This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE INFOCOM 2007 proceedings.

A Min-Plus System Interpretation of Bandwidth Estimation

Jörg Liebeherr
Department of ECE
University of Toronto

Markus Fidler
Multimedia Communications Lab
Technische Universität Darmstadt

Shahrokh Valaee
Department of ECE
University of Toronto

***Abstract*— Significant research has been dedicated to methods that estimate the available bandwidth in a network from traffic measurements. While estimation methods abound, less progress has been made on achieving a foundational understanding of the bandwidth estimation problem. In this paper, we develop a min-plus system theoretic formulation of bandwidth estimation. We show that the problem as well as previously proposed solutions can be concisely described and derived using min-plus system theory, thus establishing the existence of a strong link between network calculus and network probing methods. We relate difficulties in network probing to potential non-linearities of the underlying systems, and provide a justification for the distinctive treatment of FIFO scheduling in network probing.**

I. Introduction

The benefits of knowing how much network bandwidth is available to an application has motivated the development of techniques that infer the available bandwidth from traffic measurements [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12]. Even though the number of techniques available today is significant and much empirical experience has been gained, less progress has been made towards a foundational understanding of measurement based methods for estimating the available bandwidth. Recent stochastic analyses point out that an improved understanding of the principles of bandwidth estimation could lead to better methods [13], [14].

In this paper, we pursue a different avenue to reason about available bandwidth estimation. We view bandwidth estimation as the analysis of a deterministic min-plus linear system. This approach enables us to give mathematical derivations that show how existing bandwidth estimation methods infer information about a network. Also, we are able to reason which bandwidth estimation methods can extract the most information from a network. Finally, we can show that some key difficulties encountered when measuring available bandwidth become evident in a system theoretic view, and that heuristics that are applied in practice can be explained in terms of min-plus system theory.

We view bandwidth estimation as the problem of determining an unknown function that describes the available bandwidth, based on measurements of a single sequence of probing packets or passive measurements of a single sample path of arrivals. Given a set of (deterministic) timestamps that record the transmission times of probing packets and their arrival times at the destination, we show how and how much information can be extracted about the network.

We show that estimating the available bandwidth in a general network corresponds to solving a max-min optimization problem. The problem becomes more tractable when the network satisfies the property of 'min-plus linearity'. We show that many existing estimation techniques can be accurately characterized if we interpret them as analyzing a network with linear input-output relationships. We explain why available bandwidth estimation is difficult if the underlying network uses FIFO scheduling by showing that the input-output relation of FIFO systems is decomposable into a min-plus linear and a disjoint non-linear region. Here, the crossing of these regions coincides with the available bandwidth.

The arguments in this paper draw from known relationships between linear system theory and the network calculus. The success in describing relatively complex probing schemes using min-plus algebra hints at a possibly stronger link between bandwidth estimation and network calculus. A limitation of our work is that we only consider a single packet trace or sequence of probing packets. Since, in principle, a system theoretic approach does not preclude a statistical analysis, where each probe is interpreted as a random sample, we believe that this limitation can be eventually removed.

The remainder of this paper is structured as follows. In Section II, we discuss bandwidth estimation methods and other related work. In Section III, we review the min-plus linear system interpretation of the deterministic network calculus. In Section IV, we formulate bandwidth estimation as the solution to an inversion problem in min-plus algebra. In Section V, we derive solutions to compute the inversion, and relate them to probing schemes from the literature. In Section VI, we justify how these probing schemes can be applied in networks that are not min-plus linear. We present brief conclusions in Section VII.

II. Available Bandwidth Estimation Techniques

The goal of bandwidth estimation is to infer from measurements a reliable estimate of the unused capacity at a multi-access link, a single switch, or a network path. The available bandwidth of a network is often specified as $A = \min_i(C_i - \lambda_i)$, where C_i and λ_i are the capacity and total traffic, respectively, on link i of a network path. The majority of estimation methods monitor the transmission of control (probe) packets. We call these methods active monitoring or probing schemes. An alternative approach is to take passive

0743-166X/07/$25.00 ©2007 IEEE

measurements by monitoring live data traffic in a network. The latter group is the preferred approach for measurement based admission control (MBAC), which seeks to determine if a network has sufficient resources to support minimal service requirements [15], [16]. In comparison to passive measurements, probing schemes have an additional degree of freedom since they can determine the transmission pattern of probing packets.

Active monitoring techniques typically generate probing traffic as packet pairs or packet trains. Packet pairs consist of two packets with a defined spacing, and packet trains consist of more than two packets. Since it was first suggested in [17], [18], packet pair probing has evolved significantly, and has been used for estimating the bottleneck capacity (e.g., *Bprobe* [1], *CapProbe* [10]) as well as the available bandwidth (e.g., *Spruce* [2]). The rationale behind these methods builds on the relation of packet dispersion and available bandwidth resources, i.e., packet pairs with a defined gap may be spaced out on slow or loaded links and thus carry information about the network path. Packet train methods (e.g., *PBM* [11], *Cprobe* [1], *pathrate* [12]) seek to improve the accuracy of bandwidth estimation over packet pairs.

More recently proposed schemes, including *pathload* [4], [5], *pathvar* [9], *TOPP* [3], *PTR/IGI* [6], *pathchirp* [7], and *BFind* [8], adaptively vary the rate of probing traffic to induce congestion in the network. This has been found to increase the fidelity of estimation methods. For example, *pathload* uses a sequence of constant rate packet trains, and increases the transmission rate of consecutive trains until they converge to the available bandwidth. *Pathchirp* uses packet trains, referred to as chirps, with an exponentially decreasing inter-packet gap. Here, the network is probed over a range of rates similar to *pathload*, however, the rate scan is performed within a single packet train.

Some estimation techniques are designed with an assumption that the network as a whole exhibits the behavior of a single queueing system with cross traffic. Often it is assumed that the network behaves as a single FIFO system [2], [3], [6], [7], [19]. This is justified by the particular packet dispersion of FIFO systems which is matched by empirical data [19]. Since a flow in an overloaded FIFO system may receive a share of the capacity that exceeds the available bandwidth, it has been found that the best estimates are obtained if the probing traffic increases the load close to, but not beyond, an overloaded state. An analytical investigation in [14], [20] showed that probing schemes can be further improved by accounting for the random fluctuations of traffic.

We note that links between network calculus and bandwidth estimation, have been made before mostly in the context of MBAC [21], [22], [23].

III. Min-Plus Linear System Theory for Networks

This section reviews the linear system representation of networks and introduces needed concepts and notation. We consider a continuous-time setting.

