

When is Delayed Feedback More Effective?

Jitendra K. Singh and Vinod Sharma

Dept. of Electrical Communication Engineering, Indian Institute of Science, Bangalore-560012, India

E-mail: vinod@ece.iisc.ernet.in, Fax: (+)91-80-3600563

Abstract:— We consider the problem of the effectiveness of the delayed feedback in stochastic systems. Our motivation is the congestion control and scheduling of channels in communication networks via feedback. We conclude that regeneration length and correlation coefficient can provide information on the effectiveness of the delayed feedback. As a consequence the buffer length, the traffic intensity, and the burstiness of traffic influence the effectiveness of delayed feedback.

1 Introduction

It has been known for a long time that feedback in a system (deterministic or stochastic) can be very effectively used to improve the system performance. This has been demonstrated in feedback amplifiers, tracking targets, stabilising systems and also for congestion control in communication network (e.g. via flow control in TCP/IP networks and ABR service in ATM networks, and via scheduling of wireless channels in Cellular systems). However, it has also been variously demonstrated that if the feedback information is delayed, it becomes less effective in improving the system performance. Then it is of importance to know that if the information is delayed by time (say) Δ in a particular system, will the feedback be still useful and if yes how effective will it be.

The delay in feedback can introduce a lot more complicated behaviour than without delay. It can affect the deterministic systems differently than the stochastic systems. For example, in the deterministic systems delay can introduce oscillations which usually will not be seen in stochastic systems. Also, continuous time deterministic systems may behave differently from their discrete counterparts. The feedback delay has been much more systematically studied in the deterministic systems (see e.g. Fendick *et al.* [3], Yin and Hluchyj [11] for network applications and Hale [4] for a general theory), than for the stochastic systems (see however Fendick and Rodrigues [2], Pazhyannur and Agrawal [6]).

Our immediate motivation to study the general question of the effect of delay in the feedback is the congestion control and scheduling in communication networks. In long haul wired networks (Internet or ATM based) and satellite networks, the propagation delay can be very large

as compared to the queue dynamics in the switches and routers. Thus feedback may reach the flow controllers (e.g. a TCP connection or an ABR source) with significant delay. Also, if the uplink of a satellite network is scheduled by the satellite among the various earth stations, based on the queue length information at the earth stations, the delay in the information will be significant. Various studies mentioned above are motivated by these systems.

In this paper we study the effect of feedback delay on stochastic systems, considering several network models as examples. We start with the general idea that at time k the feedback sent at time $k - \Delta$, $\Delta > 0$ will be more useful if the system state at time k ($\equiv Y_k$) has more dependence on $Y_{k-\Delta}$. A key problem now is to define a useful concept of dependence which can be measured (in some sense for the system of interest to us) and then can provide practically useful information for efficient system design.

We consider in the next section the concepts of *mixing*, *autocorrelation* and *regeneration length* as measures of dependence. We study their interrelationship and then discuss their applicability for the following representative network examples: a queue with state dependent service rate, two queues with service to the longer queue and an ABR source with a bottleneck queue. As a result of our investigations we conclude for a given delay Δ , the feedback is more effective if

- i. buffer at the switches is longer.
- ii. ρ (traffic intensity) of the system increases (keeping the distributions same)
- iii. the second moment of number of arrivals in a slot increases.
- iv. burstiness of input traffic increases.

In addition as expected, the feedback becomes less effective as Δ increases. Furthermore, for a given system the largest Δ for which feedback will be effective can be estimated by looking at the autocorrelation of the queue length for the system with $\Delta = 1$.

The paper is organized as follows. The theory and network examples are provided in section 2. The simulation results are provided in section 3.

