ON THE ANALYSIS AND COMPENSATION OF NETWORK INDUCED COMMUNICATION DELAYS FOR DISTRIBUTED CONTROL SYSTEMS

By

EMMANUEL VYERS

Bachelor of Science

Florida Institute of Technology

Melbourne, Florida

1995

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the degree of
MASTER OF SCIENCE
December, 1997

OKLAHOMA STATE UNIVERSITY

ON THE ANALYSIS AND COMPENSATION OF NETWORK INDUCED COMMUNICATION DELAYS FOR DISTRIBUTED CONTROL SYSTEMS

Thesis Approved:

Gary E. Young
Thesis Advisor

Wayne B. Powell
Dean of the Graduate College

PREFACE

The control of physical systems with a computer is becoming commonplace. In addition, computer networking is now used to perform spatially distributed interrelated functions. Sensors, controllers and actuators can then be interconnected via the network to form a distributed control loop. However, due to the asynchronous nature of network communications, time varying transport delays are introduced into the feedback control loop. These network induced delays significantly degrade the performance of control systems. This is illustrated and discussed in this paper. Further, different compensation strategies proposed by control researchers are reported. Finally, a procedure has been derived and exemplified to integrate time varying network induced delays within the control design method traditionally used with time invariant systems.

I sincerely thank my advisor, Dr. Gary Young, for his supervision of this project. Through his support I was able to synthesize this review on the promising field of networked distributed control systems.

TABLE OF CONTENTS

Chapter Page

1 **INTRODUCTION** 1

1.1 The Problem 2
1.2 Objective and Contribution of the Research 2

2 **BACKGROUND ON NETWORKED CONTROL SYSTEMS** 4

2.1 The Feedback Loop Situation 4
2.2 Characteristics and Impact of the Network Data Latency 10
2.3 Dynamic Characterization of a Delayed Control Systems 17

3 **REVIEW OF THE LITERATURE TOWARD ANALYSIS AND COMPENSATION** 26

3.1 The P-step Observer 26
3.2 Optimal Compensation with Stochastic Delay Assumption 32
3.3 Robust Analysis of Time Varying delays 38

4 **A NEW INTERPRETATION TO DESIGN** 41

4.1 Proposition 41
4.2 Scope and Limitations 44
4.3 Example. 46

5 **CONCLUSIONS** 49

BIBLIOGRAPHY 51

APPENDIX 53

The Controller Area Network Communication Protocol for Distributed Control systems 53

LIST OF FIGURES

Figure Page

1. Three Methods of Forming a Closed Loop Over the Serial Bus 5
2. Distributed Delays in a Control System 6
3. The Variable Transport Delay in a Sampled Data System 8
4. Data Loss on the Measurement Signal. 9
5. Jitter on the Control Signal 10
6. Effect of Jitter in the Control Signal on the Output of the System 15
7. Effect of Data Loss in the Measurement Signal on the Output of the System . . 15
8. Combined Effect of Jitter and Data Loss on the Output of the System 15
9. A Slower Controller on the Situation of Figure 8 16
10. Initial Situation of a Delay Differential Equation 19
11. Elimination of the Time Variations by Data Buffering 27
12. Schematic Representation of a Three-step Predictor/Controller 29
13. Modeling Closed Loop Time-Varying delays as Disturbances 42
14. Configuration of the Distributed Delay Disturbances 45
15. Compensated System Response Under Jitter and Data Loss 48

CHAPTER 1

INTRODUCTION

In control engineering, a number of systems implement distributed interrelated functions incorporating more than one control device. The complexity of these functions requires an exchange of data that is invariably cumbersome and expensive when realized through hard wired signal lines. An efficient way to overcome these limitations is to network the system components using a serial data bus. Substantial savings are then achieved by reducing installation and maintenance costs while improving the system reliability. Along with reduced wiring, increased flexibility and lower cost, the serial bus system also enables the use of intelligent input/output devices. However, from the point of view of the control system designer, trading a dedicated point-to-point connection for a multiplexed network results in the adverse effect of introducing transport delays. These network induced communication delays represent a threat to the stability and dynamic performances of feedback control systems. As the result, a prime interest is to examine the possible compensation techniques. The core of this work is therefore composed of three sections, respectively, chapters 2, 3 and 4. An overview of the necessary background in relation to the characterization of the time varying network induced delays is presented first. A survey of the different techniques that have been proposed by researchers toward analysis and compensation of the network induced delays is given next. Finally a new interpretation to design in the face of network induced transport delays is introduced and discussed.

1.1 The Problem

Substantial advantages can be derived from networking a distributed control system. However, due to the asynchronous nature of network communications, time varying transport delays are introduced into the feedback loop. The compensation for constant time-delays has been extensively treated and is straightforward (cf. section 2.3). Comparatively, little work has been done to characterize and compensate the effects of time varying delays. Indeed, the stability of time varying systems cannot be predicted from the eigenvalues of the characteristic equation of the time invariant system. Further, in discrete time, one of the problem in the compensation is that time variations typically occur each sampling period. More precisely, the time varying nature of these delays affect the control signal by introducing jitter at the input of the process and producing data loss on the measurement data. This means that the traditional frequency domain analysis which is suitable for linear time-invariant systems may not be valid for analyzing the dynamic performances and stability of a closed loop distributed control system subject to time varying transport delays. Accordingly the sensitivity of feedback control systems to the network induced time variations can vary substantially and is difficult to determine.

1.2 Objective and Contribution of the Research

The objective of this research is therefore to provide a perspective on the problem of time-varying network induced transport delays for distributed control systems. This is accomplished by reporting several compensation solutions from the literature. These solutions represent a significant improvement over the uncompensated alternative but are invariably complex to implement and analyze. Clearly it would be advantageous to still be able to use the wide array of classical time invariant design methods despite the presence of the network induced time variations. With this idea in mind, an attempt has been made to try to capture the detrimental effects of the network induced delays into fictitious signals that once fed into the loop would account for the time variation effects and enable time invariant analysis. The particular contribution of this work is based on the development and discussion of such an interpretation. Specifically, jitter and data loss caused by the network are respectively assimilated as fictitious disturbance and noise signals. Qualitative reasoning on the controller's sensitivity to network induced delays can then be made with conventional time invariant methods.

CHAPTER 2

BACKGROUND ON NETWORKED CONTROL SYSTEMS

To properly understand the distribution of the control function over the network, many elements have to be considered. First a schematic representation of the feedback loop along with an illustration of the control flow is given in order to present and qualify the particular effects of the network induced transport delays. Then the characteristics of the data latency and its impact on control systems are discussed qualitatively and illustrated with an example. Finally, a short remainder on the compensation of both constant and variable delays is given.

2.1. The Feedback Loop Situation

There are three methods of forming a closed loop over a serial bus. The traditional approach is for the feedback loop to be closed by a remote controller. Then, both the measurement and the control signal travel back and forth on the bus in between the field and the control room. With the advent of smart instrumentation, the control strategy can also be moved away from the central controller and executed by the field instruments either at the measurement site or at the actuator location. By allowing devices to communicate directly with one another, extra passes through a central controller can be eliminated, thus allowing a loop to be closed with minimum elapsed time, minimum bandwidth, and minimum risk of error or other failures. Figure 1 illustrates these three alternative as a), b) and c).

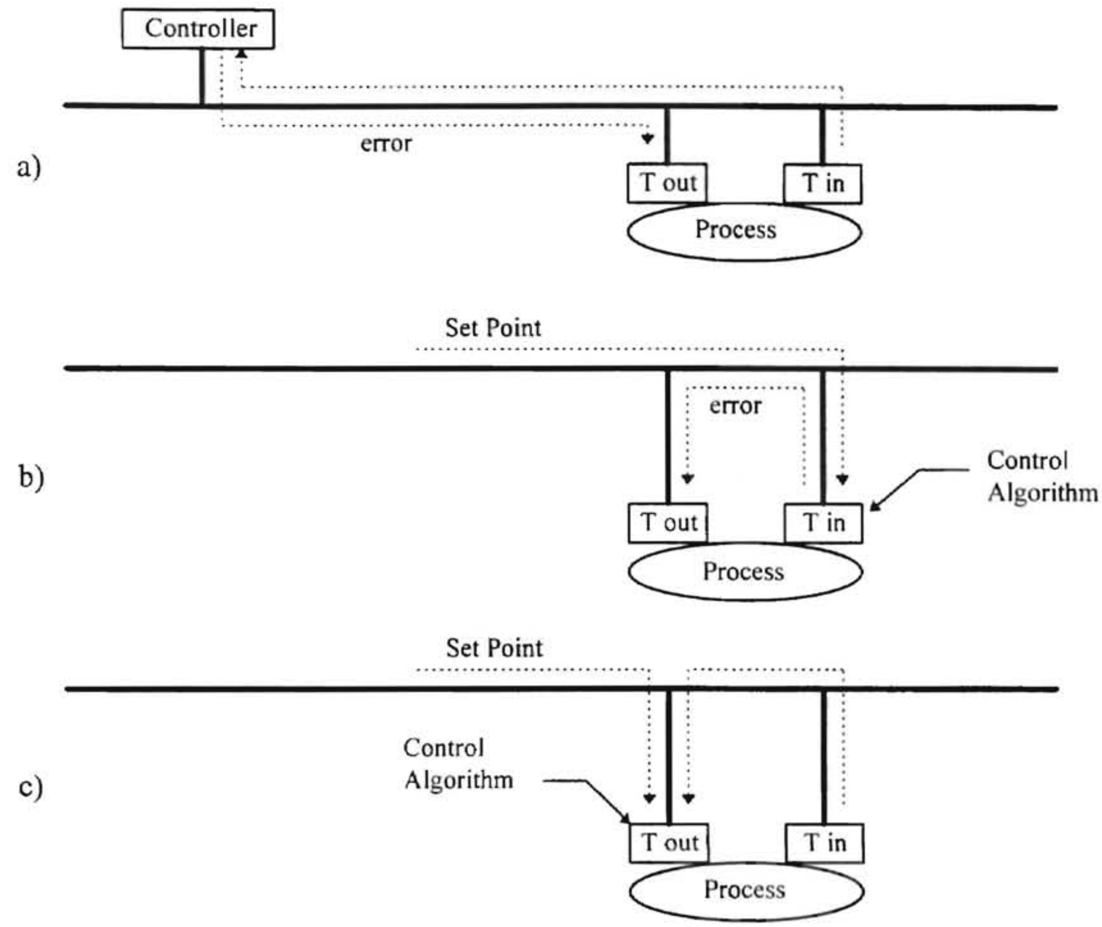


Fig. 1. Three Methods of Forming a Closed Loop Over the Serial Bus.

Let us examine the case where the controller is remote. Then the network transport delay will affect both the control signal and the feedback measurement as shown on Figure 2. The system under consideration consists of a continuous-time plant and a discrete-time controller that share the same data communication network with other subscribers.

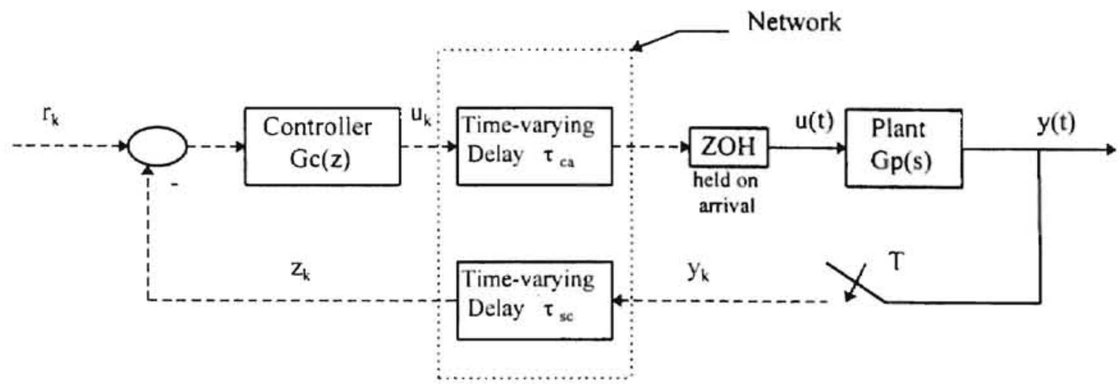


Fig.2. Distributed delays in a control system.

Note that the controller structure in Figure 2 is just one example. Nevertheless, in the later part of this report, this framework will be implicitly considered along with the following assumptions unless specified otherwise,

- There is a zero probability of data loss.
- The plant noise is defined and bounded.
- The plant is completely reachable and observable.
- Sampling and Zero Order Hold are ideal processes.
- There is no delay in the process of sensor signal generation.
- The delay Δ_p in the processing of the control signal is constant.
- Network induced delays are bounded to one sampling period.
- The actuator operates as a continuous-time device.

The network induced delays are only one source of delay among others of significance, namely,

- Delays in the dynamic of the system itself, typically caused by a system with mass transportation or by inertia in the actuation where it is difficult to obtain a measurement without delay.
- Delays inherent to sampling in digital control systems, due to periodic operation and zero order hold (ZOH) in the control system. This delay can be approximated by T/2, given the sampling period, T.
- The processing delay at the controller, is the time required by the computer to produce the expected control signal.
- The network data latency, defined as the difference between the instant of arrival of the message at the transmitter queue and the instant of reception of its last bit at the destination terminal.
- The detection delay, this delay is due to a lack of synchronization, either between cooperating periodic activities or a periodic and an event based activity. It is the time between the instant the last bit of the message has been received and the instant the message is actually picked up.

