

Optimal Sampling of Overflow Paths in Jackson Networks

By JOSE BLANCHET
Columbia University

July 21, 2011

Abstract

We consider the problems of computing overflow probabilities at level N in any subset of stations in a Jackson network and of simulating sample paths conditional on overflow. We construct algorithms that take $O(N)$ function evaluations to estimate such overflow probabilities within a prescribed relative accuracy and to simulate paths conditional on overflow at level N . The algorithms that we present are optimal in the sense that the best possible performance that can be expected for conditional sampling involves $\Omega(N)$ running time. As we explain in our development, our techniques have the potential to be applicable to more general classes of networks.

1 Introduction

Our focus in this paper is on the development of optimal (in a strong sense) simulation algorithms for overflow probabilities at level N in a busy period for an important class of overflow sets. We concentrate on overflows at level N within a busy period in any arbitrary subset of stations in the network (i.e. the probability that, starting from a fixed position, say for instance the origin, the sum of the queue lengths in a given subcollection of stations reaches level N before returning to the origin). However, the methods that we shall discuss can be easily adapted to more general overflow sets. These extensions are explained in Section 7.

Our algorithms allow to compute such overflow probabilities in $O(N)$ function evaluations. Actually, if the initial condition is the origin, our algorithm runs in $O(1)$ function evaluations. Moreover, we also describe how to generate conditional paths to overflow in $O(N)$ running time (see Theorem 1 in Section 3). It is very important to emphasize that the performance of these algorithms is optimal in the sense that the best possible running time for a path generation algorithm given overflow at level N involves at least $\Omega(N)$ function evaluations. There are very few algorithms in

the rare event simulation literature for stochastic networks that exhibit this type of optimal performance.

Efficient rare-event simulation methodology for Jackson networks has been the focus of many papers in the literature in recent years (see [24], [2], [29], [16], [21], [19], [13], [23] [14], [12], [11]; also the survey papers of [20] and [6] for additional references). Most of the development in the literature concentrates on the design and analysis of estimators that are shown to be weakly efficient (for efficiency concepts in rare event simulation see [3], [7]). Testing weak efficiency not only provides theoretical justification for the excellent empirical performance that well designed estimators exhibit in practice, but also it gives insight into ideas that are often applicable in a wide class of problems. Nevertheless, the fact is that *weakly efficient estimators can only guarantee that a subexponential number of replications* (as a function of N) suffice to estimate overflow probabilities of interest within a prescribed relative accuracy. On the other hand it is possible to setup an associated linear system of equations for computing overflow probabilities in Jackson networks which requires $O(N^d)$ variables where d is the dimension of the network (see for instance [4]). Although this implies that polynomial time algorithms are available for computing overflow probabilities in Jackson networks, it is clear that the computational burden can be substantial even for networks of moderate size.

Importance sampling and splitting are the two most popular approaches that have been used to design efficient rare-event simulation estimators for Jackson networks. Importance sampling requires the specification of a so-called change-of-measure whose aim is to mimic as much as possible the conditional distribution of the network given overflow (such conditional distribution is precisely the optimal change-of-measure in the sense of variance minimization). Since the conditional distribution is not accessible in analytical form (which is required to calculate the form of the associated importance sampling estimator) one uses so-called exponential changes-of-measure or exponential tiltings, which are motivated by their role in the asymptotic description of the required conditional distribution by means of large deviations theory.

The use of large deviations theory in the design of efficient rare event simulation estimators can be traced back to the work of Siegmund [28]. Since then, a wealth of papers have been written on the design of importance sampling using exponential changes-of-measure (earlier references include [8], [27], the survey paper [20] contains dozens of references on this and related topics; see also the texts [7], [3]). The work in [17] shows, however, that following the deterministic fluid description given by large deviations theory directly, in a state-independent fashion, could even result in variance increase.

A substantial amount of work has been put in understanding how to apply large deviations ideas to the design of efficient importance sampling estimators for Jackson networks. The papers of [24], [2] are early references in this direction. The first paper develops heuristics based on large deviation techniques and applies them following the deterministic fluid description to the most likely path to overflow in a state-

independent fashion. The second reference uses reversibility to identify the most likely path to overflow in fluid scale (also known as the optimal path) and it also uses exponential changes-of-measure to track the optimal path in a state-independent fashion. The idea of using reversibility to describe the most likely path in fluid scale has been used frequently in the literature in special cases of Jackson networks (see, for instance [19]). The work of [18], [22] and [14] provide rigorous support for the use of reversibility in the description of asymptotic conditional distributions via large deviations in the setting of arbitrary Jackson networks. Reversibility ideas in the setting of sharp large deviations asymptotics for overflow probabilities have been applied in [1].

Glasserman and Kou [16] showed that state-independent sampling is particularly troublesome in the queueing setting and they were able to develop sampling strategies that were shown to be efficient under certain parameter ranges. [11] shows in a simple example of a tandem network with two stations that following such state-independent simulation one can actually obtain infinite variance.

The fact that state-dependent importance sampling seemed necessary for the design of efficient samplers sparked the development of a series of papers by P. Dupuis, H. Wang and their students. Their work, based on a game / control theoretic approach involves finding a subsolution to a non-linear PDE called the Isaacs equation. They first applied their ideas to construct weakly efficient estimators for tandem networks (see [13]). Then, they extended the construction of their sampler to arbitrary Jackson networks (see [14]). In the later reference, they show that their sampler is weakly efficient regardless of the overflow set assuming that the network is initially empty. The fact that the same sampler can be used for virtually any overflow set is a very convenient feature, but, as the authors point out, the price to pay is that for specific overflow sets the sampler can exhibit diminished performance in practice.

Another class of algorithms based on splitting have been used successfully in practice (see [29] for an early reference on the use of splitting to rare event simulation in networks). The work of [15] provides some complexity analysis for splitting methods in rare events. The work of [12] shows how to construct algorithms with guaranteed subexponential complexity using splitting. Such construction relies on the specification of an appropriate importance function (a critical element in the design of splitting algorithms that is analogous to the selection of a suitable importance sampling distribution) which must correspond to a subsolution to an Isaacs equation.

We mentioned earlier that reversibility is a feature that has been exploited to describe the asymptotic conditional paths of the network given overflow. Large deviations theory exploits this feature to provide a description in fluid scale (i.e. neglecting contributions that are of size $o(N)$) of the optimal path to overflow. The problem when translating this description into an importance sampling algorithm is precisely that the conditional distribution of the network up to the time of overflow is only specified asymptotically in fluid scale. It is because of this lack of fine enough specification that the construction of efficient samplers is complicated. Indeed, attempting

to construct an importance sampling distribution backwards in time from the reversed process poses the problem of how to start the process (the initial condition of the reversed process lies precisely in the overflow set). On the other hand, in any attempt to mimic the conditional distribution to overflow forward in time one needs to “protect” the importance sampling distribution on those rogue overflow paths that are unlikely from a large deviations description but could contribute substantially (in relative terms) to the variance of the estimator.

The forward sampling is the one for which a subsolution to the Isaacs equation is required in order to control the behavior of the likelihood ratio precisely on those rogue sample paths. The backward sampling idea has been explored empirically by [23], but without a precise guidance on the selection of the initial distribution in the overflow set and without a complexity analysis of the associated running time. Not surprisingly, given the theory presented here, the empirical performance observed in [23] is excellent whenever the initial distribution imposed in the overflow set happens to be appropriately selected.

The algorithm that we present here is related to the backward sampling idea. Our basic strategy rests on two observations. First, that the conditional distribution of the network at the time of overflow can be approximated by the conditional steady state distribution (conditioned on the overflow set). Second, that a stable Jackson network, starting from any overflow state at level N can empty completely prior to returning to the overflow set with some probability that is bounded away from zero uniformly in N (see (8) and Proposition 4). Although in this paper we focus on Jackson networks only, we believe that the basic ideas can be potentially applied to more general settings, for instance to the context of overflows in Kelly networks. These generalizations are the subject of current research and they will be reported in the future.

The rest of the paper is organized as follows. In Section 2 we discuss basic notions of Jackson networks and provide a specific statement of our problems. Section 3 describes our simulation strategies and also summarizes our main results in Theorem 1. Section 4 discusses closed formulae and conditional sampling of the steady-state distribution of the Jackson network conditional on overflow sets. Section 5 provides a useful representation for the conditional distribution of the Jackson network given overflow in terms of the associated time-reversed stationary process. Section 6 further explores properties implied by such representation for the conditional distribution. Some extensions are discussed in Section 7 and finally, in Section 8, we discuss some numerical experiments.

2 Assumptions, Notation and Problem Statement

Let $Y = (Y_n : n \geq 0)$ be the embedded discrete time Markov chain corresponding to a Jackson network with d stations. The state space is then $\mathcal{S} = \mathcal{Z}_+^d$, where $\mathcal{Z}_+ = \{0, 1, 2, \dots\}$. The transition dynamics of a Jackson network are specified as

follows (see [25] p. 92). The arrival rates are given by the vector $\lambda = (\lambda_1, \dots, \lambda_d)^T$ and service rates are given by $\mu = (\mu_1, \dots, \mu_d)^T$. (By convention all of the vectors in this paper are taken to be column vectors and T denotes transposition.) A job that leaves station i joins station j with probability $Q(i, j)$ and it leaves the system with probability

$$Q(i, 0) \triangleq 1 - \sum_{j=1}^d Q(i, j).$$

The matrix $Q = \{Q(i, j) : 1 \leq i, j \leq d\}$ is called the routing matrix. We shall consider open Jackson networks, which satisfy the following conditions:

- i) Either $\lambda_i > 0$ or $\lambda_{j_1} Q(j_1, j_2) \dots Q(j_k, i) > 0$ for some j_1, \dots, j_k .
- ii) Either $Q(i, 0) > 0$ or $Q(i, j_1) \dots Q(j_k, 0) > 0$ for some j_1, \dots, j_k .

Assumptions i) and ii) are standard in the description of open Jackson networks. The first assumption simply says that any given station can eventually receive arriving customers to the system, either directly or after routing through the network. The second assumption indicates that a customer currently at any station can eventually leave the system either directly or via routing through the network.

The total flow coming to station i satisfies

$$\phi_i = \lambda_i + \sum_{j=1}^d \phi_j Q(j, i),$$

for all $i \in \{1, \dots, d\}$. Given conditions i) and ii), the previous system of equations has a unique solution (see [25] p. 93). The quantity $\rho_i = \phi_i / \mu_i \in (0, 1)$ is the so-called traffic intensity at station i .

If $Y_\infty = (Y_\infty(1), \dots, Y_\infty(d))^T$ follows the stationary distribution of Y , then Y_∞ has the product form

$$P(Y_\infty = (n_1, \dots, n_d)) = \prod_{j=1}^d P(Y_\infty(j) = n_j),$$

where $Y_\infty(j)$ is geometrically distributed; more precisely, $P(Y_\infty(j) = k) = (1 - \rho_j) \rho_j^k$ for $k \geq 0$ and $j = 1, \dots, d$ (see [25] p. 95).

Throughout our development we shall take advantage of the associated reversed Markov chain associated to Y , which we shall denote by $Y' = (Y'_n : n \geq 0)$. The chain Y' has a transition kernel $(K'(y, x) : y, x \in \mathcal{S})$ defined via

$$K'(y, x) = \pi(x) K(x, y) / \pi(y).$$

It turns out that Y' is also the embedded discrete time Markov chain corresponding to a stable Jackson network with the same stationary distribution as Y . The parameters

of the network corresponding to Y' are specified as follows (see [25] p. 95). The vectors of arrival and service rates $\lambda' = (\lambda'_1, \dots, \lambda'_d)^T$ and $\mu' = (\mu'_1, \dots, \mu'_d)^T$ satisfy

$$\lambda'_i = \phi_i Q(i, 0), \quad \mu'_i = \mu_i, \quad i = 1, \dots, d$$

and the corresponding routing matrix is given by $Q'(i, j) = \phi_j Q(j, i) / \phi_i$ for $i, j = 1, \dots, d$ and $Q'(i, 0) = \lambda_i / \phi_i$.

