MATHEMATICAL ENGINEERING
TECHNICAL REPORTS

Polynomial-time Randomized Approximation and Perfect Sampler for Closed Jackson Networks with Single Servers

Shuji KIJIMA and Tomomi MATSUI

METR 2005–12 May 2005

DEPARTMENT OF MATHEMATICAL INFORMATICS
GRADUATE SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY
THE UNIVERSITY OF TOKYO
BUNKYO-KU, TOKYO 113-8656, JAPAN

WWW page: http://www.i.u-tokyo.ac.jp/mi/mi-e.htm

Polynomial-time Randomized Approximation and Perfect Sampler for Closed Jackson Networks with Single Servers

(Extended Abstract)

Shuji KIJIMA and Tomomi MATSUI
Department of Mathematical Informatics,
Graduate School of Information Science and Technology,
University of Tokyo, Bunkyo-ku, Tokyo 113-8656, Japan
kijima@misojiro.t.u-tokyo.ac.jp
http://www.simplex.t.u-tokyo.ac.jp/~tomomi/

May 31, 2005

Abstract

In this paper, we propose the first fully polynomial-time randomized approximation scheme (FPRAS) for basic queueing networks, closed Jackson networks with single servers. Our algorithm is based on MCMC (Markov chain Monte Carlo) method. Thus, our scheme returns an approximate solution, of which the size of error satisfies a given error rate. We propose two Markov chains, one is for approximate sampling, and the other is for perfect sampling based on monotone coupling from the past algorithm.

1 Introduction

In this paper, we propose a randomized approximation scheme for basic queueing networks, closed Jackson networks with single servers. Our scheme is based on the MCMC (Markov chain Monte Carlo) method, and returns an approximate value of the normalizing constant of the steady-state distribution of the number of customers at nodes. The complexity of our algorithm is bounded by a polynomial of n and $\ln K$ where n is the number of nodes and K is the number of customers in a network. Thus, our scheme is a (weakly) polynomial time approximation scheme. We propose two Markov chains, both of which are rapidly mixing. One is for approximate sampling, while the other is for perfect sampling.

The Jackson network is proposed by Jackson in 1957 [10], and is one of the basic and significant model in queueing network theory. It is well known that the steady-state distribution of a Jackson network is a product form [10]. By computing the normalizing constant of the product form solution, we can obtain significant evaluated value like as through put, rates of utilization of stations, and so on.

There is well-known Buzen's algorithm [4], which computes the normalizing constant. However the running time of Buzen's algorithm is pseudo-polynomial time that depends on the number of customers in a closed network. Chen and O'Cinneide [5] proposed a randomized algorithm based on MCMC, but their algorithm is weakly polynomial-time in some very special cases. Ozawa [18] proposed a perfect sampler for closed Jackson networks with single servers, however his chain mixes in pseudo-polynomial-time.

When we construct a randomized approximation algorithm based on MCMC, we need to consider the accuracy of the obtained value. For any given parameter ε and δ, satisfying $0 < \varepsilon < 1$, $0 < \delta < 1$, an FPRAS (Fully Polynomial-time Randomized Approximation Scheme) provides an algorithm which finds an approximate solution Z satisfying

$$\Pr[|Z - A| \leq \varepsilon A] \geq 1 - \delta$$

where A is the exact solution, and whose running time is bounded by a polynomial of the input size of the instance (the number of nodes and *logarithm* of the number of customers), ε^{-1} and $\ln(\delta^{-1})$ [13].

In this paper, we are concerned with a closed Jackson network with single servers. Thus, we assume that we are given a strongly connected network, each node has a single server, a class of customers is unique, and no customer leaves or enters the network. Then we propose an FPRAS based on MCMC for calculating the normalizing constant. We propose two new Markov chains, both of which have a product form solution of a closed Jackson network as a unique stationary distribution. Both of our chains mix in polynomial-time of input size. Here we note that our chains are not a simulation of a queueing network, but just have a unique stationary distribution which is the same as a product form solution for a network.

1.1 Product-form Solution

We denote the set of real numbers (non-negative, positive real numbers) by $\mathbb{R}$ ($\mathbb{R}_+$, $\mathbb{R}_{++}$), and the set of integers (non-negative, positive integers) by $\mathbb{Z}$ ($\mathbb{Z}_+$, $\mathbb{Z}_{++}$), respectively. In queueing network theory, it is well known that a closed Jackson network has a product form solution. Let $n \in \mathbb{Z}_{++}$ be the number of nodes and $K \in \mathbb{Z}_+$ be the number of customers in a closed Jackson network. Let us consider the set of non-negative integer points

$$\Xi(K) \stackrel{\text{def.}}{=} \left\{ \boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{Z}_+^n \mid \textstyle\sum_{i=1}^n x_i = K \right\}$$

in an $n-1$ dimensional simplex. We abbreviate $\Xi(K)$ to Ξ, if there is no confusion. Let W be the transition probability matrix for a closed Jackson network system. Here we note that W is ergodic, so 1 is an eigenvalue and corresponding eigenvector is unique, excluding constant factor. Let $\theta \in \mathbb{R}_{++}^n$ be an eigenvector for W with corresponding to the eigenvalue 1, i.e., $\theta W = \theta$. Given a vector $\mu = (\mu_1, \mu_2, \ldots, \mu_n) \in \mathbb{R}_{++}^n$ of the expected service time on nodes, the steady-state distribution $J : \Xi \to \mathbb{R}_{++}$ for the closed Jackson network is product form defined by

$$J(\boldsymbol{x}) = \frac{1}{G(K)} \prod_{i=1}^n \alpha_i^{x_i} \left(\equiv \frac{1}{G(K)} \prod_{i=1}^n \left(\frac{\theta_i}{\mu_i} \right)^{x_i} \right)$$

where $\alpha_i \stackrel{\text{def.}}{=} \theta_i/\mu_i$ and $G(K) \stackrel{\text{def.}}{=} \sum_{\boldsymbol{x} \in \Xi(K)} \prod_{i=1}^n \alpha_i^{x_i}$ is the normalizing constant [10].

2 Randomized Approximation Scheme

In the following we consider a closed Jackson network with n nodes, K customers and parameters $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{Z}_{++}$, which has the product form solution $J(\boldsymbol{x}) = (1/G(K)) \prod_{i=1}^n \alpha_i^{x_i}$ for any $\boldsymbol{x} \in \Xi(K)$.

2.1 Rapidly Mixing Markov Chain

Now we propose a new Markov chain $\mathcal{M}_{\mathrm{A}}(K)$ with state space $\Xi(K)$. A transition of $\mathcal{M}_{\mathrm{A}}(K)$ from a current state $X \in \Xi(K)$ to a next state X' is defined as follows. First, we choose a pair of distinct indices $\{j_1, j_2\}$ uniformly at random. Next, put $k = X_{j_1} + X_{j_2}$, and choose $l \in \{0, 1, \ldots, k\}$ with probability

$$\frac{\alpha_{j_1}^l \alpha_{j_2}^{k-l}}{\sum_{s=0}^k \alpha_{j_1}^s \alpha_{j_2}^{k-s}} \qquad \left(\equiv \frac{\alpha_{j_1}^l \alpha_{j_2}^{k-l} \prod_{j \notin \{j_1, j_2\}} \alpha_j^{X_j}}{\sum_{s=0}^k \alpha_{j_1}^s \alpha_{j_2}^{k-s} \prod_{j \notin \{j_1, j_2\}} \alpha_j^{X_j}} \right)$$

0	$\alpha_i^0\alpha_j^k/A$	$\alpha_i^1\alpha_j^{k-1}/A$	$\cdots$	$\alpha_i^k\alpha_j^0/A$	1

0	$\alpha_i^0\alpha_j^{k+1}/A'$	$\alpha_i^1\alpha_j^k/A'$	$\alpha_i^2\alpha_j^{k-1}/A'$	$\cdots$	$\alpha_i^{k+1}\alpha_j^0/A'$	1

Figure 1: A figure of alternating inequalities for a pair of indices (i,j) and a non-negative integer k. In the figure, $A \stackrel{\text{def.}}{=} \sum_{s=0}^{k}\alpha_i^s\alpha_j^{k-s}$ and $A' \stackrel{\text{def.}}{=} \sum_{s=0}^{k+1}\alpha_i^s\alpha_j^{k+1-s}$ are normalizing constants.

and set

$$X_i' = \begin{cases} l & (\text{for } i = j_1), \\ k-l & (\text{for } i = j_2), \\ X_i & (\text{otherwise}). \end{cases}$$

The Markov chain $\mathcal{M}_{\mathrm{A}}(K)$ is irreducible and aperiodic, so ergodic, hence has a unique stationary distribution. Also, $\mathcal{M}_{\mathrm{A}}(K)$ satisfies the detailed balance equation, thus the stationary distribution is the product form solution $J(\boldsymbol{x})$.

Given a pair of probability distributions ν_1 and ν_2 on a finite state space Ω, the *total variation distance* between ν_1 and ν_2 is defined by $\mathrm{d_{TV}}(\nu_1,\nu_2) \stackrel{\text{def.}}{=} \frac{1}{2}\sum_{x\in\Omega}|\nu_1(x)-\nu_2(x)|$. The *mixing time* of an ergodic Markov chain is defined by $\tau(\varepsilon) \stackrel{\text{def.}}{=} \max_{x\in\Xi}\{\min\{t \mid \forall s \geq t,\ \mathrm{d_{TV}}(\pi, P_x^s) \leq \varepsilon,\ (0 < \varepsilon < 1)\}\}$ where π is the stationary distribution and P_x^s is the probability distribution of the chain at time period $s \geq 0$ with initial state x at time period 0. In the following, we discuss the mixing time of $\mathcal{M}_{\mathrm{A}}(K)$.

Here, we consider the cumulative distribution function $g_{ij}^k : \{0,1,\ldots,k\} \to \mathbb{R}_+$ defined by

$$g_{ij}^k(l) \quad \stackrel{\text{def.}}{=} \quad \frac{\sum_{s=0}^{l}\alpha_i^s\alpha_j^{k-s}}{A_{ij}^k} = \begin{cases} \frac{\alpha_i^{l+1}-\alpha_j^{l+1}}{\alpha_i^{k+1}-\alpha_j^{k+1}}\cdot\alpha_j^{k-l} & (\alpha_i \neq \alpha_j), \\ \frac{l}{k+1} & (\alpha_i = \alpha_j), \end{cases}$$

for $l \in \{0,1,\ldots,k\}$, where $A_{ij}^k \stackrel{\text{def.}}{=} \sum_{s=0}^{k}\alpha_i^s\alpha_j^{k-s}$ is a normalizing constant. We also define $g_{ij}^k(-1) \stackrel{\text{def.}}{=} 0$, for convenience. We can simulate the Markov chain $\mathcal{M}_{\mathrm{A}}(K)$ efficiently by using the function g_{ij}^k as follows. First, choose a pair $\{i,j\}$ of indices with the probability $2/(n(n-1))$. Next, put $k = X_i + X_j$, generate an uniformly random real number $\Lambda \in [0,1)$, choose an integer l satisfying $g_{ij}^k(l-1) \leq \Lambda \leq g_{ij}^k(l)$, and set $X_i' = l$ and $X_j' = k-l$, keeping the value of the other indices. We can execute a transition of $\mathcal{M}_{\mathrm{A}}$ efficiently by employing ordinary binary search technique.

The following is a key lemma in this paper.

