Speed scaling in fork-join queues: a comparative study

Andrea Marin
Università Ca' Foscari Venezia
Venezia, Italy
marin@unive.it

Sabina Rossi
Università Ca' Foscari Venezia
Venezia, Italy
sabina.rossi@unive.it

Carey Williamson
Università Ca' Foscari Venezia
University of Calgary
carey@cpsc.ucalgary.ca

ABSTRACT

Frequency scaling plays an important power-saving role in computer systems. In fork-join systems, dynamic adaptation of the server speeds can significantly reduce system power consumption while maintaining high throughput. In previous work, we studied a rate adaptation policy that dynamically chooses server speeds based on the difference in join-queue lengths, with each server knowing only its own join-queue length and that of one other server. In this work, we increase the information available to each server, and choose speeds based on the knowledge of the join-queue lengths of two other servers. We show that, under a specific canonical configuration of the service rates, the new system has exactly the same throughput and subtask dispersion as before, but with reduced power consumption. We use time-reversal analysis to derive the exact stationary performance of this new model under saturation conditions, and use simulation to study more general cases.

1. INTRODUCTION

Fork-join queueing models are important abstractions of computer systems in which jobs are split into a set of tasks that are processed in parallel by independent servers. The served tasks are eventually merged back together before leaving the system. Examples of this type of computation are MapReduce, RAID (Redundant Array of Independent Disks), and parallel database queries. Figure 1 shows a fork-join queue. Jobs arrive externally, and are forked into K tasks that are stored in the fork-queues while they wait for a free server. Each server fetches tasks from its fork-queue, and once it finishes its work, the served tasks are stored in the join-queues. Once all the sibling tasks of a job are served, the join occurs, and the jobs then leave the system.

In fork-join queueing systems, we can reduce power consumption using speed scaling methods. In single server systems, the speed is usually dynamically set based on the number of jobs in the queue [1]. In fork-join systems, however, it makes sense to slow down the servers whose join-queues contain more tasks than the others, since a job can leave the system only after its last task completes.

DOI: 10.1145/1235

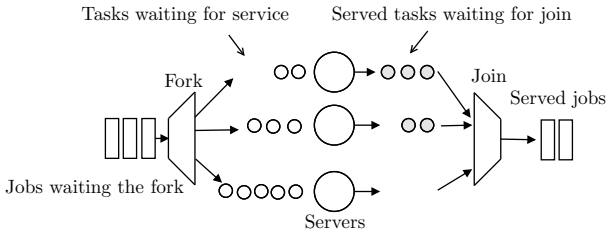


Figure 1: Sketch of a fork-join queueing system.

In practice, when the number of parallel servers is high, as in many MapReduce applications, it is difficult for a server to know the join-queue lengths at all the other servers. For this reason, in [7, 9], we presented two rate adaptation algorithms in which each server decides its speed based on local knowledge of the difference between its join-queue length and that of a neighbour. We have shown that under Markovian assumptions and saturation conditions, the algorithms lead to join-queue lengths with finite mean. In this paper, we build upon the *Bimodal* rate adaptation algorithm [7], in which each server chooses between two possible speeds according to its state (i.e., higher speed if it has a shorter join-queue than its neighbour, and slower speed otherwise).

In this work, we explore the benefits of increasing the information available for each server. Specifically, we study a *trimodal* (three-speed) rate adaptation algorithm that bases its decisions on the knowledge of the differences between its join-queue length and those of *two* other servers (analogous to power-of-d choices [4]). We provide an exact analysis of the join-queue length distribution for this new model under saturation (i.e., there is always a job waiting to be served), and use stochastic simulation to explore other cases. The exact analysis uses the definition of ρ-reversibility to derive the system's throughput, the stationary distribution of the queue length differences, and the system's power consumption. We compare the results with those previously obtained in the literature for the bimodal model. The outcomes of our investigations can be summarised as follows:

- We introduce a canonical trimodal rate adaptation algorithm, which decides the servers' speed based on the join-queue lengths of two other neighbours. If the server has a shorter join-queue than both neighbours, then it works at the fastest speed. If its join-queue length matches or exceeds those for both neighbours,

it works at the slowest speed. Otherwise, it works at a speed that is the mean of the previous two.

- We provide an exact analysis of the canonical trimodal algorithm under Markovian assumptions and saturation. Although the CTMC underlying the trimodal model is very different from that of the bimodal model, and does not have the same time-reversal properties, we show that the system throughput and the distribution of the join-queue lengths are identical. We found this result quite surprising and intriguing.

- We prove that, although the canonical trimodal model does not improve control of the join-queue length compared to the bimodal model, it guarantees lower power consumption. That is, the extra information available to servers allows the trimodal system to achieve the same performance as the bimodal system, but with lower power cost.

- Finally, we use stochastic simulations to study the system in heavy load and compare the result with those obtained by the analytical model. Moreover, we study the trimodal algorithm when the intermediate speed is chosen differently from the mean of the lowest and highest speeds. In fact, simulations suggest that other settings for the intermediate speed provide interesting tradeoffs between system throughput, join-queue length, and power consumption.

Related work

Fork-join systems have been widely studied in the queueing theory literature, but very few exact analytical results are available. Under the Flatto-Hahn-Wright assumptions (independent Poisson arrivals and exponential service times), the solution for $K = 2$ servers is known [2, 14]. The wide adoption of distributed computation by modern data centers has re-invigorated interest in fork-join systems. Some recent works have introduced accurate approximate analyses [6, 11] for servers with constant speed, while in [12] the authors use job statistics to help determine the servers' rates in such a way that the join-queue lengths are reduced.

Similar to the work proposed here, in [7] we studied a model in which servers adapt their speeds according to the difference between their join-queue length and that of another server. The analytical results rely on an important property of the Markov chain underlying the queueing system, i.e., it is ρ-reversible. Informally, a ρ-reversible Markov chain has the property that its time-reversed process is stochastically indistinguishable from the original one if we rename the states according to function ρ [8]. This property, in the form known as *dynamic reversibility*, has been previously used to study the kinetics of polymer crystallization in a similar way to what we do in this work [3]. In this paper, we extend the results of [7] by computing the power consumption of the bimodal algorithm and the marginal distribution of the join-queue length differences. Moreover, we introduce the trimodal model in which the algorithm adapts the servers' rates based on the differences between their join-queue lengths. Although the trimodal's Markov chain is ρ-reversible under certain assumptions on the servers' rates (i.e., canonical trimodal), the renaming function is completely different from that of the bimodal, and hence its analysis is entirely new.

Paper structure

The paper is structured as follows. Section 2 summarizes prior results for the bimodal model, and presents new results on its symmetry and power consumption. Section 3 introduces the new trimodal model and presents its mathematical analysis under the saturation assumption. Section 4 supplements this analysis with simulation results that explore more general cases. Finally, Section 5 concludes the paper, with formal proofs provided in the Appendix.

2. THE BIMODAL MODEL

In [7], we studied a queue length control mechanism that can be applied to fork-join queues, and provided an exact analysis under the assumption of iid exponential service times and saturation. The latter assumption requires the fork queues to always have at least one task, and is a good approximation for the heavy load case studied in Section 4. In [7], each server has knowledge of the difference between its own join-queue length and that of a neighbour. The neighbourhood relation is defined in a circular way, i.e., if each of the $K \geq 2$ servers is labelled with a number $1, \ldots, K$, then the neighbour of server k is k^+, where $k^+ = (k \mod K)+1$. Let $X_K^b(t) = (n_1, \ldots, n_K)(t)$ be the state of the system at time t, where $n_k = \ell_k - \ell_{k^+}$ and ℓ_k is the join-queue length at server k. If the next job completion occurs at time $t+\Delta t$ at server k, then $X_K^b(t+\Delta t) = X_K^b(t)+\mathbf{e}_k-\mathbf{e}_{k^-}$, where $\mathbf{e}_k$ is the K-dimension vector with a 1 in position k and all other elements are 0, and $k^- = ((k+K-2) \mod K)+1$. When the service times are state-dependent exponential random variables, the stochastic process $X_K^b(t)$ is a CTMC. The *bimodal* model of [7] is defined in such a way that each server operates at a high speed μ if its join-queue is shorter than that of its neighbour, and at a low speed η, otherwise.

Definition 1 (Bimodal model). *The* bimodal model *is the process $X_K^b(t)$ with state space*

$$\mathcal{S}_K = \{\mathbf{n} = (n_1, \ldots, n_K) : n_i \in \mathbb{Z} \wedge \sum_{k=1}^{K} n_k = 0\} \qquad (1)$$

and transition rates for $h \to 0^+$ defined as follows:

$$Pr\{X_K^b(t+h) = \mathbf{n}-\mathbf{e}_{k^-}+\mathbf{e}_k \mid X_K^b(t) = \mathbf{n}\} = \lambda(n_k)h+o(h)$$

$$Pr\{X_K^b(t+h) = \mathbf{n} \mid X_K^b(t) = \mathbf{n}\} = 1-\left(\sum_{k=1}^{K}\lambda(n_k)\right)h+o(h)$$

where:

$$\lambda(n_k) = \begin{cases}\mu & \text{if } n_k < 0\\ \eta & \text{if } n_k \geq 0\end{cases}. \qquad (2)$$

This model has the following stationary performance indices [7]. Let $\pi_K^b(\eta,\mu)$ be the stationary distribution (when it exists) for the bimodal model. The balance index is a measure of the effectiveness of this control mechanism in maintaining short join-queues. The throughput is the expected number of joins per unit time in steady-state.

