INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2013, VOL. 59, NO. 2, PP. 141–149
Manuscript received January 12, 2013; revised April, 2013. DOI: 10.2478/eletel-2013-0017

# A Contribution to the System-Theoretic Approach to Bandwidth Estimation

Andrzej Borys, Katarzyna Wasielewska, and Dariusz Rybarczyk

***Abstract*—The network calculus provides a theoretical background for description of traffic in computer networks. Using this tool in explanation of the so-called pathchirp method of measuring the available bandwidth, the validity and range of application of some relationships exploited are verified in this paper. The derivations are carried out in a wider context than that considered in a recent paper by Liebeherr et al. published in IEEE/ACM Transactions on Networking on network bandwidth estimation, providing thereby new insights and outcomes. These results, summarized in a table, show a means of bounding the service curve, depending upon its convexity or non-convexity property assumed and upon the linearity or non-linearity of a network considered. Moreover, it is shown here that the nonlinear network example analyzed by Liebeherr et al. can be viewed equivalently as a linear parametric network. For this network, the behaviour of the cross traffic is considered in a more detail, too.**

***Keywords*—network calculus, rate chirps, bandwidth estimation, nonlinear and/or linear parametric traffic systems**

## I. Introduction

AVAILABLE bandwidth estimation is needed for several important tasks in computer networks. A route selection in some routing processes, traffic management and embedded QoS mechanisms – these are just some of many. Available bandwidth means unused bandwidth along an end-to-end connection. In the recent years, we observe a particular growth in diversity of overlay networks. It requires continuous research work on quality of service providing. There are many methods of available bandwidth estimation that are passive or use active probing, as for example IGI [1], Pathchirp [2], Pathload [3], TOPP [4]. Note that we require of these probing tools that they are accurate, fast and do not cause a noticeable load in the network.

All modern methods of bandwidth estimation in networks published in the literature can be nicely described using the network calculus technique, as it has been shown by Liebeherr et al. in [5]. The network calculus is an approach to network traffic analysis and optimization that uses the so-called min-plus algebra [6]. One of the fundamental notions in it is the service curve of a traffic system (node, path). This curve, as Liebeherr et al. [5] show, expresses the system's available bandwidth. If it is a straight line, this corresponds to a constant bit rate (constant bandwidth). However, it can be also a more complex function; in this case, it describes a variable rate (bandwidth). So, the problem of finding the available bandwidth can be formulated equivalently as a problem of finding a service curve. And in such a way, it is dealt with in [5], and also here.

It is assumed that the service curve of a given traffic system (node, path) is unknown. So our task is to evaluate it by applying a single probing packet sequence or multiple sequences to the system's input, and afterwards carrying out some measurements. Here, we consider only one of the possible methods that use a single packet train, but with the geometrically decreasing lengths of the packets consisting of the same number of bits. Note that this corresponds to increasing the transmission rate, giving the name to that sequence: rate chirp. Furthermore, we stress on the fact that the rate chirp is introduced in measurements intentionally, that is these measurements have an active character.

In this paper, we extend the results regarding the rate chirp method, presented by Liebeherr et al. [5], by consideration of the possible cases of service curve estimation: when the service curve is convex or non-convex, and when the traffic system is linear or nonlinear. The detailed derivations for all the four possibilities are presented, and the final results are summarized in a table.

The second problem discussed in [5], that is of nonlinear behaviour of a traffic system, is dealt with here, too. We show in this paper that the nonlinear network example analyzed in [5] can be viewed equivalently as a linear parametric system. It is a valuable observation. And, on this occasion, the behaviour of the cross traffic in the aforementioned network example is considered in a more detail, too.

This paper is organized as follows. Section II presents the main network calculus operations needed in understanding the material of the next sections. In Section III, the problem of network bandwidth estimation is formulated in terms of the min-plus algebra. Further, in the next section, the Legendre transform is shortly described. It is used in derivations of this paper for description of the min-plus systems in the so-called rate domain. The Section V contains the main results regarding solutions of the estimation problem in the four cases mentioned before. In the next section, the original results referring to an equivalent description of a nonlinear traffic system as a parametric linear one are presented. Additionally, the results regarding the behaviour of the cross traffic in such a system are given. The last section contains summary and conclusions.

A. Borys is with the Institute of Telecommunications, Faculty of Telecommunications and Electrical Engineering, University of Technology and Life Sciences, Bydgoszcz, Poland (e-mail: andrzej.borys@utp.edu.pl).

K. Wasielewska is with the Institute of Applied Informatics, The State School of Higher Professional Education in Elblag, Poland (e-mail: k.wasielewska@pwsz.elblag.pl).

D. Rybarczyk is with the OPEN SYSTEM Dariusz Rybarczyk, Bydgoszcz, Poland (e-mail: dariusz.rybarczyk@hotmail.com).

## II. Network Calculus Technique

Network calculus is a mathematical framework for analyzing performance guarantees in computer networks. This technique is based on min-plus algebra in which operation of addition is replaced by calculation of minimum, but operation of multiplication is replaced by the addition. In this approach, we consider a network as a min-plus linear or nonlinear system that converts input signals into output signals. If we want to provide guarantees to data flows, we have to limit the traffic that is sent by source nodes. It can be done by arrival curves.

Let $A(t)$ be an input traffic function, and $D(t)$ be the corresponding output traffic function. These functions are also called the arrival and departure traffics, respectively; they are illustrated in Fig. 1. $A(t)$ and $D(t)$ represent the cumulative number of bits observed in the interval $[0, t]$, with 0 meaning the beginning time instant and $t$ the ending time instant, respectively, of the traffic probe considered. Furthermore, denote by $S(t)$ a service curve relating the traffics $A(t)$ and $D(t)$.

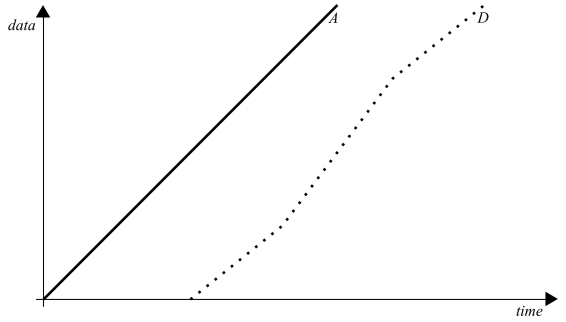


Fig. 1. Example of arrival and departure traffic functions.

