Performance evaluation of a distributed credit-based fairness mechanism for slotted WDM rings

CH. LINARDAKIS, H. C. LELIGOU, J. D. ANGELOPOULOS
Department of Electronic and Electrical Engineering
National Technical University of Athens
Heroon Polytechneiou 9, 15773 Zografou, Athens
GREECE

Abstract: - In slotted WDM metro rings employing spatial re-use the problem of fairness is inherent and quite significant. To allow all nodes around the ring to equally share the available bandwidth a novel mechanism based on access credits and executed in a distributed way is proposed. Its performance is evaluated by simulations to show its effectiveness even under highly asymmetric loading and is compared to the well-known token-based SAT mechanism.

Key-Words: - MAC, spatial re-use, fairness, MAN, WDM.

1 Introduction

WDM rings of metropolitan dimensions are increasingly deployed to collect traffic from access systems. The emerging new WDM components in conjunction with the accelerating growth of IP-based traffic has fuelled intensive research effort on systems that can handle optical payload switching even if the control signalling remains in the electrical domain [1], [2]. For medium-sized networks this approach can give particularly promising results in the context of the ring topology where the control is exercised by a MAC protocol.

The European IST DAVID project on which this work is based, has developed [3] in one of its tasks, a slotted WDM packet-mode ring featuring fully dynamic traffic control [4] while employing the same format for all kinds of encapsulated traffic to allow for easy burst operation and optical switching. Variable packets are accommodated by use of a train of slots not necessarily concatenated. Up to 32 wavelengths running at 10 Gbps can be available on each ring with a slot size of 10000 bits (1μs). The slots on all wavelengths are synchronised, therefore creating simultaneous slots in all wavelengths (multi-slots) [5].

The ring is a shared medium requiring a Medium Access Control (MAC) protocol [5-8], to arbitrate access to its slots regulating both the time and wavelength dimensions. By devoting one wavelength exclusively to MAC control information, it is possible to base all access decision on the contents of this channel, which is processed in the electrical domain in all transit nodes. In contrast, all other data remain in the optical domain all the way around the ring and are not buffered, re-formatted or processed, except at the edge routers. The control information indicating the destination node address, the status bit (slot occupied or not), priority, fairness control bits, etc., which is contained in the control channel is organised in a way establishing a one-to-one correspondence with the available data slots.

Access to slots is based on the empty slot protocol: nodes monitor the control channel to find the slots destined for them so they can receive the payload from the corresponding wavelength leaving the slot empty (thus enabling slot reuse). Nodes cannot alter traffic in transit; they can only seek an empty slot to place their data by checking the control channel [5].

An indispensable part of the control of such a system is a distributed fairness mechanism in order to ensure that all nodes get a fair share of the total available system bandwidth [6]. Fairness issues arise in any shared-medium system [8], but it is particularly important in the ring topology with spatial re-use as is the system under consideration. The spatial re-use, which allows for a doubling of the effective transport capacity of the ring in the case where the packets destinations follow a uniform distribution, aggravates the inherent in all rings unfairness [8-10] since the downstream neighbour nodes of a destination node receiving a lot of traffic are strongly favoured finding many empty slots compared to all other nodes. The action of the closed-loop controls embedded in the TCP protocol further aggravates the fairness problem of the ring. Although these mechanisms have been designed to allow flows sharing a bottleneck to converge towards a fair share based on the max-min criterion [11], this is only true in centralized multiplexers when all TCP flows go through the same buffer and suffer similar loss probabilities. In the case of a distributed multiplexer such as the WDM ring, where flows do not share the same buffer space (e.g. 10 flows may go through the buffer of one node while in another node only one flow may be present at a particular time) any bandwidth unbalance will go out of control.

Connections that first suffer losses will further reduce their rates at the TCP source, leaving those with already better access advantage at a further improved state. The target is to bring the bandwidth enjoyed by each node to the ratios agreed by Service Level Agreements (SLAs) during service provisioning. Given the latest developments for enhancing IP services, a more general definition on fairness than the max-min criterion of traditional best-effort networks is required. To give an example, when one node provides access to the gateway of a big customer (e.g. University or Corporation) with a 155Mbps interface and a service agreement for a minimum guaranteed bandwidth of 100Mbps and in another node there is a cluster of residential and SME customers subscribing to an ISP with a 34Mbps I/F and a service level agreement guaranteeing 10Mbps, the notion of fairness must be enhanced to allow for the much higher tariff paid by the first customer. Thus, the notion of weighted proportional fairness is adopted for this system. More on this extension of fairness can be found in [11].

