

# Network Access Protocols for Real-Time Distributed Control Systems

ASOK RAY, SENIOR MEMBER, IEEE

**Abstract**—Analysis and design for real-time distributed data communication and control system networks is an evolving field which requires interactions between the disciplines of communication systems and control systems engineering. Concepts of pertinent network parameters are introduced and an analytical base for system design is established. A simulation methodology has been developed for performance evaluation of real-time control system networks for distributed dynamical processes like autonomous manufacturing, electric power and chemical plants, and aircraft and spacecraft. The network-induced delays, resulting from data latencies, data loss, and missynchronism between system components, degrade the dynamic performance of the controlled process and are a source of potential instability. The impact of delays induced by different network access protocols on the control system performance are investigated. The results of combined discrete-event and continuous-time simulation are presented to this effect.

## I. INTRODUCTION

**D**ISTRIBUTED information processing is essential for control and operation of complex dynamical processes like autonomous manufacturing, large-scale electric power and chemical plants, and advanced aircraft and spacecraft. This requires a number of spatially dispersed computers and intelligent terminals to perform, in real time, diverse but interrelated functions ranging from closed-loop control of essential process variables to routine maintenance support and information display [1]–[4]. The requirements for a communication network for a distributed control system may vary for specific applications. For example, in an autonomous manufacturing environment, the prime objective of distributed data communication and control system (DDCCS) networks is the integration of computer-controlled complexes of robots, automated machine tools, material-handling devices, and guided vehicles on the shop floor, with information processing across local and remote stations. Similarly, in nuclear power plants, distributed control systems could regulate the essential and time-dependent process variables, like reactor thermal power, feedwater flow, and electric power, as well as provide the necessary support for plant monitoring, fuel scheduling, and preventive maintenance. The major challenge in an advanced

aircraft is to design a fly-by-wire (or fly-by-light) integrated control system that will ensure safe and reliable missions.

Two basic methods are available for interconnecting the components of a distributed information-process and control system [3], [4]: 1) point-to-point dedicated connections, and 2) asynchronous time-division multiplexed (TDM) networking. The major advantages of multiplexed networks over point-to-point connections include reduced wiring and power requirements, flexibility of operations, evolutionary design process, and ease of maintenance, diagnostics, and monitoring. However, the data latency in multiplexed communication networks could introduce unacceptable delays and thereby degrade the dynamic performance of the control system.

The sensor and control data in a DDCCS network are subject to time-varying delays due to data latency of messages in the communication network [5], [6] in addition to the sampling time delay that is inherent in digital control systems. This is evident in very fast processes (e.g., flight dynamics in tactical aircraft [5]). The effects of data latency on the dynamic performance and stability of feedback control systems are often ignored in relatively slow processes, such as those encountered in manufacturing and processing plants. However, as the number of users on the network increases, the augmented traffic causes a larger data latency, to a point where its impact on the performance of some of the control loops (sharing the network) can no longer be ignored. The detrimental effects of data latency on the system dynamic performance are further aggravated by missynchronism between the control-loop components as well as by loss of messages resulting from the saturation of buffers at the terminals and the data corruption by noise in the network medium.

In a distributed control system environment, the network traffic would consist of both real-time and non-real-time data packets. The real-time data are usually generated as short messages and must be communicated on a periodic basis as the sampling interval  $T$  of a control system is predetermined according to the dynamic characteristics of the controlled process. For example,  $T$  could be very small, that is, on the order of 50 ms, for communications between two robots working on the same transfer station. The non-real-time traffic is generated by the design and administrative information, such as production orders, part drawings, and status reports, and is transmitted nonperiodically and much less frequently. These messages are usually long and do not need to be processed within the constraints of a real-time environment. Whereas data latency is critical for real-time messages, non-real-time mes-

Paper IUSD, approved by the Industrial Control Committee of the IEEE Industry Applications Society for presentation at the 1986 Industry Applications Society Annual Meeting, Denver, CO, September 28–October 2. This work supported in part by the National Science Foundation under research grant DMC-8707648. Manuscript released for publication December 1, 1987.

The author is with the Mechanical Engineering Department, 208 Mechanical Engineering Building, The Pennsylvania State University, University Park, PA 16802.

IEEE Log Number 8820182.

sages need the assurance of accurate delivery via the shared communication medium.

Occasional losses of (real-time) sensor and control data in a feedback loop can be tolerated whereas the additional delay due to retransmission of erroneous signals may seriously degrade the system dynamic performance. Therefore the real-time data may not use the acknowledgment option; messages with detected errors would be simply discarded. On the other hand, the acknowledgment option needs to be adopted for non-real-time messages to ensure accurate delivery of data. Under these circumstances the real-time data should have a higher priority over the non-real-time data. This may delay the delivery of the non-real-time data but will not compromise the data integrity because of the built-in packet retransmission capabilities of the network protocol.

The mixing of real-time unacknowledged data and delayable acknowledged data is similar to what is encountered in packetized voice and data communication networks [7]. The preceding concept can be extended to integrated voice and data communications if the real-time data packets are replaced by voice packets.

