AIMD Penalty Shaper to Enforce Assured Service for TCP Flows

Emmanuel Lochin[1], Pascal Anelli[2], and Serge Fdida[1]

[1] LIP6 - Université Paris 6, [2] IREMIA - Université de la Réunion, France
{emmanuel.lochin, serge.fdida}@lip6.fr, pascal.anelli@univ-reunion.fr

Abstract. Many studies explored the guaranteed TCP throughput problem in DiffServ networks. Several new marking schemes have been proposed in order to solve this problem. Even if these marking schemes give good results in the case of per-flow conditioning, they need complex measurements. In this paper we propose a conditioning method to reduce these complex measurements and an AIMD[1] Penalty Shaper (APS) which is able to profile a set of TCP flows so as to improve its conformance to a desired target rate. The main novelty of this shaper is that the shaping applies an AIMD penalty delay which depends on the out-profile losses in a DiffServ network. This penalty shaping can be used with any classic conditionner such as a token bucket marker (TBM) or a time sliding window marker. We made an evaluation of the APS on a real testbed and showed that the proposed scheme is easily deployable and allows for a set of TCP flows to achieve its target rate.

Key words: Bandwith allocation, Edge to Edge performance, Quality of service, Assured Service, TCP, Experimentation with real testbeds.

1 Introduction

The Differentiated Services architecture [1] proposes a scalable means to deliver IP Quality of Service (QoS) based on handling of traffic aggregates. This architecture advocates packet tagging at the edge and lightweight forwarding routers in the core. Core devices perform only differentiated aggregate treatment based on the marking set by the edge devices. Edge devices in this architecture are responsibles for ensuring that user traffic conforms to traffic profiles. The service called Assured Service (AS) built on top of the AF PHB is designed for elastic flows. The minimum assured throughput is given according to a negotiated profile with the user. Such traffic is generated by adaptive applications. The throughput increases as long as there are available resources and decreases when a congestion occurs. The throughput of these flows in the assured service breaks up into two parts. First, a fixed part that corresponds to a minimum assured throughput. The packets of this part are marked like inadequate for loss (colored green or marked IN). Second, an elastic part which corresponds to an

---

[1] Additive Increase Multiplicative Decrease

opportunist flow of packets (colored red or marked OUT). These packets are conveyed by the network on the principle of "best-effort" (BE). In the event of congestion, they will be dropped first. Thanks to an AIMD Penalty Shaper, we show that it is possible to provide service differentiation between two source domains, on a set of TCP flows, based on its marking profile. In this paper we evaluate the solution with long-lived TCP flows. The proposed solution provides the advantage of neither needing RTT[2] evaluation nor loss probability estimation. The solution takes care of the behavior of TCP flows only. Consequently, as it is easily deployable, it has been experimented on a real testbed.

2 Related work

There have been a number of studies that focused on assured service for TCP flows but also on the aggregate TCP performance. In [2], five factors have been studied (RTT, number of flows, target rate, packet size, non responsive flows) and their impact has been evaluated in providing a predictable service for TCP flows. In an over-provisionned network, target rates are achieved regardless of these five factors. This result is corroborated by [3]. However the distribution of the excess bandwidth depends on these five factors. When responsive TCP flows and non-responsive UDP flows share the same class of service, there is unfair bandwidth distribution and TCP flow throughtputs are affected. The fair allocation of excess bandwidth can be achieved by giving different treatment to out-of-profile traffic of two types of flows [3]. Recently [4] demonstrates the unfair allocation of out-of-profile traffic and concludes that the aggregate that has the smaller/larger target rate occupies more/less bandwith than its fair-share regardless of the subscription level. In [5], a fair allocation of excess bandwidth has been proposed based on a traffic conditioner. The behavior of the traffic conditioner has a great impact on the service level, in terms of bandwidth, obtained by TCP flows. Several markers have been proposed to improve throughtput insurance, [6], [7], [8], [9]. These algorithms propose to mark aggressive TCP flows severely out-of-profile so that they are preferentially dropped. Even if these marking strategies work well in simulation, their main disavantage is their implementation complexity. Indeed, these algorithms need to measure a flow's RTT, its loss probability or have a per-state information of the flows.

