Proceedings of the 32nd Conference
on Decision and Control
San Antonio, Texas • December 1993

Subadditivity and Stability of a Class of Discrete-Event Systems

Paul Glasserman*
Graduate School of Business
Columbia University
New York, NY 10027

David D. Yao* †
IEOR Department
Columbia University
New York, NY 10027

Abstract

We investigate the stability of discrete-event systems modeled as generalized semi-Markov processes with event times that satisfy $(\max, +)$ recursions. We show that there exists for each event a *cycle time,* which is the long-run average time between event occurrences. We characterize the rate of convergence to this limit, bounding the error for finite horizons. The main tools we use are $(\max, +)$ matrix products, the subadditive ergodic theorem, and martingale inequalities. We discuss connections with these different fields, with the general theory of random matrix products, and with recent results for discrete-event systems modeled as Petri nets.

1 Introduction

Two seemingly unrelated areas of research in discrete-event systems have expanded to the point where they share some interesting overlap. One avenue of work originates in the sample-path analysis of stochastic systems, especially through perturbation analysis and stochastic monotonicity results; the other originates in the subject of deterministic $(\max, +)$-linear systems. In one direction, the type of structure used for perturbation analysis in Glasserman [6] and for stochastic comparisons in, e.g., Shanthikumar and Yao [16] has been further developed in Glasserman and Yao [7, 8, 9], and shown in [9] to imply $(\min, \max, +)$-recursions for stochastic event times. In the other direction, randomness has been introduced to the deterministic $(\max, +)$-linear systems of Cohen et al. [3], for example in Baccelli [1] and Olsder et al. [12]. What emerges from this intersection of techniques is a class of discrete-event systems covering many examples and possessing many interesting properties.

In this paper, we build on the $(\max, +)$ structure developed for *generalized semi-Markov processes* (GSMP) in Glasserman and Yao [9] and use it to establish stability results. Letting α denote a type of event and letting $T_\alpha(n)$ denote the time of the n-th occurrence of that event, we give conditions under which the limit of $n^{-1}T_\alpha(n)$ as $n \to \infty$ exists and is a finite constant independent of initial conditions. This limit is the long-run average time between occurrences of α, its *cycle time,* and is thus a key measure of performance, reflecting a type of growth rate. We next examine the rate of convergence of $n^{-1}T_\alpha(n)$ to its limit, bounding the error for finite n.

The $(\max, +)$ structure contributes to these results through the *subadditivity* of max, making it possible to invoke Kingman's [10] subadditive ergodic theory. We use this theory together with results from the field of random matrix products. These matrix products arise from vectorization of $(\max, +)$ recursions.

That connections exist among growth rates, subadditivity, and random matrix products is not new. Kingman [10] includes products of random matrices as a key application of his subadditive ergodic theorem. Cohen [4] links various generalizations of matrix multiplication (namely, $(\max, +)$, $(\min, +)$, $(\max, \times)$, and $(\min, \times)$) with examples from operations research, building in part on deterministic models of Cunninghame-Green [5]. A specific $(\max, +)$ application in [4] is the existence of a cycle time for interconnected machines. In a deep further development of these techniques, Baccelli [1] uses subadditivity to prove the existence of cycle times in a class of stochastic Petri nets.

Our work differs from earlier results in several important respects. Not least of these is the GSMP setting. GSMPs have emerged as an important class of models for discrete-event systems, and identifying the right structure and conditions to apply subadditivity is not altogether straightforward. In analyzing cycle times, we derive some extensions of Cohen's [4] results for random matrix products to suit the more general class of matrices we encounter; these are of independent interest. Also, we give a complete characterization of the limiting matrix for our setting; related but different limits are analyzed in Baccelli [1]. Our analysis of convergence rate uses a martinagle inequality in a method developed for probabilistic analysis of combinatorial problems in Rhee and Talagrand [15]; this method has not previously been used with discrete-event systems or random matrices.

The rest of this paper is organized as follows. Section 2 introduces the GSMP framework, then provides the necessary set-up for $(\max, +)$ recursions. Section 3 includes a brief review of subadditive ergodic theory, then establishes necessary results on the $(\max, +)$-products of random matrices. These results are used to establish the existence of cycle times. Section 4 examines the rate of convergence, giving error bounds for finite horizons.

This paper is an overview of a full paper, which contains all proofs (omitted here). There we also study the stability of *delays,* i.e., differences between event times, and develop the connection with stochastic difference equations.

2 Event Time Recursions

We use the framework of Glasserman and Yao [7, 8, 9]; further references to the GSMP literature are given there.

A GSMP is based on a generalized semi-Markov *scheme* (GSMS), denoted by $\mathcal{G} = (\mathbf{S}, \mathbf{A}, \mathcal{E}, p)$ and consisting of the following ingredients: a countable state space $\mathbf{S}$; a finite set $\mathbf{A}$ of events; a mapping $\mathcal{E}$ from elements of $\mathbf{S}$ to subsets of $\mathbf{A}$; and a collection $p = \{p(\cdot; s, \alpha), s \in \mathbf{S}, \alpha \in \mathcal{E}(s)\}$ of probability mass functions on $\mathbf{S}$. The elements of $\mathbf{S}$ represent physical states of a system and the elements of $\mathbf{A}$ are events that potentially change the state. The active events in state $s \in \mathbf{S}$ are the elements of the *event list* $\mathcal{E}(s) \subseteq \mathbf{A}$. We always assume that

$$\mathbf{A} \subseteq \bigcup_{s \in \mathbf{S}} \mathcal{E}(s), \tag{1}$$

so $\mathbf{A}$ contains no extraneous elements. Upon the occurrence of event α in state s, $\alpha \in \mathcal{E}(s)$, the system moves to state s' with probability $p(s'; s, \alpha)$. In this paper, we treat only schemes with

*Supported in part by NSF Grant MSS-92-16490.
†Supported in part by NSF Grant ECS-89-96230.

deterministic routing; these are schemes in which each mass function $p(\cdot;s,\alpha)$ is concentrated on a single point. In this case, we replace p with a function ϕ, where $\phi(s,\alpha)=s'$ if $p(s';s,\alpha)=1$.

