On occupancy based randomized routing schemes in large systems of shared servers

Thirupathaiah Vasantam
Electrical and Computer Engineering
University of Waterloo, Canada
Email: tvasanta@uwaterloo.ca

Ravi R. Mazumdar
Electrical and Computer Engineering
University of Waterloo, Canada,
Email: mazum@uwaterloo.ca

Abstract*—Recently there has been a great interest in randomized load balancing schemes for large systems of parallel servers. Various policies such as SQ(d) (shortest queue amongst d randomly sampled), threshold based policies etc. have been studied via a mean-field approach. All these policies are special cases of occupancy based routing decisions. In this paper we present a unified mean-field approach that holds for any routing scheme that only depends on the server occupancy in a system with a large numbers of processor sharing servers as an archetype of shared resource systems. The mean-field equations we obtain hold for general job length distributions unlike most recent works that assume exponentially distributed job lengths. We then show that the probability measure of occupancy defined on the set of non-negative integers $\mathcal{Z}_+$ obtained from a fixed-point of the mean-field also satisfies the stationary mean-field equations under the assumption that the job lengths are exponential with the same average length. If the mean-field under exponential case has unique fixed-point, then the fixed point is *insensitive* to the job length distribution. The approach is via a measure-valued Markov process approach.

I. Introduction

With the emergence of large cloud computing infrastructure and the more recent fog computing paradigm the issues of matching jobs to resources to minimize the turn around or latency is becoming very crucial. In typical settings there could be 100,000 servers and as it is well known from the classical literature, latency is minimized by sending jobs to the servers that are most lightly loaded. This is referred to as Join the Shortest Queue (JSQ) scheme [1]. In the 1990's in seminal work by Vvedenskaya *et al* [2] and Mitzenmacher [3], it is shown that the delay performance of JSQ routing in a large-scale system with N parallel servers under Poisson job arrival process with rate $N\lambda$ is closely approximated by a randomized scheme that selects the server with the shortest queue among $d \geq 2-$ uniformly sampled (out of N) queues. Furthermore, the most improvement over uniform routing is obtained when $d = 2$ leading to the term "The Power of Two". The analysis was based on approximating the large system by its mean-field limit carried out assuming that the service time distributions are exponential with mean $\frac{1}{\mu}$ ($\lambda < \mu$). One of the key observations was that in the mean-field limit any finite set of queues could be considered statistically independent.

Subsequent work [4]–[6] extended the mean-field analysis for SQ(d) routing to loss system models and processor sharing (PS) models because of their importance in modeling Amazon EC2 and Microsoft Azure system architectures. All of these works assumed the job lengths to be exponentially distributed. In [4]–[6], it was shown that the mean-field analysis could be extended to heterogeneous systems, where it was observed through simulations that the insensitivity property in the case of loss and PS models holds when number of servers $N \to \infty$. The insensitivity property means that the stationary distribution depends only on the average service time. Although it is known in the classical context of these models, but it requires proof in this context because the underlying Markov process is not a standard birth-death type process but a nonlinear Markov process whose arrival rates themselves depend on the distribution. This question has been recently settled for loss models in [7] and for PS models in [8].

Recently there has been a flurry of activity on randomized routing schemes that are now referred to as PUSH type and PULL type. In PUSH type, the dispatcher selects a server from a set of d randomly selected servers as the destination server by exploiting their occupancy information. On the otherhand, in PULL type [9]–[11], servers notify the dispatcher about their state subject to some occupancy constraints. A typical PULL type strategy is the Join Idle Queue (JIQ) in which idle servers notify the dispatcher and an incoming request is now uniformly matched to an idle server in the list at the dispatcher. If there are no idle servers then the destination server is selected uniformly at random from N servers. Of course, in lightly loaded systems the number of idle servers is $O(N)$. For lightly loaded systems JIQ will out perform SQ(d). However as d increases the SQ(d) scheme in the limit behaves as the JSQ or even the JIQ scheme [11], [12]. Realizing that all such schemes are just routing schemes based on the queue or server occupancy state at the arrival of a request it is natural to seek a mean-field analysis for general occupancy based routing strategies that subsume all these different policies.

In this paper we restrict our attention to PUSH type of schemes that are based on a random sampling of a finite number of servers and the destination server is chosen from this set by using the occupancy information of the sampled servers. This routing decision is reflected as a mapping from the server states to probabilities, that will be stated later precisely in the paper. Finally, we demonstrate our conclusions by using two occupancy based policies such as SQ(d) and a threshold type. In the threshold type, requests are matched

978-0-9883045-5-0/18/$31.00 ©2018 ITC
DOI 10.1109/ITC30.2018.00013

uniformly to servers below a threshold occupancy amongst the sampled servers if there are any available, otherwise a standard JSQ policy is followed. Although a little reflection shows that the SQ(d) policy will be better in terms of minimizing the average response time but other policies that have less computational cost, energy consumption than the SQ(d), but result in almost same performance as the SQ(d) are worth studying. For example, in [3], a threshold based policy is studied in which a sampled server is chosen as the destination server if its occupancy is less than or equal to the chosen threshold value otherwise another server is sampled and the same procedure is repeated until d servers are sampled. If all the d sampled servers have occupancy greater than the threshold value, then the shortest queue is chosen as the destination server. This policy results in descent system performance depending on the threshold but with less computational cost than the SQ(d) policy. Recently, for better energy usage, in [13], the SQ(d) policy is investigated for a large-scale multi-server system in which a server is turned off when it becomes idle and turned on again when its occupancy reaches a pre-specified threshold value. The policy in [13] can be improved further for better energy usage by choosing a shortest queue among working sampled servers with number of jobs less than the threshold value if there are any available as the destination server, otherwise the shortest queue over all the d sampled servers can be chosen as the destination server. Therefore it is worth studying randomized load balancing schemes under general framework.

Since in practice the exponential job length hypothesis is not valid, for example, the distributions are Log-normal in call centers [14], and Gamma distributed in automatic teller machines (ATMs) [15] etc., we study the case of general job length distributions. The particular model we consider in this paper is the shared server model, where servers employ PS. The PS model where the processing speed of a server is equally shared by all the progressing jobs well approximates the round robin schemes used in practice [16]. There is another reason why the PS model is interesting. It is known to be insensitive under random thinning and so showing insensitivity of the mean-field under very general occupancy based randomized routing can then lead us to conclude that insensitivity is a generic property of the PS service discipline rather than the precise input to a queue.

In [17], the SQ(d) policy is investigated for PS, FCFS, and LIFO-PR systems with general service time distributions under the assumption of asymptotic independence of any finite set of servers and the existence of unique limiting stationary distribution as stated in an *ansatz*. Due to the assumption of asymptotic independence of servers for PS models, insensitivity immediately follows from reversibility in the limiting system since arrival process to each individual server is a Poisson process which is a necessary condition to have insensitivity. So far the proof of *ansatz* remains an open problem except for FCFS models with decreasing hazard rate functions proved in [18]. The proof is based on montonic behavior of the system in this case. As stated in [18], the proof techniques cannot be extended to processor sharing models as we do not have ordering between states unlike FCFS systems.

For queuing systems with general service times, to Markovianize the system behavior, we need to consider the occupancy state as well as the age or residual service times of the jobs in service. This results in a bi-dimensional Markov process defined on $Z_+ \times \Re_+$ for FCFS systems. In the processor sharing case the Markov process is defined on $\bigcup_{n\in Z_+}(n, \Re_+^n)$ since we need to track both occupancy and age of each progressing job. The analysis is much more difficult because the ages do not increase linearly at the rate of processing speed of the server but the rate depends on the server occupancy or the first coordinate of the state. This is also a mathematical reason why the PS model is interesting.

Recently, FCFS systems with generally distributed service times have been revisited in [19] under the SQ(d) policy via a mean-field approach. They established the mean-field limit by considering ages of jobs. However, the analysis is only restricted to the finite time or transient case and no results are given for the stationary case.

