ARTICLE IN PRESS

Optical Switching and Networking ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Optical Switching and Networking

journal homepage: www.elsevier.com/locate/osn

Admission control in Flow-Aware Networking (FAN) architectures under GridFTP traffic☆

César Cárdenas [a,*], Maurice Gagnaire [a], Víctor López [b], Javier Aracil [b]

[a] *École Nationale Supérieure des Télécommunications (Telecom ParisTech), 46 rue Barrault-Paris Cedex 13, France*

[b] *Dept. Ingeniería Informática, Universidad Autónoma de Madrid, Calle Francisco Tomás y Valiente, 11-Madrid, Spain*

ARTICLE INFO

Article history:
Received 9 April 2008
Accepted 17 May 2008
Available online xxxx

Keywords:
Quality of service
Flow-Aware networking
Grid services

ABSTRACT

Computing and networking resources virtualization is the main objective of Grid services. Such a concept is already used in the context of Web-services on the Internet. In the next few years, a large number of applications belonging to various domains (biotechnology, banking, finance, car and aircraft manufacturing, nuclear energy etc.) will also benefit from Grid services. Admission control is a key functionality for Quality of Service (QoS) provision in IP networks, and more specifically for Grid services provision. Service differentiation (DS) is a widely deployed technique on the Internet. It operates at the packet level on a best-effort mode. Flow-Aware Networking (FAN) that operates at the scale of the IP flows relies on implicit flow differentiation through priority fair queuing (PFQ). It may be seen as an alternative to DS. A Grid session may be seen as a succession of parallel TCP/IP flows characterized by data transfers with much larger volume than usual TCP/IP flows. In this paper, we propose an extension of FAN for the Grid environment called Grid over FAN (GoFAN). We compare, by means of computer simulations, the efficiency of Grid over DS (GoDS) and GoFAN. Two variants of GoFAN architectures based on different fair queuing algorithms are considered. As a first step, we provide two short surveys on QoS for Grid environment and on QoS in IP networks respectively.

1. Introduction

Grid networks consist of large-scale distributed systems that share heterogeneous resources (computing, storage, network components and equipment, sensors, etc.), and make possible the creation of virtual organizations (utility-computing, utility-storage, virtual laboratories, etc.) [1]. Furthermore, these capabilities enable powerful, flexible, pervasive and cost-effective services to the users. The term *Grid* has been adopted as an analogy to the power Grid. Since the widespread of the Internet, the growth of users and the increasing demand for high-demanding applications, Grid services will be progressively deployed in Internet networks (e.g. GoIP) in the years to come. However, the large installed base of Internet services, equipment and providers slows down network development and makes the introduction of disruptive technology difficult. To solve this problem, overlay network technologies, like Grid networks, appear to be very promising [2].

Quality of Service (QoS) is a key issue for Grid services provisioning [4], and admission control mechanisms are very important to achieve this [3]. Most current Grid

☆ A short summarized version of this paper was presented at the First Symposium on Advanced Networks and Telecom Systems (ANTS) 2007 in Mumbai, India, in December 2007. The authors would like to thank the support from CONACYT, SEP and Monterrey Tech Querétaro Campus, Mexico, the European Union VI Framework Programme e-Photon/ONe+ Network of Excellence (FP6-IST-027497) and 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.

* Corresponding author. Tel.: +33 524422383155; fax: +33 52448238 3278.

E-mail addresses: ccardena@itesm.mx, cardenas@telecom-paristech.fr (C. Cárdenas).

1573-4277/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.osn.2008.05.003

services are provided over best-effort (BE) networks. Thus, QoS architectures originally developed for IP such as DiffServ (DS) have been adapted to Grid environments: GARA [6], NRSE [7], G-QoSM [8], GNRB [9] and [10–17]. Nevertheless, none of those proposals has been widely adopted yet. Therefore, QoS provisioning for GoIP still remains a challenge.

Flow-Aware Networking (FAN) architectures were proposed in [18–20] as a potential alternative for QoS provisioning in Internet networks. FAN overcomes the difficulties of DS and IntServ (IS). To this end, FAN employs per-flow admission control and implicit flow differentiation through priority fair queuing (no packet marking and explicit classification as in DS, no resource reservation as in IS).

In our previous work [23,24] we compared DS against one of the second generation FAN architectures under Grid traffic (Go2GFAN); the scheduling algorithm was based on Priority Fair Queuing (PFQ) [19]. The metrics were average GridFTP session delay and average GridFTP goodput. Our results showed that FAN approach can also be considered as a promising solution for QoS provisioning in a Grid environment. In another work [25], we make an extensive comparison of the two FAN architectures under GridFTP traffic. The work presented here complements our previous results [23]. First, we give a short overview of QoS architectures for a Grid environment, then we compare the other 2GFAN (PDRR or Priority Deficit Round Robin) architecture against DS, and finally we compare 2GFAN (PFQ) against 2GFAN (PDRR) when admission control is applied to Grid sessions.

This work is organized as follows. In Section 2, we survey current QoS architectures for the Grid environment. Then, in Section 3 we recall the main standards for QoS in IP networks before going to Section 4 where we describe the FAN architectures. Our main previous results and related work are discussed in Section 5. In Section 6 we describe our experiments. Then, in Section 7 we discuss the results of our computer simulations. The last section concludes this work.

2. Quality of service architectures for grid environment

Currently, almost all Grid services are being supported by undifferentiated, nondeterministic, best effort IP services. Grid networks must support many large-scale data-intensive applications requiring high volume and high performance data communications. Grid network performance is measured by the support for high-volume data flows and by the capacity of the network to control fine-grained applications [4]. Some efforts to provide QoS in Grid networks are: GARA [6], NRSE [7], G-QoSM [8], and GNRB [9]. Which are describe as follows.

General-purpose Architecture for Reservation and Allocation (GARA) [6] (a.k.a. Pre-GRAM) is a prototype intended to integrate Grid environments and networks services. GARA provides a uniform QoS for different types of Grid resources, it allows advance and online reservation of such resources. Some functionalities of GARA are part of the Globus Tool Kit (GTK).[1] Through Application Programming Interfaces (APIs), GARA links Grid services to Layer 3 services and allows the DS-based router interfaces to ensure application requirements are fulfilled by network resources and controlled by Grid services. GARA signaling and per-flow state overhead cause scalability problems.

