Directed FCFS infinite bipartite matching

Gideon Weiss

Received: 1 December 2019 / Revised: 9 November 2020 / Accepted: 11 November 2020 /
Published online: 1 December 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract
We consider an infinite sequence consisting of agents of several types and goods of several types, with a bipartite compatibility graph between agent and good types. Goods are matched with agents that appear earlier in the sequence using FCFS matching, if such are available, and are lost otherwise. This model may be used for two-sided queueing applications such as ride sharing, Web purchases, organ transplants, and for parallel redundant service queues. For this model, we calculate matching rates and delays. These can be used to obtain waiting times and help with design questions for related service systems. We explore some relations of this model to other FCFS stochastic matching models.

Keywords Dynamic online matching · FCFS matching · Matching rates · Matching delays · Two-sided queues · Parallel service systems

Mathematics Subject Classification 60K25 · 68M20 · 90B22

1 Introduction
Classical queueing models have random streams of customer arrivals and fixed sets of servers. Recently, however, more and more applications are better modeled by two-sided queues, where the fixed sets of servers are replaced by streams of arriving servers. The simplest example of such a system is the taxi stand, where cabs and passengers arrive at random times and depart at the moment that they meet [10]. More general systems where two-sided models seem appropriate include ride sharing systems, organ transplants, Web postings and searches, job markets, trading and auctions. These models share a more symmetric role of the so-called customers and servers, and
in particular, in these models service durations are replaced by inter-arrival times. In the simplest taxi stand model, the symmetric version where both cabs and passengers can wait is always either transient or null-recurrent, and to be useful the model needs to include finite waiting rooms or abandonments. In particular, one may assume that cabs always wait, while passengers abandon and are lost if they do not find a cab at the stand. This seems also to be the model appropriate for ride sharing systems like Uber and Lyft [9,15,18], for transplants of perishable organs or blood transfusions [25], perishable inventory systems [14,21], trading systems [6,13,17], and models of the sharing economy [22]. Our results in this paper are relevant for the following two-sided parallel service model: Agents of several types arrive to the system and wait for goods, and goods of several types arrive and match and depart with a compatible agent, or, if no match is immediately available for an arriving good, it is lost. Compatibility is determined by a bipartite graph. The system is controlled by a matching policy—which of the available compatible agents is assigned the arriving good. In this paper, we focus on the first come, first served (FCFS) policy: The arriving good is assigned to the longest waiting compatible agent. We do not, however, study this system directly, instead we consider the following discrete time matching model.

We consider a matching system in which agents of several types, indexed by $c_i \in C = \{c_1, \ldots, c_I\}$, are matched with several types of goods, indexed by $s_j \in S = \{s_1, \ldots, s_J\}$, subject to a bipartite compatibility graph $G$, such that $(s_j, c_i) \in G$ if agent type $c_i$ is compatible with good type $s_j$. We assume an infinite sequence of agents and goods, which is chosen i.i.d. from $C \cup S$. Matching is done on a first come, first served (FCFS) basis, where goods are only matched to agents that appear earlier in the sequence, and goods that find no compatible match are lost. We call this a directed matching model, to distinguish it from a symmetric model where goods are not lost, and can be matched to agents that appear later in the sequence. This model is an exact discrete time analog of the two-sided parallel service system if arrivals of agents and goods form independent Poisson processes. Moreover, this discrete model has the important advantage that it covers the most general arrival processes, stationary or time varying, as long as types of successive arrivals are i.i.d. Data for this more abstract model are very compact; it consists of the compatibility graph and the frequencies of the types.

The directed matching model was introduced in [2], where the following results were derived:

- The matching is stable if and only if there are enough compatible goods for every subset of agent types (see Definition 2).
- Under stability, there is almost surely a unique matching for the bi-infinite sequence of agents and goods (see Uniqueness Theorem 2).
- Exchanging the positions of each matched pair in the bi-infinite sequence results in an i.i.d sequence, in which the same matches are now directed FCFS in reversed time (see Reversibility Theorem 3).
- The stationary distribution of unmatched agents at every stage along the sequence is derived up to a normalizing constant (see Product Form Theorem 1).

Several questions were left unanswered in [2] and motivate the current paper. Our main results are:
We find a detailed Markov description of the matching process, obtain its stationary distribution, which is of a very simple intuitive form, and find a simple characterization of the possible states (see Sect. 3).

We calculate the normalizing constant which was not obtained in [2] (see Sect. 4).

We calculate the matching rates, i.e., long-range frequencies of \((s_j, c_i) \in G\) matches, and the fraction of unmatched goods of each type (see Sect. 5).

We obtain the distributions of delays, i.e., the distances between matched agent and good pairs, and derive explicit expressions for their generating functions and first and second moments (see Sect. 6).

We show that in heavy traffic, the directed bipartite matching model converges in a certain sense to the symmetric bipartite matching model of [1,3,8,19] (see Sect. 7).

Our model in this paper is closely related to the model in [1] and to the follow-up models in [16,19,20]. The model of [1] again consists of agent and goods types \(C\) and a bipartite compatibility graph, but instead of a single sequence of both agents and goods, there are two sequences, one of agents and one of goods, and matching is FCFS; however, goods may be matched to agents appearing later in the sequence. The model of [19] consists of a single set of types and a general compatibility graph. In addition to FCFS, several other policies are considered in [16,20]. For these models, as well as for the model in this paper, a uniqueness and reversibility theorem and stationary distributions are obtained. The methods are similar, but their execution is different enough to require separate derivations. We will discuss the relations between these models in Sect. 7.

Several papers that consider parallel service systems with customers and servers of several types and bipartite compatibility graph are closely related to the matching model of this paper and of [1,19]. The stationary distribution of the parallel service system with Poisson arrivals and exponential service times, derived in [4], converges to the bipartite matching model of [3] in heavy traffic. In [12], Gardner et al. introduce another kind of FCFS policy, called a redundancy policy: each server maintains its own FCFS queue, and customers make independent copies of their job and join the queues of all the compatible servers, so that they may be served simultaneously by more than one server. The customer then departs when his service at the first of the compatible servers is complete. This is also called cancel-on-complete in [7], to distinguish it from the model of [4] which is referred to as cancel-on-start. If service times are exponential, then simultaneous service until first completion incurs no loss, and with Poisson arrivals the distribution of the queue lengths is shown to be of product form. Adan et al. [2] compare the policies of [4] and [12] and discuss their relations to our FCFS directed bipartite matching model. Surprisingly, Adan et al [2] find that the stationary Markov chain that describes the redundancy service system is the same as the one describing the unmatched agents in our model. Further research on this type of model is surveyed in [11].

In [5], a heuristic based on calculation of matching rates for the symmetric FCFS bipartite matching model is used in the design of parallel service systems under many server scaling. A similar design heuristic based on the calculations of matching rates
for the directed FCFS matching model may be useful for design of two-sided systems such as ride sharing.

We now briefly describe our results in the following sections. In Sect. 2, we define the directed FCFS bipartite matching model and summarize the results of [2], which form the basis for our analysis in this paper.

In Sect. 3, we consider the following detailed Markov chain that describes the evolution of the directed FCFS matching: We perform matching of all goods up to time $N$ (i.e., position $N$ in the sequence), and with each match we exchange the positions of the matched agent and good. We then define the state as the list of items in the sequence, starting with the first unmatched agent and ending at the $N$th item. The list includes unmatched agents, unmatched and lost goods, and matched and exchanged agents and goods. We show in Theorem 4 that the probability of such a list is simply the product of the frequencies of the items, multiplied by a normalizing constant. This simplicity is deceptive, since it hides dependence on the constraints of what states are actually admissible (i.e., states that occur with positive probability). We next obtain a simple characterization of admissible states in Theorem 5. The results in this section are analogous to similar results in [1] and [19].

In Sect. 4, we consider a more concise Markov chain associated with our model. We again consider making all the matches up to a discrete time $N$ and list the agents that remain unmatched. The list of these unmatched agents up to $N$ is a Markov chain, and its stationary product form distribution was given in [2], following the derivation in [12]. A more concise Markov chain includes just the identities of which types appear in the list of unmatched agents, in their order of appearance in the list, and the count of how many additional unmatched agents are between them. This is also a Markov chain, and we derive its stationary distribution in Theorem 6. We use this more tractable Markov chain to calculate the normalizing constant, which is also the normalizing constant for the detailed process and the process of [2].

In Sect. 5, we derive expressions for the matching rates (Definition 4), i.e., the long-range fraction of matches for each pair $s_j, c_i \in \mathcal{G}$, and the fraction of goods of each type that remain unmatched. It is conjectured that the calculation of the matching rates (as well as the normalizing constant) is $\#P$-complete. Nevertheless, the calculation is straightforward, and we list an algorithm to perform it.

In Sect. 6, we derive the distribution of delays, i.e., for each $s_j, c_i \in \mathcal{G}$, the distance in the sequence between the agent and the matched good. We find that these distributions are a mixture of convolutions of independent geometric random variables. We obtain explicit expressions for the generating functions and for the mean and variance of these delays. We list an enhancement to the algorithm of Sect. 5 to calculate these means and variances. These can be used to obtain waiting times in the corresponding queueing models.

In Sect. 7, we show that in heavy traffic the directed FCFS bipartite matching model converges to the symmetric matching model of [1,3,8]. We also obtain the limiting matching rates in light traffic.

In Sect. 8, we illustrate our results for a simple example. We demonstrate the advantage of pooling service by bipartite matching over discriminating service.

Some of the proofs are delayed to the Appendix.
2 The model and previous results

We consider a single infinite sequence of agents and goods, which is generated as follows: Each successive item in the list is an agent with probability \( \frac{\lambda}{\lambda + \mu} \), or a good with probability \( \frac{\mu}{\lambda + \mu} \). If it is an agent, its type is \( c_i \in C = \{c_1, \ldots, c_I\} \) with probability \( \alpha_{c_i} \), and if it is a good, its type is \( s_j \in S = \{s_1, \ldots, s_J\} \) with probability \( \beta_{s_j} \), so that in summary, the elements of the sequence are i.i.d. \( (z^n)_{n=-\infty}^{\infty} \), and

\[
\begin{align*}
P(z^n = c_i) &= \frac{\lambda}{\lambda + \mu} \alpha_{c_i} = \frac{\lambda_{c_i}}{\lambda + \mu}, \quad c_i \in C, \quad n \in \mathbb{Z}, \\
P(z^n = s_j) &= \frac{\mu}{\lambda + \mu} \beta_{s_j} = \frac{\mu_{s_j}}{\lambda + \mu}, \quad s_j \in S, \quad n \in \mathbb{Z},
\end{align*}
\]

where \( \lambda_{c_i}, \mu_{s_j} \) are as defined by (1). The compatibility graph \( \mathcal{G} \) is a bipartite graph between \( C, S \), with \( (s_j, c_i) \) if agent \( c_i \) can use good \( s_j \). We define the matching as follows:

**Definition 1** (Directed FCFS matching) We say that agents and goods are matched according to directed FCFS matching if the following holds:

(i) \( n > m \),

(ii) \( (s_j, c_i) \in \mathcal{G} \),

(iii) For all \( k < m \) such that \( z^k = c_i' \) and \( (s_j, c_i') \in \mathcal{G} \), there exists \( z^\ell \) with \( \ell < n \), such that \( z^\ell \) is matched with \( z^k \).

If \( z^n = s_j \) has no match that satisfies these conditions, then \( z^n \) remains unmatched.

Figure 1 illustrates this matching: There are four types of agents and three types of goods, with the illustrated bipartite compatibility graph, and a sequence of randomly occurring agents and goods, \( z_1, \ldots, z_{15} \). In this sequence, \( z_1 = c_1 \) is matched with \( z_4 = s_3 \), \( z_2 = s_2 \) has no match and is lost, etc. At the end of the sequence, \( z_{10} = c_4, z_{12} = c_1, z_{13} = c_1, z_{15} = c_3 \) are as yet unmatched and waiting for goods that will appear later in the sequence.

This matching model was introduced in [2], and the following definitions and three theorems are quoted from that paper for completeness.

**Notation** We let \( \mathcal{S}(c_i) \) denote the subset of good types compatible with \( c_i \), and \( \mathcal{C}(s_j) \) denote the subset of agent types compatible with \( s_j \). For \( C \subset \mathcal{C}, S \subset \mathcal{S} \), we let \( \mathcal{C}(C) = \bigcup_{c_i \in C} \mathcal{S}(c_i), \mathcal{S}(S) = \bigcup_{s_j \in S} \mathcal{C}(s_j) \), and denote by \( \overline{\mathcal{S}}(S) = \overline{\mathcal{C}(S)} \) (where \( \overline{A} \) denotes complement of the set \( A \)) those agent types that are compatible only with...
good types in $S$. We also let $\alpha_C = \sum_{c_i \in C} \alpha_{c_i}$, $\beta_S = \sum_{s_j \in S} \beta_{s_j}$ and $\lambda_C = \sum_{c_i \in C} \lambda_{c_i}$, $\mu_S = \sum_{s_j \in S} \mu_{s_j}$.

**Definition 2** (Stability condition) We say that the stability condition holds if for all $C \subseteq C$ there are enough goods to supply the demand, i.e., the following holds:

$$\lambda_C < \mu_{S(C)}.$$  \hfill (2)

The following Markov chain was defined in [2]: Assume all the possible directed FCFS matches of goods up to $z_N$ have been made. Define $X(N)$ as the ordered sequence $c_1, \ldots, c_L$, such that $c_\ell = z_{n_\ell}$, $n_1 < \cdots < n_L \leq N$, are the only unmatched agents remaining, ordered as in the sequence. In the example in Fig. 1, the state is $X(15) = (c_4, c_1, c_1, c_3)$.

**Notation** Throughout the paper, we use symbols with superscripts to denote random variables which are elements in a sequence, and we use subscripts for the actual values, i.e., $z^n$ is the $n$th item in the sequence, which can have as value a type of agent $c_i$ or a type of good $s_j$. Similarly, we may have in the sequence of unmatched agents $c^m$ denoting the random item in position $m$ which may have the value $c_i$.