Analysis and design of an integrated network for real-time DDCCS require interactions between the disciplines of communication systems and control systems engineering. It may be appropriate to bring out the notions of *delay*, as it is used in the two disciplines in somewhat different manners. In communication systems, the delay is primarily referred to as *queueing delay* and *data latency*, which are associated with only those messages that successfully arrive at the destination terminal [5]; messages that are corrupted by noise or deleted due to queue saturation at the transmitter buffer of the source terminal are not considered for this purpose. In real-time control systems the delay is related to the question: How old is the data that is currently used?. When no messages are rejected, the two notions of delay are identical; otherwise they are different.

Although sufficient research papers in modeling and simulation of communication protocols have been published [8], the significance of network-induced delays relative to the stability of feedback control systems apparently has not been addressed except in a few cases [5], [6]. A systematic methodology needs to be developed for design of real-time control systems that are subject to network-induced delays. A generic methodology, based on simulation analysis, has been developed for evaluating the system dynamic performance. Further analytical work in this evolving area is under active research [9], [10]. The major objectives of this paper are

- to analyze the characteristics of network-induced delays in distributed communication and control systems; and
- to introduce a simulation methodology for evaluating the system dynamic performance and present pertinent results of simulation experiments for diverse network traffic.

The paper is organized in five sections. Section II introduces concepts of pertinent network parameters and establishes an analytical base for system design. The structure of a simulation methodology that has been developed for performance evaluation of real-time distributed control systems

is described in Section III. Simulation results under diverse traffic are presented and discussed in Section IV. The summary and conclusions of this evolving research are presented in Section V.

## II. PERFORMANCE ANALYSIS OF DISTRIBUTED CONTROL SYSTEM NETWORKS

This section emphasizes analytical modeling of network-induced delays with the objective of evaluating the dynamic performance of distributed control systems. Other pertinent parameters for DDCCS performance evaluation, such as network reliability and data frame error rate, have not been addressed in this paper. The reason is that reliability and availability are largely hardware-dependent and are usually analyzed during the hardware design phase, when detailed specifications become available. Data frame errors depend on the raw bit error rate in the network medium and the error detection algorithm provided in the protocol (e.g., cyclic redundancy check). Raw bit error rates in (coaxial cable) network media in a typical industrial environment range from  $10^{-12}$  to  $10^{-8}$  [11]; corresponding values for optical fiber media are significantly lower. Although message retransmissions or rejections (as a result of detected frame errors) apparently have no significant bearing on data latency and throughput, under these circumstances it is the undetected frame errors that degrade control system reliability. Therefore the issue of undetected frame error rate should be examined along with reliability.

The following conditions were assumed for analytical modeling of the DDCCS network.

- 1) Each terminal has one transmitter queue and one receiver queue.
- 2) First in, first out (FIFO) ranking is adopted at each terminal's transmitter queue, and its buffer capacity is limited to an *a priori* assigned constant; in the event of queue saturation, the first message (i.e., the message at the front end of the queue) is rejected and the stack is pushed down to accommodate the new arrival at the rear end of the queue.
- 3) The receiver queue capacity at each terminal is one.

Now we proceed to develop analytical models for network performance evaluation by introducing pertinent definitions.

*Definition 1:* Network traffic is defined to be deterministic if the message length and the message interarrival time are constants.

*Definition 2:* Network traffic is defined to be simultaneous if it is deterministic and if the messages arrive at all terminals simultaneously at each sampling period.

*Definition 3:* For a message, the word count  $W_c$  is defined as

$$W_c(L) = \begin{cases} \text{int}(L/W_d), & \text{if } \text{rem}(L, W_d) = 0 \\ \text{int}(L/W_d) + 1, & \text{if } \text{rem}(L, W_d) > 0 \end{cases}$$

where

$L$  length (in bits) of the data part of a message,

excluding the bits due to forming and overhead,  
 $W_d$  number of data bits per word,  
 $\text{int}^*(*)$  the integer part of \*

and  $\text{rem}(a, b) \triangleq a - [\text{int}(a, B)]b$ .

**Definition 4:** Frame length  $L'$  (bits) of a message is defined as

$$L' = W_f W_c(L) + \Omega$$

where

$W_f$  length (bits) of a formatted word,  
 $\Omega$  overhead (bits) associated with a message.

**Definition 5:** Offered traffic  $G$  is defined as

$$G = (E[L]N)/RE[T]$$

where

$N$  number of active terminals in the network,  
 $R$  data transmission rate in bits/s,  
 $T$  message interarrival time at a terminal,  
 $E[*]$  expected value of \*.

**Definition 6:** Cycle time  $\tau$  is defined as

$$\tau = \sum_{i=1}^n [L'_i/R] + N\sigma$$

where  $\sigma$  equals the average bus idle time prior to the beginning of a message transmission and subscript  $i$  corresponds to the terminal  $i$ .

**Remark 1:** For a given traffic, cycle time may be interpreted as the token time required to complete the transmission of one message from each of the  $N$  terminals.