Network Resource Scheduling Entity (NRSE) [7] tries to overcome the difficulties of GARA by storing per-flow/per-application states only at the end-host involved in the communication. Service demands can also be online or in advance. A drawback of NRSE is that the API is not clearly defined.

Grid Quality of Service Management (G-QoSM) [8] is a framework to support QoS management under the Open Grid Service Architecture (OGSA). G-QoSM supports many types of resources.

Grid Network-aware Resource Broker (GNRB) [9] is a centralized and enhanced per-domain Grid Resource Broker with the capabilities provided by a Network Resource Manager. GNRB allows requests of the network status and can reserve network resources. A problem may arise when the number of administrative domains rises since the GNRB may become a bottleneck. Also, the administrative domain is very sensitive to GNRB failure.

A new concept for QoS provisioning in Grid networks based on a Virtual Machine approach is in development [26]. It provides very fine grain reservations of CPU time, disk and network bandwidth. The main idea is to reserve the resources and to run the jobs on top of them. Other advanced QoS concepts and architectures have been tested in experimental platforms: Equivalent Differentiated Services (EDS) [14], programmable networks [15], active networks [16], DiffServ-IntServ [17].

3. Quality of service in IP networks

Native IP technology is connectionless and only offers Best Effort (BE) services. Two paradigms have been proposed to improve QoS in IP networks: IntServ (IS) and DiffServ (DS). IntServ (IS) is based on the concept of flow defined as a packet stream that requires a specified QoS level and it is identified by the quintuple "IP source address, IP destination address, Protocol, TCP/UDP source port, TCP/UDP destination port". QoS is reached by the appropriate tuning of different mechanisms: resource reservation, admission control, packet scheduling and buffer management. Both packet scheduling and buffer management act on per-flow basis. The state of the flows must be maintained in the routers and periodically updated by means of a resource reservation signaling system. Since it needs to detect each single flow, the cost and complexity increase with the number of flows, therefore, IS lacks scalability.

DiffServ has been proposed to solve the scalability problems of IntServ. DS classify an aggregation of the traffic in 64 different classes by means of a label in the DS Code Point (DSCP) field of the IPv4 packet header. Identification is performed at edge nodes. The DSCP

[1] http://www.globus.org/.

Please cite this article in press as: C. Cárdenas, et al., Admission control in Flow-Aware Networking (FAN) architectures under GridFTP traffic, Optical Switching and Networking (2008), doi:10.1016/j.osn.2008.05.003

specifies a forwarding behavior (Per-Hop Behavior; PHB) within the DS domain. The same DSCP may have different meanings in consecutive domains and negotiations are needed. The class selector PHB offers three forwarding priorities: Expedited Forwarding (EF), Assured Forwarding (AF) and Best Effort (BE). Packets marked with the highest drop precedence are dropped with lower probability than those characterized by the lowest drop precedence. Although DS does not suffer from scalability problems, it is not able to provide the required end-to-end QoS to IP flows [43]. To overcome the limitations of IS and DS, the Flow-Aware Networking (FAN) approach [18] was proposed.

4. Flow-Aware networking architectures

Flow-Aware Networking (FAN) architectures are required mainly because current QoS provisioning architectures for Internet networks have the following difficulties: native IP is widespread and has no QoS guarantees; IntServ (IS), as explained in the previous paragraph, is considered too complex and not scalable because RSVP needs refreshing mechanisms that introduce resource management complexities; DiffServ (DS) is scalable but has a limited number of bits for identifying individual flows. Moreover, within an IP Diffserv-based network, the QoS offered is per packet-class, another motivation for FAN is that Internet traffic at the packet level can be approximated to a self-similar process but the design of traffic control mechanisms (e.g. Token Bucket configuration) from this approach seems to be very complex [28]. Furthermore, it has been shown that flow-based traffic models represent the QoS perceived by the user better than packet-based traffic models [29]. Another reason for FAN is the current increment of new real-time applications (IPTV, VoIP, P2P, etc.) requiring more restricted QoS; then the flow-based approach can naturally take into account this new class of flows. The last but not least motivation for FAN is that since the current Internet user explosion there is a need for QoS provisioning architectures easily adapted to BE interfaces and FAN can be implemented by putting together an admission control and the scheduler in each BE port.

A first generation of FAN (1GFAN) was proposed in [18] as a new approach to offer IP-QoS at flow level. A flow can be considered a stream of packets with same header attributes and with a maximum inter-packet spacing and are classified explicitly (like in DS). Second generation FAN (2GFAN), performs through per-flow priority fair queuing, implicit classification (no packet marking as in DS, no resource reservation as in IS) of flows into either streaming (high-priority) or elastic (low-priority), and defines a proactive per-flow admission control mechanism. Then, 2GFAN combines two flow-based traffic control mechanisms: Per-flow Fair Queuing (pfFQ) and Per-flow Admission Control (pfAC). pfFQ ensures that link bandwidth is shared equitably between contending flows and pfAC ensures the scheduler performs correctly even in overload by keeping the rate at pfFQ above a minimum rate, called the fair rate. On high capacity links fair queuing is enough to guarantee low packet delay and loss for real-time flows (whose rate is less than the fair rate). 2GFAN seeks

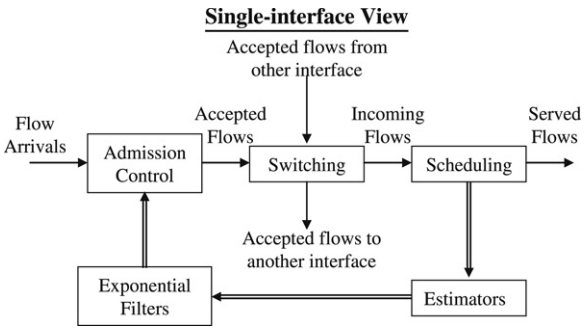

Fig. 1. Flow-Aware networking (FAN) mechanisms.

two objectives: on the one hand, it gives preference to streaming flows attempting to minimize the delay and loss (signal conservation) they experience but, at the same time, it aims to assure a minimum throughput rate to elastic flows (throughput conservation). 2GFAN simplifies network operations leading to potentially significant cost reductions in the IP backbone because it increases network efficiency. It requires no change on the existing protocols and new protocols. It can be implemented as an individual device connected to each BE router interface. An accepted flow is protected during all its transmission time, if the time interval between two packets of that flow keep below a timeout value. To this aim, accepted flows are registered in a list called the Protected Flow List (PFL). Fig. 1 shows one interface of a FAN router.

