Y. Halevi
Mechanical Engineering Department,
Technion, Haifa, Israel

A. Ray
Mechanical Engineering Department,
The Pennsylvania State University,
University Park, PA 16802
Mem. ASME

Performance Analysis of Integrated Communication and Control System Networks[1]

This paper presents statistical analysis of delays in Integrated Communication and Control System (ICCS) networks [1–4] that are based on asynchronous time-division multiplexing. The models are obtained in closed form for analyzing control systems with randomly varying delays. The results of this research are applicable to ICCS design for complex dynamical processes like advanced aircraft and spacecraft, autonomous manufacturing plants, and chemical and processing plants.

I Introduction

Distributed information processing is essential for control and management of complex dynamical processes like advanced aircraft and spacecraft, autonomous manufacturing plants, and chemical and processing plants [3–8]. Computer networking is a reliable and efficient means fc· communicating between spatially distributed components of such processes. The role of Integrated Communication and Control Systems (ICCS) is to coordinate and perform inter-related functions ranging from closed-loop control of essential process variables to routine maintenance support and information display. In ICCS, a control loop is closed via the common medium of the network which mulitplexes digital data from sensor to controller and from controller to the actuator along with traffic from other control loops and plant management and decision-making functions [1–4]. Due to asynchronous time-division multiplexing of the protocol, as it is commonly used in ICCS networks, randomly varying delays are introduced within the control system. The notion of network-induced delays in ICCS, as explained in [1], is different from that in conventional local area networks (LANs) [9, 10] as the delays degrade dynamic performance of the feedback control system(s) and are a source of potential instability.

Although ample research work in modeling of communication protocols has been reported [9–11], the significance of network-induced delays relative to stability of feedback control systems has not been apparently addressed except in a few cases [1–4]. Analysis and design of ICCS require interactions between the disciplines of communication systems and control systems. It has been shown in [1–4] how network-induced delays can degrade the stability of a feedback control system where the components are interconnected via a common medium. The dynamic performance of ICCS networks was evaluated in [3, 4] using combined discrete-event and continuous-time simulation under diverse traffic conditions as well as for different protocols. Finite-dimensional models of (randomly varying) delayed systems were developed in [1, 2] from the point of view of control systems design. Section 4 of [2] elucidates certain critical steps in ICCS design where the knowledge of statistical characteristics of network-induced delays is essential.

Since ICCS networks must accommodate a combination of periodic and nonperiodic traffic, the message arrival process may not be Poisson. Therefore, the M/G/1 queueing modeling approach [9–11], that uses the concept of imbedded Markov chain, is not generally applicable to statistical analysis of delays in ICCS networks. On the other hand, G/G/1 queueing modeling does not provide a closed-form solution [11] and hence is not suitable for analyzing the control systems that are subjected to randomly varying delays. The major contribution of this paper is statistical modeling (as closed form solutions) of network-induced delays for ICCS design from the points of view of both communication and control systems. The work reported in this paper complements the earlier work [1, 2] which provides the necessary background.

The paper is organized in three sections including the introduction. Section II which forms the main body of this paper presents concepts of pertinent network parameters and establishes an analytical base for ICCS network design. The general case of random message lengths and inter-arrival times at individual terminals is dealt with in the first part of Section II. Then network-induced delays and associated parameters are analytically derived for two special configurations of control systems: One is the case of identical sampling frequency of the sensor and controller within a feedback control loop; the other addresses the situation where the sensor sampling frequency is larger than that of the controller, which is a viable option for ICCS design [2]. Terminologiés, specific to ICCS, are explained in the nomenclature which is distributed in two places in Section II. Assumptions are stated and definitions are introduced wherever necessary. Twelve propositions and one supporting lemma present the analytical results which are focused toward statistical analysis of network-induced delays. The significance of the derived analytical results relative to ICCS network design are discussed as remarks that follow the

[1]This work was supported in part by:
NASA Lewis Research Center Grant No. NAG 3-823.
National Science Foundation Research Grant No. DMC-8707648.
Contributed by the Dynamic Systems and Control Division for publication in the JOURNAL OF DYNAMIC SYSTEMS, MEASUREMENT, AND CONTROL. Manuscript received by the Dynamic Systems and Control Division July 29, 1988.

individual propositions. Section III presents the summary and conclusions.

II Analytical Modeling of Network Performance Parameters

Nomenclature

- N = number of terminals in the network
- T = message inter-arrival time (for example, sampling interval in a control loop) at a terminal
- l = message length (in time units) including overhead, i.e., the message length (in bits) divided by the transmission rate (bits/unit time)
- θ = queueing delay, i.e., waiting time of a message at a source terminal
- q = cumulative length (in units of time) of all waiting messages at the queue of a source terminal at a given instant
- σ = the sum of the station response time to a source terminal and the propagation delay from its predecessor. (Note that $(\sigma_1 + \sigma_2 + \ldots + \sigma_N)$, where the subscripts indicate terminal numbers, is called ring latency [10] of the network.)

