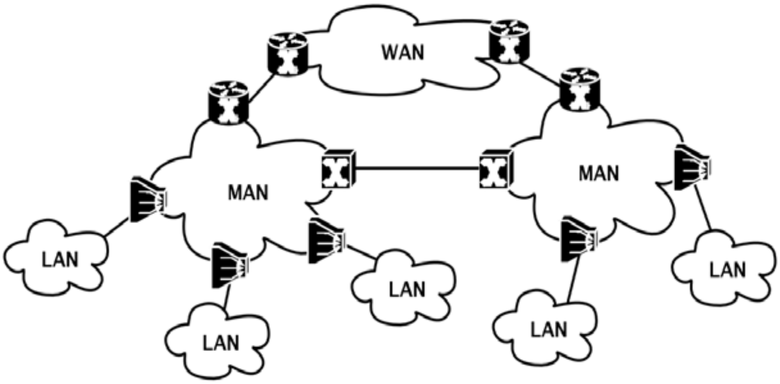


Fig. 1. Global view of metropolitan area network

Designed to be a network that services different access networks and high-bandwidth end-users (e.g., financial organizations, banking, large Internet Service Providers (ISP), etc.), a MAN needs to provide much higher bandwidth capacity and connectivity than a LAN. It should provide reliable and available services as well. The reliability and availability requirements for MAN can be expressed through the very fast restoration of service in the event of failures, today in around 50 ms. Added to that, a MAN should support services with end-to-end security such as virtual private networks (VPN) or virtual LAN (VLAN). With the ever-growing of data and video traffic today, a MAN should also offer high scalability to accommodate rapid change in terms of bandwidth demand in the metro area.

The advent of optical technology is changing the face of the MAN. There can be no doubt that optical networks offer promises to build excellent MANs that meet many of the requirements stated above. The main interest of optical technology is that it provides huge capacity for the network. In addition, optical fibers offer a much more reliable medium of transmission than copper cables. Therefore a MAN based on optical technology will provide a common and solid infrastructure over which varied services could be delivered. The evolution of optical networks is summarized in Fig. 2.

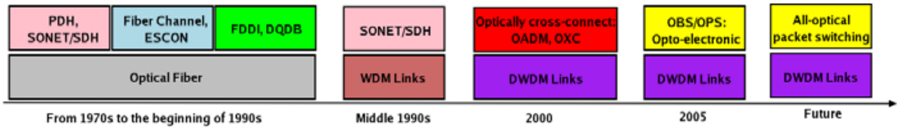


Fig. 2. Evolution of optical networks

In the 1980s, to transmit data at higher bit rate, most carriers (or service providers) deployed unidirectional optical fibers in their backbone infrastructures with point-to-point mesh topology. The processing of switching and multiplexing all remained in electronic domain. During the first half of the same decade, the Plesiochronous Digital Hierarchy (PDH) network was widely deployed. This technology has primarily focused on multiplexing digital voice circuits. The Synchronous Optical Network (SONET) systems were introduced later in North America. A closely related standard, Synchronous Digital Hierarchy (SDH), has been adopted in Europe, Japan and for most submarine links. These systems were developed to overcome some drawbacks of the PDH systems such as lack of multiplexing flexibility, point-to-point limited supervision, etc. Besides, at the end 1980s and the beginning of the 1990s, a number of standards for MAN has been proposed, namely Fiber Distributed Data Interface (FDDI) token ring, an ANSI standard [1], and Distributed Queue Dual Bus (DQDB) IEEE standard [2]. Those networks are packet-based and use ring topology. Other enterprise networks such as Fiber Channel and ESCON were developed to offer storage services to enterprises. They constitute another class of networks, existing in parallel with SONET/SDH systems.

As the demand for wider bandwidth has been increasing continuously, the Wavelength Division Multiplexing (WDM) technology was introduced in optical networks at the middle of the 1990s. WDM increases transmission bit rate by several orders of magnitude. Moreover, in addition to the considerable increase of transmission rate, the optical Add/Drop Multiplexer (OADM), providing wavelength-based add/drop function, and the optical cross-connect (OXC), offering wavelength-based cross-connect function, are the key elements of this generation of optical networks. OADM and OXC were introduced at the very beginning of the year 2000.

Recently, many researchers have studied the feasibility of performing multiplexing and switching optically. In reality, it is advantageous to adopt optical switching and routing means over electronic counterparts, as they are more amenable at higher bit rates and do not require optical-to-electrical (O/E) and electrical-to-optical (E/O) conversions. The feasibility of optical packet switching and optical burst switching networks has been studied (e.g., [3], [4], [5], and [6]).Thus the latest generation of optical networks that could appear at the end of this decade should be all-optical packet switching networks.

2.2 Challenges in Optical Metro Network

The strong increase of demand in terms of bandwidth and value-added services in both access and metropolitan networks today is pushing the service providers to innovate their existing network infrastructures. Although WDM provides huge transmission capacity, there is always a need of providing more sophisticated services to customers such as interactive games and video-on-demand and telephony over IP, which generate much more revenues than raw-bandwidth service offered by WDM. As new applications are constantly being generated, notably in the metro area in which several organizations and enterprises are directly involved, a new generation of MAN is strongly required to replace the obsolete and inappropriate MAN.

The main requirements which are challenging the metro network limits are: a sporadic incoming traffic from LANs with different QoS constraints, different client formats, cost sensitivity and finally the necessity for a scalable architecture to support increasing traffic. The rapid increasing volume of data traffic in metro network is challenging the flexibility of existing transport infrastructure based on circuit-oriented technologies like SDH/SONET. In spite of its high reliability, high QoS and standardization, SDH/SONET lacks the efficiency of transporting data traffic. On one hand, the solutions developed to adapt packet-based traffic to circuit-switching environment (i.e. Packet over SDH/SONET (PoS) and IP over ATM (IPoATM)) use inefficiently the transport infrastructure bandwidth. On the other hand, although ATM over SDH/SONET (ATMoS) offers a predictable end-to-end transport service, it leads to complex interfaces (i.e. SAR algorithms), 10 percent of capacity consumed with transport overhead, and some bandwidth loss when ATM cells are placed in fixed size SDH/SONET frames. Furthermore, the complex algorithms used to transport packet-based traffic over such architecture become difficult to implement at high-speed processing.

Over the last decade, the introduction of optical technology in metropolitan area was studied intensively. The fibre technology is considered the response to the problems of bandwidth requirements and QoS because of its huge capacity. Optical networking technology, through WDM, may represent the solution for meeting increasing bandwidth demands. Unfortunately, the introduction of all-optical WDM equipments and architectures in the metro environment for replacement of the traditional architectures, as initially envisioned, has been slowed down significantly by the current economic situation. As a result, the cost of components and the design of the network as a whole on the other hand has become extremely important and cost versus performance trade-offs now play the central role in the design and engineering of metro networks. The metropolitan networking environment with its constant and rapidly changing bandwidth requirements, cost sensitivity and different customer traffic need present a challenge for the deployment of the latest optical technologies.

2.2.1 Traffic Evolution

The traffic is changing in communication networks, notably in the metro regions. The evolution of the global traffic is led by many elements. First of all, the voice traffic obviously plays an important role in the global traffic today. Some recent observations have pointed out that although the volume of mobile voice and voice over IP continuously increases, the total volume of voice traffic tends to decrease due to the strong decrease of fixed voice traffic.

Beside this traditional traffic, the volume of multimedia video traffic becomes more and more important. The emergence of new multimedia applications such as high quality video conferencing, high-definition television on demand, etc, leads to a considerable growth of video traffic volume (e.g., video traffic is about 60% of the total traffic volume). As a result, there is a need of upgrading the access networks to very high-speed in order to offer enough bandwidth for multimedia end-users. For instance, in Asia, notably in Japan, one plans to provide up to 1 Gbs per user using FTTH (Fiber-To-The-Home) technology, in which 80% bandwidth will be reserved for high quality multimedia applications. The MAN, which connects those access networks, must be upgraded in terms of capacity and quality of service consequently.

Another element that considerably contributes to the evolution of traffic today is the evolution of Pear-to-Pear (P2P) traffic. Recent measures shown that P2P volume continues to grow rapidly (e.g., around 20% of global internet traffic in the year 2004 [7]). This growth may break the asymmetric bandwidth assumption, and also shifts traffic exchange to domestic networks due to P2P nature. This means that most of traffic exchanges in the future may be local exchange (i.e., the major traffic will circulate inside the same MAN), and may does not reach backbone network.

Last, but not least, we cannot ignore the growth of Internet traffic over the world, which mainly consists of data traffic such as WEB browsing, file transfer, hosting, etc. Globally, Internet traffic volume doubles each year. But the Internet growth rate is different in each continent. For example, in the year 2004, the Internet growth rate is about 400% in Asia, 200% in America but only 80% in Europe [8]. The proliferation of this traffic, in addition to the P2P traffic, causes an unprecedented growth of data traffic in the global traffic.

