Stabilizing Policies for Probabilistic Matching Systems

Burak Büke and Hanyi Chen

School of Mathematics, The University of Edinburgh
King's Buildings, Edinburgh, EH9 3JZ, UK
E-mail:B.Buke@ed.ac.uk, H.Chen-29@sms.ed.ac.uk

Abstract

In this work, we introduce a novel queueing model with two classes of users, where users wait in the system to match with a candidate from the other class, instead of accessing a resource. This new model is useful for analyzing the traffic in web portals that match people who provide a service with people who demand the service, e.g. employment portals, matrimonial and dating sites and rental portals. We provide a Markov chain model for these systems and derive the probability distribution of the number of matches up to some finite time given the number of arrivals. We prove that if no control mechanism is employed these systems are unstable for any set of parameters, and suggest four different classes of control policies to assure stability. Contrary to the intuition that the rejection rate should decrease as the users get more likely to match, we show that for certain control policies the rejection rate is insensitive to the matching probability. Even more surprisingly, we show that for reasonable policies the rejection rate may be an increasing function of the matching probability. We also prove insensitivity results related to the average queue lengths and waiting times.

1 Introduction

Due to the advances in technology in the last decade, it has become widely adopted by society to use the internet for business and personal interactions. Web portals that serve as a meeting point for people who provide a specific service/product and people who demand the service/product are becoming increasingly popular. Employment and rental portals, dating and matrimonial sites are some examples of such systems. In this work we introduce a new queueing model, which we call a *probabilistic matching system*, to model the traffic in these web portals.

To understand the dynamics of user behavior in probabilistic matching systems, consider an employment portal as an example. There are two classes of users, employers and employees, in an employment system. Employers arrive at the system at random times to hire an employee. Upon arrival, they first check the resumés of the employees in the portal. If they can find suitable candidates, they hire one of the candidates and close the position. If there are no suitable candidates, they create a job posting on the website and wait for a suitable candidate to apply for the position. Similarly, when a potential employee arrives at the system, she applies for the existing postings and leaves the system if she gets hired. If she cannot find a suitable job, she posts her resumé on the website and waits until a suitable job becomes available. As an important feature of this system, each given employer-employee pair matches with some given probability. Hence, the operators of a probabilistic matching system do not have any control of who matches with whom.

There are several different types of matching systems studied in the literature. Adan and Weiss [1] and Caldentey et al. [4] consider a matching system where customers are matched with servers. Several different types of customers and servers arrive at the system according to a stochastic process, and each customer type can be served by a subset of server types and each server type can serve a subset of customer types. Adan and Weiss [1] study the stability of the system under the first-come-first-serve policy and derive explicit product-form equations for the matching rate for some specific configurations. Bušić et al. [3] develop necessary and sufficient conditions for the stability of different matching policies for systems where customers and servers form a bipartite graph and study the computational complexity of deciding whether a given policy is stable. In a recent and different line of work, Gurvich and Ward [9] consider a queueing system with multiple types of jobs that can match with each other. Their goal is to minimize the holding cost by developing a dynamic policy to decide which types should match at any given time.

The concept of "type" plays an important role in the work presented above, and differs from the "class" concept used in this work. In [1], [3] and [4], customers and servers can be viewed as two classes of users of the matching system. These classes are further divided into types according to their special properties, e.g. there are several different types of customers and servers. A special subclass of these systems is the double-ended queue (see e.g. Kashyap [11] and Liu et al. [14]), where there is exactly one type of user from each class. In the aforementioned models, once the types of two users (a server and a customer) are known, they match or fail to match deterministically. In many real world systems, this assumption fails to hold and special attention is needed on an individual basis. For example, when a company is hiring for a position, it does not just hire any person who has the desired background, but rather the personal qualifications of the candidates play an important role in addition to their competences in the field. Hence, each candidate should be considered individually, rather than being classified into a specific type. In this work, we incorporate this individuality by assuming that matchings occur probabilistically. We consider probabilistic matching systems with two classes (e.g. customers-servers, or employers-employees). Upon arrival each user scans the queue of the other class and may match with each individual in the queue independently with probability q. If there are more than one matching users in the other queue, one is chosen uniformly at random. Similar to the aforementioned work, we also assume that the matching procedure does not take time and happen instantaneously.

The matching probability q is a key factor in the analysis of probabilistic matching systems. If the matching probability $q = 1$ and there are two users from different classes have arrived, they match and leave the system immediately and the users from different classes cannot co-exist in the system. Hence, the system can be modeled as a one-dimensional continuous time random walk. However, when the matching probability is less than 1, we need a two-dimensional process, as users from different classes can be present in the system at the same time.

We start our analysis by analyzing the transient behavior of the matching process, i.e., the counting process for the number of matched pairs up to a finite time t. We first ask the following question: "What is the probability that exactly k matchings have occured, if we know that exactly m class-1 and n class-2 users have arrived?" To the best of our knowledge, this basic probability distribution has not been studied in the literature and in Section 3 we provide the explicit equation for the mass function of this distribution. Unfortunately, this equation is fairly complicated, which indicates the difficulty in completely characterizing the counting process for the number of matched pairs.

Next, we study the stability of the probabilistic matching systems by defining a system to be stable when it is ergodic. The earlier line of work on "assembly-like queues" [10] or "queues with paired customers" [13] exhibits similarities to probabilistic matching systems in regards to stability. These systems operate similarly to the probabilistic matching systems where pairs from different classes match with probability $q = 1$. However, in these systems the matching procedure (or assembly) takes some time and requires a resource, whereas the matchings are assumed to occur instantaneously in a probabilistic matching system. Harrison [10] studied the waiting time processes for the assembly-like queues and showed that these systems are not stable, regardless of the balance between input and service rates. When the matching probability $q = 1$, a probabilistic matching system is modeled as a one-dimensional random walk on integers, which is known to be unstable. More specifically, it is null recurrent or transient depending on whether the arrival rates are equal or not, respectively. Using a coupling argument, we show that this implies instability of the matching systems with $q < 1$. We also show that when arrival rates are equal, a probabilistic matching system is null recurrent even when $q < 1$.

To stabilize probabilistic matching systems, we suggest four different classes of admission control policies: (i) the simple threshold policy, (ii) accept-the-shortest-queue policy, (iii) functional threshold policy and (iv) the one-sided threshold policy. The simple threshold policy puts constant bounds on the number of users that can be present in the system from each class. As the resulting state space is finite, the simple threshold policy stabilizes the matching systems with $q = 1$. However, if the matching probability is less than 1, this policy yields absorbing states which indicates that some users experience an infinite waiting time. To avoid absorbing states, the accept-the-shortest-queue and functional threshold policies use a "moving" threshold and try to "balance" the number of users from each class. The accept-the-shortest-queue admits users if they belong to a class with the minimum number of users in the system and functional threshold policy admits users if the number of users from that class is less than a function of the number of users from the other class. With very mild conditions on the threshold function to prevent absorbing states, we show that both policies stabilize the system for any set of arrival rates and any positive matching probability. This result is closely related to the work of Latouche [13] on queues with paired customers. Latouche [13] studies the stability of these systems with state-dependent arrival rates, and characterizes the stationary distributions using matrix analytic methods when the system is stable. He defines the "excess" as the difference between the numbers of the two classes of users, and concludes if the state-dependent arrival process keeps the excess bounded then the queues with paired customers can be stable. The stability of the accept-the-shortest-queue policy relies on the same idea and guarantees an excess with absolute value always less than one. However, the functional threshold policy allows the excess to be an increasing function of number of users in the system, and suggests that for the probabilistic matching systems the stability can be achieved by keeping the excess only "under control", instead of keeping it strictly bounded. Our last policy, one-sided threshold policy, relies on the assumption that one of the classes has a higher arrival rate than the other, and only rejects users from the class with higher arrival rate if they exceed a certain constant threshold. This policy also stabilizes probabilistic matching systems with any set of arrival rates that satisfy the above assumption, any positive matching probability and any non-negative threshold value.

Our results for the performance measures of probabilistic matching systems under these policies are even more intriguing. One may think that, as the matching probability increases, the users match more easily and hence, the departure rate of matched pairs should increase, or equivalently the long run rejection rate should decrease. However, contrary to this initial intuition, we prove that under the accept-the-shortest-queue policy and a subclass of functional threshold policies, the long run rejection rate does not depend on the matching probability. Even more surprisingly, we observe that under most reasonable functional threshold policies *the long run rejection rate actually increases as the matching probability increases*! More specifically, when the threshold function is chosen so that the operators of the system become more eager to admit users of a class as users of the other class accumulate in the system, then the long run percentage of rejected users is an increasing function of the matching probability. In Section 5 we intuitively explain this phenomenon based on the geometry of the state space. This explanation is closely related to the "excess" (as defined by Latouche [13]) and clarifies why the long run rejection rate is independent of the matching probability under the accept-the-shortest-queue policy. We further show that the behavior of the difference between average queue lengths of the two classes is closely related to the behavior of the long run rejection rate. We prove that for the cases which the rejection rate is independent of the matching probability, the difference of queue lengths also does not depend on the matching probability. However, we observe that if the rejection rate is an increasing function of the matching probability, then the difference between average queue lengths is a decreasing function of the matching probability and vice versa. We finally show that under the one-sided threshold policy, the long run rejection rate is independent of both the matching probability and the threshold value, the explanation of which is however completely different from the above and relies on the well-known arrival-departure theorem.

The paper is organized as follows. We introduce the mathematical model for probabilistic matching systems in Section 2 and we analyze the transient behavior of the system by concentrating on the process counting the number of matches up to time t in Section 3. Section 4 discusses the stability of probabilistic matching systems and introduces four stabilizing admission control policies. We present simulation results analyzing the long run rejection rates, average queue lengths and average waiting times under the suggested policies in Section 5. Finally, we present open problems and future research directions in Section 6.

2 Mathematical Model

In this section we present a continuous-time Markov chain (CTMC) model for the probabilistic matching systems. The basic assumptions of our model are as follows:

1. The arrival processes of class-1 and class-2 users are independent Poisson processes with rates $0 < \lambda_1 < \infty$ and $0 < \lambda_2 < \infty$, respectively.

2. Each class-1 and class-2 user pair matches with probability q ($0 < q \leq 1$), independent of other users.

3. When a class-1 user arrives, she checks whether there are any class-2 users in the queue matching with her. If there are matching class-2 users, she chooses one of them uniformly at random and then they leave the system together. If there is no matching class-2 user in the system, she joins the queue. A similar mechanism applies when a class-2 user arrives.

4. Once a suitable match is found, the matched pair leaves the system immediately, i.e., the matching procedure does not take any time.

5. Users do not abandon the system without being matched.

Under these assumptions, the system can be modelled as a two-dimensional CTMC $\{X^q(t) = (X_1^q(t), X_2^q(t)), t \geq 0\}$, where $X_i^q(t)$ is the number of class-i users, $i = 1, 2$, in a system with matching probability q at time t. The state space is $\mathbb{S} = \mathbb{N}^2$, where $\mathbb{N}$ is the set of non-negative integers. Since the arrival processes are Poisson processes and the matchings of each pair of class-1 and class-2 users are independent, we have the Markov property.

The probability that a given class-1 and class-2 user pair does not match is $1 - q$. Hence, due to independence of matchings, a class-1 user finding j class-2 users waiting in the system upon arrival does not match with anyone and joins the queue with probability $(1 - q)^j$ or she leaves the system with a matching class-2 user with probability $1 - (1 - q)^j$. For notational simplicity, we define $r = 1 - q$ as the probability of not matching for each pair and state the generator matrix as follows:

$$Q_{(n_1,n_2)(n_1',n_2')} = \begin{cases} \lambda_1 r^{n_2} & \text{if } n_1' = n_1 + 1 \text{ and } n_2' = n_2, \\ \lambda_2 r^{n_1} & \text{if } n_1' = n_1 \text{ and } n_2' = n_2 + 1, \\ \lambda_1(1 - r^{n_2}) & \text{if } n_1' = n_1 \text{ and } n_2' = n_2 - 1 \geq 0, \\ \lambda_2(1 - r^{n_1}) & \text{if } n_1' = n_1 - 1 \geq 0 \text{ and } n_2' = n_2, \\ -(\lambda_1 + \lambda_2) & \text{if } n_1' = n_1 \text{ and } n_2' = n_2, \\ 0 & \text{otherwise.} \end{cases}$$

As the generator matrix suggests, it is convenient to use the probability of not matching, r, in the equations. However, the probability of matching q is a more intuitive quantity to refer to in the natural language. Hence, we use q and r notation together in the remainder of the paper, assuming that the relation $q + r = 1$ is obvious.

When $q = 1$, class-1 users and class-2 users cannot co-exist in the system at the same time. Hence, the system can be modeled as a one-dimensional CTMC, where $X^1(t) = k \geq 0$ when there are k class-1 users and $X^1(t) = -k \leq 0$ when there are k class-2 users in the system at time t. This CTMC is a continuous time random walk on integers with rates $Q_{n,n+1} = \lambda_1$ and $Q_{n,n-1} = \lambda_2$.

We can also represent the process $\{X_i^q(t), t \geq 0\}, i = 1, 2$, as the difference of two counting processes. Let $A_i(t)$ and $M^q(t)$ be the number of arrivals from class-i and the number of matched pairs up to time t, respectively. Then, $X_i^q(t) = A_i(t) - M^q(t), i = 1, 2$, for all $t \geq 0$. This representation is useful when the Markov chain techniques are difficult to be applied directly.

3 Transient Behavior of the System

We first study the transient behavior of the system to gain probabilistic insight about the situation at time t. We focus on the counting process $\{M^q(t), t \geq 0\}$, and provide an explicit equation to the probability of observing exactly k matched pairs up to time t.

