Andrzej Borys
Gdynia Maritime University
Poland

SOME RELATIONSHIPS BETWEEN EFFECTIVE AND LEFTOVER SERVICE CURVES

Network service curve plays a fundamental role in the techniques called network calculus. For example, it is very useful in analysis of admission control algorithms and scheduling procedures. Its basic form is assumed to be time-invariant and independent of cross traffic. However, in practice, when cross traffic must be taken into account, another form of this curve has to be used. It is called leftover service curve that depends upon the cross traffic intensity and indicates how much bandwidth is actually left over for through traffic. Outside the field of network calculus, a service curve was also devised that depends upon cross traffic, similarly as the leftover service curve. It was named $\in$-effective service curve and proved to be useful for ad hoc networks. In this paper, we discuss the relationships existing between the aforementioned service curves. In particular, we show that under some assumptions the $\in$-effective service curve can be viewed as the result of interpolation of the leftover service curve obtained in measurements. The discussions are provided with the relevant derivations.

Keywords: *Network calculus, relationships between effective and leftover service curves.*

INTRODUCTION

Network calculus [1, 2, 3] stands for a family of calculation tools that are an alternative to the methods using queueing theory for solving network performance problems. In this approach, the so-called network service curve [3, 4] plays a crucial role. This curve is used in finding solutions to a variety of sometimes difficult problems regarding, for example, network admission and scheduling algorithms. In the literature, two variants of the network service curve are considered, deterministic and stochastic ones. They are used in the deterministic and stochastic calculi, respectively.

In this paper, we exploit the first one. Its basic form is time-invariant and independent of cross traffic. However, because in many cases occurring in practice, the influence of cross traffic must be taken into account another form of it is used. It is then named a leftover service curve [3] that varies according to the cross traffic intensity changes and indicates how much bandwidth is actually left over in a network (or a network node) for servicing the through traffic. As such, it becomes inherently time-dependent. Then, in order to avoid calculation problems, an alternative function in form of the deterministic time-invariant leftover service curve is often used. And this does not lead to poor results; for more details of this approach see, for example [3].

Applying a little bit different techniques than those used in network calculus, Valaee and Li in [5] constructed another service curve that depends upon the cross traffic. It was named by them the $\in$-effective service curve.

Fidler in [3] explains the above curve as that which comes out from a certain optimization task formulated in the max-plus algebra [2]. One obtains then, as solutions, the successive times of servicing data portions consisting of *k* consecutive packets with *k* changing from, say, 1 to *K*. But, a drawback related with these solutions is that they depend upon an auxiliary parameter. The set of servicing times obtained in such a way is used afterwards to get an inverted function. And this inverted function forms $\in$-effective service curve under assumption that the so-called small parameter $\in$ (it will be defined precisely in the next section) equals zero. For more details, see [3].

Here, we present another interpretation of the $\in$-effective service devised by Valaee and Li [5]. It seems to be more convincing than that presented in [3]. In our interpretation, we show that the $\in$-effective service curve can be treated as a result of interpolation of the leftover service curve using some data. These data, which one can get by carrying out computer simulations or by performing measurements in a network, regard the values of waiting times of data portions consisting of *k* consecutive packets to be serviced (with *k* changing from, say, 1 to *K*, as above). The theoretical results presented in this paper are illustrated by showing some outcomes from simulations performed with the use of NS3 simulator. They show how servicing the through traffic depends upon the cross traffic at a given network node or a network processing device (as for example a router). Moreover, we observe here, similarly as in [5], that the $\in$-effective service curves achieved in simulations do not differ much from the straight lines (after leaving an initial range).

1. NOTIONS OF LEFTOVER AND EFFECTIVE SERVICE CURVES

Using network calculus terminology [2, 3], we can relate network traffic flows by exploiting the notion of a service curve. In this framework, the traffic flows are expressed with the use of a measure called the cumulative traffic [2, 3]. This is a number of bits (or packets) that enters or leaves a traffic system or device in the period from 0 to *t*. So it is a function of time and it can be viewed as belonging to the function class $\mathcal{F}$ [2]. Note that these are such functions that are wide-sense increasing and have values identically equal to zero for negative times (i.e. for $t < 0$), where the property of being wide-sense increasing means that a function $f(t)$ possesses this property if and only if $f(\tau) \le f(t)$ holds for all $\tau \le t$.