3 The AIMD Penalty Shaper (APS)

Let $r(i)_{AS}$ be the assured rate of the flow i (*i.e.* in-profile packets throughput), n the number of AS TCP flows in the aggregate at the bottleneck level and C the link capacity. Precisely, this capacity corresponds to a bottleneck link in the network. If a number of i flows cross this link, the total capacity allocated for assured service R_{AS} is : $\sum_{i=1}^{n} r(i)_{AS}$. Let C_{AS} be the ressource allocated to the assured service.

$$R_{AS} < C_{AS} \tag{1}$$

---

[2] Round Trip Time

Equation (1) means an **under-subscription** network. In this case, there is excess bandwidth in the network. If $R_{AS} \geqslant C_{AS}$, this is an **over-subscription** network and there is no excess bandwidth. This configuration is the worst case for the AS. This service must provide an assurance until the over-subscription case is reached. Afterwards, not enough ressources are available and the service is downgraded.

$$\textit{TCP Throughput} = \frac{W_{max} * \text{Maximum Segment Size}}{RTT * \sqrt{p}} \tag{2}$$

In a well-dimensioned network, inequity from (1) should be respected. When there are losses in the network, it corresponds to the losses of out-profile packets, and not in-profile packets. It means that a light congestion appears in the network and some out-profile packets must be dropped. In order to increase the loss probability of the opportunist flows, new conditionners presented in section 2 are based on increasing the out-profile part of the most aggressive traffic. Then, the loss probability raises and the TCP throughput of the opportunist traffic decreases. It's a logical behaviour because the latter has a reject probability higher than the non-opportunist traffic. [10] gives a model of TCP throughput represented by the equation (2). With W_{max} is the TCP maximum window size and p the loss probability. Changing the p value from the equation (2) thanks to a marking strategy is complex. Indeed, it is necessary to evaluate the loss probability of the network and estimate an RTT for each flow. As opposed to the marking strategy adopted by new conditionners, we propose a delay based shaper. This shaper applies a delay penalty to a flow if there are out-profile packets losses in the network and if it outperforms its target rate. The basic idea is that the penalty is a function of the out-profile packet losses. Instead of raising the p value, from equation (2), of the most opportunist flow, the AIMD Penalty Shaper raises a delay penalty to the flow. It results in a growth of the RTT. Mathematicaly, as shown in (2), increasing RTT value is similar to increasing p value in term of TCP throughput. [11] has shown that limiting out-profile packets is a good policy to achieve a target rate. Indeed, by avoiding packets dropping we avoid TCP retransmission. This is an efficient solution to optimize the bandwidth usage. Thus, our goal is to reduce out-profile losses by applying a delay penalty to the flows that are the most opportunist in the network. Therefore, when a RIO[3] [12] router in the core network is dropping out-profile packets, it marks the ECN flag [13] of the in-profile packets enqueued in the RIO queue. In a well-dimensioned network, there is no in-profile packet loss. Then, the edge device can be aware that there is a minimum of one flow or set of flows which are opportunists in the network. This opportunist traffic is crossing the same path. The edge device evaluates its sending rate thanks to a Time Sliding Window (TSW) algorithm [14]. If its sending rate is higher than its target rate, it considers that its traffic may be opportunist. Then, it applies a penalty to the incoming traffic until the network feedback that there are out-profile packets losses. This penalty allows a raise of the RTT and consequently, decrease

---

[3] RED with IN and OUT

```
K = 10ms
FOR each observation period T
 TSW gives an evaluation of the throughput : throughput_measured
 IF throughput_measured < target_rate OR there are no out-profile losses
 THEN reduce the penalty delay
      current_penalty = current_penalty - ((i/2) * K)
      i = 1
 ELSE
     raise the penalty delay
     current_penalty = current_penalty + (i * K)
     i = i * 2;
 ENDIF
ENDFOR
```