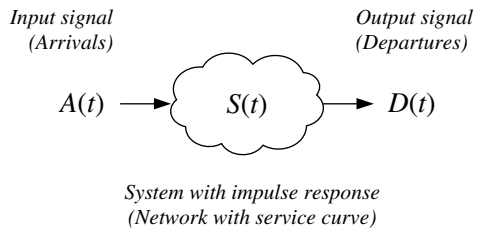


Fig. 1. Linear Time-Invariant system and min-plus linear network.

Classical linear system theory deals with linear time-invariant (LTI) systems with input signal $A(t)$ and output signal $D(t)$ (see Fig. 1). Linear means that for any two pairs of input and output signals (A_1, D_1) and (A_2, D_2), any linear combination of input signals $b_1A_1(t) + b_2A_2(t)$ results in the linear combination of output signals $b_1D_1(t) + b_2D_2(t)$. Time-invariant means that for any pair of inputs and outputs (A, D), a time-shifted input $A(t-\tau)$ results in a shifted output $D(t-\tau)$.

Let $S(t)$ be the impulse response of the system, that is, the output signal generated by the system if the input signal is a unity (Dirac) impulse at time zero. The basic property of an LTI system is that it is completely characterized by its impulse response, where the output of the system is expressed as the convolution of the input signal and the impulse response:

$$D(t) = \int_{-\infty}^{\infty} A(\tau)S(t-\tau)d\tau =: A * S(t).$$

A. Min-Plus Algebra in the Network Calculus

A significant discovery of networking research from the 1990's is that networks can often be viewed as linear systems, when the usual algebra is replaced by a so-called min-plus algebra [24], [25], [26]. In a min-plus algebra [27], addition is replaced by a minimum (we write infimum) and multiplication is replaced by an addition. Similar to LTI systems, a min-plus linear system is a system that is linear under the min-plus algebra. This means that a min-plus linear combination of input functions $\inf\{b_1 + A_1(t), b_2 + A_2(t)\}$ results in the corresponding linear combination of output signals $\inf\{b_1 + D_1(t), b_2 + D_2(t)\}$. In min-plus system theory, the burst function

$$\delta(t) = \begin{cases} \infty\ , & \text{if } t > 0\ , \\ 0\ , & \text{otherwise}\ , \end{cases} \tag{1}$$

takes the place of the Dirac impulse function.

Let $S(t)$ be the impulse response, that is, the output when the input is the burst function $\delta(t)$. Any time-invariant min-plus linear system is completely described by its impulse response, and the output of any min-plus linear system can be expressed as a linear combination of the input and shifted impulse responses by

$$D(t) = \inf_{\tau}\{A(\tau) + S(t-\tau)\} =: A * S(t).$$

As in [24], [25], [26] we use the convention that input and output signals in the min-plus linear system theory are non-decreasing non-negative functions. In analogy to LTI systems,

this operation is referred to as convolution of the min-plus algebra [27].[1] Conversely, if there exists a function $S(t)$ such that $D(t) = A * S(t)$ for all pairs (A, D), then it follows that the system is min-plus linear.

The min-plus convolution shares many properties with the usual convolution, e.g., it is commutative and associative. The associativity of min-plus convolution is of particular importance since it implies an easy way of concatenating systems in series. Given a tandem of two min-plus linear systems $S_1(t)$ and $S_2(t)$, the output can be computed iteratively as $D(t) = (A * S_1) * S_2(t)$ and, with associativity, $D(t) = A * (S_1 * S_2)(t)$ holds. This leads to the important observation that the tandem system is equivalent to a single system with impulse response $S(t) = S_1 * S_2(t)$.

The observation that some networks can be adequately modeled by a min-plus linear system led to the min-plus formulation of the network calculus [24], [25], [26]. Here, a system is a network element or entire network, input and output functions A and D are arrivals and departures, respectively, and the impulse response S, called the *service curve*, represents the service guarantee by a network element. Network elements that are known to be min-plus linear include a work-conserving constant rate link ($S(t) = C\,t$, where C is the link capacity), a shaper ($S(t) = \sigma + \rho\,t$, where σ is a burst size and ρ is a rate), and a rate-latency server ($S(t) = r\,(t - d)_+$, where r is a rate, d is a delay, and $(x)_+ = \max(x, 0)$), and their concatenations.

The relevance of the network calculus as a tool for the analysis of networks results from an extension of its formal framework to networks that do not satisfy the conditions of min-plus linearity. Nonlinear systems implement more complex mappings Π of arrival to departure functions $D(t) = \Pi(A)(t)$. In the network calculus, these are replaced by linear mappings that provide bounds of the form $D(t) \geq A * \underline{S}(t)$ or $D(t) \leq A * \overline{S}(t)$ ([26], pp. *xviii*). Here, $\underline{S}$ is referred to as a *lower service curve* and $\overline{S}$ is referred to as an *upper service curve*, indicating that they are bounds on the available service. In a min-plus linear system, the service curve S is both an upper and a lower service curve ($S = \underline{S} = \overline{S}$), which is therefore frequently referred to as *exact service curve*.

B. Legendre transform in Min-Plus Linear Systems

In linear system theory, the Fourier transform of $f(t)$, denoted by $\mathcal{F}_f(\omega)$, establishes a dual domain, the frequency domain, for analysis of LTI systems. In the frequency domain, the Fourier transform turns the convolution to a multiplication, that is, $\mathcal{F}_{f*g}(\omega) = \mathcal{F}_f(\omega) \cdot \mathcal{F}_g(\omega)$.

In min-plus linear systems, the *Legendre transform*, also referred to as convex Fenchel conjugate, plays a similar role. The Legendre transform of a function $f(t)$ is defined as

$$\mathcal{L}_f(r) = \sup_{\tau}\{r\tau - f(\tau)\}.$$

Since r can be interpreted as a rate, one may view the domain established by the Legendre transform as a rate domain. The Legendre transform takes the min-plus convolution to an addition [27], [28], that is,[2]

$$\mathcal{L}_{f*g} = \mathcal{L}_f + \mathcal{L}_g\ . \tag{2}$$

Other properties of the Legendre transform that we exploit in this paper are that, for convex functions f, we have

$$\mathcal{L}(\mathcal{L}_f) = f\ . \tag{3}$$

In other words, a convex function f can be recovered from $\mathcal{L}_f$ by reapplying the Legendre transform [28]. In general, we only have

$$\mathcal{L}(\mathcal{L}_f) \leq f \quad \text{and} \quad \mathcal{L}(\mathcal{L}_f) = \text{conv}_f\ , \tag{4}$$

where conv_f denotes the convex hull of f, defined as the largest convex function smaller than f.