We start our derivations in this section with recalling the notion of the strict service curve [2, 3]. To this end, consider a traffic system with the input and output cumulative traffics denoted, respectively, by $A(t)$ and $D(t)$. The relation between them can be expressed for all the times $0 \le \tau \le t$ falling into a continuously backlogged period as [2, 3]

$$D(t) \ge D(\tau) + S(t-\tau), \tag{1}$$

where the function of time $S(t)$ means the so-called strict service curve [2, 3]. Note that in the literature one assumes this function to belong to the $\mathcal{F}$-class of functions, too. Furthermore, parameter τ in (1) means the beginning of the last backlogged period which occurred before *t*. In other words, we assume here that the system considered was fully empty directly before τ. That is it contained then no through as well as no cross traffic.

It was proved [2] that if a traffic system possesses the strict service curve then it has also the conventional service curve property. This means that the following inequality [2, 3]

$$D(t) \ge (A \otimes S)(t) = \inf_{0 \le \tau \le t} \left(A(\tau) + S(t-\tau) \right) \tag{2}$$

holds with $S(t)$ being the service curve that is identical with the strict service curve occurring in (1). In (2), the symbol $\otimes$ means the operation of min-plus convolution [2], [3]. Moreover, the symbol inf therein stands for the operation of finding infimum value, but $A(t)$ means the cumulative traffic at the system's input.

We assume here that the traffic at the system's input consists of two components

$$A(t) = A^t(t) + A^c(t), \tag{3a}$$

where $A^t(t)$ means the input through traffic, but $A^c(t)$ is the input cross traffic. The latter one is that which after servicing leaves another system output (or outputs). Accordingly, $D^t(t)$ and $D^c(t)$ mean the output through traffic and the output cross traffic. That is we have

$$D(t) = D^t(t) + D^c(t) \tag{3b}$$

for the aggregated output traffic.

Substituting (3a) and (3b) into (1), and after rearranging the corresponding components, we obtain

$$D^t(t) \ge D^t(\tau) + S(t-\tau) - \left(D^c(t) - D^c(\tau) \right) \tag{4}$$

for all the times $0 \le \tau \le t$ falling into a continuously backlogged interval. In the next step, using the causality condition relating D^c with A^c, i.e. $D^c(t) \le A^c(t)$, and afterwards applying the envelope bound on A^c, i.e. $A^c(t) - A^c(\tau) \le E^c(t-\tau)$, in (4), we arrive at

$$D^t(t) \geq D^t(\tau) + S(t-\tau) - E^c(t-\tau), \tag{5}$$

where $E^c(t)$ means an envelope [3, 4] of the input cross traffic in the system considered. Further, comparison of (5) with (1) shows that the function

$$S^t(t) = S(t) - E^c(t) \tag{6}$$

plays a kind of service curve for the through traffic at the system. For this, however, a slight modification of (6) is needed because of possibility of occurrence of the negative values of $S^t(t)$ for some times t. In order to ensure that all the values of a hypothetical service curve will be nonnegative, a function

$$S^t(t) = \left[S(t) - E^c(t)\right]^+ \tag{7}$$

is used instead of that given by (6). For more explanation, see [3]. In [3], this function is named the leftover service curve. Further, the symbol $\left[x(t)\right]^+$ in (7) stands for finding the maximum value in the set $\left\{x(t), 0\right\}$ for each time instant.

The leftover service curve $S^t(t)$ given by (7) determines the amount of service that is left over by the cross traffic for the through traffic at a given system. In other words, it is cross-traffic dependent. It can retain the property of a strict service curve or do not retain, as shown in [3, 6]. Independently of this, we will call however in what follows a leftover service curve any one that is cross-traffic dependent. (It can be simply also named a cross-traffic dependent service curve).