All this goes to prove that the traffic pattern is excessively changing, the voice traffic volume becomes minor compared to multimedia and data traffic volume. The MAN is facing with a real challenge: mainly transporting sporadic video and data traffic, while still being able to offer high level of quality of service for the voice traffic as it represents most revenues of service providers.

2.2.2 Limits of Traditional Circuit Switching Networks

Traditionally, metro networks are based on SONET/SDH rings employing circuit switching technology. Professional users are usually connected directly to SONET/SDH rings through their ADMs, while residential users are connected to those same rings indirectly via the central offices and the access networks. It is a well-known fact that SONET/SDH rings combined with TDM circuit switching and WDM wavelength switching (i.e., SONET over WDM) technologies can provide a huge network capacity with reliable services for metropolitan users. All the same, owing to the ever-growing of video and data traffic that has surpassed voice traffic today, SONET/SDH is facing with great challenges because of its disadvantages in supporting sporadic traffic.

The most important limitation of SONET/SDH is the lack of flexibility and scalability to deal with new demands of data service today. Essentially designed for transporting voice traffic, SONET/SDH networks are unsuitable for the transport of data traffic. Thanks to the constant rate and predictable behaviour of voice traffic, the establishment and provision of circuits were easy. Unfortunately, the volume of data traffic today exceeds that of voice traffic. This tendency is likely to continue for at least the next several years. In contrast to voice traffic, the data traffic is usually sporadic, changing and unpredictable. As a consequence, SONET/SDH must be able to provision and upgrade circuits *efficiently* and *rapidly* to support data traffic. However, this capability did not exist in the traditional design of SONET/SDH.

More specifically, SONET/SDH connections are usually established and dimensioned for a long term contract (months or years). This means that the traffic transported by these connections must be predictable and supposed to be unchanged. Such conditions are not applicable for data traffic due to its sporadic nature. Besides, these connections in most cases use dedicated circuits or wavelengths, with the granularity of PDH hierarchy (about several tens of Mbs) or wavelength bit rate (about several Gbs) in case of wavelength switching. These granularities are too coarse compared to the large variety of data rates that may be required by end-users. Furthermore, even if those circuits are reserved for voice traffic only, they will not be used during periods of inactivity of related users, whereas new users are asking for bandwidth. The implication is that the bandwidth reserved for each circuit/wavelength is often wasted or is used inefficiently.

New techniques are currently being developed to address many of these limitations. Generic Framing Procedure (GFP) [9] has been developed as a new framing for data accommodation into SDH/SONET and optical transport network (OTN). Virtual Concatenation (VC) [10] has been standardized for flexible bandwidth assignment of SDH/SONET paths. Link Capacity Adjustment Scheme (LCAS) [10] has been discussed for dynamic bandwidth allocation in support of virtual concatenation. One of the most important objectives of these new technologies is to enable flexible and reliable data transport over SDH/SONET, which is referred to as data over SDH/SONET (DoS).

The solution listed above may be sufficient to achieve a high utilization in backbone networks where the traffic flows are aggregates of many individual flows and are relatively smooth. In metro networks, however, the traffic is bursty/variable and it is desired to efficiently share the available capacity between the nodes at the time scale of individual packets (packet switching) or bursts of packets (burst switching). While the standardization efforts in the area of SONET/SDH are not specific to metro networks, the importance of metro gap is reflected by the large number of recently initiated standardization activities and industry forums such as IETF working group for IP over RPR (IPoRPR), the IEEE 802.17 resilient packet ring working group (RPRWG), the Metro Ethernet Forum (MEF), and the resilient packet ring alliance.

Last, but not least, SONET/SDH needs a very long time to provision a new circuit. As a matter of fact, it takes weeks to months to fulfill a new bandwidth request, and requires long-term contractual agreements as well. This way of deploying services now becomes obsolete and inefficient. The data traffic today is continuously increasing and varying, constantly generating new service requests. Consequently, the modern MAN should be able to deploy new services rapidly without long-term contracts. Of course, this entails that the modern MAN should provide finer granularity in terms of switching and upgrading relative to SONET/SDH. In this context, sub-wavelength switching technologies such as optical burst switching and optical packet switching, which provide fine switching granularity at a burst and packet levels, are widely studied by many researchers and service providers recently, in order to improve, not to say replace, their existing infrastructures.

2.3 Towards Optical Packet Switching Networks

The need of fine switching granularity for next generation MAN leads researchers to investigate the optical packet switching (OPS) technology that provides the finest granularity. Many recent projects regrouping both academic and industrial researchers have focused on the analysis of OPS, such as KEOPS in the year 1998 [3], WASPNET in the year 1999 [11], DAVID in the year 2003 [4] and ROM-EO [6] in the year 2005.

In OPS networks, packets are switched optically without being converted to electrical signal. The main advantage of this method is to bypass the electronic switching bottleneck and provide enabled switching solution that matches with WDM transmission capability. Fig. 3 shows the generic architecture of an OPS switch. Readers interested in the detail of OPS switch are warmly invited to refer to [3], [11], [12], [13], [4] and [5].

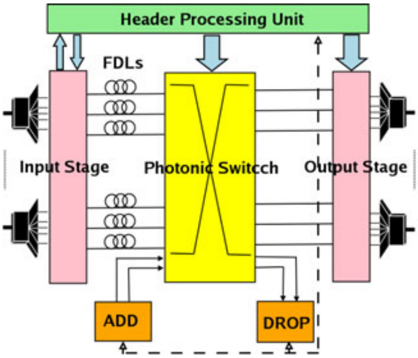


Fig. 3. A generic all-optical switch in OPS network

The main benefit of OPS relative to circuit switching approaches is that OPS results in a better network utilization since it can easily achieve a high degree of statistical multiplexing and effectively be amenable for traffic engineering. Moreover it enables packet-switching capability at high rates that cannot be contemplated using electronic devices. Nevertheless, statistical multiplexing property, which makes OPS attractive over circuit switching, introduces some important effects. Among others, the need of buffering optical packets in case of contention is typical. Contention occurs whenever two or more optical packets are trying to leave a switch from the same output port on the same wavelength. This cannot happen to circuit switching networks as an output port is always reserved for a specific circuit. How the contention is resolved has significant impact on the overall network performance.

Contention is easy to resolve in the electronic world, where electronic random access memory (RAM) is available. OPS, however, is still lacking mature technologies addressing contention issues. Owing to the lack of optical RAM, OPS typically uses fiber delay line (FDL) to delay the packets in contention and send them out at a later time. As fiber delay lines are strict fist-come-first-serve (FIFO) queues with fixed delays, they are far less efficient than electronic RAM. Switched fiber delay lines are recently developed to more efficiently resolve contention, but they require large number of fiber delay lines. Handling a large number of fibers in a switch appears to be a difficult task today. Therefore, optical RAM still remains in an early stage until now.

Other approaches are also used to address contention issue in OPS, such as deflection routing and spectral contention resolution using wavelength conversion. Deflection routing means that a packet in contention is send to a free output port other than the desired port. The disadvantage of deflection routing is that it introduces extra propagation delays and causes packets to arrive out of order. Spectral contention resolution consists in the use of wavelength converters to change the wavelength of the packets in contention. By consequent, it multiples packets can be sent simultaneously to the same output port as multiple wavelengths are generally available at each output port.

It is worthwhile to note that there currently exists another switching technology that is also gaining attention of service providers: the optical burst switching (OBS). OBS is a solution that lies between optical circuit switching and optical packet switching. It offers the switching granularity between a circuit and a packet. A burst is composed of several packets having the same destination. It provides high node throughput, hence high network utilization, and at the same time, it can be implemented with currently available technologies. It promises to be a short-term practical choice as an alternative to traditional circuit switching, before OPS becomes mature.

3 OPS Ring Network for Future MAN

Through the previous subsections of this tutorial, a clear trend one can recognize is that service providers are investigating much effort to construct common packet-based network infrastructures for the future MAN. The key elements of these infrastructures, in the long term, will be undoubtedly all-optical technologies, namely OPS and DWDM. This section is devoted to specifically introduce a good candidate for the future MAN: the optical packet switching networks using ring topology (OPSR). We begin with a discussion about the reason why the ring topology is widely adopted in MAN. Then we describe an example of OPSR network that has been studied by a number of industrials recently.

3.1 Motivation for Ring Topology

Packet switching networks using ring topology are the most common architectures in metropolitan areas. For instance, at the end of 1980s and the beginning of 1990s, there were a number of standards for MAN using ring topology such as FDDI token ring and DQDB IEEE standard. More recently, a ring architecture known as Hybrid Opto-electronic Ring Network (HORNET) has been proposed in the year 2002 [14] and, in the year 2003, a new standard for optical MAN has been approved by IEEE: the Resilient Packet Ring (RPR) [15].