A GSMS is made dynamic through the introduction of *clock times*. This is a sequence $\omega=\{\omega_\alpha(n),\alpha\in\mathbf{A},n=1,2,\ldots\}$, where $\omega_\alpha(n)\geq 0$ represents the n-th lifetime associated with event α. If the system starts in state s^0, then at time zero clocks are set for the first occurrence of the events in $\mathcal{E}(s^0)$, the initially active events. If $\alpha\in\mathcal{E}(s^0)$ then the clock for α is set to $\omega_\alpha(1)$. Clocks run at unit rate; when a clock runs out, the corresponding event occurs triggering a state transition. Just after each state transition, new clocks are set for any newly active events: at the n-th activation of event β, its clock is initialized to $\omega_\beta(n)$. Clocks for events active in the previous state that are also in the event list of the new state continue to run. A previously active event that is no longer active (other than the event that triggered the transition) is said to be interrupted.

Through this mechanism, we obtain from ω the sequence $T=\{T_\alpha(n),\alpha\in\mathbf{A},n=1,2,\ldots\}$ where $T_\alpha(n)$ is the time of the n-th occurrence of event α. If α fails to occur n times, $T_\alpha(n)=\infty$. We also set $T_\alpha(n)=0$ for $n\leq 0$ and $T_\alpha(\infty)=\infty$. If we put a probability measure on the set of ω's, then T becomes stochastic. For emphasis, to indicate random clock times we use $\xi=\{\xi_\alpha(n),\alpha\in\mathbf{A},n=1,2,\ldots\}$ and to indicate a particular realization we use ω.

2.1 Structural Conditions

A *string* of events is a finite sequence of elements of $\mathbf{A}$. A string $\sigma=\alpha^1\cdots\alpha^n$ is called *feasible* in state s^0 if there are states $s^1,\ldots,s^{n-1}$ with $\alpha^i\in\mathcal{E}(s^{i-1})$, $i=1,\ldots,n$, and

$$p(s^1;s^0,\alpha^1)p(s^2;s^1,\alpha^2)\cdots p(s^{n-1};s^{n-2},\alpha^{n-1})>0.$$

Thus, a feasible string is just a possible sequence of events. With the initial state s^0 fixed, we denote the collection of all feasible strings by $\mathcal{L}$ and call it the *language* generated by the GSMS. By convention, $\mathcal{L}$ contains the *empty string*. In a scheme with deterministic transitions, we extend $\phi(s^0,\cdot)$ from $\mathcal{E}(s^0)$ to all of $\mathcal{L}$ in the obvious way: if $\sigma\in\mathcal{L}$ is non-empty, then $\phi(s^0,\sigma)$ is the state reached from s^0 through the sequence of events σ, and $\phi(s^0,\cdot)$ applied to the empty string returns s^0. For any string σ (feasible or not), we denote by $[\sigma]$ the vector $([\sigma]_\alpha,\alpha\in\mathbf{A})$ where $[\sigma]_\alpha$ records the number of occurrences of α in σ. The vector $[\sigma]$ is called the *score* of σ.

With this notation, we give the following structural conditions:

Definition 2.1 A scheme with deterministic transitions is

(i) *non-interruptive,* if $\{\alpha,\beta\}\subseteq\mathcal{E}(s)$ implies $\beta\in\mathcal{E}(\phi(s,\alpha))$, for all $s\in\mathbf{S}$ and all $\alpha,\beta\in\mathbf{A}$;

(ii) *permutable,* if $[\sigma^1]=[\sigma^2]$ implies $\mathcal{E}(\phi(s^0,\sigma^1))=\mathcal{E}(\phi(s^0,\sigma^2))$, for all $\sigma^1,\sigma^2\in\mathcal{L}$;

(iii) *strongly permutable,* if $[\sigma^1]=[\sigma^2]$ implies $\phi(s^0,\sigma^1)=\phi(s^0,\sigma^2)$, for all $\sigma^1,\sigma^2\in\mathcal{L}$.

In a non-interruptive scheme the occurrence of one event cannot deactivate other events; a clock, once set, always runs out at its scheduled time. Permutability requires that changing the order of events (while preserving feasibility) not change the event list reached. Strong permutability is indeed stronger, requiring that a change in the order of events not change the *state* reached. Taken together, properties (i) and (iii) are equivalent to the commuting condition of Glasserman [6], and properties (i) and (ii) are equivalent to condition (M) of Glasserman and Yao [8]. Various alternative formulations of these conditions (and their consequences) are discussed in [7].

For the present analysis, the most important consequence of these structural conditions is a set of recursions for the event times. To develop these, we need to introduce the *score space* associated with a GSMS, which is just the set of scores of feasible strings. Let $m=|\mathbf{A}|$. By ordering the elements of $\mathbf{A}$, we can identify any score $[\sigma]$ with an element of $\mathbf{Z}_+^m$. The score space is then

$$\mathcal{N}=\{\mathbf{x}\in\mathbf{Z}_+^m:\exists\sigma\in\mathcal{L},[\sigma]=\mathbf{x}\}.$$

The elements of $\mathcal{N}$ are called feasible scores.

For any $\alpha\in\mathbf{A}$ and any $n=1,2,\ldots$, consider the set of strings leading to the n-th occurrence of α; these are the strings σ for which $[\sigma]_\alpha=n-1$ and $\alpha\in\mathcal{E}(\phi(s^0,\sigma))$. This set may be empty (if α cannot occur n times) and may consist only of the empty string (if $n=1$ and $\alpha\in\mathcal{E}(s^0)$). Denote by $\mathcal{N}_{\alpha,n}$ the set of scores of these strings. We say that an element $\mathbf{x}$ of $\mathcal{N}_{\alpha,n}$ is *minimal* if there does not exist $\mathbf{y}\in\mathcal{N}_{\alpha,n}$ with $\mathbf{y}\neq\mathbf{x}$ and $\mathbf{y}\leq\mathbf{x}$, where $\leq$ denotes the usual componentwise partial order. Each non-empty $\mathcal{N}_{\alpha,n}$ contains a finite number of minimial elements $\mathbf{x}^i(\alpha,n)$, $i=1,\ldots,J_{\alpha,n}$. If $\mathcal{N}_{\alpha,n}$ is empty, set $J_{\alpha,n}=1$ and $\mathbf{x}^1_\beta(\alpha,n)=\infty$ for all $\beta\in\mathbf{A}$. From [9], we have

Theorem 2.2 In a non-interruptive, permutable scheme, for all ω the event times satisfy

$$T_\alpha(n)=\omega_\alpha(n)+\min_{1\leq i\leq J_{\alpha,n}}\max_{\beta\in\mathbf{A}}\{T_\beta(\mathbf{x}^i_\beta(\alpha,n))\}. \tag{2}$$

The indices on the right do not depend on ω.