**Definition 7:** For a given traffic and normalized cycle time,  $G'$  is defined as the ratio of the expected values of the cycle time and the message interarrival time, i.e.,  $G' = E[\tau]/E[T]$ .

**Remark 2:** For a given  $G$ , individual protocols may load the medium to different levels and thus influence the performance of the DDCCS to different degrees. Therefore  $G$  is used as a parameter for selection of network access protocols instead of  $G'$ . A limit of  $G$  above which a given protocol is expected to overload the medium, resulting in message rejection, needs to be specified.

**Definition 8:** The critical offered traffic  $G_{cr}$  for a protocol is defined as the largest offered traffic for which no message frame is rejected, due to queue saturation for deterministic traffic under steady state.

**Remark 3:**

$$G' \leq 1 \text{ if } G = G_{cr}.$$

**Definition 9:** Queuing delay  $\delta_q$  of a successfully transmitted message is the difference between the instant of arrival of the message at the transmitter queue of the source terminal and the instant of transmission of its first bit on the medium.

**Definition 10:** Data latency  $\delta$  of a successfully transmitted message is defined as the difference between the instant of arrival of the message at the transmitter queue of the source

terminal and the instant of reception of its last bit at the destination terminal.

**Remark 4:** Data latency and queuing delay are approximately related as

$$\delta = \delta_q + L'/R$$

where  $R$  equals the data transmission rate in bits/s.

In a feedback control loop, messages are transmitted via the network medium from the sensor terminal to the controller terminal and from the controller terminal to the actuator terminal. Thus the control system in Fig. 1 is subject to time-varying delays due to data latency, and its dynamic performance is dependent on the traffic in the network. Additional definitions of network delay parameters are introduced below.

**Definition 11:** Sensor-controller delay  $\Theta_{sc}^k$  of a successful message at the  $k$ th sensor sampling instant is defined as the interval between the instant of its arrival at the sensor terminal's transmitter buffer and the instant when the controller starts processing this message.

**Remark 5:** If the sensor and controller have the same sampling time  $T$ ,  $\Theta_{sc}$  can be expressed in terms of the sensor-to-controller data latency  $\delta_{sc}$ , and the time skew  $\Delta_s$  between the sensor and controller sampling instants ( $0 < \Delta_s \leq T$ ) as

$$\Theta_{sc}(\delta_{sc}, \Delta_s) = \begin{cases} \Delta_s & \text{for } \delta_{sc} < \Delta_s \\ \ell T + \Delta_s & \text{for } (\ell - 1)T + \Delta_s \leq \delta_{sc} < \ell T + \Delta_s \end{cases}$$

where  $\ell$  is a positive integer.

**Definition 12:** Controller-actuator delay  $\Theta_{ca}^j$  of a successful message at the  $j$ th controller sampling instant is the interval between the instant when the controller starts processing the available sensor data and the instant of arrival of the last bit of the processed control command at the actuator.

**Remark 6:** In contrast with the sensor data, which may wait at the controller's receiver queue before being processed by the controller, the control command acts upon the plant immediately after arriving at the actuator terminal. This happens because the controller is scheduled to generate signals at constant intervals whereas the actuator operations are essentially asynchronous.

**Remark 7:** Since the delay  $\Theta_{sc}$  is time-varying, the controller may use the sensor data generated at the current or earlier samples. The delay  $\Theta_{ca}$  to which the control command sequence is subject is also time-varying. This implies that the intervals between successive arrivals of control commands at the actuator terminal may not be a constant.

**Remark 8:** The characteristics of the delay  $\Theta_{sc}$  are different from those of  $\Theta_{ca}$  in the case where more than one sensor data arrive at the controller during one of its sampling intervals. Since the controller operates in discrete time, unlike the actuator, which is essentially a continuous-time device, only the most recent sensor data is accepted. This phenomenon is called *message rejection* at the controller's receiver. On the other hand, if no fresh sensor data is available during a sampling interval, the previous sensor data is used

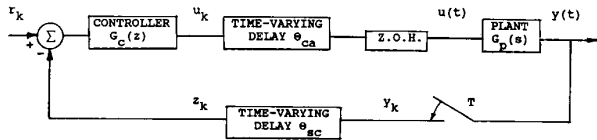


Fig. 1. Delayed control system.

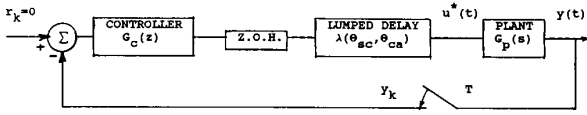


Fig. 2. Lumped delayed system.

to compute the control command. This phenomenon is called *vacant sampling* at the controller.

*Remark 9:* Since the delay  $\Theta_{sc}$  is time-varying, the controller may use the sensor data generated at the current or earlier samples. The delay  $\Theta_{ca}$  to which the control command sequence is subject is also time-varying. This implies that the intervals between successive arrivals of control commands at the actuator terminal may not be a constant.

Under random traffic conditions the control system delays  $\Theta_{sc}$  and  $\Theta_{ca}$  are stochastic processes. Even for deterministic traffic, i.e., periodic traffic with constant message lengths, these delays are both time-varying and therefore, in general, may not be lumped together. However, if the digital control algorithm is linear time-invariant, these two delays could be lumped under certain conditions. A statement of this property is given below.

