

TIMING CHARACTERISTICS OF HIERARCHICAL DISCRETE EVENT SYSTEMS*

Kevin M. Passino
 Department of Electrical Engineering
 The Ohio State University
 2015 Neil Ave.
 Columbus, OH 43210

Panos J. Antsaklis
 Department of Electrical Engineering
 University of Notre Dame
 Notre Dame, IN 46556

ABSTRACT

A discrete event system (DES) is a dynamical system whose evolution in time develops as the result of the occurrence of physical events at possibly irregular time intervals. Although many DES's operation is asynchronous, others have dynamics which depend on a clock or some other complex timing schedule. Here we provide a formal representation of the advancement of time for logical DES via *interpretations of time*. We show that the interpretations of time along with a *timing structure* provide a framework to study principles of the advancement of time for hierarchical DES (HDES). In particular, it is shown that for a wide class of HDES the *event rate* is higher for DES at the lower levels of the hierarchy than at the higher levels of the hierarchy. Relationships between event rate and *event aggregation* are shown. We define a measure for event aggregation and show that there exists an inverse relationship between the amount of event aggregation and the event rate at any two successive levels in a class of HDES. Next, we study how to *design* the timing structure to ensure that there will be a decrease in the event rate (by some constant factor) between any two levels of a wide class of HDES. It is shown that if the *communications* between the various DES in the HDES satisfy a certain admissibility condition then there will be a decrease in the event rate. These results for HDES constitute the main results of this paper since they provide the first mathematical characterization of the relationship between event aggregation and event rates of the HDES and show how to design the interconnections in a HDES to achieve event rate reduction. Examples are provided to illustrate the results.

1.0 INTRODUCTION

In our main results we show that the *interpretations of time* which characterize the advancement of time in DES (introduced in Section 2) along with a *timing structure* provide a framework to study principles of the advancement of time for hierarchical DES (HDES). In Section 3 (Theorem 1) it is shown that for a wide class of HDES the *event rate* is higher for DES at the lower levels of the hierarchy than at the higher levels of the hierarchy. Relationships between event rate and *event aggregation* are shown. We define a measure for event aggregation and show that a high amount of event aggregation will result in a much lower event rate at higher levels in a certain class of HDES while a low amount of event aggregation will result in higher event rates (Theorem 2). Event rate reduction in HDES is often desirable so that the processors implementing the higher level controls are permitted adequate time before they must attend to the lower level systems. Next, we study how to *design* the timing structure to ensure that there will be a decrease in the event rate (by some constant factor) between any two levels of a wide class of HDES. It is shown that if the *communications* between the various DES in the HDES satisfy a certain admissibility condition then there will be a decrease in the event rate (Theorem 3). Hence, we show that one only needs to restrict the *interconnections* in the HDES to achieve event rate reduction. The results are illustrated with a conventional discrete event control system and a manufacturing system. Some of the results in this paper were originally reported in [4,6,8,9] and an expanded version of this paper (which includes the proofs) is given in [18].

We focus on timing characteristics of single DES or HDES which have as components DES that can be accurately modelled with

$$P=(X,U,Y,\delta,\lambda,X_0) \quad (1)$$

where,

- (i) X is the set of plant states x ,
- (ii) U is the set of plant inputs u ,
- (iii) Y is the set of plant outputs y ,
- (iv) $\delta:U \times X \rightarrow \mathcal{P}(X)$ is the plant state transition function ($\mathcal{P}(X)$ denotes the power set of X),
- (v) $\lambda:U \times X \rightarrow Y$ is the plant output function, and
- (vi) $X_0 \subset X$ is the set of possible initial plant states.

The plant state transition function (a partial, point to set function) specifies for each current input u and state x the set of possible next states $x' \in \delta(u,x)$. The output function specifies, for the current input u and state x , the current output symbol $y = \lambda(u,x)$. Formally, P is equivalent to a directed graph with node set X and edges $x \rightarrow x'$ labelled with "u/y" for each triple (u,x,x') such that $x' \in \delta(u,x)$ and $y = \lambda(u,x)$. The model P is similar to a standard automaton but X , U , and Y are not required to be finite. A *run* of P is defined as a sequence of triples $(u_0, x_0, y_0), (u_1, x_1, y_1), (u_2, x_2, y_2), \dots$ such that $x_0 \in X_0$, u_0 is the initial input, $x_{k+1} \in \delta(u_k, x_k)$, and $y_k = \lambda(u_k, x_k)$. Notice that since it is possible that $\delta(u_k, x_k) = \emptyset$ for all $u_k \in U$ at some $x_k \in X$, a run may have a finite length. We note that our results are not restricted solely to the use of the DES model (1). The above model was chosen so that the results here would directly apply to a wide class of systems that can be represented with "logical DES models" (e.g., General and Extended Petri nets [10], finite automata, and other DES models [11,17]). The development of results analogous to ours for hierarchical timed or performance models for DES is an important research direction.