In the linear traffic systems, we have the following relation between $A(t)$ and $D(t)$ [2,3,10]

$$D = A \otimes S, \tag{1}$$

where $\otimes$ is a convolution operator defined by

$$(A \otimes S)(t) = \inf_{0 \le \tau \le t} \{A(\tau) + S(t - \tau)\}. \tag{2}$$

For the non-linear system, instead of (1), we write

$$D \ge A \otimes S. \tag{3}$$

Note that writing $D \ge A \otimes S$ as in (3) means that there are certainly such times $t$ for which $D > A \otimes S$ holds. But, in this case, the times for which $D = A \otimes S$ may, eventually, also occur. Opposite to this, when we write $D = A \otimes S$ means that this equality holds for all times and for all pairs $A$, $D$.

## III. Bandwidth Estimation Problem in Min-Plus Algebra

We view a network as a min-plus linear or non-linear system that converts input signals (arrivals) into output signals (departures) according to a fixed but unknown service curve $S$ [5]. The service curve expresses the available bandwidth.

The estimation method considered here is a measurements oriented one. That is, first, a scenario for carrying out the needed measurements is thought out in such a scheme. Then, the measuring concept is implemented, and the data obtained in measurements are recorded. Finally, using this data, the estimation of the service curve is performed.

Note that two points in this general scheme are crucial: the choice of the form of the probing input traffic sequence and the choice of the means of estimation. We will discuss the choices made for the pathchirp method in Section V.

In the literature [5], one defines and uses the lower and upper bounds for the cumulative traffic. That is, considering the output traffic function $D(t)$ and the service curve $S(t)$, we can write $D(t) \ge (A \otimes \underline{S})(t)$ and $D(t) \le (A \otimes \overline{S})(t)$, where $\underline{S}$ stands for a lower, but $\overline{S}$ for an upper service curve, respectively. More precisely, among those nondecreasing functions that lowerbound $D(t)$, we choose for $\underline{S}(t)$ the one that maximizes the bound $(A \otimes \underline{S})(t)$. Similarly, for $\overline{S}$, we choose this nondecreasing function that minimizes the upper bound $(A \otimes \overline{S})(t)$. So, in this context, we understand the service curve $S(t)$ as $\underline{S}$ in the case of ineq. (3). Further, in the case of eq. (1), we have $S(t) = \underline{S} = \overline{S}$. (Such the service curve is called an exact service curve [5].)

Now take into account the difference between the left- and right-hand sides in (1) or (3) and denote it by $U(t)$. Using this, we can write

$$U(t) = D(t) - (A \otimes S)(t) = 0 \tag{4}$$

in the case of (1), and

$$U(t) = D(t) - (A \otimes S)(t) \ge 0 \tag{5}$$

in the case of (3), respectively.

Looking at (4) and (5), we see that it is possible to treat both the cases in the same way, formulating the approximation (estimation) problem for $S(t)$ as: *given $A(t)$ and $D(t)$, search for a possibly maximal approximate (estimate) $S_m(t)$ of $S(t)$ such that $U(t)$ containing $S_m$ instead of unknown $S$ is always greater or equal to zero, and it is minimized towards zero.*

## IV. Legendre Transform in Min-Plus Systems

In min-plus systems, the Legendre transform of a function $f(t)$ for $t \ge 0$ is defined as

$$\mathcal{L}_f(r) = \sup_{\tau \ge 0} \{r\tau - f(\tau)\}$$

where $r$ is interpreted as a rate and $r \ge 0$. This transform converts a convolution to an addition. That is we get

$$\mathcal{L}_{f*g} = \mathcal{L}_f + \mathcal{L}_g. \tag{6}$$

In general, the following two relations

$$\mathcal{L}(\mathcal{L}_f) \le f \tag{7}$$

and

$$\mathcal{L}(\mathcal{L}_f) = \text{conv}_f \tag{8}$$

hold, where $\text{conv}_f$ means the convex hull of $f$ that is defined as the largest convex function smaller than $f$. And the convex function $f$ can be obtained by performing the Legendre transform twice. That is

$$\mathcal{L}(\mathcal{L}_f) = f. \tag{9}$$

Moreover, the Legendre transform reverses an inequality order in the sense explained below

$$f \geq g \Rightarrow \mathcal{L}_f \leq \mathcal{L}_g. \tag{10}$$

If $g$ in (10) is convex, the relation therein becomes an equivalency.

## V. Rate Chirps Method in Min-Plus System

The means of building and shape of the probing traffic in any active method of network bandwidth estimation is important because it has influence on the method accuracy and time of estimation. The pathchirp method (rate chirps technique) [2] uses only a single sequence of packets as the probing traffic. This makes it faster than the methods exploiting multiple sequences of packets transmitted with different rates (changing at some range), as for example the rate scanning method [3]. The variable rate is realized in the pathchirp method by construction of the packet sequence as a packet train with a geometrically decreasing lengths of the packets consisting of the same number of bits. It is called then a chirp packet train; such a one is illustrated in Fig. 2. For more details, see for example [2].

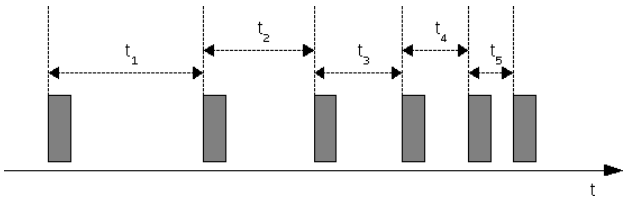


Fig. 2. Illustration of the chirp packet train with decreasing time intervals $t_i$ between the first bits of the consecutive packets.