*Proposition 1:* The time-varying delays  $\Theta_{sc}$  and  $\Theta_{ca}$  in Fig. 1 can be equivalently lumped together as  $\lambda$  in Fig. 2, provided that

- 1) the sensor and controller have identical sampling periods  $T$ ,
- 2) there is no message rejection at the sensor and controller terminals,
- 3)  $\lambda(t) \geq 0$  for  $\forall t$ ,
- 4)  $\lambda(kT + \Phi^k) = \Phi^k$ , and
- 5)  $\lambda(kT + \phi) \leq \phi \quad \forall \phi \in (\Phi^k, T + \Phi^{k+1})$ ,

where

$$\Phi^k = \Theta_{sc}^k + \Theta_{ca}^k.$$

*Proof of Proposition 1:* The lumped and unlumped delayed systems in Figs. 1 and 2 are equivalent with respect to the input/output relation if, for a given input  $y(t)$  to the controller, the outputs  $u(t)$  and  $u^*(t)$  are identical. The function  $u(t)$  is piecewise constant and is given by

$$u(t) = u_k \quad \text{if } kT + \Phi^k < t \leq (k+1)T + \Phi^{k+1}.$$

The function  $u^*(t)$  is given by

$$u^*(t) = u_k \quad \text{if } kT \leq (t - \lambda(t)) < (k+1)T.$$

For  $u(t)$  and  $u^*(t)$  to be equivalent, the time ranges in the preceding equations must be the same. Condition 4 guarantees

$$u(kT + \Phi^k) = u_k$$

and condition 5) assures that this value remains constant as

long as  $t < (k+1)T + \Phi^{k+1}$ . This follows from the fact that

$$t - \lambda(t) = kT + \phi - \lambda(kT + \phi) \geq kT$$

at the instant  $t = kT + \phi$ .

Thus the input to the plant is  $u(t) = u_k$ .

*Remark 10:* As can be seen from the conditions 4) and 5) in Proposition 1,  $\lambda(t)$  is not unique, and any admissible function is equally applicable.

*Remark 11:* If  $\lambda(t)$  is a constant, or can be approximated by a constant for all  $t$ , even though  $\Theta_{sc}$  and  $\Theta_{ca}$  could be individually time-varying, then the control system can be designed using conventional frequency-domain techniques.

*Proposition 2:* For periodic traffic and  $G > G_{cr}$ , i.e.,  $\tau < T$ , implying that some messages are rejected due to queue saturation, let  $\tau = mT + \vartheta$ , where  $m$  is a positive integer and  $\vartheta \in (0, T]$ . Then

- a) the probability of message rejection is  $(\tau - T)/\tau$ ;
- b) following a successful message either  $(m-1)$  or  $m$  messages are rejected. The probability of  $m$  rejections is  $P_m = \vartheta/T$ .

*Proof of Proposition 2a):* The expected value of the number of messages that arrive at each terminal during the interval  $T\tau$  is  $\tau$ . The expected value of the number of messages that are successfully transmitted during the same period is  $T$ . Hence the probability of a message to be successfully transmitted in  $T/\tau$  and that of rejection is  $(\tau - T)/\tau$ .

*Proof of Proposition 2b):* Let the messages periodically arrive at a certain terminal at the instants  $jT$ . Upon successful transmission of a message at  $t_i = kT + \alpha$ ,  $\alpha \in (0, T]$ , the next successful message is transmitted at  $t_i + \tau = (k+m)T + \alpha + \vartheta$ . If  $(\alpha + \vartheta) > T$ , then  $m$  messages are rejected; otherwise,  $(m-1)$  messages are rejected.

To obtain the probability of message rejections, we make use of the following lemma to show that the probabilities of the events  $\{(\alpha + \vartheta) > T\}$  and  $\{(\alpha + \vartheta) \leq T\}$  are  $\vartheta/T$  and  $(T - \vartheta)/T$ , respectively.

*Lemma for Proposition 2:* For  $\tau > T$ , if  $\alpha_j$  is defined as

$$\alpha_j = \text{rem}(\alpha_0 + j\tau, T)$$

where  $\text{rem}(\cdot, \cdot)$  is as stated in Definition 3, then

- i)  $\alpha_0 = 0$  implies that  $\alpha_j$  is discretely uniformly distributed in  $(0, T]$ ;
- ii) if  $\alpha_0$  is continuously uniformly distributed in  $(0, T]$ , then  $\alpha_j$  is also continuously uniformly distributed in  $(0, T]$ .

*Proof of Lemma i):* Let  $\gamma$  be the greatest common divisor of  $\tau$  and  $T$ , i.e.,  $\tau = l\gamma$  and  $T = \mathcal{P}\gamma$ , where  $l$  and  $\mathcal{P}$  are co-prime integers and  $l > \mathcal{P}$ . Then  $\alpha_j = \text{rem}(jl, \mathcal{P})\gamma$  and consequently  $\alpha_j$  assumes  $\mathcal{P}$  distinct values  $j\gamma$  for  $j = 1, 2, \dots, \mathcal{P}$ . Since  $\alpha_{\mathcal{P}+i} = \alpha_i$ , the sequence  $\{\alpha_j\}$  is cyclic, with a period of  $\mathcal{P}$ , and the distribution of  $\alpha_j$  is discretely uniform.