Definition 2 (Balance index). *The balance index for stable bimodal models is defined as follows:*

$$B_K^b(\eta,\mu) = \sum_{\mathbf{n}\in\mathcal{S}_K}\pi_K^b(\mathbf{n},\eta,\mu)p(\mathbf{n})\,,$$

where $p(\mathbf{n})$ is the sum of positive components of $\mathbf{n}$, i.e., $p(\mathbf{n}) = \sum_{k=1}^{K} n_k \delta_{n_k>0}$ where δ_P is 1 if proposition P is true, and 0 otherwise.

DEFINITION 3 (THROUGHPUT). *The throughput for stable bimodal models is defined as follows:*

$$T_K^b(\eta,\mu) = \frac{1}{K} \sum_{\mathbf{n}\in\mathcal{S}_K} \pi_K^b(\mathbf{n},\eta,\mu) \sum_{k=1}^{K} \lambda(n_k)\,.$$

2.1 Previous results on the bimodal model

In [7], we studied the process $X_K^b(t)$ of Definition 1 underlying the bimodal model. Here, we summarise previously proven results, which we will use subsequently.

- For any finite K, the CTMC $X_K^b(t)$ is ergodic if and only if $\eta < \mu$.
- Under the stability condition, the CTMC $X_K^b(t)$ is dynamically reversible under the renaming:

  $$\rho(n_1,\ldots,n_K) = (n_K,\ldots,n_1)$$

  and its stationary distribution is:

  $$\pi_K^b(\mathbf{n},\eta,\mu) = \frac{1}{G_K^b(\eta,\mu)} \left(\frac{\eta}{\mu}\right)^{p(\mathbf{n})}\,,$$

  where the normalizing constant $G_K^b(\eta,\mu)$ is:

  $$G_K^b(\eta,\mu) = 1 + \sum_{j=1}^{K-1} \binom{K}{j}\binom{K-1}{j-1}(K-j) \cdot \beta(\eta/\mu, K-j, 1-K)\,,$$

  where β is Euler's incomplete Beta-function:

  $$\beta(z.a.b) = \int_0^z u^{a-1}(1-u)^{b-1}du\,.$$

- In stability, the balance index is given by:

  $$B_K^b(\eta,\mu) = \frac{1}{G_K^b(\eta,\mu)} \left(\frac{\eta/\mu}{1-\eta/\mu}\right)^K \cdot \sum_{j=1}^{K-1} \binom{K}{j}\binom{K-1}{j-1}(K-j)\left(\frac{\mu}{\eta}\right)^j\,,$$

  which can be normalised on K.
- In stability, the throughput of the K servers is:

  $$T_K^b(\eta,\mu) = \frac{1}{G_K^b(\eta,\mu)}\Bigg(K\eta + \sum_{j=1}^{K-1}(j\eta+(K-j)\mu)\binom{K}{j}\binom{K-1}{j-1} \cdot (K-j)\;\beta\left(\eta/\mu, K-j, 1-K\right)\Bigg)\,.$$

  and T_K^b/K is the throughput of the system.
- Despite the formulation in terms of the incomplete β-function, the normalising constant and the stationary performance indices can be computed exactly with a finite number of elementary operations.

2.2 New results on the bimodal model

In this section, we derive some novel results for the bimodal model. We start with a corollary of the result on the stationary distribution:

COROLLARY 1. *If $X_K^b(t)$ is the process in Definition 1, and $\mathbf{n} \in \mathcal{S}_K$, then:*

$$\pi_K^b(\mathbf{n},\eta,\mu) = \pi_K^b(-\mathbf{n},\eta,\mu)\,.$$

The proof is trivial, since the sum of positive components of $\mathbf{n}$ equals the opposite of the sum of the negative components. A second important observation is that the marginal distribution must be symmetric with respect to state $\mathbf{0}$. More formally, for $n \in \mathbb{Z}$, the marginal distribution of the bimodal model is defined as:

$$\pi_K^{b*}(n,\eta,\mu) = \sum_{\substack{\mathbf{n}\in\mathcal{S}_K:\\ \mathbf{n}=(n,n_2,\ldots,n_K)}} \pi_K^b(\mathbf{n},\eta,\mu)\,.$$

Moreover, we consider the following definitions:

$$\pi_K^{b+}(\eta,\mu) = \sum_{n=1}^{\infty} \pi_K^{b*}(n,\eta,\mu)\,, \quad \pi_K^{b-}(\eta,\mu) = \sum_{n=1}^{\infty} \pi_K^{b*}(-n,\eta,\mu)\,.$$

The next corollary proves the symmetry of π^{b*} with respect to state $\mathbf{0}$. This property is useful when computing a closed-form expression for $\pi_K^{b*}(n,\eta,\mu)$.

COROLLARY 2. *For the process $X_K^b(t)$ of Definition 1:*

1. *$\pi_K^{b*}(n,\eta,\mu) = \pi_K^{b*}(-n,\eta,\mu)$ for all $n \in \mathbb{N}$,*
2. *$\pi_K^{b+}(\eta,\mu) = \pi_K^{b-}(\eta,\mu)$.*

Proof. For $n \in \mathbb{N}$ with $n \neq 0$, consider two subsets of $\mathcal{S}_K$:

$$S_K^n = \{\mathbf{n} \in \mathcal{S}_K \mid \mathbf{n} = (n, n_2, \ldots, n_K)\}$$

consisting of the states $(n, n_2, \ldots, n_K)$, i.e., with fixed first component and arbitrary remaining ones, and

$$S_K^{-n} = \{\mathbf{n} \in \mathcal{S}_K \mid \mathbf{n} = (-n, n_2, \ldots, n_K)\}\,.$$

Observe that for each state $\mathbf{n} \in S_K^n$, there exists a corresponding opposite state $-\mathbf{n} \in S_K^{-n}$, and vice versa. Since $\mathbf{n}$ and $-\mathbf{n}$ have the same stationary probability (by Corollary 1), the first statement is proved. The second one follows from the first straightforwardly. □

The following result expresses the marginal distribution for the model. The proof is given in the Appendix.

THEOREM 1. *Let $X_K^b(t)$ be the process of Definition 1 with $K \geq 2$. The marginal stationary distributions of $X_K^b(t)$ with rates η and μ such that $0 < \eta < \mu$ are:*

$$G_K^b(\eta,\mu)\pi_K^{b*}(0,\eta,\mu) = 1 + \sum_{j=1}^{K-2} \binom{K-1}{j}\binom{K-2}{j-1} \cdot (K-j-1)\beta\left(\eta/\mu, K-j-1, 2-K\right)\,,$$

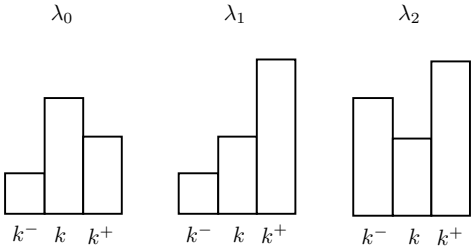


Figure 2: Example configurations of the trimodal model.

and, for $n > 0$:

$$G_K^b(\eta,\mu)\pi_K^{b*}(n,\eta,\mu) = G_K^b(\eta,\mu)\pi_K^{b*}(-n,\eta,\mu) = \binom{n+K-2}{K-2}\left(\frac{\eta}{\mu}\right)^n + \sum_{j=1}^{K-2}\binom{K-1}{j}\binom{K+n-2}{j-1}\left(\frac{\eta}{\mu}\right)^{K-j-1+n} \cdot {}_2F_1\left(K+n-1, K-j-1, K+n-j, \frac{\eta}{\mu}\right).$$

The aggregated probabilities of observing a positive (resp. negative) state in one of the servers is:

$$G_K^b(\eta,\mu)\pi_K^{b+}(\eta,\mu) = G_K^b(\eta,\mu)\pi_K^{b-}(\eta,\mu) = \sum_{j=1}^{K-1}\binom{K-1}{j}\binom{K-1}{j-1}(K-j)\left(\frac{\eta}{\mu}\right)^{K-j} \cdot \beta\left(\eta/\mu, K-j, 1-K\right).$$

From the marginal distributions, we can derive the system's power consumption. Specifically, we assume that the power consumption of a computational unit scales with a factor α, where $2 \le \alpha \le 3$ (see, e.g., [10]). Thus, the power consumption of the bimodal model is given by Corollary 3.

COROLLARY 3. *Let $X_K^b(t)$ be the bimodal model of Definition 1 with $K \ge 2$. The power consumption of the bimodal model with rates η and μ such that $0 < \eta < \mu$ is:*

$$P_K^b(\eta,\mu) = K\left(\pi_K^{b+}(\eta,\mu)(\eta^\alpha + \mu^\alpha) + \pi_K^{b*}(0,\eta,\mu)\eta^\alpha\right).$$

Proof. The proof follows straightforwardly from Corollary 2 and by the symmetry of the system. □

3. THE TRIMODAL MODEL

In this section, we study a new model in which each server k is aware of the differences between its join-queue length and those of its two neighbours, k^- and k^+. Three cases are possible: the k-th join-queue is longer than or equal to those of its two neighbours; its length is in between those of its neighbours; and the k-th join-queue is the shortest. We use three speeds λ_0, λ_1, and λ_2 for these cases, respectively. Figure 2 shows examples of each. Let $X_K^{tr}(t) = (n_1, \ldots, n_K)(t)$ be the process defined as $X_K^b(t)$ except for the transition rates that are as in Definition 4. Again, under state-dependent exponential service times, the process $X_K^{tr}(t)$ is a CTMC.