Main contributions: For general class of occupancy based randomized load balancing policies, we make following contributions: Let $(\overline{\nu}_t^N, t \geq 0)$ be a measure-valued Markov process that tracks the fraction of servers lying in each possible server state. We prove that $(\overline{\nu}_t^N, t \geq 0)$ converges in distribution to a unique deterministic measure-valued process $(\overline{\nu}_t, t \geq 0)$ referred to as the mean-field limit. As a consequence of this, we show asymptotic independence of servers (propagation of chaos) for any compact interval of time. In particular, we show that $\overline{\nu}_t$ represents the server state probability distribution for a server in the limiting system (system with $N \to \infty$). We then observe that the mean-field equations (MFEs) represent dynamics of a classical single server system with Poisson job arrival process having intensity $\lambda\Phi_n(\overline{\nu}_t)$ when there are n progressing jobs at time t and current distribution is $\overline{\nu}_t$. The function $\Phi_n(\cdot)$ is defined later in Section III. We then show that every fixed-point of the partial differential equations (PDEs) that describe the mean-field limit corresponds to a probability measure for occupancy on $\mathcal{Z}_+$ is also a fixed point of the MFEs under exponential distributions having same mean. Finally, we present two occupancy based policies and provide simulation results.

Organization of the paper: The rest of the paper is organized as follows: We first give preliminaries in Section II where we introduce system model, the class of occupancy based randomized load balancing policies, the required notation, and a Markovian description of the system by using a state descriptor. We then present main results of this paper in Section III where we also discuss two occupancy based policies SQ(d) and a threshold based policy. The proofs of results are given in Appendix. The simulation results that support insensitivity of the fixed-point of the mean-field limit under considered occupancy based policies are given in Section IV. Finally, we conclude in Section V with a discussion on future work.

II. Preliminaries

In this section, we introduce the system model, required notation, and then a system state descriptor that we use to model the system evolution by a Markov process.

System Model: We first introduce the system model and the class of occupancy (number of progressing jobs) information based routing policies.

Consider a large-scale processor sharing system containing N identical servers each with capacity of unit processing rate. Therefore, from processor sharing service discipline, a job that is in progress at a server with n jobs is processed at the rate of $\frac{1}{n}$. We assume that the job arrival process is a Poisson process with rate $N\lambda$. Furthermore, there is a central job dispatcher that routes each incoming job to one of the servers according to a predefined routing policy. It makes the routing decision based on the occupancy information of d randomly chosen servers with replacement and one of the sampled servers is chosen as the destination server. We denote this class of routing policies by OB(d) policies.

Definition 1. *The OB(d) Policies:*
Upon an arrival, a set of d servers are sampled uniformly at random[1] *referred to as the potential destination servers. Let $(l_1, \cdots, l_d)$, $(n_1, \cdots, n_d)$ be a sampled potential servers list with corresponding number of jobs where l_i denotes the i^{th} sampled potential server with number of progressing jobs equal to n_i. Further, let $\mathcal{Z}_+$ be the set of non-negative integers and let $\mathcal{M}_1(\mathcal{Z}_+^d)$ be the set of all probability measures on $\mathcal{Z}_+^d$. Any policy $\Lambda \in OB(d)$ is a mapping $\Lambda : \mathcal{Z}_+^d \mapsto \mathcal{M}_1(\mathcal{Z}_+^d)$ such that*

$$\Lambda((n_1, \cdots, n_d)) = (p_1, \cdots, p_d). \tag{1}$$

The term p_i denotes the probability with which i^{th} sampled server having n_i jobs is chosen as the destination server when $(n_1, \cdots, n_d)$ is the vector of number of jobs at the potential servers. Further, the following two properties hold:

- *If $n_i = n_j$, then*
$$p_i = p_j, \tag{2}$$
that is, if two potential servers have same number of progressing jobs, then the probabilities with which they are chosen as the destination servers are same.
- *If σ is a permutation on $(n_1, \cdots, n_d)$, then*
$$\Lambda(\sigma((n_1, \cdots, n_d))) = \sigma((p_1, \cdots, p_d)). \tag{3}$$

The job lengths are generally distributed with cumulative distribution function $G(\cdot)$ and density function $g(\cdot)$, respectively. Further, we assume that the average job length is equal to $\frac{1}{\mu}$. The hazard rate function of the job length distribution is denoted by $\beta(\cdot) = \frac{g(\cdot)}{\bar{G}(\cdot)}$. Also, we must have $\lambda < \mu$ in order for the system to be in stable region.

Notation: Let $\mathcal{Z}, \mathcal{R}, \mathcal{Z}_+$, and $\mathcal{R}_+$ be the set of integers, real numbers, non-negative integers, and non-negative real numbers, respectively. For any given metric space U, we denote the space of bounded measurable functions, bounded continuous functions, and continuous functions with compact support by $\mathcal{K}_b(U), \mathcal{C}_b(U)$, and $\mathcal{C}_k(U)$, respectively. Further, let $\mathcal{C}^1(U), \mathcal{C}_b^1(U)$, and $\mathcal{C}_k^1(U)$ be the space of continuously differentiable, continuously differentiable with bounded first derivatives, and continuously differentiable with compact support functions, respectively.

Let $\|f\|$ for $f \in \mathcal{K}_b(U)$ be defined as

$$\|f\| = \sup_{x \in U} |f(x)|. \tag{4}$$

The space $\mathcal{C}_b(U)$ is equipped with the topology induced by the norm $\|\cdot\|$, i.e., the sequence $\{f_n, n \geq 1\} \in \mathcal{C}_b(U)$ converges to $f \in \mathcal{C}_b(U)$ if $\|f_n - f\|$ as $n \to \infty$. Let us define a norm $\|\cdot\|_1$ on $\mathcal{C}_b^1(U)$ as follows:

$$\|f\|_1 = \|f\| + \|f'\|, \tag{5}$$

where $f'$is the first derivative of $f \in \mathcal{C}_b^1(U)$. We then assume that the space $\mathcal{C}_b^1(U)$ is associated with the topology induced by the norm $\|\cdot\|_1$. Further, for any $f \in \mathcal{C}_b^1(\mathcal{R}_+^n)$, let f_Σ^1 be defined as

$$f_\Sigma^1(x_1, \cdots, x_n) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial f(x_1, \cdots, x_n)}{\partial x_i}. \tag{6}$$

For given metric space U, let the Borel σ-algebra be denoted by $\mathcal{B}(U)$. Further, let $\mathcal{M}_F(U)$ and $\mathcal{M}_1(U)$ be the space of finite non-negative and probability measures defined on the space U, respectively. We denote the measure value of the set $B \in \mathcal{B}(U)$ *w.r.t.* the measure $\eta \in \mathcal{M}_F(U)$ by $\eta(B)$ and let $\eta(\{x\})$ be the measure value at an element $x \in U$. Also, let δ_x be the Dirac measure with unit mass at $x \in U$. We now define a set of probability measures $\mathcal{M}_1^N(U)$ as follows

$$\mathcal{M}_1^N(U) = \{\nu \in \mathcal{M}_1(U) : N\nu(B) \in \mathcal{Z}_+, \forall B \in \mathcal{B}(U)\}. \tag{7}$$

We next define for any $\phi \in \mathcal{C}_b(U)$, $\nu \in \mathcal{M}_F(U)$,

$$\langle \nu, \phi \rangle = \int_{y \in U} \phi(y)\, d\nu(y). \tag{8}$$

We assume that the space $\mathcal{M}_F(U)$ is equipped with the weak topology, *i.e.*, a sequence $\{\nu_n\}$ in $\mathcal{M}_F(U)$ is said to converge weakly to $\nu \in \mathcal{M}_F(U)$ if and only if $\langle \nu_n, \phi \rangle \to \langle \nu, \phi \rangle$ as $n \to \infty$ for all $\phi \in \mathcal{C}_b(U)$. We recall that the space $\mathcal{M}_F(U)$ equipped with the weak topology a Polish space when U is a Polish space.

