A Multi-layer Recovery Strategy in FAN over WDM Architectures

Jerzy Domżał[1], Robert Wójcik[1], Krzysztof Wajda[2], Andrzej Jajszczyk[4]
Department of Telecommunications, AGH University of Science and Technology,
al. Mickiewicza 30, 30-059 Kraków, Poland, e-mail: {jdomzal, robert.wojcik, wajda, jajszczyk}@kt.agh.edu.pl
Víctor López[1], Jose Alberto Hernández, Javier Aracil[3]
Dept. Ingeniería Informática, Universidad Autónoma de Madrid,
Calle Francisco Tomás y Valiente, 11 - Madrid (Spain), email: {victor.lopez, jose.hernandez, javier.aracil}@uam.es
César Cárdenas[1], Maurice Gagnaire[2]
Department of Networks and Computer Science,
TELECOM ParisTech (formerly ENST Paris and/or Télécom Paris), e-mail: {gagnaire, cardenas}@telecom-paristech.fr
[1] *Student Member, IEEE*, [2] *Member, IEEE*, [3] *Senior Member, IEEE*, [4] *Fellow, IEEE*

***Abstract*—Network operators are migrating towards IP over WDM architectures. In such multi-layer networks, it is necessary to efficiently use the resources available from both layers in order to provide coordinated recovery strategies. Thanks to the development of the control plane (GMPLS and ASON), it is feasible to set up and tear down lightpaths automatically, so the WDM layer itself can support failure recovery. This paper describes a multi-layer recovery strategy in a FAN/WDM (Flow-Aware Networking/Wavelength Division Multiplexing) architecture. We propose using the EHOT (Enhanced Hold-Off Timer) algorithm to control network operation after link or node failure. Although FAN operates only on the IP level, the presented analysis shows that it is possible to ensure sufficiently low (less than 50 ms) recovery times in FAN working over an intelligent optical layer. Additionally, the paper shows the motivation for FAN networks and presents the results of carefully selected simulation experiments which allow for evaluating the duration of outages in data transmission under various conditions.**

***Index Terms*—Flow-Aware Networking; Wavelength Division Multiplexing; recovery strategy; multi-layer networks; QoS; Enhanced Hold-Off Timer**

I. Introduction

With the current technology trends, network operators are gradually migrating towards an IP over WDM paradigm, mainly to benefit from the larger transmission capacity offered by optical fibers. In addition to the improvements in the physical layer, the control plane has been standardized thanks to GMPLS (Generalised Multiprotocol Label Switching) [1] and ASON (Automatic Switched Optical Network) [2]. Using these proposals, it is possible to manage IP/MPLS and optical equipment with the same protocols. The scientific community has proposed multi-layer capable routers [3] that can deal with IP/MPLS traffic and optical services.

Flow-Aware Networking (FAN) is a recent and promising approach to QoS provisioning in IP networks. However, FAN was designed to operate at the IP level without any information about the underlying layers. This is the reason why the authors in [4] defined the concept of Multi-layer Flow-Aware Networking (MFAN) as a router architecture which is capable of handling optical resources provided by WDM technology within the FAN architecture. However, [4] studies the congestion avoidance problem using extra resources provided by the optical layer. This work defines and evaluates an algorithm that allows MFAN nodes to support failure recovery.

This paper describes a multi-layer recovery strategy in FAN/WDM (Flow-Aware Networking/Wavelength Division Multiplexing) architecture. The EHOT (Enhanced Hold-Off Timer) algorithm is used to control network operation after link or node failure. This solution, first proposed in [5], is the improved version of a well known HOT (Hold-Off Timer) algorithm [6] which ensures better coordination between layers. The analysis presented in this paper shows that it is possible to achieve sufficiently low (less than 50 ms) recovery times in FAN working over an intelligent optical layer. There are some proposals that allow for ensuring similar recovery times like, e.g., fast reroute described in [7]. While this solution meets the requirements of short outages in transmission, it has some restrictions, e.g., it works only in packet networks using the SONET/SDH layer and MPLS. Moreover, using this mechanism, traffic is redirected after a network element failure, which may cause the problems in FAN. Flows that are sent through the backup route, have to be accepted in all FAN routers which may not happen immediately. In consequence, this may prolong the break in transmission. Our proposal gives the possibility to repair the failed link or node in the optical layer. Of course, if it is impossible to recover from the failure in optical layer, traffic has to be redirected in the IP layer, e.g., using the fast reroute mechanism. Based on this explanation, we can conclude that MFAN with EHOT inherits the advantages of fast reroute by additional possibilities of network recovery. The paper shows the motivation for FAN networks and presents the results of carefully selected simulation experiments which allow for evaluating the duration of outages in data transmission under various conditions.

The remainder of this document is organized as follows: Section II presents the Flow-Aware Networking, main assumptions, principles, and methods of QoS provisioning. Section III describes why is it important to evaluate the cooperation

978-1-4244-5048-0/09/$26.00©2009 IEEE

between IP and optical layers with regard to network resilience capabilities. Simulation results and their analysis are presented in Section IV, while Section V concludes the paper.

II. Flow-Aware Networking

In the past years, there have been many attempts to introduce a Quality of Service architecture to the IP-based networks. IETF came up with two ready-to-use proposals, namely: IntServ [8] in 1994 and DiffServ [9] in 1998. Unfortunately, due to certain limitations, they have not been adapted to common use. IntServ is known for its utter lack of scalability as it requires constant signalling for a resource reservations (the RSVP protocol) and maintaining the state of each flow on every router in the network. DiffServ was to be an answer to scalability issues of IntServ. Although scalable, DiffServ is criticized mainly for its granularity, complexity and flow aggregations. As an aggregate is the main entity for which QoS is provided, the assurances on the flow level cannot be achieved. Moreover, [10] reveals that the end-to-end delay in DiffServ may increase infinitely unless the link utilizations are kept under a certain level. Finally, the appropriateness of previously proposed QoS architectures is questioned in [11].