DEFINITION 4 (TRIMODAL MODEL). *The* trimodal model *is the process $X_K^{tr}(t)$ with state space $\mathcal{S}_K$ as in Equation (1), and transition rates for $h \to 0^+$ defined as follows:*

$$Pr\{X_K^{tr}(t+h) = \mathbf{n} - \mathbf{e}_{k^-} + \mathbf{e}_k \mid X_K^{tr}(t) = \mathbf{n}\} = \lambda(n_{k^-}, -n_k)h + o(h), \quad (3)$$

$$Pr\{X_K^{tr}(t+h) = \mathbf{n} \mid X_K^{tr}(t) = \mathbf{n}\} = 1 - \left(\sum_{k=1}^{K}\lambda(n_{k^-}, -n_k)\right)h + o(h), \quad (4)$$

where:

$$\lambda(x,y) = \begin{cases}\lambda_0 & \text{if } x \le 0 \wedge y \le 0\\ \lambda_1 & \text{if } (x \le 0 \wedge y > 0) \vee (x > 0 \wedge y \le 0)\\ \lambda_2 & \text{if } x > 0 \wedge y > 0\end{cases}. \quad (5)$$

We can define the balance index and the throughput for the trimodal model as we have done for the bimodal model. Let $\pi_K^{tr}(\mathbf{n}, \lambda_0, \lambda_1, \lambda_2)$ be the stationary distribution of the trimodal model when it exists.

DEFINITION 5 (BALANCE INDEX). *The balance index for stable trimodal models is defined as follows:*

$$B_K^{tr}(\lambda_0,\lambda_1,\lambda_2) = \sum_{\mathbf{n}\in\mathcal{S}_K}\pi_K^{tr}(\mathbf{n},\lambda_0,\lambda_1,\lambda_2)p(\mathbf{n}),$$

where $p(\mathbf{n})$ is the sum of positive components of $\mathbf{n}$.

DEFINITION 6 (THROUGHPUT). *The throughput for stable trimodal models is defined as follows:*

$$T_K^{tr}(\lambda_0,\lambda_1,\lambda_2) = \frac{1}{K}\sum_{\mathbf{n}\in\mathcal{S}_K}\pi_K^{tr}(\mathbf{n},\lambda_0,\lambda_1,\lambda_2)\sum_{k=1}^{K}\lambda(n_{k^-}, -n_k)$$

3.1 Analysis of the trimodal model

Each server in the trimodal model has more information than in the bimodal model, and hence we expect better performance. The problem is how to regulate the three speeds $\lambda_0, \lambda_1, \lambda_2$. Here we study the specific configuration $\lambda_1 = (\lambda_0 + \lambda_2)/2$, i.e., when the join-queue length is in between those of the two neighbours, the service speed is the average of the minimum (λ_0) and maximum (λ_2) speeds. For this specific configuration, we prove a very surprising result: the stationary distribution, throughput, and balance index are exactly the same as those of the bimodal model with $\eta = \lambda_0$ and $\mu = \lambda_2$, even though the infinitesimal generator for $X_K^{tr}(t)$ is very different from that of $X_K^b(t)$. However, the trimodal model has a key advantage with respect to the bimodal, namely lower power consumption. Indeed, although the configuration $\lambda_1 = (\lambda_0+\lambda_2)/2$ does not improve throughput or lower the balance index, we can reduce power consumption compared to the bimodal model. Hereafter, we refer to the trimodal model with $\lambda_1 = (\lambda_0+\lambda_2)/2$ as *canonical trimodal model*, and exclude λ_1 from the notation for the stationary distribution and performance indices.

THEOREM 2. *The canonical trimodal model for $K \ge 2$ is stable if $\lambda_0 < \lambda_2$. In this case, its stationary distribution is:*

$$\pi_K^{tr}(\mathbf{n},\lambda_0,\lambda_2) = \frac{1}{G_K^{tr}(\lambda_0,\lambda_2)}\left(\frac{\lambda_0}{\lambda_2}\right)^{p(\mathbf{n})}, \quad (6)$$

where $G_K^{tr}(\lambda_0,\lambda_2) = G_K^b(\lambda_0,\lambda_2)$.

To prove Theorem 2, we follow a constructive approach by deriving the expression for the stationary distribution from the properties of ρ-reversible CTMCs, as reported the Appendix. First, we show that $X^{tr}(t)$ is ρ-reversible, under the assumption of ergodicity.

THEOREM 3. *Let $K \geq 2$. If $X_K^{tr}(t)$ is ergodic, then it is ρ-reversible (also dynamically reversible) under renaming:*

$$\rho(\mathbf{n}) = \mathbf{n}^R \tag{7}$$

where $\mathbf{n}^R = (-n_1, \ldots, -n_{K-1}, -n_K)$.

Before proving Theorem 3, we study some properties of random walks in $X_K^{tr}(t)$. Let u be a path starting from $\mathbf{n}$ and characterised by the arrivals of jobs at servers $(c_1, c_2, \ldots, c_T)$ with $c_i \in \{1 \ldots K\}$ and $T \in \mathbb{N}^+$:

$$u : \mathbf{n} \xrightarrow{\lambda_{c_1}} \mathbf{n}+\mathbf{e}_{c_1}-\mathbf{e}_{c_1^-} \xrightarrow{\lambda_{c_2}} \cdots \xrightarrow{\lambda_{c_t}} \mathbf{n}+\sum_{w=1}^{t}\mathbf{e}_{c_w}-\sum_{w=1}^{t}\mathbf{e}_{c_w^-} \cdots \xrightarrow{\lambda_{c_T}} \mathbf{n}+\sum_{w=1}^{T}\mathbf{e}_{c_w}-\sum_{w=1}^{T}\mathbf{e}_{c_w^-} .$$

The next proposition allows us to state that for each path u, we can define a reversed path u^R according to the renaming specified in Theorem 3.

PROPOSITION 1. *For each transition $\mathbf{n} \xrightarrow{\lambda_c} \mathbf{n}+\mathbf{e}_c-\mathbf{e}_{c^-}$ in the transition graph of $X_K^{tr}(t)$, there exists an* inverse *transition $(\mathbf{n}+\mathbf{e}_c-\mathbf{e}_{c^-})^R \xrightarrow{\lambda'_c} \mathbf{n}^R$ where:*

- *if $\lambda_c = \lambda_1$ then $\lambda'_c = \lambda_1$ and $p(\mathbf{n}+\mathbf{e}_c-\mathbf{e}_{c^-}) = p(\mathbf{n})$,*
- *if $\lambda_c = \lambda_0$ then $\lambda'_c = \lambda_2$ and $p(\mathbf{n}+\mathbf{e}_c-\mathbf{e}_{c^-}) = p(\mathbf{n})+1$,*
- *if $\lambda_c = \lambda_2$ then $\lambda'_c = \lambda_0$ and $p(\mathbf{n}+\mathbf{e}_c-\mathbf{e}_{c^-}) = p(\mathbf{n})-1$.*

PROOF. The proof is trivial. Indeed it is sufficient to observe that the inverse transition adds one unit in position c and removes one from position c^-. □

We are now in position to prove Theorem 3.

PROOF OF THEOREM 3. In order to prove that $X_K^{tr}(t)$ is dynamically reversible with respect to the renaming given by Equation (7), we have to prove that conditions $(K1)$ and $(K2)$ of Lemma 1 in the Appendix are satisfied.

In order to verify condition $(K1)$, we need to compute the residence time in each state $\mathbf{n} \in \mathcal{S}$. Indeed, the residence time in $\mathbf{n} \in \mathcal{S}$ is exponential with rate:

$$\sum_{k=1}^{K}\lambda(n_{k^-}, -n_k) = K\lambda_0 + \frac{(\lambda_2-\lambda_0)}{2}\sum_{i=1}^{K}\delta_{n_i\neq 0} .$$

where $\delta_{n_i\neq 0}$ is 1 if $n_i \neq 0$, and 0 otherwise. To simplify the proof, we use γ to denote the quantity $(\lambda_2-\lambda_0)/2$. Hence, the definition of $\lambda(x,y)$ can be expressed as follows:

$$\lambda(x,y) = \begin{cases} \lambda_0 & \text{if } x \leq 0 \wedge y \leq 0 \\ \lambda_0+\gamma & \text{if } (x \leq 0 \wedge y > 0) \vee (x > 0 \wedge y \leq 0) \\ \lambda_0+2\gamma & \text{if } x > 0 \wedge y > 0 \end{cases} .$$

The proof is by induction on K. If $K = 2$, then the residence time of $\mathbf{n}$ is exponentially distributed with rate:

$$\begin{aligned}\lambda(n_2,-n_1)+\lambda(n_1,-n_2) &= 2\lambda_0+\gamma(\delta_{n_2>0}+\delta_{n_1<0}+\delta_{n_1>0}+\delta_{n_2<0}) \\ &= 2\lambda_0+\gamma(\delta_{n_2\neq 0}+\delta_{n_1\neq 0}) \\ &= 2\lambda_0+\frac{(\lambda_2-\lambda_0)}{2}(\delta_{n_2\neq 0}+\delta_{n_1\neq 0}) .\end{aligned}$$