Another property that will be used is that the Legendre transform reverses the order of an inequality, i.e.,

$$f \geq g \Rightarrow \mathcal{L}_f \leq \mathcal{L}_g\ . \tag{5}$$

The statement is an equivalency when g is convex. Applications of the Legendre transform in the network calculus have been previously studied in [21], [29], [30], [31].

IV. A Min-Plus Algebra Formulation of the Bandwidth Estimation Problem

We view a network as a time-invariant min-plus linear or non-linear system that converts input signals (arrivals) into output signals (departures) according to a fixed but unknown service curve S. The service curve of the network expresses the available bandwidth, which can be a constant-rate or a more complex function. Measurements of a network probe, defined as a sequence of at least two packets, can be characterized by an arrival function $A^p(t)$ and a departure function $D^p(t)$, where the functions represent the cumulative number of bits seen in the interval $[0, t]$. These functions are constructed from timestamps of the transmission and reception of packets, and from knowledge of the packet size. In Fig. 2 we illustrate a network probe consisting of five packets of equal size with fixed spacing between consecutive packets. The vertical distance between arrivals and departures can be viewed as a virtual backlog $B(t) = A^p(t) - D^p(t)$. The horizontal distance can be viewed as a virtual delay $W(t)$.

Representing the network by a min-plus linear system, we interpret a probing scheme as trying to determine from a specific sample of functions A^p and D^p an *a priori* unknown lower service $\underline{S}^u$, such that $D \geq A * \underline{S}^u$ holds for all pairs (A, D) of arrival and departure functions. Since the estimate of the available bandwidth should not be overly pessimistic, the goal of a probing scheme is to select a maximal $\underline{S}^u(t)$, i.e., there is no other lower service curve larger than $\underline{S}^u(t)$ that satisfies the definition.[3] One problem in devising a probing

[1] We re-use the symbol for notational simplicity. The context makes this slight abuse of notation non-ambiguous.

[2] Whenever possible, from now on we use shorthand notation f to mean '$f(t)$ for all $t \geq 0$', and $\mathcal{L}_f$ to mean '$\mathcal{L}_f(r)$ for all $r \geq 0$'.

[3] We define a partial ordering of service curves, such that $S_1 \leq S_2$ iff. $S_1(t) \leq S_2(t)$ for all t.

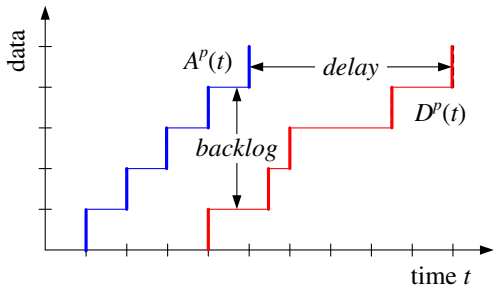


Fig. 2. Example arrival and departure function of a probe of five packets.

scheme lies in the selection of the probing pattern, i.e., a function A^p that reveals a maximal service curve.

Taking a step back and looking at the overall problem, bandwidth estimation is trying to find a service curve $\underline{S}^u$ that has the best worst-case performance. This corresponds to expressing $\underline{S}^u$ as the solution to the following optimization problem:

$$\begin{array}{ll} \text{MAXIMIZE} & \underline{S} \\ \text{SUBJECT TO} & D(t) \geq \inf_\tau\{A(\tau) + \underline{S}(t-\tau)\}, \\ & \qquad \forall t \geq 0, \text{ for all pairs } (A, D). \end{array}$$

This problem has the structure of a max-min optimization, which is fundamentally hard. In addition, since service curves only form a partial ordering, there may not be an optimal solution, but only solutions that cannot be further improved.

The bandwidth estimation problem is easier when the network can be described by a min-plus linear system. As we will see in Section VI, some non-linear networks, such as FIFO systems, are min-plus linear under low load conditions. Recalling that a system is min-plus linear if it can be described by an exact service curve, the bandwidth estimation problem is reduced to solving the inversion of

$$D(t) = A * S^u(t) \text{ for all } t \geq 0.$$

If we can take a measurement of A^p and D^p which solves the equation for S^u, then, due to min-plus linearity, we have a solution for all possible arrival and departure functions. From Section III, we can infer that a solution is obtained by using the burst function of Eq. (1) as probing pattern, i.e., $A^p(t) = \delta(t)$. This follows since the service curve is the impulse response of a min-plus system, that is, $D^p(t) = \delta * S^u(t) = S^u(t)$. However, sending a probe as a burst function is not practical, since it assumes the instantaneous transmission of an infinite sized packet sequence. While a burst function can be approximated by a sufficiently large back-to-back packet train, a high-volume transmission of probes consumes network resources and interferes with other packet traffic. More importantly, the service curve of a burst function (or its approximation), may cause some networks that operate in a min-plus linear regime to become non-linear. The observation that large packet trains can lead to unreliable estimates has been noted in the literature [12].

In the next section, we present derivations for three bandwidth estimation methods in min-plus linear systems. We are able to relate two of these methods to previously proposed probing schemes. Some schemes can be applied to certain non-linear systems.

We conclude this section with remarks on some general aspects of probing schemes and their representations in min-plus linear system theory.

• **Timestamps and asynchrony of clocks:** When clocks at the sender and receiver of a probing packet are perfectly synchronized, and the sender includes the transmission time into each probing packet, the receiver can accurately construct the functions A^p and D^p. In practice, however, clocks are not synchronized. When clocks have a fixed offset (but no drift), the arrival function A^p can be viewed as being time-shifted by an unknown offset T. In the min-plus algebra a time-shift can be expressed by a convolution, i.e., $A^p(t-T) = A^p * \delta_T(t)$ where $\delta_T(t) = \delta(t-T)$. Here, the convolution of arrival function and service curve becomes $(A^p * \delta_T) * \underline{S}^u$, which due to associativity and commutativity of the convolution operation, can be rewritten as $A^p * (\underline{S}^u * \delta_T)$. Hence, when the offset is fixed but unknown, even an ideal probing scheme can only compute a service curve that is a time-shifted version of the actual service curve of the network. Drifting clocks make the problem harder. Many bandwidth estimation schemes circumvent the problem of asynchronous clocks by returning probes to the sender [1], [8], or by only recording time differences of incoming probes [2], [3], [4], [6], [7]. A moment's consideration shows that knowledge of the differences between the transmission and arrival of probing packets has the same limitations as dealing with an unknown clock offset T between the sender and receiver of probing packets.

