www.czasopisma.pan.pl PAN www.journals.pan.pl

jet INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2018, VOL. 65, NO. 1, PP. 133-138
Manuscript received September 11, 2018; revised January, 2019. DOI: 10.24425/ijet.2019.126293

Bandwidth Estimation Using Network Calculus in Practice

Katarzyna Wasielewska and Andrzej Borys

***Abstract*—An available bandwidth at a link is an unused capacity. Its measuring and/or estimation is not simple in practice. On the other hand, we know that its continuous knowledge is crucial for the operation of almost all networks. Therefore, there is a continuous effort in improving the existing and developing new methods of available bandwidth measurement and/or estimation. This paper deals with these problems. Network calculus terminology allows to express an available bandwidth in terms of a service curve. The service curve is a function representing a service available for a traffic flow which can be measured/estimated in a node as well as at an end-to-end connection of a network. An Internet traffic is highly unpredictable what hinders to a large extent an execution of the tasks mentioned above. This paper draws attention to pitfalls and difficulties with application of the existing network calculus methods of an available bandwidth estimation in a real Internet Service Provider (ISP) network. The results achieved in measurements have been also confirmed in simulations performed as well as by mathematical considerations presented here. They give a new perspective on the outcomes obtained by other authors and on their interpretations.**

***Keywords*—bandwidth estimation, network calculus, quality of service, Internet traffic**

I. Introduction

An available bandwidth estimation in computer networks is still an important and interesting research problem. In practice, results of this estimation can be used in different fields like: network traffic engineering, Quality of Service (QoS) and Service Level Agreement (SLA) evaluation, congestion control, route selection process and others. It is well-know that environmental conditions on an end-to-end path in the Internet can change dynamically and in most cases in an unpredictable way. Therefore, an available bandwidth estimation is a nontrivial task. The estimation results are influenced by many factors like intermediary devices, medium properties as well as a cross-traffic. There are a lot of methods developed for an available bandwidth estimation but no one is perfect.

Formally, the available bandwidth B on a route at time t means that unused bandwidth, which can be utilized by an application without any influence of the transmission quality of flows occurring on this route. The available bandwidth depends on time t. If there occur n nodes on an end-to-end path, then the available bandwidth B on this path in time t is given by

$$B(t) = \min_{i=1,\dots,n} \{A_i(t)\}, \tag{1}$$

where A_i is the available bandwidth in the node i.

There are different methods of the available bandwidth estimation. Among them, active methods rely on sending probing packets to a network and making measurements after these packets reach a destination. On the other hand, passive methods require an analysis of the captured traffic. Among the active methods, we have the following ones: IGI [1], Spruce [2], Pathload [3], pathChirp [4], and others.

Good tutorials on the active methods and their comparisons are presented in [2], [5], [6], [7], [8]. Although these methods provide more accurate measurement results, they require some installations on both sides of an end-to-end path measured, what is not always possible. Furthermore, an installed software may affect the operation of other applications (and thereby the usage of the available bandwidth).

On the other hand, the passive methods can be more practical, in particular in the ISP environment. They allow to perform more quickly measurements, the results of which are used afterwards in a bandwidth estimation procedure. Furthermore, they do not influence a network operation as well as do not need any changes in network configuration and in its control mechanisms. Their working principle and basic task is simply to capture a traffic for a further analysis.

Evaluating suitability of different available bandwidth estimation approaches, we must bear in mind the fact that any network changes continuously. For example, some changes may occur with the time elapsing in its configuration, number of users, access schemes, and control mechanisms. Useful available bandwidth estimation methods are those ones that respond correctly to the above phenomena. Obviously, the measurement and estimation results are then time-dependent. Moreover, it is also worth noting here that the existing available bandwidth estimation methods deliver better results for no cross-traffic paths. We know however, on the other hand, that the latter cannot be guaranteed in the ISP networks [5].

The network calculus (NC) techniques [9] deliver some solutions regarding the problem formulated shortly above. In NC, an available bandwidth is described by a service curve – a function which expresses a service available to a traffic flow [10]. The service curve can be estimated using a passive measurement, what is a natural method in real networks

The work of the second author of this paper was supported in part by the grant AMG DS/450/2018.

K. Wasielewska is with the Institute of Applied Informatics, The State University of Applied Sciences in Elblag, Poland (e-mail: k.wasielewska@ieee.org).