When $q = 1$, the probability of having k matched pairs up to time t is trivial. As every class-1 (class-2) user matches upon arrival if there are any class-2 (class-1) users in the system, $M^1(t) = \min\{A_1(t), A_2(t)\}$. Hence, $\mathbb{P}[M^1(t) = 0] = e^{-\lambda_1 t} + e^{-\lambda_2 t} - e^{-(\lambda_1+\lambda_2)t}$, and for $k > 0$,

$$\begin{aligned} \mathbb{P}[M^1(t) = k] = {} & \frac{e^{-\lambda_1 t}(\lambda_1 t)^k}{k!} \left(1 - \sum_{i=0}^{k-1} \frac{e^{-\lambda_2 t}(\lambda_2 t)^i}{i!}\right) \\ & + \frac{e^{-\lambda_2 t}(\lambda_2 t)^k}{k!} \left(1 - \sum_{i=0}^{k} \frac{e^{-\lambda_1 t}(\lambda_1 t)^i}{i!}\right). \end{aligned}$$

However, when $0 < q < 1$, the problem is significantly more complicated.

We now define

$$P_{k,m,n}^q \equiv \mathbb{P}[M^q(t) = k | A_1(t) = m, A_2(t) = n], \tag{1}$$

for $0 < q \leq 1$, then using the law of total probability, we write

$$\mathbb{P}[M^q(t) = k] = e^{-(\lambda_1+\lambda_2)t} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} P_{k,m,n}^q \frac{\lambda_1^m \lambda_2^n t^{m+n}}{m!\, n!}. \tag{2}$$

The quantity $P_{k,m,n}^q$ is of interest on its own right. For example, one may be interested in the probability of exactly k people getting hired, when there are m companies hiring and n employees looking for jobs. To find an explicit equation for $P_{k,m,n}^q$ we need to solve a three dimensional recursion. Unfortunately, generating function methods are not easy to use to solve this recursion as the coefficients involve powers. Hence, we resort to more direct methods for solving the recusion. Theorem 1 presents an explicit formula for the desired conditional probability.

Theorem 1. *Suppose $P_{k,m,n}^q$ is as defined in (1) and without loss of generality assume $m \leq n$. Then,*

(i) when $q = 1$, $P_{k,m,n}^1 = 1$ if $k = m$ and 0 otherwise.

(ii) when $0 < q < 1$,

$$P_{k,m,n}^q = \begin{cases} r^{mn}, & k = 0, \\ a_{k,m} r^{(m-k)n} \prod_{i=0}^{k-1} (1 - r^{n-i}), & 1 \leq k \leq m, \\ 0, & \textit{otherwise}, \end{cases} \tag{3}$$

where

$$\begin{aligned} a_{k,m} = & \sum_{l>0} \left(\sum_{d_1+d_2+\ldots+d_l=k} r^k \frac{(-1)^l}{\prod_{i=1}^{l} \prod_{j=1}^{d_i} (1 - r^j)} \right) + r^{-mk} \frac{r^{k^2}}{\prod_{j=1}^{k} (1 - r^j)} \\ & + \sum_{i=1}^{k-1} \frac{r^{-mi} r^{ki}}{\prod_{j=1}^{i} (1 - r^j)} \sum_{l>0} \left(\sum_{d_1+\ldots+d_l=k-i} r^{k-i} \frac{(-1)^l}{\prod_{w=1}^{l} \prod_{j=1}^{d_w} (1 - r^j)} \right), \end{aligned}$$

with indexes $d_1, d_2, \ldots$ belonging to $\mathbb{N}_+ = \{1, 2, 3, \ldots\}$.

Proof. See Appendix 8.1 □

Theorem 1 reveals that the elementary conditional probability of k matches succeed given that m class-1 users and n class-2 users have arrived has a fairly complicated expression. This implies that even the one dimensional distribution for the matching process is far from trivial, which further indicates the difficulty in the calculation of finite dimensional distributions to completely characterize $\{M^q(t), t \geq 0\}$ and the transient behavior of probabilistic matching systems. Next, we concentrate on the steady-state behavior of probabilistic matching systems.

4 Stability Analysis and Stabilizing Policies

In this paper, a system is defined to be stable if it is ergodic, i.e., positive recurrent, and unstable if it is either null recurrent or transient. In this section, we first show that an uncontrolled probabilistic matching system is unstable for any set of arrival rates, λ_1 and λ_2, and matching probability q. Then, we suggest admission control policies to stabilize the system and study the performance of these policies.

4.1 Instability of the Uncontrolled System

To study the stability of the uncontrolled system, we first prove that it is unstable when the matching probability $q = 1$. Then, using a coupling argument, we show that this also implies the instability of the systems where $0 < q < 1$.

Theorem 2. *An uncontrolled probabilistic matching system is unstable for any set of arrival rates λ_1 and λ_2, and matching probability $0 < q \leq 1$. More specifically, it is null recurrent if $\lambda_1 = \lambda_2$ and transient if $\lambda_1 \neq \lambda_2$.*

Proof. When $q = 1$, the system can be modelled as a one-dimensional continuous time random walk which is known to be null recurrent if $\lambda_1 = \lambda_2$ and transient if $\lambda_1 \neq \lambda_2$. To prove the case where $0 < q < 1$, we use a coupling argument. Let $0 < \theta_1 < \theta_2 < \cdots$ be the sequence of occurrence times of events following a homogeneous Poisson process with rate $\lambda_1 + \lambda_2$, and, to determine the class of arrivals, define a sequence of independent discrete random variables $\{\tau_n, n \in \mathbb{N}\}$ which takes the values 1 or 2 with probabilities $\frac{\lambda_1}{\lambda_1+\lambda_2}$ and $\frac{\lambda_2}{\lambda_1+\lambda_2}$, respectively. Let $\{U_n, n \in \mathbb{N}\}$ be a sequence of independent uniform(0,1) random variables, and $I(A)$ be the indicator function for event A which takes value 1 if A occurs and 0 otherwise. Finally define $\tilde{X}^q_{0,i} = 0, i = 1, 2$, and for $n \in \mathbb{N}$,

$$
\begin{aligned}
\tilde{A}_{n,i} &= \sum_{j=1}^{n} I(\{\tau_j = i\}), i = 1, 2, \\
\tilde{M}^q_n &= \sum_{j=1}^{n} \left(I(\{\tau_j = 1\}) I(\{U_j > r^{\tilde{X}^q_{j-1,2}}\}) + I(\{\tau_j = 2\}) I(\{U_j > r^{\tilde{X}^q_{j-1,1}}\}) \right), \\
\tilde{X}^q_{n,i} &= \tilde{A}_{n,i} - \tilde{M}^q_n, \\
\tilde{X}^q_i(t) &= \tilde{X}^q_{n,i}, \forall t \in [\theta_n, \theta_{n+1}) \text{ and } i = 1, 2, \\
\tilde{X}^q(t) &= (\tilde{X}^q_1(t), \tilde{X}^q_2(t)), \forall t \in [\theta_n, \theta_{n+1}).
\end{aligned}
$$

The process $\{\tilde{X}^q(t), t \geq 0\}$ is stochastically equivalent to $\{X^q(t), t \geq 0\}$. If the matching probability is one, i.e., $r = 1 - q = 0$ and $\tilde{X}^1_{n,1} = 0$ ($\tilde{X}^1_{n,2} = 0$) then arriving class-2 (class-1) user cannot match upon arrival. Hence, we take the indeterminate form 0^0 to be 1 for notational convenience.

Let $r_1 > r_2$ (or equivalently $q_1 < q_2$) and define

$$
n^* = \min \left\{ n : \tilde{X}^{q_1}_{n,1} < \tilde{X}^{q_2}_{n,1} \text{ or } \tilde{X}^{q_1}_{n,2} < \tilde{X}^{q_2}_{n,2} \right\}.
$$

By definition $n^* > 0$, and without loss of generality, we can assume $\tilde{X}^{q_1}_{n^*,1} < \tilde{X}^{q_2}_{n^*,1}$, which implies $\tilde{X}^{q_1}_{n^*-1,1} = \tilde{X}^{q_2}_{n^*-1,1}$. The number of class-$i$ arrivals is independent of the matching probability

q and the number of matched individuals (departures) is the same for $\tilde{X}^q_{n,1}$ and $\tilde{X}^q_{n,2}$. Hence, $\tilde{X}^{q_1}_{n^*-1,1} = \tilde{X}^{q_2}_{n^*-1,1}$ implies $\tilde{X}^{q_1}_{n^*-1,2} = \tilde{X}^{q_2}_{n^*-1,2}$. If $\tau_{n^*} = 1$, then $U_{n^*} > r_1^{\tilde{X}^{q_1}_{n^*-1,2}}$ and $U_{n^*} \leq r_2^{\tilde{X}^{q_2}_{n^*-1,2}}$, and if $\tau_{n^*} = 2$, then $U_{n^*} > r_1^{\tilde{X}^{q_1}_{n^*-1,1}}$ and $U_{n^*} \leq r_2^{\tilde{X}^{q_2}_{n^*-1,1}}$ which is not possible in either cases and leads to a contradiction. Hence, we conclude that such an n^* does not exist and $\tilde{X}^{q_1}_i(t) \geq \tilde{X}^{q_2}_i(t), i = 1, 2$, holds for every $t \geq 0$. This implies that for any $q < 1$, if $\{\tilde{X}^1_i(t), t \geq 0\}$ is transient, then $\{\tilde{X}^q_i(t), t \geq 0\}$ is also transient, and if $\{\tilde{X}^1_i(t), t \geq 0\}$ is null recurrent, then $\{\tilde{X}^q_i(t), t \geq 0\}$ is either null recurrent or transient.

Now we show that $\{X^q(t), t \geq 0\}$ is recurrent when $\lambda_1 = \lambda_2$ and $0 < q < 1$. Let $X^q_n = (X^q_{n,1}, X^q_{n,2})$ denote the corresponding embedded DTMC. Using Theorem 2.2.1 from Fayolle et al. [7] pg. 26, if we can find a finite set H and a positive function $f(i,j)$ such that for $(i,j) \notin H$,

$$\mathbb{E}[f(X^q_{n+1,1}, X^q_{n+1,2})|(X^q_{n,1}, X^q_{n,2}) = (i,j)] - f(i,j) \leq 0$$

and $f(i,j) \to \infty$ as $i + j \to \infty$, the result follows. The transition probabilities for the embedded DTMC are

$$p_{(i,j)(k,l)} = \begin{cases} r^j/2 & \text{if } k = i+1 \text{ and } l = j, \\ r^i/2 & \text{if } k = i \text{ and } l = j+1, \\ (1-r^j)/2 & \text{if } k = i \text{ and } l = j-1 \geq 0, \\ (1-r^i)/2 & \text{if } k = i-1 \geq 0 \text{ and } l = j, \\ 0 & \text{otherwise.} \end{cases}$$

We choose $f(i,j)$ to be

$$f(i,j) = \begin{cases} 1 & \text{if } i = j = 0, \\ i + j - 2 & \text{if } i > 0, j > 0 \text{ and } i + j > 2, \\ i + j & \text{otherwise.} \end{cases}$$

Clearly, $f(i,j)$ is positive and $f(i,j) \to \infty$ as $i + j \to \infty$. Let $H_1 = \{(i,j) : 0 \leq i, j \leq 2\}$ and for any $(i,j) \notin H_1$ we have

$$\mathbb{E}[f(X^q_{n+1,1}, X^q_{n+1,2})|(X^q_{n,1}, X^q_{n,2}) = (i,j)] - f(i,j)$$
$$= \begin{cases} 0 & \text{if } i = 0, j > 2 \text{ or } i > 2, j = 0, \\ -r^j & \text{if } i = 1, j > 2, \\ -r^i & \text{if } i > 2, j = 1, \\ r^i + r^j - 1 & \text{if } i > 2, j > 2. \end{cases}$$

Let $H_2 = \{(i,j) : 0 \leq i, j \leq \frac{\ln(1-r^2)}{\ln r}\}$, then for any $(i,j) \notin H = H_1 \cup H_2$, we have $\mathbb{E}[f(X^q_{n+1}) - f(X^q_n)|X^q_n = (i,j)] \leq 0$ and the result follows. □

Theorem 2 indicates that for any set of parameters, users of an uncontrolled probabilistic matching system experiences arbitrarily long waiting times. In the next section, we suggest admission control policies to stabilize probabilistic matching systems and analyze some performance measures for those policies.

4.2 Stabilizing Policies

In this section we analyze four different admission control policies to stabilize probabilistic matching systems. The policies specify the conditions under which an arriving user is allowed to enter the system and we call an admission control policy a *stabilizing policy* if it makes the probabilistic matching system ergodic. The first policy we study is the *simple threshold* policy which relies on limiting the number of users from each class in the system. The simple threshold policy stabilizes probabilistic matching systems with matching probability $q = 1$, whereas fails to provide a finite expected waiting time for users for systems with $0 < q < 1$. Analyzing the reasons behind this failure, we then suggest the *accept-the-shortest-queue* (ASQ) policy which relies on balancing the system by accepting only from the class with minimum number of users in the system. ASQ policy manages to stabilize systems with any set of arrival rates and matching probability, but may lead to very poor performance measures. Hence, we introduce the functional threshold (FT) policy by relaxing the definition of "balancing". The *one-sided threshold* (OST) policy is employed if a class of users has a higher arrival rate than the other, and imposes an upper bound on the number of users belonging to the class with higher arrival rate while always accepting the other users.

In addition to proving the stability of probabilistic matching systems under the suggested policies, we analyse how these policies affect some key performance measures. As the admission control policies are based on rejecting users under specific conditions, the long run percentage of rejected users is an important measure to assess the performance of a specific policy. A basic result in queueing theory states that the throughput of a stable system, which is defined to be the long run average rate at which users leave the system, is equal to the long run average rate at which users are admitted to the system. This implies that there is a one-to-one relationship between long run percentage of rejected users and throughput. For completeness, we prove this result for probabilistic matching systems in Theorem 3. We refer the reader to Asmussen [2] and El-Taha and Stidham [6] for more results of similar flavour.