• **Packet pairs:** The arrival and departure functions of a packet pair have each only three points, i.e., the origin and the two timestamps related to the packet pair. If it can be assumed that the service curve has a certain shape, e.g., a rate-latency curve $S(t) = r \cdot (t-d)_+$, the service curve can be recovered. In the absence of such an assumption, packet pair methods may not be able to recover more complex service curves.

V. Min-Plus Theory of Network Probing Methods

In this section, we derive bandwidth estimation methods as solutions to finding an unknown service curve for a min-plus system. We make a number of idealizing assumptions. First, we consider a fluid flow view of traffic and service. This assumption can be relaxed at the cost of additional notation. Unless stated otherwise, we assume that the network represents a min-plus linear system. This assumption will be partially relaxed in Section VI. We generally assume that accurate timestamps for transmission and arrival of probes are feasible. If measurements only record time differences between events or include an unknown clock offset between sender and receiver, the computed service curves need to be time shifted by some constant value.

A. Passive Measurements

We first try to answer the question: *How much information about the available bandwidth can be extracted from passive*

measurements of traffic? To provide an answer we first introduce the deconvolution operator of the min-plus algebra, which is defined for two functions f and g by

$$f \oslash g(t) = \sup_{\tau}\{f(t+\tau) - g(\tau)\}.$$

The deconvolution operation is *not* an inverse to the convolution ($g \neq f \oslash (f * g)$), however, it has aspects of such an inverse. This is expressed in the following duality statement from [26], which states that for functions f, g and h, the following equivalency holds:[4]

$$f \leq g * h \quad \Leftrightarrow \quad h \geq f \oslash g. \tag{6}$$

We will exploit this property to formulate the following lemma.

Lemma 1 *For two non-decreasing non-negative functions g and h, we have*

$$((h * g) \oslash g) * g = h * g \ .$$

Proof: Let us define $f = g * h$ and $\tilde{h} = f \oslash g$. From Eq. (6) we can conclude that $f \leq g * \tilde{h}$. By definition of f, we see from Eq. (6) that $h \geq f \oslash g$. By definition of $\tilde{h}$, this gives us $h \geq \tilde{h}$. From $h \geq \tilde{h}$ and $f = g * h$ we get $f \geq g * \tilde{h}$. Combining the two statements gives us $f = g * \tilde{h}$. Now, by inserting the definition $\tilde{h} = f \oslash g$, we obtain $f = g * (f \oslash g)$. Inserting the definition $f = g * h$ yields $g * h = g * ((g * h) \oslash g)$. Reordering the expression using commutativity of the min-plus convolution completes the proof. ■

The lemma justifies the following passive measurement scheme. Let us denote the arrival and departure functions measured from a traffic trace of one or more flows by A^{tr} and D^{tr}. By assumption of linearity, we know that $D^{tr} = A^{tr} * S$ holds, but the shape of S is unknown. Suppose we compute a function $\tilde{S}$ from the trace as the deconvolution of the departures and the arrivals, i.e., we set

$$\tilde{S} = D^{tr} \oslash A^{tr} \ . \tag{7}$$

Then, we can conclude with Lemma 1 that

$$D^{tr} = A^{tr} * \tilde{S} \ . \tag{8}$$

With the duality property from Eq. (6) we obtain with Eq. (7) and Eq. (8) that

$$\tilde{S} \leq S \ .$$

Hence, the estimate of the service curve $\tilde{S}$ is a lower service curve, such that for all pairs of arrival and departure functions (A, D), we have $D \geq A * \tilde{S}$. Since, from Eq. (8), $\tilde{S}$ can completely reconstruct the departure function from the arrival function, we can conclude that $\tilde{S}$ is the best possible estimate of the actual service curve that can be justified from measurements of A^{tr} and D^{tr}, in the sense that it extracts the most information from the measurements.

[4]We use shorthand notation $f = g * h$ to mean '$f(t) = (g * h)(t)$ for all $t \geq 0$'.

TABLE I
PARAMETERS OF ON-OFF SOURCES.

Scenario	high load			low load		
Burstiness	high	med	low	high	med	low
Number of sources	1	5	25	1	5	25
Source peak rate [Mbps]	200	40	8	200	40	8
Total average rate [Mbps]	20	20	20	10	10	10

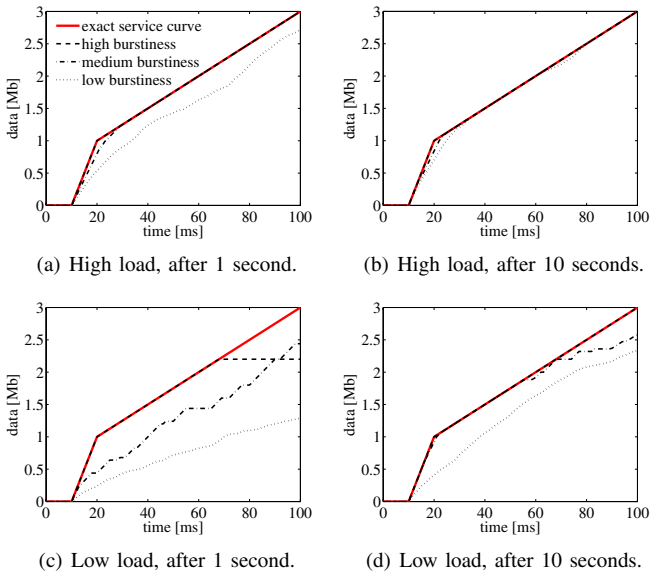
The main drawback of this method is that it can only be applied to linear networks. For networks that do not satisfy min-plus linearity, i.e., that can only be described by a lower service curve ($D \geq A * \underline{S}$) or upper service curve ($D \leq A * \overline{S}$), $\tilde{S}$ only computes a (not useful) lower bound for an upper service curve $\overline{S}$. As another remark, note that Lemma 1 does not help us with designing a probing scheme, since it does not provide guidance how to select the traffic A^p for the network probes.

For illustration of the passive measurement scheme, we now present numerical results of an (idealized) fluid flow scenario of a min-plus linear system, which is governed by a service curve $S(t) = (b + rt) * (R[t - T]^+)$. The system represents a network that regulates the ingress with a leaky-bucket with parameters b and r, and the service corresponds to a latency-rate service curve with delay T and rate R. We set $b = 0.75$ Mb, $r = 25$ Mbps, $R = 100$ Mbps, and $T = 10$ ms.