If $K > 2$, then by the inductive hypothesis we have that the rate of the residence time of $\mathbf{n}$ is:

$$\begin{aligned}\sum_{k=1}^{K}\lambda(n_{k^-},-n_k) &= (K-1)\lambda_0+\gamma\sum_{i=2}^{K}\delta_{n_i\neq 0} \\ &\quad -\lambda(n_K,-n_2)+\lambda(n_1,-n_2)+\lambda(n_K,-n_1) \\ &= (K-1)\lambda_0+\gamma\sum_{i=2}^{K}\delta_{n_i\neq 0}-\lambda_0-\gamma\delta_{n_K>0}-\gamma\delta_{n_2<0} \\ &\quad +\lambda_0+\gamma\delta_{n_1>0}+\gamma\delta_{n_2<0}+\lambda_0+\gamma\delta_{n_K>0}+\gamma\delta_{n_1<0} \\ &= K\lambda_0+\gamma\sum_{i=1}^{K}\delta_{n_i\neq 0} .\end{aligned}$$

Now, condition $(K1)$ is easy to verify. Since $\mathbf{n}^R$ has the same components of $\mathbf{n}$, but with different signs, they have the same number of non-zero components. Hence:

$$\sum_{k=1}^{K}\lambda(n_{k^-},-n_k) = \sum_{k=1}^{K}\lambda(n^R_{k^-},-n^R_k) .$$

In order to prove condition (K2), let u be a cycle from state $\mathbf{n}$ back to the same state $\mathbf{n}$, and u^R be the corresponding reverse cycle from $\mathbf{n}^R$ back to $\mathbf{n}^R$. We define $\psi(u)$ as the product of the transition rates that appear in u. Then, we prove that $\psi(u) = \psi(u^R)$. From Proposition 1, we know that for each λ_1 transition $\mathbf{n}_i \xrightarrow{\lambda_1} \mathbf{n}_i+\mathbf{e}_c-\mathbf{e}_{c^-}$ in the cycle u, there is a transition $(\mathbf{n}_i+\mathbf{e}_c-\mathbf{e}_{c^-})^R \xrightarrow{\lambda_1} \mathbf{n}_i^R$ in the cycle u^R, i.e., cycles u and u^R contain the same number of λ_1 transitions. Moreover, if we denote by $p(\mathbf{n})$ the sum of positive components of $\mathbf{n}$, then it holds that λ_1 transitions preserve this number, i.e., $p(\mathbf{n}_i) = p(\mathbf{n}_i+\mathbf{e}_c-\mathbf{e}_{c^-})$. Furthermore, for each λ_0 transition $\mathbf{n}_i \xrightarrow{\lambda_0} \mathbf{n}_i+\mathbf{e}_c-\mathbf{e}_{c^-}$ in u, there is a λ_2 transition $(\mathbf{n}_i+\mathbf{e}_c-\mathbf{e}_{c^-})^R \xrightarrow{\lambda_2} \mathbf{n}_i^R$ in u^R, and $p(\mathbf{n}_i+\mathbf{e}_c-\mathbf{e}_{c^-}) = p(\mathbf{n}_i)+1$. Finally, for each λ_2 transition $\mathbf{n}_i \xrightarrow{\lambda_2} \mathbf{n}_i+\mathbf{e}_c-\mathbf{e}_{c^-}$ in u, there is a λ_0 transition $(\mathbf{n}_i+\mathbf{e}_c-\mathbf{e}_{c^-})^R \xrightarrow{\lambda_0} \mathbf{n}_i^R$ in u^R, and $p(\mathbf{n}_i+\mathbf{e}_c-\mathbf{e}_{c^-}) = p(\mathbf{n}_i)-1$. Since u is a cycle, i.e., a sequence of transitions from state $\mathbf{n}$ back to the same state $\mathbf{n}$, then the total change in $p(\mathbf{n})$ is zero, so there must be an equal number of λ_0 and λ_2 transitions in cycle u. Hence, by Proposition 1, cycles u and u^R contain the same number of λ_0 and λ_2 transitions. □

Finally, we can prove Theorem 2.

PROOF OF THEOREM 2. From the fact that $X_K^{tr}(t)$ is dynamically reversible, we can derive the expression of the invariant measure associated with state $\mathbf{n}$ with respect to a reference state $\mathbf{0}$ as given by Lemma 2 in the Appendix. Let u be an arbitrary path from state $\mathbf{0}$ to state $\mathbf{n}$, and let u^R be its reversed path according to Proposition 1. Then:

$$\frac{\pi(\mathbf{n})}{\pi(\mathbf{0})} = \frac{\psi(u^R)}{\psi(u)} .$$

Consider an arbitrary state $\mathbf{n}$ and let T be the minimum possible number of arrivals that takes the model from state $\mathbf{n}$ to state $\mathbf{0}$. Notice that T is well-defined. We proceed by induction on T. If $T = 1$, then $\mathbf{n} = \mathbf{0} - \mathbf{e}_c + \mathbf{e}_{c^-}$ for some $1 \leq c \leq K$, and hence we have:

$$u : \mathbf{n} \xrightarrow{\lambda(1,1)} \mathbf{0} \qquad u^R : \mathbf{0} \xrightarrow{\lambda(0,0)} \mathbf{n}^R$$

which verifies Equation (6). If $T > 1$, then by Lemma 1 in the Appendix we can allow any arrival to get one step closer to the reference state $\mathbf{0}$. We choose c such that $n_c < 0$ and $n_{c^-} \geq 0$. In this case, we have:

$$\begin{aligned} u &: \mathbf{n} \xrightarrow{\lambda(n_{c^-},-n_c)} \mathbf{n} + \mathbf{e}_c - \mathbf{e}_{c^-} \\ u^R &: (\mathbf{n} + \mathbf{e}_c - \mathbf{e}_{c^-})^R \xrightarrow{\lambda(-n_{c^-}+1,n_c+1)} \mathbf{n}^R . \end{aligned}$$

Hence, by the inductive hypothesis, we have:

$$\pi(\mathbf{n} + \mathbf{e}_c - \mathbf{e}_{c^-}) = \frac{1}{G_K^{tr}(\lambda_0,\lambda_2)} \left(\frac{\lambda_0}{\lambda_2}\right)^{p(\mathbf{n})} \cdot \frac{\lambda(n_{c^-},-n_c)}{\lambda(-n_{c^-}+1,n_c+1)} . \quad (8)$$

If $\lambda(n_{c^-},-n_c) = \lambda_0$, then $\lambda(-n_{c^-}+1,n_c+1) = \lambda_2$, so we can rewrite Equation (8) as follows:

$$\begin{aligned} \pi(\mathbf{n} + \mathbf{e}_c - \mathbf{e}_{c^-}) &= \frac{1}{G_K^{tr}(\lambda_0,\lambda_2)} \left(\frac{\lambda_0}{\lambda_2}\right)^{p(\mathbf{n})} \frac{\lambda_0}{\lambda_2} \\ &= \frac{1}{G_K^{tr}(\lambda_0,\lambda_2)} \left(\frac{\lambda_0}{\lambda_2}\right)^{1+p(\mathbf{n})} . \end{aligned}$$

We know that if $\lambda(n_{c^-},-n_c) = \lambda_0$, then $p(\mathbf{n} + \mathbf{e}_c - \mathbf{e}_{c^-}) = p(\mathbf{n}) + 1$, so we have:

$$\pi(\mathbf{n} + \mathbf{e}_c - \mathbf{e}_{c^-}) = \frac{1}{G_K^{tr}(\lambda_0,\lambda_2)} \left(\frac{\lambda_0}{\lambda_2}\right)^{p(\mathbf{n}+\mathbf{e}_c-\mathbf{e}_{c^-})} .$$

If $\lambda(n_{c^-},-n_c) = \lambda_1$, then $\lambda(-n_{c^-}+1,n_c+1) = \lambda_1$, so we can rewrite Equation (8) as follows:

$$\begin{aligned} \pi(\mathbf{n} + \mathbf{e}_c - \mathbf{e}_{c^-}) &= \frac{1}{G_K^{tr}(\lambda_0,\lambda_2)} \left(\frac{\lambda_0}{\lambda_2}\right)^{p(\mathbf{n})} \frac{\lambda_1}{\lambda_1} = \\ &\frac{1}{G_K^{tr}(\lambda_0,\lambda_2)} \left(\frac{\lambda_0}{\lambda_2}\right)^{p(\mathbf{n})} . \end{aligned}$$

Indeed, if $\lambda(n_{c^-},-n_c) = \lambda_1$, then $p(\mathbf{n} + \mathbf{e}_c - \mathbf{e}_{c^-}) = p(\mathbf{n})$, so we have:

$$\pi(\mathbf{n} + \mathbf{e}_c - \mathbf{e}_{c^-}) = \frac{1}{G_K^{tr}(\lambda_0,\lambda_2)} \left(\frac{\lambda_0}{\lambda_2}\right)^{p(\mathbf{n}+\mathbf{e}_c-\mathbf{e}_{c^-})} .$$