*Proof of Lemma ii):* In this case  $\alpha_j$  assumes the  $\mathcal{P}$  distinct values of  $\alpha_0 + k\gamma$  with equal probability, where  $-\text{int}[\alpha_0/\gamma] \leq k \leq \mathcal{P} - \text{int}[\alpha_0/\gamma] - 1$ . Since  $\alpha_0$  is continuously

uniformly distributed in  $(0, T]$ ,  $\alpha_j$  is also continuously uniformly distributed in  $(0, T]$ .

### III. SIMULATION MODELS FOR DDCCS NETWORK PERFORMANCE EVALUATION

Protocols based on asynchronous time-division multiplexing (TDM) [11] are suitable for real-time DDCCS networks that are subject to a combination of periodic, nonperiodic, and bursty traffic. The advantages of distributed controlled-access protocols, like token bus and token ring, over random-access protocols, like CSMA/CD, have been presented in our recent publications [5], [6].

Two standard distributed controlled-access protocols, namely the Society of Automotive Engineers (SAE) token ring [12] and linear token bus [13], which offer larger data communication speeds and less complexity than the respective IEEE Std. 802 family protocols [14], [15], were selected as the candidate protocols for the DDCCS network. The performance of the SAE linear token bus and token ring protocols was compared with respect to the (centralized controlled-access) MIL Std. 1553B protocol [16], which has been used extensively in digital avionic systems of fly-by-wire military aircraft control systems.

Simulation modes for the SAE linear token bus and token ring and the MIL Std. 1553B protocols were developed for DDCCS performance evaluation using combined discrete-event and continuous-time techniques. Discrete-event models are suitable for representing the time-ordered sequences of operations that are encountered in network protocols. On the other hand, continuous-time models are essential for solving the differential and difference equations associated with plant and controller dynamics. The coupling of two types of simulation permits performance evaluation of the network as a dynamic element of the closed-loop control system. This allows for monitoring the effects of delays introduced by the network.

The simulation program for the DDCCS consists of two subsystem models: 1) a discrete-event model of the network and 2) a continuous-time model of process dynamics and a discrete-time model of the controller. The program structure is modular, i.e., any one of the protocol models (e.g., SAE linear token bus, SAE token ring, or MIL Std. 1553B) can be inserted in the simulation program while operating on the same control system model, and vice versa. A schematic diagram for the combined discrete-event and continuous-time model structure of the DDCCS network is shown in [6]. The network subsystem model consists of two independent but interacting submodels: 1) the message generation submodel and 2) the protocol submodel.

The message-generation submodel has an identical structure for all types of protocols and is driven by an external pool of messages that arrive at the network system either periodically or at random intervals of time (Poisson arrival, for example). Similarly, the message lengths can be either constants or randomly distributed (exponential, for example). When a new message arrives at the system from the external message generator, the message attributes are defined to establish the message identity in the following ways.

- Time of arrival—this is the instant at which the arrival of a message at the transmitter queue is recorded.
- The message information length (overhead not included).
- The source terminal, i.e., the terminal from which the message is generated.
- The destination terminal—this could be any terminal on the network other than the source terminal.
- The message priority, if applicable.

The protocol submodel essentially represents the algorithm of the network access protocol under consideration. Different submodels, each of which identically enters the simulation program as a subroutine, have been developed for SAE linear token bus, SAE token ring, and MIL Std. 1553B protocols. Although internal algorithms of the individual protocol submodels are different, their interactions with the message generation submodel and the models of the plant and controller are identical. For example, the attributes of the generated messages are captured by the protocol submodel, which in turn regulates the delays for exchange of messages between the plant and controller terminals. The interactions between the plant and controller models involve exchange of sensor and control signals which undergo time-varying delays introduced by scheduled events in the network model.

Key considerations in the choice of a language for the DDCCS simulation are: 1) combined discrete-event and continuous-time simulation capability; 2) programming flexibility and software portability; 3) verification and debugging capability; 4) built-in statistical testing capability; and 5) automatic ordering of scheduled events. After considering a number of simulation languages [17], SIMAN, which offers all of these key features [18], was chosen as the software for coding and implementation of the DDCCS model.

### IV. SIMULATION RESULTS AND DISCUSSION

The network model was operated under steady-state conditions while disturbances were applied to the control system model to observe the impact of the network-induced delays on system dynamics. The following network configuration was employed as the basic model for simulation experiments and subsequent comparisons of the results.