As traffic trace, we use an arrival sample path that represents the aggregate arrivals from a set of statistically independent On-Off traffic sources. In the *On state*, a source generates traffic with a given peak rate, and in the *Off state* no data is generated. Every millisecond, an active source leaves the On state with probability p, and an Off source becomes active with probability q. This choice of traffic enables us to evaluate the sensitivity of the passive measurement method with respect to the burstiness of the trace, the fraction of available bandwidth that is utilized by the flows, and the length of the measurement period.

The parameters are depicted in Table I. In the *high load* setting, we set $p = 0.09$ and $q = 0.01$, resulting in a total arrival rate of 20 Mbps. In *low load*, we set $p = 0.19$ and $q = 0.01$, which leads to an average total traffic rate of 10 Mbps. We control the burstiness of the traffic by increasing the number of flows, and accordingly decrease the peak rate of each flow. Due to statistical multiplexing, an aggregate of multiple On-Off sources is less bursty than a single flow with the same peak and average rate. In our plots burstiness levels of *high*, *medium*, and *low* correspond to a trace with 1, 5, and 25 sources.

In Fig. 3(a)-3(d) we show the estimates of the lower service curves $\tilde{S}$ obtained with the above method, and compare them to the actual service curve S, indicated as a thick (red) line in each graph. The length of the measurement is taken for 1 second (plots on the left), 10 seconds (plots on the right). In all plots, we see that burstier traffic leads to better estimates of the service curve. This is expected since the burstiest traffic, i.e., a burst impulse, can perfectly recover S (see Section IV).



(a) High load, after 1 second. (b) High load, after 10 seconds.

(c) Low load, after 1 second. (d) Low load, after 10 seconds.

Fig. 3. Bandwidth estimation for passive measurement of a traffic trace.

For the same reason, the estimates improve when the traffic trace has a higher utilization of the available bandwidth. Observe that all estimates improve with increasing length of the evaluation period. This follows from the definition of the supremum in the min-plus deconvolution operation.

Since the presented method is ideal in the sense that it computes the largest service curve (available bandwidth) that can be justified from a given traffic trace, our method will perform no worse than existing methods, e.g., from the MBAC literature [15].

B. Rate Scanning

Next we consider probing schemes where sequences of packet trains are transmitted at different rates, such as *pathload* [4], [5]. We refer to these techniques as rate-scanning techniques. We provide a justification for this technique using min-plus system theory.

Given arrival and departure functions A and D, using the earlier definition of backlog, the maximum backlog can be computed as

$$B_{max} = \sup_t \{A(t) - D(t)\}.$$

If the arrivals are a constant rate function, that is, $A(t) = rt$, and the network satisfies min-plus linearity, we can write B_{max} as a function of r as follows:

$$\begin{aligned} B_{max}(r) &= \sup_t \{rt - \inf_\tau \{r\tau + S(t-\tau)\}\} \\ &= \sup_t \{\sup_\tau \{r(t-\tau) - S(t-\tau)\}\} \\ &= \sup_t \{rt - S(t)\}. \end{aligned}$$

The first line uses that the departures in min-plus linear system theory can be characterized by $D = A * S$. The second line moves the infimum in front of the subtraction, where it becomes a supremum. The third line is simply a substitution.

Recalling the definition of the Legendre transform from Subsection III-B, the right hand side of the last equation can be written as the Legendre transform of S, that is, $B_{max}(r) = \mathcal{L}_S(r)$. This relation has been observed in [29], [32], [31]. We take a further step by applying this relation in the reverse transform. Due to Eq. (3), we have for convex service curves S that

$$S(t) = \mathcal{L}(\mathcal{L}_S)(t) = \mathcal{L}_{B_{max}}(t) = \sup_r \{rt - B_{max}(r)\} \ .$$

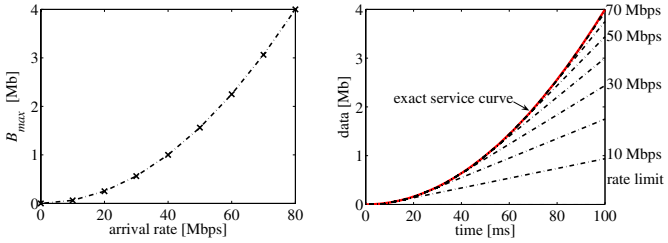
Thus, every convex service curve can be recovered by measurements of the maximum backlog B_{max} by constant-rate probe traffic that is sent at varying rates. (For service curves that are not convex one recovers, using Eq. (4), a lower bound for the service curve.) The interpretation of rate scanning is that each constant bit rate stream with rate r reveals one point $B_{max}(r)$ of the service curve in the Legendre domain $\mathcal{L}_S(r)$.

In practice, a rate scanning method specifies a *rate increment*, which sets the increase of the rate between packet trains, a *rate limit*, which sets the maximum rate at which the network is scanned, and the length of the packet trains. The service curve calculated via rate scanning consists of piecewise linear segments. The choice of the rate increment determines the length of the segments, and, in this way, the accuracy of the computed service curve. Without offering a proof or further data, we note that (under loose assumptions) rate scanning is capable of tracking a convex service curve, up to a time where the derivative of the service curve reaches the rate limit. The higher the maximum rate, the more information about the service curve is recovered.

A criterion for picking the rate limit suggested by our derivations is to stop rate scanning when increasing the scanning rate does not yield an improvement of the service curve. This criteria may fail when the underlying network is not min-plus linear. Pathload [4], [5] is a prototypical example of a rate scanning method. It uses an iterative procedure which varies the rate r of consecutive packet trains until measured delays indicate an increasing trend. Such a trend is interpreted as reflecting that the rate has exceeded the available bandwidth. In Section VI we remark that this criteria can be justified in non-linear systems that behave linearly at low loads.

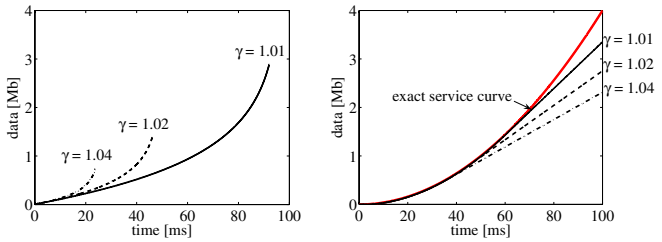
The number of packets in a packet train must be large enough so that the maximum backlog can be accurately measured. Under the assumptions of the rate scanning technique, i.e., the arrival function is a constant rate function and the service curve is convex, the backlog $B(t)$ is a concave function. Hence, if increases to the backlog caused by the packets of a train are sufficiently small, additional packets of the train do not provide new information.