Finally, if $\lambda(n_{c^-},-n_c) = \lambda_2$, then $\lambda(-n_{c^-}+1,n_c+1) = \lambda_0$, so we can rewrite Equation (8) as follows:

$$\begin{aligned} \pi(\mathbf{n} + \mathbf{e}_c - \mathbf{e}_{c^-}) &= \frac{1}{G_K^{tr}(\lambda_0,\lambda_2)} \left(\frac{\lambda_0}{\lambda_2}\right)^{p(\mathbf{n})} \frac{\lambda_2}{\lambda_0} \\ &= \frac{1}{G_K^{tr}(\lambda_0,\lambda_2)} \left(\frac{\lambda_0}{\lambda_2}\right)^{p(\mathbf{n})-1} , \end{aligned}$$

since if $\lambda(n_{c^-},-n_c) = \lambda_2$ then $p(\mathbf{n} + \mathbf{e}_c - \mathbf{e}_{c^-}) = p(\mathbf{n}) - 1$ then we have:

$$\pi(\mathbf{n} + \mathbf{e}_c - \mathbf{e}_{c^-}) = \frac{1}{G_K^{tr}(\lambda_0,\lambda_2)} \left(\frac{\lambda_0}{\lambda_2}\right)^{p(\mathbf{n}+\mathbf{e}_c-\mathbf{e}_{c^-})} . \quad \square$$

It is intriguing that $\pi_K^b(\mathbf{n},\lambda_0,\lambda_2) = \pi_K^{tr}(\mathbf{n},\lambda_0,\lambda_2)$ for all $\mathbf{n} \in \mathcal{S}_K$. This means that if the servers can work at two speeds $\lambda_0 < \lambda_2$, or at three speeds $\lambda_0 < (\lambda_0+\lambda_2)/2 < \lambda_2$, then the stationary probabilities of the join-queue length differences are identical. This fact has important immediate consequences that we summarise in the following corollaries.

COROLLARY 4. *The canonical trimodal model with the rates $\lambda_0 < \lambda_2$ has the same join-queue length distribution as the bimodal model with the same rates $\lambda_0 < \lambda_2$.*

PROOF. It is easy to observe that, under immediate join, there exists a bijection between the state of the systems encoded with the join-queue length differences and that encoded with the absolute join-queue length. As a consequence, the corollary is proved. $\square$

COROLLARY 5. *The canonical trimodal model with rates $\lambda_0 < \lambda_2$ has the same balance index as the bimodal model with the same rates $\lambda_0 < \lambda_2$:*

$$B_K^b(\lambda_0,\lambda_2) = B_K^{tr}(\lambda_0,\lambda_2) .$$

PROOF. Based on Definitions 2 and 5, the balance indices are derived directly from the stationary distribution. Thus, by Theorem 2, the result follows straightforwardly. $\square$

Moreover, it straightforwardly holds that the marginal probabilities of observing a positive, negative, or any arbitrary components in the canonical trimodal model are the same as the bimodal model with parameters λ_0, λ_2. However, it is not straightforward to compare the throughputs of the canonical trimodal model and the bimodal one. In fact, Definitions 3 and 6 are quite different. The following theorem states (quite surprisingly) that the throughputs of the bimodal and canonical trimodal model with the same parameters are the same.

THEOREM 4. *Consider the bimodal model with rates $\lambda_0 < \lambda_2$ and the canonical trimodal model with the same rates λ_0, λ_2. It holds that:*

$$T_K^b(\lambda_0,\lambda_2) = T_K^{tr}(\lambda_0,\lambda_2) .$$

PROOF. Since the system is stable, the total throughput of the system corresponds to the throughput of a single server. This can be computed as follows:

$$\begin{aligned} T_K^{tr}(\lambda_0,\lambda_2) &= \pi_K^{\leq,\geq}(\lambda_0,\lambda_2)\lambda_0 \\ &+ \left(\pi_K^{\leq,<}(\lambda_0,\lambda_2) + \pi_K^{>,\geq}(\lambda_0,\lambda_2)\right) \frac{\lambda_0+\lambda_2}{2} \\ &+ \pi_K^{>,<}(\lambda_0,\lambda_2)\lambda_2 , \end{aligned}$$

where:

$$\begin{aligned} \pi_K^{\leq,\geq}(\lambda_0,\lambda_2) &= \sum_{\substack{\mathbf{n}=(n_1,n_2,\cdots,n_K)\in\mathcal{S}_K: \\ n_1\leq 0, n_2\geq 0}} \pi_K^{tr}(\mathbf{n},\lambda_0,\lambda_2) , \\ \pi_K^{\leq,<}(\lambda_0,\lambda_2) &= \sum_{\substack{\mathbf{n}=(n_1,n_2,\cdots,n_K)\in\mathcal{S}_K: \\ n_1\leq 0, n_2< 0}} \pi_K^{tr}(\mathbf{n},\lambda_0,\lambda_2) , \\ \pi_K^{>,\geq}(\lambda_0,\lambda_2) &= \sum_{\substack{\mathbf{n}=(n_1,n_2,\cdots,n_K)\in\mathcal{S}_K: \\ n_1> 0, n_2\geq 0}} \pi_K^{tr}(\mathbf{n},\lambda_0,\lambda_2) , \\ \pi_K^{>,<}(\lambda_0,\lambda_2) &= \sum_{\substack{\mathbf{n}=(n_1,n_2,\cdots,n_K)\in\mathcal{S}_K: \\ n_1> 0, n_2< 0}} \pi_K^{tr}(\mathbf{n},\lambda_0,\lambda_2) . \end{aligned}$$

Analogously, the throughput for the bimodal model is:

$$T_K^b(\lambda_0,\lambda_2) = \left(\pi_K^{b+}(\lambda_0,\lambda_2) + \pi_K^{b*}(0,\lambda_0,\lambda_2)\right)\lambda_0 + \pi_K^{b-}(\lambda_0,\lambda_2)\lambda_2\,.$$

We next establish that $\pi_K^{\leq,<}(\lambda_0,\lambda_2) = \pi_K^{>,\geq}(\lambda_0,\lambda_2)$. We define the following bijection f between a state $\mathbf{n}$, where $n_1 \leq 0$ and $n_2 < 0$, and a state where $n_1 > 0$ and $n_2 \geq 0$:

$$f(\mathbf{n}) = (-n_2, -n_1, -n_3, \ldots, -n_K)\,.$$

Since $\pi_K^{tr}(\mathbf{n},\lambda_0,\lambda_2) = \pi_K^{tr}(f(\mathbf{n}),\lambda_0,\lambda_2)$, because for any state the sum of positive components equals the sum of negative components, we conclude that $\pi_K^{\leq,<}(\lambda_0,\lambda_2) = \pi_K^{>,\geq}(\lambda_0,\lambda_2)$. Therefore, we can write:

$$\begin{aligned}T_K^{tr}(\lambda_0,\lambda_2) &= \pi_K^{\leq,\geq}(\lambda_0,\lambda_2)\lambda_0 + \pi_K^{>,\geq}(\lambda_0,\lambda_2)(\lambda_0+\lambda_2) + \pi_K^{>,<}(\lambda_0,\lambda_2)\lambda_2\\ &= (\pi_K^{b-}(\lambda_0,\lambda_2) + \pi_K^{b*}(0,\lambda_0,\lambda_2) - \pi_K^{\leq,<}(\lambda_0,\lambda_2))\lambda_0 + \pi_K^{>,\geq}(\lambda_0,\lambda_2)\lambda_0 + \pi_K^{>,\geq}(\lambda_0,\lambda_2)\lambda_2 + (\pi_K^{b+}(\lambda_0,\lambda_2) - \pi_K^{>,\geq}(\lambda_0,\lambda_2))\lambda_2\,.\end{aligned}$$

Since, by Corollary 2, $\pi_K^{b-}(\lambda_0,\lambda_2) = \pi_K^{b+}(\lambda_0,\lambda_2)$ and, as proved above, $\pi_K^{\leq,<}(\lambda_0,\lambda_2) = \pi_K^{>,\geq}(\lambda_0,\lambda_2)$, we have:

$$\begin{aligned}T_K^{tr}(\lambda_0,\lambda_2) &= (\pi_K^{b+}(\lambda_0,\lambda_2) + \pi_K^{b*}(0,\lambda_0,\lambda_2))\lambda_0 - \pi_K^{>,\geq}(\lambda_0,\lambda_2)\lambda_0 + \pi_K^{>,\geq}(\lambda_0,\lambda_2)\lambda_0 + \pi_K^{>,\geq}(\lambda_0,\lambda_2)\lambda_2 + (\pi_K^{b-}(\lambda_0,\lambda_2) - \pi_K^{>,\geq}(\lambda_0,\lambda_2))\lambda_2\,.\\ &= (\pi_K^{b+}(\lambda_0,\lambda_2) + \pi_K^{b*}(0,\lambda_0,\lambda_2))\lambda_0 + \pi_K^{b-}(\lambda_0,\lambda_2)\lambda_2\\ &= T_K^b(\lambda_0,\lambda_2)\,.\end{aligned}$$

□

Up to this point, we have shown that the canonical trimodal model does not seem to give any benefit with respect to the bimodal despite its additional features. In practice, an important benefit of the trimodal model is reduced power consumption, as given by the following theorem.

THEOREM 5. *Let the power consumption of a server depend on its speed λ according to the $P(\lambda) = \lambda^\alpha$, with $\alpha \geq 1$. Then, if $\lambda_0 < \lambda_2$, the power consumption of the canonical trimodal model is always less than or equal to that of the bimodal model, with equality holding only if $\alpha = 1$.*

The proof is given in the appendix. The impact of the power saving will be analysed in Section 4.

4. EVALUATION RESULTS

In this section, we study the bimodal and trimodal rate adaptation policies using simulation. This allows us to study the system without the saturation assumption. We assume a Poisson arrival process with intensity λ. Tasks have iid exponential service times.

