Extension of the Flow-Aware Networking (FAN) architecture to the IP over WDM environment

V. López*, C. Cárdenas†, J. A. Hernández*, J. Aracil*, M. Gagnaire†
* Universidad Autónoma de Madrid, Spain
Email:{Victor.Lopez,Jose.Hernandez,Javier.Aracil}@uam.es
† École Nationale Supérieure des Télécommunications , Paris, France
Email:{Cesar.Cardenas,Maurice.Gagnaire}@enst.fr

***Abstract*—Backbone networks are migrating to IP over WDM architectures. In such multi-layer network configurations, it is necessary to combine efficiently the resources of both layers in order to provide enhanced Quality of Service (QoS) to the end-users. In the context of existing IP networks, Flow-Aware Networking (FAN) has been proposed in order to provide QoS guarantees to multiplexed IP flows within an IP router. FAN is based on implicit admission control and per-flow scheduling. In this paper, we propose a new node architecture that extends the FAN concept to IP over WDM overlayed networks in which both optical and electronic resources are available. Three different policies are introduced to decide on which criteria an IP flow arriving at a node must be bifurcated from the standard FAN architecture to be forwarded onto a transparent lightpath up to its destination. The performance of the three proposed policies are discussed in terms of goodput and of queueing delay.**

***Index Terms*—Quality of Service; FAN; Multi-Layer FAN; Traffic engineering; IP over WDM.**

I. Introduction

The provisioning of Quality of Service (QoS) to applications implemented at the end-nodes is one of the key issues in the engineering of the Next-Generation Internet. Besides network resource overprovisioning, two main approaches for QoS support in IP networks have been proposed in the literature: IntServ and DiffServ. The former is well-known for its lack of scalability due to the soft state of the virtual circuits on which rely this technique [1]. The latter requires a signaling channel and a control plane based on complex algorithms to address packet marking and metering. Such a mechanisms lead to an expensive approach for QoS provisioning. In this context, a new approach, called Flow-Aware Networking (FAN [2], [3], [4], [5]) has been proposed as a promising technology to manage congestion control in IP networks and to provide QoS to applications.

Essentially, FAN operates at packet level and implicitly distinguishes between two types of flows: streaming (or priority flows), and elastic (or non-priority flows). Streaming flows typically refer to voice or video applications (UDP), whereas elastic flows typically refer to TCP/IP sessions. The FAN architecture is designed in order to meet two main objectives: (1) minimize the queueing delay suffered by streaming flows in the routers; and (2) intend to provide a minimum fair rate to the elastic flows. When FAN cannot satisfy these minimum requirements, it rejects incoming flows. By using such an admission control policy, FAN keeps into service the already admitted flows, thus assuring a minimal QoS under overloaded conditions.

In a first approach, FAN was conceived to operate at the IP level without any information about the underlying layers. With the current technology trends, network operators are gradually migrating to an IP over WDM paradigm, mainly to benefit of a larger transmission capacity on each optical fiber. Unlike current protocol stacks such as IP/ATM/SDH, the IP and WDM layers remain relatively independent. This is the reason why we propose an adaptation of the FAN concept to IP/WDM architectures, and by extension to multi-layer capable routers including optical and electronic switching [6].

We define the concept of Multi-layer FAN (MFAN) as a router architecture enabling to handle optical resources provided by WDM technology

The authors would like to thank the support from the European Union VI Framework Programme e-Photon/ONe+ Network of Excellence (FP6-IST-027497). They are also thankful to Abdesselem Kortebi from France Telecom R&D for his support in the FAN implementation and to Mrs Sara Oueslati from France Telecom R&D for our fruitful discussions.

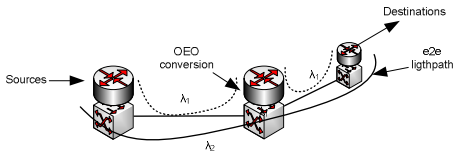

Fig. 1. End-to-end transparent lightpaths and hop by hop opaque lightpaths.