In order to describe the system evolution, we now introduce the following notation. We consider that the state of a server is given by $(n, a_1, \cdots, a_n)$ where n denotes the number of progressing jobs and a_i denotes the age of the i^{th} progressing job. The age of a progressing is equal to the amount of service received since its arrival and hence, if a job arrives at time t, then its age a at time $t + h$ is given by

$$a = \int_{s=t}^{t+h} \frac{1}{k(s)} ds, \tag{9}$$

[1]As $N \to \infty$, sampling potential servers with or without replacement would yield same mean-field limit and hence, we use sampling with replacement to simplify the notation.

where $k(s)$ denotes the instantaneous number of progressing jobs at the server at which the job is processed at time s. Let E_n be the set of all possible states of a server when there are n progressing jobs, *i.e.*,

$$E_n = \{(n, a_1, \cdots, a_n) : a_i \in \mathcal{R}_+, 1 \leq i \leq n\}. \tag{10}$$

Further, when server has no progressing jobs, the server state lies in the set

$$E_0 = \{0\}. \tag{11}$$

Hence, the state of a server lies in the set

$$E = \cup_{n \geq 0} E_n. \tag{12}$$

Without loss of generality, an element of the form $(n, u_1, \cdots, u_n) \in E$ for $n \geq 0$ is denoted by $\underline{u}$. For $\underline{u} = (n, u_1, \cdots, u_n)$ and $\underline{v} = (m, v_1, \cdots, v_m)$, we now define the metric

$$d_E(\underline{u}, \underline{v}) = \begin{cases} \sum_{i=1}^{n} |u_i - v_i| & \text{if } n = m \\ \infty & \text{otherwise.} \end{cases} \tag{13}$$

We say that a sequence $\{\underline{u}_n, n \geq 1\}$ in E converges to $\underline{u} \in E$ if $\lim_{n\to\infty} d_E(\underline{u}_n, \underline{u}) = 0$.

For any Borel set $B \in \mathcal{B}(E)$, the indicator function of B is defined as

$$I_{\{B\}}(\underline{u}) = \begin{cases} 1 & \text{if } \underline{u} \in B \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

For any measure $\nu \in \mathcal{M}_F(E)$ restricted to the space E_0 is a Dirac measure with mass at (0). Further, for $n \geq 1$, we say that the measure $\nu \in \mathcal{M}_F(E)$ is absolutely continuous at $\underline{x} \in E_n$ if $\nu(\{\underline{x}\}) = 0$ and the measure ν is called absolutely continuous with respect to Lebesgue measure if $\nu(\{\underline{y}\}) = 0$ for all $\underline{y} \in E_n$, $n \geq 1$. Any function $f : E \mapsto \mathcal{R}$ is said to be differentiable if for every $i \geq 1$, $\frac{\partial f(i, x_1, \cdots, x_i)}{\partial x_j}$ is defined for every $1 \leq j \leq i$ at every $(x_1, \cdots, x_i) \in \mathcal{R}_+^i$. Hence a function of the type $f = I_{\{E_n\}}$, for $n \geq 1$, that we use frequently in our analysis is differentiable.

We next define the following two functions that are used in establishing the mean-field limit. The first function $\Xi : E \mapsto \mathcal{R}$ is defined as

$$\Xi(n, x_1, \cdots, x_n) = n, \tag{15}$$

for $(n, x_1, \cdots, x_n) \in E$. The second function, $\Upsilon : E \mapsto \mathcal{R}$ is defined as

$$\Upsilon(n, x_1, \cdots, x_n) = \begin{cases} 0 & \text{for } n = 0, \\ x_1 + \cdots + x_n & \text{otherwise.} \end{cases} \tag{16}$$

In describing the time evolution of the system, we use the following transition operators on functions and measures. For $b > 0$, let us define an operator $\tau_b^+ : E \mapsto E$ as

$$\tau_b^+(n, u_1, \ldots, u_n) = \begin{cases} 0 & \text{if } n = 0, \\ (n, v_1, \cdots, v_n) & \text{otherwise,} \end{cases} \tag{17}$$

where $v_i = u_i + \frac{b}{n}$, for $1 \leq i \leq n$. Further, for any $b > 0, f \in \mathcal{K}_b(E)$, we define a mapping

$$\tau_b : \mathcal{K}_b(E) \to \mathcal{K}_b(E) \tag{18}$$

satisfying

$$\tau_b f(\underline{u}) = f(\tau_b^+ \underline{u}), \tag{19}$$

for $\underline{u} \in E$. For $b > 0$, let $\tau_b \nu \in \mathcal{M}_F(E)$ be the shifted measure defined such that for any Borel set $B \in \mathcal{B}(E)$, we have

$$\tau_b \nu(B) = \nu(\tau_b^+(B)). \tag{20}$$

For $\nu \in \mathcal{M}_F(E)$, the measure $\tau_b \nu \in \mathcal{M}_F(E)$ satisfies

$$\langle \tau_b \nu, f \rangle = \langle \nu, \tau_b f \rangle \tag{21}$$

for all $f \in \mathcal{K}_b(E)$. The existence of the unique measure $\tau_b \nu$ satisfying equation (21) follows from the Riesz-Markov-Kakutani theorem [20]. For the measure-valued Markov process $(\overline{\nu}_t^N, t \geq 0)$ that describes the system evolution, we can derive the expression of the generator of the process $(\overline{\nu}_t^N, t \geq 0)$ using equation (21). By using equation (21), the change in the process $(\overline{\nu}_t^N, t \geq 0)$ in a given small time interval can be considered as a change in the function f. Based on this idea, with the help of the class of functions of the type $\overline{\nu}_t^N \mapsto \langle \overline{\nu}_t^N, \phi \rangle$ for $\phi \in \mathcal{C}_b^1(E)$, we obtain the generator of the Markov process $(\overline{\nu}_t^N, t \geq 0)$.

For $\nu \in \mathcal{M}_F(E)$, $\langle \nu, \phi \rangle$ is a continuous linear operator on the space of functions $\phi \in \mathcal{C}_b(E)$, we define

$$\|\nu\| = \sup_{\phi \in \mathcal{C}_b(E)} \frac{|\langle \nu, \phi \rangle|}{\|\phi\|}. \tag{22}$$

Let $\mathcal{D}_{\mathcal{H}}([0,T]), \mathcal{D}_{\mathcal{H}}([0,\infty))$ be the càdlàg[2] functions that take values in Polish space $\mathcal{H}$ defined on $[0,T]$, $[0,\infty)$, respectively. The space of the continuous functions that take values in $\mathcal{H}$ defined on $[0,T]$, $[0,\infty)$ are denoted by $\mathcal{C}_{\mathcal{H}}([0,T]), \mathcal{C}_{\mathcal{H}}([0,\infty))$, respectively. The spaces $\mathcal{D}_{\mathcal{H}}([0,T]), \mathcal{D}_{\mathcal{H}}([0,\infty))$ are equipped with the Skorokhod J_1-topology and in that case, they are Polish spaces. For two local martingales $(M_t^1, t \geq 0)$ and $(M_t^2, t \geq 0)$, the covariation and quadratic variation in $\mathcal{D}_{\mathcal{R}}([0,T])$ are denoted by $(< M_\cdot^1, M_\cdot^2 >_t, t \geq 0)$, $(< M_\cdot^1 >_t, t \geq 0) = (< M_\cdot^1, M_\cdot^1 >_t, t \geq 0)$, respectively.