In [5], another service curve, which depends upon the cross traffic, was defined. We can say that it is an experiment-oriented one because it uses data obtained by measuring the traffic in a real network or by generating the traffic in a simulated system and registering it in the system's chosen nodes. These data are needed for the curve construction. More precisely, a batch of probing packets is sent at the system's input in the presence of cross traffic. The probing packets constitute the through traffic. Generally, they are not serviced immediately, but must wait for some time; the lengths of waiting times depend upon the cross traffic intensity. Further, the waiting times (called also delays) are measured and, after releasing from a certain constant delay, are recorded. Assume that we apply a probing sequence consisting of K consecutive packets. Then, the successive sums of the aforementioned waiting times can be expressed as

$$\Delta^{(k)} = \sum_{i=1}^{k} w_i, \quad k = 1, \ldots, K, \tag{8}$$

where w_i means the waiting time of the i-th probing packet. Further, describe the summed delays also in a probabilistic way as

$$T_k^{\varepsilon} = \inf\left(\tau \middle| \Pr\left(\Delta^{(k)} > \tau\right) \leq \varepsilon\right), \quad k = 1, \ldots, K, \tag{9}$$

where $\Pr\left(\Delta^{(k)} > \tau\right) \le \varepsilon$ stands for such a situation in which the probability of occurrence of $\Delta^{(k)} > \tau$ is less than ε. Using the consecutive parameters T_k^{ε} calculated by applying (8) and (9), the so-called $\in$-effective service curve $S_{\varepsilon}(t)$ can be defined [5] as a function consisting of K segments described by

$$S_{\in}(t) = S_{\in}(t,k) = \frac{t - T_{k-1}^{\in}}{T_k^{\in} - T_{k-1}^{\in}} + k - 1, \;\; T_{k-1}^{\in} \le t \le T_k^{\in}, \;\; k = 1,2,...,K, \tag{10}$$

with the index *k* denoting the *k*-th segment, and with the initial value of $T_0^{\varepsilon} = 0$. Note further that the segment description $S_{\varepsilon}(t,k)$ given by (10) with the index *k* changing from 1 to *K* is in fact also the description of the whole function $S_{\varepsilon}(t)$. The traffic expressed by $S_{\varepsilon}(t)$ is counted in packets. To express it in bits, $S_{\varepsilon}(t)$ must be simply multiplied by the packet length (in bits). In what follows, we denote this length by l_B (assumed the packets are of the same length).

It is worth noting that the values of T_k^{ε} in (9) can be estimated by the actual values of $\Delta^{(k)}$. That is we can write

$$T_k = T_k^{\varepsilon=0} \cong \Delta^{(k)}, \quad k = 1,...,K\,. \tag{11}$$

The simplification given by (11) is important from the practical point of view because the probability functions $\Pr\left(\Delta^{(k)} > \tau\right)$ are often not known at all. But with the use of (11), estimation of the function (10) becomes easy.

Observe further that the summed delays defined by (8) or (9) are the characteristic points of the $\in$-effective service curve $S_{\varepsilon}(t)$ and change with the cross traffic intensity. Their values become larger with the cross traffic intensity increase. Looking at (10), we see that this has such an effect that the slopes of the straight lines of the segments of $S_{\varepsilon}(t)$ are smaller. This causes the curve $S_{\varepsilon}(t)$ to move nearer to the *Ot* axis.

Finally, it seems that the service curve $S_{\varepsilon}(t)$ can approximate in some way the leftover service curve expressed by (7). We will consider this topic in the next section.

2. EFFECTIVE SERVICE CURVE AS SOLUTION OF AN INTERPOLATION PROBLEM

Here, we present interpretation of the $\in$-effective service curve $S_{\varepsilon}(t)$ as such a curve that is obtained by means of linear interpolation of a leftover service curve, not necessarily strict one. In this interpolation, the experimental data (obtained by performing real measurements or carrying out appropriate traffic simulations) are used. To begin, let us approximate the function of the system's input through traffic $A^t(t)$, consisting of a batch of K probing packets, similarly as in [8]. That is as shown in Fig. 1.