Lemma 2.1 *The function g_{ij}^k satisfies the* alternating inequalities*,*

$$g_{ij}^{k+1}(l) \leq g_{ij}^k(l) \leq g_{ij}^{k+1}(l+1), \quad \forall k \in \{1,\ldots,K\},\ \forall l \in \{1,\ldots,k\}.$$

Proof: First, we prove the former inequality $g_{ij}^{k+1}(l) \leq g_{ij}^k(l)$ as follows,

$$\frac{g_{ij}^k(l)}{g_{ij}^{k+1}(l)} = \frac{\sum_{s=0}^{l}\alpha_i^s\alpha_j^{k-s}}{A_{ij}^k}\cdot\frac{A_{ij}^{k+1}}{\sum_{s=0}^{l}\alpha_i^s\alpha_j^{k+1-s}} = \frac{\sum_{s=0}^{k+1}\alpha_i^s\alpha_j^{k+1-s}}{\alpha_j\sum_{s=0}^{k}\alpha_i^s\alpha_j^{k-s}} = \frac{\sum_{s=0}^{k+1}\alpha_i^s\alpha_j^{k+1-s}}{\sum_{s=0}^{k}\alpha_i^s\alpha_j^{k+1-s}} \geq 1.$$

Next, we prove the latter inequality $g_{ij}^{k}(l) \leq g_{ij}^{k+1}(l+1)$ as follows (please see Appendix for detail),

$$\frac{g_{ij}^{k+1}(l+1)}{g_{ij}^{k}(l)} = \frac{\left(\sum_{s=0}^{k} \alpha_i^s \alpha_j^{k-s}\right)\left(\sum_{s=0}^{l} \alpha_i^s \alpha_j^{k+1-s} + \alpha_i^{l+1}\alpha_j^{k-l}\right)}{\left(\sum_{s=0}^{k} \alpha_i^s \alpha_j^{k+1-s} + \alpha_i^{k+1}\right)\left(\sum_{s=0}^{l} \alpha_i^s \alpha_j^{k-s}\right)}$$

$$= \frac{\left(\alpha_i^{l+1}\alpha_j^{k-l}\right)^{-1} \alpha_j \left(\sum_{s=0}^{k} \alpha_i^s \alpha_j^{k-s}\right)\left(\sum_{s=0}^{l} \alpha_i^s \alpha_j^{k-s}\right) + \sum_{s=0}^{k} \alpha_i^s \alpha_j^{k-s}}{\left(\alpha_i^{l+1}\alpha_j^{k-l}\right)^{-1} \alpha_j \left(\sum_{s=0}^{k} \alpha_i^s \alpha_j^{k-s}\right)\left(\sum_{s=0}^{l} \alpha_i^s \alpha_j^{k-s}\right) + \sum_{s=k-l}^{k} \alpha_i^s \alpha_j^{k-s}} \geq 1.$$

Thus we obtain the claim. □

The above alternating inequalities imply the following.

Theorem 2.2 *For $0 < \forall\varepsilon < 1$, the mixing time $\tau(\varepsilon)$ of Markov chain $\mathcal{M}_{\mathrm{A}}(K)$ satisfies*

$$\tau(\varepsilon) \leq \frac{n(n-1)}{2} \ln(K\varepsilon^{-1}).$$

Proof: Let $G = (\Xi, \mathcal{E})$ be an undirected simple graph with vertex set Ξ and edge set $\mathcal{E}$ defined as follows. A pair of vertices $\{\boldsymbol{x}, \boldsymbol{y}\}$ is an edge of G if and only if $(1/2)\sum_{i=1}^{n} |x_i - y_i| = 1$. Clearly the graph G is connected. We define the length $l_{\mathrm{A}}(e)$ of every edge $e \in \mathcal{E}$ by $l_{\mathrm{A}}(e) \stackrel{\text{def.}}{=} 1$. For each pair $(\boldsymbol{x}, \boldsymbol{y}) \in \Xi^2$, we define the distance $d_{\mathrm{A}}(\boldsymbol{x}, \boldsymbol{y})$ by the length of the shortest path between $\boldsymbol{x}$ and $\boldsymbol{y}$ on G with respect to l_{A}. Clearly, the diameter of G, defined by $\max_{\boldsymbol{x},\boldsymbol{y}\in\Xi}\{d_{\mathrm{A}}(\boldsymbol{x}, \boldsymbol{y})\}$, is bounded by K.

We define a joint process $(X, Y) \mapsto (X', Y')$ for any pair $\{X, Y\} \in \mathcal{E}$. Pick a distinct pair of indices $\{i_1, i_2\}$ uniformly at random. Then set $k_X = X_{i_1} + X_{i_2}$ and $k_Y = Y_{i_1} + Y_{i_2}$, generate a uniform random number $\Lambda \in [0, 1)$, choose l_X and l_Y satisfying $g_{i_1 i_2}^{k_X}(l_X - 1) \leq \Lambda < g_{i_1 i_2}^{k_X}(l_X)$ and $g_{i_1 i_2}^{k_Y}(l_Y - 1) \leq \Lambda < g_{i_1 i_2}^{k_Y}(l_Y)$, and set $X'_{i_1} = l_X$, $X'_{i_2} = k_X - l_X$, $Y'_{i_1} = l_Y$, and $Y'_{i_2} = k_Y - l_Y$.

Now we show that

$$\text{for any pair } \{X, Y\} \in \mathcal{E},\ \mathrm{E}[d_{\mathrm{A}}(Y', Y')] \leq \beta \cdot d_{\mathrm{A}}(X, Y),\ \text{where } \beta = 1 - \frac{2}{n(n-1)}.$$

Since $\{X, Y\} \in \mathcal{E}$, there exists a distinct pair of indices $\{j_1, j_2\}$ satisfying $|X_j - Y_j| = 1$ for $j \in \{j_1, j_2\}$, and $|X_j - Y_j| = 0$ for $j \notin \{j_1, j_2\}$.

Case 1: When neither indices j_1 nor j_2 are chosen, i.e., $\{i_1, i_2\} \cap \{j_1, j_2\} = \emptyset$, we put $k = X_{i_1} + X_{i_2}$, and it is easy to see that $\Pr(X'_{i_1} = l) = \Pr(Y'_{i_1} = l)$ for any $l \in \{0, \ldots, k\}$ since $Y_{i_1} + Y_{i_2} = k$. Thus $X'_{i_1} = Y'_{i_1}$ and $X'_{i_2} = Y'_{i_2}$ hold. Hence $d_{\mathrm{A}}(X', Y') = d_{\mathrm{A}}(X, Y)$.

Case 2: When both indices j_1 and j_2 are chosen, i.e., $\{i_1, i_2\} = \{j_1, j_2\}$, in the same way as Case 1, both $X'_{i_1} = Y'_{i_1}$ and $X'_{i_2} = Y'_{i_2}$ hold. Hence $d_{\mathrm{A}}(X', Y') = 0$.

Case 3: When exactly one of j_1 and j_2 is chosen, i.e., $|\{i_1, i_2\} \cap \{j_1, j_2\}| = 1$, without loss of generality, we can assume that $i_1 = j_1$ and that $X_{i_1} = Y_{i_1} + 1$. Let $k + 1 = X_{i_1} + X_{i_2}$. Then $Y_{i_1} + Y_{i_2} = k$ obviously. We parameterize the transition of Markov chain $\mathcal{M}_{\mathrm{A}}(K)$ with the uniformly random number $\Lambda \in [0, 1)$ by using the function $g_{i_1 i_2}^{k+1}$ and $g_{i_1 i_2}^{k}$. Set $X_{i_1} = l_1$ such that $g_{i_1 i_2}^{k+1}(l_1 - 1) \leq \Lambda < g_{i_1 i_2}^{k+1}(l_1)$ and $Y_{i_1} = l_2$ such that $g_{i_1 i_2}^{k}(l_2 - 1) \leq \Lambda < g_{i_1 i_2}^{k}(l_2)$ for the common random number Λ. From Lemma 2.1, the alternating inequalities holds, and we can see that $l_1 = l_2$ or $l_1 = l_2 + 1$. In either case, we can set $[X'_{i_1} = Y'_{i_1} + 1$ and $X'_{i_2} = Y'_{i_2}]$ or $[X'_{i_1} = Y'_{i_1}$ and $X'_{i_2} = Y'_{i_2} + 1]$. Hence $d_{\mathrm{A}}(X', Y') = d_{\mathrm{A}}(X, Y)$.

Considering that Case 2 occurs with probability $2/(n(n-1))$, we obtain that

$$\mathrm{E}[d_{\mathrm{A}}(X',Y')] \leq \left(1 - \frac{2}{n(n-1)}\right) d_{\mathrm{A}}(X,Y).$$

Since the diameter of G is bounded by K, Path Coupling Theorem (Theorem 2.3) implies that the mixing time $\tau(\varepsilon)$ of $\mathcal{M}_{\mathrm{A}}(K)$ satisfies

$$\tau(\varepsilon) \leq \frac{n(n-1)}{2} \ln(K\varepsilon^{-1}).$$

□

The Path Coupling Theorem proposed by Bubbly and Dyer is a useful technique for bounding the mixing time.

Theorem 2.3 (Path Coupling Theorem [3]) *Let $\mathcal{M}$ be a finite ergodic Markov chain with state space Ω. Let $H = (\Omega, \mathcal{E})$ be a connected undirected graph with vertex set Ω and edge set $\mathcal{E} \subseteq \binom{\Omega}{2}$. Let $l : \mathcal{E} \to \mathbb{R}_{++}$ be a positive length defined on the edge set. For any pair of vertices $\{x, y\}$ of H, the distance between x and y, denoted by $d(x,y)$ and/or $d(y,x)$, is the length of a shortest path between x and y, where the length of a path is the sum of the lengths of edges in the path. Suppose that there exists a joint process $(X,Y) \mapsto (X',Y')$ with respect to $\mathcal{M}$ whose marginals are a faithful copy of $\mathcal{M}$ and satisfying*

$$0 < \exists\beta < 1,\ \forall\{X,Y\} \in \mathcal{E},\ \mathrm{E}[d(X',Y')] \leq \beta d(X,Y).$$

Then the mixing time $\tau(\varepsilon)$ of the Markov chain $\mathcal{M}$ satisfies $\tau(\varepsilon) \leq (1-\beta)^{-1}\ln(\varepsilon^{-1}D/d)$ where $d \stackrel{\text{def.}}{=} \min\{d(x,y) \mid \forall x, \forall y \in \Omega\}$ and $D \stackrel{\text{def.}}{=} \max\{d(x,y) \mid \forall x, \forall y \in \Omega\}$. ■

The above theorem differs from the original theorem in [3] since the integrality of the edge length is not assumed. We drop the integrality and introduced the minimum distance d. Theorem 2.3 can be proved by a slight modification of the original proof.

2.2 Monte Carlo Integration

In this section, we give an FPRAS for calculating the normalizing constant $G(K)$ of product form solution for a closed Jackson network. Our approximation scheme is a standard Jerrum-Sinclair type recursive algorithm [13, 12] but we should be careful at some points.

Here we suppose that $K \geq 1$ and arrange the indices of nodes in network satisfying the following condition.

Condition 1 $\alpha_1 = \max_i \alpha_i$.

We define a set $\Xi'(K) \subset \Xi(K)$ by $\Xi'(K) \stackrel{\text{def.}}{=} \left\{\boldsymbol{x} \in \Xi(K) \mid x_1 \geq \left\lceil \frac{K}{n} \right\rceil\right\}$. Let $G'(K) \stackrel{\text{def.}}{=} \sum_{\boldsymbol{x}\in\Xi'(K)} \prod_{i=1}^{n} \alpha_i^{x_i}$, then it is not difficult to see that Condition 1 implies $G'(K)/G(K) \geq 1/n$.

