Working Group: G. Bianchi
Internet Draft University of Palermo, Italy
N. Blefari-Melazzi

Document:

University of Perugia, Italy

draft-bianchi-blefari-admcontr-over-af-phb-00.txt

Category: Informational March 2000

Per Flow Admission Control over AF PHB Classes

Status of this Memo

1. Abstract

This memo shows that an AF PHB class, as defined in RFC 2597, is capable of supporting explicit per flow admission control. Admission control is obtained by a suitable use of two out of the three drop precedence levels. Level 1 is dedicated to conforming accepted traffic, i.e. traffic that has passed admission control test and conforms to edge-policing functions. Level 2 is used as an implicit signaling pipe and acts as the core network support for the admission control procedure.

Table of Contents

2. Introduction

There is a growing feeling [RFC2990, RFC2998] that the basic DiffServ architectural model [RFC2474, RFC2475] lacks the capability of providing service accuracy. Quoting [RFC2990] both the Integrated Services architecture and the Differentiated Services architecture have some critical elements in terms of their current definition which appear to be acting as deterrents to widespread deployment... There appears to be no single comprehensive service environment that possesses both service accuracy and scaling properties . Also, in [RFC2998], it is pointed out that further refinement of the QoS architecture is required to integrate DiffServ network services into an end-to-end service delivery model with the associated task of resource reservation .

To this purpose, [RFC2990] recommends to define an admission control function which can determine whether to admit a service differentiated flow along the nominated network path . In fact, without per flow admission control, prevention of overload in a given service class, e.g. by means of pure inter-domain service level agreements, does not appear to be an easy task. Upon overload in a given service class, all flows in that class suffer a potentially harsh degradation of service.

The Assured Forwarding Per Hop Behavior (AF PHB, [RFC2597]) has been devised to provide different levels of forwarding assurances. The example services presented in the appendix of [RFC2597] show that the primary intent of AF is to promote packet loss differentiation, either among different traffic classes, e.g., marked with different drop levels, as well as within the same traffic class, e.g. marking traffic conforming to a policy specification with a lower drop level than non conforming traffic.

While low loss and low latency traffic support, such as that achievable with a strict admission control function, appears to be out of the scopes of the AF model, the class-based structure of the AF PHB has intrinsic (unforeseen?) capabilities to support per flow admission control. Our key idea (originally proposed - although not specifically envisioned over an AF PHB class - in [BBM01], under the name GRIP - Gauge&Gate Reservation with Independent Probing) is to rely the decision to admit a new connection upon the successful

delivery of probes, tagged with a higher AF drop level mark than accepted traffic, independently generated by end points during each flow setup phase. It is quite interesting to remark that the idea of pushing traffic control to the edge and basing the connection request acceptance/refusal on packet loss detection is close to what TCP congestion control technique does, but it is used in the novel context of admission control.

3. Related work

Several literature proposals describe a promising novel framework for admission control, generally referred to as Endpoint Admission Control (EAC - see [BKS00] and references therein contained). In

EAC, the explicit decisions whether to accept or refuse a connection request are taken by edge devices, rather than by devices within the network (e.g. core routers or bandwidth brokers). The driving idea of EAC schemes is to convey the congestion status of network nodes to the end-points, either in an implicit manner (i.e. by means of endpoint traffic measurements [BOR99, ELE00]), or explicitly (e.g. by requiring that network devices mark packets in a way that is load dependent [KEL00]). Both approaches have their pitfalls. Endpoint measurements appear unreliable when performed over the very short time frame given by the flow setup time. Explicit congestion notification mechanisms, instead, require to actively involving core routers in the process of marking packets, an activity that appears contrary to the spirit of DiffServ. Moreover, to be effective, such solutions need to be supported by all DiffServ routers at once, and a uniform marking protocol needs to be specified, e.g. by identifying a set of bits (one or more) in the IP packet header.

An apparently un-related issue is how to perform admission control in a core router. We argue that this problem does not stay in the fact that the network has to be oblivious of individual flows. In fact recent literature has shown that admission control schemes driven by load measurements are extremely robust and efficient [GRO99, BJS00]. These schemes do not base the accept/reject decision on per-flow state information and related traffic specifications. Instead, they operate on the basis of per-node aggregate traffic measurements carried out at the packet level. The robustness of these schemes stays in the fact that, in suitable conditions (e.g. flow peak rates small with respect to link capacities), they are barely sensitive to uncertainties on traffic profile parameters. As a consequence, it seems that scalable decisions can be independently taken by the routers.