2 Theory and Examples

We start with the general idea that the state feedback of a system with delay $\Delta (\geq 0)$ is more efficient if the system has a *longer memory*. To be specific, for a stochastic system, if the system state at time t is Y_t , then if Y_t and $Y_{t+\Delta}$ have *more dependence*, the state feedback with delay Δ will be more effective. In particular, if Y_t is independent of $Y_{t+\Delta}$ then information about Y_t at $t + \Delta$ may not be useful for system control. However, for most systems, there is no constant Δ beyond which the system state becomes independent. For example, even for the simplest system of a two state irreducible Markov chain the *dependence* (defined appropriately) decays exponentially but we do not attain independence for any given Δ . Nevertheless for stationary processes various concepts of *mixing* are defined which are related to the rate of decay of some measure of dependence. Then one can make qualitative statements like if for one stochastic system the mixing coefficients (say the $\alpha(n)$ in a strongly mixing process) decay exponentially while for another, they decay only polynomially, then we should expect the feedback to be more effective in the second system at least for large Δ . But usually for queueing systems, it may not be easy to estimate the mixing coefficient decay rate. If we restrict the class of stochastic processes to the regenerative processes, then it is shown in [10] that if the regeneration length τ satisfies $E[\tau^{\beta+1}] < \infty$ for an $\beta > 0$ then $\alpha(n) \leq cn^{-\beta}$. The conditions for $E[\tau^{\beta+1}] < \infty$ for various networking systems are available in the literature and are much more tractable than directly obtaining the decay rates of the mixing coefficients. Further more, regenerative processes constitute a large enough class to encompass most stochastic models encountered in networks (including the long range dependent models).

Another reason for considering the distributions and moments of τ is that for a regenerative process if Y_k and $Y_{k+\Delta}$ are in different regeneration cycles then they are independent of each other and hence the feedback will be useless. Of course, τ is random in general and hence for a given Δ one can assume Y_k and $Y_{k+\Delta}$ to be independent only with a certain probability. Since $P[\tau \geq \Delta] \leq E[\tau^\beta]/\Delta^\beta$, a finiteness of higher moments of τ will generally imply that the probability Y_k and $Y_{k+\Delta}$ are in different regeneration cycles is higher and hence the feedback will be less effective. Of course one need not consider only the moments of τ . If we can show that the regeneration length of one system is (stochastically) larger than that of another, we would expect the feedback of the first system to be more effective.

Our reasoning above is qualitative/asymptotic and hence can only provide some guidelines to the possible

outcome which we will verify via simulations in the next section. Also for finite Δ , quantitative comparisons can be at variance with the results obtained via this reasoning. Thus we also consider the correlation coefficient ρ_Δ between Y_k and $Y_{k+\Delta}$. This is a common measure of dependence. One reason we have not emphasized this so far is that for the systems of interest to us, we don't know studies which provide exact results or even qualitative trends on correlation of Y_k and $Y_{k+\Delta}$ (unlike for τ). But our simulation results in the next section will suggest that the correlation coefficient is a better indicator of effectiveness of the delayed feedback than τ . Of course a larger τ should suggest a larger correlation. Based on correlation of Y_k and Y_{k+m} for $\Delta = 1$, we will give a reasonable estimate of the largest Δ for a system, for which one can expect useful feedback information.

We explore these ideas on three different stochastic systems related to networks. In the rest of this section we describe these networks. In section 3 we will verify our ideas via simulations.

2.1 Discrete Queue with Feedback

Consider a discrete queue with the time axis slotted. In slot k , A_k new packets arrive at the queue which has an infinite buffer. Let X_k be the queue length at the end of slot k . This queue gets a service in slot k if $X_{k-\Delta}$ is non-zero. Thus,

$$X_{k+1} = (X_k - \min(X_{k-\Delta}, c))^+ + A_{k+1} \quad (1)$$

where c is the maximum number of packets serviced in a slot.

This queueing system is motivated by satellite networks where this may be the queue at a user on an earth-station. The slots on the uplink are allocated by a scheduler on the satellite which allocates slots based on the queue length information it received Δ time back. This queue was studied in [10] and [7] and found to perform satisfactorily. In [10] the effect of delay Δ was also studied via simulations and a formula for the stationary queue length distribution found. We study this queue along with the system where the buffer length B is finite. We apply our ideas to explain and predict for this queue the qualitative behaviour of throughput and the channel utilization. We measure channel utilization by $EU = P[X_k > 0 | X_{k-\Delta} > 0]$, which we consider as a measure of feedback effectiveness. If the channel is utilized efficiently then the feedback information is useful. For our simulations we have taken $c = 1$. The sequence $\{A_k\}$ is assumed *iid* or an ON-OFF source in our simulations.

