On Transience and Recurrence in Irreducible Finite-State Stochastic Systems

PETER W. GLYNN, Stanford University
PETER J. HAAS, IBM Research

Long-run stochastic stability is a precondition for applying steady-state simulation output analysis methods to a discrete-event stochastic system, and is of interest in its own right. We focus on systems whose underlying stochastic process can be represented as a Generalized Semi-Markov Process (GSMP); a wide variety of stochastic systems fall within this framework. A fundamental stability requirement for an irreducible GSMP is that the states be "recurrent" in that the GSMP visits each state infinitely often with probability 1. We study recurrence properties of irreducible GSMPs with finite state space. Our focus is on the "clocks" that govern the occurrence of events, and we consider GSMPs in which zero, one, or at least two simultaneously active events can have clock-setting distributions that are "heavy tailed" in the sense that they have infinite mean. We establish positive recurrence, null recurrence, and, perhaps surprisingly, possible transience of states for these respective regimes. The transience result stands in strong contrast to Markovian or semi-Markovian GSMPs, where irreducibility and finiteness of the state space guarantee positive recurrence.

25

Categories and Subject Descriptors: G.3 [**Probability and Statistics**]: Stochastic Processes; I.6.6 [**Simulation and Modeling**]: Simulation Output Analysis

General Terms: Theory, Algorithms

Additional Key Words and Phrases: Discrete-event stochastic systems, generalized semi-Markov processes, hazard rate, heavy tails, recurrence, regular variation, stability

ACM Reference Format:
Peter W. Glynn and Peter J. Haas. 2015. On transience and recurrence in irreducible finite-state stochastic systems. ACM Trans. Model. Comput. Simul. 25, 4, Article 25 (May 2015), 19 pages.
DOI: http://dx.doi.org/10.1145/2699721

1. INTRODUCTION

We are pleased to honor the achievements of Donald L. Iglehart, who has played a major role in creating a modern theory of stochastic simulation. An important development of this theory is the recognition that the Generalized Semi-Markov Process (GSMP)—originally proposed by König et al. [1974] for the study of complex service systems—is an exceptionally well suited mathematical model of the underlying stochastic process of a complex discrete-event simulation, and therefore worthy of serious study by simulation researchers. (The underlying process records the state of the system as it evolves over continuous time.) Work by Professor Iglehart, his students, and his colleagues has established the GSMP, which formalizes the usual variable-time-advance simulation procedure, as the standard mathematical framework for investigating fundamental questions about discrete-event simulation [Glynn 1989; Haas and Shedler

Authors' addresses: P. W. Glynn, Department of Management Science and Engineering, Stanford University, Stanford, CA 94305; email: glynn@stanford.edu; P. J. Haas, IBM Almaden Research Center, 650 Harry Road, San Jose, CA 95120-6099; email: phaas@us.ibm.com.

DOI: http://dx.doi.org/10.1145/2699721

1987; Henderson and Glynn 2001; Iglehart and Shedler 1983, 1984; Shedler 1993; Whitt 1980]. Although there are classes of discrete-event simulations that cannot be modeled as GSMPs [Haas and Shedler 1987; Miyazawa 1993], a rich collection of complex manufacturing, computer, transportation, telecommunication, health care, logistic, and workflow simulations can be captured within the GSMP framework using a remarkably simple set of building blocks.

In the context of steady-state output analysis for discrete-event simulations, a fundamental question is whether the underlying stochastic process is stochastically stable over time. Such long-run stochastic stability ensures that questions about the "steady state" of a given system are well posed, and is a precondition for applying steady-state simulation output analysis techniques such as regenerative, spectral, or standardized time series methods. Steady-state stability is important in its own right, since lack of such stability can indicate problematic behavior in the system of interest or deficiencies in the modeling process. In this article, we study long-run stability for systems in which the underlying stochastic process can be represented as a GSMP.

A basic requirement for steady-state output analysis is that the states of the GSMP be "recurrent," in that the GSMP visits each state infinitely often (i.o.) with probability 1. For example, the regenerative method for simulation output analysis [Crane and Iglehart 1975; Glynn and Iglehart 1993; Shedler 1993] requires that each regeneration point be almost surely (a.s.) finite, which often amounts to the requirement that a specified "single state" is recurrent. Other methods, such as the method of standardized time series, assume that the GSMP obeys a Functional Central Limit Theorem (FCLT) (see Glynn and Iglehart [1990]). Such an FCLT is implied by a "Harris recurrence" condition (Section 2.5) that in turn implies recurrence for each state [Glynn and Haas 2006]. In this article, we study recurrence properties of GSMPs. We focus on GSMPs having a finite state space and in which any state can potentially be reached from any other state via a sequence of event occurrences; GSMPs having this latter property are called "irreducible" (see Definition 2.2 in the following).

Lack of recurrence for a GSMP with a countably infinite state space often means that the process drifts off to the far reaches of the space, never to return. Such outward drift can correspond, for example, to overflowing queues or buffers in telecommunication, service, and manufacturing models. In GSMPs with finite or infinite state spaces, transience of certain states can lead to effective reducibility in a nominally irreducible system. In this latter scenario, any state can be reached from any other state in principle, but if such reachability hinges on passage through transient states, then certain states will eventually become unreachable from each other with probability 1. Consequently, the steady-state or limiting distribution of the GSMP, if it exists, may depend upon the initial state of the system in a potentially complex manner, making steady-state estimation difficult. Transience of certain states can also correspond to "starvation" (withholding of needed resources) or overprovisioning of some process or activity within a complex system, as discussed later in this section.

A GSMP makes a state transition when one or more events associated with the state occur. For each such "active" event, a clock records the time until the event is scheduled to occur. These clocks determine when the next state transition occurs and which of the scheduled events actually trigger this state transition. When all events are "simple" (Section 2.2) and each clock is set according to an exponential distribution, the GSMP reduces to a continuous-time Markov chain [Haas 2002, Section 3.4] and classical theory applies (see, e.g., Asmussen [2003]). In particular, if the GSMP is irreducible with finite state space, then recurrence of each state is assured. Indeed, each state is "positive recurrent"—in that the expected number of state transitions between visits is finite—and the expected hitting times of each state (in continuous time) are finite. Similarly, if each state has at most one active event, then the GSMP is a semi-Markov process, and again the finiteness of the state space and the irreducibility property

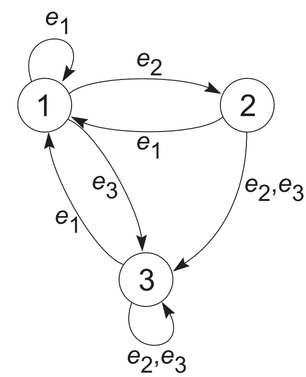

Fig. 1. State-transition diagram for GSMP with a transient state.

together guarantee that each state is positive recurrent. If, moreover, each clock-setting distribution has finite mean, then the expected hitting times of each state are finite.

When the clock-setting distribution functions can be arbitrary and multiple events can be active in a state, the behavior of the GSMP becomes much more complex and sometimes counterintuitive. For example, it is possible to construct a GSMP having an infinite state space in which the expected time between successive state transitions increases linearly but, with probability 1, an infinite number of state transitions occur in a finite time interval [Haas 2002, p. 90]. In what follows, we show that even when the state space is constrained to be finite, nonintuitive behavior can occur in a non-Markovian GSMP when the means of the clock-setting distributions can be infinite. Specifically, we study GSMPs in which zero, one, or at least two simultaneously active events can have such "heavy-tailed" clock-setting distributions. We establish discrete-time positive recurrence, null recurrence, and, perhaps surprisingly, possible transience of states for these respective regimes.

The transience result is established by means of a simple example, specifically, a GSMP having three states, three events, and a state-transition diagram as in Figure 1. Events e_1 and e_2 have heavy-tailed clock-setting distributions, whereas e_3 has a light-tailed distribution and occurs at frequent intervals. Observe that the only way in which the GSMP can hit state 2 is if event e_1 occurs and then event e_2 occurs without an intervening occurrence of frequent event e_3. We show (Theorem 5.1) that, with probability 1, such a scenario occurs only a finite number of times, because the clocks for e_1 and e_2 are almost never small at the same time and hence these events almost never occur close together in time.

This GSMP may be viewed as a simplified model of a starvation scenario in which, for example, resources are used by two processes, each of which alternates between busy periods with heavy-tailed durations and short idle periods. The resources are available for use by a third process—which periodically submits a resource request—only when both heavy-tailed processes are idle. For example, the resource might represent a tool in a manufacturing cell. Theorem 5.1 essentially implies that, with probability 1, the third process will only be granted the resource a finite number of times because there will only be a finite number of times at which both heavy-tailed processes become idle roughly simultaneously. Alternatively, the GSMP can be viewed as a stylized model of overprovisioning. For example, suppose that a resource is permanently dedicated to handling the simultaneous or near-simultaneous occurrence of two events, each of which is recurring but with a heavy-tailed distribution for the interevent times. In an emergency response simulation, one event might be a major forest fire and another event a magnitude 6.0 earthquake; in a reliability simulation, the events may correspond to failures of two highly reliable parts. Our results imply that with probability 1 such events will occur jointly at most a finite number of times.