Remark 1: The station response time at a terminal is the electronic delay due to protocol execution and is an identical constant for all terminals under normal operating conditions. The propagation delay between a pair of terminals is negligible in small local area networks with low transmission rates but it could be significant for high transmission rates. For example, in the 100 Mbps SAE token bus protocol [12], the station response time is specified to be bounded within 0.5 μs; if the cable distance between pairs of terminals within a small network (e.g., ICCS network in an aircraft [3]) is assumed to be in the range of 2m and 80m, then the resulting σ at a terminal should be approximately between 0.51 and 0.9 μs. Nevertheless, σ plays an important role in the performance analysis of network-induced delays and this is shown later. ■

Assumption 0: The key assumption is that random processes that are incurred in the network operations are ergodic with probability one [13]. ■

Assumption 1: The random variables T and l are mutually statistically independent at and between all terminals. Furthermore, σ_i is a constant $\forall i$ but $\sigma_i \neq \sigma_j$ is possible for $i \neq j$. ■

Definitions and General Results

Definition 1: Offered traffic G is defined as the expected value of the lengths (in units of time) of messages that arrive at network terminals per unit time. ■

Proposition 1:

$$G = \sum_{i=1}^{N} E[l_i]/E[T_i]$$

where the subscript i denotes the ith terminal.

Proof: Let the length of the jth message transmitted by the ith terminal be denoted as l_i^j. Consider an interval Υ, during which the ith terminal receives $n^i(\Upsilon)$ messages. As $\Upsilon \to \infty$ (and also $n^i \to \infty$)

$$\left(\sum_{i=1}^{N} \sum_{j=1}^{n^i} l_i^j\right)/\Upsilon \to G$$

Since $\Upsilon/n^i \to E[T_i]$ and $\left(\sum_{j=1}^{n^i} l_i^j\right)/n^i \to E[l_i]$, the proof is complete. ■

Remark 2: If $E[T_i]$ is identical for all terminals, then

$$G = NE[l]/E[T]$$

where $E[l]$ is the expected value of message lengths over all terminals. Furthermore, if l_i and T_i are constants for all i, then

$$G = \sum_{i=1}^{N} l_i/T_i$$ ■

Definition 2: Critical offered traffic G_{cr} is the largest offered traffic for which, assuming infinite queue capacity, all queues are bounded under steady states. ■

Remark 3: Definition 2 implies that if $G > G_{cr}$, then some of the queues become unbounded and a steady state does not exist. ■

Remark 4: $G < G_{cr}$ does not guarantee absence of message rejections for random traffic if the queue capacity at a terminal is finite. But it does guarantee no message rejections under steady states for identical l_i and T_i at all terminals even if the queue capacity is limited to 1. (See Propositions 1 and 2 in [3].) ■

Definition 3: Cycle time τ is defined as the time interval between two consecutive opportunities to transmit waiting message(s) at a given terminal. ■

Remark 5: The random variable τ is a network parameter related to all terminals. ■

Remark 6: In a token passing protocol, the random variable τ is the interval between two successive token arrivals at any terminal. ■

Now we develop a relationship between $E[\tau]$ and G in terms of σ_i.

Proposition 2: For $G < G_{cr}$, $E[\tau]$ is given as

$$E[\tau] = \sum_{i=1}^{N} \sigma_i/(1-G)$$

Proof: Let the system complete k cycles during a period Υ. Then,

$$\Upsilon = k\sum_{i=1}^{N} \sigma_i + \Upsilon G - \sum_{i=1}^{N} E[q_i] \text{ as } \Upsilon \to \infty$$

The first term on the right is the total time due to ring latency [10], the second term is the total length (in units of time) of messages that arrive in $(0, \Upsilon)$, and the third term is the total length of messages accumulated in the queues. The difference between the second and third terms yields the total transmission time. Since $G < G_{cr}$ and all q_i's are finite, dividing the equation by k and using the notation $E[\tau] = \Upsilon/k$ as $\Upsilon \to \infty$, we obtain

$$E[\tau] = \sum_{i=1}^{N} \sigma_i + E[\tau]G - \left(\sum_{i=1}^{N} E[q_i]\right)/k$$

Since $k \to \infty$ as $\Upsilon \to \infty$, the last term goes to zero and the result follows. ■

Remark 7: The implication of Proposition 2 is that, for a given G, $E[\tau]$ is directly proportional to the ring latency $\sum_{i=1}^{N} \sigma_i$ of the network [10]. For a large N, the ring latency may have a major impact on τ (and hence on network delay performance) even if G is small. Under these circumstances, an implicit token passing protocol will perform better than an explicit taken passing protocol because of a smaller σ at each terminal

[9]. The explicit token passing bus protocol, as recommended for avionic applications [3] and real-time chemical processes [4] that are usually characterized by small N, may not be ideal for large Computer-Integrated Manufacturing (CIM) processes where N is likely to be very large [7, 8]. This is especially important if a high-speed fiber optic protocol is adopted for CIM applications where σ may be not be insignificant relative to $E[l]$. ∎