2.0 CHARACTERIZING THE ADVANCEMENT OF TIME IN DES

When a physical plant is modelled via (1), the meaning of the advancement of time must be defined. If Z is an arbitrary set, then Z^* denotes the set of all finite strings of elements from Z . If Z and Z' are arbitrary sets then $Z^{Z'}$ denotes the set of all functions mapping Z' to Z . Let \mathbb{N} denote the set of natural numbers. In order to discuss timing issues for P , an *index set* J and *index sequences*

$$\alpha \in J^* \cup J^{\mathbb{N}} \quad (2)$$

are utilized similar to the approach in [12]. The index set J is thought of as a set of times. Let \mathbb{R}^+ denote the set of strictly positive real numbers and $\mathbb{R}_+ = \mathbb{R}^+ \cup \{0\}$, the set of non-negative reals. Note that \mathbb{N} or \mathbb{R} could be candidates for the set J . For convenience, we assume that $J = \mathbb{R}_+$. The index sequences $\alpha \in J^* \cup J^{\mathbb{N}}$ are sequences of time instants that can be of finite or infinite length. For $\alpha \in J^* \cup J^{\mathbb{N}}$ let $|\alpha|$ denote the cardinality of the size of the string α . Note that either $\alpha: \mathbb{N} \rightarrow J$ or $\alpha: [0, a] \rightarrow J$ where $[0, a] \subset \mathbb{N}$, and $\alpha(k)$ simply denotes an element in J . An index sequence (function) $\alpha \in J^* \cup J^{\mathbb{N}}$ is said to be *admissible* if:

- (i) it is order preserving, i.e.,
 - (a) if $\alpha \in J^{\mathbb{N}}$, then for all $k_1, k_2 \in \mathbb{N}$, $k_1 \leq k_2$ implies that $\alpha(k_1) \leq \alpha(k_2)$,
 - (b) if $\alpha \in J^*$, then for all $k_1, k_2 \in \mathbb{N}$ with $k_1, k_2 \in [0, |\alpha| - 1]$, $k_1 \leq k_2$ implies that $\alpha(k_1) \leq \alpha(k_2)$, and
- (ii) it is injective and if $\alpha \in J^{\mathbb{N}}$ then $\alpha(k) \rightarrow \infty$ as $k \rightarrow \infty$.

Following [12], the state of the plant $x \in X$ is associated with the index $\alpha(k)$

* The authors gratefully acknowledge the partial support of the Jet Propulsion Laboratory. Please address all correspondence to K. Passino (email: passino@eagle.eng.ohio-state.edu).

for some $\alpha \in J^* \cup J^{\mathbb{N}}$ and is denoted with $x(\alpha(k))$, meaning "the state at time $\alpha(k)$ ". Similarly, inputs $u \in U$ and outputs $y \in Y$ are associated with that same index and denoted with $u(\alpha(k))$ and $y(\alpha(k))$ respectively. The transition to a state in the set $\delta(u,x)$ can be thought of as leading to the next state, with "next" quantified with the index sequence α as $\alpha(k+1)$. With this, the transition function is given as $x(\alpha(k+1)) \in \delta(u(\alpha(k)), x(\alpha(k)))$ which is often abbreviated as $x_{k+1} \in \delta(u_k, x_k)$. Similarly, the output is often denoted with $y_k = \lambda(u_k, x_k)$ for $k \in \mathbb{N}$. Each run of $P(u_0, x_0, y_0), (u_1, x_1, y_1), \dots$ has an associated index sequence $\alpha \in J^* \cup J^{\mathbb{N}}$, $\alpha = \alpha(0), \alpha(1), \dots$ specifying the time instants at which the triples are defined. Notice that if for some $x_k \in X$ and all $u_k \in U$, $\delta(u_k, x_k) = \emptyset$, then $\alpha(k+1)$ is undefined and in this case α has finite length (in this situation we say that P is "deadlocked").

A DES often *activates* or *triggers* other DES to act. For instance, in the case where P represents a plant, P may trigger a controller to generate an input to P . In this case, the trigger often represents certain changes that occur in the plant. For instance, *events* can be used as the trigger. Similar to [11], we let $E \subset X \times X$ denote the set of *events* e , where

$$E = \{(x, x') \in X \times X : x' \in \delta(u, x)\} \quad (3)$$

An event $e = (x, x')$ is said to *occur* if the state transition from x to $x' \in \delta(u, x)$ takes place. For convenience we shall assume that the event occurs (is defined) at the time instant $\alpha(k+1)$ where the next state is defined. Due to the injective part of the admissibility requirement for α the variables x , u , and y are defined at time instants which are distinct from one another. By condition (ii) of the admissibility requirement when state transitions occur it is guaranteed that time will advance (although it may be a very small amount) and if an infinite number of events occur this will take an infinite amount of time. The other important implication is that using the definition of events E in (3) it is automatically assumed that events occur at distinct times, i.e., simultaneous events are not allowed because the index sequences are required to be admissible. Suppose for a moment that condition (ii) of the admissibility requirement is omitted so that for $\alpha \in J^* \cup J^{\mathbb{N}}$ it is possible that $\alpha(k+1) = \alpha(k)$ for any $k \in \mathbb{N}$ such that $\alpha(k)$ and $\alpha(k+1)$ are defined. This will allow events to occur simultaneously at a particular time instant. In fact, for $\alpha \in J^{\mathbb{N}}$ it will allow even an infinite number of events to occur at one time instant resulting in the possibility that time will not advance. Normally, to treat simultaneous events, only a finite number of events are allowed to occur at a single time instant; hence, other events representing the case that "several events occur at once" can often be defined. So the problem of dealing with simultaneous events is often transformed to the case where only a single event occurs at each time instant so that time is guaranteed to advance and admissibility can be assumed (e.g., this can be done for Petri nets [10]).

