Proceedings of the 14th International Conference on Automation & Computing, Brunel University, West London, UK, 6 September 2008

Optimising Wireless Network Control System Traffic – Using Queuing Theory

Alison Carrington (nee Griffiths), Chris Harding, and Hongnian Yu
Faculty of Computing, Engineering and Technology
Staffordshire University
Stafford, United Kingdom
{A.L.Carrington},{C.A.Harding},{H.Yu}@staffs.ac.uk

***Abstract*—The network delay has a huge impact to the quality of a networked control system. This paper investigates modelling the network delay incurred by control system packets traversing a Mobile Ad Hoc Network from a plant to a controller using the queuing theory. Control systems can become unstable if information is lost or delayed so in this paper we compare voice traffic modelling with control packet data characteristics and show that with correct scheduling then delay can be minimised and kept stable until the network becomes congested**

Keywords- Networked Control System; Queuing Theory; WNCS; OPNET;

I. INTRODUCTION

A Networked Control System (NCS) traditionally consists of a wired based communication medium, either direct connections between the plant and controller using dedicated cables or by employing a bus based technology such as token ring or Ethernet. Recent research has investigated using wireless networks between the plant and a backbone wired network technology such as Ethernet to the controller.

This paper investigates using a wireless network that does not rely on any wired infrastructure as the communication medium. Mobile Ad Hoc Networks (MANETs) are dynamic infrastructure less wireless networks where each node within the network is required to forward and route packets, nodes can also leave and enter the network in real time due to their mobility.

The work in [1] describes an algorithm that can be used to calculate the delay between the plant and the controller over a multi-hop MANET, but this work does not provide a method of calculating a value for the transmission delays.

This paper employs tele-traffic queuing theory to calculate the transmission delay between neighbouring nodes; this can further be extended so that it can be used to mathematically model the multi-hop delay algorithm. This work addresses the problem of using constant arrival rates rather than the Poisson arrival rates that are commonly used within tele-traffic theory for both congested and non-congested conditions [2].

This paper contains the following; Section II provides a literature survey that introduces WNCS, MANETs and queuing theory. Section III shows the physical system that is to be modelled. Section IV presents the Poisson approximation equations as well as introducing the new mathematical model for calculating the WNCS queue size and delay. Whilst section V introduces the three test cases and simulation parameters used and the results for congested and non-congested systems are presented and discussed in section VI. Finally section V presents the conclusion and future work.

II. LITERATURE SURVEY

Work carried out by [3] shows that when the arrival at the queue is constant and of a set packet size then equations that assume Poisson arrival rates are not accurate when calculating the queue size or queue delay.

Existing measurements of the average queue delay within voice networks have to contend with the bursty nature of the network. Burst networks have to contend with two main phases within the communication over the network, periods of large amount of communication (the burst phase) and silent phases, as such measurement calculations must take into account the average number of packets passing through the queue [4]. The work carried out in [5] extended that in [6] to include the multi-circuit case and provides techniques for modelling the two phases of voice traffic.

The Random Early Detection (RED) [7][8] protocol for gateways detects incipient congestion by computing the average queue size and dropping random packets when the queue size reaches a threshold, this helps to reduce congestion at gateways. Tele-traffic theory as shown in [9] can be used to calculate the average queue size when using Erlang's Poisson inter-arrival arguments and packet sizes [2].

A NCS contains a plant and a controller that communicate over a real time communications network. The plant sends its state to the controller at a constant rate at a set period, also known as the plant Sampling Period (T_{sp}). This means that calculations for the average queue size and queue delay must reflect the constant inter-arrival arguments at the queue and are therefore not based on Poisson arrival rates [1][10].

III. PHYSICAL WIRELESS NETWORKED CONTROL SYSTEM

The plant transmits its status once every T_{sp}, it arrives at the controller after a time $\tau_{latency.}$ This is negatively fedback and compared with the reference model, the error signal is fed to the controller as shown in Figure 1.