In this paper, we generally work with $\mathcal{H}-$valued stochastic processes where $\mathcal{H} = \mathcal{M}_F(E)$. Further, the considered stochastic processes are random elements defined on $(\Omega, \mathbb{F}, \mathcal{P})$ with sample paths lying in $\mathcal{D}_{\mathcal{H}}([0,\infty))$. The space $\mathcal{D}_{\mathcal{H}}([0,\infty))$ is equipped with the Borel $\sigma-$algebra generated by the open sets under the Skorokhod J_1- topology [21]. A sequence $\{X_n\}$ of $\mathcal{H}$-valued càdlàg processes defined on $(\Omega_n, \mathbb{F}_n, \mathcal{P}_n)$ is said to converge in distribution to a $\mathcal{H}$-valued càdlàg process X defined on $(\Omega, \mathbb{F}, \mathcal{P})$ if, for every bounded, continuous, real valued functional $F : \mathcal{D}_{\mathcal{H}} : [0,\infty) \to \mathcal{R}$, we have

$$\lim_{n\to\infty} \mathbb{E}_n(F(X_n)) = \mathbb{E}(F(X)) \tag{23}$$

[2]Also referred to as RCLL (right continuous with left limits).

where the expectation operators $\mathbb{E}_n, \mathbb{E}$ are defined with respect to $\mathcal{P}_n, \mathcal{P}$, respectively. The convergence of $\{X_n\}$ in distribution to X is denoted by $X_n \Rightarrow X$. We use $\mathcal{L}(X)$ to denote the Law of a random variable X.

Markovian Description: We now describe the system evolution by introducing a state descriptor. For this purpose, we incorporate age of each progressing job at each server into state descriptor in modeling the system evolution as an evolution of a Markov process. If a server has n progressing jobs with i^{th} job having age $a_i, 1 \leq i \leq n$ at time t, then we say that it lies in state $(n, a_1, \cdots, a_n)$ at time t. If a server has no job at time t, then the server state is equal to (0). Once we define a server state, we now define a system state descriptor that we use in describing the system evolution.

Definition 2. *System State Descriptor:*
Let $s_t^i = (n_i, a_{1,i}, a_{2,i}, \cdots, a_{n_i,i})$ be the state of i^{th} server at time t, then we define the system state at time t as follows

$$\nu_t^N = \sum_{i=1}^{N} \delta_{(s_t^i)} \tag{24}$$

where $\delta_{(n,x_1,\cdots,x_n)}$ denotes the Dirac measure with unit mass at $(n, x_1, \cdots, x_n)$. Therefore $\nu_t^N(\{(m, z_1, \cdots, z_m)\})$ is equal to the number of servers with state $(m, z_1, \cdots, z_m)$ at time t.

Once we define the state descriptor, we now look at the time evolution of the the measure-valued process $(\nu_t^N, t \geq 0)$. For $h > 0$, given ν_t^N, to know the value of ν_{t+h}^N, we need to track how each server state changes. In the time interval $(t, t+h]$, either there can be no event (arrival or departure) or some events(arrivals and departures) can occur in the system. In our analysis, by assuming h is a small value, we consider that multiple events do not occur in the interval $(t, t+h]$. Furthermore, we can model departure events by using hazard rate function $\beta(x) = \frac{g(x)}{\bar{G}(x)}$ that defines the instantaneous rate of departure of a job conditioned on its age value equal to x. Precisely, if a job achieves age x at time t, then it departs in the interval $(t, t+y]$ with probability given by $\frac{G(x+y)-G(x)}{\bar{G}(x)} = \beta(x)y + o(y)$. When the number of progressing jobs does not change at a server with state say $(n, a_1, \cdots, a_n)$ at time t in the interval $(t, t+h]$, then its state becomes $\tau_h^+(n, a_1, \cdots, a_n)$ at time $t+h$.

III. Main Results

In this section, we present main results of this paper and the proofs are given in Appendix. Due to space constraints, we only give rough sketch of the proofs. In the rest of the paper, we consider a policy $\Lambda \in OB(d)$, according to which, let $q_{(n_1,\cdots,n_{d-j})}^{(m)}$ for $1 \leq j \leq d$ be the probability that a particular potential server with m jobs is chosen as the destination server when there are j potential servers with m jobs and n_k is the number of progressing jobs at the potential server with k^{th} smallest index (by index b we mean b^{th} potential server) among the $d-j$ potential servers whose number of progressing jobs is not equal to m.

In the following Lemma, we first provide expression for the probability of choosing a server with state $(m, x_1, \cdots, x_m)$ as destination server upon an arrival under a OB(d) routing policy Λ.

Lemma 1. *Suppose $\nu_t^N = \eta$, let $Q_n(\frac{\eta}{N}) = \frac{\eta(\{E_n\})}{N}$ represents the fraction of servers with n jobs at time t. Then according to policy Λ, if a job arrives at time t, then the probability that the destination server lies in state $(m, x_1, \cdots, x_m)$ is equal to*

$$p_r(m, x_1, \cdots, x_m; \eta) = \frac{\eta\{(m, x_1, \cdots, x_m)\}}{N} \Phi_n\left(\frac{\eta}{N}\right), \tag{25}$$

where

$$\Phi_n\left(\frac{\eta}{N}\right) = \sum_{j=1}^{d} \sum_{(n_1,\cdots,n_{(d-j)})} j q_{(n_1,\cdots,n_{d-j})}^{(m)} \binom{d}{j} \times \left(Q_m\left(\frac{\eta}{N}\right)\right)^{j-1} \prod_{i=1}^{d-j} Q_{n_i}\left(\frac{\eta}{N}\right). \tag{26}$$

The proof is given in Appendix A.

By using the filtration defined as

$$\mathcal{F}_t^N = \sigma(\nu_s^N(B) : s \leq t, B \in \mathcal{B}(E)), \tag{27}$$

we now construct a Martingale process with the help of Dynkin's formula. We use this martingale process in establishing the convergence of $(\frac{\nu_t^N}{N}, t \geq 0)$. The generator of the Markov process $(\nu_t^N, t \geq 0)$ is denoted by $A^N(\cdot)$.

Proposition 1. *The process $(\nu_t^N, t \geq 0)$ is a Feller-Dynkin process in $\mathcal{D}_{\mathcal{M}_F(E)}([0, \infty))$. Let $\phi \in \mathcal{C}_b^1(E)$, then the process $(M_t^N(\phi), t \geq 0)$ defined as*

$$M_t^N(\phi) = \langle \nu_t^N, \phi \rangle - \langle \nu_0^N, \phi \rangle - \int_{s=0}^{t} A^N \langle \nu_s^N, \phi \rangle ds \tag{28}$$

is a square integrable $\mathcal{F}_t^N$-martingale and it is right continuous with left limits (RCLL) process.

The proof follows mutatis mutandis from the proof of Propositions 1 and 2 of [22].

Our main aim in this paper is to obtain the limit of the normalized process $(\bar{\nu}_t^N, t \geq 0)$ defined as

$$\bar{\nu}_t^N = \frac{\nu_t^N}{N} \tag{29}$$

when $N \to \infty$. We make following assumptions in the rest of the paper in our analysis.

Assumption 1. *The hazard rate function $\beta(\cdot)$ satisfies*

$$\beta \in \mathcal{C}_b(\mathcal{R}_+) \text{ and } \|\beta\| < \infty. \tag{30}$$

Assumption 2. *The sequence of initial measures of the normalized measure-valued processes $(\bar{\nu}_t^N, t \geq 0)$ satisfy*

$$(\bar{\nu}_0^N, \langle \bar{\nu}_0^N, \Xi \rangle, \langle \bar{\nu}_0^N, \Upsilon \rangle) \Rightarrow (\bar{\nu}_0, \langle \bar{\nu}_0, \Xi \rangle, \langle \bar{\nu}_0, \Upsilon \rangle) \tag{31}$$

where $\bar{\nu}_0 \in \mathcal{M}_1(E)$ is absolutely continuous measure satisfying $\langle \bar{\nu}_0, \Xi \rangle < \infty$ and $\langle \bar{\nu}_0, \Upsilon \rangle < \infty$.

We next state our main result of the paper.

