On the Allocation of Network Service Curves for Bandwidth/Delay-Decoupled Scheduling Disciplines

Jens B. Schmitt
Multimedia Communications (KOM), Department of Electronic Engineering & Information Technology
Darmstadt University of Technology, Germany
Jens.Schmitt@KOM.tu-darmstadt.de

***Abstract* -- Providing Quality of Service (QoS) guarantees in packet-switched networks like the Internet has been and still is an important research area. In particular, it is important to ensure strict (deterministic) guarantees for highly time-sensitive data flows. In this paper, we discuss how to efficiently allocate service curves for deterministically guaranteed services. We focus on bandwidth/delay-decoupled service disciplines and their respective service curves due to their distinct advantages over purely rate-based schedulers. Resource-optimal network service curves for deterministic service flows scheduled by bandwidth/delay-decoupled service disciplines are derived and their performance is discussed by numerical examples.**

I. Introduction

A. Background & Motivation

Future multi-service IP (Internet Protocol) networks will carry a diverse set of traffic types stemming from very different applications, nowadays transported over specialized legacy networks. This set of traffic types ranges from simple best-effort services for (time-)uncritical applications like ftp or email to deterministic services which give strict guarantees on delay and loss characteristics for (time-)critical applications as, e.g., the control of a nuclear plant or telemedicine applications. The latter services are important, if a future multi-service IP-based Internet really shall take over the traffic served by specialized legacy networks. Certainly, deterministic services may not constitute the major part of the overall carried traffic of a future multi-service Internet, yet their high significance and their expensive implementation makes them nevertheless attractive for optimization with respect to resource allocation.

B. Scheduling Disciplines for Deterministic Services

To enable deterministic services requires packet scheduling disciplines in routers to give guarantees on minimum service quantums dedicated to deterministic service flows[1] even under worst-case traffic arrival patterns. A large number of scheduling disciplines differing mostly in their accuracy and complexity has been proposed (a good, although a little bit dated overview is given in [1]). In general, there are two broad classes of schedulers: *rate-based* and *delay-based* schedulers. For rate-based schedulers there is work which generalizes and formalizes their behaviour in a class of so called latency-rate (*LR*) schedulers [2] (in [3] there is an almost identical framework which is called guaranteed rate schedulers). *LR* schedulers guarantee a deterministic service flow a certain minimum rate after a certain latency due to deviation from the perfect fluid model of a GPS (Generalized Processor Sharing) server. Delay-based schedulers, on the other hand, guarantee a certain maximum delay a node may hold a packet. In [4], it is shown that delay-based schedulers are superior to rate-based schedulers with respect to resource allocation for the single node case mainly because of their decoupling of bandwidth and delay requirements. However, as first shown in the seminal work by Parekh and Gallagher [5, 6] rate-based schedulers benefit from the “pay burst only once” principle in the multiple node case, whereas for delay-based schedulers it is still an open issue how to optimally partition the end-to-end delay budget in the multiple node case. This global effect may overcompensate the local superiority of delay-based schedulers.

Nevertheless, it can be noted that an important insight for the resource allocation optimization of deterministic services is that delay and bandwidth need to be decoupled to achieve efficient resource allocations. While in purely rate-based approaches this decoupling has not been taken into account, recent work has done so even without loosing the nice global behaviour of rate-based scheduling, see [7, 8]. The latter proposes a bandwidth-delay decoupled scheduling scheme, which is based on a scheduling discipline using non-linear service curves. In principle, they, like other approaches to decouple bandwidth and delay, propose piecewise linear service curves. Yet, none of these has dealt with an optimal choice of parameters of the piecewise linear service curve for a given (regulated) traffic flow. This is what we focus on in our work.