The states of this Markov chain are described by finite sequences of agent types. This is a countable state Markov chain (the set of finite sequences of agent types is a countable set, even if the set of types is countably infinite). Its transition probabilities are

$$P(X(N+1) = c^1, \ldots, c^L | X(N) = c^1, \ldots, c^L) = \frac{\mu_{S(\{c^1, \ldots, c^L\})}}{\lambda + \mu},$$

$$P(X(N+1) = c^1, \ldots, c^{i-1}, c^{i+1}, \ldots, c^L | X(N) = c^1, \ldots, c^L) = \frac{\mu_{S(\{c^1, \ldots, c^{i-1}\}) \cap S(c^i)}}{\lambda + \mu}.$$  \hfill (3)

It turns out that this Markov chain is the jump chain of the redundancy service model studied by [12], where the following is proved, using partial balance.

**Theorem 1** (Product Form Theorem) The Markov chain $X(N)$ is ergodic if and only if the stability condition (2) holds. Its stationary distribution is given, up to a normalizing constant, by

$$\pi_X(c^1, \ldots, c^L) \propto \prod_{\ell=1}^L \frac{\lambda_{c_\ell}}{\mu_{S(\{c^1, \ldots, c^\ell\})}}.$$  \hfill (4)

It is clear that for any sequence $z^1, z^2, \ldots$, the directed FCFS matching is unique. The corresponding Markov chain is $X(N)$, $N = 0, 1, \ldots$ starting from no unmatched agents, i.e., $X(0) = \emptyset$. However, if we prepend to the sequence an earlier item $z^0, z^1, z^2, \ldots$, and start from $X(-1) = \emptyset$, we usually will get a different matching, i.e., the process starting from $X(-1) = \emptyset$ will be different from that starting with
Theorem 2 (Uniqueness Theorem) Let $\ldots, z^{-1}, z^0, z^1, \ldots$ be a sequence of agent and good types chosen i.i.d. as above. If the stability condition (2) holds, then almost surely there exists a directed FCFS matching of goods to cover all the agents, and this matching is unique.

The proof is given in [2]. The main steps are to show forward coupling, then to show backward coupling, and then to conclude that a Loynes construction will yield the unique matching almost surely. The fact that under the condition of stability, almost surely every realized sequence will have a unique FCFS matching from $-\infty$ to $\infty$ is remarkable. It is also the basis for the next theorem.

For the next result, which is also analogous to a similar result in [1,19] and is proven in [2], we define the following exchange transformation:

Definition 3 (Exchange Transformation) For the directed FCFS matching, after all possible matches were made on $\ldots, z^{-1}, z^0, z^1, \ldots$, we define the exchanged sequence $\ldots, \tilde{z}^{-1}, \tilde{z}^0, \tilde{z}^1, \ldots$ as follows: If $z^m = c_i$ was matched to $z^n = s_j$, where $m < n$, then in the exchanged sequence we will have $\tilde{z}^m = s_j$, $\tilde{z}^n = c_i$. If $z^n = s_j$ was unmatched, then $\tilde{z}^n = z^n$.

Notation Throughout the paper, we will denote with $\sim$ an item in the sequence that was matched and exchanged, i.e., $\tilde{z}^n$ is the item in position $n$ in the exchanged sequence. For the values of $\tilde{z}^n$, we will denote the types of exchanged items by $\tilde{c}_i$ if it is an agent of type $c_i$ and by $\tilde{s}_j$ if it is a good of type $s_j$.

Figure 2 illustrates the exchange transformation for a block of items in the sequence which is perfectly matched (all agents are matched).

As a consequence of the stability and the unique matching over $\mathbb{Z}$, we obtain:

Theorem 3 (Reversibility Theorem) Assume the stability condition (2) holds. The sequence $\ldots, \tilde{z}^{-1}, \tilde{z}^0, \tilde{z}^1, \ldots$ obtained from the sequence $\ldots, z^{-1}, z^0, z^1, \ldots$ by the exchange transformation is an i.i.d. sequence. The unique directed matches for the new sequence performed in reversed time result in exactly the reversed matches of the original sequence, almost surely.
3 Detailed Markov chains with multi-Bernoulli stationary distributions

We now define two additional Markov chains associated with the directed FCFS infinite bipartite matching model. These processes describe the evolution of matching and exchanging items in the sequence \((x^n)_{n=1}^{\infty}\). We define detailed goods matching process \(U(N)\), to describe matching and exchanging all goods up to (and including) \(N\) with agents that appear earlier in the sequence (with some of the goods left unmatched and lost, and some of the agents still unmatched). We define the detailed agents matching process \(W(N)\), to describe matching and exchanging all agents up to (and including) \(N\) with goods that appear later in the sequence. For \(U(N)\), we perform directed FCFS matches and exchanges of all compatible pairs \(z^m = c_i, z^n = s_j, m < n \leq N\), while for \(W(N)\) we perform directed FCFS matches and exchanges of all compatible pairs \(z^m = c_i, z^n = s_j, m < n\), for all agents \(m \leq N\), including some \(z^n = s_j\) for which \(n > N\).

Specifically, \(U(N)\) is defined as follows: If \(X(N) = \emptyset\), i.e., all agents appearing up to \(N\) have been matched to goods appearing up to \(N\), then \(U(N) = \emptyset\). Otherwise, let \(N\) be the location in the sequence of the first unmatched agent, then \(U(N) = u = (u_1, \ldots, u_L)\), where \(L = N - N + 1\), and \(U(N)\) records the elements in the sequence from \(N\) to \(N\), after matching and exchanging of goods to earlier agents, and also exchanging the remaining unmatched goods up to \(N\) with themselves. The elements of \(u\) are then of types \(c_i\) or \(\tilde{c}_i\) or \(\tilde{s}_j\) and are constructed as follows:

- \(u_\ell = c_i\) if \(z^{N+\ell-1} = c_i\) is unmatched,
- \(u_\ell = \tilde{s}_j\) if \(z^{N+\ell-1} = s_j\) remains unmatched (it is exchanged with itself),
- \(u_\ell = \tilde{c}_j\) if \(z^{N+\ell-1}\) has originally been a good and has been matched and exchanged with an agent of type \(c_j\) that appeared earlier in the sequence,
- \(u_\ell = \tilde{s}_j\) if \(z^{N+\ell-1}\) has originally been an agent and has been matched and exchanged with a good of type \(s_j\) that has appeared later in the sequence, in a position \(\leq N\).
- Note that \(u_1\) is always an agent \(c_i\), the first of the unmatched agents.

Similarly, \(W(N)\) is defined as follows: If \(X(N) = \emptyset\), i.e., all agents appearing up to \(N\) have been matched to goods appearing up to \(N\), then \(W(N) = \emptyset\). Otherwise, let \(\overline{N}\) be the location in the sequence of the last matched and exchanged good (we have \(\overline{N} > N\) since \(X(N) \neq \emptyset\), then \(W(N) = w = (w_1, \ldots, w_K)\) where \(K = \overline{N} - N\) and \(W(N)\) records the elements in the sequence from \(N + 1\) to \(\overline{N}\), after matching and exchanging of agents from locations up to \(N\) with later goods. The elements of \(w\) are of types \(c_i\) or \(\tilde{c}_i\) or \(s_j\) and are constructed as follows:

- \(w_\ell = c_i\) if \(z^{N+\ell} = c_i\) is as yet unmatched, because only agents up to \(N\) have been matched.
- \(w_\ell = s_j\) if \(z^{N+\ell} = s_j\) has not been matched to an agent appearing up to \(N\),
- \(w_\ell = \tilde{c}_j\) if \(z^{N+\ell}\) has originally been a good and has been matched and exchanged with an agent of type \(c_j\) that appeared in the sequence in position \(\leq N\).
- Note that the last element \(w_K\) is always a matched and exchanged agent \(\tilde{c}_i\). It is in the position of the latest good that was used in matching all agents up to \(N\).
Figure 3 illustrates the detailed states. It also serves as illustration of Proposition 1. We discuss Fig. 3 in more detail after the proof of Proposition 1.

Both $U(N)_{N \in \mathbb{Z}}$ and $W(N)_{N \in \mathbb{Z}}$ are discrete time and space Markov chains. They contain more information than the chain $X(N)$. They are of interest because, as we will show, their stationary distribution is of an extremely simple nature: Each agent appearing in $U(N)$ or in $W(N)$ is of type $c_i$ or $\tilde{c}_i$ with probability $\alpha_{c_i}$, and each good is of type $s_j$, $\tilde{s}_j$ with probability $\beta_{s_j}$, and this is independent of all other elements appearing in $U(N)$, $W(N)$. We prove this in Theorem 4. In a way this is not entirely surprising, since the sequence $z^n$ is i.i.d. and so is also the sequence $\tilde{z}^n$, so we can expect this to show also during the evolution of the matching, though this does not explain the independence of the $z$ and $\tilde{z}$ terms. The simplicity is somewhat misleading, since all possible states of $U(N)$, $W(N)$ need to satisfy some structural restrictions. We characterize the possible states in Theorem 5. These results are again similar to those in [1,19], using similar proof methods, but there are enough differences to require a careful drafting of the proofs, which we put in the Appendix.

Our main motivation for studying the detailed processes is that $U(N)$ is essential in deriving the distribution of the link lengths, in Sect. 6.

To study $U(N)$, $W(N)$, we use the reversibility theorem. By the reversibility Theorem 3 under the condition (2), almost surely, if we take the sequence $(z^n)_{-\infty < n < \infty}$ and do all the directed FCFS matches and exchanges we obtain the sequence $(\tilde{z}^n)_{-\infty < n < \infty}$ which is again i.i.d., and if we do directed FCFS matches and exchanges in reversed time on the sequence $(\tilde{z}^n)_{-\infty < n < \infty}$ we will retrieve the original sequence $(z^n)_{-\infty < n < \infty}$. In fact, the pairs that are matched are the same in both sequences.

We consider the sequence $(\tilde{z}^n)_{-\infty < n < \infty}$ and define goods and agents detailed matching processes for it in reversed time: $\tilde{U}(N+1)$, $\tilde{W}(N+1)$ are obtained when we do matching for the sequence $(\tilde{z}^n)_{-\infty < n < \infty}$, where for $\tilde{U}(N+1)$ we perform directed FCFS matches and exchanges in reversed time, of pairs $\tilde{z}^n = \tilde{s}_j$, $\tilde{z}^m = \tilde{c}_i$, $m > n$, for all goods in positions $n > N$, while for $\tilde{W}(N+1)$ we perform directed FCFS matches.
and exchanges in reversed time of pairs $\tilde{z}^n = \tilde{s}_j$, $\tilde{z}^m = \tilde{c}_i$, $m > n$, for all agents in positions $m > N$.

**Notation** For any vector $u$ consisting of items $c, \tilde{c}, s, \tilde{s}$, we denote by $\tilde{u}$ the exchanged vector in which we exchange between $c, \tilde{c}$ and between $s, \tilde{s}$. We denote by $\overset{←}{\tilde{u}}$ the reversed vector which is $u$ written in reversed order. We will also use the notation $\overset{←}{\tilde{u}}$ for exchanging and reversing.

**Proposition 1** For almost all $(z^n)_{-\infty < n < \infty}$ and corresponding $(\tilde{z}^n)_{-\infty < n < \infty}$, and for $-\infty < N < \infty$, we have

$$
\begin{align*}
U(N) &= u = (u^1, \ldots, u^L) \\
U(N + 1) &= u' = (u'^1, \ldots, u'^L)
\end{align*}
$$

$\text{then}$

$$
\begin{align*}
\tilde{W}(N + 2) &= \overset{←}{\tilde{u}'} = (u'^L, \ldots, u^1), \\
\tilde{W}(N + 1) &= \overset{←}{\tilde{u}} = (u^L, \ldots, u^1).
\end{align*}
$$

and similarly,

$$
\begin{align*}
W(N) &= w = (w^1, \ldots, w^K) \\
W(N + 1) &= w' = (w'^1, \ldots, w'^K)
\end{align*}
$$

$\text{then}$

$$
\begin{align*}
\overset{←}{\tilde{U}}(N + 2) &= \overset{←}{\overset{←}{\tilde{w}'} = (w'^K', \ldots, w^1'), \\
\overset{←}{\tilde{U}}(N + 1) &= \overset{←}{\overset{←}{\tilde{w} = (w^K, \ldots, w^1)}.
\end{align*}
$$

**Proof** By definition, $U(N)$ will have all goods in the sequence $z^n$ up to $N$ matched and exchanged and will list the resulting items from $N$ to $\infty$, while all items in positions $> N$ are still unmatched and not exchanged. On the other hand, $\tilde{W}(N + 1)$ will have all the agents in the exchanged sequence $\tilde{z}^n$ from $N + 1$ to $\infty$ matched and exchanged back to their values in $\tilde{z}^n$, and list the items from $N$ to $\infty$ in reversed order. In particular, the first item in $U(N)$ is the earliest unmatched agent $z^N = c_i$, and it could only be matched later with $z^m = s_j$ which is in position $m > N$, while in $\tilde{W}(N + 1)$ $\tilde{z}^m = \tilde{c}_i$ is matched and exchanged back with $\tilde{z}^N = \tilde{s}_j$. Similarly, any unmatched agent $z^n = c_i, n < N$ in $U(N)$ has a match $z^m$ with $m > N$, hence $\tilde{z}^n = \tilde{s}_j$, and $\tilde{z}^m = \tilde{c}_i$, and in $\tilde{W}(N + 1)$ they will be matched and exchanged back. All other elements of $U(N)$ are exchanged $\tilde{c}$ or $\tilde{z}$ which are in their original positions in $\tilde{W}(N + 1)$. Hence, $\tilde{W}(N + 1)$ lists the same items as $U(N)$ in reversed order. By exactly the same argument, $\tilde{W}(N + 2)$ lists the same items as $U(N + 1)$ in reversed order. The statement (6) can be proved similarly, and in fact it follows directly from the reversibility Theorem 3. □

Figure 3 illustrates this proposition. It has a section of the sequence $z^n$ on the first row, the second row shows the state $U(N)$, and the third row the state $W(N)$. The bottom row has the matched and exchange section of the sequence, $\tilde{z}^n$. The reader can check that when performing matching and exchanging in of $\tilde{z}^n$ in reversed time we obtain $\overset{←}{\tilde{U}}(N + 1)$ and $\overset{←}{\tilde{W}}(N + 1)$ in the third and second row.