The pair $I = (A, J)$ where J is an index set and $A \subset J^* \cup J^{\mathbb{N}}$ will be referred to as an *interpretation of time* since it specifies the meaning of the advances in time for the occurrence of state transitions, i.e. it specifies the time instants where the variables of the DES P are defined. In general, a system P is said to have a particular interpretation of time $I = (A, J)$ as long as the time instants associated with the elements of the runs of P are elements of J and the index sequences associated with the runs of P are elements of A . The *admissible interpretation of time* will be denoted with $I_{ad} = (A_{ad}, J_{ad})$ where J_{ad} is an index set and

$$A_{ad} = \{\alpha \in J^* \cup J^{\mathbb{N}} : \alpha \text{ is admissible}\}. \quad (4)$$

Most often we can choose $J_{ad} = J = \mathbb{R}_+$ and this is what we will assume here. It is common to discuss the timing characteristics of DES relative to a *clock*. By a

"clock" we mean a device which has a fixed interval $T \in \mathbb{R}^+$ between *ticks* and which does not stop ticking (if there is deadlock, the clock keeps ticking but no events occur). Next we provide definitions for several standard interpretations of time used in DES studies:

Definition 1: The *asynchronous interpretation of time* is $I_a = (A_a, J_a)$ where $J_a = \mathbb{R}_+$ and $A_a = \{\alpha \in A_{ad} : \alpha(0) = 0\}$.

Definition 2: The *partially asynchronous interpretation of time* is $I_{pa} = (A_{\gamma\beta}, J_{pa})$ with $J_{pa} = \mathbb{R}_+$ and $A_{\gamma\beta} = \{\alpha \in A_a : \alpha(k) + \gamma \leq \alpha(k+1) \leq \alpha(k) + \beta\}$ for $\gamma, \beta \in \mathbb{R}^+$ where $\beta \geq \gamma$.

Definition 3: The *general synchronous interpretation of time* is $I_s = (A_T, J_s)$ with $J_s = \mathbb{R}_+$ and $A_T = \{\alpha \in A_a : \alpha(k+1) = \alpha(k) + nT \text{ where } n \in \mathbb{N} - \{0\}\}$ with $T \in \mathbb{R}^+$.

When $n=1$ we shall refer to I_s simply as the *synchronous interpretation of time*.

3.0 TIMING CHARACTERISTICS OF HIERARCHICAL DES

The formation of a control theory for HDES is just beginning [15-17] even though such systems occur quite frequently. Some principles of the evolution of time in hierarchical systems have been postulated but not fully investigated in [1,13,14,5,7,2]. As in [3] what these researchers have recognized is that "systems usually operate at the higher rates at the lower levels in a hierarchical system". We shall verify this intuition for a wide class of HDES here.

3.1 A Hierarchical DES Model

We shall focus on HDES that have as components two types of DES, G_j , $1 \leq j \leq m$, and P_i , $1 \leq i \leq n$, all defined via (1) except with different timing characteristics. We introduce what we call a *timing structure* which will define how the various components of the HDES influence (are influenced by) the timing characteristics of other components of the HDES. The definition of the timing structure is based on the interpretations of time defined in Section 2.0 and what will be called *input* and *output triggers*. Each P_i , $1 \leq i \leq n$, in the HDES has timing characteristics that are simply specified via their own interpretation of time denoted with $I_{pi} = (A_{pi}, J_{pi})$. Roughly speaking, each G_j , $1 \leq j \leq m$, has timing characteristics that depend on other P_i and G_l via the timing structure as we now discuss in more detail.

Let E_{pi} denote the set of events for P_i , and E_{gj} , the set of events for G_j both defined in a similar manner to the events E for P in (3). Let C_{pi} (C_{gj}) denote the set of *communications* that can be transmitted from P_i (G_j) via the timing structure to other G_l . The *output triggers* for the P_i (resp., for G_j) are defined via

$$\phi_i: E_{pi}^* \rightarrow C_{pi} \quad 1 \leq i \leq n \quad (\text{resp., } \psi_j: E_{gj}^* \rightarrow C_{gj} \quad 1 \leq j \leq m) \quad (5)$$

(or restrictions of these maps). If $\phi_i(e) = c$ or $\psi_j(e') = c'$ then c and c' are *communications* that are said to *occur* due to the occurrence of event string e or e' (e triggers communication c). We use the standard notation for *concatenation*, e.g., if $e, e' \in E_{pi}^*$ then ee' denotes the concatenation of e and e' . The time instant at which the communication c (c') occurs is the same time instant that $e \in E_{pi}$ ($e' \in E_{gj}$) occurs where $\phi_i(ee) = c$ ($\psi_j(e'e) = c'$). The *input triggers* for the G_j are defined by the τ_j maps for j , $1 \leq j \leq m$, where

$$\tau_j: C_{p1} \times \dots \times C_{pn} \times C_{g1} \times \dots \times C_{gj-1} \times C_{gj+1} \times \dots \times C_{gm} \rightarrow \{0,1\} \quad (6)$$

and $\tau_j(\cdot) = 1$ ($=0$) indicates that an event $e_{gj}(\alpha(k+1)) \in E_{gj}$ where $e_{gj}(\alpha(k+1)) = (x_{gj}(\alpha(k)), x_{gj}(\alpha(k+1)))$ is forced (not) to occur in G_j . The τ_j indicate which P_i and G_l communicate with G_j via the timing structure. Equation 6 indicates the form for the τ_j maps; the absence of a C_{pi} or C_{gl} from the cross product in the domain of τ_j indicates that P_i or G_l does not communicate with G_j via τ_j . It is assumed that the τ_j maps form a "tree

structured" timing structure as we describe next.