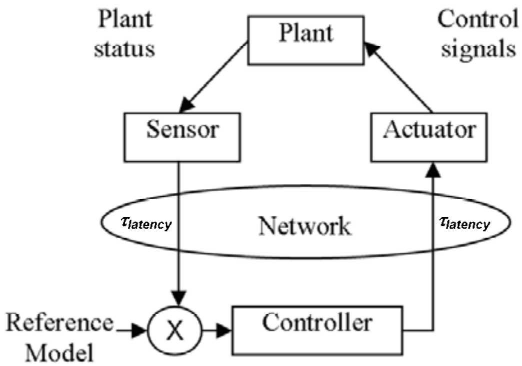


Figure 1. NCS Generic closed loop model

The controller then transmits the control signal to the plant which instructs it to change its input parameters after $\tau_{latency}$ has elapsed, so it can update its status accordingly, unless the signal is lost or delayed. This process repeats ad-infinitum.

The plant and controller may not actually be situated in close proximity due to logistical or safety reasons, therefore the input signal may incur some delay or even become lost between transmission and reception this is dependent on the transmission medium.

In this work the plant and controller are nodes within a Mobile Adhoc NETwork or MANET, an example topology is shown in Figure 2.

MANETs consist of several nodes which transfer information over an infrastructure less wireless network; each node contains a complete or partial topological map of the network [11]. A packet may be transmitted over a number of unknown hops before it reaches its destination and consequently packets may take different routes due to the mobility of the nodes and the routing protocol parameters.

Consider the simple peer-to-peer communication used with direct communication between the plant and the controller illustrated in Figure 3. The delay between the plant and the controller, $\tau_{latency}$ can therefore be calculated.

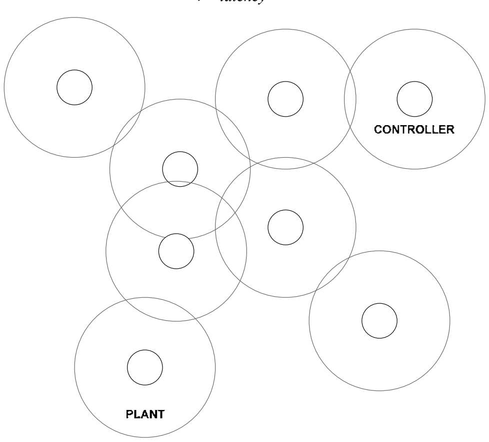


Figure 2. Multi-hop MANET network

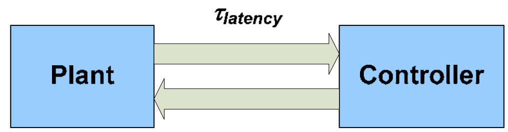


Figure 3. Peer-to-Peer communication with latency

A plant sends its current state to the controller within a constant sized packet at a pre-defined interval T_{sp}, the time taken to traverse the link is dependent on the capacity of that link. If there are m nodes in the route between the plant and the controller then there are $m-1$ links to cross. For a single link it is possible to calculate τ_{latency} using (1).

$$\tau_{latency} = \tau_{proc} + \tau_{queue} + \tau_{tx} + \tau_{prop} \qquad (1)$$

where τ_{proc} is the processing delay, τ_{queue} is the queuing delay, τ_{tx} is the transmission delay and τ_{prop} is the propagation delay of a particular link. For this paper a single hop is considered and the only delay to be taken into account is queuing delay hence (1) simplifies to:

$$\tau_{latency} = \tau_{queue}$$

In general a MANET may contain multiple hops, the packet latency can be calculated by summing the physical delays of each link (2).

$$\tau_{latency} = \sum_{i=0}^{m-2} \left(\tau_{proc_i} + \tau_{queue_i} + \tau_{tx_i} + \tau_{prop_i} \right) \quad (2)$$

The multiple node wireless system will be considered in later work. The single node system is illustrated in Figure 4.