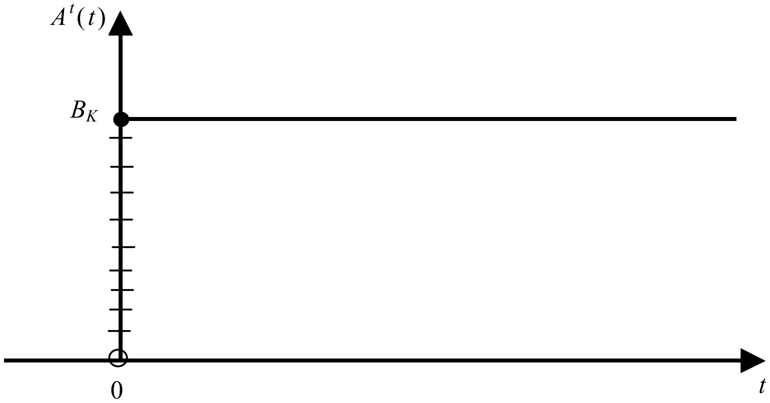

Fig. 1. The cumulative traffic function approximating the input through traffic $A^t(t)$, consisting of a batch of K probing packets.

In Fig. 1, $B_K = K \cdot l_B$ means the total length in bits of the whole batch of K probing packets. Moreover, note that the function $A^t(t)$ in Fig. 1 is assumed to be right-continuous.

Assume now that the traffic system considered behaves approximately linearly when servicing the through traffic. That is we can write

$$D^t(t) \cong (A^t \otimes S^t)(t) = \inf_{0 \le \tau \le t} \left(A^t(\tau) + S^t(t-\tau) \right), \tag{12}$$

where the function $S^t(t)$ means now and in what follows any leftover service curve (i.e. a time-independent and cross-traffic dependent service curve) fulfilling (12). As we know from the previous section and the literature cited therein, the function given by (7) can play this role in many cases.

Further, consider such the times t for which the following holds.

$$S^t(t) \le B_K \tag{13}$$

Applying $A^t(t)$ as given in Fig. 1 in (12) and analyzing then the inner expression on the right-hand side of (12), i.e. $A^t(\tau) + S^t(t-\tau)$, for the times $0 \le \tau \le t$, we find easily that its minimal value equals $S^t(t)$ for each t for which (13) holds. So this results in

$$D^t(t) \cong S^t(t). \tag{14}$$

Note that (14), under the conditions of the scenario described above, can be also interpreted as

$$\tilde{S}^t(t) = D^t(t), \tag{15}$$

where $\tilde{S}^t(t)$ means an estimating function of the function $S^t(t)$. This estimating function for the range of times $0 \le t \le t_{\max} = \left(S^t\right)^{-1}(B_K)$, where $\left(S^t\right)^{-1}$ means the inverted function $S^t(t)$, is obtained as the output through traffic function $D^t(t)$.

Note however that we do not arrive at a continuous function $D^t(t)$ when working with the data taken from the real traffic measurements or computer simulations of real networks. This is because we do not have to do then with single bits, but with the whole packets. That is we register and record arrival and departure times of the data portions (consisting of bits), i.e. packets, as sketched for $D^t(t)$ in Fig. 2.

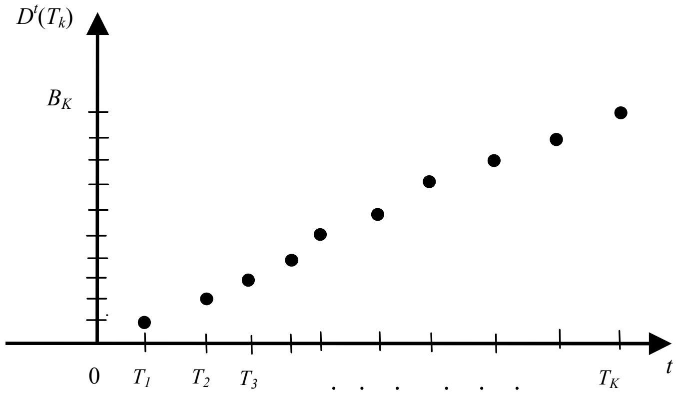

Fig. 2. Visualization of the departing times of packets building the cumulative output through traffic function $D^t(t)$