One can show that the system (1) is stable if $E[A_k] < 1$. For this system we take the regeneration epochs the

times k when $X_{k-\Delta} = 0, \dots, X_k = 0$. We have shown in [10] that for this system $E[\tau^\beta] < \infty$ (for A_k iid) if in addition $E[A_k^\beta] < \infty$ for any $\beta \geq 1$. It is also a necessary condition for $E[\tau^\beta] < \infty$.

If in (1), $\min(X_{k-\Delta}, c)$ is replaced by c then we have the following conclusions:

- i. For the finite buffer case, if the buffer length B increases, the corresponding regeneration length τ_B also increases stochastically *i.e.*, if $B_1 < B_2$ then $P[\tau_{B_1} \geq x] \leq P[\tau_{B_2} \geq x]$ for all x (denoted by $\tau_{B_1} \leq_{st} \tau_{B_2}$).
- ii. For a given $B \leq \infty$, if $A_k \leq_{st} A'_k$ then $\tau_B \leq_{st} \tau'_B$.
- iii. For the ON-OFF source, if the β^{th} moment (for $\beta \geq 1$) of ON and OFF periods is finite and the β^{th} moment of number of arrivals is finite, then $E[\tau^\beta] < \infty$.

One expects that similar conclusions hold for the system(1). Then as mentioned above we expect similar behaviour for channel utilization and correlation. We verify these conclusions for system (1), via simulations in section 3.

In [10] we also study a deterministic version of this queue (by replacing $A_{k+1} \equiv a > 0$). For $c > a$ we find that this system has a unique global attractor for all $\Delta \geq 0$. But a corresponding deterministic fluid queue model displays oscillations for $\Delta > 0$, although it has a unique global attractor for $\Delta = 0$

2.2 Serve the Longest Queue

In this section we consider two discrete queues with a single server. The arrival stream for queue $i, i = 1, 2$ is $\{A_k(i)\}$ and its queue length is $X_k(i)$. In slot k we serve the queue which has a longer queue at time $k - \Delta$. Thus the equations for the queue length are

$$X_{k+1}(i) = (X_k(i) - 1(X_{k-\Delta}(i) \geq X_{k-\Delta}(j)))^+ + A_{k+1}(i), \quad i = 1, 2, \quad j \neq i. \quad (2)$$

When $X_{k-\Delta}(1) = X_{k-\Delta}(2)$, in our simulations we serve any one queue with equal probability. This queueing discipline has been studied extensively and is a natural scheduling policy for channel sharing. Of course we have considered two queues for simplicity. The results should extend to more than two queues.

For this system also, we expect (i) – (iii) in section 2.1 to hold and then also their consequences. We measure the effectiveness of the feedback control by channel utilization as defined in section 2.1 and by the stationary mean number of packets in the system. We verify these conclusions via simulations in section 3.

A deterministic version of this system is studied in [10] where it is shown that $\Delta > 0$ may lead to oscillations.

2.3 ABR Source with a Bottleneck Queue

In this section we briefly describe a system which models an ABR service connection (in ATM networks) along with a bottleneck queue. For details of ABR service see [1] and analysis and model, Sharma and Kuri [9], (and reference therein). In contrast to [9], our model here is discrete.

The ABR source is an infinite data source which sends packets to a bottleneck queue. There is propagation delay of Δ_1 , in the forward direction. Another exogenous (uncontrolled) source also sends the packets to the queue. This source may represent the superposition of other (CBR, VBR etc.) sources sharing that channel. The queue length information is fed back to the ABR source at periodic intervals. It reaches the ABR source after a delay of Δ_2 . Let $\Delta = \Delta_1 + \Delta_2$. For given constants $0 < V_T < DV_T$, if the queue length X_n satisfies $V_T < X_n \leq DV_T$, then there is a mild congestion and we set a binary bit $NI = 1$ (otherwise 0). If $X_n > DV_T$ then there is high congestion and we set $CI = 1$ (otherwise 0). Also, MCR denotes the minimum rate, and PCR the peak rate for the ABR services. Two constants RIF and RDF are also appropriately fixed. Then the arrival rate λ_n after the n^{th} feedback for the ABR source is specified as

$$\lambda_n = \begin{cases} \min(\lambda_{n-1} + RIF \cdot PCR, PCR) & \text{if } NI = 0, CI = 0 \\ \max(\lambda_{n-1} - \lambda_{n-1} \cdot RDF, MCR) & \text{if } NI = 0, CI = 1 \\ \lambda_{n-1} & \text{if } NI = 1, CI = 0 \\ \max(\lambda_{n-1} - \lambda_{n-1} \cdot RDF, MCR) & \text{if } NI = 1, CI = 1. \end{cases} \quad (3)$$

