British Journal of Mathematics & Computer Science
4(24): 3476-3488, 2014
ISSN: 2231-0851

SCIENCEDOMAIN *international*
www.sciencedomain.org

Analytical Derivation of Latency in Computer Networks

Adelakun A. Adebiyi[1*], Samuel N. John[1] and Charles Ndujuiba[1]

[1]Department of Electrical and Information Engineering, Covenant University, Ota, Ogun State, Nigeria.

Article Information

DOI: 10.9734/BJMCS/2014/10770
Editor(s):
(1) Qiang Duan, Information Sciences & Technology Department, The Pennsylvania State University Abington, USA.
(2) Victor Carvalho, Polytechnic Institute of Cávado and Ave (IPCA), Barcelos, Portugal, Catholic University of Portugal (UCP), Braga, Portugal and Lusiada University (ULP), Vila Nova de Famalicão, Portugal.
(3) Tian-Xiao He, Department of Mathematics and Computer Science, Illinois Wesleyan University, USA.
Reviewers:
(1) Anonymous, College of Engineering and Technology, Chandigarh, India.
(2) Anonymous, Federal University of Technology, Nigeria.
(3) Anonymous, Payame Noor University, Iran.
(4) Ecir Uğur Küçüksille, Computer Engineering, Süleyman Demirel University, Turkey.
Complete Peer review History: http://www.sciencedomain.org/review-history.php?iid=699&id=6&aid=6364

Received: 08 April 2014
Accepted: 25 July 2014
Published: 06 October 2014

Original Research Article

Abstract

This paper presents mathematical method of estimating the latency of a corporate computer network through broad classification into propagation, serialization and queue delays. The parameters of interest considered are the sending rates, arrival times, connection bandwidth and the speed of travelling in the medium. Simple and easily determined factors that are essential for computer networks' quality of service QoS were used to derive the expressions for computing latencies. The expressions were tested with randomly varying packet sizes, variable mean service rates of nodal devices and varying packet arrival rates.

Keywords: Congestion, collision, latency, computer network

1 Introduction

The ever increasing patronage and usage of the internet as the major tool communication vis-à-vis the convergence with other communication systems has called for a dedicated and controlled monitoring of its performances both from the users and the operators. A prominent parameter of measurement is the Quality of service QoS, which specifies the internet's performance in terms of its latency and throughput. The bandwidth of the user is another factor of interest. Even though there is a wide distinction between the available bandwidth and the capacity bandwidth [1], the

*Corresponding author: adisa.adelakun@covenantuniversity.edu.ng;

two impact the quality of the channel's performances with respect to the volume of data transfer per unit time. The shared nature (CSMA, CSMA-CD) of the internet paths creates a survival of the fittest for users' data traversing the network path and also the ease of accessing distant servers. Efforts have therefore been made to increase the bandwidth as a measure of improvement of the QoS on the network rather than thorough scrutiny of the "clogs" on the network paths. These clogs are created by legitimate users, network nodal devices (with differs processing capacities, speeds and even buffers) and attackers. An idealized network connection is supposed to offer a seamless traffic flow between the sender and the destination where the communication delay approaches zero [2]. The sharp deviation from the ideal spurs researches into the causes and effects of network impediments. The protocols of internet have built-in tools that navigate the network paths and report the status of the network [3,4]. One of such tools is the Simple Network Management Protocol (SNMP) which can give information about the traffic level of the network path. Internet Control Message Protocol (ICMP) and the SNMP have been widely employed in tools for the determination of end to end information of the network path [2,3,4,5]. Typical tools from these are the ping and trace route, Network latency view etc., whose typical screenshot is displayed in Fig. 1. These tools offer a quick measurement of the internet network performance by sending probing packets into the network and initiating timing, recording the sent packets, received packets and dropped packets [6,7,8].

Congestion and collision represent the prominent banes of the computer network; they are responsible for the elapses time from when a packet is thrown into the network and when it is received at the destination in a fully functional computer network. They are inevitable especially in the bottleneck links where the processing capacity and the buffer storage capacity of the nodal device are over-stressed beyond the designed capacity. Researches are therefore directed toward alleviating the impact of these challenges in order to obtain low latency and high throughput [8,9]. The instantaneous status/condition of the shared network path is essential for any control measure's efficacy. Hence, many of the deployed tools utilized the condition on the nodal devices to effectively determine the performance measures.

This paper presents a quick method of computing internet performance measure-latency; using the knowledge of the network path parameters namely number of packets in transit, delays at the nodal devices majorly owing to queuing, and propagation delay among others. This will in no small measure assist network designers in having a foresight of the network performance even while designing.