We are interested in estimating overflow probabilities during a busy period in any subset of the network. To be precise, let us define

$$A_N = \{y : v^T y \geq N\} \tag{1}$$

where v is a binary vector (i.e. $v \in \{0, 1\}^d$) which encodes a particular subset of the network. For any set $B \subseteq \mathcal{S}$ define the *first passage time to the set B* via

$$T_B = \inf\{k \geq 0 : Y_k \in B\}.$$

We also define σ_B as the *first return time to set B* . In particular,

$$\sigma_B = \inf\{k \geq 1 : Y_k \in B\}.$$

Note the distinction between return times (strictly positive) and first passage times (which can take zero value). Our goal is to efficiently estimate

$$u_N(x) = P_x(T_{A_N} < \sigma_{\{0\}})$$

and to sample the process $(Y_n : 0 \leq n \leq T_{A_N})$ given that $T_{A_N} < \sigma_{\{0\}}$. Throughout our development we assume that x is fixed and independent of N , but our results hold as long as $\|x\| = O(1)$ provided N is sufficiently large. In addition, as we shall explain, our methods can be extended in a straightforward way to the case of overflow in the union of subsets of the network. These and other extensions will be explained in detail in Section 7.

Before we proceed to describe our algorithms in the next subsection, we note that associated return and first passage times for the reversed process are defined using analogous notation. For instance, we define the first return time $\sigma'_B = \inf\{k \geq 1 : Y'_k \in B\}$ for $B \subseteq \mathcal{S}$.

3 General Strategy

First describe our algorithm for sampling from $(Y_n : 0 \leq n \leq T_{A_N})$ given that $T_{A_N} < \sigma_{\{0\}}$, assuming that $Y_0 = x$ and that $v^T x < N$. The justifications behind the algorithms are given later in the paper.

Algorithm A (Conditional Sampling Given an Overflow)

Step 1: Sample Y'_0 from $\pi(\cdot)$ conditional on $v^T Y'_0 = N$.

Step 2: If $Y'_0 = y$ is such that $P_y(\sigma'_{A_N} = 1) = 1$, then go back to Step 1. Otherwise, given Y'_0 simulate $(Y'_n : 0 \leq n \leq \sigma'_{\{0\}})$ conditional on the event that $\sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \sigma'_{A_N}$.

Step 3: Given the path $(Y'_n : 0 \leq n \leq \sigma'_{\{0\}})$ from Step 2, compute $\xi'_{\{x\}} = \max\{0 \leq k \leq \sigma'_{\{0\}} : Y'_k = x\}$. Let $Y_n = Y'_{\xi'_{\{x\}} - n}$ for $n \in \{0, 1, \dots, \xi'_{\{x\}}\}$, set $T_{A_N} = \xi'_{\{x\}}$ and output $(Y_n : 0 \leq n \leq T_{A_N})$.

As we will see, the previous algorithm can be executed in $O(N)$ function evaluations. The idea is to take advantage of the fact that Step 1 involves a static distribution (i.e. not a stochastic process). We will design a procedure to execute Step 1 in $O(1)$ function evaluations, the procedure is summarized in Proposition 2. Then, as we shall see, the conditioning in Step 2 will not typically involve a rare event and thus this step can be easily executed by running Y' in $O(N)$ function evaluations (because it takes $\Theta(N)$ steps in expectation for the process Y' to empty given an initial condition of order $O(N)$).

We now provide our algorithm for estimating $u_N(x)$, which takes advantage of the following identity which is proved in Proposition 3,

$$u_N(x) = \frac{P(v^T Y_\infty = N) P_\pi(\sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \sigma'_{A_N} | v^T Y'_0 = N)}{\pi(0) P_0(T_{\{x\}} < T_{A_N} \wedge \sigma_{\{0\}})}. \quad (2)$$

The corresponding algorithm for estimating $u_N(x)$ proceeds as follows.

Algorithm B (Estimating Overflow Probabilities)

Step 1: Estimate $P_\pi(v^T Y'_0 = N)$ via Monte Carlo using m i.i.d. replications of the estimator Z_N defined in (19) in Proposition 2 and call the estimate $\widehat{R}_0(m, N)$.

Step 2: Estimate the probability that $P_\pi(\sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \sigma'_{A_N} | v^T Y'_0 = N)$ via crude Monte Carlo using a suitably large number of, say, m i.i.d. replications (use Algorithm A to generate the Bernoullies) and call the estimate $\widehat{R}_1(m, N)$.

Step 3: Estimate $P_0(T_{\{x\}} < T_{A_N} \wedge \sigma_{\{0\}})$ via crude Monte Carlo using m i.i.d. replications and call the estimate $\widehat{R}_2(m, N)$.

Step 4: Using the estimators in Steps 1, 2 and 3 (obtained independently from each other) estimate $u_N(x)$ using the estimator

$$\widehat{u}_N(x, m) := \frac{\widehat{R}_0(m, N) \widehat{R}_1(m, N)}{\pi(0) \widehat{R}_2(m, N)}, \quad (3)$$

based on (2).

An alternative method to Algorithm B is to use importance sampling. That is, instead of estimating the probability in Step 2 by sampling Y' given that $v^T Y'_0 = N$

one might simply estimate $P_\pi(T'_{\{x\}} \leq T'_{\{0\}} < \sigma'_{A_N}; v^T Y'_0 = N)$ applying a suitable importance sampling scheme to Y'_0 . This variation is straightforward to implement in view of the fact that our approach to sampling Y'_0 in Step 1 of Algorithm A takes advantage of a carefully designed acceptance / rejection procedure. We then can use the proposal distribution as importance sampling distribution. Note that the resulting estimator in Step 4 is biased, unless $x = 0$, in which case the estimator becomes unbiased because $\widehat{R}_2(m, N) = 1$. Regardless, the estimator given by Algorithm B achieves a prescribed level of relative error for estimating $u_N(x)$ with $O(N)$ function evaluations as $N \rightarrow \infty$ (for fixed x).

Remark 1 *Actually, as we shall see in our proof of Lemma 2 in Section 6, if $x = 0$, an estimate with prescribed relative accuracy and $O(1)$ running time can be developed using a slight modification of Algorithm B. In particular, suppose that $x = 0$, and replace*

$$P_\pi(\sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \sigma'_{A_N} | v^T Y'_0 = N) = P_\pi(\sigma'_{\{0\}} < \sigma'_{A_N} | v^T Y'_0 = N) \quad (4)$$

in the numerator of (2) by

$$P_\pi(\sigma'_{\widetilde{A}_{N-\kappa}^c} < \sigma'_{A_N} | v^T Y'_0 = N) \quad (5)$$

where $\widetilde{A}_{N-\kappa} = \{y \in S : y(1) + \dots + y(d) \geq N - \kappa\}$ for some $\kappa > 0$. That is, replace (4) by the probability of reaching a population of size at most $N - \kappa$ before returning to A_N . It is not difficult to see that the relative error made by such replacement is of order $O(\exp(-\kappa\delta'))$ for some $\delta' > 0$. Our numerical experiments in the last section take advantage of this observation.

Before we develop the theory behind the validity and efficiency of Algorithms A and B it is useful to discuss at a high level why these algorithms have excellent complexity properties.

In order to obtain a fast execution of Algorithm A for large N it is crucial to show that Step 2 indeed can be performed with an expected computational cost of order $\Theta(N)$ as $N \rightarrow \infty$. In simple words, the cost of executing Step 2 involves understanding the probability that *the time reversed Markov chain Y' hits the origin, passing through state x , before returning to A_N given an initial position in A_N sampled according to π .*

First, as we will see, the set of states $y \in A_N$ for which it is possible to hit the origin, passing through state x , before coming back to a point in A_N has a probability under π that is of the same order as $P(Y_\infty \in A_N)$. More precisely, if we define $D_N := \{y \in A_N : P_y(\sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \sigma'_{A_N}) > 0\}$ we will show that there exists $\delta' > 0$ and $N_0 > 0$ such that if $N \geq N_0$,

$$\delta' P(Y_\infty \in A_N) \leq P(Y_\infty \in D_N). \quad (6)$$

In fact, it is not difficult to conclude that

$$D_N = \{y \in A_N : P_y(\sigma'_{A_N} = 1) < 1\} \subset \{y \in A_N : v^T y = N\}. \quad (7)$$

To see this note that if $y \in A_N$ satisfies $P_y(\sigma'_{A_N} = 1) < 1$ then there is a station encoded by the vector v with at least one customer in it, and with the property that it is possible for a customer, in the next step, to transition from such station to a station not encoded by v or to leave the network. Once we have $N-1$ customers in the stations encoded by v , then we can carefully move these customers one-by-one through the network to have them leave the system. Once a sufficiently large number of customers have left the whole system, because of the structure of Jackson networks it is clear that it is possible for the process Y' to visit x no later than 0. Observation (7) makes evident that D_N does not really depend on x . To gain a better understanding of D_N suppose for instance that Y is a two-node tandem network and that $v^T = (1, 1)$ (i.e. we look at the total population overflow); a transition diagram for both Y and Y' is given in Section 8. Then, $D_N = \{(y(1), y(2)) : y(1) \geq 1, y(1) + y(2) = N\}$.

We will also show that $\delta', c_* > 0$ can be chosen so that for every $y \in D_N$ with $\sum_{i=1}^d y(i) \leq N + c_*$ we have that

$$P_y(\sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \sigma'_{A_N}) > \delta'. \quad (8)$$

Inequalities (6) and (8) lie at the heart of the fast execution of Algorithm A. The fact that Algorithm A indeed provides samples from the conditional distribution of $(Y_n : 0 \leq n \leq T_{A_N})$ given that $T_{A_N} < \sigma_{\{0\}}$ and $Y_0 = x$ follows, as we will see, from (2).

Now we explain what is involved in the analysis behind (6) and (8). For simplicity suppose that $v = (1, \dots, 1)^T$, that is we consider the case of total population overflow at level N . Inequality (6) can be shown observing that each element in D_N must have at least a customer in the “exterior” nodes of the network described by Y' (i.e. a customer that can leave the whole network in the next step, for instance, if Y' encodes a two node tandem network the downstream station must start with at least one customer). Since there are at most d stations, inequality (6) then can be shown to follow (as we do in Proposition 4) from standard properties of sums of geometric and negative binomial distributions.

The most interesting part of the analysis involves (8). This inequality is shown first by starting with a configuration $y \in D_N$ and then intersecting with the event that a sufficiently large number of customers, say κ , leave the network before any new customer arrives to the network. Such an event clearly has a probability that is bounded away from zero. Then, the total number in system is process with negative drift when observed at sufficiently large time steps, provided that we start with an initial population of size $N - \kappa$. We then show using a standard large deviations consideration that we can make the probability of returning to the set A_N arbitrarily

close to zero by picking κ large. In other words, if we start from any configuration with $N - \kappa$ customers in the network, then with high probability (depending on κ) we will be near the origin before having N customers in the network. In particular, with high probability we will reach a suitable compact set near the origin, containing x , before reaching set A_N . However, once we reach a suitable compact set, it is clear that the probability of reaching state x not later than reaching state 0 is bounded away from zero (just intersect with the event that arrivals and departures occur in a suitable specific order throughout the whole network).