This seemingly complicated expression has a simple interpretation. Each $\mathbf{x}^i(\alpha,n)$ represents an alternative set of precedents for the n-th occurrence of α, in the sense that once every event β has occurred $\mathbf{x}^i_\beta(\alpha,n)$ times, α is activated for the n-th time. This happens at the maximum of $T_\beta(\mathbf{x}^i_\beta(\alpha,n))$, $\beta\in\mathbf{A}$. However, α is activated for the n-th time as soon as the first complete set of precedents is met; hence, we take the minimum over the alternative sets indexed by $i=1,\ldots,J_{\alpha,n}$. Finally, in a non-interruptive scheme, event α occurs $\omega_\alpha(n)$ time units after the n-th activation of α.

We obtain some simplification through the following:

Definition 2.3 A deterministic scheme is *irreducible* if for every pair of states s and s' there exists a string σ with $\phi(s,\sigma)=s'$. The scheme is *event-irreducible* if for every state s and event α there exists a string σ with $\alpha\in\mathcal{E}(\phi(s,\sigma))$.

It is easy to see that under our standing assumption (1), irreducibility implies event-irreducibility. Moreover, under event-irreducibility, every $\mathcal{N}_{\alpha,n}$ is non-empty. Therefore, In an event-irreducible scheme, all indices in (2) are finite, for all α and n.

We turn, now, to a class of systems for which the (min,max,+) recursions in (2) simplify to (max,+) recursions. Clearly, this simplification occurs when $\mathcal{N}_{\alpha,n}$ contains just one minimal element, in which case the min becomes superfluous. As shown in [9], the score space $\mathcal{N}$ and its subsets $\mathcal{N}_{\alpha,n}$ are automatically closed under componentwise maximum (denoted by $\vee$) if the scheme is non-interruptive and permutable; this is but one manifestation of the *antimatroid* property that comes from these structural conditions. If each $\mathcal{N}_{\alpha,n}$ were closed under componentwise minimum (denoted by $\wedge$), it would have a unique minimal element. So, the key additional property we need is closure under $\wedge$. For that we have the following condition from [9].

(CX) If $\sigma^1,\sigma^2,\sigma^3\in\mathcal{L}$, then $[\sigma^3]\geq[\sigma^1]\wedge[\sigma^2]$ implies

$$[\mathcal{E}(\phi(s^0,\sigma^1))\cap\mathcal{E}(\phi(s^0,\sigma^2))]\backslash A\subseteq\mathcal{E}(\phi(s^0,\sigma^3)),$$

where $A=\{\alpha:[\sigma^3]_\alpha>[\sigma^1]_\alpha\wedge[\sigma^2]_\alpha\}$.

From [9] we also have

Theorem 2.4 In a scheme satisfying (CX), $\mathcal{N}$ and every $\mathcal{N}_{\alpha,n}$ are closed under $\wedge$. Consequently, the event times satisfy

$$T_\alpha(n)=\omega_\alpha(n)+\max_{\beta\in\mathbf{A}}\{T_\beta(\mathbf{x}_\beta(\alpha,n))\}, \tag{3}$$

where $\mathbf{x}(\alpha,n)$ is the (unique) minimal element of $\mathcal{N}_{\alpha,n}$.

Later, we will need what might be considered an explicit solution to the recursion (3). This alternative representation of the the event times depends on a notion of *longest path* to a pair (α, n), $\alpha \in \mathbf{A}$, $n = 1, 2, \ldots$. A path to (α, n) is a sequence $\{(\beta_{i_1}, k_{i_1}), \ldots, (\beta_{i_r}, k_{i_r})\}$ with $k_{i_1} = 1$, $\beta_{i_1} \in \mathcal{E}(s^0)$, $\beta_{i_r} = \alpha$ and $k_{i_r} = n$, in which the k_{i_j}-th occurrence of β_{i_j} activates event $\beta_{i_{j+1}}$ for the $k_{i_{j+1}}$-th time, $j = 1, \ldots, r-1$. More formally, the activation condition is

$$\begin{aligned} [\beta_{i_1} \cdots \beta_{i_j}] &\geq \mathbf{x}(\beta_{i_{j+1}}, k_{i_{j+1}}) \\ [\beta_{i_1} \cdots \beta_{i_{j-1}}] &\not\geq \mathbf{x}(\beta_{i_{j+1}}, k_{i_{j+1}}). \end{aligned}$$

A path corresponds to a *triggering sequence* in [6].

Let $\Pi(\alpha, n)$ denote the collection of all paths to (α, n). In an irreducible scheme, every $\Pi(\alpha, n)$ is non-empty. It follows from (3) by a simple inductive argument that we have

Proposition 2.5 In an irreducible scheme satisfying (CX), the event times satisfy

$$T_\alpha(n) = \max_{\pi \in \Pi(\alpha, n)} \sum_{(\beta, k) \in \pi} \omega_\beta(k). \tag{4}$$

If we think of $\omega_\beta(k)$ as the length of step (β, k) in a path π, then (4) states that $T_\alpha(n)$ is the length of the longest path to (α, n). For any pair (β, k), $\beta \in \mathbf{A}$, $k = 1, 2, \ldots$, define $\Pi_{\beta,k}(\alpha, n) = \{\pi' = (\beta_{i_1}, k_{i_1}) \cdots (\beta_{i_r}, k_{i_r}) \ : \ \pi\pi' \in \Pi(\alpha, n) \text{ for some } \pi \in \Pi(\beta, k)\}$. Thus, the elements of $\Pi_{\beta,k}(\alpha, n)$ are the tails of paths to (α, n) that pass through (β, k). This set is empty if and only if $k > \mathbf{x}_\beta(\alpha, n)$. If $k \leq \mathbf{x}_\beta(\alpha, n)$, then some path to (α, n) passes through (β, k), so

$$T_\alpha(n) \geq T_\beta(k) + \max_{\pi \in \Pi_{\beta,k}(\alpha, n)} \sum_{(\beta', k') \in \pi} \omega_{\beta'}(k').$$