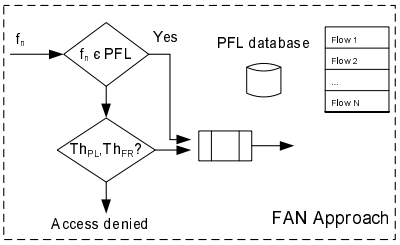

Fig. 2. FAN Admission Control Flow Diagrams.

within the FAN architecture. When a new data flow arrives at a MFAN node, this node first tries to deal with the packets of this flow at the IP layer. If the IP layer is already busy, the flow is transferred to the optical layer whereas this layer benefits of available resources. Figure 1 illustrates the architecture of the considered IP/WDM network. FAN operates at the IP layer, IP packets being routed and forwarded via hop by hop lightpaths from source to destination. If an incoming flow is bifurcated from the IP layer to the optical layer, one assumes that one or several transparent lightpaths are pre-established between the source and the destination. The optical switching technology used at the WDM Layer is based on Wavelength Selective Switches (WSS) themselves on Reconfigurable Add-Drop Multiplexors (ROADM). Data flows to be transmitted at the optical layer are buffered in a simple FIFO queue at the source MFAN node. The transmission technique at the optical layer is based on optical bursts with random size. By simplification one assumes that an optical burst corresponds to an IP packet. As it is depicted in Figure 1, an intermediate node between the source node and the destination node may insert along the lightpath between this pair of nodes its own optical bursts. Similarly, an intermediate node may extract upstream optical bursts for which it is the destination. Such a transparent optical circuit linking a source to multiple destinations is known as a light-trail [7], [8]. Traffic inserted at an intermediate node along a light-trail may be viewed as a cross-traffic for the optical layer. We can then estimate that MFAN nodes are located at the ingress and egress nodes of a transparent optical cloud.

The remainder of this paper is organized as follows. In Section II, we introduce the concept of FAN. In Section III, we describe the MFAN architecture. A set of three policies is proposed to manage the transfer of an IP flow from the electrical layer to the optical layer. In Section IV, we outline the benefits and drawbacks of MFAN from simulation. Finally, section V concludes this work and proposes a few perspectives.

II. Flow Aware Networking

Flow-Aware Networking (FAN) was proposed in [3] as a new approach to offer Quality of Service to the Next-Generation Internet, based on implicit classification and admission control of incoming flows, and flow-based scheduling.

Essentially, FAN performs such implicit classification of flows into either streaming (high-priority) or elastic (low-priority), and defines an admission control mechanism which seeks two objectives. On the one hand, it gives preference to streaming flows on attempts to minimise the delay they experience but, at the same time, it aims at assuring a minimum useful data rate (also known as goodput) to elastic flows.

To this end, FAN defines two parameters: The Priority Load (PL) and Fair Rate (FR). Fair Rate is an estimation of the bandwidth that an incoming flow would receive if admitted, while Priority Load is an estimation of the service rate of streaming packets in the queue.

Incoming flows are denied access to the system, when the FAN architecture can not guarantee a given performance level (delay and fair rate). This admission control mechanism is depicted in fig. 2.

The complete process is as follows: When a packet arrives at the system, the admission control finds the flow it belongs to, namely f_n, and evaluates whether such f_n is in its inner Protected Flow List (PFL). This list stores the ids of each flow already accepted by FAN and transmitted over the IP layer. If $f_n \in$ PFL, then the packet is served. Otherwise, the packet is part of a new flow which must pass through the FAN admission control process. When so, it is tested whether $PL < Th_{PL}$ and $FR > Th_{FR}$, that is, if a given QoS guarantees defined by the Th_{PL} and Th_{FR} thresholds are maintained or not. If this is the case, the new flow is accepted; otherwise, it is rejected.

Although flows already accepted are somehow protected, only those flows which transmit at a lower rate than Th_{FR} are treated as streaming flows (high-priority). All the others are considered

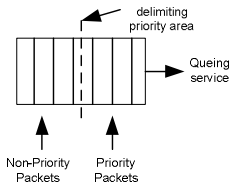

Fig. 3. Priority Fair Queueing architecture.

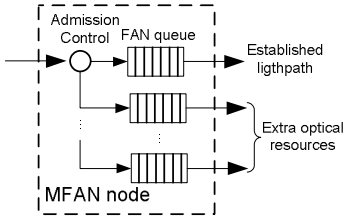

Fig. 4. Multilayer Flow-Aware Networking (MFAN) node architecture.

as elastic flows and receive less preference. This is done in order to avoid flows which abuse from the system resources.