The Plant consists of a number of sources which represent multiple systems that each generate traffic at a constant arrival rate of λ_i messages/sec which is the inverse of the sampling time of the particular plant, T_{spi} for i from 1 to the number of sources, n. This is an application of the model of a d/d/1 queue as presented in [3]. The messages are passed to an infinite buffer, which is modelled as a first-in first-out (FIFO) queue. The packets a serviced at a rate of C bits/second (3).

$$C = \frac{(d_r)}{\mu^{-1}} \qquad (3)$$

Where d_r is the data rate of the system and $\mu^{-1}$is the packet size (bits/message). Traditionally tele-traffic theory the n processes are Poisson processes and the size of the packets also have a Poisson distribution [2]. Using the work in [9] the average queue delay is shown in (4).

$$\bar{\tau}_{queue} = \frac{1}{\mu C - \lambda} \qquad (4)$$

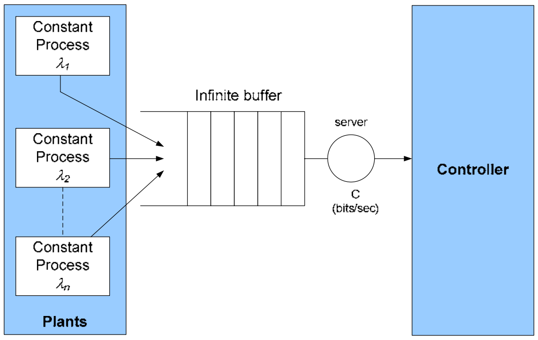


Figure 4. Mathematical Equivalent d/d/1 System model

Also the average queue size is given by (5):

$$\overline{P}_{queue} = \frac{\rho}{1-\rho} \tag{5}$$

where ρ is the traffic intensity parameter and is $\frac{\lambda}{\mu C}$, this value indicates the level of congestion, hence low values of ρ indicate low traffic levels and ρ greater than or equal to 1 indicates congestion. The plant sources are not Poisson processes but are constant, the Poisson approximation equations for an m/m/1 queue obtained from (4) and (5) will be compared with the derived d/d/1 queue with constant inter-arrival arguments and packet sizes.

IV. D/D/1 Mathematical Model

Consider the simple model it shows a single plant source, queue and sink controller processes as shown in Figure 5. This is implemented with OPNET Modeller simulation tool,.

One packet arrives every T_{sp} and it is serviced after one processing time, τ_{proc} by the queue as shown in Figure 6. a). τ_{proc} is calculated using:

$$\tau_{proc} = \frac{1}{\mu C}$$

Substituting for C from equation (3) to obtain:

$$\tau_{proc} = \frac{1}{\mu^2 . d_r} \tag{6}$$

When there are n multiple sources then n packets arrive simultaneously as shown in Figure 6. b).

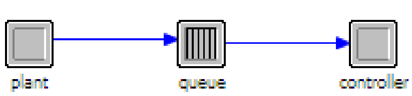


Figure 5. queuing model with 1 source

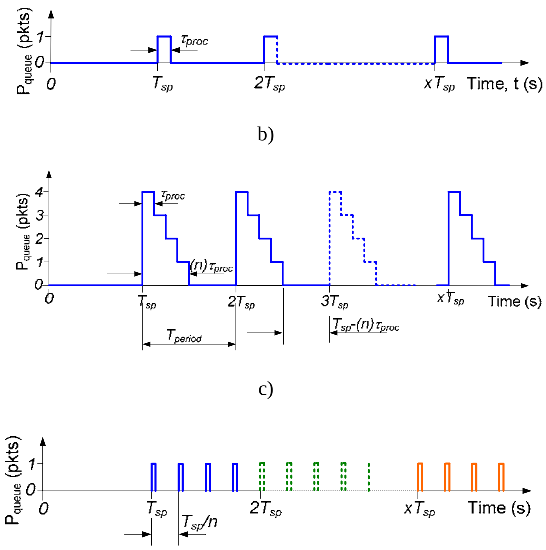


Figure 6. queue operation a) single b) simultaneous c) staggered sources

n packets arrive every T_{sp}, they are queued and each packet is served after a minimum delay of τ_{proc} and multiples thereof. It would be convenient to schedule the packets to arrive separately and hence each packet would only be queued for the minimum time of τ_{proc}, this is illustrated in Figure 6. c). This can be achieved by staggering the start times of the sources such that, the start time of the i^{th} source is given (7).