We do not have equality because the longest path to (α, n) may not pass through (β, k). However, if we can identify a set of pairs $(\beta_1^*, k_1^*), \ldots, (\beta_m^*, k_m^*)$ such that every path in $\Pi(\alpha, n)$ passes through at least one of the (β_i^*, k_i^*), $i = 1, \ldots, m$, then in particular the longest path must pass through one of these points and we obtain equality:

Corollary 2.6 Suppose $\{(\beta_i^*, k_i^*), i = 1, \ldots, m\}$ have the property that for every $\pi \in \Pi(\alpha, n)$ there is some $(\beta_i^*, k_i^*) \in \pi$, $i = 1, \ldots, m$. Then

$$T_\alpha(n) = \max_{i=1,\ldots,m} \{T_{\beta_i^*}(k_i^*) + \max_{\pi \in \Pi_{\beta_i^*, k_i^*}(\alpha, n)} \sum_{(\beta', k') \in \pi} \omega_{\beta'}(k')\}, \tag{5}$$

with a max over empty set taken to be $-\infty$.

The minimal score $\mathbf{x}(\alpha, n)$ provides a set of pairs through which every path to (α, n) must path: take the set $\{(\beta, \mathbf{x}_\beta(\alpha, n)), \beta \in \mathbf{A}\}$. For this choice, (5) simplifies to (3).

2.2 Homogeneous Minimal Elements

The availability of recursions like (3) is a powerful tool in analyzing the convergence of $T_\alpha(n)/n$, but (3) by itself is not quite enough. In a stochastic setting, the clock times $\{\xi_\alpha(n), \alpha \in \mathbf{A}, n = 1, 2, \ldots\}$ should stabilize (we will assume stationarity), and the minimal elements $\mathbf{x}(\alpha, n)$ should also, in some sense, stabilize as n increases. To see what can happen with arbitrary indexing, consider the following scheme:

$$\xrightarrow{\alpha}\xrightarrow{\beta}\xrightarrow{\alpha}\xrightarrow{\alpha}\xrightarrow{\beta}\xrightarrow{\beta}\xrightarrow{\alpha}\xrightarrow{\alpha}\xrightarrow{\alpha}\xrightarrow{\alpha}\cdots$$

The lengths of runs of each event are powers of two. This scheme trivially satisfies (CX) and its event times can be represented as sums of clock times. Connecting this with (3), we have $\mathbf{x}_\alpha(\alpha, n) = n - 1$ and

$$\mathbf{x}_\beta(\alpha, 2^n) = \cdots = \mathbf{x}_\beta(\alpha, 2^{n+1} - 1) = 2^n - 1.$$

Suppose all clock times are identically equal to 1. Then $T_\alpha(2^n) = 2^{n+1} - 1$ and $T_\alpha(2^{n+1} - 1) = 2^{n+1} + 2^n - 2$, so $\limsup T_\alpha(n)/n = 2$ whereas $\liminf T_\alpha(n)/n = 1.5$. This discrepancy is a consequence of the increasingly long runs of α's and β's. To rule out this type of behavior, we introduce a condition on the minimal elements which, in any case, appears to be satisfied in most applications:

Definition 2.7 A GSMS satisfying (CX) has *homogeneous* minimal elements if for all $\alpha, \beta \in \mathbf{A}$ and all $n = 1, 2, \ldots$,

$$\mathbf{x}_\beta(\alpha, n) = \max\{\mathbf{x}_\beta(\alpha, n+1) - 1, 0\}. \tag{6}$$

A more straightforward condition would be $\mathbf{x}_\beta(\alpha, n+1) = \mathbf{x}_\beta(\alpha, n) + 1$, but this version does not account for the possibility that $\mathbf{x}_\beta(\alpha, n+1) = 0$. Most queueing models that satisfy (CX) have homogeneous minimal elements. From a practical standpoint, the only potentially interesting models ruled out by (6) are those that require, say, exactly two occurrences of β between occurrences of α. Such a model could be incorporated by replacing the -1 on the right side of (6) with -2 or, more generally, $-k$. That option complicates the exposition so we do not pursue it further.

It follows from (6) that, for all α and β, either $\mathbf{x}_\beta(\alpha, n) = 0$ for all n, or else there exists an integer $u_{\alpha\beta}$ such that $\mathbf{x}_\beta(\alpha, n) = n - u_{\alpha\beta}$ for all sufficiently large n. To unify these two cases, we set

$$u_{\alpha\beta} = \sup_{n \geq 1}\{n - \mathbf{x}_\beta(\alpha, n)\}.$$

Recalling our convention that $T_\alpha(n) = 0$ for $n \leq 0$ and extending it to $n = -\infty$, from (3) we get

Corollary 2.8 In a scheme satisfying (CX) with homogeneous minimal elements, the event times satisfy

$$T_\alpha(n) = \omega_\alpha(n) + \max_\beta\{T_\beta(n - u_{\alpha\beta})\} \tag{7}$$

We next obtain a compact representation of the above recursion using (max,+) algebra, as in the work of [1, 3, 4, 5, 11]. For real numbers a, b define $a \oplus b = \max(a, b)$ and $a \otimes b = a + b$. Extend $\mathbf{R}$ to include $-\infty$ with $a \oplus -\infty = a$ for all a. Clearly, $a \otimes 0 = a$ for all a. If v is a (row) vector over $\mathbf{R} \cup \{-\infty\}$ and A is a matrix, define $v \otimes A$ by replacing $+, \times$ in ordinary vector-matrix multiplication with $\oplus, \otimes$; more specifically,

$$(v \otimes A)_i = \bigoplus_j (v_j \otimes A_{ji}) = \max_j (v_j + A_{ji}).$$

Extend this to matrix multiplication in the usual way.

Proposition 2.9 In a scheme satisfying (CX) with homogeneous minimal elements, the sequence $\{\tilde{T}(n), n \geq 0\}$ satisfies

$$\tilde{T}(n) = \tilde{T}(n-1) \otimes A(n-1) \tag{8}$$

for a sequence of matrices $\{A(n), n \geq 0\}$ that have the following property: if ω' denotes the shifted sequence defined by $\omega'_\alpha(n) = \omega_\alpha(n+1)$, then $A(n+1, \omega) = A(n, \omega')$

Note that since for all α and k we have $T_\alpha(k) \geq T_\alpha(k-1) + \omega_\alpha(k)$, the matrices $\{A(n), n \geq 0\}$ may be modified so that every diagonal entry is equal to some clock time and therefore greater than $-\infty$.