II. Network Calculus for Deterministic Services

A. The Setting for LR Schedulers and TSpec-Regulated Traffic

The mathematics of deterministic services, commonly called network calculus, are originally based on the work of Cruz [10] on arrival and service curves, yet have been refined and deepened in several other works [11, 12, 13]. In brief, while arrival curves describe the worst-case behavior of a source within given time intervals, service curves specify the

1. Let flow denote an arbitrary logically associated packet stream.

minimal service that is provided by a queue service discipline. By combining these two concepts it is possible to derive deterministic guarantees on loss and delay under the worst-case scenario of a greedy source and a fully loaded server. A typical and often used arrival curve is the so-called TSpec(*r*,*b*,*p*,*M*) [14], defined by the following arrival curve

$$a^{TSpec}(t) = \begin{cases} M + pt & t < T \\ b + rt & t \geq T \end{cases}, \tag{1}$$

where $T = (b - M)/(p - r)$ may be considered as burst duration. The TSpec is essentially a dual token bucket, where a controlled burstiness is accounted for by the first bucket characterized by peak rate *p* and maximum packet size *M*, and the long-term behaviour is captured by the second token bucket characterized by average rate *r* and a bucket size *b*.

A typical (linear) service curve for deterministic services [9] based on latency-rate schedulers

$$s^{linear}(t) = R(t - L)^{+} = \begin{cases} 0 & t \leq L \\ R(t - L) & t > L \end{cases}, \tag{2}$$

where $L = C/R + D$ and *R* is the service rate assigned to a data flow by the respective queue service discipline, assuming that the stability condition $R \geq r$ holds. Here, the *C* and *D* terms represent the rate-dependent respectively rate-independent deviations of a *packet-based* scheduler from the perfect fluid model as introduced by [5][2]. These *error terms* are summed up along the data transmission path for each router during an advertisement phase.

B. Delay Bound and Required Service Rate

Applying network calculus we can compute a worst-case delay bound based on the horizontal deviation between arrival and service curve:

$$\begin{aligned} d_{max} &= h(a, s) \\ &= sup_{s \geq 0}(inf\{H: H \geq 0 \wedge a(t) \leq s(t + H)\}) \end{aligned} \tag{3}$$

For the TSpec as arrival curve and the linear service curve defined in (2), we note that the horizontal deviation needs to be taken on at either the origin or the burst duration *T* depending upon whether the service rate is faster or slower than the peak rate of the flow (see Figure 1 for an illustration). This observation results in

$$d_{max} = \begin{cases} \frac{T(p - R)}{R} + \frac{M + C}{R} + D & p \geq R \geq r \\ \frac{M + C}{R} + D & R \geq p \geq r \end{cases} \tag{4}$$

From the perspective of the application desiring a maximum queuing delay d_{max}, the service rate *R* that has to be reserved at the routers being traversed follows directly:

$$R = \begin{cases} \frac{pT + M + C}{d_{max} + T - D} & d_{max} \geq \bar{d} \\ \frac{M + C}{d_{max} - D} & d_{max} < \bar{d} \end{cases}, \tag{5}$$

$$\text{where } \bar{d} = \frac{M + C}{p} + D \tag{6}$$

[2] For many schedulers, in particular for WFQ (Weighted Fair Queueing) also known as PGPS (Packet-by-packet Generalized Processor Sharing) $C = M$ and $D = M/c$, where *c* is the link speed.

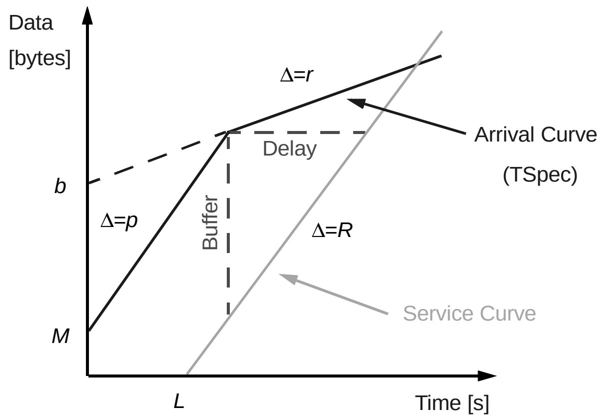

Figure 1: Network calculus for *LR* schedulers.