$$\tau_{start_i} = (i-1)\frac{T_{sp}}{n}, \quad for\ i = 1\ to\ n \tag{7}$$

Assuming all T_{sp} are constant, otherwise (7) is calculated for i=1 to the number sources with same T_{sp} for each particular T_{sp},.this situation will be considered in further work.

All things being equal there is a limit to the number of packets that the queue can service in any given T_{sp}. If the time between subsequent arrivals is less than τ_{proc} then no queuing occurs, this limit can be calculated by (8).

$$n \le \frac{T_{sp}}{\tau_{proc}} \tag{8}$$

If condition (8) is met then $\rho < 1$, otherwise congestion occurs. Figure 7. illustrates the effect of condition described in (8) not being met.

The queue will not clear before the next batch of control packets arrive and hence the number of packets in the queue will keep increasing with time and hence so will the packet delay. Clearly when arrivals are staggered the queue time will increase but not as drastically as the non-staggered start times. If n is the number of sources, generating packets at a constant sample period, then n packets will be generated every sampling period and arrive at the queue at the same time.

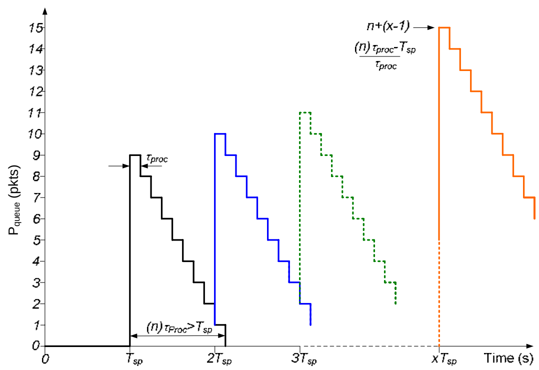


Figure 7. congested queue operation

Then a packet will suffer a maximum queue delay of $n(\tau_{proc})$ but the mean queue delay when $\rho < 1$ is (9).

$$\bar{\tau}_{queue} = \frac{1}{n} \lim_{n \to T_{period}} \sum_{i=0}^{n} \left[(i) \left(\tau_{proc} \right)_i \right] \qquad (9)$$

The average number of packets in the queue can be calculated using (10).

$$\bar{P}_{queue} = \frac{1}{n} \lim_{n \to \infty} \sum_{i=0}^{n} \left[P_{queue}(t_i) \right] \qquad (10)$$

when $\rho < 1$ then n tends to $T_{period,}$ which is the same as T_{sp} for this case. For multiple T_{sp} T_{period} can be calculated, for example if T_{sp1}=0.02s and T_{sp2}=0.03s then:

$$3T_{SP1} = 2T_{SP2} = T_{period}$$
$$3(i)T_{SP1} = 2(i)T_{SP2}$$

Otherwise when $\rho >= 1$, n tends to infinity and congestion occurs and $\bar{P}_{queue}$ will increase with a rate that is related to the value of ρ.

V. Simulation Parameters

The three test cases are described in TABLE I. These cases represent an un-scheduled WNCS (case 1), an optimally-scheduled WNCS (case 2) and finally voice traffic for the same conditions (case 3). TABLE II. shows the common input parameters, for cases 1 and 2, each plant source creates a μ^{-1} 64 bit packet every T_{sp} 0.02 seconds, giving a λ of 50 pkts/s per source. For case 3 these values are exponentially distributed forming a Poisson distribution. The d_r is equal to a wireless LAN [12]. The service rate, C is calculated using (3) and is constant for cases 1 and 2 but a mean value for case 3. The number of plant sources was increased to the point at which congestion (8) is no longer satisfied, when n is 54 and $\rho > 1$, this is shown in TABLE III.