While the first three listed sources of delays can generally be treated as constant values, the network data latency and the detection delays stand out as truly time-varying characteristics. These time-varying delays are the primary focus of this report. Since the delay from the sensor to the controller is time-varying, the controller may use sensor data generated at the current or earlier samples. At the same time, even if the controller generates the commands at a constant rate, the interval between their successive arrivals at the actuator terminal may not be constant. Figure 3 is an illustration of this situation. On this diagram, the sensor sampling instant and the point where the controller picks up the measurement are assumed to be synchronized and time shifted by an amount Δs called the time skew. The control signal is generated Δp time later and transmitted to the actuator for immediate actuation. As the transmission delay is symbolized by a dropping dotted line, one can see that, variations will affect the control-flow and on occasion, samples will be either recycled or rejected. The effects on the measurement data have been characterized in the literature (Halevi and Ray 1988) (Ray and Halevi 1988) by two particular phenomena, Vacant Sampling and Message Rejection.

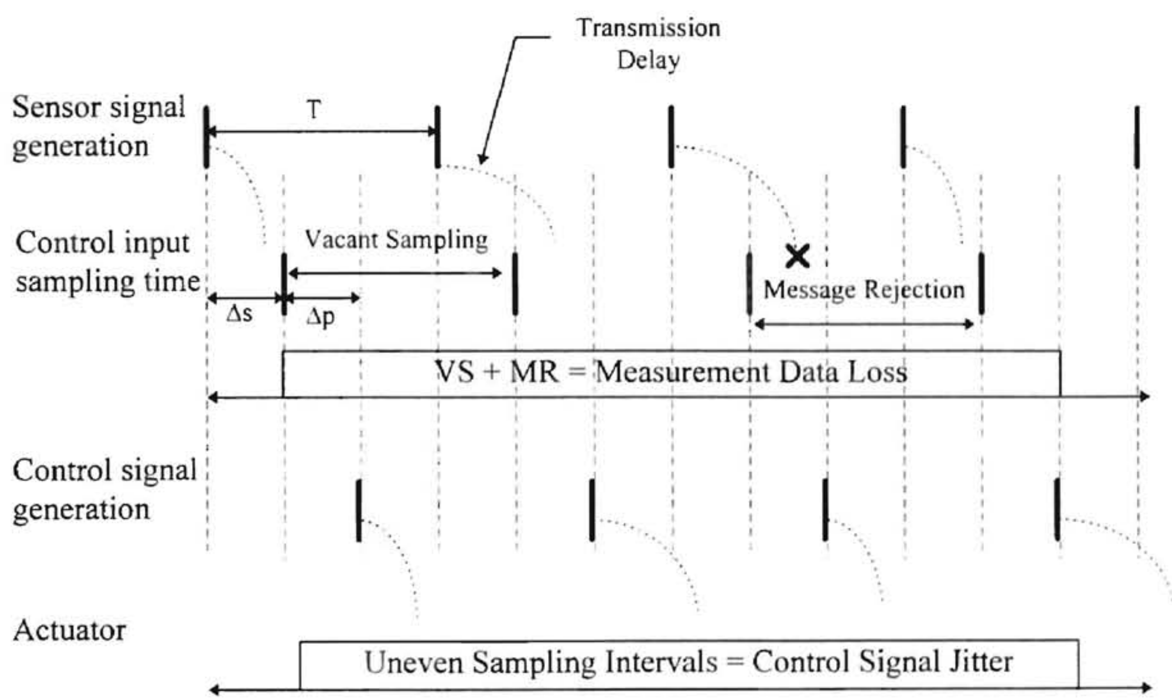


Fig. 3. The Variable Transport Delay in a Sampled Data System

Vacant sampling, is the case where, no fresh sensor message arrives at the controller during its j^{th} sampling period. The old sensor data is used at the $(j+1)^{st}$ sampling instant for computing the control signal so that, $z_{j+1}=y_j$. Vacant sampling only occurs when an on time delivery is followed by a late delivery. Message rejection, is the case where, two sensor messages arrive at the controller during its j^{th} sampling period. The former sensor data is discarded and the latter arrival is used for the computation of the control signal so that, $z_{j+1}=y_{j+1}$. Message rejection only occurs when a late delivery is followed by an on time delivery. The process of samples being recycled is named data loss and illustrated in Figure 4.

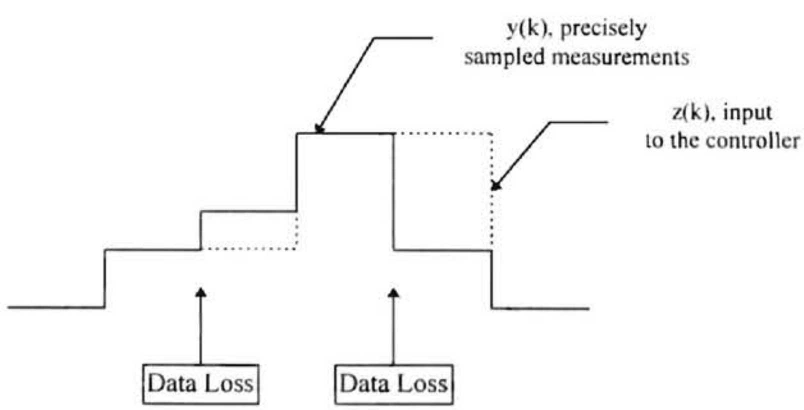


Fig. 4. Data Loss on the Measurement Signal

The effects of the data latency on the control signal are different since the delayed samples from the controller are fed into a continuous system. As a direct consequence of the variable network delays, the plant then receives samples with a variable time interval. This effect is designated as jitter at the input of the process and is illustrated in Figure 5.

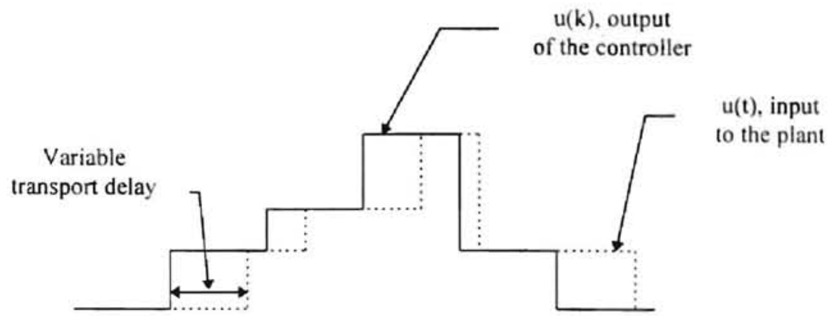


Fig. 5. Jitter on the Control Signal

2.2. Characteristics and Impact of the Network Data Latency

Toward completely digital distributed control systems, peer-to-peer network communications between smart transducers seems to be the most logical communication scheme. The principal feature of a smart transducer is its ability to off-load some of the processing functions from the control system. Clearly, complete distribution of data acquisition and control functions within the field devices is expected to be geographically distributed. Therefore, intelligent field devices are no longer limited to the task of making measurements and driving final control elements. One is then to expect that, the traffic induced by a control system application is both periodic and aperiodic. The periodic traffic is typically time critical and generated by feedback control loops. The aperiodic traffic is composed of control and non-control information. The aperiodic control information may be viewed as one-time messages to announce an alarm, often in relation to an emergency. The aperiodic non-control information is the result of the augmented role that smart instruments play. This information generally contributes to the management of the whole system but does not have a real-time significance. The characteristic of the data latency is undoubtedly an important factor in the design of a compensation algorithm. The data latency is significantly affected by the intensity and the distribution of the network traffic. Extensive simulations have been made where the bus load was generated for random traffic with Poisson arrival and exponentially distributed message length (Wang, Lu, Hedrick and Stone 1992), (Tindell, Burns and Wellings 1994). These simulations and analytical results may give the control system designer some idea about the relationship between data latency, bus load and priority setting. The assumption of randomness of the bus traffic and message generation rate is however questionable. Indeed, in this particular case, This cannot be all true since periodic traffic is introduced by the multiple control loops operating with a fixed sampling time on the network. Halevi and Ray (1988) mentioned that, the network traffic can be approximately periodic. Further, for certain applications, such as token bus, the delay sequences have been shown to be periodic. Specifically, Ray (1987) showed an actual profile of queuing delays for the SAE linear token bus that exhibits a triangular wave form pattern at steady state. Clearly, the characteristics of the bus load is dependent on the number, the type of stations, and possible misynchronizations. If the number of stations is small, with a combination of periodic and sporadic transmitters, the resulting bus load will conserve its periodic characteristic. The outcome is a traffic that is qualified as quasi-periodic. Halevi and Ray (1988) therefore suggested that, as a first step in the analysis, one should consider periodic traffic and accommodate for quasi-periodic traffic in a second instance. Ray and Halevi (1988) also indicated that the variations in the network traffic pattern are usually slow relative to the dynamics of the control system. Consequently, because the characteristic of the bus traffic is not randomly distributed (Ludvigson 1990), average performance calculations may not provide an adequate method for evaluating bus performances for distributed control systems. Consequently the characteristics of the network data latency are bounded to be time varying. Nonetheless, very little information exists toward the identification of these characteristics.

In the context of a control loop, it is reasonable to assume that the transport delay is by design bounded by one sampling period. Delays of more than one sampling period are not of much practical significance, firstly because the network design should not allow such overload even under the worst conditions. Secondly, any unbounded delay goes against the concept of closed loop operation. Accordingly, neither vacant sampling nor sample rejection can happen twice in succession. In other words, the occurrence of vacant sampling implies that only sample rejection can happen next. Truly, the problem is time varying but discrete in behavior. These observations enable us to define four possible cases over a two sampling instant window. Case #1, the first sample is late the second sample is on time (vacant sampling), Case #2, the data at both instants are on time, Case #3, the first sample is on time the second sample is late (vacant sampling), Case #4, the data at both instants are late. Cases 1, 2, 3 and 4 can be seen in their respective order by taking the sensor sampling instants by pair starting from the left on Figure 3. Truly, vacant sampling is the constraining element as the measurement delay is increased and the controller has to recycle old data. As an illustration of the effects of the network induced time varying delays, the feedback loop situation of Figure 2 has been simulated. The nominal continuous time plant model is chosen to be,

$$G(s) = \frac{1000}{s(0.5s+1)} \tag{1}$$

and a proportional state feedback controller is designed using Akermann's formula in order to generate a system response with a desired damping of 0.7 and a desired natural frequency of 56 rad/s. The observer poles are chosen to be 3 times faster than the system poles. Within the simulation algorithm, the sampling of the feedback loop has been set to 0.01 second, however, the dynamics of the plant is solved at a sampling rate of 0.001. In other words, the output of the plant is solved 10 times in between each controller sampling instant to simulate continuous operation of the plant. The data loss on the measurement signal is a sequence that repeats according to the data loss vector [0 0 1 0 1 0], where 0 symbolizes an on time transmission and 1, the loss of one sample. The delay on the control signal is also a repeating sequence driven by the jitter vector, [.1 .5 .2 .7 .1 .8], where the numbers represent a fraction of the sampling period T=0.01. There is no documentation available on the characteristic of the network data latency in the context of distributed control systems. Therefore, these two sequences have been arbitrarily chosen to generate a sufficient alternation and yield an average close to 0.4 T. The following three figures display the output of the control system subjected to distributed delays in the condition of the simulation described above. The resulting responses have been obtained by setting the initial states to zero and tuning the set point in order to get a steady state of one. On Figure 6, The continuous line is the nominal step response and the dashed line represents the step response when only the control signal is subjected to jitter. Figure 7 shows the same response when only data loss is affecting the system. Figure 8 exemplifies the compounded effect of both control jitter and data loss on the response of the system.