Let each DES component P_i , $1 \leq i \leq n$, or G_j , $1 \leq j \leq m$, of the HDES represent a node (e.g., denoted with boxes as in Figure 3.1) of a directed graph δ and let the τ_j define the arcs (e.g., denoted with shaded arcs in Figure 3.1) that connect the P_i and G_j to G_j in the following manner: Let $i_1, i_2, \dots, i_l, l_1, l_2 \in N$. If there exists i (l) such that $\tau_j: C_{p_1}^{i_1} \times \dots \times C_{p_i}^{i_l} \times \dots \times C_{g_1}^{l_1} \times \dots \times C_{g_l}^{l_2} \rightarrow \{0,1\}$ then there exists an arc pointing from P_i to G_j (G_l to G_j). In this paper we focus on HDES which have a tree structured timing structure, i.e., the case where δ has no closed cycles. In this way we eliminate the possibility that some G_j can directly force its own events to occur via the timing structure. Notice that the P_i , $1 \leq i \leq n$, are the "leaves" of the tree structured timing structure.

Intuitively, the ϕ_i and ψ_j specify what each DES will communicate (C_{p_i} and C_{g_j}) to the other DES in the HDES. The τ_j define communication channels (the arcs and paths in δ) and where the communications are distributed in the HDES. Next, we define the time instants at which events occur when they are forced to do so by other DES components of the HDES via the timing structure.

Whereas the interpretation of time is always specified for the P_i , $1 \leq i \leq n$, the interpretations of time for the G_j are specified in terms of the other G_l and the P_i via the timing structure. Let $\alpha_{c_{p_i}}(k+1)$ and $\alpha_{c_{g_l}}(k+1)$ denote the time instants at which communications $c_{p_i} \in C_{p_i}$ and $c_{g_l} \in C_{g_l}$ occur respectively. Suppose that at some time instant $\alpha_{g_j}(k+1)$, $\tau_j(\cdot)=1$ so that $e_{g_j}(\alpha_{g_j}(k+1)) \in E_{g_j}$ occurs. This time instant at which $e_{g_j}(\alpha_{g_j}(k+1))$ occurs is given by

$$\alpha_{g_j}(k+1) = \max(\alpha_{c_{p_i}}(k+1), \alpha_{c_{g_l}}(k+1)) : \exists \text{ an arc in } \delta \text{ from } P_i \text{ or } G_l \text{ to } G_j \quad (7)$$

and corresponds to the time instant at which the last communication accessible to G_j occurred and caused $\tau_j(\cdot)=1$. Each time a communication occurs which forces $\tau_j(\cdot)=1$, an event occurs in G_j ; hence the "1" represents a pulse sent to G_j via τ_j which forces an event to occur. Hence, if $\tau_j(\cdot)$ is set equal to 1 at some time instant, an event in G_j must occur at that time instant (unless G_j is deadlocked); if every communication in a sequence of communications all cause $\tau_j(\cdot)=1$ then there is one event occurrence in G_j for each communication in the sequence. The interpretation of time for any G_j is found by executing all possible runs (in all possible orders) of the P_i , $1 \leq i \leq n$, and G_l for which there exists a path in δ from P_i or G_l to G_j . Then via (7), the time instants and hence index sequences and interpretations of time for the G_j are specified. We shall study HDES where there is at least one P_i and the interpretations of time for the G_j can be uniquely defined in terms of the P_i .

Note that although we consider only tree structured timing structures we place no restrictions on the manner in which the DES inputs U and outputs Y are connected. This allows our results to apply to a large class of HDES with a wide variety of input/output connecting structures. Tree structured timing structures allow us to study properties of what has been called a "time scale hierarchy" [2]. In this hierarchy a DES component is "higher in the hierarchy" than another DES component if its timing characteristics can be influenced by the other DES (i.e., there exists a path in δ from one to the other).

3.2 Lower Event Rates at Higher Levels in the HDES

To analyze the timing characteristics of HDES we study one fundamental component (shown in Figure 3.1) of the HDES defined above. Even though we consider only P_i at the lower level, it requires only a simple modification to consider a mix of P_i and G_j at the lower level and all of our results below are still valid. Moreover, our results easily generalize to the fully interconnected HDES

by repeated application of the derived relationships which pertain to the two levels in Figure 3.1.

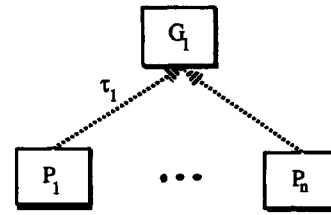


Figure 3.1 Hierarchical DES with Single-Branch

Let the admissible interpretation of time for P_i be $I_{p_i} = (A_{p_i}, J_{p_i})$, $1 \leq i \leq n$, with $J_{p_i} \in \mathbb{R}_+$ and for G_1 be $I_{g_1} = (A_{g_1}, J_{g_1})$.

Definition 4: The event occurrence rate (event rate) in P_i or G_j is the number of events that occur in the time interval $T_u = (r_1, r_2]$ where $(r_1, r_2] \subset \mathbb{R}^+$ and it will be denoted with $\#(P_i, T_u)$ and $\#(G_j, T_u)$ respectively.

Notice that if P_i has a synchronous interpretation of time with $T \in \mathbb{R}^+$ and we choose T_u such that $lr_2 - r_1 = T$ then $\#(P_i, T_u) = 1$, i.e., there is 1 event occurrence in the time interval T_u no matter what the particular values of r_1 and r_2 are. If P_i has an asynchronous interpretation of time then no matter how T_u is chosen it is possible that $\#(P_i, T_u) = 0$, since we cannot guarantee that an event will occur in the given time interval T_u . In fact, we do not know how many events will occur in T_u . It would appear that our definition of event rate is too restrictive. This is, however, not the case since the focus here is on comparing the event rates of different DES components in the HDES and this comparison is made relative to T_u , an interval of the real time line.