TABLE I. Test Cases

case	T_{sp}	distribution, pkt size	start time
1	const, 0.02	const 64bits	same
2	const, 0.02	const, 64bits	staggered
3	exp, 0.02	exp, 64bits	same

TABLE II. Common Input Parameters

T_{sp} (s)	d_r (bits/s)	μ^{-1} (bits)	C (bits/s)	μC (pkts/s)
0.02	11,000,000	64	171,875.00	2,685.55

TABLE III. Variable No. of Sources Input Parameters

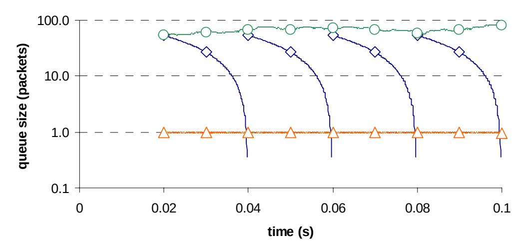
n (srcs)	1	53	54
λ (pkts/s)	50	2650	2700
ρ	0.019	0.99	1.01

$\bar{\tau}_{queue}$ and $\bar{P}_{queue}$ can be simulated and also calculated using (9) and (10) respectively for cases 1 and 2 and (4) and (5) for case 3. The equations have been validated and simulations were run for 1hour of simulation time and all results agree with calculated values, with an accuracy in the order of 10^{-8}.

VI. Simulation Results

Figure 8. shows the queue size for the 3 cases for a) ρ = 0.99 before congestion occurs and b) ρ = 1.01 when congestion occurs. The congestion causes the queue size to increase for case 3 in both a) and b). Whereas for case 1 there is queuing however the queue is cleared before the next batch of 53 samples arrive, however for 54 samples there is 1 packet left in the queue when the new batch arrives, this means at the arrival of each subsequent batch the size of the queue increases. For case 2 the queue size is minimal at ρ < 1 and there is a slight increase for 54 sources when ρ=1.01, however this has much less of an impact than that the of other cases two cases.

a)



b)

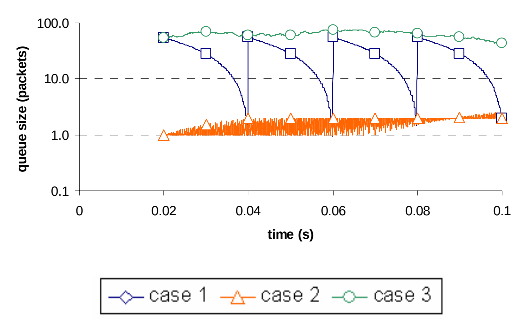


Figure 8. queue size: a) ρ = 0.99, b) ρ = 1.01

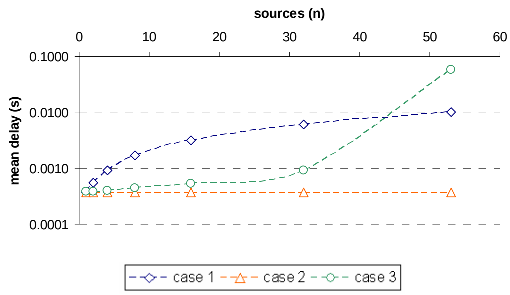
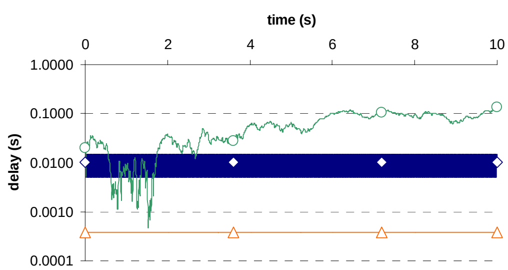