Fig. 1. APS algorithm

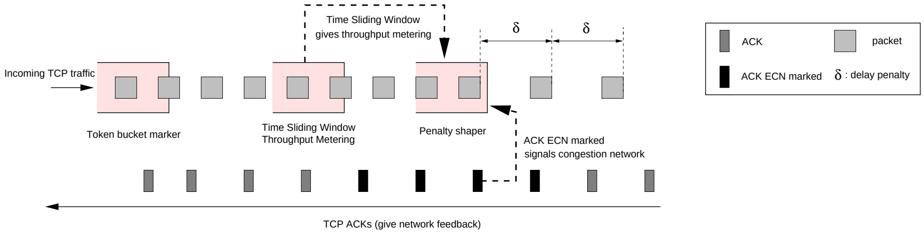


Fig. 2. conditioning with APS

the TCP throughput. We choose to use an AIMD penalty instead of a linear penalty in order to be in conformity with the TCP congestion control. If there is no loss anymore, the penalty decreases and the TCP throughput raises. The algorithm presented in figure 1 explains how the AIMD penalty is calculated and applied. As explained on figure 2, once incoming TCP traffic is shaped, it passes through a marker such as a TBM. This conditioner is setup on the edge device such an ingress edge router. Many are the conditioners presented in section 2 which will never leave the framework of simulation because of their conditioning constraints. We chose to make the traffic conditioning in the following way : each client emitting one or more flows towards one or more destinations will have one traffic profile per destination. As shown on figure 3, client A forces the edge router to setup three different traffic conditioners. Two conditioners with a contract rate of $4Mbits/s$ and one conditioner with a contract rate of $2Mbits/s$. The main advantage of this solution is that the conditioning can be made on flows with similar RTTs (i.e. in the same order of magnitude). This solution doesn't depend on the complex problem of RTT estimation necessary to the functionement of the conditioners presented in section 2. The solution of traffic shaping coupled to a conditioner/marker such as the TBM should be easily deployable and scalable.

4 Experimental testbed

As shown in figure 4, we use the well-known dumbbell topology. The testbed is composed of computers running FreeBSD. On the edge routers, the token

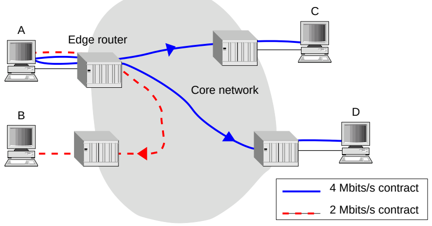


Fig. 3. Trafic conditioning sample

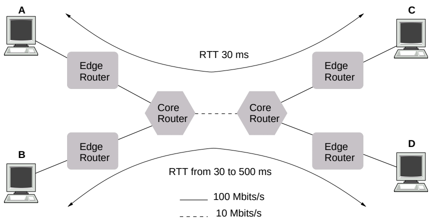


Fig. 4. Experimental testbed

bucket marker from ALTQ[4] developement and the AIMD Penalty Shaper based on Dummynet[5]. On the core routers, a RIO queue with the ECN marking functionnality from ALTQ developement. Thus, the RIO queue is able to mark the ECN flag of the in-profile packets if it detects out-profile losses in its queue. We use the `Iperf 1.7.0`[6] traffic generator and two transmitting machines and two receivers for measurements. The main parameters and hypothesis are : traffic generation is carried out in the following way: A to C (A, C) and B to D (B, D), after 120 seconds, `Iperf` gives an average throughput of the flow ; each AS flow is transmitted as TCP, packets have a size of 1024 bytes ; `Iperf` uses a TCP maximum window size $Wmax = 64packets$; each set of flows between two hosts is conditionned by one TBM with or without APS ; b parameter of the TBM is set to one packet ; r parameter is set to the desired target rate ; the delay penalty is set to $10ms$ and the observation period to $1sec$. It means that each second the algorithm gives an estimation of the throughput and evaluates the penalty delay ; we use a non-overlapping RIO with parameters : $(min_{out}, max_{out}, p_{out}, min_{in}, max_{in}, p_{in}) = (1, 63, 0.1, 64, 128, 0.02)$, the queue size corresponds to $2 * Wmax$.

5 Performance evaluation of the AIMD Penalty Shaper

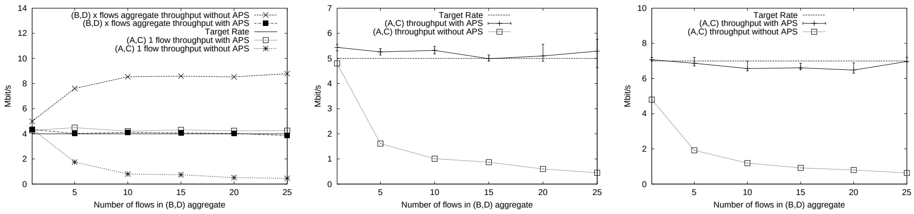
This section presents the results obtained in a real testbed with the APS. We evaluate the performance of the APS when TCP traffic have the same or a different number of flows and identical or different RTTs. In these tests, the total capacity allocated for the assured service is $R_{AS} = 8Mbits/s$. The ressource allocated to the assured service is $C_{AS} = 10Mbits/s$ that corresponds to the bottleneck capacity. This is an under-subscription network because there are $2Mbits/s$ of excess bandwidth.