Proposition 3: The critical offered traffic is given as

$$G_{cr} = 1 - \sum_{i=1}^{N} \sigma_i / (\min_i E[T_i])$$

Proof: Assuming infinite queue capacity, no queue saturation occurs under steady state if and only if the average number of message arrivals at every terminal is less than or equal to the number of opportunities to transmit (e.g., number of token arrivals in a token bus protocol). This implies that $E[T_i] \geq E[\tau] \forall i$. Therefore, the critical inequality is obtained by use of Proposition 2 as

$$\min_i E[T_i] \geq \sum_{i=1}^{N} \sigma_i / (1-G) \text{ or}$$

$$G \leq \left(1 - \left(\sum_{i=1}^{N} \sigma_i / \min_i E[T_i]\right)\right).$$

Proof follows from Definition 2. ∎

Remark 8: The above result is useful for obtaining an estimate of the critical offered traffic G_{cr}. The implication is that an ICCS network should be designed such that G is smaller than G_{cr} by a safe margin [3]. ∎

We define the conditional cycle time with the objective of evaluating the expected value of the queueing delay θ[3, 4].

Definition 4: The conditional cycle time t is defined as the cycle time (see Definition 3) during which at least one message arrives at any one of the terminals. ∎

Propositon 4: For steady-state operations of the network, t and τ are related as

$$E[t] = E[\tau] + \mathrm{Var}[\tau]/E[\tau]$$

where $\mathrm{Var}[\tau]$ is the variance of τ.

Proof: Let the random variables t and τ assume one of the discrete values in the sequence $\{\Upsilon_j, j = 1, 2, \ldots, J\}$. Let a time period Υ be partitioned into subintervals of lengths Υ_j with multiplicity $n_j, j = 1, 2, \ldots, J$. As $\Upsilon \to \infty$, the probability p_j of a message arrival in the subinterval Υ_j is given as

$$p_j = \frac{n_j \Upsilon_j}{\Upsilon} = \frac{n_j \Upsilon_j}{\left(\sum_j n_j\right) E[\tau]} = \pi_j \Upsilon_j / E[\tau]$$

where π_j is the probability that τ assumes the value Υ_j. As $J \to \infty$, the distribution functions of t and τ are related as

$$dF_t(\phi) = (\phi / E[\tau])\, dF_\tau(\phi)$$

Therefore, $E[t] = \int \phi dF_t(\phi) = \int \phi (\phi / E[\tau]) dF_\tau(\phi)$

$$= \left(\int \phi^2 dF_\tau(\phi)\right) / E[\tau] = E[\tau] + \mathrm{Var}[\tau]/E[\tau]$$ ∎

Corollary to Proposition 4: The expected value of queueing delay at a terminal is $E[\theta] = (E[\tau] + \mathrm{Var}[\tau])/E[\tau])/2$.

Proof: Since the process of a message arrival at a terminal is independent of the instant of its transmission, the queueing delay θ is equally likely to take any value within the period of t (see Definition 4), i.e., uniformly distributed between 0 and t. Therefore, $E[\theta] = E[t]/2$. ∎

Remark 9: The above procedure does not assume any specific distributions of the message arrival process and message length. However, the first two moments of τ, as required in the analytical expression for $E[\theta]$ in Corollary to Proposition 4, can only be evaluated under certain specific conditions. On this basis, the problem of evaluating expected values of network-induced delays within a control loop is addressed later. ∎

Now we deal with the situation when the network could be overloaded due to unexpected increase in traffic such that $G > G_{cr}$.

Proposition 5: Let the terminals be ordered according to the expected values of their message inter-arrival times in a nondecreasing sequence, i.e., $E[T_{i+1}] \geq E[T_i]$. Then, under steady states, πth terminal's queue will be saturated or unbonded if the same happens at $(\rho - 1)$st terminal and

$$E[T_\pi] < \frac{\sum_{i=1}^{N} \sigma_i + \sum_{i=1}^{\rho-1} E[l_i]}{1 - \sum_{i=\rho}^{N} E[l_i]/E[T_i]}$$

where ρ is the minimum index such that $T_\pi = T_\rho$. (Note that $\pi = \rho$ if $T_\pi > T_{\pi-1}$.)

Proof: If the πth terminal saturates, then all terminals with smaller sampling intervals saturate. Let the system complete k cycles during a period Υ. With Υ and k approaching ∞, it follows from the proof of Proposition 2 that

$$\Upsilon = k \sum_{i=1}^{N} \sigma_i + k \sum_{i=1}^{\rho-1} E[l_i] + \Upsilon \sum_{i=\rho}^{N} E[l_i]/E[T_i] - \sum_{i=\rho}^{N} E[q_i]$$

The second term in the right denotes message transmission time by the terminals with saturated queues, which transmit at every cycle, i.e., k times in Υ. The message transmission time in the remaining terminals is given as the difference between the third and fourth terms. $E[\tau]$ is obtained in the same way as in the proof of Proposition 2. Queue at the πth terminal will be unbounded if $E[T_\pi] < E[\tau]$. ∎

Corollary to Proposition 5: Given that only terminals 1, ..., j saturate, $E[\tau]$ is given by

$$E[\tau] = \frac{\sum_{i=1}^{N} \sigma_i + \sum_{i=1}^{j} E[l_i]}{1 - \sum_{i=j+1}^{n} E[l_i]/E[T_i]}$$