Figure 9. mean delay vs sources without congestion, $\rho < 1$

$\bar{\tau}_{queue}$ is plotted against the number of sources for the 3 cases without congestion when ρ < 1 in Figure 9. It can be seen that the mean delay increases linearly for case 1 from 0.372ms to 10ms when n=53 or ρ = 0.99. Case 2 mean delay is constant and minimal at 0.372ms for all ρ < 1. Case 3 is slightly higher than case 1 until n=32 or ρ = 0.60 then it increases and is greater than case 1 at n=45 or ρ = 0.84 this is due to the exponential distribution of inter-arrival rates and packet sizes, enabling congestion to occur at some instances and not at others.

The mean delay of packets is plotted against time (to 10s) for all 3 cases for a) no congestion and b) with congestion as shown in Figure 10. When ρ = 0.99 case 1 fluctuates about 10ms and again case 2 shows a constant delay of 0.375ms, whereas case 3 shows a higher value that increases with time, it reaches a mean value 0.1s after 10s and congestion is already occurring.

a)



b)

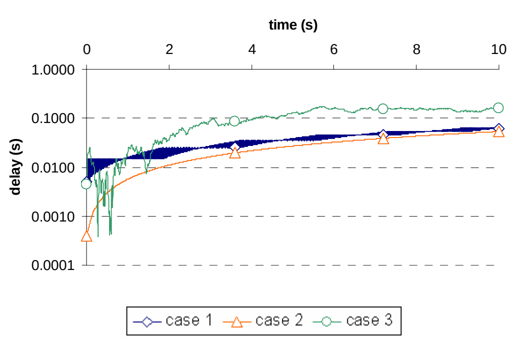


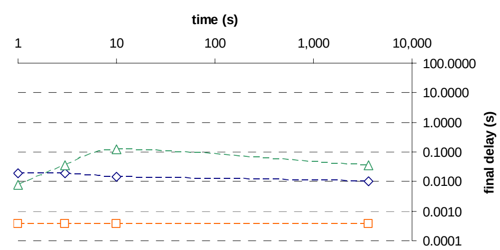
Figure 10. delay vs time: a) ρ = 0.99, b) ρ = 1.01

When ρ = 1.01 case 1 shows a linear increase with time and is less than all of the other cases. Case 2 also shows a linear increase but the delays have a variance of results than in case 1. Case 3 shows a higher increase with time with a much higher variance.

The final delay is plotted against time for both a) no-congestion and b) congestion is shown in Figure 11. When ρ = 0.99 case 1 is a constant 0.372ms, which is 27 times faster than case 2, that varies slightly about 10ms. Case 3 is much more variable hence the need for the approximations defined by (4) and (5), these are 157 times slower than case 2 results at 58.41ms. Hence these approximations are not suitable for predicting delays for d/d/1 queues.

When ρ = 1.01, the final delay increases with time for all cases, refer to Figure 11. b. This is the expected result however it occurs at different rates for the cases. Case 2 is 3.00ms after 1s, which is 2.91 times faster than case 3 and 4.72 times faster than case 1. Initially case 1 has a worse performance than case 3, again this is due to the inter-arrival arguments and packet sizes being Poisson in case 3. After 3s case 3 has the worst performance, after 1hour then case 1 and 2 are almost equal at 9.645s and 9.635s respectively, these are extremely high values compared to the non-congested case and a WNCS system would soon become unstable in these conditions, hence the need for this optimisation to occur. Case 3 is 2.52 times slower than case 2 at 24.297s after 1hour. As time is increasing the performance difference between cases 1 and 2 is reducing, this is known as the Central Limit theorem, however the variance of the values is higher for case 1.

a)



b)