Van Jacobson [10] pioneered the work in what is now called TCP/IP congestion control. Congestion control algorithm detects and control congestion at every stage of packet communication and relay the status of the path through ACK (acknowledgement message)to the sender.

Queue management brought to the fore the control of packets congestion at the router by setting a maximum packet length. If this window size is exceeded, the router begins to drop the subsequent packet to return to the set window capacity. This method is known as the drop-tail.

Floyd and Jacobson [11] improved on the drop-tail technique with the design of Random Exponential Detection (RED). Its objectives were to minimize packet loss, queuing delay, and maintain global synchronization of packet sources as well as maintain high link utilization. But unfortunately, because of the large buffer space requires by RED to function effectively, the buffer adds considerable to end-to-end delay and jitter.

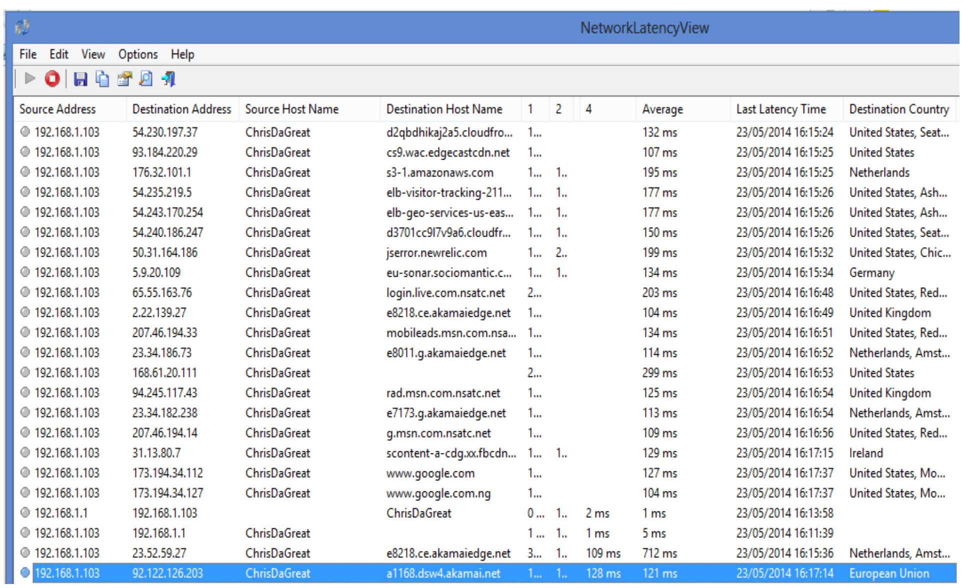
Random Exponential Marking (REM) [12] aims to achieve both high utilization and negligible loss and delay in a simple and scalable manner. The key idea is to decouple congestion measure from performance measure such as loss, queue length or delay. While congestion measure indicates excess demand for bandwidth and must track the number of users, performance measure should be stabilized around their targets independently of the number of users.

The BLUE algorithm [13] resolves some issues of RED. It uses flow and queue events to modify congestion notification rate. It maintains a single probability to mark or drop packets. If the queue is sufficiently large due to buffer overflow, it increases the probability, thus, increasing the rate at which it sends back congestion notification to the source. If otherwise, the probability is decreased. This makes BLUE to be intelligent in its operation.

Another version of RED proposed [14] is called Hazard rate packet dropping function in RED, "HERED". It reduces the packet dropping rate of the traditional RED at light traffic *load* while the dropping rate ***becomes*** more aggressive at heavy load. These and many others are classified as Active Queue Management have been presented in research works with proven potency to mitigate congestion, reduce queues and hence reduces packet delays on the nodal devices.

Many These tools have been adequately applied in various works to improve the performance measures of the network.

This paper is organized as follows. Section II gives detail description of bottlenecks and their locations on a corporate computer network. It also explains the mathematical modeling of delays. Section IV illustrated our derived model for latency and throughput on a typical network path. Section V presents the derived latency and throughput for a multi-host, multi- node network. Section VI presents the conclusion.