Proof: The proof follows directly from the expression

$$\Upsilon = k \sum_{i=1}^{N} \sigma_i + k \sum_{i=1}^{\rho-1} E[l_i] + \Upsilon \sum_{i=\rho}^{N} E[l_i]/E[T_i] - \sum_{i=\rho}^{N} q_i E[l_i]$$

as $\Upsilon \to \infty$. ∎

Remark 10: Proposition 5 and its Corollary show how the saturated terminals affect the rest of the system. For example, if only a few terminals that execute decision-making functions are saturated under an emergency situation, the average performance of other terminals that conduct routine functions are not likely to be significantly affected. On the other hand, if multiple functions are assigned to a single terminal, all of them will suffer if that terminal saturates. This fact should be taken into account in ICCS network design. ∎

II.1 Time Delay in a Control Loop of ICCS

Nomenclature

- ϑ = sum of the propagation delay from source to destination and the electronic delay due to protocol execution beyond the source terminal. Furthermore, ϑ_i is a constant $\forall i$ but $\vartheta_i \neq \vartheta_j$ is possible for $i \neq j$. [Note: ϑ is different from σ.]
- δ = data latency of a message; $\delta = (\theta + l + \vartheta)$ where θ is the queueing delay at the source terminal and l is the message length (in time units). Sensor-to-controller data latency is denoted as δ_{sc} and controller-to-actuator data latency as δ_{ca}.
- Δ_s = time skew between the sampling instants of sensor and controller
- Δ_p = processing delay at the controller computer for executing the control law
- Θ_{sc} = sensor to controller time delay
- Θ_{ca} = controller to actuator time delay
- Θ_l = lumped network-induced delay in the loop; $\Theta_l = \Theta_{sc} + \Theta_{ca}$
- Θ_t = total delay in the control system that includes Θ_l and the effect of sampling

Remark 11: Since the station response time is a constant for a given protocol and the relative positions of sensor, controller, and actuator are fixed, ϑ_{sc} and ϑ_{ca} can be treated as constants. Δ_p is also a constant for a specific control law. ■

Remark 12: The sensor and control signals usually have fixed lengths. However, these terminals may serve other functions and additional information may be concatenated with sensor and control data within a message. Therefore, l_{sc} and l_{ca} are treated, in general, as random variables. ■

The following assumptions are based on normal operating conditions of ICCS networks [1, 2].

Assumption 2. The sampling interval T of the controller is almost identical to that of the sensor. Therefore, Δ_s varies very slowly and is considered to be a constant over a finite time window. Following Appendix A of [2], Δ_s is uniformly distributed in $[0, T)$. ■

Assumption 3. The control signal acts upon the plant immediately upon its arrival at the actuator (see Section 3 of [1]). ■

Assumption 4. The network is lightly loaded, i.e., $G < G_{cr}$ implying that $E[\tau] < E[T]$. ■

Assumption 5. Sup $\delta_{sc} < T$ and Sup $\delta_{ca} < T$ with probability 1. ■

Under the assumptions 2 to 5, Θ_{sc} and Θ_{ca} are given below [1].

$$\Theta_{sc} = \Delta_s + \mathfrak{J}(\delta_{sc} - \Delta_s)T$$

where the unit step function $\mathfrak{J}(x) := \begin{cases} 0 \text{ for } x < 0 \\ 1 \text{ for } x \geq 0 \end{cases}$

and $\Theta_{ca} = \Delta_p + \delta_{ca}$

Remark 13: In general $\Theta_{sc} > \delta_{sc}$ because the sensor waits at the receiver buffer of the controller before being processed at the next sampling interval. ■

Proposition 6: $E[\Theta_{sc}] = T/2 + E[\delta_{sc}]$

Proof: Let $f_\delta(\cdot)$ be the probability density function of δ_{sc}. Under assumption 2, $f_\delta(\phi) = 0$ for $\phi \notin [0, T)$. Let $f_{\delta\Delta}(\cdot, \cdot)$ be the joint probability density function of δ_{sc} and Δ_s which are independent, i.e., $f_{\delta\Delta}(\phi, \varphi) = f_\delta(\phi) f_\Delta(\varphi) = (1/T) f_\delta(\phi)$, $\phi, \varphi \in [0, T)$.

$$E[\Theta_{sc}] = \left(\int_0^T \int_0^T (\varphi + T\mathfrak{J}(\phi - \varphi)) f_\delta(\phi) d\varphi d\phi\right)/T$$

$$= \left(\int_0^T \varphi d\varphi \int_0^T f_\delta(\phi) d\phi\right)/T + \int_0^T f_\delta(\omega) \int_0^T \mathfrak{J}(\omega - \varphi) d\varphi d\omega$$

$$= \left(\int_0^T \varphi d\varphi\right)/T + \int_0^T \omega f_{sc}(\omega) d\omega = T/2 + E[\delta_{sc}] \quad ■$$

