Dynamic server allocation with observation delays

Shubhanshu Shekhar
shshekha@eng.ucsd.edu

1 Introduction

Consider a queueing system with N parallel queues and one server, with each queue having its own arrival and connectivity processes $(a_i(t))_{i=1}^N$ and $(C_i(t))_{i=1}^N$. The queue update rule is given by

$$X(t+1) = (X(t) - u(t).C(t))^+ + a(t)$$

where $u(t)$ is the control action at time t and $C(t)$ denotes the connectivity vector. The control action takes values in the set $\mathcal{A} = \{e_0, e_1, \ldots, e_N\}$ where e_i is the i^{th} standard normal unit vector, and e_0 is the all zero vector. The operation $u(t).C(t)$ above denotes the term-wise product.

In this paper, we study the problem of server allocation to the queues in the case where the current queue length information is not known to the controller. Instead the controller at time t has access to the queue length information from time $t - D$ for a fixed $D \in \mathbb{N}$. The paper contains the following results:

1. In Sec. 3.1, we show that the optimal policy with $D = 1$ is selecting the queue with the longest deterministic queue backlog. The term *deterministic queue length* was used in [1] to denote the quantity $V_i(t) = (X_i(t-D) - u(t-D))^+$, because for the case of $D = 1$, the current (unobserved) queue length can be written as $X_i(t) = V_i(t) + a_i(t-1)$.

2. We then show in Sec. 3.2 that for the $D = 2$ case, the distribution of the random variables $X_i(t)$ for $i = 1, 2, \ldots, N$ conditioned on the current observation $Z(t)$, can be sorted according to the usual stochastic ordering, and the policy which chooses the queue which is largest according to this stochastic ordering is optimal.

3. In Sec. 3.3, we finally provide the optimal policy for the case of general observation delay $D > 0$, under the assumption that $N \geq D + 1$. We show that even for arbitrary constant delay, the optimal policy is to choose the queue for which the conditional distribution of the current queue length given the current observations stochastically dominates all other queues. Unlike the two previous cases, we use backward induction [2] to prove this result.

4. The three results above are true under the assumption of symmetric arrival and connectivity processes. In Sec. 4, we use simulations to study the behavior of two heuristic policies in the case where the arrivals are not symmetric.

2 Observation Model and Notations

We now formally describe the observation model and the control objective. Let us consider a system of N parallel queues and one server. Each queue has an associated arrival process $\big(a_i(t)\big)_{t=-(D-1)}^{\infty}$ and a connectivity process $\big(C_i(t)\big)_{t=-(D-1)}^{\infty}$. In this paper, we provide the results for the case of symmetric arrivals and connectivity process. For simplicity, we present the results for the case of *i.i.d.* Bernoulli arrivals, i.e. $a_i(t) = Bernoulli(\lambda)$ for all $i \in [N]$ and $t \in \mathbb{N}$. These results can be extended to the case of arrival processes with finite number of arrivals at each time step.

We assume that at any time $t \in \mathbb{N}$, the observation available to the controller is $Z(t) = \{X(t-D), C(t-D), C(t-D+1), \ldots, C(t), u(t-D), u(t-D+1), \ldots, u(t-1)\}$. The aim of the controller is to obtain a policy which chooses the actions based on the available information at all times t, so as to minimize the following cost function

$$J^\pi(Z(1)) = \sum_{t=1}^{\infty} \beta^{t-1} \mathbb{E}[\sum_{i=1}^{N} X_i^\pi(t) | Z(t)] \tag{1}$$

For a system with observations $Z(t)$ and queue length vectors $X(t)$, we will use the following notations for the rest of the paper:

- we use $\bar{X}(t)$ to represent a random vector which has the same distribution as the conditional distribution of the current queue length vector $X(t)$ given the current observations $Z(t)$. In other words, for any set B, we have $Pr(\bar{X}(t) \in B) = \mathbb{E}[1_B(X(t))|Z(t)]$.
- We use $d_i(t)$ to denote the number of times the queue i was successfully allocated the server in the last D time steps, i.e.,
$$d_i(t) = \sum_{s=t-D}^{t-1} u_i(s) C_i(s)$$
- For two real valued random variables X and Y, we use $X \preccurlyeq_{st} Y$ to represent the usual stochastic ordering, i.e.,
$$X \preccurlyeq_{st} Y \Leftrightarrow 1 - F_X(x) \leq 1 - F_Y(x) \quad \forall x \in \mathbb{R}$$

Definition 1. Let γ_1 and γ_2 be two admissible policies and let $X(t)$ and $Y(t)$ respectively denote the queue length random vectors under these two policies. Then, we say that policy γ_1 *dominates* γ_2 if the following condition is satisfied for all $t \in \mathbb{N}_0$.

$$\sum_{i=1}^{N} \bar{X}_i(t) \preccurlyeq_{st} \sum_{i=1}^{N} \bar{Y}_i(t) \tag{2}$$

where $\bar{X}(t)$ and $\bar{Y}(t)$ are random vectors having the same distributions as the conditional distributions of the current queue length vector given the observations at time t under policies γ_1 and γ_2 respectively.

