This paper was presented as part of the High-Speed Networks 2011 (HSN 2011) Workshop at IEEE INFOCOM 2011

Ethernet Burst Transport for Next Generation Optical Metro Networks

Angelo Germoni and Patrizia Testa
Co.Ri.TeL.
Consorzio di Ricerca sulle TeLecomunicazioni
Via Anagnina 203, Roma, Italy
Email: angelo.germoni@coritel.it

Roberto Sabella
Ericsson Research
Ericsson Telecomunicazioni
Pisa, Italy

Marco Listanti
DIET
University of Rome ”Sapienza”
Roma ,Italy

***Abstract*—The main requirement for the Next Generation Transport Network infrastructure is a flexible and efficient support of different services, demanding for several levels of Quality of Service (QoS) and resilience. In order to have an effective utilization of network resources, and the ability to react to traffic demand changes with time, such multi-service next generation transport networks, should be, to some extend, self-adapting. This requirement are pushing the migration from the traditional legacy circuit based transport networks towards integrated packet optical solutions. The need to introduce packet flexibility into the optics world relying on huge and reliable static pipes, without impacting the scalability of the nodes has lead to multilayer solutions such as current MSPP and POTP platforms based on multiple switching layers (i.e. packet, OTN and optical). This however requires complex control plane functionalities that limit their effectiveness and flexibility. This paper presents a new approach for next generation optical packet transport, based on a pure Layer 2 switching, that is Ethernet compliant since it does not require changes in Ethernet frame format and main Ethernet switch functionalities. It relies on a burst transmission structure that allows to reduce packet processing without introducing underlaid switching layers and consequently to scale switch forwarding functionalities. It could be regarded as a concrete step towards the realization of self-adapting networks. Some relevant simulation results are reported to discuss the main characteristics of such a new transport solution and assess the feasibility of the concept.**

I. Introduction

Transport networks are facing significant challenges in order to scale in capacity and meet more stringent requirements. The necessity to support services so as they are affordable for users and profitable for operators requires that savings must be made in both network deployment and operation. Given the uncertainty about the size, shape and timing of the new requirements, is crucial to maintain network flexibility to be able to respond to changes. The close coupling of traditional transport networks and the services provided over them means they are inflexible when it comes to introducing new services. In order to improve flexibility and benefit from finer granularity in traffic multiplexing and better capacity utilization, a strict integration between packet and optical layer is required [1]. Different approaches for a packet-based transport network, considered the pillar of next generation multi-service infrastructure, are emerging such as PBB-TE and MPLS-TP [2]–[4]. Such technologies are able to replicate SDH carrier class performance and provide tunnel switching, allowing to remove coupling between transport and services, and aggregation of flows. On the other hand Ethernet technology is assuming higher networking responsibility and high speed Ethernet switches are continuing to evolve to accommodate changes in networked applications and to pave the way for the next generations of Ethernet at 100 Gbps and 1 Tbps. With this migration towards a fully packet based architecture, packet processing at very high rate has becoming a key challenge for network elements [5] requiring to provide Ethernet switches with efficient aggregation capabilities and more scalable forwarding functionalities. Hardware duplication could represent an approach to solve this issue, but this impacts cost, footprint and power consumption that are yet approaching their physical limits. Alternative solutions focus on packet processing reduction, such as increase Minimum Frame size, increase MTU (introducing for instance Jumbo Frames) or aggregate frames in a larger one [6]. However, all these approaches require to modify the Ethernet frame structure.

In this paper, we propose an aggregation method, that does not need to modify the Ethernet frame format, based on a burst transmission structure [7]. A burst consists of a variable number of consecutive frames of the same flow separated by a proprietary inter-frame gap (IFG) and preceded by an Ethernet Burst Control Frame. Such an approach allows to reduce forwarding tables dimension and packet processing, being burst classification performed by just processing the control frame. The adoption of the Burst Control Frame introduces an additional signaling in the Ethernet network and can be exploited to support smart traffic management functionalities. In this paper we introduce, a dynamic and flexible load-balancing over aggregated Ethernet links performing burst reordering at each transit node on the basis of burst sequence numbers carried by the Bust Control Frame.
The paper is structured as follows. Section II introduces the main Ethernet Burst Transport features; section III gives an example of a metro network scenario implementing such solution considering both the node and network perspectives. In section IV the results of some relevant simulation results are reported and finally in section V some conclusion are drawn.