Corollary to Proposition 6: If $E[\theta_{sc}] = E[\theta_{ca}] = E[\theta]$ and l_{sc} and l_{ca} are constants, then

$$E[\Theta_l] = T/2 + 2E[\theta] + \Delta_p + l_{sc} + l_{ca} + \vartheta_{sc} + \vartheta_{ca}$$

$$E[\Theta_t] = T + 2E[\theta] + \Delta_p + l_{sc} + l_{ca} + \vartheta_{sc} + \vartheta_{ca}$$

Proof: The first equation follows directly from Proposition 6 by using the relationships $\Theta_l = \Theta_{sc} + \Theta_{ca}$ and $\delta = \theta + l + \vartheta$. The second equation follows from the fact that when the sampling process is viewed as a time varying delay, its average is $T/2$. ■

Remark 14: Assumption 5 may be replaced by the weaker one that neither the sensor's queue nor the controller's queue saturate. If other terminals on the network saturate, $E[\theta]$ is affected but Proposition 6 and its Corollary still hold. ■

The queueing delays are the only random variables in computation of Θ_l and Θ_t if l, ϑ, and Δ_p are assumed to be constants. By corollary to Proposition 4, $E[\theta]$ depends only on the first two moments of the cycle time τ. By Proposition 2, $E[\tau]$ is obtained in terms of the offered traffic G and the known network parameters σ_i, $i = 1, 2, \ldots\ldots, N$. We present the following definition and a proposition to find Var[τ] for obtaining $E[\theta]$.

Definition 5: M_i is the probability of message transmission by the ith terminal transmits in a cycle. ■

Lemma 1: Given that the queue capacity is 1 at the ith terminal,

$$M_i = \begin{cases} 1 & \text{for } E[T_i] < E[\tau] \\ E[\tau]/E[T_i] & \text{for } E[T_i] \geq E[\tau] \end{cases}$$

Proof: If the ith queue is not saturated, then $\Upsilon/E[\tau]$ cycles and $\Upsilon/E[T_i]$ transmitted messages are expected over a period Υ. If its queue saturates, a terminal transmits at every cycle. ■

Proposition 7: $\text{Var}[\tau] = \sum_{i=1}^{N} M_i(1 - M_i)(E[l_i])^2$

Proof: The distribution is binomial where the ith terminal has a probability M_i to transmit a message for a duration of $E[l_i]$. ■

II.1.1 Identical Traffic at all Terminals.

We specialize now to the case where all N terminals have statistically identical and independent traffic. Proposition 8 is a summary of the results obtained in the previous propositions under this specific condition.

Proposition 8: If $E[T_i] = E[T]$, $E[l_i] = E[l]$, and $\sigma_i = \sigma$ $\forall i$, then

(i) $G = NE[l]/E[T]$
(ii) $E[\tau] = N\sigma/(1 - G)$
(iii) $G_{cr} = (1 - N\sigma/E[T])$
(iv) $E[t] = E[\tau] + N(E[T] - E[\tau])(E[l]/E[T])^2$ or alternatively, $E[t] = G^2 E[T]/N + (N - G^2)\sigma/(1 - G)$

(v) $E[\Theta_{sc}] = (1 + G^2/N)E[T]/2 + [(N - G^2)/(2(1 - G))]\sigma + \vartheta_{sc} + E[l]$

(vi) $E[\Theta_{ca}] = G^2E[T]/(2N) + [(N - G^2)/(2(1 - G))]\sigma + \vartheta_{ca} + E[l] + \Delta_p$

(vii) $E[\Theta_l] = (1 + 2G^2/N)E[T]/2 + [(N - G^2)/(1 - G)]\sigma + \vartheta_{sc} + \vartheta_{ca} + 2E[l] + \Delta_p$

(viii) $E[\Theta_t] = (1 + G^2/N)E[T] + [(N - G^2)/(1 - G]\sigma + \vartheta_{sc} + \vartheta_{ca} + 2E[l] + \Delta_p$

Proof: Parts (i)-(iii) are obtained directly from Propositions 1 to 6. Part (iv) is derived from Proposition 7 as follows.

$$\begin{aligned}\mathrm{Var}[\tau] &= N(E[\tau]/E[T])(1 - E[\tau]/E[T])E[l]^2\\ &= NE[\tau](E[T] - E[\tau])(E[l]/E[T])^2\\ &= E[\tau](E[T] - E[\tau])G^2/N\end{aligned}$$

Then (iv) is obtained as

$$\begin{aligned}E[t] &= E[\tau] + \mathrm{Var}[\tau]/E[\tau] = E[\tau] + (E[T] - E[\tau])G^2/N\\ &= N\sigma/(1 - G) + (E[T] - N\sigma/(1 - G))G^2/N\\ &= \sigma(N - G^2)/(1 - G) + G^2E[T]/N\end{aligned}$$

Proofs of (v) and (vi) follow by substituting $E[\theta] = E[t]/2$ from Corollary to Proposition 4. (vii) is obtained from the relationship $\Theta_l = \Theta_{sc} + \Theta_{ca}$ and (viii) follows from the relationship $E[\Theta_t] = E[T]/2 + E[\Theta_l]$. ■