As an example, consider m queues in tandem with finite intermediate buffers. Node 1 draws new jobs from an infinite supply; jobs completed at node m leave the system immediately. The buffer between nodes i and $i+1$, $i = 1, \ldots, m-1$, has room for k_i jobs, including one in service at node $i+1$. If upon completion of service at node i a job finds the downstream buffer full, it remains at node i which then becomes blocked. This and more general blocking mechanisms are consistent with (CX), as discussed in [7].

The dynamics of this system are conveniently summarized through the matrix $\mathbf{u}$, where

$$u_{ij} = \begin{cases} 1 + k_{i+1} + \cdots + k_j, & i < j; \\ 1, & i = j; \\ 0, & i > j. \end{cases}$$

For $i < j$, u_{ij} bounds the number of jobs that may be completed by node i but not yet by node j; server i becomes blocked when the limit u_{ij} is reached for some $j = i+1, \dots, m$.

Let β_i denote service completion at node i, $i = 1, \dots, m$. Then, taking $s^0 = (0, \dots, 0)$, we have $\mathbf{x}_{\beta_j}(\beta_i, n) = n - u_{ij}$, and

$$\begin{aligned} & T_{\beta_i}(n) = \omega_{\beta_i}(n) \\ + \quad & \max\{T_{\beta_{i-1}}(n), T_{\beta_i}(n-1), T_{\beta_j}(n-u_{ij}), j = i+1, \dots, m\}. \end{aligned}$$

Since the n-th service completion at node j precedes the n-th service completion at node i whenever $j < i$, and since $u_{ij} = 0$ for $j < i$, we may rewrite the above as

$$T_{\beta_i}(n) = \omega_{\beta_i}(n) + \max\{T_{\beta_j}(n-u_{ij}), j = 1, \dots, m\},$$

which has precisely the form of (7) with $u_{\beta_i \beta_j} = u_{ij}$. Using the fact that upstream service completions always precede downstream service completions, we obtain

$$T_{\beta_i}(n) = \max_{k=1,\dots,i} \max_{j=k,\dots,m} \{T_{\beta_j}(n-u_{kj}) + \sum_{r=k}^{i} \omega_r(n)\}.$$

3 Subadditivity and Stability

We now use the framework of GSMPs satisfying (CX) with homogeneous minimal elements to establish the existence of cycle times, i.e., of $\lim_{n\to\infty} T_n(\alpha)/n$. We use the linear recursion (8) together with subadditivity in a stochastic setting.

3.1 The Subadditive Ergodic Theorem

For background, we include in this section a statement of Kingman's subadditive ergodic theorem. Before doing so, we state an elementary result. A sequence $\{a_1, a_2, \dots\}$ of real numbers is called *subadditive* if

$$a_{m+n} \le a_m + a_n, \quad m, n = 1, 2, \dots.$$

If $\{a_n, n \ge 1\}$ is subadditive, then $\{a_n/n, n \ge 1\}$ has a limit as $n \to \infty$, possibly equal to $-\infty$. Cohen [4] includes a proof, citing an exercise in Pólya and Szegö [13].

Kingman's [10] result is formulated in terms of *subadditive processes*. These are processes $X = \{X_{mn}, m = 0, 1, \dots, n = m+1, m+2, \dots\}$ satisfying the following conditions:

(S1) If $i < j < k$, then $X_{ik} \le X_{ij} + X_{jk}$, a.s.

(S2) The joint distributions of the process $\{X_{m+1,n+1}, n > m\}$ are the same as those of $\{X_{mn}, n > m\}$.

(S3) The expectation $g_n = \mathsf{E}[X_{0n}]$ exists and satisfies $g_n \ge -cn$ for some finite constant c and all $n = 1, 2, \dots$.

A consequence of (S1), (S3) and the elementary result given above is that $\gamma = \lim_{n\to\infty} g_n/n$ exists and is finite. We can now state Kingman's subadditive ergodic theorem:

Theorem (Kingman [10]). If X is a subadditive process, then the finite limit

$$\zeta = \lim_{n\to\infty} X_{0n}/n$$

exists almost surely, and $\mathsf{E}[\zeta] = \gamma$

Condition (S2), on the shift $\{X_{mn}\} \mapsto \{X_{m+1,n+1}\}$, is a stationarity condition. If all events defined in terms of X that are invariant under this shift have probability zero or one, then X is *ergodic*. In this case, as discussed by Kingman [10, p.885], the limiting random variable ζ is almost surely constant and equal to γ. It is this version of the result that we will use. Notice that the limit provided by Kingman's theorem holds in expectation, as well as almost surely: $\lim_{n\to\infty} n^{-1}\mathsf{E}[X_{0n}] = \mathsf{E}[\zeta]$.

3.2 Products of Random Matrices

Cohen [4] gives an excellent account of connections between subadditive ergodic theory and products of random matrices, and considers, among other settings, the case of $(\max, +)$ matrix multiplication. For purposes of reference and comparison, we paraphrase his Theorem 4:

Theorem (Cohen [4]). Let $\{A(n), n = 1, 2, \dots\}$ be a stationary and ergodic sequence of random $d \times d$ real matrices and let

$$P(m,n) = A(m+1) \otimes \cdots \otimes A(n), \quad m+1 \le n. \tag{9}$$

If $-\infty < \mathsf{E}[A_{ij}(n)] < \infty$ for all $1 \le i, j \le d$, then the finite limit

$$\lim_{n\to\infty} n^{-1} P(0,n)_{ij} = \gamma$$

exists almost surely, is a constant, and is independent of i and j.

If the matrices in the recursion (8) satisfied the hypotheses of Cohen's theorem, we would immediately be able to conclude that

$$\lim_{n\to\infty} n^{-1}\tilde{T}(n) = \lim_{n\to\infty} n^{-1}\mathbf{0} \otimes (A(0) \otimes \cdots \otimes A(n-1)) = [\gamma],$$

where $\mathbf{0}$ is the zero vector and $[\gamma]$ is a vector with all entries equal to γ. Unfortunately, even if the matrices in (8) are stationary and ergodic, we have seen that they typically include entries equal to $-\infty$ required to effect a permutation of certain entries of $\tilde{T}(n)$. Permutations were required to obtain a first-order recursion. So, we need a generalization of Cohen's result for the types of matrices arising in our setting.