Theorem 1. *Suppose the sequence $\{\overline{\nu}_0^N\}$ satisfies the assumption 2, then for every $T>0$, $(\overline{\nu}_t^N, 0\le t\le T) \Rightarrow (\overline{\nu}_t, 0\le t\le T)$ where the process $(\overline{\nu}_t, 0\le t\le T)$ is the unique solution of the following equation with the initial point $\overline{\nu}_0\in\mathcal{M}_1(E)$, for all $\phi\in\mathcal{C}_b(E)$, we have*

$$\begin{aligned}\langle \overline{\nu}_t,\phi\rangle &= \langle \overline{\nu}_0,\tau_t\phi\rangle + \int_{r=0}^{t}\Bigg(\sum_{n=1}^{\infty}\sum_{j=1}^{n}\int_{x_1}\cdots\int_{x_n}\frac{\beta(x_j)}{n}\\ &\times\left(\tau_{t-r}\phi(n-1,x_1,\cdots,x_{j-1},x_{j+1},\cdots,x_n)\right.\\ &\left.-\tau_{t-r}\phi(n,x_1,\cdots,x_n)\right)\,d\overline{\nu}_r(n,x_1,\cdots,x_n)\\ &+\lambda\Bigg[\left(\overline{\nu}_r(\{0\})\Phi_0\left(\overline{\nu}_r\right)\left(\tau_{t-r}\phi(1,0)-\tau_{t-r}\phi(0)\right)\right)\\ &+\sum_{n=1}^{\infty}\sum_{j=1}^{n+1}\int_{x_1}\cdots\int_{x_n}\frac{1}{(n+1)}\Phi_n\left(\overline{\nu}_r\right)\\ &\times\left(\tau_{t-r}\phi(n+1,x_1,\cdots,x_{j-1},0,x_j,\cdots,x_n)\right.\\ &\left.-\tau_{t-r}\phi(n,x_1,\cdots,x_n)\right)d\overline{\nu}_r(n,x_1,\cdots,x_n)\Bigg]\Bigg)\,dr. \end{aligned}\tag{32}$$

We call equation (32) as mean-field equation and its unique solution is referred to as the mean-field limit.

The proof is given in Appendix B.

Once we establish the mean-field limit, for given $T\ge 0$, we can prove the propagation of chaos by the same arguments as in [6] for every $t\le T$ and hence the proof is omitted. As a consequence, at time t, the distribution of a server state in the limiting system coincides with $\overline{\nu}_t$.

We now convert the integral form of MFEs (32) into partial differential equations. Since the mapping $t\mapsto\overline{\nu}_t$ is continuous (from continuity of $t\mapsto\langle\overline{\nu}_t,\phi\rangle$ for $\phi\in\mathcal{C}_b(E)$ and $\mathcal{C}_b(E)$ is a separating class of $\mathcal{M}_1(E)$ [23]) and $\overline{\nu}_0$ is absolutely continuous *w.r.t.* Lebesgue measure, we have that at every $t\ge 0$, $\overline{\nu}_t$ is also absolutely continuous. Further, let $\overline{\nu}_t(\{0\})$ be denoted by $p_t(0)$ and $p_t(n,x_1,\cdots,x_n)$ be the Radon-Nikodym derivative of the measure $\overline{\nu}_t$ *w.r.t.* Lebesgue measure at $(n,x_1,\cdots,x_n)$. Let a process $P_t=(P_t(\underline{u}),\underline{u}\in E)$ be defined as

$$\begin{aligned}&P_t(n,y_1,\cdots,y_n)\\ &\quad=\int_{x_1=0}^{y_1}\cdots\int_{x_n=0}^{y_n}p_t(n,x_1,\cdots,x_n)\,dx_1\cdots dx_n.\end{aligned}\tag{33}$$

Therefore the probability that a server has n progress jobs in the limiting system at time t and further, i^{th} job ($1\le i\le n$) has age atmost y_i is equal to $P_t(n,y_1,\cdots,y_n)$. We can obtain evolution equations of $P_t(n,y_1,\cdots,y_n)$ from mean-field equation (32) using the fact that there exists sequence of bounded continuous functions monotonically converging to indicator functions and monotone convergence theorem. After simplifications, we obtain the following result.

Corollary 1. *The process $P_t=(P_t(\underline{u}),\underline{u}\in E)$ satisfies the PDEs*

$$\frac{dP_t(0)}{dt}=\int_{y=0}^{\infty}\beta(y)\left(\frac{\partial P_t(1,y)}{\partial y}\right)dy-\lambda\Phi_0(P_t)P_t(0),\tag{34}$$

for $n\ge 1$,

$$\begin{aligned}\frac{dP_t(n,y_1,\cdots,y_n)}{dt}&=-\sum_{i=1}^{n}\frac{1}{n}\frac{\partial P_t(n,y_1,\cdots,y_n)}{\partial y_i}\\ &+\sum_{j=1}^{n+1}\int_{x_j=0}^{\infty}\frac{\beta(x_j)}{n+1}\\ &\times\left(\frac{\partial P_t(n+1,y_1,\cdots,y_{j-1},x_j,y_j,\cdots,y_n)}{\partial x_j}\right)dx_j\\ &-\sum_{j=1}^{n}\int_{x_j=0}^{y_j}\frac{\beta(x_j)}{n}\\ &\times\left(\frac{\partial P_t(n,y_1,\cdots,y_{j-1},x_j,y_{j+1},\cdots,y_n)}{\partial x_j}\right)dx_j\\ &+\sum_{j=1}^{n}\lambda\Phi_{n-1}(P_t)P_t(n-1,y_1,\cdots,y_{j-1},y_{j+1},\cdots,y_n)\\ &-\lambda\Phi_n(P_t)P_t(n,y_1,\cdots,y_n),\end{aligned}\tag{35}$$

where $\Phi_n(P_t)$ is same as equation (26) except that $Q_n(P_t)=P_t(n,\infty,\cdots,\infty)$.

The above mean-field PDEs (34)-(35) represents the dynamics of a single server processor sharing system with Poisson job arrival process with rate $\lambda\Phi_n(P_t)$ when server has n jobs and current distribution is P_t at time t. Therefore MFEs represent dynamics of a non-linear Markov process.

We now state our main result on the stationary behavior of the mean-field limit. Let us first define the class of fixed-points of the mean-field $\mathcal{Y}$[3] as: suppose $\overline{R}_n(\boldsymbol{\theta})=\sum_{j=n}^{\infty}\theta(n,\infty,\ldots,\infty)$, then $\mathcal{Y}=\{\boldsymbol{\theta}:\lim_{n\to\infty}\overline{R}_n(\boldsymbol{\theta})=0\}$.

Theorem 2. *Any fixed-point $\pi=(\pi(\underline{u}),\underline{u}\in E)\in\mathcal{Y}$ satisfies*

$$\pi(n,y_1,\cdots,y_n)=\Gamma_n\mu^n\prod_{i=1}^{n}\int_{x_i=0}^{y_i}\overline{G}(x_i)\,dx_i.\tag{36}$$

where $\Gamma=(\Gamma_n,n\ge 0)$ is a fixed-point of the mean-field limit when job lengths are exponentially distributed with mean $\frac{1}{\mu}$. In particular, if the mean-field of the occupancy process under exponential case has unique fixed-point, then there exists unique fixed-point under general job length distributions. Furthermore, as $\int_{x=0}^{\infty}\overline{G}(x)\,dx=\frac{1}{\mu}$, the fixed-point is insensitive since

$$\pi(n,\infty,\cdots,\infty)=\Gamma_n.\tag{37}$$

The proof is given in Appendix C.

We now demonstrate our results for two occupancy based policies.

