Perspective

On the Average Rates of Data Communications' Traffic Flows Problem

Monday Ofori Eyinagho and Samuel Oluwole Falaki

Department of Electrical and Computer Engineering, Afe Babalola University, Ado-Ekiti, Nigeria

Abstract

Background and Objective: The guaranteed and controlled load services' provisioning of, the present Internet operates on the principle of admission control, in which a flow is set-up that conforms to IETF's (Internet Engineering Task Force's) T-SPEC or traffic specification. A parameter of this specification is the flow's sustainable rate. Similarly, the Committed Information Rate (CIR) is one of the parameters defined by Carrier Ethernet or Transport Ethernet services' providers in their Bandwidth Profiles, for billing for bandwidth usage and engineering network resources. Presently, no practically known method(s) or formula(s) for determining a flow's sustainable rate or CIR exists. This study describes the derivation of one such formula. **Methodology:** An empirical, practically utilizable formula for calculating the average rate of any data communications' traffic flow was derived and a numerical example to illustrate the application of the formula was given. **Results:** Numerically obtained values showed that, for the same maximum delay, the longer the distance between the source and destination of a flow, the higher the average rate that should be specified for the flow; the higher the maximum delay, the smaller the average rate that should be specified: Both are physically realistic situations. **Conclusion:** The formula is quite simple for practical applications. With it, Applications can communicate average rates' values that should be provisioned for, by nodes on the paths of the flows to the different destinations.

Key words: Bandwidth profile, committed information rate, guaranteed and controlled-load services, maximum end-to-end delay, sustainable rate of data traffic

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

Voice, video and an increasing variety of data sessions require upper bounds on delay and lower bounds on rate[1-4]. In the context of service provisioning, the Internet supports the guaranteed and controlled-load services[5]. While the guaranteed services class is designed for real-time traffics that need guaranteed maximum end-to-end delays, the controlled-load services class is designed for traffics that does not require this guarantee but are, nevertheless, sensitive to, overloaded networks and to the danger of losing packets. File transfer sessions are examples of traffics that do not need real guarantee, while real-time audio and video transfers are examples of traffics that need real guarantees. The two classes of services operate on the principle of 'admission control', in which a flow is set-up, which must conform to the IETF's (Internet Engineering Task Force's) T-SPEC. The admission control principle operates as follows[5-7]:

- In order to receive either type of service, a flow must first perform a reservation during a flow set-up
- A flow must conform to an arrival curve of the form $\sigma(t) = \min(M + pt, rt + b)$. In Intserv jargon, the 4-uple (p, M, r, b) is called a T-SPEC or traffic specification. In order words, the T-SPEC is declared during the reservation phase. Here, M = maximum packet size, p = peak rate, b = burst tolerance, r = sustainable rate
- All routers (and in most cases, layer 3 switches) along the path accept or reject the reservation. With the guaranteed service, routers accept the reservation only if they are able to provide a service curve guarantee and enough buffer for loss-free operation. The service curve is expressed during the reservation phase

Similarly, Carrier Ethernet service providers define a Bandwidth Profile, which allows the service providers to bill for bandwidth usage and engineer their networks' resources to provide performance assurances for in-profile Service Frames. The Committed Information Rate (CIR), which is defined as the average rate in bits/sec., that a network is committed to deliver frames over a predetermined time period, is one of the parameters of this profile. Therefore, a major issue that has to be dealt with by both services' providers and managers of organizations' Local and/or Campus Area Networks is, agreeing on a value for this parameter when a communication session is to be set up. The challenge therefore, is: How would a flow's r as stated

previously or what is termed ρ, the long-term average rate of traffic flow for a communication session[8-10] also known as the committed information rate, in the context of bandwidth profile for Carrier Ethernet be determined? We have not seen any empirical approach(s) or formula(s) for determining this parameter reported in literature; that is, there does not seem to be available, generally known method(s) or formula(s) for determining and assigning a value for this parameter. For example, the first step in the end-to-end delay algorithm enumerated by Georges *et al.*[11] states that: Identify all streams on each station and determine the initial leaky bucket values (ρ is one of the leaky bucket parameters). The paper did not however, give any empirical method for initially determining ρ. In this study, authors report the derivation of an empirical formula that can be used to determine this parameter.