- The network consists of 31 terminals or nodes (this is the upper limit for MIL Std. 1553B).
- Terminal 1 operates as the sensor and actuator terminal, with its transmitter queue serving the sensor and its receiver queue serving the actuator in the closed-loop control system.
- Terminal 2 operates as the controller terminal, with its transmitter queue handling actuator commands and the receiver queue handling sensor data.
- Terminals 1 and 2 have periodic traffic with fixed-length messages, with the data part  $L = 64$  bits and a sample period of 10 ms.
- Terminals 3 to 31 are modeled as ordinary nodes of the network, where the expected value of the message inter-arrival time was set to 10 ms and message lengths were varied to regulate the offered traffic in the network.
- The digital control system is single-input single-output with unity feedback. The plant transfer function in Fig.

TABLE I  
STEADY-STATE PERFORMANCE OF NETWORK ACCESS PROTOCOLS

	Average Data Latency ( $\mu s$ )								
	MIL Std. 1553B			SAE Linear Token Bus			SAE Token Ring		
	$G = 0.2$	$G = 0.7$	$G = 1.2$	$G = 0.2$	$G = 0.7$	$G = 1.2$	$G = 0.2$	$G = 0.7$	$G = 1.2$
$Q = 1$	3231	5308	5508	1077	3659	5404	1116	3690	5407
$Q = 2$	3231	15 309	15 508	1077	3659	15 404	1116	3690	15 407
Exponentially Distributed Message Length and Constant Interarrival Time									
$Q = 1$	3101	5258	5477	1079	3687	5338	1123	3714	5334
$Q = 2$	3101	15 240	15 482	1079	3708	15 178	1123	3738	15 228
Constant Message Length and Exponentially Distributed Interarrival Time									
$Q = 1$	2095	3842	6206	110	499	2338	127	508	2356
$Q = 2$	3367	9393	15 555	110	567	7684	127	577	7720
Exponentially Distributed Message Length and Exponentially Distributed Interarrival Time									
$Q = 1$	2066	3807	5914	118	668	2500	135	676	2535
$Q = 2$	3260	9189	15 131	119	820	7899	136	835	8096

1 was selected as  $Gp(s) = 1/[(0.3s + 1)(0.03s + 1)]$ , and the analog equivalence of the transfer function of the digital controller as  $Gc(s) = 7(s + 5)/s$ , where  $s$  is the Laplace transform variable.

#### A. Steady-State Performance of the Network

Under each traffic condition and for each protocol the simulation model was run for 2 million  $\mu s$ , i.e., 2 s. Built-in statistical testing routines provided in SIMAN [18] were employed to obtain the average values, standard deviations, and confidence intervals at different percentiles of the data latency. Because of the time-varying nature of the network's operations, these steady-state statistical parameters, except the average values, have no major bearing on the DDCCS network design.

In many applications, such as routine information exchange between the design office and factory floor in an autonomous manufacturing environment, message interarrival time  $T$  and message length  $L$  are random variables. Performance evaluation of the DDCCS network was conducted under four different traffic scenarios: 1) constant  $T$  and constant  $L$ , i.e., deterministic traffic; 2) constant  $T$  and exponentially distributed  $L$ ; 3) exponentially distributed  $T$  and constant  $L$ ; and 4) both  $T$  and  $L$  being exponentially distributed. Simulation results for different combinations of constant and exponentially distributed cases of  $T$  and  $L$  are listed in Table I for the three protocols under consideration.

Each protocol was simulated at two values of queue limits:  $Q = 1$  and  $Q = 2$ , and three values of offered traffic:  $G = 0.2$ ,  $G = 0.7$ , and  $G = 1.2$ . On the basis of Definitions 3 and 4 in Section II and the data provided in Appendix C of [5], the critical offered traffic  $G_{cr}$  (see Definition 8) for  $N = 31$  and  $T = 10$  ms was computed to be 0.993, 0.986, and 0.523 for the SAE token bus, SAE token ring, and MIL Std. 1553B, respectively.

Based on the simulation results provided in Table I, the following inferences can be drawn.

- For all protocols with  $G < G_{cr}$ , the average data latency

is improved with larger variability in traffic. The randomness of  $T$  contributes more to performance improvement than that of  $L$ . However, no general conclusion can be drawn for  $G > G_{cr}$ , especially if  $G$  significantly larger than  $G_{cr}$ .

- SAE linear token-bus and token-ring protocols yield almost equivalent performances, which are superior to that of the MIL Std. on the average.
- In terms of average data latency, the SAE linear token bus and token ring exhibit their best performance under Poisson arrival of messages with constant lengths.
- Results indicate that the linear token-bus and token-ring protocols clearly excel beyond the MIL Std. 1553B in terms of average data latency. The reason for the relatively poor performance of the MIL Std. is its longer bus idle time and larger overhead due to its message forming structure. The linear token bus yields a slightly better performance than the token ring, but the difference may not be of statistical significance.

The steady-state performance data, i.e., the average data latencies, in Table I are significant from a network design point of view. However, for control system design it is the time-varying characteristics of the delays that also need to be taken into account.

#### B. Dynamic Performance of the Distributed Control System

Since the data latency characteristics of the SAE token ring and linear token bus are similar, to a large extent, under normal operating conditions of the network, the transient responses of the digital control system in Fig. 1 were generated with the linear token-bus and MIL Std. 1553B protocols under different traffic conditions. Transient responses under identical conditions were obtained for an equivalent centralized digital controller, where the control loop is not subject to any delay except the usual sampling time delay. These responses were used as a reference for evaluating the individual protocols in the DDCCS.