It has been shown by Liebeherr et al. in [5] that the estimation procedure of service curve estimation applied in the pathchirp method can be explained in terms of the network calculus, shortly saying, as performing twice the Legendre transform. In [5], the validity of this result has been restricted to only one case in which it is assumed that the estimated service curve is convex and the network analyzed is linear (in the sense of the min-plus algebra). Here, we consider the next three possible cases and show that the results then achieved may be also reasonable. In particular cases, their accuracy may not even be worse than that we get in the case considered in [5]. A simple criterion of the reasonableness of an estimate achieved is given.

Let us now list all the four possible cases:

1) system is min-plus linear and $S$ is convex,
2) system is min-plus linear and $S$ is non-convex,
3) system is min-plus non-linear and $S$ is convex,
4) system is min-plus non-linear and $S$ is non-convex.

In what follows, we show how the estimation procedure works in each of the aforementioned cases. As the final result of derivations, in each case, we get an information on means of bounding the actual service curve by its estimate.

Case 1. This case has been described in [5]. It is reasonable to remind it here for the sake of completeness. If a system is linear, then (1) holds. Applying the Legendre transform to (1), we obtain

$$\mathcal{L}_D = \mathcal{L}_{A*S} = \mathcal{L}_A + \mathcal{L}_S. \tag{11}$$

Further, rearranging the terms in (11), we get

$$\mathcal{L}_D - \mathcal{L}_A = \mathcal{L}_S. \tag{12}$$

In the next step, applying the Legendre transform to both sides of (12) gives

$$\mathcal{L}(\mathcal{L}_S) = \mathcal{L}(\mathcal{L}_D - \mathcal{L}_A). \tag{13}$$

And finally, because of the assumed convexity of $S$ in this case, we can use (7) on the right-hand side of (13). This results in

$$S = \mathcal{L}(\mathcal{L}_D - \mathcal{L}_A). \tag{14}$$

In what follows, interpretation and usage of (14) is more general than that given above. The right-hand side of this equation is treated as an approximating (estimating) formula for calculation of the approximate (estimate) of a service curve $S$, indepedently of the fact whether the aforementioned assumptions regarding linearity and convexity are fullfilled or not, and in the cases of modifications undertaken, too.

Let us now consider an ideal chirp packet train understood as such a one that allows transmission of its consecutive packets with the rates going to infinity. Denote the arrivals and departures associated with this chirp packet train by $A^{chrp}$ and $D^{chrp}$, respectively, and by $\tilde{S}$ the service curve estimate calculated for it using (14). That is

$$\tilde{S} = \mathcal{L}(\mathcal{L}_{D^{chrp}} - \mathcal{L}_{A^{chrp}}). \tag{15}$$

Since in this case $S$ is convex, we have obviously $\tilde{S} = S$.

In practice, realization of the transmission rates going to infinity, for the chirp train packets, is not possible. This is because the chirp packets cannot be transmitted faster than the maximal rate of data sending by the packet sender. For this reason, some modifications of the probing scheme based on rate chirps have been made in [5]. For practical purposes, the following extrapolations of the chirp arrival and departure functions have been undertaken [5]

$$\tilde{A}^{chrp}(t) = \begin{cases} A^{chrp} & \text{if } 0 \leq t \leq t^A_{max} \\ \infty & \text{if } t > t^A_{max} \end{cases} \tag{16}$$

$$\tilde{D}^{chrp}(t) = \tag{17}$$

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

where it is assumed that a packet chirp is transmitted in a time interval $[0, t^A_{max}]$ and $D$ is observed over an interval $[0, t^D_{max}]$. Note that after setting the value of the arrival function to $\infty$ past the last measurement at time $t^A_{max}$, we meet the requirement that $\mathcal{L}_A(r) < \infty$. An example of the convex function $\tilde{A}^{chrp}$ is sketched in Fig. 3.

Let us now denote by $S^*$ an estimate of $S$ obtained by substituting the functions $\tilde{A}^{chrp}(t)$ and $\tilde{D}^{chrp}(t)$ given by (16) and (17) into estimating formula (14). We will show that the following

$$S^* \leq S = \tilde{S}$$

holds. That is $S^*$ is a lower bound of $S$.

Further, denote the departure function that corresponds to the arrival function $\tilde{A}^{chrp}$ by $D^*$. Obviously, this function is not identical with $\tilde{D}^{chrp}$.

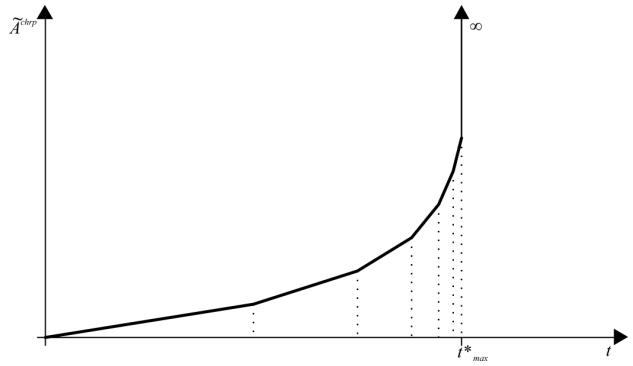


Fig. 3. Example of convex function $\tilde{A}^{chrp}$.

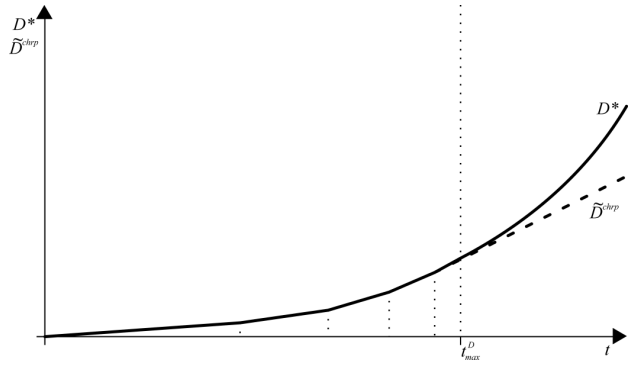


Fig. 4. Illustration of the relation between the functions $D^*$ and $\tilde{D}^{chrp}$. In the range from from 0 to $t^D_{max}$, they overlap each other.

The pairs ($\tilde{A}^{chrp}$,$D^*$) fulfill eq. (14). That is

$$S = \mathcal{L}(\mathcal{L}_{D^*} - \mathcal{L}_{\tilde{A}^{chrp}})$$

holds.