Our aim is to integrate the effectiveness of measurement-based

approaches with the simplicity of EAC. In other words, we want to implicitly convey the status of core routers (evaluated by means of aggregate measurements) to the end points so that the latter devices can take learned admission control decisions. The main open issue is then how to define and handle the signaling information needed to exchange information between end-points and core routers, without relying on explicit signaling and thus violating the DiffServ paradigm.

4. AF PHB class management and implicit signaling

Four AF PHB classes have been standardized, each composed of three drop levels. In what follows, we will use the notation AFxj to indicate packet marks belonging to the AF class x, with drop level j. Conforming to [RFC2597], within a class x, if i<j, the dropping probability of packets labeled AFxi is lower than that of packets labeled AFxj. Quoting [RFC2597], an AF implementation MUST detect and respond to long-term congestion within each class by dropping packets, while handling short term congestion (packet bursts) by queueing packets. This implies the presence of a smoothing or filtering function that monitors the instantaneous congestion level

and computes a smoothed congestion level. The dropping algorithm uses this smoothed congestion level to determine when packets should be discarded .

Let us now focus our attention to a specific module in charge of handling AF traffic belonging to a given class x. A particular implementation of the AF specification is depicted in Figure 1 (for simplicity, the drop level AFx3 is neglected until Appendix A).

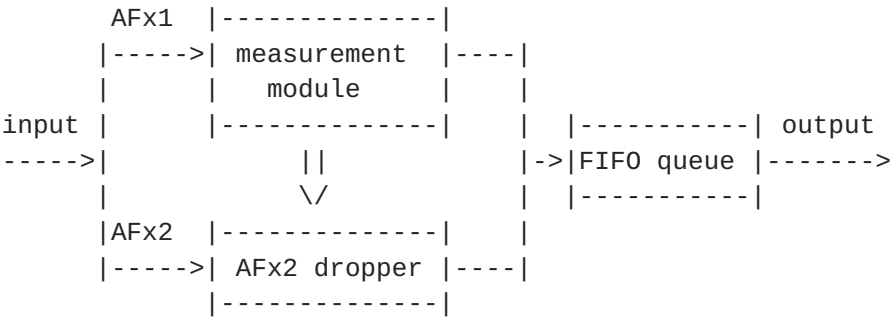

Figure 1: logical handling of an AF class x

A measurement module is devised to run-time measure the aggregate AFx1 traffic. The measurement module depicted in the figure does not interact with the AFx1 packets forwarding, i.e., these packets are forwarded to the FIFO buffer placed at the output regardless of the

measurements taken. On the basis of such measurements, this module triggers a suitable dropping algorithm on the AFX2 traffic. With respect to the general AF PHB operation, we are restricting the AFX2 dropping algorithm to depend only on AFX1 traffic measurements.

The simplest dropping algorithm is represented by a gate (smoother dropping algorithms for AFx2 packets - e.g. RED-like algorithms - may be considered to improve stability). When the measurement module does not detect congestion on the AFx1 traffic, being the notion of congestion implementation-dependent, it keeps the gate opened (we call this ACCEPT state). When the gate is open no AFx2 packet is dropped. Conversely, the measurement module keeps the gate closed (REJECT state) when congestion is detected, i.e. it enforces a 100% drop probability over AFx2 packets. Note that this operation does not violate the AF drop level relationship, as AFx1 dropping probability is lower than the AFx2 one.

While the above description is simply a particular implementation of an AF class, we now show its interpretation in terms of implicit signaling, which has important consequences for the definition of an overlay admission control function. In fact, let us assume that: i) the considered AF class x, is devoted to the support of QoS aware flows, requiring an admission control procedure; ii) traffic labeled AFx1 is generated by flows which have already passed an admission control test, iii) AFx2 packets are signaling packets injected in the network by flows during the setup phase (in principle, one AFx2 packet per flow).

According to the described operation, an AFx2 packet is delivered to its destination ONLY IF it encounters all the routers along the path in the ACCEPT state. This operation provides an implicit binary signaling pipe, semantically equivalent to a one-bit explicit congestion notification scheme.

5. End Point Admission Control operation

The above interpretation of the AF PHB allows deploying an admission control function, whose operation is concentrated in the flow end points, and which does not relies on state management and explicit signaling exchange in DiffServ network routers (both core routers within a DS domain and border routers among different DS domains).