Considering that α_i ($\forall i \in \{1,2,\ldots,n\}$) is independent of K, it is easy to see that

$$\begin{aligned} G'(K) &= \textstyle\sum_{\boldsymbol{x}\in\Xi'(K)} \left(\alpha_1^{\lceil K/n\rceil} \cdot (\prod_{i=2}^{n} \alpha_i^{x_i}) \cdot \alpha_1^{x_1 - \lceil K/n\rceil}\right) \\ &= \textstyle\sum_{\boldsymbol{x}\in\Xi(K-\lceil K/n\rceil)} \left(\alpha_1^{\lceil K/n\rceil} \cdot (\prod_{i=2}^{n} \alpha_i^{x_i}) \cdot \alpha_1^{x_1}\right) \\ &= \alpha_1^{\lceil K/n\rceil} \cdot G(K - \lceil K/n\rceil). \end{aligned}$$

Thus, we can compute $G(K)$ by

$$G(K) = \frac{G(K)}{G'(K)} \cdot \alpha_1^{\lceil K/n \rceil} \cdot G(K - \lceil K/n \rceil)$$

if we know the value of $G(K)/G'(K)$ and $G(K - \lceil K/n \rceil)$. By applying the above recursively, $G(0) = 1$ implies that

$$G(K) = G(0) \prod_{j=1}^{R} \frac{G(K_{j-1})}{G(K_j)} = \prod_{j=1}^{R} \left(\alpha_1^{K_{j-1}-K_j} \cdot \frac{G(K_{j-1})}{G'(K_{j-1})} \right) = \alpha_1^K \prod_{j=0}^{R-1} \frac{G(K_j)}{G'(K_j)},$$

where we define $K_0 \stackrel{\text{def.}}{=} K$, and $K_j \stackrel{\text{def.}}{=} K_{j-1} - \left\lceil \frac{K_{j-1}}{n} \right\rceil$ for $j = 1, 2, \ldots$, while $K_j \geq 0$, and let $R \in \mathbb{Z}_{++}$ be the minimum index satisfying $K_R = 0$.

Lemma 2.4 *The number of recursions R satisfies that $R \leq n \ln K + 1$ for any $K \in \mathbb{Z}_{++}$.*

Since we already have an approximate sampler via the Markov chain $\mathcal{M}_{\mathrm{A}}(K)$, we only need to estimate $G(K_j)/G'(K_j)$ for $j \in \{0, 1, \ldots, R-1\}$ by the Monte Carlo method. The whole algorithm is as follows,

Algorithm 1 (Randomize Approximation Scheme with approximate sampler)

Step 1. Set $j = 1$, $K' = K$.
  While $K' \geq 1$,
    do$\rightarrow$
      Generate Q_{A} samples, each of which is obtained by simulating $\mathcal{M}_{\mathrm{A}}(K')$ for $T_{\mathrm{A}}(K')$ steps.
      Let U_j be the number of samples which satisfy $x_1 \geq K'/n$.
      Set $Z_j := (U_j + 1)/(Q_{\mathrm{A}} + 1)$.
      Set $K' := K' - \left\lceil \frac{K'}{n} \right\rceil$ and set $j := j + 1$.
    $\leftarrow$od
Step 2. Set $Z := \alpha_1^K \prod_j (1/Z_j)$. Output Z.

Our algorithm generates Q_{A} samples by simulating $\mathcal{M}_{\mathrm{A}}(K_j)$ for $T_{\mathrm{A}}(K_j)$ steps for each sample. By setting $Q_{\mathrm{A}} = 144nR^2\varepsilon^{-2}\ln(2R/\delta)$ and $T_{\mathrm{A}}(K') = \left\lceil \frac{n(n-1)}{2} \ln \frac{6nRK'}{\varepsilon} \right\rceil$, we obtain the following theorem.

Theorem 2.5 *If we set $Q_{\mathrm{A}} = 144nR^2\varepsilon^{-2}\ln(2R/\delta)$ and $T_{\mathrm{A}}(K') = \left\lceil \frac{n(n-1)}{2} \ln \frac{6nRK'}{\varepsilon} \right\rceil$, then our randomized approximation scheme (Algorithm 1) returns Z satisfying*

$$\Pr\left[|Z - G(K)| \leq \varepsilon G(K)\right] \geq 1 - \delta.$$

In the proof of above theorem, we need the following modified Chernoff bound [14].

Lemma 2.6 *Let X_i $(1 \leq i \leq M)$ be i.i.d. random variables such that $X_i = 1$ with probability p, and $X_i = 0$ with probability $1 - p$. Let $U = \sum_{i=1}^{M} X_i$ and $0 < \lambda < 1$. If $M \geq (4 + 2\sqrt{3})/p\lambda$, then*

$$\Pr\left[\left|\frac{U+1}{M+1} - p\right| \geq \lambda p\right] \leq 2\mathrm{e}^{-\frac{1}{4}\lambda^2 Mp}$$

hold.

Also we can estimate the error of bias.

Theorem 2.7 *If we set* $Q_{\mathrm{A}} = 144nR^2\varepsilon^{-2}\ln(2R/\delta)$ *and* $T_{\mathrm{A}}(K') = \lceil \frac{n(n-1)}{2} \ln \frac{6nRK'}{\varepsilon} \rceil$, *then the bias of the expectation of the obtained approximate solution is bounded as follows,*

$$\frac{|\mathrm{E}[Z] - G(K)|}{G(K)} \leq \frac{\varepsilon}{4} + R\mathrm{e}^{-120R^2\varepsilon^{-2}\ln(2R/\delta)} \leq \left(\frac{1}{4} + \frac{1}{10^{36}}\right)\varepsilon.$$

Proofs of the above results appear in Appendix section.

3 Perfect Sampler

3.1 Monotone Coupling from the Past

Here we review CFTP briefly. Suppose that we have an ergodic Markov chain $\mathcal{M}$ with a finite state space Ω and a transition matrix P. The transition rule of the Markov chain $X \mapsto X'$ can be described by a deterministic function $\phi : \Omega \times [0,1) \to \Omega$, called *update function*, as follows. Given a random number Λ uniformly distributed over $[0,1)$, update function ϕ satisfies that $\Pr(\phi(x,\Lambda) = y) = P(x,y)$ for any $x, y \in \Omega$. We can realize the Markov chain by setting $X' = \phi(X, \Lambda)$. Clearly, update functions corresponding to the given transition matrix P are not unique. The result of transitions of the chain from the time t_1 to t_2 ($t_1 < t_2$) with a sequence of random numbers $\boldsymbol{\lambda} = (\lambda[t_1], \lambda[t_1+1], \ldots, \lambda[t_2-1]) \in [0,1)^{t_2-t_1}$ is denoted by $\Phi_{t_1}^{t_2}(x, \boldsymbol{\lambda}) : \Omega \times [0,1)^{t_2-t_1} \to \Omega$ where $\Phi_{t_1}^{t_2}(x,\boldsymbol{\lambda}) \stackrel{\text{def.}}{=} \phi(\phi(\cdots(\phi(x,\lambda[t_1]),\ldots,\lambda[t_2-2]),\lambda[t_2-1])$. We say that a sequence $\boldsymbol{\lambda} \in [0,1)^{|T|}$ satisfies the *coalescence condition*, when $\exists y \in \Omega$, $\forall x \in \Omega$, $y = \Phi_T^0(x, \boldsymbol{\lambda})$.

Suppose that there exists a partial order "$\succeq$" on the set of states Ω. A transition rule expressed by a deterministic update function ϕ is called *monotone* (with respect to "$\succeq$") if $\forall \lambda \in [0,1)$, $\forall x, \forall y \in \Omega$, $x \succeq y \Rightarrow \phi(x,\lambda) \succeq \phi(y,\lambda)$. We also say that a chain is *monotone* if the chain has a *monotone* update function. Here we suppose that there exists a unique pair of states $(x_{\max}, x_{\min})$ in partially ordered set $(\Omega, \succeq)$, satisfying $x_{\max} \succeq x \succeq x_{\min}$, $\forall x \in \Omega$.

With these preparations, a standard monotone Coupling From The Past algorithm is expressed as follows.

Algorithm 2 (Monotone CFTP Algorithm [19])

Step 1. Set the starting time period $T := -1$ to go back, and set $\boldsymbol{\lambda}$ be the empty sequence.

Step 2. Generate random real numbers $\lambda[T], \lambda[T+1], \ldots, \lambda[\lceil T/2 \rceil - 1] \in [0,1)$, and insert them to the head of $\boldsymbol{\lambda}$ in order, i.e., put $\boldsymbol{\lambda} := (\lambda[T], \lambda[T+1], \ldots, \lambda[-1])$.

Step 3. Start two chains from $x_{\max}$ and $x_{\min}$, respectively, at time period T, and run each chain to time period 0 according to the update function ϕ with the sequence of numbers in $\boldsymbol{\lambda}$. (Here we note that every chain uses the common sequence $\boldsymbol{\lambda}$.)

Step 4. [Coalescence check] The state obtained at time period 0 is denoted by $\Phi_T^0(x, \boldsymbol{\lambda})$.
  (a) If $\exists y \in \Omega$, $y = \Phi_T^0(x_{\max}, \boldsymbol{\lambda}) = \Phi_T^0(x_{\min}, \boldsymbol{\lambda})$, then return y.
  (b) Else, update the starting time period $T := 2T$, and go to Step 2.

Theorem 3.1 (Monotone CFTP Theorem [19]) *Suppose that a Markov chain defined by an update function* ϕ *is monotone with respect to a partially ordered set of states* $(\Omega, \succeq)$, *and* $\exists x_{\max}, \exists x_{\min} \in \Omega$, $\forall x \in \Omega$, $x_{\max} \succeq x \succeq x_{\min}$. *Then the monotone CFTP algorithm (Algorithm 2) terminates with probability 1, and obtained value is a realization of a random variable exactly distributed according to the stationary distribution.* ■

Theorem 3.1 gives a (probabilistically) finite time algorithm for infinite time simulation.

3.2 Monotone Markov Chain

In this section we propose new Markov chain $\mathcal{M}_{\rm P}$. The transition rule of $\mathcal{M}_{\rm P}$ is defined by the following *update function* $\phi : \Xi \times [1, n) \to \Xi$. For a current state $X \in \Xi$, the next state $X' = \phi(X, \lambda) \in \Xi$ with respect to a random number $\lambda \in [1, n)$ is defined by

$$X'_i = \begin{cases} l & (\text{for } i = \lfloor\lambda\rfloor), \\ k - l & (\text{for } i = \lfloor\lambda\rfloor + 1), \\ X_i & (\text{otherwise}), \end{cases}$$

where $k = X_{\lfloor\lambda\rfloor} + X_{\lfloor\lambda\rfloor+1}$ and $l \in \{0, 1, \ldots, k\}$ satisfies

$$g^k_{\lfloor\lambda\rfloor(\lfloor\lambda\rfloor+1)}(l-1) < \lambda - \lfloor\lambda\rfloor \leq g^k_{\lfloor\lambda\rfloor(\lfloor\lambda\rfloor+1)}(l).$$

Our chain $\mathcal{M}_{\rm P}$ is a modification of $\mathcal{M}_{\rm A}$, obtained by restricting to choose only a consecutive pair of indices. Clearly, $\mathcal{M}_{\rm P}$ is ergodic. The chain has a unique stationary distribution $J(\boldsymbol{x})$ defined in Section 2.

In the following, we show the monotonicity of $\mathcal{M}_{\rm P}$. Here we introduce a partial order "$\succeq$" on Ξ. For any state $\boldsymbol{x} \in \Xi$, we introduce *cumulative sum vector* $c_{\boldsymbol{x}} = (c_{\boldsymbol{x}}(0), c_{\boldsymbol{x}}(1), \ldots, c_{\boldsymbol{x}}(n)) \in \mathbb{Z}^{n+1}_+$ defined by

$$c_{\boldsymbol{x}}(i) \stackrel{\text{def.}}{=} \begin{cases} 0 & (\text{for } i = 0), \\ \sum_{j=1}^{i} x_j & (\text{for } i \in \{1, 2, \ldots, n\}). \end{cases}$$

For any pair of states $\boldsymbol{x}, \boldsymbol{y} \in \Xi$, we say $\boldsymbol{x} \succeq \boldsymbol{y}$ if and only if $c_{\boldsymbol{x}} \geq c_{\boldsymbol{y}}$. Next, we define two special states $x_{\max}, x_{\min} \in \Xi(K)$ by $x_{\max} \stackrel{\text{def.}}{=} (K, 0, \cdots, 0)$ and $x_{\min} \stackrel{\text{def.}}{=} (0, \ldots, 0, K)$. Then we can see easily that $\forall \boldsymbol{x} \in \Xi(K)$, $x_{\max} \succeq \boldsymbol{x} \succeq x_{\min}$.