The concept of Flow-Aware Networking as a novel approach to assure quality of service in packet networks was initially introduced in [12] and, then, presented as a complete system in 2004 [13]. Bearing in mind all the shortcomings of the previous architecture attempts, FAN is designed to be scalable, simple, efficient and feasible. The goal of FAN is to enhance the current IP network by improving its performance under heavy congestion. To achieve that, certain traffic management mechanisms to control link sharing are introduced, namely: measurement-based admission control [14] and fair scheduling with priorities [13], [15]. The former is used to keep the flow rates sufficiently high, to provide a minimal level of performance for each flow in case of overload. The latter realizes fair sharing of link bandwidth, while ensuring negligible packet latency for flows emitting at lower rates.

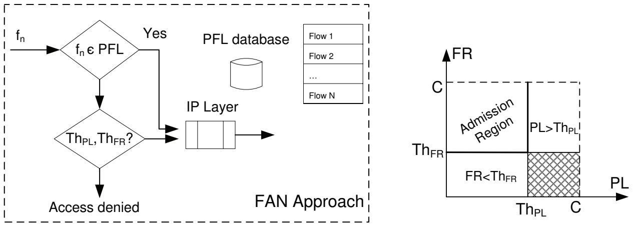
In FAN, admission control and service differentiation are implicit. There is no need for a priori traffic specification, as well as there is no class of service distinction. However, streaming and elastic flows are implicitly identified inside a FAN domain. This classification is based solely on the current flow peak rate. All flows emitting at lower rates than the current fair rate are referred to as streaming flows, and packets of those flows are prioritized. The remaining flows are referred to as elastic flows. Nevertheless, if a flow, firstly classified as streaming, surpasses the Maximum Transfer Unit (MTU) value during a given time interval, it is degraded to the elastic flow. The distinctive advantage of FAN is that both streaming and elastic flows achieve good enough quality of service without any mutual detrimental effect.

For the reasons described above, FAN is an enhancement to the existing IP network. The standard interconnection device in FAN networks is called a Cross-Protect router (XP router). The term ‘cross-protection’ implies that two congestion control mechanisms, i.e., measurement-based admission control and scheduling, cooperate and protect each other to achieve good performance and scalability. The admission control block limits the number of active flows in the XP router, which essentially improves the queuing algorithm functionality, and reduces its performance requirements. It is vital that queuing mechanisms operate quickly, as for extremely high speed links the available processing time is strictly limited. On the other hand, the scheduling block provides admission control with the information on congestion status on the outgoing interfaces. The information is derived based on, for example, current queues occupancy. The cross-protection, therefore, contributes to a shorter required flow list and queue sizes that significantly improve FAN scalability [16].

A. Measurement-Based Flow Admission Control

Admission control is responsible for accepting or rejecting the incoming packets belonging to the flows, based on the current congestion status. Admitted and currently in-progress flows are registered in a Protected Flow List (PFL). If the flow identity of a newly arriving packet is already on the PFL, the packet is forwarded unconditionally. If not, the flow is subject to admission control. If the outgoing link is congested, the packet is simply discarded. In the absence of congestion, the packet is forwarded, and its flow ID is added to the PFL. The ID may be removed from the PFL only after a specified time period of flow inactivity.

The admission control block, in FAN, realizes the measurement based admission control (MBAC) functionality [17]. As FAN does not use signalling of any kind, it implies that every decision taken by the node is autonomous, based solely on the latest measurements performed by the node itself. All of the above makes MBAC, in FAN, implicit.



(a) (b)

Fig. 1. FAN Admission Control: (a) Flow Diagram, (b) Admission Policy

Fig. 1a summarizes the described admission control routine for FAN architectures. Fig. 1b introduces the admission policy applied by the XP router. There are two parameters, based upon which, the admission control block makes its decisions, namely: priority load (PL) and fair rate (FR). A new flow may be admitted only when the current FR is greater and PL is lower than their threshold values, i.e., Th_{FR} and Th_{PL}, respectively (admission region in Fig. 1). If either boundary is exceeded, we refer to the link state as congested. Both congestion indicators are described in the next section.

B. Flow Scheduling Algorithms

As mentioned, currently there are two per-flow fair queuing algorithms proposed for FAN architectures: Priority Fair

Queuing (PFQ) and Priority Deficit Round Robin (PDRR). Both scheduling algorithms have, logically, one priority queue and a secondary queuing system. In addition, an Active Flow List (AFL) is maintained by each. This list is similar to the PFL defined above, but it also stores 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.

Both scheduling algorithms are intended to realize fair sharing of link bandwidth to elastic flows and priority service to streaming flows. The latter (PDRR) was primarily suggested to speed up commercial adoption since it improves the algorithm complexity from $O(log(N))$ to $O(1)$; where N is the number of flows in the AFL. However, it has been shown that both scheduling algorithms have similar performance [18].

PFQ as defined in [13] is based on the SFQ (Start-time Fair Queuing) algorithm [19] and is used to give preference to streaming over elastic flows. PFQ is built on a PIFO (Push In First Out) queue, which stores packet information (flow identifier, size and memory location) and time stamp, the latter determined by the SFQ algorithm. The PIFO queue is split into two areas delimited by a priority pointer (see Fig. 2), 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 (strict and exhaustive scheduling policy). Finally, the queue stores elastic and streaming packet count statistics, which are further used to compute the values of PL and FR. The computational complexity of PFQ is $O(logN)$, where N is the number of active flows in the queuing system.