Some typical polices: One can obtain the MFEs for any OB(d)

[3]This class contains the fixed-points under which the average occupancy is finite.

policy by using its $q^{(m)}_{(n_1,\cdots,n_{d-j})}$ function in equation (26) from general form of MFEs (34)-(35). We now discuss two policies that belong to OB(d) below:

Definition 3. *SQ(d) Policy: An arrived job is routed to the server with shortest occupancy among d randomly chosen servers and ties are broken uniformly at random. Therefore*

$$q^{(m)}_{(n_1,\cdots,n_{d-j})} = \begin{cases} \frac{1}{j} & \text{if } n_i > m, \forall 1 \le i \le d-j, \\ 0 & \text{otherwise.} \end{cases} \tag{38}$$

As a consequence, upon simplification, we get that

$$\lambda\Phi_n(P_t) = \lambda \frac{\overline{R}_n(P_t)^d - \overline{R}_{n+1}(P_t)^d}{\overline{R}_n(P_t) - \overline{R}_{n+1}(P_t)}, \tag{39}$$

where $\overline{R}_n(P_t) = \sum_{j=n}^{\infty} P_t(j, \infty, \cdots, \infty)$.

Definition 4. *Threshold based policy: For given integer valued parameter α, from set of d sampled servers with number of progressing jobs less than or equal to α, a server is randomly chosen as destination server if there are any available. Otherwise, server with minimum number of jobs is chosen as destination server and ties are broken uniformly at random. Therefore, let* $a = \sum_{i=1}^{d-j} I_{\{n_i \le \alpha\}}$, *then*

$$q^{(m)}_{(n_1,\cdots,n_{d-j})} = \begin{cases} \frac{1}{j+a} & \text{if } m \le \alpha, \\ \frac{1}{j} & \text{else if } n_i > m, \forall 1 \le i \le d-j, \\ 0 & \text{otherwise.} \end{cases} \tag{40}$$

As a result, upon simplification, if $m \le \alpha$, *then*

$$\lambda\Phi_m(P_t) = \lambda \frac{1 - \overline{R}^d_{\alpha+1}(P_t)}{1 - \overline{R}_{\alpha+1}(P_t)} \tag{41}$$

On the other hand, if $m > \alpha$, *then*

$$\lambda\Phi_m(P_t) = \lambda \frac{\overline{R}_m(P_t)^d - \overline{R}_{m+1}(P_t)^d}{\overline{R}_m(P_t) - \overline{R}_{m+1}(P_t)}. \tag{42}$$

IV. Simulation Results

In this section, to support the insensitivity of the fixed-point of the mean-field limit under SQ(d) policy and threshold based policy defined in Section III, we present simulation results. Similar to [4] for SQ(d), we can show that the mean-field limit of the tail occupancy process for exponential job lengths case under the threshold based policies is quasi-montonic and there exists unique globally asymptotic stable fixed-point of the mean-field. Let ϑ^N_i be the stationary probability that there are exactly i jobs in progress at a server and let us define $\vartheta^N = (\vartheta^N_i, i \ge 0)$. In our simulation results, we sample d servers without replacement whereas in our analysis we assume that servers are sampled with replacement. We then compute ϑ^N by using PASTA property for different type of job length distributions by simulating the system upto 1500000 job arrivals. Suppose $\vartheta^{(exp)} = (\vartheta^{(exp)}_i, i \ge 0)$ be the fixed point of the mean-field limit of the occupancy process, then we compute the total variation distance between ϑ^N and $\vartheta^{(exp)}$ defined as $d_{TV}(\vartheta^N, \vartheta^{(exp)}) = \sum_i \left|\vartheta^N_i - \vartheta^{(exp)}_i\right|$.

We assume that the parameters λ, μ, d, and α are fixed at $0.7, 1, 2$, and 2, respectively. We consider exponential (Exp), constant (Const), power-law (PL), mixed-Erlang (ME) distributions. The power-law distribution has CDF $G(y) = 1 - \frac{1}{3y^{\frac{3}{2}}}$ for $y \ge \frac{1}{3}$ and zero otherwise. The mixed-Erlang distribution has i ($1 \le i \le 3$) exponential phases with probability p_i and each exponential phase has rate μ'. We choose $p_1 = .3, p_2 = 0.3, p_3 = 0.4$ and μ' is computed by using the average job length given by $\sum_{i=1}^{3} \frac{ip_i}{\mu'} = \frac{1}{\mu}$. Then from Table I, with $N = 300$, the fixed-point of the mean-field limit under exponential case is close to ϑ^N for different job length distributions having same average job length. This supports both insensitivity of the mean-field limit and the approximation of the stationary distribution of large N system by the fixed-point of the mean-field.

Table I
$d_{TV}(\vartheta^N, \vartheta^{(exp)})$ FOR DIFFERENT JOB LENGTH DISTRIBUTIONS

Policy	Exp	$Const$	PL	ME
SQ(d)	0.0022	0.0046	0.0064	0.0056
Threshold	0.0039	0.0026	0.0007	0.0015

V. Conclusions and Future Work

In this paper we presented a unified mean-field analysis for general occupancy dependent PUSH based policies from which the known results for particular cases like SQ(d) can be recovered. The proof techniques can be extended to the load balancing policies in other system architectures with general service time distributions. In the future work, we will extend the analysis to PULL based policies.

References

[1] W. Winston, "Optimality of the shortest line discipline," *Journal of Applied Probability*, vol. 14(1), pp. 181–189, 1977.

[2] N. D. Vvedenskaya, R. L. Dobrushin, and F. I. Karpelevich, "Queueing system with selection of the shortest of two queues: an asymptotic approach," *Problems of Information Transmission*, vol. 32, no. 1, pp. 20–34, 1996.

[3] M. Mitzenmacher, "The power of two choices in randomized load balancing," *PhD Thesis, Berkeley*, 1996.

[4] A. Mukhopadhyay and R. R. Mazumdar, "Analysis of randomized join-the-shortest-queue (JSQ) schemes in large heterogeneous processor sharing systems," *IEEE Transactions on Control of Network Systems*, vol. 3(2), pp. 116–126, 2016.

[5] ——, "Rate-based randomized routing in large heterogeneous processor sharing systems," in *Proceedings of 26th International Teletraffic Congress (ITC 26)*, 2014.

[6] A. Mukhopadhyay, R. R. Mazumdar, and F. Guillemin, "The power of randomized routing in heterogeneous loss systems," in *Teletraffic Congress (ITC 27), 2015 27th International*, 2015, pp. 125–133.

[7] T. Vasantam, A. Mukhopadhyay, and R. R. Mazumdar, "Insensitivity of the mean-field limit of loss systems under Power-of-d routing," 2017. [Online]. Available: http://arxiv.org/abs/1708.09328

[8] ——, "The mean-field behavior of processor sharing systems with general job lengths under the SQ(d) policy," May 2018, submitted.

[9] Y. Lu, Q. Xie, G. Kliot, A. Geller, J. R. Larus, and A. Greenberg, "Join-idle-queue: A novel load balancing algorithm for dynamically scalable web services," *Perform. Eval.*, vol. 68, no. 11, pp. 1056–1071, Nov. 2011.