Theorem 1: $\sum_{i=1}^n \#(P_i, T_u) \geq \#(G_1, T_u) \geq 0$ for all T_u .

Theorem 1 states the intuitively clear fact that the timing structure can mask events and hence remove the time instants at which events occur in higher levels of the hierarchy. This means that the event rate is lower in DES at the higher levels of the HDES and higher in lower levels of the HDES no matter what the interpretations of time are for the P_i , $1 \leq i \leq n$.

Remark 1: Repeated application of Theorem 1 to the multi-level hierarchy in Figure 3.2 results in $\#(P_1, T_u) \geq \#(G_1, T_u) \geq \dots \geq \#(G_m, T_u) \geq 0$ for all T_u . This result supports the studies in [3] where the author assumes a synchronous interpretation of time and that the event rates can be split into "spectra" according to the level in the hierarchy. It also shows that in the more general case, e.g., for asynchronous P_1 , the event rates in DES at the higher levels are also greater than or equal to the event rates at the lower levels.

Example 1: (Conventional Discrete Event Control System) Consider the controlled DES shown in Figure 3.3. We have $\phi_1: E_{p_1}^* \rightarrow C_{p_1}$ and $\tau_1: C_{p_1} \rightarrow \{0,1\}$ and for the standard control configuration it is most often assumed that for all $e \in E_{p_1}^*$ such that $e = e'e$ ($e \in E_{p_1}$), $\tau_1(\phi_1(e'e)) = 1$ so that each time an event occurs in P_1 , G_1 is forced to act by having an event in G_1 occur (it is normally assumed that one always exists). Clearly, then if $I_{p_1} = (A_{p_1}, J_{p_1})$ is the interpretation of time for P_1 and $I_{g_1} = (A_{g_1}, J_{g_1})$ for G_1 where $J_{g_1} = J_{p_1}$, then $A_{g_1} = A_{p_1}$. The interpretation of time for the plant and controller are the same.

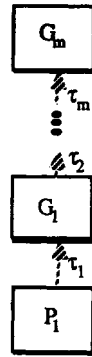


Figure 3.2 Multi-Level Hierarchical DES with m+1 Levels

In this way we think of specifying the interpretation of time for G_1 by I_{p1} and P_1 via τ_1 and ϕ_1 . Via Theorem 1, for general ϕ_1 and τ_1 we see that we can expect fewer events to occur in G_1 than in P_1 since P_1 may not communicate the occurrence of an event or G_1 may not recognize the communication. ■

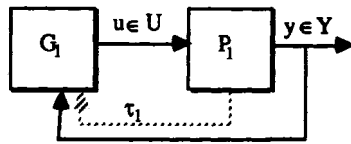


Figure 3.3 Discrete Event Control System

Remark 1 and Example 1 illustrate the generality of Theorem 1; the result applies to hierarchical DES currently being studied (in addition to the work in [3] the result also applies to the work in [15-17]), and standard discrete event control systems. Example 2 (a manufacturing system) in Section 3.4 is used to further illustrate the use of Theorem 1.

3.3 Event Rates and Aggregation in a Class of HDES

Next we study how aggregation affects the event rates in HDES. Again we shall focus on the HDES component shown in Figure 3.1 but note that the result easily generalizes to a fully interconnected HDES. Let $E_{a_i}^* \subset E_{p_i}^*$ and $\phi_i': E_{a_i}^* \rightarrow C_{p_i}$ for all i , $1 \leq i \leq n$, denote restrictions of ϕ_i . We use ϕ_i' maps for aggregation rather than the ϕ_i ; if $e \in E_{p_i}^*$, $e \notin E_{a_i}^*$ then ϕ_i is said to *ignore* e (rather than *mask* e). Let $B_j \subset N - \{0\}$ for j , $1 \leq j \leq n$.

Definition 5: $\{P_j = (X_j^I, U_j^I, Y_j^I, \delta_j^I, \lambda_j^I, X_{0j}^I), \phi_j'; 1 \leq j \leq n\}$ satisfies the $(\pi^1, \pi^2, \dots, \pi^n)$ -event aggregation property if for each j , $1 \leq j \leq n$,

- (i) There exists a family of sets $X_{ij} \subset X_j^I$, $i \in B_j$ such that
 - (a) $X_{ij} \cap X_{kj} = \emptyset$ for all $i \neq k$, and $X_{0j} \cap X_{ij} = \emptyset$ for $i \in B_j$,
 - (b) If P_j first enters a state $x \in X_{ij}$ for some $i \in B_j$, it will take (for all possible runs) at least $\pi_j > 0$ state transitions before the state of P_j , say x' , is such that $x' \in X_{ij}$.
- (ii) $\phi_j': E_{a_j}^* \rightarrow C_{p_j}$ where $E_{a_j}^* = \{e \in E_{p_j}^*; e = e', \text{ where } e = (x, x'), \text{ and for some } i \in B_j, x \in X_{ij}, x' \in X_{ij}\}$.