Source Address	Destination Address	Source Host Name	Destination Host Name	1	2	4	Average	Last Latency Time	Destination Country
192.168.1.103	54.230.197.37	ChrisDaGreat	d2qbdhikaj2a5.cloudfro...	1...			132 ms	23/05/2014 16:15:24	United States, Seat...
192.168.1.103	93.184.220.29	ChrisDaGreat	cs9.wac.edgecastcdn.net	1...			107 ms	23/05/2014 16:15:25	United States
192.168.1.103	176.32.101.1	ChrisDaGreat	s3-1.amazonaws.com	1...	1..		195 ms	23/05/2014 16:15:25	Netherlands
192.168.1.103	54.235.219.5	ChrisDaGreat	elb-visitor-tracking-211...	1...	1..		177 ms	23/05/2014 16:15:26	United States, Ash...
192.168.1.103	54.243.170.254	ChrisDaGreat	elb-geo-services-us-eas...	1...	1..		177 ms	23/05/2014 16:15:26	United States, Ash...
192.168.1.103	54.240.186.247	ChrisDaGreat	d3701cc9l7v9a6.cloudfr...	1...	1..		150 ms	23/05/2014 16:15:26	United States, Seat...
192.168.1.103	50.31.164.186	ChrisDaGreat	jserror.newrelic.com	1...	2..		199 ms	23/05/2014 16:15:32	United States, Chic...
192.168.1.103	5.9.20.109	ChrisDaGreat	eu-sonar.sociomantic.c...	1...	1..		134 ms	23/05/2014 16:15:34	Germany
192.168.1.103	65.55.163.76	ChrisDaGreat	login.live.com.nsatc.net	2...			203 ms	23/05/2014 16:16:48	United States, Red...
192.168.1.103	2.22.139.27	ChrisDaGreat	e8218.ce.akamaiedge.net	1...			104 ms	23/05/2014 16:16:49	United Kingdom
192.168.1.103	207.46.194.33	ChrisDaGreat	mobileads.msn.com.nsa...	1...			134 ms	23/05/2014 16:16:51	United States, Red...
192.168.1.103	23.34.186.73	ChrisDaGreat	e8011.g.akamaiedge.net	1...			114 ms	23/05/2014 16:16:52	Netherlands, Amst...
192.168.1.103	168.61.20.111	ChrisDaGreat		2...			299 ms	23/05/2014 16:16:53	United States
192.168.1.103	94.245.117.43	ChrisDaGreat	rad.msn.com.nsatc.net	1...			125 ms	23/05/2014 16:16:54	United Kingdom
192.168.1.103	23.34.182.238	ChrisDaGreat	e7173.g.akamaiedge.net	1...			113 ms	23/05/2014 16:16:54	Netherlands, Amst...
192.168.1.103	207.46.194.14	ChrisDaGreat	g.msn.com.nsatc.net	1...			109 ms	23/05/2014 16:16:56	United States, Red...
192.168.1.103	31.13.80.7	ChrisDaGreat	scontent-a-cdg.xx.fbcdn...	1...	1..		129 ms	23/05/2014 16:17:15	Ireland
192.168.1.103	173.194.34.112	ChrisDaGreat	www.google.com	1...			127 ms	23/05/2014 16:17:37	United States, Mo...
192.168.1.103	173.194.34.127	ChrisDaGreat	www.google.com.ng	1...			104 ms	23/05/2014 16:17:37	United States, Mo...
192.168.1.1	192.168.1.103		ChrisDaGreat	0 ...	1..	2 ms	1 ms	23/05/2014 16:13:58	
192.168.1.103	192.168.1.1	ChrisDaGreat		1 ...	1..	1 ms	5 ms	23/05/2014 16:11:39	
192.168.1.103	23.52.59.27	ChrisDaGreat	e8218.ce.akamaiedge.net	3...	1..	109 ms	712 ms	23/05/2014 16:15:36	Netherlands, Amst...
192.168.1.103	92.122.126.203	ChrisDaGreat	a1168.dsw4.akamai.net	1...	1..	128 ms	121 ms	23/05/2014 16:17:14	European Union

Fig. 1. Screenshot of Captured data of Host-to-clients latencies [3]

2 Congestion and Collision in Computer Junctions

Packet transmission in a LAN follows protocols that encourage fairness and congestion and collision avoidance of the packets from the senders. The supposed equitable access to the shared medium is often violated thereby causing absymal performance of the network.Many factors have been advanced as degrading the performance of bandwidth over the years. Such factors include and not limited to; *(i)* inadequate traffic management *(ii)* poor caching *(iii)* poor compression [15]. Packet loss occur majorly due to congestion , a few percentage of the losses due to damage (<<1%). A well-managed traffic on a computer network will substantially reduce the risk of information collision and hence, the need for data retransmission. Caching and compression equally reduce the need for fetching the data from the source server every time it is needed thus minimizing congestion at the gateways as well as reducing collision thereby reducing the latency. Fig. 2 highlights the locations of the bottlenecks on a typical corporate computer network.