These positive recurrence, null recurrence, and transience results have a rough analogy to the theory of random walks on multidimensional integer lattices—see Section 6—and illustrate the richly complex behavior of non-Markovian GSMPs relative to their Markovian or semi-Markovian counterparts. Our results extend to other discrete-event formalisms, such as stochastic Petri nets [Balbo and Chiola 1989; Chiola 1991; Chiola and Ferscha 1993; Haas 2002; Haas and Shedler 1986] and event graphs [Schruben 1983]. A preliminary version of these results, in the setting of stochastic Petri nets, appeared in Glynn and Haas [2012].

2. GSMP PRELIMINARIES

We briefly review the standard GSMP model and then describe the "hazard rate" representation of a GSMP. We next define a useful set of irreducibility and "positive density" conditions and then discuss some notions of recurrence in a GSMP.

2.1. The GSMP Model

Following Shedler [1993], let $E = \{e_1, e_2, \ldots, e_M\}$ be a finite set of *events* and S be a finite or countably infinite set of *states*. For $s \in S$, let $s \mapsto E(s)$ be a mapping from S to the nonempty subsets of E; here $E(s)$ denotes the set of all events that are scheduled to occur when the process is in state s. An event $e \in E(s)$ is said to be *active* in state s. When the process is in state s, the occurrence of one or more active events triggers a state transition. Denote by $p(s'; s, E^*)$ the probability that the new state is s' given that the events in the set E^* ($\subseteq E(s)$) occur simultaneously in state s. A "clock" is associated with each event. The clock reading for an active event indicates the remaining time until the event is scheduled to occur. These clocks determine which of the active events actually trigger the next state transition. The most general GSMP formulation allows clocks to run down at state-dependent "speeds"; for simplicity, we restrict attention throughout to GSMPs in which all speeds are equal to 1. (Our results extend to GSMPs in which all speeds are positive.) Let $C(s)$ be the set of possible *clock-reading vectors* when the state is s:

$$C(s) = \{\, c = (c_1, \ldots, c_M) : c_i \in [0, \infty) \text{ and } c_i > 0 \text{ if and only if } e_i \in E(s)\,\}.$$

Here the ith component c_i of a clock-reading vector $c = (c_1, \ldots, c_M)$ is the clock reading associated with event e_i. Beginning in state s with clock-reading vector $c = (c_1, \ldots, c_M) \in C(s)$, the time $t^*(s, c)$ to the next state transition—also called the *holding time* in s—is given by

$$t^*(s, c) = \min_{\{\, i : e_i \in E(s)\,\}} c_i. \tag{1}$$

The set of events $E^*(s, c)$ that trigger the next state transition is given by

$$E^*(s, c) = \left\{\, e_i \in E(s) : c_i - t^*(s, c) = 0 \,\right\}.$$

At a transition from state s to state s' triggered by the simultaneous occurrence of the events in the set E^*, a finite clock reading is generated for each *new event* $e' \in N(s'; s, E^*) = E(s') \setminus (E(s) \setminus E^*)$. The clock reading for a new event e' is generated according to a *clock-setting distribution function* $F(\,\cdot\,; s', e', s, E^*)$, independently of the clock readings for the other new events. We assume that $F(0; s', e', s, E^*) = 0$, so that new clock readings are a.s. positive, and that $\lim_{x\to\infty} F(x; s', e', s, E^*) = 1$, so that each new clock reading is a.s. finite. For each *old event* $e' \in O(s'; s, E^*) = E(s') \cap (E(s) \setminus E^*)$, the old clock reading is kept after the state transition. For $e' \in (E(s) \setminus E^*) \setminus E(s')$, event e' (which was active before the events in E^* occurred) is canceled and the clock reading is discarded. When E^* is a singleton set of the form $E^* = \{ e^* \}$, we write $p(s'; s, e^*) = p(s'; s, \{ e^* \})$, $O(s'; s, e^*) = O(s'; s, \{ e^* \})$, and so on. The GSMP is a continuous-time stochastic process $\{ X(t) : t \geq 0 \}$ that records the state of the system at time t.

Formal definition of the process $\{ X(t) : t \geq 0 \}$ is in terms of a general state space Markov chain $\{ (S_n, C_n) : n \geq 0 \}$ that describes the process at successive state-transition times. Heuristically, S_n represents the state and $C_n = (C_{n,1}, \ldots, C_{n,M})$ represents the clock-reading vector just after the nth state transition (see Shedler [1993] for a formal definition of the chain). The chain takes values in the set $\Sigma = \bigcup_{s \in S} \bigl(\{ s \} \times C(s) \bigr)$. Denote by μ the *initial distribution* of the chain; for a subset $B \subseteq \Sigma$, the quantity $\mu(B)$ represents the probability that $(S_0, C_0) \in B$. We use the notations P_μ and E_μ to denote probabilities and expected values associated with the chain, the idea being to emphasize the dependence on the initial distribution μ; when the initial state of the underlying chain is equal to some $(s, c) \in \Sigma$ with probability 1, we write $P_{(s,c)}$ and $E_{(s,c)}$. Typically, the GSMP is initialized by selecting an initial state s_0 according to a discrete distribution ν over S, and then generating a clock reading for each $e \in E(s_0)$ from an initial clock-setting distribution function $F_0(\,\cdot\,; e, s_0)$. We assume such an initialization procedure throughout. The symbol P^n denotes the n-step *transition kernel* of the chain: $P^n((s, c), A) = P_{(s,c)}\{ (S_n, C_n) \in A \}$ for $(s, c) \in \Sigma$ and $A \subseteq \Sigma$.

We construct a continuous-time process $\{X(t) : t \geq 0\}$ from the chain $\{ (S_n, C_n) : n \geq 0 \}$ in the following manner. Let ζ_n $(n \geq 0)$ be the nonnegative, real-valued time of the nth state transition: $\zeta_0 = 0$ and $\zeta_n = \sum_{j=0}^{n-1} t^*(S_j, C_j)$ for $n \geq 1$. We focus throughout on GSMPs having finite state space S, in which case an argument as in Theorem 3.3.13 of Haas [2002] shows that $P_\mu\{ \sup_{n \geq 0} \zeta_n = \infty \} = 1$. The GSMP is then defined by setting $X(t) = S_{N(t)}$ for $t \geq 0$, where $N(t) = \sup \{ n \geq 0 : \zeta_n \leq t \}$ is the number of state transitions that occur in the interval $(0, t]$. By construction, the GSMP takes values in the set S and has piecewise constant, right-continuous sample paths.

We focus throughout on GSMPs in which, with probability 1, events never occur simultaneously. In this setting, for $n \geq 1$, denote by $e_n^* = e^*(S_{n-1}, C_{n-1})$ the nth event to occur, and for $n \geq 0$ denote by $t_n^* = t^*(S_n, C_n)$ the holding time in state S_n. Observe that the sequence $U = (S_0, t_0^*, e_1^*, S_1, t_1^*, e_2^*, \ldots)$ completely specifies $\{X(t) : t \geq 0 \}$ along with the sequence of trigger events.

2.2. Simple Events and Heavy-Tailed Events

To simplify the exposition, we focus throughout on GSMPs in which each event e' is *simple* in that there exists a distribution function $F(\,\cdot\,; e')$ such that $F(\,\cdot\,; s', e', s, E^*) \equiv F(\,\cdot\,; e')$ and $F_0(\,\cdot\,; e', s) \equiv F(\,\cdot\,; e')$ for all s', s, and E^*. Note that the assumption of simple events entails no loss of generality for a finite-state GSMP, because any such GSMP with clock-setting distribution functions having an explicit dependence on s', s, and e^* can be "mimicked" by a GSMP with clock-setting distribution functions of the foregoing simple form by (a) using "complex" states of the form $z = (s', s, e^*)$ that record the prior state and current trigger event in the original GSMP and (b) using "complex" events of the form $u = (e'; s', s, e^*)$ that similarly record data about the state transition at which the clock for e' is set. That is, when the original GSMP is in state s and the occurrence of e^* triggers a state transition to s', the "simple" GSMP makes a state transition from $z = (s, s_-, e_-)$ to $z' = (s', s, e^*)$, where s_- denotes the state prior to s and e_- denotes the event that triggered the transition from s_- to s. If $e' \in N(s'; s, e^*)$ in the original GSMP then, at the corresponding transition from z to z' as described previously, the clock for event $u = (e'; s', s, e^*)$ is set according to $F(\,\cdot\,; u)$, which coincides with the distribution function $F(\,\cdot\,; s', e', s, e^*)$ from the original GSMP. The notion of mimicry can be formalized as in Haas [2002, Sec. 4].

For a GSMP with simple events, let H be the (possibly empty) subset of events in E such that $e \in H$ if and only if $\int_0^\infty t\,dF(t;e) = \infty$. Thus H is the set of "heavy-tailed" events. Set $\eta(H) = \max_{s \in S} |E(s) \cap H|$, so that $\eta(H)$ is the maximum number of heavy-tailed events that can be active simultaneously. In the following sections, we consider three possible scenarios: $\eta(H) = 0$, $\eta(H) = 1$, and $\eta(H) \geq 2$.

In applications, a heavy-tailed clock reading might correspond to the time required to process or transmit a computer file [Resnick and Rootzén 2000], the time between extreme geophysical events [Benson et al. 2007], or the time to deal with a financial loss [Moscadelli 2004], insurance claim [Powers 2010], extreme event in a highly optimized physical system [Carlson and Doyle 1999], or natural disaster such as a forest fire [Holmes et al. 2008]—in the latter examples we assume that these times are proportional to the loss amount, claim amount, magnitude of physical-system deviation, number of acres burned, and so on. In general, infinite-mean distributions are used to model situations in which there is a relatively large chance of seeing values much greater than any seen previously.