METHODOLOGY

The method adopted in this study is neither analytical, nor experimental but emphasizes on deriving a practically, utilizable formula for computing the parameter ρ, the average rate of any data-communication's traffic flow; which is the same thing as CIR (Committed Information Rate). According to Cruz[10], if R(t) represents the instantaneous rate of traffic flow to a network element, then $\forall x,y$; $y \geq x$, $x \geq 0$, $\int_x^y R(t)dt$, a cumulative function, is defined as the number of bits seen on a flow in the interval [x, y]. Furthermore, Cruz[10] stated that:

$$\int_x^y R(t)dt \leq \sigma + \rho(y - x) \qquad (1)$$

where, Eq. 1 implies that there is an upper bound on the amount of traffic contained in any interval [x, y] that is equal to a constant σ plus a quantity that is proportional to the length of the interval. The constant of proportionality ρ determines an upper bound to the long-term average rate of traffic flow, if such an average rate exists[10]. Taking the upper bound in Eq. 1, we have:

$$\int_x^y R(t)dt = \sigma + \rho(y - x) \qquad (2)$$

Equation 3 was shown[12] to give the maximum delay of a Protocol Data Unit (PDU) through a node (switch or router):

$$D_{jmax}(\sec) = \frac{L}{C_{N-1}} + (N-1) \times \left(\frac{L}{2 \times \sum_{i=1}^{N} C_i} \right) + \frac{\sigma_j}{C_{out}} + \frac{L}{C_{out}} \qquad (3)$$

Where:

D_{jmax} = Maximum delay in sec for a PDU to cross node j, having, N input/output ports

$C_{i,i}$ = 1, 2, 3, $\cdots$, N = bit rates of ports 1, 2, 3,$\cdots$,N in bps = channel rates of input ports in bps

C_{out} = Bit rate of the Nth output link in bps = output port (line) rate of the Nth port

C_{N-1} = Bit rate of the $(N-1)^{th}$ input port in bps

L = Maximum length in bits of a PDU, for example, a packet, for packet switched networks, a cell for an ATM network

σ_j = Maximum amount of traffic in bits that can arrive in a burst to an input port of node j

This equation is composed of four components, which are $\frac{L}{C_{N-1}}$ = Maximum PDU Forwarding Delay, $(N \times 1) \times \left[\frac{L}{2} \times \sum_{i=1}^{N} C_i \right]$ = Maximum PDU Routing or Switching delay (RSD)+Simultaneous arrivals of PDUs delay (SAD), $\frac{\sigma_j}{C_{out}}$ = Maximum PDU Queuing Delay and L/C_{out} = Maximum PDU Transmission Delay. The first two delays, that is, PDU forwarding delay and PDU routing and switching delay, constitutes what is generally called processing delay; which is, the time required for nodal equipment to perform the necessary processing and switching[13,14] of a PDU.

Eyinagho[15] reduced Eq. 3 to 4:

$$D_{jmax} = \frac{\sigma_j}{C_{out}} + \frac{5L}{2C_{out}} \qquad (4)$$

where, Eq. 4 gives the maximum delay of a PDU across a node, j and $5L/2C_{out}$ the intercept on the delay axis (as shown in Fig. 1) gives the minimum delay (D_{min}) of a PDU across the node.