**Proposition 2** For the stationary versions of the detailed Markov process $U(N)$, the reversed transition probabilities are given by

$$
P(U(N + 1) = u') P(U(N) = u | U(N + 1) = u')
\begin{align*}
&= P(W(N) = \overset{←}{\tilde{u}}) P(W(N + 1) = \overset{←}{\tilde{u}} | W(N) = \overset{←}{\tilde{u}}).
\end{align*}
$$

© Springer
Proof

\[ P \left( U(N+1) = u' \right) P \left( U(N) = u | U(N+1) = u' \right) = P \left( U(N) = u, U(N+1) = u' \right) = P \left( \tilde{W}(N+1) = \tilde{u}, \tilde{W}(N+2) = \tilde{u}' \right) = P \left( \tilde{W}(N+2) = \tilde{u}' \right) P \left( \tilde{W}(N+1) = \tilde{u} | \tilde{W}(N+2) = \tilde{u}' \right) = P \left( W(N) = \tilde{u}' \right) P \left( W(N+1) = \tilde{u} | W(N) = \tilde{u}' \right). \] (8)

The first and third equalities are by definition of conditional probability. The second equality holds for the stationary processes by Proposition 1 and Eq. (5). The last equality follows from the reversibility Theorem 3, since the Markov chain \( W(N) \) driven by the sequence \( z^n \) has the same stationary distribution and the same transition probabilities as the Markov chain \( \tilde{W}(N) \) driven by the sequence \( \tilde{z}^n \) in reversed time. \( \square \)

Theorem 4 (Multi-Bernoulli Form) If the stability conditions (2) holds, then the stationary distribution of the detailed Markov chains is given by

\[ \pi_U(u_1, \ldots, u_L) = \pi_W(u_1, \ldots, u_L) = B \left( \frac{\lambda}{\lambda + \mu} \right)^{\text{agents}} \left( \frac{\mu}{\lambda + \mu} \right)^{\text{goods}} \prod_{c_i \in \mathcal{C}} \alpha_{c_i}^{s_j} \prod_{s_j \in \mathcal{S}} \beta_{s_j}^{z^{c_i} + z^{s_j}}, \] (9)

where \( B \) is the normalizing constant given by (14), and we are counting the number of agents and of goods overall, and of each type separately, in \( u_1, \ldots, u_L \).

Proof (Outline of the proof) That \( \pi_U(u_1, \ldots, u_L) = \pi_W(u_1, \ldots, u_L) \) follows from Proposition 1. By definition, \( \pi_U(\emptyset) = \pi_W(\emptyset) = \pi_X(\emptyset) \). We will show that \( \pi_X(\emptyset) = B \) and show that \( B \) is given by (14) in Sect. 4. It remains to prove (9) for \( \pi_U \). For a stationary Markov chain, this is equivalent to verifying that

\[ \pi_U(u) P \left( U(N+1) = u' | U(N) = u \right) = \pi_U(u') P \left( U(N) = u | U(N+1) = u' \right). \] (10)

We have found in Proposition 2 that the reversed transition probabilities for \( U \) are given by transition probabilities of \( W \). Also, we note that in the form of equation (9), \( \pi_U(u) = \pi_U(\tilde{u}) = \pi_U(u) = \pi_U(\tilde{u}) \), with similar equalities for \( W \). Therefore, to prove (10) we will verify that the expressions in (9) satisfy

\[ \frac{\pi_U(u')}{\pi_U(u)} = \frac{P \left( U(N+1) = u' | U(N) = u \right)}{P \left( W(N+1) = \tilde{u} | W(N) = \tilde{u}' \right)}. \] (11)

and to do so we need to calculate transition probabilities for \( U \) and for \( W \) from \( N \) to \( N+1 \). These transition probabilities hinge on the identity of \( x = z^{N+1} \) and the corresponding \( \tilde{x} = z^{N+1} \). We start from \( U(N) = u = (u_1, \ldots, u_L) \). There are four cases to verify, which we do in Appendix A.1 \( \square \)
The stationary distribution of the Markov chains $U(N)$ and $W(N)$ looks very simple: It seems to be the distribution of a sequence of independent trials with outcomes in $\mathcal{C} \cup \mathcal{S}$ with fixed probabilities; we call that a multi-Bernoulli sequence. However, this simplicity is very deceptive: To say anything more about the distribution, we need to know which are the possible states that may occur and we need to find a way to count them. The next theorem gives a characterization of the valid states of $U$ and $W$.

**Theorem 5 (Detailed State Characterization)** A state $U$ in $\mathcal{C}$ is simple: It seems to be the distribution of a sequence of independent trials with outcomes $\mathcal{C}_k$ and $\mathcal{S}_k$, and for any $k < l$ if $u_k = c_i$ and $u_{\ell} = s_j$, then $(s_j, c_i) \notin \mathcal{G}$. Similarly, a state $W$ in $\mathcal{W}$ is simple: It seems to be the distribution of a sequence of independent trials with outcomes $\mathcal{C}_k$ and $\mathcal{S}_k$, and for any $k > l$ if $w_k = c_i$ and $w_{\ell} = s_j$, then $(s_j, c_i) \notin \mathcal{G}$.

The proof is given in Appendix A.2.

### 4 A concise Markov chain and calculation of the normalizing constant

The Markov chain $U(N)$ and its reversal partner $W(N)$ contain very detailed information about the evolution of the directed FCFS matching. Much of this information is irrelevant, in particular the $\tilde{c}$ items and $\tilde{s}$ items in $U(N)$ refer to the past of the matching process and have no influence on future matching. The only relevant part for the future is the ordered list of unmatched agents, which is given by the Markov chain $X(N)$, introduced in [2], where its stationary distribution (Theorem 1, Eq. (4)) is obtained using a partial balance argument from [12]. We have briefly summarized results on $X(N)$ in Sect. 2.

The transitions of $X(N)$ are very simple: Given $X(N) = (c_1, \ldots, c_L)$ and $z^{N+1}$, the state $X(N + 1)$ is determined uniquely. On the other hand, the states are given by finite words of arbitrary length, which is awkward for calculations. We next define a more concise Markov chain with simpler, finite-dimensional states.

We define the Markov chain $Y(N)_{N \in \mathbb{Z}}$ whose states are of the form $Y(N) = y = (C_1, n_1, \ldots, C_k, n_k)$, where $k \leq I$, $(C_1, \ldots, C_k)$ is a permutation of a subset $C \subseteq \mathcal{C}$, and $n_j$ are nonnegative integers. In particular, for $k = 0$ the state is denoted as $Y(N) = \emptyset$ or $Y(N) = 0$. The states of $Y(N)_{N \in \mathbb{Z}}$ are derived from the states of $X(N)_{N \in \mathbb{Z}}$ as follows: If $X(N) = \emptyset$, then also $Y(N) = \emptyset$. If $X(N) = x = (c_1, \ldots, c_L)$, then $C_1, \ldots, C_k$ are the types of agents appearing in $x$, in the order in which they appear, and $n_j$ is the number of terms in $x$ between the first appearance of $C_j$ and the first appearance of $C_{j+1}$, for $j = 1, \ldots, k - 1$, and $n_k$ is the number of terms in $x$ from the first appearance of $C_k$ up to the last term $c_L$. The number of unmatched agents for $Y(N) = (C_1, n_1, \ldots, C_k, n_k)$ is $k + \sum_{j=1}^{k} n_j$. Note that now, given $Y(N)$ and $z^{N+1}$ we can determine the next match, but the next state is not deterministic, several possible states $Y(N + 1)$ may result. This Markov chain is similar to a Markov chain used in [4, 23].
Theorem 6 If the stability condition (2) holds, then the stationary distribution of \( Y(N) \) is given by

\[
\pi_Y(C_1, n_1, \ldots, C_k, n_k) = B \prod_{\ell=1}^{k} \frac{\lambda_{C_\ell}}{\mu \mathcal{Y}((C_1, \ldots, C_\ell))} \left( \frac{\lambda_{(C_1, \ldots, C_\ell)}}{\mu \mathcal{Y}((C_1, \ldots, C_\ell))} \right)^{n_\ell},
\]

where \( B \) is a normalizing constant, which is also equal to the probability of a perfect match, i.e., \( \pi_Y(\emptyset) = B \). The value of \( B \) is given by (14).

Proof We note that \( Y(N) = (C_1, n_1, \ldots, C_k, n_k) \) includes all states \( X(N) = (c^1, \ldots, c^L) \) where \( L = k + \sum_{\ell=1}^{k} n_\ell \), where \( C_\ell, \ell = 1, \ldots, k \), appear at least once, in that order, and where between the appearance of \( C_\ell \) and \( C_{\ell+1} \) there are any combinations of \( n_\ell \) agents of types \( C_1, \ldots, C_\ell \). Hence, by (4)

\[
\pi_Y(C_1, n_1, \ldots, C_k, n_k) = B \prod_{\ell=1}^{k} \frac{\lambda_{C_\ell}}{\mu \mathcal{Y}((C_1, \ldots, C_\ell))} \left( \frac{\lambda_{(C_1, \ldots, C_\ell)}}{\mu \mathcal{Y}((C_1, \ldots, C_\ell))} \right)^{n_\ell}
\]

We also get the probability of a perfect match, \( \emptyset \), from the balance equation:

\[
\pi_X(0) = \pi_X(0) \frac{1}{\lambda + \mu} + \sum_{i=1}^{k} \pi_X(c_i) \frac{\mu \mathcal{Y}(c_i)}{\lambda + \mu},
\]

\[
\pi_X(0) \lambda = \sum_{i=1}^{k} \pi_X(c_i) \mu \mathcal{Y}(c_i) = B \sum_{i=1}^{k} \frac{\lambda_{c_i}}{\mu \mathcal{Y}(c_i)} = B \lambda,
\]

so we get \( \pi_X(\emptyset) = B \), and from this also \( \pi_Y(\emptyset) = B \).

We also obtain the stationary probabilities of observing a state in which only agents of types \((C_1, \ldots, C_k)\) are unmatched, and they appear for the first time in that order, aggregated over all the values of \( n_1, \ldots, n_k \), which we denote as \((C_1, \cdot, \ldots, C_k, \cdot)\):

\[
\pi_Y(C_1, \cdot, \ldots, C_k, \cdot) = B \prod_{\ell=1}^{k} \frac{\lambda_{C_\ell}}{\mu \mathcal{Y}((C_1, \ldots, C_\ell))} \left[ \sum_{n_\ell=0}^{\infty} \left( \frac{\lambda_{(C_1, \ldots, C_\ell)}}{\mu \mathcal{Y}((C_1, \ldots, C_\ell))} \right)^{n_\ell} \right]
\]

\[
= B \prod_{\ell=1}^{k} \frac{\lambda_{C_\ell}}{\mu \mathcal{Y}((C_1, \ldots, C_\ell))} - \lambda_{(C_1, \ldots, C_\ell)}
\]

Together with \( \pi_Y(\emptyset) = B \), this gives us the stationary distribution of the unmatched types in \( X(N) \) or in \( U(N) \), in the order in which they appear for the first time. These

\[\square\]
probabilities play a central role in the calculation of matching rates (Sect. 5) and in deriving the distribution of delays (Sect. 6).

We now calculate the normalizing constant, \( B = \pi_Y(0) \). We need to sum over all subsets of \( \mathcal{C} \), and for each subset \( C \subseteq \mathcal{C} \) we need to sum over all the permutations of \( C \). (Altogether, the summation has \( f(I) = \sum_{k=0}^{I} \frac{I!}{(I-k)!} \) terms, for example, \( f(0), f(1), \ldots, f(10) = 1, 2, 5, 16, 65, 326, 1597, 13700, 109601, 986410, 9864101; \) we note that \( n! < f(n) < (n+1)! \).)

\[
B = \left( \sum_{k=0}^{I} \sum_{C \subseteq \mathcal{C}, |C| = k} \left( \sum_{(C_1, \ldots, C_k) \in \mathcal{P}(C)} \prod_{\ell=1}^{k} \left( \frac{\lambda_{C_\ell}}{\mu_{\mathcal{C}}(\{C_1, \ldots, C_\ell\})} - \lambda_{\{C_1, \ldots, C_\ell\}} \right) \right) \right)^{-1}
\]

(14)

where the term for \( k = 0 \) equals 1. We strongly suspect that calculation of \( B \) is \#P-complete. (This is similar to the conjecture that the normalizing constant in symmetric FCFS infinite bipartite matching is \#P-complete, made in [3].)

We note that the stationary distribution (4) of \( X(N) \) and (12) of \( Y(N) \) can be obtained directly from the distribution of \( U(N) \). In Appendix A.3, we show how to calculate \( \pi_Y(C_1, \ldots, C_\ell, \cdot) \) directly from Theorems 4 and 5.

5 Calculation of the matching rates

**Definition 4** (Matching Rates) In a stable FCFS directed bipartite matching, we define matching rates as follows: \( r_{s_j,c_i} \) is the fraction of matches which are of goods type \( s_j \) and agent type \( c_i \), and \( r_{s_j,\emptyset} \) is the fraction of goods which are of type \( s_j \) and are not matched.