In [5], it has been shown that $\tilde{D}^{chrp} \leq D^*$ in case 1. This is illustrated in Fig. 4, and means that $\tilde{D}^{chrp}$ is a lower bound for departures that correspond to arrivals $\tilde{A}^{chrp}$. From the order-reversing property of Legendre transform (10) applied to the last inequality, we get

$$\mathcal{L}_{\tilde{D}^{chrp}} \geq \mathcal{L}_{D^*}$$

and further

$$\mathcal{L}_{\tilde{D}^{chrp}} - \mathcal{L}_{\tilde{A}^{chrp}} \geq \mathcal{L}_{D^*} - \mathcal{L}_{\tilde{A}^{chrp}}. \tag{18}$$

Finally, applying (10) once again, now to (18), we obtain

$$S^* = \mathcal{L}(\mathcal{L}_{\tilde{D}^{chrp}} - \mathcal{L}_{\tilde{A}^{chrp}}) \leq S. \tag{19}$$

Inequality (19) states that the estimate $S^*$ of the service curve $S$ is its lower bound in case 1.

Case 2. In this and the remaining two cases, we repeat the successive steps regarding the pairs ($A^{chrp}$,$D^{chrp}$), ($\tilde{A}^{chrp}$,$D^*$), and ($\tilde{A}^{chrp}$,$\tilde{D}^{chrp}$) as it was done in case 1. So now, because the network considered is linear similarly as in case 1, we have

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

However, in this case, opposite to the previous one $S$ is assumed to be non-convex. Therefore, it follows from (7) that

$$S \geq \mathcal{L}(\mathcal{L}_S) = \mathcal{L}(\mathcal{L}_{D^{chrp}} - \mathcal{L}_{A^{chrp}}) = \tilde{S} \tag{20}$$

where $\tilde{S}$ is the convex hull of $S$, according to (8). So recapitulating (20), we write

$$S \geq \tilde{S}. \tag{21}$$

(21) means that the equality between $S$ and $\tilde{S}$ does not hold now. It is clear from the definition of $\tilde{A}^{chrp}$ that

$$\tilde{A}^{chrp} \geq A^{chrp}.$$

Hence, we receive from (10)

$$\mathcal{L}_{\tilde{A}^{chrp}} \leq \mathcal{L}_{A^{chrp}}.$$

Note now that because $S$ is non-convex each of the following inequalities

$$\tilde{D}^{chrp} \leq D^* \text{ or } \tilde{D}^{chrp} \geq D^*$$

can hold. However, considering the above inequalities in this case and case 4, we restrict ourselves to only such the situations that have practical meaning. That is when we write $\tilde{D}^{chrp} \leq D^*$, it means that this inequality holds for all times $t \geq 0$. And similarly, writing the inverse inequality $\tilde{D}^{chrp} \geq D^*$ means its fulfillment for all times $t \geq 0$. In other words, when one of the above inequalities is satisfied, this excludes satisfaction of the second one (except, of course, of such times for which $\tilde{D}^{chrp} = D^*$ holds). Referring to Fig. 4, two examples of such the functions $D^*$ as described above are illustrated in Fig. 5.

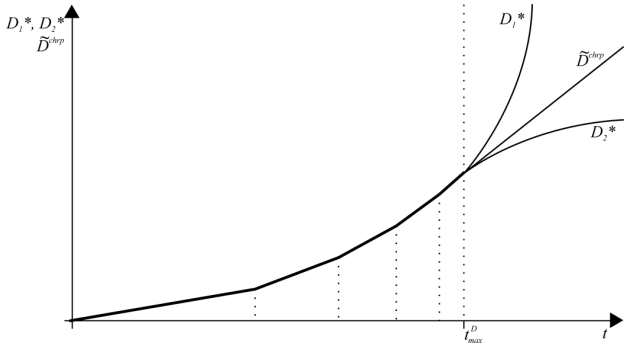


Fig. 5. Illustration of the class of functions $D^*$ considered: an example of $D_1^*$ being convex and another one being non-convex.

It means that

$$\mathcal{L}_{\tilde{D}^{chrp}} \geq \mathcal{L}_{D^*} \text{ or } \mathcal{L}_{\tilde{D}^{chrp}} \leq \mathcal{L}_{D^*}$$

are possible. Subtract now $\mathcal{L}_{\tilde{A}^{chrp}}$ on both sides of the above inequalities. We obtain

$$\mathcal{L}_{\tilde{D}^{chrp}} - \mathcal{L}_{\tilde{A}^{chrp}} \geq \mathcal{L}_{D^*} - \mathcal{L}_{\tilde{A}^{chrp}} \tag{22}$$

or the inverse inequality. Applying then (10) to (22), we get

$$\mathcal{L}(\mathcal{L}_{\tilde{D}^{chrp}} - \mathcal{L}_{\tilde{A}^{chrp}}) \leq \mathcal{L}(\mathcal{L}_{D^*} - \mathcal{L}_{\tilde{A}^{chrp}}) \tag{23}$$

or the inverse inequality. In the next step, using the definition of $S^*$ given on the left-hand side of (19), and applying (13) with (7) in (23) or in the inverse inequality, we get

$$S^* \leq \text{conv}_S \leq S \tag{24}$$

or

$$S^* \geq \text{conv}_S \leq S \tag{25}$$

respectively. In summary of case 2, we can rewrite (21), (24), and (25) as

$$S \geq \tilde{S} \text{ and } S^* \leq \text{conv}_S \leq S,$$

or

$$S \geq \tilde{S} \text{ and } S^* \geq \text{conv}_S \leq S.$$

Looking at the above relations, we see that the ideal chirp approximate $\tilde{S}$ of $S$ is its lower bound. With regard to the (practical) calculated estimate $S^*$ of $S$, however, we are not certain whether we get a lower or an upper bound of $S$.

Case 3. If the system is non-linear, then $D \geq A \otimes S$. Therefore, instead of (13), we have now

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

This relation is true in particular for pairs $(A^{chrp}, D^{chrp})$.