4.1. Measurement-based flow admission control

To accomplish their tasks, the admission control of 2GFAN is based on two congestion measures of threshold type: Priority Load (PL) and Fair Rate (FR) [19]. PL is the service rate of the priority queue and FR is the service rate a new TCP flow can get when using fair queuing. PL is estimated every tenths of milliseconds (packet timescale) and FR is estimated every hundredths of milliseconds (flow timescale). The fair rate measure is equivalent to the available throughput available for a new TCP connection and is estimated using the TCP phantom technique [19]. The priority load estimator represents the amount of bytes served by the priority queue during the sampling period. Packets of flows emitting at less than the FR are given priority. Incoming flows are denied access to the system when the 2GFAN architecture can not guarantee a given performance level in terms of delay and fair rate. This admission control mechanism is depicted in Fig. 2 along with its mathematical structure in Fig. 3.

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 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 admission control process. Here it is tested whether $PL < PL_{Th}$ and $FR > FR_{Th}$, that is, whether the QoS guarantees defined by the PL_{Th} and FR_{Th} thresholds are maintained or not. If this is the case, the new flow is accepted;

Please cite this article in press as: C. Cárdenas, et al., Admission control in Flow-Aware Networking (FAN) architectures under GridFTP traffic, Optical Switching and Networking (2008), doi:10.1016/j.osn.2008.05.003

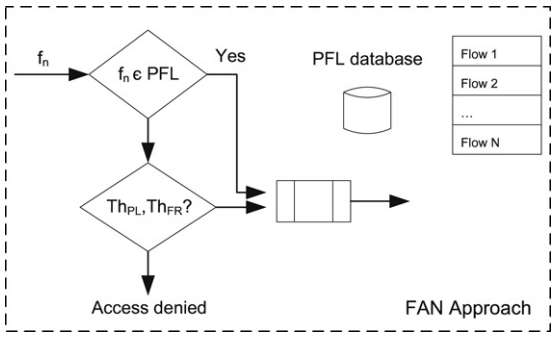

Fig. 2. FAN admission control flow diagrams.

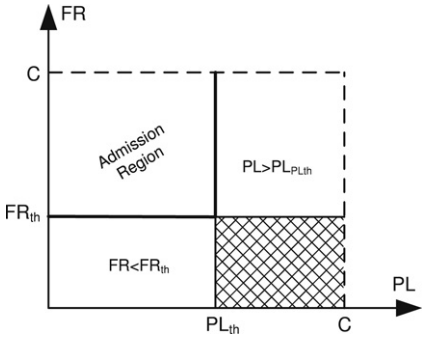

Fig. 3. Proactive Flow-based admission control policy.

otherwise, it is rejected. Although flows already accepted are somehow protected, only those flows which transmit at a lower rate than FR_{Th} are treated as streaming flows (high-priority). All the others are considered as elastic flows and receive less preference. This is done in order to avoid flows which abuse the system resources. Finally, Per-flow fair queuing scheduling algorithms are used to give preference to streaming over elastic flows.

4.2. Flow scheduling algorithms

Currently, there are two per-flow priority fair queuing algorithms proposed in 2GFAN architectures (on high capacity links fair queuing is enough to guarantee low packet delay and loss for real-time flows). Priority Fair Queuing (PFQ) and Priority Deficit Round Robin (PDRR). They have one priority queue and a secondary queuing system. In addition, an Active Flow List (AFL) is maintained by each queuing system. 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. This list may be thought to pose scalability problems. However, as shown in [30], this is not the case, and 2GFAN scales well.

4.2.1. Priority Fair Queuing (PFQ)

PFQ as defined in [19] is based on the Start-time Fair Queuing algorithm [31] and is used to give preference to streaming over elastic flows. Basically, PFQ is a PIFO (Push

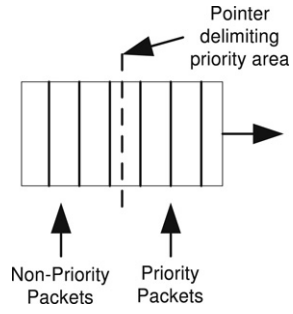

Fig. 4. Priority Fair Queue system (PFQ).

In First Out) queue, which stores packet information (flow identifier, size and memory location) and a time stamp, the latter determined by the SFQ algorithm. The PFQ queue is split into two areas delimited by a priority pointer (see Fig. 4), whereby streaming flows are temporarily 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 $O(\log(N))$, where N is the number of active flows on the scheduler.

4.2.2. Priority Deficit Round Robin (PDRR)

Priority Deficit Round Robin (PDRR) policy, as defined in [20] inherits the $O(1)$ complexity and fairness properties of DRR while improving latency by the use of a priority queue for low rate flows. PDRR is split into two queuing systems (see Fig. 5), a priority FIFO queue and a DRR queuing system (one FIFO queue per flow). The priority policy of PDRR is strict and exhaustive (like in PFQ). Streaming flows are enqueued in the priority queue, and the elastic flows are enqueued individually in the DRR queuing system. In addition, the AFL stores data for flows that have packets in the queue. These data include the flow identity, the current deficit count DC_i, flow quantum Q_i and pointers realizing a FIFO linked list of queued packets for that flow. An additional parameter $ByteCount(i)$ is used to determine whether flow packets should or should not be sent to the priority queue. AFL entries are visited in a certain order in each scheduling round. This order is defined by a pointer in each AFL entry indicating the next flow to be visited. A packet arriving at an active flow will be given priority while $ByteCount(i) \leq Q_i$; otherwise it is placed at the end of the flow queue. These operations (and the removal of inactive flows from AFL) ensure that packets of flows emitting less than one quantum per round realize low packet latency.