5.1 Impact of the aggregates' aggressiveness in an under-subscribed network

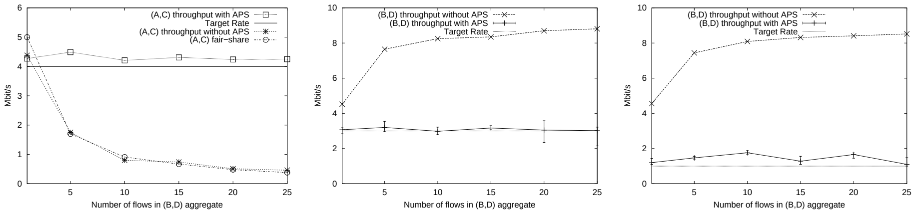
Even if there is a different number of flows in the aggregates, the APS is able to reach the desired target rate. Results are presented in figure 5. When two aggregates with different number of flows are in a network, the higher outperforms the

---

[4] `http://www.csl.sony.co.jp/person/kjc/`
[5] `http://info.iet.unipi.it/~luigi/ip_dummynet/`
[6] `http://dast.nlanr.net/Projects/Iperf/`



(a) Test1 TR=4Mb/s (b) Test2 TR=5Mb/s (c) Test3 TR=7Mb/s



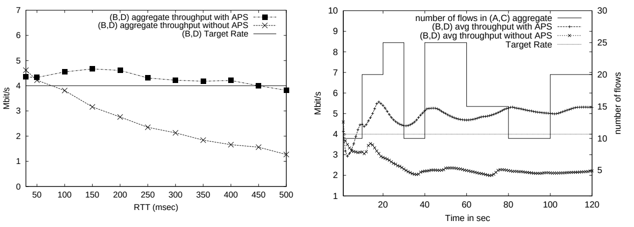
(d) Test1 TR=4Mb/s (e) Test2 TR=3Mb/s (f) Test3 TR=1Mb/s

Fig. 5. TCP throughput versus aggregates' aggressiveness with various target rates

smaller [2]. In these tests, (A, C) has one flow and (B, D) has a variable number of flows ranging from 1 to 25. the RTT is set to $30ms$. After repeating each experiments five times, we calculate the average throughput value (for information purpose, min/max values are represented on the second and third test). In the first test, the target rate of both is set to $r(A, C)_{AS} = r(B, D)_{AS} = 4Mbits/s$. Figure 5 (a) shows the throughput obtained by both aggregates. For clarification, we draw on figure 5 (d) the throughput obtained by the (A, C) aggregate alongside the fair-share curve. Figure 5 (d) shows that the TBM stays close to the fair-share and that we obtain the desired target rate with the APS. Figures 5 presents the same scenario but the target rate on figures 5(b) and (e) is set to $r(A, C)_{AS} = 5Mbits/s$ and $r(B, D)_{AS} = 3Mbits/s$ and on figures 5(c) and (f) : $r(A, C)_{AS} = 7Mbits/s$ and $r(B, D)_{AS} = 1Mbits/s$. So, the second and the third tests illustrate both the case where the aggregates have near target rates and the case where they have distant target rates under under-subscription conditions. With APS, the target rate of TCP can be controlled and has a value over the target or near the target (in our worst case).

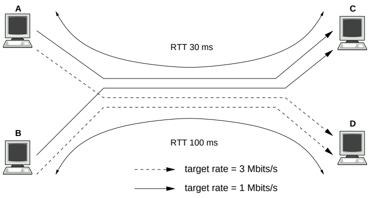
5.2 Impact of the RTT in an under-subscribed network

Even if there is a high number of flows in the aggregate and a high RTT difference, the APS is able to reach the target rate requested by an aggregate. The target rate for (A, C) and (B, D) is $r(A, C)_{AS} = r(B, D)_{AS} = 4Mbits/s$ Figure 6(a) shows the throughput of a 10 flow aggregate (B, D) in competition with a 10 flow aggregate (A, C). For the (A, C) aggregate, the RTT is equal to



(a) (b)

Fig. 6. TCP throughput versus RTT



agregate's name	# flows	without APS	with APS
AC	5	2.12 / 1	1.42 / 1
AD	5	2.54 / 3	3.24 / 3
BC	5	2.04 / 1	1.04 / 1
BD	5	2.64 / 3	3.04 / 3
AC	10	2.19 / 1	1.24 / 1
AD	10	2.45 / 3	3.15 / 3
BC	10	2.27 / 1	1.01 / 1
BD	10	2.45 / 3	3.13 / 3