The first simulation experiment is designed to verify the conjecture that the balance index in the canonical trimodal model is equal to that of the bimodal model with equivalent rates. We consider a system with $K = 5$ servers, and rates $\lambda_0 = \eta = 3$ and $\lambda_2 = \mu = 5$. The maximum allowed throughput under the saturation assumption is $T_5^b(3,5)/5 = T_5^{tr}(3,5)/5 = 3.85477$. Then, we perform 15 independent simulation runs for different arrival rates, i.e., $\lambda = \rho T_5^b(3,5)$ with $\rho \in \{0.5, 0.7, 0.9, 0.95, 0.98, 0.99\}$. Finally, we constructed confidence intervals with 98% confidence level for the balance index, as shown in Table 1. The first observation is that the mean balance index values are very similar for the bimodal and trimodal models, and the confidence intervals overlap. In fact, this holds for all the values of ρ considered. This suggests that the analytical result that we derived on the balance index for the bimodal and canonical trimodal models may hold even when the model is not saturated, or differ only negligibly. The second observation is that when ρ approaches 1, the balance index approaches the theoretical value derived in Section 3. This provides cross-validation for the theoretical framework and the simulator. Similar observations apply for other values of K that we have considered.

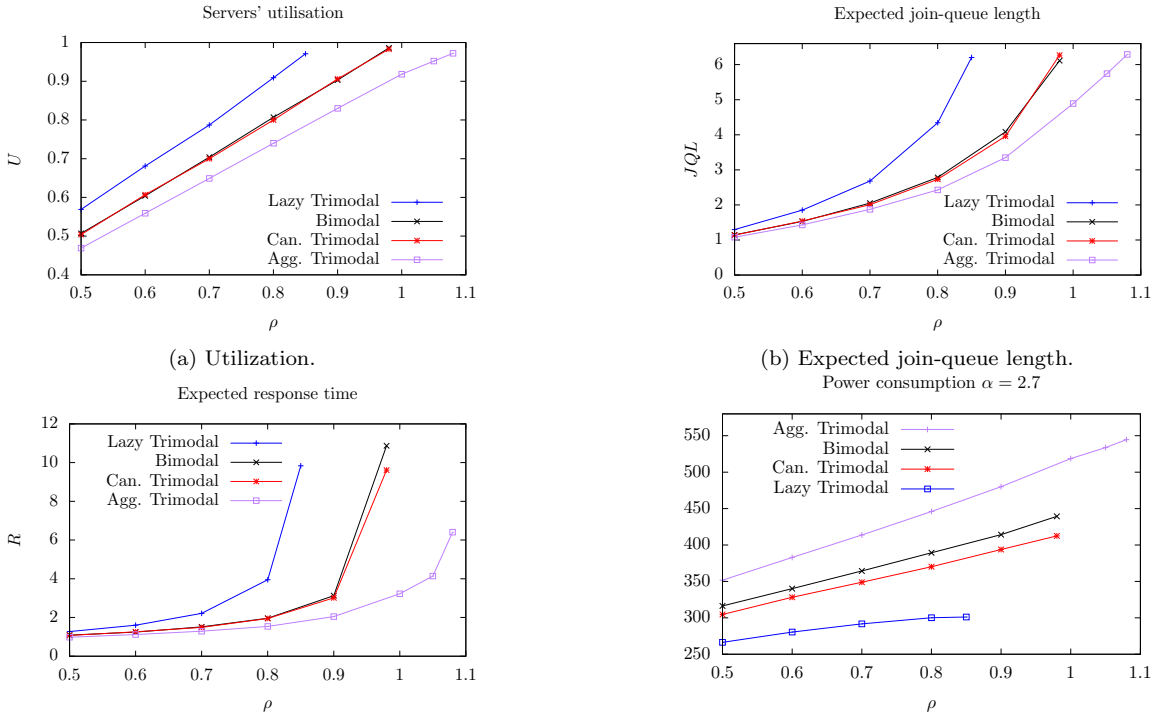
Table 1: Simulation comparison of balance index (98% C.I.) for canonical trimodal and bimodal without saturation. Analytic result for saturation shown in **bold**.

Bal.	**Can. Trimodal**			**Bimodal**		
ρ	avg	min	max	avg	min	max
0.50	0.351	0.350	0.353	0.353	0.352	0.354
0.70	0.571	0.568	0.575	0.575	0.571	0.580
0.90	1.056	1.043	1.069	1.055	1.056	1.065
0.95	1.262	1.251	1.273	1.259	1.244	1.274
0.98	1.418	1.402	1.433	1.429	1.411	1.447
0.99	1.462	1.442	1.481	1.475	1.458	1.491
1	**1.541**	n/a	n/a	**1.541**	n/a	n/a

The second experiment compares four different systems: bimodal, canonical trimodal, and two other configurations of the trimodal algorithm. One of the latter, called *lazy trimodal*, has $\lambda_1 = \lambda_0$, while the other, called *aggressive trimodal*, has $\lambda_1 = \lambda_2$. We consider $K = 10$, for which the computed maximum throughput of the canonical trimodal model with rates 3 and 5 is 3.8655. For all the models, the simulated job arrival rate is $\lambda = 3.8655\rho$. Figure 3a shows that the servers reach their maximum utilisation U at different arrival rates. Specifically, while the bimodal and the canonical trimodal have their maximum utilisation for $\rho = 1$ and show identical behaviour, the lazy and aggressive trimodal have lower and higher maximum throughput, respectively. The observation is confirmed by Figures 3b and 3c, which show the expected join-queue length (JQL) and the expected response time (R), respectively. Figure 3d shows the power consumption of the four models. The plot confirms Theorem 5 and shows that the canonical trimodal model has the same performance in terms of join-queue length, response time, and utilisation as the bimodal one, but with lower power consumption. For the lazy and aggressive trimodal models, there are tradeoffs between power consumption and other performance indices. For example, the aggressive trimodal policy increases the maximum throughput by almost 10% compared to the canonical trimodal, but does so with 13% higher power consumption.

5. CONCLUSION

In this paper, we have considered two rate adaptation algorithms for fork-join queues. The first, called *bimodal*,



(a) Utilization.

(b) Expected join-queue length.

(c) Expected response time.

(d) Power consumption.

Figure 3: Simulation results for bimodal and trimodal models.

changes the speed of a server according to the difference between its join-queue length and that of another server, while the second, called *trimodal*, chooses the speed of a server according to the differences between its join-queue length and those of two other servers. We have shown, analytically and by simulation, that the canonical trimodal model has the same performance as the bimodal in terms of subtask dispersion, utilisation, and expected join-queue length, but with lower power consumption. We explored by simulation other configurations of the server speeds, and discussed the tradeoffs between performance indices and power consumption. This may inform the design of a trimodal scheme that dynamically changes its speed to optimise the tradeoffs between response time, join-queue length, and power consumption according to the intensity of the incoming traffic.

6. REFERENCES

[1] L. Andrew, M. Lin, and A. Wierman. Optimality, fairness, and robustness in speed scaling designs. *SIGMETRICS Perform. Eval. Rev.*, 38(1):37–48, 2010.

[2] L. Flatto and S. Hahn. Two parallel queues created by arrivals with two demands I. *SIAM Journal on Applied Mathematics*, 44(5):1041–1053, 1984.

[3] D. J. Gates and M. Westcott. Kinetics of polymer crystallization I. discrete and continuum models. *Proc. of Royal Society London A*, 416:443–461, 1988.

[4] T. Hellemans and B. Van Houdt. On the power-of-d-choices with least loaded server selection. *Proc. ACM Meas. Anal. Comput. Syst.*, 2(2):27:1–27:22, June 2018.

[5] F. Kelly. *Reversibility and stochastic networks*. Wiley, New York, 1979.

[6] W. R. KhudaBukhsh, S. Kar, A. Rizk, and H. Koeppl. Provisioning and performance evaluation of parallel systems with output synchronization. *ACM Transactions on Performance Evaluation of Computer Systems TOMPECS*, 4(1):6:1–6:31, 2019.

[7] A. Marin and S. Rossi. Fair workload distribution for multi-server systems with pulling strategies. *Performance Evaluation*, 113:26–41, 2017.

[8] A. Marin and S. Rossi. On the relations between Markov chain lumpability and reversibility. *Acta Inf.*, 54(5):447–485, 2017.

[9] A. Marin and S. Rossi. Power control in saturated fork-join queueing systems. *Performance Evaluation*, 116:101–118, 2017.

[10] T. Rauber and G. Rünger. Energy-aware execution of fork-join-based task parallelism. In *In Proc. of the IEEE International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems, (MASCOTS)*, pages 231–240, 2012.

[11] A. Rizk, F. Poloczek, and F. Ciucu. Computable bounds in fork-join queueing systems. *ACM SIGMETRICS Performance Evaluation Review*, 43(1):335–346, 2015.

[12] I. Tsimashenka, W. J. Knottenbelt, and P. G. Harrison. Controlling variability in split-merge systems and its impact on performance. *Annals of Operational Research*, 239(2):569–588, 2016.

[13] P. Whittle. *Systems in Stochastic Equilibrium*. John Wiley & Sons, Inc., New York, NY, USA, 1986.

[14] P. E. Wright. Two parallel processors with coupled inputs. *Advances in Applied Probability*, 24(4):986–1007, 1992.