In this case, $S$ is convex. So, it follows from the above, (8), and the definition of $\tilde{S}$

$$S = \mathcal{L}(\mathcal{L}_S) \leq \mathcal{L}(\mathcal{L}_{D^{chrp}} - \mathcal{L}_{A^{chrp}}) = \tilde{S}.$$

It means that $\tilde{S}$ is an upper bound of $S$.

Here, as in case 1, we have $D^{\tilde{chrp}} \leq D^*$. This fact and the assumed convexity of $S$ cause that the next steps to be carried out are the same as the corresponding ones in case 1. In consequence, we get $S^* \leq S$. So, in summary, we can write

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

Thereby, we have shown that the calculated estimate $S^*$ of $S$ is its lower bound in case 3. Moreover, the ideal chirp approximate $\tilde{S}$ is then an upper bound of $S$.

Case 4. In this case, $S$ is non-convex. So, we have

$$S \geq \mathcal{L}(\mathcal{L}_S) \leq \mathcal{L}(\mathcal{L}_{D^{chrp}} - \mathcal{L}_{A^{chrp}}) = \tilde{S}.$$

In other words, we have shown that

$$S \geq \text{conv}_S \text{ and } \tilde{S} \geq \text{conv}_S.$$

So, we can say nothing about the relation between $S$ and $\tilde{S}$. And we conclude that both the inequalities

$$S \geq \tilde{S} \text{ or } S \leq \tilde{S}$$

are possible in case 4.

Searching for a relation between $S^*$ and $S$ goes in this case along the corresponding lines of case 2 and with the use of relation $\mathcal{L}(\mathcal{L}_S) \leq \mathcal{L}(\mathcal{L}_D - \mathcal{L}_A)$ applied to the pairs $(\tilde{A}^{chrp}, D^*)$. We arrive then at the result that both the following inequalities

$$S^* \leq \mathcal{L}(\mathcal{L}_{D^*} - \mathcal{L}_{\tilde{A}^{chrp}}) \geq \text{conv}_S \leq S$$

or

$$S^* \geq \text{conv}_S \leq S$$

are possible. They show how complicated is this case. From the above, we conclude only that each of the two cases

$$S^* \leq S \text{ or } S^* \geq S$$

can happen in case 4.

TABLE I
SUMMARY OF STUDY CASES

| | Linear system $D = A \otimes S$ | Non-linear system $D \geq A \otimes S$ |
|---|---|---|
| Service curve S convex | $S^* \leq S$<br>lower bound calculated | $S^* \leq S$<br>lower bound calculated |
| Service curve S non-convex | $S^* \leq S$<br>or<br>$S \leq S^*$<br>not known whether lower or upper bound calculated | $S^* \leq S$<br>or<br>$S \leq S^*$<br>not known whether lower or upper bound calculated |

Summary of all the cases has been presented in Tab. I.

Let us now substitute $S^*$ instead of $S$ in (4) and (5) and denote the resulting $U(t)$ as $U^*(t)$, that is

$$U^*(t) = D(t) - (A \otimes S^*)(t).$$

Further, note that for the two cases of the first line of Tab. I, we have $S^* \leq S$. Performing the convolution operation on both sides of this inequality (using one of the rules given in [6]), we get $A \otimes S^* \leq A \otimes S$. Hence, when (4) or (5) holds, then certainly $U^*(t) \geq 0$ is satisfied for all times $t \geq 0$. This means that the estimate $S^*$ calculated in case 1 and 3 is found in accordance with the lines of the estimation procedure described at the end of Section III.

With regard to the second line of Tab. I, the situation is more complicated. However, intuitively, we can assume that when $U^*(t) \geq 0$ for each $t \geq 0$ holds, then the calculated estimate $S^*$ is reasonable in cases 2 and 4.

One remark more: It can happen in practice that we know nothing about a network (that is about its linearity and of its service curve convexity). In such a general case, the condition of $U^*(t) \geq 0$ for each time $t \geq 0$ can play a very useful role of the estimate reasonableness, indepedently of whether $S^*$ calculated constitutes a lower or an upper bound of $S$ (which is unknown for us).

## VI. NONLINEARITY

In [5], it has been shown how to model the FIFO system with capacity $C$, illustrated schematically in Fig. 6, in a wide range of traffic intensities. The model proposed by Liebeherr et al. [5] consists of two separate descriptions. First of them holds for small (mild) traffic loads (intensities) and is given by (1).

But, the second expressed by (3) is used, when the loads are large (leading to network overloading). In other words, the FIFO system works in a linear region of operation, when its traffic loading is small (mild). But, when this loading is high, it works in non-linear region of operation.

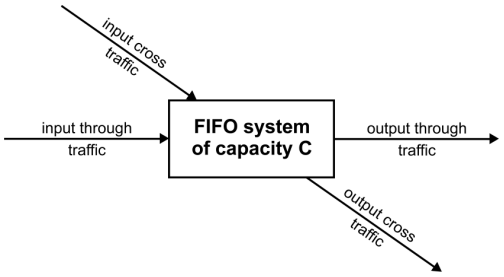


Fig. 6. A scheme of the FIFO system with capacity $C$ and depicted input and output through and cross traffics.

So, the FIFO system after the model presented in [5] works linearly or nonlinearly, depending upon the value of network loading (traffic intensity). In this section, we present an alternative model of the FIFO system which is linear in both the regions of values of the traffic intensities mentioned above. Characteristic for it is, however, the fact that it is non-parametric for mild loads. But, for larger loads, being still linear, it becomes parametric one. Then, it begins to depend upon the parameters of the cross traffic.

To verify experimentally the relation given in [5] for the output through traffic at the FIFO system working with the constant bit rate (CBR) through and cross traffics, we have built a measuring setup shown in Fig. 7.

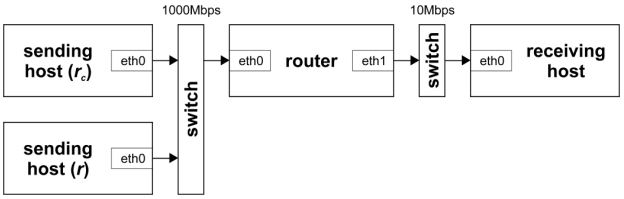


Fig. 7. The measuring setup used in measurements of the output through traffic for different rates of the input through traffic, in the presence of the constant rate $r_c$ of the input cross traffic.

