Assignment #2 for Computer Networks

Savvas C. Nikiforou
Department of Computer Science and Engineering
University of South Florida
Tampa, FL 33620

***Abstract* – The purpose of this assignment is to compare the queueing behavior of real network traffic input to an infinite size queue versus the M/M/1. To simulate the real network a trace file was used, which was taken on the USF backbone 100-Mbps Ethernet connected to the router to the Internet. The results of the experiment show the difference between the real network simulation and the M/M/1 simulation, which grow bigger as the server utilization increases. The difference in the results between the two simulations is due to the assumption that the network traffic has a Poisson arrival time as well as packet length. It is clearly shown that it is not really the case but instead the packet size as well as the arrival time histograms is showing a very bursty behavior. Another thing that it is not taken into account for the M/M/1 model is the self-similarity of the data.**

I. INTRODUCTION

The purpose of this paper is to calculate how closely the behavior of a real network is to that of the more theoretical M/M/1.

In order to model network traffic most of the time many assumptions are usually made. This is necessary because computer networks are unpredictable. It is very important though to be able to predict the behavior of the traffic in a network because this will help to maximize utilization of resources in the network. An assumption that is made very often is that packet sizes and arrival times are Poisson processes. This type of processes is predictable and easy to work with, but they don't always describe the network under testing very precisely. Recent work shows that LAN traffic is much better mo deled using statistically *self-similar* processes because they have much different theoretical properties than Poisson processes [2]. For self-similar traffic, there is no natural length for a "burst"; traffic bursts appear on a wide range of time scales [1]. Another way to predict packet arrivals in a computer network is to use techniques similar to that predicting memory paged references in a paged memory computer [3]. It is observed that page references are correlated such that the probability of a page being referenced decreases as the time to its previous reference increases. Similarly if we find that the probabilities of packets going to different destinations in a computer network are not the same then we may use different strategies than if we assume the probabilities to be the same [3].

II. Previous Work

Jain in [1] suggests a new model for computer network traffic and argues that a packet train model is more close to the traffic in a real network than the Poisson distribution assumed in other models such as the M/M/1 queue. A histogram of the packet arrival times was used to show that the traffic does not have a Poisson distribution. For the traffic to be Poisson the histogram has to be exponential. The one of the real network though it was not exponential and at times it had lots of bursts. The measurements on real traffic led to a new model of arrival, which was named the *train model*. It suggests that packets flowing very close to each other most probably they have the same destination as the first packet in the bunch. This is similar to a train where all the cars have the same destination as the locomotive that is pulling them. Larger gap in time than a pre specified maximum suggests a different packet train with a destination different than the previous train.

The main reason that the M/M/1 model is not very similar to real networks is because of the actual real traffic characteristics, which do not obey the Poisson distribution.

[1] Paxson et al discusses why the failure of the Poisson model in the wide area networks.

III. BACKGROUND

The *M/M/1 Model:*

The M/M/1 model is a single server model with an infinite size queue. M stands for a Markovian and the first M is referred to the Arrival rate whereas the second refers to the Service Distribution. The system has a single server with an infinite capacity of a queue and the number of possible customers is unlimited. The M/M/1 queue displays exponential service time. In the M/M/1 model the probability of an arrival is independent of the previous one. i.e the traffic is independent. A steady state is reached when the number of arriving customers is less than rate at which the server can provide service to them. The queue discipline is first come first served.

The M/M/1 queue id described the following set of formulae.

- $L_q = \rho^2/(1 - \rho)$, where L_q = length of the queue
- $W_q = (\lambda/\mu^2) / (1 - \rho)$ where W_q = wait in queue

IV. THE SETUP FOR THE SIMULATION

To compare the behavior of the M/M/1 and real network several simulation programs were used. Values were gathered for both the real network and the M/M/1 for different utilization values varying from 10 to 98%. In order to gather results for different utilization values the link speed (or media rate) was varied for both simulation programs.

The histogram of the inter arrival times is shown in Figure 1. As it can be seen from the histogram the time distribution is not a Poisson process. To be such the logarithmic histogram must be a straight line. Instead what we observe here is that the Histogram has a lot of bursts and lots of high values at small time stamps. At higher time stamps it has many bursts, which is quite different from the theoretical Poisson arrival, which is a straight line. This affects the results of the simulation a great deal.