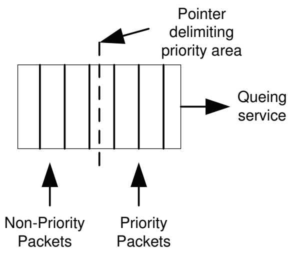


Fig. 2. Priority Fair Queueing system (PFQ)

It has been proved in [16] that fair queuing is scalable since complexity does not increase with link capacity. Moreover, fair queuing is feasible, as long as link loads are not allowed to attain saturation levels, which is asserted by admission control. Compared to other QoS architectures, FAN scalability, due to the lack of signalling and ideal data handling complexity, is not matched by any other architecture [20]. Finally, FAN is a solution which conforms to net neutrality paradigms, as the differentiation is based only on the internal, implicit node decisions. This way, services in a network may be differentiated, while still, maintaining fairness and neutrality.

C. Flow Scheduling Measurements

As mentioned, the two parameters used for flow admission criteria, i.e., FR and PL, are estimated in the scheduling block of the XP router. Both indicators are measured periodically. Considering the time scales of their respective congestion phenomena, the PL parameter is updated at several tens of milliseconds and the FR parameter is updated at several hundreds of milliseconds [13].

PL represents the sum of the lengths of priority packets transmitted in a certain time interval, divided by the duration of that interval, and normalized with respect to the link capacity. To estimate PL, a counter is incremented on the arrival of each priority packet by its length in bytes. Let $pb(t)$ be the value of this counter at time t, (t_1, t_2) a measurement period (in seconds) and C the link capacity. Then, an estimation of PL is:

$$PL = \frac{(pb(t_2) - pb(t_1)) \times 8}{C(t_2 - t_1)} \quad (1)$$

FR indicates approximately the throughput achieved by any flow that is continuously backlogged. In other words, it is the rate available to each flow at the moment. To estimate FR, a virtual flow emitting single byte packets, inserted between real packets in an order dictated by the scheduling algorithms, is considered. For PFQ, in a busy period, the number of bytes transmitted by the queue is given by the evolution of the $virtual\_time$ parameter. In an idle period, the virtual flow emits at the link capacity. Let $v(t)$ be the value of $virtual\_time$ at time t, (t_1, t_2) the measurement period (in seconds), S the total idle time during this interval and C the link capacity. The estimation of the FR for PFQ is:

$$FR = \frac{max\{S \times C, (vt(t_2) - vt(t_1)) \times 8\}}{(t_2 - t_1)} \quad (2)$$

Exponential filters are applied after both measurements. These formulas are for the PFQ scheduling algorithm. Since this work is based on the PFQ scheduling, the formulas for the PDRR scheduling are not presented. Nevertheless, we suggest referring to [15] should it be of the readers interest.

D. Motivation to work on FAN

Flow-based QoS architectures have attracted much attention, mainly due to their appropriate target of service differentiation, i.e., flows. It is worth mentioning that flow-based architectures have been tested [21], [22], patented [23], [24], standardized [25] and commercialized [26]. Furthermore, they have been chosen as a basis for QoS provisioning in Next Generation Networks (NGN) [27]. Particularly, ITU-T has adopted the flow-state-aware transport technology for the provision of QoS in NGN [28]. Furthermore, in [22], the authors compared flow-based and packet-based routers. The results showed that flow-based approach offers enhanced performance in terms of packet processing. FAN architectures have recently received more attention from the Grid community.[1] For instance, the authors in [29]–[31] have evaluated FAN architectures under Grid traffic and showed that FAN outperforms DiffServ architectures under their Grid environment. In conclusion, FAN is a promising approach for QoS provisioning.

III. Motivation for multi-layer recovery strategy

There are no protection and restoration mechanisms in Flow-Aware Networks defined so far. After a link or node failure, the traffic of broken flows is redirected according to the routing algorithm to reach the destination. It means that the flows previously accepted in the FAN router have to compete again with other flows for access to the network resources on the new route. It may take a lot of time and lengthen the transmission time significantly. Unfortunately, voice or real time video applications have to begin their transmission immediately and do not tolerate long delays or outages in transmission. The solution of the long acceptance time problem for such connections is presented in [32] and [33]. The paper shows that congestion control mechanisms for FAN analyzed in [34] and [35] ensure short acceptance times of new streaming flows in a congested FAN link. When using the EFM (Enhanced Flushing Mechanism), RAEF (Remove Active Elastic Flows) or RBAEF (Remove and Block Active Elastic Flows) it is also possible to decrease the restoration time of broken streaming flows. All these mechanisms work based on total or partial cleaning (also called flushing) of the PFL content in congestion. After such an action the congestion is eliminated and new flows may be accepted in the admission control block and begin to send the packets.

When a network element fails, rerouting might be needed. The rerouted flows are treated as new in the XP routers on their new path. If the broken flows are rerouted to the congestion-less link they are accepted immediately. However, when links on the new path are congested, the flows need to wait for the network resources to be available. This waiting time (usually a few seconds) is satisfactory for new streaming flows [36], still, it is too long when the flow was in progress before the outage, and the service does not tolerate breaks in transmission. Moreover, the link may be congested for a long time giving no possibility for rerouted flows to continue the transmission.