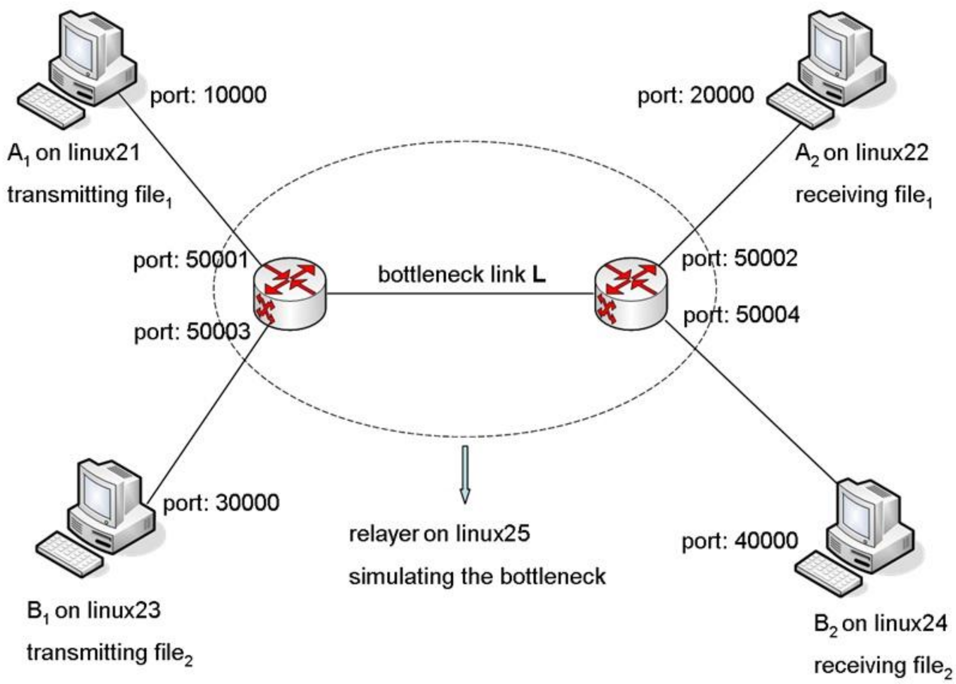


Fig. 2. Bottleneck path in computer network

3 Bit/Packet Arrival Probabilty

We will employ basic queue theory to determine the probability of packet arrival at a port of a multi-ports gateway equipment. The port was further subjected to varying load capacities such that its capacity bandwidth is exceeded. The bit arrival at such port has probability expression [16] shown in equation (1);

$$P_j(t) = \frac{(\lambda t)^j}{j!} e^{-\lambda t} \tag{1}$$

Where $P_j(t)$ represents the probability of bit arrival at time t in seconds, λis the packet arrival rate at port$j$. The packet arrival rate is determined by the total number of packet arriving per specified unit of time.

$$\lambda = \frac{\alpha(t)}{t} \tag{2}$$

α is the total number of packets.

The gateway equipment's buffer (switch) has a function to reduce the dropped packets by placing them (packets) in queue while the arrival rate of the packets exceeds the processing capacity of the equipment. In several cases, the queue length is given by the product of average delay and the average arrival time in a ***bufferless*** situation. Otherwise, the buffer capacity has to be considered before any queue could build-up. The average delay for a waiting time of $\gamma$is thus computed as,

$$T = \frac{\gamma(t)}{\alpha(t)} \tag{3}$$

This gives the Average Queue length N of equation (4)

$$N = \frac{\gamma(t)}{T} \tag{4}$$

4 Derived Model for Latency on a Network Path

Delays along computer network path were classified into three broad sub-groups namely; propagation delays, serialization/transmission delay and queue delay as are shown in Fig. 3 [17]. These delays affect the transit time of packets from the sending end to the destination. Nodal delay resulting from check bit errors and determination of the output link is considered negligible.

The propagation delay, if the speed of travelling of the packet in any medium is taken to be 2/3 of the speed of light in air, is given as

$$P_D = \frac{x}{0.67c}(Seconds) \tag{6}$$

The serialization delay when N number of packets succeeded in traversing a network path of bandwidth A_i is

$$S_D = \frac{N_i}{A_i} \tag{7}$$

The available bandwidth is a function of the utilization factor of the capacity bandwidth i.e. not all the available bandwidth is fully utilized per time.

Available bandwidth becomes:

$$A_i = (1-\rho)C_i$$

Where ρ is the average utilization factor and have value $< \rho \leq 1$. A value of one signifies a congested network which is a critical state. But

$$\rho = \frac{\lambda}{\mu}$$

i.e. the ratio of average arrival rate λ to the mean processing time μ.