2.3. Hazard-Rate Representation

It is sometimes convenient to use an alternative construction of a GSMP based on "hazard rates" (see, e.g., Glasserman [1991, Ch. 6] or Glynn [1989]). Consider a GSMP for which each clock-setting distribution $F(x;e)$ has a density function $f(x;e)$, and define the corresponding *hazard-rate function* by $h(x;e) = f(x;e)/\bar{F}(x;e)$, where $\bar{F} = 1 - F$ and we take $0/0 = 0$. With probability 1, events never occur simultaneously for such a GSMP. Also observe that if $h(x;e) \geq \underline{h}$ for some $\underline{h} > 0$ and all $x \in (0,\infty)$, then $f(x;e) > 0$ for all $x \in (0,\infty)$ and $F(\,\cdot\,;e)$ has finite moments of all orders.

The hazard-rate construction rests on the following fact. Let F be a distribution function having a density f and hazard rate h, and let Q be distributed according to a unit exponential distribution function. The random variable τ defined by $\tau = \inf\{\, t \geq 0 : Q - \int_0^t h(x)\,dx = 0 \,\}$ has distribution function F. That is, we can simulate a clock-setting distribution F for an event e by sampling from a unit exponential clock-setting distribution and running the corresponding new clock reading down to 0 at a time-varying rate given by h. To see this, observe that $h(t) = f(t)/\bar{F}(t) = -d/dt(\ln \bar{F}(t))$, so that $\int_0^x h(t)\,dt = -\ln \bar{F}(x)$ and

$$P\{\,\tau \geq x\,\} = P\left\{ \int_0^x h(t)\,dt \leq Q \right\} = \exp\left(-\int_0^x h(t)\,dt \right) = \bar{F}(x)$$

for $x \in \Re$. More generally, if we set $\tau = \inf\{\, t \geq 0 : Q - \int_a^t h(x)\,dx = 0 \,\}$ for some $a > 0$, then $P\{\,\tau \geq x\,\} = \bar{F}(x)/\bar{F}(a) = P\{\, Y > x \mid Y > a \,\}$ for $x \geq a$, where Y is distributed according to F. Thus τ is distributed as a clock reading for an event e with clock-setting distribution F in a conventionally defined GSMP, conditional on the fact that the current "age" of the clock in such a GSMP is a. (The *age* of a clock is the amount of time that the clock has been running down.)

Continuing further, it follows from the memoryless property of the exponential distribution that if a clock for an event e has a unit exponential clock-setting distribution, then—conditional on the past history of GSMP states, holding times, and trigger events—the clock reading $\tilde{C}$ at a state-transition time $\tilde{\zeta}$ also has a unit exponential distribution (provided that the corresponding event is active at time $\tilde{\zeta}$). This assertion can be established rigorously by directly applying Lemma 3.4.10 in Haas [2002]; indeed, the argument shows that if multiple such clocks are active at $\tilde{\zeta}$, then the clock readings are independent and identically distributed (i.i.d). exponential, whether the clocks run down at a constant or time-varying rate. Thus, denoting by $\tilde{\zeta}_\alpha$ the time when the clock for e was most recently set, we can take the previous $a = \tilde{\zeta} - \tilde{\zeta}_\alpha$ and $Q = \tilde{C}$ to see that

$$\tau = \inf\left\{ t \geq 0 : \int_{\tilde{\zeta}}^{t} h(x - \tilde{\zeta}_\alpha)\, dx = \tilde{C} \right\} = \inf\left\{ t \geq \tilde{\zeta}_\alpha : \tilde{C} - \int_{\tilde{\zeta}-\tilde{\zeta}_\alpha}^{t-\tilde{\zeta}_\alpha} h(x)\, dx = 0 \right\}$$

satisfies $P\{\, \tau \geq x \,\} = P\{\, \tilde{\zeta}_\alpha + Y > x \mid \tilde{\zeta}_\alpha + Y > \tilde{\zeta} \,\}$ where $Y \overset{D}{\sim} F$ as before. That is, the random time τ is distributed precisely as the scheduled occurrence time for event e in a conventional GSMP, assuming that the clock for e was set according to F at time $\tilde{\zeta}_\alpha$ and given that e occurs after observation time $\tilde{\zeta}$.

These results suggest an alternate algorithm for simulating a GSMP in which new clocks are always set according to a unit exponential distribution but clocks run down at time-varying rates specified by their hazard-rate functions. Specifically, consider the algorithm given in the following, where we use tildes to distinguish the hazard-rate construction from the standard GSMP construction. In the algorithm, the quantity $\alpha(n, i)$ denotes the random index of the most recent event-occurrence time at or prior to $\tilde{\zeta}_n$ at which the clock for $e_i \in E(\tilde{S}_n)$ was set; by convention, $\alpha(n, i) = 0$ for $e_i \notin E(\tilde{S}_n)$.

(0) (Initialization) Set $n = 0$, $\tilde{\zeta}_0 = 0$, and $\alpha(0, i) = 0$ for $e_i \in E$. Select $\tilde{S}_0$ according to ν. For each $e \in E(\tilde{S}_0)$, generate $\tilde{C}_{n,i}$ according to a unit exponential distribution $F(x) = (1 - \exp(-x))I[x \geq 0]$.
(1) For $e_i \in E(\tilde{S}_n)$, set $\tau_{n,i} = \inf\{\, t \geq 0 : \int_{\tilde{\zeta}_n}^{t} h(x - \tilde{\zeta}_{\alpha(n,i)}; e_i)\, dx = \tilde{C}_{n,i} \,\}$. Then set $\tilde{\zeta}_{n+1} = \min_{\{i: e_i \in E(\tilde{S}_n)\}} \tau_{n,i}$, $\tilde{t}^*_n = \tilde{\zeta}_{n+1} - \tilde{\zeta}_n$, and $\tilde{e}^*_{n+1} = e_i$ such that $\tau_{n,i} = \tilde{\zeta}_{n+1}$.
(2) Generate $\tilde{S}_{n+1}$ according to $p(\,\cdot\,; \tilde{S}_n, \tilde{e}^*_{n+1})$.
(3) For each $e_i \in N(\tilde{S}_{n+1}; \tilde{S}_n, \tilde{e}^*_{n+1})$, set $\alpha(n+1, i) = n+1$ and generate $\tilde{C}_{n+1,i}$ according to a unit exponential distribution.
(4) For each $e_i \in O(\tilde{S}_{n+1}; \tilde{S}_n, \tilde{e}^*_{n+1})$, set $\alpha(n+1, i) = \alpha(n, i)$ and $\tilde{C}_{n+1,i} = \tilde{C}_{n,i} - \int_{\tilde{\zeta}_n}^{\tilde{\zeta}_{n+1}} h(x - \tilde{\zeta}_{\alpha(n,i)}; e_i)\, dx$.
(5) For each $e_i \in (E(\tilde{S}_n) \setminus \{\tilde{e}^*_n\}) \setminus E(\tilde{S}_{n+1})$, set $\tilde{C}_{n+1,i} = \alpha(n+1, i) = 0$.
(6) Set $n \leftarrow n+1$ and go to Step 1.

In Step 1, we can see from the previous discussion that the $\tau_{n,i}$ random variables are distributed as the scheduled occurrence times (in a standard GSMP) of the events currently active at $\tilde{\zeta}_n$, so that the minimum of these times ($\tilde{\zeta}_{n+1}$) is correctly distributed as the time of the next event occurrence and $\tilde{\zeta}_{n+1} - \tilde{\zeta}_n$ as the holding time in the current state. After generating clock readings for the new events in Step 3, the clock for each old event e_i—which was originally set to a value of $\tilde{C}_{\alpha(n,i),i}$—is decremented in Step 4 to a value of

$$\begin{aligned}
\tilde{C}_{n+1,i} &= \tilde{C}_{\alpha(n,i),i} - \int_0^{\tilde{\zeta}_{n+1}-\tilde{\zeta}_{\alpha(n,i)}} h(x; e_i)\, dx \\
&= \tilde{C}_{\alpha(n,i),i} - \int_0^{\tilde{\zeta}_n-\tilde{\zeta}_{\alpha(n,i)}} h(x; e_i)\, dx - \int_{\tilde{\zeta}_n-\tilde{\zeta}_{\alpha(n,i)}}^{\tilde{\zeta}_{n+1}-\tilde{\zeta}_{\alpha(n,i)}} h(x; e_i)\, dx \\
&= \tilde{C}_{n,i} - \int_{\tilde{\zeta}_n-\tilde{\zeta}_{\alpha(n,i)}}^{\tilde{\zeta}_{n+1}-\tilde{\zeta}_{\alpha(n,i)}} h(x; e_i)\, dx = \tilde{C}_{n,i} - \int_{\tilde{\zeta}_n}^{\tilde{\zeta}_{n+1}} h(x - \tilde{\zeta}_{\alpha(n,i)}; e_i)\, dx.
\end{aligned}$$

Using this algorithm, we can define a continuous-time process $\{\, \tilde{X}(t) : t \geq 0 \,\}$ in a manner analogous to the usual definition of a GSMP by setting $\tilde{X}(t) = \tilde{S}_{\tilde{N}(t)}$, where $\tilde{N}(t) = \sup\{\, n \geq 0 : \tilde{\zeta}_n \leq t \,\}$.