There are two main groups of internal routing protocols: distance vector, e.g., RIP (Routing Information Protocol) and link-state, e.g., OSPF (Open Shortest Path First). They are necessary in networks but, unfortunately, they have some disadvantages. RIP is simple but its biggest disadvantage is the maximum hop count of 15, which causes the scalability problems. On the other hand, the OSPF protocol is used in large networks but is more complex (it has to calculate a metric based on many parameters). Almost all routing algorithms have the same drawback, i.e., they are slow to find the route when topology changes. In OSPF it may take even a few seconds from a link or node failure to the installation of a new route [37]. It is possible to reduce the restoration time in the IP layer, e.g., by implementing the MPLS functionality below it, but it is still very hard to ensure that the outage of connectivity is shorter than 50 ms.

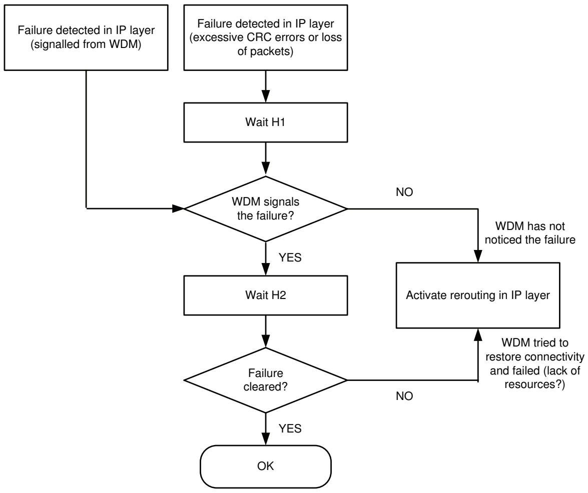

Failure detected in IP layer (signalled from WDM)
Failure detected in IP layer (excessive CRC errors or loss of packets)
Wait H1
WDM signals the failure?
NO
WDM has not noticed the failure
YES
Wait H2
Activate rerouting in IP layer
WDM tried to restore connectivity and failed (lack of resources?)
Failure cleared?
NO
YES
OK

Fig. 3. Enhanced Hold-Off Timer (EHOT) approach

Most of link or node failures may be repaired and if the outage of the network element is short enough the rerouting action is not needed. In this paper, we propose a multi-layer strategy for FAN/WDM architectures. WDM ensures fast failure detection and restoration of a network link. In [38] it is shown that the protection or restoration actions in WDM networks, in most cases, take a few milliseconds. After this time the broken transmission may be continued.

The multi-layer strategy may be implemented in two ways, as the uncoordinated or coordinated approach. The simplest solution is to run the different mechanisms in parallel and independently from each other. In such a case the rerouting algorithm in the IP (FAN) layer is activated when failure occurs. At the same time the protection or restoration mechanisms in the optical transport layer try to restore the connectivity. Changes made by the routing algorithm and the restoration mechanisms in the lower layer may lead to significant performance degradation as well as potential network instability and unnecessary reduction of the network capacity. It is a well known problem in current multi-layer networks. The solution to it is to provide interworking between the IP and optical layers. In this paper, we propose to use the EHOT algorithm, which is presented in Fig. 3. The IP layer can detect a failure in different ways, itself or from lower layers. In particular, it is able to distinguish if the information is from the optical layer or not. This possibility is used by the EHOT algorithm, which is based on dividing the entire Hold-Off Timer into two parts: H1 (short) and H2 (long). The first one (H1) is activated when the IP layer detects a failure and gives the chance for the optical layer to decide if it is possible to recover connectivity at that level. If the answer is positive the H2 timer is activated and the recovery mechanism in the lower layer begins. The

[1] "It's a very promising technology and has significant potential, addressing a number of issues in a way no one else is today." Joe Mambretti, EETimes, 08/06/2007

restoration procedure in the optical layer encompasses both fault localization and the recovery mechanisms (i.e., dedicated path protection or restoration). If the optical layer is unable to solve the failure (basically due to the unavailability of resources) during the H2 period, then the rerouting algorithm in the IP layer is launched. The same situation takes place if there is no positive answer from the optical layer during the H1 time. The main assumption of the algorithm is that the WDM is required to signal both signal degradation and signal failure to its client layer while the IP layer is able to accept such signals. The operation of the algorithm is simple and it is easy to predict that it allows for decreasing the recovery time in comparison to the HOT algorithm when the failure occurs in the IP layer. The simulation results of the EHOT algorithm in the RPR (Resilient Packet Rings) over OTN (Optical Transport Networks) multi-layer network presented in [5] confirm this statement. Moreover, the cost of using our proposal is essentially the same as without it.

IV. Simulation scenarios and results

In this section we present the results of carefully selected simulation experiments carried out in the ns-2 simulator. In order to analyze the differences between basic FAN and the multi-layer FAN/WDM (MFAN) strategy we simulated the traffic distribution in two scenarios described below. In the first case, there is only one FAN link on the backup route, while in the second one the traffic is sent through the backup route with two FAN links. The impact of the EHOT algorithm on the transmission in the network was also analyzed. Basic simulation parameters are presented below. The detailed description of the simulation environment is presented in [39].

A. FAN in case of link failure

The network topology, which was used in our analysis, is presented in Fig. 4. The traffic is generated in the source node (marked in Fig. 4 as S). It is destined to different nodes (D1 and D2 in Fig. 4) accordingly to the analyzed scenario.

In the first part of our research we simulated two different scenarios to show the disadvantages of FAN when the link fails. In both of them the traffic is sent to the destination node (hereafter called D1) connected to router R3. While the cost of each link is the same, the route from the source to the destination is the same in both analyzed routing algorithms (distance vector and link state). Under normal conditions (without failures) the traffic to the destination node is sent through routers R1, R2 and R3. In the first analyzed case the traffic is also sent to the destination node (hereafter called D2) connected to router R5. The traffic to node D2 is sent through routers R1, R4 and R5. The goal of such a traffic assignment is to cause congestion in both FAN links and to check what happens when link L2 fails at 200 s time instant. The simulation duration was set to 500 s to allow for observing the acceptance times of the redirected flows in the R4 router.