The admission control function is triggered by explicit signaling exchanges between the source terminal (SRC in figure 2) and the access router (ARin) to the local DS domain. Assume that a one way connection is offered from user terminal SRC to user terminal DST,

in principle belonging to different DS domains (see figure 2). Upon setup request, the ARin node starts a connection setup attempt by sending, in principle, just one packet labeled AFx2 through the network. In the same time, a probing phase timeout is started.

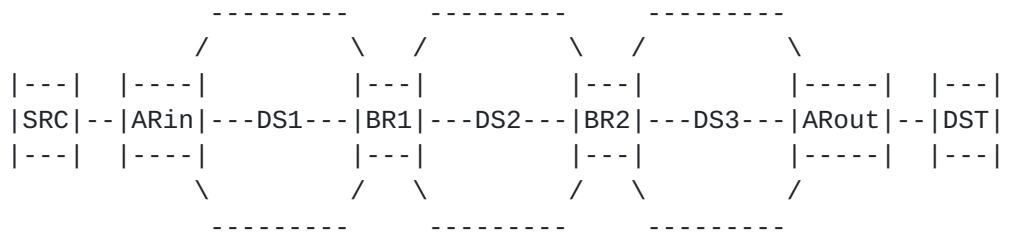

Figure 2: Sample Network Configuration

We recall that, when an AFx2 packet reaches the destination, it implicitly conveys the information that all routers encountered across the path have been locally declared themselves in the ACCEPT state, i.e. capable of admitting new flows without QoS degradation. It is up to each specific router/domain to quantitatively determine its QoS target, and link this QoS target to the runtime measurements carried on the AFx1 traffic (see Appendix C for further discussion).

If the AFx2 packet eventually reaches the ARout destination node, an explicit signaling is executed between the Arout and the DST node to verify that the DST terminal is capable (and willing) of accepting the call. We remark that a by-product of this operation is to allow a receiver capability negotiation function procedure, which is deemed necessary in [RFC2990].

Once this test is performed with successful response, the ARout router replies with a feedback packet, labeled AFx1. The choice of the AFx1 mark to transmit the feedback packet is motivated by the fact that, for one way flows, it is not necessary to test the reverse path (from DST to SRC), and thus the AFx1 tag provides better guarantees that the feedback packet is correctly received at

the SRC. Indeed, when full duplex flows are considered, the feedback packet shall be tagged as AFx2, since in this case reverse path testing is also necessary.

The decision whether to admit or reject the incoming call is driven by the eventual reception of the feedback. When a feedback packet is received in response, the source node labels the flow as accepted, and starts transmitting packets marked as AFx1. Conversely, by not receiving a feedback packet within the probing phase timeout, the source node is made able to implicitly determine that at least one

router along the path has declared itself not capable of accommodating additional flows, and thus the source node can abort the flow setup attempt (or reiterate the setup attempt according to some suitable backoff mechanism).

6. Appendix A: Possible roles of the AFx3 level

In the above description, the AFx3 drop level appears in principle unnecessary. However, it may be convenient to use this level. A first possibility is that the AFx3 level be used to mark non-conforming packets, which can be eventually delivered if network resources are available. Second, AFx3 packet marking can be enforced over flows that have not successfully passed the described admission control test. This allows deploying a service model where high QoS is provided to flows that pass the admission control test, while best effort delivery is provided to initially-rejected flows. These latter flows may occasionally retry the setup procedure, by simply marking occasional packets as AFx2 (e.g. by adhering to a suitable backoff procedure), and may eventually receive the upgraded AFx1 marking when network resources become available (as testified by the eventual reception of an AFx1 feedback).

The usage of the AFx3 level as described above is targeted to increase the link utilization. However, [RFC2597] requires the drop probability for AFx3 to be greater (or at most equal) than AFx2. This implies that the link utilization is bounded by the possibly strict mechanism that triggers AFx2 packets dropping: when AFx2 packets receive a 100% dropping probability, all AFx3 packets must also be dropped to conform to the [RFC2597] specification. A more effective mechanism would consist in implementing a dropping algorithm for the AFx3 traffic not directly related to the AFx2 drop algorithm. However, this usage of the AFx3 level does not conform with the AF specification, since the AFx3 dropping probability may be eventually lower than AFx2.