In the case that fewer receivers than wavelengths are used in the system, the problem has been extensively studied in [6] using a fairness mechanism based on an extension of [9,11] and will not be considered here. For the needs of this paper, each node is assumed to be equipped with a set of fixed transmitters and receivers equal to the number of wavelengths.

To overcome the unfairness problem and allocate the available bandwidth in a proportional way, a distributed mechanism has been designed and presented in [12]. In this paper, first the mechanism is briefly presented for completeness reasons, and then simulations results proving the proportional allocation of bandwidth and allowing for its performance assessment are presented. Also, its performance is compared to the performance of the SAT algorithm presented in [6], to illustrate how suitable for the proliferating IP traffic the presented algorithm is. Conclusions are drawn in section 4.

2 The Credit-based fairness algorithm

The mechanism proposed below enforces the weighted proportional fairness on the traffic of the ring and provides a tool for the operator to apply the suitable weights according to its provisioning policies. It keeps a log of the number of packets sent by each node making possible to choke those going ahead and reduce the difference. The action is of the ON-OFF (or bang-bang) type for simplicity of implementation. Thus over the long term the number of packets sent is made to comply with the weighted proportions (and any small difference is only temporary until compensation is exercised).

The scheme uses a 24 bit credit counter (CC) at each node, which holds the number of credits allocated to the node (an equal number of packets can be transmitted). The credits are generated according to rates allocated to the node at service provisioning time. A full wavelength channel corresponds to one credit per slot. However, no credits are generated above the number of actual packets (expressed in slots) that are queued in the node (i.e. a “use it or lose it” policy is followed).

To prevent overflow, when a node’s CC reaches a high credit threshold (HCT), it sets the STOP bit in the control channel which travels around the ring signalling to all nodes to stop credit generation. This is equivalent to an equal decrement of bandwidth allocation (which would have not been satisfied anyway) for all nodes. Nodes with a value above zero continue to send until their credit counter drop to a value a bit above zero called Blocking Credit Threshold (BCT). At the moment the credit generation stops, it is obvious that nodes that were not favoured by asymmetries will have a high value of CC (particularly the one which initiated the stop signal) while the favoured ones a lower value. So the former will sooner or later be forced one after another to stop transmitting. This will give the opportunity to those lagging behind to catch up since the number of empty slots circulating will increase. Utilization will not suffer much, since this occurs when at least some nodes have loaded buffers.

Once the node that blocked the credit generation is relieved (reaches a low credit threshold (LCT)), it sets a START bit in the control channel signalling the beginning of credit generation again. The stop and start messages run in the same direction for a full ring rotation thus making sure equal loss of credits for all nodes. If more than one node initiates the stopping, still a full circle will be covered both in the stopping and the re-starting process so no nodes will be handicapped.

It is obvious that as traffic fluctuates, the total offered load at times exceeds the ring capacity resulting in losses. The TCP congestion avoidance mechanisms jump into action and adjust transmission rate at the detection of losses hence placing more emphasis on the evolution of queue size of the nodes. By not allowing nodes to send above their credit limit, the fairness enforcement mechanism restores the effectiveness of TCP controls over the distributed multiplexing of the ring. Both mechanisms residing in different layers work in concert to establish fair bandwidth share over the concatenated links including the ring.