Theorem 3. *Consider a probabilistic matching system under a stabilizing policy and let $A_i^e(t)$ and $M(t)$ to be the processes counting the number of type i users admitted to the system and the number of matched pairs up to time t respectively. Then,*

$$\lim_{t\to\infty} \frac{A_1^e(t)}{t} = \lim_{t\to\infty} \frac{A_2^e(t)}{t} = \lim_{t\to\infty} \frac{M(t)}{t}.$$

Proof. Without loss of generality, we assume that $X(0) = (0, 0)$. Define $\theta_0 = 0$,

$$\theta_j = \inf\{t \geq \theta_{j-1} : X(t) = (0,0) \text{ and } \exists s \text{ where } \theta_{j-1} < s < t, X(s) \neq (0,0)\}$$

for $j \in \mathbb{Z}_{>0}$ and $J(t) = \max\{j : \theta_j \leq t\}$. Then, for $i = 1, 2$

$$\begin{aligned}
\frac{A_i^e(t) - M(t)}{t} &= \sum_{j=1}^{J(t)} \frac{A_i^e(\theta_j) - A_i^e(\theta_{j-1}) - M(\theta_j) + M(\theta_{j-1})}{t} \\
&\quad + \frac{A_i^e(t) - A_i^e(\theta_{J(t)}) - M(t) + M(\theta_{J(t)})}{t} \\
&= \frac{A_i^e(t) - A_i^e(\theta_{J(t)}) - M(t) + M(\theta_{J(t)})}{t} \\
&\leq \frac{A_i(\theta_{J(t)+1}) - A_i(\theta_{J(t)})}{t} \\
&= \frac{A_i(\theta_{J(t)+1}) - A_i(\theta_{J(t)})}{J(t)+1} \frac{J(t)+1}{t}.
\end{aligned}$$

The first equality above uses the fact that $A_i^e(\theta_{J(t)}) = \sum_{j=1}^{J(t)} A_i^e(\theta_j) - A_i^e(\theta_{j-1})$ and $M(\theta_{J(t)}) = \sum_{j=1}^{J(t)} M(\theta_j) + M(\theta_{j-1})$. The second equality follows as $A_i^e(\theta_j) = M(\theta_j)$ for all j by the definition of θ_j. Finally the inequality in the third step follows by using $M(\theta_{J(t)})-M(t) < 0$ and then realizing that $A_i^e(t) - A_i^e(\theta_{J(t)})$ is the number of accepted users in $(\theta_{J(t)}, t]$, whereas $A_i(\theta_{J(t)+1}) - A_i(\theta_{J(t)})$ is the total number of accepted and rejected users in the same time window. As $t \to \infty$, the second term on the righthand side converges to a finite number by ergodicity and the elementary renewal theorem. Now, let $A_{i,j} = A_i(\theta_{j+1}) - A_i(\theta_j)$ and $\mathbb{E}[A_{i,j}] = \lambda_i \mathbb{E}[\theta_{j+1} - \theta_j] < \infty$ (see e.g., Corollary V.6.7 in Çınlar [5]). Then for any $\epsilon > 0$, $\limsup_{t\to\infty} \frac{A_i^e(t)-M(t)}{t} \le \epsilon$ if and only if $\mathbb{P}[A_{i,j} > \epsilon j \text{ infinitly often}] = 0$ which follows from Borel-Cantelli lemma. □

Theorem 3 implies that if c_i is the long run proportion of rejected users from class-i, then

$$\lim_{t\to\infty} \frac{A_1^e(t)}{t} = (1-c_1)\lambda_1 = (1-c_2)\lambda_2 = \lim_{t\to\infty} \frac{A_2^e(t)}{t} \tag{4}$$

which is also equal to the throughput of the system. Moreover, using Poisson-arrivals-see-time-averages (PASTA) property (see El-Taha and Stidham [6]), c_i also equals to the total stationary probability of being at a state where class-i users are rejected.

4.2.1 The Simple Threshold Policy

The simple threshold policy imposes a constant upper bound on the number of users in the system for each class. The aim is to reduce the process $\{X^q(t), t \ge 0\}$ to a finite state space irreducible CTMC which would be always ergodic.

Definition 4. A simple threshold policy is an admission control policy which admits a class-i user arriving at time t if and only if $X_i^q(t-) \le N_i$, where $i = 1, 2$ and $0 \le N_i < \infty$.

Theorem 5. *When $q = 1$, the simple threshold policy stabilizes a probabilistic matching system for any $0 < \lambda_i < \infty$, and $0 \le N_i < \infty$ where $i = 1, 2$.*

Proof. The system with $q = 1$ is reduced to an irreducible one-dimensional CTMC with finite state space $\mathbb{S} = \{-N_2 - 1, -N_2, ..., -1, 0, 1, ..., N_1, N_1 + 1\}$ under the simple threshold policy and hence ergodic (see Kulkarni [12], pg. 82, Theorem 3.7 and pg. 285, Theorem 6.10). □

The admitted users leave the system only when there is a continuing intake of users from the other class. When $q = 1$, users of different classes cannot co-exist in the system. Hence, when the number of users of a certain class reaches its maximum, the arrivals from the other class are always admitted. However, when $0 < q < 1$, the resulting CTMC is no longer irreducible. Once the system reaches state $(N_1 + 1, N_2 + 1)$, no users are admitted under the simple threshold policy and hence no users leave the system resulting in infinite waiting times. Therefore, better control mechanisms are needed when $0 < q < 1$.

Now, we investigate the long run average proportion of rejected users when $q = 1$ and the simple threshold policy is employed. Even though, the simple threshold policy is not applicable when $q < 1$, it provides us with an interesting benchmark to compare other control policies.

Theorem 6. *Let c_i be the long run average proportion of rejected class i users for $i = 1, 2$. If $q = 1$ and the simple threshold policy is employed, then*

$$c_1 = \begin{cases} \frac{1}{N_1+N_2+3}, & \text{for } \lambda_1 = \lambda_2, \\ \frac{1-\frac{\lambda_2}{\lambda_1}}{1-(\frac{\lambda_2}{\lambda_1})^{N_1+N_2+3}}, & \text{for } \lambda_1 \ne \lambda_2, \end{cases} \tag{5}$$

and c_2 is given by interchanging λ_1 and λ_2 in (5).

Proof. Let $\{X^{1,ST}(t), t \geq 0\}$ be the resulting stochastic process, then c_1 and c_2 correspond to the stationary probabilities for states $-N_1 - 1$ and $N_2 + 1$, respectively. The process $\{X^{1,ST}(t) + N_1 + 1, t \geq 0\}$ is stochastically identical to an $M/M/1/N_1 + N_2 + 2$ queueing process and the result follows (see, Gross and Harris [8] pg. 77). □

By considering the extreme cases $N_1 = N_2 = 0$ and $N_1, N_2 \to \infty$ in equation (5), we obtain bounds on c_1. In particular,

$$\max(1 - \frac{\lambda_2}{\lambda_1}, 0) \leq c_1 \leq \frac{\lambda_1^2}{\lambda_1^2 + \lambda_1\lambda_2 + \lambda_2^2}. \tag{6}$$

These bounds act as a benchmark when evaluating other policies.

4.2.2 Accept-the-Shortest-Queue (ASQ) Policy

In the previous section we have seen that assuring continuing intakes of both classes of users plays a key role in avoiding any absorbing states. Therefore, rather than imposing strict bounds, the ASQ policy tries to maintain a balance between different classes by only admitting users belonging to the shorter queue. As the system size increases, both classes of users get more likely to find a match upon arrivals.

Definition 7. The accept-the-shortest-queue policy is an admission control policy which admits a class-i user arriving at time t if and only if $X_i^q(t-) = \min_{j \in \{1,2\}} X_j^q(t-), i = 1, 2$.

Let $\{X^{q,ASQ}(t), t \geq 0\}$ be the CTMC representing the probabilistic matching system under the ASQ policy. Then the state space is $\mathbb{S} = \{-1, 0, 1\}$ when $q = 1$, and $\mathbb{S} = \{(i,j) \in \mathbb{N}^2 : |i - j| \leq 1\}$ when $0 < q < 1$. We now prove the stability of the system under the ASQ policy using Foster's criterion (see e.g., Fayolle et al. [7], pg. 29, Theorem 2.2.3 or Kulkarni [12], pg. 95. Theorem 3.10).

Theorem 8. *A probabilistic matching system is stable for any set of arrival rates λ_1 and λ_2 and matching probability $0 < q \leq 1$ under the ASQ policy.*

Proof. When $q = 1$, the state space is finite and the result follows. When $0 < q < 1$, define $\{X_n^{q,ASQ}, n \in \mathbb{N}\}$ to be the corresponding embedded DTMC and $f(i,j) = i + j + 1$. The transition probabilities for the embedded DTMC are

$$p_{(i,j)(k,l)} = \begin{cases} \lambda_1 r^j/(\lambda_1 + \lambda_2) & \text{if } k = i+1 \text{ and } l = j = i, \\ \lambda_2 r^i/(\lambda_1 + \lambda_2) & \text{if } k = i = j \text{ and } l = j+1, \\ \lambda_1(1 - r^j)/(\lambda_1 + \lambda_2) & \text{if } k = i = j \text{ and } l = j - 1 \geq 0, \\ \lambda_2(1 - r^i)/(\lambda_1 + \lambda_2) & \text{if } k = i - 1 \geq 0 \text{ and } l = j = i, \\ r^j & \text{if } k = l = j = i+1, \\ 1 - r^j & \text{if } k = l = i = j - 1 \geq 0, \\ r^i & \text{if } k = l = i = j+1, \\ 1 - r^i & \text{if } k = l = j = i - 1 \geq 0. \\ 0 & \text{otherwise.} \end{cases}$$

Then, $f(i,j)$ is positive for all states and

$$E[f(X^{q,ASQ}_{n+1,1}, X^{q,ASQ}_{n+1,2})|X^{q,ASQ}_n = (i,j)] - f(i,j) = \begin{cases} 2r^i - 1, & \text{if } i > j, \\ 2r^j - 1, & \text{if } i \le j. \end{cases}$$

Let $m = \min\{i \in \mathbb{N} : r^i < \frac{1}{2}\}$. Then, for the finite set $H = \{(i,j) \in \mathbb{N} \times \mathbb{N}, 0 \le i,j \le m\} \subset S$ and $\epsilon = \frac{1}{2} - r^m > 0$, we have $|E[f(X^{q,ASQ}_{n+1,1}, X^{q,ASQ}_{n+1,2})|X^{q,ASQ}_n = (i,j)] - f(i,j)| < 2$, if $(i,j) \in H$, and $E[f(X^{q,ASQ}_{n+1,1}, X^{q,ASQ}_{n+1,2})|X^{q,ASQ}_n = (i,j)] - f(i,j) < -2\epsilon$ if $(i,j) \notin H$. Thus, all conditions in Foster's Criterion are satisfied, and so the system is ergodic. □ □

The ergodicity of the CTMC guarantees the existence of stationary probabilities. Theorem 9 provides an explicit representation for the stationary probabilities.

Theorem 9. *The stationary probabilities under the ASQ when $0 < q < 1$ are*

$$p_{i,j} = a_{ij} p_{0,0}, \ \textit{for } (i,j) \in \mathbb{S},$$

where

$$a_{ij} = \begin{cases} 1 & \textit{if } i = j = 0 \\ \dfrac{r^{i^2}}{\left[\prod_{k=1}^{i}(1-r^k)\right]^2} & \textit{if } i \ge 1, j = i \\ \dfrac{\lambda_1}{\lambda_2(1-r)} & \textit{if } i = 1, j = 0 \\ \dfrac{\lambda_2}{\lambda_1(1-r)} & \textit{if } i = 0, j = 1 \\ \dfrac{\lambda_2 r^{i(i+1)} p_{0,0}}{\lambda_1 \prod_{k=1}^{i}(1-r^k) \prod_{k=1}^{i+1}(1-r^k)} & \textit{if } i \ge 1, j = i+1 \\ \dfrac{\lambda_1 r^{i(i-1)} p_{0,0}}{\lambda_2 \prod_{k=1}^{i}(1-r^k) \prod_{k=1}^{i-1}(1-r^k)} & \textit{if } j \ge 1, i = j+1 \end{cases}$$

and $p_{0,0} = \dfrac{1}{1 + \sum_{i=1}^{\infty} \sum_{j=i-1}^{i+1} a_{ij}}$.

Proof. See Appendix 8.2. □ □

In principle, the performance measures of a probabilistic matching system under the ASQ policy can be calculated using the stationary probabilities presented in Theorem 9. In the next section we generalize the ASQ policy and use another method to calculate the long run percentage of rejected users and present some insights about average queue lengths and waiting times for a more general set of policies including the ASQ policy.

4.2.3 The Functional Threshold Policy

In this section we generalize the idea of applying a "moving" threshold behind the ASQ policy. Instead of applying the threshold as the number of users in the other queue, the functional threshold (FT) policy sets the threshold to be a *function* of the number of users in the other queue.

Definition 10. An admission control policy is called a functional threshold policy if it admits a class-i user arriving at time t when $X_i^q(t-) \leq h(X_j^q(t-))$, $i, j = 1, 2, i \neq j$, where $h(\cdot) : \mathbb{N} \to \mathbb{R}$ is a non-decreasing function and satisfies $x \leq h(x) < \infty$, for all $x \geq 0$.

The threshold function $h(\cdot)$ makes the FT policy more flexible compared to the ASQ Policy. When the threshold function is set to be $h(x) = x$, the functional threshold policy is equivalent to the ASQ policy. The condition $x \leq h(x) < \infty$, for all $x \geq 0$ prevents selecting inappropriate threshold functions (e.g., $h(x) = N$, as seen in the ST policy), and implies that $\min_{j \in \{1,2\}} X_j^q(t-) \leq X_i^q(t-) \leq h(X_i^q(t-)), i = 1, 2$. Hence, we always accept the users from class with the shortest queue and assure a continuing intake to avoid absorbing states.