Theorem 3.2 *Markov chain $\mathcal{M}_{\rm P}$ is monotone on the partially ordered set $(\Xi(K), \succeq)$, i.e., $\forall\lambda \in [1, n)$, $\forall X, \forall Y \in \Xi(K)$, $X \succeq Y \Rightarrow \phi(X, \lambda) \succeq \phi(Y, \lambda)$.*

Outline of proof: We say that a state $X \in \Xi$ *covers* $Y \in \Xi$ (at j), denoted by $X \cdot\!\succ Y$ (or $X \cdot\!\succ_j Y$), when

$$X_i - Y_i = \begin{cases} +1 & (\text{for } i = j), \\ -1 & (\text{for } i = j + 1), \\ 0 & (\text{otherwise}). \end{cases}$$

We show that if a pair of states $X, Y \in \Xi$ satisfies $X \cdot\!\succ_j Y$, then $\forall\lambda \in [1, n)$, $\phi(X, \lambda) \succeq \phi(Y, \lambda)$. We denote $\phi(X, \lambda)$ by X' and $\phi(Y, \lambda)$ by Y' for simplicity. For any index $i \neq \lfloor\lambda\rfloor$, it is easy to see that $c_X(i) = c_{X'}(i)$ and $c_Y(i) = c_{Y'}(i)$, and so $c_{X'}(i) - c_{Y'}(i) = c_X(i) - c_Y(i) \geq 0$ since $X \succeq Y$. We can show that $c_{X'}(\lfloor\lambda\rfloor) \geq c_{Y'}(\lfloor\lambda\rfloor)$, with considering the following three cases, Case 1 [$\lfloor\lambda\rfloor \neq j - 1$ and $\lfloor\lambda\rfloor \neq j + 1$], Case 2 [$\lfloor\lambda\rfloor = j - 1$], and Case 3 [$\lfloor\lambda\rfloor = j + 1$].

For any pair of states X, Y satisfying $X \succeq Y$, it is easy to see that there exists a sequence of states $Z_1, Z_2, \ldots, Z_r$ with appropriate length satisfying $X = Z_1 \cdot\!\succ Z_2 \cdot\!\succ \cdots \cdot\!\succ Z_r = Y$. Then applying the above property repeatedly, we obtain that $\phi(X, \lambda) = \phi(Z_1, \lambda) \succeq \phi(Z_2, \lambda) \succeq \cdots \succeq \phi(Z_r, \lambda) = \phi(Y, \lambda)$. $\square$

Since $\mathcal{M}_{\rm P}$ is a monotone chain, we can design a perfect sampler based on monotone CFTP. We could also employ Wilson's read once algorithm [20] and Fill's interruptible algorithm [7, 8], each of which also gives a perfect sampler.

3.3 Expected Running Time

Here, we assume a condition, which gives expected polynomial time monotone CFTP algorithm.

Condition 2 *Parameters are arranged in non-increasing order i.e.,* $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n$.

The following is a main result of this paper.

Theorem 3.3 *Under Condition 2, the expected running time of our perfect sampler is bounded by* $\mathrm{O}(n^3 \ln K)$*, where* n *is the number of nodes and* K *is the number of customers in a closed Jackson network.*

We can show Theorem 3.3 by estimating the expectation of *coalescence time* $T_* \in \mathbb{Z}_{++}$ defined by $T_* \stackrel{\text{def.}}{=} \min\{t > 0 \mid \exists y \in \Xi,\ \forall x \in \Xi,\ y = \Phi^0_{-t}(x, \mathbf{\Lambda})\}$. Note that T_* is a random variable.

Outline of proof: Let $G = (\Xi, \mathcal{E})$ be the graph defined in the proof of Theorem 2.2 in Section 2. For each edge $e = \{X, Y\} \in \mathcal{E}$, there exists a unique pair of indices $j_1, j_2 \in \{1, 2, \ldots, n\}$ called the *supporting pair* of e, satisfying

$$|X_j - Y_j| = \begin{cases} 1 & \text{for } j \in \{j_1, j_2\}, \\ 0 & \text{otherwise.} \end{cases}$$

We define the length of $l_{\mathrm{P}}(e)$ of an edge $e = \{X, Y\} \in \mathcal{E}$ by $l_{\mathrm{P}}(e) \stackrel{\text{def.}}{=} (1/(n-1)) \sum_{i=1}^{j^*-1} (n-i)$, where $j^* = \max\{j_1, j_2\} \geq 2$ and $\{j_1, j_2\}$ is the supporting pair of e. Note that $1 \leq \min_{e \in \mathcal{E}} l_{\mathrm{P}}(e) \leq \max_{e \in \mathcal{E}} l_{\mathrm{P}}(e) \leq n/2$. For each pair $X, Y \in \Xi$, we define the distance $d_{\mathrm{P}}(X, Y)$ be the length of a shortest path between X and Y on G. Clearly, the length between $x_{\max}$ and $x_{\min}$ is bounded by Kn. Here we denote $X' = \phi(X, \lambda)$ and $Y' = \phi(Y, \Lambda)$ with uniform real random number $\Lambda \in [1, n)$, then Condition 2 implies

$$\mathrm{E}[d_{\mathrm{P}}(X', Y')] \leq \left(1 - \frac{1}{n(n-1)^2}\right) d_{\mathrm{P}}(X, Y)$$

for any pair $\{X, Y\} \in \mathcal{E}$, and we can prove the claim. ■

4 Concluding Remarks

We proposed FPRAS for closed Jackson networks with single servers. Our scheme is based on MCMC, and we proposed two rapidly mixing Markov chains. One is for approximate sampling, while the other is for perfect sampling. Though we omit the details, we can also construct an FPRAS using the perfect sampler in the same way as with the approximate sampler. At that time, we can show that it suffices to take $36nR^2\varepsilon^{-2}\ln(2R/\delta)$ samples in each iteration.

One of future works is to extend our FPRAS to closed Jackson networks with multiple serves. Extension to closed BCMP networks is another.

References

[1] D. Aldous, Random walks on finite groups and rapidly mixing Markov chains, in Séminarie de Probabilitiés XVII 1981/1982, vol. 986 of Springer-Verlag Lecture Notes in Mathematics, D. Dold and B. Eckmann, (ed.), Springer-Verlag, New York, 1983, 243–297.

[2] R. Bubley, Randomized Algorithms: Approximation, Generation, and Counting, Springer-Verlag, New York, 2001.

[3] R. Bubley and M. Dyer, Path coupling : a technique for proving rapid mixing in Markov chains, Proceedings of the 38th Annual Symposium on Foundations of Computer Science (FOCS 1997), 223–231.

[4] J. P. Buzen, Computational algorithms for closed queueing networks with exponential servers, Communications of the ACM, **16** (1973), 527–531.

[5] W. Chen and C. A. O'Cinneide, Towards a polynomial-time randomized algorithm for closed product-form networks, ACM Transactions on Modeling and Computer Simulation, **8** (1998), 227–253.

[6] M. Dyer and C. Greenhill, Polynomial-time counting and sampling of two-rowed contingency tables, Theoretical Computer Sciences, **246** (2000), 265–278.

[7] J. Fill, An interruptible algorithm for perfect sampling via Markov chains, The Annals of Applied Probability, **8** (1998), 131–162.

[8] J. Fill, M. Machida, D. Murdoch, and J. Rosenthal, Extension of Fill's perfect rejection sampling algorithm to general chains, Random Structures Algorithms, **17** (2000), 290–316.

[9] O. Häggström, Finite Markov Chains and Algorithmic Application, London Mathematical Society, Student Texts **52**, Cambridge University Press, 2002.

[10] J. R. Jackson, Networks of waiting lines, The Journal of Operations Research Society of America, **5** (1957), 518–521.

[11] J. R. Jackson, Jobshop-like queueing systems, Management Science, **10**(1) (1963), 131–142.

[12] M. Jerrum, Counting, Sampling and Integrating: Algorithms and Complexity, Lectures in Mathematics, ETH Zürich, Birkhauser Verlag, Basel, 2003.

[13] M. Jerrum and A. Sinclair, The Markov chain Monte Carlo method: an approach to approximate counting and integration, in Approximation Algorithm for NP-hard Problems, D. Hochbaum, (ed.), PWS, 1996, 482–520.

[14] S. Kijima and T. Matsui, Approximate counting scheme for $m \times n$ contingency tables , IEICE Transactions on Information and Systems, vol. **E87-D** (2004), pp. 308–314.

[15] S. Kijima and T. Matsui, Polynomial time perfect sampling algorithm for two-rowed contingency tables, Random Structures Algorithms, *to appear*.

[16] T. Matsui and S. Kijima, Polynomial time perfect sampler for discretized Dirichlet distribution, METR 2003-17, Mathematical Engineering Technical Reports, University of Tokyo, 2003. (available from http://www.keisu.t.u-tokyo.ac.jp/Research/techrep.0.html)

[17] T. Matsui, M. Motoki, and N. Kamatani, Polynomial time approximate sampler for discretized Dirichlet distribution, 14th International Symposium on Algorithms and Computation (ISAAC 2003), LNCS, **2906** (2003) 676–685.

[18] T. Ozawa, Perfect simulation of a closed Jackson network, ORSJ Queueing symposium, Hikone, 2004.

[19] J. Propp and D. Wilson, Exact sampling with coupled Markov chains and applications to statistical mechanics, Random Structures Algorithms, **9** (1996), 223–252.

[20] D. Wilson, How to couple from the past using a read-once source of randomness, Random Structures Algorithms, **16** (2000), 85–113.

APPENDIX

A Proofs

Lemma 2.1 The function g_{ij}^k satisfies the *alternating inequalities*,

$$g_{ij}^{k+1}(l) \le g_{ij}^k(l) \le g_{ij}^{k+1}(l+1), \quad \forall k \in \{1,\dots,K\},\ \forall l \in \{1,\dots,k\}.$$

Proof: First, we prove the former inequality $g_{ij}^{k+1}(l) \le g_{ij}^k(l)$ as follows,

$$\begin{aligned}
\frac{g_{ij}^k(l)}{g_{ij}^{k+1}(l)} &= \frac{\sum_{s=0}^{l} \alpha_i^s \alpha_j^{k-s}}{A_{ij}^k} \frac{A_{ij}^{k+1}}{\sum_{s=0}^{l} \alpha_i^s \alpha_j^{k+1-s}} = \frac{A_{ij}^{k+1}}{\alpha_j A_{ij}^k} \\
&= \frac{\sum_{s=0}^{k+1} \alpha_i^s \alpha_j^{k+1-s}}{\alpha_j \sum_{s=0}^{k} \alpha_i^s \alpha_j^{k-s}} = \frac{\sum_{s=0}^{k+1} \alpha_i^s \alpha_j^{k+1-s}}{\sum_{s=0}^{k} \alpha_i^s \alpha_j^{k+1-s}} \ge 1.
\end{aligned}$$