Consider the sequences $U = (S_0, t^*_0, e^*_1, S_1, t^*_1, e^*_2, \ldots)$ and $\tilde{U} = (\tilde{S}_0, \tilde{t}^*_0, \tilde{e}^*_1, \tilde{S}_1, \tilde{t}^*_1, \tilde{e}^*_2, \ldots)$, where U is defined for the standard GSMP as at the end of Section 2.1 and

$\tilde{U}$ is the corresponding sequence for the hazard-rate construction. We have the following result.

PROPOSITION 2.1. *U and $\tilde{U}$ are identically distributed, as are $\{X(t) : t \geq 0\}$ and $\{\tilde{X}(t) : t \geq 0\}$.*

The proof of the proposition combines the foregoing results in a straightforward inductive argument very similar to the proof of Theorem 3.4.21 in Haas [2002]. When using the hazard-rate construction in what follows, we suppress the tilde notation.

2.4. Irreducibility and Positive Density Conditions

In the following sections, we restrict attention to a certain class of "irreducible" finite-state GSMPs with simple events whose clock-setting distributions satisfy a "positive density" condition. To make our assumptions precise, we give a couple of definitions. For a GSMP with state space S and event set E and for $s, s' \in S$ and $e \in E$, write $s \xrightarrow{e} s'$ if $p(s'; s, e) > 0$ and write $s \to s'$ if $s \xrightarrow{e} s'$ for some $e \in E(s)$. Also write $s \rightsquigarrow s'$ if either $s \to s'$ or there exist states $s_1, s_2, \ldots, s_n \in S$ ($n \geq 1$) such that $s \to s_1 \to \cdots \to s_n \to s'$.

Definition 2.2. A GSMP is said to be *irreducible* if $s \rightsquigarrow s'$ for each $s, s' \in S$.

Recall that a nonnegative function G is a *component* of a distribution function F if G is not identically equal to 0 and $G \leq F$. If G is a component of F and G is absolutely continuous, so that G has a density function g, then we say that g is a *density component* of F.

Definition 2.3. Assumption PD(q) holds for a specified GSMP and real number $q \geq 0$ if

(1) the state space S of the GSMP is finite;
(2) the GSMP is irreducible; and
(3) there exists $\bar{x} \in (0, \infty)$ such that each clock-setting distribution function $F(\,\cdot\,;, e')$ of the GSMP has finite qth moment and a density component that is positive and continuous on $(0, \bar{x})$.

Observe that Assumption PD(q) implies Assumption PD(r) whenever $q \geq r$. Moreover, Assumption PD(0) imposes finite-state, irreducibility, and positive density conditions, but does not impose any conditions on the means of the clock-setting distributions. The role of the positive density assumption is to rule out pathological situations in which, for example, the clocks for events e_i and e_j are always set at the same time at a transition to a given state s, but the maximum possible clock reading for e_i is less than the minimum possible clock reading for e_j, so that e_i always occurs before e_j. Thus we might have $p(s'; s, e_j) > 0$ for some state s', but, with probability 1, e_j will never actually trigger a transition from s to s', so that the GSMP may be reducible with probability 1 even though it is nominally irreducible in the sense of Definition 2.2.

2.5. Recurrence in GSMPs

Consider a GSMP having state space S and defined in terms of an underlying Markov chain $\{(S_n, C_n) : n \geq 0\}$ with state space Σ. The underlying chain is said to be ϕ-*irreducible* if ϕ is a nontrivial measure on subsets of Σ and, for each $(s, c) \in \Sigma$ and subset $A \subseteq \Sigma$ with $\phi(A) > 0$, there exists $n > 0$—possibly depending on both (s, c) and A—such that $P^n((s, c), A) > 0$. The chain is *Harris recurrent* with recurrence measure ϕ if the chain is ϕ-irreducible and $P_{(s,c)}\{(S_n, C_n) \in A \text{ i.o.}\} = 1$ for all $(s, c) \in \Sigma$ and $A \subseteq \Sigma$ with $\phi(A) > 0$. A Harris recurrent chain admits an invariant distribution π_0 that is unique up to constant multiples. If $\pi_0(\Sigma) < \infty$, then $\pi(\,\cdot\,) = \pi_0(\,\cdot\,)/\pi_0(\Sigma)$ is the unique invariant probability distribution for the chain. A Harris recurrent chain that admits such a probability distribution is called *positive Harris recurrent*. Roughly speaking, Harris recurrence means that any "dense enough" set of states is hit infinitely often with probability 1; positive Harris recurrence means that the expected time between successive hits is finite. A GSMP state $s \in S$ is *recurrent* if and only if $P_\mu\{\, S_n = s \text{ i.o.} \,\} = 1$. Observe that if $\{(S_n, C_n) : n \geq 0\}$ is Harris recurrent with a recurrence measure ϕ such that $\phi(\{s\} \times C(s)) > 0$ for each $s \in S$, then each $s \in S$ is a recurrent state of the GSMP.

3. NO HEAVY-TAILED EVENTS

The following Proposition 3.1 gives the basic recurrence result when Assumption PD(0) holds and every clock-setting distribution has finite mean. Denote by $J_n(s)$ and $T_n(s)$ the nth hitting time of state s (in discrete and continuous time, respectively) by the processes $\{S_n : n \geq 0\}$ and $\{X(t) : t \geq 0\}$.

PROPOSITION 3.1. *Suppose that Assumption PD*(1) *holds for a GSMP, so that all clock-setting distributions have finite mean. Then, for any initial distribution* μ, *(i) every state of the GSMP is recurrent and (ii)* $E_\mu[J_n(s)] < \infty$ *and* $E_\mu[T_n(s)] < \infty$ *for* $s \in S$ *and* $n \geq 0$.

The assertion in (ii) can be viewed as a form of "positive" recurrence.

PROOF. By Theorem 6.4 in Haas [1999], the underlying chain is positive Harris recurrent with recurrence measure equal to $\bar{\phi}$, where $\bar{\phi}$ is the unique measure on subsets of Σ such that

$$\bar{\phi}\big(\{\, s \,\} \times [0, x_1] \times [0, x_2] \times \cdots \times [0, x_M]\big) = \prod_{\{i : e_i \in E(s)\}} \min(x_i, \bar{x})$$

for all $s \in S$ and $x_1, x_2, \ldots, x_M \geq 0$. (If, for example, a set $B \subseteq \Sigma$ is of the form $B = \{\, s \,\} \times A$ with $E(s) = E$, then $\bar{\phi}(B)$ is equal to the Lebesgue measure of the set $A \cap [0, \bar{x}]^M$.) Applying this result to sets of the form $\{s\} \times C(s)$ for $s \in S$ establishes that every state of the GSMP is recurrent.

To establish the remaining assertions, fix $s \in S$ and $n \geq 0$. By a standard "splitting" construction—see, for example, Glynn and Haas [2006, Prop. 5.1.1]—the positive Harris recurrence of the underlying chain implies that there exists a positive integer q and a sequence $\{\theta(k) : k \geq 0\}$ of "od-regeneration points" such that (a) the points divide the sample paths of the chain into one-dependent stationary cycles of length at least q and (b) $S_{\theta(k)-q} = s$ for $k \geq 0$. For a function $g : \Sigma \mapsto \Re$, set $Y_k(g) = \sum_{j=\theta(k-1)}^{\theta(k)-1} g(S_j, C_j)$ for $k \geq 0$, where we take $\theta(-1) = 0$. The random variables $\{\, Y_k(g) : k \geq 1 \,\}$ are identically distributed, and Theorem 5.2.1 in Glynn and Haas [2006] implies that $E_\mu[Y_0(|g|)] \vee E_\mu[Y_1(|g|)] < \infty$ if

$$\sup_{(s,c) \in \Sigma} |g(s, c)|/(1 + t^*(s, c)) < \infty. \tag{2}$$

(Recall that t^* is the holding-time function.) For a nonnegative function g that satisfies (2), define $U_n(g) = \sum_{i=0}^{\gamma(n)-1} g(S_i, C_i)$, where $\gamma(n)$ is the random index corresponding to the nth hitting time of state s. Observe that $U_n(g) \leq \sum_{i=0}^{\theta(n)-1} g(S_i, C_i) = \sum_{k=0}^{n} Y_k(g)$, so that $E_\mu[U_n(g)] \leq E_\mu[Y_0(g)] + nE_\mu[Y_1(g)] < \infty$. The final two assertions of the proposition now follow by taking $g \equiv 1$ and $g = t^*$, respectively. □

4. ONE HEAVY-TAILED EVENT

In this section we focus on GSMPs with simple events for which $\eta(H) = 1$. We first show by means of a couple of examples that the hitting times for a specified state may

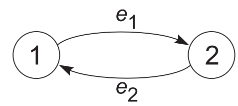

Fig. 2. State-transition diagram for a simple GSMP.

or may not have finite means. We then show that, in any case, each state in a GSMP with $\eta(H) = 1$ is recurrent.

First consider a GSMP with state space $S = \{1, 2\}$ and event set $E = \{e_1, e_2\}$ such that $E(j) = \{e_j\}$ for $j = 1, 2$. Thus exactly one event is active in each state. Suppose that $p(2; 1, e_1) = p(1; 2, e_2) = 1$ and that $P_\mu\{X(0) = 1\} = 1$ (see Figure 2). Also suppose that at least one of the clock-setting distributions for e_1 and e_2 has infinite mean. Set $T_0 = 0$ and $T_n = \inf\{t > T_{n-1} : X(t-) = 2 \text{ and } X(t) = 1\}$, so that T_n is the nth hitting time for state 1. Then each T_n is a.s. finite (because each new clock reading is a.s. finite as per Section 2.1) but $E_\mu[T_n] = \infty$ for $n > 0$. Each hitting time for state 2 is also a.s. finite with infinite mean.