A. Borys is with the Department of Marine Telecommunications, Faculty of Electrical Engineering, Gdynia Maritime University, Gdynia, Poland, (email: a.borys@we.am.gdynia.pl).

because it does not require injection of any packet probes into the network.

In [11] (as well as in its extended earlier version [12]), network calculus methods for available bandwidth estimation have been presented. In more detail, the authors of [11] and [12] have formulated the problem of available bandwidth estimation as a problem of finding an unknown service curve of a system with the use of the min-plus algebra [13]. They have discussed the following three methods: passive measurements, a rate scanning, and so-called rate chirps. The first approach mentioned allows the estimation based on the traffic transmitted in a network while the other two require injecting an additional traffic into the network. It is worth noting at this point that the latter two methods can influence the network performance and their application is not simple in a real network. This paper presents some results of experiments carried out with the use of passive measurements in the Internet.

The paper is organized as follows. Section II contains a short description of the network calculus method based on passive measurements that has been presented in [11]. Application of this method in an ISP network and results of experiments performed in it are presented in Section III. Additionally, outcomes of simulations carried out, which confirm the experimental ones mentioned above, are provided in Section IV. Moreover, a mathematical evidence of the results presented in the previous two sections is given (Section V). The paper ends with Section VI that contains conclusions.

II. Passive Measurements Method

In [11], an available bandwidth estimation method based on the use of a service curve estimator $\tilde{S}$ has been proposed. The service curve estimator $\tilde{S}$ derived in [11] is given by

$$\tilde{S} = D^p \oslash A^p, \quad (2)$$

where $D^p(t)$ means a traffic departure function and $A^p(t)$ is a traffic arrival function. We understand the latter as coming from a traffic trace of one or more flows. Both the functions express a sum of bits in the outgoing and incoming traffic, respectively, in the time $[0, t]$. Furthermore, $\oslash$ is a deconvolution operator defined in the sense of the min-plus algebra in the following way:

$$f \oslash g(t) = \sup_{\tau \geq 0}\{f(t+\tau) - g(\tau)\} \quad (3)$$

for functions f and g. In [11], it has been shown that the estimator $\tilde{S}$ given by (2) can be regarded as the best possible estimate of an actual service curve that can be deduced from measurements of A^p and D^p. Furthermore, observe that the deconvolution operation given by (3) computes the largest available bandwidth that can be justified on the basis of a given traffic trace [11]. This means that this method cannot be worse than any existing Measurement-Based Admission Control (MBAC) method. In order to achieve the best possible estimate, the authors of [11] maximized the use of a link capacity by multiplying the number of flows or they lowered the transmission capacity of a link. In this way, they obtained the best approximation when the intensity of the transmitted traffic has reached the maximum.

This paper presents a new interpretation of the aforementioned result presented in [11]. It relies upon that the estimator $\tilde{S}$ does not in fact estimate an available bandwidth but only provides us with the information whether a user or users gets/get a required bandwidth or not. This interpretation has been proved by experiments we carried out in a real ISP network, simulations as well as mathematical considerations, which are presented in the next sections.

III. Tests in a Real ISP Network

We have used and verified the service curve estimator $\tilde{S}$ in a real Internet network. The ISP network topology exploited in our experiments is presented in Fig. 1. In fact, it is only a part of a larger ISP network. Fig. 1 shows intermediary devices on the end-to-end path and a cross traffic. RouterA in Fig. 1 is a border router, which is connected with an another Internet service provider by a fiber optic link (100 Mb/s). Network devices are connected by wire and wireless links. The wireless links worked very stable - around the bandwidth of maximum 150 Mb/s and delay of maximum 1 ms. All routers experienced cross traffic. Additionally, the network in Fig. 1 used symmetrical and unsymmetrical links and had individual and commercial customers. For particular customers, there were limitations on download/upload rates; there occurred also some other limitations. Further, it is worth noting that packets in the network of Fig. 1 travelled always the same routes.

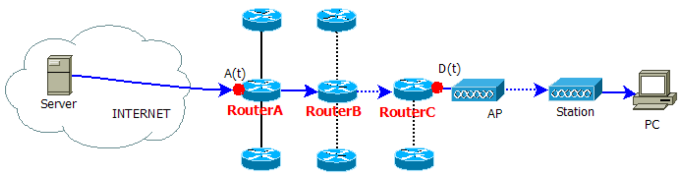


Fig. 1. The physical topology of an ISP network used in experiments.