Theorem 2: If $\{P_j = (X_j^I, U_j^I, Y_j^I, \delta_j^I, \lambda_j^I, X_{0j}^I), \phi_j'; 1 \leq j \leq n\}$ satisfies the $(\pi^1, \pi^2, \dots, \pi^n)$ -event aggregation property and $T_n = (r_1, r_2)$ and $|r_2 - r_1|$ is sufficiently large, then

$$\sum_{j=1}^n \left\{ \frac{\#(P_j, T_n)}{\pi_j} + 1 \right\} \geq \#(G_1, T_n). \quad (8)$$

Remark 2: For the multi-level HDES in Figure 3.2, if the (π^j) -event aggregation property holds for each successive level and $T_n = (r_1, r_2)$ with $r_1 = 0$ then for all $r_2 > 0$, and for j , $1 < j \leq m$,

$$\frac{\#(G_j, T_n)}{\pi_j} \geq \#(G_{j+1}, T_n).$$

The ϕ_j , ψ_j , and τ_j can be viewed as maps that cause event aggregation; consequently, Theorem 2 and Remark 2 provide a relationship between event aggregation and event rates for one class of HDES. If there is a high measure of aggregation at level j (large π^j) then there will be a lot fewer events occurring at level $j+1$. This illustrates that there is an inverse relationship between event aggregation and event rate between two levels of a HDES. In general, hierarchical systems researchers have observed a similar inverse relationship between "time scale density" ("time granularity") and "model abstractness" [2,13]. The above results provide the first mathematical validation of these researcher's intuition about relationships between event aggregation and event rates for a class of HDES. Example 2 in Section 3.4 is used to illustrate the use of Theorem 2.

3.4 HDES Timing Structure Design for Event Rate Reduction

Theorem 2 provides a characterization of how event rates are affected by aggregation for one class of HDES. In this section we study the problem of how to *design* the timing structure to ensure that there is a decrease (by a some constant factor) in the event rate between any two levels of a *wide* class of HDES. A reduction in event rate is often desirable so that the processors implementing the higher level controls are permitted adequate time before they must attend to the lower level systems (e.g., take actions based on the occurrence of an event string).

Let

$$S_{p_i} \subset E_{p_i}^* \quad (S_{g_j} \subset E_{g_j}^*) \quad (9)$$

and the communications be defined by

$$C_{p_i} = \mathbb{P}(S_{p_i}) \quad (C_{g_j} = \mathbb{P}(S_{g_j})). \quad (10)$$

The notation $e \in e'$ will be used to denote the fact that e is a substring of e' . Consider the case where the output trigger is defined so that $s \in \phi_i(e)$ if $s \in e$ and $s \in S_{p_i}$ ($s' \in \phi_i(e)$ if $s' \in e$ and $s' \in S_{g_j}$). This output trigger initiates a communication the first time an event string occurs and $\# \phi_i(e) \geq \# \phi_i(e')$ if $|e| \geq |e'|$. Hence, if the same event string occurs twice (or more) in some run this fact cannot be reported by this output trigger. Similar problems can exist if we define the output trigger so that $s \in \phi_i(e)$ if $e = e_1 e_2$, $p \geq l_2 > 0$, and $s \in e_2$. This output trigger does, however, have the interesting property that it will "forget" about event strings in the past (depending on the choice for p). Here, we shall define the output trigger so that

$$s \in \phi_i(e) \text{ if } e = e's \text{ and } s \in S_{p_i} \quad (11)$$

($s' \in \psi_j(e)$ if $e = e's'$ and $s' \in S_{g_j}$). By definition, if $\phi_i(e) = \emptyset$ ($\psi_j(e) = \emptyset$), a "null communication" occurs which cannot directly cause an event occurrence in any other DES in the HDES. These assumptions about C_{p_i} (C_{g_j}) and ϕ_i (ψ_j) in (9-11) are only mildly restrictive since it is possible that there can be a different communication representing each possible set of finite event strings that have just occurred. Moreover, there will be no particular assumptions about the τ_j maps and the definition for the output triggers via (11) and communications via (9-10) is quite practical since each component DES is allowed to communicate the fact that sequences of events have just occurred; other DES in the HDES can then act based on such behavior.

The design of the timing structure will entail choosing the proper S_{pi} and hence the communications C_{pi} that can occur between the various DES in the HDES. It is shown that by restricting the choice of what communications are allowed one can achieve a decrease in the event rate at the higher levels of the HDES. In this way we achieve event rate reduction by restricting the manner in which the DES communicate and not by making particular assumptions about the dynamics of each component DES (as was done for Theorem 2). As in Sections 3.2 and 3.3 we shall focus only on the HDES of Figure 3.1 and the results easily generalize to fully interconnected tree-structured HDES. First, we introduce a fundamental property of communications within the HDES:

Definition 5: The set S_{pi} is said to be γ_i -admissible if for all $s, s' \in S_{pi}$ such that $s=ab, s'=cd$, and $b=c$ with $|b|=|c| \geq 0$ it is the case that $|d| \geq \gamma_i > 0$.

A similar definition can be given for the S_{gi} . Clearly, there may not exist $S_{pi} \subset E_{pi}^*$ such that S_{pi} is γ_i -admissible for some given γ_i ; but there always exists some $\gamma_i > 0$ such that S_{pi} is γ_i -admissible. Hence for some DES one may be able to achieve more event rate reduction than for others and γ_i -admissibility characterizes this property. Intuitively, if the behavior of some DES P_i is such that it generates event strings which do not frequently cause communications to other DES then γ_i is large. Next, we provide several examples of $S_{pi} \subset E_{pi}^*$ that are γ_i -admissible:

- (1) Assume that for all $s \in S_{pi}$, $|s| \geq \gamma_i$. If for all $s \in S_{pi}$ and all $e \in s$ where $e \in E_{pi}$, there does not exist $s' \in S_{pi}$, $s' \neq s$, such that $e \in s'$ then S_{pi} is γ_i -admissible.
- (2) If for all $s \in S_{pi}$ there exists $e \in E_{pi}$, and s_2 such that $s=es_2$ and $|s_2| \geq \gamma_i - 1$, and there does not exist $s' \in S_{pi}$, $s' \neq s$, such that $e \in s'$ then S_{pi} is γ_i -admissible (similarly for $s=s_2e$), and more generally:
- (3) If for all $s \in S_{pi}$ there exists s_1, s_2 such that $s=s_1s_2$, $|s_1| \geq 1$, and $|s_2| \geq \gamma_i - 1$, and there does not exist $s' \in S_{pi}$, $s' \neq s$, such that $s_1 \in s'$ then S_{pi} is γ_i -admissible.