A continuous version of this system is exhaustively studied in [9]. However, using similar techniques one can show that if the traffic intensity of the exogenous system is less than 1 and its packets arrive as an *iid* stream then there is a unique stationary distribution for the queue length process. Also, if the $E[A_k^{\beta+1}] < \infty$ for some $\beta > 0$, then $E[\tau^\beta] < \infty$ for this system.

3 Simulation Results

We provide the simulation results for the three systems described above. We consider $\{A_k\}$ to be *iid* with Bernoulli, Poisson, geometric, and a fat tailed distribution $p(n) = 1/s \cdot (n+1)^{-3}$, $n \geq 0$, where $s = \sum_{n=0}^{\infty} (n+1)^{-3}$. We also consider $\{A_k\}$ to be an ON-OFF stream with ON and OFF periods having geometric distributions.

We denote by $B, E[q], E[\tau], \rho_k$, and $E[U]$ the buffer capacity, the mean queue length, the mean regeneration length, the correlation coefficient with lag k and the channel utilization.

Because of the lack of space we have included tables and figures only for the geometric and ON-OFF case. But

the conclusions hold for other above mentioned distributions also. Figures for some of the above distributions are available in [10]. For the single queue of section 2.1 the conclusions are as follows. Keeping all other parameters constant,

- i As B increases, ρ_Δ , $E[U]$, $E[q]$ and $E[\tau]$ increase (Table 1).
- ii For infinite B , as $E[A_k]$ increases, ρ_Δ , $E[U]$, $E[q]$ and $E[\tau]$ increase (Fig. 1).
- iii As Δ increases, ρ_Δ and $E[U]$ decrease but $E[q]$ and $E[\tau]$ increase (Fig. 2). This shows that ρ_Δ is a better indicator of effectiveness of feedback, than $E[\tau]$. Part of the reason $E[\tau]$ increases is because as Δ changes the definition of regeneration epoch changes. This changes $E[\tau]$ more than it should.
- iv As second moment of $\{A_k\}$ increases, $E[q]$, ρ_Δ , $E[U]$ increase but $E[\tau]$ decreases. For the distribution with fat tail even though correlation is much more than for other distributions, the channel utilization is not so much higher (see Figures 6, 7 and 8 for the two queue system). The decrease of $E[\tau]$ is surprising in view of the qualitative results discussed in section 2.1
- v For very large Δ , all distributions provide the same performance. At that Δ X_k and $X_{k-\Delta}$ become almost uncorrelated for all distributions (see Fig. 8 for the two queue system).
- vi Looking at the correlation plot for $\Delta = 1$ (Fig. 3) one can roughly estimate the Δ up to which the feedback is effective. For example, for $\rho = 0.1, 0.2, 0.4$ we can get the largest Δ to be 2, 2, 6 respectively. Now from Fig. 4 we observe that the utilization beyond these Δ becomes below $(E[A])^{1/2}$, the utilization obtained when X_k and $X_{k-\Delta}$ are independent.

For the ON-OFF source all the above conclusions hold. In addition as the means of the ON and OFF periods increase (keeping $E[A_k]$ same, but now the burstiness increases), $E[q]$, ρ_Δ and $E[U]$ increase but $E[\tau]$ decreases (Fig. 5).

For the two queue system of section 2.2 we have following conclusions.

- i All the above results for single queue hold for this system also (Figures 6, 7 and 8).
- ii The queue with the higher second moment has more $E[q]$, and ρ_Δ , $E[U]$ than for the other queue for the same $E[A_k]$.

For the ABR service we have the following conclusions. (We do not provide the figures because of lack of space).