Summarizing then, the *important features behind the efficiency* of Algorithms A and B are that the steady-state distribution is tractable when conditioned on the set D_N and that the time reversed process will naturally evolve towards a compact set containing both x and 0. The compact set must also be such that it is possible to visit x no later than 0 starting from any attainable point (other than 0) in the compact set. These properties are in place for the case of Jackson networks, as we will show in the rest of the paper. We believe, however, that these properties are also applicable in significant generality and we plan to investigate this in future work.

The following formal statement summarizes the most important properties of Algorithms A and B. The proof links together all the future developments in the paper.

Theorem 1

1. *Algorithm A provides unbiased samples of the random variables $(Y_n : 0 \leq n \leq T_{A_N})$ given that $Y_0 = x$ and $T_{A_N} < \sigma_{\{0\}}$. Moreover, the expected number of random variables required to run Algorithm A is of order $O(N)$ as $N \rightarrow \infty$ (assuming that x is fixed or $x = O(1)$ as $N \rightarrow \infty$).*
2. *If $x = 0$, the estimator given by Algorithm B is unbiased and strongly efficient in the sense that there exists a constant $c \in (0, \infty)$ such that for all $m \geq 1$, $E\hat{u}_N(0, m)^2 \leq cu_N(0, m)^2$.*
3. *Regardless of the value of x (as long as $x = O(1)$ as $N \rightarrow \infty$), Algorithm B takes $O(N)$ function evaluations to be executed. Moreover, if we choose $m = O(\varepsilon^{-2}\delta^{-1})$ we can guarantee that*

$$P(|\hat{u}_N(0, m) - u_N(0)| \geq \varepsilon u_N(0)) \leq \delta. \tag{9}$$

Consequently, Algorithm B provides an estimate with prescribed relative error with $O(N)$ function evaluation. If in addition $x = 0$, Step 2 of Algorithm B can be implemented in $O(\log(1/\varepsilon))$ function evaluations with an overall error of 2ε in (9), therefore providing an estimate with prescribed relative error in $O(\varepsilon^{-2} \log(1/\varepsilon)\delta^{-1})$ function evaluations.

Proof. About part 1), first Proposition 3 guarantees that indeed Algorithm A provides unbiased samples from the target conditional distribution. The procedure to simulate Y_∞ given that $Y_\infty \in A_N$ in $O(1)$ function evaluations is based on acceptance / rejection and it is summarized in Proposition 2. The calculation of the running time is standard, the results of [9] and [10] yield that the expected hitting time to the origin starting from a position with $O(N)$ total population involves $O(N)$ expected steps.

Regarding part 2), the fact that the estimator of Algorithm B is unbiased follows from representation (2), which in turn is proved in Proposition 3. Strong efficiency follows because of three reasons. First, Proposition 2 ensures that the estimator $R_0(m, N)$ is strongly efficient to estimate $P(v^T Y_\infty = N)$. In addition, we need to verify that: i) that $u_N(x) = \Theta(P(Y_\infty \in A_N))$ and ii) that the probability in Step 2 of Algorithm B is bounded away from zero. Part i) is proved in Corollary 2 and in Proposition 4. Part ii) is also established in Proposition 4.

Part 3) indicates that Algorithm B can be executed in $O(N)$ function evaluations (given that the value of m is selected). This running time follows from the results of [9] and [10] as in part 1). The error bounds involving (9) is quite standard, follows from strong efficiency and Chebyshev's inequality. The observation given in Remark 1 above about the error of order $O(\exp(-\kappa\delta'))$ introduced by the replacement in (5) can be verified as in the proof of Lemma 2. This remark then implies that Step 2 can indeed be implemented with $O(\log(1/\varepsilon))$ function evaluations incurring in a bias of size ε . The rest of the analysis follows from elementary manipulations and Chebyshev's inequality, see for instance Theorem 2, part 2) in [5]. ■

4 An Efficient Sampler for Conditional Sums of Geometric Random Variables

In this section we explain how to simulate Y_∞ conditional on $v^T Y_\infty = N$ in $O(1)$ function evaluations. Assume that $v^T Y_\infty = Y_\infty(i_1) + \dots + Y_\infty(i_s)$ for some $s \geq 1$ (i.e. we are interested in the overflow for the subset of the network corresponding to the stations $\{i_1, \dots, i_s\}$). We recall that the $Y_\infty(i)$'s are independent geometric random variables and re-express $v^T Y_\infty$ using the fact that the sum of i.i.d. geometric random variables is negative binomial. To use this fact, we will group the stations within the set $\{i_1, \dots, i_s\}$ that share the same value of their traffic intensity. We proceed to do this grouping next.

Suppose that out of the set $\{i_1, \dots, i_s\}$ of stations, there is a subset say $\{i_1^{(1)}, \dots, i_{m_1}^{(1)}\}$ of size m_1 whose elements share the same traffic intensity equal to ρ_{t_1} , a subset $\{i_1^{(2)}, \dots, i_{m_2}^{(2)}\}$ of size m_2 with stations that share the same traffic intensity equal to $\rho_{t_2} < \rho_{t_1}$ and so forth, up to a subset of stations $\{i_1^{(k)}, \dots, i_{m_k}^{(k)}\}$ which share the same traffic intensity ρ_{t_k} . Assume that $\rho_{t_k} < \dots < \rho_{t_2} < \rho_{t_1}$ and that $m_1 + \dots + m_k = s$

with $m_i > 0$ for $i \in \{1, \dots, k\}$. Then we define

$$M_j = Y_\infty(i_1^{(j)}) + \dots + Y_\infty(i_{m_j}^{(j)}),$$

for $j = 1, \dots, k$. The random variable M_j is negative binomial with parameters m_j and $p_j = (1 - \rho_{t_j})$ (denoted $M_j =_d NBin(m_j, p_j)$) and it satisfies

$$P(M_j = l) = \binom{l + m_j - 1}{m_j - 1} p_j^{m_j} (1 - p_j)^l, \quad l = 0, 1, 2, \dots$$

Observe that the M_j 's are independent for $j = 1, \dots, k$ and that $v^T Y_\infty = M_1 + \dots + M_k$.

The next proposition provides basic asymptotic properties for the distribution of the sum of independent negative binomial distributions. Some of these results have very likely appeared in the literature but we were not able to locate a reference for it, so we provide a proof at the end of the section. The exact distribution for the case of sums of geometric random variables with different parameters can be obtained explicitly and it is parallel to that of so-called hypo-exponential distributions (see [26] p. 309).

Proposition 1 *Suppose that M_1, M_2, \dots, M_k are independent random variables such that $M_j =_d NBin(m_j, p_j)$ and $p_1 < p_2 < \dots < p_k$, then there is an explicit constant w^* such that*

$$P(M_1 + \dots + M_k = N) = w^* P(M_1 = N) (1 + o(1)) \quad (10)$$

as $N \rightarrow \infty$.

Remark 2 *Although the value of w^* is not of interest to us here, a precise probabilistic representation is given in the proof of this result.*

Now we shall explain how to efficiently sample Y_∞ given that $v^T Y_\infty = N$.

Algorithm C (Algorithm for Sampling I.I.D. Geometrics Conditioned on Sums)

We first assume that we can simulate M_1, \dots, M_k given that $v^T Y_\infty = M_1 + \dots + M_k = N$.

Step 1: Given M_j , sample $Y_\infty(i_1^{(j)}), \dots, Y_\infty(i_{m_j}^{(j)})$ as follows. If $m_j = 1$ set $Y_\infty(i_{m_j}^{(j)}) = M_j$. If $m_j > 1$ then sample $m_j - 1$ elements, $r_1^{(j)} < \dots < r_{m_j-1}^{(j)}$, of the set $\{1, \dots, m_j + M_j - 1\}$ without replacement.

Step 2: Let $r_0^{(j)} = 0$ and $r_{m_j}^{(j)} = M_j + m_j$ and simply put $Y_\infty(i_l^{(j)}) = r_l^{(j)} - r_{l-1}^{(j)} - 1$ for $l = 1, \dots, m_j$. Apply

Step 3: Apply Steps 1 and 2 for all $j = 1, \dots, k$.

Step 4: Simulate $Y_\infty(i)$ for $i \neq i_l^{(j)}$ for $l = 1, \dots, m_j$ and $j = 1, \dots, k$ independently, each from its nominal distribution (i.e. $P(Y_\infty(i) = z) = (1 - \rho_i)\rho_i^z$ for $z = 0, 1, \dots$).

We observe that given the values of M_1, \dots, M_k , which are assumed to be known in the previous procedure, the simulation of Y_∞ can be done in $O(1)$ function evaluations as a function of N .

Now, let us set the stage to explain how to simulate M_1, \dots, M_k given that $M_1 + \dots + M_k = N$. In order to understand our strategy, first let us assume $k = 2$. We now provide an argument that allows us to understand the conditional distribution of M_1 given that $M_1 + M_2 = N$. Note that then we define

$$\begin{aligned}
& P(M_1 = N - l | M_1 + M_2 = N) \\
&= \frac{P(M_1 = N - l) P(M_2 = l)}{P(M_1 + M_2 = N)} \\
&= \binom{m_1 + N - l - 1}{m_1 - 1} \binom{m_2 + l - 1}{m_2 - 1} \frac{p_1^{m_1} q_1^{N-l} p_2^{m_2} q_2^l}{P(M_1 + M_2 = N)} \\
&\longrightarrow \binom{m_2 + l - 1}{m_2 - 1} (q_2/q_1)^l (1 - q_2/q_1)^{m_2} = P(M' = l) \tag{11}
\end{aligned}$$

as $N \rightarrow \infty$, where $M' \stackrel{d}{=} NBin(m_2, 1 - q_2/q_1)$. In other words, conditional on $M_1 + M_2 = N$, we have the approximation in distribution given by $M_1 \approx N - M'$. This observation is the basis for our simulation method to generate M_1, \dots, M_k given that $M_1 + \dots + M_k = N$.

For the case $k \geq 2$ we first will apply an exponential change-of-measure that effectively reduces the sampling problem back to the case $k = 2$ and, on top of this exponential change-of-measure, we will use the insight obtained from (11). We shall pursue this program now.

Again we suppose that $p_1 < p_2 < \dots < p_k$ and let $q_j = 1 - p_j$ for $1 \leq j \leq k$, thus $q_1 > q_2 > \dots > q_k$. For $j \geq 3$ define $\eta_j = \log(q_2/q_j) > 0$ and put $\eta_j = 0$ for $1 \leq j \leq 2$; so that $E \exp(\eta_j M_j) = (p_j/p_2)^{m_j}$ for $j \geq 3$. Consider the exponentially tilted distribution associated to η_j 's, namely,

$$\begin{aligned}
& \widehat{P}(M_1 = l_1, \dots, M_k = l_k) \\
&= \prod_{j=1}^k \left(\frac{\exp(\eta_j l_j)}{E \exp(\eta_j M_j)} P(M_j = l_j) \right) \\
&= P(M_1 = l_1) P(M_2 = l_2) \prod_{j=3}^k \left(\frac{p_2^{m_j} \exp(\eta_j l_j)}{p_j^{m_j}} P(M_j = l_j) \right)
\end{aligned}$$

Under $\widehat{P}(\cdot)$ the M_j 's are independent and the marginal distributions of M_1 and M_2 remain unchanged, but for $j \geq 3$, $M_j \stackrel{d}{=} NBin(m_j, p_2)$. Therefore, if we let

$$M = M_2 + \dots + M_k,$$

then under $\widehat{P}(\cdot)$ we have that $M =_d N \text{Bin}(m, p_2)$, where $m = m_2 + \dots + m_k$.