[10] M. Mitzenmacher, “Analyzing distributed join-idle-queue: A fluid limit approach,” *CoRR*, vol. abs/1606.01833, 2016. [Online]. Available: http://arxiv.org/abs/1606.01833
[11] A. L. Stolyar, “Pull-based load distribution in large-scale heterogeneous service systems,” *Queueing Syst. Theory Appl.*, vol. 80, no. 4, pp. 341–361, Aug. 2015.
[12] D. Mukherjee, S. Borst, J. S. H. van Leeuwaarden, and P. Whiting, “Asymptotic optimality of power-of-*d* load balancing in large-scale systems,” *CoRR*, 12/2016. [Online]. Available: http://arxiv.org/abs/1612.00722
[13] Q. Bu and L. L. Y. Q. Zhao, “Mean field approximations to a queueing system with threshold-based workload control scheme,” *ArXiv e-prints*, Feb. 2018.
[14] L. Brown, N. Gans, A. Mandelbaum, A. Sakov, H. Shen, S. Zeltyn, and L. Zhao, “Statistical analysis of a telephone call center,” *Journal of the American Statistical Association*, vol. 100, no. 469, pp. 36–50, 2005.
[15] P. Kolesar, “Stalking the endangered cat: A queueing analysis of congestion at automatic teller machines,” *Interfaces*, vol. 14, no. 6, pp. 16–26, 1984.
[16] S. F. Yashkov and A. S. Yashkova, “Processor sharing: A survey of the mathematical theory,” *Automation and Remote Control*, vol. 68, no. 9, pp. 1662–1731, 2007.
[17] M. Bramson, Y. Lu, and B. Prabhakar, “Randomized load balancing with general service time distributions,” in *Proceedings of ACM SIGMETRICS*, 2010, pp. 275–286.
[18] ——, “Asymptotic independence of queues under randomized load balancing,” *Queueing Systems*, vol. 71, no. 3, pp. 247–292, 2012.
[19] R. Aghajani and K. Ramanan, “The hydrodynamic limit of a randomized load balancing network,” *ArXiv e-prints*, Jul. 2017.
[20] V. Varadarajan, “On a theorem of F. Riesz concerning the form of linear functionals,” *Fund. Math.*, vol. 46, pp. 209–220, 1959).
[21] P. Billingsley, *Convergence of probability measures*, 2nd ed. John Wiley & Sons, Inc., New York, 1999.
[22] L. Decreusefond and P. Moyal, “A functional central limit theorem for the M/GI/∞ queue,” *Ann. Appl. Probab.*, vol. 18, no. 6, pp. 2156–2178, 12 2008.
[23] S. N. Ethier and T. G. Kurtz, *Markov Processes: Characterization and Convergence*. John Wiley and Sons Ltd, 1985.
[24] A. Jakubowski, “On the Skorokhod topology,” *Annales de l’I.H.P. Probabilités et Statistiques*, vol. 22, no. 3, pp. 263–285, 1986.
[25] P. Robert, *Stochastic Networks and Queues*, ser. Stochastic Modelling and Applied Probability Series. New York: Springer-Verlag, 2003.
[26] O. Kallenberg, *Random measures*. Akademie-Verlag, 1983.
[27] S. L. Brumelle, “A generalization of Erlang’s loss system to state dependent arrival and service rates,” *Mathematics of Operations Research*, vol. 3, no. 1, pp. 10–16, 1978.

Appendix

A. Proof of Lemma 1

Proof. Each server can be chosen as a potential server with probability $\frac{1}{N}$. Now out of d potential servers, say j potential servers are having m jobs. Furthermore, let k potential servers out of j servers with m jobs are lying in state $(m, x_1, \cdots, x_m)$. Also, let n_i be the number of progressing jobs at the potential server with i^{th} smallest index out of the remaining $d-j$ potential servers. Then according to the policy Λ, the probability that out of d potential servers, there are j servers with m jobs and among them there are k servers lying in state $(m, x_1, \cdots, x_m)$, and the remaining $d-j$ servers have number of jobs equal to $n_1, \cdots, n_{d-j}$, respectively, is given by,

$$\binom{d}{j}\binom{j}{k} k q^{(m)}_{(n_1,\cdots,n_{d-j})} \left(\frac{\eta(\{(m,x_1,\cdots,x_m)\})}{N}\right)^k \times \left(\frac{\eta(E_m)-\eta(\{(m,x_1,\cdots,x_m)\})}{N}\right)^{j-k} \prod_{i=1}^{d-j}\left(\frac{\eta(E_{n_i})}{N}\right).$$

Then by summing over all possible values of j, k, and $(n_1, \cdots, n_{d-j})$, we get equation (25). □

B. Proof of Theorem 1

Proof. In this proof, we call any solution of the mean-field equation (32) for given initial point as the mean-field solution. By using standard arguments, using Gronewall's inequality and the norm defined in equation (22), we get that if $(\overline{\nu}^1_t, t \geq 0), (\overline{\nu}^2_t, t \geq 0)$ are two solutions with initial points $\overline{\nu}^1_0, \overline{\nu}^2_0$, respectively, then $\overline{\nu}^1_0 = \overline{\nu}^2_0$ implies $\langle \overline{\nu}^1_t, \phi \rangle = \langle \overline{\nu}^2_t, \phi \rangle$ for all $\phi \in \mathcal{C}_b(E)$. Since $\mathcal{C}_b(E)$ is a separating class of $\mathcal{M}_1(E)$ [23], we have $\overline{\nu}^1_t = \overline{\nu}^2_t$.

The existence of a solution for MFEs follows from the proof of the convergence of $\{\overline{\nu}^N_t, t \geq 0\}$ in $\mathcal{D}_{\mathcal{M}_1(E)}([0, \infty))$.

We now look at the convergence of the sequence of $\{\overline{\nu}^N_t, t \geq 0\}$. The crux of the proof is to first show that the sequence of processes $(\overline{\nu}^N_t, t \geq 0)$ is relatively compact. After that, we show every limit point $(\chi_t, t \geq 0)$ has sample paths that satisfy MFEs almost surely. Since for every limiting point, the initial measure is the deterministic measure $\overline{\nu}_0$, we get that all limiting points have almost surely identical sample paths coinciding with the unique mean-field solution. The unique mean-field solution is referred to as the the mean-field limit denoted by $(\overline{\nu}_t, t \geq 0)$.

Let $\overline{\mathcal{F}}^N_t$ be the natural filtration associated with the process $(\overline{\nu}^N_t, t \geq 0)$. Then, from Proposition 1, the following process $(\overline{M}^N_t(\phi), t \geq 0)$ is a RCLL square integrable $\overline{\mathcal{F}}^N_t$−martingale: for $\phi \in \mathcal{C}^1_b(E)$,

$$\overline{M}^N_t(\phi) = \langle \overline{\nu}^N_t, \phi \rangle - \langle \overline{\nu}^N_0, \phi \rangle - \int_{s=0}^t \overline{A}^N \langle \overline{\nu}^N_s, \phi \rangle \, ds, \quad (43)$$

where $\overline{A}^N(\cdot)$ is the generator of the Markov process $(\overline{\nu}^N_t, t \geq 0)$ given by,

$$\begin{aligned}
\overline{A}^N \langle \overline{\nu}^N_s, \phi \rangle &= \langle \overline{\nu}^N_s, \phi'_\Sigma \rangle \\
&+ \Bigg(\sum_{n=1}^{\infty} \sum_{j=1}^{n} \int_{x_1} \cdots \int_{x_n} \frac{\beta(x_j)}{n} \\
&\times (\phi(n-1, x_1, \cdots, x_{j-1}, x_{j+1}, \cdots, x_n) - \phi(n, x_1, \cdots, x_n)) \\
&\times d\overline{\nu}^N_s(n, x_1, \ldots, x_n) \\
&+ \lambda \Bigg[\left(\overline{\nu}^N_s(\{0\}) \Phi_0 \left(\overline{\nu}^N_s\right) (\phi(1,0) - \phi(0)) \right) \\
&+ \sum_{n=1}^{\infty} \sum_{j=1}^{n+1} \int_{x_1} \cdots \int_{x_n} \frac{1}{(n+1)} \Phi_n \left(\overline{\nu}^N_s\right) \\
&\times (\phi(n+1, x_1, \cdots, x_{j-1}, 0, x_j, \cdots, x_n) - \phi(n, x_1, \cdots, x_n)) \\
&\times d\overline{\nu}^N_s(n, x_1, \cdots, x_n) \Bigg] \Bigg). \quad (44)
\end{aligned}$$

To establish the relative compactness of the sequence $(\overline{\nu}^N_t, t \geq 0)$, we use Jakubowski's criteria given in Theorem 4.6 of [24] from which we need to establish two necessary and sufficient conditions. The first condition is to show that the sequence $(\langle \overline{\nu}^N_t, \phi \rangle, t \geq 0)$ for $\phi \in \mathcal{C}^1_b(E)$ is relatively compact in $D_{\mathcal{R}}([0,\infty))$. To prove this, we need to establish the two sufficient conditions given in [25, Theorem C.9]. These conditions are verified to be true by showing $\mathbb{E}\left[< \overline{M}_{\cdot}^{N}(\phi) >_T\right] < \infty$ and exploiting Doob's inequality, and the proof is similar to [22].

