RJ 10138 (95013) January 20, 1999
Computer Science

Research Report

ESTIMATION OF DELAYS IN NON-REGENERATIVE STOCHASTIC PETRI NETS

Peter J. Haas

IBM Research Division
Almaden Research Center
650 Harry Road
San Jose, CA 95120-6099

IBM Research Division
Yorktown Heights, New York • San Jose, California • Zurich, Switzerland

ESTIMATION OF DELAYS IN NON-REGENERATIVE STOCHASTIC PETRI NETS

Peter J. Haas

IBM Research Division
Almaden Research Center
650 Harry Road
San Jose, CA 95120-6099
e-mail: `peterh@almaden.ibm.com`

ABSTRACT: Many long-run delay characteristics of computer, communication, manufacturing, and transportation systems can be specified as time-average limits of sequences of delays in stochastic Petri nets (SPN's). We consider SPN's with general firing times in which sequences of delays are determined from the marking changes of the net using the method of start vectors. In this setting, time-average limits typically must be estimated using simulation. Previous work on estimation methods for delays has focused on SPN's in which there exists a sequence of regeneration points for the marking process. For such nets, point estimates and confidence intervals for time-average limits can be obtained using the regenerative method for simulation output analysis or recently proposed variants of the regenerative method. This paper is concerned with SPN's for which regenerative methods are not applicable. We provide conditions on the building blocks of an SPN and start-vector mechanism under which the sequence of delays is an "od-regenerative" process with finite cycle-sum moments. Although the od-regeneration points usually cannot be determined explicitly, the existence of these points implies that time-average limits are well defined and the sequence of delays obeys a multivariate functional central limit theorem. It then follows from results of Glynn, Iglehart, and Muñoz that methods based on standardized time series can be used to obtain strongly consistent point estimates and asymptotic confidence intervals for time-average limits and functions of time-average limits. In particular, the method of batch means is applicable.

Keywords: stochastic simulation, standardized time series, batch means, stochastic Petri nets, passage times, delays, discrete-event stochastic systems, Markov chains, stability

1. Introduction

Performance assessment of computer, communication, manufacturing, and transportation systems often involves analysis of long-run delay characteristics. Examples of such characteristics include the long-run average time to produce an item in a flexible manufacturing system, the long-run fraction of queries in a database system that require more than a specified amount of time to compute, and the long-run variability in the time required to transmit a message packet from one node to another in a communication network. When the system of interest is modeled as a stochastic Petri net (SPN), each of these performance characteristics can be expressed as a time-average limit of the form $\lim_{n\to\infty}(1/n)\sum_{j=0}^{n-1} f(P_j)$, where f is a real-valued function and $P_0, P_1, \ldots$ is a sequence of delays determined by the marking changes of the net. For SPN's with generally-distributed firing times (Haas and Shedler [21]) such time-average limits typically cannot be computed analytically or numerically, but must be estimated using simulation. This paper deals with estimation methods for time-average limits.

A delay in an SPN usually is computed as the length of a corresponding "passage-time interval" (that is, a random time interval) whose start (left endpoint) and termination (right endpoint) each coincide with a marking-change epoch. We focus on sequences of delays that can be specified using the method of "start vectors" introduced by Haas and Shedler [24, 25]. This method uses a recursively-generated sequence of random vectors to provide the link between the starts and terminations of individual passage-time intervals. Heuristically, the nth such "start vector" records the start of passage-time intervals for all ongoing passage times and newly started passage times at the nth marking change. The values of the starts and the order of the starts in the start vector together summarize the history of the marking process and comprise sufficient information to measure individual lengths. At each marking change, the current time may be inserted, old starts may be deleted, and the components of the start vector may be permuted according to a mechanism that depends explicitly on the current marking, new marking, and set of transitions that simultaneously fire and trigger the marking change. Deleted starts are subtracted from the current time to compute delays. This method for specifying and measuring delays avoids the need to "tag" individual entities in the system (such as customers) by using either distinguishable tokens (as in colored SPN's [7, 23, 40]) or additional places and transitions [24, Example 2.5]. The method is particularly effective when (as is often the case in practice) there can be more than one ongoing passage time at a time point and passage times need not terminate in the order in which they start. In this paper, we actually consider a slight extension of the original method of start vectors in which passage-time intervals of length zero are permitted.

Formal definition of the marking process of an SPN with general firing times is in terms of a general state space Markov chain that describes the net at successive marking changes. Previous work on estimation methods for delays in SPN's has focused on nets for which there exists an infinite sequence of random indices, called regeneration points, that decomposes sample paths of this underlying chain into independent and identically distributed (i.i.d.) "cycles." For example, if there exists a marking s in which exactly one transition e is enabled and if s is hit infinitely often (i.o.) with probability 1, then the successive times at which the marking is s and transition e fires form a sequence of regeneration points. (Such a marking is sometimes called a "single state.") Each regeneration point for the chain gives rise to a regeneration point (in continuous time) for the marking process. When there are no ongoing passage times at any regeneration point for the marking process, the sequence of delays is a regenerative process in discrete time. Under suitable moment conditions, the standard regenerative method for analysis of simulation output (see, for example, [37, Chapter 3]) then can be used [34] to obtain strongly consistent point estimates and asymptotic confidence intervals for time-average limits of the sequence of delays. When there are ongoing passage times at each regeneration point, sample paths of the process $\{ P_j : j \geq 0 \}$ can be decomposed into one-dependent and identically distributed (o.i.d.) cycles, and either an extension of the standard regenerative method or a "multiple runs" method can be used to obtain point estimates and confidence intervals; see [24, 25]. Also, as discussed in [24], indirect estimation methods based on Little's Law can be used in the special case in which we wish to estimate the limiting average delay $\lim_{n\to\infty}(1/n)\sum_{j=0}^{n-1} P_j$. Regenerative-type estimation methods for delays also have been established in the setting of colored SPN's [23], networks of queues (see [37] and references therein), and generalized semi-Markov processes (GSMP's) [25, 37].

This paper deals with methods for estimation of time-average limits when there is no apparent sequence of regeneration points, so that the above estimation methods are not applicable. In particular, we consider estimation methods based on standardized time series (STS). We use the term "standardized time series" in the sense of Glynn and Iglehart [13] and refer not only to the original STS-area and STS-maximum methods introduced by Shruben [38], but to any estimation method based on "cancelling out the variance constant" as described in [13]. For example, the well-known method of batch means (with the number of batches independent of the simulation run length) is an STS method, as is the method of spaced batch means and STS-weighted-area method; see Conway [8], Brillinger [5], Fox, Goldsman, and Swain [11], and Goldsman, Meketon, and Shruben [15].

It is nontrivial to determine for a specific SPN, start-vector mechanism, and function f whether the output process $\{ f(P_j) : j \geq 0 \}$ obeys a strong law of large numbers (SLLN) so that the time-average limit is well defined. It is even harder to determine whether STS methods can be used to estimate the time-average limit, assuming that it exists. Birkhoff's Ergodic Theorem provides general conditions under which a stochastic process obeys an SLLN; see, for example, Section 6.2 in Durrett [9]. A key condition of the Ergodic Theorem is that the process be stationary. Brillinger [5] establishes the validity of the batch means method under the stationarity assumption and Shruben [38] establishes an analogous result for the STS-maximum and STS-area methods. In the simulation setting, however, it is usually impossible to select the initial marking and clock readings of an SPN so as to ensure stationarity. Glynn and Iglehart [13] avoid the stationarity assumption by showing that STS methods are applicable when the output process obeys a *functional central limit theorem* (FCLT). Such a limit theorem asserts that a sequence of stochastic processes converges in distribution to a Brownian motion, generalizing the ordinary CLT assertion that a sequence of random variables converges in distribution to a normal random variable. The discussion in [13] does not address the key problem of determining for a specific sequence of delays whether an FCLT holds.

In this paper we provide conditions on the building blocks of an SPN under which time-average limits exist and STS methods are applicable. The idea is to build upon the results in [18, 19, 20], which provide sufficient conditions for the existence of a sequence of "od-regeneration points" that decompose the underlying chain of a marking process into o.i.d. cycles. A stochastic process that permits such decomposition is called an "od-regenerative process;" see Glynn [12] and Sigman [39]. Under a suitable strengthening of the conditions in [18, 19, 20], we show that the output process $\{ f(P_j) : j \geq 0 \}$ inherits the od-regenerative property and the sum of the process over a cycle has finite moments of all orders. Unlike in [24], the cycles of the output process usually cannot be determined explicitly, and neither the regenerative method nor its variants can be applied. The mere existence of these cycles, however, implies that the output process obeys an FCLT. It then follows from [13] that STS methods such as the method of batch means can be used to obtain strongly consistent point estimates and asymptotic confidence intervals for time-average limits. Moreover, an extension of the method of batch means due to Muñoz and Glynn [29] can be used to obtain point estimates and confidence intervals for functions of time-average limits.

Besides the studies cited above, most of the previous work related to delays in SPN's has been restricted to specific types of delays, such as "cycle times" between successive firings of a fixed transition, and particular kinds of SPN's, such as "stochastic marked graphs" or SPN's with exponentially distributed firing times; see, for example, Baccelli and Schmidt [3], Campos, Colom, Chiola, and Silva [6], Molloy [28], Muppala, Trivedi, Mainkar, and Kulkarni [31], and Natkin [32]. The primary emphasis has been on the development of exact and approximate analytical expressions and bounds for average delays. Baccelli, Cohen, Olsder, and Quadrat [2] derive sufficient conditions for time-average convergence of sequences of cycle times in stochastic marked graphs with general firing times. We consider time-average limits for a broad collection of delays in a class of SPN's that has the same modeling power (see Haas and Shedler [22]) as GSMP's. Glynn [12] considers sequences of delays determined by an underlying "Harris recurrent" Markov chain [27], and shows that sample paths of the corresponding output process can be decomposed into o.i.d. cycles. The discussion in [12] applies to sequences for which there is at most one ongoing passage time at any time point; neither criteria for the applicability of the results to specific systems nor the connection to STS estimation methods are considered.

The remainder of this paper is organized as follows. In Section 2 we review our SPN formalism and the method of start vectors. We then summarize some basic properties of od-regenerative processes in Section 3 and provide conditions on the building blocks of an SPN under which the underlying chain of the marking process has od-regenerative structure. Section 4 contains the key result of the paper, which shows that the conditions of Section 3 together with a mild regularity condition on the start-vector mechanism imply that there exists a sequence of od-regeneration points for the output process $\{ f(P_j) \colon j \geq 0 \}$ and the sum of the $f(P_j)$'s over a cycle has finite moments of all orders. In Section 5 we discuss the implications of this result for estimation of time-average limits. Section 6 contains our conclusions.

2. Delays in SPN's

In this section, we recall the definition of an SPN with general firing times. We also review (and extend slightly) the method of start vectors for specification and measurement of delays in SPN's.

2.1. Stochastic Petri Nets with General Firing Times

Following [21, 22], let $D = \{ d_1, d_2, \ldots, d_L \}$ be a finite set of *places*, $E = \{ e_1, e_2, \ldots, e_M \}$ be a finite set of *transitions*, and $E' \subset E$ a (possibly empty) set of *immediate transitions*. The transitions in $E - E'$ are called *timed transitions*. Also let $I(e), L(e), J(e) \subseteq D$ be the sets of *normal input places*, *inhibitor input places*, and *output places*, respectively, for transition $e \in E$. Denote by G the finite or countably infinite set of *markings*. For $s \in G$ we write $s = (s_1, s_2, \ldots, s_L)$, where s_j is the number of *tokens* in place $d_j \in D$. Set $E(s) = \big\{ e \in E \colon s_j \geq 1 \text{ for } d_j \in I(e) \text{ and } s_j = 0 \text{ for } d_j \in L(e) \big\}$. The set $E(s)$ (assumed nonempty) is the set of transitions that are *enabled* when the marking is s. A transition $e \in E - E(s)$ is *disabled* when the marking is s. Define the set S' of *immediate markings* by $S' = \{ s \in G \colon E(s) \cap E' \neq \varnothing \}$ and the set S of *timed markings* by $S = G - S' = \{ s \in G \colon E(s) \cap E' = \varnothing \}$. According to this definition, an element of the marking set is an immediate marking if at least one immediate transition is enabled. The marking changes when one or more enabled transitions *fire*.

For $E^* \subseteq E(s)$, denote by $p(s'; s, E^*)$ the probability that the new marking is s' given that the marking is s and the transitions in the set E^* fire simultaneously. The new-marking probabilities are constrained by the requirement that a transition remove at most one token from each normal input place and deposit at most one token in each output place when it fires. The token count of a place may increase or decrease by more than 1 when transitions fire simultaneously.

Remark. The results in this paper also apply to SPN's in which a "weight" or "priority" is associated with each transition in order to ensure that only one transition fires when several clocks run down to zero simultaneously. Such "conflict-resolution" mechanisms can be viewed as special cases of our general marking-change mechanism.

Often in applications the stochastic mechanism for removing and depositing tokens is degenerate. A transition is said to be *deterministic* if, with probability 1, it removes exactly one token from each normal input place and deposits exactly one token in each output place whenever it fires (and no other transitions fire simultaneously).