If $l_1 + \dots + l_k = N$, with $l_j \geq 0$ for $1 \leq j \leq k$ then

$$\begin{aligned}
& P(M_1 = l_1, \dots, M_k = l_k | M_1 + \dots + M_k = N) \\
&= \frac{P(M_1 = l_1, \dots, M_k = l_k)}{P(M_1 + \dots + M_k = N)} \\
&= \frac{\widehat{P}(M_1 = l_1, \dots, M_k = l_k) \exp(-\sum_{j=3}^k \eta_j l_j)}{P(M_1 + \dots + M_k = N)} \prod_{j=3}^k \left(\frac{p_j}{p_2}\right)^{m_j} \\
&= \frac{\widehat{P}(M_1 = l_1, \dots, M_k = l_k | M_1 + M = N) \exp(-\sum_{j=3}^k \eta_j l_j)}{\widehat{P}(M_1 + M = N)^{-1} P(M_1 + \dots + M_k = N)} \prod_{j=3}^k \left(\frac{p_j}{p_2}\right)^{m_j} \quad (12) \\
&= \frac{\widehat{P}(M_1 = N)}{\widehat{P}(M_1 + M = N)} \times \frac{\binom{m_1+l_1-1}{m_1-1}}{\binom{m_1+N-1}{m_1-1}} \times \frac{p_2^{m_2} \widehat{P}(M'' \leq N) \exp(-\sum_{j=3}^k \eta_j l_j)}{\widehat{P}(M_1 + M = N)^{-1} P(M_1 + \dots + M_k = N) (1 - q_2/q_1)^m} \prod_{j=3}^k p_j^{m_j} \quad (13)
\end{aligned}$$

Now observe that whenever $l_1 + \dots + l_k = N$, with $l_j \geq 0$ for $1 \leq j \leq k$ our observation in (11) yields for $0 \leq l \leq N$

$$\widehat{P}(M_1 = N - l | M_1 + M = N) \approx \widehat{P}(M'' = l)$$

where $M'' =_d N \text{Bin}(m, 1 - q_2/q_1)$ under $\widehat{P}(\cdot)$. We will apply importance sampling to the distribution of M_1 given that $M_1 + M = N$ using that of $N - M''$. Note that if $l = l_2 + \dots + l_k$ and $l_1 = N - l$ then

$$\begin{aligned}
& \widehat{P}(M_1 = N - l, \dots, M_k = l_k | M_1 + M = N) \\
&= \frac{\widehat{P}(M_1 = N - l) \widehat{P}(M = l)}{\widehat{P}(M_1 + M = N)} \times \frac{\widehat{P}(M_2 = l_2) \dots \widehat{P}(M_k = l_k)}{\widehat{P}(M = l)} \\
&= \widehat{P}(M_1 = N - l | M_1 + M = N) \times \widehat{P}(M_2 = l_2, \dots, M_k = l_k | M = l). \quad (14)
\end{aligned}$$

On the other hand,

$$\begin{aligned}
& \widehat{P}(M_1 = N - l | M_1 + M = N) \\
&= \frac{1}{\widehat{P}(M_1 + M = N)} \binom{m_1 + N - l - 1}{m_1 - 1} \binom{m + l - 1}{m - 1} p_1^{m_1} q_1^{N-l} p_2^m q_2^l \\
&= \frac{p_1^{m_1} q_1^N \binom{m_1+N-1}{m_1-1}}{\widehat{P}(M_1 + M = N)} \times \frac{\binom{m_1+N-l-1}{m_1-1}}{\binom{m_1+N-1}{m_1-1}} \times \frac{\binom{m+l-1}{m-1} p_2^m \left(\frac{q_2}{q_1}\right)^l}{\widehat{P}(M'' \leq N)} \widehat{P}(M'' \leq N) \\
&= \frac{\widehat{P}(M_1 = N)}{\widehat{P}(M_1 + M = N)} \times \frac{\binom{m_1+N-l-1}{m_1-1}}{\binom{m_1+N-1}{m_1-1}} \times \frac{p_2^m \widehat{P}(M'' \leq N)}{(1 - q_2/q_1)^m} \widehat{P}(M'' = l | M'' \leq N). \quad (15)
\end{aligned}$$

We now are ready to define a distribution that allows to sample according to M_1, \dots, M_k given that $M_1 + \dots + M_k = N$ (which corresponds, together with Algorithm C, to Step 1 of Algorithm A). Let $l_1, l_2, \dots, l_k \geq 0$ set $l = l_2 + \dots + l_k$ and assume that $l_1 = N - l$. Define a probability measure $P_N^{\&}(\cdot)$ under which

$$\begin{aligned} & P_N^{\&}(M_1 = l_1, M_1 = l_2, \dots, M_k = l_k) \\ &= \widehat{P}(M'' = N - l_1 | M'' \leq N) \times \widehat{P}(M_2 = l_2, \dots, M_k = l_k | M = l). \end{aligned} \quad (16)$$

We will explain how to sample from $\widehat{P}(\cdot)$ in Steps 1 and 2 in Algorithm D below.

Representations (12), (14), and (15) allow us to obtain the following relation (recall that $l = N - l_1$)

$$\begin{aligned} & \frac{P(M_1 = l_1, M_2 = l_2, \dots, M_k = l_k | M_1 + \dots + M_k = N)}{P_N^{\&}(M_1 = l_1, M_1 = l_2, \dots, M_k = l_k)} \\ &= \frac{\widehat{P}(M_1 = N)}{P(M_1 + \dots + M_k = N)} \times p_2^{m_2} \times \prod_{j=3}^k p_j^{m_j} \times \frac{\widehat{P}(M'' \leq N)}{(1 - q_2/q_1)^m} \\ & \times \frac{\binom{m_1 + N - l - 1}{m_1 - 1}}{\binom{m_1 + N - 1}{m_1 - 1}} \times \exp\left(-\sum_{j=3}^k \eta_j l_j\right) \end{aligned} \quad (17)$$

Ratio (17) is the key element in the following acceptance / rejection procedure.

Algorithm D (Sampling Independent Negative Binomials Conditioned on Sums)

Step 1: First simulate $M'' =_d NBin(m, 1 - q_2/q_1)$ conditional on $M'' \leq N$ and set $M'' = l$. Then let $M_1 = N - l$.

Step 2: Simulate (M_2, \dots, M_k) with distribution $\widehat{P}(M_2 = \cdot, \dots, M_k = \cdot | M_2 + \dots + M_k = l)$. Let $M_j = l_j$ for $2 \leq j \leq k$. The sampling procedure can be done as follows. Consider i.i.d. random variables $(G_{i,j} : 2 \leq i \leq k, 1 \leq j \leq m_i)$ with distribution $Geo(p_2)$. Note that under $\widehat{P}(\cdot)$ we can represent $M_i = \sum_{j=1}^{m_i} G_{i,j}$ for $i \leq k$, and $M = M_2 + \dots + M_k = \sum_{i=2}^k \sum_{j=1}^{m_i} G_{i,j}$. So, sampling the $G_{i,j}$'s is done using a procedure that is completely analogous to Step 1 in Algorithm C.

Step 3: Generate a uniform random variable U independent of the M_j 's. If

$$\binom{m_1 + N - 1}{m_1 - 1} U \leq \binom{m_1 + N - l - 1}{m_1 - 1} \exp\left(-\sum_{j=3}^k \eta_j l_j\right),$$

then output (M_1, \dots, M_k) . Otherwise, go to Step 1.

The next result summarizes the considerations leading to Algorithm D and its complexity properties.

Proposition 2 *Let*

$$\kappa(N) = \frac{\widehat{P}(M_1 = N)}{P(M_1 + \dots + M_k = N)} \times p_2^{m_2} \times \prod_{j=3}^k p_j^{m_j} \times \frac{1}{(1 - q_2/q_1)^m}. \quad (18)$$

The expected number of proposals required to obtain an acceptance (i.e. the number of times one cycles through Step 1 in Algorithm D) is equal to $\kappa(N) = O(1)$ as $N \rightarrow \infty$. Moreover, we have that if we consider the importance sampling estimator

$$Z_N = \frac{dP}{dP_N^{\&}}(M_1, \dots, M_k), \quad (19)$$

under $P_N^{\&}(\cdot)$ then, using $E_N^{\&}(\cdot)$ to denote the expectation operator associated to $P_N^{\&}(\cdot)$, we have that

$$E_N^{\&}(Z_N^2) \leq P(M_1 + \dots + M_k = N)^2 \kappa(N)^2. \quad (20)$$

Equivalently, Z_N is strongly efficient as $N \rightarrow \infty$.

Proof. Steps 1 and 2 simulate (M_1, \dots, M_k) from the distribution $P_N^{\&}(\cdot)$. So, if $M_1 = N - l$ and $M_j = l_j$ for $j \geq 2$ we have, in view of (17) and the definition of $\kappa(N)$ that

$$\begin{aligned} & \frac{P(M_1 = l_1, M_2 = l_2, \dots, M_k = l_k | M_1 + \dots + M_k = N)}{P_N^{\&}(M_1 = l_1, M_2 = l_2, \dots, M_k = l_k)} \\ &= \kappa(N) \frac{\binom{m_1 + N - l - 1}{m_1 - 1}}{\binom{m_1 + N - 1}{m_1 - 1}} \times \exp\left(-\sum_{j=3}^k \eta_j l_j\right). \end{aligned} \quad (21)$$

On the other hand we have that

$$\binom{m_1 + N - l - 1}{m_1 - 1} \leq \binom{m_1 + N - 1}{m_1 - 1} \quad (22)$$

for any $0 \leq l \leq N$; this follows easily noting that the left hand side is interpreted as the number of subsets of size $m_1 - 1$ out of a set of size $m_1 + N - l - 1$. Since $l \geq 0$ this number is smaller than the number of subsets of size $m_1 - 1$ represented by the right hand side, which are chosen out of a set of size $m_1 - N + 1$. Therefore, combining (22) with (21) and using the fact that $\eta_j \geq 0$ we obtain that

$$\frac{P(M_1 = l_1, M_2 = l_2, \dots, M_k = l_k | M_1 + \dots + M_k = N)}{P_N^{\&}(M_1 = l_1, M_2 = l_2, \dots, M_k = l_k)} \leq \kappa(N). \quad (23)$$

The fact that $\kappa(N)$ is the expected number of proposals to obtain an acceptance is a basic fact of acceptance / rejection algorithms (see [3]). The strong efficiency of Z_N also follows immediately from inequality (23). The fact that $\kappa(N) = O(1)$ follows from Proposition 1. ■

We finish the section with the proof of Proposition 1.