It is important to note that for a given S_{pi} that might be chosen in the design of a timing structure it is not difficult to test whether or not S_{pi} is in fact γ_i -admissible for some given γ_i (of course this may be computationally intensive).

Theorem 3: If S_{pi} is γ_i -admissible for all $i, 1 \leq i \leq n$, then

$$\sum_{i=1}^n \left\{ \frac{\#(P_i, T_U)}{\gamma_i} \right\} \geq \#(G_1, T_U) \text{ for all } T_U. \quad (12)$$

Theorem 3 shows that if each S_{pi} , for $i, 1 \leq i \leq n$, is γ_i -admissible then there results a special type of aggregation between two levels of a HDES so that event rate reduction can be obtained. It is important that this aggregation is achieved via conditions on the communications and not on the structure of the $P_i, 1 \leq i \leq n$. Of course, for a given set of lower level P_i one may be able to achieve lower event rates than for another set of P_i ; Theorem 3 shows how to design the communications in the timing structure to achieve event rate reduction for a given set of P_i .

Example 2: (Manufacturing System) A simple manufacturing system will be used to illustrate the results from Sections 3.2, 3.3, and 3.4. We consider a manufacturing system which consists of a machine that can process parts of two types, one at a time. The machine outputs each type part into a particular output bin and the machine can be idle. Let $X = \{MI, M_1, M_2, OUT_1, OUT_2\}$ be the set

of states where MI means "machine idle", M_i means the machine is busy processing part type i , and OUT_i means that the machine outputs part type i . Let $U = \{u_1, u_2\}$ where u_i means input part type i into the machine. Let $Y = \{y_b, y_{1d}, y_{2d}\}$ where y_b indicates that the machine is busy processing a part of either type, and y_{id} indicates that the machine is done processing a part of type i . The transition function δ and the output function λ for the manufacturing system are specified via the bottom of Figure 3.4. We let $X_0 = \{MI\}$ and consider $P_1 = (X, U, Y, \delta, \lambda, X_0)$ to be our plant.

There is a higher level mechanism which forces the alternate processing of one part of type 1 and then two parts of type 2. This device is pictured in the top of Figure 3.4 and will be referred to as G_1 . We have $G_1 = (X_g, U_g, Y_g, \delta_g, \lambda_g, X_{0g})$ and $X_g = \{x_1, x_2, x_3\}$, $U_g = Y - \{y_b\}$, $Y_g = U$, and $X_{0g} = \{x_1\}$. Also, initially the input to the plant is u_1 . Notice that G_1 completely ignores output y_b as it is unimportant in coordinating the alternation of processing. (Hence, the inputs and outputs are not connected in a conventional manner where in G_1 , $\delta_g(u, x)$ must be defined for all u .)

Suppose that we let $\tau_1: C_{p1} \rightarrow \{0,1\}$, $C_{p1} = E_{p1}$, and $\phi_1(e) = e$ for all $e \in E_{p1}$ such that $e = (OUT_i, MI)$ for $i=1,2$ (otherwise, $\phi_1(e) = \emptyset$ so that no event occurs in G_1). If the manufacturing system operates asynchronously (synchronously) then the coordination mechanism will operate asynchronously (with a general synchronous interpretation of time). Via Theorem 1, $\#(P_1, T_U) \geq \#(G_1, T_U)$ for all T_U . In particular we see that since event strings ending with (MI, M_i) and (M_i, OUT_i) for $i=1,2$ are masked, a greater number of events will occur in the plant P_1 (lower level system) than in the controller G_1 (higher level system). Hence, for this simple manufacturing system the event rate at the higher level is lower no matter what the interpretation of time in P_1 is. Also, notice that if we choose $B_1 = \{1,2\}$, and $X_{i1} = \{M_i, OUT_i\}$ for $i \in B$, then if we use the input trigger $\phi_1^1: E_{a1}^* \rightarrow \{0,1\}$, a restriction of ϕ_1 , and initial state as defined above, the conditions of Theorem 2 are satisfied with $\pi^1 = 2$ so $\#(P_1, T_U)/2 + 1 \geq \#(G_1, T_U)$ for the proper T_U and all τ_1 . Hence, via Theorem 2 we find an inverse relationship between event aggregation and event rate for this simple manufacturing system.