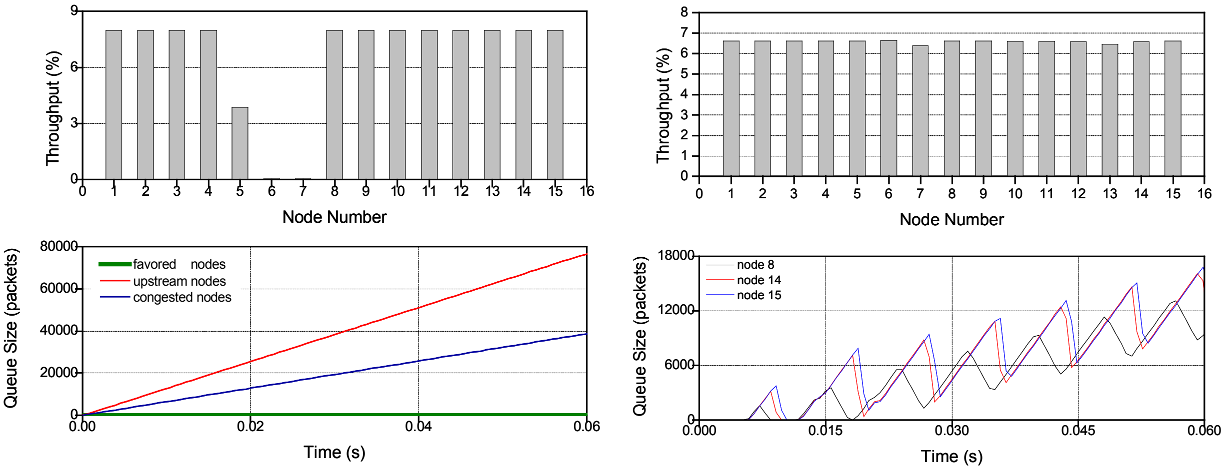

Figure 1: The nodes throughput under 120% total offered load without any fairness mechanism enforced (left side) and with the proposed mechanism enforced at right.

To allow for better performance assessment, the simulation results will be compared to the case where the SAT algorithm is employed. The SAT (Satisfied) is a one-bit flag travelling in the control channel. Each node is allowed to transmit K packets between two consecutive SAT receptions. This implies that when a node receives the SAT, it will not release the SAT until it has transmitted K packets or its queue has become empty. (K is measured in packets and can differ from node to node.)

It is stressed that all parameters used by both algorithms are handled by the system operator and provide the tool to enforce the contracted SLAs. The implementation cost of both algorithms is very low, since few registers are only required, assuming that the data manager of the node will announce the arrival and the transmission of packets.

3 Performance evaluation

To study the performance of the fairness algorithm under realistic scenarios, given the lack of analytical tools due to the high system complexity, a series of computer simulations have been carried out. Fifteen (15) nodes were modelled, with 30μs time distance between each node pair as would be the case for a metro ring of 90km, with evenly spaced nodes. Hence, the propagation time for a packet transmitted from node j to node j+1 is equal to 30μs while from node 1 to node 15 is equal to 420μs. In each of the 16 data wavelengths, data were transmitted at 10Gbps.

In the first scenario, which will be used as a benchmark to evaluate more realistic situations, no fairness mechanism is enforced. Constant Bit Rate sources inject traffic loading the ring at 120% of its capacity (160Gbps). All packets are destined to node 8, hence cancelling the benefits of slots re-use. (In this case, the maximum ring throughput is equal to the ring capacity.) Node 7 is expected to find almost all slots occupied and to experience the highest delay of all nodes, suffering the unfairness.

In Figure 1, the throughput that each node experiences is shown at the top of the figure while at the bottom the evolution of the queue size is depicted. The results when the proposed mechanism is (is not) enforced are shown at the right (left) hand side. Although the total throughput remains close to 100% in both cases, nodes sitting before node 8 are distressed, as shown at left when no fairness mechanism is employed, while node 8 and 9 find all slots empty and hence are highly favored. This is also proved by the evolution of queue size: favored nodes have constantly empty queues while the queue size of the congested nodes increases. The more dramatic increase is measured for nodes close to node 8 (denoted as upstream nodes in the figure). The queue size increase leads to packet loss, which will be detected by the TCP closed loop flow control and force the source to decrease its packet generation rate, further aggravating the problem. The proposed mechanism rectifies the unfairness in access: as shown at the right side of the same figure, all nodes enjoy the same throughput and their queues exhibit similar behaviour. Using the SAT fairness mechanism, unfairness is also rectified, achieving the same total throughput, although small differences between nodes exist.

It is worth stressing that adjusting the threshold the differences in queue sizes between congested and favored nodes can be eliminated. Decreasing HCT, queue size variation decreases as well, at the expense of total system throughput. In general, in the highly bursty and asymmetric telecommunications real life, efficiency improves when longer-term control actions are allowed.