The foregoing example shows that the hitting times to a state may be infinite when $\eta(H) = 1$. To see that hitting times need not be infinite, define an event e to be *uninfluential* if $e \in E(s)$ and $p(s; s, e) = 1$ for $s \in S$. The occurrence of an uninfluential event e does not change the state of the GSMP and does not cause the scheduling or cancellation of any event other than e. Observe that a GSMP having one or more uninfluential events behaves essentially identically to a GSMP in which the uninfluential events are not present. More precisely, the original GSMP can be "strongly mimicked"—in a sense almost identical to Haas [2002, Ch. 4]—by a GSMP in which the uninfluential events have been dropped. If the original GSMP satisfies Assumption PD(0) and only uninfluential events have infinite means, then the mimicking GSMP satisfies Assumption PD(1) and the conclusion of Proposition 3.1 holds. Moreover, the transience and recurrence behavior of states in the two GSMPs is identical. Less trivial examples can be constructed along the same lines.

We now consider the question of recurrence in a GSMP with simple events for which $\eta(H) = 1$. For definiteness and ease of exposition, we assume that $e_1 \in H$ and $e_2, e_3, \ldots, e_m \notin H$. We strengthen Assumption PD(0) by requiring that the clock-setting distributions for events $e_2, e_3, \ldots, e_m$ have hazard rates that are bounded from above and below. (Recall that such clock-setting distributions have density functions that are positive on $(0, \infty)$ and have finite moments of all orders.)

THEOREM 4.1. *Suppose that Assumption PD*(0) *holds for a GSMP. Also suppose that all events are simple, that* $e_1 \in H$, *and that, for* $2 \le i \le m$, *the clock-setting distribution function* $F(\cdot\,; e_i)$ *has a hazard rate* $h(\cdot\,; e_i)$ *with* $0 < \underline{h}_i \le h(t; e_i) \le \overline{h}_i < \infty$ *for* $t \ge 0$. *Then each state s of the GSMP is recurrent.*

To prove this result, we use the hazard-rate construction throughout, with the slight modification that the heavy-tailed event e_1 is set according to $F(\cdot\,; e_1)$ and runs down at unit rate. We assume throughout that there are no "single states" s^* such that $E(s^*) = \{e_1\}$. This assumption also entails no loss of generality. If the GSMP contains such a state s^*, we can construct a modified GSMP $\{X^*(t) : t \ge 0\}$ such that (a) s^* does not belong to the state space of the modified GSMP and (b) each state in $\{X(t) : t \ge 0\}$ is recurrent if each state in $\{X^*(t) : t \ge 0\}$ is recurrent. In the modified GSMP, the transition probabilities are altered so that $p^*(s'; s, e) = p(s'; s, e) + p(s^*; s, e)p(s'; s^*, e_1)$ for all s', s, and e. It is as if we simply take sample paths of the original GSMP and delete all intervals in which the GSMP is in state s^*.

As detailed in Section A of the Appendix, our proof makes repeated use of a conditional geometric trials lemma: if $\{\mathcal{F}_n\}$ is an increasing sequence of σ-fields and $\{A_n\}$ a sequence of events such that $A_n \in \mathcal{F}_n$, then $P\{A_n \text{ i.o.}\} = 1$ if $\inf_n P\{A_n \mid \mathcal{F}_{n-1}\} > \delta$ for some $\delta > 0$ (see Haas [2002, p. 88] or Hall and Heyde [1980, Cor. 2.3]). The idea is to choose the $\{A_n\}$ sequence so that the recurrence of these events implies the recurrence of the GSMP states. Specifically, a "trial" begins at a state transition at which the clock for e_1 has just been set to a value in an interval $[\underline{u}, \overline{u}]$. Denote by s the state just after this state transition and let $s' \in S$ be a specified state (possibly depending on s) that satisfies $e_1 \in E(s')$; we allow s' to coincide with s. The trial is a "success"—that is, an event A_n occurs—if (a) while the e_1 clock runs down, the occurrence of events in $E \setminus \{e_1\}$ causes the GSMP to visit all states reachable from s and end up in s', all within $\underline{u}$ time units, and (b) upon arriving in s' the clock reading for each event $e \in E(s') \setminus \{e_1\}$ equals or exceeds $\overline{u}$, guaranteeing that e_1 triggers the transition from state s'. The reason for the requirement in (b) is that not all states in S may be reachable from s via occurrences of events in $E \setminus \{e_1\}$; when this is the case, the state s' is chosen such that the occurrence of e_1 in s' causes the chain to jump to a previously inaccessible part of the state space (with positive probability). An additional complication is that e_1 may be canceled and rescheduled one or more times during the excursion described previously. For a trial to be a "success" we therefore also require that (c) the clock for e_1 be set to a value in $[\underline{u}, \overline{u}]$ at each such rescheduling; this requirement guarantees that there will be no occurrences of e_1 to interrupt the excursion through the states driven by the events in $E \setminus \{e_1\}$ and that, when the GSMP arrives in s', event e_1 is still guaranteed to trigger the next state transition. The boundedness of the hazard rates for the events in $E \setminus \{e_1\}$ allows us to bound the probability of a "success" away from 0 uniformly in n, so that there are infinitely many successes with probability 1 by the geometric trials lemma.

We conjecture that the conclusion of the theorem remains true even in the absence of the bounded-hazard-rate assumption. As discussed previously, the upper and lower hazard-rate bounds are jointly used to bound away from 0 the probability of a success-that is, to bound the probability that certain events occur either "soon enough" or "late enough" to ensure a complete tour of reachable states before the next occurrence of e_1—so that the sum of success probabilities diverges and hence successes (and thus visits to each state s) occur infinitely often with probability 1 by the Borel-Cantelli lemma. The Borel-Cantelli argument only requires, however, that the success probabilities converge to 0 slowly enough that their sum diverges, which implies that strict bounding may not be needed. Moreover, there are many other scenarios under which all of the states are visited besides the one given in the proof, and the probability of the union of such scenarios might be large enough so as to lead to recurrence without the need for hazard-rate bounds.

5. TWO OR MORE HEAVY-TAILED EVENTS

We continue to assume that all events are simple, but now suppose that $\eta(H) \geq 2$. Arguments almost identical to those in the previous section show that the hitting times for a specified state s may or may not have finite means. The key result of the current section is that a GSMP can have transient states when $\eta(H) \geq 2$. It may seem surprising that, in the presence of a finite state space and irreducibility, some sort of additional moment condition appears necessary to ensure recurrence. We establish the transience result by means of an example; as mentioned in Section 1, our example GSMP can be seen as a stylized model of situations corresponding to resource starvation or overprovisioning.

The GSMP has state space $S = \{1, 2, 3\}$ and event set $E = \{e_1, e_2, e_3\}$. The set of active events is given by $E(s) = E$ for $s = 1, 2, 3$. The state-transition probabilities

Table I. Summary of Recurrence Results for GSMPs with Simple Events

$\eta(H)$	Recurrent?	Positive Recurrent?	"Random Walk" Dimension
0	Yes	Yes	0
1	Yes?	Maybe	1 or 2
≥ 2	Maybe	Maybe	≥ 2

are given by $p(1; s, e_1) = 1$ for $s = 1, 2, 3$, $p(2; 1, e_2) = p(3; 2, e_2) = p(3; 3, e_2) = 1$, and $p(3; s, e_3) = 1$ for $s = 1, 2, 3$ (see the state-transition diagram in Figure 1).

For each event e_i, the clock-setting distribution function is of the form $F(\,\cdot\,; s', e_i, s, E^*) \equiv F_i(\cdot)$. In particular,

$$F_1(t) = 1 - (1 + t)^{-\alpha}, \quad t \geq 0,$$

$$F_2(t) = 1 - (1 + t)^{-\beta}, \quad t \geq 0,$$

and

$$F_3(t) = t/a, \quad t \in [0, a],$$

where $a \in (0, \infty)$ and $\alpha, \beta \in (0, 1)$ with $\beta > 1/2$ and $\alpha + \beta < 1$. Denote by f_i the density function of F_i and observe that f_1, f_2, and f_3 are positive and continuous on $(0, a)$. If each clock-setting distribution had a finite mean, then Assumption PD(1) would hold and each state of the GSMP would be recurrent by Proposition 3.1. The clock-setting distributions F_1 and F_2 have infinite mean, however, and we show in the following that state $s = 2$ is transient.

As mentioned earlier, the only way in which the GSMP can hit state 2 is if event e_1 occurs and then event e_2 occurs without an intervening occurrence of event e_3. That is, if T_n denotes the nth time at which event e_1 occurs and E_n denotes the first event to occur after T_n, then state 2 is recurrent if and only if $P_\mu\,\{E_n = e_2 \text{ i.o.}\} = 1$. Roughly speaking, the clock readings for e_1 and e_2 must simultaneously be close to 0 infinitely often with probability 1. The following result, whose proof is given in Section B of the Appendix, shows that this condition does not hold.

THEOREM 5.1. *Under the foregoing assumptions on the clock-setting distribution functions,* $P_\mu\{E_n = e_2 \textit{ i.o.}\} = 0$.