Finally, a Priority Fair Queueing (PFQ) algorithm, as defined in [4] (which is based on the Start Fair Queueing algorithm [9]), is used to give preference to streaming flows over elastic flows.

Basically, PFQ is a PIFO (Push In First Out) queue, which stores packet information (flow identifier, size and memory location) and timestamp, the latter determined by the SFQ algorithm. The PFQ queue is split into two areas delimited by a priority pointer (see fig. 3), whereby streaming flows are temporally stored at the priority queue area (at the head of the queue), and the elastic flows are stored at the tail of the queue. Preference is given to the priority area since it is served before the non-priority area. Finally, the queue stores elastic and streaming packet count statistics, which are further used to compute the values of PL and FR.

In addition, an Active Flow List (AFL) is maintained by the PFQ. This list is similar to the PFL defined above, but it also saves the amount of packets transmitted per flow in the recent past. The flows with the greatest amount of transmitted packets (also known as greatest "backlog") may be discarded under severe congestion conditions. It has been shown in [4], [10] that the AFLs does not suffer from scalability problems.

III. Extension of Flow-Aware Networking to an IP over WDM network architecture

Figure 4 proposes an architecture for the Multilayer Flow-Aware Networking (MFAN) node. As it is illustrated, an MFAN node contains an optical layer and its subsequent queue, which can be used for routing traffic when the IP layer is near congestion. Let us remind that the electrical traffic aggregation of multiple flows in this queue is carried out on the basis of an optical burst mode transmission and an end-to-end light-trail. In other terms, the aim of the MFAN architecture enables to accept new flows at the transparent optical layer which would otherwise be rejected at the opaque IP layer.

Essentially, it is assumed that the QoS provided by FAN at the IP level is sufficient. For this reason, the extended Multilayer FAN node uses the IP resources whenever it is possible (that is, $PL < Th_{PL}$ and $FR > Th_{FR}$), and requests extra optical resources once either PL or FR falls out of their ranges. Therefore, the MFAN solution does not try to improve the QoS provided to incoming flows, but to increase the network resource utilisation making an efficient use of the optical resources.

In the MFAN architecture, the flows that use the optical resources are stored in the PFLλ list. This list is looked up when incoming packets arrive at the MFAN node, to see whether such packet belongs to an already accepted flow or is part of new flow arrival (see fig. 5). If it is a new flow, it is first tried to be routed over the electronic layer (whether $PL < Th_{PL}$ and $FR > Th_{FR}$). If the flow is denied access to the IP layer, it then tries the optical layer, first checking whether there is any free wavelength, and secondly evaluating the optical queue threshold (OQ_{th}). If it is successful, then the flow is accepted. The following defines three different policies about what to do when the optical layer accepts a flow in a situation of congested electronic layer.

Newest-flow policy. Those incoming flows, which cannot be accepted by the FAN queue, are sent over the optical layer, only if the occupancy of the optical queue is below a given threshold OQ_{th}. Admission control is used also in the other two policies.

Most-Active-flow policy. The flow with the greatest "backlog" in the AFL (existing flow) is bifurcated to the optical layer, thus releasing some space in the electrical layer for new incoming flows. As FAN provides information about the implicit classification, no streaming flows are sent through the optical queue. The reason is that the

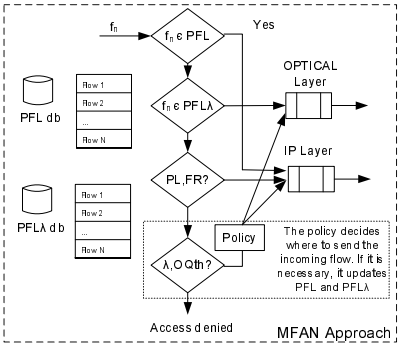

Fig. 5. MFAN Admission Control Flow Diagram.

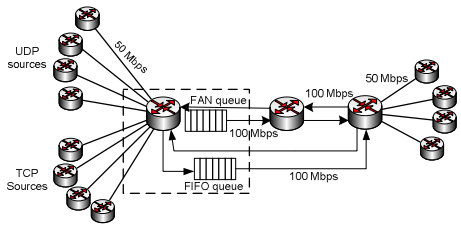

Fig. 6. MFAN Simulation Scenario.

optical queue can not give them priority (let us recall that the optical queue is a simple FIFO queue).