4.3. Flow scheduling performance measurements

To estimate priority load 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

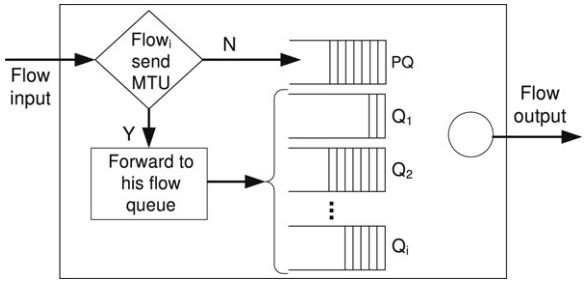

Fig. 5. Priority Fair Queue System (PDRR).

measurement period (in seconds) and C the link capacity. An estimation of the priority load is:

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

To estimate the fair rate consider that a virtual flow emitting single byte packets is inserted between real packets in an order dictated by the scheduling algorithms. For PFQ, in a busy period, the number of bytes transmitted by the queue is given by the evolution of the *virtual_time*. 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 fair rate for PFQ is:

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

For PDRR, by considering a flow continuously backlogged with a quantum of MTU (Maximum Transmission Unit), the fair rate is obtained by dividing the number of bytes this flow should transmit over the measurement period. If the number of bytes (fairBytes) that the virtual flow should send during the time interval (t_1, t_2) (incremented by MTU each time each time the virtual flow arrives), the estimation of the fair rate for PDRR is:

$$FR = \frac{\max\{S \times C, fairBytes \times 8\}}{(t_2 - t_1)}. \quad (3)$$

Exponential filters are applied after both measures.

5. Previous results and FAN extensions

We have evaluated 2GFAN (PFQ) against DS [23,24]. Our results shown that for a given average job size, 2GFAN (PFQ) enables lower average access delays than DS. Also, we observed that for a given offered load, the benefit of 2GFAN (PFQ) over DS is even more noticeable in presence of cross traffic. At the opposite, for a given offered load and a given average job size, the achievable average goodput is lower for 2GFAN (PFQ) than for DS but for high job size, the achievable goodput remaining stable under 2GFAN (PFQ) while it collapses under DS and is accentuated in the presence of cross-traffic. We conclude that 2GFAN (PFQ) is a very good candidate for IP-based Grid services provisioning. Moreover, in [25], we compared the two 2GFAN architectures. Our results shown that PFQ ensures better QoS performances than the PDRR system even if the last gives better rejection rate and its computational complexity is inferior. This is because the admission control protects the accepted flows but alone does not guarantee the QoS performances.

5.1. Optical grids and extensions of FAN to IP over WDM environments

Grid services demand multigranular architectures for QoS provisioning [39]. Therefore, the Optical Burst Switching (OBS) paradigm has been considered a good approach by the Grid community (e.g. GOBS) [41,4]. Based on these facts, we proposed in other works a multigranular architecture called Multi-layer FAN (MFAN) [32,33]. MFAN is an extension of 2GFAN (PFQ) architecture over a WDM environment by including an optical layer upon congested IP layer. Three different policies concerning the choice of which flows are moved to the optical layer were analyzed. The simulations show that the best possible choice, in terms of delay and goodput experienced by the flows, is to switch the most-active and oldest 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. Currently, we are evaluating MFAN architecture under Grid traffic (e.g. GoMFAN).

5.2. Extensions of FAN to Grid environments

The most important development for QoS provisioning in Grid networks has been GARA [6]. The last version of GARA propose several resource managers: at network level the DiffServ architecture, the CPU scheduling algorithm is the DSRT (Dynamic Soft Real Time), the disk access algorithms are DPSS (Distributed Parallel Storage Server) or GRIO (Guaranteed Rate I/O). Therefore, FAN is intended to contribute in GoFAN as network resource manager. To this end our work will be extended in the future taking into account more resource managers as those mentioned above and perhaps some others. In [22] we presented the firsts steps towards this goal. We proposed how FAN can be adapted to the Grid environment, an approach that we call Virtual FAN (VFAN) and intends to virtualize FAN routers. Since GARA manages admission control, scheduling, and configurations for Grid resources, including network resources then FAN could help GARA to manage admission control for Layer 3 services. An algorithm missing in the [5] proposal. Also, GARA implements advanced and online resource reservations, here the PFL of FAN can help to accomplish these tasks at the IP level.

5.3. Related work

FAN architectures have been tested [34], patented [35,36], standardized [37] and commercialized.[2] In addition, in [34] authors compare flow-based and packet-based routers; flow-based approach offers enhanced

[2] http://www.anagran.com/.

Please cite this article in press as: C. Cárdenas, et al., Admission control in Flow-Aware Networking (FAN) architectures under GridFTP traffic, Optical Switching and Networking (2008), doi:10.1016/j.osn.2008.05.003

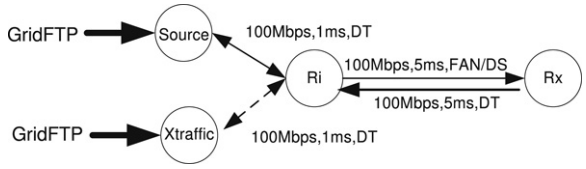

Fig. 6. Simulation topology.

performance in terms of packet processing. Also, in [38], the authors show that flow level bandwidth guarantees are achievable with two of their proposed admission control schemes, they achieved an order of magnitude in jitter and latency in individual flows. Even more recently FAN architectures have received more attention from the Grid community.[3], [4] All the above show that FAN is a promising approach for Voice over IP (VoIP), Multimedia over IP (MMoIP) and for Grid over IP (GoIP).

6. Experimental setup

6.1. Topology

Our simulation topology is single domain. It consist of an access router connected to an egress router through a bottleneck link of 100 Mbps and 5 ms. A GridFTP source is connected to the access router through links of 100 Mbs and 1 ms; a similar GridFTP source was used as a cross traffic in other studies [25,24]. In the bottleneck link, the outbound queue is based either on FAN and/or on DS. The inbound queue is drop tail (DT). Access queues are DT in both directions. To make fair comparisons DS was configured to mimic FAN. Fig. 6 shows our simulation topology.