Next, we prove the latter inequality $g_{ij}^k(l) \le g_{ij}^{k+1}(l+1)$ as follows,

$$\begin{aligned}
\frac{g_{ij}^{k+1}(l+1)}{g_{ij}^k(l)} &= \frac{A_{ij}^k}{A_{ij}^{k+1}} \frac{\sum_{s=0}^{l+1} \alpha_i^s \alpha_j^{k+1-s}}{\sum_{s=0}^{l} \alpha_i^s \alpha_j^{k-s}} = \frac{\left(\sum_{s=0}^{k} \alpha_i^s \alpha_j^{k-s}\right)\left(\sum_{s=0}^{l+1} \alpha_i^s \alpha_j^{k+1-s}\right)}{\left(\sum_{s=0}^{k+1} \alpha_i^s \alpha_j^{k+1-s}\right)\left(\sum_{s=0}^{l} \alpha_i^s \alpha_j^{k-s}\right)} \\
&= \frac{\left(\sum_{s=0}^{k} \alpha_i^s \alpha_j^{k-s}\right)\left(\sum_{s=0}^{l} \alpha_i^s \alpha_j^{k+1-s} + \alpha_i^{l+1}\alpha_j^{k-l}\right)}{\left(\sum_{s=0}^{k} \alpha_i^s \alpha_j^{k+1-s} + \alpha_i^{k+1}\right)\left(\sum_{s=0}^{l} \alpha_i^s \alpha_j^{k-s}\right)} \\
&= \frac{\left(\sum_{s=0}^{k} \alpha_i^s \alpha_j^{k-s}\right)\left(\alpha_j \sum_{s=0}^{l} \alpha_i^s \alpha_j^{k-s}\right) + \alpha_i^{l+1}\alpha_j^{k-l}\left(\sum_{s=0}^{k} \alpha_i^s \alpha_j^{k-s}\right)}{\left(\alpha_j \sum_{s=0}^{k} \alpha_i^s \alpha_j^{k-s}\right)\left(\sum_{s=0}^{l} \alpha_i^s \alpha_j^{k-s}\right) + \alpha_i^{k+1}\left(\sum_{s=0}^{l} \alpha_i^s \alpha_j^{k-s}\right)} \\
&= \frac{\left(\alpha_i^{l+1}\alpha_j^{k-l}\right)^{-1} \alpha_j \left(\sum_{s=0}^{k} \alpha_i^s \alpha_j^{k-s}\right)\left(\sum_{s=0}^{l} \alpha_i^s \alpha_j^{k-s}\right) + \sum_{s=0}^{k} \alpha_i^s \alpha_j^{k-s}}{\left(\alpha_i^{l+1}\alpha_j^{k-l}\right)^{-1} \alpha_j \left(\sum_{s=0}^{k} \alpha_i^s \alpha_j^{k-s}\right)\left(\sum_{s=0}^{l} \alpha_i^s \alpha_j^{k-s}\right) + \alpha_i^{k-l}\alpha_j^{l-k} \sum_{s=0}^{l} \alpha_i^s \alpha_j^{k-s}} \\
&= \frac{\left(\alpha_i^{l+1}\alpha_j^{k-l}\right)^{-1} \alpha_j \left(\sum_{s=0}^{k} \alpha_i^s \alpha_j^{k-s}\right)\left(\sum_{s=0}^{l} \alpha_i^s \alpha_j^{k-s}\right) + \sum_{s=0}^{k} \alpha_i^s \alpha_j^{k-s}}{\left(\alpha_i^{l+1}\alpha_j^{k-l}\right)^{-1} \alpha_j \left(\sum_{s=0}^{k} \alpha_i^s \alpha_j^{k-s}\right)\left(\sum_{s=0}^{l} \alpha_i^s \alpha_j^{k-s}\right) + \sum_{s=k-l}^{k} \alpha_i^s \alpha_j^{k-s}} \ge 1.
\end{aligned}$$

Thus we obtain the claim. □

Lemma 2.4 The number of recursion R satisfies that $R \le n \ln K + 1$ for any $K \in \mathbb{Z}_{++}$.

Proof: If $n = 1$, then $R = 1$ hence we obtain the claim. If $n \ge 2$ and $K = 1, 2$ then $R = 1, 2$, respectively, hence we also obtain the claim. In the following, we consider the case $n \ge 2$ and $K \ge 3$. We define R' by

$$R' \overset{\text{def.}}{=} \min\left\{ r \;\middle|\; K\left(\frac{n-1}{n}\right)^r < 1 \right\},$$

then clearly $R \le R'$, since $K' - \lceil K'/n \rceil \le K'(n-1)/n$ for any $K' \in \mathbb{Z}_{++}$. Thus it is enough to show that

$$K \left(\frac{n-1}{n} \right)^{n \ln K} \le 1.$$

With considering $\ln K > 0$,

$$\left(\frac{n-1}{n} \right)^{n \ln K} = \left(\left(1 - \frac{1}{n} \right)^{n} \right)^{\ln K} \le \left(\frac{1}{\mathrm{e}} \right)^{\ln K} = 1/K.$$

Thus we obtain the claim. $\square$

Theorem 2.5 If we set $Q_{\mathrm{A}} = 144nR^2\varepsilon^{-2}\ln(2R/\delta)$ and $T_{\mathrm{A}}(K') = \left\lceil \frac{n(n-1)}{2} \ln \frac{6nRK'}{\varepsilon} \right\rceil$, then our randomized approximation scheme (Algorithm 1) returns Z satisfying

$$\Pr\left[|Z - G(K)| \le \varepsilon G(K) \right] \ge 1 - \delta.$$

Proof: In the following, we denote $\omega_j \stackrel{\text{def.}}{=} G'(K_j)/G(K_j)$ and $\widehat{\omega}_j \stackrel{\text{def.}}{=} \mathrm{E}[U_j/Q_{\mathrm{A}}]$ for $j \in \{1, 2, \ldots, R\}$, for simplicity.

(i) We show that $1 \le \forall j \le R$, the inequality $\widehat{\omega}_j \ge 1/n - \varepsilon/(6nR)$ holds. By the hypothesis of the theorem, $|\omega_j - \widehat{\omega}_j| \le \varepsilon/(6nR)$. Since $\omega_j \ge 1/n$, we can see that $\widehat{\omega}_j \ge \omega_j - \varepsilon/(6nR) \ge 1/n - \varepsilon/(6nR)$ for $1 \le j \le R$.

(ii) We show that $|\omega_j - \widehat{\omega}_j| \le \frac{\varepsilon \widehat{\omega}_j}{6R - \varepsilon}$. By using the result of (i), we have

$$|\omega_j - \widehat{\omega}_j| \le \frac{\varepsilon}{6nR} = \frac{\varepsilon}{6nR} \cdot \frac{1}{\widehat{\omega}_j} \cdot \widehat{\omega}_j \le \frac{\varepsilon}{6nR} \cdot \frac{\widehat{\omega}_j}{\left(\frac{1}{n} - \frac{\varepsilon}{6nR} \right)} = \frac{\varepsilon}{6R - \varepsilon} \cdot \widehat{\omega}_j.$$

(iii) We show that $\Pr[|Z_j - \widehat{\omega}_j| > (\varepsilon/(6R - \varepsilon))] \le \delta/R$. By employing the modified Chernoff bound in Lemma 2.4, we have

$$\Pr\left[\left| \tfrac{U+1}{M+1} - p \right| \ge \lambda p \right] \le 2\mathrm{e}^{-\frac{1}{4}\lambda^2 M p}.$$

By substituting the parameters $\lambda = \varepsilon/(6R - \varepsilon)$ and $p = \widehat{\omega}_j \ge 1/n - \varepsilon/(6nR)$, clearly $Q_{\mathrm{A}} \ge (4 + 2\sqrt{3})/(p\lambda)$ holds. We put $M = Q_{\mathrm{A}} = 144nR^2\varepsilon^{-2}\ln(2R/\delta)$ and $U = U_j$. Then we obtain that $Z_j = \frac{U_j + 1}{Q_{\mathrm{A}} + 1}$ satisfies

$$\begin{aligned} \Pr\left[|Z_j - \widehat{\omega}_j| > \tfrac{\varepsilon}{6R - \varepsilon} \widehat{\omega}_j \right] &\le 2\mathrm{e}^{-\left(\frac{\varepsilon}{6R-\varepsilon}\right)^2 \frac{1}{4} 144nR^2\varepsilon^{-2}\left(\ln \frac{2R}{\delta}\right)\widehat{\omega}_j} \\ &\le 2\mathrm{e}^{-\left(\frac{\varepsilon}{6R-\varepsilon}\right)^2 36nR^2\varepsilon^{-2}\left(\ln \frac{2R}{\delta}\right)\left(\frac{6R-\varepsilon}{6nR}\right)} \le 2\mathrm{e}^{-\left(\frac{6R}{6R-\varepsilon}\right)\left(\ln \frac{2R}{\delta}\right)} \le 2\mathrm{e}^{-\left(\ln \frac{2R}{\delta}\right)} = \frac{\delta}{R}. \end{aligned}$$

(iv) We show that $|(Z_1 \cdots Z_R)^{-1} - (\omega_1 \cdots \omega_R)^{-1}| \le \varepsilon(\omega_1 \cdots \omega_R)^{-1}$ with probability higher than $1 - \delta$. By the result of (ii), we obtain that

$$\frac{6R - 2\varepsilon}{6R - \varepsilon} \cdot \widehat{\omega}_j \le \omega_j \le \frac{6R}{6R - \varepsilon} \cdot \widehat{\omega}_j \tag{1}$$

By the result of (iii), we obtain that

$$\frac{6R - \varepsilon}{6R} \cdot Z_j \le \widehat{\omega}_j \le \frac{6R - \varepsilon}{6R - 2\varepsilon} \cdot Z_j, \tag{2}$$

with probability higher than $1-\delta/R$. By combining (1) and (2) with considering $Z_j > 0$, ω_j and Z_j satisfy

$$\frac{6R-2\varepsilon}{6R} \leq \frac{\omega_j}{Z_j} \leq \frac{6R}{6R-2\varepsilon} \quad \text{with probability higher than } 1-\delta/R.$$

The above inequality holds for each $1 \leq j \leq R$ and each Z_j follows i.i.d., thus with probability higher than $1-\delta$, the inequalities

$$\left(1-\frac{\varepsilon}{3R}\right)^R \leq \frac{\omega_1 \cdots \omega_R}{Z_1 \cdots Z_R} \leq \left(\frac{1}{1-\frac{\varepsilon}{3R}}\right)^R \tag{3}$$

hold. The right hand side of inequalities (3) satisfies,

$$\left(\frac{1}{1-\frac{\varepsilon}{3R}}\right)^R \quad = \quad \left(1+\frac{\varepsilon}{3R-\varepsilon}\right)^R \leq \left(1+\frac{\varepsilon}{3R-1}\right)^R \leq \mathrm{e}^{\frac{R}{3R-1}\varepsilon} \leq \mathrm{e}^{\frac{1}{2}\varepsilon} \leq 1+\varepsilon$$

The left hand side satisfies,

$$\left(1-\frac{\varepsilon}{3R}\right)^R = \left(\frac{1}{1-\frac{\varepsilon}{3R}}\right)^{-R} \geq \frac{1}{1+\varepsilon} \geq 1-\varepsilon.$$

Thus, the inequality (3) is transformed to

$$1-\varepsilon \leq \frac{\omega_1 \cdots \omega_R}{Z_1 \cdots Z_R} \leq 1+\varepsilon$$

accordingly. Then we have the result that with the probability higher than $1-\delta$,

$$|(Z_1 \cdots Z_R)^{-1} - (\omega_1 \cdots \omega_R)^{-1}| \leq \varepsilon(\omega_1 \cdots \omega_R)^{-1}.$$

(v) Since $\alpha_1^K(\omega_1 \cdots \omega_R)^{-1} = G(K)$ and $\alpha_1^K(Z_1 \cdots Z_R)^{-1} = Z$, we obtain the desired result that

$$\Pr\left[|Z - G(K)| \leq \varepsilon G(K)\right] \geq 1-\delta.$$

□

Lemma 2.6 Let X_i $(1 \leq i \leq M)$ be i.i.d. random variables such that $X_i = 1$ with probability p, and $X_i = 0$ with probability $1-p$. Let $U = \sum_{i=1}^M X_i$ and $0 < \lambda < 1$. If $M \geq (4+2\sqrt{3})/p\lambda$, then

1. $\Pr\left[p - \frac{U+1}{M+1} \geq \lambda p\right] \leq \mathrm{e}^{-\frac{1}{2}\lambda^2 Mp}$,
2. $\Pr\left[\frac{U+1}{M+1} - p \geq \lambda p\right] \leq \mathrm{e}^{-\frac{1}{4}\lambda^2 Mp}$, and
3. $\Pr\left[\left|\frac{U+1}{M+1} - p\right| \geq \lambda p\right] \leq 2\mathrm{e}^{-\frac{1}{4}\lambda^2 Mp}$,

hold.