The tests conducted and described in this paper were based on the following methodology:

1. The Internet traffic was captured on the selected interfaces of ISP routers.
2. The Internet traffic was filtered.
3. The $A(t)$ and $D(t)$ functions were generated on the basis of the measured traffic (time series of the input and output traffic, respectively).
4. The values of the service curve $\tilde{S}$ were calculated from (2).

We carried out tests for different flows (single streams, sets of streams) and separately for different kinds of traffic (TCP, UDP and HTTP).

First, we captured an Internet traffic downloaded by a customer (PC in Fig. 1). The maximal download rate for that customer was set on 6.41 Mb/s. Fig. 2 presents a load caused by all traffic of this customer. For the purpose of evaluation, we selected four streams (green color is related with an aggregate of four streams) and performed calculations for 200 seconds (from the 100th to the 300th second). Furthermore, note that the green curve versus time in Fig. 2 runs close to the red one. That means that the former curve describes approximately well also the total user traffic.

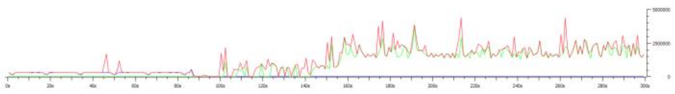


Fig. 2. Load of network by downloaded traffic (red color – total traffic, green color – aggregated traffic, blue color – one of another streams).

In short, the traffic that was exploited in the experiments discussed in this paper can be characterized as follows: amount of data: 265.86 Mb (24273 packets), average rate: 1.329 Mb/s, maximal rate: about 1.7 Mb/s. Note that the latter means that the available bandwidth was equal to about 4.71 Mb/s (6.41 Mb/s minus 1.7 Mb/s).

Fig. 3 presents the arrival and departure functions as well as the service curve estimator $\tilde{S}$ for the case described above.

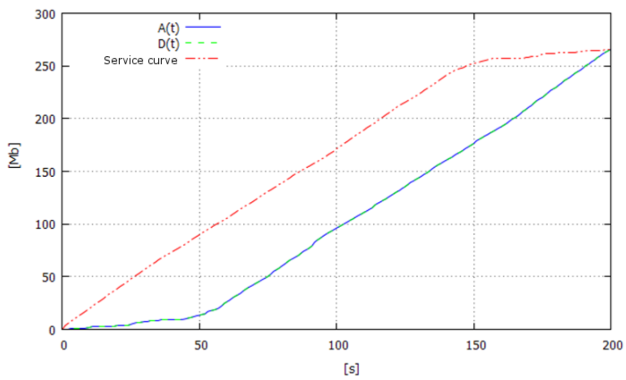


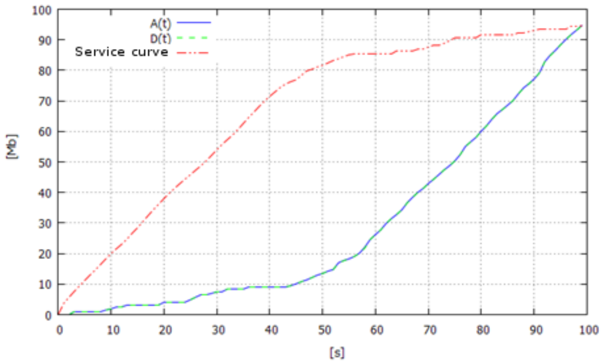
Fig. 3. The arrival and departure functions as well as the service curve estimator $\tilde{S}$ for the aggregated flow discussed in the text.

As it is seen in Fig. 3, the service curve estimator $\tilde{S}$ has a slope similar to the one that possess the arrival and departure functions. So, this means that the available bandwidth estimated, in this case, is far from the value of 4.71 Mb/s. Also, it is worth noting here that the service curve estimator connects with the departure function $D(t)$ at the end of the estimation period (at time t=200 s in Fig. 3).

The above result and observation have been verified in another test. To perform it, we have divided the traffic sample into two parts – each one lasting 100 seconds. Afterwards, we have carried out estimation. Fig. 4 presents graphs computed separately for the first 100 seconds - (a), and for the next 100 seconds - (b).