The FIFO system of Fig. 7 had the capacity $C = 10$ Mbps. In our all measurements carried out, this system experienced the CBR cross traffic with the rate $r_c = 5$ Mbps sent in 625-byte packets. The input through traffic applied to the system was in accordance with the linear relation $A(t) = rt$, where the rates $r$ changed as follows: $1, 2, 3, 4, 5, 6, 7, 8, 9$, and $10$ Mbps. The probing traffic packets had the lengths ranging from $125$ to $1250$ bytes, depending on the probing rate $r$. For each of the above rate, five measurements were carried out and afterwards the median filtering was applied to eliminate possible "outliers".

The results of the measurements described above are summarized in Tab. II. Using this data, the experimental relationship between the input through traffic rates and the throughput for the output through traffic has been found and is illustrated in Fig. 8.

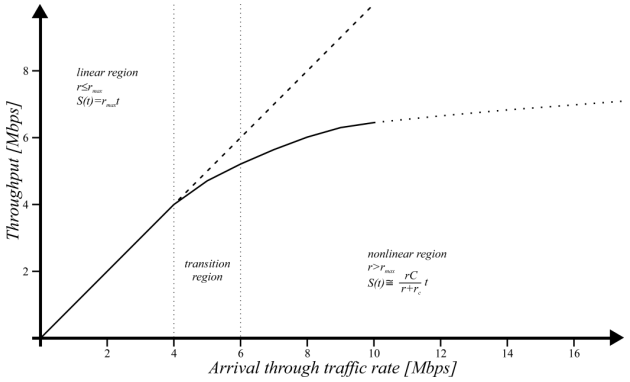


Fig. 8. Experimental relationship between throughput and arrival through traffic rate.

TABLE II
PROBING TRAFFIC ARRIVAL RATE AND THROUGHPUT

| Arrival rate [Mbps] | Throughput [Mbps] |
|---|---|
| 1,000 | 1,000 |
| 2,000 | 2,000 |
| 3,000 | 3,000 |
| 4,000 | 3,999 |
| 5,000 | 4,709 |
| 6,000 | 5,211 |
| 7,000 | 5,642 |
| 8,000 | 6,016 |
| 9,000 | 6,301 |
| 10,00 | 6,451 |

In [7], the output through traffic for the scenario described above and exploited in the measurements resulting in the data of Tab. II and curve of Fig. 8 is modeled by the following relations

$$D(t) = \begin{cases} rt & \text{if } r \le C - r_c \\ \frac{r}{r+r_c} Ct & \text{if } r > C - r_c. \end{cases} \tag{26}$$

The relations (26) are interpreted in [5] in the following way: as the probing traffic rate does not exceed the threshold rate $r_{max} = C - r_c$, the output traffic depends upon it linearly. That is when $r \le r_{max}$ holds. Outside this range, that is for $r > r_{max}$, it is assumed to behave nonlinearly.

In what follows, we show that the FIFO system considered behaves linearly in both the regions of values of $r$ mentioned above. To this end, consider first the range of $r \le r_{max}$ and apply (1) and (2) with $A(t) = rt$ and a presumed form of the service curve $S(t) = r_{max} t$. In consequence, we get

$$D(t) = \inf_{0 \le \tau \le t} \{r\tau + r_{max}(t - \tau)\} =$$

$$= \inf_{0 \le \tau \le t} \{r\tau + r_{max} t - r_{max}\tau)\} =$$

$$= \inf_{0 \le \tau \le t} \{r_{max} t + r\tau - r_{max}\tau)\} =$$

$$= \inf_{0 \le \tau \le t} \{r_{max} t + \tau(r - r_{max})\} =$$

$$= r_{max} t + t(r - r_{max}) = rt$$

because $r - r_{max} \leq 0$. Hence, we conclude that the presumed $S(t) = r_{max}t$ describes correctly the considered region of operation.

Let us now take into account the range of $r > r_{max}$, applying (1) and (2) as before with $A(t) = rt$, but a new presumed form of the service curve

$$S(t) = \frac{r}{r + r_c} Ct. \tag{27}$$

As a result, we obtain in this case

$$\begin{aligned} D(t) &= \inf_{0 \leq \tau \leq t} \{A(\tau) + S(t - \tau)\} = \\ &= \inf_{0 \leq \tau \leq t} \left\{ r\tau + \frac{r}{r + r_c} C(t - \tau) \right\} = \\ &= \inf_{0 \leq \tau \leq t} \left\{ r\tau + \frac{rCt}{r + r_c} - \frac{rC\tau}{r + r_c} \right\} = \\ &= \inf_{0 \leq \tau \leq t} \left\{ \frac{rCt}{r + r_c} + \tau(r - \frac{rC}{r + r_c}) \right\} = \\ &= \inf_{0 \leq \tau \leq t} \left\{ \frac{rCt}{r + r_c} + \tau \frac{r}{r + r_c}(r + r_c - C) \right\} = \\ &= \inf_{0 \leq \tau \leq t} \left\{ \frac{rCt}{r + r_c} + \tau \frac{r}{r + r_c}(r - r_{max}) \right\} = \\ &= \frac{rCt}{r + r_c} \end{aligned}$$

because now $r > r_{max}$. So the chosen $S(t)$ given by (27) is correct for the above range of values of $r > r_{max}$.

In summary, we conclude that the FIFO system treated as a linear network possesses the following service curve

$$S(t) = \begin{cases} r_{max}t & \text{if } r \leq r_{max} \\ \frac{r}{r + r_c} Ct & \text{if } r > r_{max}. \end{cases} \tag{28}$$

Observe from (28) that the slope of the linear function $S(t)$ of a variable $t$ for the range $r > r_{max}$,

$$\frac{rC}{r + r_c},$$

depends upon the parameter $r$ characterizing the input through traffic. So, the FIFO system behaves then as a linear parametric system (its characteristic is "modulated" by the parameters of the input traffic).