To make things clear, \( \sum_{s_j \in \mathcal{S}} \left( r_{s_j,\emptyset} + \sum_{c_i \in \mathcal{C}} r_{s_j,c_i} \right) = 1 \). We also have:

- For goods of type \( s_j \), the fraction that will remain unmatched and the fraction that will be matched with agents of type \( c_i \) are given by

  \[
  \eta_{s_j}(\emptyset) = \frac{r_{s_j,\emptyset}}{r_{s_j,\emptyset} + \sum_{c_i \in \mathcal{C}} r_{s_j,c_i}}, \quad \eta_{s_j}(c_i) = \frac{r_{s_j,c_i}}{r_{s_j,\emptyset} + \sum_{c_i \in \mathcal{C}} r_{s_j,c_i}}.
  \]

- For agents of type \( c_i \), the fraction that will be matched with goods of type \( s_j \) is given by

  \[
  \theta_{c_i}(s_j) = \frac{r_{s_j,c_i}}{\sum_{s_j \in \mathcal{S}} r_{s_j,c_i}}.
  \]

The total fraction of unmatched goods is, as can be obtained directly by considering total rates of all agents and goods, is

\[
r_{\emptyset} = \sum_{s_j \in \mathcal{S}} \frac{\mu_{s_j}}{\mu} \eta_{s_j,\emptyset} = \frac{\mu - \lambda}{\mu}.
\]
For the next theorem, we need the discrete time version of PASTA, Bernoulli events see time average—BASTA, which we formulate here:

**Proposition 3 (BASTA)** Let \( X(N) \) be a discrete time stochastic process, let \( B(N) \) be i.i.d. \( P(B(N) = 1) = p, \) \( P(B(N) = 0) = 1 - p, \) \( N = 1, 2, \ldots, \) and assume that \( B(N) \) is independent of \( X(M), \) \( M \leq N \) (but, as is often the point of the theorem, \( B(N) \) is not independent of \( X(M), \) \( M > N \)). Let \( U(N) \) be an indicator of the event \( X(N) \in E \) for some event \( E. \) Let

\[
\bar{U}(N) = \frac{1}{N} \sum_{m=1}^{N} U(m), \quad \bar{A}(N) = \frac{\sum_{m=1}^{N} U(m)B(m)}{\sum_{m=1}^{N} B(m)}.
\]

Then, if one of the limits \( \lim_{N \to \infty} \bar{U}(N) \) and \( \lim_{N \to \infty} \bar{A}(N) \) exists, so does the other, and they are equal.

In particular, if \( X(N) \) is a discrete time ergodic stationary Markov chain, and \( B(N) \) as above is independent of \( X(N) \) but may determine \( X(N + 1), \) then the distribution of \( X(N) \) at times that \( B(N) = 1 \) is the same as the distribution of \( X(N) \) at any time. This proposition hinges on the memoryless property of the discrete Geometric distribution, similar to the memoryless property of the exponential distribution which is the basis of PASTA. The proof of this proposition follows similar steps to the proof of PASTA by Wolff [24].

**Theorem 7** In a stable system, the matching rates for \( (s_j, c_i) \in \mathcal{G} \) are given by

\[
\begin{aligned}
r_{s_j, c_i} &= \frac{\mu_{s_j}}{\mu} \sum_{k=1}^{I} \sum_{C \subseteq \mathcal{G}, |C| = k} \sum_{(C_1, \ldots, C_k) \in \mathcal{P}(C)} \pi_Y(C_1, \cdot, \ldots, C_k, \cdot) \\
&\times \left( 1(C_1 = c_i) \lor \bigvee \left\{ (C_1 \cap \mathcal{G}(s_j) = \emptyset \land C_2 = c_i) \right. \right. \\
&\left. \lor \cdots \lor 1(C_1, \ldots, C_{k-1} \cap \mathcal{G}(s_j) = \emptyset \land C_k = c_i) \right) \right) \quad (15)
\end{aligned}
\]

The rates of unmatched goods are given by

\[
\begin{aligned}
r_{s_j, \emptyset} &= \frac{\mu_{s_j}}{\mu} \sum_{k=1}^{I} \sum_{C \subseteq \mathcal{G}, |C| = k} \sum_{(C_1, \ldots, C_k) \in \mathcal{P}(C)} \pi_Y(C_1, \cdot, \ldots, C_k, \cdot) \\
&\times \left( 1((C_1, \ldots, C_k) \cap \mathcal{G}(s_j) = \emptyset) \right) \quad (16)
\end{aligned}
\]

**Proof** We consider the stationary version of \( Y(N), \) By the discrete time version of PASTA as in Proposition 3, at the moment that \( z^{N+1} \in \mathcal{I} \) the state of \( Y(N) \) at time \( N \) is the stationary state of the process. For each value of \( z^{N+1} \in \mathcal{I}, \) we wish to calculate the probability that it matches with \( c_i \) or that it remains unmatched. These probabilities will be the matching rates. In particular, \( P(z^{N+1} = s_j | z^{N+1} \in \mathcal{I}) = \beta_{s_j} \), and the distribution of \( Y(N) \) at times when \( Z^{N+1} = s_j \) is also the stationary distribution of
Y(N). So \( r_{s_j,c_i} \) is calculated as the probability that, conditional on \( z^{N+1} = s_j \), an \((s_j, c_i)\) match occurs, times \( \beta_{s_j} = \frac{\mu_{s_j}}{\mu} \). If \( z^{N+1} = s_j \), and \( Y(N) = (C_1, \ldots, C_k) \), then the match for \( z^{N+1} \) is deterministic: It either matches with the first compatible \( C_\ell \), or it remains unmatched. In particular, the match will be with \( c_i \) if and only if for some \( 1 \leq \ell \leq k \), \( C_\ell = c_i \), and all of \( C_1, \ldots, C_{\ell-1} \) are not compatible with \( s_j \). If all of \( C_1, \ldots, C_k \) are incompatible with \( s_j \), then \( z^{N+1} \) will remain unmatched.

Hence, to obtain \( r_{s_j,c_i} \) we need to add up all the stationary probabilities of states \((C_1, \ldots, C_k, \cdot)\) for which \( s_j \) will match with \( c_i \), and multiply this sum by \( \frac{\mu_{s_j}}{\mu} \), which is \( P(z_{N+1} = s_j \mid z_N = \mathcal{S}) \). Similarly, for \( r_{s_j,\emptyset} \) we add the probabilities for \((C_1, \ldots, C_k, \cdot)\) for which \( s_j \) remains unmatched and multiply by \( \frac{\mu_{s_j}}{\mu} \). These are exactly the expressions given in (15) and (16).

As already stated in several previous papers, it is very likely that the calculation of \( B \) is \#P-complete. Calculation of \( B \) directly from (14) requires adding up \( \sum_j j! \binom{n}{j} \) terms, which is still quite feasible for a small number of agent types \(|\mathcal{C}| = I\). We note that the terms added to calculate \( B \) are also the terms needed to add for calculating the matching rates. We formulate the following algorithm for calculation of the matching rates:

Algorithm 1: Calculation of normalizing constant and matching rates

1. Initialize: \( B := 1; r_{s_j,c_i} := 0; \)
2. For \( A \subseteq \mathcal{C}, A \neq \emptyset \):
   - For \((C_1, \ldots, C_{|A|}) \in \text{Permutations}(A)\):
     - \( X = 1; C_{set} := \emptyset; S_{set} := \emptyset; \lambda_{set} := 0; \mu_{set} := 0; \text{Match}(s_j) := \emptyset. \)
     - For \( l = 1, |A| \):
       - \( C_{set} := C_{set} \cup C_\ell; \Delta S_{set} := \mathcal{S}(C_\ell) \setminus S_{set}; S_{set} := S_{set} \cup \mathcal{S}(C_\ell); \lambda = \lambda_{C_\ell}; \lambda_{set} := \lambda_{set} + \lambda; \mu_{set} := \mu_{set} + \mu \Delta S_{set}; \)
       - \( R_\ell := \frac{\mu_{set} - \lambda_{set}}{\lambda_{set}}; X := X \lambda R_\ell. \)
     - \( B := B + X. \)
   - For \( s_j \in |A| \):
     - If \( C_\ell \in \mathcal{Q}(s_j) \land \text{Match}(s_j) = \emptyset \) then:
       - \( \text{Match}(s_j) := C_\ell; r_{s_j,C_\ell} := r_{s_j,C_\ell} + X. \)
     - \( B := 1/B; r_{s_j,c_i} := B \frac{\mu_{s_j}}{\mu} r_{s_j,c_i}; r_{s_j,\emptyset} := \frac{\mu_{s_j}}{\mu} - \sum_{c_i \in \mathcal{C}} r_{s_j,c_i}. \)

We explain the algorithm here: Line 1 is initialization. Lines 2–8 are a loop over all subsets of agent types and permutations of the subsets. Within this loop, for each \( C_1, \ldots, C_k \), there are two internal loops: Lines 3–5 contain a loop over \( C_\ell \), for the calculation of the ratios \( \frac{\lambda_{C_\ell}}{\mu \mathcal{S}(C_1, \ldots, C_\ell) - \lambda_{C_1, \ldots, C_\ell}} \) and their successive products to obtain \( X \propto \pi_Y(C_1, \ldots, C_k, \cdot) \), then line 6 accumulates the calculated values of \( X \); the second loop, lines 7–8, is over all \( C_\ell \) and \( s_j \in \mathcal{S} \); it uses the value \( X \) and accumulates
it to \( r_{s_j,c_i} \) if \( Y(N) = (C_1, \ldots, C_k, \cdot) \) and \( Z^{N+1} = s_j \) results in an \((s_j, c_\ell)\) match. Line 9 completes the calculation of the normalizing constant and matching rates.

6 The distribution and the moments of delays

**Definition 5** (Delays) If \( z^m = c_i \) is matched to \( z^n = s_j \), then we define the \((s_j, c_i)\) delay between them as \( L_{s_j,c_i} = n - m \). If \( z^m = c_i \) is matched to \( z^n \in \mathcal{S} \), we define the delay of \( c_i \) by \( L_{c_i} = n - m \).

In this section, we obtain the distribution of the delays. It turns out that this distribution is a mixture of convolutions of geometric random variables. We will derive expressions for its generating function, and based on the simple form of the distribution we also obtain expressions for its moments. We will also translate the delays to waiting times, if the arrivals of all types form independent Poisson processes.

**Theorem 8** Consider the process \( U(N) \), and assume that the corresponding \( Y(N) = (C_1, \ldots, C_k, \cdot) \). Denote by \( D_\ell(C_1, \ldots, C_k) \) the number of items in \( U(N) \) from the first appearance of \( C_\ell \) to the first appearance of \( C_{\ell+1} \), \( \ell = 1, \ldots, k - 1 \), and from the first appearance of \( C_k \) to the end, i.e., if \( U(N) = u = (u_1, \ldots, u_L) \), and \( u_{m_\ell} = C_\ell, u_{m_\ell+1} = C_{\ell+1} \) are the first appearances of agent types \( C_\ell, C_{\ell+1} \) in \( u \), then \( D_\ell(C_1, \ldots, C_{\ell+1}) = m_{\ell+1} - m_\ell \), \( \ell = 1, \ldots, k - 1 \), and \( D_k(C_1, \ldots, C_k) = L + 1 - m_k \). The distribution of \( D_\ell(C_1, \ldots, C_k) \) is a geometric distribution over \( m = 1, 2, \ldots, \), given by

\[
P(D_\ell(C_1, \ldots, C_k) = m) = \frac{\mu_{\mathcal{S}}((C_1, \ldots, C_{\ell+1})) - \lambda(C_{\ell+1}, \ldots, C_k)}{\lambda + \mu} \left( 1 - \frac{\mu_{\mathcal{S}}((C_1, \ldots, C_{\ell+1})) - \lambda(C_{\ell+1}, \ldots, C_k)}{\lambda + \mu} \right)^{m-1}
\]

Furthermore, \( D_\ell(C_1, \ldots, C_k) \), \( \ell = 1, \ldots, k \), are independent.

**Proof** Note that by the condition (2) for a stable system, \( 0 < \frac{\mu_{\mathcal{S}}((C_1, \ldots, C_{\ell+1})) - \lambda(C_{\ell+1}, \ldots, C_k)}{\lambda + \mu < 1} \). We fix \( C_1, \ldots, C_k \) and, for notational convenience, use \( D_{\ell} \) to denote \( D_\ell(C_1, \ldots, C_k) \).

We calculate the probability of the states of \( U(N) \), with \((C_1, \cdots, C_\ell, \cdots, C_k, \cdots)\), where we sum up over all possible values of each of the \( m_\ell \) - 1 items in the interval between the first appearance of \( C_\ell \) and the first appearance of \( C_{\ell+1} \), i.e., we will calculate \( P(D_\ell = m_1, \ldots, D_k = m_k \cap Y = C_1, \cdots, C_k, \cdot) \).