Observe from Fig. 4 that the departure function $D(t)$ connects with the service curve estimator $\tilde{S}$ at the end of the estimation period, as before. Here, this occurs at t=100 ms (a), and t=200 (b), respectively. Additionally, regarding the traffic average rates and the estimated bandwidths, we see that the results are also similar. This follows from the data gathered in Table I for the two calculation periods lasting 100 ms that were mentioned above. Really, we see that an average rate for the traffic period lasting altogether 200 seconds equals 1.329 Mb/s ((0.946+1.712)/2=1.329), and this is the same value we had before.

(a)



(b)

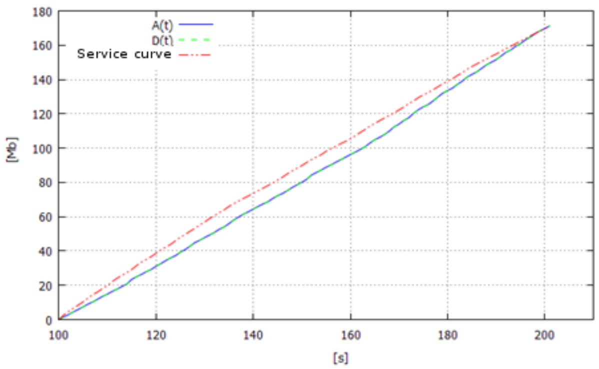


Fig. 4. The arrival and departure functions as well as the associated estimated service curves for the first 0-100 seconds (a), and the next 100-200 seconds (b).

TABLE I
CHARACTERISTICS TRAFFIC PERIODS ANALYZED

Traffic unit	Duration [s]	Data [Mb]	Average rate [Mb/s]
Streams A-D	100 (0-100)	94.64	0.946
Streams A-D	100 (100-200)	171.23	1.712

We have verified that a maximal rate (peak rate) in the trial shown in Fig. 2 has occurred in a time close to the 192nd second - 4 000 000 bits transmitted during one second. This means that maximal rate (in the short-term sense) was then equal to 3.815 Mb/s. Therefore, an available bandwidth for this period was minimum 2.595 Mb/s. And similarly as in the two previous cases discussed before, our calculations did not confirm this fact also here.

We have performed a whole series of traffic tests to verify whether the procedure of the available bandwidth estimation discussed in detail above is useful in practice. More results have been presented in [14]. All of these tests have given negative results. So, the only reasonable conclusion in this situation is to say that the function defined by (2) is not an appropriate estimator for the available bandwidth estimation in computer networks.

It follows from our observations of the service curve estimators $\tilde{S}(t)$ obtained in our experiments that they deliver rather information about the potential possibilities of bandwidth usage by a traffic.

IV. Simulations

To confirm the results and observations discussed in the previous section, we have also performed some simulations. First, in these simulations, we have used the following forms of the arrival and departure functions: $A(t) = 4t$ and $D(t) = 2t$ for $t \geq 0$. Note this is a case in which the traffic comes to a system faster than it can be served. The first curve in Fig. 5 in its left corner illustrates the service curve estimator $\tilde{S}$, obtained according to (1), for the aforementioned arrival and departure functions. Observe that in this case the service curve estimator $\tilde{S}$ overlaps with the departure function $D(t)$.