The conclusions drawn above are illustrated in Fig. 9, where the functions $A(t)$, $D(t)$, and $S(t)$ are sketched for three cases: mildly-loaded (linear), transition point ($r = r_{max}$), and over-loaded (non-linear) ones, for comparison.

Further, observe from (26) that the modeled relation between the input through traffic rate $r$ and its throughput is as follows

$$\text{throughput} = \begin{cases} r & \text{if } r \leq r_{max} \\ \frac{rC}{r + r_c} & \text{if } r > r_{max}. \end{cases} \tag{29}$$

The function given by (29) is sketched in Fig. 10.

Observe that the theoretical (modeled) curve of Fig. 10 differs slightly from the experimental one shown in Fig. 8. The largest differences between the curves occur in the neighbourhood of $r = r_{max}$, where we have to do with the transition

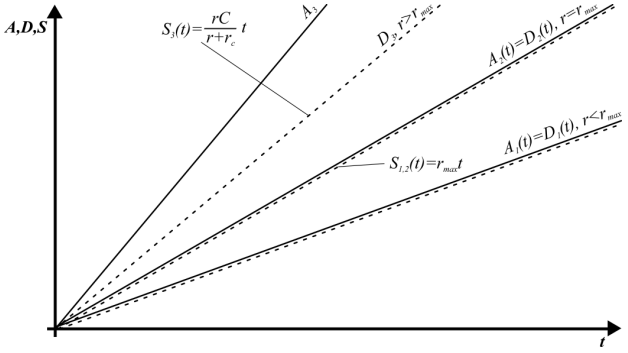


Fig. 9. Comparison of the functions $A(t)$, $D(t)$, and $S(t)$ for the FIFO system for three cases: mildly-loaded (linear), transition point ($r = r_{max}$), and over-loaded (non-linear) ones.

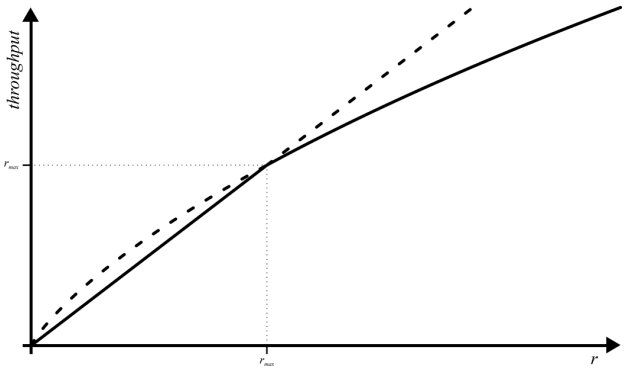


Fig. 10. Sketch of the modeled function of throughput versus $r$.

between the linear and non-linear regions of operation. In the model discussed, this transition region is very sharp, in fact reducted to only one point $r = r_{max}$. Opposite to this, as Fig. 5 shows, the real transition region means a mild passing.

Coming back to (28), note that this relation can be expressed equivalently as

$$S_p(t) = [r_{max} \vee \frac{rC}{r + r_c}]t \tag{30}$$

where the symbol $\vee$ means a larger of two given values. Evidently, $S_p(t)$ given by (30) depends on $r$, being the parameter of the input through traffic. This is in opposition to the form of the service curve

$$S_{\text{fifo}}(t) = [C - r_c]^+ t = [r_{max}]^+ t \tag{31}$$

where $[x]^+ = \max(0, x)$, provided in [5]. The non-parametric service curve given by (31), $S_{\text{fifo}}(t)$, when applied in (1) with (2) or in (1) with (3), depending upon the value of the rate $r$, separates from each other the linear and non-linear regions of operation according to the model of [5]. This is shown in detail in Appendix.

Let us now formally check whether the parametric model of the FIFO system given by (30) is really linear. For this purpose, we use the linearity principle as defined in [5], starting with formulation of the input through traffic in the following form

$$(a + A_1(t)) \wedge (b + A_2(t))$$

where $A_1(t) = r_1 t$ and $A_2(t) = r_2 t$, and $\wedge$ means the infimum operation. Then, we calculate the following

$$D_\wedge(t) = \inf_{0\le\tau\le t}\{(a + A_1(\tau)) \wedge (b + A_2(\tau)) + S_p(t-\tau)\} =$$
$$= \inf_{0\le\tau\le t}\{\inf_2\{a + A_1(\tau) + S_p(t-\tau),$$
$$b + A_2(\tau) + S_p(t-\tau)\}\} =$$
$$= \inf_2\{\inf_{0\le\tau\le t}\{a + A_1(\tau) + S_p(t-\tau)\},$$
$$\inf_{0\le\tau\le t}\{b + A_2(\tau) + S_p(t-\tau)\}\} =$$
$$= \inf_2\{a + \inf_{0\le\tau\le t}\{A_1(\tau) + S_p(t-\tau)\},$$
$$b + \inf_{0\le\tau\le t}\{A_2(\tau) + S_p(t-\tau)\}\} =$$
$$= \inf_2\{a + D_1(t), b + D_2(t)\}$$

where

$$D_1(t) = r_1 t \text{ or } D_1(t) = \frac{r_1}{r_1 + r_c} Ct$$

and

$$D_2(t) = r_2 t \text{ or } D_2(t) = \frac{r_2}{r_2 + r_c} Ct$$

accordingly. In summary, we can rewrite the final result of the above derivation as

$$D_\wedge((a + A_1(t)) \wedge (b + A_2(t))) = (a + D_1(t)) \wedge (b + D_2(t)).$$

This means that the linearity principle is fulfilled in the case of the parametric model.

Finally, it would be also interesting to consider, even shortly, relations regarding servicing the cross traffic. Note that such the relations have not been discussed in [5] at all. To this end, let us start with finding a service curve for the cross traffic in the linear region of operation. Having the previous derivations for the through traffic in mind, it seems that the function

$$S_{cl}(t) = r_{maxc} t$$

where $r_{maxc} = C - r$, is a proper service curve for the cross traffic path in the linear region of operation of the FIFO system. We check this in what follows. For $r \le C - r_c$, we have

$$(r_c t) \otimes S_{cl}(t) = \inf_{0\le\tau\le t}\{r_c\tau + r_{maxc}(t-\tau)\} =$$
$$= \inf_{0\le\tau\le t}\{(r_c - r_{maxc})\tau + r_{maxc} t\} =$$
$$= (r_c - r_{maxc})t + r_{maxc} t = r_c t.$$

Therefore

$$D_c(t) = (r_c t) \otimes S_c l(t)$$

really holds.