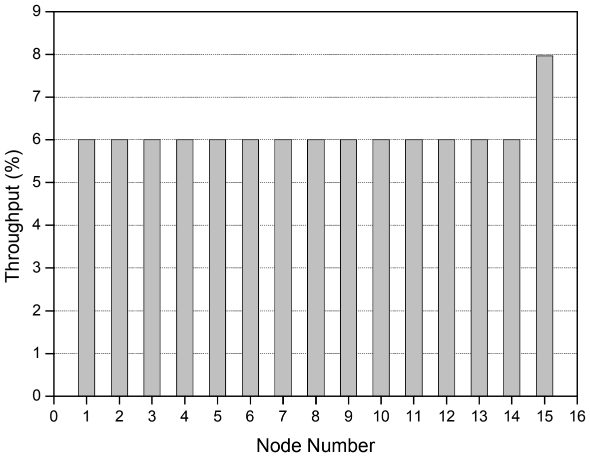

Figure 2: Nodes throughput for scenario 2, where node 15 is assigned higher access rights

In the 2nd scenario, the target is to show how the proposed mechanism can proportionally allocate the available bandwidth. In more detail, the total offered load is the same (120%) with each node injecting the same amount of traffic but node 15 is allocated double access rights than every other node (the credit generation period is the same for all nodes but two credits are generated by node 15 while one for every other node). Hence, node 15 generates traffic equal to 8% of the ring capacity and its access rights are equal to 12.5% (2/16), while any other node generates the same amount of traffic and has rights to access just the 6,25% of the ring capacity.

As shown in Figure 2, the access rights cover all the generated traffic for node 15, which hence achieves a throughput of 8% while all other nodes equally share the left-over bandwidth, enjoying throughput equal to 6% (2% below their source rate). This happens because the source rate for node 15 is well below its access rights which is not the case for all other nodes. It is worth stressing that should the source of node 15 generate traffic at higher rate than the node's access rights, it would also experience congestion.

In the last scenario, sources of ON-OFF type have been used. The ON and OFF periods were exponentially distributed as well as the packet inter-arrival timers during the ON period. The ratio between mean ON and mean OFF values was equal to 6. All packets were destined to node 8, exactly as in the previous scenarios. As regards the fairness algorithm parameters, these were chosen as follows: HCT=7200, LCT=1500, BCT=0, for all nodes. The results will be compared to the results of SAT algorithm for K=7200.

In this scenario, all nodes enjoy the same throughput (not shown here due to lack of space), using either the proposed credit-based or the SAT mechanism. Figure 3 shows the evolution in queue size.

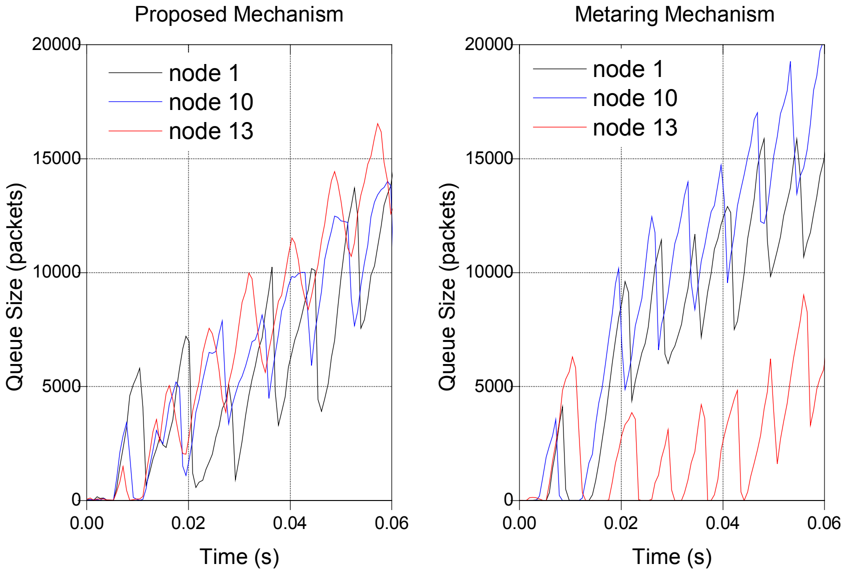

Figure 3: The evolution of queue size for three nodes in the last scenario (VBR sources)

Using the SAT mechanism, the queue of node 13 remains at significant lower levels of occupancies than the other two nodes, while using the proposed mechanism uniformity is more evident. Thus, the TCP flow control mechanisms will impose similar effects to all nodes.