The capacity of FAN links (L3 and L5) was set to 100 Mbit/s and the capacity of the rest of the links, with

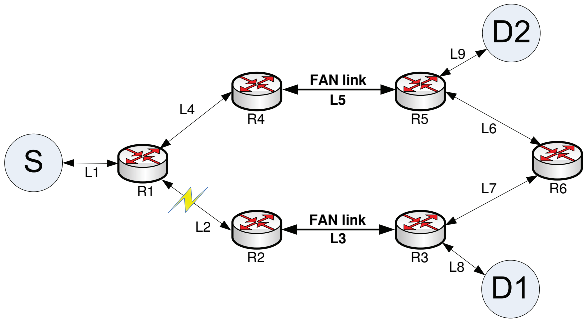


Fig. 4. Basic simulated network topology

the FIFO queues, was set to 1 Gbit/s. We provided the traffic pattern with the Pareto distribution for calculating the volume of the elastic traffic (number of bytes a source sends) directed to both destination nodes. The exponential distribution for generating the time intervals between beginnings of the transmissions of the elastic flows as well as for generating the start times of streaming flows was used. We decided to analyze the VoIP connections realizing the Skype service. The packet size was set to 100 bytes and the transmission rate was set to 80 kbit/s for each of the streaming flows. We made our simulation runs in various conditions changing the number of elastic and streaming flows. We analyzed the acceptance time of each streaming flow in the AC block of router R2 (before failure) and router R4 (after failure) and the number of accepted flows in L3 and L5 links before and after the L2 link failure. The measurement interval for the PL parameter was set to 50 ms while the FR values were estimated every 500 ms. The thresholds Th_{PL} (maximum allowed value of the PL) and Th_{FR} (minimum allowed value of the FR) were set to 70% and 5% of the link capacity, respectively, and the *pfl_flow_timeout* parameter was set to 20 s. 95% confidence intervals were calculated by using the Student's t-distribution.

The mean acceptance time of streaming flows in the R2 router before the L2 link failure in function of the number of elastic flows active in background is presented in Fig. 5. The mean number of active flows in steady state in link L3 and the number of flows accepted in link L5 after link L2 failure are presented in Fig. 6. As we can see, the acceptance time of new streaming flows (also referred to as *waiting_time*) in the congested link is independent of the number of elastic flows being active in background. It is also independent of the number of streaming flows, which want to begin the transmission [35]. The streaming flows are accepted in the routers of both FAN links after tens of seconds. It means that they have to wait for a long time before they begin to send their packets. Moreover, their transmission is broken for a long time if the link failure occurs and the traffic is redirected to the congested link. Our analysis does not consider the time consumed by the routing algorithm which might lengthen the outage in connectivity for additional seconds or even minutes. The presented values are unacceptable and have to be shortened. It is possible to decrease the *waiting_time*

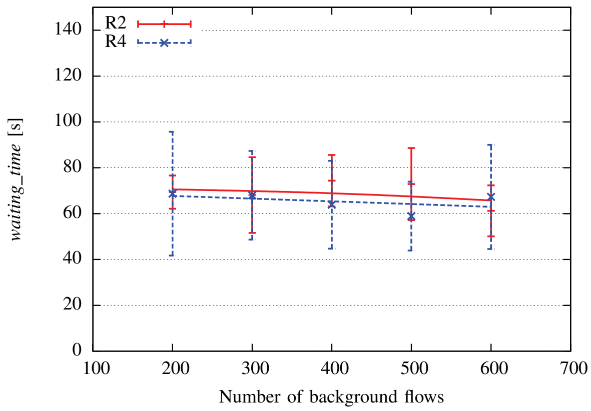


Fig. 5. The acceptance time of streaming flows in routers R2 (under normal circumstances) and R4 (after L2 failure); L5 is congested before the failure

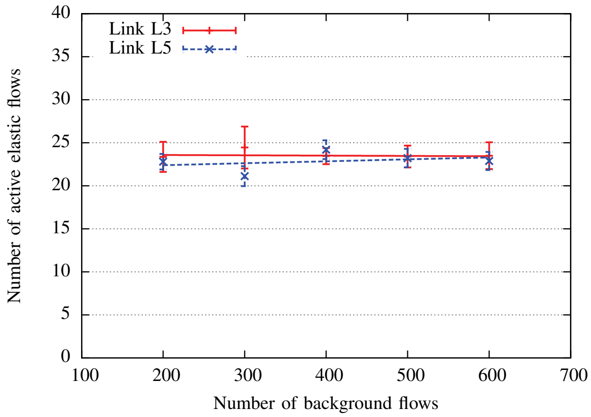


Fig. 6. The number of active elastic flows accepted in links L3 (under normal circumstances) and L5 (after L2 failure); L5 is congested before the failure

of streaming flows by implementing the congestion control mechanisms proposed for FAN in [34] and analyzed in the failure scenarios in [5]. Unfortunately, it is impossible to ensure the recovery time shorter than 50 ms in this solution. Moreover, the number of flows accepted in the AC block after the flushing action of the PFL content is significantly greater than that in the basic FAN architecture. In Fig. 6, we can see that the mean number of active flows in the FAN link is almost constant and does not exceed 30.

The scenario in which there is more than one FAN link on the backup route and all of them are congested in the moment when L2 link fails (see Fig. 7) was also analyzed. In such a case the additional source S1 sends the traffic to node D3. The mean volume of a single TCP flow (provided by the Pareto distribution) and the mean inter-arrival times between elastic flows (provided by the exponential distribution) were increased twice with regard to those for traffic sent to node D2. The results of 40 simulation runs, 10 for each case (basic FAN links and FAN links with the respective congestion control mechanisms) are presented in Tab. I.