APPENDIX

A. RHO-REVERSIBILITY

We recall the concept of *ρ-reversibility* that we use to study the models presented in this paper.

A stationary CTMC $X(t)$ is said to be *reversible* if it is stochastically indistinguishable from $X(\tau - t)$ for all $\tau, t \in \mathbb{R}$. The concept of ρ-reversibility [8] extends the notion of reversibility by requiring that the time-reversed process $X(\tau - t)$ is stochastically indistinguishable from $X(t)$ when we apply a renaming ρ to its states. More precisely, if $\mathcal{S}$ is the state space of $X(t)$, the renaming function ρ is an arbitrary bijection from the state space to itself. When ρ is an involution, i.e., $\rho(\rho(s)) = s$ for all $s \in \mathcal{S}$, then we say that the system is dynamically reversible [5, 13]. Clearly, when ρ is the identity, $X(t)$ is reversible. Given the renaming ρ, proving that $X(t)$ is ρ-reversible can be structurally done by means of the Kolmogorov's criteria as stated below. For $s, s' \in \mathcal{S}$, we denote by $q(s, s')$ the transition rate from s to s', with $s \neq s'$.

LEMMA 1 (KOLMOGOROV'S CRITERIA). *Let $X(t)$ be a stationary CTMC with state space $\mathcal{S}$ and ρ be a renaming of $\mathcal{S}$. Then, $X(t)$ is ρ-reversible if and only if:*

($K1$) *for each $s \in \mathcal{S}$, $\sum_{\substack{s' \in \mathcal{S} \\ s' \neq s}} q(s, s') = \sum_{\substack{s' \in \mathcal{S} \\ s' \neq \rho(s)}} q(\rho(s), s')$,*

($K2$) *for any finite sequence of states $s_1, \ldots, s_n$ with $s_i \in \mathcal{S}$, we have*

$$q(s_1, s_2)q(s_2, s_3) \cdots q(s_{n-1}, s_n)q(s_n, s_1) = q(\rho(s_1), \rho(s_n))q(\rho(s_n), \rho(s_{n-1})) \cdots q(\rho(s_2), \rho(s_1)) .$$

Informally, ($K1$) requires that the residence time in a state and in its renaming are stochastically identical, while ($K2$) requires that, given any cycle of transitions in the CTMC, the product of its rates equals the product of the rates of the inverse cycle in the renamed CTMC. Analogously to standard reversibility, there exists an efficient way for computing the stationary distribution of ρ-reversible chains.

LEMMA 2 (STATIONARY DISTRIBUTION). *Let $X(t)$ be a ρ-reversible CTMC with state space $\mathcal{S}$, $\boldsymbol{\pi}$ its stationary distribution and let $r, s \in \mathcal{S}$. Then, for each sequence of transitions taking the chain from state r to state s*

$$r \equiv s_1 \xrightarrow{q(s_1,s_2)} s_2 \xrightarrow{q(s_2,s_3)} \cdots \xrightarrow{q(s_{n-1},s_n)} s_n \equiv s ,$$

we have:

$$\pi(s) = \pi(r) \frac{\prod_{i=1}^{n-1} q(\rho(s_{i+1}), \rho(s_i))}{\prod_{i=1}^{n-1} q(s_i, s_{i+1})} .$$

B. PROOFS

B.1 Proof of Theorem 1

Let us consider the case $n = 0$. Without loss of generality, we want to sum the stationary probabilities of all the states in $\mathcal{S}_K$ of the form $\mathbf{n} = (0, n_2, \ldots, n_K)$. Notice that all the states belonging to this set that have the same sum of the positive components have the same stationary probability. Given that the $K - 1$ variable components of the states in this class have exactly j non-negative components that sum to p, the number of states sharing the same stationary probabilities are:

$$\binom{p+j-1}{j-1}\binom{p-1}{K-j-2}\binom{K-1}{j},$$

where the first binomial coefficient counts the number of non-negative integer solutions to the equation $x_1 + x_2 + \ldots + x_j = p$, and the second one the number of positive solutions to $x_1 + x_2 + \ldots + x_{K-j-1} = p$. Finally, the third binomial coefficient counts the number of possible configurations that have exactly j non-negative components in the $K - 1$ last positions of the state vector.

Thus, the marginal distribution of the binomial model can be computed as follow:

$$G_K^b(\eta, \mu)\pi_K^{b*}(0, \eta, \mu) = 1 + \sum_{j=1}^{K-2} \sum_{p=K-j-1}^{\infty} \binom{p+j-1}{j-1} \cdot \binom{p-1}{K-j-2}\binom{K-1}{j}\left(\frac{\eta}{\mu}\right)^p ,$$

where 1 accounts for the state consisting of all 0s.

By considering the Pochhammer's symbol $(y)_n$ defined as:

$$(y)_n = y(y+1)\cdots(y+n-1) .$$

and the Taylor's expansion of the incomplete Beta-function:

$$\beta(x, a, b) = x^a \sum_{n=0}^{\infty} \frac{(1-b)_n}{n!(a+n)} x^n ,$$

then further simplifications bring to:

$$\begin{aligned} G_K^b(\eta, \mu)\pi_K^{b*}(0, \eta, \mu) &= 1 + \sum_{j=1}^{K-2} \sum_{w=0}^{\infty} \binom{w+K-2}{j-1} \cdot \binom{w+K-j-2}{K-j-2}\binom{K-1}{j}\left(\frac{\eta}{\mu}\right)^{w+K-j-1} \\ &= 1 + \sum_{j=1}^{K-2} \binom{K-1}{j} \sum_{w=0}^{\infty} \frac{(w+K-2)!}{(j-1)!(w+K-j-1)!} \cdot \frac{(w+K-j-2)!}{(K-j-2)!w!}\left(\frac{\eta}{\mu}\right)^{w+K-j-1} \\ &= 1 + \sum_{j=1}^{K-2} \binom{K-1}{j}\left(\frac{\eta}{\mu}\right)^{K-j-1} \cdot \sum_{w=0}^{\infty} \frac{(K-2)!(K-1)_w}{(j-1)!(w+K-j-1)(K-j-2)!w!}\left(\frac{\eta}{\mu}\right)^w \end{aligned}$$

$$\begin{aligned} &= 1 + \sum_{j=1}^{K-2} \binom{K-1}{j}\binom{K-2}{j-1}(K-j-1) \cdot \left(\frac{\eta}{\mu}\right)^{K-j-1} \sum_{w=0}^{\infty} \frac{(K-1)_w}{(w+K-j-1)w!}\left(\frac{\eta}{\mu}\right)^w \\ &= 1 + \sum_{j=1}^{K-2} \binom{K-1}{j}\binom{K-2}{j-1}(K-j-1) \cdot \beta\left(\eta/\mu, K-j-1, 2-K\right) . \end{aligned}$$

Let us now consider the case $n > 0$. Recall that, by Corollary 2, $\pi_K^{b*}(n,\eta,\mu) = \pi_K^{b*}(-n,\eta,\mu)$. Consider $\pi_K^{b*}(-n,\eta,\mu)$. As before we want to sum the stationary probabilities of all the states of the form $\mathbf{n} = (-n, n_2, \ldots, n_K)$ for $n \neq 0$ and $n \in \mathbb{N}$. Notice that all the states belonging to this set that have the same sum of the positive components have the same stationary probability. Given that the $K-1$ variable components of the states in the class have exactly j non-negative components that sum to p, the number of states sharing the same stationary probabilities are:

$$\binom{p+j-1}{j-1}\binom{p-n-1}{K-j-2}\binom{K-1}{j},$$

where the first binomial coefficient counts the number of non-negative integer solutions to the equation $x_1+x_2+\ldots+x_j = p$, and the second one the number of positive solutions to $x_1+x_2+\ldots+x_{K-j-1} = p-n$. Finally, the third binomial coefficient counts the number of possible configurations that have exactly j non-negative components in the $K-1$ last positions of the state vector.

Hereafter we consider the Gauss hypergeometric function ${}_2F_1$, defined as

$${}_2F_1(a,b,c,z) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n\ n!} z^n .$$

For $|z| < 1$ and generic parameters a, b and c, the above infinite sum is convergent.