A *clock* is associated with each transition. The clock reading for an enabled transition indicates the remaining time until the transition is scheduled to fire. These clocks, along with the speeds at which the clocks run down, determine which of the enabled transitions trigger the next marking change (according to a "race model"). Denote by $r(s,e)$ (≥ 0) the *speed* (finite, deterministic rate) at which the clock associated with transition e runs down when the marking is s. We do not allow zero speeds for immediate transitions; such transitions always fire the instant that they become enabled. By convention, $r(s,e) = 1$ if $e \in E' \cap E(s)$. We assume that $r(s,e) > 0$ for some $e \in E(s)$. Let $C(s)$ be the set of possible *clock-reading vectors* when the marking is s: $C(s) = \{ c = (c_1, \ldots, c_M) \colon c_i \geq 0 \text{ and } c_i > 0 \text{ if and only if } e_i \in E(s) \}$. (The ith component of a clock-reading vector $c = (c_1, \ldots, c_M)$ is the clock reading associated with transition e_i.) Beginning in marking s with clock-reading vector $c = (c_1, \ldots, c_M) \in C(s)$, the time $t^*(s,c)$ to the next marking change is given by

$$t^*(s,c) = \min_{\{ i \colon e_i \in E(s) \}} c_i / r(s, e_i), \tag{2.1}$$

where $c_i / r(s, e_i)$ is taken to be $+\infty$ when $r(s, e_i) = 0$. The set of transitions $E^*(s,c)$ that fire simultaneously and trigger the next marking change is given by

$$E^*(s,c) = \{ e_i \in E(s) \colon c_i - t^*(s,c) r(s, e_i) = 0 \}. \tag{2.2}$$

Observe that $E^*(s,c) = E(s) \cap E'$ whenever the marking s is immediate.

At a marking change from s to s' triggered by the simultaneous firing of the transitions in the set E^*, a finite clock reading is generated for each *new transition* $e' \in N(s'; s, E^*) = E(s') - (E(s) - E^*)$. Denote the *clock-setting distribution function* (that is, the distribution function of such a new clock reading) by $F(\,\cdot\,; s', e', s, E^*)$. For $e' \in E'$, we require that $F(0; s', e', s, E^*) = 1$ for s, s' and E^* so that immediate transitions always fire instantaneously. For $e' \in E - E'$, we require that $F(0; s', e', s, E^*) = 0$ for s, s' and E^* so that timed transitions never fire instantaneously. For each *old transition* $e' \in O(s'; s, E^*) = E(s') \cap \bigl(E(s) - E^*\bigr)$, the old clock reading is kept after the marking change. For $e' \in (E(s) - E^*) - E(s')$, transition e' (that was enabled before the transitions in E^* fired) becomes disabled and the clock reading is discarded.

Denote by $X(t)$ the marking of the SPN at time t. Formal definition of the *marking process* $\{ X(t) \colon t \geq 0 \}$ of an SPN with general firing times is in terms of a general state space Markov chain $\{ (S_n, C_n) \colon n \geq 0 \}$ that describes the net at successive marking changes; see [21] for details. Heuristically, $S_n = (S_{n,1}, \ldots, S_{n,L})$ represents the marking and $C_n = (C_{n,1}, \ldots, C_{n,M})$ represents the clock-reading vector just after the nth marking change. (For $1 \leq i \leq M$, $C_{n,i}$ is the clock reading for transition e_i.) Denote by Σ the state space of the chain. Also denote by μ the *initial distribution* of the chain; that is, for a subset $B \subseteq \Sigma$, the quantity $\mu(B)$ represents the probability that $(S_0, C_0) \in B$. We assume throughout that the marking process is initialized by selecting an initial timed marking $s_0 \in S$ according to a probability distribution p and then setting the clock for each enabled transition $e \in E(s_0)$

independently according to a distribution function $F(\,\cdot\,; e, s_0)$. Equivalently, we assume that the initial distribution satisfies

$$\mu\bigl(\{s_0\} \times [0, a_1] \times \cdots \times [0, a_M]\bigr) = p(s) \prod_{e_i \in E(s_0)} F_0(a_i; e_i, s_0) \tag{2.3}$$

for $s_0 \in S$ and $a_1, a_2, \ldots, a_M \geq 0$. For ease of exposition, we assume that for each $s_0 \in S$ and $e \in E(s_0)$ there exists s and E^* such that $F(\,\cdot\,; e, s_0) = F(\,\cdot\,; s_0, e, s, E^*)$.

Let ζ_n ($n \geq 0$) be the nonnegative, real-valued time of the nth marking change and assume throughout that

$$P_\mu\Bigl\{ \sup_{n \geq 0} \zeta_n = \infty \Bigr\} = 1 \tag{2.4}$$

so that only a finite number of marking changes occur in any finite time interval. This will be the case if G is finite and the SPN is "irreducible" as defined in Section 3; see [17, 19].

Remark. Here and elsewhere, we use the notation P_μ when displaying probabilities of events involving the chain or marking process in order to emphasize the fact that (S_0, C_0) is distributed according to μ. When the initial state is equal to (s, c) with probability 1, we write $P_{(s,c)}$. We similarly use the notation E_μ and $E_{(s,c)}$ to denote expectations of random quantities computed from the chain.

The class of SPN's described above has the same modeling power as GSMP's. For any GSMP there exists an SPN with a marking process that has the same finite-dimensional distributions. Conversely, for any marking process there exists a GSMP that has the same finite-dimensional distributions; see Haas and Shedler [22].

Let $\{\, \nu(n) \colon n \geq 0 \,\}$ be the indices of the successive marking changes at which the new marking is timed: $\nu(-1) = -1$ and

$$\nu(n) = \inf\{\, n > \nu(n-1) \colon S_n \in S \,\} \tag{2.5}$$

for $n \geq 0$. Define the *embedded chain* $\{\, (S_n^+, C_n^+) \colon n \geq 0 \,\}$ by setting

$$(S_n^+, C_n^+) = (S_{\nu(n)}, C_{\nu(n)})$$

for $n \geq 0$. Since each random index $\nu(n)$ is a stopping time with respect to the underlying chain $\{\, (S_n, C_n) \colon n \geq 0 \,\}$ and is almost surely (a.s.) finite by (2.4), it follows from the strong Markov property for $\{\, (S_n, C_n) \colon n \geq 0 \,\}$ that $\{\, (S_n^+, C_n^+) \colon n \geq 0 \,\}$ is indeed a well-defined general state space Markov chain. Denote by Σ^+ the state space of the embedded chain: $\Sigma^+ = \{\, (s, c) \in \Sigma \colon s \in S \,\}$. Observe that under the assumption in (2.3) the initial distribution of the embedded chain coincides with the initial distribution of the underlying chain.

2.2. Passage-Time Intervals and Start Vectors

A sequence of delays in an SPN is specified by means of nonnegative random variables $\{ A_j : j \geq 0 \}$ and $\{ D_j : j \geq 0 \}$ defined on the same probability space as the underlying chain $\{ (S_n, C_n) : n \geq 0 \}$. We restrict attention to sequences $\{ A_j : j \geq 0 \}$ and $\{ D_j : j \geq 0 \}$ such that $A_j = \zeta_{\alpha(j)}$ and $D_j = \zeta_{\beta(j)}$ for $j \geq 0$, where $\alpha(j)$ and $\beta(j)$ are a.s. finite random indices. That is, we restrict attention to passage times that start and terminate only at marking changes. We also focus on sequences for which the $\alpha(j)$'s are nondecreasing, so that delays are enumerated in start order. The $\beta(j)$'s need not be nondecreasing, however, reflecting the fact that there can be more than one ongoing passage time at a time point and passage times need not terminate in the order in which they start.

Methods based on tagging for measuring individual delays in an SPN with indistinguishable tokens may require a large number of additional places and transitions; see [24, Example 2.5]. The approach to specifying and measuring delays given in [24] avoids the need for tagging. The idea is to use a sequence of real-valued random vectors, called *start vectors*, to construct the sequences $\{ A_j : j \geq 0 \}$, and $\{ D_j : j \geq 0 \}$. The sequence $\{ V_n : n \geq 0 \}$ of start vectors is determined by the sample paths of the chain $\{ (S_n, C_n) : n \geq 0 \}$ and provides the link between the starts and terminations of the individual passage-time intervals. The nth start vector V_n records the starts of passage-time intervals for all ongoing passage times and newly started passage times at time ζ_n, that is, all starts $A_j = \zeta_{\alpha(j)}$ such that $\alpha(j) \leq n < \beta(j)$. Usually (but not necessarily) the positions of the starts in the start vector correspond to the locations in the system of entities, such as jobs or customers, for whom a passage time is underway. We assume that the current marking determines the length of the start vector and denote this length by $\phi(s)$ when the current marking is s. Some components of V_n may be equal to -1. As discussed below, lengths are never computed for passage-time intervals with negative starts, so that negative components of a start vector can be used to ensure that specified terminations do not result in the computation of a delay. The negative components typically serve as placeholders and correspond to entities that are in the system at time 0 but for whom no passage time is underway. The initial start vector is a specified vector, denoted $v_0(S_0)$, determined by the initial marking S_0 and having components that are equal to 0 or -1. Take $v_0(S_0)$ to be the empty vector $\varnothing$ when $\phi(S_0) = 0$.

Whenever the transitions in the set E^* fire simultaneously and trigger a marking change from s to s', a new start vector is obtained from the current start vector by

(i) inserting the current time at zero or more positions specified by an index vector $i_\alpha(s'; s, E^*)$;

(ii) deleting components at zero or more positions specified by an index vector $i_\beta(s'; s, E^*)$; and then

(iii) permuting the components according to an index vector $i_\pi(s'; s, E^*)$.

Components are deleted one at a time in the order in which the indices appear in the vector $i_\beta(s'; s, E^*)$. For each nonnegative component that is deleted, the length of a passage-time interval is computed by subtracting the deleted component from the current time. These deleted components are the left endpoints of passage-time intervals for the passage times that terminate at the current time. Deleted components equal to -1 are not used to compute lengths of passage-time intervals and are simply discarded.

Remark. In the formulation of start vectors given in [24], steps (i) and (ii) are interchanged. The current formulation permits specification of delays that are equal to 0. Such delays occur, for example, when a job arrives at an empty queue and experiences no queueing delay.

For a real-valued nonempty vector $v = (v_1, v_2, \ldots, v_k)$ and a nonempty index vector $i = (i_1, i_2, \ldots, i_l)$, denote by $\mathrm{Del}(v, i)$ the vector of length $k - l$ obtained from v by deleting the components at positions $i_1, i_2, \ldots, i_l$. Similarly, denote by $\mathrm{Ins}(v, i, \zeta)$ the vector of length $k + l$ obtained from v by inserting the value $\zeta \in \Re$ to the right of the components at positions $i_1, i_2, \ldots, i_l$. For example, if $v = (v_1, v_2, v_3, v_4, v_5)$ and $i = (0, 2, 2, 3, 5)$, then $\mathrm{Ins}(v, i, \zeta) = (\zeta, v_1, v_2, \zeta, \zeta, v_3, \zeta, v_4, v_5, \zeta)$. (If $v = \varnothing$, set $\mathrm{Ins}(v, i, \zeta) = (\zeta, \zeta, \ldots, \zeta)$, where the vector on the right side is of length l.) Finally, for a vector $i = (i_1, i_2, \ldots, i_k)$ of distinct indices with $1 \le i_1, i_2, \ldots, i_k \le k$, set $\mathrm{Per}(v, i) = (v_{i_1}, v_{i_2}, \ldots, v_{i_k})$ so that $\mathrm{Per}(v, i)$ is the vector of length k obtained from v by permuting the components according to the index vector i. By convention, $\mathrm{Del}(v, \varnothing) = \mathrm{Ins}(v, \varnothing, \zeta) = \mathrm{Per}(v, \varnothing) = v$.

The sequence $\{ V_n : n \ge 0 \}$ is generated recursively. Set $V_0 = v_0(S_0)$. Then set

$$\begin{aligned} V'_n &= \mathrm{Ins}\bigl(V_{n-1}, i_\alpha(S_n; S_{n-1}, E^*_{n-1}), \zeta_n\bigr), \\ V''_n &= \mathrm{Del}\bigl(V'_n, i_\beta(S_n; S_{n-1}, E^*_{n-1})\bigr), \end{aligned}$$

and

$$V_n = \mathrm{Per}\bigl(V''_n, i_\pi(S_n; S_{n-1}, E^*_{n-1})\bigr)$$

for $n \ge 1$. (Here $E^*_k = E^*(S_k, C_k)$ for $k \ge 0$, so that E^*_{n-1} is the set of transitions that trigger the nth marking change.)

Construct the sequence $\{ P_j : j \ge 0 \}$ in terms of the sequence $\{ V_n : n \ge 0 \}$ as follows. Denote by A'_j ($j \ge 0$) the jth nonnegative component that is deleted from a start vector in the sequence $\{ V_n : n \ge 0 \}$ and by D'_j the time at which A'_j is deleted. Then $[A'_0, D'_0], [A'_1, D'_1], [A'_2, D'_2], \ldots$ is the sequence of passage-time intervals, enumerated in order of increasing terminations. If there are no immediate transitions, obtain the sequence $\{ (A_j, D_j) : j \ge 0 \}$ by rearranging the sequence $\{ (A'_j, D'_j) : j \ge 0 \}$ in order of increasing starts and set $P_j = D_j - A_j$ for $j \ge 0$. If there is at least one immediate transition, two passage times can start at the same point in continuous time but at different marking changes. In this case we determine for $j \ge 0$ the random index $\alpha'(j)$ for which $A'_j = \zeta_{\alpha'(j)}$ (see [24] for details) and obtain the sequence $\{ (\alpha(j), A_j, D_j) : j \ge 0 \}$ by rearranging the sequence $\{ (\alpha'(j), A'_j, D'_j) : j \ge 0 \}$ in order of increasing value of the random indices. Then we compute P_j as before.