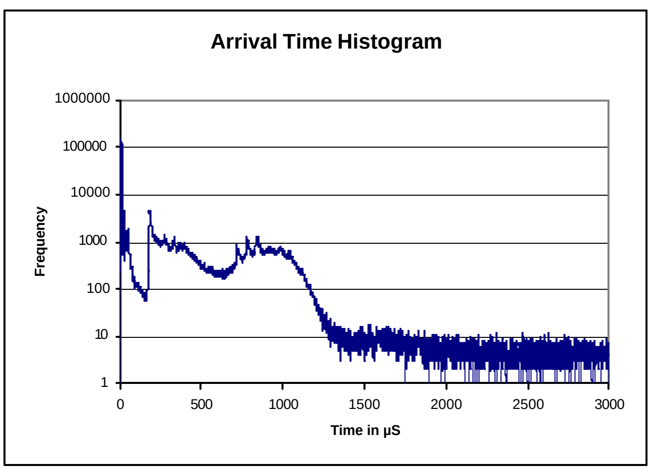


Figure 1. Inter arrival time histogram

For the simulations the CSIM libraries and simulation tools were used. The CSIM program is available from Mesquite Software. It is a simulation tool that is used to simulate artificial traffic based on the values of arrival time and packet length.

In Figure 2 the packet size histogram is presented where again we see that the distribution is not Poisson and it has lots of bursts as in the case of the time arrivals.
For the simulations in this paper the incoming traffic was used. This is important to be mentioned because having used the outgoing traffic the results might be different. This again happens because the traffic coming into a closed network is not necessarily the same as the traffic leaving the network.

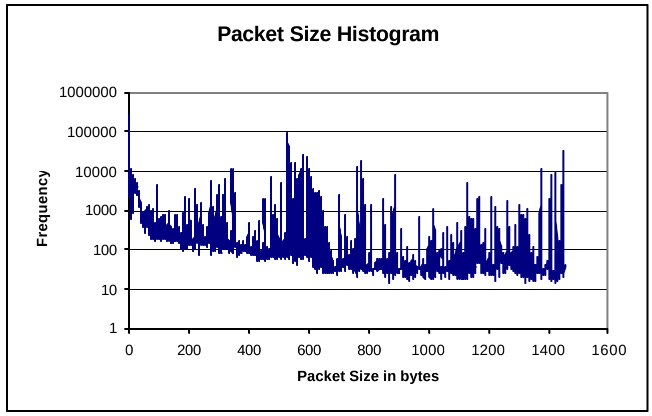


Figure 2. Packet Size Histogram

V RESULTS

Real Network Simulation				M/M/1 Results		
?(%)	Q_L	T_R	T_S	Q_L	T_R	T_S
10	0.183	0.061	0.034	0.111	0.000	0.000
20	0.424	0.142	0.067	0.250	0.001	0.001
30	0.699	0.235	0.101	0.429	0.001	0.001
40	1.0195	0.342	0.134	0.667	0.002	0.001
50	1.395	0.468	0.168	1.000	0.003	0.001
60	1.856	0.623	0.201	1.500	0.004	0.002
70	2.461	0.826	0.235	2.333	0.007	0.002
80	3.374	1.134	0.269	4.000	0.012	0.002
90	5.241	1.759	0.302	9.000	0.027	0.003
92	5.951	1.997	0.309	11.422	0.034	0.003
95	7.619	2.557	0.319	18.921	0.056	0.003
98	11.184	3.754	0.329	43.910	0.131	0.003

Table 1 Simulation Results

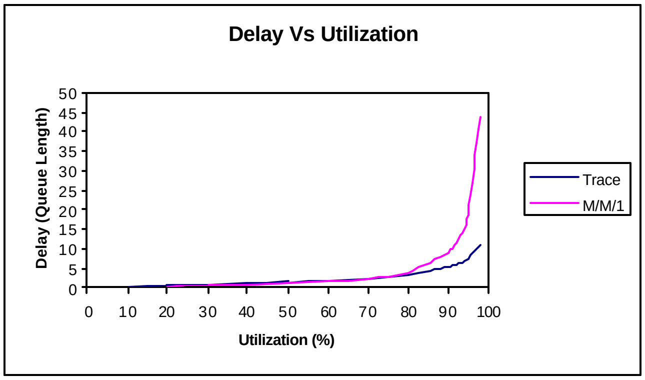
Running the simulation for the real network as well as the M/M/1 we gathered results on mean service time, response time utilization and queue length.