978-1-4244-9920-5/11/$26.00 ©2011 IEEE

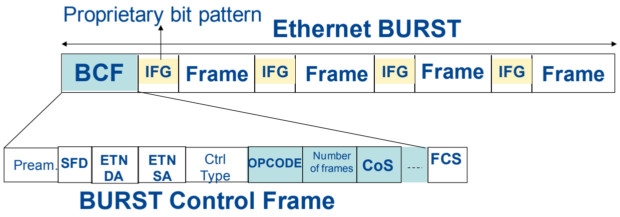

Fig. 1. The Burst Control Frame Structure, IFG

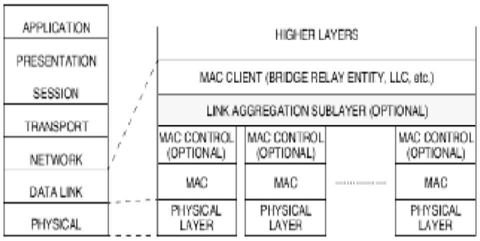

Fig. 2. OSI reference model

II. ETHERNET BURST TRANSPORT FEATURES

The basic idea of the proposed Ethernet Burst Transport (EBT) solution is to introduce new functionalities in Ethernet Switches, requiring slight standard modifications, with the aim of providing a low-cost carrier grade Ethernet transport. Differently from the current packet transport network solutions based on multi-layer switching technologies it allows, in a more scalable and flexible manner, to reduce transit traffic processing and to guarantee the support of QoS and bandwidth efficiency. It exploits Ethernet technology that is simple and low cost to improve network flexibility and facilitate end-to-end transport of the traffic. It therefore contributes to the evolution of Ethernet as a packet transport technology. According to the proposed solution, transmission of each node on the transport network is structured in bursts. A burst, shown in Fig. 1, consists of a variable number of Ethernet frames of the same flow separated by a proprietary IFG and preceded by an Ethernet Burst Control Frame (BCF).

A. The Burst Control Frame Structure and the Inter Frame Gap

The BCF carries information necessary for burst frames classification and forwarding such as its Ethernet Transport Node (ETN) MAC destination address, Class of Service (CoS), etc. That allows the transport node to determine if to drop the burst or to transmit it on the network by just inspecting the BCF, i.e. the following Ethernet frames in the burst will be opportunely queued without being processed. The BCF, shown in Fig 1, has the Ethernet frame structure with a specific value in the control type field that identifies it. The Ethernet data frames are left unmodified; they are just separated by a proprietary bit sequence in the IFG. That makes our proposal compliant with the Ethernet Standard. The adoption of the proposed transmission structure allows to avoid processing of transit traffic and consequently reducing the number of lookups in the forwarding table. Access rate to the forwarding table represents one of the most critical factors that impact Ethernet switches performance. In addiction reduction of frame processing improves scalability and allows to reduce power consumption and costs. The proprietary IFG allows frames of the same burst to be recognized even if the BCF is lost and helps the nodes handle classification and forwarding of the bursts frames, if the BCF is discarded, by inspecting only the first non-corrupted frame of the burst. It contains a specific pattern of idle bits similar to the one adopted in the IEEE 802.3z Gigabit Ethernet for the implementation of the Half Duplex frame bursting capability.

B. MAC control sublayer for burst transport support

A MAC control sublayer deputed to process the control frame need to be introduced in the IEEE 802.1D reference model as in figure 2. When the MAC layer recognize a control frame from its type field, the frame is passed to the MAC control sublayer without processing the frame. MAC control sublayer then will recognize in the OPCODE that the frame is a BCF and will process all the classifying information related to the burst. Then the MAC control will disable MAC layer classification of the incoming frames belonging to the same burst.

C. Dynamic Load Balancing for Efficient Bandwidth Utilization

Ethernet Link Aggregation, [8], allows to have multiple parallel link to increase the link speed beyond the limits of any one single interface. Conventional distribution algorithm, deputed to split flows among the bonded links are based on a static mapping of the incoming flows on a single sub-link of the aggregated one (usually on L3 hashes) in order to avoid frame reordering. This results in an inefficient utilization of the aggregated bandwidth and in a poor load balancing among the bonded links. Moreover the maximum granularity of a single flow remains limited to the sub-link capacity.

The EBT solution allows for a dynamic distribution of flows among different aggregated links by exploiting burst sequence number information carried by the BCF for burst reordering at each transit nodes. The adoption of a flow splitting mechanism allows for the support of flow granularity higher than the single sub-link capacity. This solution relies on the following traffic management mechanisms [9]: 1) a splitting flow mechanism that at the EBT source node controls flows distribution on the aggregated sub-links on the basis of the average traffic load; 2) a distributed request/grant scheduler, that ensures to respect of the burst ordering. This mechanism allows that different subsequent bursts could be at most transmitted simultaneously on different sub-links; 3) a dynamic output sub-link selection algorithm that burst by burst send the eligible burst on the least congested sub-link.

