QoS of Optical Packet Metro networks

Annie Gravey, Philippe Gravey, Michel Morvan, Bogdan Uscumlic, Lida Sadeghioon

Telecom Bretagne, France. firstname.familyname@telecom-bretagne.eu

Abstract: Metro networks support increasing traffic volumes and evolving traffic profiles. Revisiting metro networks architecture, this paper shows that both optical transparency and sub-wavelength granularity can be achieved, while still ensuring transport network QoS levels.

OCIS codes: 060.1155, 060.4256, 060.4259, 060.4261, 060.4262, 060.4265.

1. Requirements on MANs

Several forecasts have emphasized that distribution/aggregation networks, also called Metro Area Networks (MAN), are particularly impacted by traffic evolution. They have to support a sharp increase in utilization, together with a strong demand on functions (such as video content distribution) that were not used by traditional telephony services.

Future MANs should fulfil several requirements. Flexibility, facilitating a quick adaptation to varying traffic demands, in terms of supported protocols and traffic profiles, is mandatory. An efficient support of both fine granularity and large volumes of traffic demands, for uplink and downlink traffics, is necessary, as MANs have to interconnect both low activity nodes (e.g. small DSLAMs) and high activity nodes (e.g. data centres). MANs should provide methods for isolating different clients' flows, together with an excellent QoS, including high reliability and fast protection. Lastly, energy efficiency is a must, in order to limit Operational Expenditures (OPEX).

The above requirements seem difficult to be achieved simultaneously. MAN networks currently rely on optical links; in opaque networks, nodes operate in the electronic layer (e.g. Ethernet), while transparent networks are fully optical between source and destination. Currently deployed transparent networks present a coarse (wavelength) granularity, which precludes an efficient MAN usage. As contention avoidance is easily implemented in opaque networks (thanks to electronic buffers), the current MANs are often Ethernet rings implementing specific protection protocols [1,2]. The main issues with these networks are their high energy consumption on the one hand and their fixed Ethernet packet granularity that is adequate for Metro Access areas, but is too fine for Metro Core areas.

Optical packet/burst switching (OPS/OBS) have been for many years considered as potential options to combine sub-wavelength granularity and optical transparency. However, technological immaturity and doubts on their ability to achieve a high multiplexing gain together with a QoS similar to the one provided by electronic switching have prevented their use in operational networks. The key issue is contention avoidance, as optical buffering is not a currently viable option. This paper shows that a specific OPS architecture, namely a Packet Optical Add and Drop Multiplexing (POADM) ring [3], can meet most of these requirements.

The paper is organized as follows: the next section presents the POADM ring architecture together with the main features of an appropriate MAC for multi-ring POADM networks. We then put the emphasis on the importance of a suitable network dimensioning, as a mandatory step for providing simple operation and transport network level QoS. This is illustrated by some quantitative evaluation results regarding both delivered QoS and dimensioning cost of POADM networks. Lastly, we show how POADM rings can be used to directly support Metro Ethernet services and thus allow collapsing transport network layers.

2. POADM ring architecture and label-based operation

A POADM ring is fully transparent, and is thus more energy efficient than an opaque ring. It operates on fixed size optical slots in order to facilitate multiplexing efficiency, and relies on electronically processed control information that is carried on a separate wavelength. Multiple data channels, carried on different wavelengths, are controlled by a single control channel; this naturally segregates user data from control data, which protects the network from many attacks. Control and data packets are synchronized. Each node has a single fast tunable transmitter, which can select a different wavelength at each time slot to send a data packet. One or several receivers allow the node to receive multiple optical data packets per time slot, whereas it can send only one data packet per time slot. Data channels are shared between all nodes, contrarily with alternate implementations [4] where a wavelength is used by a single destination.

Each node has one or several client interfaces. An interface includes an adaptation layer, which first encapsulates client data into variable size Service Data Units (SDUs) and then aggregates multiple SDUs in fixed size Protocol Data Units (PDUs). Labels at SDU levels identify client flows, whereas labels at PDU level are used to identify the flows between source and destination nodes. Multiple client protocols (Ethernet, IP, MPLS...) can thus transparently share a single POADM ring. Control PDUs carry both general information (e.g. OAM messages) and information specific to each data channel; this specific information is relative to whether a data PDU is carried or not, and, if a data PDU is carried, its characteristics, including a label stack.