Using the pairs of discrete values $\{T_k, D^t(T_k)\}$ shown in Fig. 2, we can interpolate the other points of the curve $D^t(t)$ by using, for example, a linear interpolation formula having a general form

$$y^*\left(x^*\right) = f\left(x_{k-1}\right) + \frac{x^* - x_{k-1}}{x_k - x_{k-1}}\left(f\left(x_k\right) - f\left(x_{k-1}\right)\right), \tag{16}$$

where the value x^* lies on the Ox between the values x_{k-1} and x_k (with $x_k > x_{k-1}$). Furthermore, $f(x_{k-1})$ and $f(x_k)$ denote the values of the function $f(x)$ at the points x_{k-1} and x_k, respectively. We assume that the pairs $\{x_{k-1}, f(x_{k-1})\}$ and $\{x_k, f(x_k)\}$ are known from the measurements (simulations). And $y^*(x^*)$ is the interpolated value of the function $f(x)$ at the point x^*.

Applying now (16) in our case, i.e. substituting in this expression $x^* = t$, $y^*(x^*) = D^t(t)$, $x_{k-1} = T_{k-1}$, $x_k = T_k$, $f(x_{k-1}) = D^t(T_{k-1}) = k-1$ (what means that we count here the traffic in number of packets), $f(x_k) = D^t(T_k) = k$, with the initial values $x_0 = T_0 = 0$ and $f(x_0) = D^t(T_0) = 0$, we get

$$D^t(t) = k - 1 + \frac{t - T_{k-1}}{T_k - T_{k-1}}, \quad T_{k-1}^{\varepsilon} \le t \le T_k^{\varepsilon} \tag{17}$$

with k changing in the range $k = 1,...,K$.

Comparison of (10) with (17) shows that these expressions are identical. Hence, we can write

$$D^t(t) = S_{\varepsilon}(t) = S_{\varepsilon}(t,k), \; T_{k-1}^{\varepsilon} \le t \le T_k^{\varepsilon}, \; k = 1,...,K. \tag{18}$$

That is the so-called ϵ-effective service curve $S_{\varepsilon}(t)$ elaborated in [5] can be viewed as a solution of an interpolated problem as described above in this section.

The procedure described above of getting the linearly interpolated function $D^t(t)$, which uses the data presented in Fig. 2, is illustrated in Fig. 3.

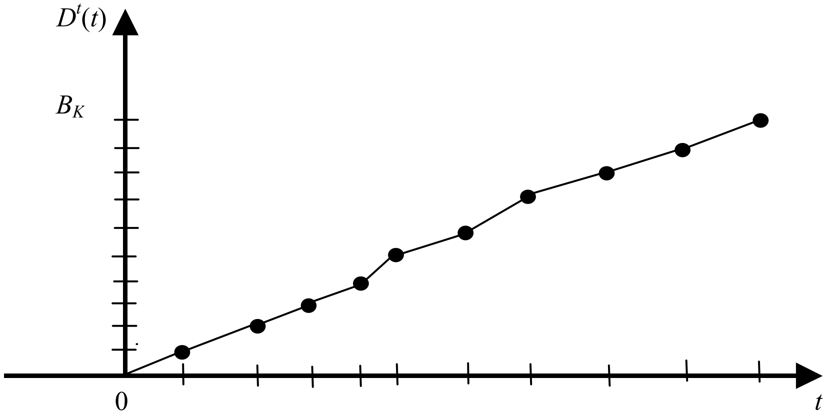

Fig. 3. Visualization of the interpolation process for getting the interpolated function $D^t(t)$

And finally, relating (18) with (14), and (15), we can write

$$\tilde{S}^t(t) = S_{\varepsilon}(t). \tag{19}$$

Equation (19) states that the estimating function of the leftover service curve for the through traffic obtained in the linear interpolation process using the data gathered in measuring (simulation) of the departing through traffic (as described above) is equal to the so-called ϵ-effective service curve devised in [5].

3. SIMULATIONS WITH NS3

For illustration, we present in this section some results of simulations carried out with the use of the network simulator NS3. These results were also presented by this author and his coworker in another paper [7] on applications of the leftover and $\in$-effective service curves in admission control in ad hoc networks. The scenario modeled here was similar to that described in [5]. A network simulated had 25 mobile nodes placed on the area of 250 × 250 m^2. Moreover, the MAC layer applied was defined according to the IEEE 802.11 standard with the CSMA/CA access procedure. Further, the simulations were performed for different intensities of the cross traffic. The data about the traffic in the respective nodes were registered and used afterwards to obtain the $\in$-effective service curves.