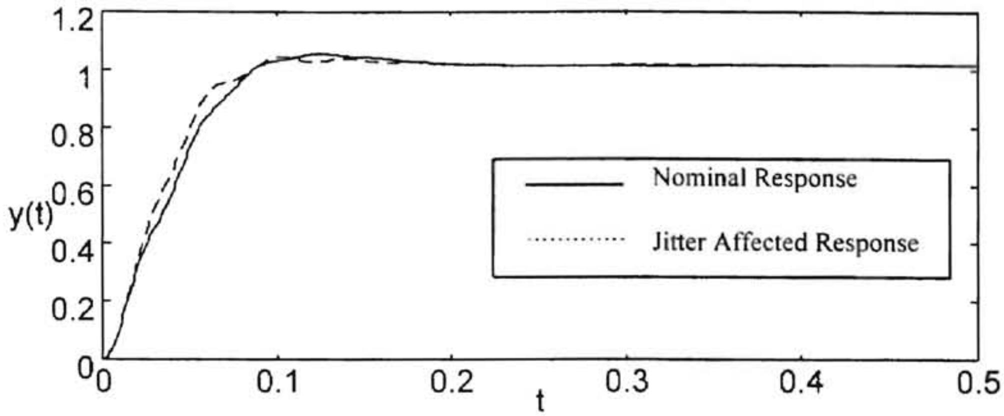


Fig. 6. Effect of Jitter in the Control Signal on the Output of the System

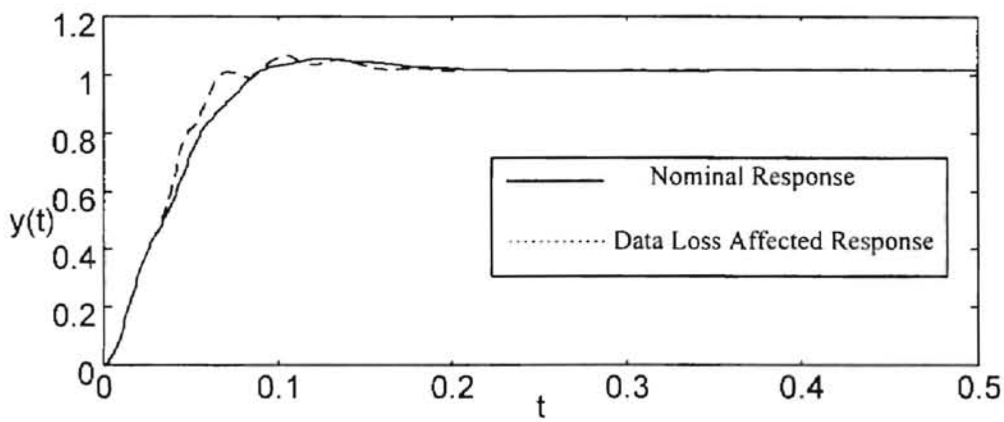


Fig. 7. Effect of Data Loss in the Measurement Signal on the Output of the System

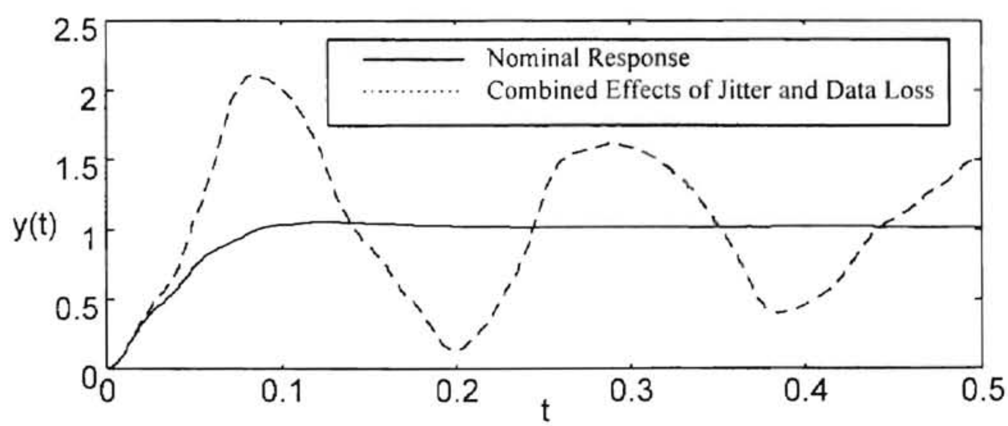


Fig. 8. Combined Effect of Jitter and Data Loss on the Output of the System

Qualitatively speaking, it has been observed on the above example and also has been indicated by Torngren (1996) that, a sudden change in the measurement delay (vacant sampling) introduces a disturbance in the system to which the controller will attempt to provide compensation. The compensated control signal, however, introduced a new disturbance in the system when the delay returns to its original value (sample rejection). More to the point, the detrimental effects are directly proportional to the speed of the controller. For instance, in Figure 9 only the natural frequency of the desired response has been relaxed to 21 rad/s leading to a lower gain controller. The response is indeed slower as the rise time is almost doubled, but, at the same time the sensitivity to variable delays in the feedback loop is reduced.

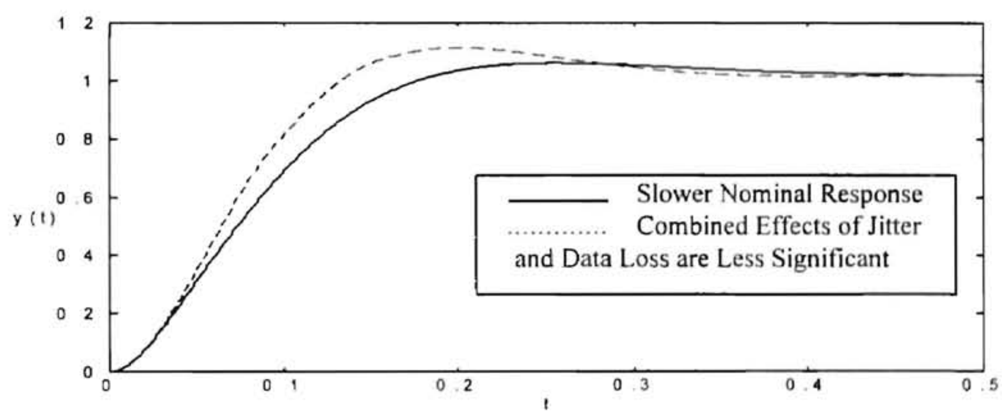


Fig.9. A Slower Controller on the Situation of Figure 8.

In an attempt to quantify the effect of data loss and jitter, the square of the difference between nominal and actual response has been summed up from 0 to 0.5 second. The result is a performance index that is inversely proportional to the quality of the control function. That performance index has been used to asses the effects of the loss of only one sample on the whole step response. Four different trials have been run with that one particular sample lost at the respective instant, .05, .07, .09, and .011 seconds. The resulting normalized performance index for these four trials yielded the values 1, .18, .13 and .004. This shows that the location were the data loss occur is of major significance. Moreover, it has been observed that adding jitter to the same particular situations typically increased the performance index approximately and consistently by a factor of 14.

From these results and more individual experimentation we make the following points,

- The compounding effects of jitter and data loss are much larger than the summation of the respective individual effects.
- The controller speed is a determinant factor to the sensitivity toward time varying delays.
- The location along the step response where a sample is lost is of major significance to the resulting control performance.
- Although not illustrated here, it has been observed in this particular simulation that the observer dynamics have relatively little effect on the response.

2.3 Dynamic characterization of a delayed control systems

A time-delay is a system which delays a signal but otherwise does not change it. In the s-domain it is characterized by the transfer function, $e^{-\tau s}$. The consequence is very detrimental to a control system by the way of introducing phase lag into the system. The gain of a time delay is constant and equal to one and the phase lag grows exponentially with frequency. The resulting effect on control systems is to decrease the stability margin. A stable system is therefore difficult to design particularly if a high feedback gain is desired. Elements of the characterization of delayed control system are presented in this section. As defined before, the delays introduced by the multiplexed data network are variable. The majority of the reported work deals with constant delays, but modeling and compensation of the time varying delays cannot readily be done in discrete-time control systems. One has therefore to characterize the specifics of the situation at hand, namely the feedback control loop closed via the communication network, to construct a compensation solution.

Time delays are difficult to handle in the continuous-time case as the transfer function becomes non-rational. A formulation of the system by ordinary differential equations is no longer possible. The dynamic characterization of the system has then to be done by a system of delay-differential equations. The delay-differential equations (DDE) are a special class of differential equations called functional differential equations. The application of DDE's to control system design is extensively explained by Oguztoreli (1966). In a DDE, the derivative of an unknown function, x, has a value at t that is related to x as a function of some other function at t. For example:

$$\dot{x}(t) = A\,x(t-\tau) \qquad \tau > 0$$

Or more generally:

$$\dot{x}(t) = f(t, x(.)) \tag{2}$$

It is interesting to note that, since a DDE can describe a process with after effects (previous history of the system), the initial data now includes an initial function ϕ in place of an initial condition. The solution is therefore: $x(t)=x(t,t_0, \phi)$ with I.C.= $\phi(t_0)$ and the initial situation has to be seen according to Figure 10.

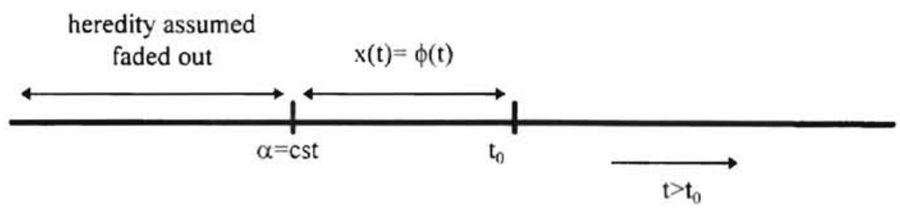


Fig. 10. Initial Situation of a DDE

For control systems, even with variable delays a suitable model equation could be viewed as,

$$\dot{x}(t) = f(t, u(t), d(t), x(.)) \tag{3}$$

where: u(t) is the control signal

d(t) is the disturbance signal

x(.) is a function defined on $[\alpha, t]$

In this case the change in the state x is affected by u(t), d(t), but is also dependent upon how x is affected by some mechanism x(.). The solution can exhibit exponential growth or decay but can also be oscillatory. DDE's also make the case for a possible optimization problem, namely, trying to maximize performances by choosing the best {ϕ , u} pair. A good example of the use of DDE's to the analysis of control systems can be found in a work by Hirai and Satoh (1980). In that short paper a first order system is written by,

$$\dot{x}(t) + \alpha\ x(t) + \beta\ x(t - L(.)) = 0 \tag{4}$$

where L(.) is a variable delay which is a function of time t. L(t) is chosen to be a particular arbitrary "saw tooth" delay wave form. Proof has been made that the time varying system is unstable even if the time invariant system is stable.

Time delays can also be studied in continuous-time using analytical approximations. We consider here three approximations studied by Wang, Lundstrom and Skogestad (1994). The approximations correspond to the power series expansions of respectively the numerator (*zero*), denominator (*pole*) and a combination of the two referred to as, the *Pade* approximation.

$$\text{Zero, } e^{-\tau s} \approx 1 - \tau\ s \tag{5}$$

$$\text{Pole, } e^{-\tau s} \approx \frac{1}{1 + \tau\ s} \tag{6}$$

$$\text{Pade, } e^{-\tau s} \approx \frac{1 - \frac{\tau}{2} s}{1 + \frac{\tau}{2} s} \tag{7}$$

The relative accuracy of these approximations has been qualified by Wang, Lundstrom and Skogestad (1994) and compared based on how well they predict the smallest delay required to destabilize a certain closed loop system. The outcome is, *zero* is always conservative while *pole* is overestimating the stability margin. *Pade* is clearly the best approximation while still a bit optimistic. They also pointed out that, theoretically, it is possible to get arbitrary high accuracy by dividing the delay in n parts $\left(e^{-\frac{\tau}{n}s}\right)^n$ and use any approximation for each of the smaller delays $e^{-\frac{\tau}{n}s}$.

A delayed control system is easier to handle in discrete-time. The approach is called state augmentation and has been described in Astrom and Wittenmark (1990) and Franklin and Powell (1994). Let the system be described by,

$$\dot{x} = Ax(t) + Bu(t-\tau) \tag{8}$$

The general solution is,

$$x(t) = e^{A(t-t_0)}x(t_0) + \int_{t_0}^{t} e^{A(t-s)}Bu(s-\tau)ds \tag{9}$$

If we let t_0 =kT and t= kT + T, then,

$$x(k+1) = e^{AT}x(k) + \int_{kT}^{kT+T} e^{A(k+1-s)}Bu(s-\tau)ds \tag{10}$$

The next step consist in breaking the integral into two parts as follow,

$$x(k+1) = e^{AT}x(k) + \int_{kT}^{kT+\tau} e^{A(k+1-s')}Bds'u(k-1) + \int_{kT+\tau}^{kT+T} e^{A(k+1-s')}Bds'u(k) \tag{11}$$

In this form, the control signal u is constant in each part and sampling of the continuous system gives,

$$x(k+1) = \Phi x(k) + \Gamma_0 u(k) + \Gamma_1 u(k-1) \tag{12}$$

where,

$$\Phi = e^{AT}, \quad \Gamma_0 = \int_0^{T-\tau} e^{As} ds\, B, \quad \Gamma_1 = e^{A(T-\tau)} \int_0^{T} e^{As} ds\, B \tag{13}$$

The state space model is therefore,

$$\begin{bmatrix} x(k+1) \\ u(k) \end{bmatrix} = \begin{bmatrix} \Phi & \Gamma_1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x(k) \\ u(k-1) \end{bmatrix} + \begin{bmatrix} \Gamma_0 \\ I \end{bmatrix} u(k) \tag{14}$$

Now if the delay is longer than T, one needs to separate the system delay into an integral number of sampling periods plus a fraction, such that,

$\tau = lT - mT \quad with, \quad l \geq 0 \quad and, \quad 0 \leq m \leq 1$

Equation (12) than becomes,

$$x(k+1) = \Phi(x(k) + \Gamma_0 u(k-l+1) + \Gamma_1 u(k-l) \tag{15}$$

Thus the state space model is therefore,

$$\begin{bmatrix} x(k+1) \\ u(k-l+1) \\ \vdots \\ u(k-1) \\ u(k) \end{bmatrix} = \begin{bmatrix} \Phi & \Gamma_1 & \Gamma_0 & \cdots & 0 \\ 0 & 0 & I & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & I \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} x(k) \\ u(k-l) \\ \vdots \\ u(k-2) \\ u(k-1) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ I \end{bmatrix} u(k) \tag{16}$$

This state augmentation approach consists of incorporating the delay in the plant model (namely A, B and C). Provided that the sampled system is reachable, the dynamics of the augmented system can be controlled with linear state feedback. One can then synthesize the controller gains that must accommodate for the loss of phase margin due to the delay. The new model that incorporates these delays rapidly becomes cumbersome and complete controllability and observability may be lost as additional states are added. It is therefore more convenient to keep the induced delays outside the plant model. This alternative approach has lead to the multistep prediction scheme presented in section 3.1.