Hence, the serialization delay of any packet in a network path is

$$S_D = \frac{N_i * \mu}{(\mu - \lambda)C_i} \tag{8}$$

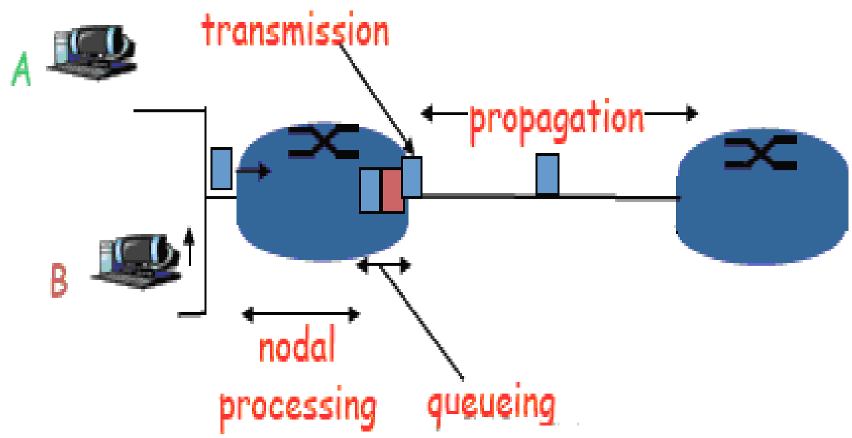


Fig. 3. Locations of the generalized delays on the network

Equation (9) is the queue delay in a single- source single –node network path. In deriving the queue delay for a multi-nodes network path, packets from all the sources must be considered to have an accurate estimation of the queue length at the node. Therefore, using Little's formula,

$$T = \frac{1}{\gamma_E} \sum_{i=1}^{L} \lambda_i T_i \tag{9}$$

$$\gamma_E = \sum_{i=1}^{M} \sum_{j=1}^{M} \gamma_{i \to j} \tag{10}$$

Where

γ_E = total workload in packets per second

$\gamma_{i \to j}$ = workload between source i and destination j,

M = total number of sources and destinations

L = total number of links and

T is the total waiting time.

The queue delay is therefore given by

$$Q_D = \frac{1}{\gamma_E} \sum_{1}^{L} \left\{ \frac{1}{\frac{R_B}{k} - \lambda} \right\} = \frac{1}{\gamma_E} \sum_{1}^{L} \left\{ \frac{k}{R_B - \lambda * k} \right\} \tag{11}$$

Where k the average packet size and R_B is the medium bandwidth given by the function in equation (12)

$$R_B = \begin{cases} A_i & \rho << 1 \\ C_i & \rho \geq 1 \end{cases} \tag{12}$$

The estimate of the total delays is thus the summation of the propagation, serialization and queue delays on the network path.

$$P_D + S_D + Q_D = \frac{x}{0.67c} + \frac{N * \mu}{(\mu - \lambda)C_i} + \frac{1}{\gamma_E} \sum_{1}^{L} \left\{ \frac{k}{R_B - \lambda * k} \right\} \tag{13}$$

Packet transmission in a LAN follows protocols that encourage fairness and congestion and collision avoidance of the packets from the senders. Each source therefore senses shared physical medium before transmitting into the medium. The overall packet in transit is dependent on the protocol of the shared medium. For instance, the traditional Ethernet medium capacity is 10Mbps, fast Ethernet is 100Mbps. The CSMA/CD protocol of Ethernet ensures that the packets in transit do not exceed the maximum bandwidth capacity. This requires that all the transmitting hosts' must co-ordinate their sending rate in such a way that the average packet arrival rate at the switch does not exceed the average service rate/processing time of the switch. In a typical computer network, a host sends a packet to another host which may reside either on the same network or other network. Fig. 4 indicates hosts resident on the same network topology. The figure shows packet exchange between hosts ith and nth with the following parameters.

Distance of each host from the sink x

Inter-arrival rate at the node/ sending rate of each host λ

Mean service rate of each of the nodal devices μ

Number of packets sent is host dependent N

Packet sizes of each of the hosts$k$.