One may wander why ring topology is widely used in MAN. There are actually several factors that have made ring the topology of choice. First of all, bidirectional rings inherently provide fast restoration time after events of outage such as link cut or node failure. For example, in case of a single link cut, packets can be simply redirected to transmit on the other direction of the ring to reach destinations. The highest record of restoration time in today's MAN is around 50 ms, which is typically guaranteed by SONET/SDH rings. RPR also provides this resilience feature. The fast restoration time plays an important role in providing good quality of services (QoS) to customers. It actually contributes to the degree of network availability, a typical requirement that MAN's providers must take into account when designing their networks.

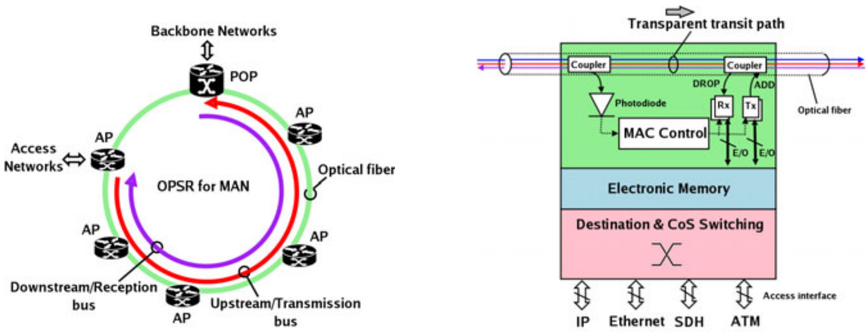
There is a second argument that cannot be ignored, namely the *statistical multiplexing* of data traffic flowing from different nodes over the shared medium of transmission (also known as *collecting function* of a ring). As a MAN typically interconnects a variety of enterprises and organizations of small, medium or big size, its overall traffic is rather sporadic and heterogeneous. If the MAN uses point-to-point or meshed topology to connect their nodes, there will be cases where some links are underused, while the others are overloaded. Thus it would be better if the underused links could support part of traffic from the overloaded links. The ring topology with shared medium of transmission can effectively deal with this problem without complicated routing algorithms. As a matter of fact, since the total bandwidth of the network is shared by ring nodes, the effect of statistical multiplexing will clearly improve the resource utilization. In addition to this feature, a similar utility that a ring could provide is spatial reuse. The idea is roughly described as follows: bandwidth occupied by traffic of upstream nodes will be released at destination nodes, hence will be reused by other nodes after those destination nodes.

As far as the cost of network infrastructure is concerned, ring topology in the metro areas provides very efficient utilization of optical fibers, hence effectively reduces the infrastructure cost by using small number of optical fibers. In point-to-point or mesh topologies, the number of optical fibers is usually proportional to the number of links between nodes, whereas only one fiber may be sufficient to connect all nodes in a ring topology.

Finally, it is worthwhile to note that ring topology simplifies the routing functionality of ring nodes. Since there are either one (for unidirectional rings) or two (for bidirectional rings) paths for routing traffic, the complexity of routing function is obviously reduced relative to mesh topologies.

3.2 Example of an OPS Ring

We now introduce the architecture of an OPSR that uses opto-electronic components. The concept of this architecture mainly comes from on a recent experimental architecture, the Dual Bus Optical Ring Network (DBORN) [16]. The idea is to design an optical network that satisfies requirements for the next generation MAN, while



a) *Global OPSR architecture* b) *OPS node*

Fig. 4. OPSR architecture

tremendously reducing the network cost by employing advanced technologies in optical networking.

As shown in Fig. 4, the OPSR logically consists of two unidirectional buses: a transmission (upstream) bus that provides a shared transmission medium for carrying traffic from several access point (AP) nodes to a point of presence (POP) node; and a reception (downstream) bus carrying traffic from the POP node to all AP nodes. Thus, an AP node always "writes" to the POP node employing the transmission bus and "listens" for the POP node using the reception bus. The traffic emitted on the transmission bus by an AP node is first received by the POP node, then is either switched to the reception bus to reach its destination node, or is routed to other MAN or backbone networks. The easiest way to implement these separated transmission/reception buses is to use two optical fibers: one for the transmission bus, the other for the reception bus. However, this approach will require two more fibers for the purpose of protection in case of fiber cut. Another way is to employ different wavelength bands. In this case, only two optical fibers are needed, one main fiber for normal usage and another for protection.

In this architecture, each AP node consists of two parts (Fig. 4). The electronic part has interfaces with client access networks such as IP, ATM or Ethernet. It is equipped with electrical memory where electronic packets are buffered before being converted to optical signals. The optical part mainly consists of Packet-OADM performing packet add/drop function optically. On the transmission bus, transit optical packets flowing from upstream AP nodes pass through downstream AP nodes optically (i.e., without O/E/O conversion), thanks to the use of optical couplers (Fig. 4.b). Hence transit packets at an AP node are not received or processed by the node. This is the well-known concept of passive optical network (PON), which highly reduces the number of transmitters and receivers at AP nodes. An AP node uses burst mode transmitters (Tx), which works in an asynchronous mode, to insert its optical packet on the transmission bus. Similarly, it employs burst mode receivers (Rx) to drop packets on the reception bus. Therefore, AP nodes in this architecture only need equipment (i.e., optical couplers and burst mode transponders - BMT) to process the packets to and from its local customers. On the contrary, conventional architectures, notably the ones using O/E/O conversion, such as RPR, require significantly more equipment in AP nodes because each node must receive, process and retransmit all optical packets that pass through. In addition to the advantage of offering low number of equipments, this architecture offers an important feature: Tx and Rx in an AP node are out of the transit line, making the upgrade of the node easier, and lessening the service interruption's probability.

4 Optical Packet Format: Fixed versus Variable

In optical packet switching networks, the client layer information is adapted from its original format to optical payloads and then transported transparently through the optical area. At the outgoing border, the client information are extracted from optical payloads, reconverted to their initial format, and forwarded to per-hop or final destination. Therefore, the influence of packet format is to be discussed at two levels: at the border edge and in the core of the optical networks.

4.1 Packet Format in the Core of Optical Network

The impact of packet format in the core of optical networks was intensively studied in the last few years. Numerous works have analyzed the interaction between the packet format and the performance of optical switch fabrics situated at the heart of optical networks. When the access protocol is based on variable length optical packets, the optical payload corresponds to an Ethernet Layer 2 frame. When using fixed format, the optical payload corresponds to one or several Ethernet Layer 2 frames. According to the optical packet creation policies, several upper layer services may be multiplexed onto single optical transport unit. This implies additional electronic information for packet delineation at outgoing edge. The Ethernet PDU trailers can be used for packet delineation. We note that in case of fixed format optical packet, the optical payload should be large enough to accommodate the maximum PDU (e.g., Ethernet MTU).

Here, the format of optical packet raises two questions: how to manage "optical" memories with the presence of fixed and variable packet formats and the hardware limits of optical technology in handling different formats of optical packet.

Regarding the resources contention, the packet format influences the design of Contention Resolution Mechanisms (CRMs) and sizes optical memories. In case of fixed format, the CRMs latency is limited to exploiting the WDM dimension and to exploiting multi-terabit switching planes. The variable format adds a supplementary temporal dimension to contention resolution algorithms. When all output resources are occupied, the optical payloads are buffered by mean of optical memory organized as feed-back or recirculation lines (i.e., FDL). When the optical packets have a fixed format, the FDL is equivalent to a normal queue of which traffic issues are well managed. Variable format approach leads to more complex issue [17], where the packet loss probability has shown to be a convex function of the delay granularity of FDL. Therefore, the optimum value of elementary delay line depends on distribution of optical packet sizes.

The optical processing requires a strict delineation of optical packets. When the network is slotted there is a necessity for synchronization in core nodes. The precision of synchronization is very important and can be achieved only in order of nanosecond. The utilization of variable format brings no constraints for synchronization during the optical processing. In case of switching optical payloads, the packet format with fixed size seems to be more efficient since a synchronous approach offers the possibility to adopt large switching plane organized in "pages" of constant size (at least it is the case in fast-switching ATM fabrics).

4.2 Packet Format at Edge of Optical Network

In case of optical payloads with fixed format, the complexity is placed at the edge. Optical payloads are filled with the client packets usually structured as packets with variable format. If the length of the optical packet payload is too short, the client layer packet is segmented and sent in several optical packets. At the outgoing edge, the packets are reassembled. Bandwidth loss comes from the additional electronic header carrying the additional information needed to reassemble the packet, and also from the last optical packet presumably partial filled. On the contrary, if the optical packet is large enough, several blocks are aggregated in it. The details of the optical payload content should be stored in the optical header which leads to bandwidth loss and some bandwidth from optical payload may not be used. An additional delay is incurred, corresponding to the time to fill the optical packet.