Along the lines of state augmentation, a methodology for the characterization and compensation of delays both in the input and output variable has been developed by Halevi and Ray (1988). Halevi and Ray assumed the situation presented in section 3.2, the sensor data is subject to a transport delay that generates, vacant sampling and sample rejection. Accordingly, z, the delayed measurement data is, $z_j = y_{j-p(j)}$ where y is the sensor data and p(j) a non-negative integer bounded by μ. In addition, the control input data is also subject to a time-varying delay. This implies that even if the controller generates the command at a constant rate, the interval between their successive arrival at the actuator is not constant. The model therefore uses the facts that the input to the process is piecewise constant. In other words u(t) assumes at most (l+1) different values in the interval [kT,(k+1)T) where changes occur at the instants $kT+t_i^k$, with i=1,2,3,...,l. Thus, the solution of the state equation can be reformulated accordingly,

$$x(k+1) = e^{AT}x(k) + \int_0^T e^{A(T-s)}Bu(s)ds = e^{AT}x(k) + \sum_{i=0}^{l} B_i^k u(k-i) \qquad (17)$$

where,

$$B_i^k = \int_{t_i^k}^{t_{i-1}^k} e^{A(T-s)} B ds \quad with \quad t_{-1}^k = T, \quad t_l^k = 0 \tag{18}$$

For simplicity, the control law is chosen to be purely proportional such that,

$$u_k = K(r_k - z_k) \tag{19}$$

Where r_k is the reference signal, and z_k the delayed sensor data as defined previously. The state vector of the augmented state system is then,

$$X_k = \begin{bmatrix} x_k & y_{k-1} & \cdots & y_{k-\mu} & u_{k-1} & \cdots & u_{k-l} \end{bmatrix}$$

Ray and Halevi's work (1988) presents a generalized formulation of the state augmented model. A simplified version of this model is given here, in a sense that the control law is proportional feedback, the reference signal is constant and equal to zero, the bound on the variable delay is equal to one and the control signal can assume only three different values per sampling period (L=2). The compensated system then becomes,

$$\begin{bmatrix} x_{k+1} \\ y_k \\ u_k \\ u_{k-1} \end{bmatrix} = \begin{bmatrix} \Phi - B_0^k K(1-p(k)) & -B_0^k Kp(k) & B_1^k & B_2^k \\ C & 0 & 0 & 0 \\ -K(1-p(k)) & -Kp(k) & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_k \\ y_{k-1} \\ u_{k-1} \\ u_{k-2} \end{bmatrix} \tag{20}$$

where,

$$B_0^k = 1 - \exp(t_0^k - T), \quad B_1^k = \exp(t_0^k - T) - \exp(t_1^k - T), \quad B_2^k = \exp(t_1^k - T) - \exp(-T)$$

This compensation algorithm has been simulated on a simple system by Ray and Halevi (1988). Let us recall that the time skew Δs between measurement and controller acquisition of the input data determines the distribution and proportion of vacant sampling and sample rejection. In Ray's work, a series of simulations were conducted for different combinations of Δs and K to result in the definition of a stability region. Further, the simulations showed that the stability analysis cannot be made solely on the basis of average values of the time varying delays. Indeed, three different delay sequences of the same average characteristic revealed different stability ranges.

CHAPTER 3

REVIEW OF THE LITERATURE TOWARD ANALYSIS AND COMPENSATION

The first concept is the so called p-step delay compensation observer. The compensation scheme eliminates the time variations by mean of buffering. The resulting time invariant delay of several sampling periods as seen from the output of the buffer is compensated for by a multistep prediction method. Encouraging simulation results later prompted the development of a Loop Tranfer Recovery (LTR) synthesis method to identify an observer gain that minimizes the H_2 norm of the sensitivity error transfer matrix. The second concept reported in section 3.2 from Ray (1994), is also along the lines of the situation exposed in 1988, and consists of an output feedback control law in a stochastic setting. In this case the delays are considered to be stochastic and knowledge of the their probability repartition is assumed to be available. Finally, section 3.3 reports the work of Lundstrom and Skogestad (1994) toward the construction of an uncertainty model suitable for robust control analysis.

3.1 The P-step Observer

With the idea that it is more convenient to keep the induced delays outside the plant model, Luck and Ray (1990) proposed to eliminate the time variations in the network transport delay using buffering and estimation. In this concept, the idea is to monitor the data when it is generated and to keep track of the delay associated with it. The time variations are eliminated by buffering the data and always presenting samples that have the same age at the controller. In this conditions the control function only has to compensate for a fixed amount of delay. Note that, the delayed data is used regardless of whether the current data is available. If more fresh data is available, it has to be kept in a first-in first-out buffer for later processing. Referring to the control structure shown, in Figure 11, the network delays τ_{sc} and τ_{ac} are bounded by r and q units of sampling period respectively. The data is kept in a first in-first out buffer to absorb the time variations both at the controller and actuator sites. Accordingly, the model can now be treated with two constant time delays r and q respectively in place of the variable network delays previously τ_{sc} and τ_{ac}.

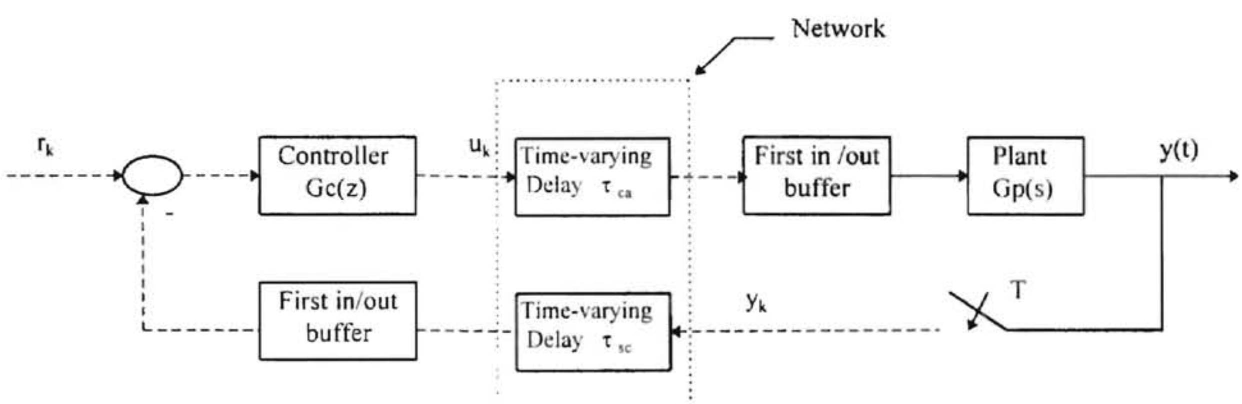


Fig. 11. Elimination of the time variations by data buffering.

The control problem is therefore one of compensation of a constant delay of multiple sample interval. For that aim the algorithm proposed by Luck and Ray (1990) consists of using an observer to estimate the delayed states and then predict the current state using the state variable model of the plant recursively. The basic equations used in the multistep delay compensation scheme are listed below, consider the plant,

$$x_{k+1} = Ax_k + Bu_k; \quad y_k = Cx_k \tag{21}$$

The observer model is,

$$z_{k+1/1} = Az_{k/1} + Bu_k + L_k(y_k - Cz_{k/1}) \tag{22}$$

The p-step predictor is,

$$z_{k+1/p} = Az_{k/p-1} + Bu_k \tag{23}$$

With the predictive control law,

$$u_k = \Gamma_k z_{k/p} \tag{24}$$

Where,

$z_{k/r}$ = prediction of x_k based on the measurement history $\{y_{k-r}, y_{k-r-1}, \ldots\}$

Luck and Ray (1990), have implemented and verified the resulting closed loop equations,

$$\begin{bmatrix} x_{k+1} \\ e_{k-p+1} \end{bmatrix} = \begin{bmatrix} A + B\Gamma_k & -B\Gamma_k\Theta_k \\ 0 & (A - L_{k-p}C) \end{bmatrix} \begin{bmatrix} x_k \\ e_{k-p} \end{bmatrix} \tag{25}$$

where,

$$e_k = x_k - z_{k/1} \tag{26}$$

and,

$$\Theta_k = \begin{cases} \prod_{j=1}^{p} \left(A - L_{k-p-j-1}C\right) + \sum_{i=1}^{p-1} \left(A^{i-1} L_{k-i} C \prod_{j=1}^{p-i} \left(A - L_{k-p+j-1}C\right) \right) & if\ p \geq 2 \\ I & if\ 0 \leq p < 2 \end{cases} \tag{27}$$

Schematic representation of a three-step predictor/controller is given in Figure 12.

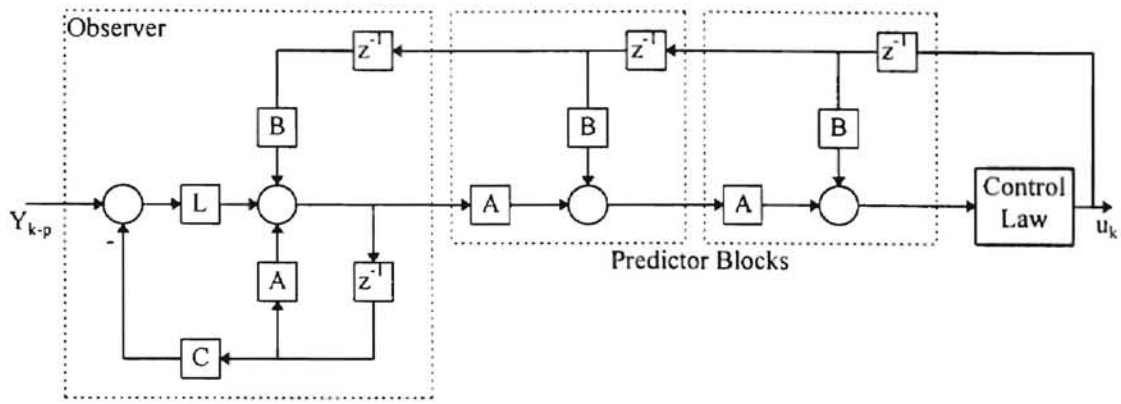


Fig. 12. Schematic representation of a three-step predictor/controller

The number of predicted steps could be obtained as the sum of the specified bound, namely p=r+q, if the joint statistics of r and q are known, p could be computed more precisely. Extensive simulations of the compensation scheme can be found in Luck and Ray (1994), the experimental results come from the a d.c. motor assembly interfaced to a microcomputer. Data from the A/D converter is stored in a buffer to generate a constant delay. A proportional-integral (PI) controller is used with and without the delay compensation algorithm. In that particular paper, the authors clearly showed that the dynamics of the motor is considerably improved with the observer. Further, the predictive properties of the observer-based control algorithm were able to cancel a limit cycle problem happening at low reference input of the uncompensated system. The delay compensator has also been investigated for a simulated flight control system within a network environment. The design showed superior performance but also yielded a steady state error. Indeed the observer based controller produces a small steady state error because the integrator acts upon the estimate of the state and not on the true system output. Some of the robustness issues of the delay compensator for structured uncertainties have been covered. Nevertheless, the effect of the modeling uncertainty upon the performance of the predictor controller have been studied using the gain matrix that were originally designed for the non-delayed system. Toward robust compensation, Shen and Ray (1993) proposed to synthesize the control system for delay compensation by extending the concept of loop transfer recovery (LTR) to the multistep observer described in Luck and Ray (1990). The uncertain communication delays are lumped at the plant input in the form of an input multiplicative term. The idea is to tune the loop transfer recovery matrix such that the error transfer matrix, the difference between the actual and target sensitivity matrix, is minimized. Let us initially consider the LTR concept for the regular one step observer. In a first stage, the target loop transfer matrix is designed by selecting full-state feedback gain for a given performance index. Next, the loop transfer matrix and the sensitivity matrix of the compensator are calculated by incorporating an observer in the loop.

Consider the plant,

$$x_{k+1} = Ax_k + Bu_k, \quad y_k = Cx_k \tag{28}$$

with the full-state feedback law,

$$u_k = -Fx_k \tag{29}$$

The plant transfer matrix is then,

$$G(z) = C\Phi(z)B, \quad \text{where } \Phi(z) = (zI - A)^{-1} \tag{30}$$

The target loop transfer matrix is,

$$H(z) = F\Phi(z)B \tag{31}$$

The target sensitivity is,

$$S(z) = [I + H(z)]^{-1} \tag{32}$$

The observer introduced in the loop to compensate the one step delay has the transfer matrix,

$$G_1(z) = F(zI - A + BF + LC)^{-1}L = F[I + \Phi(z)(BF + LC)]^{-1}\Phi(z)L \tag{33}$$

The loop transfer matrix with the compensating observer then becomes,

$$\begin{aligned} L_1(z) = G_1(z)G(z) &= F[I + \Phi(z)(BF + LC)]^{-1}\Phi(z)LC\Phi(z)B \\ &= [I + E_1(z)]^{-1}[H(z) - E_1(z)] \end{aligned} \tag{34}$$

where $E_1(z)$ is the 1-step error matrix with,

$$E_1(z) = F[zI - A + LC]^{-1}B \tag{35}$$

The resulting 1-step sensitivity matrix is,

$$S_1(z) = [I + L_1(z)]^{-1} = [I + H(z)]^{-1}[I + E_1(z)] \tag{36}$$

Finally the relative sensitivity error is,

$$E_1(z) = S(z)^{-1}[S_1(z) - S(z)] \tag{37}$$

The same reasoning as been applied in Luck and Ray (1990) to the p-step delay compensator. the loop transfer matrix is derived first. Then, the error of the sensitivity matrix relative to that of the target loop is formulated. The approach used in the synthesis of the p-step delay compensator is to minimize the loop recovery error where the gain of the observer is set to a prescribed value. In other words, the key is to identify an L that minimizes the relative sensitivity error transfer matrix. The procedure is then as follows.