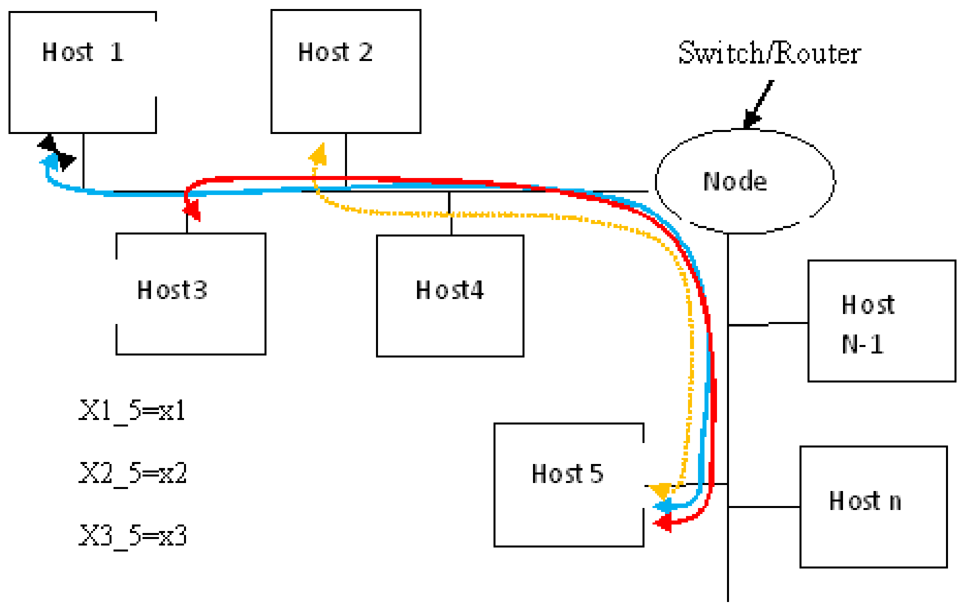


Fig. 4. LAN segment for analysis of Multi-host, multi-hop

The latency from a host ith to nth on the network;

$$Latency_{[i\to n]} = \sum [P_D, Q_D, S_D]$$

$$Latency_{[i\to j]} = \frac{x_{[i\to j]}}{0.67c} + \frac{N_{[i\to j]} \times \mu}{(\mu - \lambda)C_i} + \frac{1}{\gamma_E}\sum_{1}^{L}\left\{\frac{k}{R_B - \lambda * k}\right\} \tag{14}$$

The following are the assumption for equation 13:

1. Processing time of nodal devices is uniform i.e. $\mu =$ uniform/constant
2. Packet arrival rate at the nodal devices is uniform/constant
3. Nodal devices are sufficiently buffered such that dropped packets are negligible.

For insufficient buffer space, the packet through the nodal device is reduced to $(N - \delta)$ where δ defined the number of dropped packets across the nodal device. Under such condition, equation 13 becomes

$$Latency_{[i\to j]} = \frac{x_{[i\to j]}}{0.67c} + \frac{(N - \delta)_{[i\to j]} \times \mu}{(\mu - \lambda)C_i} + \frac{1}{\gamma_E}\sum_{1}^{L}\left\{\frac{k}{R_B - \lambda * k}\right\} \tag{15}$$

Packet routed through a less congested node because of congestion of the nearest node is accounted for by estimating the fraction of the packet $(k_m : m = 1,2,3,...)$ through the individual nodes, the distances $(x : x_\tau = path\_taken)$ to the destination and if the nodes' processing time $(\mu_n : n = 1,2,3,...)$ vary, the equation becomes

$$Latency_{[i \to j]} = \frac{x_{\tau[i \to j]}}{0.67c} + \frac{(N)_{[i \to j]} \times \mu_n}{(\mu_n - \lambda)C_i} + \frac{1}{\gamma_E} \sum_{1}^{L} \left\{ \frac{k_m}{R_B - \lambda * k_m} \right\} \tag{16}$$

5 Results and Discussion

The derived expressions were evaluated for various test scenarios. In the first case, a constant mean service rate μ was implemented in all the nodal devices and the various latencies of the paths were recorded as shown in Table 1.

The latency caused by propagation was linearly related to the distance while that of the serialization was majorly dependent on the volume of packet in transit and the workload of the nodal device in question. The results for randomly generated packet sizes were presented in Figs. 5 and 6 with variable mean service rate and fixed packet arrival rate at the nodal devices.

Table 1. Latencies at constant mean service rate,1Gbpsbandwidth

Distance (metres)	Number of packets(N)	Propagation Delay (µs)	Serialization Delay (µs)	Queue delay (Qd)
100	1000000	49.8	6.8	7.5E-10
150	3000000	74.6	12.4	1.43E-10
200	25000000	99.5	46.7	2.94E-12
500	11000000	24.9	55.6	1.61E-11
1000	15000000	49.8	71.4	6.98E-12
2000	102000000	99.5	76.9	1.23E-12
4500	27000000	22.4	25.4	1.27E-11
6000	31000000	29.9	368	8.24E-12
8000	1000000	39.8	173	1.5E-09
15000	205000000	74.6	76.4	2.39E-12
25000	31000000	124.0	728	5.42E-12
50000	45000000	249.0	2900	1.5E-12