To do that, we calculate first

\[
\frac{P(D_1 = m_1 + 1 \cap Y = C_1, \cdot)}{P(D_1 = m_1 \cap Y = C_1, \cdot)} = \frac{P(C_1, \cdots, m_1 + 1)}{P(C_1, \cdots, m_1)} = \sum_{u \in \{C_1, \mathcal{S}(C_1), \cdots, i=1, \ldots, l\}} \frac{m_1}{P(C_1, \cdots, u) P(C_1, \cdots)}
\]

\[
= \frac{1}{\lambda + \mu} \left( \lambda_{C_1} + \mu_{\mathcal{S}(C_1)} + \lambda \right) = 1 - \frac{\mu_{\mathcal{S}(C_1)} - \lambda_{C_1}}{\lambda + \mu},
\]

\( \square \) Springer
from which we obtain
\[
P(D_1 = m_1 \cap Y = C_1, \cdot) = B \frac{\lambda C_1}{\lambda + \mu} \left( 1 - \frac{\mu \mathcal{S}(C_1) - \lambda C_1}{\lambda + \mu} \right)^{m_1 - 1}.
\]

Proceeding in the same way for \( D_2, \ldots, D_k \), we obtain that
\[
P(D_1 = m_1, \ldots, D_k = m_k \cap C_1, \cdot, \ldots, C_k, \cdot) = B \prod_{\ell=1}^{k} \frac{\lambda C_\ell}{\lambda + \mu} \left( 1 - \frac{\mu \mathcal{S}(C_1, \ldots, C_\ell) - \lambda C_1, \ldots, C_\ell}{\lambda + \mu} \right)^{m_\ell - 1}.
\]

Dividing by \( \pi_Y(C_1, \cdot, \ldots, C_k, \cdot) \) given in Eq. (13), we obtain
\[
P\left( D_1 = m_1, \ldots, D_k = m_k \mid Y(N) = (C_1, \cdot, \ldots, C_k, \cdot) \right) = \prod_{\ell=1}^{k} \frac{\mu \mathcal{S}(C_1, \ldots, C_\ell) - \lambda C_1, \ldots, C_\ell}{\lambda + \mu} \left( 1 - \frac{\mu \mathcal{S}(C_1, \ldots, C_\ell) - \lambda C_1, \ldots, C_\ell}{\lambda + \mu} \right)^{m_\ell - 1}
\]
which is what we needed to show. \( \square \)

By definition, if \( Y(N) = (C_1, \cdot, \ldots, C_k, \cdot) \), with \( C_\ell = c_\ell \), and \( z^{N+1} = s_j \), so that \( s_j \) matches with \( C_\ell \) (i.e., \( s_j \notin \mathcal{S}(C_1, \ldots, C_{\ell-1}) \) and \( s_j \in \mathcal{S}(C_\ell) \)), then
\[
L_{s_j, c_\ell} = \sum_{h=\ell}^{k} D_h(C_1, \ldots, C_k)
\]
and this will happen with probability \( \pi_Y((C_1, \cdot, \ldots, C_k, \cdot) \frac{\mu s_j}{\mu}) \). We obtain directly:

**Theorem 9** The delays \( L_{s_j, c_\ell} \) and \( L_{c_\ell} \) are distributed as mixtures of convolutions of geometric random variables, with probability generating functions

\[
G_{s_j, c_\ell}(Z) = \frac{1}{r_{s_j, c_\ell}} \frac{\mu s_j}{\mu} \sum_{k=1}^{1} \sum_{C \subseteq \mathcal{E}, |C|=k} \sum_{(C_1, \ldots, C_k) \in \mathcal{P}(C)} \sum_{\ell=1}^{k} \pi_Y(C_1, \cdot, \ldots, C_k, \cdot) \
\times 1((C_\ell = c_\ell) \wedge (s_j \notin \mathcal{S}(C_1, \ldots, C_{\ell-1}))) \times \prod_{h=\ell}^{k} \frac{Z^{\mu \mathcal{S}(C_1, \ldots, C_h) - \lambda C_1, \ldots, C_h}}{1 - Z^{\mu \mathcal{S}(C_1, \ldots, C_h) - \lambda C_1, \ldots, C_h}}.
\]

\[
G_{c_\ell}(Z) = \sum_{s_j \in \mathcal{S}(c_\ell)} \frac{r_{s_j, c_\ell}}{\sum_{s_h \in \mathcal{S}(c_\ell)} r_{s_h, c_\ell}} G_{s_j, c_\ell}(Z).
\]

\( \square \) Springer
Proof We again use discrete time PASTA to state that when we consider \( z^{N+1} = s_j \), \( Y(N), U(N) \) are distributed according to their stationary distribution. For a particular \( s_j \) and \( c_i \), if for \( Y = (C_1, \ldots, C_k, \cdot) \) such that \( z^{N+1} = s_j \) will match with \( C_\ell = c_i \), the delay is \( \sum_{h=\ell}^k D_h(C_1, \ldots, C_k) \), which is the sum of independent geometric random variables, with parameters given by Theorem 8. This has generating function as given by the product expression in (18).

For \( L_{s_j,c_i} \), we need to take the weighted sum of these generating functions of the sum of geometrics. The sum is over \( Y = (C_1, \ldots, C_k, \cdot) \), for which \( s_j \) matches with \( C_\ell = c_i \), and that is expressed by \( \pi_Y(C_1, \ldots, C_k, \cdot) \) multiplied by the indicator function. The added terms \( \frac{\mu s_j}{\lambda} \) and \( 1/r_{s_j,c_i} \) ensure that the weights add up to 1.

The generating function of \( L_{c_i} \) is a weighted sum of the \( G_{s_j,c_i}(Z) \). □

**Corollary 1** The mean and variance of \( L_{s_j,c_i} \) are given by

\[
E(L_{s_j,c_i}) = \frac{1}{\rho_{s_j,c_i}} \frac{\mu s_j}{\rho} \sum_{k=1}^{J} \sum_{C \subseteq \mathcal{E} : |C|=k} \sum_{(C_1,\ldots,C_k) \in \mathcal{P} (C)} \sum_{\ell=1}^{k} \pi_Y(C_1,\ldots,C_k,\cdot) \times 1((C_\ell = c_i) \land (s_j \notin \mathcal{P} (C_1,\ldots,C_{\ell-1}))) \times \sum_{h=\ell}^{k} \left( \frac{\mu_{\mathcal{P} (C_1,\ldots,C_k)} - \lambda_{C_1,\ldots,C_k}}{\lambda + \mu} \right)^{-1},
\]

\[
Var(L_{s_j,c_i}) = \left\{ \frac{1}{\rho_{s_j,c_i}} \frac{\mu s_j}{\rho} \sum_{k=1}^{J} \sum_{C \subseteq \mathcal{E} : |C|=k} \sum_{(C_1,\ldots,C_k) \in \mathcal{P} (C)} \sum_{\ell=1}^{k} \pi_Y(C_1,\ldots,C_k,\cdot) \times 1((C_\ell = c_i) \land (s_j \notin \mathcal{P} (C_1,\ldots,C_{\ell-1}))) \times \sum_{h=\ell}^{k} \left( \frac{1 - \mu_{\mathcal{P} (C_1,\ldots,C_k)} - \lambda_{C_1,\ldots,C_k}}{\lambda + \mu} \right)^{-1} \right\}^2 + \left( \sum_{h=\ell}^{k} \left( \frac{\mu_{\mathcal{P} (C_1,\ldots,C_k)} - \lambda_{C_1,\ldots,C_k}}{\lambda + \mu} \right)^{-1} \right)^2 \right\} + \left( \sum_{h=\ell}^{k} \left( \frac{\mu_{\mathcal{P} (C_1,\ldots,C_k)} - \lambda_{C_1,\ldots,C_k}}{\lambda + \mu} \right)^{-1} \right)^2 \right\} \times -E(L_{s_j,c_i})^2.
\]

The mean and variance of \( L_{c_i} \) can be obtained from these directly.

**Proof** This follows from properties of geometric random variables:

\[
X \sim Geometric(p), \quad P(X = n) = p(1 - p)^{n-1}, \quad n = 1, 2, \ldots,
\]

\[
G_X(Z) = E(Z^X) = \frac{Zp}{1 - Z(1 - p)}, \quad E(X) = p^{-1}, \quad Var(X) = \frac{1 - p}{p^2}.
\]

Conditional on \( Y = (C_1, \ldots, C_k) \), the delay \( L_{s_j,c_i} \) is a sum of independent geometric random variables, with mean and variance that are the sums of the means and variances of the geometric random variables. The unconditional mean follows directly. The unconditional variance is the expectation of the conditional variances plus the variance of the conditional means. □

Calculation of means and variances of the delays is similar to the calculation of the matching rates. We provide here an algorithm for these calculations. It follows the steps of Algorithm 1 and only requires addition of some calculations to line 8, and the additional line 9a.
Algorithm 2: Calculation of means and variances of delays

1a Initialize: \( E_{s_j, c_i} := 0; \quad E^2_{s_j, c_i} := 0; \quad V_{s_j, c_i} := 0. \)

Repeat steps 2–7 of Algorithm 1.

8a If \( C_\ell \in \mathcal{C}(s_j) \land \text{Match}(s_j) = \emptyset \) then:

\[
\text{Match}(s_j) := C_\ell; \quad r_{s_j, C_\ell} := r_{s_j, C_\ell} + X.
\]

\( AE := 0; \quad AV := 0. \)

8b For \( h = l, |A|: \)

\[
P_h = 1/(\lambda + \mu) R_h; \quad AE := AE + 1/P_h, \quad AV := AV + (1 - P_h)/P_h^2.
\]

8c \( E_{s_j, C_\ell} := E_{s_j, C_\ell} + X AE; \quad E^2_{s_j, C_\ell} := E^2_{s_j, C_\ell} + X AE^2; \quad V_{s_j, C_\ell} := V_{s_j, C_\ell} + X AV. \)

9a \( E(L_{s_j, c_i}) := B \frac{1}{r_{s_j, c_i}} \frac{\mu_{s_j}}{\mu} E_{s_j, c_i}; \quad E^2_{s_j, c_i} := B \frac{1}{r_{s_j, c_i}} \frac{\mu_{s_j}}{\mu} E^2_{s_j, c_i}; \quad V_{s_j, c_i} := B \frac{1}{r_{s_j, c_i}} \frac{\mu_{s_j}}{\mu} V_{s_j, c_i}. \)

\( \text{Var}(L_{s_j, c_i}) := V_{s_j, c_i} + E^2_{s_j, c_i} - E(L_{s_j, c_i})^2. \)

We explain the algorithm: Inside the loop of line 8, for \( (C_1, \ldots, C_k) \) we locate matches \((s_j, C_\ell)\) in the if statement and then calculate the sum from \( l \) to \( k \) of the conditional mean and variance of \( D_\ell(C_1, \ldots, C_k) \), multiplied by \( X \propto \pi_Y(C_1, \ldots, C_k) \). We accumulate these over all \( (C_1, \ldots, C_k) \), to calculate the unconditional mean of the conditional variances, and the unconditional variance of the conditional means, leading to the means and variances in line 9a.

Delays can be translated to waiting times. Assume arrivals of agents and goods are independent Poisson processes, with rates \( \lambda, \mu_{s_j} \), if \( z^m = c_i \) is matched to \( z^n = s_j \), then we denote the \((s_j, c_i)\) wait between them as \( W_{s_j, c_i} \), which is the sum of \( n - m \) exponential random variables, with rate parameter \( \lambda + \mu \). If \( z^m = c_i \) is matched to \( z^n \in \mathcal{S} \), we denote the wait of \( z^m \) by \( W_{c_i} \). We then have:

**Corollary 2** (Waiting Times) Under the Poisson assumption the waiting times are distributed as mixtures of convolutions of independent exponential random variables. The moment generating function of \( W_{s_j, c_i} \) is given by

\[
M_{s_j, c_i}(s) = \frac{1}{r_{s_j, c_i}} \frac{\mu_{s_j}}{\mu} \sum_{k=1}^l \sum_{C \subseteq \mathcal{S}, |C| = k} \sum_{(C_1, \ldots, C_k) \in \mathcal{P}(C)} \sum_{\ell=1}^k \pi_Y(C_1, \ldots, C_k, \cdot) 
\times 1 \left((C_\ell = c_i) \land (s_j \notin \mathcal{S}(C_1, \ldots, C_{\ell-1}))\right) \times \prod_{h=l}^k \frac{\mu \mathcal{S}(C_1, \ldots, C_h) - \lambda C_1, \ldots, C_h}{\mu \mathcal{S}(C_1, \ldots, C_h) - \lambda C_1, \ldots, C_h - s}
\]

\[ (22) \]

\( \mathcal{S} \) Springer
with mean and variance

\[
E(W_{s_j,c_j}) = \frac{1}{r_{s_j,c_j}} \frac{\mu_{s_j}}{\mu} \sum_{k=1}^{l} \sum_{C \subseteq \mathcal{G}, |C|=k} \sum_{(C_1,\ldots,C_k) \in \mathcal{P}(C)} \sum_{\ell=1}^{k} \pi_{\ell}(C_1,\ldots,C_k) \\
\times 1(\{C_\ell = c_j\} \land (s_j \notin \mathcal{I}(C_1,\ldots,C_{\ell-1})) \} \times \frac{1}{\mu_{\mathcal{I}}(C_1,\ldots,C_k) - \lambda_{C_1,\ldots,C_k}}.
\]

(23)

\[
Var(W_{s_j,c_j}) = \left\{ \frac{1}{r_{s_j,c_j}} \frac{\mu_{s_j}}{\mu} \sum_{k=1}^{l} \sum_{C \subseteq \mathcal{G}, |C|=k} \sum_{(C_1,\ldots,C_k) \in \mathcal{P}(C)} \sum_{\ell=1}^{k} \pi_{\ell}(C_1,\ldots,C_k) \right\} - E(W_{s_j,c_j})^2.
\]

(24)

The mean and variance of \(W_{c_j}\) can be obtained from these directly.

**Proof** Recall that if \(U \sim \text{Exp}(\theta)\) then its moment generating function is \(M_U(s) = \frac{\theta}{\theta - s}\), its mean is \(E(U) = \frac{1}{\theta}\), and its variance is \(Var(U) = \frac{1}{\theta^2}\).

If \(U_i \sim \text{Exp}(\theta)\) i.i.d., and \(Y \sim \text{Geom}(p)\), then \(V = \sum_{i=1}^{Y} U_i \sim \text{Exp}(\theta p)\). Hence, each \(D_i(C_1,\ldots,C_k)\) contributes an exponential wait with rate parameter \((\mu_{\mathcal{I}}(C_1,\ldots,C_k) - \lambda_{C_1,\ldots,C_k})\).

\[\square\]

7 Light and heavy traffic, relation to symmetric FCFS matching

In this section, we consider our system when the overall traffic intensity, given by \(\lambda/\mu\), is close to 0, i.e., the system is in light traffic, or close to and below 1, i.e., the system is in heavy traffic. We wish to obtain the limiting matching rates under these conditions.