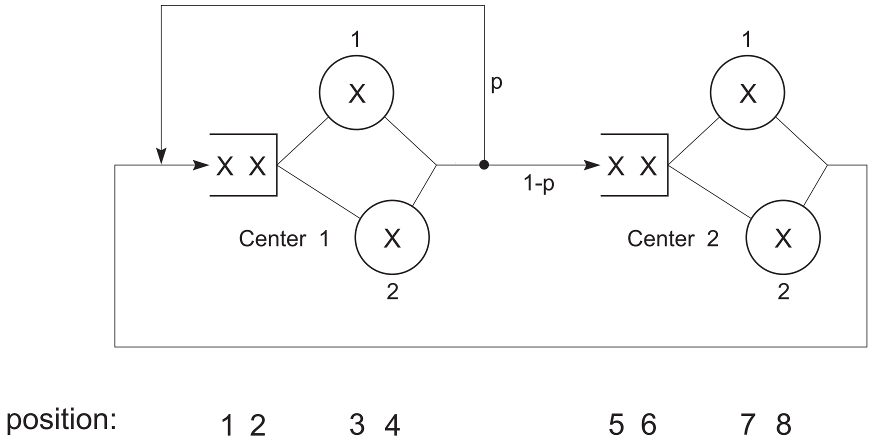

Figure 1: Positions of jobs in cyclic queues with feedback.

Denote by $V_{n,i}$ the ith component of the vector V_n for $1 \leq i \leq \phi(S_n)$, and set

$$K = \inf\{\, n \geq 0\colon V_{n,i} \neq -1 \text{ for } 0 \leq i \leq \phi(S_n) \,\}\,. \tag{2.6}$$

The total number of newly started passage times (of positive duration) and ongoing passage times at the nth marking change is less than or equal to $\phi(S_n)$ for $0 \leq n < K$ and equal to $\phi(S_n)$ for $n \geq K$.

Example 2.7. (Cyclic queues with feedback) Consider a closed network of queues with two service centers, two servers at each center, and N (> 2) jobs numbered $1, 2, \ldots, N$. At each center, jobs form a single queue and are served by one of two servers, numbered 1 and 2, according to a first-come, first-served queueing discipline; whenever a job joins an empty queue and both servers are idle, there is an immediate start of service by server 1. With fixed probability p $(0 < p < 1)$, upon completion of service at center 1 a job joins the tail of the queue at center 1 and with probability $1 - p$ moves to center 2. Upon completion of service at center 2, the job moves to center 1. Successive service times by server j at center i are i.i.d. as a positive random variable $L_{i,j}$ with continuous distribution function. (Under this assumption no two completions of service ever occur simultaneously.)

To specify the position of each job in the network, we can conceptually order the N jobs in a "job stack" as in Chapter 4 of [37]. The jobs at center 1 are closer to the top of the stack than the jobs at center 2. At each center, jobs waiting in queue are closer to the top of the stack than jobs in service; the jobs waiting in queue appear in the job stack in the order in which they join the tail of the queue, the latest to join being closest to the top of the stack. Moreover, a job being served by server 1 appears closer to the top of the stack than a job being served by server 2. The job at the top of the job stack is said to be in position 1, the next job in position 2, and so forth; see Figure 1 for $N = 8$ jobs.

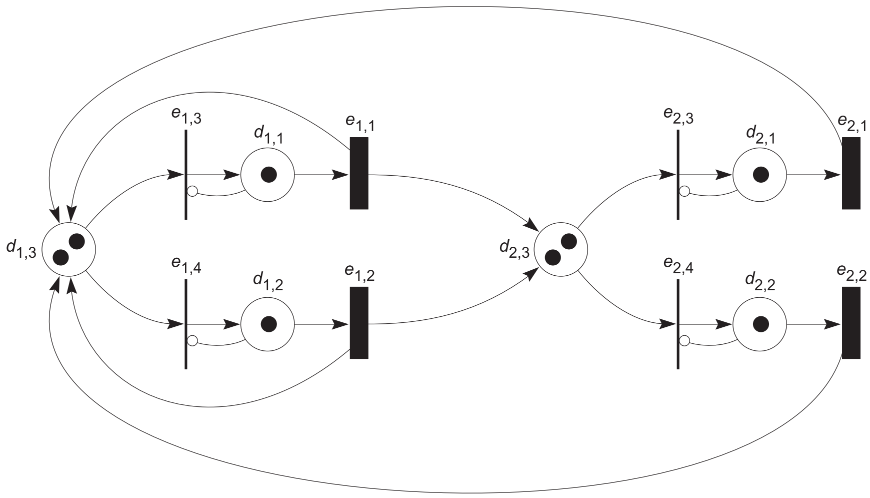

Figure 2: SPN representation of cyclic queues with feedback.

This system can be specified as an SPN with a finite marking set; see Figure 2. For ease of exposition, we use multiple subscripts to denote both transitions and places. Transition $e_{i,j}$ = "completion of service by server j at center i" for $i \in \{1,2\}$ and $j \in \{1,2\}$, and transition $e_{i,j}$ = "start of service by server $(j-2)$ at center i" for $i \in \{1,2\}$ and $j \in \{3,4\}$. Place $d_{i,j}$ $(i, j \in \{1,2\})$ contains either zero or one token; there is a token in place $d_{i,j}$ if and only if a service is underway by server j at center i. There are n tokens $(0 \le n \le N-2)$ in place $d_{i,3}$ $(i \in \{1,2\})$ if and only if there are n jobs waiting in queue at center i. All of the transitions except $e_{1,1}$ and $e_{1,2}$ are deterministic. Whenever transition $e_{1,j}$ fires $(j \in \{1,2\})$, a token is removed from place $d_{1,j}$; with probability p a token is deposited in place $d_{1,3}$ and with probability $1-p$ a token is deposited in place $d_{2,3}$. Formally, $p(s'; s, \{e_{1,1}\}) = p$ whenever $s = (1, s_{1,2}, s_{1,3}, s_{2,1}, s_{2,2}, s_{2,3}) \in S$ and $s' = (0, s_{1,2}, s_{1,3}+1, s_{2,1}, s_{2,2}, s_{2,3})$, and $p(s'; s, \{e_{1,1}\}) = 1-p$ whenever $s = (1, s_{1,2}, s_{1,3}, s_{2,1}, s_{2,2}, s_{2,3}) \in S$ and $s' = (0, s_{1,2}, s_{1,3}, s_{2,1}, s_{2,2}, s_{2,3}+1)$; otherwise, $p(s'; s, \{e_{1,1}\}) = 0$. We define $p(s'; s, \{e_{1,2}\})$ similarly. When there is a simultaneous firing of the transitions in the set $E^* = \{e_{i,3}, e_{i,4}, e_{3-i,j}\}$ $(i \in \{1,2\}$ and $j \in \{3,4\})$, one token is removed from each of places $d_{1,3}$ and $d_{2,3}$ and one token is deposited in each of places $d_{i,1}$ and $d_{3-i,j-2}$. For example, $p(s'; s, \{e_{1,3}, e_{1,4}, e_{2,4}\}) = 1$ when $s = (0, 0, s_{1,3}, s_{2,1}, 0, s_{2,3}) \in S'$ and $s' = (1, 0, s_{1,3}-1, s_{2,1}, 1, s_{2,3}-1)$. All speeds are equal to 1. The clock-setting distribution functions for timed transitions are given by $F(x; s', e_{i,j}, s, E^*) = P\{L_{i,j} \le x\}$ for $i \in \{1,2\}$ and $j \in \{1,2\}$. Suppose that at time 0 all jobs are at center 1 and there is a start of service by each of the servers at center 1. Then the initial distribution μ is of the form (2.3) with $p(s_0) = 1$ and $F_0(x; e_{1,j}, s_0) = P\{L_{1,j} \le x\}$ for $j \in \{1,2\}$, where $s_0 = (1, 1, N-2, 0, 0, 0)$.

Consider the passage-time intervals from whenever a job completes service at center 2 (and moves to center 1) to when the job next completes service at center 2. Individual delays in the foregoing SPN can be specified and measured using the method of start vectors. The start vector V_n records for each of the N jobs in the network the most recent time during the interval $[0, \zeta_n]$ at which there was a completion of service at center 2 and the job moved to center 1. If a job has never completed service at center 2 during the interval $[0, \zeta_n]$ the corresponding component of V_n is equal to -1. The components of the start vector are ordered from left to right according to increasing positions (as defined above) of the corresponding jobs in the network.

For $s = (s_{1,1}, \ldots, s_{2,3}) \in G$, $s' = (s'_{1,1}, \ldots, s'_{2,3}) \in G$ and $E^* \subseteq E(s)$, set $\phi(s) = N$. Set $i_\alpha(s'; s, E^*) = (0)$ if $E^* = \{ e_{2,1} \}$ or $\{ e_{2,2} \}$; otherwise, set $i_\alpha(s'; s, E^*) = \varnothing$. Set $i_\beta(s'; s, E^*) = (N+1)$ if $E^* = \{ e_{2,2} \}$ or if $E^* = \{ e_{2,1} \}$ and $s_{2,2} = 0$, and set $i_\beta(s'; s, E^*) = (N)$ if $E^* = \{ e_{2,1} \}$ and $s_{2,2} = 1$; otherwise, set $i_\beta(s'; s, E^*) = \varnothing$. Finally, set

$$i_\pi(s'; s, E^*) = \begin{cases} (s_{1,3}+1, 1, 2, \ldots, s_{1,3}, s_{1,3}+2, s_{1,3}+3, \ldots, N) & \text{if } E^* = \{ e_{1,1} \},\ s_{2,3} = s'_{2,3},\ s_{1,3} > 0; \\ (s_{1,3}+2, 1, 2, \ldots, s_{1,3}, s_{1,3}+1, s_{1,3}+3, \ldots, N) & \text{if } E^* = \{ e_{1,2} \},\ s_{2,3} = s'_{2,3},\ s_{1,1} = 1,\ s_{1,3} > 0; \\ (2, 1, 3, 4, \ldots, N) & \text{if } E^* = \{ e_{1,2} \},\ s_{2,3} = s'_{2,3},\ s_{1,1} = 1,\ s_{1,3} = 0; \\ (1, 2, \ldots, s_{1,3}-1, s_{1,3}+1, s_{1,3}, s_{1,3}+2, \ldots, N) & \text{if } E^* = \{ e_{1,4} \},\ s_{1,1} = 1,\ s_{1,3} > 1; \\ (2, 1, 3, 4, \ldots, N) & \text{if } E^* = \{ e_{1,4} \},\ s_{1,1} = 1,\ s_{1,3} = 1; \\ (1, 2 \ldots, N-2, N, N-1) & \text{if } E^* = \{ e_{2,4} \},\ s_{2,1} = 1; \\ \varnothing & \text{otherwise.} \end{cases}$$

Observe that when a job completes service at center 2 and joins the queue at center 1, the new start vector is obtained from the current start vector by inserting the current time to the left of the first component and then deleting the rightmost component. The deleted component is subtracted from the current time to compute a delay (provided that the component is nonnegative). In the above specification of i_π, the first three cases handle the cyclic permutation that occurs when a job completes service at center 1 and joins the tail of the queue at center 1. The fourth and fifth cases handle the situation in which, at center 1, a job starts to undergo a service by server 2 and thereby "overtakes" a job currently undergoing a service by server 1. The sixth case handles the analogous situation at center 2. For all other situations, no permutation is needed. If the initial distribution μ is as described above and the system behaves as if at time 0 a job has just completed service at center 2 and joined the tail of the queue at center 1, then we set $V_0 = v_0(s_0) = (0, -1, -1, \ldots, -1)$. That is, exactly one passage time starts at time 0; this start corresponds to the job that has just joined the queue at center 1. □

3. OD-Regenerative Processes

In this section we review some pertinent properties of od-regenerative processes. We then provide conditions on the building blocks of an SPN and a function $f\colon \Sigma \mapsto \Re$ under which there exists a sequence of od-regeneration points for the underlying chain $\{ (S_n, C_n)\colon n \geq 0 \}$ of the marking process and the sum of the $f(S_n, C_n)$'s over a cycle has finite moments of all orders. For the sequence $\{\theta(k)\colon k \geq 0\}$ of random indices defined below, set $\delta_k = \theta(k) - \theta(k-1)$ for $k \geq 1$.

Definition 3.1. The stochastic process $\{Z_n\colon n \geq 0\}$ with state space Γ is an *od-regenerative process* in discrete time if there exists an increasing sequence $\{\theta(k)\colon k \geq 0\}$ of a.s. finite random times such that the post-$\theta(k)$ process $\{Z_{\theta(k)+n}, \delta_{k+n+1}\colon n \geq 0\}$

(i) is distributed as the post-$\theta(0)$ process $\{Z_{\theta(0)+n}, \delta_{n+1}\colon n \geq 0\}$ for $k \geq 1$; and

(ii) is independent of the pre-$\theta(k-1)$ process $\{Z_0, Z_1, \ldots, Z_{\theta(k-1)-1}; \delta_1, \delta_2, \ldots, \delta_{k-1}\}$ for $k \geq 2$.

Our definition follows [39]. The random indices $\{\theta(k)\colon k \geq 0\}$ are called *od-regeneration points* for the process $\{Z_n\colon n \geq 0\}$ and serve to decompose sample paths of $\{Z_n\colon n \geq 0\}$ into o.i.d. cycles. The quantity δ_k is the length of the kth such cycle. For an $\Re^l$-valued function f ($l \geq 1$) defined on Γ, define the cycle sum $Y_k(f) = \sum_{n=\theta(k-1)}^{\theta(k)-1} f(Z_n)$ for $k \geq 1$. Observe that $\{(Y_k(f), \delta_k)\colon k \geq 1\}$ is a sequence of o.i.d. random vectors. An od-regenerative process in continuous time is defined in an analogous manner.

Remark. If, in addition, the cycle-lengths (though not necessarily the cycles themselves) are mutually independent, then the process is called an "od-equilibrium" process; see [19]. If the cycles are mutually independent, then the od-regenerative process reduces to a classical regenerative process as in [37].