Proof of Proposition 1. We use a change-of-measure based on $P_N^{\&}(\cdot)$, defined in (16). The corresponding likelihood ratio is obtained in (17). Using the definition of $\kappa(N)$ in (18) we obtain

$$\begin{aligned} P(M_1 + \dots + M_k = N) &= E_N^{\&}(Z_N) \\ &= \widehat{P}(M_1 = N) \times p_2^{m_2} \times \prod_{j=3}^k p_j^{m_j} \times \frac{\widehat{P}(M'' \leq N)}{(1 - q_2/q_1)^m} \\ &\quad \times \widehat{E} \left(\frac{\binom{m_1+N-M''-1}{m_1-1}}{\binom{m_1+N-1}{m_1-1}} \times \exp\left(-\sum_{j=3}^k \eta_j M_j\right) \middle| M'' \leq N, M = M'' \right), \end{aligned}$$

where $M = M_2 + \dots + M_k$, $M'' =_d N \text{Bin}(m, 1 - q_2/q_1)$ and M'' is independent of the M_j 's for $j \geq 2$ under $\widehat{P}(\cdot)$. By the bounded convergence theorem (which applies in view of (22) and the fact that $\eta_j \geq 0$) we have that

$$\begin{aligned} &\widehat{E} \left(\frac{\binom{m_1+N-M''-1}{m_1-1}}{\binom{m_1+N-1}{m_1-1}} \times \exp\left(-\sum_{j=3}^k \eta_j M_j\right) \middle| M'' \leq N, M = M'' \right) \\ &\longrightarrow \widehat{E} \left(\exp\left(-\sum_{j=3}^k \eta_j M_j\right) \middle| M = M'' \right) \end{aligned}$$

as $N \rightarrow \infty$. Recall that $\widehat{P}(M_1 = N) = P(M_1 = N)$, thus we conclude

$$P(M_1 + \dots + M_k = N) = \frac{P(M_1 = N)}{(1 - q_2/q_1)^m} p_2^{m_2} \prod_{j=3}^k p_j^{m_j} \widehat{E} \left(\exp\left(-\sum_{j=3}^k \eta_j M_j\right) \middle| M = M'' \right) (1 + o(1))$$

as $N \rightarrow \infty$ as claimed. ■

5 First Passage Times and Reversibility

As we indicated in Section 3, an idea that lies at the center of our algorithms is the representation based on time reversibility given in (2). The objective of this section is to justify such representation and to verify the validity of Algorithm A, in the sense that it indeed provides an unbiased sample of $(Y_n : 0 \leq n \leq T_{A_N})$ given that $T_{A_N} < \sigma_{\{0\}}$, assuming that $Y_0 = x$. This verification is a consequence of Proposition 3 below.

Recall the following definitions. Given any two distinct elements $x, y \in \mathcal{S}$, define $T_{\{x\}} = \inf\{n \geq 0 : Y_n = x\}$ and $\sigma_{\{y\}} = \inf\{n \geq 1 : Y_n = y\}$. Note from the definitions that $P_x(T_{\{x\}} < \sigma_{\{x\}}) = 1$. Finally, we define $T'_{\{0\}} = \inf\{n \geq 0 : Y'_n = 0\}$, $\sigma'_{A_N} = \inf\{n \geq 1 : Y'_n \in A_N\}$ and $\sigma'_{\{x\}} = \inf\{n \geq 1 : Y'_n = x\}$.

Proposition 3 Suppose that $v^T x < N$, then

$$\begin{aligned} P_x(T_{A_N} < \sigma_{\{0\}}) &= \frac{P_\pi(Y'_0 \in A_N, \sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \sigma'_{A_N})}{\pi(0) P_0(T_{\{x\}} < T_{A_N} \wedge \sigma_{\{0\}})} \\ &= \frac{P_\pi(v^T Y'_0 = N, \sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \sigma'_{A_N})}{\pi(0) P_0(T_{\{x\}} < T_{A_N} \wedge \sigma_{\{0\}})}. \end{aligned} \quad (24)$$

Moreover,

$$\begin{aligned} P_0 \left[(Y_0, \dots, Y_{T_{A_N}}) \in \cdot \mid T_{\{x\}} < T_{A_N} < \sigma_{\{0\}} \right] \\ = P_\pi \left[(Y'_{\sigma'_{\{0\}}}, \dots, Y'_0) \in \cdot \mid v^T Y'_0 = N, \sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \sigma'_{A_N} \right]. \end{aligned}$$

In particular, since by the strong Markov property the law of $(Y_{T_{\{x\}}}, \dots, Y_{T_{A_N}})$ given $T_{\{x\}} < T_{A_N} < \sigma_{\{0\}}$ is the same as the law of $(Y_0, \dots, Y_{T_{A_N}})$ given $Y_0 = x$ and $T_{A_N} < \sigma_{\{0\}}$, we conclude that

$$\begin{aligned} P_x \left[(Y_0, \dots, Y_{T_N}) \in \cdot \mid T_{A_N} < \sigma_{\{0\}} \right] \\ = P_\pi \left[(Y'_{\xi'_{\{x\}}}, \dots, Y'_0) \in \cdot \mid v^T Y'_0 = N, \sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \sigma'_{A_N} \right], \end{aligned}$$

where $\xi'_{\{x\}} = \max\{0 \leq k \leq \sigma'_{\{0\}} : Y'_k = x\}$.

Proof. Consider any path $Y_0 = 0, Y_1 = y_1, \dots, Y_k = x, \dots, Y_l = y_l$ for which $T_{\{x\}} < T_{A_N} < \sigma_{\{0\}}$ with $T_{\{x\}} = k \geq 0$ and $T_{A_N} = l > k$ (note that $l > k$ is a consequence of assuming that $v^T x < N$). Note that $P_0[Y_1 = y_1, \dots, Y_k = x, \dots, Y_l = y_l] > 0$ implies that $v^T y_l = N$ and

$$\begin{aligned} P_0[Y_1 = y_1, \dots, Y_k = x, \dots, Y_l = y_l] \\ = K(0, y_1) \times \dots \times K(y_{l-1}, y_l) \\ = \frac{K(0, y_1)}{K'(y_1, 0)} \times \dots \times \frac{K(y_{l-1}, y_l)}{K'(y_l, y_{l-1})} K'(y_l, y_{l-1}) \times \dots \times K'(y_1, 0) \\ = \frac{\pi(y_l)}{\pi(0)} K'(y_l, y_{l-1}) \times \dots \times K'(y_1, 0). \end{aligned} \quad (25)$$

We let $y'_j = y_{l-j}$ for $j = 0, 1, \dots, l$ and conclude that

$$\begin{aligned} P_0[Y_1 = y_1, \dots, Y_k = x, \dots, Y_l = y_l] \\ = \frac{\pi(y'_0)}{\pi(0)} K'(y'_0, y'_1) \times \dots \times K'(y'_{l-1}, 0). \end{aligned} \quad (26)$$

Observe that the path $Y'_0 = y'_0, \dots, Y'_l = y'_l = 0$ satisfies $\sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \sigma'_{A_N}$ with $v^T y'_0 = N$. Adding over y_1, \dots, y_l we obtain

$$\pi(0) P_0(T_{\{x\}} < T_{A_N} < \sigma_{\{0\}}) = P_\pi[v^T Y'_0 = N, \sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \sigma'_{A_N}]. \quad (27)$$

Because the Y is irreducible and positive recurrent we have that $P_0(T_{\{x\}} < \infty) = 1$ and, therefore, the strong Markov property yields

$$P_0(T_{\{x\}} < T_{A_N} < \sigma_{\{0\}}) = P_0(T_{\{x\}} < T_{A_N} \wedge \sigma_{\{0\}})P_x(T_{A_N} < \sigma_{\{0\}}).$$

We then conclude that

$$\begin{aligned} P_x(T_{A_N} < \sigma_{\{0\}}) &= \frac{P_\pi[v^T Y'_0 = N, \sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \sigma'_{A_N}]}{\pi(0) P_0(T_{\{x\}} < T_{A_N} \wedge \sigma_{\{0\}})} \\ &= \frac{E_\pi[I(v^T Y'_0 = N) P_{Y'_0}(\sigma'_{\{x\}} \leq \sigma'_{\{0\}} < T'_{A_N})]}{\pi(0) P_0(T_{\{x\}} < T_{A_N} \wedge \sigma_{\{0\}})}, \end{aligned}$$

which yields the first part of the proposition. For the second part of the proposition just note that (25) and (26) together with identity (27) yield

$$\begin{aligned} &\frac{P_0[Y_1 = y_1, \dots, Y_l = y_l] I(T_{\{x\}} < T_{A_N} < \sigma_{\{0\}}, T_{A_N} = l)}{P_x(T_{A_N} < \sigma_{\{0\}}) P_0(T_{\{x\}} < T_{A_N} \wedge \sigma_{\{0\}})} \\ &= \frac{P_{y'_0}[Y'_{l-1} = y_1, \dots, Y'_0 = y_l] \pi(y_l) I(\sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \sigma'_{A_N}, \sigma'_{\{0\}} = l)}{\pi(0) P_x(T_{A_N} < \sigma_{\{0\}}) P_0(\sigma_{\{x\}} < T_{A_N} \wedge T_{\{0\}})} \\ &= \frac{P_{y'_0}[Y'_{l-1} = y'_{l-1}, \dots, Y'_0 = y'_0] \pi(y'_0) I(\sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \sigma'_{A_N}, \sigma'_{\{0\}} = l)}{P_\pi(v^T Y'_0 = N; \sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \sigma'_{A_N})}. \end{aligned}$$

So, we obtain that the law of $(Y_n : 0 \leq n \leq T_N)$ given $T_{\{x\}} < T_{A_N} < \sigma_{\{0\}}$ coincides with that of $(Y'_{\sigma'_{\{0\}}-n} : 0 \leq n \leq \sigma'_{\{0\}})$ given that $v^T Y'_0 = N$ is drawn from $\pi(\cdot)$ and also that $\sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \sigma'_{A_N}$. The last part of the proposition follows, as indicated in the statement itself, from the strong Markov property and by interpreting the meaning of $\sigma_{\{x\}}$ in terms of the time reversed process. ■

Using the previous proposition we can easily obtain as a corollary an upper for $u_N(x)$ in terms of $\pi(\cdot)$

Corollary 2 *Given $x \in \mathcal{Z}_+^d$, there exists a constant $c_0 \in (0, \infty)$ such that*

$$u_N(x) \leq c_0 P_\pi(v^T Y'_0 = N). \quad (28)$$

Proof. Clearly we have from Proposition 3 that

$$u_N(x) \leq \frac{P_\pi(v^T Y'_0 = N)}{\pi(0) P_0(T_{\{x\}} < T_{A_N} \wedge \sigma_{\{0\}})}.$$

Since $T_{A_N} \nearrow \infty$ we have (by the monotone convergence theorem) that

$$P_0(T_{\{x\}} < T_{A_N} \wedge \sigma_{\{0\}}) \longrightarrow P_0(T_{\{x\}} < \sigma_{\{0\}}) > 0$$

(the last inequality follows from irreducibility and positive recurrence). Inequality (28) follows easily from this observation. ■

6 Time-reversed Paths in Jackson Networks Starting from Overflow Sets

We shall use $\|x\|$ to denote the l_1 norm, that is, the sum of the absolute value of the components of the vector x . If $x \in \mathcal{Z}_+^d$ then $\|x\|$ is just the total population of the network at state x .

The main result of this section is given next. Its proof is given at the end of the section. Note, we use $\delta', \delta_0, \delta_1, \dots$ and $c, c', c_*, c_0, c_1, \dots$ as generic constants (independent of N) whose value can even change from line to line.

Proposition 4 *Given $x \in \mathcal{Z}_+^d$, $v^T x < N$,*

i) There is $\delta', N_0 > 0$ such that

$$\delta' < P_\pi(\sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \sigma'_{A_N} | v^T Y'_0 = N)$$

for all $N \geq N_0$.

ii) In addition, there exists a constant $c_1 \in (0, \infty)$ such that

$$u_N(x) \geq c_1 P_\pi(v^T Y'_0 = N).$$

The first part of the previous proposition is crucial to justify the computational complexity of Algorithm A, and the second part, combined with Corollary 2, is fundamental in showing the strong efficiency properties of the estimator given by Algorithm B.

In order to provide the proof of Proposition 4, we need auxiliary results which follow easily from the available literature.