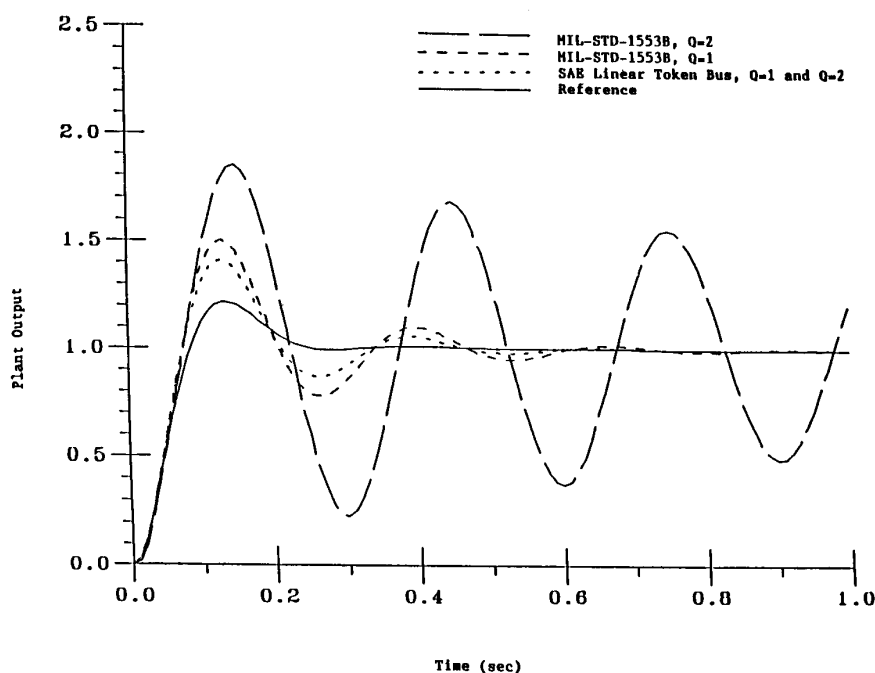


Fig. 3. Transient response of plant output at  $G = 0.7$ .

For simultaneous traffic (see Definition 2 in Section II) in all terminals (except the controller terminal 2), transient responses of the output were obtained for a unit step increase in the reference input from an initial steady-state condition. In real applications the terminals may not be synchronized, and some of them may generate random traffic; therefore the traffic would not be simultaneous. However, simulation results with simultaneous traffic allow a better insight of the characteristics of network-induced delays and their impact on the control system performance. A typical scenario is discussed below.

Fig. 3 shows transient responses for  $G = 0.7$ , where  $G$  exceeds  $G_{cr}$  for MIL Std. 1553B only, i.e., some messages are lost due to queue saturation. The performance of the MIL Std. becomes much worse, leading to instability, if  $Q$  is increased to 2 for the transmitter buffers of all terminals, including those for the sensor and the controller. The cause of the degradation is an additional queuing delay of 10 ms in each terminal in which a message never arrives at an empty buffer. This phenomenon is analytically derived in Proposition 2 of [5].

Simulation experiments were carried out for random traffic with Poisson arrival and exponentially distributed message lengths in terminals 3 to 31 to investigate the effects of stochastic delays on the dynamic performance of the feedback loop in the DDCCS network. Transient responses for the plant output with random traffic were found to have characteristics similar to those with deterministic traffic. Additional results illustrating the impact of network-induced delays on dynamic performance of aircraft and chemical-plant control systems are reported in earlier publications [5], [6].

The overshoot and settling time of transients for the SAE linear token bus are, in general, superior to those for MIL Std. 1553B under identical network traffic, because MIL Std.

1553B suffers from larger data latencies due to a larger message overhead and idle time. This also entails a relatively smaller critical offered traffic for MIL Std. 1553B, resulting in a larger number of lost messages whenever offered traffic exceeds its critical value.

The preceding observations are generic in nature and are applicable to DDCCS networks for power and chemical plants, autonomous manufacturing processes, and aircraft. Although the process dynamics in different applications may vary widely, the concept of dimensionless offered traffic and the resulting delays relative to the sampling period are similar in all cases.

## V. CONCLUSION

The paper presents the results of an ongoing research in the area of analysis and design for real-time distributed data communication and control systems (DDCCS) networks. This is an evolving field which requires interaction between the disciplines of communication systems and control systems engineering. In this respect, definitions of pertinent DDCCS parameters are introduced and propositions for system analysis have been developed and proved.

A generic methodology that builds upon the concept of combined discrete-event and continuous-time simulation has been developed for performance evaluation of real-time distributed control system networks for autonomous manufacturing, electric power and chemical plant, and aerospace applications. The simulation program is modular in the sense that diverse network protocol models and dynamic models of the plant and controller can be incorporated within the program without any need for modifying its structure.