The results for the two simulations are summarized in Table 1, where ? is the utilization; Q_L is the mean queue length; TR is the mean response time; and TS is the mean service time.

We can see that the response time increases as the utilization increases and the same is true for the mean queue length.

Another observation is that the results for the real network are really close to those of M/M/1 for low utilization but as the utilization goes up so does the difference in the values, especially for the queue length.

On the other hand the service time does not change dramatically as it is expected but although. Graph 1 represents the delay Vs the utilization for the two simulations. We can see that M/M/1 performs slightly better for small utilization than the real network, but as the utilization increases the real network has a better



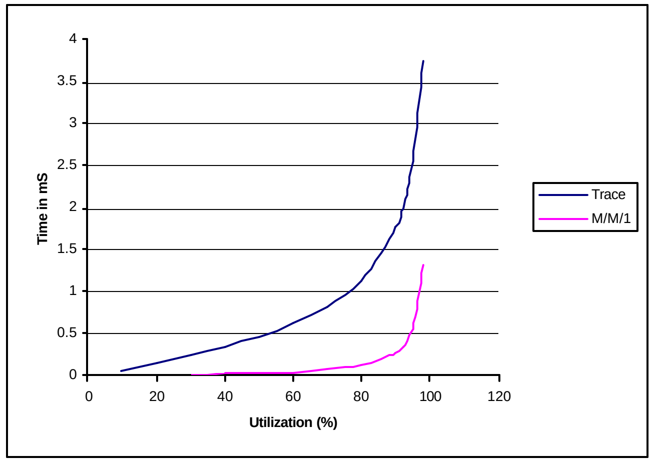
Graph 1 Delay vs Utilization

performance. The change in performance takes place little after the 70% utilization.

It is interesting to see the difference in percentile delay for the two systems. These results are presented in Table 2 and Graph 2 shows the difference graphically. The point where the values are changing from positive to negative is where the quality of performance changes for the two systems. This is the same place where the two curves meet in Graph one.

?(%)	Trace Delay	M/M/1 Delay	Difference (%)
10.00	0.18	0.11	64.86
20.00	0.42	0.25	69.60
30.00	0.70	0.43	62.94
40.00	1.02	0.67	52.85
50.00	1.40	1.00	39.50
60.00	1.86	1.50	23.73
70.00	2.46	2.33	5.49
80.00	3.37	4.00	-15.65
90.00	5.24	9.00	-41.77
92.00	5.95	11.42	-47.90
95.00	7.62	18.92	-59.73
98.00	11.18	43.91	-74.53

Table 2 The difference in Delay



Graph3. Response time Vs Utilization

In graph 3 the response time for the two models is plotted against the utilization. It is obvious that the performance of the theoretical M/M/1 model is superior to that of the real network. This is because of the irregularity of the time arrivals in the real network.

In order to achieve different utilization values the link speed was varied. Then the value used for the speed for each utilization value was also used to calculate the packet delay, by dividing the packet size in bits by the link speed in Mbps. This was giving the delay time in seconds. Then it was easy to calculate the mean variance and 99% values of the delay times by simply running one of the tools on Dr. Christensen’s Tool’s page. As it can be seen from the results the mean delay increases as the utilization increases. The Results are shown in Table 3. Mean delay variance and Standard Deviation are displayed in Table 4. As it can be seen from the results all the values increase as the Utilization increases. This means that as the server becomes more busy the delay of the packets in the queue increases with it. The values are in µS. This is expected in a way because as the server utilization increases the jobs that have to be performed by the server are more therefore more system handling is also needed from the server site. As the Utilization approaches 100% the mean delay time increases dramatically.