An important example of an od-regenerative process is a *Harris recurrent* Markov chain $\{Z_n\colon n \geq 0\}$ with a (not necessarily countable) state space Γ and *l-step transition kernel* P^l for $l \geq 1$; see [27]. (For $z \in \Gamma$, $A \subseteq \Gamma$, and $l \geq 1$, $P^l(z, A)$ is the probability, starting in state z, that after l steps the chain hits a state that is an element of A.) Roughly speaking, a Harris chain is characterized by the property that any "dense enough" set of states (as indicated by positivity of a specified "recurrence measure") is hit by the chain infinitely often with probability 1. For such a chain, the od-regenerative property follows from the fact that there exists a set $\mathcal{C} \subset \Gamma$ such that $P\{Z_n \in \mathcal{C} \text{ i.o.}\} = 1$ and

$$P^r(z, \cdot) = b\lambda(\cdot) + (1-b)Q(z, \cdot), \qquad z \in \mathcal{C} \tag{3.2}$$

for some $r \geq 1$, $b \in (0,1)$, probability distribution λ, and transition kernel Q; see Section VI.3 in [1] and also [27, Theorem 5.2.3]. The decomposition in (3.2) permits construction of a version of the chain together with a sequence $\{\theta(k)\colon k \geq 0\}$ of random indices that serve as od-regeneration points. The idea is to generate successive states of the chain (according to the initial distribution μ and one-step transition kernel $P = P^1$) until the first time $M \geq 0$ such that $Z_M \in \mathcal{C}$. At this hitting time, generate an independent Bernoulli random variable I such that $P\{I = 1\} = 1 - P\{I = 0\} = b$. If $I = 1$, then generate Z_{M+r} according to λ; if $I = 0$, then generate Z_{M+r} according to $Q(Z_M, \cdot)$. Next, generate the intermediate states $Z_{M+1}, Z_{M+2}, \ldots, Z_{M+r-1}$ according to an appropriate conditional distribution (conditioned on the endpoint values Z_M and Z_{M+r}). Now iterate this procedure starting from state Z_{M+r}. The successive times $\theta(0), \theta(1), \ldots$ at which the state of the chain is distributed according to λ form a sequence of od-regeneration points. (It actually follows from the above construction that $\{\theta(k)\colon k \geq 0\}$ is a sequence of od-equilibrium points.)

Under mild regularity conditions, an od-regenerative process obeys both an SLLN and an FCLT. Let $\{Z_n\colon n \geq 0\}$ be an od-regenerative process with state space Γ and od-regeneration points $\{\theta(k)\colon k \geq 0\}$, and define δ_k and $Y_k(f)$ as above for $k \geq 1$ and $f\colon \Gamma \mapsto \Re^l$. Also set

$$\pi(f) = \frac{E\left[Y_1(f)\right]}{E\left[\delta_1\right]} = \frac{E\left[\sum_{n=\theta(0)}^{\theta(1)-1} f(Z_n)\right]}{E\left[\delta_1\right]}. \tag{3.3}$$

Observe that $\pi(f)$ exists and is finite if $\pi(|f|) < \infty$.

Proposition 3.4. *Suppose that $E\left[\delta_1\right] < \infty$ and define π as in* (3.3)*. Then $\pi(|f|) < \infty$ and*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(Z_i) = \pi(f) \text{ a.s.}$$

for any function $f\colon \Gamma \mapsto \Re^l$ $(l \geq 1)$ such that $E\left[Y_1(|f|)\right] < \infty$.

For $l \geq 1$, denote by $D^l[0,1]$ the space of (measurable) $\Re^l$-valued functions on $[0,1]$ that are right-continuous and have limits from the left and by $\Rightarrow$ weak convergence on $D^l[0,1]$; see Billingsley [4] or Ethier and Kurtz [10]. (Weak convergence on $D^l[0,1]$ generalizes to a sequence of $\Re^l$-valued random functions—i.e., a sequence of $\Re^l$-valued stochastic processes—the usual notion of convergence in distribution of a sequence of $\Re^l$-valued random variables.) Given a function $f\colon \Gamma \mapsto \Re^l$ $(l \geq 1)$, define a sequence of $D^l[0,1]$-valued random functions $U_1(f), U_2(f), \ldots$ by setting

$$U_n(f)(t) = \frac{1}{\sqrt{n}} \sum_{j=0}^{\lfloor nt \rfloor} \bigl(f(Z_j) - \pi(f)\bigr) \tag{3.5}$$

for $0 \leq t \leq 1$ and $n \geq 0$, where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x and π is defined as in (3.3). Denote by $W^{(l)} = \{\bigl(W_1^{(l)}(t), W_2^{(l)}(t), \ldots, W_k^{(l)}(t)\bigr)\colon 0 \leq t \leq 1\}$ a standard l-dimensional Brownian motion on $[0,1]$; see, for example, [29, Section 2]. Denote by x^t the transpose of a vector x.

Proposition 3.6. *Suppose that $E\left[\delta_1^2\right] < \infty$ and that $f\colon \Gamma \mapsto \Re^l$ $(l \geq 1)$ satisfies $E\left[Y_1(|f|)^t Y_1(|f|)\right] < \infty$. Then there exists an $l \times l$ matrix $Q(f)$ such that $U_n(f) \Rightarrow Q(f) W^{(l)}$ as $n \to \infty$ for any initial distribution of the process, where $U_n(f)$ is defined by* (3.5).

See [19] for proofs of Propositions 3.4 and 3.6. (In [19] the results are actually stated for od-equilibrium processes, but the proofs apply unchanged.)

Now consider an SPN with underlying chain $\{ (S_n, C_n) \colon n \geq 0 \}$ together with a function $f \colon \Sigma \mapsto \Re$. Theorem 3.8 below provides conditions (encapsulated in the "positive-density assumption" PDE given below) under which the underlying chain is an od-regenerative process and the sum of the $f(S_n, C_n)$'s over a cycle has finite moments of all orders. For $s \in S$ and $s' \in G$, write $s \rightarrow s'$ if $p(s'; s, \{ e \}) r(s, e) > 0$ for some $e \in E(s)$. Similarly, for $s \in S'$ and $s' \in G$, write $s \rightarrow s'$ if $p\big(s'; s, E(s) \cap E'\big) > 0$. Finally, write $s \rightsquigarrow s'$ if either $s \rightarrow s'$ or there exist markings $s^{(1)}, s^{(2)}, \ldots, s^{(n)} \in G$ $(n \geq 1)$ such that $s \rightarrow s^{(1)} \rightarrow \cdots \rightarrow s^{(n)} \rightarrow s'$. We say that the SPN is *irreducible* if $s \rightsquigarrow s'$ for each $s, s' \in G$. Following [27], denote by $\mathcal{G}^+$ the set of distribution functions on $[0, \infty)$ that have a convergent LaPlace-Stieltjes transform. That is, $F \in \mathcal{G}^+$ if and only if there exists $a_F > 0$ such that $\int_0^\infty e^{ux} \, dF(x) < \infty$ for $u \in [0, a_F]$. Observe that each distribution function $F \in \mathcal{G}^+$ has finite moments of all orders.

Definition 3.7. *Assumption PDE* is said to hold for a specified SPN if

(i) the marking set G is finite;

(ii) the SPN is irreducible;

(iii) all speeds are positive; and

(iv) there exists $0 < u \leq \infty$ such that each clock-setting distribution function $F(\,\cdot\,; s', e', s, E^*)$ with $e' \in E - E'$ belongs to $\mathcal{G}^+$ and has a density function that is positive and continuous on $(0, u)$.

Remark. In [19] we define Assumption PD(q) (where $q \geq 0$) identically to Assumption PDE except that each clock-setting distribution, rather than being required to belong to $\mathcal{G}^+$, is only required to have finite qth moment. Observe that if Assumption PDE holds, then Assumption PD(q) holds for $q \geq 0$.

In general, for $f \colon \Gamma \mapsto \Re^l$, $g \colon \Gamma \mapsto \Re$ and $A \subseteq \Gamma$, write $f = O(g; A)$ if $\sup_{x \in A} |f_j(x)|/|g(x)| < \infty$ for $1 \leq j \leq l$. (Here we take $0/0 = 0$.) When $A = \Gamma$, simply write $f = O(g)$. In the SPN setting, a function $f \colon \Sigma \mapsto \Re^l$ $(l \geq 1)$ is said to be *polynomially dominated* if f is bounded on $\Sigma - \Sigma^+$ and $f = O(h_q; \Sigma^+)$ for some $q \geq 0$, where $h_q(s, c) = 1 + \max_{1 \leq i \leq M} c_i^q$ for $s \in S$, $c = (c_1, c_2, \ldots, c_M) \in C(s)$, and $q \geq 0$.

Theorem 3.8. *Suppose that Assumption PDE holds. Then there exists a sequence $\{ \theta(k) \colon k \geq 0 \}$ of od-regeneration points for $\{ (S_n, C_n) \colon n \geq 0 \}$. Moreover,*

$$Y_1(|f|) = \sum_{n=\theta(0)}^{\theta(1)-1} |f(S_n, C_n)|$$

has finite moments of all orders for any polynomially dominated function $f \colon \Sigma \mapsto \Re$.

Remark. Observe that the od-regenerative cycle-length $\delta_1 = \theta(1) - \theta(0)$ has finite moments of all orders under the conditions of the theorem.

The proof of Theorem 3.8 is given in Appendix A. The idea is to first prove the result for the embedded chain $\{ (S_n^+, C_n^+) : n \geq 0 \}$ by establishing a "geometric drift" criterion for stability. This drift criterion implies that the embedded chain is Harris recurrent, and hence we can construct a sequence of od-regeneration points as described previously. Moreover, a new result due to Roberts and Tweedie [35] implies that the cycle lengths, and consequently the cycle sums, have finite moments of all orders. We then observe that the od-regeneration points for the embedded chain also can be taken as od-regeneration points for the underlying chain. To complete the proof, we use geometric-trials arguments similar to those in [19, Appendix B] to extend the finite-moment results from the embedded chain to the underlying chain.

4. Stable Sequences of Delays

Theorem 3.8 provides conditions on an SPN under which the underlying chain $\{ (S_n, C_n) : n \geq 0 \}$ is an od-regenerative process and cycle sums of the form $Y_k(|f|)$ have finite moments of all orders. It is natural to seek additional conditions under which these stability properties are inherited by a sequence of delays $\{ P_j : j \geq 0 \}$ determined from the marking changes of the net using the method of start vectors. As shown below, all that is needed is a mild regularity condition on the building blocks of the start-vector mechanism. This condition, which we call Assumption T, is needed to ensure that every passage time terminates with probability 1.

Denote by $\mathcal{X}$ the set of all infinite-length sequences $(s^{(0)}, E^{(0)}, s^{(1)}, E^{(1)}, \dots)$ such that, for all $j \geq 0$, $E^{(j)} = \{ e^{(j)} \}$ for some $e^{(j)} \in E(s^{(j)})$ if $s^{(j)} \in S$, $E^{(j)} = E(s^{(j)}) \cap E'$ if $s^{(j)} \in S'$, and $p(s^{(j+1)}; s^{(j)}, E^{(j)}) > 0$. For an element $x = (s^{(0)}, E^{(0)}, s^{(1)}, E^{(1)}, \dots) \in \mathcal{X}$, recursively define a sequence of vectors $v_0, v_1, \dots$ by setting $v_0(x) = \bigl(-1, -1, \dots, -1\bigr)$ (where $v_0(x)$ contains $\phi(s^{(0)})$ components) and then setting

$$
\begin{aligned}
v_n'(x) &= \mathrm{Ins}\bigl(v_{n-1}(x), i_\alpha(s^{(n)}; s^{(n-1)}, E^{(n-1)}), 0\bigr), \\
v_n''(x) &= \mathrm{Del}\bigl(v_n'(x), i_\beta(s^{(n)}; s^{(n-1)}, E^{(n-1)})\bigr),
\end{aligned}
$$

and

$$
v_n(x) = \mathrm{Per}\bigl(v_n''(x), i_\pi(s^{(n)}; s^{(n-1)}, E^{(n-1)})\bigr)
$$

for $n \geq 1$. Denote by $\iota(x)$ the smallest integer n such that $v_n(x)$ contains no negative components; if such an integer n does not exist, then set $\iota(x) = \infty$.

Definition 4.1. *Assumption T* is said to hold for a specified SPN and sequence of start vectors if there exists $x = (s^{(0)}, E^{(0)}, s^{(1)}, E^{(1)}, \dots) \in \mathcal{X}$ such that $\iota(x) < \infty$.

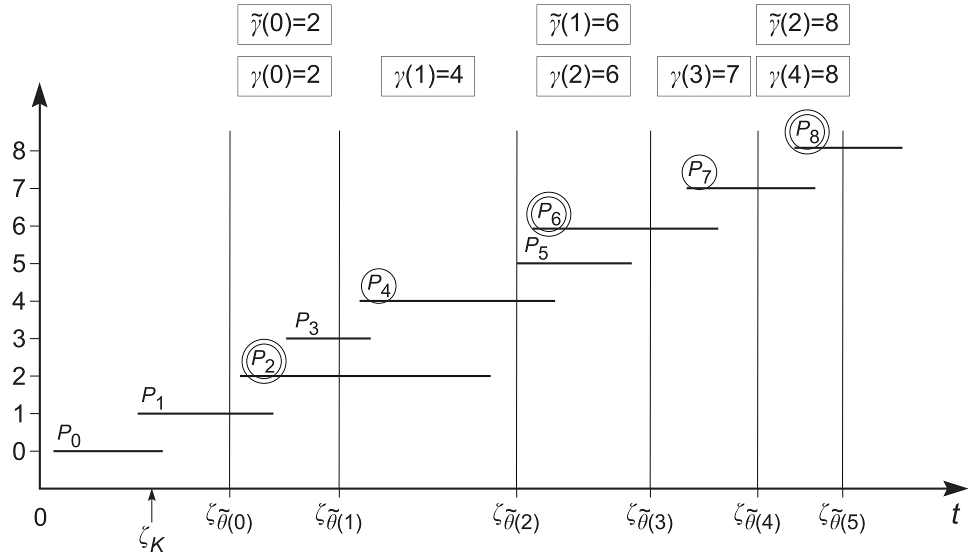

Figure 3: Definition of one-dependent cycles.