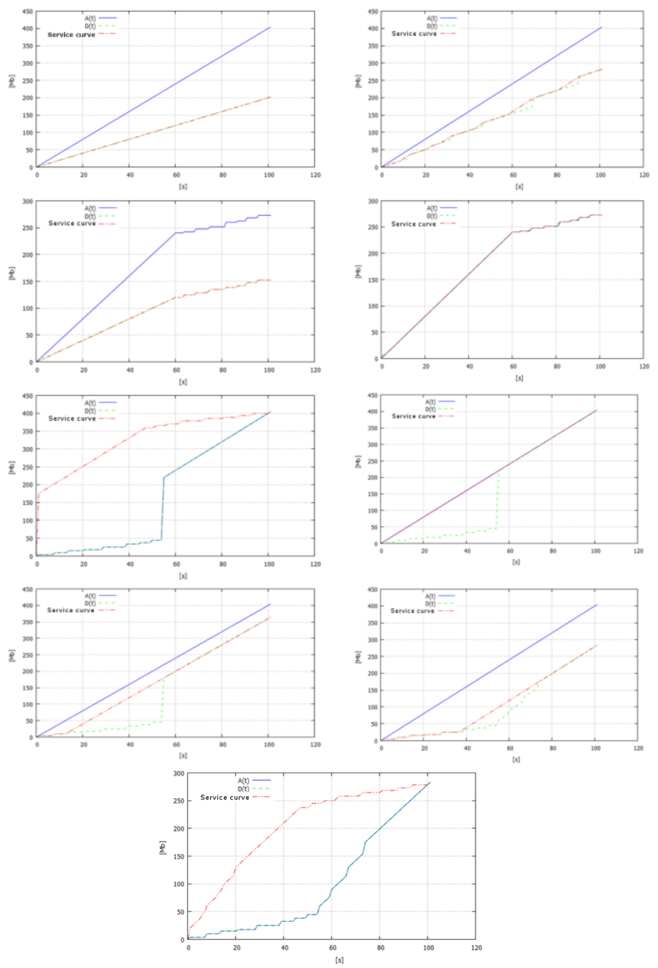


Fig. 5. Examples of service curves estimators corresponding to different arrival and departure functions – obtained in simulations.

The remaining diagrams in Fig. 5 present another examples of the service curve estimators obtained with the use of different randomly chosen arrival and departure functions. In these simulations, we did not however use any traffic generators. We only kept the condition that $A(t) \geq D(t)$, which must hold for all $t \geq 0$. At this point, we would like also to point out that the results of simulations presented in Fig. 5 can describe different traffic instances as, for example, those existing in a node or on an end-to-end connection (as well).

The main conclusion following from different forms of the functions describing the estimators $\tilde{S}(t)$, which are presented in Fig. 5 for different scenarios, is that they cannot be interpreted as an available bandwidth estimation method. It should be mentioned that obtained results do not depend on capacity of node or path (if a capacity is big enough, obviously).

V. Mathematical Evidence

The results obtained in a real Internet network and simulations, which were presented in the previous two sections, are confirmed here using some mathematical arguments.

To this end, suppose that a node (or a network) is a linear or nonlinear system described in the min-plus algebra. (Note that in according to [12] a system is linear if its load does not exceed some threshold; or, in other words, a system behaves as a nonlinear one if the traffic volume in it exceeds its capacity.) Further, an unknown service curve S describes the process of transforming an incoming traffic to the system's output, resulting in an outgoing traffic. This service curve S expresses an available bandwidth for the traffic and it can be a constant bit rate or more complex function. Further, suppose that the system's traffic is in form of a sequence of packets described by the functions A^p and D^p defined at the beginning of section II. Assume also that a system considered is time-invariant.

Now, we will assume that the incoming traffic to a system has a constant rate. That is we assume that the system's incoming traffic for $t \geq 0$ is described by an arrival function that is given by

$$A^p(t) = Rt, \tag{4}$$

where R is a rate of the incoming traffic.

In what follows, we consider two variants of the system's service curve: a straight line beginning at $t = 0$ and a time-shifted one (describing a constant rate servicing with a delay). And we start with the first one given by

$$S(t) = rt \tag{5}$$

for $t \geq 0$, where r is a rate of traffic processing in this system. The analysis presented in [15] shows that then the system's departure function assumes the following form:

$$D^p(t) = \begin{cases} rt & \text{for } \ r < R, \\ Rt = rt & \text{for } \ r = R, \\ Rt & \text{for } \ r > R. \end{cases} \tag{6}$$

Furthermore, assume similarly as in [15] that the network considered is linear.

Now, we can use the above knowledge for calculation of the service curve estimator $\tilde{S}$ of the system considered. That is by applying (2), we get

Case 1. If $r < R$, then

$$\tilde{S}(t) = \sup_{\tau}\{D^p(t+\tau) - A^p(\tau)\} = \sup_{\tau}\{r(t+\tau) - R\tau\} = \\ = \sup_{\tau}\{(r-R)\tau + rt\} = rt. \tag{7}$$

Case 2. If $r = R$, then

$$\tilde{S}(t) = \sup_{\tau}\{D^p(t+\tau) - A^p(\tau)\} = \sup_{\tau}\{r(t+\tau) - R\tau\} = \\ = \sup_{\tau}\{rt\} = rt = Rt. \tag{8}$$

Case 3. If $r > R$, then

$$\tilde{S}(t) = \sup_{\tau}\{D^p(t+\tau) - A^p(\tau)\} = \sup_{\tau}\{R(t+\tau) - R\tau\} = \\ = \sup_{\tau}\{(R-R)\tau + Rt\} = Rt. \tag{9}$$

for the three cases distinguished in (6).