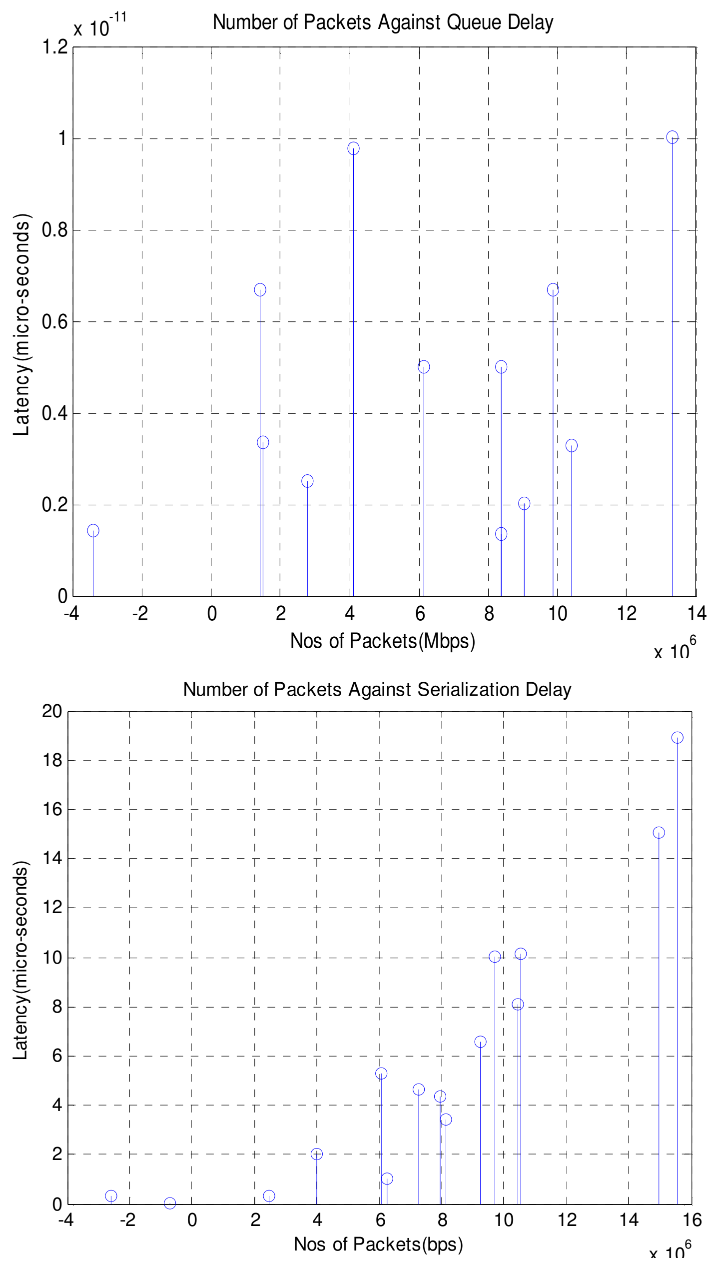


Fig. 5. Variable packets sizes plotted against queue delay and serialization delay