C. Buffer Requirements

Again, by applying the basic results from network calculus we can compute the buffer requirements at internal routers as the vertical deviation between arrival and service curve:

$$B = v(a, s) = sup_{t \geq 0}\{a(t) - s(t)\} \tag{7}$$

For the TSpec as arrival curve and the linear service curve defined in (2), we note that the vertical deviation needs to be taken on at either the latency of the service curve *L* or the burst duration *T* depending upon whether the service rate is faster or slower than the peak rate of the flow and whether the burst duration is shorter or longer than the scheduler latency (see Figure 1 for an illustration). These observations result in

$$B = \begin{cases} M + \frac{(p - R)(b - M)}{p - r} + C + RD & p \geq R \geq r, L \leq T \\ M + p\left(\frac{C}{R} + D\right) & R \geq p \geq r, L \leq T \\ b + r\left(\frac{C}{R} + D\right) & L > T \end{cases} \tag{8}$$

III. Bandwidth-Delay Decoupling of Service Curves

A. "Low Bandwidth / Short Delay"-Type of Flows

To use a linear service curve, i.e., a simple latency-rate curve, as done in the formulas above may lead to wasteful resource allocations, in particular, for "low bandwidth / short delay"-type of flows. For example, consider a data flow with TSpec = (2000, 1000, 8000, 500) [in bytes resp. bytes/s]. Let us assume this flow crosses 5 hops (all with MTU = 9188 bytes and link speed c = 155 Mb/s) all doing PGPS (Packet-by-packet Generalized Processor Sharing) [5]. Then we have $C = 5M$ = 2500 bytes and $D = MTU/c$ = 2.371 ms. Let us further assume the application desires a maximum queuing delay of d_{max} = 100 ms. Then we obtain from the formulas given above that R = 30729 bytes/s $\approx 4p = 16r$. That means in order to ensure a fairly strict delay bound the rate assignment of a flow is extremely over-provisioned in relation to its bandwidth requirements.

B. Decoupling Delay and Bandwidth Assignments

The solution to the above problem is to decouple the bandwidth assignment from delay goals which can only be achieved by non-linear service curves. The most simple non-linear service curve that can achieve bandwidth-delay decoupling is a continuous piece-wise linear service curve consisting of two linear segments, i.e., a service curve of the form

$$s^I(t) = \begin{cases} R_s(t-L)^+ & t < I \\ R_l t + f & t \geq I \end{cases} \quad (9)$$

with $f = (R_s - R_l)I - R_s L$ to ensure continuity.

Here I denotes the inflection point which separates the two phases of the service curve, with the first phase where a service rate R_s, which we call the short-term rate, is allocated to achieve a flow's delay goal, and the second phase where rate R_l, which we call the long-term rate, is assigned to ensure its bandwidth requirements. This service curve may thus be considered as a slight, yet effective generalization of the latency-(single-)rate scheduling scheme towards *latency-two-rates (L2R) service curves* and a corresponding class of schedulers.

While other non-linear service curves also have the bandwidth/delay-decoupling characteristic we further on focus on service curves as defined in (9) due to their simplicity and consequent attractiveness for actual implementation. In fact, we have integrated this kind of scheduling discipline available in the ALTQ framework [15] as a traffic control module into the experimental KOM RSVP engine [16].

IV. Optimal Bandwidth-Delay Decoupled Service Curves

A. Resource-Optimality for L2R Schedulers

If we deal with optimality we need to define optimal with respect to what and under which constraints. We take a per-flow perspective, i.e., we do not focus on the global optimization problem of competing flows and how to assign resources between these (although this is certainly interesting), but we try to minimize the resource consumption of a single flow. In particular, we try to minimize the bandwidth assignment for *L2R* schedulers while keeping the delay bound and buffer requirements the same as for the traditional *LR* schedulers as presented in Section II. This essentially means we fix the short-term-rate R_s of the *L2R* service curve to be the service rate R of the *LR* service curve as defined in (5). As a further constraint, we require the *L2R* service curves to have as the long-term rate the average rate r as specified by the application's TSpec. This requirement makes *L2R* optimal in the sense that they have minimal impact on delay-uncritical flows in, e.g., a best-effort service class, although it has to be noted that concurrent deterministic service flows might benefit from a more aggressive choice for the long-term-rate R_l. Since the scheduler's latency term L is dependent upon the actual scheduling algorithm used, it is not a design parameter for our resource optimization problem and thus the inflection point I remains as the sole variable to be optimized. Obviously, the smaller we can choose I in (9) the more rate resources we can save for other (less delay-critical) flows (possibly from other service classes).