Remark 15: Using the relationship $E[l] = E[T]G/N$ from (i) in Proposition 8, the network-induced delays can be expressed in terms of G and N as follows:

$$E[\Theta_{sc}] = (N + G^2 + 2G)E[T]/(2N) + (N - G^2)\sigma/(2(1 - G)) + \vartheta_{sc}$$

$$E[\Theta_{ca}] = (G^2 + 2G)E[T]/N + (N - G^2)\sigma/(2(1 - G)) + \vartheta_{ca} + \Delta_p$$

and expressions for $E[\Theta_l]$ and $E[\Theta_t]$ follow accordingly. The above equations reveal that the average delays cannot be solely determined by the offered traffic G but they also explicitly depend on N. ■

Remark 16: Reduction in $E[T]$ has two effects that are mutually opposing. On one hand, the delay due to sampling decreases with T. On the other hand, the queuing delay θ increases since G is inversely proportional to $E[T]$. Therefore $E[\Theta_t]$ may have a minimum which means that there may exist an upper bound beyond which increasing the sampling interval may cause larger delays in the control loop.

$$\partial G/\partial E[T] = -G/E[T]$$

$$\begin{aligned}dE[\Theta_t]/dE[T] &= 1 + G^2/N - (\partial G/\partial E[T])(2G(1 - G) - N + G^2)\sigma/(1 - G)^2\\ &= 1 - G^2/N - (1 + (N - 1)/(1 - G)^2)G\sigma/E[T]\end{aligned}$$

As $G \to 1$, $\partial[E[\Theta_t]/\partial E[T] < 0$ but in addition G must satisfy $G < G_{cr}$; therefore, a minimum may or may not exist. ■

Remark 17: Suppose that $E[T]$ and G are given as constant parameters which imply constant $NE[l]$ but it does not determine N and $E[l]$ separately. The problem under consideration is which combinations of N and $E[l]$ lead to acceptably small delays.

$$\Delta\Theta_t/\Delta N = -G(G + 2)E[T]/N^2 + \sigma/(1 - G)$$

$$\begin{aligned}\Delta\Theta_t/\Delta N > 0 &\Rightarrow \sigma/(1 - G) > G(G + 2)E[T]/N^2\\ &\Rightarrow N > [G(G + 2)(1 - G)E[T]/\sigma]^{1/2}\end{aligned}$$

where $\Delta\Theta_t/\Delta N$ is the increment of Θ_t relative to the integer N. ■

Since $\sigma << E[T]$, for moderate traffic (i.e., G not very close to 1) with N not being very large, $\Delta\Theta_t/\Delta N < 0$. This implies that, in terms of delay for a given G and $E[T]$, it is better to have more terminals with short messages than few terminals with long messages. On the other hand, for N being larger than the critical value for which $\Delta\Theta_t/\Delta N > 0$, an increase in N accompanied by an appropriate decrease in $E[l]$ will increase the delay. The observation has a significant bearing on ICCS network design for selecting the number of terminals as well as the number of subscribers that individual terminals support.

Next we derive the probability density function $f_\theta(\cdot)$ of θ with the objective of evaluating the probability P_{vs} of vacant sampling at the controller [1, 2]. (Vacant sampling is defined as the phenomenon of no sensor data arrival at the controller's receiver during a sampling interval.) Evaluation of f_θ and P_{vs} are required for selection of Δ_s for optimal performance of ICCS as delineated in Section 4 of [2].

Proposition 9: The probability density function $f_\theta(\cdot)$ of the queueing delay is given by

$$f_\theta(\xi) = \left(\sum_{i=0}^{N} \pi_i \mathfrak{I}[(N\sigma + iE[l]) - \xi]\right)/E[\tau]$$

where $\pi_i = \binom{N}{i}(E[\tau]/E[T])^i(1 - E[\tau]/E[T])^{N-i}$, and

$\mathfrak{I}[\cdot]$ is the unit step function as defined earlier.

Proof: By assumption 4, $E[\tau] < E[T]$. Probability of message transmission from a given terminal during a cycle is $E[\tau]/E[T]$. The probability that exactly i terminals transmit during one cycle is π_i as defined above. If all terminals have identically distributed messages, then $N + 1$ possible values of the $E[t]$ (see Definition 4), namely, $\Upsilon_i = N\sigma + iE[l]$, $i = 0, 1, 2, \ldots, N$ exist. Following the proof of Proposition 4, the probability of $\{E[t] = \Upsilon_i\}$ is $\pi_i\Upsilon_i/E[\tau]$. Using $f_\theta(\xi \mid t = \Upsilon_i) = \mathfrak{I}(\Upsilon_i - \xi)/\Upsilon_i$, we have

$$f_\theta(\xi) = \sum_{i=0}^{N} f_\theta(\xi \mid t = \Upsilon_i)\pi_i\Upsilon_i/E[\tau] = \sum_{i=0}^{N} \pi_i \mathfrak{I}(\Upsilon_i - \xi)/E[\tau]$$ ■

Corollary to Proposition 9: The distribution function of θ is

$$F_\theta(\xi) = \sum_{i=0}^{N} \pi_i(\xi + (\Upsilon_i - \xi)\mathfrak{I}(\xi - \Upsilon_i))/E[\tau]$$