First, the target loop is designed assuming no delay with full state feedback. Second, the observer gain L is calculated by solving the steady state Riccati equations for a fictitious measurement noise ρ, where ρ is a tunable scalar parameter. This parameter ρ is then adjusted so that the maximum singular value of the compensated loop transfer matrix are below those of the target loop transfer matrix to satisfy the requirement for stability robustness. The minimum singular values of the compensated loop transfer matrix represent the lower bounds of the performance. As expected, Ray observed that it is impossible to fully recover the target loop characteristics by tuning the observer gain when a predictive state estimator is used. In addition, for a fixed requirement on the stability robustness, performance decreases as the compensated delay is increased.

3.2 Optimal Compensation with Stochastic Delay Assumption

In the design presented in section 3.1, the state estimate was consistently obtained on past measurements regardless of whether the sensor data is delayed or not. It is similar to having the measurement data always delayed by a specific number of samples. However if the probability of delayed arrival of measurement data is small, then the cost of introducing a constant delay may be excessive. As an alternative, Ray Liou and Shen (1993) and Ray (1994) proposed an estimation algorithm that uses the most up-to-date sensor data at each sampling instant. Consequently output feedback control under randomly varying distributed delays has been formulated as an alternative to the deterministic approach used in the p-step observer. The control system under consideration is as described in section 2.1 and illustrated on Figure 2. The major assumption is that the statistics of the induced delays are white and independent, and knowledge of the their probability repartition is assumed to be available. The approach consists of a combined state estimation and state feedback. The state estimation and state feedback control laws are synthesized separately on the principle of optimality and then integrated together. The state estimation filter is formulated when the sensor data arrival is either timely or delayed by one sampling period and the probability of vacant sampling is made to be very small. Based on the concept of Linear Quadratic Gaussian, (LQG) the estimation filter assumes that, the plant is subject to random disturbances, the sensor data is contaminated with noise and the measurement delay is a random sequence from the set $\{0,1\}$. The regulator follows the structure of the conventional linear quadratic regulator (LQR) control law and is formulated by Ray (1994) in the presence of randomly varying delays from the controller to the actuator and full state feedback. Accordingly, the control architecture is reported from the referenced publications in two parts. The estimator is presented first and accounts for the variable delays, plant noise and sensor noise. The regulator then assumes full state feedback and variable delays in the control signal but no plant and measurement noise.

The state estimator illustrated here has been extracted from the work of Ray, Liou and Shen (1994). It is a modification of the conventional minimum variance state estimator to account for the effects of sensor to controller randomly varying delays. The plant model is expressed as.

$$\xi_{k+1} = \Phi_{k+1,k}\xi_k + w_k$$

$$y_k = H_k\xi_k + v_k \tag{38}$$

$$z_k = (1-p_k)y_k + p_k y_{k-1}$$

where,

ξ_k the plant state at instant k

p_k the random delay from the set $\{0,1\}$.

w_k the noise vector

z_k is the delayed sensor data (either on time $z_k=y_k$ or late $z_k=y_{k-1}$)

If the measurement history Z_j up to the j^{th} instant is available, the conditional estimate is,

$$\hat{\xi}_{k|j} = E\{\xi_k | Z_j\} \text{ for } j<k \tag{39}$$

the state estimation error is,

$$e_{k|j} = (\hat{\xi}_{k|j} - \xi_k) \text{ for } j<k \tag{40}$$

and the conditional error covariance is,

$$\Sigma_{k|j} = E\{e_{k|j}e_{k|j}^T | Z_j\} \tag{41}$$

Note that E{.} represents the expectation value with respect to the statistic of the plant noise, sensor noise and the variable delay sequence. The problem of finding an optimal estimate of the state is solved by minimizing the following cost functional at each sensor sampling instant.

$$J_k = E\{e_{k|k}^T e_{k|k}\} \tag{42}$$

The objective is therefore to synthesize a sequence of filter gain matrices, $\{K_k\}$ for k=1,2,3..., that would minimize the cost functional for each k. This is accomplished by considering the recursive structure,

$$\begin{aligned} \hat{\xi}_{k} &= L_k \hat{\xi}_{k|k-1} + K_k z_k \\ \hat{\xi}_{k-1} &= \Phi_{k,k-1} \hat{\xi}_{k-1|k-1} \end{aligned} \tag{43}$$

where, the gain matrices L_k and K_k of the above filter are derived in a sequel.

The linear quadratic regulator is compensated for control signal delays varying but bounded. Similarly to the methodology of state augmentation presented before from Halevi and Ray (1988), the input to the plant is considered piecewise constant during a sampling interval in the controller frame to take into account the effects of controller to actuator delays. The augmented state vector is then composed of the plant states plus the control input at discrete instant of time,

$$x_k = [\xi_k^T \; u_{k-1}^T]^T \quad \in \Re^{n+m} \tag{44}$$

where,

$\xi \in \Re^n$ is the plant states

$u_k \in \Re^m$ represent u(t) at discrete instants of time

Following the same methodology introduced in section 4.2.(Halevi and Ray 1988), the augmented plant model is then,

$$A = \begin{bmatrix} \Phi((k+1)T, kT) & b_k^1 \\ 0 & 0 \end{bmatrix}; \quad B = \begin{bmatrix} b_k^0 \\ I_m \end{bmatrix} \tag{45}$$

where,

$$b_k^0 = \int_{kT+t^k}^{(k+1)T} \Phi[(k+1)T,\lambda]d\lambda \tag{46}$$

$$b_k^1 = \int_{kT}^{(k+1)T} \Phi[(k+1)T,\lambda]b(\lambda)d\lambda - b_k^0$$

Note that relation (46) is a stochastic process because the time period t^k limiting the integration are random. The relationship for optimal control is derived recursively by minimizing the following cost functional over a finite-time horizon of N sampling intervals,

$$J_k(x_k,u_k) = \frac{1}{2}\left[x_k^T Q_k x_k + u_k^T R_k u_k\right] + \{E\{J_{k+1}^{\bullet}(x_{k+1}|x_k)\} \tag{47}$$

where,

$$J_k^{\bullet}(x_k) = J_k(x_k,u_k^{\bullet}) \tag{48}$$

and $u_k^{\bullet}$ is the optimal state feedback law at the k[th] sample.

For k=N, we assume that the terminal state is reached and there is no need for any control. Therefore,

$$J_N^{\bullet}(x_N) = J_N(x_N,u_N^{\bullet}) = \frac{1}{2}x_N^T P_N x_N \tag{49}$$

where $P_N = S$, the final state penalty matrix, is given. Practically, the optimal control law is given via a recursive relationship. Let the control law at the k[th] stage be

$$u_k^{\bullet}(x_k) = -F_k x_k \tag{50}$$

for k<N and the resulting performance cost

$$J_k^{\bullet}(x_k) = \frac{1}{2}x_k^T P_k x_k \tag{51}$$

where

$$F_k = \left[R_k + E\left\{B_k^T P_{k+1} B_k\right\}\right]^{-1} E\left\{B_k^T P_{k+1} A_{k+1,k}\right\} \quad (52)$$

$P_k = Q_k + E\left\{A_{k+1,k} P_{k+1}\left(A_{k+1,k} - B_k F_k\right)\right.$ with P_N=S and each equation is evaluated backward starting from N-1. The control law is computed off-line to numerically obtain a steady-state value of the gain matrix F on a finite-time horizon. The expected values are numerically generated based on the known probability distribution of the controller-actuator delay.

The integration of the state estimator and the state feedback controller is now executed. The estimate obtained with the predictive filter replaces the actual plant states part of the augmented state vector in the formulation of the optimal control law. The combined state estimation and state feedback control law is then obtained by changing the plant states in the augmented state vector of the full state feedback control law by their estimated counterpart. The feedback control law is then,

$$u_k(Z_k) = -F_k[\hat{\xi}_{k|k}^T u_{k-1}^T]^T \quad (53)$$

for k<N and where Z_k is the history of the delayed measurements used for generating the control signal and F_k is the state feedback control gain given in equations (52).

The proposed compensation technique has been simulated by Ray (1994) on the unstable longitudinal motion dynamics of an advanced aircraft. It been found that even though the delay compensated LQG algorithm only offers a suboptimal solution when variable delays are present, the state estimator together with the feedback regulator are capable of compensating for randomly varying delays. The control performance generally degraded with larger noise covariance but no evidence of instability was found. Finally the delay compensated regulator was found to be sensitive to both structured and unstructured plant uncertainties. As pointed out by Ray (1994), This is expected because LQG has poor robustness properties and the injected delays further deteriorated the robustness of the compensated regulator.

3.3 Robust Analysis of Time Varying Delays

All networked distributed control systems have to be digitally implemented. Consequently, it is logical to design the compensation mechanism on the basis of the discrete time sampled data system. This approach has been exemplified in the previous sections by explicitly considering periodic operation and zero order hold. Nevertheless, even though continuous time analysis is not really relevant to the case of networked distributed control systems, it may bring additional elements of understanding to the problem of variable delays. Furthermore, a continuous system can still be translated to a discrete time implementation with sufficient over-sampling. Robust design for instance, is easier to study in continuous time. In μ synthesis and H_∞ design, it is interesting to view uncertain or time varying delays as model uncertainties. Therefore, in this section, attention will be given to the translation of variable delays into an uncertainty model.

In section 2.3 rational approximations have been introduced for time delays, they were respectively called, *zero*, *pole* and *Pade*. These approximations can also be used to represent a delay uncertainty of the type,

$$e^{-\tau\Delta s}, \quad -1 \le \Delta \le 1 \tag{54}$$

Wang, Lundstrom and Skogestad (1994) indicated that, to represent a complex perturbation only *zero* can be used since *pole* and *Pade* would produce an infinite uncertainty for frequencies above respectively $1/\tau$ and $2/\tau$. Another alternative, commonly used in robust design is to let the Δ uncertainty in the denominator of the Pade approximation equal 1 and let the remaining Δ become complex, such that,

$$e^{-\tau\Delta s} \approx 1 - \frac{\tau s}{1+\frac{\tau}{2}s}\Delta = 1 + w(s)\Delta \qquad |\Delta| \le 1 \tag{55}$$

Then the uncertain plant can be written,

$$g_p(s) = g(s)(1 + w(s)\Delta(s)) \qquad |\Delta| \le 1 \tag{56}$$

To account for varying network delays using an uncertainty model, the uncertainty has to be with respect to the average data latency. Accordingly the delay to be considered would include a constant and an uncertain part, or,

$$e^{-(\tau_A+\tau_V\Delta)s} \qquad -1 \le \Delta \le 1 \tag{57}$$

where, τ_A is the average latency

τ_V is the variation with respect to the average

Two alternatives to design have been proposed by Wang, Lundstromand Skogestad (1994). The first consists of leaving the average delay in the nominal model and let the uncertainty model account for the varying part. The second option leaves the nominal model delay free and the time delay uncertainty is as (57). Comparable results have been shown with both approaches by Wang, Lundstrom and Skogestad (1994). They however pointed out that the delay free (nominal model without delay) design problem is easier to handle.

CHAPTER 4

A NEW INTERPRETATION TO DESIGN

4.1. Proposition

An interesting approach introduced by Torngren (1996) is motivated by the fact that vacant sampling and sample rejection seem to affect the control system just as disturbances would. With this idea in mind, a feedback control system with variable delays could be viewed as a time-invariant system exposed to disturbances that are introduced by the communication system. The control signal and the measurement data are nevertheless affected differently. The control signal is only affected by the communication delay, it is variable and causes periodically generated control samples to arrive at the input of the process with a varying sampling period (jitter). The measurement data on the other hand is affected by the communication delay and the detection delay, the effects have been characterized before as vacant sampling and sample rejection. One can see on the left hand side of Figure 13, the measurement data subject to vacant sampling and message rejection assimilated as the disturbance signal v_τ . On the other side of the Figure 13, we assume the control signal to be already compensated for a constant transport delay τ_c. Once again the variability of the transport delay can be translated into a disturbance signal w_τ .

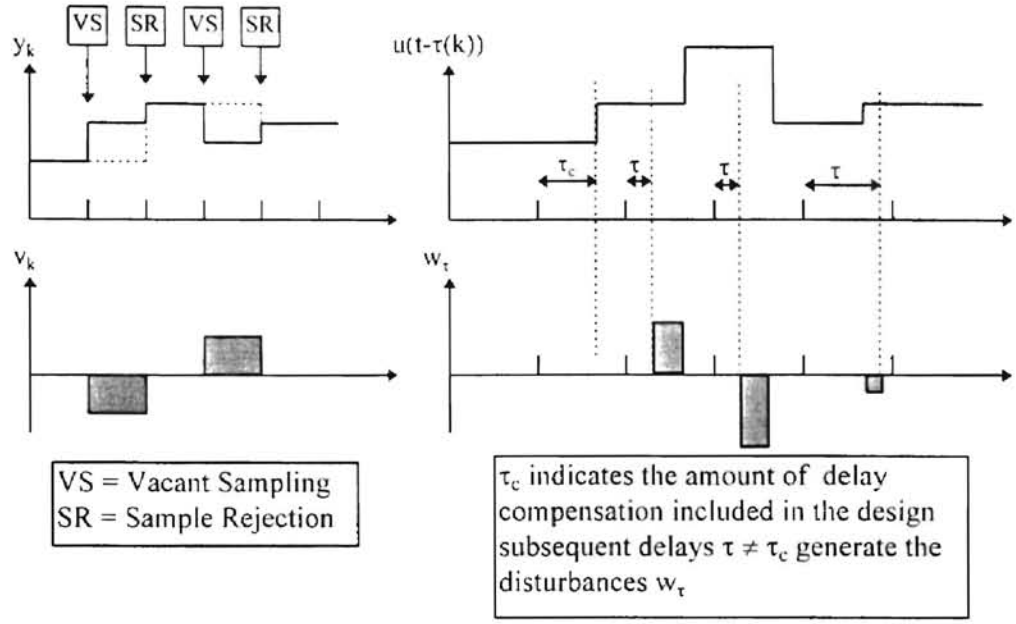


Fig. 13. Modeling closed loop time-varying delays as disturbances

The characterization of v_k can be made as follows.