Thus, the marginal distribution of the binomial model for $n > 0$ (resp., $n < 0$) can be computed as follow:

$$\begin{aligned} &G_K^b(\eta,\mu)\pi_K^{b*}(n,\eta,\mu) = G_K^b(\eta,\mu)\pi_K^{b*}(-n,\eta,\mu) \\ &= \binom{n+K-2}{K-2}\left(\frac{\eta}{\mu}\right)^n + \sum_{j=1}^{K-2}\sum_{p=K-j-1+n}^{\infty}\binom{p+j-1}{j-1} \\ &\qquad \cdot \binom{p-n-1}{K-j-2}\binom{K-1}{j}\left(\frac{\eta}{\mu}\right)^p \end{aligned}$$

where the first addend accounts for $j = K-1$. Then

$$\begin{aligned} &G_K^b(\eta,\mu)\pi_K^{b*}(n,\eta,\mu) = G_K^b(\eta,\mu)\pi_K^{b*}(-n,\eta,\mu) \\ &= \binom{n+K-2}{K-2}\left(\frac{\eta}{\mu}\right)^n + \sum_{j=1}^{K-2}\sum_{w=0}^{\infty}\binom{w+K+n-2}{j-1} \\ &\quad \cdot \binom{w+K-j-2}{K-j-2}\binom{K-1}{j}\left(\frac{\eta}{\mu}\right)^{w+K-j-1+n} \\ &= \binom{n+K-2}{K-2}\left(\frac{\eta}{\mu}\right)^n + \sum_{j=1}^{K-2}\binom{K-1}{j}\left(\frac{\eta}{\mu}\right)^{K-j-1+n} \\ &\cdot \sum_{w=0}^{\infty}\frac{(w+K+n-2)!}{(j-1)!(w+K+n-j-1)!}\frac{(w+K-j-2)!}{(K-j-2)!w!}\left(\frac{\eta}{\mu}\right)^w \\ &= \binom{n+K-2}{K-2}\left(\frac{\eta}{\mu}\right)^n + \sum_{j=1}^{K-2}\binom{K-1}{j}\left(\frac{\eta}{\mu}\right)^{K-j-1+n} \\ &\qquad \cdot \sum_{w=0}^{\infty}\frac{(K+n-2)!(K+n-1)_w}{(j-1)!(K+n-j-1)!(K+n-j)_w} \\ &\qquad \cdot \frac{(K-j-2)!(K-j-1)_w}{(K-j-2)!w!}\left(\frac{\eta}{\mu}\right)^w \\ &= \binom{n+K-2}{K-2}\left(\frac{\eta}{\mu}\right)^n + \sum_{j=1}^{K-2}\binom{K-1}{j}\binom{K+n-2}{j-1} \\ &\cdot \left(\frac{\eta}{\mu}\right)^{K-j-1+n}\sum_{w=0}^{\infty}\frac{(K+n-1)_w(K-j-1)_w}{(K+n-j)_w\ w!}\left(\frac{\eta}{\mu}\right)^w \\ &= \binom{n+K-2}{K-2}\left(\frac{\eta}{\mu}\right)^n + \sum_{j=1}^{K-2}\binom{K-1}{j}\binom{K+n-2}{j-1} \\ &\cdot \left(\frac{\eta}{\mu}\right)^{K-j-1+n}{}_2F_1\left(K+n-1, K-j-1, K+n-j, \frac{\eta}{\mu}\right). \end{aligned}$$

Finally, in order to compute the aggregated probabilities of observing a negative state in one of the servers we want to sum the stationary probabilities of all the states of the form $\mathbf{n} = (n_1, n_2, \ldots, n_K)$ where there is at least one negative n_i. Notice that all the states belonging to this set that have the same sum of the positive components have the same stationary probability. Given that the K variable components of the states in the class have exactly j non-negative components that sum to p, the number of states sharing the same stationary probabilities are:

$$\binom{p+j-1}{j-1}\binom{p-1}{K-j-1}\binom{K-1}{j},$$

where the first binomial coefficient counts the number of non-negative integer solutions to the equation $x_1 + x_2 + \ldots + x_j = p$, and the second one the number of solutions to $x_1 + x_2 + \ldots + x_{K-j} = p$. Finally, the third binomial coefficient counts the number of possible configurations that have exactly j non-negative components in the $K-1$ last positions of the state vector.

Thus, the aggregated probability of observing a positive (resp., a negative) state in one of the servers can be computed as follow:

$$\begin{aligned} &G_K^b(\eta,\mu)\pi_K^{b^+}(\eta,\mu) = G_K^b(\eta,\mu)\pi_K^{b^-}(\eta,\mu) = \\ &\sum_{j=1}^{K-1}\sum_{p=K-j}^{\infty}\binom{p+j-1}{j-1}\cdot\binom{p-1}{K-j-1}\binom{K-1}{j}\left(\frac{\eta}{\mu}\right)^p \end{aligned}$$

$$
\begin{aligned}
&= \sum_{j=1}^{K-1} \sum_{w=0}^{\infty} \binom{w+K-1}{j-1} \binom{w+K-j-1}{K-j-1} \\
&\qquad \cdot \binom{K-1}{j} \left(\frac{\eta}{\mu}\right)^{w+K-j} \\
&= \sum_{j=1}^{K-1} \binom{K-1}{j} \left(\frac{\eta}{\mu}\right)^{K-j} \sum_{w=0}^{\infty} \frac{(w+K-1)!}{(j-1)!(w+K-j)!} \\
&\qquad \cdot \frac{(w+K-j-1)!}{(K-j-1)!w!} \left(\frac{\eta}{\mu}\right)^{w} \\
&= \sum_{j=1}^{K-1} \binom{K-1}{j} \left(\frac{\eta}{\mu}\right)^{K-j} \sum_{w=0}^{\infty} \frac{(K-1)!K_w}{(j-1)!(w+K-j)!} \\
&\qquad \cdot \frac{(w+K-j-1)!}{(K-j-1)!w!} \left(\frac{\eta}{\mu}\right)^{w} \\
&= \sum_{j=1}^{K-1} \binom{K-1}{j} \binom{K-1}{j-1} (K-j) \left(\frac{\eta}{\mu}\right)^{K-j} \\
&\qquad \cdot \sum_{w=0}^{\infty} \frac{K_w}{(w+K-j)w!} \left(\frac{\eta}{\mu}\right)^{w} \\
&= \sum_{j=1}^{K-1} \binom{K-1}{j} \binom{K-1}{j-1} (K-j) \left(\frac{\eta}{\mu}\right)^{K-j} \\
&\qquad \cdot \beta\Big(\eta/\mu, K-j, 1-K\Big) .
\end{aligned}
$$

□

B.2 Proof of Theorem 5

The case $\alpha = 1$ is trivial and immediately follows from Theorem 4. Let us assume $\alpha > 1$. We can write the power consumption of a single server for the standard trimodal model as:

$$
\begin{aligned}
P_K^{tr}(\lambda_0, \lambda_2) &= \pi_K^{\leq,\geq}(\lambda_0, \lambda_2)\lambda_0^{\alpha} \\
&\quad + \left(\pi_K^{\leq,<}(\lambda_0, \lambda_2) + \pi_K^{>,\geq}(\lambda_0, \lambda_2)\right) \left(\frac{\lambda_0+\lambda_2}{2}\right)^{\alpha} \\
&\quad + \pi_K^{>,<}(\lambda_0, \lambda_2)\lambda_2^{\alpha} \\
&= \pi_K^{\leq,\geq}(\lambda_0, \lambda_2)\lambda_0^{\alpha} + \pi_K^{\leq,<}(\lambda_0, \lambda_2)2^{1-\alpha}(\lambda_0+\lambda_2)^{\alpha} \\
&\quad + \pi_K^{>,<}(\lambda_0, \lambda_2)\lambda_2^{\alpha} .
\end{aligned}
$$

Thanks to the relations introduced in the proof of Theorem 4, we can write:

$$
\begin{aligned}
P_K^{tr}(\lambda_0, \lambda_2) &= \pi_K^{b-}(\lambda_0, \lambda_2)\lambda_0^{\alpha} + \pi_K^{b*}(0, \lambda_0, \lambda_2)\lambda_0^{\alpha} \\
&\quad - \pi_K^{\leq,<}(\lambda_0, \lambda_2)\lambda_0^{\alpha} + \pi_K^{>,\geq}(\lambda_0, \lambda_2)2^{1-\alpha}(\lambda_0+\lambda_2)^{\alpha} \\
&\quad + \pi_K^{b+}(\lambda_0, \lambda_2)\lambda_2^{\alpha} - \pi_K^{>,\geq}(\lambda_0, \lambda_2)\lambda_2^{\alpha}
\end{aligned}
$$

Since we have that:

$$
\begin{aligned}
P_K^{b}(\lambda_0, \lambda_2) &= \pi_K^{b-}(\lambda_0, \lambda_2)\lambda_0^{\alpha} + \pi_K^{b*}(0, \lambda_0, \lambda_2)\lambda_0^{\alpha} \\
&\quad + \pi_K^{b+}(\lambda_0, \lambda_2)\lambda_2^{\alpha} ,
\end{aligned}
$$

to prove that $P_K^{tr}(\lambda_0, \lambda_2) < P_K^{b}(\lambda_0, \lambda_2)$, we need to show that:

$$
2^{1-\alpha}(\lambda_0+\lambda_2)^{\alpha} - \lambda_2^{\alpha} - \lambda_0^{\alpha} < 0 .
$$

Define the following function in the variables λ_0 and λ_2:

$$
y = 2^{1-\alpha}(\lambda_0+\lambda_2)^{\alpha} - \lambda_2^{\alpha} - \lambda_0^{\alpha} ,
$$

and observe that for $\lambda_0 = \lambda_2$ we have $y = 0$. We may consider the set of planes with equations $\lambda_0 + \lambda_2 - d = 0$, for all $d \geq 0$. The intersection of this plane with function y is function g in variable λ_0 defined as:

$$
g = 2^{1-\alpha} d^{\alpha} - \lambda_0^{\alpha} - (d-\lambda_0)^{\alpha} ,
$$

whose domain is $\lambda_0 < d$. We have:

$$
g' = \alpha(d-\lambda_0)^{\alpha-1} - \alpha\lambda_0^{\alpha-1} ,
$$

and:

$$
g'' = \alpha(\alpha-1)(-(d-\lambda_0)^{\alpha-2} - \lambda_0^{\alpha-2}) ,
$$

i.e., g is concave on its domain and its derivative is zero for $\lambda_0 = d/2$. Therefore, $g \leq 0$ in its domain for all $d > 0$ and it is zero only for $\lambda_0 = d/2$, i.e., $\lambda_0 = \lambda_2$, as required. □