Proof: 1. If $p - \frac{U+1}{M+1} \geq \lambda p$ then $p - \frac{U}{M} \geq \lambda p$. This implies that

$$\Pr\left[p - \frac{U+1}{M+1} \geq \lambda p\right] \leq \Pr\left[p - \frac{U}{M} \geq \lambda p\right] \leq \mathrm{e}^{-\frac{1}{2}\lambda^2 Mp},$$

where the last inequality is obtained by Chernoff bound.
2. By using Chernoff bound, the condition $M \geq (4+2\sqrt{3})/p\lambda$ implies that

$$\begin{aligned}
\Pr\left[\frac{U+1}{M+1} - p \geq \lambda p\right] &\leq \Pr\left[\frac{U}{M} + \frac{1}{M} - p \geq \lambda p\right] = \Pr\left[\frac{U}{M} - p \geq \left(\lambda - \frac{1}{Mp}\right) p\right] \\
&\leq \mathrm{e}^{-\frac{1}{3}\left(\lambda - \frac{1}{Mp}\right)^2 Mp} = \mathrm{e}^{-\frac{1}{3}\left(1 - \frac{1}{Mp\lambda}\right)^2 \lambda^2 Mp} \leq \mathrm{e}^{-\frac{1}{4}\lambda^2 Mp}.
\end{aligned}$$

3. By using the result of 1 and 2, clearly we have

$$\begin{aligned}
\Pr\left[\left|\frac{U+1}{M+1} - p\right| \geq \lambda p\right] &= \Pr\left[p - \frac{U+1}{M+1} \geq \lambda p\right] + \Pr\left[\frac{U+1}{M+1} - p \geq \lambda p\right] \\
&\leq \mathrm{e}^{-\frac{1}{2}\lambda^2 Mp} + \mathrm{e}^{-\frac{1}{4}\lambda^2 Mp} \leq 2\mathrm{e}^{-\frac{1}{4}\lambda^2 Mp}.
\end{aligned}$$

□

Theorem 2.7 If we set $Q_{\mathrm{A}} = 144nR^2\varepsilon^{-2}\ln(2R/\delta)$ and $T_{\mathrm{A}}(K') = \lceil \frac{n(n-1)}{2} \ln(\frac{6nRT'}{\varepsilon}) \rceil$, then the bias of the expectation of the obtained approximate solution is bounded as follows,

$$\frac{|\mathrm{E}[Z] - G(K)|}{G(K)} \leq \frac{\varepsilon}{4} + R\mathrm{e}^{-120R^2\varepsilon^{-2}\ln(2R/\delta)} \leq \left(\frac{1}{4} + \frac{1}{10^{36}}\right)\varepsilon.$$

Proof: Since $Z_1, \ldots, Z_R$ are independent,

$$\mathrm{E}[Z] = \alpha_1^K \mathrm{E}\left[\prod_{i=1}^{R} \frac{1}{Z_i}\right] = \alpha_1^K \prod_{i=1}^{R} \mathrm{E}\left[\frac{1}{Z_i}\right].$$

Now, we have

$$\begin{aligned}
\mathrm{E}\left[\frac{1}{Z_i}\right] &= \sum_{U_i=0}^{Q_{\mathrm{A}}} \frac{Q_{\mathrm{A}}+1}{U_i+1}\binom{Q_{\mathrm{A}}}{U_i}\widehat{\omega_i}^{U_i}(1-\widehat{\omega_i})^{Q_{\mathrm{A}}-U_i} = \frac{1}{\widehat{\omega_i}}\sum_{U_i=0}^{Q_{\mathrm{A}}}\binom{Q_{\mathrm{A}}+1}{U_i+1}\widehat{\omega_i}^{U_i+1}(1-\widehat{\omega_i})^{Q_{\mathrm{A}}-U_i} \\
&= \frac{1}{\widehat{\omega_i}}\left\{1 - (1-\widehat{\omega_i})^{Q_{\mathrm{A}}+1}\right\}.
\end{aligned}$$

Let $\gamma_i \stackrel{\text{def.}}{=} (1-\widehat{\omega_i})^{Q_{\mathrm{A}}+1}$, then the equality

$$\mathrm{E}[Z] = \alpha_1^K \prod_{i=1}^{R} \frac{1}{\widehat{\omega_i}}(1-\gamma_i) \tag{4}$$

holds, where we note that $0 \leq \gamma_i < 1$. From (4), $|G(K) - \mathrm{E}[Z]|$ satisfies

$$\begin{aligned}
|G(K) - \mathrm{E}[Z]| &= \left|\alpha_1^K \textstyle\prod_{i=1}^R \frac{1}{\omega_i} - \alpha_1^K \prod_{i=1}^R \frac{1}{\widehat{\omega_i}}(1-\gamma_i)\right| \\
&= \alpha_1^K \left|\textstyle\prod_{i=1}^R \frac{1}{\omega_i}(1-\gamma_i) + \left(\prod_{i=1}^R \frac{1}{\omega_i}\right)\left\{1 - \prod_{i=1}^R (1-\gamma_i)\right\} - \prod_{i=1}^R \frac{1}{\widehat{\omega_i}}(1-\gamma_i)\right| \\
&\leq \alpha_1^K \left|\textstyle\prod_{i=1}^R \frac{1}{\omega_i} - \prod_{i=1}^R \frac{1}{\widehat{\omega_i}}\right| \prod_{i=1}^R (1-\gamma_i) + \alpha_1^K \left(\prod_{i=1}^R \frac{1}{\omega_i}\right)\left\{1 - \prod_{i=1}^R (1-\gamma_i)\right\} \\
&\leq \alpha_1^K \left(\textstyle\prod_{i=1}^R \frac{1}{\omega_i}\right)\left|1 - \prod_{i=1}^R \frac{\omega_i}{\widehat{\omega_i}}\right| + \alpha_1^K \left(\prod_{i=1}^R \frac{1}{\omega_i}\right)\left\{1 - \prod_{i=1}^R (1-\gamma_i)\right\} \\
&= G(K)\left|1 - \textstyle\prod_{i=1}^R \frac{\omega_i}{\widehat{\omega_i}}\right| + G(K)\left\{1 - \prod_{i=1}^R (1-\gamma_i)\right\}.
\end{aligned}$$

If $1 - \prod_{i=1}^{R}(\omega_i/\widehat{\omega}_i) \leq 0$, we have that

$$\begin{aligned}\left|1 - \prod_{i=1}^{R} \frac{\omega_i}{\widehat{\omega}_i}\right| &= \prod_{i=1}^{R} \frac{\omega_i}{\widehat{\omega}_i} - 1 \leq \left(1 + \frac{\varepsilon}{6R - \varepsilon}\right)^R - 1 = 1 + \sum_{k=1}^{R} \binom{R}{k} \left(\frac{\varepsilon}{6R - \varepsilon}\right)^k - 1 \\ &\leq \sum_{k=1}^{R} \left(\frac{R}{6R - \varepsilon}\right)^k \varepsilon^k \leq \varepsilon \sum_{k=1}^{R} \left(\frac{1}{5}\right)^k \leq \varepsilon \sum_{k=1}^{\infty} \left(\frac{1}{5}\right)^k \leq \frac{\varepsilon}{4},\end{aligned}$$

since $R \geq 1 \geq \varepsilon$. Otherwise, $1 - \prod_{i=1}^{R}(\omega_i/\widehat{\omega}_i) > 0$ implies that

$$\left|1 - \prod_{i=1}^{R} \frac{\omega_i}{\widehat{\omega}_i}\right| = 1 - \prod_{i=1}^{R} \frac{\omega_i}{\widehat{\omega}_i} \leq 1 - \left(1 - \frac{\varepsilon}{6R - \varepsilon}\right)^R \leq \left(1 + \frac{\varepsilon}{6R - \varepsilon}\right)^R - 1 \leq \frac{\varepsilon}{4}.$$

Thus

$$|G(K) - \mathrm{E}[Z]| \leq \frac{\varepsilon}{4} G(K) + G(K) \left\{1 - \prod_{i=1}^{R}(1 - \gamma_i)\right\}.$$

Since $0 \leq \gamma_i < 1$ $(\forall i)$, it is easy to show that $1 - \prod_{i=1}^{R}(1 - \gamma_i) \leq \sum_{i=1}^{R} \gamma_i$ by induction on R. Acordingly, we have

$$\begin{aligned}& 1 - \prod_{i=1}^{R}(1 - \gamma_i) \leq \sum_{i=1}^{R} \gamma_i = \sum_{i=1}^{R}(1 - \widehat{\omega}_i)^{Q_{\mathrm{A}}+1} \leq R \left\{1 - \left(\tfrac{1}{n} - \tfrac{\varepsilon}{6nR}\right)\right\}^{Q_{\mathrm{A}}} \\ = \;& R \left\{1 - \left(\frac{1 - \frac{\varepsilon}{6R}}{n}\right)\right\}^{Q_{\mathrm{A}}} \leq R \left(1 - \frac{5}{6n}\right)^{Q_{\mathrm{A}}} \leq R \left(1 - \frac{5}{6n}\right)^{144nR^2\varepsilon^{-2}\ln(2R/\delta)} \\ \leq \;& R \left(\mathrm{e}^{-1}\right)^{\frac{5}{6n}144nR^2\varepsilon^{-2}\ln(2R/\delta)} = \mathrm{e}^{-120R^2\varepsilon^{-2}\ln(2R/\delta)}.\end{aligned}$$

Hence the bias is bounded as follows

$$|G(K) - \mathrm{E}[Z]| \leq \frac{\varepsilon}{4} G(K) + R\mathrm{e}^{-120R^2\varepsilon^{-2}\ln(2R/\delta)} G(K) \leq G(K) \left(\frac{\varepsilon}{4} + R\mathrm{e}^{-120R^2\varepsilon^{-2}\ln(2R/\delta)}\right).$$

□

Theorem 3.2 Markov chain $\mathcal{M}_{\mathrm{P}}$ is monotone on the partially ordered set $(\Xi(K), \succeq)$, i.e., $\forall \lambda \in [1, n)$, $\forall X, \forall Y \in \Xi(K)$, $X \succeq Y \Rightarrow \phi(X, \lambda) \succeq \phi(Y, \lambda)$.

Proof: We say that a state $X \in \Xi$ *covers* $Y \in \Xi$ (at j), denoted by $X \cdot\!\!\succ Y$ (or $X \cdot\!\!\succ_j Y$), when

$$X_i - Y_i = \begin{cases} +1 & (\text{for } i = j), \\ -1 & (\text{for } i = j + 1), \\ 0 & (\text{otherwise}). \end{cases}$$

We show that if a pair of states $X, Y \in \Xi$ satisfies $X \cdot\!\!\succ_j Y$, then $\forall \lambda \in [1, n)$, $\phi(X, \lambda) \succeq \phi(Y, \lambda)$. We denote $\phi(X, \lambda)$ by X' and $\phi(Y, \lambda)$ by Y' for simplicity. For any index $i \neq \lfloor \lambda \rfloor$, it is easy to see that $c_X(i) = c_{X'}(i)$ and $c_Y(i) = c_{Y'}(i)$, and so $c_{X'}(i) - c_{Y'}(i) = c_X(i) - c_Y(i) \geq 0$ since $X \succeq Y$. In the following, we show that $c_{X'}(\lfloor \lambda \rfloor) \geq c_{Y'}(\lfloor \lambda \rfloor)$.