Consider the delayed output,

$$z_k = y_{k-p(k)} \tag{58}$$

where,

y = output vector

z = delayed output

p (k)= variable sensor to controller delay belonging to the set {0,1}

Then the following situations may occur.

$$p(k-1) = 0;\ p(k) = 0 \Rightarrow z_k = z_{k-1} + \Delta y_k$$

$$p(k-1) = 1;\ p(k) = 1 \Rightarrow z_k = z_{k-1} + \Delta y_{k-1} \tag{59}$$

$$p(k-1) = 0;\ p(k) = 1 \Rightarrow z_k = z_{k-1}$$

$$p(k-1) = 1;\ p(k) = 0 \Rightarrow z_k = z_{k-1} + \Delta y_k + \Delta y_{k-1}$$

where,

$$\Delta y_k = (y_k - y_{k-1}) \tag{60}$$

$$\Delta y_{k-1} = (y_{k-1} - y_{k-2})$$

- In the first situation of expression (59) the system does not suffer any delay.
- In the second situation the whole system is simply phase lagged by one sampling period: z_k is off the actual value by $\Delta y_k - \Delta y_{k-1}$.
- In situation 3, a vacant sampling has occurred: z_k if off the actual value by $-\Delta y_k$.
- Situation 4, illustrates a message rejection where z_k if off the actual value by Δy_{k-1}.

Along these lines it is suggested that, from the above situation, the effects of the delay variability on the measurement data can be expressed by,

$$z_k = z_{k-1} + \Delta y_k + v_k \tag{61}$$

where v_k is a train of impulses with varying amplitude (amplitude related to the rate of change of y(k)). It is these impulses that distort the control signal. Under the assumption that the sampling period is usually much faster than the system, let $\Delta y_k \approx \Delta y_{k-1}$ and v_k

could be viewed as a linear function of $(y_k - y_{k-1})$. Moreover, it seems reasonable to see v_k as a signal bounded by,

$$|v_k| < \Delta y_k \tag{62}$$

The characterization of w_k is obtained differently. It is assumed that the control signal is already compensated for the average communication delay τ_c. The actual communication delay is time varying and denoted $\tau(k)$. The disturbance model w_τ is then a function of $[\tau(k) - \tau_c]$ and the control signal $u(k)$. This disturbance model in continuous-time is expressed as,

$$w_\tau(t) = u(t - \tau(k)) - u(t - \tau_c) \tag{63}$$

and is illustrated on the left hand side of Figure 13. The effect on the control system is the result of the time integral $w_\tau(t)$ during a period.

4.2 Scope and Limitations

The disturbance signal interpretations may be used to evaluate and shape the design of networked distributed control systems. For instance, if the time varying characteristics of the data latency are known from the study of the network characteristics, the disturbance model will be known. These disturbances may be concentrated at certain frequencies and one might want to emphasize regulation at those frequencies. The disturbance signal interpretation illustrated above can be viewed entering the system in the configuration of Figure 14.

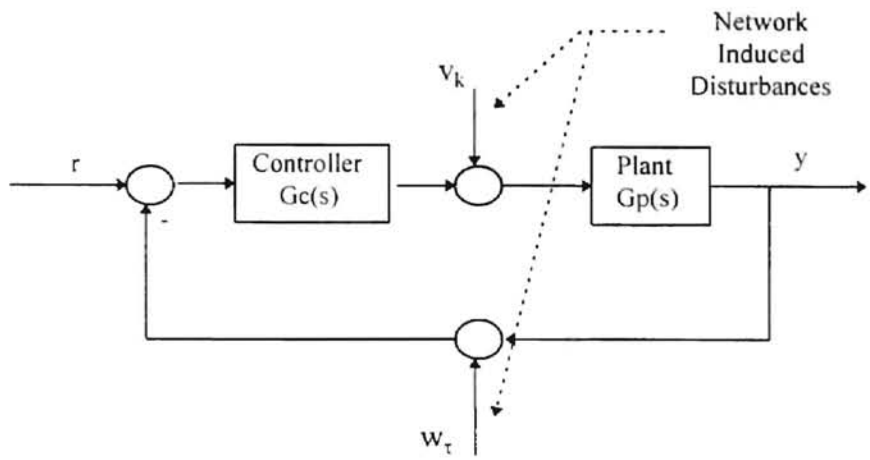


Fig. 14. Configuration of the Distributed Delay Disturbances

Provided that the characteristics of v_k and w_τ are known or can be estimated, the control configuration is now a familiar one. The frequency loop shaping approach then becomes an alternative to design. Accordingly the detrimental effects of the network induced delay can be integrated in the design tradeoff, high loop gain for tracking and small loop gain to reduce the effects of the network delays. The main obstacle to that objective is that the frequency characteristic of the network data latency typically include both high and low frequencies. Furthermore, the frequency characteristic probably varies with the number of active network stations and their respective synchronization changes. With these limitations in mind, the example treated in section 2.2 is considered again for improvement on the system's response.

4.3 Example

Recall the plant transfer function of the example in section 2.2.

$$G(s) = \frac{1000}{s(0.5s+1)} \tag{64}$$

with the combined effect of jitter on the control signal and data loss on the measurement. it was not possible to achieve a stable system response that would satisfy the specified requirements of 0.7 damping ratio and a 56 rad/s natural frequency. It is the intention of this section to show that this situation can be improved with prior knowledge of the characteristics of the induced network delays. For the benefit of shaping the loop gain to better reject the disturbances caused by the variable transport delays, the jitter and data loss sequences have been examined in the frequency domain. As a reminder, the control signal jitter is driven by the vector [.1 .5 .2 .7 .1 .8] where each element represents the transport delay as a fraction of the sampling period T=0.01. Similarly, measurement samples are lost due to late arrival for each 'one' in the repeating vector [0 0 1 0 1 0]. These two vectors have the same frequency spectrum when repeated at the sampling period of 0.01 second. Their common frequency spectrum exhibit three distinct peaks, respectively at 103, 207 and 314 rad/s. It is these relatively high frequencies that have to be targeted and suppressed by rolling off the loop gain. For that matter, two poles have been added to the compensation. The poles are located at -100 on the s-plane with the intention to provide a -40dB per decade of gain loss after 100 rad/sec. The corresponding transfer function is.

$$Gc(s) = \frac{1}{(0.01+s)^2} \tag{65}$$

or equivalently in the z-domain.

$$Gc(z) = \frac{0.2642z + 0.1353}{z^2 - 0.7358z + 0.1353} \tag{66}$$

in the context of the stepwise simulation it is easier to deal with the equivalent difference equation,

$$y(k)=0.2642\ x(k-1)+0.1353\ x(k-2)+0.7358\ y(k-1)-0.1353\ y(k-2) \tag{67}$$

Incorporation of this low pass transfer function into the simulation algorithm improved the response of the system by reducing the effects of the jitter and data loss. With no other changes to the conditions of the simulation, the system response is now stable as shown on Figure 15. In this figure, the solid line represents the same nominal response, the dot-dash line is the uncompensated system response, and the dashed line is the compensated system response. The compensated response is an improvement over the simple proportional state feedback that was unstable, nevertheless, the required damping ratio of 0.7 is not achieved. One has to note that we are dealing with a border line situation where both the sampling period, the plant dynamics, and the jitter and data loss patterns are combined to generate a challenging control situation. Clearly, decreasing the sampling period would have a much more beneficial effect. The point of this example is to show that the detrimental effects of the network induced delays can be integrated in the design process, reduced or possibly suppressed by frequency shaping the loop transfer function.

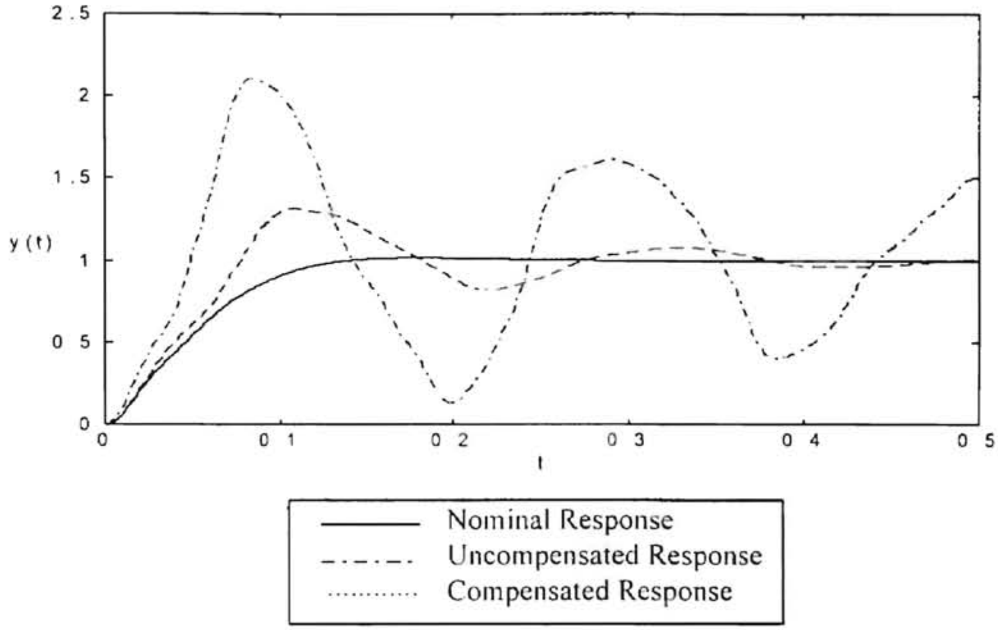


Fig. 15. Compensated System Response Under Jitter and Data Loss

It is important to remember that, this reasoning relies on the questionable assumption that the frequency characteristic of the network data latency is known and constant. Furthermore, the detrimental effects can only be attenuated provided that they are not located in a frequency band that necessitates a high gain for other reasons, such as, tracking requirements.

CHAPTER 5

CONCLUSIONS

In many networked real-time distributed systems, the measurement and control signals within the feedback loop are subject to time-varying network induced transport delays. The time-varying nature of these delays has been shown to cause jitter on the control signal and data loss on the measurement data. The detrimental effects of these two phenomena have been simulated in the context of a networked feedback loop. Interestingly, jitter and data loss proved to have a multiplicative detrimental effect upon each other. Moreover, and as expected, the controller fastness is a determinant factor to the sensitivity of the controller performance toward time variations. Last, for a single sample loss, the instant of occurrence has a great significance on the effect of the performance. This report therefore clearly provides the reader with a perspective on the problem of transport delays in the context of a distributed control loop and a report on the current research status. In addition, it has been shown that successful qualitative reasoning can be made by interpreting the time variations as disturbances introduced by the communication system. Nevertheless, a major limiting factor in the design of any compensation algorithm is the absence of information about the characteristics of the network data latency. Consequently, further work is needed, first in the characterization and identification of the network performances particularly when subject to quasi-periodic traffic and secondly, in the integration of these characteristics into the compensation design possibly using the disturbance interpretation provided here. To that aim a short report on the Controller Area Network (CAN) of widespread use for distributed control systems is given in the appendix.

BIBLIOGRAPHY

Astrom, K.J., and Wittenmark, B., *Computer Controlled Systems, Theory and Design* Prentice Hall, 1990

Bosh, "CAN Specification" Version 2.0 1991. *Robert Bosh GmbH*, Postfacch 50, D-7000 Stuttgart 1

Franklin, G.F., Powell, J.D., and Workman, M.L., *Digital Control Systems* Addison Wesley, 1994

Gergeleit, M., and Streich, H., "Implementing a Distributed High-Resolution Real-time Clock Using the CAN-Bus"

Halevi, Y., and Ray, A., "Integrated Communication and Control Systems: Part I Analysis," *ASME Journal of dynamic systems, Measurement and Control* Dec 1988 pp. 367-373.

Hirai, K., and Satoh, Y., "Stability of a System with a Variable Time Delay" *IEEE Transaction on Automatic Control*, Vol. AC-25, No. 3, June 1980, pp. 552-554.

Liou, L. W., and Ray, A., "Integrated Communication and Control Systems: Part III Non-Identical Sensor and Controller Sampling," *ASME Journal of dynamic systems, Measurement and Control* Sept. 1990.

Luck, R., and Ray, A., "An Observer-Based Compensator for Distributed Delays," *Automatica*, Vol. 26, No. 5, pp. 903-908, Sept. 1990.