B. Intuitive Choice of Inflection Point

Under all the prerequisites discussed in the previous section, a simple, intuitive choice of parameters for the service curve assignment of a flow is to choose the inflection point at $I = T + d_{max}$. This would definitely ensure that the delay bound is met by the short-term rate assignment before the service curve bends towards the long-term rate to meet the bandwidth guarantee. This intuitive choice of the inflection point results in the following service curve allocation

$$s^{simple}(t) = \begin{cases} R(t-L)^+ & t < T + d_{max} \\ rt + (R-r)(T+d_{max}) - RL & t \geq T + d_{max} \end{cases} \quad (10)$$

with R as in (5).

C. Optimal Choice of Inflection Point

However, taking into account that for the case $R \geq p$ (or, alternatively $d_{max} \leq \bar{d}$) the delay bound is taken on at the origin, we can actually improve (reduce) the service curve allocation by shifting the inflection point I further left (i.e., earlier). More accurately, I is chosen such that $s^I(T + d_{max}) = a(T)$. Thus, the delay bound is still ensured, because the horizontal deviation between arrival and service curve is now taken on at both possible locations, the origin and the burst duration T (see also Figure 2 for an illustration). The corresponding service curve is then given by:

$d_{max} > \bar{d}$:

$$s^{opt}(t) = \begin{cases} R(t-L)^+ & t < T + d_{max} \\ rt + (R-r)(T+d_{max}) - RL & t \geq T + d_{max} \end{cases} \quad (11)$$

with $R = \dfrac{pT + M + C}{d_{max} + T - D}$

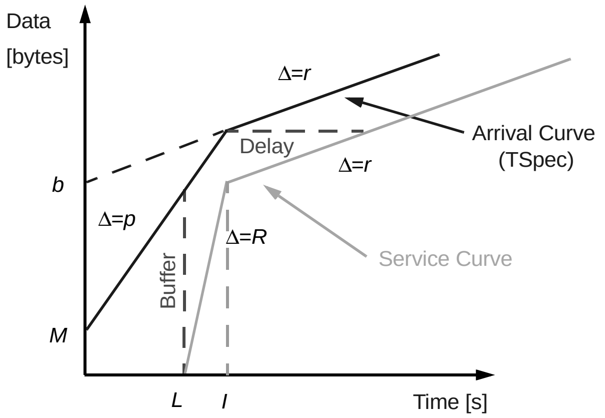

Figure 2: Network calculus for *L2R* schedulers.

$d_{max} \le \bar{d}$:

$$s^{opt}(t) = \begin{cases} R(t-L)^{+} & t < \dfrac{b - rd_{max} + RL}{R - r} \\ rt + b - rd_{max} & t \ge \dfrac{b - rd_{max} + RL}{R - r} \end{cases} \quad (12)$$

$$\text{with } R = \frac{M + C}{d_{max} - D} \quad (13)$$

Note that for $d_{max} > \bar{d}$ the optimal service curve is identical to the simple service curve of the previous section, but they differ for the case where $d_{max} \le \bar{d}$.

The service curve, s^{opt}, is optimal in the sense that the inflection point *I* is "shifted" as far left as possible without spending more resources neither on the short time-scales nor on the long time-scales of the service curve than the linear service curve, s^{linear}, and the simple non-linear service curve, s^{simple}. As discussed above, this restriction with respect to the short-term rate allocation on the optimal service curve is reasonable since the short time-scales of service have to deal with satisfying delay bounds and are thus the rate resources most heavily contended for.

The service curve s^{opt} is especially designed to just meet the delay bound also achieved by the *LR* service curve. A further requirement placed on it was that the buffer requirements in routers should not be increased. In fact, it can be shown that the buffer requirements are the same as for the *LR* scheduler service curve presented in Section II:

Theorem: The horizontal deviation to a TSpec for s^{opt} is the same as for s^{linear}.

Proof: As for the simple *LR* service curve the only locations where the vertical deviation *h* can be taken on is the scheduler latency *L* or the burst duration *T*. We need to consider different cases:

Case 1: $R < p$

s_{opt} is identical to s_{linear} over $[0, T + d_{max}]$, i.e., their vertical deviation must be the same since *L* and *T* fall within this interval. For *T* this is obvious, for *L* it is because *L* needs to be smaller than d_{max}, since otherwise the delay bound could not be satisfied.