A classical label based operation is implemented in each node thanks to multiple local tables: Forwarding Equivalence Class (FEC) tables, as in MPLS, identify client flows, Client To SDU (CTS) tables bind a label to each client flow, a SDU To PDU (STP) table, binds one or several SDU labels to a PDU label, a Forwarding Information Table (FIT) identifies the set of actions to perform on those PDUs that have a label entry in the FIT. Incoming PDUs marked with an unknown label are passed transparently. For a given PDU label, the FIT can specify one or several of the following actions: insertion, reception, erasure. For example, a source node shall insert PDUs on the ring; the destination node of a unicast flow shall receive the PDU and erase it in order to facilitate spatial reuse; on the other hand, a destination node of a multicast flow shall only receive the PDU while letting it pass for downstream destinations; a multicast extract node shall erase the PDU (and possibly not receive it, if it is e.g. the multicast source node); a hub node that interconnects two or more rings through O/E/O conversion shall receive and erase a PDU from one ring, and insert it on the other ring, having potentially operated on the label stack (swapping, pushing or popping labels).

The FEC, CTS, STP and FIT tables can be static (i.e. configured manually or by the management plane) or dynamic (i.e. configured by control messages carried in control PDUs).

3. The cost of QoS in Transport Networks

A transport network relies on physical resources that have been provisioned, based on conservative traffic predictions, as making new resources available involves lengthy processes such as deploying new optical fibres. In the past, transport networks relied on sets of permanent, fixed rate circuits established between each (source, destination) pair. The circuit's capacity could greatly exceed the required peak traffic rate, depending on the available granularity of circuits. A MAN build with POADM rings still relies on optical fibres, but replaces fixed rate circuits by fixed rate virtual circuits, which potentially allows optimizing bandwidth usage.

However, the need to provision resources based on predicted peak traffic rates should still be taken into account. We thus assume that the MAN operator monitors network usage in order to regularly update a "Traffic Matrix" T that recaps peak traffic requirements. Assuming that the network operator provisions enough resources to support T, we propose to use a simple "opportunistic" PDU insertion scheme: a node can insert a PDU on the first free timeslot present on a convenient data channel. As traffic arriving to a node is aggregated, the insertion process can be modelled by a a simple discrete time Geo/Geo/1 queue [5], where PDU generation at a node is modelled by a Bernoulli process and the distribution of the number of slots between two free timeslots is geometrically distributed [6].

As long as client layers conform with T (conformance can be enforced at each node by a simple access control procedure), the performance delivered by POADM rings can be shown to comply to QoS objectives set for Carrier Ethernet in a MAN [7]. Indeed, latency is mostly due to geographical distance, as POADM nodes are optically transparent for transit data PDUs; one can safely assume that the distance between nodes in a MAN is less than 1000km, which yields a latency less than 10ms. Jitter, on the other hand, is due to the insertion process. Using a Geo/Geo/1 model, one can show that even in the pessimistic case where only one data channel is used, jitter is less than 0.5ms as shown in Fig. 1. Even if a flow has to cross 3 POADM rings, and therefore be inserted 3 times, the end-to-end jitter is still less than 2ms. Lastly, POADM operation is loss free, except in case of link or node failure.

It is worthwhile noting that "fairness" is not an issue for a properly dimensioned POADM ring: although some labelled flows may receive a better level of QoS than others, all receive Carrier Ethernet levels of QoS.

POADM multi-ring networks can protect labelled flows with mechanisms mimicking the classical 1+1 and 1:1 protection schemes. Bidirectional rings are necessary; a protected flow is allocated a single label on both directions of the ring. A "working" and a "backup" flow are always sent for a "premium" flow (relying on 1+1 protection), and the destination node only receives one of these flows. For a "regular" flow, relying on 1:1 protection, only one of those flows is sent, the sending of the backup flow is triggered by the reception of a failure notification carried in control PDUs. Failure detection is very rapid, as it is detected when a control packet is not received; failure notification is directly sent by the two nodes adjacent to the failure, and all nodes are made aware of the failure in a few ms (depending on the ring size). Each node stores in a local Protection Information Table (PIT) the set of flows that are affected by each potential failure. When a failure notification is received, each node can then apply the appropriate protection mode for each impacted flow. With these procedures, the unavailability duration for a given flow due to a fibre failure is less than the round trip time between source and destination nodes, i.e. less than 10ms.

When dimensioning a POADM ring with protected labelled flow, one should derive from T and the protection mode for each labelled flow each matrix T_l corresponding to the traffic matrix in case of failure of link l. It can be shown that protecting a POADM ring may double the cost of a network, and that 1:1 is slightly less expensive than 1+1 protection, as supplementary resources planned for regular traffic are shared between all regular flows, making this mechanism similar to 1:N circuit-switched protection; this is illustrated by Fig. 2.