Lemma 1

i) Suppose that $Y'_0 := y'_n/n$ converges to y' as $n \rightarrow \infty$. Then, $Y'_{\lfloor n \rfloor}/n \rightarrow y'(\cdot)$ in probability under the uniform norm on $[0, 1]$. The system $y'(\cdot)$ is piecewise linear. Moreover, there exists $\delta' > 0$ and $c', t_0 \in (0, \infty)$ such that $\|y'(t)\| - \|y'(0)\| < -\delta'$ uniformly over $\|y'(t)\| \geq c'$ and $t \geq t_0$.

ii) There exists m' and δ' such that

$$\overline{\lim}_{M \rightarrow \infty} \sup_{\{x: \|x\| \geq M\}} E_x \left(\frac{\|Y'_{m'}\| - \|x\|}{m'} \right) < -\delta'.$$

iii) There exists $\theta > 0$ such that

$$\overline{\lim}_{M \rightarrow \infty} \sup_{\{x: \|x\| \geq M\}} E_x \exp(\theta (\|Y'_{m'}\| - \|x\|)) < \infty.$$

iv) For each $\delta'_0 > 0$ sufficiently small there exists $\theta > 0$ such that

$$\overline{\lim}_{M \rightarrow \infty} \sup_{\{x: \|x\| \geq M\}} E_x \exp(\theta (\|Y'_{m'}\| - \|x\|)) \leq \exp(-\delta'_0).$$

Remark 3 *So far we have been working with the processes Y and Y' directly, without introducing any scaling. Nevertheless, in this result we introduce a time and space fluid scaling parameterized by n . This scaling will be useful because the initial condition of the time-reversed process in our algorithm is of order $O(N)$ and N is large, so it will be convenient to select n to be either a fraction of N or a large constant.*

Proof. Property i) is standard in the stability analysis of queueing networks (see for instance [9], [10]). Part iii) is immediate since $|||Y'_{m'}|| - ||x|| \leq m'$. For part ii) first note that we can apply i) to conclude that there exist positive integers $M, c'' > 1$ and constants $\delta', m' \in (0, \infty)$ (depending in principle on c'' but not on M) such that

$$\sup_{\{x: M \leq \|x\| \leq M+c''\}} E_x \left(\frac{|||Y'_{m'}|| - ||x||}{m'} \right) < -\delta'.$$

The uniform integrability required to take expectations on the limits described in property i) is satisfied because the random variables are clearly bounded. We note that the supremum above does not change when one varies M (provided it stays sufficiently large). So, the supremum can be taken over $\{x : \|x\| \geq M\}$ rather than $\{x : M \leq \|x\| \leq M + c''\}$ given that the evaluation of the expectation depends on finitely many possibilities for the initial condition x which are determined by the boundaries. So, property ii) follows. Finally, to show iv) we let

$$\psi(x, \theta) = \log E_x \exp(\theta[|||Y'_{m'}|| - ||x||]).$$

Combining properties ii) and iii) we have that there exists M_0 sufficiently large so that for all x with $\|x\| \geq M_0$ and each small enough $\theta > 0$

$$\psi(x, \theta) \leq -\delta'\theta + c_1\theta^2$$

for some $c_1 \in (0, \infty)$. Part iv) follows by appropriately selecting θ suitably close to zero. ■

Using the previous lemma we can provide the proof of the following useful result.

Lemma 2 *We can select $\delta', c_* > 0$ such that*

$$\inf_{\{y: y \in C_N, y \in D_N\}} P_y(\sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \sigma'_{A_N}) > \delta',$$

where $C_N = \{y : \|y\| \leq N + c_*\}$ and $D_N = \{y \in A_N : P_y(\sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \sigma'_{A_N}) > 0\}$.

Proof. We shall write

$$\tau'_N = \inf\{n \geq 1 : |||Y'_n|| \geq N\}. \tag{29}$$

Since $\tau'_N \leq \sigma'_{A_N}$, it suffices to show that

$$\inf_{\{y: y \in C_N, y \in D_N\}} P_y(\sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \tau'_N) > \delta'.$$

Let $y \in C_N \cap D_N$. By suitably conditioning on a sufficiently large number of customers leaving the network before any new customer arrives to the network we obtain that there exists $\gamma \in (0, 1)$ such that

$$P_y(\sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \tau'_N) \geq \gamma P_z(\sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \tau'_N),$$

where z is chosen so that $N - 2\kappa \leq \|z\| \leq N - \kappa$ for some $\kappa > 2c_*$ to be selected momentarily. Now define $C = \{y : \|y\| \leq c_0\}$ for a given constant $c_0 > 1$ and note that (thanks to the structure of Jackson networks which allows us to isolate a specific sequence involving finitely many arrivals and departures) we have

$$\inf_{y: c_0/2 \leq \|y\| \leq c_0} P_y(\sigma'_{\{x\}} \leq \sigma'_{\{0\}}) > \delta''$$

for some $\delta'' > 0$ (depending on c_0 but independent of N). Consider $\sigma'_C = \inf\{n \geq 1 : Y'_n \in C\}$ and observe that

$$P_z(\sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \tau'_N) \geq P_z(\sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \tau'_N, \sigma'_C < \tau'_N). \quad (30)$$

Let us study $P_z(\tau'_N \leq \sigma'_C)$. First, we observe that

$$\begin{aligned} P_z(\tau'_N \leq \sigma'_C) \\ = P_z(\tau'_N \leq \sigma'_C, \tau'_N \leq \delta'N) + P_z(\tau'_N \leq \sigma'_C, \tau'_N > \delta'N). \end{aligned}$$

It follows that

$$\begin{aligned} P_z(\tau'_N \leq \sigma'_C, \tau'_N > \delta'N) \\ \leq P_z(\tau'_N \leq \sigma'_C, \tau'_N > \delta'N, \|Y'_{N\delta'} - y'_{N\delta'}\| \leq N\delta'\varepsilon) \\ + P_z(\|Y'_{N\delta'} - y'_{N\delta'}\| > N\delta'\varepsilon) \\ \leq E_z[P_{Y'_{N\delta'}}(\tau'_N \leq \sigma'_C); \|Y'_{N\delta'} - y'_{N\delta'}\| \leq N\delta'\varepsilon] \\ + P_z(\|Y'_{N\delta'} - y'_{N\delta'}\| > N\delta'\varepsilon). \end{aligned}$$

By virtue of part i) in Lemma 1 the previous two expectations can be made arbitrarily small as $N \nearrow \infty$ uniformly over z in the region $N - 2\kappa \leq \|z\| \leq N - \kappa$. So, the most interesting part of the analysis behind $P_z(\tau'_N \leq \sigma'_C)$ involves estimating $P_z(\tau'_N \leq \sigma'_C, \tau'_N \leq \delta'N)$. Nevertheless,

$$\begin{aligned} P_z(\tau'_N \leq \sigma'_C, \tau'_N \leq \delta'N) \\ \leq P_z(\tau'_N \leq \delta'N) \leq \sum_{j \leq \delta'N} P_z(\|Y'_j\| \geq N). \end{aligned}$$

Note that $\|Y'_j\| \geq N - \delta'dN - 2\kappa$ for all $j \leq \delta'N$ so we have that

$$\begin{aligned} & P_z(\|Y'_j\| - \|y\| \geq \kappa) \\ & \leq \exp(-\kappa) E_z \exp(\theta[\|Y'_j\| - \|y\|]). \end{aligned}$$

We also can write

$$\begin{aligned} & \|Y'_j\| - \|Y'_{j-m'}\| + \|Y'_{j-m'}\| - \|Y'_{j-2m'}\| \\ & + \dots + \|Y'_{m'+j \bmod m'}\| - \|Y'_{j \bmod m'}\| \\ & + \|Y'_{j \bmod m'}\| - \|z\|. \end{aligned} \tag{31}$$

Moreover, by virtue of property iv) of Lemma 1 we have that

$$E_{Y'_{j-m'}} \exp(\theta[\|Y'_j\| - \|Y'_{j-m'}\|]) \leq \exp(-\delta_1)$$

for some $\delta_1 > 0$ uniformly in $Y'_{j-m'} \geq N(1 - \delta'd) - 2\kappa$. Consequently, using (31) and iterating the previous inequality we obtain

$$E_z \exp(\theta[\|Y'_j\| - \|y\|]) \leq c_2 \exp(-\delta_1 \lfloor j/m' \rfloor).$$

for a constant c_2 depending only on m' . We conclude that one can choose κ sufficiently large so that for all N large enough

$$P_z(\tau'_N \leq \sigma'_C) \leq \varepsilon$$

uniformly over $N - 2\kappa \leq \|z\| \leq N - \kappa$. On the other hand,

$$\begin{aligned} & P_z(\sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \tau'_N, \sigma'_C < \tau'_N) \\ & = E_z[I(\sigma'_C < \tau'_N) P_{Y'_{\sigma'_C}}(\sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \tau'_N)], \end{aligned}$$

and we have that

$$P_{Y'_{\sigma'_C}}(\sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \tau'_N) \longrightarrow P_{Y'_{\sigma'_C}}(\sigma'_{\{x\}} \leq \sigma'_{\{0\}}) > 0.$$

as $N \nearrow \infty$. Therefore, given $\varepsilon > 0$, for all N sufficiently large we have that

$$P_z(\sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \tau'_N, \sigma'_C < \tau'_N) \geq P_{Y'_{\sigma'_C}}(T_{\{x\}} < T'_{\{0\}})(1 - 2\varepsilon) > 0$$

uniformly over $N - 2\kappa \leq \|z\| \leq N - \kappa$. The previous inequality combined with (30) yields the result. ■

Finally, we are ready to provide the proof of Proposition 2

Proof of Proposition 4. Define C_N as in Lemma 2. We first will verify that there exists $c'' > 0$ such that

$$P_\pi(v^T Y'_0 = N, Y'_0 \in C_N \cap D_N) \geq c'' P_\pi(v^T Y'_0 = N). \tag{32}$$

Note that

$$\begin{aligned}
& P_\pi (v^T Y'_0 = N, Y'_0 \in C_N, Y'_0 \in D_N) \\
& \geq P_\pi (v^T Y'_0 = N, Y'_0 \in C_N, Y'_0 \in D_N \mid Y'_0(1) \geq 1, \dots, Y'_0(d) \geq 1) \\
& \times P(Y'_0(1) \geq 1) P(Y'_0(2) \geq 1) \dots P(Y'_0(d) \geq 1).
\end{aligned}$$

In turn, conditional on $Y'_0(1) \geq 1, \dots, Y'_0(d) \geq 1$ we have that $Y'_0 \in D_N$. Moreover, because of the memory less property of the geometric distribution we have that $Y'_\infty(j)$ given that $Y'_\infty(j) \geq 1$ has the same distribution as $Y'_\infty(j) + 1$ and therefore

$$\begin{aligned}
& P_\pi (v^T Y'_0 = N, Y'_0 \in C_N, Y'_0 \in D_N \mid Y'_0(1) \geq 1, \dots, Y'_0(d) \geq 1) \\
& = P_\pi (v^T Y'_0 = N - \|v\|, \|Y'_0\| \leq N + c_* - d) \\
& = P_\pi (v^T Y'_0 = N - \|v\|, \|Y'_0\| - v^T Y'_0 + v^T Y'_0 \leq N + c_* - d) \\
& = P_\pi (v^T Y'_0 = N - \|v\|, \|Y'_0\| - v^T Y'_0 \leq c_* - d + \|v\|) \\
& = P_\pi (v^T Y'_0 \geq N - \|v\|) P_\pi (\|Y'_0\| - v^T Y'_0 \leq c_* - d + \|v\|),
\end{aligned}$$

where the last equality follows from the independence between $v^T Y'_0$ and $\|Y'_0\| - v^T Y'_0$. We conclude that for all N large enough

$$\begin{aligned}
& P_\pi (v^T Y'_0 = N, Y'_0 \in C_N, Y'_0 \in D_N) \\
& \geq a_0 P_\pi (v^T Y'_0 = N - \|v\|),
\end{aligned}$$

where

$$a_0 = P_\pi (\|Y'_0\| - v^T Y'_0 \leq c_* - d + \|v\|) \rho_1 \dots \rho_d$$

and c_* is chosen larger than d . Finally, it follows easily from basic properties of the negative binomial distribution and from Proposition 1 that for some $a_1 > 0$,

$$\frac{P_\pi (v^T Y'_0 = N - \|v\|)}{P_\pi (v^T Y'_0 = N)} \geq a_1$$

for all integers $N \geq \|v\|$.