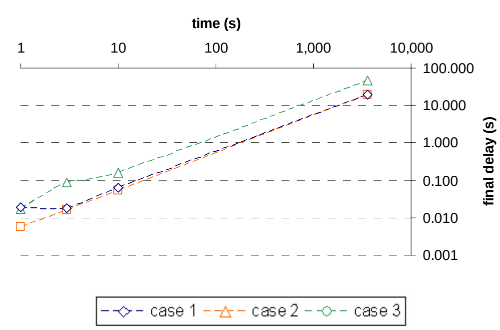


Figure 11. final delay vs time: a) ρ = 0.99, b) ρ = 1.01

VII. Conclusion

The paper proposes a method to successfully simulate and calculate average queue length and time for a single node with many different sources for the 3 different test cases. It has been shown both theoretically and through simulations that the Erlang, Poisson process approximations of an m/m/1 queue are not suitable for predicting results for d/d/1 queues. A d/d/1 system can be optimised should the arrival times of the sources be staggered and there is a distinct point at which performance degrades, i.e. when the traffic intensity parameter $\rho = 1$ and congestion occurs, this is in contrast to the m/m/1 model in which performance slowly degrades as ρ tends towards 1. The developed equations can be used to accurately predict the network delay which plays an important role in the quality of NCS.

Further work includes adding external traffic sources at each intermediate node within the model and varying the amount of external traffic. This will enable us to analyse the effect of control traffic for various different control systems, this is similar to SNR (signal-to-noise ratio) measurements in Telecommunication systems. Different network topologies could be measured and more layers of the ISO protocol stack for MANETs could be included and compared to the theoretical expectations. The harmful effects of the radio link can also be included and performance degradations can also be evaluated. This work has provided a solid benchmark system for future work as different queuing mechanisms can also be modelled, including priority queuing. This is useful as some control problems are more prone to become unstable due to delayed samples than others.

References

[1] C. A. Harding, H. Yu and A. L. Griffiths, “Delay Algorithm for WNCS over MANETs in MATLAB and SIMULINK”, Chinese Automation and Computing Society Conference, Loughborough University, UK. September 2006.

[2] E. Cinlar, “Superposition of Point Processes,” Stochastic Point Processes (P. A. W. Lewis ed.) Wiley, 1972, pp. 549-606

[3] A. E. Eckberg, “The Single Server Queue with Periodic Arrival Process and Deterministic Service Times”, IEEE Transactions on Communications (pre-1998). Vol. 27. Issue 3. March 1979.

[4] V. S. Frost and B. Melamed, “Traffic Modelling For Telecommunications Networks”, Communication Magazine, March 1994, Vol. 32, Issue 3.

[5] T. Bially, B. Gold, and S. Seneff, “A Technique for Adaptive Voice Flow Control in Integrated Packet Networks”, IEEE Transactions on Communications. March 1980. Vol. 28, Issue 3.

[6] C. Weinstein, “Fractional Speech and Talker Activity Model for TASI and for Packet-Switched Speech”, IEEE Transactions on Communication. August 1978. Vol. 28. Issue 8.

[7] S. Floyd and V. Jacobson, “Random Early Detection for Congestion Avoidance”, IEEE/ACM Transactions on Networking (TON), Vol. 1, Issue 4. August 1993.

[8] T. Bonald, M. May and J. Bolot, “Analytic Performance of RED Performance”, Ninetieth Annual Joint Conference of the IEE Computer and Communication Societies. Vol. 3, March 2000.

[9] A. L. Griffiths and R. Carrasco, “IP Multiple Access between 3G/4G Mobile Radio and Fixed Packet Switched Networks”, PhD Thesis, Staffordshire University, 2004

[10] M. S. Hasan, C. A. Harding, H. Yu and A. L. Griffiths, “Modelling delay and packet drop in networked control systems using network simulator NS2”, International Journal of Automation and Computing, vol 2. 2005. pp 187-194.

[11] C. Siva Ram Murthy and B. S. Manoj, “Ad Hoc Wireless Networks Architectures and Protocols”, Prentice Hall. ISBN:013147023X.

[12] F. Lian, J. R. Moyne and D. M. Tilbury, “Performance evaluation of control networks: Ethernet, ControlNet and DeviceNet”, IEEE Control Systems Magazine, February 2001.