6.2. Grid traffic vs. Internet traffic

In this section we provide some reasons why Grid traffic differs from Internet traffic, then we explain how we considered Grid traffic in our experiments. First, to the best of our knowledge, no Grid traffic model had been published when we executed our experiments [39]. In general, Grid traffic consists of short (Grid Service calls) and bulk data transfers, which may be very large compared to Internet traffic. Moreover, Grid applications have a larger probability of showing some workflow aspects than Internet applications. In Grid environments a set of nodes participates towards a common goal and they are expected to remain available for a long time. Grid applications may be able to specify their communication processes in advance therefore a node may know (in advance) when another node will send something. In Grid networks, a scheduler decides where applications go. Finally and perhaps the most important difference in Grid environments is that Grid traffic is mostly generated by machines and not by humans. On the other hand, in [21] it has been shown that Internet traffic at the packet level can be approximated by a self-similar stochastic process (the probability distribution function of the job size follows a Pareto law and arrivals are correlated).

6.2.1. Grid traffic over FAN (GoFAN) characterization

Our Grid traffic model is based in the fact that the software architecture of Globus Tool Kit (GTK), the most used software platform within the Grid community, offers a transport service named GridFTP [40] which consists of sending a set of parallel TCP connections per Grid session at the same time. Because of the lack of a Grid traffic model, in our previous studies and in this work we assumed that our Grid traffic is composed only by GridFTP sessions that arrive following a stationary Poisson process with intensities of 5, 10, 15 and 20 arrivals per minute. We assume that Grid job sizes follow an exponential distribution with mean of 100 MB, the average packet size being 1000 Bytes.

6.3. Operation and management policies

GridFTP configuration is end-user specific, the authors in [42] have shown that throughput between 90% and 95% can be reached using between 4 and 6 parallel TCP connections, independently of the loss policy. Therefore, we decided to keep per-flow loss policy. In operational networks every time a GridFTP session arrives the number of parallel TCP connections vary. To evaluate the impact of the number of parallel TCPs we assume its number is equal for all GridFTP sessions during the simulation but we test two bounds (3 and 9 parallel TCP/IP flows). We assume that job sizes are divisible. We decide to apply a policy of equal quantity per-flow within a GridFTP session. Also, we applied a total GridFTP session admission policy instead of partial admission. Furthermore, a single per-flow scheduling policy was applied. In FAN, the FR threshold was configured with the value of 0.25 and the PL threshold with the value of 0.8. To simplify the configuration of FAN we considered both estimation periods of identical value of 100 ms [19]. The maximum TCP window size was of 5000 packets.

6.4. DiffServ configuration

We chose two physical queues and two virtual queues. Scheduling is configured as strict priority (like in FAN). The policer (smoother) consists in a Time Slide Window with Two Color Marking (TSW2CM). The Committed Information Rate (CIR) is equal to the FR estimator of the FAN and updated at the same time interval (100 ms). The packet rejection probability is estimated with the size of every virtual queue (RIO-D). RED parameters are fixed at 0.6 and 0.8 of each virtual queue size and the maximal probability is 0.5. The default queue weight is 0.002. In this DS configuration, packets that do not meet CIR are deprecated to the second virtual queue (they lose priority). In FAN, an accepted flow sending more than FR is deprecated to second priority.

[3] "Research Consortium Demostrates High Performance Flow Switching Network Enabled by Anagran at SC07 Conference", Anagran news, November 12, 2007.

[4] "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.

Please cite this article in press as: C. Cárdenas, et al., Admission control in Flow-Aware Networking (FAN) architectures under GridFTP traffic, Optical Switching and Networking (2008), doi:10.1016/j.osn.2008.05.003

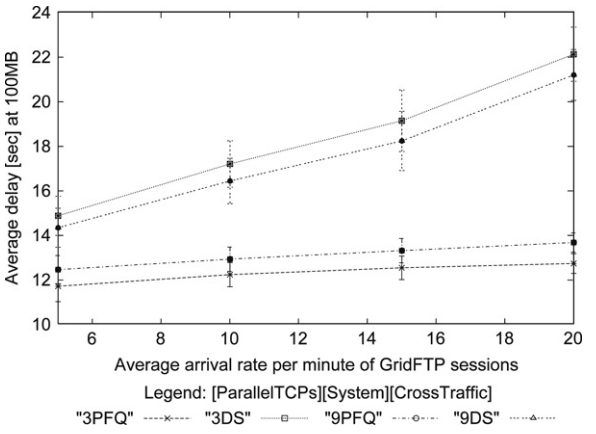

Fig. 7. Average delay of GridFTP sessions with average job size of 100 MB.

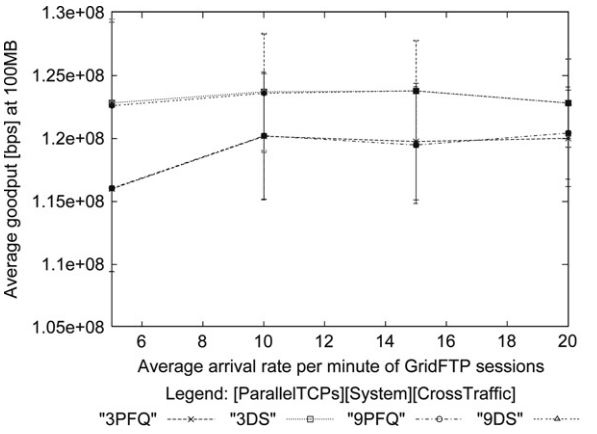

Fig. 8. Average goodput of GridFTP sessions with average job size of 100 MB.

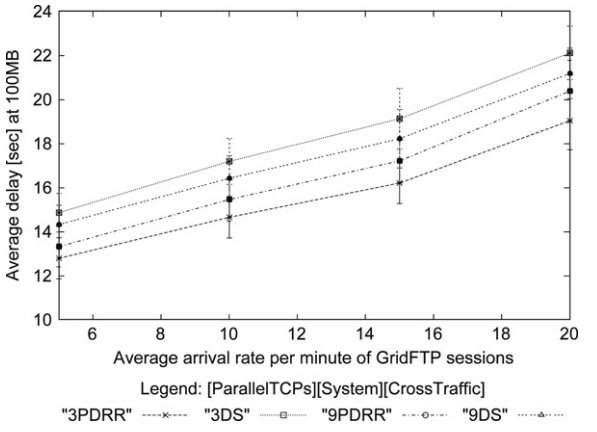

Fig. 9. Average delay of GridFTP sessions with average job size of 100 MB.