Assume that a PDU crosses M-nodes in transiting from the source of the PDU to the destination of the PDU. Let D_{1max}, D_{2max},..., D_{Mmax}, be the maximum delay of a PDU in node1, node2,... and node M, respectively as shown in Fig. 2 with:

$$D_{jmax} = \frac{\sigma_j}{C_{out}} + \frac{5L}{2C_{out}} \quad j = 1, 2,...,M$$

Then the total maximum end-to-end delay (D_{totmax}) is given by Eq. 5:

$$\begin{aligned} D_{totmax} &= D_{1max} + D_{2max} + ... + D_{Mmax} \\ &= \frac{\sigma_1}{C_{out}} + \frac{5L}{2C_{out}} + \frac{\sigma_2}{C_{out}} + \frac{5L}{2C_{out}} + \cdots + \frac{\sigma_M}{C_{out}} + \frac{5L}{2C_{out}} \\ &= \frac{\sigma_1 + \sigma_2 + ... + \sigma_M}{C_{out}} + \frac{5ML}{2C_{out}} \end{aligned} \qquad (5)$$

$$D_{totmax} = \frac{\sigma_{tot}}{C_{out}} + \frac{5ML}{2C_{out}} \qquad (6)$$

where, $\sigma_1 + \sigma_2 + ... \sigma_M = \sigma_{tot}$ = the total of the maximum burst traffic that can be held by the nodes on the origin to destination path.

But Cruz[10] enunciated and proved that, if D_{max} = maximum delay offered by a network element to a traffic stream, R_{in}= rate of the traffic stream as it enters the network
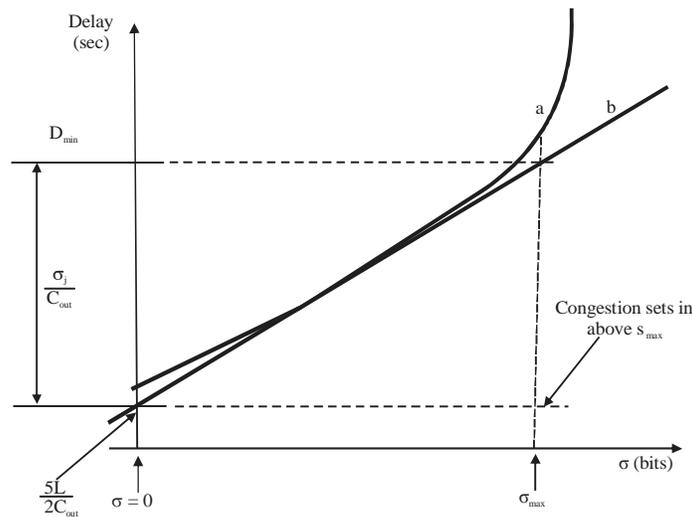

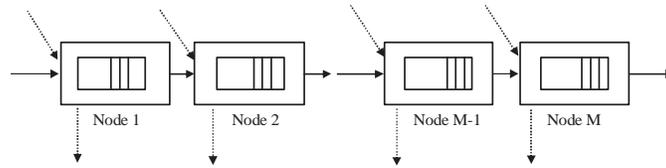Fig. 1: Delay versus burst traffic arrivals to a node (switch or router)

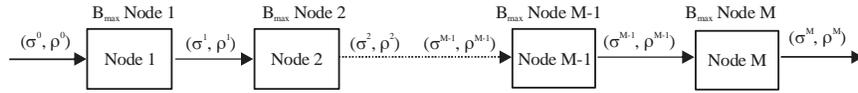Fig. 2: Origin-destination of a PDU traversing M-nodes

Fig. 3: Burstiness evolution along a switched network

element, R_{out} = rate of the traffic stream as it exits the network element and $R_{in} \sim b_{in}$ with $D_{max} < +\infty$, then $R_{out} \sim b_{out}$; where $b_{out}(x)$ is defined for all $x \geq 0$ as:

$$b_{out}(x) = b_{in}(x+D_{max}) \qquad (7)$$

As an example of Eq. 7, if:

$$R_{in} \sim (\sigma, \rho), \text{ then } R_{out} \sim (\sigma + \rho D_{max}, \rho) \qquad (8)$$

This is known as the burstiness evolution concept. The intuition that motivates this concept is that, if D_{max} is small, the output stream will closely resemble the input stream and cannot be much, more bursty than the input stream. According to this concept, also expounded by Georges *et al.*[11]; if the traffic entering a network is not too bursty, the traffic flowing in the network will also not be too bursty. This is illustrated in Fig. 3. Given, therefore, a method for calculating the upper bounds for delay in various types of network elements, the importance of Eq. 7 is that, it gives a basis for analyzing the network as a whole. Cruz[10] stated further that Eq. 7 can be improved upon if the structure of the network element is known. This has already been done for switching nodes (switches and routers) as indicated by Eq. 4.