To prepare for our main result, Theorem 4.3 below, we construct the sequence of random indices $\{\, \tilde{\gamma}(k) \colon k \geq 0 \,\}$ that will serve as od-regeneration points for the process $\{\, P_j \colon j \geq 0 \,\}$. To this end, we first construct recursively a distinguished sequence $\{\, \zeta_{\tilde{\theta}(k)} \colon k \geq 0 \,\}$ of od-regeneration points for the marking process. These latter points are constructed to ensure that all passage times that start during a cycle terminate before the end of the next cycle. (In general, the $\tilde{\theta}(k)$'s are spaced farther apart than the $\theta(k)$'s constructed for the proof of Theorem 3.8.) Suppose that Assumption PDE holds. Then by results in [19] (specifically, Theorem 4.5 and Proposition 3.9), there exists $\mathcal{C} \subseteq \Sigma^+$, $r \geq 1$, $b \in (0,1)$, λ, and Q such that a decomposition as in (3.2) holds for the embedded chain $\{\, (S_n^+, C_n^+) \colon n \geq 0 \,\}$. Define a sequence of Bernoulli random variables $\{\, I_n : n \geq 0 \,\}$ by setting $I_n = 1$ with probability b and $I_n = 0$ with probability $1 - b$ whenever $(S_n, C_n) \in \mathcal{C}$; set $I_n = 0$ whenever $(S_n, C_n) \notin \mathcal{C}$. Also set $L_n = \inf\{\, l > n \colon I_l = 1 \,\}$ and $M(n;r) = \inf\{\, m \geq 1 \colon \sum_{i=n+1}^{n+m} 1_S(S_i) = r \,\}$ for $n \geq 0$. (Here and elsewhere $1_A(x)$ denotes the indicator function that equals 1 if $x \in A$ and equals 0 otherwise.) To define the sequence $\{\, \zeta_{\tilde{\theta}(k)} \colon k \geq 0 \,\}$, take $\tilde{\theta}(-1) = 0$. Then, given $\tilde{\theta}(k)$, wait until all ongoing passage times and newly started passage times at the $\tilde{\theta}(k)$th marking change terminate. Denote the distinguished marking change at which the last of these passage times terminates by $N(k)$; take $N(-1) = K$, where K is defined as in (2.6). Then take $\tilde{\theta}(k+1) = L_{N(k)} + M(L_{N(k)}; r)$. Thus, after waiting until all ongoing passage times and newly started passage times have terminated, we wait until the next time that the underlying chain hits the set $\mathcal{C}$ and the associated Bernoulli random variable equals 1, and then take $\tilde{\theta}(k+1)$ as the rth subsequent marking change at which the new marking is timed. We may assume that, as in the construction of Section 3 and proof of Theorem 3.8, each $(S_{\tilde{\theta}(k)}, C_{\tilde{\theta}(k)})$ is generated according to the distribution λ and the intermediate states $\{ (S_n, C_n) \colon L_{N(k-1)} < n < \tilde{\theta}(k) \}$ are generated according to an appropriate conditional distribution. Finally, set $\gamma(-1) = -1$ and

$$\gamma(k) = \inf\{ j > \gamma(k-1) \colon \alpha(j-1) < \tilde{\theta}(m) < \alpha(j) \text{ for some } m \geq 0 \}$$

for $k \geq 0$ (take $\alpha(-1) = 0$), and set

$$\tilde{\gamma}(k) = \gamma(2k) \tag{4.2}$$

for $k \geq 0$. These ideas are illustrated in Figure 3. In the figure, the delays $P_{\gamma(k)}$ are circled; concentric circles indicate the delays $P_{\tilde{\gamma}(k)}$. Two complete cycles $\{ P_2, P_3, P_4, P_5 \}$ and $\{ P_6, P_7 \}$ are displayed, as well as the beginning of a third cycle $\{ P_8, \dots \}$. Observe that the delays in the third cycle are detemined by the evolution of the marking process from time $\zeta_{\tilde{\theta}(4)}$ onwards, and hence are independent of the delays $\{ P_j \colon j \leq 5 \}$, which are detemined by the evolution of the marking process prior to time $\zeta_{\tilde{\theta}(3)}$.

We now state our key result. In the following, a real-valued function f is said to be *polynomially dominated* if $f = O(x^r + 1)$ for some $r \geq 0$.

Theorem 4.3. *Let $\{ P_j \colon j \geq 0 \}$ be a sequence of delays determined from the underlying chain $\{ (S_n, C_n) \colon n \geq 0 \}$ of a marking process using the method of start vectors. Suppose that both Assumption PDE and Assumption T hold. Then*

(i) each P_j is a.s. finite;

(ii) the random indices $\{ \tilde{\gamma}(k) \colon k \geq 0 \}$ defined by (4.2) form a sequence of od-regeneration points for $\{ P_j \colon j \geq 0 \}$; and

(iii) the cycle sum $\tilde{Y}_1(|f|) = \sum_{j=\tilde{\gamma}(0)}^{\tilde{\gamma}(1)-1} |f(P_j)|$ has finite moments of all orders for any polynomially dominated function f.

The proof of Theorem 4.3 is given in Appendix B. The idea is to use Assumption T to construct a set $A \subseteq \Sigma^k$ for some $k \geq 1$ such that all ongoing passage times at time ζ_n terminate by time ζ_{n+k} whenever $Z_n = (S_n, C_n, \dots, S_{n+k-1}, C_{n+k-1}) \in A$. It follows from Assumption PDE that the embedded chain is Harris recurrent, and a geometric-trials argument then shows that $P_\mu \{ Z_n \in A \text{ i.o.} \} = 1$. The first assertion of the theorem then follows directly. To establish the second assertion, we first show that the random indices $\{ \tilde{\theta}(k) \colon k \geq 0 \}$ form a sequence of od-regeneration points for the underlying chain $\{ (S_n, C_n) \colon n \geq 0 \}$. It then follows from this od-regenerative structure and the definition of the start-vector mechanism that the cycles of the process $\{ f(P_j) \colon j \geq 0 \}$ induced by the random indices $\{ \tilde{\gamma}(k) \colon k \geq 0 \}$ are o.i.d.. To prove the third assertion, we use a geometric-trials argument to show that the random number of $\theta(k)$-cycles (as in Theorem 3.8) that occur within a single $\tilde{\theta}(k)$-cycle has finite moments of all orders. The desired result then follows from this fact, the finite-moment properties of the $\theta(k)$-cycles guaranteed by Theorem 3.8, and a new extension of Theorem I.5.2 in Gut [16] from i.i.d random variables to o.i.d. random variables.

Combining Theorem 4.3 with Propositions 3.4 and 3.6, we obtain the following corollary.

Corollary 4.4. *Let $\{ P_j : j \geq 0 \}$ be a sequence of delays determined from the underlying chain of a marking process using the method of start vectors, and let $f : \Re_+ \mapsto \Re^l$ $(l \geq 1)$ be polynomially dominated. Suppose that both Assumption PDE and Assumption T hold. Then*

(i) there exists a finite constant $\pi(f) \in \Re^l$ such that $\lim_{n\to\infty}(1/n)\sum_{j=0}^{n-1} f(P_j) = \pi(f)$ a.s.; and

(ii) there exists an $l \times l$ matrix $Q(f)$ such that $U_n(f) \Rightarrow Q(f)W^{(l)}$ as $n \to \infty$ for any initial distribution μ, where $\Rightarrow$ denotes weak convergence on $D^l[0,1]$ and

$$U_n(f)(t) = \frac{1}{\sqrt{n}} \sum_{j=0}^{\lfloor nt \rfloor} \big(f(P_j) - \pi(f) \big) \tag{4.5}$$

for $0 \leq t \leq 1$ and $n \geq 0$.

5. Estimation Methods

In this section we focus on estimation of time-average limits and functions of time-average limits when the marking process has no apparent sequence of regeneration points.

5.1. Standardized Time Series

Let f be a real-valued function that is polynomially dominated. Under the conditions of Theorem 4.3, the time-average limit $\pi(f)$ of the output process $\{ f(P_j) : j \geq 0 \}$ is well-defined and finite, and we can obtain point estimates and confidence intervals for $\pi(f)$ using STS methods. To this end, set $\overline{Y}_n(t) = (1/n)\sum_{j=0}^{\lfloor nt \rfloor} f(P_j)$ for $0 \leq t \leq 1$ and $n \geq 1$, and set $\hat{\pi}_n = \overline{Y}_n(1)$. By Corollary 4.4(i), the point estimator $\hat{\pi}_n$ is strongly consistent for $\pi(f)$. To obtain asymptotic confidence intervals for $\pi(f)$, we follow the approach of Glynn and Iglehart [13]. Denote by $C[0,1]$ the set of continuous real-valued functions defined on $[0,1]$ and, for a mapping $\xi : C[0,1] \mapsto \Re$, let $D(\xi)$ be the set of discontinuity points for ξ. (If $x, x_1, x_2, \ldots \in C[0,1]$ with $\lim_{n\to\infty} \sup_{0\leq t\leq 1} |x_n(t) - x(t)| = 0$, then $x \in D(\xi)$ if and only if the relation $\lim_{n\to\infty} \xi(x_n) = \xi(x)$ fails to hold; see [4, Section 8].) As in [13], let ξ be a mapping from $C[0,1]$ to $\Re$ such that

(i) $\xi(ax) = a\xi(x)$ for $a \in \Re_+$ and $x \in C[0,1]$;

(ii) $\xi(x - be) = \xi(x)$ for $b \in \Re$ and $x \in C[0,1]$, where $e(t) = t$ for $0 \leq t \leq 1$;

(iii) $P\{\xi(W) > 0\} = 1$; and

(iv) $P\{W \in D(\xi)\} = 0$.

Also let $\sigma(f)$ be a nonnegative constant such that

$$U_n(f) \Rightarrow \sigma(f)W \tag{5.1}$$

as $n \to \infty$, where $U_n(f)$ is given by (4.5) and $W = \{W(t)\colon 0 \le t \le 1\}$ is a standard one-dimensional Brownian motion; the existence of $\sigma(f)$ follows from Corollary 4.4(ii) with $l = 1$. We focus throughout on the nondegenerate case in which $\sigma(f) > 0$. Finally, set $\xi_n = \xi(\overline{Y}_n)$. Arguing essentially as in [13], we find that the convergence in (5.1), the properties in (i), (ii), and (iv), and the Continuous Mapping Theorem [4, Theorem 5.1] together imply that $\sqrt{n}\xi_n = \xi\Big(\sqrt{n}\big(\overline{Y}_n - \pi(f)e\big)\Big) = \xi\big(U_n(f)\big) \Rightarrow \sigma(f)\xi(W)$ and $\sqrt{n}\big(\hat{\pi}_n - \pi(f)\big) = U_n(f)(1) \Rightarrow \sigma(f)W(1)$ as $n \to \infty$, where $\Rightarrow$ denotes ordinary convergence in distribution and the two sequences converge jointly. Using the property in (iii), it follows that

$$\frac{\hat{\pi}_n - \pi(f)}{\xi_n} \Rightarrow \frac{\sigma(f)W(1)}{\sigma(f)\xi(W)} = \frac{W(1)}{\xi(W)} \tag{5.2}$$

as $n \to \infty$. Choosing z_γ so that $P\{-z_\gamma \le W(1)/\xi(W) \le z_\gamma\} = \gamma$ (where $0 < \gamma < 1$), we obtain the asymptotic $100\gamma\%$ confidence interval

$$[\hat{\pi}_n - \xi_n z_\gamma, \hat{\pi}_n + \xi_n z_\gamma] \tag{5.3}$$

for $\pi(f)$. A key feature of this confidence interval is that there is no need (as in regenerative simulation) to estimate $\sigma^2(f)$; the variance constant has been "cancelled out" in (5.2).

Of course, it is necessary to determine the distribution of $W(1)/\xi(W)$ in order to choose appropriate values of z_γ. Observe in this connection that by definition of Brownian motion, $W(1)$ has a standard (mean 0, variance 1) normal distribution. Moreover, it is shown in [13] that $W(1)$ is independent of $\xi(W)$.

As pointed out in [13], the well-known *batch means* confidence interval is obtained as a special case of the interval in (5.3) by fixing $m \ge 2$ and taking

$$\xi(x) = \left[\frac{m}{m-1}\sum_{i=1}^{m}\Big(x(i/m) - x\big((i-1)/m\big) - x(1)/m\Big)^2\right]^{1/2}. \tag{5.4}$$

It is shown in [13] that when ξ is defined by (5.4), the conditions in (i)–(iv) hold and the limiting random variable $W(1)/\xi(W)$ has a Student's t distribution with $m-1$ degrees of freedom. Thus, setting

$$\overline{X}_n(i) = \overline{X}_n(i;f) = \frac{1}{n/m}\sum_{j=(i-1)n/m}^{(in/m)-1} f(P_j) \tag{5.5}$$

for $1 \leq i \leq m$ and $n > 0$, we find that the interval in (5.3) is an asymptotic $100\gamma\%$ confidence interval for $\pi(f)$ when z_γ is the $(1+\gamma)/2$ quantile of the Student's t distribution with $m-1$ degrees of freedom and

$$\xi_n = \frac{1}{\sqrt{m}} \left[\frac{1}{m-1} \sum_{i=1}^{m} \left(\overline{X}_n(i) - \frac{1}{m} \sum_{j=1}^{m} \overline{X}_n(j) \right)^2 \right]^{1/2}.$$

According to the foregoing formulas, the batch means confidence interval based on simulation of delays $P_0, P_1, \ldots, P_n$ is obtained by decomposing the sequence $f(P_0), f(P_1), \ldots, f(P_n)$ into m disjoint "batches" (intervals) of length n/m. (Typically, $10 \leq m \leq 30$; see [36].) The "batch mean" $\overline{X}_n(i)$ is the average of the $f(P_j)$'s over the ith such interval. In forming the confidence interval, the batch means $\{\, \overline{X}_n(i) \colon 1 \leq i \leq n/m \,\}$ are treated as if they are independent, normally distributed random variables. Although in general the independence and normality assumptions do not hold exactly, the resulting error in the confidence interval becomes negligible as the length of each of the m batches (equivalently, the length n of the simulation) becomes large. Other choices for ξ lead to alternative confidence intervals; see, for example, [11, 13].