A large optical packet seems to be a good solution since 50 percent of Internet packets are small packets, so the aggregation is an advantage as it does not constraint the capacity of core nodes. This solution improves the effective bandwidth used. However this solution may prevent the clients from efficiently using the network resources in case of low and unbalanced network load conditions, when a timer mechanism is used, this avoids the increase in delay introduced by optical packet creation process. There are many timer-based mechanisms that were proposed in the literature ([18], [19] and [20]).

In case of variable length optical packet, segmentation may still be needed when the upper layer packet size is larger than the network Maximum Transmission Unit (MTU). But the use of variable format avoids the complex processes of aggregation (at egress node) and packet extraction from optical payload (at ingress node). The gain in throughput is presumably related to the optical header size. The optical packet creation delay is given by the electronic to optical conversion and depends on the wavelength bit rate.

As far as the E/O interface is concerned, an optical packet with fixed format guarantees better performance than the variable format. First of all, this solution provides a good efficiency of adapting client information to optical layer only when large optical payloads are used. Therefore, a SAR mechanism was proposed in order to assure good utilization of network resources regardless the size of optical payload. Here, the segmentation is done "on demand" implying lower overhead and less complexity. Next, a fixed size of optical payloads may avoid scalability problems when the pattern of electronic packet size changes. Furthermore, an eventual system upgrade (i.e. passing from a wavelength modulated at 10 Gbps to 40 Gbps) leads to an increase in utilization of network resources.

Unfortunately, the optical systems employing fixed size of transport unit contain several parameters to be configured and this is not an easy task in presence of traffic with different QoS constraints. One important point is defining the size of optical packet which has an impact on the performance. Works done in [21], [22], [23] highlighted that the delay exhibited by clients leads to a complex behavior, where the queuing delay is a convex function of system utilization. At low loads, the optical packet creation may lead to poor delay characteristics. Furthermore, queuing delay grows up as optical payload increases. For example, a voice flow of 64 Kbps fills up a payload of 5 μs (i.e. 50 Kbit at 10 Gbps) in 781 *m*s! To avoid such situation, timer mechanisms should be deployed in order to efficiently prevent electronic information from starvation. Yet, a timer mechanism introduces a high quantity of overload [23] and hence underutilizes system resources. In congestion region, optical payloads partially filled introduce an overload that influences negatively the system performance. Thus, it is essential to use large optical payloads.

A variable format of optical packet leads to simple E/O interface. There are two parameters involved in the performance of optical packet creation. The first one is represented by the distribution of packet sizes in client networks. We have shown [23] that the performance of the interface based on transport unit with variable format is very sensitive to electronic packet size distribution. An eventual change in this distribution may lead to serious problems of scalability. The second one is the wavelength capacity and hence depends on transport infrastructure. Here, an increase in optical channel capacity reduces the efficiency of transporting upper layer clients and thus requires an upgrade of system resources.

The last step in electing the format of packet to be deployed in next generation of metropolitan optical packet switching networks, questions about the interaction between the packet format and the end-to-end network performance. Therefore, it is expressed by the interaction between packet format and the performance of MAC protocol. In this context, the discussion is reduced to an analysis of slotted and unslotted schemes of access protocol [24].

5 Performance Issues on MAC Protocol

5.1 Optical MAC Protocol

5.1.1 Traditional MAC

The MAC protocol is specially required to control the access of multiple nodes to the shared resources. Its main function is to detect or avoid collisions in packet insertion process to efficiently exploit the available bandwidth of the medium. Globally, there are two major categories of MAC protocols that are extensively deployed in the networks, namely static channel allocation using Time Division Multiplexing (TDM), and dynamic channel allocation using Carrier Sense Multiple Access (CSMA). In TDM, bandwidth is divided into fixed time slots which are allocated to different users. Every user can transmit on his allocated time slots only. This leads to a waste of bandwidth when reserved time slots are not used because the user has no data to transmit. TDM also requires global clock synchronization on the network. What is more, this protocol may do not work well with data traffic due to its sporadic nature.

The CSMA protocol is used when a node can listen for the occupation state of the carrier (transmission medium) and act accordingly. This protocol can work either in the synchronous mode (i.e., *slotted* CSMA) or in the asynchronous mode (i.e., *unslotted* CSMA). There are several versions of this protocol, including the CSMA with Collision Detection (CSMA/CD) and CSMA with Collision Avoidance (CSMA/CA). The main advantage of those protocols is their simplicity that translates into low cost of implementation, management and maintenance. Moreover, they are appropriate to support sporadic data traffic. In particular, the unslotted CSMA offers the capability of supporting variable length packets without complex and costly segmentation/assembly processes.

5.1.2 Optical Unslotted (OU-)CSMA/CA Protocol

The OU-CSMA/CA protocol is a derivation of the CSMA/CA, which has been widely used in wireless networks. In addition to the simplicity property inherited from CSMA, the CSMA/CA protocol provides more efficient resource utilization thanks to avoidance of collision. Recently, there is a number of works that extends this protocol to the optical domain, such as in HORNET's first version [14] and DBORN [16].

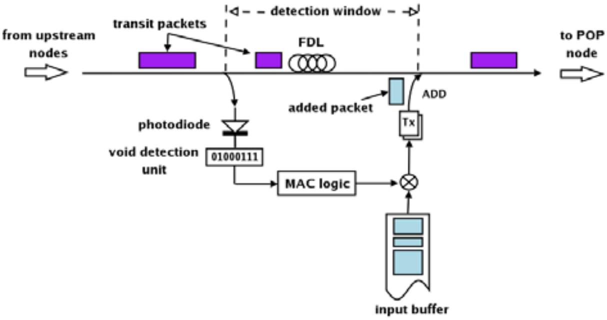


Fig. 5. OU-CSMA/CA schema

The network that uses OU-CSMA/CA protocol obviously supports variable length packets thanks to its asynchronous nature.

Fig. 5 schematizes a possible implementation of the OU-CSMA/CA protocol. The OU-CSMA/CA protocol functions based on the detection of idle periods (we also call it *voids*) on the transmission wavelength shared by several nodes. To detect activity on one wavelength, a node uses low bit rate photodiode, typically 155 MHz. Thanks to an optical coupler; the incoming signal is separated into two identical signals: the main transit signal that goes straightforwardly, and its copy that is received by the photodiode. The photodiode sends control information to the MAC logic which will inform the transmitters about whether it can transmit. The avoidance of collision is performed by employing a fiber delay line (FDL), which creates on the transit line a fixed delay between the control and the add/drop functions. This FDL should be long enough to provide the MAC logic with sufficient time to listen and to measure the medium occupancy. The FDL storage capacity should be at least larger than the maximum transmission unit (MTU) of the transport protocol used. For instance, if the network is used to transport Ethernet packets, then the storage capacity of the FDL should be greater than 1500 bytes, which is the Ethernet MTU.

The OU-CSMA/CA is appropriate to be used in a ring network such as the OPSR network above described, where there is problem of bandwidth sharing among ring nodes (e.g., on the transmission bus). Due to interdependence among ring nodes, an exact performance analysis of OPSR networks using OU-CSMA/CA protocol is difficult. However approximate methods may be used to assess the performance of such system. Among others, authors of [25], [26] have approximately analyzed the performance of this system using priority queuing theory. Globally, the authors have identified two main performance characteristics of OU-CSMA/CA protocol in OPSR networks. The first one is unfairness among ring nodes due to positional priority: upstream nodes (i.e., the nodes closest to the beginning of the shared transmission medium) might monopolize all the bandwidth, and prevent downstream nodes from transmitting. For instance, Fig. 6 (extracted from [25]) shows that the mean response time is likely to increase rapidly as we move downstream on the shared bus.

The second one is the bandwidth fragmentation due to the asynchronous nature of the OU-CSMA/CA protocol, resulting in low resource utilization. Indeed, the asynchronous transmission of packets at upstream nodes may fragment the global

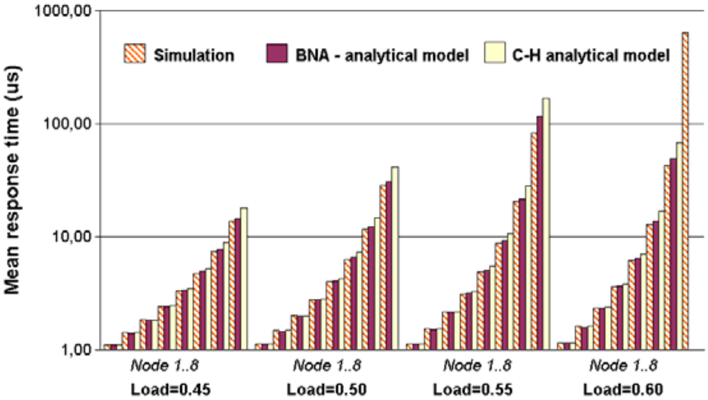


Fig. 6. Mean response time at ring nodes as function of ring load: "real-life" Internet traffic assumption

bandwidth into small voids unusable for the transmission of downstream nodes. The implication is that the network resource is used inefficiently, and the acceptable ring load of such system is limited to some 60% of the total network bandwidth as shown in Fig. 6.