For light traffic, we have:

**Proposition 4** If \(\rho = \lambda/\mu \to 0\), then \(r_{s_j,\emptyset} \to 1\) and for \((s_j, c_j) \in \mathcal{G},\) \(\frac{r_{s_j,c_j}}{1-r_{s_j,\emptyset}} \to \frac{\mu_{s_j}}{\sum_{s_k \in \mathcal{G} \setminus \{c_j\}} \mu_{s_k}}\).

**Proof** Clearly, when \(\rho \to 1\) most goods will be lost, so \(r_{s_j,\emptyset} \to 1\). If \(\lambda/\mu \to 0\), then for agent \(z^N = c_i\), with high probability there will be no earlier unmatched agents prior to \(Z^N\), i.e., \(X(N-1) = \emptyset\), and \(z^N\) will be followed by a large number of goods of all types. \(z^N\) will then match with the first good with which it is compatible, and this good will be of type \(s_j\) with probability \(\frac{\mu_{s_j}}{\sum_{s_k \in \mathcal{G} \setminus \{c_j\}} \mu_{s_k}}\). To obtain \(\frac{r_{s_j,c_j}}{1-r_{s_j,\emptyset}}\), this needs to be multiplied by \(\alpha_{c_j}\).

\[\square\]

We next define a property which ensures stability for all \(0 < \lambda/\mu < 1\).
Definition 6 (Complete Resource Pooling) For given compatibility graph \( G \) and frequency vectors \( \alpha, \beta \), we say that the system has complete resource pooling (CRP) if the following three equivalent conditions hold for all subsets \( C \subseteq \mathcal{C}, C \neq \emptyset, \mathcal{C} \), and \( S \subseteq \mathcal{S}, S \neq \emptyset, \mathcal{S} \):

\[
\alpha_C < \beta_{\mathcal{C}(C)}, \quad \beta_S < \alpha_{\mathcal{C}(S)}, \quad \alpha_{\mathcal{S}(S)} < \beta_S.
\]  

This condition is also the necessary and sufficient condition for stability of the symmetric FCFS bipartite matching model of [1,3,8]. By Theorem 2, CRP implies that our system is stable for all \( 0 < \lambda / \mu < 1 \). If CRP does not hold, then our system will be stable if and only if

\[
0 < \lambda / \mu < \min_{C \neq \emptyset, \mathcal{C}} \frac{\beta_{\mathcal{C}(C)}}{\alpha_C}.
\]

As we shall see in the next theorem, in heavy traffic our model behaves more and more like the symmetric FCFS bipartite matching model of [1,3,8]. For clarity, we will denote quantities relating to our directed FCFS matching model by superscript \( D \) and quantities relating to the symmetric FCFS bipartite matching model by superscript \( B \).

It was noted in Moyal, Busic, and Mairesse [19] that in a FCFS matching model with a single sequence of arrivals of both agents and goods, if agents as well as goods are allowed to wait until they find a match the system can never be stable, even if the data satisfy the property of CRP, since the total number of agents minus the total number of goods in the sequence starting from empty behaves like an unconstrained one-dimensional random walk, which is always either transient if CRP does not hold, or null recurrent if CRP holds. Therefore, the number of unmatched agents or goods diverges in the sense that if we define:

\[
\Delta(N) = \text{absolute difference between the number of unmatched agents and unmatched goods by time } N, \text{ then } P(\Delta(N) > n) \rightarrow 1 \text{ as } N \rightarrow \infty.
\]

However, in the symmetric FCFS matching model of [1,3,8] there are two sequences, one only of agents and the other only of goods. Starting from empty, by time \( N \) the number of agent arrivals equals the number of goods arrivals, and if CRP holds then the system is stable, and the number of unmatched agents or goods converges to a stationary distribution.

For our system, we defined the Markov processes \( X^D(N) = (c^1, c^2, \ldots, c^L) \) that list all unmatched agents at time \( N \) (Sect. 2). For the symmetric FCFS matching model in [1], we define the Markov process \( X^B(N) = (c^1, c^2, \ldots, c^L) \) which lists unmatched agents after all goods up to \( N \) have been matched to agents earlier or later than \( N \), and we include in this list all the unmatched agents starting from the first unmatched agent and continuing until at least one agent of each type is included in the list. The process \( X^B(N) \) is a slight modification of the so-called natural process defined and studied in Section 5.5 of [1]. The stationary distribution of \( X^D(N) \) was given in (4). The stationary distribution of \( X^B(N) \) can be derived in the same way that the stationary distribution of the natural process in [1] is derived.
The stationary distributions of $X^D(N)$ (by Theorem 1) and of $X^B(N)$ (see [1] Section 5.5) are
\[
\pi^D_X(c^1, \ldots, c^L) \propto \left(\frac{\lambda}{\mu}\right)^L \prod_{\ell=1}^{L} \frac{\alpha_{c^\ell}}{\beta_{\mathcal{I}}(\{c^1, \ldots, c^\ell\})},
\]
\[
\pi^B_X(c^1, \ldots, c^L) \propto \prod_{\ell=1}^{L} \frac{\alpha_{c^\ell}}{\beta_{\mathcal{I}}(\{c^1, \ldots, c^\ell\})}.
\] (26)

We also defined the Markov process $Y^D(N) = (C_1, n_1, \ldots, C_k, n_k)$ (Sect. 4). We define similarly the process $Y^B(N) = (C_1, n_1, \ldots, C_{I-1}, n_{I-1}, C_I)$ which is obtained from $X^B(N)$ by listing the types of agents in their order of appearance and counting the number of unmatched agents between each pair. Note that $C_I$ is determined as the one type not included in $C_1, \ldots, C_{I-1}$, and it is followed by the infinite sequence of agents of all types not yet matched. The stationary distribution of $Y^B(N)$ is derived similarly to the derivation of the stationary distribution of $Y^D(N)$ in Theorem 6. We have
\[
\pi^D_Y(C_1, n_1, \ldots, C_k, n_k) \propto \prod_{\ell=1}^{k} \left(\frac{\lambda}{\mu}\right) \frac{\alpha_{C_\ell}}{\beta_{\mathcal{I}}(\{C_1, \ldots, C_\ell\})} \left(\frac{\lambda}{\mu}\right) \frac{\alpha_{\{C_1, \ldots, C_\ell\}}}{\beta_{\mathcal{I}}(\{C_1, \ldots, C_\ell\})} n_\ell,
\]
\[
\pi^B_Y(C_1, n_1, \ldots, C_{I-1}, n_{I-1}, C_I) \propto \alpha_{C_I} \prod_{\ell=1}^{I-1} \frac{\alpha_{C_\ell}}{\beta_{\mathcal{I}}(\{C_1, \ldots, C_\ell\})} \left(\frac{\alpha_{\{C_1, \ldots, C_\ell\}}}{\beta_{\mathcal{I}}(\{C_1, \ldots, C_\ell\})}\right) n_\ell.
\] (27)

Theorem 10 Assume CRP holds. When $\frac{\lambda}{\mu} \not\to 1$, the following holds for the stationary distribution of $Y^D(N)$:

(i) The probability that $Y^D(N) = (C_1, n_1, \ldots, C_k, n_k)$, where $k < I$, converges to 0, i.e., the set of waiting agents always includes agents of all types.

(ii) The probability that $Y^D(N) = (C_1, n_1, \ldots, C_I, n_I)$ and $n_I$ is less than any constant goes to 0, i.e., while $n_1, \ldots, n_{I-1}$ remain stable, $P(n_i > m) \to 1$ for all $m > 0$.

(iii) Conditional on $k = I$, $n_I > 0$, the process $Y^D_0(N) = (C_1, n_1, \ldots, C_I)$, obtained from $Y^D(N)$ by deleting $n_I$ from the state, is a Markov process.

(iv) The stationary distribution and the transition probabilities of $Y^D_0(N)$ converge to the stationary distribution and the transition probabilities of $Y^B(N)$ in total variation distance.

Proof (i) We note that $\frac{\alpha_{\{C_1, \ldots, C_\ell\}}}{\beta_{\mathcal{I}}(\{C_1, \ldots, C_\ell\})} \to 1$. By Theorem 8, all the $n_i$ have geometric distributions. For $\ell < I$, we have by CRP
\[
\frac{\lambda}{\mu} \frac{\alpha_{\{C_1, \ldots, C_\ell\}}}{\beta_{\mathcal{I}}(\{C_1, \ldots, C_\ell\})} < \frac{\alpha_{\{C_1, \ldots, C_\ell\}}}{\beta_{\mathcal{I}}(\{C_1, \ldots, C_\ell\})} \leq \max_{C \subset \mathcal{E} ; C \neq \emptyset , \mathcal{E}} \frac{\alpha_C}{\beta_{\mathcal{I}}(C)} = \text{constant} < 1,
\]
while for $k = I$ we have $\frac{\lambda}{\mu} \frac{\alpha_{\{C_1, \ldots, C_I\}}}{\beta_{\mathcal{I}}(\{C_1, \ldots, C_I\})} = \frac{\lambda}{\mu} \not\to 1$. It follows that most of the sum of terms needed to compute the normalizing constant $B$ is concentrated in the terms for which $k = I$.

(ii) When $k = I$, $n_I$ is distributed as a geometric random variable and so the probability that $n_I$ is less than any constant $A$ is $1 - \left(\frac{\lambda}{\mu}\right)^A \to 0$.

(iii) If $k = I$ and $n_i >> 0$, the possible transitions from $Y^D(N)$ are: If $z^{N+1} = c_i$, it augments $n_i$. Otherwise, if $z^{N+1} = s_j$ it will match with one of $C_1, \ldots, C_I$. 

\[\square\] Springer
say $C_\ell$, and as a result the order of $C_1, \ldots, C_I$ and number of unmatched agents between them will change. At worst, $C_\ell$ appears only once, in which case it become the new $C_I$ and $n_I$ decreases by a geometric number of terms. We do not need the information on $n_I$ except to know it is positive enough.

(iv) It is seen immediately that $\pi_{Y_D}(C_1, n_1, \ldots, C_I) \to \pi_{Y_B}(C_1, n_1, \ldots, C_I)$. The description of the transitions in part (iii) shows also that the conditional transitions of $\pi_{Y_D}$ when $n_I \gg 0$ are the same as those of $Y_B$. \hfill \Box

It follows immediately that:

**Corollary 3** Assume CRP holds. When $\frac{\lambda}{\mu} \not\sim 1$, the matching rates of the directed FCFS bipartite matching model converge to the matching rates of the symmetric FCFS bipartite matching model of [1], while the fraction of unassigned goods converges to 0.

### 8 A simple illustrative example

We consider a simple system, with three types of agents and three types of goods and $G = \{(s_1, c_1), (s_1, c_2), (s_2, c_1), (s_2, c_3), (s_3, c_2), (s_3, c_3)\}$, with $\alpha = (.3, .5, .2)$, $\beta = (.3, .3, .4)$, as illustrated in Fig. 4.

With the help of Algorithms 1 and 2, we can calculate matching rates and means and variances of delays. For $\mu = 1, \lambda = 0.7$ and $\rho = \lambda / \mu = 0.7$, we obtain

$$
\begin{align*}
E(L_{s_j, c_i}) &= \begin{bmatrix}
    c_1 & c_2 & c_3 \\
    s_1 & 7.63 & 7.64 & 0 \\
    s_2 & 7.14 & 0 & 6.38 \\
    s_3 & 0 & 7.40 & 6.45 \\
\end{bmatrix}, & \sigma(L_{s_j, c_i}) &= \begin{bmatrix}
    c_1 & c_2 & c_3 \\
    s_1 & 6.14 & 6.30 & 0 \\
    s_2 & 5.97 & 0 & 5.41 \\
    s_3 & 0 & 6.21 & 5.46 \\
\end{bmatrix}
\end{align*}
$$

From this calculation, we see that the loss probabilities for the various goods may be quite different. We also obtain the following for the delays of the various types of
agents:

\[ E(L_{c_1}) = 7.35, \quad E(L_{c_2}) = 7.50, \quad E(L_{c_3}) = 6.38, \]
\[ \sigma(L_{c_1}) = 6.05, \quad \sigma(L_{c_2}) = 6.25, \quad \sigma(L_{c_3}) = 5.44. \]

Assuming now that arrivals are Poisson, with rates \( \lambda, \mu \), then using Corollary 2 we have

\[ E(W_{c_1}) = 4.33, \quad E(W_{c_2}) = 4.41, \quad E(W_{c_3}) = 3.75. \]
\[ \sigma(W_{c_1}) = 3.90, \quad \sigma(W_{c_2}) = 4.01, \quad \sigma(W_{c_3}) = 3.53. \]

We can see from these results that the FCFS matching policy achieves balanced supply of goods, the expected delays for the three types of agents are quite close to each other, and even the expected delays for individual \( (s_j, c_i) \in \mathcal{G} \) are quite close. We also see that the standard deviation of delays is close to the expectation, which indicates that the distribution of delays and waiting times is close to exponential.

By Proposition 4 and Theorem 10, we know the limits of the matching rates as \( \rho = \frac{\lambda}{\mu} \) approaches 0 or 1. Figure 5 shows how the matching rates for \( (s_j, c_i) \in \mathcal{G} \) vary as \( 0 < \rho < 1 \).

We see that the variations in values of matching rates are not very large. In very light traffic, each agent receives goods immediately, and is not influenced by the agents of the other types, so the matching rates are as if each type of agent has its own set of goods without sharing. In heavy traffic, there is CRP and sharing of all the resources of the goods. The change over the range of values of \( \rho \) is monotone. For a possible explanation of how the rates change, consider agent type \( c_2 \). In light traffic, it receives \( \frac{3}{7} \) from \( s_1 \) and \( \frac{4}{7} \) from \( s_3 \). In heavy traffic, goods of types \( s_1 \) are shared with \( c_1 \), and goods of types \( s_3 \) are shared with \( c_3 \), but the requirements of \( c_1 \) are 0.3 and those of \( c_3 \)
are only 0.2, so more goods of type $s_3$ then of type $s_1$ are diverted to $c_2$, and so $r_{s_1,c_2}$ decreases and $r_{s_3,c_2}$ increases with $\rho$.