Note that a GSMP with $\eta(H) \geq 2$ need not have transient states. For example, if only uninfluential events have infinite means, then Proposition 3.1 implies that every state s is recurrent (see Section 4).

6. SUMMARY AND CONCLUSION

Table I summarizes our results. The finite-mean requirement in Proposition 3.1 ensures that each state s is recurrent, and indeed "positive recurrent" in the sense of the proposition. Theorem 4.1 asserts that recurrence (though not positive recurrence) is still ensured if at most one heavy-tailed event can be active at any time point. (We imposed a bounded-hazard-rate assumption to facilitate the proof, but we conjecture that the conclusion holds in the absence of this assumption, hence the "?" in the second row of the table.) The results in Section 5 show that the requirement of at most one active heavy-tailed event is "almost necessary" in that a GSMP can have transient states when the requirement is relaxed. A simple necessary moment condition for recurrence appears elusive, as does a simple sufficient condition weaker than those in Proposition 3.1 and Theorem 4.1. We expect that obtaining weaker conditions for recurrence would involve analysis of the detailed structure of the GSMP under consideration. Indeed, GSMPs contain networks of queues as special cases, and recurrence theory for such networks is quite intricate [Bramson 2008].

The rightmost column of the table illustrates a rough analogy to the theory of random walks on the d-dimensional integer lattice. As is well known [Chung and Fuchs 1951], the origin is positive recurrent (trivially) when $d = 0$, null recurrent when $d = 1$ or $d = 2$, and transient when $d \geq 3$. Our transience example centers around an event that occurs only when multiple clock readings are "small" simultaneously, which roughly corresponds to a multidimensional random walk being close to the origin. A heavy-tailed clock-setting distribution-that is, the distribution of the time required for the clock reading to run down to 0—can be viewed as analogous to the distribution of the time required for a one- or two-dimensional random walk to return to the origin. The case $\eta(H) = 1$ therefore roughly corresponds to a random walk situation on the one-dimensional or two-dimensional lattice. The corresponding random walk analogs for the other cases are displayed in the rows above and below. Thus, in retrospect, it might not be totally surprising that positive recurrence can be established in the absence of heavy-tailed events, null recurrence can be established with one heavy-tailed event, and transience can occur with two or more heavy-tailed events.

Ultimately, the results in this article serve to illustrate the rich and complex behavior that can occur in non-Markovian GSMPs, especially those with the sort of heavy-tailed clock-setting distributions that arise in financial, insurance, internet-traffic, reliability, and geophysical modeling. Much work remains to be done in gaining a fundamental understanding of this class of stochastic models. The behavioral complexity of GSMP models described here also highlights the importance of simulation as a tool for studying such models in the context of practical system design and decision-making.

APPENDIX

A. PROOF OF THEOREM 4.1

We begin the formal proof of Theorem 4.1 by establishing lower bounds for probabilities of a pertinent class of events, defined as follows. Set $K(s) = \{j : j \geq 2 \text{ and } e_j \in E(s)\}$ for $s \in S$, and set $\tilde{t}^*_n = (\min_{i\in K(S_n)} \tau_{n,i}) - \zeta_n$, so that $\tilde{t}^*_n$ is the holding time in state S_n if event e_1 does not trigger the state transition out of S_n. Also define the event

$$\begin{aligned} B_n(s', e_i, x, z_1, z_2) = \big\{&S_{n+1} = s',\ e^*_{n+1} = e_i,\ \tilde{t}^*_n \leq x, \\ &\text{and } C_{n+1,1} \in [z_1 \wedge z_2, z_1 \vee z_2] \text{ if } e_1 \in N(s'; S_n, e^*_{n+1})\big\}, \end{aligned}$$

where $u \wedge v$ ($u \vee v$, respectively) denotes the minimum (maximum, respectively) of u and v. The sequence of states $\sigma = (s^{(1)}, e^{(2)}, s^{(2)}, \ldots, e^{(k)}, s^{(k)})$ is called *feasible* if $e^{(j)} \in E(s^{(j-1)})$ and $p(s^{(j)}; s^{(j-1)}, e^{(j)}) > 0$ for $2 \leq j \leq k$. For any $k \geq 1$ and feasible sequence $\sigma = (s^{(1)}, e^{(2)}, s^{(2)}, \ldots, e^{(k)}, s^{(k)})$ such that $e_1 \in E(s^{(1)}) \cap E(s^{(k)})$, set

$$B_n(\sigma, x, y) = \bigcap_{l=2}^{k} B_{n+l-2}\big(s^{(l)}, e^{(l)}, x/(k-1), x, y\big) \cap \big\{\tilde{t}^*_{n+k-1} > y\big\}.$$

The occurrence of event $B_n(\sigma, x, y)$ implies that, starting from state $s^{(1)}$, the GSMP visits successive states $s^{(2)}, s^{(3)}, \ldots, s^{(k)}$ in x time units or less, since each of the $(k-1)$ holding times is less than or equal to $x/(k-1)$. Moreover, upon arriving in $s^{(k)}$, the clock reading for each event $e \in E(s^{(k)}) \setminus \{e_1\}$ exceeds y time units. Finally, the clock for event e_1 is set to a value in the interval $[x \wedge y, x \vee y]$ whenever e_1 becomes a new event during the passage through the states in σ. (Thus if $x < y$ and the clock reading for e_1 initially lies in the interval $[x, y]$ when the GSMP is in state $s^{(1)}$, then e_1 is guaranteed to be the trigger event when the GSMP makes a transition from state $s^{(k)}$.)

The following Lemma A.1 provides lower bounds on the conditional probability of event $B_n(\sigma, x, y)$. To prepare for this lemma, we need to introduce some additional notation. Set $K(s,i) = K(s) \setminus \{i\}$ and $h(s,i) = \underline{h}_i + \sum_{j\in K(s,i)} \overline{h}_j$. Then set

$$\begin{aligned}\delta(s, s', e_i, x, z_1, z_2) \;=\; & \frac{\underline{h}_i}{\sum_{j\in K(s,i)} \overline{h}_j}(1-\exp(-h(s,i)x))p(s';s,e_i)\\ & \times(I[e_1 \in N(s';s,e_i)](F(z_1\vee z_2;e_1) - F(z_1\wedge z_2;e_1))\\ & + I[e_1 \notin N(s';s,e_i)]),\end{aligned}$$

where $I[A]$ denotes the indicator of event A. For a feasible sequence σ, set

$$\delta(\sigma, x, y) = \exp\left(-y \sum_{i\in K(s^{(k)})} \overline{h}_i\right) \prod_{l=2}^{k} \delta\big(s^{(l-1)}, s^{(l)}, e^{(l)}, x/(k-1), x, y\big).$$

Finally, for $n \ge 1$, denote by $\mathcal{F}_n$ the σ-field generated by (S_0, t_0^*, e_1^*), $(S_1, t_1^*, e_2^*), \ldots, (S_{n-1}, t_{n-1}^*, e_n^*)$, $S_n, C_{n,1}$, where these variables are defined as in Section 2.3.

LEMMA A.1. *Under the conditions of Theorem 4.1,*

$$P\{B_\gamma(\sigma, x, y) \mid \mathcal{F}_\gamma\} \ge I\big[S_\gamma = s^{(1)}\big]\delta(\sigma, x, y)\ a.s. \tag{3}$$

for any random index γ that is a stopping time with respect to $\{\mathcal{F}_n : n \ge 1\}$.

PROOF. Fix $n \ge 1$ and, as in Section 2.3, denote by ζ_n the time of the nth state transition, by $\alpha(n,i)$ the index of the most recent time at or prior to ζ_n at which the clock for e_i was set, and by $\tau_{n,i} = \inf\{t : \int_{\zeta_n}^t h(x - \zeta_{\alpha(n,i)}; e_i)\,dx = C_{n,i}\}$ the time at which event $e_i \in E(S_n)$ is scheduled to occur.

Using the boundedness of the hazard rates, the memoryless property of the exponential distribution and Lemma 3.4.10 in Haas [2002], we have

$$\begin{aligned}P\{\tau_{n,i} > x \mid \mathcal{F}_n\} &= P\left\{\int_{\zeta_n}^{x} h_i(u - \zeta_{\alpha(n,i)}) < C_{n,i} \;\Big|\; \mathcal{F}_n\right\}\\ &\ge P\{(x-\zeta_n)\overline{h}_i < C_{n,i} \mid \mathcal{F}_n\}\\ &= \exp\big(-\overline{h}_i(x-\zeta_n)\big) \text{ a.s.}\end{aligned}$$

Thus, conditioned on $\mathcal{F}_n$, the quantity $\tau_{i,n} - \zeta_n$ is stochastically bounded below by $\overline{\mathcal{E}}_i$, where $\overline{\mathcal{E}}_i$ denotes an exponentially distributed random variable with mean $1/\overline{h}_i$. In a similar manner, it can be shown that $\tau_{i,n} - \zeta_n$ is conditionally stochastically bounded above by $\underline{\mathcal{E}}_i$, where $\underline{\mathcal{E}}_i$ denotes an exponentially distributed random variable with mean $1/\underline{h}_i$.