Case 2: $R \ge p$

Case 2.1: $I > T$

First, note that $I > L$ per definition. s_{opt} and s_{linear} are identical over $[0, I]$, i.e. their vertical deviation must be the same since *T* and *L* fall within this interval.

Case 2.2. $I \le T$

This case is a little more complicated because s_{opt} and s_{linear} are no longer identical over an interval containing both possible locations for the vertical deviation to be taken on. For *L* it applies that $s_{opt}(L) = s_{linear}(L)$. If we can now show that

$$a_{TSpec}(L) - s_{opt}(L) \ge a_{TSpec}(T) - s_{opt}(T) \quad (14)$$

it follows that the vertical deviation for that case is taken on at *L*. Thus, since s_{opt} and s_{linear} are identical at *L* they have the same vertical deviation.

Hence, it remains to be shown that (14) is valid:

Call $x(t) = a_{TSpec}(t) - s_{opt}(t)$, we can thus rewrite (14) as

$$x(L) - x(T) \ge 0 \quad (15)$$

Note that

$$x(L) = M + C + pD \quad (16)$$

and

$$x(T) = rT + b - rT + b + rd_{max} \le r\left(\frac{M + C}{p} + D\right) \quad (17)$$

where the inequation captures the fact that for $R \ge p$ the largest d_{max} possible is $\bar{d} = \frac{M + C}{p} + D$.

If we now use (16) and (17) in (15) and note that $r \le p$ we obtain

$$\begin{aligned} x(L) - x(T) &\ge M + C + pD - \frac{r}{p}M - \frac{r}{p}C - rD \\ &\ge (p - r)D \ge 0 \end{aligned} \quad (18)$$

and can thus confirm (14).

❒

V. Discussion: Numerical Example

Reconsider the "low bandwidth, short delay"-type of flow taken as an example above in Section III.A, i.e., its TSpec = (2000, 1000, 8000, 500) [in bytes resp. bytes/s], its delay requirement d_{max} = 100 ms, and *C* = 2500 bytes and *D* = 2.371 ms. For such a flow the different service curve allocations are shown in Figure 3.

It is obvious that the linear service curve approach wastes a

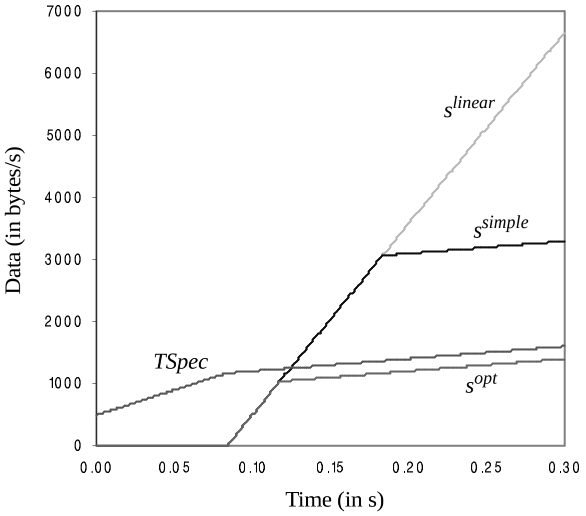

Figure 3: Different service curves for the given example flow.

lot of rate resources, the fact which led to the consideration of non-linear service curves. In comparison of the simple vs. the optimal non-linear service curve it can also be observed that the savings in terms of the shifting of the inflection point are considerable for this example: *I* is located about 65 ms earlier for the optimal than for the simple non-linear service curve. Let us perform a simple back-of-the-envelope calculation to illustrate the benefit of the optimal service curve allocation:

Assume we have a 10 Mb/s link and we have 40 flows of the above type (the maximum number that can be accepted). For the linear service curve no further guaranteed traffic could be accepted. For the simple *L2R* service curve, e.g., an additional 10 flows each with TSpec = (117000, 7800, 117000, arbitrary) and a delay requirement of 250 ms could be admitted on the link. For the optimal service curve, the same amount of additional flows could be admitted but now even with a delay bound of approx. 185 ms. Stated differently, the optimal *L2R* service curve allocation in this example would allow to admit more delay-sensitive flows compared to the simple *L2R* service curve and even more so compared to the linear service curve of *LR* schedulers.