It is perhaps worthwhile to compare the bipartite matching of this example to what would be the results of eliminating the pooling effect. Assume each type of agent is matched with only one type of good, as follows: $G_1 = \{(s_1, c_1), (s_2, c_3), (s_3, c_2)\}$. (These edges are marked by heavier lines in Fig. 4.) In Fig. 6, we plot the average waiting times for each type of agent, for our system under the pooled bipartite matching, compared with the waiting times for discriminating matching according to $G_1$. Under discriminating matching, each agent–good pair behaves like an M/M/1 queue with expected waiting time $\frac{1}{\mu_{s_j} - \lambda_{c_i}}$. We see that the pair $(s_3, c_2)$ becomes unstable for $\rho > 0.8$, the pair $(s_2, c_3)$ has discriminating traffic intensity $< \frac{2}{3}$, but pooled waiting time is still shorter than discriminating for all $\rho < 0.8$. For the pair $(s_1, c_1)$, where discriminating traffic intensity varies from 0 to 1, the ratio of waiting times starts at $\approx 2$ in light traffic and converges to $\approx 3.3$ as $\rho \searrow 1$. This is slightly better than the ratio we expect from pooling 3 i.i.d. M/M/1 queues. It is very noteworthy to observe that pooled service results with similar waiting times for all types of agents.

**A Appendix: Completion of proofs**

**A.1 Completion of the Proof of Theorem 4**

**Proof** We start from $U(N) = u = (u_1, \ldots, u_L)$. There are four cases to check:

Case (i): $x = c_i$

If $U(N) = u = (u_1, \ldots, u_L)$ and $z^{N+1} = c_i$, then $U(N+1) = u' = (u_1, \ldots, u_L, c_i)$. The probability of the transition $u \to u'$ is $\frac{\lambda}{\lambda + \mu} \alpha_{c_i}$.

In that case, $W(N) = \tilde{u}^{\gamma} = (\tilde{c}_i, \tilde{u}_L, \ldots, \tilde{u}_1)$. This means that $z^{N+1}$ has already been matched and exchanged and is now $\tilde{z}^{N+1} = \tilde{x} = \tilde{c}_i$. Therefore, all agents up to $N + 1$ have been matched, and $W(N + 1) = \tilde{\tilde{u}} = (\tilde{u}_L, \ldots, \tilde{u}_1)$. The probability of the transition $\tilde{u}' \to \tilde{\tilde{u}}$ is 1.
We now check
\[
\frac{\pi_U(u_1, \ldots, u_L, c_i)}{\pi_U(u_1, \ldots, u_L)} = \frac{\lambda}{\lambda + \mu} \alpha_{c_i},
\]
\[
P(U(N + 1) = (u_1, \ldots, u_L, c_i) | U(N) = (u_1, \ldots, u_L)) = \frac{\lambda}{\lambda + \mu} \alpha_{c_i}.
\]

Case (ii): \(x = s_j\), and \(s_j\) is incompatible with the unmatched agents in \((u_1, \ldots, u_L)\)

If \(U(N) = u = (u_1, \ldots, u_L)\) and \(z^{N+1} = s_j\) and \(s_j\) is incompatible with any of the unmatched agents in \((u_1, \ldots, u_L)\), then \(s_j\) remains unmatched and is replaced by \(\tilde{s}_j\) (matched with itself), so that \(U(N + 1) = u' = (u_1, \ldots, u_L, \tilde{s}_j)\). The probability of this transition is 1.

In that case, \(W(N) = \tilde{u}' = (s_j, \tilde{u}_L, \ldots, \tilde{u}_1)\). This means that the \(N + 1\) element is \(z^{N+1} = s_j\). Since all agents prior to \(N\) have been matched, \(s_j\) will remain unmatched, and in the transition from \(W(N)\) to \(W(N + 1)\) there will be no change in \((\tilde{u}_L, \ldots, \tilde{u}_1)\), and so \(W(N + 1) = (\tilde{u}_L, \ldots, \tilde{u}_1) = \tilde{u}\). The probability of this transition is again 1.

We now check
\[
\frac{\pi_U(u_1, \ldots, u_L, \tilde{s}_j)}{\pi_U(u_1, \ldots, u_L)} = \frac{\mu}{\lambda + \mu} \beta_{s_j},
\]
\[
P(U(N + 1) = (u_1, \ldots, u_L, \tilde{s}_j) | U(N) = (u_1, \ldots, u_L)) = \frac{\mu}{\lambda + \mu} \beta_{s_j}.
\]

Case (iii): \(x = s_j\), which is matched FCFS with \(u_\ell = c_i\), where \(l > 1\)

If \(U(N) = u = (u_1, \ldots, u_L)\) and \(z^{N+1} = s_j\) and \(s_j\) is compatible with \(u_\ell = c_i\) and is incompatible with all unmatched agents among \(u_1, \ldots, u_{\ell-1}\), then \(s_j\) is matched and exchanged with \(u_\ell\), so \(U(N + 1) = u' = (u_1, \ldots, u_{\ell-1}, \tilde{s}_j, u_\ell+1, \ldots, u_L, \tilde{c}_i)\). The probability of this transition is 1.

In that case, \(W(N) = \tilde{u}' = (c_i, \tilde{u}_L, \ldots, \tilde{u}_\ell+1, \tilde{s}_j, \tilde{u}_{\ell-1}, \ldots, \tilde{u}_1)\). Because \(z^{N+1} = c_i\), it needs to be matched and exchanged in the transition from \(W(N)\) to \(W(N + 1)\). Clearly, it can be matched and exchanged with \(s_j\) which is between \(\tilde{u}_{\ell+1}, \tilde{u}_{\ell-1}\). We claim that \(c_i\) cannot be matched to any of \(\tilde{u}_L, \ldots, \tilde{u}_{\ell+1}\). Recall that it was left unmatched in \(U(N)\). Assume some \(\ell + 1 \leq k \leq L\), \(\tilde{u}_k = s_{j'}\) is compatible with \(c_i\). In that case, \(u_k = \tilde{s}_{j'}\) and there are two possibilities: Either \(z^{N+k-1} = s_{j'}\) could not be matched and was exchanged to \(\tilde{s}_{j'}\), or \(z^{N+k-1} = c_i\) that was matched and exchanged with \(s_{j'}\) that appeared later in the sequence but before \(N + 1\). But in either of these cases \(s_{j'}\) should then have been matched and exchanged with \(z^{N+\ell-1} = c_i\), which is in contradiction to \(u_\ell = c_i\) being unmatched in \(U(N)\). It follows that in this case in the transition from \(W(N)\) to \(W(N + 1)\), \(c_i\) will be matched and exchanged with \(s_j\) and \(W(N + 1) = (\tilde{u}_L, \ldots, \tilde{u}_{\ell+1}, \tilde{c}_i, \tilde{u}_{\ell-1}, \ldots, \tilde{u}_1)\), which is \(\tilde{u}\). The probability of this transition is 1.
We now check
\[ \frac{\pi_U(u_1, \ldots, u_{\ell-1}, \tilde{s}_j, u_{\ell+1}, \ldots, u_L, \tilde{c}_i)}{\pi_U(u_1, \ldots, u_{\ell-1}, c_i, u_{\ell+1}, \ldots, u_L)} = \mu \\frac{\lambda}{\lambda + \mu} \beta_{s_j}, \]
\[ \frac{P(U(N+1) = (u_1, \ldots, u_{\ell-1}, \tilde{s}_j, u_{\ell+1}, \ldots, u_L, \tilde{c}_i) \mid U(N) = (u_1, \ldots, u_{\ell-1}, c_i, u_{\ell+1}, \ldots, u_L))}{P(W(N+1) = (\tilde{u}_L, \ldots, \tilde{u}_{\ell+1}, \tilde{c}_i, \tilde{u}_{\ell-1}, \ldots, \tilde{u}_1) \mid W(N) = (c_i, \tilde{u}_L, \ldots, \tilde{u}_{\ell+1}, s_j, \tilde{u}_{\ell-1}, \ldots, \tilde{u}_1))} = \mu \\frac{\lambda}{\lambda + \mu} \beta_{s_j}. \]

Case (iv): \( x = s_j \), which can match FCFS to \( u_1 = c_i \).

In Cases (i)–(iii), the reversed transition was \( \tilde{u}^{(i)} \rightarrow \tilde{u} \) with probability 1. In Case (iv), this is no longer so, and we will need to calculate the probability of this transition.

Assume \( U(N) = u = (c_i, u_2, \ldots, u_{\ell-1}, c_i', u_{\ell+1}, \ldots, u_L) \), where \( u_2, \ldots, u_{\ell-1} \) contain no unmatched agents, and also it does not contain any goods compatible with \( c_i \). Assume \( z^{N+1} = s_j \), and \( s_j \) is compatible with \( u_1 = c_i \). Then, \( s_j \) is matched and exchanged with \( u_1 \). (Recall that \( u_1 \) is always an unmatched agent.) The resulting state then is \( U(N+1) = u' = (c_i', u_{\ell+1}, \ldots, u_L, \tilde{c}_i) \), because now \( z^{N+1} = s_j \) is the first unmatched agent. The probability of this transition \( u \rightarrow u' \) is \( \mu \frac{\lambda}{\lambda + \mu} \beta_{s_j} \).

In that case, \( W(N) = \tilde{u}' = (c_i, \tilde{u}_L, \ldots, \tilde{u}_{\ell+1}, \tilde{c}_i') \). Because \( z^{N+1} = c_i \), it needs to be matched and exchanged in the transition from \( W(N) \) to \( W(N+1) \). By the same argument as for Case (iii), \( c_i \) is not compatible with any unmatched goods in \( \tilde{u}_L, \ldots, \tilde{u}_{\ell+1} \).

We now consider the possibility that conditional on \( W(N) = \tilde{u}' \), the next transition will be to the state \( W(N+1) = \tilde{u} = (\tilde{u}_L, \ldots, \tilde{u}_{\ell+1}, \tilde{c}_i', \tilde{u}_{\ell-1}, \ldots, \tilde{u}_2, s_j) \). This will happen if and only if \( (z^{N+1}, \ldots, z^{N+\ell-1}) = (\tilde{u}_{\ell-1}, \ldots, \tilde{u}_2, s_j) \). Denote by \( \theta \) the probability of this event. Then
\[ \theta = P(W(N+1) = \tilde{u} \mid W(N) = \tilde{u}') = P((z^{N+1}, \ldots, z^{N+\ell-1}) = (\tilde{u}_{\ell-1}, \ldots, \tilde{u}_2, s_j)) \]
\[ = \left( \frac{\lambda}{\lambda + \mu} \right)^{\text{agents}} \left( \frac{\mu}{\lambda + \mu} \right)^{\text{goods}} \prod_{c_i \in \mathcal{C}} \beta_{s_j}^{z_{c_i}+z_{\tilde{c}_i}} \times \frac{\mu}{\lambda + \mu} \beta_{s_j}, \]
where we are counting the number of agents and of goods overall, and of each type separately, in \( \tilde{u}_{\ell-1}, \ldots, \tilde{u}_2 \).

We now check
\[ \frac{\pi_U(c_i, u_{\ell+1}, \ldots, u_L, \tilde{c}_i)}{\pi_U(c_i, u_2, \ldots, u_{\ell-1}, c_i', u_{\ell+1}, \ldots, u_L)} = \theta^{-1} \frac{\mu}{\lambda + \mu} \beta_{s_j}, \]
\[ \frac{P(U(N+1) = (c_i, u_{\ell+1}, \ldots, u_L, \tilde{c}_i) \mid U(N) = (c_i, u_2, \ldots, u_{\ell-1}, c_i', u_{\ell+1}, \ldots, u_L))}{P(W(N+1) = (\tilde{u}_L, \ldots, \tilde{u}_{\ell+1}, \tilde{c}_i', \tilde{u}_{\ell-1}, \ldots, \tilde{u}_2, \tilde{c}_i) \mid W(N) = (c_i, \tilde{u}_L, \ldots, \tilde{u}_{\ell+1}, \tilde{c}_i))} = \theta^{-1} \frac{\mu}{\lambda + \mu} \beta_{s_j}. \]
A.2 Proof of Theorem 5

Proof We prove the theorem for $U$. The result for $W$ follows from the reversibility and Proposition 1.

The necessity is easy to see. By definition, $(u_1, \ldots, u_L)$ consists only of elements $c_i, \tilde{c}_i, \tilde{s}_j$. Assume to the contrary that for some $k < l$, $u_k = c_i$ and $u_l = \tilde{s}_j \in \mathcal{S}(c_i)$. If $u_\ell = \tilde{s}_j$, then either $z_{N+\ell-1} = s_j$ and $z_{N+\ell-1}$ remained unmatched and was exchanged with itself, or $z^L_{N+\ell-1} = c_i$ and it was matched and exchanged with $z_{N+\ell-1}$ for some $\ell < m \leq N$. But in either of these cases $z_{N+k-1} = c_i$ should have been matched and exchanged with $s_j$, which is a contradiction.

More subtle is the proof that any $(u_1, \ldots, u_L)$ that satisfies the conditions can occur with positive probability as $U(N)$. For such $u = (u_1, \ldots, u_L)$, we construct $\tilde{u} = (\tilde{u}_1, \ldots, \tilde{u}_L)$ and $w$ as follows:

(i) Leave all unmatched elements $u_m = c_i$ in place, i.e., let $\tilde{u}_m = c_i$, this includes $\tilde{u}_1$ since $u_1$ is the first unmatched agent.

(ii) Perform directed FCFS matching in reversed time on the elements $\tilde{c}_i$ and $\tilde{s}_j$ in $u$, and exchange their places, so that if $u_k = \tilde{c}_i$ is matched with $u_\ell = \tilde{s}_j$ (where $k > l$) we let $\tilde{u}_k = s_j$ and $\tilde{u}_\ell = c_i$.