5.2. Functions of Time-Average Limits

Let $f \colon \Re \mapsto \Re^l$ $(l \geq 1)$ be polynomially dominated, and consider a sequence of delays $\{\, P_j \colon j \geq 0 \,\}$ specified using the method of start vectors. Suppose that Assumption PDE and Assumption T hold, so that $\lim_{n\to\infty}(1/n)\sum_{j=0}^{n-1} f(P_j) \to \pi(f) = \big(\pi_1(f), \pi_2(f), \ldots, \pi_l(f)\big)$ a.s. by Corollary 4.4. In this section we focus on the problem of estimating a performance measure of the form $g\big(\pi(f)\big) = g\big(\pi_1(f), \pi_2(f), \ldots, \pi_l(f)\big)$, where $g \colon \Re^l \mapsto \Re$ is a nonlinear function. As shown in the next section, long-run variances and long-run conditional relative frequencies are performance measures of this type.

Following Muñoz and Glynn [29], fix $m \geq 2$ and set

$$J_n(i) = mg\Big(m^{-1} \sum_{j=1}^{m} \overline{X}_n(j)\Big) - (m-1)g\Big((m-1)^{-1} \sum_{j \neq i} \overline{X}_n(j)\Big)$$

for $1 \leq i \leq m$, where $\overline{X}_n(i)$ is defined as in (5.5). Then set $\hat{\pi}_n^{(J)} = (1/m)\sum_{i=1}^{m} J_n(i)$. The following result shows that $\hat{\pi}_n^{(J)}$ is strongly consistent for $g\big(\pi(f)\big)$; the proof is essentially the same as that of Theorem 5.6 in [19].

Theorem 5.6. *Suppose that Assumption PDE and Assumption T hold, that f is polynomially dominated, and that g is differentiable in a neighborhood of $\pi(f)$. Then $\hat{\pi}_n^{(J)} \to g\big(\pi(f)\big)$ a.s. as $n \to \infty$.*

To obtain confidence intervals, observe that by Corollary 4.4(ii) there exists an $l \times l$ matrix $Q(f)$ such that

$$U_n(f) \Rightarrow Q(f)W^{(l)} \tag{5.7}$$

as $n \to \infty$ for any initial distribution μ, where $U_n(f)$ is defined by (4.5) and $W^{(l)}$ is a standard l-dimensional Brownian motion. The matrix $Q(f)$ is non-singular except in degenerate cases, and we assume that $Q(f)$ is nonsingular throughout. It follows from the convergence in (5.7) and Theorem 2 in [29] that $\sqrt{m}\big(\hat{\pi}_n^{(J)} - \pi(f)\big)/s_n^{(J)} \Rightarrow t_{(m-1)}$ as $n \to \infty$, where $t_{(m-1)}$ denotes a random variable having a Student's t distribution with $m-1$ degrees of freedom and $s_n^{(J)} = \big[\big(1/(m-1)\big)\sum_{i=1}^{m}\big(J_n(i) - \hat{\pi}_n^{(J)}\big)^2\big]^{1/2}$. Thus

$$\left[\hat{\pi}_n^{(J)} - \frac{s_n^{(J)} t_{\gamma,m-1}}{\sqrt{m}}, \hat{\pi}_n^{(J)} + \frac{s_n^{(J)} t_{\gamma,m-1}}{\sqrt{m}}\right]$$

is an asymptotic $100\gamma\%$ confidence interval for $g\big(\pi(f)\big)$, where $t_{\gamma,m-1}$ is the $(1+\gamma)/2$ quantile of the Student's t distribution with $m-1$ degrees of freedom.

Remark. A number of point and interval estimators are considered in [29]; the "jackknifed" estimators presented here are designed to combat bias problems than can arise when estimating nonlinear functions of time-average limits.

5.3. An Example

Consider the closed network of queues in Example 2.7 and assume that $N > 4$. Suppose that for $j \in \{1,2\}$ the random variable $L_{1,j}$ has a uniform distribution on $[0, w_j]$, where $w_j > 0$ is a specified constant, and that random variables $L_{2,1}$ and $L_{2,2}$ each have a log-normal distribution. Under these distributional assumptions, there is no apparent sequence of regeneration points for the underlying chain or the marking process. Indeed, $|E(s)| > 1$ for all $s \in G$, so that there is no single state, and there is no "memoryless property" to exploit because there are no exponential distributions available. Each clock-setting distribution function, however, belongs to $\mathcal{G}^+$ and has a density function that is positive and continuous on the interval $(0, u)$, where $u = \min(w_1, w_2)$. It is straightforward to show that the marking process is irreducible, for example, by showing for each $s \in G$ that $s \to \tilde{s}$ and $\tilde{s} \to s$, where $\tilde{s} = (0,0,0,1,1,N-2)$ is the marking that corresponds to all jobs being at center 2 with both of the center 2 servers busy. Thus Assumption PDE holds. Assumption T also holds: if x is any element $(s^{(0)}, E^{(0)}, s^{(1)}, E^{(1)}, \ldots) \in \mathcal{X}$ such that $s^{(0)} = \tilde{s}$, $E^{(0)} = \{e_{2,2}\}$, $E^{(1)} = \{e_{1,3}, e_{1,4}, e_{2,4}\}$, $E^{(2)} = \{e_{2,1}\}$, $E^{(3)} = \{e_{1,4}, e_{2,3}\}$, $E^{(4)} = \{e_{2,1}\}$, $(E^{(5)}, E^{(6)}, \ldots, E^{(4+2(N-4))}) = (\{e_{2,3}\}, \{e_{2,1}\}, \{e_{2,3}\}, \{e_{2,1}\}, \ldots, \{e_{2,3}\}, \{e_{2,1}\})$, and $E^{(5+2(N-4))} = \{e_{2,2}\}$, then $\iota(x) = 6 + 2(N-4) < \infty$. This scenario starts with all jobs at center 2. Then, at center 2, there is a service completion by server 2, $N-2$ successive service completions by server 1, and then a final service completion by server 2.

Since Assumption PDE and Assumption T hold, any limit of the form

$$\pi(f) = \lim_{n\to\infty} (1/n) \sum_{j=0}^{n-1} f(P_j)$$

is guaranteed to exist, provided that the (real-valued) function f is polynomially dominated. Moreover, we can use STS methods as in Section 5.1 to obtain strongly consistent point estimates and asymptotic confidence intervals for $\pi(f)$. For example, if $f(x) = 1_{(u,\infty)}(x)$, then $\pi(f)$ is the long-run fraction of delays that exceed u time units. If $f(x) = cx$ for some constant a, then $\pi(f)$ is the long-run average cost due to delays when the cost per time unit of delay is c.

We also can use batch-means techniques of Section 5.2 to estimate functions of time-average limits of the form $g(\pi(f))$, where $f\colon \Re \mapsto \Re^l$ ($l \geq 1$) is polynomially dominated and g is differentiable in a neighborhood of $\pi(f)$. For example, if $f(x) = (x, x^2)$ and $g(x,y) = y - x$, then $g(\pi(f))$ is the long-run variance of the sequence of delays. If $f(x) = \big(1_{(a,\infty)}(x), 1_{(b,\infty)}(x)\big)$ with $b > a$ and $g(x,y) = y/x$, then $g(\pi(f))$ is the long-run fraction of delays in $\mathcal{D}_a$ that exceed b time units, where $\mathcal{D}_a$ is the set of delays that exceed a time units.

5.4. Remarks

In both Theorem 4.3 and Corollary 4.4 we require that each clock-setting distribution function for a timed transition belong to $\mathcal{G}^+$ and hence have finite moments of all orders. We conjecture that this moment condition can be relaxed considerably. Specifically, we conjecture that for a time-average limit $\lim_{n\to\infty}(1/n)\sum_{j=0}^{n-1} f(P_j)$ with $f = O(x^r)$, it suffices that each clock-setting distribution have finite moments of order most $2r+1$. The idea is to establish a drift condition as in [19] rather than the stronger geometric drift condition in (A.1). To complete the proofs, it then suffices to show that for any nonnegative function $g = O(x^s)$ ($s > 1$), the cycle variable $Y_1(g) = \sum_{n=\theta(k-1)}^{\theta(k)-1} g(S_n^+, C_n^+)$ has a finite kth moment($k \geq 2$) provided each clock-setting distribution has a finite moment of order $(ks + k + 1)$. For $k = 2$ this assertion follows from the drift condition in [19] together with [27, Lemma 17.5.2]; it should be possible to extend [27, Lemma 17.5.2] to cover the case $k > 2$.

Let f be a $\Re^l$-valued function for some $l > 1$. Under Assumption PDE and Assumption T, the techniques of Muñoz [30] can be used to construct confidence regions for multi-dimensional limits of the form $\lim_{n\to\infty}(1/n)\sum_{j=0}^{n-1} f(P_j)$.

The task of measuring individual delays can be avoided when the performance measure of interest is the limiting average delay $\lim_{n\to\infty}(1/n)\sum_{j=0}^{n-1} P_j$, i.e., when the performance function f is simply the identity function. Set $Z_n = (S_n, C_n, S_{n+1}, C_{n+1})$ for $n \geq 0$ and define functions f_1 and f_2 on $\Sigma \times \Sigma$ by $f_1(s,c,s',c') = \mathrm{length}\big(i_\alpha(s'; s, E^*(s,c)\big)$ and $f_2(s,c,s',c') = \phi(s)$. Using results in [19], it can be shown that under Assumption PDE there exist finite constants l and λ such that, with probability 1. $\lim_{n\to\infty}(1/n)\sum_{j=0}^{n-1} f_1(Z_j) = \lambda$ and $\lim_{n\to\infty}(1/n)\sum_{j=0}^{n-1} f_2(Z_j) = l$. Here λ is the long-run average rate at which passage times start and l is the long-run average number of passage times underway. Using Assumption T and a version of Little's Law due to Glynn and Whitt [14], an argument almost identical to the proof of Equation (3.3) in [25] then shows that $\lim_{n\to\infty}(1/n)\sum_{j=0}^{n-1} P_j = l/\lambda$ a.s.. Thus the limiting average delay can be expressed as a ratio of two time-average limits computed from the underlying chain. Under Assumption PDE the "jackknifed batch means" method given in [19] for obtaining point estimates and confidence intervals for such ratios is applicable; there is no need to measure individual delays. It follows from the results in [19] that this method actually is applicable when the moment condition in Assumption PDE is weakened to require only that each clock-setting distribution have finite third moment.

It can be shown that under Assumption PDE and Assumption T the estimation methods described in this paper extend to functions f that depend both upon the delay and upon the states of the marking process during the corresponding passage-time interval. For example, our methods extend to functions of time-average limits of the form $\tilde{\pi}(f) = \lim_{n\to\infty}(1/n)\sum_{j=0}^{n-1} f(P_j, S_{\alpha(j)})$ where, as before, $\zeta_{\alpha(j)}$ is the start of jth passage-time interval. As an application, consider the closed network of queues in Example 2.7 and suppose we wish to estimate the limiting average delay only for those passage times that start with an arrival to an empty queue. This limit is of the form $\tilde{\pi}(f_1)/\tilde{\pi}(f_2)$, where $f_1(p,s) = p1_A(s)$, $f_2(p,s) = 1_A(s)$, and $A = \{\,(s_{1,1},\ldots,s_{2,3}) \in G\colon s_{1,3} = 1 \text{ and } s_{1,1} = s_{1,2} = 0\,\}$. Many other limits of this general type (and functions of such limits) can be handled within our framework.

6. Conclusion

For discrete-event systems that can be modeled as SPN's with generally-distributed firing times, we have shown that STS methods can be used to obtain strongly consistent point estimates and asymptotic confidence intervals for a wide variety of delay characteristics, provided that Assumption PDE and Assumption T hold.

Our prior results in [18, 19] yield estimation methods for time-average limits of the form $\lim_{t\to\infty}(1/t)\int_0^t f\big(X(u)\big)\,du$, where f is a real-valued function and $\{\,X(t)\colon t\geq 0\,\}$ is the marking process (assumed non-regenerative) of an SPN. Such limits correspond to performance measures such as long-run utilization and availability, but not to delay characteristics. Our new results can be viewed as extending the estimation methods in [18, 19] to time-average limits of delays. Our results also extend the results in [24] on estimation of delays to the setting of non-regenerative SPN's.

To apply our results, it is necessary to verify that Assumption PDE and Assumption T hold. To this end, it typically is straightforward to verify the positive-density and moment conditions on the clock-setting distributions and the positivity requirement on the speeds, since the clock-setting distributions and speeds are specified by the modeler. Moreover, it is usually not hard to determine whether there exists a marking and a subsequent firing sequence that would cause all ongoing or newly started passage times to terminate. It then remains to determine whether the marking set is finite and whether the SPN is irreducible. (These properties also need to be verified when computing steady state performance measures analytically or numerically for more tractable SPN's such as nets with exponential clock-setting distributions.) When the marking set G is specified explicitly (see, for example, [21, Example 2.8]), determining whether $|G| < \infty$ is trivial. In practice, however, G is often defined implicitly as the set of markings reachable (in the sense of the relation $\rightsquigarrow$) from some specified set of initial markings; in this case it can be nontrivial to determine whether G is finite. Under various restrictions on the form of the new-marking probabilities and speeds, finiteness and irreducibility can be checked (at least in principle) by in effect constructing "coverability graphs" using an algorithm similar to that given in Peterson [33, Section 4.2.1]. This approach is applicable, for example, when all speeds are positive, no inhibitor arcs are present, and each transition is deterministic. In general, however, the problem of determining finiteness and irreducibility can be difficult or impossible: the marking set can be so large that the computational costs of the coverability analysis are prohibitive or there may exist no algorithm that is guaranteed to terminate. The problem of determining whether $|G| < \infty$, for example, is undecidable over the class of all SPN's. On the other hand, there are many SPN models of practical interest for which finiteness and irreducibility can be verified based on the analyst's understanding of the system under study. In such models the difficulty in determining time-average limits arises not so much from the size or complexity of the marking set or SPN graph, but from the fact that the clock-setting distributions are non-exponential. In any case, prior to the current work the only available sufficient conditions for applicability of STS methods were stated in terms of the output process $\{ f(P_j) \colon j \geq 0 \}$ and were virtually impossible to verify directly. The results in this paper reduce the problem of verifying the applicability of STS methods to standard analysis problems for SPN's.