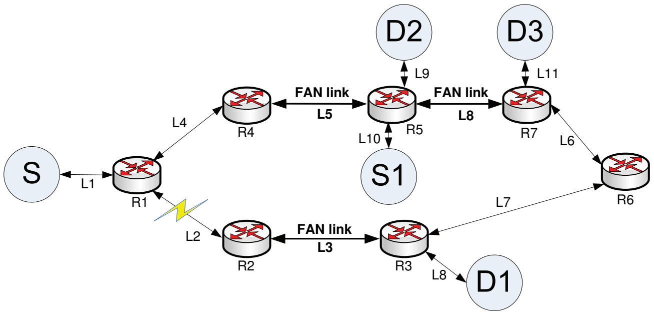


Fig. 7. Simulated network topology with two FAN links on backup route

TABLE I
THE *waiting_time* VALUES ON BACKUP ROUTE (FIG. 7)

Mechanism / Parameter	***waiting_time* at router R4**	***waiting_time* at router R5**
Basic FAN	68.70 s ± 27.01 s	135.54 s ± 32.64 s
EFM / *pfl_flushing_timer*	1.85 s ± 0.29 s	2.38 s ± 0.16 s
RAEF / *active_time*	1.69 s ± 0.11 s	2.07 s ± 0.19 s
RBAEF / *active_time*	1.44 s ± 0.09 s	2.15 s ± 0.21 s

The *pfl_flushing_timer* parameter (set to 5 s in our experiment) is the minimum time period between two flushing actions in the EFM. The *active_time* (also set to 5 s) is the key parameter for the RAEF and RBAEF mechanisms. The identifiers of elastic flows, being active for at least *active_time*, are removed from the PFL in congestion. Moreover, in the RBAEF mechanism they are also blocked for 1 s.

The results show that the congestion control mechanisms allow for decreasing the acceptance time of redirected streaming flows on the backup route. We have to note that even using one of those mechanisms, it is impossible to accept the streaming flows on the new route in time less than 50 ms. Moreover, we can see that each congested FAN link on the backup route may increase the total *waiting_time* as it can reject new flows. The analysis shows that rerouting in the IP layer is not a fast process.

In the next analyzed case (in topology from Fig. 4) the traffic is sent only to the destination node connected to router R3. It means that after a failure the traffic is redirected from the FAN link L3 to the empty FAN link L5. The results of similar analysis as in the previous case are presented in Fig. 8 and Fig. 9. We can see that the redirected streaming flows are accepted immediately in the admission control block of the R4 router. This situation was to be predicted. The more interesting feature to notice is that the number of accepted elastic flows in the L5 link after L2 failure increases linearly with the number of redirected elastic flows. It means that in the congestion-less network the streaming flows are quickly accepted but in case of a link failure the network shows scalability problems. Unfortunately, the time needed by routing algorithms to recover the connectivity after a failure cannot be reduced in an easy way which is the second problem

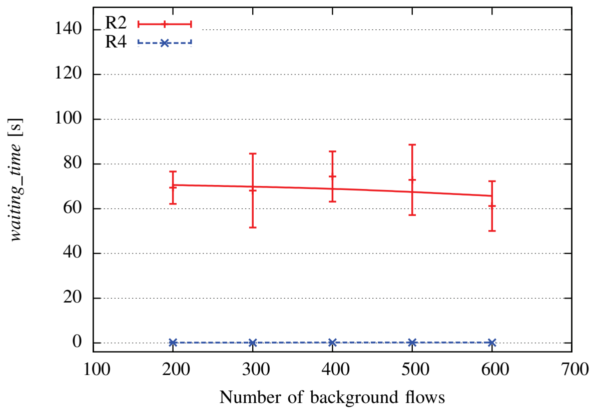


Fig. 8. The acceptance time of streaming flows in routers R2 (under normal circumstances) and R4 (after L2 failure); L5 is empty before the failure

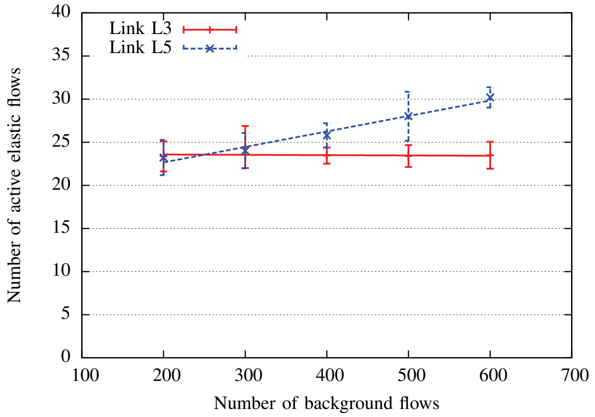


Fig. 9. The number of active elastic flows accepted in links L3 (under normal circumstances) and L5 (after L2 failure); L5 is empty before the failure

observed in this case. Moreover, the scalability problems of the congestion control mechanisms also remain unsolved if the traffic is redirected to the uncongested link (in such cases the number of accepted flows at time instant may be high).