Luck, R., and Ray, A., "Experimental Verification of the Delay Compensation Algorithm for Integrated Communication and Control Systems" *International Journal of Control*, 1994, Vol. 59, No. 6, pp. 1357-1372

Ludvigson, M. T., "Thoughts on High Speed Data Bus Performance", *National Aerospace & Electronic Conference* 1990 (NAECON) p. 163-168

Oguztoreli, M. N., *Time-Lag Control Systems* Academic Press, 1966

Ray, A. "Performance Evaluation of Medium Access Control Protocols for Distributed Digital Avionics" *ASME Journal of Dynamic Systems, Measurement, and Control,* Vol. 109 Dec.1987. pp 370-375.

Ray, A., and Halevi. Y.. "Integrated Communication and Control Systems: Part II Design Considerations," *ASME Journal of dynamic systems, Measurement and Control* Dec 1988 pp. 374-381.

Ray, A., Liou, L.W., and Shen. J. H., " State Estimation Using Randomly Delayed Measurements", *ASME journal of Dynamic Systems, Measurement and Control*, Vol. 115, March 1993.

Ray, A., "Output Feedback Control Under Randomly Varying Distributed Delays", *Journal of Guidance, Control and Dynamics*, Vol. 17, No. 4, July-August 1994.

Shen, J.H., and Ray A.. "Discrete-time LTR Synthesis of the Delayed Control System" *Proceedings of the American Control Conference*, 1992

Shen, J. H., and Ray, A.. "Extended Discrete-time LTR Synthesis of Delayed Control Systems" *Automatica*, Vol. 29. No. 2, 1993, pp. 431-438.

Tindell, K., Burns, A., and Wellings, A. "Calculating Controller Area Network Message Response Time" *IFAC Distributed Computer Control Systems*. Toledo, Spain, 1994.

Torngren, M., "Modelling and Design of Distributed Real-time Control Systems: an Automatic Control Perspective" *DAMEK-Mechatronics, Departement of Machine design*, The Royal Institute of Technology, Stockholm, Sweden. 1996.

Wang, Z., Lu, H., Hedrick, and G.. Stone, M.. "Message Delay Analysis for CAN Based Networks" *Oklahoma State University*. 1992.

Wang, Z., Lu, H., and Stone. M., "A Message Priority Assignment Algorithm for CAN Based Networks " *Oklahoma State University*. 1992.

Wang, Z., Lundstrom, P.. and Skogestad, S., "Representation of Uncertain Time Delays in the H_∞ Framework" *International Journal of Control*, Vol. 59, No. 3, 1994 pp. 627-638.

Wittenmark, B., Nilsson, J., and Torgren, M., "Timing Problems in Real-time Control Systems" *Proceedings of the American Control Conference*, Seattle, Washington, June 1995.

Zeltwanger, H., "An Inside Look at the Fundamentals of CAN" *Control Engineering*, January 1995, pp. 81-87.

Zhan, Z., and Freudenberg, J.S., "On Discrete-time Loop Transfer Recovery" *Proceedings of the American Control Conference*, 1991 pp. 2214-2219

APPENDIX

The Controller Area Network Communication Protocol for Distributed Control System

The Controller Area Network (CAN) protocol is a serial network originally developed by Robert Bosh GmbH (Bosh 1991) to provide the car industry with a communication bus for in-car electronics. It has been chosen as the communication protocol of interest for its real-time capabilities. Specifically, the priority at which the message is transmitted is incorporated into the identifier of each message. Bus access conflicts are then resolved by bitwise arbitration on the identifiers involved by each station. This mechanism fulfills the requirement of rapid bus allocation and reliability by decentralized bus control. These arguments along with low cost make CAN the preferred solution for many control system applications.

Principle of Operation

CAN uses a content-oriented addressing scheme where each message is labeled by an identifier that is unique throughout the network. The identifier defines the type of data transmitted but most importantly defines a static message priority. For rapid bus allocation when several stations wish to send messages, the CAN protocol uses bitwise arbitration. A node can start transmitting its identifier at any time when the bus is silent. During arbitration every transmitter compares the level of the bit transmitted with the level that is monitored on the bus. If any node transmits a '0' (dominant bit) then all nodes read back a zero. A unit sending '1' (recessive bit) but reading a '0', will automatically withdraw from the contention. As the result, the bus continually tracks the winner, the identifier with the lowest binary number. All losers automatically become receivers of the highest priority message and will reattempt transmission at the next idle period.

The Message Frame Formats

The CAN protocol supports two message frame formats, standard and extended. The only difference is the length of the identifier, in the standard format the length is 11 bits and in the extended format, 29 bits. The two formats can be seen in the appendix. A standard CAN message frame consists of seven different bit fields. A message begins with a *start of frame field* to indicate the beginning of a message frame. This is followed by the *Arbitration field* which contains the identifier and the remote transmit request bit (RTR). The RTR bit indicates whether it is a data frame or a request frame. Next, comes the *control field* Containing two dominant bits reserved for future use, and a count of the data bytes, the data length code. In the middle of the frame, the *data field* carries a maximum of 8 bytes of data and is followed by the *CRC field*. As part of the error checking mechanism, the CRC field includes a fifteen-bit cyclic redundancy check code and a recessive delimiter bit. Following, is the *ACKnowledge field* which is over written by dominant bits upon successful reception by other nodes. The end of the message is indicated by the *End of Frame field* consisting of seven recessive bits. At last, the *intermission field* is composed of three recessive bits enabling the CAN nodes to prepare for the next task. Also shown in the appendix is the error frame, interspace space and overload frame. The error frame is a mean by which a station can flag an error and abort transmission. The overload frame can only be initiated during the intermission field to delay any subsequent message frame.

The Error Detection Mechanism

Error detection is implemented at the bit level as well as at the message level with several mechanisms capable of distinguishing and correct sporadic and permanent errors. Resulting from noise corruption or spikes in the communication medium, transient errors are detected and corrected by abortion and retransmission of the message. Permanent errors are likely to be caused by bad connections, defective transducers or long lasting external disturbances. They are self-contained by shutting down the station that is blocking the bus. The *Cyclic Redundancy Check (CRC)* is computed on the basis of the message content. All receivers perform similar calculations and flag any errors. In addition, Certain predefined bit values such as CRC and ACK delimiters, EOF bit field, and intermission will also trigger an error flag if invalid. Likewise, If a transmitter has not been acknowledged an error is flagged. Finally, all transmitters compares the bit level of the bus with the level it transmitted and flags an error if the two are not the same (ACK bit and arbitration excepted). In order to maintain bit synchronization in between stations a minimum number of bit transitions has to occur. Therefore, after five identical bit levels have been transmitted. the transmitter will automatically inject a bit of opposite polarity. Receivers of the message will automatically delete that extra bit. Inside the CAN node, a register called the error count is dedicated to summing the number of receive and transmit errors. In order for the error count to remain low, every good message decrements the register. When the error count reaches 128 the node switches to the error passive mode. In this mode the station can still transmit and receive but can no longer flag errors. when the error count reaches 255 the node switches to bus off mode. In this mode, the device will cease to be active on the bus

The CAN Controller

The communication controller has many functional blocks. Of particular interest are, the CAN controller, the RAM and the CPU interface logic. The CAN controller controls the data stream between the RAM and the bus line. The RAM provides storage for 15 message objects and various control and status registers. The CPU interface logic provides a flexible interfacing to many commonly used microcontrollers. Each message object has 8 bytes of data, an identifier and control data segments. In addition each message object can be configured to either transmit or receive. To initiate a transmission. the transmission request bit has to be written to the message object. In a same fashion a message object can be configured to receive a message. In that case, Acceptance filtering is performed by matching the identifier of the incoming message against the identifiers of all message objects. A message is accepted and the receive interrupt activated only if a match is found. Note that. a message object can store only one message, any message that has not been picked up will be overwritten.

The Design Implementation of CAN

CAN is commonly modeled as a single channel queuing system. The bus is the server and all the spatially distributed waiting messages form a single queue. One has to note that, a message object can only contain one message. When another message with the same identifier is queued then the content of the message object is overwritten and destroyed. The queue therefore, can contain no more than one message with the same identifier. The priority scheme as described by CAN only affects the order inside the queue. In other words, depending upon its priority, a transmission request will enter the queue at different levels.

Cooperating activities in a distributed control system can occur synchronously or asynchronously. Synchronous operation enable the application to operate in a deterministic timely fashion according to a strategy defined off-line. This operating mode is supported by static scheduling and ensures predictable bus loading. Hence, in this approach, the time variations can be minimized or eliminated. As the result, synchronous operation guaranties predictability of the data latency and constant detection delays. Consequently, several attempts have been made to provide for synchronous operation on the CAN communication medium. A simple approach consist of using a high-priority signal at prescribed interval to reset the respective sensor and actuator clocks and to maintain a loose synchronization between sensor, controller and actuator. One such endeavor described within the DIRECT project (Gergeleit and Streich) consist of a protocol that synchronizes accurate local clocks via the CAN-bus network. An accuracy of about 20 microseconds is obtained using a reasonably small amount of bandwidth (< 20 messages / second). The system designer can then benefit from the distributed real-time clocks to schedule synchronous operations. Another concept, supported by the CAN-in-Automation (CiA) forum, "CAN Open", is a protocol that allows synchronous data transfer over the CAN bus. A very high priority synchronization telegram is sent on a set time period to all devices. On reception, those devices that are configured to respond to it send data onto the bus. Once again, this allows static scheduling and ensures predictable bus loading. Moreover, it is very well suited to the type of periodic operations often found in control systems, namely, sampling and actuation. For instance, after receiving the measurement data synchronized to one synchronization telegram, the controller can send its control signal back to the actuator on reception of the next synchronization telegram. Equally important, sampling and actuation messages can be configured to occur on a common multiple of the synchronization period. Synchronous and predictable operation is a significant advantage. On the other hand it provides very little room for sporadic transmissions which may occur occasionally but with a high degree of urgency. Hence, all choices must be made conservatively to cater for every possible demand. Last, it does require more intelligence to be installed in every communicating devices to support the synchronization mechanism. The synchronous solution is therefore complex, inflexible, slightly inefficient and costly. The unsynchronized alternative would include sensors that send periodic information blindly and oversampled actuators that operate on reception. In particular, it is believed that the low cost of CAN makes it very suitable to that alternative. This unsynchronized choice is time varying in behavior and will need to rely more on control engineering solutions for compensation of the time variations. Accordingly, this approach calls for the modification of the existing control systems or the development of a new control structure. Even though the design of a modified control law to compensate for the time variations represents an additional effort. It will pay off with a highly enhanced bus utilization together with an increased robustness to other possible disruptions or even malfunctions of the network, truly expanding the flexibility and reliability of the system.

The Expression of the Transport Delay in CAN

According to the CAN arbitration protocol and the queuing model presented above, the time a station must wait for the bus to become idle is called the waiting time. The waiting time is the time needed for the current message utilizing the bus to finish plus the time needed to transfer all higher priority messages waiting in the queue and arriving during the waiting time. Typically, one station will have already gained control of the bus when the transmission is requested. The longest time a station must wait for the bus to become idle is the time needed to transmit the largest CAN frame (full 8-byte extended message) and is called the blocking time. The time that the request must wait until the current transmission releases the bus can vary from 0 to the blocking time. Secondly. The request has to wait for all of the higher priority messages in the queue to go through first. Recall that messages can enter the queue at any level depending on their respective priority. Consequently, One has also to account for those messages that have been generated with higher priority during the waiting time. Only then, a station can finally seize the bus and transmit its message. The receiving station will acknowledge reception after the last bit of that message has been transmitted. The transmission delay or data latency, is then composed of the waiting time plus the time needed to transmit that message, or $D_i = W_i + F_i$ with D, W, F being respectively the transmission delay, the waiting time, and the length of the frame for the i^{th} priority message.

Rather than dealing with all possible transmission times, one solution is to consider just the longest possible, or worst-case situation. Thus, the maximum delay analysis is concerned with the worst case scenario. That is, the request happens when a message has just seized the bus and all individual stations, with higher priority, are generating their maximum rate of messages. Let W_i be the maximum waiting time for i^{th} priority message, F_k the length of the frame for the k^{th} priority message, M_k the maximum message generation rate for the k^{th} priority message, IFS the time required for an inter frame space and B the bandwidth of the bus. Then, the maximum waiting time for the i^{th} priority message has been given by Wang, Lu, Hedrick and Stone (1992) as,

$$Wi = (F_{max} + IFS - 1) + \sum_{k=0}^{i-1}(F_k + IFS) + \sum_{k=0}^{i-1}(F_k + IFS)\ M_k \frac{W_i}{B} \tag{68}$$

The three terms at the right of the equality are respectively from left to right, the blocking time, the time necessary to transmit all higher priority messages, and the time required to transmit higher priority messages coming during the waiting time. The above formulation can be solved for the maximum waiting time. The maximum transmission delay for a particular message is then obtain by adding the maximum waiting time to the time needed to physically transmit the message in relation with the request. One has to note that the maximum delay analysis is a deterministic value that would provides absolute certitude on the upper bound of the transport delay if we had considered the error recovery mechanism, overload frames, and remote transmission requests. The reader should refer to Tindell, Burns, and Wellings (1994) for the proper analysis on the cost of error handling and remote transmission requests. Since in normal mode of operation these omissions represent a negligible overhead, the simplified maximum delay formulation is satisfactory. Nevertheless information on the upper bound is only good for robustness analysis. Commonly the actual performance is likely to be much better due to interference averaging of all the communication transactions.