5.2 Enhanced MAC Protocol

Since the combination of OU-CSMA/CA and OPSR network provides limiting performance as shown above, many researches have been carried out to find new access schemes enhancing the performance of such system. This subsection describes some new enhanced access schemes that provide remarkably higher network performance.

5.2.1 Packet Concatenation Mechanism

The transmission efficiency obviously appears one of great importance in exploiting network transmission resources. The transmission efficiency can be roughly defined as the ratio of useful bandwidth (i.e., the bandwidth occupied by client payload) to the effective bandwidth that the network must use to transport client payload. Of course, the effective bandwidth includes the client payload as well as possible overheads (headers and guard bands) needed for routing, signaling and header processing purposes. Thus, in order to increase the transmission efficiency of a given network, it is essential to reduce the total volume of overheads used to transport the volume of client data.

In the OPSR network considered, ring nodes use burst mode transceivers (BMT) to only communicate with the POP node on the transmission bus. Thus the optical header is reduced to a simple bit pattern for synchronization purpose. Therefore, in [27], a new mechanism called Modified Packet Bursting (MPB) has been proposed to increase the transmission efficiency of such network by suppressing unnecessary optical overheads. The concept of MPB mechanism relies on the basic concept of packet bursting in Gigabit Ethernet (GPB [28]), but with further improvements.

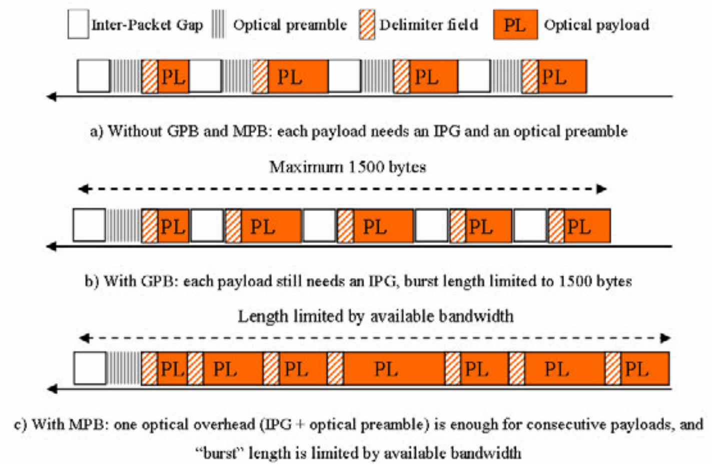


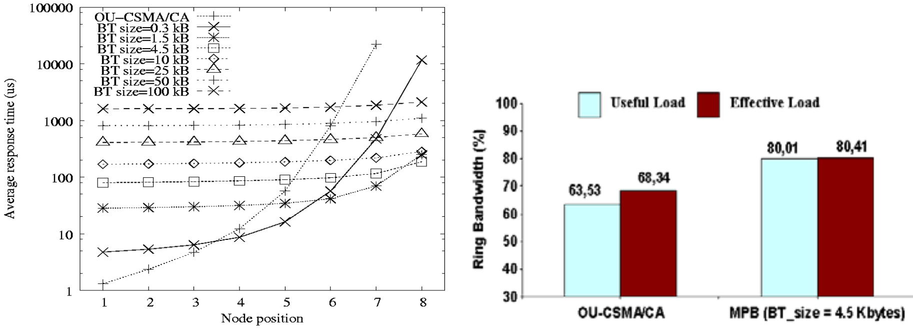
Fig. 7. Sequence of payloads transmitted consecutively with and without MPB

With MPB, in order to maximize the transmission efficiency, payloads (having the same destination) transmitted by a given node are concatenated to form a big “optical burst” consisting of one optical overhead followed by several payloads.

The principle of MPB is explained in Fig. 7. With MPB, if a node has a number of payloads to transmit, the first payload is transmitted with an optical overhead to initialize the “burst”; other payloads are then transmitted consecutively without any gap or optical overhead. Only one delimiter field is required at the beginning of each payload to separate two consecutive payloads. At the receiver side, the optical overhead is processed and consecutive payloads are extracted thanks to that delimiter field. An important point in this concept is that the “burst” length can be extended as long as the node has client payloads to transmit (including those arrive after the transmission of the first payloads) and the bandwidth is still available. The “burst” ends only when the node buffer becomes empty or there is no more available bandwidth.

Additionally, MPB also introduces the *bursting timer* principle to improve network performance. Actually, MPB uses a *bursting timer* (with *BT_size* parameter) defining the period of gathering client payloads for MPB. The longer the duration of this *gathering period*, the higher the probability of having more than one client payload cumulated in the buffer of the transmission node, hence the higher the transmission efficiency thanks to the concatenation of these payloads. Of course, a trade-off between transmission efficiency and access delay introduced by the bursting timer should be considered in MPB.

Regarding the performance of MPB versus OU-CSMA/CA scheme, it was pointed out that MPB effectively reduce the wasted bandwidth due to optical overhead, while providing satisfying performance for all ring nodes. For example, the performance results shown in Fig. 8 indicate that with a big enough *bursting timer* value, MPB may provide almost the same small mean response time at all nodes, while successfully transferring all client offered traffic with a negligible volume of optical overheads (e.g. around 0.4% against 5% in case with OU-CSMA/CA).



a) Avg. resp. time at each node vs. *BT_size* b) Useful and eff. loads: CSMA/CA vs. MPB

Fig. 8. MPB performance under offered ring load of 0.80 and real-life Internet traffic assumption

5.2.2 Fair Access Protocols

In previous sections we have identified that OU-CSMA/CA scheme suffers from unfair bandwidth sharing due to positional priority, and bandwidth fragmentation problem due to asynchronous packet insertion. For unslotted ring networks using CSMA-type protocols, there are few works in the literature addressing their fairness issue, due to the complexity in analyzing such systems by either simulation or mathematical methods. There are generally two ways of performing fairness control in a network: *centralized* and *distributed fairness schemes*. The former usually requires ring nodes to exchange control information among them. This may need dedicated control channel for transmitting feedback messages (e.g. [29]) or ring nodes must be able to handle transit traffic (e.g. [15] and [30]). This may also require a master node that controls the overall network state and parameterizes the operation of ring nodes (e.g. [26]). The main drawback of the centralized scheme for controlling fairness is that it may take a long time (due to control information exchange) for reacting to the change of network state.

Regarding the distributed fairness scheme, there are few works employing this approach, mostly owing to the difficulty of guaranteeing fairness among network nodes without knowing the global network state. Indeed, a distributed fairness scheme requires each node in the network to be able to improve the global network fairness while operating independently based on local knowledge only. The advantage of such approach is that a distributed fairness algorithm may immediately and adaptively act according to the change of the node state. This property makes distributed fairness scheme attractive to be used in modern MAN network where the traffic and the bandwidth demand are excessively changing/growing at an unprecedented rate.

An example of a distributed fairness scheme is the *Dynamic Intelligent MAC (DI-MAC)* protocol described in [31]. This protocol addresses both unfairness and bandwidth fragmentation issues in the OPSR network in question. The main idea of DI-MAC is to avoid inefficient bandwidth fragmentation by intelligently spacing out the transmission of local packets at upstream nodes, hence preserving usable voids for the transmission of downstream nodes.

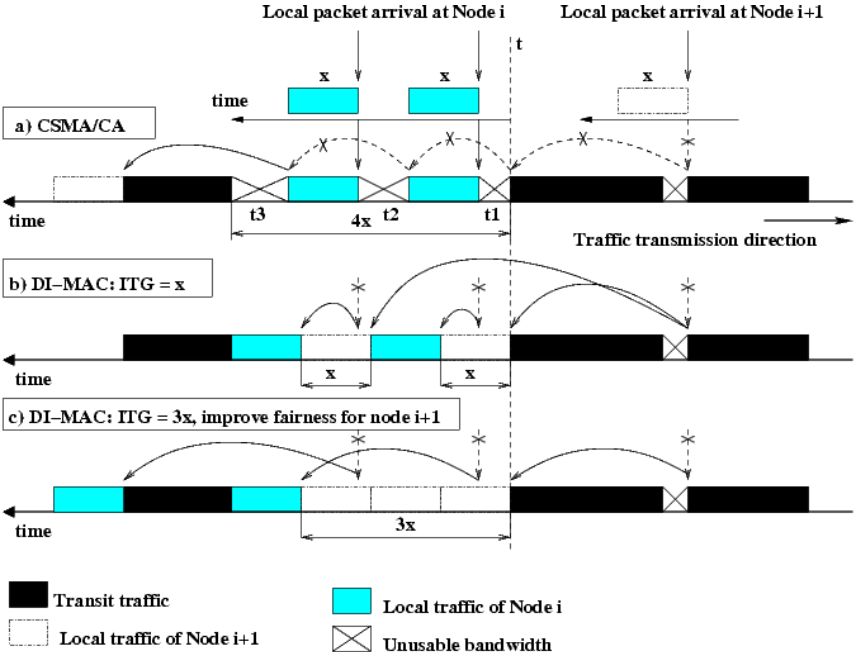


Fig. 9. Packet insertion process with OU-CSMA/CA (a) and DI-MAC (b, c)

The packet insertion process using OU-CSMA/CA at upstream nodes fragments the shared bandwidth into voids of variable lengths. Fig. 9 shows a state of the shared bandwidth at a given moment t when the node i observes a void of size $4x$, where x is the MTU of the transmission protocol used. Note that x is a metric that can be interpreted in time or bits. We suppose that in this example node i and node i+1 transmit packets of size x only, and at instant t their local buffers are empty. In OU-CSMA/CA case (Fig. 9-a), the asynchronous insertion of two packets at node i fragments the big void $4x$ into three small voids (whose size is t_1, t_2 *and* $t_3 < x$ respectively) that are not usable for the node i+1 to insert its packets of size x.