To carry out this generalization, we need some properties of ordinary (non-random) matrices in $(\max, +)$ algebra. All of these are straightforward analogs of results for non-negative matrices under standard matrix multiplication (as in Chapter 2 of Pullman [14]) with $-\infty$ playing the role usually played by 0. We write $A^{\otimes n}$ for the n-fold $\otimes$-product of A with itself.

Definition 3.1 The $d \times d$ matrix A is *reducible* if $d \ge 2$ and if there exists a permutation of the rows and columns of A under which it has the block form

$$\begin{pmatrix} B & C \\ -\infty & D \end{pmatrix},$$

where B and D are square matrices. Otherwise, A is *irreducible*. In particular, every 1×1 matrix is irreducible.

Lemma 3.2 The $d \times d$ matrix A, $d \ge 2$, is irreducible if and only if for all i and j there exists an n such that $A_{ij}^{\otimes n} > -\infty$.

Definition 3.3 The $d \times d$ matrix A, $d \ge 2$ is *periodic* if its rows and columns can be permuted to give it the block form

$$\begin{pmatrix} -\infty & B^1 & -\infty & \cdots & -\infty \\ -\infty & -\infty & B^2 & \cdots & -\infty \\ \vdots & & \ddots & \ddots & \vdots \\ -\infty & -\infty & \cdots & -\infty & B^{k-1} \\ B^k & -\infty & \cdots & -\infty & -\infty \end{pmatrix}.$$

Otherwise, A is called *aperiodic*. Every 1×1 matrix is aperiodic.

Lemma 3.4 If A is irreducible and aperiodic, and A is not the 1×1 matrix $(-\infty)$, then for some n, $A_{ij}^{\otimes n} > -\infty$ for all i, j.

Lemma 3.5 Through a permutation of its rows and columns, any matrix A can be put in the block form

$$\begin{pmatrix} A^1 & & & * \\ & A^2 & & \\ & & \ddots & \\ -\infty & & & A^K \end{pmatrix}, \tag{10}$$

where $A^1, \dots, A^K$ are irreducible, the entries below the block diagonal are $-\infty$, and the entries above the block diagonal are arbitrary.

It is not hard to see that in multiplying matrices with $(\max, +)$ algebra, the location of $-\infty$'s in the product depends only on the location of $-\infty$'s in the matrices multiplied. (The same is true of 0's when we multiply non-negative matrices in standard algebra.)

We now impose some stochastic conditions:

(A1) The matrix sequence $\{A(n), n \geq 0\}$ is stationary and ergodic.

(A2) For each i, j, the entry $A_{ij}(0)$ is integrable on the event that it exceeds $-\infty$; i.e., $\mathsf{E}[|A_{ij}(0)|; A_{ij}(0) > -\infty] < \infty$.

(A3) For each i, j the probability that $A_{ij}(0) = -\infty$ is zero or one.

Condition (A3) ensures that the location of $-\infty$'s is the same among all $\{A(n), n \geq 0\}$; results for powers of a single matrix, concerning the location of $-\infty$'s, therefore extend to products of the $\{A(n), n \geq 0\}$. Given (A3), condition (A2) simply states that each entry not identically equal to $-\infty$ is integrable.

We can now prove a preliminary generalization of Cohen's [4] Theorem 4. Let $P(m, n)$ be as in (9).

Lemma 3.6 Assume (A1)-(A3). Suppose the matrix $A(0)$ is irreducible and aperiodic. Then

$$\lim_{n\to\infty} n^{-1}P(0, n)_{ij} = \gamma$$

exists almost surely and is independent of i and j. The limit is finite unless $A(0)$ is the 1×1 matrix $(-\infty)$.

Lemma 3.6 allows matrix entries equal to $-\infty$, but it is not yet adequate for the types of matrices arising in (8); the irreducibility condition is too strong. To obtain a sufficiently general result, we need to look more closely at the matrix decomposition (10).

This decomposition of a $d \times d$ matrix A partitions the indices $\{1, \ldots, d\}$ into K classes corresponding to the K irreducible submatrices on the block diagonal. Denote these classes by $S_1, \ldots, S_K$. For fixed A, let us say that there is a path from i to j if for some n, $A_{ij}^{\otimes n} > -\infty$. Thus, for $d \geq 2$ an irreducible matrix is one in which there is a path between every pair of indices. The condition $A_{ij}^{\otimes n} > -\infty$ corresponds to the existence of a sequence $k_1, \ldots, k_{n+1}$ with $k_1 = i$, $k_{n+1} = j$, and $A_{k_r k_{r+1}} > -\infty$ for $r = 1, \ldots, n$. Let us say that there is a path from i to j through S_ℓ if, for some $k \in S_\ell$, $A_{ik}^{\otimes n_1} > -\infty$ and $A_{kj}^{\otimes n_2} > -\infty$, for some n_1, n_2. We now have

Lemma 3.7 Suppose that in the decomposition (10) of A the submatrices $A^1, \ldots, A^K$ are aperiodic (as well as irreducible). Then A has the following property: there exists an n_* such that

$$A_{ij}^{\otimes n_*} > -\infty \quad \Rightarrow \quad A_{ij}^{\otimes n} > -\infty \text{ for all } n \geq n_*;$$

$$A_{ij}^{\otimes n_*} = -\infty \quad \Rightarrow \quad A_{ij}^{\otimes n} = -\infty \text{ for all } n \geq n_*.$$

To extend this to products of random matrices, observe that if $\{A(n), n \geq 1\}$ satisfy conditions (A1) and (A3), and if $A(0)$ satisfies the condition in the lemma, then the conclusion of the lemma applies to all $P(m, m+n)$, $n \geq n_*$, for some (deterministic) n_*. This, again, follows from our remark on the location of $-\infty$'s. For each i, j, let $\rho(i, j)$ consist of those classes S_ℓ for which some (hence all) $k \in S_\ell$ satisfies $P(0, n_*)_{ik} > -\infty$ and $P(0, n_*)_{kj} > -\infty$. Thus, $S_\ell \in \rho(i, j)$ means that there are arbitrarily long paths from i to j passing through S_ℓ. We can now establish

Theorem 3.8 Suppose (A1)-(A3) hold and that in the decomposition (10) of $A(0)$ the submatrices $A^1(0), \ldots, A^K(0)$ are aperiodic (as well as irreducible). Then the matrix limit

$$\lim_{n\to\infty} n^{-1}P(0, n) = \Gamma$$

exists, almost surely. The entries of Γ satisfy

$$\Gamma_{ij} = \begin{cases} -\infty, & \rho(i, j) = \emptyset; \\ \max_{S_\ell \in \rho(i,j)} \gamma_\ell, & \text{otherwise,} \end{cases} \tag{11}$$

where γ_ℓ is the constant attached to $\{A^\ell(n), n \geq 1\}$ by Lemma 3.6.