B. MFAN in case of link failure

The multi-layer FAN/WDM architecture, presented in Fig. 10 may be the solution to the problems presented in the previous section. The EHOT algorithm shown in Fig. 3 was implemented to ensure the coordination between layers. Based on the well known parameters (partially presented in [38]) we set H1 to 5 ms and H2 to 20 ms. It means that after noticing the failure, the IP layer waits 5 ms for the decision from the optical layer if it is able to repair the failure. During this time the routing algorithm is deactivated. If there is a positive information from the optical layer, it receives 20 ms to repair the failure (we assumed that the failure is repaired in 5 ms). On the other hand, if there is no answer from that layer during 5 ms (or is negative) the traffic is rerouted in the network layer. We assumed that the L2 link is protected by an additional link in the optical layer (see Fig. 10). The simulation results show

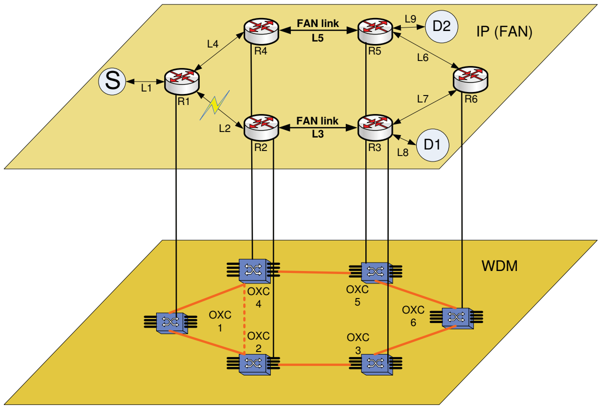


Fig. 10. The basic simulation multi-layer topology

that the transmission in the L2 link is broken for a very short time (less than 10 ms). There are no problems with redirected traffic and a too high number of accepted flows after failure. Moreover, in the case without the EHOT algorithm, the link is repaired even faster. Unfortunately, in this uncoordinated approach, the traffic is rerouted twice (after a failure to the backup link and again, when the first rerouting process ends, back to the primary link). If the optical layer is not able to repair the failure, the algorithm gives the similar results to those presented in Section IV-A.

We can assume that the multi-layer FAN/WDM strategy gives the possibilities for fast repairing the failed link in time less than 50 ms. It is very important for the streaming flows which realize the transmission of real time applications. The proposed and analyzed mechanism solves some of the well known problems of FAN. The advantages and disadvantages of the examined scenarios are summarized in Tab. II.

TABLE II
THE ADVANTAGES AND DISADVANTAGES OF FAN ARCHITECTURES, BASIC, FAN/WDM AND FAN/WDM WITH EHOT

Link failure	**Advantages**	**Disadvantages**
Basic FAN	simple (only rerouting)	relatively slow rerouting: - in congestion rerouted flows are not accepted; - in congestionless the number of accepted flows after rerouting may be too large
MFAN without EHOT	chance for fast recovery in optical layer	rerouting in IP layer and recovery in optical layer at the same time
MFAN with EHOT	chance for recovery without rerouting in the IP layer (fast and short outage in transmission and without changes in flows assignment)	complexity of the algorithm

V. Conclusions

FAN is a new network architecture proposed as an answer to the DiffServ inconveniences. It ensures implicit traffic classification. The packets of streaming flows are sent with a high probability while the elastic flows realize the best effort service. FAN is a promising solution and after some additional research may be implemented in the future Internet. There are still some problems to solve when considering FAN. One of them, presented in this paper, is to ensure the reliable transmission. There are no protection and restoration mechanisms in FAN which means that in case of link or node failure the traffic has to be redirected. The disadvantages of such a situation are described and analyzed in the paper. We propose to implement the MFAN architecture to improve the chances for making the link restoration process quicker than 50 ms. The proposed architecture with the EHOT algorithm, described in details and analyzed by simulation experiments, looks to be a good solution to the stated problem and may be used in FAN.

Acknowledgements

The work described in this paper was carried out with the support of the BONE-project ("Building the Future Optical Network in Europe"), a Network of Excellence funded by the European Commission through the 7th ICT-Framework Programme.

References

[1] IETF, "Generalized Multi-Protocol Label Switching (GMPLS) Architecture," IETF RFC 3945, October 2004.

[2] ITU-T, "Architecture for the Automatically Switched Optical Network (ASON) - Rec. 8080/Y.1304," Recommendation ITU-T Y.1304, 2001.

[3] K. Sato, N. Yamanaka, Y. Takigawa, M. Koga, S. Okamoto, K. Shiomoto, E. Oki, and W. Imajuku, "GMPLS-based photonic multilayer router (Hikari router) architecture: an overview of traffic engineering and signalingtechnology," *Communications Magazine, IEEE*, vol. 40, no. 3, pp. 96–101, 2002.

[4] V. Lopez, C. Cardenas, J. A. Hernandez, J. Aracil, and M. Gagnaire, "Extension of the flow-aware networking (FAN) architecture to the IP over WDM environment," in *Telecommunication Networking Workshop on QoS in Multiservice IP Networks, 2008. IT-NEWS 2008. 4th International*, Venice,, Feb. 2008, pp. 101–106.

[5] J. Domzal, K. Wajda, S. Spadaro, J. Sole-Pareta, and D. Careglio, "Recovery, Fairness and Congestion Control Mechanisms in RPR Networks," in *PSRT 2005*, Poznan, Poland, September 2005.

[6] P. Demeester and M. Gryseels, "Resilience in multilayer networks," *IEEE Commun. Mag.*, vol. 37, pp. 70–76, August 1999.

[7] IETF, "Fast Reroute Extensions to RSVP-TE for LSP Tunnels," IETF RFC 4090, May 2005.

[8] R. Braden, D. Clark, and S. Shenker, "Integrated Services in the Internet Architecture an Overview," IETF RFC 1633, June 1994.

[9] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An Architecture for Differentiated Services," IETF RFC 2475, December 1998.

[10] A. Charny and J. L. Boudec, "Delay Bounds in a Network with Aggregate Scheduling," in *First International Workshop of Quality of Future Internet Services (QOFIS2000)*, vol. 1922/2000, 2000, pp. 1–13.