In Fig. 9-b, DI-MAC at node i intelligently fragments the big void into usable voids for the packet insertion of downstream nodes. Indeed, the transmission of the first local packet (the packet arrives at $t + t_1$) of node i is delayed, and it is transmitted only when a large enough void (e.g., equal to x) has been reserved for the downstream node i+1. The terms "large enough" mean that the void reserved by node i should allow the insertion of a maximum size packet of node i+1, …. The transmission of the second packet, in turn, is also delayed and carried out only when node i reserves a void equal to x after the successful transmission of the first packet. In such a way, the node i+1 becomes able to transmit its packets. The void reserved by

DI-MAC, through the control of the inter-transmission of local packets, is called Inter-Transmission Gap (ITG). ITG is a local parameter of DI-MAC implemented at each ring node. In Fig. 9-b, ITG is set to x. It is clear that ITG should at least equal to the MTU of the transmission protocol used, since this helps a node to release a void that is always usable for downstream nodes.

DI-MAC also provides a simple way to address fairness issue in unslotted OPSR network, without any additional control wavelength or synchronization operation. Indeed, the fairness enhancement can be done by simply increasing the ITG value at upstream nodes. Fig. 9-c shows an example with ITG equal to $3x$, where downstream nodes have larger bandwidth for transmitting their traffic.

It is worth mentioning that to effectively provide more bandwidth for downstream nodes, the first upstream node plays an important role in the operation of DI-MAC. Indeed, the first upstream node begins the fragmentation of the shared bandwidth. Thus the packet insertion operation of all other nodes strongly depends on the length of voids left between the inter-departures of local packets at the most upstream node. For instance, it is easy to see that the bandwidth state viewed be node 2 in case with DI-MAC will be a series of occupied and free bandwidth (voids) with void length greater or equal to ITG value of node 1.

Clearly, the choice of ITG value at each node strongly influences the performance of DI-MAC. Authors of [31] have proposed a dynamic distributed algorithm for controlling the ITG value at each node. Its principle is summarized as follows. Each node must always try to increase its ITG value whenever possible to release more bandwidth for downstream nodes (if any); at the same time, a node must try to guarantee that the waiting time of its client packets does not exceed a predefined threshold W_{max}, which may be freely chosen or deduced from the QoS parameters offered by service providers (e.g., from Service Level Agreement - SLA). This means that if a node observes that its client packets have been waited for a time considered too long, it must decrease the value of its own ITG instead of increasing it. The implication is

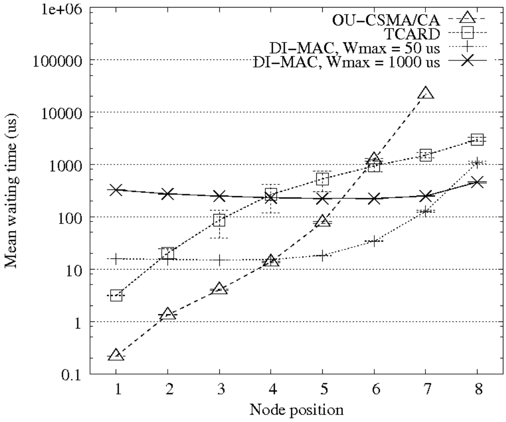


Fig. 10. Mean waiting time at each node with different access schemes: offered ring load = 0.80, uniform traffic pattern on the ring

that each ring node will be able to adjust its own ITG value that will have neither too big value nor too small value due to the ITG increasing process alternating with the ITG decreasing process. With this algorithm, DI-MAC remains fully-distributed since it requires only knowledge about the standard QoS parameters offered by service providers. And, it possesses the dynamic property since its ITG value changes with the variation of local packets waiting time, which clearly reflects the variation of bandwidth occupation state and of local packer arrival process.

Performance analysis of DI-MAC has shown that DI-MAC considerably improves the network performance (much higher than that obtained with OU-CSMA/CA) both in terms of resource utilization and performance parameters such as loss and delay. For instance, Fig. 10 plots the mean waiting time at each node obtained with DI-MAC, with OU-CSMA/CA and with a centralized mechanism TCARD [26]. With a big value of W_{max} (e.g., 1 ms), DI-MAC provides almost the same small mean waiting time for all nodes, hence offering the highest fairness level (and performance as well) to the network. Furthermore, this mechanism renders the network performance more stable and almost insensitive to network configuration and traffic change [31].

6 Guarantee of QoS in Multiservice Optical Metro Ring

Delivering services with guaranteed quality to users is an essential requirement for a metropolitan network. A metropolitan network typically interconnects enterprises, organizations and academic campus generating a variety of applications that might require quality-guaranteed services such as delay-sensitive (real-time) applications (voice, video conferencing, …), loss-sensitive applications (banking transaction, critical data transfer, …), or quality-non-guaranteed services such as best-effort Internet applications. The volume of data traffic today largely exceeds that of voice traffic. However, most revenues of service providers are still generated by voice service. Hence it points to a burning topic that MAN service providers should seek to construct novel common packet-based network infrastructure, which possesses the capability of supporting data traffic efficiently while delivering the same quality for voice service as in traditional TDM-based networks (e.g. SONET/SDH) to metropolitan users. This section is devoted to discuss some issues on the guarantee of QoS in a multiservice OPSR network.

6.1 QoS Architecture and Service Mapping

The global metropolitan traffic can be divided into two main categories: real-time traffic (e.g. TDM) and non-real-time traffic (e.g. data). The non-real-time traffic can be divided, in turn, into QoS-guarantee traffic (e.g. loss-sensitive data applications) and QoS-non-guarantee traffic (e.g. Internet best-effort traffic). These three traffic categories (or classes of service - CoS) usually cover all types of traffic that a MAN might be supposed to transport.

A metropolitan network like OPSR might interconnect many types of client networks, such as IP, SONET/SDH, ATM, Frame Relay…, each provides its own quality of service. Thus it is important to define a rule for the mapping of the services

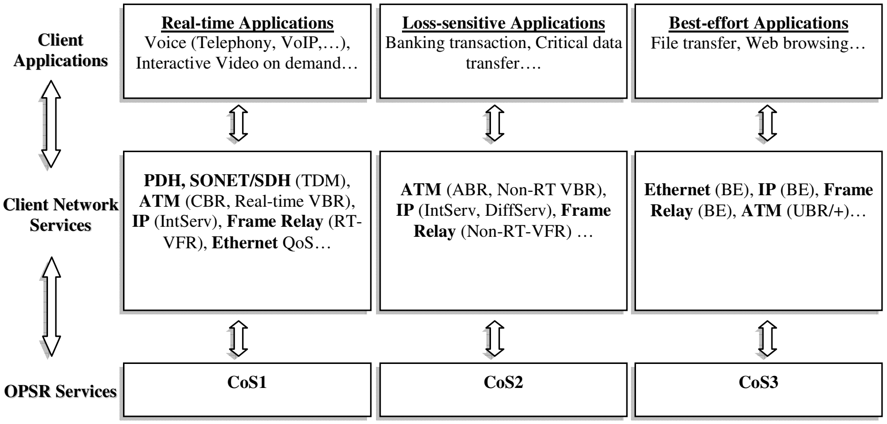


Fig. 11. Service mapping in OPSR network

provided by those client networks to the services offered by the OPSR network itself. Based on the assumptions on QoS architecture stated above, a possible service mapping solution is described in Fig. 11.

6.2 Transport of TDM Circuit across Packet-Based Networks

Currently, two main efforts exist in telecommunication research community, trying to guarantee the QoS to some types of traffic that a metro network may transport, mainly the real-time TDM traffic. The design of metropolitan architecture employing ring topology and OU-CSMA/CA protocol faces the fairness issue where the traffic transmitted at upstream nodes influences the performance of downstream nodes. To protect TDM-based flows performance from the influence of data traffic circulating on the bus, new protocols may be needed. Thus, hybrid protocols, which combines feature of CSMA/CA and token-passing-type, may represent a possible solution [32].