- i For a given $\{A_k\}$ of the exogenous traffic as Δ increases, $E[q]$ increases while $E[\tau]$, ρ_Δ , $E[U]$, and $P[V_T < X_n \leq DV_T]$ decrease. This indicates that the control becomes less effective.
- ii For a given Δ as $E[A_k]$ of the exogenous traffic increases, $E[q]$, ρ_Δ , and $E[U]$, increase but $P[V_T < X_n \leq DV_T]$ and $E[\tau]$ decrease. Similar trends hold when B increases.

References

- [1] "ATM traffic management specification 4.0," ATM Forum, April 1996.
- [2] K. W. Fendick and M. A. Rodrigues, "Asymptotic analysis of adaptive rate control for diverse routers with delayed feedback", IEEE Transactions on Information Theory Vol. 40, 1994 2006 - 2025.
- [3] K. W. Fendick and M. A. Rodrigues, Alan weiss "Analysis of rate-based control strategy for long haul data transport", Performance Evaluation 16 1992, 67 - 84.
- [4] J. K. Hale, "Theory of functional differential equations", NY 1997.
- [5] A. Mukherjee and J. C. Strikwerla, "Analysis of dynamic congestion control problem - A Fokker Planck approximation", Proc. SIGCOMM 1991 in Communication, Architecture and Protocols, Vol. 21, 1991.
- [6] R. Pazhyannur and R. Agrawal, "Analytical Numerical Results for Feedback Based Flow control of B-ISDN/ATM Networks with significant Delays.", IEEE INFOCOM 1995, 738 - 745.
- [7] S. G. Sankar, "A study of effect of information delays in reservation based bandwidth access", ME thesis, Dept. of ECE, IISc, Bangalore, 1998.
- [8] V. Sharma, "Reliable estimation via simulation", Queueing Systems, 19, 1995, 169 - 192.
- [9] Vinod Sharma and Joy Kuri, "Stability and performance of rate-based feedback flow controlled ATM networks", Queueing System 29 (1998).
- [10] J. K. Singh, "Effect of Delayed Feedback on System Performance" ME Thesis, Dept. of ECE, IISc, Bangalore, 2000.
- [11] N. Yin and Michel G. Hluchyj, "On closed loop rate control for ATM cell relay networks", Proc. IEEE INFOCOM Conf., No 15, 1997.

	$B = 5$	$B = 10$	$B = 15$	$B = 20$	$B = \infty$
$E[q]$	1.882	2.494	2.615	2.6309	2.632
$E[\tau]$	39.968	44.748	45.327	45.374	45.376
ρ_Δ	0.153	0.459	0.533	0.5468	0.549
$E[U]$	0.741	0.777	0.780	0.781	0.781