In Fig. 4(a) we present an example of the rate scanning approach for a fluid-flow service curve with a quadratic form $S(t) = 0.4t^2$. In the example, rate scanning is performed at rates $10,\ 20, \ldots, 80$ Mbps. In Fig. 4(a), we plot the maximum backlog observed for each scanning rate. The function $B_{max}(r)$ is constructed by connecting the measured data points by lines. (For any rate r exceeding the rate limit we



(a) Maximum Backlog $B_{max}(r)$. (b) Results with different rate limits.

Fig. 4. Service curve estimation with rate scanning.



(a) Rate Chirps. (b) Results with different spread factors.

Fig. 5. Service curve estimation with rate chirps.

make the conservative assumption that $B_{max}(r) = \infty$. The assumption gives us a Legendre transform for all rate values.) In Fig. 4(b), we show the service curves that are obtained with different rate limits. The higher the rate limit, the more accurate the results.

Note that both the plots of the backlog in Fig. 4(a) and the service curves in Fig. 4(b) consist of linear segments. Decreasing the increment of the rate will improve the accuracy of the service curve. If we compared the service curves from the rate scanning method with the previous subsection, we would observe that the rate scanning method with rate limit r_{max} generates similar results as passive measurement of constant rate traffic with the same rate. The difference of results would consist of inaccuracies due to the approximation of the service curve by linear segments.

C. Rate Chirps

The need of rate scanning to measure a possibly large number of packet sequences has motivated a method where measurements are done with a single packet train, with an exponentially decreasing inter-packet spacing. The approach, proposed in [7], takes inspiration from chirp signals in signal processing, which are signals whose frequencies change with time. We refer to this approach as *rate chirp*, since the decreased gap between packets corresponds to an increase of the transmission rate.

We will show that a rate chirp scheme can be justified by properties of the min-plus system theory, specifically, properties of the Legendre transform.

Suppose we have a lower service curve $\underline{S}$ satisfying $D \geq A * \underline{S}$ for all pairs (A, D). Taking the Legendre transform we obtain with the order reversing property of Eq. (5) and with Eq. (2), that

$$\mathcal{L}_D \leq \mathcal{L}_{A*\underline{S}} = \mathcal{L}_A + \mathcal{L}_{\underline{S}} \ .$$

We can re-write this as

$$\mathcal{L}_{\underline{S}} \geq \mathcal{L}_D - \mathcal{L}_A \ ,$$

as long as the difference $\mathcal{L}_D(r) - \mathcal{L}_A(r)$ is defined for all r. A sufficient condition is that $\mathcal{L}_A(r) < \infty$, since it prevents both transforms $\mathcal{L}_D$ and $\mathcal{L}_A$ from becoming infinite at the same value of r. Another application of Eq. (5) yields

$$\mathcal{L}(\mathcal{L}_{\underline{S}}) \leq \mathcal{L}(\mathcal{L}_D - \mathcal{L}_A) \ .$$

If the system is min-plus linear, that is, $D = A * S$, we get,

$$\mathcal{L}(\mathcal{L}_S) = \mathcal{L}(\mathcal{L}_D - \mathcal{L}_A) \ .$$

If S is convex, then by Eq. (3), we have $S = \mathcal{L}(\mathcal{L}_D - \mathcal{L}_A)$.

This provides us with a justification for *pathchirp* [7] as a probing method. If we depict the transmission of a packet chirp as a fluid flow function, we see that it grows to an infinite rate, thus, yielding a Legendre transform that is finite for all rates. By measuring arrivals and departures of the chirp, denoted by A^{chap} and D^{chrp}, we can compute a function $\tilde{S}$ by

$$\tilde{S}(t) = \mathcal{L}(\mathcal{L}_{D^{chrp}} - \mathcal{L}_{A^{chrp}})(t) \ . \tag{9}$$

If the network satisfies $D = A*S$ for all arrivals, then the right hand side of Eq. (9) computes $\mathcal{L}(\mathcal{L}_{\underline{S}})$. Then, with Eq. (4), we obtain $\tilde{S} \leq S$, which tells us that $\tilde{S}$ is a lower service curve that satisfies $D \geq A * \tilde{S}$ for any traffic with arrival function A and departure function D. If, in addition, S is convex we have $\tilde{S} = S$, and we can recover the service curve.

In practice, a rate chirp stops sending packets at some maximum rate. Suppose that packets of a chirp are transmitted in a time interval $[0, t^A_{max}]$, and that the observations of D are made in $[0, t^D_{max}]$. To make a practical rate chirp comply with the formal requirements of the above equations, we define the following extension:

$$\tilde{A}^{chrp}(t) = \begin{cases} A^{chrp} \ , & \text{if } 0 \leq t \leq t^A_{max} \ , \\ \infty \ , & \text{if } t > t^A_{max} \ , \end{cases}$$

$$\tilde{D}^{chrp}(t) = \begin{cases} D^{chrp} \ , \text{if } 0 \leq t \leq t^D_{max} \ , \\ D^{chrp}(t^D_{max}) + (t - t^D_{max}) \frac{dD^{chrp}}{dt}(t^D_{max}) \ , \\ \qquad \text{if } t > t^D_{max} \ . \end{cases}$$

The arrival function is simply set to ∞ past the last measurement. The departure function is continued at a rate that corresponds to its slope at the time of the last measurement. For convex service curves S, the above extensions are conservative.

In Fig. 5(a) we show several rate chirps for a network probe. The rate chirp consists of a sequence of probing packets of 1200 Bytes that are transmitted at an increasing rate, starting at 10 Mbps and growing to 200 Mbps. The rate is increased by reducing the elapsed time between the transmission of the first bit of two consecutive packets, by a constant factor γ, which is called the spread factor in [7]. Larger values for γ lead to shorter chirps that grow fast to the maximum rate. In Fig. 5(b), we show the service curves computed from the chirps in Fig. 5(a). The actual service curve is $S(t) = 0.4t^2$, indicated as a thick (red) line in the figure. It appears that a

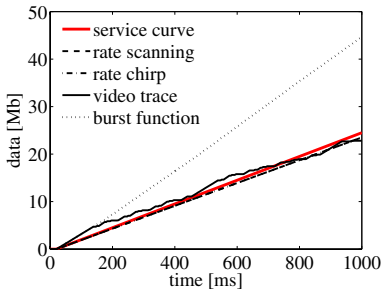


Fig. 6. Service curve estimation of a FIFO network.

path chirp with a smaller spread factor γ, which transmits more packets over longer time interval, leads to better estimates of the service curve.