VI. Conclusions

In this work, we have derived explicit formulas for optimal service curves based on bandwidth-delay decoupled service disciplines. Furthermore, we have shown their potential when compared to “intuitive / naive” choice of service curve parameters by a numerical example. Their full potential can be exploited especially in the case where “low bandwidth, short delay”-type of flows are multiplexed with less delay-critical, yet still delay-sensitive flows which can take advantage of the relative moderateness of the optimal service curves on longer time-scales.

As a future work item we perceive the investigation of globally optimal allocations for concurrent flows instead of the local per-flow perspective we pursued in this paper.

VII. References

[1] H. Zhang. Service Disciplines for Guaranteed Performance Service in Packet-Switching Networks. *Proceedings of the IEEE*, 83(10):1374–1396, October 1995.

[2] D. Stiliadis and A. Varma. Latency-Rate Servers: A General Model for Analysis of Traffic Scheduling Algorithms. *IEEE/ACM Transactions on Networking*, 6(5):611–624, October 1998.

[3] P. Goyal, S. S. Lam, and H. Vin. Determining End-to-End Delay Bounds in Heterogeneous Networks. In *Network and Operating System Support for Digital Audio and Video, 5th International Workshop, NOSSDAV’95*. Springer LNCS 1018, April 1995.

[4] J. Liebeherr, D. E. Wrege, and D. Ferrari. Exact Admission Control for Networks with Bounded Delay Service. *IEEE/ACM Transactions on Networking*, 4(6), December 1996.

[5] A. K. Parekh and R. G. Gallager. A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Single-Node Case. *IEEE/ACM Transactions on Networking*, 1(3):344–357, June 1993.

[6] A. K. Parekh and R. G. Gallager. A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Multiple Node Case. *IEEE/ACM Transactions on Networking*, 2(2):137–150, April 1994.

[7] H. Sariowan, R. L. Cruz, and G. Polyzos. SCED: A Generalized Scheduling Policy for Guaranteed Quality of Service. *IEEE/ACM Transactions on Networking*, 7(5):669–684, October 1999.

[8] I. Stoica, H. Zhang, and T. S. E. Ng. Hierarchical Fair Service Curve Algorithm for Link-Sharing, Real-Time and Priority Service. *ACM Computer Communication Review*, 27(4), October 1997. Proceedings of SIGCOMM’97 Conference.

[9] S. Shenker, C. Partridge, and R. Guerin. Specification of Guaranteed Quality of Service. Proposed Standard RFC 2212, September 1997.

[10] R. L. Cruz. Quality of Service Guarantees in Virtual Circuit Switched Networks. *IEEE Journal of Selected Areas in Communication*, 13(6):1048–1056, August 1995.

[11] J. Y. LeBoudec. Application of Network Calculus To Guaranteed Service Networks. *IEEE Transactions on Information Theory*, 44(3):1087–1096, August 1998.

[12] R. Agrawal, R. L. Cruz, C. Okino, and R. Rajan. Performance Bounds for Flow Control Protocols. *IEEE/ACM Transactions on Networking*, 7(3):310–323, June 1999.

[13] C. S. Chang. On Deterministic Traffic Regulation and Service Guarantees: A Systematic Approach by Filtering. *IEEE Transactions on Information Theory*, 44(3):1096–1107, August 1998.

[14] J. Wroclawski. The Use of RSVP with IETF Integrated Services. Proposed Standard RFC 2210, September 1997.

[15] K. Cho. A Framework for Alternate Queueing: Towards Traffic Management by PC-UNIX Based Routers. In *Proceedings of USENIX 1998 Annual Technical Conference, New Orleans, USA*, June 1998.

[16] M. Karsten, J. Schmitt, and R. Steinmetz. Implementation and Evaluation of the KOM RSVP Engine. In *Proceedings of the 20th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM’2001)*, pages 1290–1299. IEEE, April 2001.