III. METRO NETWORKS ADOPTING ETHERNET BURST TRANSPORT

Metro networks, whose architecture is shown in Fig. 3, will more likely be characterized by edge devices able to aggregate ATM, TDM and mobile traffic by means of Ethernet technology. Such devices are typically interconnected with a feeder/hub node through the access section of the metro network consisting of an optical CWDM or DWDM transport network. Such a transport network is characterized by a ring physical topology with a maximum diameter of 100 Km whose nodes are co-located with the edge and the feeder devices and interconnected with them with 1 GE or 10GE links. Feeder devices are interconnected through the core section of the metro network, an optical DWDM transport network with multi-ring physical topology, with the edge routers of the core network. DWDM links carry different wavelengths with capacity of 2.5 Gbps or 10 Gbps.

Metro networks adopting the proposed solution does not require a separate optical transport network allowing to avoid 1GE and 10GE interfaces duplication on Ethernet switches network and consequently to reduce costs deriving from architectural scalability issues. In fact they consist of Ethernet transport nodes/switches interconnected through multiple Ethernet links that may be WDM multiplexed over one or more optical fibers, as shown in the Fig. 3. The use of WDM is justified by high capacity requirements of next generation transport network dictated by the need to support new high-capacity services such as HD-IPTV.

Each Ethernet transport node, as shown in Fig. 4, is characterized by line cards with input/output ports connected to the nodes local networks (lets call them client cards) and by line cards with input/output ports connected to other nodes on the ring (lets call them network cards). Ethernet frames arriving at the local cards are queued on the basis of their destination transport node and CoS and successively assembled in bursts. A burst is formed either when a timer expires or when the size of the assembly queue reaches a pre-specified threshold. In order to support different CoSs, different values for the timer of each class can be set as a function of the maximum frame delay accepted for each class. Local cards need to know/discover the association between the customer MAC addresses and the MAC addresses of the transport nodes to which the customer network is connected in order to opportunely queue the frames and build the BCF. At the input ring card, the Physical layer signals to the MAC layer the detection of Ethernet standard or proprietary IFGs allowing to recognize the frames of the same burst. A packet received after a standard IFG is always processed by the classifier; frames preceded by a proprietary IFG are instead queued without processing according to the information carried in the first packet of the relevant burst, i.e the BCF if received correctly. At intermediate nodes burst size can be dynamically adapted to the available bandwidth in order to limit frame delay and jitter. In other words the number of frames assembled in the bursts can be determined on the basis of the bandwidth granted by the scheduler to the corresponding flow/queue.

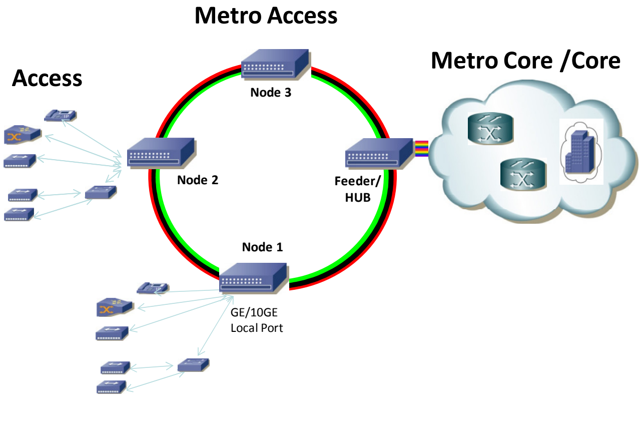

Fig. 3. Metro reference network scenario

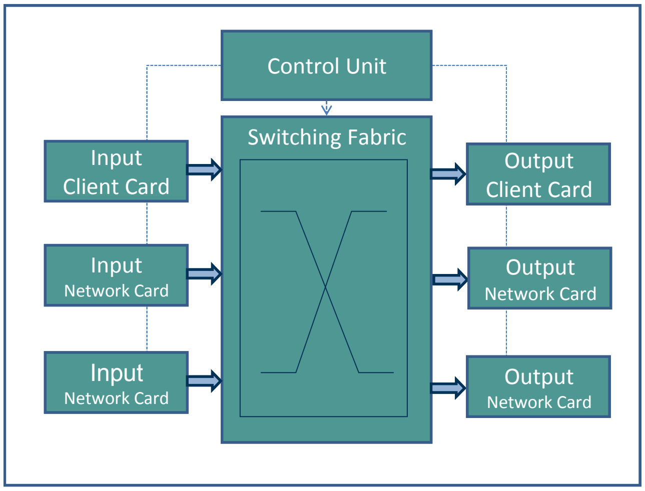

Fig. 4. Metro Transport Node Architecture.