Looking at (7) and (8), we see that in these two cases, when the incoming traffic exploits in a maximal way the possibilities offered by the system considered, the service curve estimator $\tilde{S}$ (which is interpreted in [12] as an available bandwidth) overlaps with the outgoing traffic curve. Whereas (9) shows that if the network considered is able to service the incoming traffic faster than its rate R (that is when an incoming traffic does not reach a "saturation point" of the servicing system considered), then the estimator curve $\tilde{S}(t)$ overlaps with the incoming traffic curve. So, in the latter case, it is rather not possible to find a good estimator of the available bandwidth, in particular when the slope r of $S(t)$ given by (5) is much greater than the traffic rate r.

Consider now the second service curve variant mentioned above. That is the case when it has a form of a time-shifted straight line. Then, this curve is called a rate-latency service curve and described by the following function:

$$S(t) = r[t-T]^+ = \begin{cases} r(t-T) & \text{for } t > T \\ 0 & \text{for } t \le T, \end{cases} \tag{10}$$

where T is a latency.

It has been shown in [15] that for the service curve given by (10) one obtains $D^p(t) = 0$ for $t < T$ and

$$D^p(t) = \begin{cases} r(t-T) & \text{for } r < R, \\ R(t-T) = r(t-T) & \text{for } r = R, \\ R(t-T) & \text{for } r > R. \end{cases} \tag{11}$$

for $t \ge T$.

Applying (4) and (11) in (2) gives

Case 1. If $r < R$, then

$$\tilde{S}(t) = \sup_{\tau}\{D^p(t-T+\tau) - A^p(\tau)\} = \\ = \sup_{\tau}\{r(t-T+\tau) - R\tau\} = \\ = \sup_{\tau}\{r(t-T) + (r-R)\tau\} = r(t-T). \tag{12}$$

Case 2. If $r = R$, then

$$\tilde{S}(t) = \sup_{\tau}\{D^p(t-T+\tau) - A^p(\tau)\} = \\ = \sup_{\tau}\{R(t-T+\tau) - R\tau\} = \sup_{\tau}\{R(t-T)\} = \\ = R(t-T). \tag{13}$$

Case 3. If $r > R$, then

$$\tilde{S}(t) = \sup_{\tau}\{D^p(t-T+\tau) - A^p(\tau)\} = \\ = \sup_{\tau}\{R(t-T+\tau) - R\tau\} = = \sup_{\tau}\{R(t-T)\} = \\ = R(t-T). \tag{14}$$

Observing now the results obtained in (12), (13), and (14), we see they are similar to those achieved in the previous case (of a peak-rate service curve). More precisely, for $r \ge R$, the slope of the incoming traffic curve is equal to the slope of the service curve estimator $\tilde{S}$. So, we can conclude that if a system offers some bandwidth and an incoming traffic to this system is smaller than this bandwidth, then the estimator $\tilde{S}$ is not able to estimate an available bandwidth correctly. Then, it is a bad estimator of an actual service curve S.

Now, note that we can generalize the observations made above in form of the following two theorems.

Theorem 1. Let incoming and outgoing traffic have constant rates r_A and r_D, respectively, and let $r_D \le r_A$. Then, the service curve estimator $\tilde{S}(t)$ defined by (2) has a slope of the outgoing traffic curve D.

Proof. If $r_D \le r_A$, then

$$\tilde{S}(t) = \sup_{\tau}\{D^p(t+\tau) - A^p(\tau)\} = \sup_{\tau}\{r_D(t+\tau) - r_A\tau\} = \\ = \sup_{\tau}\{(r_D - r_A)\tau + r_D t\} = r_D t,$$

and this ends the proof.

Obviously, consideration of the case when $r_D > r_A$ does not make sense because an outgoing traffic cannot be transmitted faster than the associated incoming one.