Note there are three cases covered by (11). If i and j belong to the same class, then $\rho(i, j)$ consists of at most that class and the result reduces to Lemma 3.6. If the class containing i has a higher index than that containing j, then there is no path from i to j and the limit is $-\infty$. If j's class has a higher index than i's, then for large n the longest path from i to j spends most of its time cycling through the class S_ℓ, $\ell \in \rho(i, j)$ with the largest average cycle length γ_ℓ.

As discussed in the related settings of Baccelli [1] and Cohen [4], the constants γ_ℓ appearing in (11) are analogs of *Lyapunov exponents* in ordinary products of random matrices. In [1] and [4], a corresponding connection is made with *Oseledec's multiplicative ergodic theorem*. Oseledec's theorem (and its analog in [1]) is concerned with the action of random matrix products on individual vectors. It includes a partial characterization of certain random subspaces for the limiting product associated with the Lyapunov exponents. In our (more specialized) setting, we give, instead, a complete characterization of the limiting matrix Γ itself, and show that the corresponding subspaces (corresponding to the block structure of Γ) are non-random.

3.3 Cycle Times

We now combine the representations of Section 2.2 with the convergence results of Section 3.2 to establish the existence of cycle times. To apply Theorem 3.8, we need to verify the aperiodicity condition. In general, a sufficient condition for an irreducible matrix to be aperiodic is that it have at least one diagonal entry greater than $-\infty$. From this we get

Lemma 3.9 The matrices in Proposition 2.9 may be selected to have upper-triangular representations with aperiodic, irreducible submatrices on the diagonal.

To apply Theorem 3.8, we assume that the clock times $\xi = \{\xi_\alpha(n), \alpha \in \mathbf{A}, n = 1, 2, \ldots\}$ are stationary, meaning that $\{\xi_\alpha(n + 1), \alpha \in \mathbf{A}, n = 1, 2, \ldots\}$ has the same joint distributions as ξ, and ergodic, meaning that any shift-invariant events have probability zero or one. This gives

Theorem 3.10 Consider an irreducible GSMP whose scheme satisfies (CX) with homongeneous minimal elements. Suppose the clock times are integrable, stationary and ergodic. Then the finite limit

$$\gamma_\alpha = \lim_{n\to\infty} n^{-1}T_\alpha(n)$$

exists almost surely for each $\alpha \in \mathbf{A}$ and is independent of the initial state s^0.

It is a simple consequence of this result that cycle times exist in the tandem-queues example of Section 2.2. More precisely, suppose the service times are integrable, stationary and ergodic. Then the finite limits

$$\lim_{n\to\infty} n^{-1}T_{\beta_i}(n) = \gamma_i, \; i = 1, \ldots, m,$$

exist almost surely and are independent of the initial state. Moreover, if $i < j$ then $\gamma_i \leq \gamma_j$; if $u_{ij} < \infty$, meaning that node i may be blocked by node j, then $\gamma_i = \gamma_j$.

4 Rate of Convergence

The subadditive ergodic theorem guarantees the existence of a limit for a (normalized) subadditive process, but says nothing about the rate of convergence. All ergodic theorems may be viewed as generalizations of the strong law of large numbers; convergence rates and error bounds for strong laws are provided by central limit theorems. In this section, we develop bounds to complement the convergence of sequences $\{n^{-1}T_\alpha(n), n \geq 0\}$. These bounds are formally similar to Gaussian approximations but are not based on central limit theorems (which are not generally available for subadditive sequences). Instead, they follow from a martingale inequality. The main additional assumptions we need are that the clock times are i.i.d. and bounded. Our use of this method follows the application in Rhee and Talagrand [15] to bin-packing and traveling-salesman problems.

Throughout this section, we consider the event times of an irreducible scheme satisfying (CX) with homogeneous minimal elements. Our first step bounds the difference $T_\alpha(n) - \mathsf{E}[T_\alpha(n)]$ for any α and n. Write $\xi(j)$ for the vector of clocks $(\xi_\alpha(j), \alpha \in \mathbf{A})$. Define

$$\mathcal{F}_i = \sigma\text{-algebra generated by } \{\xi(j)), j = 1, \ldots, i\},$$

and let $\mathcal{F}_0$ be the trivial σ-algebra. Clearly, $\{\mathcal{F}_n, n \geq 0\}$ is an increasing family. For fixed α and n, define

$$D_i = \mathsf{E}[T_\alpha(n)|\mathcal{F}_i] - \mathsf{E}[T_\alpha(n)|\mathcal{F}_{i-1}], \; i = 1, \ldots, n. \tag{12}$$

We always have $\mathsf{E}[T_\alpha(n)|\mathcal{F}_0] = \mathsf{E}[T_\alpha(n)]$; if, in addition, $T_\alpha(n)$ is $\mathcal{F}_n$-measurable, then $\mathsf{E}[T_\alpha(n)|\mathcal{F}_n] = T_\alpha(n)$ and, by telescoping the sum we get

$$\sum_{i=1}^{n} D_i = T_\alpha(n) - \mathsf{E}[T_\alpha(n)]. \tag{13}$$

This expresses the total error $T_\alpha(n) - \mathsf{E}[T_\alpha(n)]$ as the sum of individual errors D_i.

For this decomposition to be useful, the D_i's must have some structure. The relevant property is this:

Definition 4.1 Random variables $\{Y_n, n \geq 1\}$ form a *martingale difference sequence* (MDS) with respect to an increasing family $\{\mathcal{F}_n, n \geq 0\}$ of σ-algebras, if each Y_n is $\mathcal{F}_n$-measurable and $\mathsf{E}[Y_n|\mathcal{F}_{n-1}] = 0$.