IV. SIMULATION RESULT ANALYSIS

Simulation study has been carried out with the use of Opnet Modeler [10]. Performance of a metro network adopting the proposed Ethernet burst transport approach have been evaluated and compared with those of a standard Ethernet Switch (assumed ideal: line rate processing capabilities, no latency inside the switch) in terms of packet processing reduction, end to end delay and jitter. Jitter is evaluated as the maximum transfer delay variation experienced by consecutive frames of a given flow [11]. We simulated a metro access network with 4 nodes (two aggregated 10Gbps links, 20Km length). Clockwise unidirectional transmission has been considered for the sake of simulation model simplicity. Traffic demands have been generated through a self-similar traffic source model [12] with Hurst parameter equal to 0.952. It consists of multiplexed ON/OFF sources with heavy-tailed distributed ON and OFF periods. Mean length of 800 Bytes is assumed for Ethernet frames. We tested a Hub and Spoke traffic scenario and measured performance of the unidirectional flow from node 0

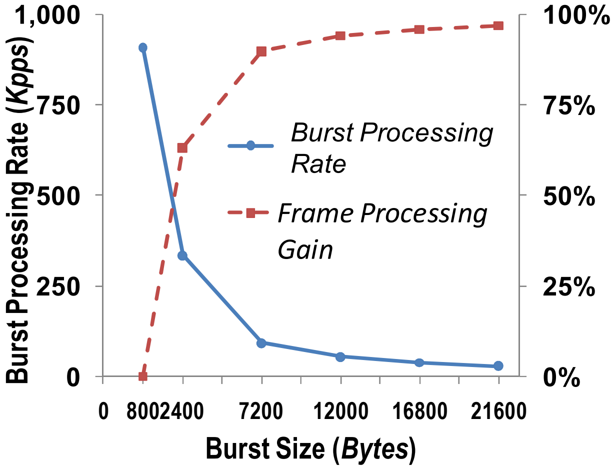

Fig. 5. Performance vs Burst Size: Frame Rate and Processing Reduction.

to node 3 assumed as target flow. Ethernet frames arriving at a client port are queued in different assembling queues, when the burst is ready it will be queued in an ”add” buffer associated to the corresponding output port according to a FIFO policy; transit bursts are queued in a ”transit queue”. Transit and Add traffic, competing for the same output port, are served with a round robin policy.

Fig. 5 shows that the adoption of burst aggregation allows to highly reduce the frame processing rate. On the left axis is reported the burst processing rate of the target flow (Kilo packets per second - Kpps), while on the right axis is reported the processing gain, both at the varying of the burst size. The first value reported shows the performance of the conventional Ethernet switch used as benchmark. We can see that when no burst mode is adopted more than 900 Kpps need to be processed. At the increase of the burst size the burst rate decreases and also for an average burst size of 7200 Bytes results a processing saving of 90%.
Figures 6 and 7 report the end to end delay and the jitter experienced by the frames of the target flow when the dynamic load-balancing is enabled and compare it with the static case. The static mapping of traffic demand on the two 10GE results in a average traffic loads respectively of 2.3 Gbps and 9 Gbps. In the dynamic mapping case results the average link load is fairly balanced at 5.5 Gbps. It is straightforward to see how EBT with dynamic load balancing leads to high processing gain without incurring in delay penalties. Jitter performance worse with burst size increasing but result acceptable. Since even for high burst size it's lower than $2ms$ required by the MEF.

EBT allows to reach very high processing saving at the cost of slightly worse transfer delay than conventional Ethernet switches; this delay is compensated by a more efficient bandwidth handling. In fact Dynamic Load Balancing on burst basis allows to split traffic with finer granularity fixing the issues of Ethernet Link Aggregation Mechanisms.