Case 1: In case that $\lfloor \lambda \rfloor \neq j - 1$ and $\lfloor \lambda \rfloor \neq j + 1$. If we put $k = X_{\lfloor \lambda \rfloor} + X_{\lfloor \lambda \rfloor + 1}$, then it is easy to see that $Y_{\lfloor \lambda \rfloor} + Y_{\lfloor \lambda \rfloor + 1} = k$. Accordingly $X'_{\lfloor \lambda \rfloor} = Y'_{\lfloor \lambda \rfloor} = l$ where l satisfies

$$g^k_{\lfloor \lambda \rfloor (\lfloor \lambda \rfloor + 1)}(l - 1) \leq \lambda - \lfloor \lambda \rfloor < g^k_{\lfloor \lambda \rfloor (\lfloor \lambda \rfloor + 1)}(l),$$

hence $c_{X'}(\lfloor \lambda \rfloor) = c_{Y'}(\lfloor \lambda \rfloor)$.

Case 2: Consider the case that $\lfloor\lambda\rfloor = j-1$. Let $k+1 = X_{j-1}+X_j$. Then $Y_{j-1}+Y_j = k$, since $X \cdot\!\succ_j Y$. From the definition of cumulative sum vector,

$$\begin{aligned} & c_{X'}(\lfloor\lambda\rfloor) - c_{Y'}(\lfloor\lambda\rfloor) = c_{X'}(j-1) - c_{Y'}(j-1) \\ = \ & c_{X'}(j-2) + X'_{j-1} - c_{Y'}(j-2) - Y'_{j-1} = c_X(j-2) + X'_{j-1} - c_Y(j-2) - Y'_{j-1} \\ = \ & X'_{j-1} - Y'_{j-1}. \end{aligned}$$

Thus, it is enough to show that $X'_{j-1} \geq Y'_{j-1}$. Now suppose that $l \in \{0,1,\ldots,k\}$ satisfies $g^k_{(j-1)j}(l-1) \leq \lambda - \lfloor\lambda\rfloor < g^k_{(j-1)j}(l)$ for λ. Then $g^{k+1}_{(j-1)j}(l-1) \leq \lambda - \lfloor\lambda\rfloor < g^{k+1}_{(j-1)j}(l+1)$, since the alternating inequalities imply that $g^{k+1}_{(j-1)j}(l-1) \leq g^k_{(j-1)j}(l-1) < g^{k+1}_{(j-1)j}(l) \leq g^{k+1}_{(j-1)j}(l+1)$. Thus we have that if $Y'_{j-1} = l$ then $X'_{j-1} = l$ or $l+1$. In other words,

$$\begin{pmatrix} X'_{j-1} \\ Y'_{j-1} \end{pmatrix} \in \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \ldots, \begin{pmatrix} k \\ k \end{pmatrix}, \begin{pmatrix} k+1 \\ k \end{pmatrix} \right\}$$

and $X'_{j-1} \geq Y'_{j-1}$ in all cases. Accordingly, we have that $c_{X'}(\lfloor\lambda\rfloor) \geq c_{Y'}(\lfloor\lambda\rfloor)$.

Case 3: Consider the case that $\lfloor\lambda\rfloor = j+1$. We can show $c_{X'}(\lfloor\lambda\rfloor) \geq c_{Y'}(\lfloor\lambda\rfloor)$ in a similar way to Case 2.

For any pair of states X, Y satisfying $X \succeq Y$, it is easy to see that there exists a sequence of states $Z_1, Z_2, \ldots, Z_r$ with appropriate length satisfying $X = Z_1 \cdot\!\succ Z_2 \cdot\!\succ \cdots \cdot\!\succ Z_r = Y$. Then applying the above claim repeatedly, we obtain that $\phi(X,\lambda) = \phi(Z_1,\lambda) \succeq \phi(Z_2,\lambda) \succeq \cdots \succeq \phi(Z_r,\lambda) = \phi(Y,\lambda)$. $\square$

Theorem 3.3 Under Condition 2, the expected running time of our perfect sampler is bounded by $\mathrm{O}(n^3 \ln K)$, where n is the number of stations and K is the number of customers in a closed Jackson network.

To show the above theorem, we need following three lemmas.

Lemma A.1 *Under Condition 2, the mixing rate τ of our Markov chain $\mathcal{M}$ satisfies*

$$\tau \leq n(n-1)^2(1 + \ln Kn).$$

Proof: Let $G = (\Xi, \mathcal{E})$ be an undirected simple graph with vertex set Ξ and edge set $\mathcal{E}$ defined as follows. A pair of vertices $\{X, Y\}$ is an edge if and only if $(1/2)\sum_{i=1}^n |X_i - Y_i| = 1$. Clearly, the graph G is connected. For each edge $e = \{X, Y\} \in \mathcal{E}$, there exists a unique pair of indices $j_1, j_2 \in \{1, \ldots, n\}$, called the *supporting pair* of e, satisfying

$$|X_i - Y_i| = \begin{cases} 1 & (i = j_1, j_2), \\ 0 & (\text{otherwise}). \end{cases}$$

We define the length $l(e)$ of an edge $e = \{X, Y\} \in \mathcal{E}$ by $l(e) \stackrel{\text{def.}}{=} (1/(n-1))\sum_{i=1}^{j^*-1}(n-i)$ where $j^* = \max\{j_1, j_2\} \geq 2$ and $\{j_1, j_2\}$ is the supporting pair of e. Note that $1 \leq \min_{e \in \mathcal{E}} l(e) \leq \max_{e \in \mathcal{E}} l(e) \leq n/2$. For each pair $X, Y \in \Xi$, we define the distance $d(X, Y)$ be the length of a shortest path between X and Y on G. Clearly, the diameter of G, i.e., $\max_{(X,Y) \in \Xi^2} d(X, Y)$, is bounded by $Kn/2$, since $d(X, Y) \leq (n/2)\sum_{i=1}^n (1/2)|X_i - Y_i| \leq (n/2)K$ for any $(X, Y) \in \Xi^2$. The definition of edge length implies that for any edge $\{X, Y\} \in \mathcal{E}$, $d(X, Y) = l(\{X, Y\})$.

We define a joint process $(X,Y) \to (X',Y')$ as $(X,Y) \to (\phi(X,\Lambda),\phi(Y,\Lambda))$ with uniform real random number $\Lambda \in [1,n)$ and the update function ϕ defined in Subsection 3.2. Now we show that

$$\mathrm{E}[d(X',Y')] \leq \beta \cdot d(X,Y) \text{ where } \beta = 1 - 1/(n(n-1)^2), \tag{5}$$

for any pair $\{X,Y\} \in \mathcal{E}$. In the following, we denote the supporting pair of $\{X,Y\}$ by $\{j_1,j_2\}$. Without loss of generality, we can assume that $j_1 < j_2$, and $X_{j_2}+1 = Y_{j_2}$.

Case 1: When $\lfloor\Lambda\rfloor = j_2 - 1$, we will show that

$$\mathrm{E}[d(X',Y') \mid \lfloor\Lambda\rfloor = j_2 - 1] \leq d(X,Y) - (1/2)(n-j_2+1)/(n-1).$$

In case $j_1 = j_2 - 1$, $X' = Y'$ with conditional probability 1. Hence $d(X',Y') = 0$. In the following, we consider the case $j_1 < j_2 - 1$. Put $k' = X_{j_2-1} + X_{j_2}$ and $k'' = Y_{j_2-1} + Y_{j_2}$. Since $X_{j_2}+1 = Y_{j_2}$, $k'+1 = k''$ holds. From the definition of the update function of our Markov chain, we have the followings,

$$\begin{array}{rcl} X'_{j_2-1} = l & \Leftrightarrow & [g^{k'}_{(j_2-1)j_2}(l-1) \leq \Lambda - \lfloor\Lambda\rfloor < g^{k'}_{(j_2-1)j_2}(l)] \\ Y'_{j_2-1} = l & \Leftrightarrow & [g^{k'+1}_{(j_2-1)j_2}(l-1) \leq \Lambda - \lfloor\Lambda\rfloor < g^{k'+1}_{(j_2-1)j_2}(l)]. \end{array}$$

Now, the alternating inequalities

$$\begin{array}{rl} 0 < & g^{k'+1}_{(j_2-1)j_2}(0) = g^{k'}_{(j_2-1)j_2}(0) \leq g^{k'+1}_{(j_2-1)j_2}(1) \leq g^{k'}_{(j_2-1)j_2}(1) \leq \cdots \\ \leq & g^{k'+1}_{(j_2-1)j_2}(k') \leq g^{k'}_{(j_2-1)j_2}(k') = g^{k'+1}_{(j_2-1)j_2}(k'+1) = 1, \end{array}$$

hold. Thus we have

$$\begin{pmatrix} X'_{j_2-1} \\ Y'_{j_2-1} \end{pmatrix} \in \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \ldots, \begin{pmatrix} k' \\ k' \end{pmatrix}, \begin{pmatrix} k' \\ k'+1 \end{pmatrix} \right\}.$$

If $X'_{j_2-1} = Y'_{j_2-1}$, the supporting pair of $\{X',Y'\}$ is $\{j_1,j_2\}$ and so $d(X',Y') = d(X,Y)$. If $X'_{j_2-1} \neq Y'_{j_2-1}$, the supporting pair of $\{X',Y'\}$ is $\{j_1, j_2-1\}$ and so $d(X',Y') = d(X,Y) - (n-j_2+1)/(n-1)$.

Lemma A.2 (proved later) implies that if $\alpha_{j_2-1} \geq \alpha_{j_2}$, then

$$\begin{array}{rl} & \Pr[X'_{j_2-1} \neq Y'_{j_2-1} \mid \lfloor\Lambda\rfloor = j_2-1] - \Pr[X'_{j_2-1} = Y'_{j_2-1} \mid \lfloor\Lambda\rfloor = j_2-1] \\ = & \sum_{l=0}^{k'} \left(g^{k'}_{(j_2-1),j_2}(l) - g^{k'+1}_{(j_2-1),j_2}(l) \right) \\ & - \sum_{l=1}^{k'} \left(g^{k'+1}_{(j_2-1),j_2}(l) - g^{k'}_{(j_2-1),j_2}(l-1) \right) - g^{k'+1}_{(j_2-1),j_2}(0) \geq 0. \end{array}$$

Hence

$$\begin{array}{rcl} \Pr[X'_{j_2-1} = Y'_{j_2-1} \mid \lfloor\Lambda\rfloor = j_2-1] & \leq & (1/2), \\ \Pr[X'_{j_2-1} \neq Y'_{j_2-1} \mid \lfloor\Lambda\rfloor = j_2-1] & \geq & (1/2). \end{array}$$

Thus we obtain that

$$\begin{array}{rcl} \mathrm{E}[d(X',Y') | \lfloor\Lambda\rfloor = j_2-1] & \leq & (1/2)d(X,Y) + (1/2)(d(X,Y) - (n-j_2+1)/(n-1)) \\ & = & d(X,Y) - (1/2)(n-j_2+1)/(n-1). \end{array}$$

Case 2: When $\lfloor\Lambda\rfloor = j_2$, we can show that $\mathrm{E}[d(X',Y') | \lfloor\Lambda\rfloor = j_2] \leq d(X,Y) + (1/2)(n-j_2)/(n-1)$ in a similar way to Case 1.

Case 3: When $\lfloor \Lambda \rfloor \neq j_2 - 1$ and $\lfloor \Lambda \rfloor \neq j_2$, it is easy to see that the supporting pair $\{j'_1, j'_2\}$ of $\{X', Y'\}$ satisfies $j_2 = \max\{j'_1, j'_2\}$. Thus $d(X,Y) = d(X',Y')$.