Fig. 4 presents some of the results achieved. Note that the simulated curves are not here straight lines as in [5]. This difference will need to be clarified by carrying out further investigations. However, observe that the curves do not differ much from the straight lines outside the initial area (similarly as in the paper by Valaee and Li [5]). Note also that the time and cumulative traffic values on the axes in Fig. 4 are normalized.

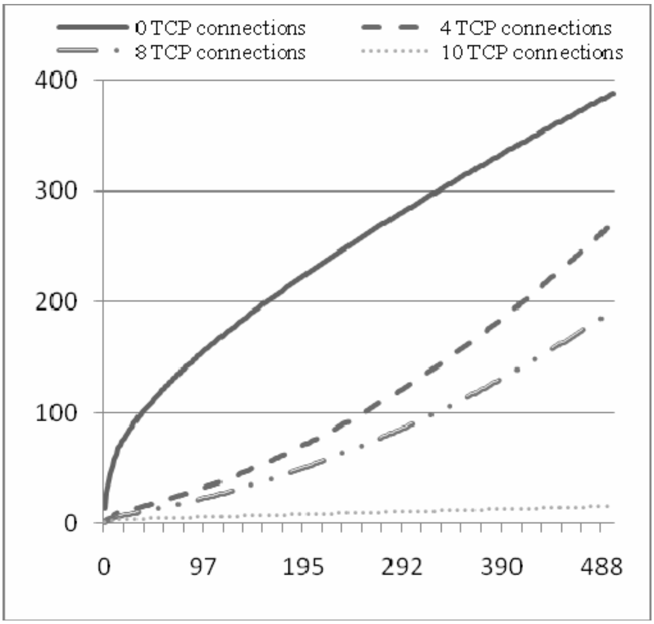

Fig. 4. The $\in$-effective service curves obtained in simulations with the use of NS3 simulator for the through traffic in the absence or presence of the crossing traffic [7]

CONCLUSIONS

The basic result achieved in this paper is that the estimating function of the leftover service curve for the through traffic, obtained in a linear interpolation process using the data gathered in measuring (simulation) of the departing through traffic (when the input through traffic consists of a sequence of probing packets according to a scenario devised in [5]), is equal to the so-called ϵ-effective service curve [5].

REFERENCES

1. Firoiu V., Le Boudec J.-Y., Towsley D. and Zhang Z.-L., *Theories and models for Internet quality of service*, Proc. IEEE, vol. 90, no. 9, pp. 1565–1591, May 2002.
2. Le Boudec J.-Y. and Thiran P., *Network Calculus. A Theory of Deterministic Queuing Systems for the Internet*, Springer Verlag, Berlin, Germany, 2004.
3. Fidler M., *A survey of deterministic and stochastic service curve models in the network calculus*, IEEE Communications Surveys & Tutorials, vol. 12, no. 1, First Quarter 2010, pp. 59–86.
4. Agrawal R., Cruz R. L., Okino C. and Rajan R., *Performance bounds for flow control protocols*, IEEE/ACM Trans. Networking, vol. 7, no. 3, pp. 310–323, June 1999.
5. Valaee S. and Li B., *Distributed call admission control for ad hoc networks*, in Proc. IEEE 56th VTC, Sep. 2002, pp. 1244–1248.
6. Bouillard A., Gaujal B., Lagrange S. and Thierry E., *Optimal routing for end-to-end guarantees using Network Calculus*, Performance Evaluation, vol. 65, no. 11, pp. 883–906, 2008.
7. Borys A. and Aleksiewicz M., *Use of leftover service curve in admission control in ad hoc networks*, in Proc. of the 11th IFAC/IEEE International Conference on Programmable Devices and Embedded Systems (PdeS 2012), pp. 33–37, May 2012.
8. Borys A., *Some remarks on practical aspects of the effective service curve use in ad hoc networks*, Gdynia Maritime University Journal, vol. 1, pp. 26–36, 2015.