The maximum delay analysis is only useful to guaranty that certain hard deadline will be meet no mater what. Nonetheless, it is the representation of a very low probability situation. Thus, the characterization of a message delay can be obtained by assuming random generation of messages those statistics are characterized by the Poison distribution. Similarly the message length is also assumed to be exponentially distributed. With these two elements in mind, the same reasoning as for the maximum delay analysis can be carried out, using expectation values in the place of deterministic upper bound values. This delay analysis model has been developed by Wang, Lu, Hedrick and Stone (1992). The total expected waiting time for the i^{th} priority message is given along the same reasoning and three-term formulation as in maximum delay analysis.

$$W_i = \rho \frac{F + IFS - 1}{2} + \frac{\lambda \sum_{k=0}^{i-1} \lambda_k}{\mu(\mu - \lambda)} F + \sum_{k=0}^{i-1} (F_k + IFS)\lambda_k \frac{W_i}{B} \quad (69)$$

The same notation is reused with the addition of,

λ_i the average message generation rate of the ith priority message,

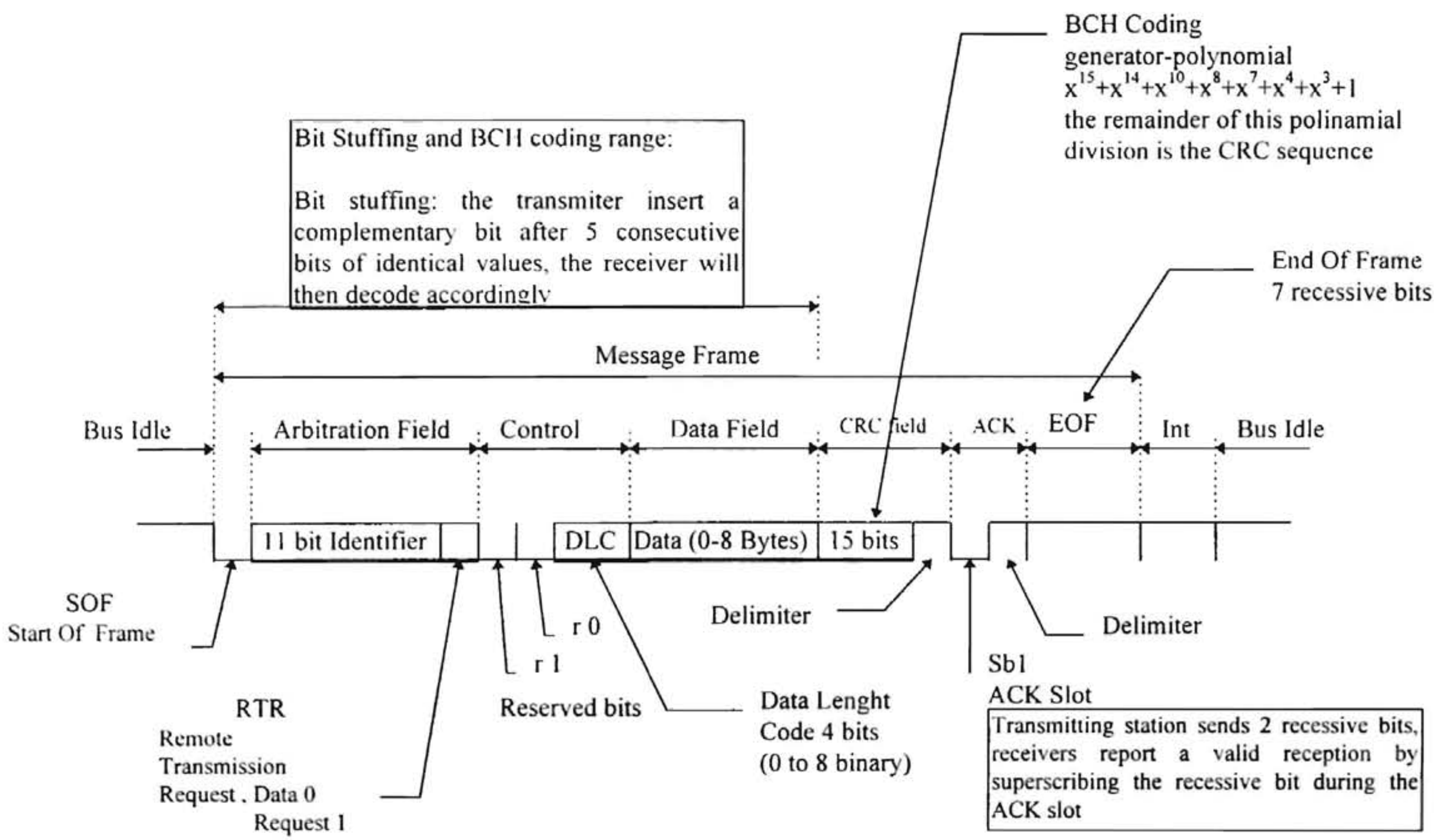
μ the expected number of messages the bus can transmit per second.

ρ the traffic intensity of the network (λ/μ).

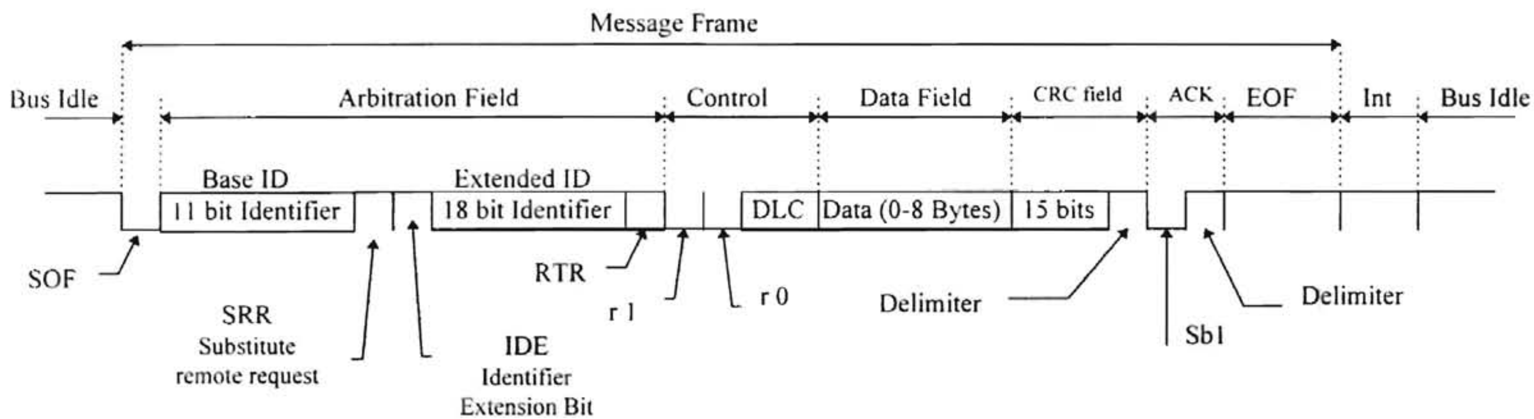
The expected delay analysis results are likely to be closer to the actual experienced delays than the maximum delay analysis. It is however more difficult to obtain an accurate expression for the expectation values than it is to obtain the upper bound. Further, It will be pointed out later in section 3.1 that average performance calculations do not provide an adequate method for evaluating the bus performance because the bus traffic is not randomly distributed.

The Priority Setting in CAN

In network scheduling terms, priority is a positive integer representing the urgency or importance assigned to a message. In CAN the urgency is in inverse order to the numeric value of the priority. In addition, priority is a static property of the sender of the message and cannot be changed dynamically. In most real-time systems, there is a robustness issue and a performance issue. To ensure robustness, for every message the allowed maximum transport delay cannot be exceeded. For performance we want the average data latency to be minimum. To guaranty that the maximum delay requirement of every message is satisfied, the priority assignment is to be carefully organized. Many techniques are available for static priority assignment such as, rate-monotonic (priority=period^{-1}), deadline monotonic (priority=deadline^{-1}), or according to the importance of each type of message. These well known techniques can provide a workable assignment but do not generate an optimum solution. In CAN the interest is not only to satisfy all the time constraints but also to create an optimum assignment that will minimize the average delay of all messages. This "conditional optimum" (or optimum subject to a set of hard constraints) assignment is based on a systematic search and sorting of a characterized set of messages and as been described by Wang, Lu and Stone (1992). The rationale of this technique relies on, putting a task into the lowest priority position and checking whether a feasible result is obtained. Feasible meaning that, no hard deadline is violated. If this fails the next higher priority position is considered, and so on.

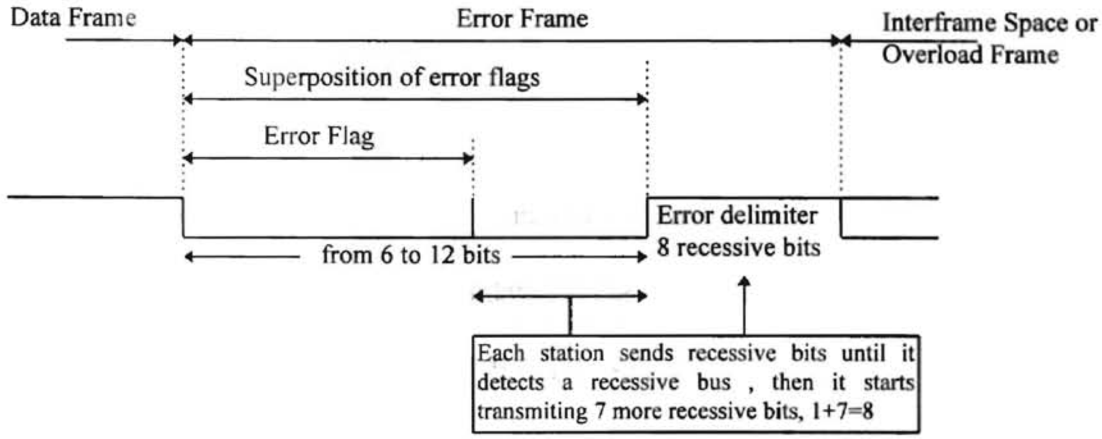


Message Frame 2.0 A Format (Standard Frame)

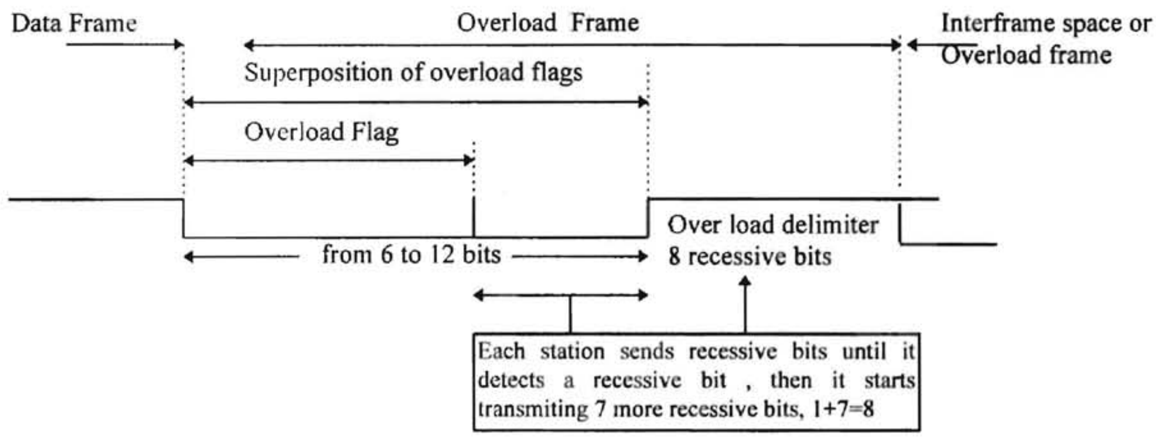


Message Frame 2.0 B Format (extended frame)

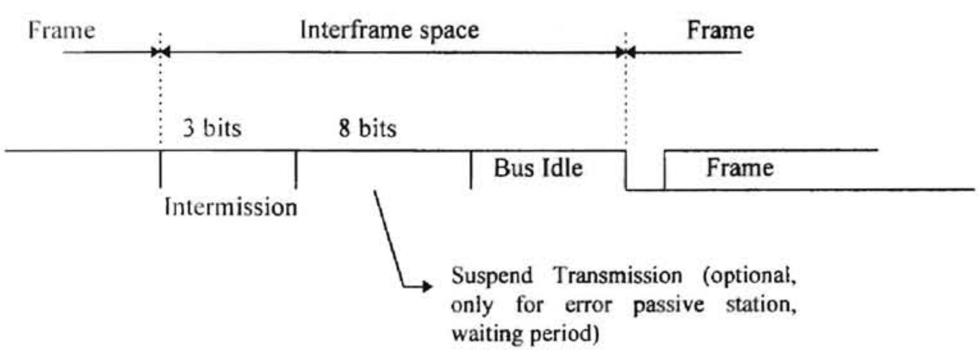
Error Frame:



Overload Frame:



Interframe Space:



VITA

Emmanuel Vyers

Candidate for the Degree of

Master of Science

Thesis: ON THE ANALYSIS AND COMPENSATION OF NETWORK INDUCED COMMUNICATION DELAYS FOR DISTRIBUTED CONTROL SYSTEMS

Major Field: Mechanical Engineering

Biographical:

Personal Data: Born in Denain, Nord, France, on September 15, 1969, the son of Roger and Raymonde Vyers.

Education: Graduated from High School, Les Eucaliptus, Nice, France in 1988; received Bachelor of Science degree in Aerospace Engineering from Florida Institute of Technology, in May 1995. Will complete the requirements for the Master of Science degree with major in Mechanical Engineering at Oklahoma State University in December 1997.