References

[1] Asmussen, S. (1987). *Applied Probability and Queues*. Wiley, New York.

[2] Baccelli, F., Cohen, G., Olsder, G. J., and Quadrat, J.-P. (1993). *Synchronization and Linearity: An Algebra for Discrete Event Systems*. Wiley, Chichester, England.

[3] Baccelli, F. and Schmidt, V. (1995). Taylor expansions for Poisson-driven $(\max, +)$-linear systems. INRIA Research Report 2494, Centre de Diffusion, INRIA, Le Chesnay Cedex, France.

[4] Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

[5] Brillinger, D. R. (1973). Estimation of the mean of a stationary time series by sampling. *J. Appl. Probab.* **10**, 419–431.

[6] Campos, J., Colom, J. M., Chiola, G., and Silva, M. (1989). Tight polynomial bounds for steady-state performance of marked graphs. In *Proc. Third Internat. Workshop Petri Nets Performance Models*, 200–209. IEEE Computer Society Press.

[7] Chiola, G., Bruno, G., and Demaria, T. (1988). Introducing a color formalism into generalized stochastic Petri nets. In *Proc. 9th European Workshop Appl. Theory Petri Nets*.

[8] Conway, R. W. (1963). Some tactical problems in digital simulation. *Management Sci.* **10**, 47–61.

[9] Durrett, R. (1991). *Probability: Theory and Examples*. Wadsworth and Brooks/Cole, Pacific Grove, CA.

[10] Ethier, S. N. and Kurtz, T. G. (1986). *Markov Processes: Characterization and Convergence*. Wiley, New York.

[11] Fox, B. L., Goldsman, D., and Swain, J. J. (1991). Spaced batch means. *Oper. Res. Lett.* **10**, 255–263.

[12] Glynn, P. W. (1982). Simulation output analysis for general state space Markov chains. Ph.D. Dissertation, Department of Operations Research, Stanford University, Stanford, CA.

[13] Glynn, P. W. and Iglehart, D. L. (1990). Simulation output analysis using standardized time series. *Math. Oper. Res.* **15**, 1–16.

[14] Glynn, P. W. and Whitt, W. (1986). A central-limit-theorem version of $L = \lambda W$. *Queueing Systems Theory Appl.* **2**, 191–215.

[15] Goldsman, D., Meketon, M. S., and Shruben, L. W. (1990). Properties of standardized time series weighted area variance estimators. *Management Sci.* **36**, 602–612.

[16] Gut, A. (1988). *Stopped Random Walks: Limit Theorems and Applications*. Springer-Verlag, New York.

[17] Haas, P. J. (1985). Recurrence and regeneration in non-Markovian simulations. Ph.D. Dissertation, Department of Operations Research, Stanford University, Stanford, CA.

[18] Haas, P. J. (1997). Estimation methods for stochastic Petri nets based on standardized time series. In *Proc. Seventh Internat. Workshop Petri Nets Performance Models*, 194–204. IEEE Computer Society Press.

[19] Haas, P. J. (1999). Estimation methods for non-regenerative stochastic Petri nets. *IEEE Trans. Software Engrg.* to appear.

[20] Haas, P. J. (1999). On simulation output analysis for generalized semi-Markov processes. *Comm. Statist. Stochastic Models* to appear.

[21] Haas, P. J. and Shedler, G. S. (1989). Stochastic Petri nets with timed and immediate transitions. *Comm. Statist. Stochastic Models* **5**, 563–600.

[22] Haas, P. J. and Shedler, G. S. (1991). Stochastic Petri nets: Modeling power and limit theorems. *Probab. Engrg. Inform. Sci.* **5**, 477–498.

[23] Haas, P. J. and Shedler, G. S. (1993). Passage times in colored stochastic Petri nets. *Comm. Statist. Stochastic Models* **9**, 31–79.

[24] Haas, P. J. and Shedler, G. S. (1995). One-dependent cycles and passage times in stochastic Petri nets. In *Proc. Sixth Internat. Workshop Petri Nets Performance Models*, 191–202. IEEE Computer Society Press.

[25] Haas, P. J. and Shedler, G. S. (1996). Estimation methods for passage times using one-dependent cycles. *Discrete Event Dynam. Systems Theory Appl.* **6**, 43–72.

[26] Loéve, M. (1977). *Probability Theory*. Springer-Verlag, New York, 4th edn.

[27] Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer-Verlag, London.

[28] Molloy, M. K. (1982). Performance analysis using stochastic Petri nets. *IEEE Trans. Comput.* **C-31**, 913–917.

[29] Muñoz, D. F. and Glynn, P. W. (1997). A batch means methodology for estimation of a nonlinear function of a steady-state mean. *Management Sci.* **43**, 1121–1135.

[30] Muñoz, D. F. and Glynn, P. W. (1997). Multivariate standardized time series for output analysis in simulation experiments. *Oper. Res.* To appear.

[31] Muppala, J. K., Trivedi, K. S., Mainkar, V., and Kulkarni, V. G. (1994). Numerical computation of response time distributions using stochastic reward nets. *Ann. Oper. Res.* **48**, 155–184.

[32] Natkin, S. (1985). Les réseaux de Petri stochastiques. *Technique et Science Informatiques* **4**, 143–160.

[33] Peterson, J. L. (1981). *Petri Net Theory and the Modeling of Systems*. Prentice Hall, Englewood Cliffs, New Jersey.

[34] Prisgrove, L. A. and Shedler, G. S. (1986). Symmetric stochastic Petri nets. *IBM J. Res. Develop.* **30**, 278–293.

[35] Roberts, G. O. and Tweedie, R. L. (1999). Bounds on regeneration times and convergence rates for Markov chains. *Stochastic Process. Appl.* to appear.

[36] Schmeiser, B. (1982). Batch size effects in the analysis of simulation output. *Oper. Res.* **30**, 556–568.

[37] Shedler, G. S. (1993). *Regenerative Stochastic Simulation*. Academic Press, New York.

[38] Shruben, L. W. (1983). Confidence interval estimation using standardized time series. *Oper. Res.* **31**, 1090–1108.

[39] Sigman, K. (1990). One-dependent regenerative processes and queues in continuous time. *Math. Oper. Res.* **15**, 175–189.

[40] Zenie, A. (1985). Colored stochastic Petri nets. In *Proc. Internat. Workshop Timed Petri Nets*, 262–271. IEEE Computer Society Press.

A. Proof of Theorem 3.8

By results in [19] (specifically, Theorem 4.5, Proposition 3.9, and Proposition 3.13), there exists a sequence $\{\, \theta^+(k) \colon k \geq 0 \,\}$ of od-regeneration points for the embedded chain $\{\, (S_n^+, C_n^+) \colon n \geq 0 \,\}$. We can assume without loss of generality that for some $r \geq 1$ the r-step transition kernel of the embedded chain satisfies a decomposition as in (3.2), and that the sample paths of the embedded chain and the od-regeneration points are generated simultaneously as described in Section 3. Moreover, we can assume that each sample path of the underlying chain $\{\, (S_n, C_n) \colon n \geq 0 \,\}$ is generated by first generating a sample path of the embedded chain and then filling in the intermediate states. That is, for $n \geq$

0 we generate zero or more intermediate states $(S_{\nu(n)+1}, C_{\nu(n)+1}), (S_{\nu(n)+2}, C_{\nu(n)+2}), \ldots,$ $(S_{\nu(n+1)-1}, C_{\nu(n+1)-1}) \in \Sigma^+$ according to the conditional joint distribution for the underlying chain, where ν is defined as in (2.5) and we have conditioned on the endpoint values $(S_{\nu(n)}, C_{\nu(n)})$ and $(S_{\nu(n+1)}, C_{\nu(n+1)})$. Under these assumptions, it is not hard to see that the desired sequence of od-regeneration points for the underlying chain can be defined by setting $\theta(k) = \nu\bigl(\theta^+(k)\bigr)$ for $k \geq 0$.

To prove the second assertion of the theorem, fix $r > 0$ and a polynomially dominated function f, assuming without loss of generality that f is nonnegative. Write $Y_1(f) = Y_1^+(f) + Y_1'(f)$, where

$$Y_1^+(f) = \sum_{n=\theta(0)}^{\theta(1)-1} f(S_n, C_n) 1_S(S_n)$$

and

$$Y_1'(f) = \sum_{n=\theta(0)}^{\theta(1)-1} f(S_n, C_n) 1_{S'}(S_n).$$

Since

$$E_\mu[Y_1^r(f)] \leq 2^{r-1} E_\mu[\bigl(Y_1^+(f)\bigr)^r] + 2^{r-1} E_\mu[\bigl(Y_1'(f)\bigr)^r]$$

by the c_r-inequality (see Loéve [26]), it suffices to show that $Y_1^+(f)$ and $Y_1'(f)$ each has finite moments of all orders. To this end, fix $q > 0$ such that $\int_0^\infty e^{qx}\, dF(x; s', e', s, E^*) < \infty$ for all s', s, e', and E^*; such a value q exists by the finiteness of the marking set G and Assumption PDE. Set

$$g_q(s, c) = \exp\Bigl(q \max_{1 \leq i \leq M} c_i\Bigr)$$

for $s \in S$ and $c = (c_1, c_2, \ldots, c_M) \in C(s)$. By an argument almost identical the proof of [19, Theorem 4.5], it can be shown that there exists an integer $m \geq 1$ such that, for all sufficiently large b,

(i) $\sup_{(s,c) \in H_b} E_{(s,c)}\left[g_q(S_m^+, C_m^+) - g_q(S_0^+, C_0^+)\right] < \infty$; and

(ii) there exists $\beta > 0$ such that

$$E_{(s,c)}\left[g_q(S_m^+, C_m^+) - g_q(S_0^+, C_0^+)\right] \leq -\beta g_q(s, c) \tag{A.1}$$

for $(s, c) \in \Sigma^+ - H_b$,

where $H_b = \{\, (s, c) \in \Sigma^+ \colon \max_{1 \leq i \leq M} c_i \leq b \,\}$. That is, the embedded chain satisfies a "geometric" Foster-type drift criterion for stability; see [27, Chapter 15]. Moreover, it follows immediately from [19, Theorem 4.5] that the embedded chain is "ϕ-irreducible" in the sense of [27]. Finally, a simple argument using (2.3) shows that $E_\mu\left[g_q(S_0^+, C_0^+)\right] < \infty$. It then follows from Theorem 2.3 in Roberts and Tweedie [35] that $\delta_1^+ = \theta^+(1) - \theta^+(0)$ has finite moments of all orders.

Remark. Strictly speaking, the result in [35] requires that the chain be "strongly aperiodic" and that the drift criterion hold for $m = 1$. As mentioned in [35], however, the strong aperiodicity assumption can be dropped. Moreover, a drift criterion with $m > 1$ can be handled by applying [35, Theorem 2.3] to the chain $\{ (S^+_{mn}, C^+_{mn})\colon n \geq 0 \}$.

Since the function f^{2r} is polynomially dominated, Propositions 3.9 and 3.13 in [19] together imply that $Y_1^+(f^{2r})$ has finite mean. Using the Cauchy-Schwarz inequality, we have

$$\begin{aligned} E_\mu\big[\big(Y_1^+(f)\big)^r\big] = E_\mu\left[\left(\sum_{n=\theta^+(0)}^{\theta^+(1)-1} f(S_n^+, C_n^+)\right)^r\right] &\leq E_\mu\Big[(\delta_1^+)^r \max_{\theta^+(0)\leq n\leq \theta^+(1)-1} f^r(S_n^+, C_n^+)\Big] \\ &\leq E_\mu^{1/2}[(\delta_1^+)^{2r}]E_\mu^{1/2}[Y_1^+(f^{2r})] \\ &< \infty. \end{aligned}$$

To complete the proof, write $Y_1'(f) = \sum_{k=0}^{\delta_1^+-1} M_k$, where

$$M_k = \sum_{n=\nu(\theta^+(0)+k)+1}^{\nu(\theta^+(0)+k+1)-1} f(S_n, C_n).$$

Since f is polynomially dominated, we have $Y_1'(f) \leq \psi \sum_{k=0}^{\delta_1^+-1} N_k$, where

$$\psi = \sup_{(s,c)\in\Sigma-\Sigma^+} f(s,c) < \infty$$

and $N_k = \nu(\theta^+(0) + k + 1) - \nu(\theta^+(0) + k) - 1$. Since δ_1^+ has finite moments of all orders as shown above and the SPN is irreducible with finite marking set, a straightforward extension of the geometric-trials argument in [19, Appendix B] shows that $\sum_{k=0}^{\delta_1^+-1} N_k$ has finite moments of all orders.