Recall that $\tilde{t}_n^* = \min_{i\in K(S_n)}(\tau_{n,i} - \zeta_n)$ is the holding time in state S_n if event e_1 does not trigger the state transition out of S_n. Then reasoning similar to that given previously yields

$$P\{\tilde{t}_n^* > x \mid \mathcal{F}_n\} = P\left\{\min_{i\in K(S_n)}(\tau_{n,i} - \zeta_n) > x \;\Big|\; \mathcal{F}_n\right\} \ge \exp\left(-x \sum_{i\in K(S_n)} \overline{h}_i\right) \text{ a.s.} \tag{4}$$

and

$$
\begin{aligned}
P\{e^*_{n+1} &= e_i \text{ and } \tilde{t}^*_n \le x \mid \mathcal{F}_n\} \\
&= P\left\{\tau_{n,i} - \zeta_n < \min_{j \in K(S_n,i)} (\tau_{n,j} - \zeta_n) \text{ and } \tau_{n,i} - \zeta_n \le x \mid \mathcal{F}_n\right\} \\
&\ge \frac{\underline{h}_i}{\sum_{j \in K(S_n,i)} \overline{h}_j} (1 - \exp(-h(S_n, i)x)) \text{ a.s.}
\end{aligned}
$$

for $i \ge 2$ such that $e_i \in E(S_n)$. We have again used Lemma 3.4.10 in Haas [2002], as well as standard properties of independent exponential random variables. We can take the foregoing calculations a step further to obtain

$$
P\{B_n(s', e_i, x, z_1, z_2) \mid \mathcal{F}_n\} \ge I[S_n = s]\delta(s, s', e_i, x, z_1, z_2) \text{ a.s.} \tag{5}
$$

for each $s \in S$. An inductive argument based on (4) and (5) shows that

$$
P\{B_n(\sigma, x, y) \mid \mathcal{F}_n\} \ge I\big[S_n = s^{(1)}\big]\delta(\sigma, x, y) \text{ a.s.}, \tag{6}
$$

and a standard argument extends (6) from a deterministic index n to a stopping time γ. □

In general, event e_1 need not be active in every state, which raises the question of whether the clock for e_1 is set infinitely often with probability 1. The next lemma answers this question in the affirmative.

Lemma A.2. *Under the conditions of Theorem 4.1,* $P\{e_1 \in N(S_n; S_{n-1}, e^*_n)\ i.o.\} = 1$.

Proof. If event e_1 is active in every state, or if the set of states $\hat{S}$ in which e_1 is not active is visited only a finite number of times, the assertion of the lemma follows trivially, so assume that $\hat{S} \ne \varnothing$ and that $P\{S_n \in \hat{S} \text{ i.o.}\} = 1$. The irreducibility assumption in Theorem 4.1 implies that for each state $s \in \hat{S}$ there exists a feasible sequence $\sigma(s) = (s^{(1)}, e^{(2)}, s^{(2)}, \ldots, e^{(k)}, s^{(k)})$ of length $k = k(s)$ such that $s^{(1)} = s$, $s^{(k)} \in S - \hat{S}$, and $s^{(l)} \in \hat{S}$ for $1 \le l \le k-1$. Define a sequence of random indices $\{\gamma(n) : n \ge 0\}$ by setting $\gamma(0) = \inf\{j \ge 0 : S_j \in \hat{S}\}$ and $\gamma(n) = \inf\{j > \gamma(n-1) + k(S_{\gamma(n-1)}) : S_j \in \hat{S}\}$. Observe that each of these random indices is a.s. finite under our assumptions, and is also a stopping time with respect to $\{\mathcal{F}_n\}$. Denote by D_n the event that, at time $\zeta_{\gamma(n)}$, the GSMP proceeds to follow the sequence of states and events specified by $\sigma(S_{\gamma(n)})$. Observe that, using Lemma A.1,

$$
P\{D_n \mid \mathcal{F}_{\gamma(n)}\} = P\{B_{\gamma(n)}(\sigma(S_{\gamma(n)}), \infty, 0) \bigm| \mathcal{F}_{\gamma(n)}\} \ge \min_{s \in \hat{S}} \delta(\sigma(s), \infty, 0) > 0 \text{ a.s.}
$$

for $n \ge 0$. By the geometric trials lemma, $P\{D_n \text{ i.o.}\} = 1$. The desired result now follows, because the occurrence of D_n implies that the clock for e_1 is set at time $\zeta_{\gamma(n)+k(S_{\gamma(n)})-1}$. □

We can now complete the proof of Theorem 4.1.

Proof. The proof proceeds by showing that, with a probability bounded away from 0, each state of the GSMP can be visited while the clock for event e_1 is running down. This proof strategy is complicated by the fact that, even though the GSMP is irreducible by assumption, the GSMP may behave in a "reducible" manner while the e_1 clock is running down. We show that, nonetheless, the GSMP will visit each "reducible" set of states infinitely often with probability 1.

For definiteness, suppose that the state space S can be decomposed into disjoint subsets V_1 and V_2 such that transitions from V_1 to V_2 and vice versa occur only when e_1 triggers a state transition, but any two states in V_i ($i = 1, 2$) are reachable from each other via a sequence of state transitions triggered by events in $E \setminus \{e_1\}$. Suppose that one possible state transition from V_1 to V_2 occurs when the state is $s' \in V_1$ and e_1 triggers a state transition. The set V_1 is irreducible with respect to the GSMP with e_1 removed, so for each state $s \in V_1$ we can construct a feasible sequence $\sigma(s, s')$ of length $k(s, s')$ that starts in s, terminates in s', contains every state in V_1 at least once, and is such that e_1 is never a trigger event.

Fix constants $0 < \underline{u} < \overline{u} < \infty$ and define a sequence of random indices $\{\gamma(n) : n \geq 0\}$ by setting

$$\gamma(0) = \inf\{j \geq 0 : e_1 \in N(S_j; S_{j-1}, e_j^*) \text{ and } C_{j,1} \in [\underline{u}, \overline{u}]\}$$

and

$$\gamma(n) = \inf\{j \geq \gamma(n-1) + k(S_{\gamma(n-1)}, s') : e_1 \in N(S_j; S_{j-1}, e_j^*) \text{ and } C_{j,1} \in [\underline{u}, \overline{u}]\}$$

for $n \geq 1$. Each $\gamma(n)$ is a.s. finite by Lemma A.2 and an (unconditional) geometric trials argument, and is also a stopping time with respect to $\{\mathcal{F}_n\}$. Whenever $S_{\gamma(n)} \in V_1$ for some $n \geq 0$, denote by A_n the event that, at time $\zeta_{\gamma(n)}$,

(1) the GSMP proceeds to follow the sequence of states and events specified by $\sigma(S_{\gamma(n)}, s')$;
(2) the GSMP reaches s' in at most $\underline{u}$ time units;
(3) every new clock reading for e_1 during this excursion lies in the range $[\underline{u}, \overline{u}]$;
(4) after the GSMP arrives in s', the clock for each event in $E(s') \setminus \{e_1\}$ is scheduled to take at least $\overline{u}$ time units to run down to 0, so that e_1 triggers the next state transition; and
(5) when e_1 triggers the foregoing state transition in state s', the GSMP jumps to V_2.

Using Lemma A.1,

$$\begin{aligned} P\{A_n \mid \mathcal{F}_{\gamma(n)}\} &= P\big\{B_{\gamma(n)}(\sigma(S_{\gamma(n)}, s'), \underline{u}, \overline{u}) \cap \{S_{\gamma(n)+k(S_{\gamma(n)})} \in V_2\} \,\big|\, \mathcal{F}_{\gamma(n)}\big\} \\ &\geq \min_{s \in V_1} \delta(\sigma(s, s'), \underline{u}, \overline{u}) p(V_2; s', e_1) \\ &> 0 \text{ a.s.} \end{aligned}$$

for $n \geq 0$, where $p(V_2; s', e_1) = \sum_{s'' \in V_2} p(s''; s', e_1) > 0$. Now suppose that $S_{\gamma(n)} \in V_2$ and consider the probability, conditional on $\mathcal{F}_{\gamma(n)}$, that the GSMP will visit every state in V_2 while the clock for event e_1 runs down and then jump to V_1 when e_1 occurs. An argument similar to the one given previously shows that this conditional probability is bounded below by some positive constant δ that does not depend on n or $S_{\gamma(n)}$. By the geometric trials lemma, both of the foregoing scenarios occur infinitely often with probability 1, so that each state in $S = V_1 \cup V_2$ is recurrent. This argument generalizes in a straightforward way to cases where S can be decomposed into three or more V_i sets, or where S is irreducible in the absence of e_1. □

B. PROOF OF THEOREM 5.1

Observe that, for each event e_i, the successive occurrence times for e_i form a renewal process; the three renewal processes thus defined are mutually independent. For $i = 1, 2, 3$, denote by $C_i(t)$ the clock reading for event e_i at time $t \geq 0$. The random variable $C_i(t)$ is the residual life at time t for the renewal process associated with the successive occurrences of e_i. Setting $B_n = \{C_2(T_n) \leq C_3(T_n)\}$, we have

$$P_\mu\{E_n = e_2 \text{ i.o.}\} = P_\mu\{C_2(T_n) \leq \min(C_1(T_n), C_3(T_n)) \text{ i.o.}\} \leq P_\mu\{B_n \text{ i.o.}\}.$$

By the Borel-Cantelli lemma, it therefore suffices to show that $\sum_{n=1}^\infty P_\mu\{B_n\} < \infty$. Conditioning on T_n and $C_3(T_n)$ and exploiting the independence of the various renewal processes yields the representation

$$P_\mu\{B_n\} = \int_0^\infty \left(\int_0^a P_\mu\{C_2(t) \le y\} dG_t(y) \right) f_1^{*n}(t)\, dt \le \int_0^\infty P_\mu \left\{ C_2(t) \le a \right\} f_1^{*n}(t)\, dt, \qquad (7)$$

where G_t is the distribution function of $C_3(t)$ and f_1^{*n} is the n-fold convolution of f_1 (and hence the density function of T_n). Denote by u_1 the renewal density function that corresponds to F_1 (see, e.g., Asmussen [2003, pp. 147–148]). Using the well known representation $u_1(t) = \sum_{n=1}^\infty f_1^{*n}(t)$, we find that

$$\sum_{n=1}^\infty P_\mu \{B_n\} \le \sum_{n=1}^\infty \int_0^\infty h(t) f_1^{*n}(t)\, dt = \int_0^\infty h(t) u_1(t)\, dt, \qquad (8)$$

where $h(t) = P_\mu\{C_2(t) \le a\}$. The nonnegativity of the terms in (8) justifies the interchange of summation and integration.