The second condition is to find a compact set $\mathbb{K}_{T,\gamma}$ for every $T,\gamma > 0$ such that

$$\liminf_{N\to\infty} \mathcal{P}(\overline{v}_t^N \in \mathbb{K}_{T,\gamma}, \forall t \in [0,T]) \geq 1-\gamma. \quad (45)$$

Using the assumption 2, we can find some M_0 such that

$$\liminf_{N\to\infty} \mathcal{P}(\max(\langle \overline{v}_0^N, \Xi\rangle, \langle \overline{v}_0^N, \Upsilon\rangle) < M_0) > 1-\gamma. \quad (46)$$

Now let $M_T = M_0(1+T) + \frac{2\lambda T}{\mu}$, then we can show

$$\liminf_{N\to\infty} \mathcal{P}(\sup_{t\in[0,T]} \langle \overline{v}_t^N, \Upsilon\rangle < M_T) > 1-\gamma. \quad (47)$$

For given $0 < \gamma < 1$, let

$$\mathcal{W}_{T,\gamma} \triangleq \{\zeta \in \mathcal{M}_1(E) : \langle \zeta, \Upsilon\rangle < M_T\}. \quad (48)$$

Then we have $\lim_{c\to\infty} \sup_{\zeta\in\mathcal{W}_{T,\gamma}} \zeta(\overline{B}) = 0$, where $B = \{(0)\} \cup (\cup_n B_n)$, $B_n = ([0,c],\cdots,[0,c]) \in \mathcal{B}(E_n)$, and $\overline{B}$ denotes the complement of B. Therefore, from Lemma A7.5 of [26], $\mathcal{W}_{T,\gamma}$ is relatively compact in $\mathcal{M}_1(E)$ and let $\mathbb{K}_{T,\gamma}$ be the closure of $\mathcal{W}_{T,\gamma}$. Then the compact set $\mathbb{K}_{T,\gamma}$ satisfies equation (45). This completes the proof of relative compactness of $(\overline{v}_t^N, t\geq 0)$.

Let $(\chi_t, t \geq 0)$ be a limiting point of a converging subsequence of $(\overline{v}_t^N, t \geq 0)$. Also, from the assumption 2, χ_0 almost surely coincides with $\overline{v}_0$. Further, from the fact that $\mathbb{E}\left[< \overline{M}_{\cdot}^{N}(\phi) >_T\right] < \infty$ and Doob's inequality, due to standard convergence criterion, we have that the sequence $(\overline{M}_t^N(\phi), t \geq 0)$ converges in distribution to null process. Now by using the continuous mapping theorem, we get

$$\begin{aligned}
\langle \chi_t, \phi\rangle = \langle \chi_0, \phi\rangle &+ \int_{s=0}^{t} \langle \chi_s, \phi'_{\Sigma}\rangle\, ds \\
&- \int_{s=0}^{t} \Bigg(\sum_{n=1}^{\infty}\sum_{j=1}^{n} \int_{x_1}\cdots\int_{x_n} \frac{\beta(x_j)}{n} \\
\times \big(\phi(n-1, x_1,\cdots,x_{j-1},x_{j+1},\cdots,x_n) &- \phi(n,x_1,\cdots,x_n)\big) \\
\times\, d\chi_s(n,x_1,\cdots,x_n) + \lambda\Bigg[&(\chi_s(\{0\})\Phi_0(\chi_s)(\phi(1,0)-\phi(0))) \\
&+ \sum_{n=1}^{\infty}\sum_{j=1}^{n+1}\int_{x_1}\cdots\int_{x_n} \frac{1}{(n+1)}\Phi_n(\chi_s) \\
\times \big(\phi(n+1,x_1,\cdots,x_{j-1},0,x_j,\cdots,x_n) &- \phi(n,x_1,\cdots,x_n)\big) \\
&\times\, d\chi_s(n,x_1,\cdots,x_n)\Bigg]\Bigg) ds. \quad (49)
\end{aligned}$$

By using fundamental theorem of calculus, it can be shown that any process $(\eta_t, t\geq 0) \in \mathcal{C}_{\mathcal{M}_1(E)}([0,\infty))$ is a solution to equation (49) iff it is a solution to the mean-field equation (32). We omit this proof due to space constraints and the proof is similar to that in [7]. For given initial point, since there exists unique mean-field solution, from assumption 2, all the limiting points coincide almost surely with the unique mean-field solution. □

C. Insensitivity: Proof of Theorem 2

Proof. Suppose $\boldsymbol{\theta} = (\theta(\underline{u}), \underline{u} \in E)$ be a fixed-point for the process $(P_t, t\geq 0)$. Let $v_n = \prod_{i=1}^{n} \frac{\lambda\Phi_{i-1}(\boldsymbol{\theta})}{\mu}$. Then we show that

$$\theta(n,y_1,\ldots,y_n) = \frac{v_n}{1+\sum_{m=1}^{\infty} v_m}\mu^n \prod_{i=1}^{n}\int_{x_i=0}^{y_i} \overline{G}(x_i)\, dx_i \quad (50)$$

and

$$\theta(0) = \frac{1}{1+\sum_{m=1}^{\infty} v_m}. \quad (51)$$

Since $\lim_{n\to\infty} \overline{R}_n(\theta) = 0$, it is verified that we have $\sum_{m=1}^{\infty} v_m < \infty$.

We prove equations (50)-(51) as follows. Consider a single server processor sharing system with job length distribution same as in the system model and the job arrival process is Poisson process with pre-specified state-dependent arrival rate $\lambda\Phi_n(\boldsymbol{\theta})$ when there are n progressing jobs at the server. Then from [27], this system is stable and let $\pi^{(single)}$ be the stationary distribution. Then $\pi^{(single)}(n,y_1,\cdots,y_n)$ is equal to right side of equation (50). However, on comparing the stationary mean-field dynamics, we get that $\boldsymbol{\theta}$ is also another stationary distribution for the considered single server system. Due to stability of the single server system, we must have $\boldsymbol{\theta} = \pi^{(single)}$ and hence, equations (50)-(51) are true.

We now assume that $\Gamma = (\Gamma_n, n\geq 0)$ where $\Gamma_n = \theta(n,\infty,\cdots,\infty))$ and $\Gamma_0 = \theta(0)$. We then have from equations (50)-(51),

$$\Gamma_n = \frac{v_n}{1+\sum_{m=1}^{\infty} v_m} \quad (52)$$

and

$$\Gamma_0 = \frac{1}{1+\sum_{m=1}^{\infty} v_m}. \quad (53)$$

Further, we can write v_n in terms of Γ as $v_n = \prod_{i=1}^{n} \frac{\lambda\Psi_{i-1}(\Gamma)}{\mu}$ where

$$\Psi_l(\Gamma) = \sum_{j=1}^{d} \sum_{(n_1,\cdots,n_{(d-j)})} jq^{(l)}_{(n_1,\cdots,n_{d-j})}\binom{d}{j}(\Gamma_l)^{j-1}\prod_{i=1}^{d-j}\Gamma_{n_i}.$$

We then have that for $n\geq 0$,

$$\lambda\Psi_n(\Gamma)\Gamma_n = \mu\Gamma_{n+1}. \quad (54)$$

Then by simplifying equation (34)-(35) under exponential job length distributions with the same mean $\frac{1}{\mu}$, a fixed-point $w = (w_n, n\geq 0)$ satisfies

$$\lambda\Psi_n(w)w_n = \mu w_{n+1}. \quad (55)$$

Hence from equations (54), (55), we have that Γ is a fixed-point of the mean-field under exponential job lengths with average length $\frac{1}{\mu}$. Further, from equations (50)-(51), we get equation (36). □