Applying (8) to Fig. 3 therefore, gives the following burstiness evolution equation:

$$(\sigma^1, \rho^1) = (\sigma^0 + \rho^0 D_{max} Node1, \rho^0)$$

$$(\sigma^2, \rho^2) = (\sigma^1 + \rho^1 D_{max} Node2, \rho^1) \qquad (9)$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$(\sigma^M, \rho^M) = (\sigma^{M-1} + \rho^{M-1} D_{max} Node\ M, \rho^{M-1})$$

From the set of Eq. 9, it is seen that, $\rho^1 = \rho^2 = \rho^3 = \cdots = \rho^{M-1} = \rho^M = \rho$ = minimum rate required to serve the flow according to T-SPEC.

It is therefore, clear that (with or without traffic shaping), σ going into node 1 would have evolved to σ_{total} (σ_{tot}) at node

M. Hence, if R_{in} = the rate of traffic stream into node 1 and R_{out} = rate of traffic stream out of node M (it should be noted that we are interested in a single traffic stream traversing from node 1 to node M, although there may be other traffic streams passing through these nodes) as shown in Fig. 2. If, t = total of the maximum delays of all the nodes from the source of the traffic stream to the destination of the stream = D_{totmax} and C_{out} = port rate of the output port through which the traffic exits the node M to the destination host, then from Eq. 2:

$$\int R(t)dt = \sigma_{tot} + \rho t = \sigma_{tot} + \rho D_{totmax} \qquad (10)$$

where, the interval of integration on the left of Eq. 10 is the maximum end-to-end delay time from the source of the traffic stream to its destination.

Dividing Eq. 10 by C_{out}, results in Eq. 11:

$$\frac{1}{C_{out}} \int R(t)dt = \frac{\sigma_{total}}{C_{out}} + \frac{\rho D_{tot\,max}}{C_{out}} \qquad (11)$$

It can be seen that the left hand side of Eq. 11 gives the maximum end-to-end delay that a PDU will encounter from the source of the traffic stream to its destination; because, the integral gives the maximum number of bits that has traversed from the origin node to the destination node during this maximum end-to-end delay time and dividing this maximum number of bits by the output port rate of node M, gives the maximum end-to-end delay time.

Therefore, the left hand sides of Eq. 6 and 11 are the same quantity. Hence, comparing Eq. 6 with 11, we have:

$$\frac{5ML}{2} = \rho D_{totmax} \Rightarrow$$

$$\rho = \frac{5ML}{2D_{tot\,max}} \qquad (12)$$

$D_{totmaxa}$ = one-way maximum end-to-end delay of an application; for example, VOIP (Voice over Internet Protocol) application.

Equation 12 therefore, is the minimum rate (in bits/sec) required to serve the flow according to T-SPEC or the CIR in the context of bandwidth profile specification of Carrier Networks (for example, Carrier Ethernet Networks). Therefore, knowing the number of nodes M, that the traffic stream of an application will traverse from the origin of the stream to its destination; L, the maximum length of a PDU and D_{totmax}, the one-way maximum end-to-end delay specification of an application, ρ can be intelligently determined and communicated by the source of the flow to all routers (and/or layer 3 switches) along the route.

RESULTS

Illustrative numerical example: The basic objective of this study is to obtain the result as given by Eq. 12. The application of this equation is now illustrated by extracting the network delay-based QoS parameters for two types of applications from ITU-Ts Standard Y.1541. They are class 0 and class 1 real- time, highly interactive applications' like VOIP and Video Conferencing. The class 0 one-way network delay performance objective is 100 ms and that of class 1 is 400 ms. Taking the extended Ethernet frame as illustrated in Fig. 4 as our PDU, L = 7+1+6+6+4+2+1500+4 = 1530 bytes = 1530×8 bits = 12240 bits.