A more interesting possible usage of the AFx3 level consists in providing a second control (probing) channel in addition to AFx2. According to this solution, AFx1 traffic measurements trigger a dropping algorithm on the AFx3 traffic, with stricter dropping conditions than the AFx2 dropping algorithm (i.e. AFx3 packets are assumed to detect congestion, and notify it via packet drop, before AFx2 packets). This AFx3 probing class could request admission for flows with e.g., higher peak rate and bandwidth requirements than flows supported via the AFx2 probing class (i.e., we are adding a

second implicit signaling pipe). Also, being the AFx3 channel more reactive to congestion conditions, its usage can be envisioned to

provide lower access priority to network resources. This would improve fairness and avoid some kinds of sources to steal a large part of network resources.

7. Appendix B: Arguments for new PHB definitions?

Although this memo leaves untouched the basic AF PHB semantic, we feel that our suggested usage of AF is different (and quite unexpected) from what intended in RFC 2597. The services that are expected to make use of admission control are RTP/UDP streams with delay and loss performance requirements, whose support is currently envisioned by means of the EF PHB. On the contrary, AF appears designed to provide better than best effort support for generic TCP/UDP traffic. Thus, our study raises the case for the transformation of the (single) EF PHB into a PHB class (i.e. by adding an associated, "paired", probing pipe with a different DSCP). An alternative is defining new "paired" PHBs.

On a different prospective, paired PHBs can be envisioned to support more general control functions than admission control. For example, the TCP fast retransmission and recovery algorithm might take advantage of isolated data packets labeled as control , and thus expected to encounter loss if (controlled) congestion is encountered in the network.

8. Appendix C: Performance issues

The described admission control semantic provides a reference framework compatible with "current" AF implementations, although poor performance are most likely provided over RED-like queueing management mechanisms. An explicit traffic measurement module implementation appears thus necessary.

Quantitative and tunable performance may be independently provided and specified by each administrative entity. Uniform implementation across a specific domain allows defining a quantitative view (e.g., a PDB) of the performance achievable within the considered domain. In this way, the refinements deemed necessary in [RFC2990] to provide service accuracy in the DiffServ architectural model can be considered as accomplished.

In fact, the performance achievable by the described end point admission control operation depends on the notion of congestion as the triggering mechanism for AFx2 packet discarding, which is left to each specific implementation. Each administrative entity may arbitrarily tune the optimal throughput-delay/loss operational point supported by its routers, by simply determining the aggregate AFx1 traffic target supported in each router. The mapping of AFx1 throughput onto loss/delay performance in turns depends on the link capacities and on the traffic flow characteristics offered on the AFx class.

With this approach, it is possible to construct PDBs offering quantitative guarantees. A building block of such PDBs is the definition of specific measurement modules and AFX2 dropping algorithms.

A generic dropping algorithm is based on suitable rules (or decision criteria). An example of a trivial decision criterion is to accept all AFx2 packets when the measured throughput is lower than a given threshold and reject all AFx2 packets when the AFx1 measurements overflow this threshold. The resulting delay performance depends upon the link capacity and the traffic model. For instance, with 32 Kbps peak rate Brady on-off voice calls offered to a 2 Mbps (20 Mbps) link, a target link utilization of 75% (92%) leads to a 99th percentile per hop delay lower than 5ms - see figure 4 in [BCP00].

Tighter forms of traffic control are possible. As a second example of a decision criterion, we demonstrated that hard (loss and/or delay) QoS guarantees can be provided, under suitable assumptions on the offered traffic (i.e., traffic sources regulated by standard Dual Leaky Buckets, as in the IntServ framework) and with ad hoc defined measurement modules in the routers [BBM01].

Finally, we note that the AFX2 dropping algorithm must not be necessarily driven by traffic measurements. In fact, it can be driven by lower layers QoS capabilities, (e.g., ATM).

9. Appendix D: Security considerations

As all admission control functions, our solution presents the risk of theft of resources through the unauthorized admission of traffic. Although, logically, user terminals are the natural nodes where the endpoint admission control should operate, this is clearly not realistic, for the obvious reason that the user may bypass the admission control test and directly send AFx1 packets. Identity authentication and integrity protection are therefore needed in order to mitigate this potential for theft of resources [RFC2990]. Administrators are then expected to protect network resources by configuring secure policers at interfaces (e.g. access routers) with untrusted customers. Similar protections must be provided at the interface between different domains. In particular, it may be necessary to restrict the access to the AF class(es) used for admission controlled traffic. For example, a DS domain should re-mark AF packets when they come from an un-trusted adjacent DS domain. In more generality, we remark that policing and conditioning rules enforced at the border routers of each domain depend on the

usage of the considered AF PHB class within the specific domain and thus have to be accounted of in the definition of each specific PDB supporting admission control.