Another approach being supported by many standard organisations is to consider that the convergence between circuit- and packet-switched services should be done at the "packet" layer and therefore they concentrate on defining the functionalities and requirements at the interface between these two worlds. The technology allowing the convergence of circuit- and packet-switched services over a packet-based network is called Circuit Emulation Service (CES). This section focuses on this technology.

6.2.1 Circuit Emulation Service

Globally, Circuit Emulation Service is a technology allowing the transport of TDM service such as PDH (E1/T1/E3/T3) as well as SONET/SDH circuit over packet switching networks. The main intention of CES is to make the packet switching network behave as a standard TDM-based SONET/SDH/PDH network as seen from the customer's point of view. Thus CES should allow customers to be able to use the same existing TDM equipment, regardless of whether their traffic is carried by standard SONET/SDH/PDH network or a packet switching network using CES.

CES is supported by a number of organisms, including the Internet Engineering Task Force (IETF) [33], the Metro Ethernet Forum (MEF) [34], the International Telecommunication Union (ITU) and the Multi-Protocol Label Switching, Frame Relay and ATM alliance (MFA forum) [35], [36]. There are no important differences between the CES standards being defined by these organisms. They actually address different layers within the network (e.g. IP, MPLS, Ethernet…), and emphasize different aspects of the CES depending on the specific services they are concerned with. It is worth mentioning that using CES to transport voice across an IP network is not like voice over IP (VoIP) technology. Indeed, VoIP is used for transporting voice only, and it requires complex signaling protocols such as H.323 or SIP (Session Initial Protocol). In contrast CES can be used for transporting voice, video, and data over many types of packet switching network (not necessary IP network). Moreover it generally does not need gateway signaling as in VoIP, and requires low processing latency. Many believe that CES will displace VoIP in the future thanks to its efficiency and flexibility.

The reference model for CES on the metro OPSR network is based on the global model for circuit emulation described in IETF RFC3985 [37] and MEF3 specification [34]. Fig. 12 presents the reference model of CES for OPSR network. Generally speaking, we have two TDM customers' edges (CE) communicating via OPSR. One CE is connected to an AP node (ingress CE), the other CE is connected to the POP node (egress CE). TDM service generated by ingress CE is emulated across the OPSR network to egress CE. The emulated TDM service between two CEs is managed by two inter-working functions (IWF) implemented at appropriate nodes.

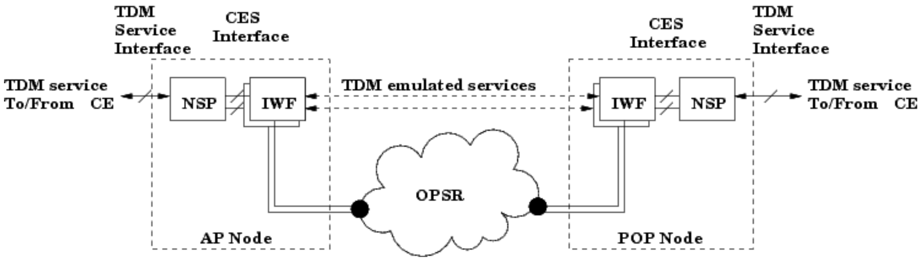


Fig. 12. Reference model for CES on metropolitan OPSR network

Operation Modes. CES has two principal modes of operation. In the first one, called "unstructured" or "structure-agnostic" emulation mode, the entire TDM service bandwidth is emulated transparently, including framing and overhead present. The frame structure of TDM service is ignored. The ingress bit stream is encapsulated into an emulated TDM flow (also called CES flow) and is identically reproduced at the egress side. The second mode, called "structured" or "structure-aware" emulation, requires the knowledge of TDM frame structure being emulated. In this mode, individual TDM frames are visible and are byte aligned in order to preserve the frame structure. "Structured" mode allows frame-by-frame treatment, permitting overhead stripping, and flow multiplexing/demultiplexing. This means that a single "structured" TDM service may be decomposed into two or more CES flows, or two or more "structured" TDM services may be combined to create a single CES flow as well.

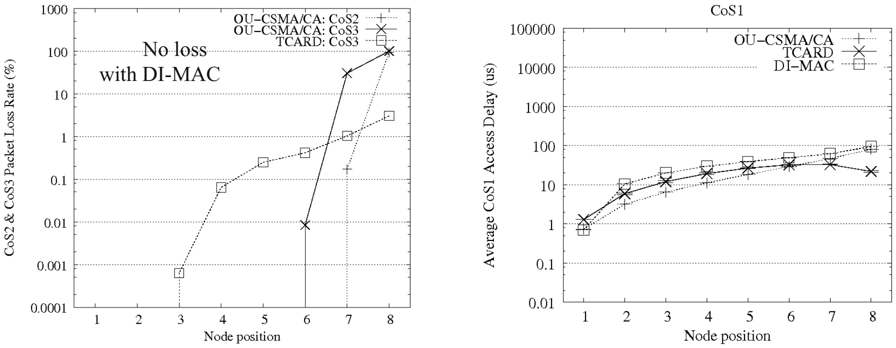
Functional Blocks. In the reference model of CES, the Native Service Processing (NSP) block performs some necessary operations (in TDM domain) on native TDM service such as overhead treatment or flow multiplexing /demultiplexing, terminating the native TDM service coming/going from/to CE. For instance, as the "unstructured" TDM service does not need framing treatment, it might not be handled by the NSP and can pass directly to the IWF block for emulation. However, the "structured" TDM service should be treated by the NSP block before going to the IWF block. Actually, the NSP could be the standard SONET/SDH framer or map per, or other propriety products.

The Inter-Working Function (IWF) block could be considered as an adaptation function that interfaces the CES application to the OPSR layer. This means that the CES technology could be considered as an application service that uses the OPSR network as a virtual wire between two TDM networks. Thus, the IWF block is responsible for ensuring a good operation of the emulated service. The main functions of IWF are to encapsulate TDM service in transport packets, to perform TDM service synchronization, sequencing, signalling, and to monitor performance parameters of emulated TDM service. Each TDM emulated service requires a pair of IWF installed respectively at ingress and egress sides.

TDM Frame Segmentation. A TDM frame would ideally be relayed across the emulated TDM service as a single unit. However, when the combined length of TDM frame and its associated header exceeds the MTU length of the transmission protocol used, a segmentation and re-assembly process should be performed in order to deliver TDM service over the OPSR network. By this consideration therefore, two segmentation mechanisms have been proposed in [38]. The first one, called *dynamic segmentation*, fragments a TDM frame into smaller segments according to void length detected on the transmission wavelength. This approach promises a good use of wavelength bandwidth, but is technically complex to implement. The second segmentation method is *static segmentation*, which segments the TDM frame according to a predefined threshold. This technique is simple to implement, and it provides resulting TDM segments with predictable size. Thus current TDM monitoring methods could be reused, simplifying the management of CES. Additionally, in [39] authors have recommended some rules to determine the segmentation threshold for TDM frames. First, the segmentation threshold should be either an integer multiple or an integer divisor of the TDM payload size. Second, for unstructured E1 and DS1 services, the segmentation threshold for E1 could be 256 bytes (i.e., multiplexing of 8 native E1 frames), and for DS1 could be 193 bytes (i.e., multiplexing of 8 native DS1 frames).

6.2.2 Performances of Circuit Emulation

Fig. 13 illustrates performance parameters of the OPSR network measured at the offered network load of 0.80 using different access schemes ([25]). Globally, OU-CSMA/CA and TCARD schemes may provide satisfying QoS for TDM service, but at the expense of degrading or even losing low priority traffic. However, DI-MAC scheme seems to support better all types of service. It effectively guarantees expected QoS for TDM service, and, at the same time provides low loss and low packet delay for lower priority services.



a) Packet loss rate of CoS2 and CoS3

b) Average access delay of CoS1

Fig. 13. Performance results on CES: offered ring load = 0.80, TDM segmentation threshold = 810 bytes

7 Summary

We have provided in this paper an overview of the optical metropolitan networks. The introduction of new optical networking into the future MAN to improve, not to say replace, the current circuit-based networks is becoming more and more urgent due to the rapid change of traffic and of user demand in today's MAN. Among many possible solutions, the OPS technology was identified as a good candidate for the future MAN thanks to its flexibility, scalability and cost-efficiency.