(iii) Leave all the remaining elements $u_m = \tilde{s}_j$ (if there are any) in place and exchange them, i.e., let $\tilde{u}_m = s_j$.

(iv) Consider the remaining elements $u_{m_1} = \tilde{c}_i_1, \ldots, u_{m_K} = \tilde{c}_i_K$ (if there are any), where $m_1 < m_2 \cdots < m_K$.

(v) For each $u_{m_k} = \tilde{c}_i_k$, let $\tilde{u}_{m_k} = s_j_k$, where we choose $s_j_k \in \mathcal{S}(c_i_k)$.

(vi) Prepend $(c_i_1, \ldots, c_i_K)$ to the constructed $(\tilde{u}_1, \ldots, \tilde{u}_L)$, to obtain $w = ((c_i_1, \ldots, c_i_K), \tilde{u}_1, \ldots, \tilde{u}_L)$.

Note that all elements $\tilde{u}_\ell$, $\ell = 1, \ldots, L$, are well defined, and note that all elements of $w$ are either $c_i \in \mathcal{C}$ or $s_j \in \mathcal{S}$.

Assume that $U(N - K - L) = \emptyset$, and let $w_{N-K-L+k} = w_k$, $k = 1, \ldots, L + K$. It is important to note that there is a positive probability for this event. We now claim that $U(N) = u$. We will go through the steps of moving from $U(N - K - L)$ to $U(N)$, using directed FCFS, and matching and exchanging agent–good pairs.

We look first for matches for $c_{i_1}, \ldots, c_{i_k}$. We start with $c_{i_1}$, the first element of $w$, i.e., $c_{i_1} = z^{N-K-L+1}$. We claim that it will be matched and exchanged with $\tilde{u}_{m_1} = s_j_1$. By construction, $(c_{i_1}, s_j_1)$ are compatible, and $\tilde{u}_{m_1} = z^{N-L+m_1}$ precedes all the other $\tilde{u}_{m_1}$ in the sequence $z^n$. We need to show that if there is a $\tilde{u}_\ell = z^{N-L+\ell} = s_j$, where $\ell < m_1$, then it must be incompatible with $c_{i_1}$. Assume to the contrary that such $\tilde{u}_\ell = z^{N-L+\ell} = s_j$ with $\ell < m_1$ is compatible with $c_{i_1}$. We examine $\tilde{u}_\ell = s_j$. By construction of $\tilde{u}$, there are two possibilities: Either for some $\ell' < \ell$, $u_{\ell'} = \tilde{s}_j$ was matched and exchanged with $u_\ell = \tilde{c}_i$ in step (ii) of the construction of $\tilde{u}$, or else $u_\ell = \tilde{s}_j$ was not matched in step (ii) and so was left in place, as $u_\ell = s_j$ in step (iii) of the construction. The two possibilities are illustrated in the following line:

- $\tilde{u} = (\tilde{u}_1, \ldots, s_j, \ldots, s_j_1, \ldots, \tilde{u}_L)$, $u = (u_1, \ldots, s_j, \ldots, s_j_1, \ldots, u_L)$
- $\tilde{u} = (\tilde{u}_1, \ldots, c_i, \ldots, s_j, \ldots, s_j_1, \ldots, \tilde{u}_L)$, $u = (u_1, \ldots, \tilde{s}_j, \ldots, c_i, \ldots, \tilde{c}_i, \ldots, u_L)$
A.3 Derivation of $\pi_\gamma(C_1, \cdot, \ldots, C_k, \cdot)$ directly from Theorems 4 and 5.

We first calculate $\pi_\gamma(C_1, \cdot)$, by summing $\pi_U(u)$ over all admissible $u$ that contain only unmatched agents of type $C_1$. A typical state of this form will be

$$u = (C_1, \ldots, \tilde{c}, \tilde{c}, \ldots, \tilde{s}_j \in \mathcal{T}(C_1), \ldots, \tilde{c}, \tilde{c}, \ldots, C_1, \ldots, \tilde{c}, \tilde{c}, \ldots, \tilde{s}_j \in \mathcal{T}(C_1), \ldots, \tilde{c}, \tilde{c}, \ldots),$$

where $\tilde{c}$ can be any exchanged agents, $\tilde{s}_j$ must all be goods incompatible with $C_1$, and there are $n \geq 1$ $C_1$s, with $m \geq 0$ $s_j \in \mathcal{T}(C_1)$ following each $C_1$, and $k \geq 0$ $\tilde{c}$s following each $C_1$ or $\tilde{s}_j$. 

\[ \square \] Springer
We therefore get, by the expression (9) for $\pi_U(u)$,

$$
\pi_Y(C_1, \cdot) = B \sum_{n=1}^{\infty} \left\{ \frac{\lambda}{\lambda+\mu} C_1 \left( \sum_{k=0}^{\infty} \frac{\lambda}{\lambda+\mu} \right)^k \right\} \left[ \sum_{m=0}^{\infty} \left( \frac{\mu}{\lambda+\mu} \beta_{(C_1)} \sum_{k=0}^{\infty} \frac{\lambda}{\lambda+\mu} \right)^m \right]^n
$$

$$
= B \sum_{n=1}^{\infty} \left\{ \frac{\lambda}{\lambda+\mu} C_1 \left( \sum_{m=0}^{\infty} \frac{\mu}{\lambda+\mu} \beta_{(C_1)} \right) \right\} \left[ \sum_{m=0}^{\infty} \left( \frac{\mu}{\lambda+\mu} \right)^m \right]^n
$$

$$
= B \sum_{n=1}^{\infty} \left\{ \frac{\lambda}{\mu} \beta_{(C_1)} \right\} \left[ \sum_{m=0}^{\infty} \left( \frac{\mu}{\lambda+\mu} \right)^m \right]^n
$$

$$
= B \sum_{n=1}^{\infty} \left\{ \frac{\lambda}{\mu} \beta_{(C_1)} \right\} = B \sum_{n=1}^{\infty} \left\{ \frac{\lambda C_1}{\mu \beta_{(C_1)} - \lambda C_1} \right\}
$$

Next, we calculate $\pi_Y(C_1, \cdot, C_2, \cdot)$. We need to sum probabilities of typical states of the form

$$
u = (C_1, \ldots, C_2, \ldots, s_j \in (C_1, C_2), \ldots, c_i \in \{C_1, C_2\}, \ldots)
$$

where the first part up to the first $C_2$ is as a typical $(C_1 \cdot)$ state, followed by the first $C_2$ and then any number of $\tilde{c}$, and any number of $c_i \in \{C_1, C_2\}$ and $\tilde{s}_j \in (C_1, C_2)$, and we then get

$$
\pi_Y(C_1, \cdot, C_2, \cdot) = B \sum_{n=1}^{\infty} \left\{ \frac{\lambda}{\lambda+\mu} C_1 \left( \sum_{k=0}^{\infty} \frac{\lambda}{\lambda+\mu} \right)^k \right\} \left[ \sum_{m=0}^{\infty} \left( \frac{\mu}{\lambda+\mu} \beta_{(C_1, C_2)} \sum_{k=0}^{\infty} \frac{\lambda}{\lambda+\mu} \right)^m \right]^n
$$

$$
\times \frac{\lambda}{\lambda+\mu} C_1 \left( \sum_{m=0}^{\infty} \left( \frac{\mu}{\lambda+\mu} \beta_{(C_1, C_2)} \right) \right) \left[ \sum_{m=0}^{\infty} \left( \frac{\mu}{\lambda+\mu} \right)^m \right]^n
$$

$$
= B \frac{\lambda C_1}{\mu \beta_{(C_1, C_2)} - \lambda C_1} \times \frac{\lambda C_1}{\mu \beta_{(C_1, C_2)} - \lambda C_1} \times \sum_{n=0}^{\infty} \left\{ \frac{\lambda C_1}{\mu \beta_{(C_1, C_2)}} \right\}^n
$$

$$
= B \frac{\lambda C_1}{\mu \beta_{(C_1, C_2)} - \lambda C_1} \times \frac{\lambda C_1}{\mu \beta_{(C_1, C_2)} - \lambda C_1} \times \sum_{n=0}^{\infty} \left\{ \frac{\lambda C_1}{\mu \beta_{(C_1, C_2)}} \right\}^n
$$

Finally, we proceed by induction, to obtain, using the same steps in the derivation

$$
\pi_Y(C_1, \cdot, \ldots, C_L, \cdot) = B \prod_{\ell=1}^{L-1} \frac{\lambda C_1}{\mu \beta_{((C_1, \ldots, C_L) \setminus \{C_{\ell+1}\})} - \lambda C_{\ell+1}} \times \frac{\lambda}{\lambda+\mu} C_\ell \left( \sum_{k=0}^{\infty} \frac{\lambda}{\lambda+\mu} \right)^k \left[ \sum_{m=0}^{\infty} \left( \frac{\mu}{\lambda+\mu} \beta_{((C_1, \ldots, C_L) \setminus \{C_{\ell+1}\})} \sum_{k=0}^{\infty} \frac{\lambda}{\lambda+\mu} \right)^m \right]^n
$$

$$
\times \sum_{n=0}^{\infty} \left\{ \frac{\lambda}{\mu \beta_{((C_1, \ldots, C_L) \setminus \{C_{\ell+1}\})} \sum_{k=0}^{\infty} \frac{\lambda}{\lambda+\mu} \right\}^n
$$

$$
= B \prod_{\ell=1}^{L} \frac{\lambda C_\ell}{\mu \beta_{((C_1, \ldots, C_L) \setminus \{C_{\ell+1}\})} - \lambda C_{\ell+1}}.
$$
References

1. Adan, I., Bušić, A., Mairesse, J., Weiss, G.: Reversibility and further properties of FCFS infinite bipartite matching. Math. Oper. Res. 43(2), 598–621 (2017)
2. Adan, I., Kleiner, I., Righter, R., Weiss, G.: FCFS parallel service systems and matching models. Perform. Eval. 127, 253–272 (2018)
3. Adan, I., Weiss, G.: Exact FCFS matching rates for two infinite multitype sequences. Oper. Res. 60(2), 475–489 (2012)
4. Adan, I., Weiss, G.: A skill based parallel service system under FCFS-ALIS - steady state, overloads, and abandonments. Stochastic Syst. 4(1), 250–299 (2014)
5. Adan, I., Boon, M., Weiss, G.: Design heuristic for parallel many server systems. Eur. J. Oper. Res. 273(1), 259–277 (2019)
6. Afèche, P., Diamant, A., Milner, J.: Double-sided batch queues with abandonment: modeling crossing networks. Oper. Res. 62(5), 1179–1201 (2014)
7. Ayesta, U., Bodas, T., Verloop, I.M.: On a unifying product form framework for redundancy models. Perform. Eval. 127, 93–119 (2018)
8. Caldentey, R., Kaplan, E.H., Weiss, G.: FCFS infinite bipartite matching of servers and customers. Adv. Appl. Probab. 41(03), 695–730 (2009)
9. Castro, F., Nazerzadeh, H., Yan, C.: Matching queues with reneging: a product form solution. ArXiv preprint ArXiv:2005.10728 (2020)
10. Dobbie, J.M.: Letter to the editor: a doubled-ended queuing problem of Kendall. Oper. Res. 9(5), 755–757 (1961)
11. Gardner, K., Righter, R.: Product forms for FCFS queueing models with arbitrary server-job compatibilities: an overview. Queue. Syst. 96(1), 3–51 (2020)
12. Gardner, K., Zbarsky, S., Doroudi, S., Harchol-Balter, M., Hyytiä, E., Scheller-Wolf, A.: Queueing with redundant requests: exact analysis. Queue. Syst. 83(3–4), 227–259 (2016)
13. Hendershott, T., Mendelson, H.: Crossing networks and dealer markets: competition and performance. J. Finance 55(5), 2071–2115 (2000)
14. Karasemen, I.Z., Scheller-Wolf, A., Deniz, B.: Managing perishable and aging inventories: review and future research directions. In: Planning Production and Inventories in the Extended Enterprise, pp. 393–436. Springer, New York (2011)
15. Lu, Y., Xiang, S., Wu, W.: Taxi queue, passenger queue or no queue? In: Proceedings of 18th International Conference on Extending Database Technology (EDBT). Brussels, Belgium, pp. 593–604 (2015)
16. Mairesse, J., Moyal, P.: Stability of the stochastic matching model. J. Appl. Probab. 53(4), 1064–1077 (2017)
17. Maglaras, C., Moallemi, C.C., Zheng, H.: Queueing dynamics and state space collapse in fragmented limit order books. Columbia Bus. School Res. Pap., (1–60) (2014)
18. Moreira-Matias, L., Garcia, J., Ardei, M., Mendes-Moreira, J., Damas, L.: Predicting taxi-passenger demand using streaming data. IEEE Trans. Intell. Transp. Syst. 14(3), 1393–1402 (2013)
19. Moyal, P., Busic, A., Mairesse, J.: A product form for the general stochastic matching model. ArXiv preprint ArXiv:1711.02620 (2018)
20. Moyal, P., Perry, O.: On the instability of matching queues. Ann. Appl. Probab. 27(6), 3385–3434 (2017)
21. Perry, D., Stadje, W.: Perishable inventory systems with impatient demands. Math. Methods Oper. Res. 50(1), 77–90 (1999)
22. Sutherland, W., Jarrahi, M.H.: The sharing economy and digital platforms: a review and research agenda. Int. J. Inf. Manag. 43, 328–341 (2018)
23. Visschers, J., Adan, I., Weiss, G.: A product form solution to a system with multi-type jobs and multi-type servers. Queue. Syst. 70(3), 269–298 (2012)
24. Wolff, R.W.: Poisson arrivals see time averages. Oper. Res. 30(2), 223–231 (1982)
25. Zenios, S.A.: Modeling the transplant waiting list: a queueing model with reneging. Queue. Syst. 31(3–4), 239–251 (1999)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.