[11] J. Roberts, "Internet Traffic, QoS and Pricing," in *Proceedings of the IEEE*, vol. 92, September 2004, pp. 1389–1399.

[12] J. Roberts and S. S. Oueslati, "Quality of Service by Flow Aware Networking," *Philosophical Transactions of The Royal Society of London*, vol. 358, pp. 2197–2207, September 2000.

[13] A. Kortebi, S. Oueslati, and J. Roberts, "Cross-protect: implicit service differentiation and admission control," in *IEEE HPSR 2004*, Phoenix, USA, April 2004.

[14] S. Oueslati and J. Roberts, "A new direction for quality of service: Flow-aware networking," in *NGI*, Rome, Italy, April 2005.

[15] A. Kortebi, S. Oueslati, and J. Roberts, "Implicit Service Differentiation using Deficit Round Robin," in *ITC19*, Beijing, China, August/September 2005.

[16] A. Kortebi, L. Muscariello, S. Oueslati, and J. Roberts, "On the scalability of fair queueing," in *ACM HotNets-III*, San Diego, USA, November 2004.

[17] A. Kortebi, S. Oueslati, and J. Roberts, "MBAC algorithms for streaming flows in Cross-protect," in *EuroNGI Workshop*, Lund, Sweden, June 2004.

[18] A. Kortebi, L. Muscariello, S. Oueslati, and J. Roberts, "Evaluating the number of Active Flows in a Schedular Realizing Fair Statistical Bandwidth Sharing," in *SIGMETRICS'05*, Banff, Canada, June 2005.

[19] P. Goyal, H. M. Vin, and H. Cheng, "Start-time Fair Queuing: A Scheduling Algorithm for Integrated Services Packet Switching Networks," *IEEE/ACM Transactions on Networking*, vol. 5, pp. 690–704, October 1997.

[20] J. Joung, J. Song, and S. S. Lee, "Flow-Based QoS Management Architectures for the Next Generation Network," *ETRI Journal*, vol. 30, pp. 238–248, April 2008.

[21] N. Benameur, S. Oueslati, and J. Roberts, "Experimental Implementation of Implicit Admission Control," 2003.

[22] J. Park, M. Jung, S. Chang, S. Choi, M. Young Chung, and B. Jun Ahn, "Performance Evaluation of the Flow-Based Router Using Intel IXP2800 Network Processors," in *International Conference on Computational Science and Its Applications (ICCSA)*, 2006.

[23] S. Oueslati and J. Roberts, "Method and device for implicit differentiation of quality of service in a network," *United States Patent 2004/0213265 A1*, October 2004.

[24] S. Oueslati, J. Roberts, and N. Benameur, "Method and device for management of flow in a packet-telecommunication network," *United States Patent 2008/0212475 A1*, September 2008.

[25] ITU-T E.417, "Framework for the network management of IP-Based networks," 2005.

[26] L. G. Roberts and A. E. Henderson, "System, Methods, and Computer Program Product for Controlling Output Port Utilization," *United States Patent 2007/0171826 A1*, July 2007.

[27] J. Song, M. Chang, S. Lee, and J. Joung, "Overview of ITU-T NGN QoS Control," *IEEE Communications Magazine*, vol. 163, pp. 116–123, September 2007.

[28] ITU-T, "Requirements for the support of flow-state-aware transport technology in an NGN," Recommendation ITU-T Y.2121, January 2008.

[29] C. Cardenas, M. Gagnaire, V. Lopez, and J. Aracil, "Admission control for Grid services in IP networks," in *IEEE First Symposium on Advanced Networks and Telecommunications Systems (ANTS07)*, 2007.

[30] ——, "Performance Evaluation of the Flow-Aware Networking (FAN) architecture under Grid environment," in *20th IEEE/IFIP Network Operations and Management Symposium (NOMS08)*, 2008.

[31] C. Cardenas and M. Gagnaire, "Performance comparison of the Flow-Aware Networking (FAN) architectures under GridFTP traffic," in *23rd ACM/SIGAPP Symposium on Applied Computing (SAC08)*, 2008.

[32] J. Domzal, R. Wojcik, and A. Jajszczyk, "The Impact of Congestion Control Mechanisms on Network Performance after Failure in Flow-Aware Networks," in *Proceedings of International Workshop on Traffic Management and Traffic Engineering for the Future Internet, FITraMEn 2008*, Porto, Portugal, December 2008.

[33] A. Jajszczyk and R. Wojcik, "Emergency Calls in Flow-Aware Networks," *Communications Letters, IEEE*, vol. 11, pp. 753–755, September 2007.

[34] J. Domzal and A. Jajszczyk, "New Congestion Control Mechanisms for Flow-Aware Networks," in *IEEE ICC*, Beijing, China, May 2008.

[35] ——, "The Flushing Mechanism for MBAC in Flow-Aware Networks," in *NGI*, Krakow, Poland, April 2008.

[36] ITU-T, "Network grade of service parameters and target values for circuit-switched services in the evolving ISDN," Recommendation ITU-T E.721, May 1999.

[37] S. Pasqualini, A. Iselt, A. Kirstadter, and A. Frot, "MPLS Protection Switching Vs. OSPF Rerouting: A Simulative Comparison," in *Switching White Papers*, Siemens, January 2008.

[38] S. Ramamurthy, L. Sahasrabuddhe, and B. Mukherjee, "Survivable WDM Mesh Networks," *Journal of Lightwave Technology*, vol. 21, pp. 870–883, April 2003.

[39] http://www.kt.agh.edu.pl/∼jdomzal/sim_param_drcn09.pdf.