Theorem 11. *A functional threshold policy is a stabilizing policy for probabilistic matching systems with any set of arrival rates* $\lambda_1 > 0$ *and* $\lambda_2 > 0$*, and matching probability* $0 < q \leq 1$.

Proof. When $q = 1$, the resulting CTMC is irreducible with a finite state space and hence stable. When $0 < q < 1$, the state space of the CTMC is $\mathbb{S} = \{(i, j) \in \mathbb{N} \times \mathbb{N} : i \leq h(j + 1), j \leq h(i + 1)\}$ and we apply Foster's criterion on the embedded DTMC. To write down the transition probabilities for the embedded DTMC, we assume $i \leq j$ and consider the following cases:

When $h(i) + 1 \leq j \leq \max\{h(i) + 1, h(i + 1)\}$,

$$p_{(i,j)(k,l)} = \begin{cases} r^j & \text{if } k = i + 1 \text{ and } j = l, \\ 1 - r^j & \text{if } k = i \text{ and } l = j - 1, \\ 0 & \text{otherwise.} \end{cases}$$

When $i \leq j \leq h(i)$,

$$p_{(i,j)(k,l)} = \begin{cases} \frac{1}{\lambda_1+\lambda_2}\lambda_2 r^i & \text{if } k = i \text{ and } j = l + 1, \\ \frac{1}{\lambda_1+\lambda_2}\lambda_2(1 - r^i) & \text{if } k = i - 1 \text{ and } l = j, \\ \frac{1}{\lambda_1+\lambda_2}\lambda_1 r^j & \text{if } k = i + 1 \text{ and } l = j, \\ \frac{1}{\lambda_1+\lambda_2}\lambda_1(1 - r^j) & \text{if } k = i \text{ and } l = j - 1, \\ 0 & \text{otherwise.} \end{cases}$$

The transition probabilities for the states where $i > j$ can be obtained by interchanging i and j. Next, we define $f(i, j) = i + j + 1$, then $f(i, j)$ is positive for all states and

$$\begin{aligned} &\mathbb{E}[f(X_{n+1,1}^{q,FT}, X_{n+1,2}^{q,FT}) | X_n^{q,FT} = (i, j)] - f(i, j) \\ &= \begin{cases} 2r^i - 1 & \text{if } h(i) + 1 \leq j \leq \max\{h(i) + 1, h(i + 1)\}, \\ 2r^j - 1 & \text{if } h(j) + 1 \leq i \leq \max\{h(j) + 1, h(j + 1)\}, \\ \dfrac{\lambda_2(2r^i - 1) + \lambda_1(2r^j - 1)}{\lambda_1 + \lambda_2} & \text{if } i \leq j \leq h(i) \text{ or } j \leq i \leq h(j). \end{cases} \end{aligned}$$

Let $m = \min\{i : r^i < \frac{1}{2}\}$ and $\epsilon = 1 - 2r^m > 0$. Then, $(i, j) \notin H = \{(k, l) \in \mathbb{N} \times \mathbb{N} : k \leq m, l \leq m\}$, implies $\mathbb{E}[f(X_{n+1,1}^{q,FT}, X_{n+1,2}^{q,FT}) | X_n^{q,FT} = (i, j)] - f(i, j) < -\epsilon$. Hence, the result follows. □ □

Under the functional threshold policy, stating and solving the global balance equations is fairly difficult. However, it is still possible to obtain insights about some key performance measures of the system by imposing some mild restrictions on the threshold function $h(x)$.

Theorem 12. *Suppose that the functional threshold policy is employed with the threshold function $h(x) = x + d$, where $d \geq 0$ is an arbitrary constant. Then, the long run percentage of rejected users, c_1 and c_2, are independent of the matching probability $0 < q \leq 1$ and*

$$c_1 = \begin{cases} \frac{1}{2[d]+3}, & \text{for } \lambda_1 = \lambda_2, \\ \frac{1-\frac{\lambda_2}{\lambda_1}}{1-(\frac{\lambda_2}{\lambda_1})^{2[d]+3}}, & \text{for } \lambda_1 \neq \lambda_2, \end{cases} \tag{7}$$

and c_2 can be obtained interchanging λ_1 and λ_2.

Proof. See Appendix 8.3 □ □

Corollary 13. *Under the ASQ policy, the long run percentage of rejected users is independent of the matching probability and*

$$c_1 = \frac{\lambda_1^2}{\lambda_1^2 + \lambda_1\lambda_2 + \lambda_2^2}, \tag{8}$$

c_2 is obtained interchanging λ_1 and λ_2.

Proof. The result follows by replacing $d = 0$ in (7) and cancelling as appropriate. □ □

One may expect that as the matching probability q increases, the users match more quickly and this yields a better throughput and hence smaller long run percentage of rejected users. However, contrary to this initial intuition, Theorem 12 indicates that under the specified subclass of functional threshold policies, the long run average percentage of rejected users does not depend q. In Section 5, we numerically show that for many reasonable threshold functions the behavior of the system is even more counter-intuitive, i.e., the long run rejection probabilities actually increase as the matching probability increases! Further discussions about the reasons behind this unexpected phenomenon are provided in Section 5.

Unlike the throughput, the long run average number of users in the system and the average waiting times depend on the matching probability q, even when $h(x) = x + d$. However, we are able to prove similar insensitivity results for the difference between the long run average numbers of users from different classes for the same class of threshold functions.

Theorem 14. *Let L_i^q denote the long run average numbers of class-i user, $i = 1, 2$ in the system and $\rho = \lambda_2/\lambda_1$. If the functional threshold policy is employed with $h(x) = x + d$, where $d \geq 0$ is a constant, then for any $0 < q \leq 1$, the difference between average queue lengths of classes, $L_1^q - L_2^q$, does not depend on the matching probability q, and we have $L_1^q - L_2^q = 0$ if $\lambda_1 = \lambda_2$ and*

$$L_1^q - L_2^q = \frac{(d+2)\rho^{2d+3} + d + 1}{1 - \rho^{2d+3}} + \frac{(1-\rho)\rho^{d+2}(\rho^{d+2} - \rho^{-d-1})}{(1-\rho^{2d+3})(\rho-1)^2} \tag{9}$$

if $\lambda_1 \neq \lambda_2$.

Proof. See Appendix 8.3. □ □

Corollary 15. *Suppose the functional threshold policy is employed with $h(x) = x + d$, where $d \geq 0$ is a constant. Let W_1^q and W_2^q denote the long run average waiting times for users, $\rho = \lambda_2/\lambda_1$ and $\lambda^e \equiv (1 - c_1)\lambda_1$, where c_1 is as in (7), then for any $0 < q \leq 1$, the difference between average* *waiting times of classes,* $W_1^q - W_2^q$*, does not depend on the matching probability* q*, and we have* $W_1^q - W_2^q = 0$ *if* $\lambda_1 = \lambda_2$ *and*

$$W_1^q - W_2^q = \frac{1}{\lambda^e}\left(\frac{(d+2)\rho^{2d+3}+d+1}{1-\rho^{2d+3}} + \frac{(1-\rho)\rho^{d+2}(\rho^{d+2}-\rho^{-d-1})}{(1-\rho^{2d+3})(\rho-1)^2}\right)$$

if $\lambda_1 \neq \lambda_2$.

Proof. Using PASTA property (see El-Taha and Stidham [6], Corollary 1.10 and Theorem 3.23), Little's Law and (4), $W_i^q = L_i^q/\lambda^e$ for $i = 1, 2$, and hence the result follows from Theorem 14. □ □

Functional threshold policy relies on rejecting both types of users in a similar fashion. We next introduce a policy which rejects only one type of user.

4.2.4 One-Sided Threshold Policy

The policies we discuss so far, reject both classes of users when the numbers of them reach certain limits. If it is known that the arrival rate of a class is less than the arrival rate of the other (e.g., $\lambda_1 < \lambda_2$), it may not be reasonable to ever reject that class of users. For example, in general, the rate of employers arriving at an employment portal is significantly less than the arrival rate of employees. Thus, each job posting is deemed valuable and the employment portal would not want to lose any employer who wishes to subscribe. In such a matching system, it is more reasonable to reject only employees when the number of them reaches a certain threshold N_2, but to always accept employers .

Definition 16. When $\lambda_1 < \lambda_2$, a one-sided threshold (OST) policy admits users of class-2 at time t if and only if $X_2(t) \leq N_2$, whereas users of class-1 are always admitted.

Theorem 17. *A probabilistic matching system with arrival rates* $\lambda_1 < \lambda_2$ *and matching probability* $0 < q \leq 1$ *is ergodic under a one-sided threshold policy which applies a finite threshold* $N_2 \geq 0$ *to class-2 users.*

Proof. When $q = 1$, the one-dimensional CTMC has a state space the set of integers from $-\infty$ to $N_2 + 1$. Since $\frac{\lambda_1}{\lambda_2} < 1$, the system is ergodic. When $0 < q < 1$, we have the state space $\mathbb{S} = \{(i,j) \in \mathbb{N} \times \mathbb{N} : i \geq 0, 0 \leq j \leq N_2 + 1\}$. The transition probabilities for the embedded DTMC under one-sided threshold policy are

$$p_{(i,j)(k,l)} = \begin{cases} \lambda_1 r^j/(\lambda_1+\lambda_2) & \text{if } k = i+1 \text{ and } l = j \leq N_2, \\ \lambda_2 r^i/(\lambda_1+\lambda_2) & \text{if } k = i \text{ and } l = j+1 \leq N_2+1, \\ \lambda_1(1-r^j)/(\lambda_1+\lambda_2) & \text{if } k = i \text{ and } 0 \leq l = j-1 \leq N_2-1, \\ \lambda_2(1-r^i)/(\lambda_1+\lambda_2) & \text{if } k = i-1 \geq 0 \text{ and } l = j \leq N_2, \\ r^j & \text{if } k = i+1 \text{ and } l = j = N_2+1, \\ 1-r^j & \text{if } k = i \text{ and } l = j-1 = N_2, \\ 0 & \text{otherwise.} \end{cases}$$

For $0 < q < 1$, we can always find a positive number a, such that, $r < a(1-r)$ (recall that $r = 1-q$). Thus, the inequality $r^j < a(1-r^j)$ holds for all j and in particular, for some $\epsilon_0 > 0$, $r^{N_2+1} - a(1-r^{N_2+1}) < -\epsilon_0$. For any state $(i,j) \in \mathbb{S}$, define $f(i,j) = i + aj + 1$ and $d(i,j) = E[f(X_{n+1,1}^{q,OST}, X_{n+1,2}^{q,OST}) - |X_n^{q,OST} = (i,j)] - f(i,j)$. Then, for all $i \geq 0$, $d(i, N_2+1) =$

$r^{N_2+1} - a(1 - r^{N_2+1}) < -\epsilon_0$. Also, for all $i \geq 0$, $d(i,0) = \frac{1}{\lambda_1+\lambda_2}(\lambda_1 - \lambda_2 + a\lambda_2 r^i + \lambda_2 r^i)$. Since, $\lambda_1 < \lambda_2$, there exists an m_1 and $\epsilon_1 > 0$ such that $d(i,0) < -\epsilon_1$. If $N_2 \geq 1$, then for $i \geq 1$ and $1 \leq j \leq N_2$

$$d(i,j) = \frac{1}{\lambda_1+\lambda_2}(\lambda_1 a r^i - \lambda_1(1-r^i) + \lambda_2 r^j - \lambda_2(1-r^j)a) < \frac{\lambda_1}{\lambda_1+\lambda_2}(r^i(a+1)-1).$$

Thus, there exists an $m_2 > 0$ and $\epsilon_2 > 0$, such that, when $i > m_2$, $d(i,j) = -\epsilon_2 < 0$. Take $m_2 = 0$ when $N_2 = 0$, and let $m = \max\{m_1, m_2\}$ and $\epsilon = \min\{\epsilon_0, \epsilon_1, \epsilon_2\}$. Then, for the finite set $H = \{(i,j) \in \mathbb{N} \times \mathbb{N}, 0 \leq i \leq m, 0 \leq j \leq N_2\}$ we have, $d(i,j) < -\epsilon, (i,j) \notin H$. The conditions of Foster's criterion are satisfied and hence the system is stable. $\square$ $\square$

Theorem 17 states that the stability of the matching system neither depends on the matching probability nor the threshold N_2. As a consequence of Theorem 3, we prove that the long run average percentage of rejected users of a probabilistic matching system under a one-sided threshold policy is also independent of the matching probability q and the threshold N_2.

Theorem 18. *Suppose a one-sided threshold policy with threshold N_2 is employed on a probabilistic matching system with arrival rates $0 < \lambda_1 < \lambda_2 < \infty$ and matching probability $0 < q \leq 1$. Then, the long run proportion of rejected users is independent of both the matching probability q and the threshold N_2 and is given by*

$$c_2 = 1 - \frac{\lambda_1}{\lambda_2}.$$

Proof. Since, no class-1 user is rejected, using Theorem 3, $\lim_{t\to\infty} \frac{A_1^e(t)}{t} = \lim_{t\to\infty} \frac{A_2^e(t)}{t} = \lambda_1$. The result follows from (4). $\square$ $\square$

Unlike the percentage of rejected users, most other performance measures, such as the average waiting time or the average queue length, depend on both the matching probability and the threshold. We analyze these quantities numerically in Section 5.

The one-sided threshold policy achieves the lower bound in (6) which is the best rejection rate possible as rejecting less users will definitely yield an unstable system. As there are always rejected users from both classes in functional threshold policy, the same performance cannot be attained under any threshold function. On the other hand, the rejection percentage under the ASQ policy is equal to the upper bound in (6), which is the worst rejection percentage possible under the simple threshold policy when $q = 1$.

