Speeding up Transaction-oriented Communications in the Internet

Tobias Küfner, Mark Doll, Götz Lichtwald, Martina Zitterbart
Institute of Telematics, University of Karlsruhe
{kuefner|doll|lichtwald|zit}@tm.uka.de

Abstract: Future mobile networks will offer a great variety of multimedia services. DiffServ is the promising framework standardized by the IETF to enhance the Internet with the essential quality of service. Unfortunately, transaction-oriented applications—the most popular class of applications—are not sufficiently supported yet. Quick Forwarding (QF) has been proposed to fill this gap between Expedited Forwarding and Assured Forwarding. In this paper,[1] we present a scalable QF Per-Domain Behaviour using a set of simple but effective traffic conditioning functions. Our approach enables ISPs to provide end-to-end services suitable for transaction-oriented communications while utilizing their network resources efficiently.

1 Introduction

Quick Forwarding (QF) [BHW01] is designed for transaction-oriented communication scenarios. We call the exchange of a request and a response message between two participants a micro transaction. It is important to realize that the processing of a message cannot start until the whole message has been received. This stresses the need to transport the messages in a bursty manner to achieve minimal delays. Fast and reliable micro transactions would for example benefit database transactions, remote procedure calls, middleware infrastructures and signaling messages. Of course, the dominant traffic—and basis for our simulative study—is web traffic, which also serves as a transport for many transaction-oriented applications, including banking, brokerage, shopping, auctions, booking and library retrieval.

Even though Expedited Forwarding (EF) [DCB$^+$02] also provides a low delay low loss service, it is targeted at non-bursty traffic. Assured Forwarding (AF) [HBWW99], on the other hand, tolerates bursts but is not especially targeted at minimizing delays. This is, where the QF per-hop behavior (PHB) adds a new class for bursty *and* delay sensitive traffic to the DiffServ Framework.

In this paper we focus on the QF per-domain behavior (PDB). We present a simple scheme for traffic conditioning and admission control at the domain boundaries that aims to min-

[1]Parts of this work were funded by the Bundesministerium für Bildung und Forschung of the Federal Republic of Germany (Förderkennzeichen 01AK045) and Siemens AG, Munich. The authors alone are responsible for the content of this paper.

imize packet loss while still keeping delays at a minimum. The scheme is scalable in the sense that the dimensioning of the traffic conditioning function for QF behavior aggregates is independent of the upstream path and the delay and loss characteristics are still achieved if QF traffic traverses multiple domains.

2 Scalable Per-Domain Behavior Design

The challenge in the design of a QF per-domain behavior (PDB) is to guarantee the low loss low delay property while keeping QF scalable in the number of micro flows. We compose a scalable PDB based on Quick Forwarding by defining the following twofold traffic conditioning. At the first-hop router QF micro flows are separated by a multi-field classifier and metered with their committed traffic profile, using a token bucket with parameters (R, B). Conforming packets are multiplexed in a FIFO queue allocated to the QF aggregate. Non-conforming packets are dropped immediately. We assume that the application can produce a micro flow according to the agreed traffic profile (R, B). At the ingress routers packets are filtered by a behavior-aggregate classifier and metered again by a traffic profile (R, B). This token bucket is part of the traffic conditioning specification (TCS) which the adjacent domains agreed upon. While R should be defined greater than the cumulated rates of all the aggregated micro flows, we propose to define B much smaller than the cumulated burst sizes and almost independent from the number of micro flows.

To ensure that all traffic stays in-profile, we employ a *token bucket shaper* (TBS) at the domain egress which delays the departure of non-conforming packets rendering them conform to the TCS agreed upon with the downstream domain. It works like a token bucket of depth B which is filled with tokens at the rate R. The size S of an arriving packet is compared to the number of available tokens T. If $T \geq S$, S tokens are taken out of the bucket and the packet is enqueued immediately. Else if $S < T$, also S tokens are taken out, but the packet is delayed for $(S - T)/R$ before it is enqueued. Of course, shaping will insert delay which is definitely not appreciated but still better than dropping. However, we suppose that shaping will be needed pretty rarely when realistic traffic is assumed. Of course, when the TBS's buffer space is depleted, QF packets have to be dropped, something that should be avoided under normal operating conditions by allocating enough memory to the TBS. Since the TBS is needed only at a domain egress, we assume the expenses are affordable.