6.5. Experiments

Simulations were run using *ns-2*.[5] Users of Grid networks commonly reserve in advance for different durations. Therefore, we run discrete time simulations assuming one hour (3600 s) of user reservation time. We checked that the first 5 min of each simulation run corresponds to the transient period for reaching the equilibrium regime. As in previous studies, arrival intensities range from [0, 20] arrivals per minute of GridFTP sessions. Because of this small arrival rate, thirty replications were carried out per scenario. We used the inverse method based on time discretization to generate the Poisson process. In terms of simulation challenges, the fact that Internet traffic job size follows a Pareto size demands extremely long-run simulations [27] while assuming a exponential probability distribution function for Grid job sizes reduce this complexity, mainly in the number of simulation runs. Simulation experiments were executed in ns-2.31 under a multiprocessor (SMP) computer with four Intel Xeon at 3.00 GHz and OS Debian 2.6.15.

7. Experiments results

7.1. 2GFAN (PFQ) vs DS

When comparing 2GFAN (PFQ) against DS on the average delay of GridFTP sessions our results show that 2GFAN (PFQ) enables the best performance in terms of average GridFTP session delay with both bounds on parallel TCP flows. The larger the number of parallel flows the longer the average delay. On the contrary for DS, the larger the number of parallel flows the shorter the average GridFTP session delay. Also, the greater the offered load, the greater the advantage of 2GFAN (PFQ) over DS. Two factors impact on this behavior, one is the number of secondary queues 2GFAN offers to the demand, the other is the admission control which rejects more GridFTP sessions protecting those already accepted. 2GFAN offer better average delay of GridFTP sessions. See Fig. 7.

When comparing both systems in the average goodput of GridFTP sessions, the DS system gives better results. Similar results were observed for both bounds in the number of parallel TCPs. Nevertheless, we can observe that this advantage is reduced as the offered load is increased. From other results [24], 2GFAN (PFQ) outperforms DS when the offered load is multiplied by five and/or a background traffic that doubles the average arrival rate is injected. In both metrics, the greater the number of parallel TCP flows the lower the DS behavior. See Fig. 8.

7.2. 2GFAN (PDRR) vs DS

When comparing 2GFAN (PDRR) against DS on the average GridFTP session delay, we observe that 2GFAN (PDRR) outperforms DS. When the number of parallel TCP flows is increased an increase in 2GFAN (PDRR) and a reduction in DS are observed. Also, as the offered load is increased the advantage of 2GFAN (PDRR) over DS is more or less maintained. These results can be explained by the fact that 2GFAN (PDRR) assigns a queue per accepted flow even if the flow pertains to a GridFTP session giving better results in the average delay. See Fig. 9.

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

Please cite this article in press as: C. Cárdenas, et al., Admission control in Flow-Aware Networking (FAN) architectures under GridFTP traffic, Optical Switching and Networking (2008), doi:10.1016/j.osn.2008.05.003

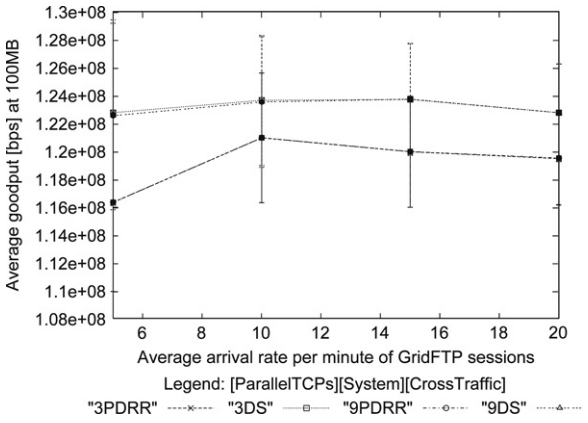

Fig. 10. Average goodput of GridFTP sessions with average job size of 100 MB.

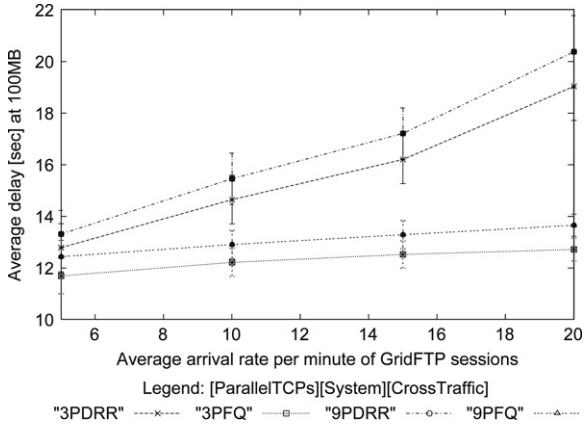

Fig. 11. Average delay of GridFTP sessions with average job size of 100 MB.

On the other hand, similar to 2GFAN (PFQ) against DS, 2GFAN (PDRR) gets lower average GridFTP goodput than DS. Nevertheless, in other results [24], 2GFAN (PDRR) also outperforms DS when the offered load is multiplied by five and the arrival rate of GridFTP sessions is doubled. See Fig. 10.

7.3. 2GFAN (PFQ) vs 2GFAN (PDRR)

When comparing both FAN architectures. 2GFAN (PFQ) outperforms 2GFAN (PDRR) in the average GridFTP session delay. When the number of parallel TCP flows is increased a similar advantage is obtained in both systems but the larger the offered load the greater the 2GFAN (PFQ) advantage. The latter architecture keeps the average GridFTP session delay more stable. See Fig. 11. The goodput remains similar for both systems. See Fig. 12.

Lastly, when comparing both architectures in the average GridFTP session rejection rate, the 2GFAN (PDRR) outperforms the 2GFAN (PFQ) architecture. This behavior is explained because 2GFAN (PDRR) is a queue per flow accepting more flows but giving lesser performance on delay and goodput measures than 2GFAN (PFQ). See Fig. 13.