For the non-linear region of operation, we expect that the function

$$S_{cn}(t) = \frac{Cr_c}{r + r_c} t$$

is proper service curve in our linear parametric model developed. And as before, we check this conjucture by writing

$$(r_c t) \otimes S_{cn}(t) = \inf_{0\le\tau\le t}\left\{r_c\tau + \frac{Cr_c}{r + r_c}(t-\tau)\right\} =$$
$$= \inf_{0\le\tau\le t}\left\{r_c\tau - \frac{Cr_c}{r + r_c}\tau + \frac{Cr_c}{r + r_c} t\right\} =$$
$$= \inf_{0\le\tau\le t}\left\{r_c\tau\frac{r + r_c - C}{r + r_c} + \frac{Cr_c}{r + r_c} t\right\}.$$

Observe now that carrying out the operation of infimum in the above relation, when $r > C - r_c$, gives

$$(r_c t) \otimes S_{cn}(t) = \frac{r_c C}{r + r_c} t.$$

Further, $D_c(t) = \frac{r_c C}{r + r_c} t$ in the non-linear region of operation (because $D_c(t) = Ct - \frac{rC}{r + r_c} t = \frac{r_c C}{r + r_c} t$). So, finally, we can write

$$(r_c t) \otimes S_{cn}(t) = D_c(t)$$

and this completes the proof of the conjecture.

## VII. Conclusion

In this paper, we have extended the results on network available bandwidth estimation presented in [5], which regards the so-called pathchirp approach, as well as the results related to bandwidth estimation in the non-linear range of operation of the FIFO system. First, we have analyzed in a wider context the fundamental equations derived for the aforementioned method, showing a wider range of its applicability than that indicated in [5]. Second, we have shown that the FIFO system working in the non-linear region of operation can be viewed alternatively as a linear parametric network. And finally, we have developed a linear parametric model for the description of the cross traffic.

## APPENDIX

In [5], the following relations

$$D(t) = (rt) \otimes S_{\text{fifo}} \text{ if } r \le C - r_c \tag{A.1}$$
$$D(t) \ge (rt) \otimes S_{\text{fifo}} \text{ if } r > C - r_c \tag{A.2}$$

valid for all $t \ge 0$ have been used for modelling the output through traffic with the use of the service curve $S_{\text{fifo}}(t)$. In this Appendix, we present the derivations showing that the relations (A.1) and (A.2) correspond to the general relations (1) with (2) and (1) with (3), respectively. In fact, for $r \le C - r_c$, we have

$$(rt) \otimes S_{\text{fifo}}(t) = \inf_{0\le\tau\le t}\{r\tau + (C - r_c)(t-\tau)\} =$$
$$= \inf_{0\le\tau\le t}\{r\tau + (C - r_c)t - (C - r_c)\tau)\} =$$
$$= \inf_{0\le\tau\le t}\{\tau(r + r_c - C) + (C - r_c)t)\}.$$

And because $r + r_c - C \le 0$ holds in this case, the infimum of the expression in the parentheses of the above relation occurs for $\tau = t$. Therefore, we get

$$(rt) \otimes S_{\text{fifo}}(t) = t(r + r_c - C) + (C - r_c)t = rt = D(t).$$

Similarly, for $r > C - r_c$ we also have

$$(rt) \otimes S_{\text{fifo}}(t) = \inf_{0\le\tau\le t}\{\tau(r + r_c - C) + (C - r_c)t)\}.$$

However, because now $r+r_c-C>0$, the infimum mentioned above occurs in this case for $\tau=0$. So, we obtain

$$(rt)\otimes S_{\text{fifo}}(t)=(C-r_c)t=r_{max}t\leq\tfrac{rC}{r+r_c}t=D(t)$$

or shortly

$$D(t)\geq(rt)\otimes S_{\text{fifo}}(t).$$

## REFERENCES


[1] N. Hu and P. Steenkiste, "Evaluation and Characterization of Available Bandwidth Techniques," in *IEEE JSAC Special Issue in Internet and WWW Measurement, Mapping, and Modeling*, 2003.

[2] V. J. Ribeiro, R. H. Riedi, R. G. Baraniuk, J. Navratil, and L. Cottrell, "PathChirp: Effcient Available Bandwidth Estimation for Network Paths," in *Passive and Active Measurement Workshop*, 2003.

[3] M. Jain and C. Dovrolis, "Pathload: A Measurement Tool for End-to-End Available Bandwidth," in *Passive and Active Measurements*, Fort Collins, CO, March 2002.

[4] B. Melander, M. Bjorkman, and P. Gunningberg, "A New End-to-End Probing and Analysis Method for Estimating Bandwidth Bottlenecks," in *Global Internet Symposium*, 2000.

[5] J. Liebeherr, M. Fidler, and S. Valaee, "A system-theoretic approach to bandwidth estimation," *IEEE/ACM Transactions on Networking*, vol. 18, no. 4, pp. 1040–1053, August 2010.

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

[7] B. Melander, M. Bjorkman, and P. Gunningberg, "First-come-first-served packet dispersion and implications for TCP," *Proceedings of IEEE GLOBECOM*, vol. 3, pp. 2170–2174, November 2002.

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

[9] A. Borys and K. Wasielewska, "On characterization and properties of greedy shapers, and variable versus fixed delay," in *Proceedings of The First European Teletraffic Seminar*, Poznan, Poland, February 2011, pp. 236–242.

[10] A. Borys, M. Aleksiewicz, and D. Rybarczyk, "On understanding the service curve and its measurement-based estimation," in *Proceedings of The First European Teletraffic Seminar*, Poznan, Poland, February 2011, pp. 232–235.

[11] M. Fidler, "A Survey of Deterministic and Stochastic Service Curve Models in the Network Calculus," *IEEE Communications Surveys and Tutorials*, vol. 12, no. 1, pp. 59–86, 2010.