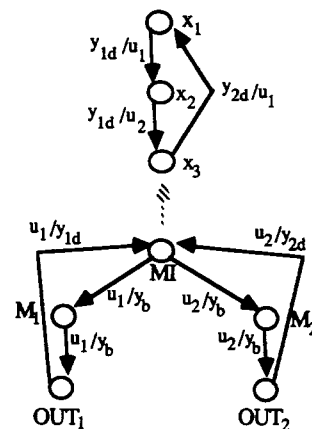


Figure 3.4 Model of a Manufacturing System

Next, we show how Theorem 3 applies to this manufacturing system. Suppose that the same models for the manufacturing system and its controller are used but with a different interconnecting timing structure. In particular, using the approach of this Section to specify the timing structure, we let $S_{p1} = ((MI, M_1)(M_1, OUT_1)(OUT_1, MI), (MI, M_2)(M_2, OUT_2)(OUT_2, MI))$,

$C_{p_i} = \mathbb{P}(S_{p_i})$, $\tau_1: C_{p_1} \rightarrow \{0,1\}$, and ϕ_1 be defined as above in (11). Notice that S_{p_1} is 3-admissible (case (1) above) so that $\#(P_1, T_u)/3 \geq \#(G_1, T_u)$ for all T_u . We see that Theorem 3 can be used to produce a tighter bound on the number of events that occur in G_1 ; hence, the design of the timing structure via the Theorem 3 results in the guarantee of an even lower event rate in G_1 . ■

4.0 CONCLUSIONS

We have provided a mathematical representation of the advancement of time in DES via index sets, index sequences, and interpretations of time. It was shown that for a wide class of HDES the event rate is higher for DES at the lower levels of the hierarchy than at the higher levels of the hierarchy. Relationships between event rate and event aggregation were shown. In order to study how aggregation effects event rates in more general HDES we studied how to design the timing structure to achieve event rate reduction. It was shown that if the communications between the various DES in the HDES satisfy a certain admissibility condition then there will be a decrease in the event rate. Hence, we showed that one only needs to restrict the interconnections in the HDES to achieve event rate reduction.

In a "time scale hierarchy" - what we have been using here - the intuition that event rates are higher for lower levels in the hierarchy has been verified here for a class of HDES, but the results here are relative to this particular type of hierarchy. If one defines a hierarchy relative to, for instance, the functional architecture of a system [2] then clearly at the higher levels of the functional architecture there may be systems that are operating at higher rates than at the lower levels of the functional architecture. It may take a re-arrangement of the system components to place the system in a time scale hierarchy so that our results apply.

5.0 REFERENCES

- [1] Albus J.S., Barbera A.J., Nagel R.N., "Theory and Practice of Hierarchical Control", Proc. of the 23rd IEEE COMPCON, pp. 19-39, 1981.
- [2] Antsaklis P.J., Passino K.M., Wang S.J., "Towards Intelligent Autonomous Control Systems: Architecture and Fundamental Issues", *J. of Intelligent and Robotic Systems*, Vol. 1, pp. 315-342, 1989.
- [3] Gershwin S.B., "Hierarchical Flow Control: A Framework for Scheduling and Planning Discrete Events in Manufacturing Systems", *Proc. of the IEEE*, Vol. 77, No. 1, pp. 195-209, 1989.
- [4] Knight J.F., Passino K.M., "Decidability for a Temporal Logic Used in Discrete Event System Analysis", *Int. Journal of Control*, 1990.
- [5] Mesarovic M., Macko D., Takahara Y., *Theory of Hierarchical Multilevel Systems*, Academic Press, NY, 1970.
- [6] Passino K.M., *Analysis and Synthesis of Discrete Event Regulator Systems*, Ph.D. Dissertation, Dept. of Electrical Eng., University of Notre Dame, Notre Dame, IN, April 1989.
- [7] Passino K.M., Antsaklis P.J., "Fault Detection and Identification in an Intelligent Restructurable Controller", *J. of Intelligent and Robotic Systems*, Vol. 1, pp. 145-161, 1988.
- [8] Passino K.M., Antsaklis P.J., "Relationships Between Event Rates and Aggregation in Hierarchical Discrete Event Systems", Proc. of the Allerton Conf. on Communication, Control, and Computing, Univ. of Illinois, Champaign-Urbana, Oct. 1990.
- [9] Passino K.M., Antsaklis P.J., "Timing Characteristics of Discrete Event Systems", Control Systems Technical Note #68, Univ. of Notre Dame, Dept. of Electrical Engineering, June 1989.
- [10] Peterson J.L., *Petri Net Theory and the Modeling of Systems*, Prentice Hall, NJ, 1981.
- [11] Ramadge P.J., Wonham W.M., "Supervisory Control of a Class of Discrete Event Processes", *SIAM J. Control and Optimization*, Vol. 25, No. 1, Jan. 1987.
- [12] Sain M.K., *Introduction to Algebraic System Theory*, Academic Press, NY, 1981.
- [13] Saridis G.N., "Intelligent Robot Control", *IEEE Trans. on Automatic Control*, Vol. AC-28, pp. 547-556, 1983.
- [14] Valavanis K.P., *A Mathematical Formulation for the Analytical Design of Intelligent Machines*, Ph.D. Dissertation, Dept. of Electrical and Computer Eng., Rensselaer Polytechnic Inst., Troy, NY, Nov. 1986.
- [15] Zhong H., Wonham W.M., "On the Hierarchical Control of Discrete-Event Systems", Proc. of the 1988 Conf. on Inf. Sciences and Systems, Princeton, NJ, March 1988.
- [16] Zhong H., Wonham W.M., "Hierarchical Control of Discrete Event Systems: Computation and Examples", Proc. of the Allerton Conf. on Communication, Control, and Computing, Univ. of Illinois, Urbana, Sept. 1989.
- [17] Zhong H., Wonham W.M., "On the Consistency of Hierarchical Supervision in Discrete-Event Systems", *IEEE Trans. on Automatic Control*, Vol. 35, No. 10, pp. 1125-1134, 1990.
- [18] Passino K.M., Antsaklis P.J., "Event Rates and Aggregation in Hierarchical Discrete Event Systems", Submitted for journal publication, 1990.