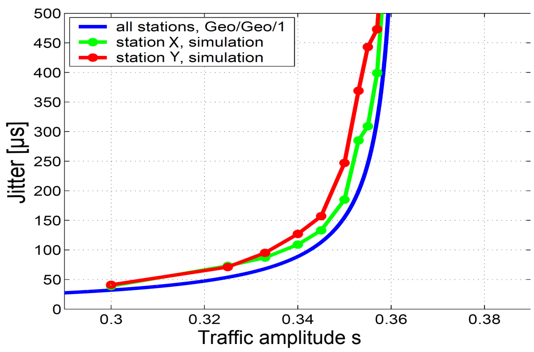

Fig. 1: 6-node POADM ring with symmetric traffic of amplitude s. Jitter in nodes X and Y [6].

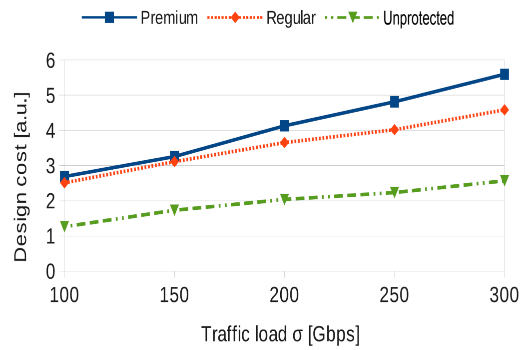

Fig. 2: 5-node POADM ring with random traffic of load σ. Design cost [8].

Assuming a fibre link failure rate less than 3 per year per 1000km, both 1+1 and 1:1 protection schemes yield an availability rate higher than 99.999%, with a restoration time significantly less than 50ms. Moreover, POADM protection schemes are flow based as in MPLS-TP [9], and not channel based as in RPR and EPR.

4. Conclusion: collapsing layers in MAN architecture

Labelling each client flow by encapsulating client information units within SDUs allows to share a single PDU layer by multiple client flows. The QoS delivered to PDUs is Carrier-Ethernet compatible, and the protection methods provided for POADM rings are as flexible as those designed for MPLS-TP. Moreover, the POADM architecture provides a clean separation between data and the control planes. The data plane is controlled by various local tables (FEC, FTS, STP, FIT and PIT) that can be static (as in the current MANs), periodically distributed by a centralized control plane (as in a SDN framework), or locally computed thanks to some distributed procedures (as in a G-MPLS framework). Lastly, control PDUs carried in the control channel can be used to disseminate OAM messages instantaneously.

POADM thus seems a good candidate to deploy a universal MAN architecture, which can be used to carry any client protocol layer. An intermediate architecture such as MPLS-TP or PBB-TE is made redundant. A fully converged architecture could thus present multiple client layers carried over POADM, which is then carried over WDM. This does not preclude using wavelength channels with OTN for very large flows that do not require fine granularity.

Acknowledgements

The research leading to this paper has received funding from the European Community's 7th Framework Programme FP7/2013-2015 under grant agreement 317762 COMBO project and from the French Ministry of Industry in the framework of the CELTIC+ SASER-Savenet project.

References

1. ITU-T, *G.8032/Y.1344 Ethernet ring protection switching*, 2012.
2. *IEEE 802.17 Resilient packet ring (RPR)*, IEEE Computer Society, 2004.
3. D. Chiaroni, *Optical packet add/drop multiplexers for packet ring networks* , 34th European Conference on Optical Communication, 2008. ECOC 2008.
4. Intune Networks *Optical Packet Switch and Transport. A Technical Introduction*, 2009 http://www.intunenetworks.com/home/shapeup/core-innovation/opst-technical-introduction/
5. J. Cao, W. Cleveland, D. Lin, and D. Sun, *Internet traffic tends toward Poisson and independent as the load increase*, in Nonlinear Estimation and Classification, D. Denison, M. Hansen, C. Holmes, B. Mallick, and B. Yu, Eds. New York, NY: Springer Verlag, Dec. 2002.
6. B. Uscumlic, A. Gravey, M. Morvan, P. Gravey, *Impact of peer-to-peer traffic on the efficiency of optical packet rings*, WOBS'2008 : international Workshop on Optical Burst/Packet Switching. London, UK, 2008.
7. *MEF 23.1, Implementation Agreement, Carrier Ethernet Class of Service Phase 2*, January 2012.
8. B. Uscumlic, L. Sadeghioon, A. Gravey, P. Gravey, *The Cost of Traffic Protection in Bidirectional Optical Packet Switching Rings*, ISCC 2013 : 18th IEEE Symposium on Computers and Communications. Croatia, 2013.
9. *MPLS-TP Ring Protection Switching (MRPS)*, draft-helvoort-mpls-tp-ring-protection-switching-05.txt, 2013.