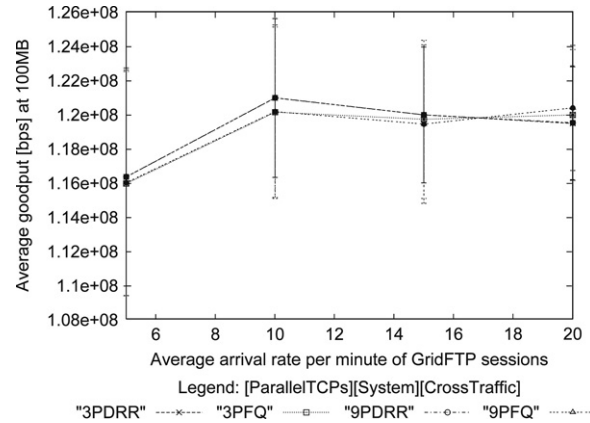

Fig. 12. Average goodput of GridFTP sessions with average job size of 100 MB.

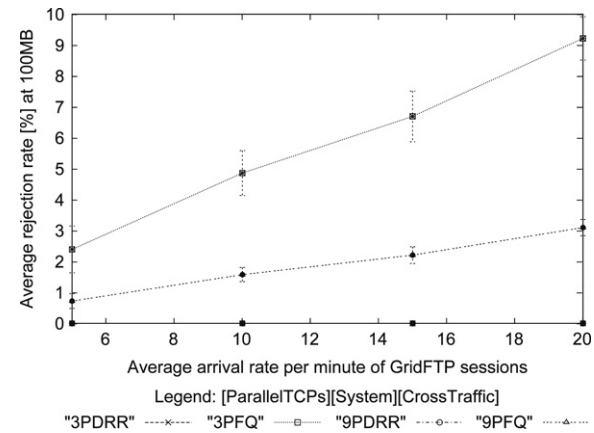

Fig. 13. Average rejection rate of GridFTP sessions with average job size of 100 MB.

8. Conclusions

The objective of this paper was to compare the GoFAN and GoDS under GridFTP traffic. As a first step, we have provided a brief survey on QoS in Grid environment and on QoS in IP networks. We have also described in detail the FAN architecture and its evolutions. We have then compared via computer simulations the suitability of the DS and 2GFAN architectures applied at IP access routers for the Grid environment. We conclude that 2GFAN architectures outperform DS in the average GridFTP session delay and the average GridFTP session goodput with increasing offered load. At this point, the admission control algorithm of 2GFAN architectures offers advantages over DS. Among both 2GFAN architectures, the one based on the PFQ algorithm offers better performance in terms of average GridFTP session delay and goodput than the one based on PDRR. Meanwhile, the performance of PDRR-based 2GFAN and of PFQ-based 2GFAN in terms of average rejection rate of GridFTP sessions are quite similar. For increasing average job size, PDRR provides the lowest performance. Such a difference can be explained by the fact that PFQ protects the priority flows better than PDRR. Since PDRR opens a queue per accepted flow, the service time available for each

Please cite this article in press as: C. Cárdenas, et al., Admission control in Flow-Aware Networking (FAN) architectures under GridFTP traffic, Optical Switching and Networking (2008), doi:10.1016/j.osn.2008.05.003

queue decreases if the number of flows increases. As a general conclusion, the previous results show that FAN is a promising approach for QoS in Grid environments and that the PFQ scheduling discipline is the best suitable, in spite of its greater computational complexity in comparison to PDRR scheduling. We are planning to proceed to a more fair comparison of DS against 2GFAN by adding admission control to DS. We also intend to extend the 2GFAN architectures by considering scheduling algorithms in the perspective of a comparison with GARA (a.k.a. Pre-GRAM).

Acknowledgements

The authors would like to thank Abdesselem Kortebi from France Telecom R&D for his support in the FAN implementation and to Mr. Jim Roberts and Mrs Sara Oueslati from France Telecom R&D for our fruitful discussions.