Table 1: Effect of buffer size for $E[A_k] = 0.5, \Delta = 5$

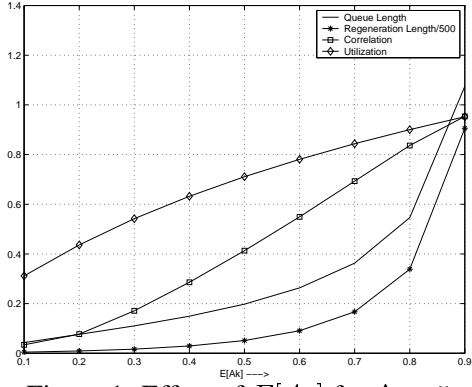


Figure 1: Effect of $E[A_k]$ for $\Delta = 5$

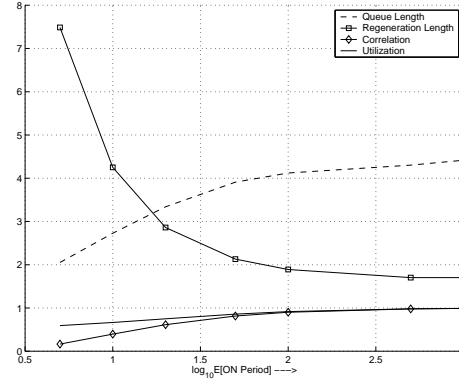


Figure 5: ON-OFF stream, $\Delta = 5$

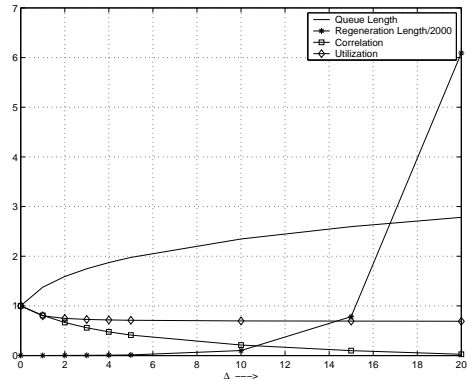


Figure 2: Effect of Δ for $E[A_k] = 0.5$

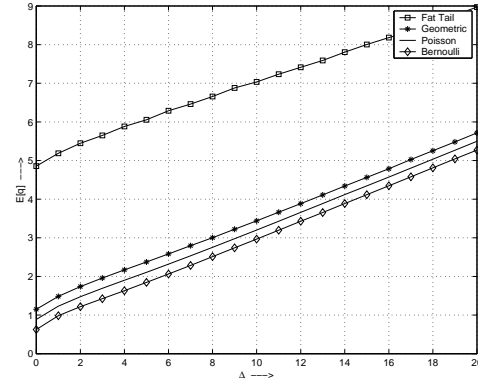


Figure 6: Two Queue, $E[q]$, effect of different dist. for $E[A_k] = 0.368$

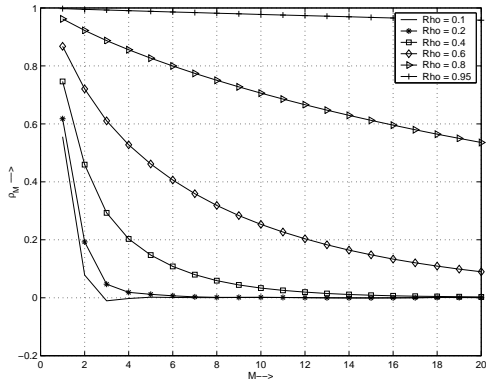


Figure 3: Estimation of affordable delay

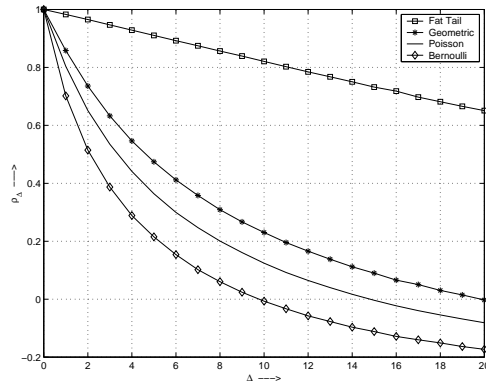


Figure 7: Two Queue, ρ_Δ , effect of different dist. for $E[A_k] = 0.368$

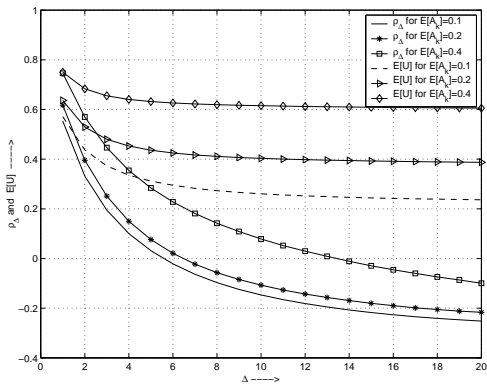


Figure 4: Effect of Δ for $E[A_k] = 0.1, 0.2, 0.4$

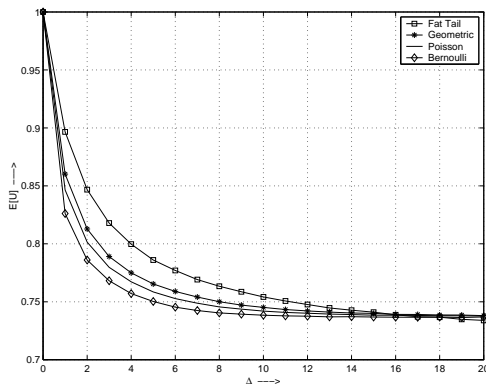


Figure 8: Two Queue, $E[U]$, effect of different dist. for $E[A_k] = 0.368$