3 Traffic Conditioning and Aggregation Effects

We derive a traffic model for transaction-oriented communication from traces and models of web traffic in [CTB98], [Ma97] and [CL99]. We compose an on/off source, which models HTTP responses only, since HTTP requests usually fit within one IP packet and thereby generate less bursty traffic. During the on-time the packets of a burst are sent at a peak rate p. The burst size and the on-time respectively is modeled by a Pareto-Cutoff distribution with parameters $\alpha, k, m > 0$ and cumulative distribution function (cdf):

$$F(x) := P(X \leq x) = \begin{cases} 1 - \left(\frac{k}{x}\right)^{\alpha} & \text{if } k \leq x < m \\ 1 & \text{if } m \leq x \end{cases} \tag{1}$$

In $[k, m)$ F is equal to the cdf of the more common Pareto distribution. The latter especially reflects the probability of large bursts in a single micro flow, which is crucial for the burstiness of the aggregate [CTB98]. By the “Cutoff” we incorporate the impact of the committed burst size B into our traffic model. If a source emits more than $m = B/(1 - R/p)$ Bytes at peak rate p, the resulting micro flow would not conform to the traffic profile (R, B) and the trailing packets would be dropped by the first-hop router. Typically the off-time between two consecutive transactions is modeled with a heavy-tailed distribution as well. This especially reflects very long off-times as they can be observed in real HTTP traces. In our scenario however, long off-times are not reasonable, because the user is expected to pay for QoS guarantees. For this reason, we model the off-times according to the committed rate R in the user’s traffic profile. After sending a burst at peak rate p an on/off source will pause until the token bucket is filled up again.

We consider a worst case topology and the maximum traffic load in order to derive the feasible delay bounds and the buffer requirements of the QF PDB. Thus, our domain model consists of n independent and identically distributed on/off sources—each modeling one QF micro flow— emitting QF packets towards a single egress router/interface.

The fundamental supposition for our approach is that packet loss can be prevented by pretty rarely applied traffic shaping at the domain boundary. In order to quantify the trade-off between required buffer space on the one hand and delay implied by traffic shaping on the other hand, we log the number of “used tokens” from the token bucket shaper (TBS). When a packet of size S passes a TBS, S tokens are taken out of the bucket, i. e. “used” by the packet. Every time a packet drops out of a token bucket shaper, we calculate the number of used tokens by subtracting the number of remaining tokens in the bucket from the bucket depth B. We log these values and plot the complementary cdf.

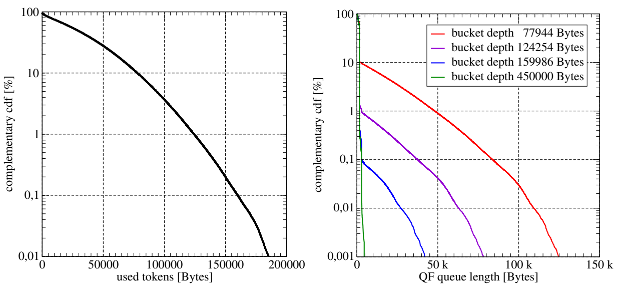

Figure 1: The used tokens from the TBS (l.), the QF queue length (c.) and the burst delay (r.) for a QF aggregate generated by 10 independent on/off sources

The distribution on the left side of Fig. 1 has been observed for the aggregation of $n = 10$ QF micro flows, each generated by an on/off source with $\alpha = 1.0$, $k = 500$ Bytes and $m = 50$ kB. The maximum arrival rate of a single micro flow is set to $p_1 = 10$ Mb/s, the committed rate is set to $R_1 = 900$ kb/s. The committed burst size is then given by $B_1 = m(1 - R_1/p_1) = 45.5$ kB. The TBS in the egress router is configured with a token rate of $R_n = 10$ Mb/s. The QF queue is served at a peak rate of $p_n = 100$ Mb/s. In the worst case n bursts of size m could arrive in parallel, resulting in an aggregated burst of size $nm = 500$ kB. Without shaping $B_n = nm(1 - R_n/p_n) = 450$ kB would need to be agreed upon in the TCS with the upstream domain. The left graph of Fig. 1 shows that B_n could be scaled down to 159986 Bytes by shaping only 0.1 % of the packets, down to 124254 Bytes by shaping 1 % and down to 77944 by shaping 10 %.

The simulation has been repeated with these three values and $B_n = 450$ kB (the worst case). The resulting distributions of the queue length are plotted in the center of Fig. 1. The left-most curve is obtained when no shaping is applied, i. e. B_n is set to the worst case. The upper three curves display the longer delays due to the smaller bucket depth B_n, experienced by 0.1 %, 1 % and 10 % of the packets. On the right side of Fig. 1 the complementary cdfs of the according burst delays are plotted. You can see, that the 1 % of the bursts with maximum size $m = 50$ kB was delayed about 40 ms when no shaping was applied. In comparison, the resulting delays for smaller bucket depths look acceptable. Note, that for B_n = 159986 Bytes 0.1 % of the packets experience significant shaping delay but less than 0.01 % of the bursts do. If the burst delay or number of delayed bursts is not suitable for an ISP however, he is able to guarantee tighter delay bounds by agreeing upon a larger bucket depth B_n with the upstream domain.