Oldest-flow policy. With this policy, the flows that have been around in the FAN queue are moved to the optical queue, thus making space for incoming ones. The age of flows is available by FAN in the PFL, thanks to the incoming order to the system. Like in the Most-Active-flow Policy, only the elastic flows are transmitted over the optical queue.

IV. Experiments and results

A. Simulation scenario description

To study the performance of the three different policies introduced above, we have simulated a scenario with four TCP/Reno and four UDP traffic sources in a two-hop network (see Figure 6) using `ns2`[1]. As illustrated, the light-trail (end-to-end transparent optical link) has been simulated by a direct connection between the first and the third nodes, whereas the IP connection traverses all three nodes. By sake of simplicity, one assumes that a single light-trail is available at the optical layer.

This scenario considers the same input traffic profile used in [4], [5] to validate FAN. Essentially, flows arrive following a stationary Poisson process, given the fact that the UDP sources (streaming flows) simulate phone calls and TCP (elastic) flow arrivals are well-known to follow this distribution (see [11]). The UDP sources are characterised by a mean rate of 64 Kbps with on/off periods of 0.5 s, and exponentially distributed duration with mean 1 min. On the other hand, the TCP job size follows a truncated Pareto distribution with tail index 1.5 and mean 1 KByte, always in the range of 8KBytes-1MBytes.

According to [5], we have considered 80% of the total traffic volume is TCP and the remaining 20% is UDP. The buffer sizes considered follow the well-known rule of $Q = \overline{RTT} \times C$, as given by [12]. The traffic load at the system was considered 110%, in order to study the admission control mechanism. Th_{PL} is set to 80% and Th_{FR} to 10%, [5].

With this configuration, we have focused on the following performance metrics: rejection ratio or percentage of incoming flows rejected at system, mean delay of the streaming packets and average goodput of the elastic flows.

Finally, the reader should note that the backbone link capacity is 100 Mbits/sec, which is much smaller than typical optical capacities, but significantly reduces the simulation time. The results obtained with this value should remain for higher capacities.

B. Implicit classification of FAN

As previously stated, FAN's implicit classification decides which flows are considered streaming (high-priority) and which others are elastic (low-priority). Following FAN's architecture, a situation with Fair Rate under a threshold indicates that the system is congested due to the elastic flows, whereas if the Priority Load threshold is exceeded the flows causing congestion are the streaming ones.

For instance, fig. 7 shows the evolution of Fair Rate and Priority Load in the scenario described in section IV-A. As shown, Fair Rate is out of its nominal range, which means that the system is heavily loaded due to the elastic traffic. This is reasonable since 80% of the simulated traffic is TCP. Thus, it makes sense to move the elastic flows to the optical queue, in order to relief the IP layer.

Such implicit classification information can be used by the Most-Active-flow and Oldest-flow policies to move to the optical layer the "most apropriate" flow in terms of congestion at the IP layer. Clearly, the Newest-flow policy makes no use of such information, since it just switches the incoming flow to the optical layer.

[1] http://www.isi.edu/nsnam/ns/

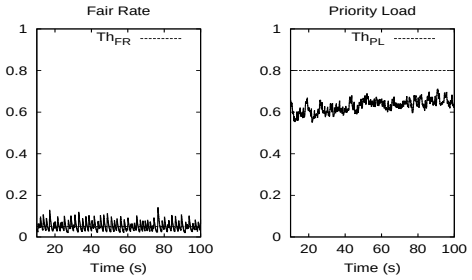

Fig. 7. Fair Rate and Priority Load evolution (Newest-flow policy and $OQ_{th} = 80\%$).

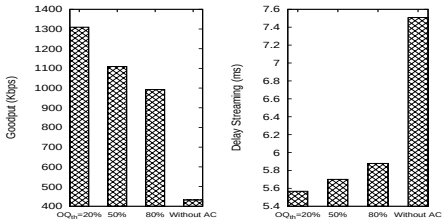

Fig. 8. Average goodput and delay with and without admission control.