(caption : goodput in Mbits/s / target rate in Mbits/s)

(a) (b)

Fig. 7. Multi TCP agregates

$30ms$ and for the (B,D) aggregate, we increase gradually the RTT from $30ms$ to $500ms$. It appears that the aggregate reaches the target rate when it is feasible (i.e. when target rate : $r(B,D)_{AS} > Wmax/RTT$). On figure 6(b), the (A,C) aggregate has an RTT of $30ms$ and number of flows varying from 10 to 25. (B,D) aggregate has 5 flows and an RTT of $100ms$. We draw the instantaneous average throughput of (B,D) aggregate in function of the number of flows in (A,C). Thanks to the APS, we can see that the (B,D) throughput stays above the target rate. Finally, in the last test, we put in competition four agregates with different RTTs, target rates and number of flows. Figure 7(a) presents the scenario and figure 7(b) the results obtained. Thanks to the APS, all the aggregates reach their target rate.

6 Conclusion and future works

In this paper, we have studied on a real testbed an AIMD Penalty Shaper which provides throughput assurance between TCP flows. This is the first proposal that uses a delay penalty which depends on the out-profile losses in a DiffServ network. The TCP throughput is guaranteed because the conditioner works with the same dynamic than TCP (AIMD). The main consequence of these measurements is that we are able to obtain the guaranteed throughput if the profiled TCP aggregates in competition have the same or different number of flows. This is true whatever the differences between their RTTs and their target rates. The proposed solution has the advantage of being easily deployable because it doesn't require complex measurements. The solution is scalable and being likely to be used with the most frequently used conditionners such as token bucket marker or time sliding window marker. We are currently deploying this proposal on a wide area network with various traffic such as long-lived and short-lived TCP flows in order to improve this mechanism in general conditions. If the results are satisfying, then this proposal allows the effective deployment of a service adapted to the TCP traffic.

References

1. Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., Weiss, W.: An architecture for differentiated services. Request For Comments 2475, IETF (1998)
2. Seddigh, N., Nandy, B., Pieda, P.: Bandwidth assurance issues for TCP flows in a differentiated services network. In: Proc. of IEEE GLOBECOM, Rio De Janeiro, Brazil (1999) 6
3. Goyal, M., Durresi, A., Jain, R., Liu, C.: Performance analysis of assured forwarding. Internet draft, IETF (2000) draft-goyal-diffserv-afstdy-00.txt.
4. Park, E.C., Choi, C.H.: Proportionnal bandwidth allocation in diffserv networks. In: Proc. of IEEE INFOCOM, Hong Kong (2004)
5. Alves, I., de Rezende, J.F., de Moraes, L.F.: Evaluating fairness in aggregated traffic marking. In: Proc. of IEEE GLOBECOM. (2000)
6. El-Gendy, M., Shin, K.: Assured forwarding fairness using equation-based packet marking and packet separation. Computer Networks **41** (2002) 435–450
7. Kumar, K., Ananda, A., Jacob, L.: A memory based approach for a TCP-friendly traffic conditioner in diffserv networks. In: Proc. of the IEEE International Conference on Network Protocols - ICNP, Riverside, California, USA (2001)
8. Feroz, A., Rao, A., Kalyanaraman, S.: A TCP-friendly traffic marker for IP differentiated services. In: Proc. of IEEE/IFIP International Workshop on Quality of Service - IWQoS. (2000)
9. Habib, A., Bhargava, B., Fahmy, S.: A round trip time and time-out aware traffic conditioner for differentiated services networks. In: Proc. of the IEEE International Conference on Communications - ICC, New-York, USA (2002)
10. Floyd, S., Fall, K.: Promoting the use of end-to-end congestion control in the Internet. IEEE/ACM Transactions on Networking **7** (1999) 458–472
11. Yeom, I., Reddy, N.: Realizing throughput guarantees in a differentiated services network. In: Proc. of IEEE International Conference on Multimedia Computing and Systems- ICMCS. Volume 2., Florence, Italy (1999) 372–376
12. Clark, D., Fang, W.: Explicit allocation of best effort packet delivery service. IEEE/ACM Transactions on Networking **6** (1998) 362–373
13. Ramakrishnan, K., Floyd, S.: A proposal to add explicit congestion notification. Request for comments, IETF (1998)
14. Fang, W., Seddigh, N., AL.: A time sliding window three colour marker. Request For Comments 2859, IETF (2000)