In another simulation, we explored the dependency of the QF PDB on the parameters α, k and m of the Pareto-Cutoff distribution. We varied α in $[1, 2]$, k in $[125, 1250]$ Bytes and m in $[12.5, 125]$ Bytes according to the models in [CTB98], [Ma97] and [CL99]. All the other simulation parameter values remained the same as in the previous simulation. The results showed, that the probability that u or more tokens are used at a point of time decreases when α is increased in $[1, 2]$, i. e. the more α tends to 1 the more buffer space is needed to prevent packet loss. The distribution of the used tokens in the TBS turned out to depend roughly linearly on the maximum burst size m, while the impact of the minimum burst size k is much smaller. This is remarkable since the mean burst size depends primarily on k and less on m. It follows that the committed burst size B_1 per micro flow is crucial for the QoS and scalability of the PDB, whereas the mean burst size—and thereby also the frequency of bursts—plays a minor role. Anyhow, the simulation results demonstrate that the proposed PDB facilitates an optimal trade-off between QoS and scalability for various burst size distributions and is not restricted to web traffic.

In order to demonstrate the scalability of proposed PDB in the number of micro flows n, we setup simulations with n inbetween 1 and 1000 traffic sources, each with the parameters from the first simulation. The token rate of the TBS R_n and the maximum service rate p_n were chosen linear to n. Our results showed, that even for an aggregation of 1000 micro flows —each with an arrival rate of $p_1 = 1$ Gb/s per micro flow— a bucket depth of $B_n = 400$ kB can be specified while shaping is only applied to 1 % of the packets. For comparison, the worst case burst size for n = 1000 is nm = 50 MB. In addition, our simulations with $p_1 = 10$ Mb/s revealed an amazing effect: While the probability, that many tokens are used at a point of time, increased for $n < 100$, as expected, the probability decreased for $n > 100$. Our explanation of this effect is based on the ratio of the maximum arrival rate per micro flow p_1 and the average service rate per aggregate R_n. As long as p_1/R_n is less than 1, packets from a single micro flow arrive less fast than non-conforming packets depart, i. e. the probability of aggregation is dominated by the probability that multiple micro flows arrive in parallel. If p_1/R_n is greater than 1, packets from a single micro flow arrive faster than non-conforming packets depart, i. e. the probability of aggregation is dominated by the probability of large bursts within a single micro flow. Simulations with $p_1 = 100$ Mb/s and $p_1 = 1$ Gb/s confirmed this proposition.

4 Conclusion

In this paper, we propose a scalable per-domain behavior for DiffServ networks to fulfill the QoS requirements of transaction-oriented applications, the currently largest class of applications in the Internet. The inherent property of transactions is that they cause bursty traffic. These bursts should be forwarded as reliable and timely as possible. Our PDB design makes use of the Quick Forwarding PHB to guarantee low delay and low loss for bursty micro flows. To keep QF scalable we apply an admission control per micro flow in first-hop routers and employ a token bucket shaper at each egress node of a DiffServ domain. The latter allows ISPs to negotiate reasonable traffic profiles for QF. We evaluated, our approach with a generic traffic model derived from traces and models of the WWW. Our simulation results demonstrated that the Quick Forwarding PDB allows an optimal trade-off between shaping delay and required buffer space. Finally, our conclusion is that ISPs are able to support and tariff end-to-end services for transaction-oriented applications based on the proposed PDB while utilizing their network resources efficiently.

References

[BHW01] Bless, R., Holzhausen, D., und Wehrle, K.: Towards Better Support of Transaction Oriented Communication in Differentiated Services Networks. In: *Proceedings of QofIS 2001*. S. 158–169. Coimbra, Portugal. September 2001. COST 263.

[CL99] Choi, H.-K. und Limb, J. O.: A Behavior Model of Web Traffic. In: *Proceedings of ICNP 1999*. S. 327–334. Toronto, Canada. 1999. IEEE.

[CTB98] Crovella, M., Taqqu, M., und Bestavros, A.: Heavy-Tailed Probability Distributions in the World Wide Web. In: *A Practical Guide To Heavy Tails*. chapter 1, S. 3–26. Chapman & Hall. New York. 1998.

[DCB+02] Davie, B., Charny, A., Bennett, J., Benson, K., Boudec, J.-Y. L., Courtney, B., Davari, S., Firoiu, V., und Stiliadis, D. An Expedited Forwarding PHB (Per-Hop Behavior). RFC 3246. March 2002. Proposed Standard.

[HBWW99] Heinanen, J., Baker, F., Weiss, W., und Wroclawski, J. Assured Forwarding PHB Group. RFC 2597. June 1999. Proposed Standard.

[Ma97] Mah, B. A.: An Empirical Model of HTTP Network Traffic. In: *Proceedings of Infocom 1997*. S. 592–600. Kobe, Japan. 1997. IEEE.