A quite obvious security hazard is flooding the network with AFx2 packets. The objective is twofold. On one side, denial of service situations can be easily created, as a massive loading of the network with AFx2 packets prevent the setup of normal connection. On the other side, the goal might be to affect fairness: the continuous

transmission of AFx2 packets at a rate higher than normal connection requests is a mean to gain faster access to resources when these are made available by a router along the path. This implies that some form of traffic conditioning and policing is necessary over AFx2 streams. While it is simple to recognize an hard attack, by monitoring the AFx2 packets crossing an edge router (the AFx2 traffic ± at most a few packets per originating connection - is minimal in normal conditions, and thus sudden increments of the AFx2 load are suspicious), it may be not straightforward for DS boundary routers to recognize smoother fairness attacks. However, note that the same fairness problem is present also in more complex reservation mechanisms, such as RSVP (malicious users can continuously require setup to increase their access possibility with respect to normal users).

Finally, all the security considerations expressed in [RFC2990] apply also to our solution.

10. References

[BBM01] G. Bianchi, N. Blefari-Melazzi: " A Migration Path for the Internet: from Best-Effort to a QoS Capable Infrastructure by means of Localized Admission Control", Lecture Notes on Computer Science, Springer-Verlag, volume 1989, January 2001. This paper and a more detailed technical report can be requested to the authors, by writing to blefari@diei.unipg.it.

[BCP00] G. Bianchi, A. Capone, C. Petrioli, Packet Management Techniques for Measurement Based End-to-end Admission Control , KICS/IEEE Journal on Communications and Networking, June 2000.

[BJS00] L. Breslau, S. Jamin, S. Schenker: "Comments on the performance of measurement-based admission control algorithms", IEEE Infocom 2000, Tel-Aviv, March 2000.

[BKS00] L. Breslau, E. W. Knightly, S. Schenker, I. Stoica, H. Zhang: "Endpoint Admission Control: Architectural Issues and Performance", ACM SIGCOMM 2000, Stockholm, Sweden, August 2000.

[BOR99] F. Borgonovo, A. Capone, L. Fratta, M. Marchese, C. Petrioli, "PCP: A Bandwidth Guaranteed Transport Service for IP networks", IEEE ICC'99, June 1999.

[ELE00] V. Elek, G. Karlsson, "Admission Control Based on End-to-End Measurements", Proc. of IEEE Infocom 2000, Tel Aviv, Israel, March 2000.

[GRO99] M. Grossglauser, D. N. C. Tse: "A Framework for Robust Measurement Based Admission Control , IEEE/ACM Transactions on Networking, Vol. 7, No. 3, June 1999.

[KEL00] F. P. Kelly, P. B. Key, S. Zachary: " Distributed Admission Control", IEEE JSAC, Vol. 18, No. 12, December 2000.

[RFC2474] K. Nichols, S. Blake, F. Baker, D. Black, "Definitions of the Differentiated Service Field (DS Field) in the Ipv4 and Ipv6 Headers", RFC2474, December 1998.

[RFC2475] S. Blade, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss, "An Architecture for Differentiated Services", RFC2475, December 1998.

[RFC2597] J. Heinanen, F. Baker, W. Weiss, J. Wroclawski, Assured Forwarding PHB Group , RFC 2597, June 1999.

[RFC2990] G. Huston, "Next Steps for the IP QoS Architecture", RFC2990, November 2000.

[RFC2998] Bernet, Y., Yavatkar, R., Ford, P., Baker, F., Zhang, L., Speer, M., Braden, R., Davie, B., Wroclawski, J. and E. Felstaine, "A Framework for Integrated Services Operation Over DiffServ Networks", RFC 2998, November 2000.

11. Author's Addresses

Giuseppe Bianchi
DIE, University of Palermo
Viale delle Scienze, Parco d'Orleans
90128 Palermo, ITALY
Tel: +39 091 6566 276
E-mail: bianchi@elet.polimi.it

Nicola Blefari-Melazzi

DIEI, University of Perugia

Via G. Duranti 93, 06125 Perugia, ITALY

Tel: +39 075 585 3630

e-mail: blefari@diei.unipg.it

12. Full Copyright Statement