References

[1] I. Foster, C. Kesselman, The Grid 2: Blueprint for a New Computing Infrastructure, Morgan Kaufmann, 2003.
[2] T. Murase, H. Shimonishi, M. Murata, Overlay network technologies for QoS control, IEEE/ACM Transaction on Networking, Special Section on Networking Technologies for Overlay Networks, Invited Paper, E89-B, N.9, 2006.
[3] S. Wright, Admission control in multi-service IP networks: A tutorial, IEEE Communications Surveys & Tutorials (2nd Quarter) (2007).
[4] F. Travostino, J. Mambretti, G. Karmous-Edwards, Grid Networks: Enabling Grids with Advanced Communication Technology, Wiley, 2006.
[5] D. Menasce, E. Casalicchio, Quality of service aspects and metrics in grid computing, in: Comp. Measurement Group Conf., 2004.
[6] I. Foster, A. Roy, V. Sander, A quality of service architecture that combines resource reservation and application adaptation, in: 8th International Workshop on Quality of Service, IWQOS.
[7] S.N. Bhatti, S. Sorensen, P. Clark, J. Crowcroft, Network QoS for Grid Systems, International Journal of High Performance Computing Applications 17 (3) (2003).
[8] R. Al-Ali, S. Sohail, O. Rana, A. Hafid, G. von Laszewski, K. Amin, S. Jha, D. Walker, Network QoS provision for distributed grid applications, International Journal of Simulation Systems, Science and Technology 5 (5) (2004).
[9] D. Adami, S. Giordano, M. Repeti, M. Coppola, D. Laforenza, N. Tonellotto, Design and implementation of a grid network-aware resource broker, in: 24th IASTED International Conference on Parallel and Distributed Computing and Networks, PDCN, 2006.
[10] V. Sander, I. Foster, A. Roy, L. Winkler, A differentiated services implementation for high-performance TCP flows, Computer Networks 34 (6) (2000) 915–929.
[11] I. Foster, M. Fidler, A. Roy, V. Sander, L. Winkler, End-to-End quality of service for high-end applications, Computer Communications 27 (14) (2004) 1375–1388.
[12] J. Leigh, O. Yu, A. Verlo, A. Roy, L. Winkler, T. DeFanti, Differentiated services experiments between the electronic visualization laboratory and argonne national laboratory, EMERGE Report.
[13] M. Rio, A. di Donato, F. Saka, N. Pezzi, R. Smith, S. Bhatti, P. Clarke, Quality of service networking for high performance grid applications, Journal of Grid Computing 1 (4) (2003) 329–343.
[14] P. Vicat-Blanc Primet, F. Echantillac, M. Goutelle, Experiments with equivalent differentiated services in a grid context, Future Generation Computer Systems 21 (4) (2005) 515–524.
[15] P. Vicat-Blanc Primet, F. Chanussot, End to End network quality of service in grid environments: The QoSINUS approach, Broadnets (2004).
[16] L. Lefevre, C. Pham, P. Primet, B. Tourancheau, B. Gaidioz, J. Gelas, M. Maimour, Active networking support for the grid, in: The IFIP-TC6 Third International Working Conference on Active Networks, 2001.
[17] K. Yang, X. Guo, A. Galis, B. Yang, D. Liu, Towards efficient resource on-demand in Grid Computing, ACM SIGOPS Operating Systems Review 37 (2) (2003) 37–43.
[18] J. Roberts, S. Oueslati, Quality of service by flow aware networking, Philosophical Transactions of The Royal Society of London Series A 358 (1773) (2000) 2197–2207.
[19] A. Kortebi, S. Oueslati, J.W. Roberts, Cross-protect: Implicit service differentiation and admission control, in: IEEE HPSR, 2004.
[20] A. Kortebi, S. Oueslati, J. Roberts, Implicit service differentiation using deficit round robin, in: ITC19, 2005.
[21] W.E Leland, S.T. Murad, W. Willigner, D.V. Wilson, On the self-similar nature of Ethernet traffic 1993, ACM/SIGCOMM.
[22] C. Cardenas, Gagnaire, VFAN: Extension of the Flow-Aware Networking (FAN) architecture to the Grid environment, 2007, 21ème Congrès DNAC - Les évolutions des réseaux IP (DNAC'07), Paris, France, November 14–16, 2007.
[23] C. Cardenas, M. Gagnaire, V. Lopez, J. Aracil, Admission control for Grid services in IP networks, in: 1st IEEE Symposium on Advanced Networks and Telecommunications Systems, ANTS'07, Bombay, India, December 7–11, 2007.
[24] C. Cardenas, M. Gagnaire, V. Lopez, J. Aracil, Performance evaluation of the Flow-Aware Networking (FAN) architecture under Grid environment, in: 20th IEEE/IFIP Network Operations and Management Symposium, NOMS'08, Salvador, Brazil, April 7–11, 2008.
[25] C. Cardenas, M. Gagnaire, Performance comparison of the Flow-Aware Networking (FAN) architectures under GridFTP traffic, 2008, 23rd ACM/SIGAPP Symposium on Applied Computing, SAC'08, Fortaleza, Ceara, Brazil, March 16–20, 2008.
[26] K. Keahey, I. Foster, I. Freeman, X. Zhang, Virtual workspaces: Achieving quality of service and quality of life in the grid, Journal of Scientific Programming, Special Issue: Dynamic Grids and Worldwide Computing 13 (4) (2005) 265–276.
[27] M.E. Crovella, A. Bestravos, Self-similarity in World Wide Web Traffic: Evidence and possible causes, IEEE/ACM Transaction on Networking 5 (6) (1997).
[28] S. Oueslati, J. Roberts, A new direction for quality of service: Flow-aware networking, EuroNGI 2005, 2005.
[29] J. Roberts, A survey on statistical bandwidth sharing, Computer Networks, International Journal of Computer and Telecommunications Networking 45 (3) (2004) 319–332.
[30] A. Kortebi, L. Muscariello, S. Oueslati, J. Roberts, Evaluating the number of active flows in a scheduler realizing fair statistical bandwidth sharing, SIGMETRICS Performance Evaluation Review 33 (1) (2005) 217–228.
[31] P. Goyal, M.V. Harrick, H. Chen, Start-time fair queueing: A scheduling algorithm for integrated services packet switching networks, IEEE/ACM Transactions on Networking 5 (5) (1996).
[32] V. Lopez, C. Cardenas, J. Hernandez, J. Aracil, Extension of the Flow-Aware Networking architectures to the IP over WDM environment, 2007, in: 4th. IEEE/QoS-IP (IT-NEWS), February 2008, Venezia, Italy.
[33] V. Lopez, C. Cardenas, J. Hernandez, J. Aracil, Multilayer Flow-Aware Networking, Elsevier Journal of Computer Networks, (2007) (Submitted).
[34] J. Park, M. Jung, S. Chang, S. Choi, M. Young Chung, 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.
[35] S. Oueslati, J. Roberts, Method and device for implicit differentiation of quality of service in a network, FR2854296, EP1478140A1, US2004/0213265A1, 2003.
[36] J. Roberts, S. Oueslati, A. Kortebi, Procede et dispositif d'ordonnancement de paquets pour leur routage dans un réseau avec détermination implicite des paquets à traiter en priorité, FR2878106, EP1813081, WO2006051244, 2006.
[37] ITU-T E.417: Framework for the network management of IP-Based networks.
[38] F.O. Sem-Jacobsen, Sven-A. Reinemo, T. Skeie, O. Lysne, Achieving flow level QoS in cut-through networks through admission control and diffserv, in: International Conference on Parallel and Distributed Processing Techniques and Applications, PDPTA, 2004.
[39] S. Volker, et al., GFD-I.037, Networking Issues for Grid Infrastructure, 2004.
[40] I. Mandrichenko, et al., GFD-47, GridFTP v2 Protocol Description, 2005.
[41] R. Nejabati, et al., GFD-128, Grid Optical Burst Switched Networks (GOBS), 2008.
[42] E. Altman, D. Barman, B. Tuffin, M. Vojnic, Parallel TCP sockets: Simple model, throughput and validation, Technical Teport MSR-TR-2005-130. 2005; shorter version presented at IEEE Infocom 2006.
[43] S. Giordano, S. Salsano, S. Van den Berghe, G. Ventre, D. Giannakopoulos, Advanced QoS provisioning in IP networks: The European premium IP projects, IEEE Communications Magazine 41 (1) (2003) 30–36.

Please cite this article in press as: C. Cárdenas, et al., Admission control in Flow-Aware Networking (FAN) architectures under GridFTP traffic, Optical Switching and Networking (2008), doi:10.1016/j.osn.2008.05.003