VI. Estimation in nonlinear systems: the FIFO system

Extending bandwidth estimation to systems that are not min-plus linear, i.e., are not described by an exact service curve, raises difficult questions. First, the formulation of bandwidth estimation for general networks from Section IV shows that the structure of the estimation problem becomes hard. More so, since in networks with non-linearities the network service may depend on the network traffic, knowledge of the available bandwidth may not help with predicting network behavior, or may even be ill-defined.

However, there are networks that can be decomposed into a min-plus linear and a disjoint non-linear region. These networks behave like a min-plus linear system at low load. Increasing the traffic rate beyond a threshold causes the network to enter the non-linear region. Given such a network, the goal of bandwidth estimation is to determine the available bandwidth of the linear region.

Consider a FIFO system with capacity C, where all traffic in the network is transmitted as constant-bit rate traffic. Suppose the FIFO queue sees (cross) traffic at a rate of r_c, and probing traffic is sent according to $A(t) = rt$. As observed in [19] and supported by simulation results therein, the departure function of the probing traffic is

$$D(t) = \begin{cases} rt \ , & \text{if } r \leq C - r_c \ , \\ \frac{r}{r+r_c} Ct \ , & \text{if } r > C - r_c \ . \end{cases} \tag{10}$$

If the probing rate is above the threshold $C - r_c$, the capacity allocated to the probe and cross traffic is proportional to their respective rates. As a result, the probing traffic gets more bandwidth when its rate is increased.

In Fig. 6 we show the results of bandwidth estimation for a FIFO system obtained in an ns-2 simulation [33]. We assume a scheduler with link capacity of 50 Mbps. Both the ingress and egress links to the scheduler have a propagation delay of each 10 ms. The cross traffic consists of 10 FTP sources that send traffic over a 25 Mbps access link. Probing traffic is sent to the queue on a 100 Mbps link, permitting us to overload the queue. We show the results for a burst impulse, passive measurements, rate scanning, and rate chirp and compare them to the actual available bandwidth. Packet sizes are set to 800 Bytes. The burst function is approximated by transmitting back-to-back probes at the maximum rate of 100 Mbps. For passive measurements we use as traffic 2 seconds from a high-bandwidth variable bit rate video trace [34] that has an average rate of 17.1 Mbps and a peak rate of 154 Mbps. Rate scanning is performed at increments of 4 Mbps, and the maximum rate is determined using the criteria given in [5]. For the rate chirp method, we run the publicly available ns-2 simulation code of *pathchirp*, with a minimum rate of 1 Mbps, a maximum rate of 80 Mbps, and a spread factor of $\gamma = 1.2$.

For FIFO, shown in Fig. 6, the burst impulse transmits at the maximum rate of 100 Mbps, and obtains most of the available bandwidth (80% would be consistent with Eq. (10), however, the TCP background traffic in the simulations does not have a constant bit rate.). The passive measurement method underestimates and sometimes overestimates the available bandwidth. This is due to the variability of the bit rate. All other schemes provide good estimates of the available bandwidth. In this example and for the chosen parameters, rate scanning and rate chirps provided the same results, even though neither scheme was tuned for this example.

We now offer a min-plus system interpretation of bandwidth estimation for FIFO for constant rate traffic. Consider the function $S_{fifo}(t) = (C - r_c)t$. From the empirical departure characterization D of a FIFO system from Eq. (10), we can verify that the following is satisfied for all $t \geq 0$:

$$D(t) = (rt) * S_{fifo} \ , \text{if } r \leq C - r_c$$
$$D(t) \geq (rt) * S_{fifo} \ , \text{if } r > C - r_c \ .$$

Therefore, S_{fifo} is an exact service curve for $A(t) = rt$ when $r \leq C - r_c$, and S_{fifo} is a lower service curve when the arrivals exceed the threshold value. In fact, S_{fifo} is the largest lower service curve for a FIFO system, and a solution to the maximization in Section IV. Any function larger than S_{fifo} may not be a lower service curve for $r > C - r_c$. This also serves as a proof that a FIFO system is not min-plus linear for $r > C - r_c$. So, we can view a FIFO network as a system that is min-plus linear at rates $r \leq C - r_c$, and crosses into a non-linear region when the rate exceeds the threshold. The crossing of these regions coincides with the available bandwidth S_{fifo}. Since a probing rate above $C - r_c$ in a FIFO system is the turning point where a backlog is created, the heuristic in *pathload* and *pathchirp* to stop measurements when increasing delays are observed can also be justified in terms of crossing the non-linear region.

We emphasize that the above statements hold only for constant rate cross and probing traffic. For variable rate traffic, e.g., TCP cross traffic and a video trace in Fig. 6, the interpretation remains to be established. Revising the notion of min-plus linearity so that short-term fluctuations of traffic do not make the system non-linear, as long as the long-term rate does not exceed the available bandwidth, is an open problem and a topic of future research.

The above considerations motivate a criteria for finding the threshold rate and the service curve for any non-linear system with disjoint linear and non-linear regions. For constant rate probe traffic $A^x(t) = xt$ and measured departures D^x, let us use any of the methods from Section V to find a function $\tilde{S}$ satisfying

$$D^x(t) = (xt) * \tilde{S}(t) \ .$$

Since an exact service curve must hold for all rates of the probe traffic, $\tilde{S}$ is not an exact service curve, if there is a rate $y < x$ such that $D^y(t) \neq (yt) * \tilde{S}(t)$ for some $t \geq 0$. In this case, x is above the threshold that makes the system leave the linear region, and the probing rate needs to be reduced. This criteria can also be used by techniques such as rate scanning and rate chirps, which start at low transmission rates and increase the probing rate. The probing stops as soon as non-linearity of the system is detected.

VII. Conclusions

We expressed measurement-based estimation of available bandwidth as a problem in min-plus linear systems, where the available bandwidth is represented by a service curve. We investigated the limits of passive measurements and derived the best service curve estimate that is justified by a given measurement trace. We showed that probing methods such as rate scanning and rate chirps can be derived in terms of the min-plus algebra. We related difficulties with network probing to non-linearities of the underlying system. The concise derivation and motivation of relatively sophisticated probing methods point to further links between network measurement approaches and min-plus system theory, which await full exploration.

Acknowledgments

M. Fidler's work was done while he was with the University of Toronto. The research in this paper is supported in part by the National Science Foundation under grants CNS-0435061, two NSERC Discovery grants, and an Emmy Noether grant from the German Research Foundation. The authors thank A. Burchard for many insights and suggestions.