5 Numerical Results

We have seen that it is rather difficult to derive explicit equations for the performance measures of the probabilistic matching systems under the suggested stabilizing policies. This is partly due to the transition rates involving powers of the matching probability. In this section, we present a numerical analysis of the performance of probabilistic matching systems under different policies. Some of our results appear to be quite counter-intuitive and we present explanations for these results.

Table 1: Long run percentage of rejected users for functional threshold policies with various threshold functions when $\lambda_1 = \lambda_2 = 1$

q	x	$2x$	x^2	$\max\{x, 5\}$
0.10	0.333	0.091	0.069	0.310
0.20	0.333	0.159	0.183	0.120
0.30	0.333	0.211	0.275	0.086
0.40	0.333	0.252	0.315	0.080
0.50	0.332	0.282	0.329	0.077
0.60	0.333	0.303	0.332	0.076
0.70	0.334	0.319	0.333	0.077
0.80	0.333	0.326	0.333	0.077
0.90	0.333	0.332	0.333	0.078
1.00	0.333	0.334	0.334	0.078

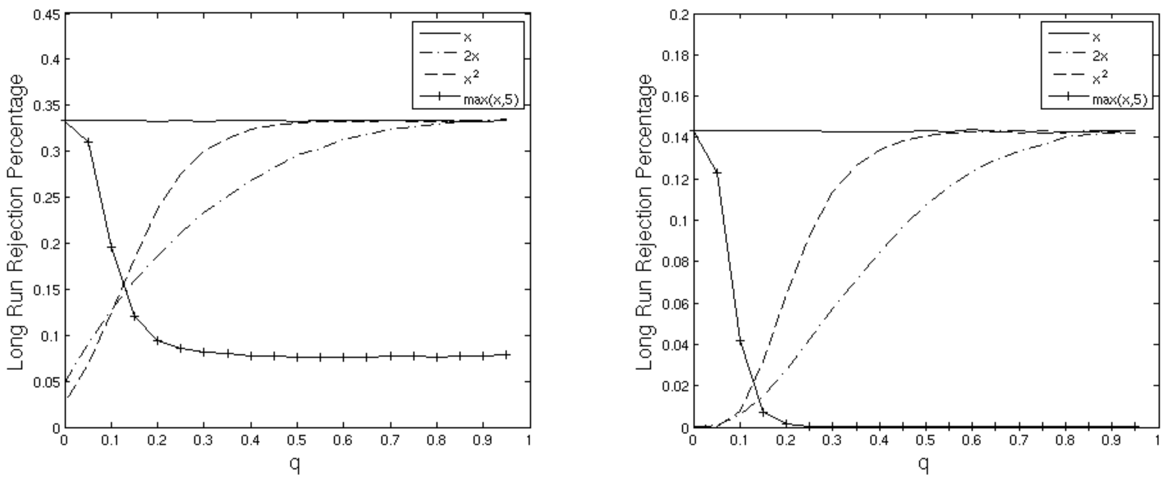
5.1 Computational Experiments on Long Run Percentages of Rejected Users

Theorem 18 shows that if one-sided threshold policy is employed the long run rejection rate is insensitive to both the matching probability and the threshold value. Similarly, Theorem 12 shows that if the functional threshold policy is employed with a specific type of threshold function, the long run rejection rate of a probabilistic matching system is insensitive to the matching probability. In this section, we present simulation results to see how the rejection rate behaves under various functional threshold policies.

We consider two probabilistic matching systems, where the first system has arrival rates $\lambda_1 = \lambda_2 = 1$ and the second system has $\lambda_1 = 1$ and $\lambda_2 = 2$, to test how the matching probability affects the long run percentages of rejected users. We simulate 10 replications in each experiment, where each replication covers 1,000,000 time units. We compare four different threshold functions: $h_1(x) = x, h_2(x) = 2x, h_3(x) = x^2$ and $h_4(x) = \max\{5, x\}$. Note that the first threshold function $h_1(x) = x$ corresponds to the ASQ policy.

Table 1 and Figure 1 summarize the corresponding results. As proven in Theorem 12, the first column of Table 1 demonstrates the insensitivity of the long run rejection probabilities to the matching probability q under the ASQ policy. The situation for general threshold functions is even more surprising. One may intuitively guess that as the matching probability increases, the users match faster and as a result better performance for rejection rates can be achieved. Contrary to this initial intuition that we would observe lower rejection percentages for higher matching probabilities, we discover that for $h_2(x) = 2x$ and $h_3(x) = x^2$ the long run rejection percentages actually *increase* as the matching probability q increases. The rejection percentages are very close to 0 when q is close to 0, and as q increases to 1, they converge to that of the ASQ policy. However, for the threshold function $h_4(x) = \max\{5, x\}$, we notice an opposite behavior which matches our initial intuition, i.e., the rejection probabilities are very close to that of the ASQ policy when q is close to 0 and it *decreases* as q increases.

We explain this surprising behavior using Figure 2 which illustrates the state space of the CTMC under various functional threshold policies. The boundaries in Figures 2(a) and 2(c) shown with bold lines correspond to the regions where a class of users are rejected. For $h_2(x) = 2x$, the shaded region corresponds to the states where users are rejected. As q decreases, we see that the



(a) $\lambda_1 = \lambda_2 = 1$

(b) $\lambda_1 = 1, \lambda_2 = 2$

Figure 1: Matching probability q vs. long run percentage of rejected class-1 users for functional threshold policies with different threshold functions

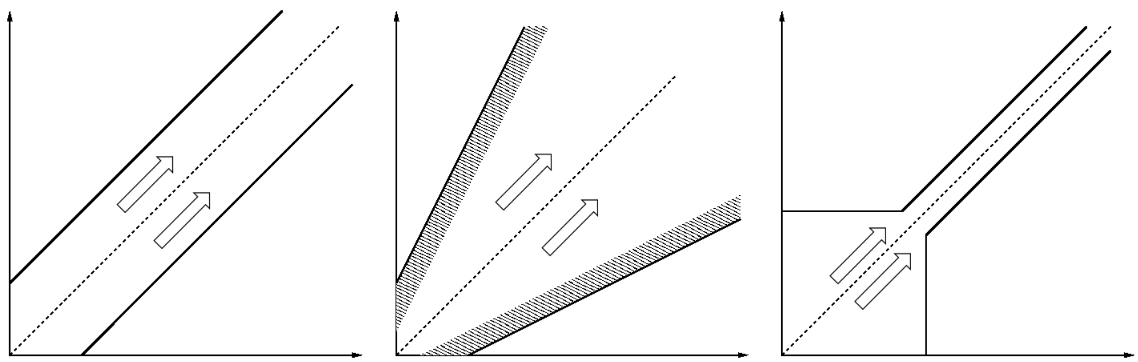
probability mass of the stationary distribution moves in the direction of the arrows shown in the figures, somewhat parallel to the diagonal illustrated by the dashed line. For the threshold function $h_2(x) = 2x$, the random walk is pushed towards a wider region as q decreases, and the proportion of time spent in rejection region decreases. For $h_4(x) = \max\{x, 5\}$, the situation is the opposite, i.e., the walk is pushed to narrower areas and the proportion of time spent in the rejection region increases as q decreases. This explanation is also in accordance with Theorem 12, as the width of the state space is constant for $h_1(x) = x + d$ as seen in Figure 2(a).

5.2 Computational Experiments on Average Queue Lengths and Average Waiting Times

We now turn to the study of long run average queue lengths (L_i) and waiting times (W_i). Our simulations use the same structure described in Section 5.1. Our first set of experiments analyze how changing the matching probability q affects our parameters under the functional threshold policy. The results are presented in Table 2 and Table 3.

As expected, we see that the long run average queue lengths decrease as q increases. The average queue lengths under all threshold functions behave similarly. The only exception is that, if q is close to 1, the average queue lengths are significantly higher under $h_4(x)$, which is expected because when $q = 1$, the number of users is bounded by 5 for this threshold function, whereas the others are bounded by 1. We also observe that the average waiting times for $h_1(x)$ and $h_4(x)$ are quite high for small q due to the poor throughput. When q is close to 1, we observe that the average waiting times under $h_4(x)$ is still higher, because even though the throughput for this threshold function is higher than the others, the average queue length is still relatively higher.

Theorem 14 proves that under specific functional threshold policies, the difference between average queue lengths of different types of customers is constant with respect to q. When $\lambda_1 = \lambda_2$,



(a) $h_1(x) = x + d$ (b) $h_2(x) = 2x$ (c) $h_4(x) = \max\{x, 5\}$

Figure 2: State space for the CTMC under functional threshold policies with various threshold functions when $\lambda_1 = \lambda_2 = 1$

Table 2: Average queue lengths and waiting times for functional threshold policies with various threshold functions when $\lambda_1 = \lambda_2 = 1$

	$h_1(x) = x$		$h_2(x) = 2x$		$h_3(x) = x^2$		$h_4(x) = \max\{x, 5\}$	
q	L_1	W_1	L_1	W_1	L_1	W_1	L_1	W_1
0.1	6.458	9.680	6.532	7.188	6.655	7.148	6.655	9.512
0.2	2.998	4.497	3.009	3.575	2.945	3.607	3.311	3.766
0.3	1.842	2.766	1.831	2.318	1.767	2.440	2.370	2.591
0.4	1.265	1.898	1.235	1.650	1.228	1.794	2.000	2.174
0.5	0.924	1.387	0.889	1.239	0.911	1.358	1.813	1.965
0.6	0.702	1.053	0.671	0.964	0.699	1.048	1.727	1.870
0.7	0.550	0.826	0.530	0.778	0.550	0.826	1.681	1.822
0.8	0.446	0.669	0.437	0.651	0.447	0.670	1.653	1.791
0.9	0.377	0.566	0.375	0.561	0.377	0.566	1.629	1.766
1.0	0.333	0.500	0.333	0.500	0.334	0.501	1.613	1.747

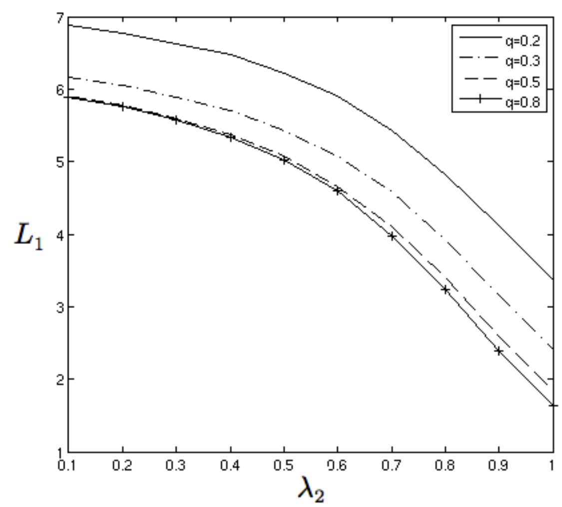
Table 3: Long run average queue length and waiting times for functional threshold policies with various threshold functions when $\lambda_1 = 1$ and $\lambda_2 = 2$

	$h_1(x) = x$						$h_2(x) = 2x$					
q	L_1	L_2	$L_2 - L_1$	W_1	W_2	$W_2 - W_1$	L_1	L_2	$L_2 - L_1$	W_1	W_2	$W_2 - W_1$
0.1	6.249	6.678	0.429	7.297	7.798	0.501	4.531	8.759	4.228	4.537	8.771	4.234
0.2	2.776	3.206	0.430	3.241	3.743	0.502	2.096	4.006	1.910	2.129	4.070	1.941
0.3	1.630	2.058	0.428	1.902	2.401	0.499	1.250	2.458	1.208	1.305	2.565	1.260
0.4	1.053	1.482	0.429	1.228	1.727	0.499	0.812	1.690	0.878	0.874	1.819	0.945
0.5	0.715	1.143	0.428	0.834	1.333	0.499	0.557	1.243	0.686	0.616	1.376	0.760
0.6	0.496	0.924	0.428	0.578	1.077	0.499	0.397	0.965	0.568	0.450	1.092	0.642
0.7	0.348	0.776	0.428	0.406	0.906	0.500	0.293	0.788	0.495	0.336	0.904	0.568
0.8	0.248	0.676	0.428	0.289	0.789	0.500	0.224	0.677	0.453	0.260	0.786	0.526
0.9	0.182	0.611	0.429	0.213	0.714	0.501	0.176	0.611	0.435	0.206	0.712	0.506
1.0	0.143	0.572	0.429	0.167	0.667	0.500	0.143	0.572	0.429	0.167	0.668	0.501
	$h_3(x) = x^2$						$h_4(x) = \max\{x, 5\}$					
q	L_1	L_2	$L_2 - L_1$	W_1	W_2	$W_2 - W_1$	L_1	L_2	$L_2 - L_1$	W_1	W_2	$W_2 - W_1$
0.1	3.475	10.467	6.992	3.473	10.458	6.985	6.243	6.851	0.608	7.118	7.811	0.693
0.2	2.063	3.928	1.865	2.131	4.058	1.927	1.898	5.158	3.260	1.911	5.195	3.284
0.3	1.367	2.199	0.832	1.507	2.422	0.915	0.604	5.084	4.480	0.603	5.079	4.476
0.4	0.969	1.502	0.533	1.109	1.719	0.610	0.214	5.061	4.847	0.214	5.060	4.846
0.5	0.689	1.143	0.454	0.801	1.328	0.527	0.085	5.041	4.956	0.085	5.038	4.953
0.6	0.490	0.923	0.433	0.571	1.075	0.504	0.039	5.032	4.993	0.040	5.037	4.997
0.7	0.345	0.775	0.430	0.403	0.905	0.502	0.024	5.023	4.999	0.024	5.018	4.994
0.8	0.248	0.676	0.428	0.289	0.789	0.500	0.018	5.021	5.003	0.018	5.020	5.002
0.9	0.183	0.610	0.427	0.213	0.711	0.498	0.016	5.017	5.001	0.016	5.017	5.001
1.0	0.143	0.571	0.428	0.167	0.667	0.500	0.015	5.013	4.998	0.015	5.009	4.994

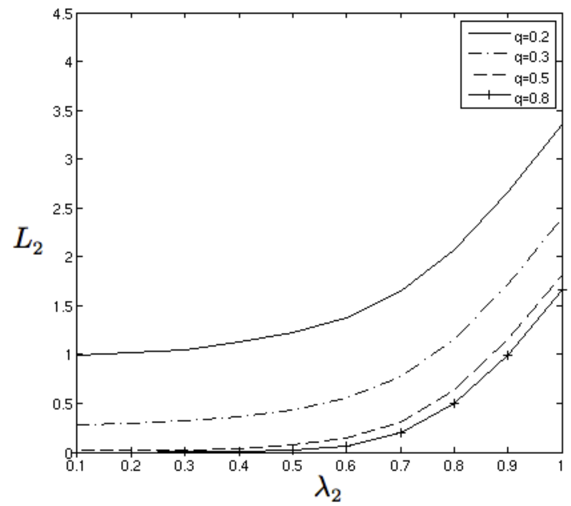
this is trivially true for any functional threshold policy, as due to symmetry $L_1 - L_2 = 0$. When $\lambda_1 = 1$ and $\lambda_2 = 2$, the observations as presented in Table 3, are parallel to those related to the rejection probabilities. For the threshold functions where the rejection probabilities are insensitive to q, we see that the difference between average queue lengths is also insensitive to q. For the threshold functions where the rejection probabilities are increasing (decreasing) with respect to q, the difference $L_2 - L_1$ is decreasing (increasing).