Tabulated ρ values for origin-destination paths, with 1, 2, 3, 4 and 5 nodes (routers or layer 3 switches), denoted as ρ_1, ρ_2, ρ_3, ρ_4 and ρ_5, respectively are as shown in Table 1, which is briefly discussed in the next sub-section.

DISCUSSION

As shown in Table 1, it can be seen that, for the same maximum delay bound, the longer the distance between the source and destination of a communication flow (that is, the more the number of nodes on the origin-destination path), the higher the minimum bits rate, that should be specified for the flow. The higher the maximum delay bound, the smaller the minimum bits rate that should be specified. It has been canversed[16] that, CIR (Committed Information Rate), which is the same as, average rate of data communication traffics flows, should be carefully engineered for a network to efficiently handle it. No suggestion was given as to how it should be engineered as it does seems that, practically

Table 1: ITU-T Y.1541 delay-bound QoS values/equivalent origin-destination p-values

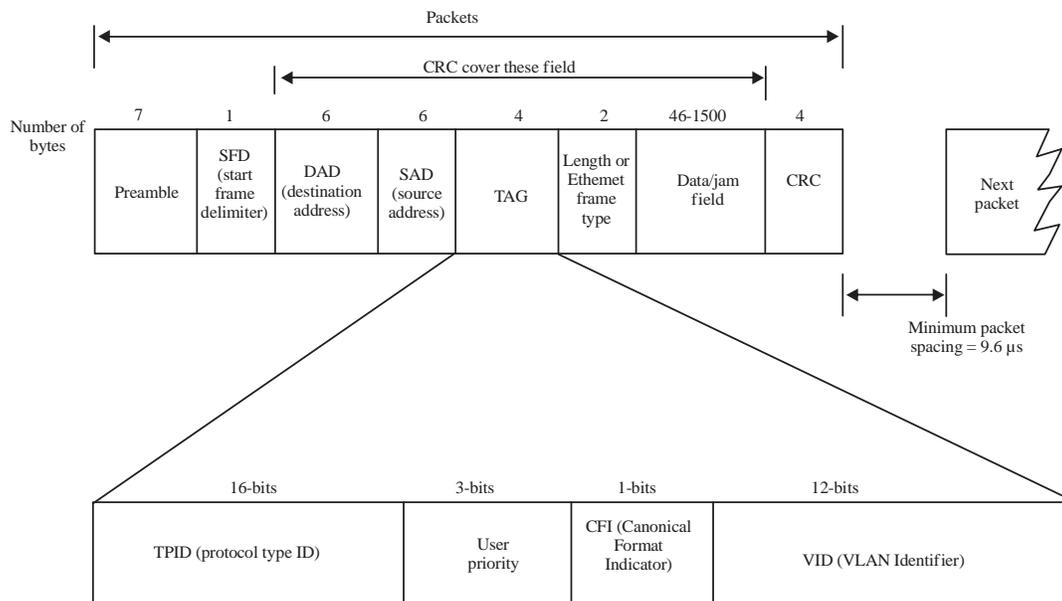| ITU-T Y.1541 delay bound	ρ_1 (Mbps)	ρ_2 (Mbps)	ρ_3 (Mbps)	ρ_4 (Mbps)	ρ_5 (Mbps)
100 ms	0.306	0.612	0.918	1.224	1.530
400 ms	0.077	0.153	0.231	0.308	0.385

Fig. 4: Extended Ethernet frame format with tag field (IEEE 802.1Q field)

utilizable formula for engineering this parameter is not available in literature. It is in this context that this study is of prime importance.

CONCLUSION

A major issue that has to be dealt with by both services' providers and managers of organizations' Local and/or Campus Area Networks, is agreeing on a value for the minimum data rate required to serve a connection or flow, also referred to as, 'Committed Information Rate' (CIR), in the context of Bandwidth profiles of Carrier Networks or, the long-term average rate of data communication traffic flow, ρ. Generally known method(s) or formula(s) for determining and assigning a value for this parameter are not however, available in literature. So it seems, in our search so far, in the literature of computer and data communication and networking. The work as reported in this study, is an attempt to proffer a formula for computing this parameter, which to a good extent, has been achieved.