C. Admission control in the optical queue

This experiment aims to show the benefits of introducing admission control in the optical queue, since this was proposed in FAN to minimise the service degradation that arises under congestion situation. In this light, fig. 8 shows the results of a simulation example that was carried out both with and without admission control. In both cases, we have considered the Newest-flow policy.

Figure 8 illustrates the average goodput for the TCP flows and the mean delay suffered by the UDP packets in the optical queue. As shown, the case without admission control offers less performance (high delay and low goodput) than when admission control is employed. Indeed, this is the case since, the more flows accepted (when no admission control is used) the more load in the queue. Furthermore, it is worth noticing that it is possible to adjust a given desired QoS just by varying the value of OQ_{th}.

D. Flow routing policies over the optical queue

This experiment aims to study the behaviour of the three policies defined in section III: Newest-flow, Most-Active-flow and Oldest-flow policies. The difference among them is the choice of which flow is to be transmitted over the optical queue. Therefore, the following focuses on the performance of the optical queue.

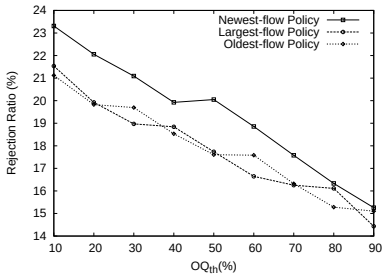

Fig. 9. Rejection ratio.

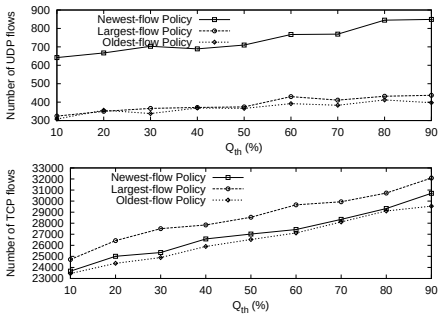

Fig. 10. Total number of UDP and TCP flows switched in the optical layer (over 100 seconds).

Figure 9 shows the flow rejection ratio by the admission control mechanism for the three policies, with OQ_{th} in the range 10% to 90% of the total queue length. As shown, the Newest-flow policy gives the greatest rejection ratio, about 2% larger than the other two policies in most of the cases. The Oldest-flow and Most-Active-flow policies show almost the same rejection ratio. In conclusion, Most-Active- and Oldest-flow policies rejects less incoming flows, thus permitting service to a greater amount of traffic than the Newest-flow policy. This is so given that the former two policies release the IP layer more than the latter, since they take the biggest flows out.

As previously stated, the simulation environment generates UDP and TCP traffic, whereas only the latter uses congestion control. In this light, fig. 10 depicts, for the three policies, the total number of UDP and TCP flows switched in the optical layer during the 100 seconds that the simulation lasts.

First of all, it is important to notice that only a few UDP flows are routed over the optical layer in the cases of Oldest- and Most-Active-flow policies. This is because, with these two policies, only some UDP flows are detected as elastic flows (false positives). On the other hand, it can be seen that the Newest-flow policy sends a greater number of UDP flows through the optical queue. This has

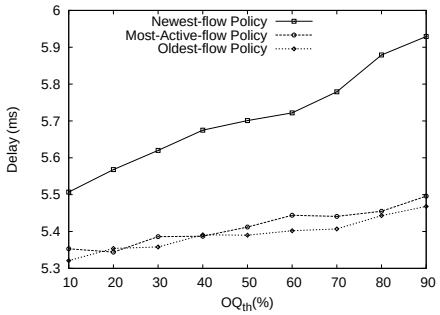

Fig. 11. Mean delay of the UDP packets in the optical queue.

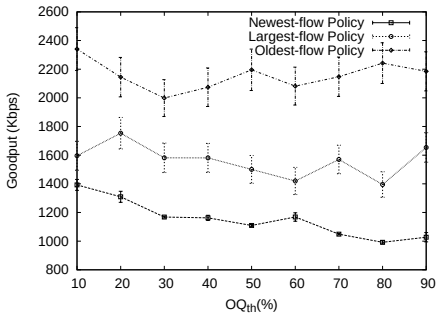

Fig. 12. TCP flows goodput in the optical queue (Confidence intervals=95%)

a tremendous impact on the performance of the optical queue since the UDP flows, which do not do congestion control, increases the overall delay in the optical queue (see fig. 11).