Next, we study the average queue lengths and the average waiting times under the one-sided threshold policy and see how they depend on N_2 and q. We assume that $\lambda_1 = 1$ and $\lambda_2 = 2$ and vary q and N_2. Since the average waiting times can be calculated as the products of the average queue lengths and the throughput (which is 1 in our case), we present the results only for the average queue lengths in Tables 4 and 5. We see that the average queue length (L_1) of the user class with the lower arrival rate is highly sensitive to the changes in matching probability and the threshold, and decreases as these quantities increase. On the other hand, the average queue length for the class with higher arrival rate (L_2) is less sensitive to the changes in the matching probability and the threshold. Under the one-sided threshold policy, the number of class-2 users is bounded by $N_2 + 1$ and we observe that average queue length is in general very close to this upper bound and increases almost linearly as N_2 increases.

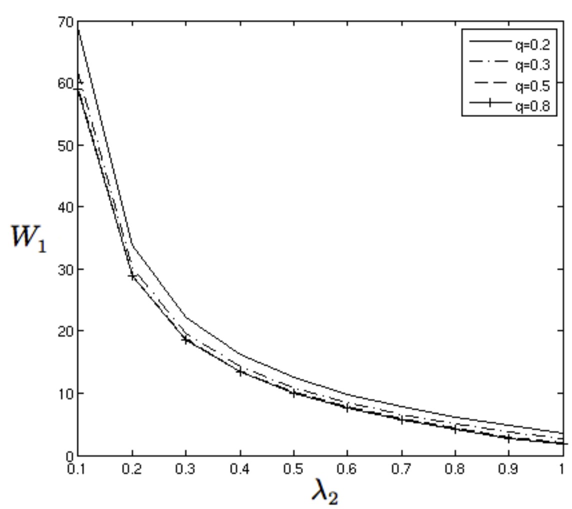
The one-sided threshold policy requires one of the arrival rates to be strictly greater than the other, e.g., $\lambda_1 > \lambda_2$. Next, we investigate how the ratio of arrival rates λ_2/λ_1 affects the average



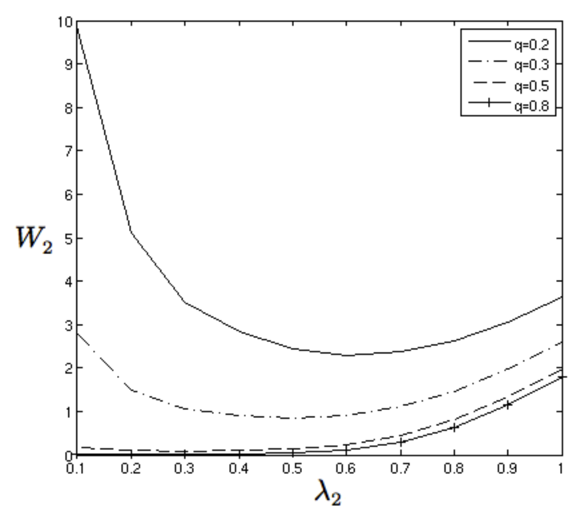
(a) L_1 vs. λ_2



(b) L_2 vs. λ_2

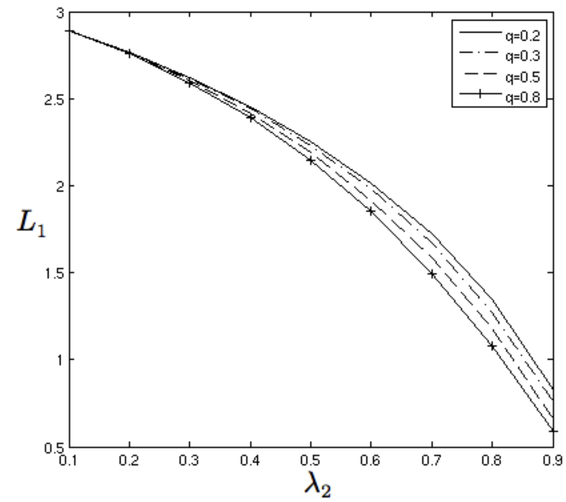


(c) W_1 vs. λ_2

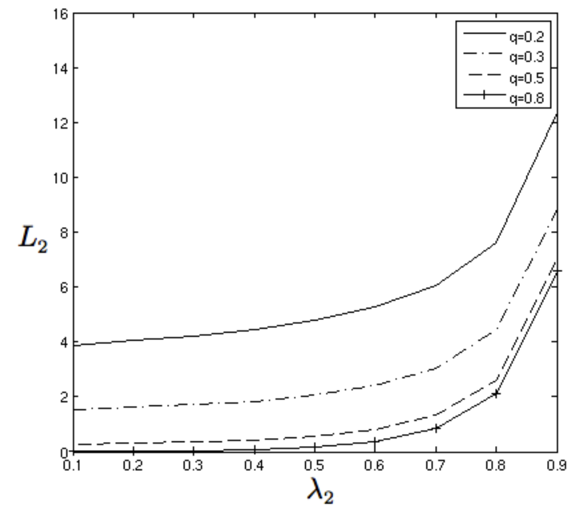


(d) W_2 vs. λ_2

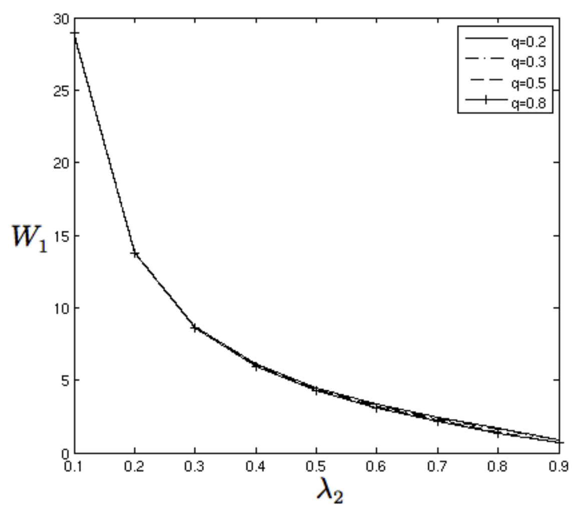
Figure 3: Average queue lengths and waiting times vs. λ_2 for various q and $\lambda_1 = 1$ under the functional threshold policy with $h(x) = x + 5$



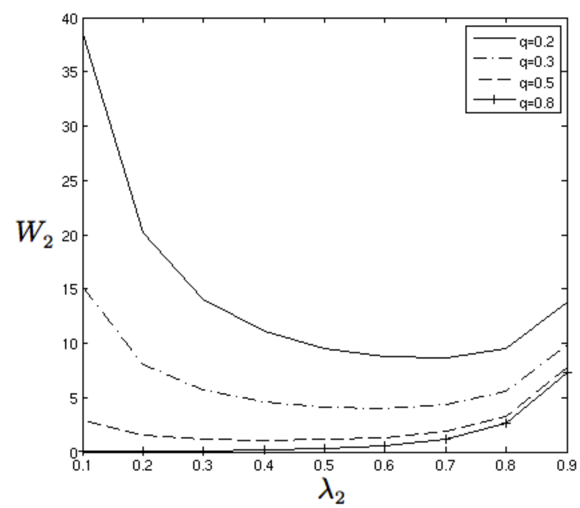
(a) L_1 vs. λ_2



(b) L_2 vs. λ_2



(c) W_1 vs. λ_2



(d) W_2 vs. λ_2

Figure 4: Average queue lengths and waiting times vs. λ_2 for various q and $\lambda_1 = 1$ under the one-sided threshold policy with $N_1 = 2$

Table 4: Long run average queue length under the one-sided threshold policy for varying q

	$N_2 = 3$		$N_2 = 5$	
q	L_1	L_2	L_1	L_2
0.1	12.847	3.219	8.851	5.152
0.2	3.429	3.192	1.933	5.116
0.3	1.327	3.167	0.604	5.086
0.4	0.597	3.139	0.213	5.066
0.5	0.291	3.121	0.086	5.043
0.6	0.161	3.100	0.041	5.032
0.7	0.102	3.085	0.025	5.024
0.8	0.077	3.075	0.019	5.019
0.9	0.067	3.067	0.017	5.016
1	0.063	3.060	0.016	5.013

Table 5: Long run average queue length under one sided threshold policy for varying N_2

	$q = 0.1$		$q = 0.4$	
N_2	L_1	L_2	L_1	L_2
0	33.726	0.499	4.128	0.500
1	21.551	1.350	1.875	1.320
2	16.149	2.270	1.023	2.211
3	12.822	3.220	0.596	3.140
4	10.557	4.182	0.354	4.096
5	8.852	5.155	0.215	5.060
6	7.537	6.128	0.131	6.039
7	6.468	7.109	0.080	7.021
8	5.599	8.102	0.048	8.018
9	4.856	9.087	0.029	9.015
10	4.262	10.075	0.018	10.007

queue lengths and waiting times under different control policies. In these experiments, we fix $\lambda_1 = 1$ and vary λ_2, while ensuring that $\lambda_1 > \lambda_2$. We present our results for functional threshold and one-sided threshold policies in Figures 3 and 4, respectively. We see that as λ_2 increases to λ_1, the average queue lengths and average waiting times for class-1 users decrease monotonically for both control policies. Similarly, we observe that the average queue lengths for class-2 users increase as λ_2 increases. Surprisingly, the average waiting times of class-2 users do not exhibit the same monotonic behavior. Under both policies, the average waiting times of class-2 users first decrease as λ_2 increases and then start increasing and this non-monotonic behavior is observed especially when the matching probability is low. To understand the reasons behind this unexpected behavior, we should remember two basic properties of the probabilistic matching systems: (i) if a user does not match with users in the system upon arrival, she should wait for new users to arrive at the system, and (ii) if too many users from the same class accumulate in the system, it becomes less likely for those users to be picked by a new arrival. When q and λ_2 are very small, an increase in λ_2 causes class-1 users to leave faster and new class-1 users can be admitted, which in return decreases the waiting time of class-2 users due to (i). However, when λ_2 is above a critical value, the negative effects of (ii) dominate the benefits of refreshing the class-1 queue, and the average waiting times increase as λ_2 increases.

6 Conclusion

In this paper we have introduced a novel queueing system where users wait in the system until they match with appropriate users of the other class. This new system can be used to model the traffic in internet portals that serve as a meeting point for suppliers and customers. The matching procedure is probabilistic, i.e., a given pair of users of different classes matches with a given probability independent of other users. We have shown that these new systems differ from conventional queueing systems significantly. We have derived an explicit formula for the conditional probability distribution for the number of matchings to characterize the transient behavior of the probabilistic matching systems. We have shown that if no control mechanism is applied, a probabilistic matching system is not ergodic for any set of parameters. We have suggested admission control policies to ensure stability and analyzed some performance measures. The simple threshold policy and one-sided threshold policy employ constant threshold values to admit users in the system. The simple threshold policy stabilizes the system for $q = 1$, but fails to stabilize the system when $q < 1$. The one-sided threshold policy stabilizes the system when one of the classes has a higher arrival rate. The ASQ and functional threshold policies rely on balancing the number of users in the system and stabilizes matching systems for any set of parameters. We have proved that under a subset of functional threshold policies, the long run proportion of rejected users and the difference between average queue lengths is insensitive to the matching probability. Even more surprisingly, we have shown that the long run proportion of rejected users is an increasing function of the matching probability for a wide subset of functional threshold policies and we have the following conjecture:

Conjecture 19. Suppose the functional threshold policy is employed with a threshold function $h(x)$ which satisfies $h(x+1) \geq h(x) + 1$. Then, the long run proportion of rejected users is a non-decreasing function of the matching probability q.

We believe that the probabilistic matching systems will provide researchers with many interesting questions and we outline a few of these here. A possible research direction is to study how a probabilistic matching system performs under user abandonments. We have seen that a direct analysis of these systems is fairly complicated and hence, we believe heavy traffic limits will be a useful tool to provide further insight. Another interesting research direction is to consider *probabilistic matching networks*, where each class have several types of users and each pair of types has a different matching probability. In the network setting we can focus on matching strategies, in addition to devising admission control policies. Furthermore, one can also consider employing admission controls using pricing mechanisms in order to maximize the financial benefits.