Many technological aspects of OPS networks have been investigated in this paper, namely the optical packet format issues, the performance improvements of MAC protocol and the guarantee of QoS in a multi-service environment. Globally, the choice of optical format has important impact on the network performance. Numerous studies have shown that there is no a unique optimal choice for optical packet format, but this choice depends on a number of factors: the traffic profile, the different QoS requirements, the network segment (edge or core), etc. After the packet format, comes the MAC protocol that is an essential issue in the design of a MAN. The aim of a MAC protocol is to control the multiple accesses to shared resources in such a way that the network resources are exploited best. A simple and cost-efficient MAC protocol that was proposed for metro OPS ring networks is the optical unslotted CSMA/CA protocol. Due to its limit in terms of performance, this protocol needs further improvements to overcome some drawbacks, namely the inefficient bandwidth utilization due to bandwidth segmentation and the unfairness among ring nodes. These issues can be resolved using packet concatenation mechanisms (e.g. Modified Packet Bursting) or / and fairness control schemes (e.g. Dynamic Intelligent MAC), which were proved efficient and robust in increasing network performance. Finally, these schemes combined with the Circuit Emulation Service technology are able to provide service distinction for multi-service optical metro networks. More specifically, with CES technology, a metro OPS network is able to support sporadic packet-based services while guaranteeing TDM-like QoS for circuit-based services.

References

1. FDDI Token Ring – Media Access Control. ANSI X3. 139-1987 American National Standards Institute (1986)
2. IEEE standard for Local and Metropolitan Area Networks - Distributed Queue Dual Bus (DQDB) Subnetwork of a Metropolitan Area Network (MAN). IEEE STD 802.6-1990 (1991)
3. Gambini, P., Renaud, M., Guillemot, C., Callegati, F., Andonovic, I., Bostica, B., Chiaroni, D., Corazza, G., Danielsen, S.L., Gravey, P., Hansen, P.B., Henry, M., Janz, C., Kloch, A., Krähenbühl, R., Raffaelli, C., Schilling, M., Talneau, A., Zucchelli, L.: Transparent Optical Packet Switching: Network Architecture and Demonstrators in the KEOPS Project. IEEE Journal on Selected Areas in Communications 16(7) (1998)
4. Stavdas, A., Sygletos, S., O'Mahoney, M., Lee, H.L., Matrakidis, C., Dupas, A.: IST-DAVID: Concept Presentation and Physical Layer Modeling of the Metropolitan Area Network. Journal of Lightwave Technology 21(2) (2003)
5. Dixit, S.: IP over WDM: Building the Next Generation Optical Internet. John Wiley & Sons, New Jersey (2003)
6. ROM-EO – Réseau Optique Multi-Service Exploitant des Techniques Electro-optiques et Opto-electroniques, `http://www.telecom.gouv.fr/rnrt/rnrt/projets/res_02_52.htm`
7. Karagiannis, T., Broido, A., Faloutsos, M., Claffy, K.C.: Transport Layer Identification of P2P Traffic. In: Internet Measurement Conference 2004 (2004)
8. TeleGeography Report of Global Internet Geography, `http://www.telegeography.net`
9. Generic Frame Protocol (GFP). G. 7041 ITU-T Standard (2001)
10. Link Capacity Adjustment Scheme (LCAS). G.7042 ITU-T Standard (2001)
11. Hunter, D.K., Nizam, M.H., Chia, M.C., Andonovicand, I., Guild, K.M., Tzanakaki, A., O'Mahony, M.J., Bainbridge, J.D., Stephens, M.F.C., Penty, R.V., White, I.H.: WASPNET A Wavelength Switched Packet Network. IEEE Communications Magazine 37(3), 120–129 (1999)
12. Yao, S., Ben Yoo, S.J., Mukherjee, B., Dixit, S.: All-Optical Packet Switching for Metropolitan Area Networks: Opportunities and Challenges. IEEE Communications Magazine (2001)
13. Jourdan, A., Chiaroni, D., Dotaro, E., Eilenberger, G.J., Masetti, F., Renaud, M.: The Perspective of Optical Packet Switching in IP-Dominant Backbone and Metropolitan Networks. IEEE Communications Magazine (2001)
14. White, I.M.: A New Architecture and Technologies for High-Capacity Next Generation Metropolitan Networks. Ph.D. Dissertation, Department of Electrical Engineering of Stanford University, CA (2002)
15. Resilient Packet Ring. IEEE 802.17-2004 IEEE Standard (2004)
16. Le Sauze, N., Dotaro, E., Dupas, A., et al.: DBORN: A Shared WDM Ethernet Bus Architecture for Optical Packet Metropolitan Network. In: Photonic in Switching Conference (2002)
17. Callegati, F.: Optical Buffers for Variable Length Packets. IEEE Communication Letters 4(9), 292–294 (2000)
18. Nguyen, T.D., Eido, T., Atmaca, T.: Impact of fixed-size packet creation timer and packet format on the performance of slotted and unslotted bus-based optical MAN. In: ICDT 2008, Bucharest, Romania (2008)
19. Nguyen, T.D., Eido, T., Atmaca, T.: DCUM: Dynamic creation of Fixed-Size Containers in Multiservice Synchronous OPS Ring Networks. In: QoSim 2009, Roma, Italy (2009)

20. Nguyen, D.T., Eido, T., Atmaca, T.: Performance of a Virtual Synchronization Mechanism in an Asynchronous Optical Networks. In: AICT 2009, Venise, Italy (2009)
21. Popa, D., Atmaca, T.: On Optical Packet Format and Traffic Characteristics. In: IEEE/Euro-NGI SAINT 2005 Workshop on Modelling and Performance Evaluation for QoS in Next Generation Internet, Trento, Italy, pp. 292–296 (2005)
22. Popa, D., Atmaca, T.: Unexpected Queueing Delay Behaviours in Optical Packet Switching Systems. In: Euro-NGI HET-NET 2005 Conference, Ilkley, UK (2005)
23. Popa, D.: Performance Issues in Metropolitan Optical Networks: Packet Format, MAC Protocol and QoS. PhD Dissertation, Telecom Sud-Paris, France (2005)
24. Atmaca, T., Kotuliak, V., Popa, D.: Performance Issues in All-Optical Networks: Fixed versus Variable Packet Format. In: Photonics in Switching 2003 (PS 2003), Versaille, France, pp. 292–296 (2003)
25. Nguyen, V.H.: Performance Study on Multi-service Optical Metropolitan Area Network: MAC Protocols and Quality of Service. PhD Dissertation, Telecommunication Networks and Services Department, Telecom Sud-Paris, France (2006)
26. Bouabdallah, N., Beylot, A.L., Dotaro, E., Pujolle, G.: Resolving the Fairness Issue in Bus-Based Optical Access Networks. IEEE Journal on Selected Areas in Communications 23(8) (2005)
27. Nguyen, V.H., Atmaca, T.: Performance analysis of the Modified Packet Bursting mechanism applied to a metropolitan WDM ring architecture. In: IFIP Open Conference on Metropolitan Area Networks: Architecture, protocols, control and management, pp. 199–215. HCMC, Viet Nam (2005)
28. IEEE 802.3z and 802.3ab - Gigabit Ethernet. IEEE Standard (1998)
29. Bhargava, A., Kurose, J.F., Towsley, D.: A hybrid media access protocol for high-speed ring networks. IEEE Journal on Selected Areas in Communications 6(6), 924–933 (1988)
30. Tang, H., Lambadaris, I., Mehrvar, H., et al.: A new access control scheme for metropolitan packet ring networks. In: IEEE Global Telecommunications Conference Workshops (GlobeCom Workshops 2004), pp. 281–287 (2004)
31. Nguyen, V.H., Atmaca, T.: Dynamic Intelligent MAC Protocol for Metropolitan Optical Packet Switching Ring Networks. In: IEEE International Conference on Communications – ICC 2006, Istanbul, Turkey (2006)
32. Popa, D., Atmaca, T.: Performance Evaluation of a Hybrid Access Protocol for an All-Optical Double-Bus Metro Ring Network. In: ICOCN Conference, Hong Kong (2004)
33. Structure-agnostic TDM over Packet, SAToP (2005), `http://www.ietf.org/internet-drafts/draft-ietf-pwe3-satop-03.txt`
34. Circuit Emulation Service definitions, framework and requirements in metro Ethernet networks (2004), `http://www.metroethernetforum.org/PDFs/Standards/MEF3.pdf`
35. TDM Transport over MPLS using AAL1 Implementation Agreement (2003), `http://www.mfaforum.org/tech/MFA4.0.pdf`
36. Emulation of TDM Circuits over MPLS Using Raw Encapsulation Implementation Agreement (2004), `http://www.mfaforum.org/tech/MFA8.0.0.pdf`
37. Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture, RFC1985 (2005), `http://www.ietf.org/rfc/rfc3985.txt?number=3985`
38. Nguyen, V.H., Ben Mamoun, M., Atmaca, T., et al.: Performance evaluation of Circuit Emulation Service in a metropolitan optical ring architecture. In: de Souza, J.N., Dini, P., Lorenz, P. (eds.) ICT 2004. LNCS, vol. 3124, pp. 1173–1182. Springer, Heidelberg (2004)
39. Structure-aware TDM Circuit Emulation over Packet Switched Network, CESoPSN (2005), `http://www.ietf.org/internet-drafts/draft-ietf-pwe3-cesopsn-06.txt`