Proof: The proof follows by integration of the density function. ■

Remark 18: The statistics of data latency are necessary for optimizing ICCS performance as discussed in Section 4 of [2]. ■

Next we calculate the probability of vacant sampling at the controller [2] as a function of the skew Δ_s between the instants of sensor and controller sampling. In the following proposition, we consider the special case when the instants of message arrival at individual terminals are random. ■

Proposition 10: If the queueing delays of any two consecutive messages at the sensor terminal are independent of each other and if the sensor message length is a constant, then the probability P_{vs} of vacant sampling is given as

$$E[P_{vs}(\Delta_s)] = \left[\sum_{i=0}^{N}\left(\pi_i(\Upsilon_i^3 - \sum_{j=i+1}^{N}\pi_j(\Upsilon_j - \Upsilon_i)^3\right)\right]\Big/(6T(E[\tau]^2))$$

where the skew Δ_s is a random variable, the constant T is the sampling period of the control loop, and $\Upsilon_i := N\sigma + il$.

Proof: Vacant sampling [1, 2] occurs at the jth sample if

$\delta_{sc} < \Delta_s$ at $j-1$ and $\delta_{sc} \geq \Delta_s$ at j where $\delta_{sc} := \theta + l + \vartheta_{sc}$. Therefore,

$$P_{vs}(\Delta_s) = F_\theta(\Delta_s - l - \vartheta_{sc})[1 - F_\theta(\Delta_s - l - \vartheta_{sc})]$$

Using Corollary to Proposition 9, we have

$$P_{vs}(\Delta s) = \sum_{i=0}^{N} \pi_i[(\Delta_s - l - \vartheta_{sc}) + (\Upsilon_i - (\Delta_s - l - \vartheta_{sc}))\mathfrak{I}(\Delta_s - l - \vartheta_{sc} - \Upsilon_i)\cdot$$

$$\left[\sum_{j=0}^{N} \pi_j[\Upsilon_j - (\Delta_s - l - \vartheta_{sc})\right]\mathfrak{I}(\Upsilon_j - (\Delta_s - l - \vartheta_{sc}))/(E[\tau])^2$$

where $\Upsilon_i = N\sigma + il$, $i = 0, 1, 2, \ldots, N$.

It has been shown in Appendix A of [2] that Δ_s is uniformly distributed in $[0, T)$. On this basis

$$E[P_{vs}] = \int_0^T P_{vs}(\phi)d\phi/T = \int_{l+\vartheta_{sc}}^{\Upsilon_N + l + \vartheta_{sc}} P_{vs}(\phi)d\phi/T$$

since $P_{vs}(\Delta_s) \neq 0$ only for $l + \vartheta_{sc} < \Delta_s < \Upsilon_N + l + \vartheta_{sc}$. Proof follows by evaluation of the integral. ■

II.1.2 Nonidentical Sampling Rates for Sensor and Controller. We consider the ICCS configuration of the sensor sampling rate being faster than that of controller [2]. This implies that the time skew Δ_s varies at every sample. Let the controller sampling period be $T_c = T$ and the sensor sampling period $T_s = T/\alpha$ given that $\alpha > 1$.

Specific scenarios of the network traffic are now considered to exemplify the methods for obtaining the system performance data such as average delays. Nevertheless the analytical procedure, presented here, is independent of these restrictions, and it can be extended to more general cases.

Assumption 6. Every terminal belongs to one of the two groups with message inter-arrival time equal to either T or T_s. ■

Assumption 7. The number of terminals is equally distributed between the two groups, i.e., $N/2$ terminals in each group. ■

On the basis of the above assumptions, Proposition 1 yields

$$G = NE[l](\alpha + 1)/(2T)$$

where $E[l]$ is the average message length over all terminals.

If $G < G_{cr}$, using the above expression for G in Proposition 2 yields

$$E[\tau] = \frac{\sum_{i=1}^{N} \sigma_i}{1 - NE[l](\alpha + 1)/(2T)}$$

Defining $p_c = E[\tau]/T$, $p_s = \alpha E[\tau]/T$, and $g_c^i = \binom{N/2}{i}(p_c)^i(1-p_c)^{(N/2)-i}$, and $g_s^i = \binom{(N/2)}{i}(p_s)^i(1-p_s)^{(N/2)-i}$ for $i \leq N/2$ and, $g_c^i = g_s^i = 0$ for $i > N/2$, we have the following result.

Proposition 11: Probability, π_j that j terminals transmit in one cycle, is given as

$$\pi_j = \sum_{i=0}^{j}\left(g_c^i g_s^{j-i}\right)$$

Proof: p_c and p_s are the probabilities of transmission from terminals with message inter-arrival times (or sampling periods) T and T_s, respectively, in one cycle. g_c^i and g_s^i are the corresponding probabilities that exactly i terminals transmit. π_j is probability that a total of j terminals from both groups transmit in one cycle. Proof follows directly. ■

Corollary to Proposition 11: The density function for queueing delay θ can be expressed as

$$f_\theta(\xi) = \left(\sum_{i=0}^{N} \pi_i \mathfrak{I}[N\sigma + iE[l] - \xi]\right)/E[\tau] \quad ■$$

For stable operation of the ICCS, the delays incurred by the sensor and the controller must be bounded in a probabilistic sense. From the point of view of ICCS network design, we modify the earlier assumption 5 as Sup $\delta_{sc} < T_s$ and Sup $\delta_{ca} < T_s$ with probability 1. Smaller δ_{sc} also reduces the probability of vacant sampling as pointed out in [2].