To show that the rightmost integral in (8) is finite, we first analyze the term $h(t)$ that appears in the integrand. Set $Q(t) = F_2(t+a) - F_2(t)$ for $t \ge 0$. A standard renewal argument [Asmussen 2003, Sec. V] shows that $h = U_2 * Q$, where $*$ denotes convolution and U_2 is the renewal function corresponding to F_2. Recall that a real-valued function G is said to be *regularly varying* at ∞ with index λ if $\lim_{t\to\infty} G(tx)/G(t) = x^\lambda$ for $x > 0$, and we write $G \in \mathrm{RV}_\lambda$ (see, e.g., Feller [1972, Sec. VIII.8] or Resnick [1987, Sec. 0.4]). Observe that $\bar{F}_2 \in \mathrm{RV}_{-\beta}$ and that $Q(t) = o(1/t)$. Moreover, Q is directly Riemann integrable [Asmussen 2003, p. 155]. It then follows from a key renewal theorem for random variables with regularly varying tails [Erickson 1970, Th. 3] that $h(t) = O(1/m_2(t))$, where

$$m_2(t) = \int_0^t \bar{F}_2(u)\, du = \frac{(1+t)^{1-\beta}}{1-\beta}.$$

Thus $h(t) = O(t^{\beta-1})$.

We now consider the term $u_1(t)$ in (8). Because $\bar{F}_1 \in \mathrm{RV}_{-\alpha}$, an argument based on Tauberian theorems shows that U_1, the renewal function corresponding to F_1, satisfies $U_1 \in \mathrm{RV}_\alpha$ (see Erickson [1970, p. 265] or Feller [1972, p. 417]). A direct application of Theorem 2 in Topchii [2010] shows that

$$\lim_{t\to\infty} u_1'(t) \left[(\alpha - 1) \frac{t^{\alpha-2} \sin(\alpha\pi)}{\pi} \right]^{-1} = 1,$$

which implies that u_1 is "ultimately monotone," that is, u_1 is decreasing on some interval of the form $[t_0, \infty)$. Theorems 2 and 4 in Feller [1972, Sec. XIII.5] then imply that $u_1 \in \mathrm{RV}_{\alpha-1}$. To complete the proof, fix $\epsilon > 0$ such that $\alpha + \beta + \epsilon < 1$. It follows from Feller [1972, p. 277] that $u_1(t) < t^{\alpha+\epsilon-1}$ for sufficiently large t, so that $h(t)u_1(t) = O(t^\gamma)$, where $\gamma = \alpha + \beta + \epsilon - 2 < -1$. The rightmost integral in (8) therefore converges.

ACKNOWLEDGMENTS

The authors would like to thank Pierre Dersin, James Hafner, Vitali Volovoi, Jim Wilson, and the two anonymous reviewers for their helpful suggestions.

REFERENCES

S. Asmussen. 2003. *Applied Probability and Queues* (2nd. ed.). Wiley, New York.

G. Balbo and G. Chiola. 1989. Stochastic Petri net simulation. In *Proceedings of the 1989 Winter Simulation Conference*, E. A. MacNair, K. J. Musselman, and P. Heidelberger (Eds.). ACM Press, New York, NY, 266–276.

D. A. Benson, R. Schumer, and M. M. Meerschaert. 2007. Recurrence of extreme events with power-law interarrival times. *Geophys. Res. Lett.* 34 (2007), L16404.

M. Bramson. 2008. *Stability of Queueing Networks*. Lecture Notes in Mathematics, Vol. 1950. Springer-Verlag.

J. M. Carlson and J. Doyle. 1999. Highly optimized tolerance: A mechanism for power laws in designed systems. *Phys. Rev. E* 60, 2 (1999), 1412–1427.

G. Chiola. 1991. Simulation framework for timed and stochastic Petri nets. *Intl. J. Comput. Simul.* 1 (1991), 153–168.

G. Chiola and A. Ferscha. 1993. Distributed simulation of Petri nets. *Parallel Distributed Techn.* 1, 3 (1993), 33–50.

K. L. Chung and W. H. J. Fuchs. 1951. On the distribution of values of sums of random variables. *Mem. Amer. Math. Soc.* 6 (1951), 1–12.

M. A. Crane and D. L. Iglehart. 1975. Simulating stable stochastic systems: III, Regenerative processes and discrete event simulation. *Oper. Res.* 23 (1975), 33–45.

K. B. Erickson. 1970. Strong renewal theorems with infinite mean. *Trans. Amer. Math. Soc.* 151 (1970), 263–291.

W. Feller. 1972. *An Introduction to Prabability Theory and Its Applications*. Vol. 2. Wiley, New York.

P. Glasserman. 1991. *Gradient Estimation via Perturbation Analysis*. Kluwer Academic, Boston.

P. W. Glynn. 1989. A GSMP formalism for discrete event systems. *Proc. IEEE* 77 (1989), 14–23.

P. W. Glynn and P. J. Haas. 2006. A law of large numbers and functional central limit theorem for generalized semi-Markov Processes. *Comm. Statist. Stochastic Models* 22 (2006), 201–231.

P. W. Glynn and P. J. Haas. 2012. On simulation of non-Markovian stochastic Petri nets with heavy-tailed firing times. In *Proceedings of the 2012 Winter Simulation Conference*, O. Rose and A. M. Uhrmacher (Eds.). WSC, 301.

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

P. W. Glynn and D. L. Iglehart. 1993. Conditions for the applicability of the regenerative method. *Manage. Sci.* 39 (1993), 1108–1111.

P. J. Haas. 1999. On simulation output analysis for generalized semi-Markov processes. *Comm. Statist. Stochastic Models* 15 (1999), 53–80.

P. J. Haas. 2002. *Stochastic Petri Nets: Modelling, Stability, Simulation*. Springer-Verlag, New York.

P. J. Haas and G. S. Shedler. 1986. Regenerative stochastic Petri nets. *Perform. Evaluation* 6 (1986), 189–204.

P. J. Haas and G. S. Shedler. 1987. Regenerative generalized semi-Markov processes. *Comm. Statist. Stochastic Models* 3 (1987), 409–438.

P. Hall and C. C. Heyde. 1980. *Martingale Limit Theory and its Application*. Academic Press, New York.

S. G. Henderson and P. W. Glynn. 2001. Regenerative steady-state simulation of discrete-event stochastic systems. *ACM Trans. Model. Comput. Simul.* 11, 4 (2001), 313–345.

T. P. Holmes, R. J. Huggett, Jr., and A. J. Westerling. 2008. Statistical Analysis of Large Wildfires. In *The Economics of Forest Disturbances: Wildfires, Storms, and Invasive Species*, T. P. Holmes, J. P. Prestemon, and K. L. Abt (Eds.). Springer.

D. L. Iglehart and G. S. Shedler. 1983. Simulation of non-Markovian systems. *IBM J. Res. Develop.* 27 (1983), 472–480.

D. L. Iglehart and G. S. Shedler. 1984. Simulation output analysis for local area computer networks. *Acta Inform.* 21 (1984), 321–338.

D. König, K. Matthes, and K. Nawrotzki. 1974. Unempfindlichkeitseigenshaften von Bedienungsprozessen. Appendix to Gnedenko and Kovalenko [1974]. (1974).

M. Miyazawa. 1993. Insensitivity and product-form decomposability of reallocatable GSMP. *Adv. Appl. Probab.* 25 (1993), 415–437.

M. Moscadelli. 2004. *The Modelling of Operational Risk: Experience with the Analysis of the Data Collected by the Basel Committee*. Technical Report 517. Banca d'Italia.

M. R. Powers. 2010. Infinite-mean losses: Insurance's "dread disease." *J. Risk Finance* 11, 2 (2010), 125–128.

S. Resnick. 1987. *Extreme Values, Regular Variation, and Point Processes*. Springer-Verlag, New York.

S. Resnick and H. Rootzén. 2000. Self-similar communication models and very heavy tails. *Ann. Appl. Probab.* 10 (2000), 753–778.

L. Schruben. 1983. Simulation modeling with event graphs. *Commun. ACM* 26, 11 (1983), 957–963.

G. S. Shedler. 1993. *Regenerative Stochastic Simulation*. Academic Press, New York.

V. A. Topchii. 2010. Derivative of renewal density with infinite moment with $\alpha \in (0, 1/2]$. *Sib. Èlektron Mat. Izv.* 7 (2010), 304–349.

W. Whitt. 1980. Continuity of generalized semi-Markov processes. *Math. Oper. Res.* 5 (1980), 494–501.

Received August 2013; revised September 2014; accepted November 2014