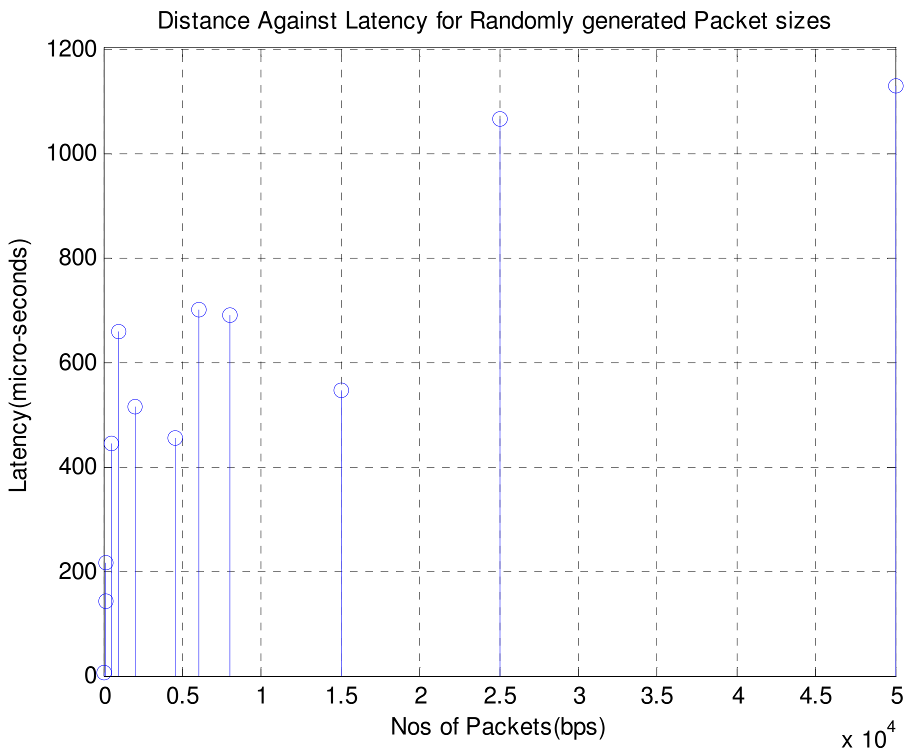


Fig. 6. Stem plot of the total network latency against variable packet sizes

6 Conclusions

In this paper, the mathematical method of estimating the latency of a corporate computer network through the broad classification of the overall network delays into propagation, serialization and queue delays was derived and presented. Expression for computing latencies of networks using only the sending rates, arrival times, connection bandwidth and the speed of travelling of the packets (assumed to be that of light) in the medium was arrived at: these pieces of information are readily obtained by network planner/engineers during network design. The available tools results can only be compared with the derived result if the distance of the path taken by the packets is the same for both (which practically is impossible since there are rules/principles guiding packet transit on network path especially when there is congestion and hence delay).

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Ravi Prasad, et al. Bandwidth Estimation: Metric, Measurement Techniques and Tools. IEEE Network, Nov/Dec 2003.
Available: www.cc.gatech.edu/home/devrolis/paper/bwest_survey.pdf

[2] Geoff Huston. Measuring IP Network performance. The Internet Protocol Journal. 2003;6(1). Accessed 10 May 2014. Available: www.cisco.com/ipj.

[3] Network Latency View v1.21. Accessed 18 May 2014.
Available: http://www.nirsoft.net/utils/network_latency_view.html

[4] Francis P, Jamin S, et al. IDMaps: A Global internet host distance estimation service. Accessed on 19 May 2014. Available: http://irl.cs.ucla.edu/papers/ton01-idmaps.pdf

[5] Brian Eriksson, et al. Estimating hop distance between arbitrary host pairs. Available: nowak.ece.wisc.edu/infocom09.pdf

[6] Michal Szymaniak,DavidPresotto, "Practical large-scale latency estimation" The International Journal of Computer and Telecommunications Networking. 2008;52(7).

[7] Tools for probing/monitoring network path Accessed 18 May 2014. Available: http://www.slac.stanford.edu/xorg/nmtf/nmtf-tools.html

[8] Ling-Jyh Chen, et al. AdHoc Probe: Path Capacity Probing in Wireless Ad Hoc Networks. Accessed 18 May 2014. Available:
http://www.iis.sinica.edu.tw/~cclljj/publication/2005/05_WICON_AdHoc_Probe.pdf

[9] John SN, Okonigene RE, Adelakun A. Impact of Latency on Throughput of a Corporate Computer Network, Proceedings: The 2010 World Congress in Computer Science, Computer Engineering, and Applied Computing (WORLDCOMP'10). Annual Summer Conference on Modeling Simulation & Visualization Methods (MSV'10), Las Vegas, Nevada, USA. 2010;282-287.

[10] Jacobson V. Congestion Avoidance and Control, ACM Computer Communication Review; Proceedings of the Sigcomm '88 Symposium in Stanford, CA. 1988;18(4):314-329.

[11] Floyd S, Jacobson V. Random early detection gateways for congestion avoidance. ACM/IEEE Transactions on Networking. 1993;1(4):397-413.

[12] Athuraliya S, Lapsley DE, Low SH. Random exponential marking for internet congestion control. IEEE/ACM Transactions on Networking. 2005;15(3):48-53.

[13] Feng W, Kandlur D, Saha D, Shin K. The BLUE Active queue management algorithms. IEEE/ACM Transactions on Networking. 2006;10(4):513-528.

[14] Abbasov B, Korukoglu S. An active queue management algorithm for reducing packet loss rate. Mathematical and Computational Applications. 2009;14(1):65-72.

[15] Akamai Technologies White Paper. Internet Bottlenecks: The case for Edge Delivery Services; 2000. Accessed 18 May 2014.
Available: http://www.cse.cuhk.edu.hk/~cslui/CSC5480/akamai-bottlenecks.pdf.

[16] Van Jacobson, Borrelli R, Coleman C. Differential Equations. Prentice-Hall Inc.; 1987.

[17] Libor Dostalek, Alena Kabelova. Understanding TCP/IP: A clear and Comprehensive Guide to TCP/IP Protocols. PACKT Publishers, Mumbai, India; 2006.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
www.sciencedomain.org/review-history.php?iid=699&id=6&aid=6364