7 Acknowledgements

The authors would like to thank Miklós Rasónyi and Slavomír Takáč for their valuable comments during the course of this research. The work of both authors was supported by the EPSRC grant EP/I017127/1 (Mathematics for Vast Digital Resources).

8 Appendix

8.1 Proof of Theorem 1

Proof of Theorem 1. When $q = 1$, the result is trivial. When $0 < q < 1$, the probability of no matchings, i.e., $k = 0$, when there are m and n class-1 and class-2 users, respectively is r^{mn}. Also, we know that $k \leq m \leq n$ and $P^q_{k,m,n} = 0$ for $k > m$. When $1 \leq k \leq m$, conditioning on whether a specific class-1 user matches with any of the n class-2 users or not, we get

$$P^q_{k,m,n} = r^n P^q_{k,m-1,n} + (1 - r^n) P^q_{k-1,m-1,n-1}. \tag{10}$$

It is clear that given $P^q_{0,m,n} = r^{mn}$ and $P^q_{k,m,n} = 0$ when $k > m$, the solution to (10) is unique. We now use induction to prove that for $m \geq 1$

$$P^q_{1,m,n} = (1 - r^n) r^{(m-1)n} \frac{1 - r^m}{r^{m-1}(1 - r)}. \tag{11}$$

For $m = 1$, $P^q_{1,1,n} = 1 - r^n$. Now, assume that (11) holds for $P^q_{k,m-1,n}$ where $2 \le m \le n$. Then,

$$\begin{aligned} P^q_{1,m,n} &= r^n P^q_{k,m-1,n} + (1-r^n) P^q_{0,m-1,n-1} \\ &= r^n (1-r^n) r^{(m-2)n} \frac{1-r^{m-1}}{r^{m-2}(1-r)} + (1-r^n) r^{(m-1)(n-1)} \\ &= (1-r^n) r^{(m-1)n} \frac{1-r^m}{r^{m-1}(1-r)}, \end{aligned}$$

and hence (11) holds for any $1 \le m \le n$.

Now, we show that solving the three-dimensional recursion (10) can be reduced to solving a two-dimensional recursion. Suppose, $\{a_{k,m}, k \ge 0, m \ge 0\}$ solves

$$a_{k,m} = \begin{cases} 1 & k = 0, \\ r^{k-m} a_{k-1,m-1} + a_{k,m-1} & 1 \le k \le m, \\ 0 & k > m. \end{cases} \tag{12}$$

Then,

$$P^q_{k,m,n} = a_{k,m} r^{(m-k)n} \prod_{i=0}^{k-1} (1 - r^{n-i}) \tag{13}$$

solves (10), where if $k = 0$, $\prod_{i=0}^{-1}(1-r^{n-i})$ is assumed to be 1. To prove this statement, first observe that (13) implies, $P^q_{0,m,n} = r^{mn}$. When $k = 1$, $a_{1,0} = 0, a_{1,1} = 1$ and $a_{1,m} = r^{-m+1} + a_{1,m-1}$ when $m \ge 2$, which implies

$$a_{1,m} = \sum_{i=0}^{m-1} \left(\frac{1}{r}\right)^i = \frac{r^m - 1}{r^{m-1}(r-1)}.$$

Now, fix k, m and n such that $n \ge m \ge 2$ and $k \ge 1$. Suppose that for $0 \le n' < n$, $0 \le m' < n'$, $0 \le k' \le m'$, $0 \le m'' < m$, $k'' \ge 0$ and $0 \le k''' \le k$, $P^q_{k',m',n'}$, $P^q_{k'',m'',n}$ and $P^q_{k''',m,n}$ given as (13) coincides with the solution of (10). Then, if $k + 1 \le m$,

$$\begin{aligned} P^q_{k+1,m,n} &= r^n P^q_{k+1,m-1,n} + (1-r^n) P^q_{k,m-1,n-1} \\ &= a_{k,m} r^{(m-k-1)n} \prod_{i=0}^{k} (1-r^{n-i}) + a_{k,m-1} r^{(m-k-1)(n-1)} \prod_{i=0}^{k} (1-r^{n-i}) \\ &= (a_{k+1,m-1} + r^{-m+k+1} a_{k,m-1}) r^{(m-k-1)n} \prod_{i=0}^{k} (1-r^{n-i}) \\ &= a_{k+1,m} r^{(m-k-1)n} \prod_{i=0}^{k} (1-r^{n-i}). \end{aligned}$$

This proves that if we can solve (12), (13) provides us with the solution of (10).

Now, we provide a solution to the recursion (12). Using (12) $m - k + 1$ times, we get

$$\begin{aligned} a_{k,m} &= r^{-m+k} a_{k-1,m-1} + r^{-m+k+1} a_{k-1,m-2} + \dots + \underbrace{a_{k-1,k-1}}_{1} + \underbrace{a_{k,k-1}}_{0} \\ &= \sum_{j=0}^{m-k} r^{-m+k+j} \cdot a_{k-1,m-1-j}. \end{aligned} \tag{14}$$

We now guess that for $0 \leq k \leq m$, $a_{k,m}$ has the following form

$$a_{k,m} = \sum_{i=0}^{k} r^{-mi} \alpha_{k,i}. \tag{15}$$

Then, $a_{0,m} = 1$ implies $\alpha_{0,0} = 1$. For $k \geq 1$, we plug (15) into both sides of (14) and we obtain:

$$\begin{aligned}
\sum_{i=0}^{k} r^{-mi} \alpha_{k,i} &= \sum_{j=0}^{m-k} r^{-m+k+j} \sum_{i=0}^{k-1} r^{-(m-1-j)i} \alpha_{k-1,i} \\
&= \sum_{i=0}^{k-1} \alpha_{k-1,i} r^{-m+k-mi+i} \sum_{j=0}^{m-k} \left(r^{(i+1)}\right)^j \\
&= \sum_{i=0}^{k-1} \alpha_{k-1,i} \frac{1 - r^{(i+1)(m-k+1)}}{1 - r^{(i+1)}} r^{-m+k-mi+i} \\
&= \sum_{i=0}^{k-1} \frac{r^{-m+k-mi+i} - r^{-ki+2i+1}}{1 - r^{i+1}} \alpha_{k-1,i} \\
&= \sum_{i=0}^{k-1} r^{-m(i+1)} \frac{r^{k+i}}{1 - r^{i+1}} \alpha_{k-1,i} + \sum_{i=0}^{k-1} \frac{-r^{-ki+2i+1}}{1 - r^{i+1}} \alpha_{k-1,i}.
\end{aligned}$$

We then shift the index of the first sum on the righthand side, and for $k \geq 1$, we get

$$\sum_{i=0}^{k} r^{-mi} \alpha_{k,i} = \sum_{i=1}^{k} r^{-mi} \frac{r^{k+i-1}}{1 - r^{i}} \alpha_{k-1,i-1} + \sum_{i=0}^{k-1} \frac{-r^{-ki+2i+1}}{1 - r^{i+1}} \alpha_{k-1,i}.$$

Now, comparing the coefficient of r^{-mi} for $0 \leq i \leq k$, we obtain,

$$\alpha_{k,0} = \sum_{i=0}^{k-1} \frac{-r^{-ki+2i+1}}{1 - r^{i+1}} \alpha_{k-1,i}, \tag{16}$$

and for every $1 \leq i \leq k$:

$$\alpha_{k,i} = \frac{r^{k+i-1}}{1 - r^{i}} \alpha_{k-1,i-1}. \tag{17}$$

Repeating (17) i times we have for $1 \leq i \leq k$,

$$\begin{aligned}
\alpha_{k,i} &= \frac{r^{k+i-1}}{(1 - r^{i})} \frac{r^{k+i-3}}{(1 - r^{i-1})} \frac{r^{k+i-5}}{(1 - r^{i-2})} \cdots \frac{r^{k-i+5}}{(1 - r^{3})} \frac{r^{k-i+3}}{(1 - r^{2})} \frac{r^{k-i+1}}{(1 - r^{1})} \alpha_{k-i,0} \\
&= \frac{r^{ki}}{\prod_{j=1}^{i} (1 - r^{j})} \alpha_{k-i,0}.
\end{aligned} \tag{18}$$

Plugging (18) into (16) and substituting $\beta_k = \alpha_{k,0}$, we obtain a new recurrence:

$$\beta_k = \sum_{i=1}^{k-1} \frac{-r^{-ki+2i+1}}{(1 - r^{i+1})} \frac{r^{(k-1)i}}{\prod_{j=1}^{i} (1 - r^{j})} \beta_{k-1-i} + \frac{-r\beta_{k-1}}{1 - r} = \sum_{i=0}^{k-1} \frac{-r^{i+1} \beta_{k-1-i}}{\prod_{j=1}^{i+1} (1 - r^{j})},$$

and replacing $l = k-1-i$, we get

$$\beta_k = \sum_{l=0}^{k-1} \frac{-r^{k-l}\beta_l}{\prod_{j=1}^{k-l}(1-r^j)}.$$

Thus by now we have a recurrence of the form: $\beta_0 = 1$ and for every $k \geq 1$:

$$\beta_k = \sum_{l=0}^{k-1} \gamma_{k,l}\beta_l, \tag{19}$$

where

$$\gamma_{i,j} = \frac{y_i}{y_j} z_{i-j}, \tag{20}$$

with $y_i = r^i$ and $z_i = \frac{-1}{\prod_{j=1}^{i}(1-r^j)}$. The recursion (19) is a 1-dimensional recurrence, and for $k \geq 1$ has a general solution :

$$\begin{aligned}
\beta_k &= \sum_{l>0} \left(\sum_{0=i_0<i_1<\ldots<i_{l-1}<i_l=k} \left(\prod_{j=1}^{l} \gamma_{i_j,i_{j-1}} \right) \right) \beta_0 \\
&= \sum_{l>0} \left(\sum_{0=i_0<i_1<\ldots<i_{l-1}<i_l=k} \gamma_{i_l,i_{l-1}}\gamma_{i_{l-1},i_{l-2}} \cdots \gamma_{i_2,i_1}\gamma_{i_1,i_0} \right).
\end{aligned} \tag{21}$$

We prove this by induction.

(i) when $k = 1$, Equation (21) implies

$$\beta_1 = \sum_{l>0} \left(\sum_{0=i_0<i_1<\ldots<i_{l-1}<i_l=1} \left(\prod_{j=1}^{l} \gamma_{i_j,i_{j-1}} \right) \right) \beta_0 = \gamma_{1,0},$$

which satisfies Equation (19). Hence Equation (21) holds for β_1.

(ii) suppose Equation (21) holds for all β_n, $1 \leq n \leq k-1$, then according to Equation (19),

$$\begin{aligned}
\beta_k &= \sum_{n=0}^{k-1} \gamma_{k,n}\beta_n \\
&= \sum_{n=0}^{k-1} \gamma_{k,n} \left(\sum_{l>0} \left(\sum_{0=i_0<i_1<\ldots<i_{l-1}<i_l=n} \left(\prod_{j=1}^{l} \gamma_{i_j,i_{j-1}} \right) \right) \right) \beta_0 \\
&= \gamma_{k,0} + \gamma_{k,1}\gamma_{1,0} + \cdots + \sum_{l>0} \left(\sum_{0=i_0<i_1<\ldots<i_{l-1}<i_l=n} \gamma_{k,n}\gamma_{n,i_{l-1}} \cdots \gamma_{i_1,0} \right) \\
&\quad + \cdots + \sum_{l>0} \left(\sum_{0=i_0<i_1<\ldots<i_{l-1}<i_l=k-1} \gamma_{k,k-1}\gamma_{k-1,i_{l-1}}\gamma_{i_{l-1},i_{l-2}} \cdots \gamma_{i_1,0} \right) \\
&= \sum_{l>0} \left(\sum_{0=i_0<i_1<\ldots<i_{l-1}<i_l=k} \left(\prod_{j=1}^{l} \gamma_{i_j,i_{j-1}} \right) \right).
\end{aligned}$$

Therefore, if Equation (21) holds for all β_n, $1 \leq n \leq k-1$, then it also holds for $\beta_n, n = k$.