B. Proof of Theorem 4.3

To establish the first assertion of the theorem, observe that under Assumption T there exists a sequence $x = (s^{(0)}, E^{(0)}, s^{(1)}, E^{(1)}, \ldots) \in \mathcal{X}$ such that $\iota(x) < \infty$. Set $Z_n = (S_n, C_n, \ldots, S_{n+l}, C_{n+l})$ for $n \geq 0$, where $l = \iota(x)$. Set

$$A = \left\{ (s_0, c_0, \ldots, s_l, c_l) \in \Sigma^l\colon s_k = s^{(k)} \text{ and } E^*(s_k, c_k) = E^{(k)} \text{ for } 1 \leq k \leq l \right\},$$

where $(s^{(0)}, E^{(0)}, s^{(1)}, E^{(1)}, \ldots)$ is the distinguished element of $\mathcal{X}$ given above and E^* is defined as in (2.2). Observe that whenever $Z_n \in A$ none of the starts in the start vector V_n are present in the start vector V_{n+l}. Under Assumption PDE the embedded chain $\{ (S_n^+, C_n^+)\colon n \geq 0 \}$ is Harris recurrent (see [19]), and it follows by a geometric-trials argument that $P_\mu \{ Z_n \in A \text{ i.o.} \} = 1$. Thus each passage time terminates with probability 1. Since each new clock reading is a.s. finite, each ζ_n and hence each delay is a.s. finite.

We now prove the second assertion of the theorem. Set $\tilde{\delta}_k = \tilde{\gamma}(k) - \tilde{\gamma}(k-1)$ for $k \geq 1$. We claim that the post-$\tilde{\gamma}(k)$ process $\{ (P_{\tilde{\gamma}(k)+n}, \tilde{\delta}_{k+n+1})\colon n \geq 0 \}$

(i) is distributed as the post-$\tilde{\gamma}(0)$ process $\{\,(P_{\tilde{\gamma}(0)+n},\tilde{\delta}_{n+1})\colon n\geq 0\,\}$ for $k\geq 1$; and

(ii) is independent of the pre-$\tilde{\gamma}(k-1)$ process $\{\,P_0,P_1,\ldots,P_{\tilde{\gamma}(k-1)-1};\tilde{\delta}_1,\tilde{\delta}_2,\ldots,\tilde{\delta}_{k-1}\,\}$ for $k\geq 2$,

so that $\{\,\tilde{\gamma}(k)\colon k\geq 0\,\}$ is indeed a sequence of od-regeneration points. To see that the above assertions hold, suppose for ease of exposition that at least one passage time starts during each interval $\bigl[\zeta_{\tilde{\theta}(k-1)},\zeta_{\tilde{\theta}(k)}\bigr)$, as in Figure 3; the argument for the general case is almost identical. Then, for $k\geq 0$, $\tilde{\gamma}(k)$ is simply the index of the first passage time that starts (strictly) after time $\zeta_{\tilde{\theta}(2k)}$.

First recall that, under the conditions of the theorem, there exists $\mathcal{C}\subseteq\Sigma^+$, $r\geq 1$, $b\in(0,1)$, λ, and Q such that a decomposition as in (3.2) holds for the embedded chain $\{\,(S_n^+,C_n^+)\colon n\geq 0\,\}$. By construction, $(S_{\tilde{\theta}(k)},C_{\tilde{\theta}(k)})$ is distributed according to λ for $k\geq 0$. Because each random variable $N(k)$ defined in the construction of the points $\{\,\tilde{\theta}(k)\colon k\geq 0\,\}$ is a stopping time (see below) with respect to the underlying chain, it follows that for $k\geq 0$ the state $(S_{\tilde{\theta}(k+1)},C_{\tilde{\theta}(k+1)})$ is independent of $\{\,(S_n,C_n)\colon n<L_{N(k)}\,\}$, where L_n is defined as in Section 4. Since each $\tilde{\theta}(k)$ also is a stopping time with respect to the underlying chain and is strictly less than $N(k)$, it follows from the strong Markov property that the random indices $\{\,\tilde{\theta}(k)\colon k\geq 0\,\}$ form a sequence of od-regeneration points for $\{\,(S_n,C_n)\colon n\geq 0\,\}$.

Next observe that for $k\geq 1$ the post-$\tilde{\gamma}(k)$ process $\{\,(P_{\tilde{\gamma}(k)+n},\tilde{\delta}_{k+n+1})\colon n\geq 0\,\}$ is determined by $\{\,(S_n,C_n)\colon n\geq\tilde{\theta}(2k)\,\}$ according to a mechanism that does not depend explicitly on k. Hence, by the od-regenerative property of the points $\{\,\tilde{\theta}(k)\colon k\geq 0\,\}$, the post-$\tilde{\gamma}(k)$ process is distributed as the post-$\tilde{\gamma}(0)$ process. Thus the assertion in (i) holds. To see that the assertion in (ii) holds, fix $k\geq 2$ and observe that, by construction, the terminations $D_0,D_1,\ldots,D_{\tilde{\gamma}(k-1)-1}$ are all less than $\zeta_{\tilde{\theta}(2k-2)}$. Thus, the collection of random variables $\{\,P_0,P_1,\ldots,P_{\tilde{\gamma}(k-1)-1};\tilde{\delta}_1,\tilde{\delta}_2,\ldots,\tilde{\delta}_{k-1}\,\}$ is determined completely by $\{\,(S_n,C_n)\colon 0\leq n<\tilde{\theta}(2k-1)\,\}$. By the od-regenerative property of $\{\,\tilde{\theta}(k)\colon k\geq 0\,\}$, the post-$\tilde{\gamma}(k)$ process is independent of $\{\,P_0,\ldots,P_{\tilde{\gamma}(k-1)-1};\tilde{\delta}_1,\ldots,\tilde{\delta}_{k-1}\,\}$; in Figure 3, for example, the delays $\{\,P_j\colon j\geq 8\,\}$ are independent of $\{\,P_j\colon 0\leq j\leq 5\,\}$. Thus, the second assertion of the theorem follows.

To prove the final assertion of the theorem, we need the following preliminary result. Consider a sequence $\{\,\mathcal{F}_n\colon n\geq 1\,\}$, where each $\mathcal{F}_n$ is a collection of random variables and $\mathcal{F}_n\subseteq\mathcal{F}_{n+1}$ for $n\geq 1$. A positive integer-valued random variable N is a *stopping time* with respect to $\{\,\mathcal{F}_n\colon n\geq 1\,\}$ if for each $k\geq 1$ the occurrence or non-occurrence of the event $\{\,N=k\,\}$ is determined by the values of the random variables in $\mathcal{F}_k$. If each $\mathcal{F}_n$ has the special form $\mathcal{F}_n=\{\,X_1,X_2,\ldots,X_n\,\}$, where $\{\,X_n\colon n\geq 1\,\}$ is a specified sequence of random variables, then N is said to be a stopping time with respect to $\{\,X_n\colon n\geq 1\,\}$.

Lemma B.1. *Let $S_N=\sum_{n=1}^{N}X_n$, where $\{\,X_n\colon n\geq 1\,\}$ is a sequence of o.i.d. random variables and N is a stopping time with respect to a nondecreasing sequence $\{\,\mathcal{F}_n\colon n\geq 0\,\}$ such that X_n is determined by $\mathcal{F}_n$ for $n\geq 1$ and independent of $\mathcal{F}_{n-2}$ for $n\geq 2$. Then for $r\geq 0$ there exists a constant b_r (depending only on r) such that*

$$E[|S_N|^r]\leq b_r E\left[|X_1|^r\right]E\left[(N+2)^r\right].$$

Lemma B.1 can be viewed as extending Theorem I.5.2 in Gut [16] from i.i.d. random variables to o.i.d. random variables.

Proof. Observe that $|S_N| \le \sum_{i=1}^{L} Y_i + \sum_{i=1}^{M} Z_i$, where $L = \lfloor N/2 \rfloor + 1$, $M = \lceil N/2 \rceil + 1$, $Y_i = |X_{2i}|$, and $Z_i = |X_{2i-1}|$. (Here $\lceil x \rceil$ denotes the smallest integer greater than or equal to x.) Using the c_r-inequality [26], it follows that

$$E\left[S_N^r\right] \le 2^{r-1} E\left[\left(\sum_{i=1}^{L} Y_i\right)^r\right] + 2^{r-1} E\left[\left(\sum_{i=1}^{M} Z_i\right)^r\right]. \tag{B.2}$$

Observe that $\{ Y_i : i \ge 1 \}$ is a sequence of i.i.d. random variables with $E\left[Y_i^r\right] = E\left[|X_1|^r\right] < \infty$ for $i \ge 1$. Also observe that, for $n \ge 1$, the value of Y_n is determined by the values of the random variables in $\mathcal{F}'_n = \mathcal{F}_{2n}$. Moreover, Y_{n+1} is independent of $\mathcal{F}'_n$. Finally, since $L = k$ ($k \ge 1$) if and only if either $N = 2k-2$ or $N = 2k-1$, it follows that L is a stopping time with respect to $\{ \mathcal{F}'_n : n \ge 1 \}$. Theorem I.5.2 in [16] then implies that there exists a constant b'_r (depending only on r) such that $E\left[\left(\sum_{i=1}^{L} Y_i\right)^r\right] \le b'_r E\left[|X_1|^r\right] E\left[L^r\right]$. An analogous argument shows that $E\left[\left(\sum_{i=1}^{M} Z_i\right)^r\right] \le b'_r E\left[|X_1|^r\right] E\left[M^r\right]$, and the desired result follows from (B.2) upon observing that $E\left[L^r + M^r\right] \le 2E\left[(L+M)^r\right] = 2E\left[(N+2)^r\right]$. □

Returning to the main proof, let f be polynomially dominated, so that $|f(x)| \le a(x^b + 1)$ for some constants $a, b > 0$. Observe that

$$\tilde{Y}_1(|f|) \le \sum_{i=1}^{3} \psi \eta_i a\big((\tau_i + \tau_{i+1})^b + 1\big)$$

where $\eta_i = \tilde{\theta}(i) - \tilde{\theta}(i-1)$, $\tau_i = \zeta_{\tilde{\theta}(i)} - \zeta_{\tilde{\theta}(i-1)}$, and $\psi = \max_{s', s, E^*} \operatorname{length}\big(i_\alpha(s'; s, E^*)\big)$. To prove the second assertion of the theorem, it therefore suffices to show that τ_1 and η_1 each have finite moments of all orders. Define a sequence of od-regeneration points $\{ \theta(k) : k \ge 0 \}$ for the underlying chain by setting $\theta(0) = \tilde{\theta}(0)$ and then defining $\theta(k)$ for $k \ge 1$ as in Section 3, that is, using the decomposition as in (3.2) but without regard for the state of the start vector. Formally, $\theta(k+1) = L'_{\theta(k)} + M(L'_{\theta(k)}; r)$ for $k \ge 0$, where $M(n; r)$ is defined as in Section 4 and $L'_n = \inf\{ l \ge n : I_l = 1 \}$ for $n \ge 0$. Denote by N' the (random) number of complete $\theta(k)$-cycles contained in the interval $[\tilde{\theta}(0), \tilde{\theta}(1)]$ and set $N = N' + 1$. Recall the definition of t^* from (2.1) and set $h(s, c) \equiv 1$ for $(s, c) \in \Sigma$. Observe that $\tau_1 \le \sum_{k=1}^{N} Y_k$ and $\eta_1 \le \sum_{k=1}^{N} Y'_k$, where

$$Y_k = \sum_{n=\theta(k-1)}^{\theta(k)-1} t^*(S_n, C_n)$$

and

$$Y'_k = \sum_{n=\theta(k-1)}^{\theta(k)-1} h(S_n, C_n).$$

Also observe that both t^* and h are polynomially dominated in the sense of Section 3. Thus each Y_k and Y'_k has finite moments of all orders by Theorem 3.8 and, by construction, $\{ Y_k \colon k \geq 1 \}$ and $\{ Y'_k \colon k \geq 1 \}$ are each sequences of o.i.d. random variables. By Lemma B.1, it suffices to show that N has finite moments of all orders. To this end, set

$$\Lambda_k = \max_{\{ j \colon \alpha(j) \leq \theta(k) < \beta(j) \}} \bigl(\beta(j) - \theta(k) \bigr)$$

for $k \geq 0$. If $\{ j \colon \alpha(j) \leq \theta(k) < \beta(j) \} = \varnothing$, set $\Lambda_k = 0$. It is a consequence of the a.s. finiteness of the delays and the od-regenerative structure of the underlying chain that there exists $m \geq 1$ such that

$$P_\mu \{ \Lambda_k < \theta(k+m) - \theta(k) \} > 0$$

for $k \geq 0$. For $l \geq 0$, set $I_l = 1$ if $\Lambda_{lm} \geq \theta\bigl((l+1)m\bigr) - \theta(lm)$; otherwise, $I_l = 0$. It follows from the od-regenerative structure of the underlying chain that $\{ I_l \colon l \geq 0 \}$ is a sequence of o.i.d. random variables. Since each delay is a.s. finite, either $P_\mu \{ I_0 = 1 \} = 0$, in which case the desired result follows immediately, or else $P_\mu \{ I_0 = 1 \} > 0$ and

$$p \stackrel{\text{def}}{=} P_\mu \{ I_1 = 1 | I_0 = 1 \} < 1.$$

In the latter case,

$$\begin{aligned}
&P_\mu \{ N' > lm \} \\
&\qquad \leq P_\mu \{ I_0 = 1, I_1 = 1, \ldots, I_{l-1} = 1 \} \\
&\qquad = P_\mu \{ I_0 = 1 \} P_\mu \{ I_1 = 1 | I_0 = 1 \} P_\mu \{ I_2 = 1 | I_1 = 1 \} \cdots P_\mu \{ I_{l-1} = 1 | I_{l-2} = 1 \} \\
&\qquad = P_\mu \{ I_0 = 1 \} p^{l-1}
\end{aligned}$$

for $l \geq 1$, so that the distribution of N'/m has geometrically decreasing tail probabilities and hence has finite moments of all orders, so that N has finite moments of all orders.