On the other hand, combining Lemma 2 with (32) we can conclude that

$$\begin{aligned}
& P_\pi (\sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \sigma'_{A_N}, v^T Y'_0 = N) \\
& \geq P_\pi (\sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \sigma'_{A_N}, Y'_0 \in A_N \cap C_N \cap D_N) \\
& \geq \delta' P_\pi (Y'_0 \in A_N \cap C_N \cap D_N) \\
& \geq \delta' a_0 a_1 P_\pi (v^T Y'_0 = N).
\end{aligned}$$

Consequently, we obtain that

$$P_\pi (\sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \sigma'_{A_N} \mid v^T Y'_0 = N) \geq \delta' a_0 a_1,$$

which is equivalent to the first part of the proposition.

The second part follows easily also from (32) and from the fact, as we explained in the proof of Corollary 2, that $P_0(T_{\{x\}} < T_{A_N} \wedge \sigma_{\{0\}}) \rightarrow P_0(T_{\{x\}} < T_{\{0\}}) > 0$ as $N \rightarrow \infty$. ■

7 Some Extensions

Before we proceed with some numerical experiments, we would like to briefly discuss a couple of extensions beyond overflow sets such as those defined by A_N in (1).

Let us consider for instance $A_N = A_N(1) \cup A_N(2) \cup \dots \cup A_N(l)$ and $A_N(i)$ is an overflow event such as (1), encoded by a binary vector v_i . We start by explaining how to sample Y_∞ given that $v_i^T Y_\infty = N$ for at least one $i \in \{1, \dots, l\}$. Let $\tilde{P}_i(\cdot)$ to denote the probability measure on Y_∞ induced by the proposal distribution underlying the sampling procedure in Proposition 2 when applied to the set $\{y : v_i^T y = N\}$. In addition, let us write $P_i^*(\cdot)$ to denote the conditional distribution of Y_∞ given that $v_i^T Y_\infty = N$ and define

$$L_i(Y_\infty) = \frac{dP_i^*}{d\tilde{P}_i}(Y_\infty) = \frac{dP}{d\tilde{P}_i}(Y_\infty) \frac{I(v_i^T Y_\infty = N)}{P(v_i^T Y_\infty = N)}.$$

Now consider a proposal distribution which is a mixture of the $\tilde{P}_i(\cdot)$'s with associated probabilities

$$\alpha_i = \frac{P(v_i^T Y_\infty = N)}{\sum_{j=1}^l P(v_j^T Y_\infty = N)} > 0,$$

and call such distribution $\tilde{P}(\cdot)$. We then have that

$$\begin{aligned} & \frac{dP}{d\tilde{P}}(Y_\infty) \frac{I(\cup_{i=1}^l \{v_i^T Y_\infty = N\})}{P(\cup_{i=1}^l \{v_i^T Y_\infty = N\})} \\ & \leq \frac{\sum_{j=1}^l P(v_j^T Y_\infty = N)}{P(\cup_{i=1}^l \{v_i^T Y_\infty = N\})} \sum_{i=1}^l L_i(Y_\infty) \leq l \sum_{i=1}^l L_i(Y_\infty). \end{aligned}$$

The right hand side is bounded as $N \nearrow \infty$ because each term in the sum has been shown to be bounded and therefore $O(1)$ number of proposals are needed to sample Y_∞ given $\cup_{i=1}^l \{v_i^T Y_\infty = N\}$. Finally, our proof of inequality (8), which is given in Lemma 2, applies to these types of overflow sets as well.

Another type of overflow sets that one could consider takes the form

$$A_N = \{y : v^T y \geq N\},$$

where $v \geq 0$ is not necessarily a binary vector. In this case we must first efficiently sample from the conditional distribution Y_∞ given that $v^T Y_\infty \geq N$. Note that we can always represent $Y_\infty(i) = \lfloor \tau_i / \beta_i \rfloor$, where $\beta_i = \log(1/\rho_i)$ and τ_i is exponentially distributed with unit mean. Therefore, we can sample first the τ_i 's conditional on $\sum_{i=1}^d v(i) \tau_i / \beta_i \geq N$; this can be easily done using a straightforward extension of our development in Section 4 replacing negative binomial by Erlang random variables

and we output $Y_\infty(i) = \lfloor \tau_i/\beta_i \rfloor$ as long as $Y_\infty \in A_N$. This sampling procedure takes $O(1)$ function evaluations because it follows without much difficulty that

$$P(Y_\infty \in A_N) = \Theta\left(P\left(\sum_{i=1}^d v(i) \tau_i/\beta_i \geq N\right)\right)$$

as $N \rightarrow \infty$. We also need to show that the associated set D_N , that is, the set of configurations $y \in A_N$ such that $P_y(\sigma'_{\{x\}} \leq \sigma'_{\{0\}} < \sigma'_{A_N}) > 0$, satisfies that

$$P(Y_\infty \in D_N) \geq \delta' P(Y_\infty \in A_N) \tag{33}$$

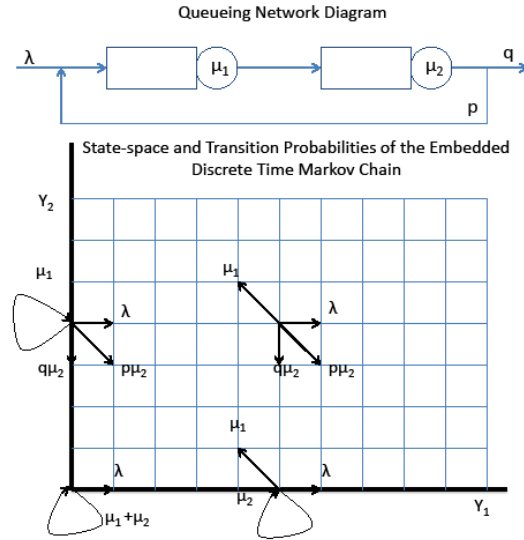
for some $\delta' > 0$. It is not difficult to see that if $y \in D_N$, then $N \leq v^T Y_\infty \leq N + \max_{i=1}^d v(i)$. In turn, it follows easily that there exists $\delta' > 0$ such that

$$\delta' P(Y_\infty \in A_N) \leq P(N \leq v^T Y_\infty \leq N + \max_{i=1}^d v(i)).$$

Finally, we have that the set D_N and the set of configurations y such that $N \leq v^T y \leq N + \max_{i=1}^d v(i)$ differ by finitely many elements. These considerations together imply (33). The analysis of inequality (8) can also be established following ideas similar to those in the proof of Lemma 2.

8 Numerical Examples

We will apply our algorithm in the context of a two node tandem network with feedback and compare against the solution given by the associated system of linear equations. Figure 8 shows the system's description both in terms of its associated queueing diagram and also in terms of a state-space diagram in which the associated discrete time Markov chain is specified. We assume without loss of generality that $\lambda + \mu_1 + \mu_2 = 1$. The embedded discrete time Markov chain induced by this system is $Y = ((Y_k(1), Y_k(2)) : k \geq 0)$. The component $Y_k(i)$ represents the number of customers present inside the i -th station (both in queue and in service) at the time of the k -th transition. Although we will not use it, just to be clear about the mathematical description of the underlying network model, the state of the queueing system at time t (in continuous time) is obtained by evaluating Y at an independent Poisson process with unit rate at time t .

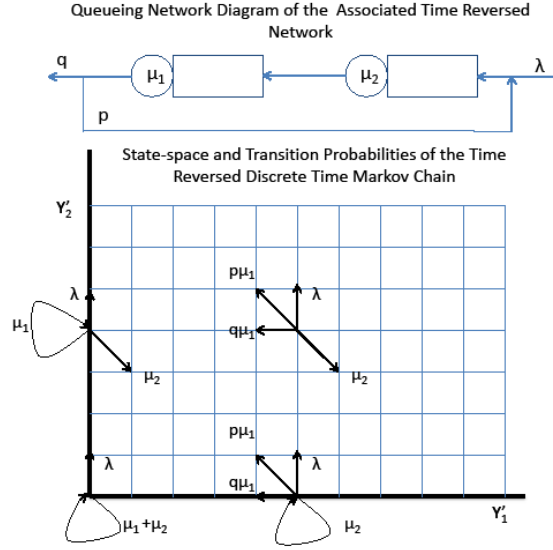


The traffic intensities, which are defined via $\rho_1 = \lambda/[\mu_1(1-p)]$ and $\rho_2 = \lambda/[\mu_2(1-p)]$ for the first and second stations are assumed to be both less than one and will be given specific values later in our discussion. If $x = (x(1), x(2))$ then

$$\pi(x) = (1 - \rho_1) \rho_1^{x(1)} (1 - \rho_2) \rho_2^{x(2)}$$

for $x(1), x(2) \in \mathcal{Z}_+$. If $Y_\infty = (Y_\infty(1), Y_\infty(2))$ denotes a random element with distribution $\pi(\cdot)$, then $Y_\infty(1)$ and $Y_\infty(2)$ are independent geometric random variables with parameters $1 - \rho_1$ and $1 - \rho_2$ respectively.

The time-reversed Markov chain associated to Y , namely $Y' = (Y'_k(1), Y'_k(2) : k \geq 0)$ corresponds to a queueing network similar to that described above. Figure 8 shows both the queueing diagram and the state-space and transition probabilities of the time reversed system.



Given $Y_0 = y_0 := (1, 0)$ we are interested in efficiently computing

$$u_N(y_0) = P_{y_0}(T_{A_N} < \sigma_{\{0\}}),$$

with $A_N = \{(y(1), y(2)) : y(1) + y(2) \geq N\}$. Note that $u_N(y_0)$ is the probability that the total population of the network (in continuous time) reaches level N before *returning* to the origin, starting from the origin.

Our numerical experiments were performed using a laptop at 2.6GHz and 3.50 GB of RAM. We assume that $\lambda = .1$, $\mu_1 = .5$, $p = .1$ and thus $\mu_2 = .4$. The goal is to compare the performance of the linear system of equations against our algorithm. We let $N = 20i$ for $i = 1, 2, \dots, 6$ and report $u_N(y_0)$ using the associated linear system of equations (which is obtained using Matlab's solver for linear systems) is reported in the row "LS". The associated CPU time for the linear system, measured in seconds, is reported in the row "LS time". The estimate produced by Algorithm B using 2000 replications is reported in row "Algo. B", we used $\kappa = 10$ in Remark 1. The CPU time corresponding to Algorithm B is reported in the row "Algo. B time" (also measured in seconds). Finally, we also report a 95% error in relative terms in row the "% error"; this line is obtained as the half-width of the associated 95% confidence interval divided by the estimate reported in Algo. B.

The following table summarizes the performance of the algorithm.