Now by adding equation (20) to (21), we have:

$$\beta_k = \sum_{l>0} \left(\sum_{0=i_0<i_1<\ldots<i_{l-1}<i_l=k} \frac{y_k}{y_0} \prod_{j=1}^{l} z_{i_j - i_{j-1}}. \right).$$

Further more, using substitution $d_j = i_j - i_{j-1} \geq 1$, where $d_1 + d_2 + \ldots + d_l = k$ we have:

$$\beta_k = \sum_{l>0} \left(\sum_{d_1+d_2+\ldots+d_l=k} \frac{y_k}{y_0} \prod_{i=1}^{l} z_{d_i} \right),$$

where indexes $d_1, d_2, ...$ are taken from $\mathbb{N} = \{1, 2, 3, ...\}$. Thus, for $k \geq 1$,

$$\beta_k = \sum_{l>0} \left(\sum_{d_1+d_2+\ldots+d_l=k} r^k \prod_{i=1}^{l} \frac{-1}{\prod_{j=1}^{d_i}(1-r^j)} \right). \tag{22}$$

Finally (18) implies for $1 \leq i \leq k$,

$$\alpha_{k,i} = \frac{r^{ki}}{\prod_{j=1}^{i}(1-r^j)} \beta_{k-i}.$$

Using (15), we have

$$\begin{aligned} a_{k,m} &= \sum_{i=0}^{k} r^{-mi} \alpha_{k,i} = (\sum_{i=0}^{k-1} r^{-mi} \alpha_{k,i}) + r^{-mk} \alpha_{k,k} \\ &= \alpha_{k,0} + (\sum_{i=1}^{k-1} r^{-mi} \alpha_{k,i}) + r^{-mk} \alpha_{k,k}. \end{aligned}$$

As a result,

$$\begin{aligned} a_{k,m} = &\sum_{l>0} \left(\sum_{d_1+d_2+\ldots+d_l=k} r^k \frac{(-1)^l}{\prod_{i=1}^{l} \prod_{j=1}^{d_i}(1-r^j)} \right) + r^{-mk} \frac{r^{k^2}}{\prod_{j=1}^{k}(1-r^j)} \\ &+ \sum_{i=1}^{k-1} r^{-mi} \frac{r^{ki}}{\prod_{j=1}^{i}(1-r^j)} \sum_{l>0} \left(\sum_{d_1+d_2+\ldots+d_l=k-i} r^{k-i} \frac{(-1)^l}{\prod_{w=1}^{l} \prod_{j=1}^{d_w}(1-r^j)} \right), \end{aligned}$$

with indexes $d_1, d_2, ...$ taken from $\mathbb{N} = \{1, 2, 3, ...\}$. $\square$ $\square$

8.2 Stationary Probabilities under the ASQ Policy.

Proof of Theorem 9. Under the ASQ policy, the rate balance equations can be written as follows:

$$
\begin{aligned}
(\lambda_1+\lambda_2)p_{0,0} &= \lambda_1(1-r)p_{0,1}+\lambda_2(1-r)p_{1,0}, && (23)\\
(\lambda_1+\lambda_2)p_{i,i} &= \lambda_1 r^i p_{i-1,i}+\lambda_2 r^i p_{i,i-1} && (24)\\
&\quad +\lambda_1(1-r^{i+1})p_{i,i+1}+\lambda_2(1-r^{i+1})p_{i+1,i},\ i\geq 1, &&\\
\lambda_2 p_{i+1,i} &= \lambda_1 r^i p_{i,i}+\lambda_1(1-r^{i+1})p_{i+1,i+1},\ i\geq 0, && (25)\\
\lambda_1 p_{i,i+1} &= \lambda_2 r^i p_{i,i}+\lambda_2(1-r^{i+1})p_{i+1,i+1},\ i\geq 0, && (26)\\
\sum_{i=0}^{\infty}\sum_{j=i-1}^{i+1} p_{i,j} &= 1. && (27)
\end{aligned}
$$

The state space has a very special structure where the removal of a state in the form $(i,i), i>0$ disconnects the transition graph. This implies a rate balance for the transitions between states $\{(i,i-1),(i-1,i)\}$ and (i,i), which implies the following detailed balance type equations:

$$
\begin{aligned}
(\lambda_1+\lambda_2)r^i p_{i,i} &= \lambda_1(1-r^{i+1})p_{i,i+1}+\lambda_2(1-r^{i+1})p_{i+1,i},\ i\geq 0, && (28)\\
(\lambda_1+\lambda_2)(1-r^i)p_{i,i} &= \lambda_1 r^i p_{i-1,i}+\lambda_2 r^i p_{i,i-1},\ i\geq 1, && (29)
\end{aligned}
$$

Equation (23) can obtained by setting $i=0$ in (28) and further, summing up (28) and (29) for a given $i\geq 1$ we have (24). Hence, any solution to the set of equations (25)-(29) also solves (23)-(27) and hence should be unique. Further, Equations (25) and (26) imply $\frac{p_{i+1,i}}{p_{i,i+1}}=\frac{\lambda_1^2}{\lambda_2^2}$. Hence, substituting $p_{i+1,i}=\frac{\lambda_1^2}{\lambda_2^2}p_{i,i+1}$ into (28) we obtain

$$
p_{i,i+1}=\frac{\lambda_2}{\lambda_1}\frac{1-r^{i+1}}{r^i}p_{i,i} \text{ and } p_{i+1,i}=\frac{\lambda_1}{\lambda_2}\frac{1-r^{i+1}}{r^i}p_{i,i}. \tag{30}
$$

Then (30) and (29) together imply $p_{i+1,i+1}=\frac{r^i r^{i+1}}{(1-r^{i+1})^2}p_{i,i}, i\geq 0$, and hence, for $i\geq 1$,

$$
p_{i,i}=\prod_{k=1}^{i}\frac{r^k r^{k-1}}{(1-r^k)^2}p_{0,0}=\frac{r^{i^2}p_{0,0}}{\left[\prod_{k=1}^{i}(1-r^k)\right]^2}. \tag{31}
$$

Substituting (31) into (30) and defining $\prod_{k=1}^{0}(1-r^k)=1$, for $i\geq 0$

$$
\begin{aligned}
p_{i,i+1} &= \frac{\lambda_2 r^{i(i+1)}p_{0,0}}{\lambda_1\prod_{k=1}^{i}(1-r^k)\prod_{k=1}^{i+1}(1-r^k)}\\
p_{i+1,i} &= \frac{\lambda_1 r^{i(i+1)}p_{0,0}}{\lambda_2\prod_{k=1}^{i}(1-r^k)\prod_{k=1}^{i+1}(1-r^k)}.
\end{aligned}
$$

Finally the result follows from plugging all $p_{i,j}$ back in (27). □ □

8.3 Insensitivity Proofs for Functional Threshold Policies with $h(x) = x + d$

Lemma 20. *Suppose that the functional threshold policy with a threshold function of the form $h(x) = x + d$, where $d \in \mathbb{N}$ is employed to stabilize a probabiltic matching system. For $(i,j) \in \mathbb{N}^2$, let $p_{i,j}$ be the stationary probability of being at state (i,j). Now, define $a_l = \sum_{i=0}^{\infty} p_{i,i+l}$ and $a_{-l} = \sum_{j=0}^{\infty} p_{j+l,j}$ Then, if $\lambda_1 \neq \lambda_2$,*

$$a_{d+1} = \frac{1 - \frac{\lambda_2}{\lambda_1}}{1 - (\frac{\lambda_2}{\lambda_1})^{2d+3}} \text{ and } a_l = (\frac{\lambda_2}{\lambda_1})^{d+1-l} a_{d+1}, \text{ for } \quad -d-1 \leq l \leq d,$$

and if $\lambda_1 = \lambda_2$, $a_l = \frac{1}{2d+3}$, $-d-1 \leq l \leq d+1$.

Proof. When $q = 1$, the process $\{X^{1,FT}(t) + d + 1, t \geq 0\}$ is stochastically equivalent to an $M/M/1/2d+2$ system and the result follows. When $0 < q < 1$, the state space can be written as $\mathbb{S} = \{(i, i+l) : i \in \mathbb{N}, -d-1 \leq l \leq d+1, i+l \in \mathbb{N}\}$, hence the global balance equations are

$$(\lambda_1 + \lambda_2) p_{0,0} = \lambda_1 (1-r) p_{0,1} + \lambda_2 (1-r) p_{1,0}, \tag{32}$$

$$\begin{aligned}(\lambda_1 + \lambda_2) p_{i,i} = {} & \lambda_1 r^i p_{i-1,i} + \lambda_2 r^i p_{i,i-1} + \lambda_1 (1 - r^{i+1}) p_{i,i+1} \\ & + \lambda_2 (1 - r^{i+1}) p_{i+1,i}, \ i \geq 1,\end{aligned} \tag{33}$$

$$\begin{aligned}(\lambda_1 + \lambda_2) p_{l,0} = {} & \lambda_1 p_{l-1,0} + \lambda_1 (1-r) p_{l+1,1} + \lambda_2 (1 - r^{l+1}) p_{l+1,0}, \\ & 1 \leq l \leq d\end{aligned} \tag{34}$$

$$\begin{aligned}(\lambda_1 + \lambda_2) p_{i+l,i} = {} & \lambda_1 r^i p_{i+l-1,i} + \lambda_1 (1 - r^{i+1}) p_{i+l,i+1} + \lambda_2 r^{i+l} p_{i+l,i-1} \\ & + \lambda_2 (1 - r^{i+l+1}) p_{i+l+1,i}, \ i \geq 1, 1 \leq l \leq d,\end{aligned} \tag{35}$$

$$\begin{aligned}(\lambda_1 + \lambda_2) p_{0,l} = {} & \lambda_2 p_{0,l-1} + \lambda_2 (1-r) p_{1,l+1} + \lambda_1 (1 - r^{l+1}) p_{0,l+1}, \\ & 1 \leq l \leq d\end{aligned} \tag{36}$$

$$\begin{aligned}(\lambda_1 + \lambda_2) p_{i,i+l} = {} & \lambda_2 r^i p_{i,i+l-1} + \lambda_2 (1 - r^{i+1}) p_{i+1,i+l} + \lambda_1 r^{i+l} p_{i-1,i+l} \\ & + \lambda_1 (1 - r^{i+l+1}) p_{i,i+l+1}, \ i \geq 1, 1 \leq l \leq d,\end{aligned} \tag{37}$$

$$\lambda_2 p_{i+d+1,i} = \lambda_1 r^i p_{i+d,i} + \lambda_1 (1 - r^{i+1}) p_{i+d+1,i+1}, \ i \geq 0, \tag{38}$$

$$\lambda_1 p_{i,i+d+1} = \lambda_2 r^i p_{i,i+d} + \lambda_2 (1 - r^{i+1}) p_{i+1,i+d+1}, \ i \geq 0, \tag{39}$$

$$\sum_{i=0}^{\infty} \sum_{j=i-d-1}^{i+d+1} p_{i,j} = 1. \tag{40}$$

We sum (33) for $i = 1$ to ∞ and then add (32) to get

$$(\lambda_1 + \lambda_2) a_0 = \lambda_1 a_{-1} + \lambda_2 a_1. \tag{41}$$

Repeating the same procedures for pairs (34) and (35), (36) and (37), (38) and (39),

$$(\lambda_1 + \lambda_2) a_l = \lambda_2 a_{l+1} + \lambda_1 a_{l-1}, \ 1 \leq l \leq d, \tag{42}$$

$$(\lambda_1 + \lambda_2) a_{-l} = \lambda_1 a_{-l-1} + \lambda_2 a_{-l+1}, \ 1 \leq l \leq d, \tag{43}$$

$$\lambda_2 a_{d+1} = \lambda_1 a_d, \tag{44}$$

$$\lambda_1 a_{-d-1} = \lambda_2 a_{-d}. \tag{45}$$

We notice that, similar to the case $q = 1$, if we replace $b_l = a_{l-d}$ in (41)-(45), we obtain the global balance equations of an $M/M/1/2d+2$ system. Hence, the result follows. □ □

Proof of Theorem 12. Using PASTA property, $c_1 = a_{d+1}$ and the result follows from Lemma 20. □ □

Proof of Theorem 14. Without lost of generality, assume $d \geq 0$ is an integer. The difference of average queue lengths can be written as

$$\begin{aligned} L_1 - L_2 &= \sum_{l=-d-1}^{d+1} l a_l \\ &= \frac{1-\frac{\lambda_2}{\lambda_1}}{1-(\frac{\lambda_2}{\lambda_1})^{2d+3}} (-\frac{\lambda_2}{\lambda_1})^{d+2} \sum_{l=-d-1}^{d+1} (-l)(\frac{\lambda_2}{\lambda_1})^{-l-1}. \end{aligned}$$

Using $\rho = \frac{\lambda_2}{\lambda_1}$,

$$\begin{aligned} \sum_{l=-d-1}^{d+1} (-l)(\frac{\lambda_2}{\lambda_1})^{-l-1} &= \sum_{l=-d-1}^{d+1} \frac{\partial}{\partial \rho} \rho^{-l} \\ &= \frac{\partial}{\partial \rho} \frac{\rho^{d+1}(1-\rho^{-2d-3})}{1-\rho^{-1}} \\ &= \frac{((d+2)\rho^{d+1} + (d+1)\rho^{-d-2})(\rho-1) - (\rho^{d+2} - \rho^{-d-1})}{(\rho-1)^2} \end{aligned}$$

Hence,

$$L_1 - L_2 = \frac{(d+2)\rho^{2d+3} + d + 1}{1-\rho^{2d+3}} + \frac{(1-\rho)\rho^{d+2}(\rho^{d+2} - \rho^{-d-1})}{(1-\rho^{2d+3})(\rho-1)^2}$$

□ □

References

[1] Adan, I., Weiss, G.: Exact FCFS matching rates for two infinite muti-type sequences. Operations Research **60**(2), 475–489 (2012)

[2] Asmussen, S.: Applied Probability and Queues. Applications of Mathematics Series. Springer, New York (2003)

[3] Bušić, A., Gupta, V., Mairesse, J.: Stability of the bipartite matching model. Advances in Applied Probability **45**(2), 351–378 (2013)

[4] Caldentey, R., Kaplan, E., Weiss, G.: FCFS infinite bipartite matching of servers and customers. Advances in Applied Probability **41**(3), 695–730 (2009)

[5] Çınlar, E.: Probability and Stochastics. Graduate Texts in Mathematics. Springer, New York (2011)

[6] El-Taha, M., Stidham, S.: Sample-Path Analysis of Queueing Systems. Kluwer Academic Publishers, New York (1999)

[7] Fayolle, G., Malyshev, V.A., Menshikov, M.V.: Topics in the Constructive Theory of Countable Markov Chains. Cambridge Univ. Press, Cambridge, UK (1995)

[8] Gross, D., Harris, C.M.: Fundamantals of Queueing Theory. Wiley Series in Probability and Statistics. John Wiley & Sons, New York (1998)

[9] Gurvich, I., Ward, A.: On the dynamic control of matching queues. (2013). Working paper.

[10] Harrison, J.M.: Assembly-like queues. J. Appl. Probab. **10**(2), 354–367 (1973)

[11] Kashyap, B.R.K.: The double-ended queue with bulk service and limited waiting space. Operations Research **14**(5), 822–834 (1966)

[12] Kulkarni, V.G.: Modeling and Analysis of Stochastic System. Texts in Statistical Science. Chapman and Hall/CRC, London, UK (1996)

[13] Latouche, G.: Queues with paired customers. J. Appl. Probab. **18**(3), 684–696 (1981)

[14] Liu, X., Gong, Q., Kulkarni, V.G.: Diffusion models for double-ended queues with renewal arrival processes (2014). Working Paper, arXiv:1401.5146