4 Conclusions

Fairness is an inherent problem in ring topologies, which is further aggravated by the congestion avoidance and flow control mechanisms of the widely used TCP protocol. In unidirectional rings with spatial re-use, nodes receiving high percentage of traffic are favored as well as their downstream neighbours, leaving other nodes to suffer the congestion. Assuming that metro nodes may connect either big organizations or highly popular servers or just a group of residential customers, the problem becomes dominant. A distributed mechanism, which allows the allocation of the available bandwidth in a weighted way, proportionally to the SLAs, has been proposed. Computer simulations show that not only the bandwidth is allocated according to the assigned weights, but also the evolution of the queue size of the nodes is very similar, causing the TCP control mechanisms to act in synergy.

Acknowledgements: The work presented in this paper has been partially funded by EU IST 1999-11742 “DAVID”.

References

[1] I. White, M. Rogge, K. Shrinkhande, L. Kazovsky, “A summary of the HORNET project: A Next-Generation Metropolitan Area Network” *IEEE Journal on Selected Areas in Communications*, Vol. 21, No 9, Movermber 2003, pp. 1478-1494.

[2] L. Dittmann, C. Develder, D. Chiaroni, F. Neri, F. Callegati, W. Koerber, A. Stavdas, M. Renaud, A. Rafel, J. Solé-Pareta, W. Cerroni, N. Leligou, Lars Dembeck, B. Mortensen, M. Pickavet, N. Le Sauze, M. Mahony, B. Berde, G. Eilenberger, *“*The European IST Project DAVID: a Viable Approach towards Optical Packet Switching*”*, *IEEE Journal on Selected Areas in Communications*, Vol. 21, No 7, September 2003.

[3] Brian Bach Mortensen, Michael S. Berger, “Optical Packet Switching Demonstrato*r*”, ECOC 2002 Conference, Copenhagen, Denmark, September 2002.

[4] Ch. Linardakis, H-C. Leligou, A. Salis, J.D. Angelopoulos, *“*Control of slotted traffic among interconnected WDM rings by explicit reservation”, Photonics in Switching 03, September 03, Versailles, France.

[5] Helen-Catherine Leligou, John D. Angelopoulos, Charalambos Linardakis and Alexandros Stavdas, “A MAC Protocol for Efficient Multiplexing QoS-Sensitive and Best-Effort Traffic in Dynamically Configurable WDM Rings”, *Computer Networks, Elsevier*, Volume 44, Issue 3 , 20 February 2004, pp. 305-317.

[6] M.A. Marsan, A. Bianco, E. Leonardi, M. Meo, F. Neri “MAC Protocols and Fairness Control in WDM Multirings with tunable Transmitters and Fixed Receivers”, *J. of Lightwave Communications,* Vol. 14, No. 6, pp 58-66, June 1996.

[7] M.A. Marsan, A. Bianco, E. Leonardi, A. Morabito, F. Neri “All optical WDM Multi-rings with Differentiated QoS”, *IEEE Comm. Mag*. Feb. 1999, Vol. 37, No. 2, pp 58-66.

[8] J. D. Angelopoulos, N. I. Lepidas, E. K. Fragoulopoulos, I.S. Venieris,“TDMA multiplexing of ATM cells in a residential access SuperPON”, *IEEE Journal on Selected Areas in Communications*, Special issue on high capacity optical transport networks, Vol.16, No. 7, September, 1998

[9] I. Cidon & Y. Ofek “Metaring-A Full Duplex Ring with Fairness and Spatial Reus*e*”, *IEEE Transactions on Communications,* Jan 1993, Vol. 41, No. 1, pp 110-119.

[10] M.J. Marsan et al, “Slot Reuse in MAC Protocols for MAN*s*”, *IEEE J. on Selected Areas in Communications*. Vol. 11, No. 8, October 1993.

[11] P. Gevros, Jon Crowcroft, P. Kirstein, Saleem Bhatti, “Congestion Control Mechanisms and the Best Effort Service Model”, *IEEE Network Magazine*, May/June 2001, Vol 15, No. 3, pp 18-26.

[12] J.D. Angelopoulos, N. Leligou, H. Linardakis, A.Stavdas, “A distributed fairness mechanism for slotted WDM ring*s*”, in “*Advances in multimedia, video and signal processing systems*”, N. Mastorakis, V. Kluev ed., pp. 267-273, WSEAS press, ISBN: 960-8052-72-6 Athens 2002.