The impact of delays induced by different network access protocols, such as the SAE linear token-bus and MIL Std.

1553B protocols, on the dynamic performance of a control system has been investigated. The following conclusions regarding DDCCS network design have been derived on the basis of analysis and simulation.

- Although integrated control systems that are interconnected by a data communication network provide for operational flexibility and evolutionary design, network-induced delays degrade the system's dynamic performance and are a source of potential instability.
- The network-induced delays are dependent on both traffic type and the access control protocol. The control system in general should have a better performance if a distributed controlled-access protocol, such as the SAE linear token bus or token ring, is used instead of a centralized controlled-access protocol like the MIL Std. 1553B. However, this conclusion is made only on the basis of data latency in which failure modes and reliability issues have not been addressed.
- The network-induced delays are time-varying and possibly stochastic. Conventional frequency-domain techniques that are routinely applied for linear time-invariant systems may not be valid for DDCCS design. Furthermore the delays from sensor to controller and from controller to actuator in a feedback control loop may not be lumped together in general.
- The network should be designed such that the offered traffic does not exceed the critical value. This implies that any combination of network design parameters (namely, the number of terminals on the network, message arrival rate, message length, medium bandwidth, and sample time) should allow for a safe margin between the offered traffic and its critical value.

#### ACKNOWLEDGMENT

The author acknowledges technical contributions of Dr. Y. Halevi, Mr. S. Lee, Mr. S. H. Hong, and Dr. P. J. Egbelu.

#### REFERENCES

- [1] N. G. Ordrey, "Integrating the automated factory," in *Data Processing Management*. New York: Auerback Publishers, Inc., 1985.
- [2] T. H. Tsai, J. W. Lane, and C. S. Lin, *Modern Control Techniques for the Processing Industries*. New York: Marcel Dekker, 1986.
- [3] A. L. Hopkins *et al.*, "System data communication structures for active-control transport aircraft," vol. II. NASA report CR 165774, June 1981.

- [4] A. Ray, R. Harper, and J. H. Lala, "Damage-tolerant multiplex bus concepts for far-term digital flight control systems," C.S. Draper Laboratory Report CSDA-R-1690, Feb. 1984.
- [5] A. Ray, "Performance evaluation of medium access control protocols for distributed digital avionics," *J. Dynamic Syst. Measurement, Contr.*, pp. 370-377, Dec. 1987.
- [6] A. Ray, "Distributed data communication networks for real-time control," *Chem. Eng. Commun.*, vol. 65, pp. 139-154, 1988.
- [7] W. L. Hoberecht, "A layered network protocol for packet-voice and data integration," *IEEE J. Selected Areas Commun.*, vol. SAC-1, no. 6, Dec. 1983, pp. 1006-1013.
- [8] *IEEE J. Selected Areas Commun., Special Issue on Network Performance and Evaluation*, A. Leon-Garcia, Ed., vol. SAC-4, no. 6, Sept. 1986.
- [9] Y. Halevi and A. Ray, "Integrated communication and control systems: Part I—Analysis," *J. Dynamic Syst. Measurement Contr.*, scheduled for Dec. 1988.
- [10] A. Ray and Y. Halevi, "Integrated communication and control systems: Part II—Design considerations," *J. Dynamic Syst. Measurement Contr.*, scheduled for Dec. 1988.
- [11] W. Stallings, *Data and Computer Communications*. New York: Macmillan 1985.
- [12] *SAE Linear Token Passing Multiplex Data Bus*. AS4074.1, version 3.0, May 1987.
- [13] *SAE High Speed Ring Bus Standard*. Issue 1, Nov. 1985.
- [14] *Token Passing Bus Access Method and Physical Layer Specifications*. IEEE Std 802.4, 1985.
- [15] *Token Ring Access Method and Physical Layer Specifications*. IEEE Std 802.5, 1985.
- [16] *MIL-STD-1553B Protocol Manual*. U.S. Govt. Printing Office, Feb. 1980.
- [17] J. Banks and J. S. Carson, "Process-interaction simulation languages," *Simulation*, vol. 44, no. 5, pp. 225-235, May 1985.
- [18] C. D. Pegden, *Introduction to SIMAN*. State College, PA: Systems Modeling Corporation, 1985.



**Asok Ray** (SM'83) received the Ph.D. degree in mechanical engineering (Northeastern University, 1976) and the master's degrees in electrical engineering, computer science, and mathematics.

He has about ten years of research and management experience at GTE Strategic Systems Division, Charles Stark Draper Laboratory, and the MITRE Corporation. He has also held research and academic positions at Carnegie-Mellon University and the Massachusetts Institute of Technology. He joined the faculty of mechanical engineering at The Pennsylvania State University as an Associate Professor in July 1985. His research experience includes real-time microcomputer-based distributed control and instrumentation, networking and communication protocols, intelligent systems design, and modeling and simulation of dynamical systems as applied to avionics, process control, and autonomous manufacturing. His current research interests include distributed control systems, knowledge-based expert systems, robotics, and communication networks. He has authored or coauthored over 100 research publications.

Dr. Ray is a senior member of the IEEE and a member of the ASME.