The probability of appearance of Case 1 is equal to $1/(n-1)$, and that of Case 2 is less than or equal to $1/(n-1)$. From the above,

$$\begin{aligned}\mathrm{E}[d(X',Y')] &\leq d(X,Y) - \frac{1}{n-1}\cdot\frac{1}{2}\cdot\frac{n-j_2+1}{n-1} + \frac{1}{n-1}\cdot\frac{1}{2}\cdot\frac{n-j_2}{n-1} = d(X,Y) - \frac{1}{2(n-1)^2} \\ &\leq \left(1 - \frac{1}{2(n-1)^2}\cdot\frac{1}{\max_{\{X,Y\}\in\mathcal{E}}\{d(X,Y)\}}\right) d(X,Y) = \left(1 - \frac{1}{n(n-1)^2}\right) d(X,Y).\end{aligned}$$

Since the diameter of G is bounded by $Kn/2$, Theorem 2.3 implies that the mixing rate τ satisfies $\tau \leq n(n-1)^2(1+\ln(Kn/2))$. $\square$

Lemma A.2 *When $\alpha_i \geq \alpha_j > 0$, the inequality*

$$\sum_{l=0}^{k}\left(g_{ij}^{k}(l) - g_{ij}^{k+1}(l)\right) - \sum_{l=1}^{k}\left(g_{ij}^{k+1}(l) - g_{ij}^{k}(l-1)\right) - g_{ij}^{k+1}(0) \geq 0.$$

holds.

Proof: We can transform the left-hand side as

$$\begin{aligned}&\textstyle\sum_{l=0}^{k}\left(g_{ij}^{k}(l) - g_{ij}^{k+1}(l)\right) - \sum_{l=1}^{k}\left(g_{ij}^{k+1}(l) - g_{ij}^{k}(l-1)\right) - g_{ij}^{k+1}(0) \\ &= \textstyle\sum_{l=0}^{k}\left(g_{ij}^{k}(l) - g_{ij}^{k+1}(l)\right) - \sum_{l=0}^{k-1}\left(g_{ij}^{k+1}(k-l) - g_{ij}^{k}(k-l-1)\right) - g_{ij}^{k+1}(0) \\ &= \textstyle\sum_{l=0}^{k-1}\left(g_{ij}^{k}(l) - g_{ij}^{k+1}(l) - g_{ij}^{k+1}(k-l) + g_{ij}^{k}(k-l-1)\right) + 1 - g_{ij}^{k+1}(k) - g_{ij}^{k+1}(0),\end{aligned}$$

and we can see that,

$$\begin{aligned}1 - g_{ij}^{k+1}(k) - g_{ij}^{k+1}(0) &= 1 - \frac{\sum_{s=0}^{k}\alpha_i^s\alpha_j^{k+1-s}}{\sum_{s=0}^{k+1}\alpha_i^s\alpha_j^{k+1-s}} - \frac{\sum_{s=0}^{0}\alpha_i^s\alpha_j^{k+1-s}}{\sum_{s=0}^{k+1}\alpha_i^s\alpha_j^{k+1-s}} \\ &= \frac{\alpha_i^{k+1}}{\sum_{s=0}^{k+1}\alpha_i^s\alpha_j^{k+1-s}} - \frac{\alpha_j^{k+1}}{\sum_{s=0}^{k+1}\alpha_i^s\alpha_j^{k+1-s}} \geq 0,\end{aligned}$$

since $\alpha_i \geq \alpha_j$ (Condition 2). Thus it is enough to show that

$$g_{ij}^{k}(l) - g_{ij}^{k+1}(l) - g_{ij}^{k+1}(k-l) + g_{ij}^{k}(k-l-1) \geq 0 \quad \text{for any } l \ (0 \leq l \leq k-1).$$

By transforming the left-hand side, we can see that

$$\begin{aligned}&g_{ij}^{k}(l) - g_{ij}^{k+1}(l) - g_{ij}^{k+1}(k-l) + g_{ij}^{k}(k-l-1) \\ &= g_{ij}^{k}(l) - g_{ij}^{k+1}(l) - \frac{\sum_{s=0}^{k-l}\alpha_i^s\alpha_j^{k+1-s}}{\sum_{s=0}^{k+1}\alpha_i^s\alpha_j^{k+1-s}} + \frac{\sum_{s=0}^{k-l-1}\alpha_i^s\alpha_j^{k-s}}{\sum_{s=0}^{k}\alpha_i^s\alpha_j^{k-s}} \\ &= g_{ij}^{k}(l) - g_{ij}^{k+1}(l) - \left(1 - \frac{\sum_{s=k-l+1}^{k+1}\alpha_i^s\alpha_j^{k+1-s}}{\sum_{s=0}^{k+1}\alpha_i^s\alpha_j^{k+1-s}}\right) + \left(1 - \frac{\sum_{s=k-l}^{k}\alpha_i^s\alpha_j^{k-s}}{\sum_{s=0}^{k}\alpha_i^s\alpha_j^{k-s}}\right) \\ &= \frac{\sum_{s=0}^{l}\alpha_i^s\alpha_j^{k-s}}{A_{ij}^{k}} - \frac{\sum_{s=0}^{l}\alpha_i^s\alpha_j^{k+1-s}}{A_{ij}^{k+1}} + \frac{\sum_{s=k-l+1}^{k+1}\alpha_i^s\alpha_j^{k+1-s}}{A_{ij}^{k+1}} - \frac{\sum_{s=k-l}^{k}\alpha_i^s\alpha_j^{k-s}}{A_{ij}^{k}} \\ &= \frac{\sum_{s=0}^{l}\alpha_i^s\alpha_j^{k-s}}{A_{ij}^{k}} - \frac{\sum_{s=0}^{l}\alpha_i^s\alpha_j^{k+1-s}}{A_{ij}^{k+1}} + \frac{\sum_{s=0}^{l}\alpha_i^{k+1-s}\alpha_j^{s}}{A_{ij}^{k+1}} - \frac{\sum_{s=0}^{l}\alpha_i^{k-s}\alpha_j^{s}}{A_{ij}^{k}}\end{aligned}$$

$$
\begin{aligned}
&= \left(\tfrac{1}{A_{ij}^{k}} - \tfrac{\alpha_j}{A_{ij}^{k+1}}\right) \textstyle\sum_{s=0}^{l} \alpha_i^s \alpha_j^{k-s} + \left(\tfrac{\alpha_i}{A_{ij}^{k+1}} - \tfrac{1}{A_{ij}^{k}}\right) \textstyle\sum_{s=0}^{l} \alpha_i^{k-s} \alpha_j^s \\
&= \tfrac{\sum_{s=0}^{l} \alpha_i^s \alpha_j^{k-s}}{A_{ij}^{k} A_{ij}^{k+1}} \left(\textstyle\sum_{s=0}^{k+1} \alpha_i^s \alpha_j^{k+1-s} - \sum_{s=0}^{k} \alpha_i^s \alpha_j^{k+1-s} \right) \\
&\quad + \tfrac{\sum_{s=0}^{l} \alpha_i^{k-s} \alpha_j^{s}}{A_{ij}^{k} A_{ij}^{k+1}} \left(\textstyle\sum_{s=1}^{k+1} \alpha_i^s \alpha_j^{k+1-s} - \sum_{s=0}^{k+1} \alpha_i^s \alpha_j^{k+1-s} \right) \\
&= \tfrac{1}{A_{ij}^{k} A_{ij}^{k+1}} \left(\alpha_i^{k+1} \textstyle\sum_{s=0}^{l} \alpha_i^s \alpha_j^{k-s} - \alpha_j^{k+1} \sum_{s=0}^{l} \alpha_i^{k-s} \alpha_j^{s} \right) \\
&= \tfrac{1}{A_{ij}^{k} A_{ij}^{k+1}} \textstyle\sum_{s=0}^{l} \left(\alpha_i^{k+1+s} \alpha_j^{k-s} - \alpha_i^{k-s} \alpha_j^{k+1+s} \right) \\
&= \tfrac{1}{A_{ij}^{k} A_{ij}^{k+1}} \textstyle\sum_{s=0}^{l} \left(\alpha_i^{k-s} \alpha_j^{k-s} \left(\alpha_i^{2s+1} - \alpha_j^{2s+1} \right) \right) \geq 0,
\end{aligned}
$$

since $\alpha_i \geq \alpha_j$ (Condition 2). Thus we obtain the claim. $\square$

Next we estimate the expectation of the coalescence time of $\mathcal{M}_{\mathrm{P}}$.

Lemma A.3 *Under Condition 2, the coalescence time T_* of $\mathcal{M}_{\mathrm{P}}$ satisfies* $\mathrm{E}[T_*] = \mathrm{O}(n^3 \ln Kn)$.

Proof: Let $G = (\Xi, \mathcal{E})$ be the undirected graph and $d(X,Y)$, $\forall X, \forall Y \in \Xi$, be the metric on G, both of which are defined in the proof of Lemma A.1. We define $D \stackrel{\text{def.}}{=} d(x_{\max}, x_{\min})$ and $\tau_0 \stackrel{\text{def.}}{=} n(n-1)^2(1 + \ln D)$. By using the inequality (5) obtained in the proof of Lemma A.1, we have

$$
\begin{aligned}
\Pr[T_* > \tau_0] &= \Pr\left[\Phi_{-\tau_0}^{0}(x_{\max}, \mathbf{\Lambda}) \neq \Phi_{-\tau_0}^{0}(x_{\min}, \mathbf{\Lambda})\right] = \Pr\left[\Phi_{0}^{\tau_0}(x_{\max}, \mathbf{\Lambda}) \neq \Phi_{0}^{\tau_0}(x_{\min}, \mathbf{\Lambda})\right] \\
&\leq \textstyle\sum_{(X,Y)\in\Xi^2} d(X,Y) \Pr\left[X = \Phi_{0}^{\tau_0}(x_{\max}, \mathbf{\Lambda}), Y = \Phi_{0}^{\tau_0}(x_{\min}, \mathbf{\Lambda})\right] \\
&= \mathrm{E}\left[d\left(\Phi_{0}^{\tau_0}(x_{\max}, \mathbf{\Lambda}), \Phi_{0}^{\tau_0}(x_{\min}, \mathbf{\Lambda})\right)\right] \leq \left(1 - \frac{1}{n(n-1)^2}\right)^{\tau_0} d(x_{\max}, x_{\min}) \\
&= \left(1 - \frac{1}{n(n-1)^2}\right)^{n(n-1)^2(1+\ln D)} D \leq \mathrm{e}^{-1} \mathrm{e}^{-\ln D} D = \frac{1}{\mathrm{e}}.
\end{aligned}
$$

By the *submultiplicativity* of coalescence time ([19]), for any $k \in \mathbb{Z}_+$, $\Pr[T_* > k\tau_0] \leq (\Pr[T_* > \tau_0])^k \leq (1/\mathrm{e})^k$. Thus

$$
\begin{aligned}
\mathrm{E}[T_*] &= \textstyle\sum_{t=0}^{\infty} t \Pr[T_* = t] \leq \tau_0 + \tau_0 \Pr[T_* > \tau_0] + \tau_0 \Pr[T_* > 2\tau_0] + \cdots \\
&\leq \tau_0 + \tau_0/\mathrm{e} + \tau_0/\mathrm{e}^2 + \cdots = \tau_0/(1 - 1/\mathrm{e}) \leq 2\tau_0.
\end{aligned}
$$

Clearly $D \leq Kn$, then we obtain the result that $\mathrm{E}[T_*] = \mathrm{O}(n^3 \ln Kn)$. $\square$

Proof of Theorem 3.3: Let T_* be the coalescence time of our chain. Clearly T_* is a random variable. Put $m = \lceil \log_2 T_* \rceil$. Algorithm 2 terminates when we set the starting time period $T = -2^m$ at $(m+1)$st iteration. Then the total number of simulated transitions is bounded by $2(2^0 + 2^1 + 2^2 + \cdots + 2^K) < 2 \cdot 2 \cdot 2^m \leq 8T_*$, since we need to execute two chains from both $x_{\max}$ and $x_{\min}$. Thus the expectation of total number of transitions of $\mathcal{M}$ is bounded by $\mathrm{O}(\mathrm{E}[8T_*]) = \mathrm{O}(n^3 \ln Kn)$. $\square$