Finally, fig. 12 shows the average goodput of the TCP flows in the optical queue, for different queueing threshold OQ_{th} values. Again, the Newest-flow policy shows the worst results (that is, low goodput values). Concerning the other two, the Oldest-flow policy presents the best results among the three policies, given that the number of TCP flows accepted is smaller than those accepted by the Most-Active-flow policy (see fig. 10 bottom).

The reason for this is that the Oldest-flow policy is more accurate at detecting the heaviest flows, since the Most-Active flow only considers the "backlog", which is a short-term measure of the heaviness of the flows.

V. Conclusions and future work

This work's contributions are two-fold: First, it proposes an extension to Flow-Aware Networking architecture by including an optical layer. This new extended architecture is a simple extension of FAN which uses the same monitoring parameters, but includes a new one (OQ_{th}) to keep FAN's admission control at the optical layer.

And, secondly, this work proposes and analyses three different policies concerning the choice of which flows are moved to the optical layer. The simulations show that the best possible choice, in terms of delay and goodput experienced by the flows, is to switch the heaviest flows found in the IP layer over the optical domain. This is possible using the Most-Active- and Oldest-flow policies which continuously monitor the current flows in the IP layer. Among these two, the latter is more accurate at detecting the heaviest flows or elephants since it monitors flows over a longer period of time.

In future work, the authors shall investigate the performance behaviour of MFAN nodes in a complex topology and its impact in the optical layer with limited resources. In addition to this, other traffic profiles, such as P2P and Grid, shall be studied.

References

[1] P. Pan and H. Schulzrinne, "Yessir: A simple reservation mechanism for the internet," in *Proc. of the 8th International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV)*, 1998.

[2] N. Benameur, A. Kortebi, S. Oueslati, and J. W. Roberts, "Selective service protection in overload: differentiated services or per-flow admission control?," in *11th International Telecommunications Network Strategy and Planning Symposium. NETWORKS*, jun 2004, pp. 217–222.

[3] J. Roberts and S. Oueslati, "Quality of service by flow aware networking," *Philosophical transactions of the Royal Society.*, 2000.

[4] A. Kortebi, S. Oueslati, and J. Roberts, "Cross-protect: implicit service differentiation and admission control," in *IEEE High Performance Switching and Routing*, apr 2004.

[5] A. Kortebi, S. Oueslati, and J. Roberts, "Implicit service differentiation using deficit round robin," in *International Teletraffic Congress*, aug 2005.

[6] K. et al. Sato, "GMPLS-based photonic multilayer router (Hikari router) architecture: an overview of traffic engineering and signaling technology," *IEEE Communications Magazine*, vol. 40, no. 3, pp. 96–101, Mar. 2002.

[7] A. Gumaste, G. Kuper, and I. Chlamtac, "Optimizing light-trail assignment to WDM networks for dynamic IP centric traffic," in *13th IEEE Workshop on Local and Metropolitan Area Networks. LANMAN*, apr 2004, pp. 113–118.

[8] Yabin Ye, H. Woesner, R. Grasso, Tao Chen, and I. Chlamtac, "Traffic grooming in light trail networks," in *IEEE Global Telecommunications Conference. GLOBECOM*, dec 2005, vol. 4.

[9] P. Goyal, H. Vin, and H. Cheng, "Start-time fair queuing: A scheduling algorithm for integrated servicespacket switching networks," in *Proc. of ACM SIGCOMM*, aug 1996.

[10] A. Kortebi, L. Muscariello, S. Oueslati, and J. Roberts, "Evaluating the number of active flows in a scheduler realizing fair statistical bandwidth sharing," *SIGMETRICS Perform. Eval. Rev.*, vol. 33, no. 1, pp. 217–228, 2005.

[11] F. P. Kelly, S. Zachary, and I. Ziedins, *Stochastic Networks: Theory and Applications*, Oxford University Press, USA, 9 1996.

[12] V. Jacobson, "Congestion avoidance and control," in *ACM SIGCOMM*, Stanford, CA, aug 1988, pp. 314–329.