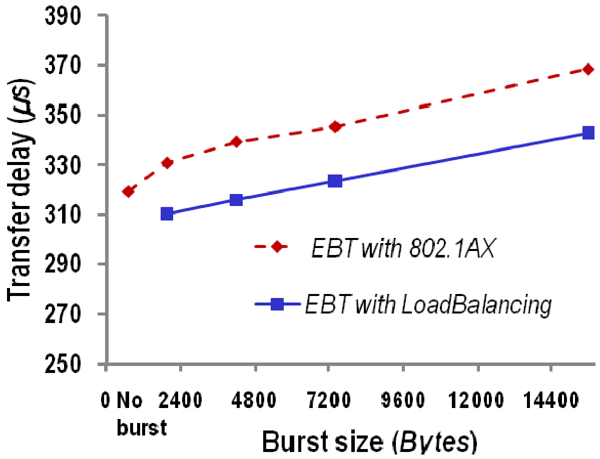

Fig. 6. Delay vs Burst Size

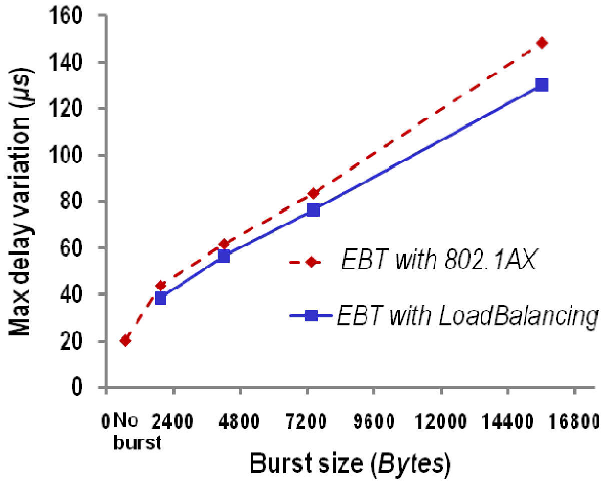

Fig. 7. Jitter vs Burst Size.

V. Conclusion

We propose an Ethernet burst transport solution that maintains the flexibility and any to any connectivity deriving from the connectionless nature of Ethernet and at the same time provides Ethernet technology with efficient aggregation capabilities. Frames of the same flows are aggregated in Ethernet compliant bursts that allow to reduce processing of transit traffic improving the switch scalability. Simulation results show that the impact of the burst aggregation on jitter and delay is largely compensated by the more efficient bandwidth utilization provided by the dynamic management and load balancing on aggregated links.

Acknowledgment

The work described in this paper was carried out with the support of the STRONGEST-project funded by the EC through the 7th ICT-FP.

References

[1] S.Perrin,"The Optical Switching Revival: Rebuilding Optical Networks for Packets", Heavy Reading VOL.7, N.3, MARCH '09;

[2] T.D Nadeau, V. Sharma, A. Gumaste, "Next-generation carrier ethernet transport technologies", IEEE Communications Magazine, Volume 46, Issue 3;

[3] D. Allan, N. Bragg, A. McGuire, A. Reid, "Ethernet as carrier transport infrastructure", IEEE Communications Magazine, Volume 44 , Issue 2;

[4] M. Beghdadi, Y. Cao, "Evolving metro transport and switching infrastructures: path to efficient Ethernet services for carrier networks", Design of Reliable Communication Networks, 2005;

[5] S. Hauger, T. Wild et al.," Packet Processing at 100 Gbps and Beyond - Challenges and Perspectives" in Proc. of ITG Photonic Networks, Leipzig, May 2009, pp. 43-52

[6] F. Feller,J. Scharf,"Increasing Packet Sizes to Mitigate Performance Issues in High-Speed Packet Networks",in Proc. of ITG Photonic Networks, Leipzig, May 2009, pp. 223-230

[7] P. Testa et al.,"Ethernet Burst Transport: a Scalable Solution for Optical Metro Networks", Proc. of ECOC '10;

[8] IEEE Std 802.1AX-2008 IEEE Standard for Local and Metropolitan Area Networks Link Aggregation;

[9] P.Testa, A.Germoni, M.Listanti,"Switching node with load balancing of burst of packets", Filed Ericsson Patent, 23 July 2010.

[10] http://www.opnet.com/solutions/network_rd/modeler.html

[11] MEF Implementation Agreement, MEF 23, Carrier Ethernet Class of Service Phase 1 10 June 2009;

[12] Lu Ying, Kang Feng-ju, Zhong Lian-jiong, Wang Zhi-Guang , "A self-similar traffic generation method and application in the simulation of mobile ad-hoc network," Computing, Communication, Control, and Management, 2009. CCCM 2009. ISECS International Colloquium on , vol.4, no., pp.229-233, 8-9 Aug. 2009