N=	20	40	60	80	100	120
LS	9.604x10 ⁻¹¹	7.274x10 ⁻²²	5.442x10 ⁻³³	4.071x10 ⁻⁴⁴	3.046x10 ⁻⁵⁵	NA
LS time	0.0313	0.1406	1.0313	4.844	14.797	NA
Algo. B	9.631x10 ⁻¹¹	7.442x10 ⁻²²	5.798x10 ⁻³³	4.032x10 ⁻⁴⁴	2.864x10 ⁻⁵⁵	2.269x10 ⁻⁶⁶
Algo. B time	0.3125	0.2977	0.3022	0.2969	0.2813	0.2989
CV Single Rep.	1.685	1.661	1.615	1.698	1.7578	1.691
% error	7.40%	7.28%	7.08%	7.04%	7.70%	7.41%

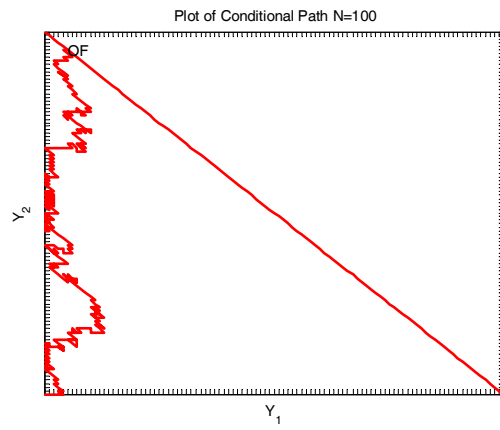
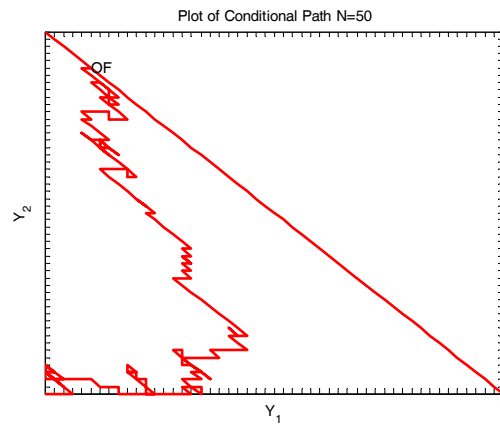
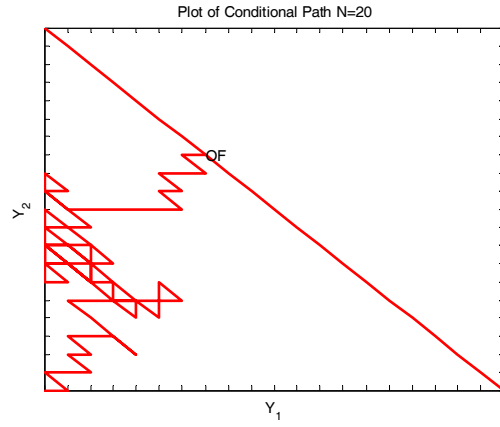
The probabilities reported are too small for practical use, but there are a couple of important points of the experiment. One is that the linear system becomes too large to handle when it reaches a size (i.e. a number of unknowns) of about 7000, which occurs when $N = 120$ (in such case the linear system is $(120 \times 121)/2 = 7260$), the other important point is that Algorithm B is preferable when the associated linear system is of size roughly equal to $25 \times 51 = 1275$. In dimension three, for instance, even the case $N = 20$ would imply a system of linear equations of size around $20^3/2 = 4000$ and in four dimensions the system would be of size 8000. Algorithm B is very stable because the Monte Carlo procedure is basically equivalent to estimating

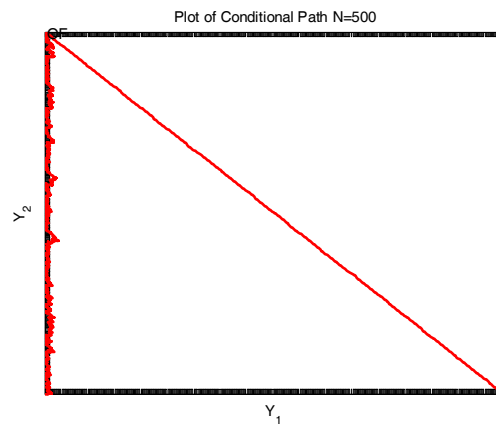
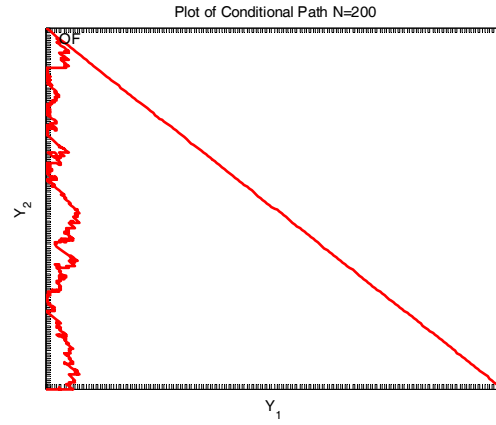
$$P_\pi(\text{reaching population } N - \kappa \text{ before returning to } A_N | Y'_0 \in A_N),$$

which in the case of the numerical experiment above is about 1/4.

We now show the outcome of associated conditional paths using Algorithm A. In each picture we draw both the line $y(1) + y(2) = N$ and the corresponding conditional sample path for different values of N . Obviously, one single picture does not appropriately convey the full characteristics of the conditional distribution of the network given the underlying overflow event. However, to give an idea of the speed at which Algorithm A runs, naive Monte Carlo even for the case $N = 20$ would take many hours (on the order of thousands) while Algorithm A runs in a tiny fraction of a second. So, thousands of replications can be easily generated in a few minutes.

We show different pictures for $N = 20, 50, 100, 200, 500$. Since $\rho_1 < \rho_2$ we have that the downstream station is the bottleneck station and therefore the theory predicts that in fluid scale as $N \rightarrow \infty$, the conditional path is a straight line traveling through the vertical axis to the overflow set. The simulated sample paths show substantial variability in the pre-limit even for events with probability of order 10^{-11} or even 10^{-55} (as is the case, according to the values reported in the table above, for $N = 20, 100$ respectively). Based on these observations we believe that fluid type approximations for conditional paths are of limited use for quantitative applications. On the other hand, because of its speed, Algorithm A could be quite handy in these types of settings.





Acknowledgement 3 *This research was partially supported by the grant NSF-0846816. The author wishes to thank the Associate Editor and two anonymous referees for their very valuable comments, which helped to improve the quality of the paper substantially. We also thank Sheldon Ross for bringing to our attention the problem of computing the distribution in Proposition 1. Insightful discussions with Karl Sigman are also acknowledged.*

References

- [1] I. Adan, R. Foley, and R. McDonald. Exact asymptotics for the stationary distribution of a markov chain: A production model. *QUESTA*, 62:311–344, 2009.
- [2] V. Anantharam, P. Heidelberger, and P. Tsoucas. Analysis of rare events in continuous time marked chains via time reversal and fluid approximation. *IBM Research Report, REC 16280*, 1990.
- [3] S. Asmussen and P. Glynn. *Stochastic Simulation: Algorithms and Analysis*. Springer-Verlag, New York, NY, USA, 2008.
- [4] J. Blanchet, K. Leder, and Y. Shi. Analysis of a splitting estimator for rare event probabilities in jackson networks. *Stochastic Systems, to appear*, 2011.
- [5] J. Blanchet and J. C. Liu. Importance sampling in ruin problems for multi-dimensional regularly varying random walks. *Journal of Applied Probability*, 47:301–322, 2010.
- [6] J. Blanchet and M. Mandjes. Rare event simulation for queues. In G. Rubino and B. Tuffin, editors, *Rare Event Simulation Using Monte Carlo Methods*, pages 87–124. Wiley, West Sussex, United Kingdom, 2009. Chapter 5.
- [7] J. Bucklew. *Introduction to Rare-event Simulation*. Springer-Verlag, New York, NY, 2004.
- [8] M. Cottrell, J.-C. Fort, and G. Malgouyres. Large deviations and rare events in the study of stochastic algorithms. *IEEE Trans. Automat. Contr.*, 28:907–920, 1983.
- [9] J. Dai. On positive harris recurrence of multiclass queueing networks: A unified approach via fluid limit models. *Annals of Applied Probability*, 5:49–77, 1995.
- [10] J. Dai and S. Meyn. Stability and convergence of moments for multiclass queueing networks via fluid limit models. *IEEE Transactions on Automatic Control*, 40:1889–1904, 1995.
- [11] P. J. De Boer. Analysis of state-independent importance sampling measures for the two-note tandem queue. *Preprint*, 2007.
- [12] T. Dean and P. Dupuis. Splitting for rare event simulation: A large deviation approach to design and analysis. Technical report, Department of Applied Mathematics, Brown University, Providence, RI, 2007.
- [13] P. Dupuis, A. Sezer, and H. Wang. Dynamic importance sampling for queueing networks. *Ann. Appl. Probab.*, 17:1306–1346, 2007.

- [14] P. Dupuis and H. Wang. Importance sampling for jackson networks. *Preprint*, 2008.
- [15] P. Glasserman, P. Heidelberger, P. Shahabuddin, and T. Zajic. Multilevel splitting for estimating rare event probabilities, 1999.
- [16] P. Glasserman and S. Kou. Analysis of an importance sampling estimator for tandem queues. *ACM TOMACS*, 5:22–42, 1995.
- [17] P. Glasserman and H. Wang. Counter examples in importance sampling for large deviations probabilities. *Ann. Appl. Prob.*, 7:731–746, 1997.
- [18] I. Ignatiouk-Robert. Large deviations of jackson networks. *Annals of Applied Probability*, 10:962–1001, 2000.
- [19] S. Juneja and V. Nicola. Efficient simulation of buffer overflow probabilities in jackson networks with feedback. *ACM Trans. Model. Comput. Simul.*, 15(4):281–315, 2005.
- [20] S. Juneja and P. Shahabuddin. Rare event simulation techniques: An introduction and recent advances. In S. G. Henderson and B. L. Nelson, editors, *Simulation*, Handbooks in Operations Research and Management Science. Elsevier, Amsterdam, The Netherlands, 2006.
- [21] D. Kroese and V. Nicola. Efficient simulation of a tandem jackson network. *ACM Trans. Model. Comput. Simul.*, 12:119–141, 2002.
- [22] K. Majewski and K. Ramanan. How large queues build up in a jackson network. *To Appear in Math. of O.R.*, 2008.
- [23] V. Nicola and T. Zaburnenko. Efficient importance sampling heuristics for the simulation of population overflow in jackson networks. *ACM Trans. Model. Comput. Simul.*, 17(2), 2007.
- [24] S. Parekh and J. Walrand. Quick simulation of rare events in networks. *IEEE Trans. Automat. Contr.*, 34:54–66, 1989.
- [25] P. Robert. *Stochastic Networks and Queues*. Springer-Verlag, Berlin, 2003.
- [26] Ross S. *Introduction to Probability Models*. Academic Press, San Diego, tenth edition, 2010.
- [27] J. S. Sadowsky and J. A. Bucklew. On large deviations theory and asymptotically efficient monte carlo estimation. *IEEE Transactions on Information Theory*, 36:579–588, 1990.

- [28] D. Siegmund. Importance sampling in the Monte Carlo study of sequential tests. *Ann. Stat.*, 3:673–684, 1976.
- [29] M. Villén-Altamirano and J. Villén-Altamirano. Restart: a straightforward method for fast simulation of rare events. In *Winter Simulation Conference*, pages 282–289, 1994.