References

[1] R. L. Carter and M. E. Crovella, "Measuring bottleneck link speed in packet switched networks," *Performance Evaluation*, vol. 27 and 28, pp. 297–318, 1996.

[2] J. Strauss, D. Katabi, and F. Kaashoek, "A measurement study of available bandwidth estimation tools," in *Proc. ACM IMC*, 2003, pp. 39–44.

[3] B. Melander, M. Björkman, and P. Gunningberg, "A new end-to-end probing and analysis method for estimating bandwidth bottlenecks," in *Proc. IEEE GLOBECOM*, Nov. 2000, pp. 415–420.

[4] M. Jain and C. Dovrolis, "Pathload: A measurement tool for end-to-end available bandwidth," in *Proc. Passive and Active Measurement Workshop*, Mar. 2002.

[5] ——, "End-to-end available bandwidth: Measurement methodology, dynamics, and relation with TCP throughput," in *Proc. ACM SIGCOMM*, Oct. 2002, pp. 295–308.

[6] N. Hu and P. Steenkiste, "Evaluation and characterization of available bandwidth probing techniques," *IEEE J. Select. Areas Commun.*, vol. 21, no. 6, pp. 879–894, Aug. 2003.

[7] V. Ribeiro, R. Riedi, R. Baraniuk, J. Navratil, and L. Cottrell, "PathChirp: Efficient available bandwidth estimation for network paths," in *Proc. Passive and Active Measurement Workshop*, Apr. 2003.

[8] A. Akella, S. Seshan, and A. Shaikh, "An empirical evaluation of wide-area internet bottlenecks," in *Proc. ACM IMC*, 2003, pp. 101–114.

[9] M. Jain and C. Dovrolis, "End-to-end estimation of the available bandwidth variation range," in *Proc. ACM SIGMETRICS*, 2005, pp. 265–276.

[10] R. Kapoor, L.-J. Chen, L. Lao, M. Gerla, and M. Y. Sanadidi, "CapProbe a simple and accurate capacity estimation technique," in *Proc. ACM SIGCOMM*, 2004, pp. 67–78.

[11] V. Paxson, "Measurements and analysis of end-to-end internet dynamics," Ph.D. dissertation, Univ. of California, Berkeley, Apr. 1997.

[12] C. Dovrolis, P. Ramanathan, and D. Moore, "What do packet dispersion techniques measure?" in *Proc. IEEE INFOCOM*, Apr. 2001, pp. 905–914.

[13] S.-R. Kang, X. Liu, M. Dai, and D. Loguinov, "Packet-pair bandwidth estimation: Stochastic analysis of a single congested node," in *Proc. IEEE ICNP*, Oct. 2004.

[14] X. Liu, K. Ravindran, and D. Loguinov, "Multi-hop probing asymptotics in available bandwidth estimation: Stochastic analysis," in *Proc. ACM IMC*, Oct. 2005.

[15] C. Cetinkaya, V. Kanodia, and E. W. Knightly, "Scalable services via egress admission control," *IEEE Trans. Multimedia*, vol. 3, no. 1, pp. 69–81, Mar. 2001.

[16] Y. Jiang, P. Emstad, A. Nevin, V. Nicola, and M. Fidler, "Measurement-based admission control for a flow-aware network," in *Proc. of 1st EuroNGI Conference on Next Generation Internet Networks Traffic Engineering*, Apr. 2005, pp. 318– 325.

[17] V. Jacobson, "Congestion avoidance and control," in *Proc. ACM SIGCOMM*, Stanford, Aug. 1988, pp. 314–329.

[18] S. Keshav, "A control-theoretic approach to flow control," in *Proc. ACM SIGCOMM*, Zurich, 1991, pp. 3–15.

[19] B. Melander, M. Björkmann, and P. Gunningberg, "First-come-first-served packet dispersion and implications for TCP," in *Proc. IEEE GLOBECOM*, vol. 3, Nov. 2002, pp. 2170–2174.

[20] X. Liu, K. Ravindran, and D. Loguinov, "What signals do packet-pair dispersions carry?" in *Proc. IEEE INFOCOM*, Mar. 2005.

[21] F. Agharebparast and V. C. M. Leung, "Slope domain modeling and analysis of data communication networks: A network calculus complement," in *Proc. IEEE ICC*, June 2006, pp. 591–596.

[22] S. Valaee and B. Li, "Distributed call admission control for ad hoc networks," in *Proc. IEEE VTC*, Sept. 2002.

[23] D. Wu and R. Negi, "Effective capacity: A wireless link model for support of quality of service," *IEEE Transactions on Wireless Communications*, vol. 2, no. 4, pp. 630–643, July 2003.

[24] R. Agrawal, R. L. Cruz, C. Okino, and R. Rajan, "Performance bounds for flow control protocols," *IEEE/ACM Trans. Networking*, vol. 7, no. 3, pp. 310–323, June 1999.

[25] C.-S. Chang, *Performance Guarantees in Communication Networks*. London, Great Britain: Springer-Verlag, 2000.

[26] J.-Y. Le Boudec and P. Thiran, *Network Calculus A Theory of Deterministic Queuing Systems for the Internet*. Berlin, Germany: Springer-Verlag, 2001.

[27] F. Baccelli, G. Cohen, G. J. Olsder, and J.-P. Quadrat, *Synchronization and Linearity: An Algebra for Discrete Event Systems*. West Sussex, Great Britain: John Wiley & Sons Ltd., 1992.

[28] R. T. Rockafellar, *Convex Analysis*. Princeton University Press, 1972.

[29] M. Fidler and S. Recker, "Conjugate network calculus: A dual approach applying the legendre transform," *Computer Networks*, vol. 50, no. 8, pp. 1026–1039, June 2006.

[30] T. Hisakado, K. Okumura, V. Vukadinovic, and L. Trajkovic, "Characterization of a simple communication network using legendre transform," in *Proc. Internationl Symposium on Circuits and Systems (ISCAS)*, May 2003, pp. 738–741.

[31] J. Naudts, "Towards real-time measurement of traffic control parameters," *Computer Networks*, vol. 34, no. 1, pp. 157–167, July 2000.

[32] R. Cruz, "A calculus for network delay, parts I and II," *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 114–141, Jan. 1991.

[33] "ns-2 network simulator," http://www.isi.edu/nsnam/ns/.

[34] G. V. der Auwera, P. T. David, and M. Reisslein, "Bit rate-variability of h.264/avc frext," Arizona State University, Tech. Rep., Apr. 2006.