SIGNIFICANCE STATEMENT

This study has shown that, the average rates of data communications' traffic flows can be practically computed by communicating applications in real-time. The developed formula that is reported in this paper will be important for solving one of the performance problems of computer and data communication networks. The fact that the formula is practically utilizable is very significant as no such formula(s), to the best of our knowledge, previously exists. The significance of the formula can be appreciated when we consider the fact that, understanding network performance has been more of an art than a science, that, there has been very little underlying theory which are actually of any use in practice and the best that has been done were to apply rules of thumb, gained from hard experience, using examples taken from the real world. Therefore; the approach adopted in this study sought to promote the merger of some level of mathematical elegance, with practically, important techniques. Authors developed formula for carrying out the engineering of the parameter ρ or CIR will serve as a basis and impetus for further researches on this problem, with a view of probably, coming up with a better formula or formulas.

REFERENCES

1.   Jonsson, M., K. Kunert and A. Kallerdahl, 2013. Analysing AFDX networks using end-to-end response time analysis. J. Inter. Net., Vol. 14. 10.1142/S0219265913500175.

2.   Mehta, V. and N. Gupta, 2012. Performance analysis of QoS parameters for wimax networks. Int. J. Eng. Innovative Technol., 1: 105-110.

3.   Bolot, J.C., 1993. Characterizing end-to-end packet delay and loss in the internet. J. High Speed Networks, 2: 305-323.

4.   Bertsekas, D. and R. Gallager, 1992. Data Networks. 2nd Edn., Prentice Hall, Englewood Cliffs, New Jersey, pp: 510.

5.   Le Boudec, J.Y. and P. Thiran, 2004. Network Calculus-A Theory of Deterministic Queuing Systems for the Internet. Springer-Verlag, Berlin, pp: 75.

6.   Stoica, I. and H. Zhang, 1999. Providing guaranteed services without per flow management. ACM Comput. Commun. Rev., 29: 81-94.

7.   Xiao, X. and L.M. Ni, 1999. Internet QoS: A big picture. IEEE Network, 13: 8-18.

8.   Ashmawi, W., R. Guerin, S. Wolf and M. Pinson, 2001. On the impact of policing and rate guarantees in diff-serv networks: A video streaming application perspective. Proceedings of the Conference on Applications, Technologies, Architectures and Protocols for Computer Communications, August 27-31, 2001, ACM., New York, USA., pp: 83-95.

9.   Cruz, R.L., 1995. Quality of service guarantees in virtual circuit switched networks. IEEE J. Sel. Areas Commun., 13: 1048-1056.

10. Cruz, R.L., 1991. A calculus for network delay. I. Network elements in isolation. IEEE Trans. Inform. Theory, 37: 114-131.

11. Georges, J.P., T. Divoux and E. Rondeau, 2005. Confronting the performances of a switched Ethernet network with industrial constraints by using the network calculus. Int. J. Commun. Syst., 18: 877-903.

12. Eyinagho, M., S. Falaki and A. Atayero, 2012. Determination of End-To-End Delays of Switched Local Area Networks. LAP LAMBERT Academic Publishing, Saarbrucken, Germany, pp: 93.

13. Gao, J., Y. Shen, X. Jiang and J. Li, 2015. Source delay in mobile ad hoc networks. Ad Hoc Networks, 24: 109-120.

14. Comer, D., 2004. Computer Networks and Intranets with Internet Applications. Pearson Prentice Hall, New Jersey, USA., pp: 224.

15. Eyinagho, M., 2016. On the minimum and maximum packet's delays determination for delay jitter computation. J. Theor. Applied Inform. Technol.

16. Metro Ethernet Forum, 2010. Understanding carrier ethernet throughput. Version 2, July 2010. https://www.mef.net/Assets/White_Papers/Understanding_Carrier_Ethernet_Throughput_-_v14.pdf