By its very definition, each D_i in (12) is $\mathcal{F}_i$-measurable. Moreover, from (12),

$$\mathsf{E}[D_i|\mathcal{F}_{i-1}] = \mathsf{E}[T_\alpha(n)|\mathcal{F}_{i-1}] - \mathsf{E}[T_\alpha(n)|\mathcal{F}_{i-1}] = 0,$$

so $\{D_i, i = 1, \ldots, n\}$ is an MDS. The MDS representation becomes useful through the following result:

Lemma 4.2 Let $\{D_i, i = 1, \ldots, n\}$ be an MDS. Then for each $t \geq 0$,

$$P(|\sum_{i=1}^{n} D_i| > t) \leq 2\exp\left(-t^2/(2\sum_{i=1}^{n} ||D_i||_\infty^2)\right), \tag{14}$$

where $||D_i||_\infty$ is the essential supremum of D_i.

This result is stated in Rhee and Talagrand [15], where a references to a proof also appears. With this lemma we have a way to bound $T_\alpha(n) - \mathsf{E}[T_\alpha(n)]$ using (13) if we can bound the D_i's.

We impose the following assumptions:

(B1) The vectors $\{\xi(n), n \geq 1\}$ are i.i.d. and integrable.

(B2) There exists a constant c such that $P(\xi_\alpha(1) \leq c) = 1$ for all $\alpha \in \mathbf{A}$.

(B3) For all $\alpha, \beta \in \mathbf{A}$, $u_{\alpha\beta} \geq 0$.

Condition (B1) strengthens our earlier assumption of stationarity and ergodicity of the clock times, but still allows dependence among the components of $\xi(n)$ for each n. Condition (B2) requires that the clock times be bounded. Condition (B3) ensures that $T_\alpha(n)$ is completely determined by $\{\xi(i), i \leq n\}$, i.e., that $T_\alpha(n)$ is $\mathcal{F}_n$-measurable. Under (B3), (13) holds. We now have

Theorem 4.3 If (B1)-(B3) hold, then for any $t \geq 0$,

$$P(|T_\alpha(n) - \mathsf{E}[T_\alpha(n)]| > t) \leq 2\exp(-t^2/(2n|\mathbf{A}|^2c^2)),$$

where c is the constant in (B2).

The main result of this section is the following:

Theorem 4.4 Suppose in addition to the conditions of Theorem 3.10 that (B1)-(B3) hold. Then for all $\epsilon > 0$ there exists an $n_0 < \infty$ such that for all $n \geq n_0$,

$$P(|n^{-1}T_\alpha(n) - \gamma_\alpha| > \epsilon) \leq 2\exp(-n\epsilon^2/(2|\mathbf{A}|^2c^2)). \tag{15}$$

The bound in (15) is useful in estimating γ_α through simulation. It can be used to construct a confidence interval for γ_α that is valid for all sufficiently large n, not just asymptotically. Moreover, using (15) obviates the sometimes difficult task of estimating an asymptotic variance: the (known) constant $|\mathbf{A}|^2c^2$ replaces the variance. Of course, when $\{\sqrt{n}(n^{-1}T_\alpha(n) - \gamma_\alpha), n \geq 0\}$ satisfies a central limit theorem, we would expect $|\mathbf{A}|^2c^2$ to be an upper bound on its variance constant. Confidence invterval halfwidths provided by (15) are $O(n^{-1/2})$ just as with a central limit theorem.

References

[1] BACCELLI, F. Ergodic Theory of Stochastic Petri Networks, *Ann. Prob.* **20**, 375-396, 1992.

[2] BACCELLI, F., AND LIU, Z. On a Class of Stochastic Recursive Sequences Arising in Queueing Theory *Ann. Prob.* **20**, 350-374, 1992.

[3] COHEN, G., DUBOIS, D., QUADRAT, J.-P., AND VIOT, M., A Linear-System-Theoretic View of Discrete-Event Processes and its Use for Performance Evaluation in Manufacturing. *IEEE Trans. Automatic Control,* **AC-30**, 210-220, 1985.

[4] COHEN, J.E., Subadditivity, Generalized Products of Random Matrices and Operations Research, *SIAM Review* **30**, 69-86, 1988.

[5] CUNNINGHAME-GREEN, R., *Minimax Algebra*, Lecture Notes in Economics and Mathematical Systems 166, Springer-Verlag, New York, 1979.

[6] GLASSERMAN, P., Structural Conditions for Perturbation Analysis Derivative Estimation: Finite Time Performance Indices, *Oper. Res.* **39**, 724-738, 1991.

[7] GLASSERMAN, P., AND YAO, D.D., Algebraic Properties of Some Stochastic Discrete Event Systems, with Applications, *DEDSTA* **1**, 7-36, 1991.

[8] GLASSERMAN, P., AND YAO, D.D., Monotonicity in Generalized Semi-Markov Processes, *Math. Oper. Res.* **17**, 1-21, 1992.

[9] GLASSERMAN, P., AND YAO, D.D., Generalized Semi-Markov Processes: Antimatroid Structure and Second-Order Properties, *Math. Oper. Res.* **17**, 444-469, 1992.

[10] KINGMAN, J.F.C., Subadditive Ergodic Theory, *Ann. Prob.* **1**, 883-909, 1973.

[11] OLSDER, G.J., Eigenvalues of Dynamic Max-Min Systems, *DEDSTA* **1**, 177-208, 1991.

[12] OLSDER, G.J., RESING, J.A.C., DE VRIES, R.E., KEANE, M.S., AND HOOGHIEMSTRA, G., Discrete Event Systems with Stochastic Processing Times, *IEEE Trans. Automatic Control,* **35**, 299-302, 1990.

[13] PÓLYA, G., AND SZEGÖ, G., *Problems and Theorems in Analysis, Vol.1,* Second Edition, Springer-Verlag, New York, 1976.

[14] PULLMAN, N.J., *Matrix Theory and its Applications: Selected Topics.* Marcel Dekker, New York, 1976.

[15] RHEE, W., AND TALAGRAND, M., Martingale Inequalities and NP-Complete Problems, *Math. Oper. Res.* **12** 177-181.

[16] SHANTHIKUMAR, J.G., AND YAO, D.D., Stochastic Monotonicity in General Queueing Networks, *J. Appl. Prob.* **26**, 413-417, 1989.