The second theorem we formulate here regards a value of the service curve estimator $\tilde{S}(t)$ at its right-hand side corner t_0 (that is in a case that occurs in practice, when the calculation of $\tilde{S}(t)$ is possible only in a finite period).

Theorem 2. Let take into account the service curve estimator defined by (2). Then, for measurement data available only for the period $[0, t_0]$, we have

$$\tilde{S}(t_0) = D^p(t_0).$$

Proof. Note that for the observation time t_0 we get the following:

$$\tilde{S}(t_0) = \sup_{\tau=0}\{D^p(t_0) - A^p(0)\} = D^p(t_0),$$

and this ends the proof.

Note that the result presented in theorem 2 strengthens our conclusion that the estimator given by (2) is not a really good estimator of an available bandwidth. Even in the cases of getting acceptable results of estimation in the middle part of the estimation period, we can obtain non-satisfactory outcomes in the vicinity of the estimation period. The tests performed in a ISP network as well as simulations we presented in the previous sections confirm the above.

Furthermore, all the results presented above allow also to say that any analysis for more complex (than the constant rate (CBR)) cases of A^p and D^p waveforms as well as of more complex traffic processing systems will lead to similar conclusions.

VI. Conclusions

In this paper, we have presented a new interpretation of a bandwidth estimation method that uses the network calculus theory. By carrying out experiments in a real Internet network as well as computer simulations, and finally proving mathematically the results of the former investigations, we have shown that the aforementioned method does not provide available bandwidth estimates. It only provides us with the information whether a user or users gets/get a required bandwidth or not.

References

[1] N. Hu, P. Steenkiste, Evaluation and characterization of available bandwidth probing techniques, IEEE Journal on Selected Areas in Communications, vol. 21, pp. 879--894, 2003.

[2] J. Strauss, D. Katabi, F. Kaashoek, A measurement study of available bandwidth estimation tools, in Proceedings of the 3rd ACM SIGCOMM conference on Internet, 2003.

[3] M. Jain, C. Dovrolis, Pathload: a measurement tool for end-to-end available bandwidth, in Proceedings of Passive and Active Measurements, pp. 14-25, 2002.

[4] V. Ribeiro, R. Riedi, R. J. Baraniuk, J. Navratil, L. Cotrell, PathChirp: efficient available bandwidth estimation for network paths, in Passive and Active Measurement Workshop, 2003.

[5] E. Goldoni, M. Schivi, End-to-end available bandwidth estimation tools, an experimental comparison, in Proceedings of the Second international conference on Traffic Monitoring and Analysis, pp. 171-182, 2010.

[6] C. D. Guerrero, M. A. Labrador, On the applicability of available bandwidth estimation techniques and tools, Computer Communications, vol. 33, no. 1, pp. 11-22, 2010.

[7] R. Prasad, M. Murray, C. Dovrolis, K. Claffy, Bandwidth estimation: metrics, measurement techniques, and tools, Network, IEEE, vol. 17, no. 6, pp. 27-35, 2003.

[8] A. Shriram et al., Comparison of Public End-to-End Bandwidth Estimation Tools on High-Speed Links, Passive and Active Network Measurement Workshop, vol. 3431, pp. 306-320, 2005.

[9] R. L. Cruz, "A calculus for network delay, Part I and II, IEEE Transactions on Information Theory, vol. 37, 1991.

[10] R. L. Cruz, "Quality of service guarantees in virtual circuit switched networks, IEEE Journal on Selected Areas in Communications, vol. 13, no. 6, pp. 1048-1056, 1995.

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

[12] J. Liebeherr, M. Fidler, S. Valaee, A System Theoretic Approach to Bandwidth Estimation, Computing Research Repository, vol. abs/0801.0455, 2008, http://arxiv.org/abs/0801.0455.

[13] J.-Y. Le Boudec, P. Thiran, Network Calculus: A Theory of Deterministic Queuing Systems for the Internet, 2050th ed.: Lecture Notes in Computer Science, 2012.

[14] Wasielewska K., Network Calculus in the practice - available bandwidth estimation, 3rd Workshop on Network Calculus (WoNeCa-3), Münster, Germany, 2016.

[15] A. Borys, M. Aleksiewicz, D. Rybarczyk, K. Wasielewska, Some principles of network calculus revisited, International Journal of Electronics and Telecommunications, vol. 57, no. 3, pp. 279-284, 2011.