Proposition 12: Sensor-to-controller delay Θ_{sc} for nonidentical sampling ($T_s = T/\alpha$, $\alpha > 1$) is given as

$$E[\Theta_{sc}] = \frac{T}{2\alpha} + \frac{E[\tau]}{2} + \frac{NE[l]^2}{4T}((1+\alpha)T - (1+\alpha^2)E[\tau]) + E[l_{sc}] + \vartheta_{sc}$$

provided that Sup $\delta_{sc} < T_s$ with probability 1.

Proof: The time skew Δ_s, between sensor and controller sampling instants, is time-varying and uniformly distributed in $[0, T_s)$ as shown in Appendix A of [2]. Therefore we have

$$\Theta_{sc} = \Delta_s + T_s\mathfrak{I}(\delta_{sc} - \Delta_s) \text{ where } \delta_{sc} = (\theta_{sc} + l_{sc} + \vartheta_{sc})$$

Proof follows by substituting the results of Propositions 6 and 4 and Corollary to Proposition 4 in the above as well as by making use of binomial distribution of messages at all terminals.

$$\text{Var}[\tau] = [p_s(1-p_s) + p_c(1-p_c)]E[l]^2N/2$$
$$= [(1+\alpha)T - (1+\alpha^2)E[\tau]]E[\tau]E[l]^2N/(2T^2) \quad ■$$

Remark 19: Proposition 12 is useful for analytically evaluating the expected values of the lumped delay Θ_l and the total system delay Θ_t. This information is necessary (although not sufficient) for stability analysis of ICCS. ■

III Summary and Conclusions

This paper presents performance analysis of integrated communication and control system (ICCS) networks as a continuation of earlier work [1-2] which provides the necessary background. Statistical models of network-induced delays, that are analytically derived in this paper, complement the finite-dimensional state-space model of the ICCS that has been reported in [1, 2]. Since the parameters of the above state-space model are stochastic processes that are dependent on the network-induced delays, the knowledge of statistics of these delays is necessary for investigating the stability and dynamic performance of ICCS. These models of network-induced delays provide critical information for ICCS design, and largely mitigate the need for repetitive simulation runs for numerically obtaining the statistical characteristics of delays.

Acknowledgment

The authors would like to thank the reviewers for their helpful suggestions which have been incorporated into this paper.

References

1 Halevi, Y., and Ray, A., "Integrated Communication and Control Systems: Part I—Analysis," ASME JOURNAL OF DYNAMIC SYSTEMS, MEASUREMENT AND CONTROL, Vol. 110, No. 4, Dec. 1988, pp. 367-373.

2 Ray, A., and Halevi, Y., "Integrated Communication and Control Systems: Part II—Design Considerations," ASME JOURNAL OF DYNAMIC SYSTEMS, MEASUREMENT AND CONTROL, Vol. 110, No. 4, Dec. 1988, pp. 374-381.

3 Ray, A., "Performance Evaluation of Medium Access Control Protocols for Distributed Digital Avionics," ASME JOURNAL OF DYNAMIC SYSTEMS, MEASUREMENT AND CONTROL, Vol. 109, No. 4, Dec. 1987, pp. 370-377.

4 Ray, A., "Distributed Data Communication Networks for Real-Time Process Control," *Chemical Engineering Communications*, Vol. 65, Mar. 1988, pp. 139-154.

5 Hopkins, A. L., et al., "System Data Communication Structures for Active-Control Transport Aircraft," Vol. II, *NASA Report CR 165774*, June 1981.

6 Lane, J. W., "Impact of Digital Control Evolution on Control Systems Design Philosophy," *Proc. Canadian Conference on Industrial Computer Systems*, Ottawa, Canada, May 1984.

7 Ray, A., "Computer Networking for Manufacturing Systems," *IEEE Network Magazine*, Vol. 2, No. 3, May 1988, pp. 40-47.

8 Ray, A., and Phoha, S., "Research Directions in Networking for Manufacturing Systems," ASME *Journal of Engineering for Industry*, Vol. 111, No. 2, May 1989, pp. 40-46.

9 Bertsekas, D., and Gallager, R., *Data Networks*, Prentice Hall, Englewood Cliffs, NJ, 1987.

10 Hammond, J. L., and O'Reilly, P. J. P., *Performance Analysis of Local Computer Networks*, Addison-Wesley, Reading, MA, 1986.

11 Kleinrock, L., *Queueing Systems, Volume I: Theory*, John Wiley, New York, 1975.

12 *SAE Linear Token Passing Multiplex Data Bus*, AS4074.1, Version 30, May 1987.

13 Stark, H., and Woods, J. W., *Probability, Random Processes, and Estimation Theory*, Prentice Hall, Englewood Cliffs, NJ, 1986.