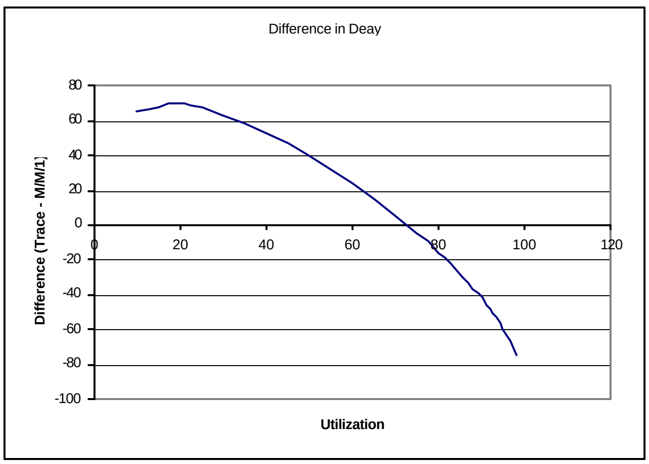
Utilization	1%	99%
10	3.144	79.323
20	6.28	158.646
30	9.43	237.97
40	12.574	317.293
50	15.71	396.616
60	18.86	475.939
70	22	555.262
80	25.14	634.585

90	28.29	713.939

Table 3 Mean Delay 99%

? (%)	Mean	Variance	Std Dev
10	33.57	868.09	29.46
20	67.14	3472.40	58.92
30	100.70	7812.95	88.39
40	134.27	13889.70	117.85
50	167.84	21702.61	147.31
60	201.41	31251.71	176.78
70	234.97	42537.09	206.24
80	268.54	55558.59	235.70
90	302.11	70316.54	265.17
92	308.82	73476.33	271.06
95	318.89	78346.40	279.90
98	328.96	83372.73	288.74

Table 4 Mean, Variance and s



Graph 2 Trace –M/M/1 Vs Utilization

VI. CONCLUSIONS

The behavior of the real network is similar to the M/M/1 for some values of utilization whereas it varies a lot for other. This is because of the real traffic characteristics. In the case of the M/M/1 model we assume that the arrival packet time as well as the packet size as both independent from the previous arrivals. In other words we assume Poisson distribution arrival time and packet length. This does not happen though with real data. If we take real data for a very long time then their distribution becomes closer to Poisson but they have a lot of bursts, which makes the real networks have a different behavior from the theoretical M/M/1 model.

The variable of interest to achieve better performance is actually the Media Rate or Link speed. We can also vary the service time in order to increase utilization but usually this may not be feasible with real networks.

VII. CONSIDERING PACKET LOSS AS A RESPONSE VARIABLE

(Extra credit part)

If we are to consider packet loss as a response variable then we have to decide how to deal with the queue size. We can approach the problem in two ways, either by limiting the number of customers that can be in the queue or by limiting the byte capacity of the queue and allowing any number of customers which their packet sum does not exceed the byte size of the queue. Each method has pros and cons. The best solution depends mostly on the traffic characteristics. If the mean packet size is very small then allowing for the byte size queue instead of the customer number makes more sense since in this way we can accommodate more customers. On the other hand if the mean packet size is large we might want to limit the number of the customers. In this case we have to make sure that the capacity of the queue will be large enough to accommodate N number (where N the number of customers allowed in the queue) times the maximum possible packet length that a packet can have. This though makes us rethink the solution of N customers queue. The reason for that is that we may not know the maximum size that a packet can have. Also in the case where many small packets need to be queued we want allow it wasting in this way bandwidth. The best solution is to limit the queue buffer size by bytes and not by packet number.

ACKNOWLEDGEMENTS

I would like to thank Dr. Christensen for Answering questions concerning the assignment. Also the T.A. of the class Mr. Hiroshi for his valuable help and availability to answer questions any time. Several hours of discussion on how to approach the solution to this assignment took place with my colleague Nikhil Kelshikar as well as other students in the class.

REFERENCES

[1] Vern Paxson and Sally Ford "Wide-Area Traffic: The failure of Poisson Modeling"

[2] Will E. Leland, Walter Willinger, Murrad S. Taqqu and Daniel V.Wilson. "On the Self -Similar Nature of Ethernet Traffic"

[3] Raj Jain and Shawn Routhier. "Packet Trains – Measurements and a New Model for Computer Network Traffic.