In the rest of this paper, use the term $A_k^{(i)}(t)$ to denote $\sum_{\tau=t-k}^{t-1} a_i(\tau)$, i.e., the sum of the last k arrivals to queue i at time t . Suppose the observation at time t is such that $X_i(t-D) - d_i(t) = -l < 0$. Of these $d_i(t)$ server allocations, let the last l allocations occur at times $t-D+k_1, t-D+k_1+k_2, \ldots, t-D+k_1+\ldots+k_l$. Then, $\bar{X}_i(t) \stackrel{d}{=} A_D^{(i)}(t) - l + I_{k_1}^{(1)}(t) + I_{k_1,k_2}^{(2)}(t) + \ldots I_{k_1,k_2,\ldots,k_l}^{(l)}(t)$. Here, $I_{k_1,k_2,\ldots,k_m}^{(m)}(t)$ for $m \in [l]$ are indicators of idling, i.e., $I_{k_1,k_2,\ldots,k_m}^{(m)}(t) = 1$ if $A_{k_m}^{(i)}(t-D+k_1+\ldots+k_m) = 0$ and $A_{k_1+k_2+\ldots+k_{m-1}}^{(i)}(t-D+k_1+\ldots+k_{m-1}) \leq m-1$.

As an example of the notations just introduced, suppose $D = 10$, $X_i(t-D) = 0$, $d_i(t) = 2$, $k_1 = 3$ and $k_2 = 4$. Then we have $\bar{X}_i(t) \stackrel{d}{=} \big((A_3^{(i)}(t-7) - 1)^+ + A_4^{(i)}(t-3) - 1\big)^+ + A_3^{(i)}(t) \stackrel{d}{=} A_{10}^{(i)}(t) - 2 + I_3^{(1)}(t) + I_{3,4}^{(2)}(t)$, where $I_3^{(1)}(t) = 1$ if $A_3^{(i)}(t-7) = 0$ and $I_{3,4}^{(2)}(t) = 1$ if $A_3^{(i)}(t-7) \leq 1$ and $A_4^{(i)}(t-3) = 0$.

Definition 2. Suppose $Z^{\gamma_1}(t)$ and $Z^{\gamma_2}(t)$ denote the observation processes under two admissible control policies γ_1 and γ_2, and let $X(t)$ and $Y(t)$ denote the corresponding queue length processes. Also let $[N]$ denote the set $\{1, 2, \ldots, N\}$. Then we define the following properties:

- **P1:** We say that $P1(Z^{\gamma_1}(t), Z^{\gamma_2}(t))$ holds, if there exist $i, j \in [N]$ and a permutation σ on $[N]$, such that

$$\begin{aligned}
&\bar{X}_{\sigma(l)}(t) = \bar{Y}_l(t), \quad \forall l \in [N] \setminus \{i, j\} \\
&\bar{X}_{\sigma(i)}(t) \preccurlyeq_{st} \bar{Y}_i(t) \\
&\bar{Y}_j(t) \preccurlyeq_{st} \bar{X}_{\sigma(j)}(t) \\
&\bar{X}_{\sigma(j)}(t) \preccurlyeq_{st} \bar{X}_{\sigma(i)}(t) \\
&\bar{X}_{\sigma(i)}(t) + \bar{X}_{\sigma(j)}(t) \preccurlyeq_{st} \bar{Y}_i(t) + \bar{Y}_j(t)
\end{aligned}$$

  This property represents the fact that based on the observations at time t, the queue under γ_1 is *more balanced* than the queue under the policy γ_2.

- **P2:** We say that $P2(Z^{\gamma_1}(t), Z^{\gamma_2}(t))$ holds if there exists a permutation σ on $[N]$ such that $\bar{X}_i(t) = \bar{Y}_{\sigma(i)}(t)$ for all $i \in [N]$.

- **P3:** We say that $P3(Z^{\gamma_1}(t), Z^{\gamma_2}(t))$ holds if $\bar{X}_{\sigma(i)}(t) \preccurlyeq_{st} \bar{Y}_i(t)$ for all $i \in [N]$ and for some permutation σ on $[N]$.

If any of the above three conditions are true at any time t, then we note that $\sum_{i=1}^{N} \bar{X}_i(t) \preccurlyeq_{st} \sum_{i=1}^{N} \bar{Y}_i(t)$. Thus for proving that a policy γ_1 dominates policy γ_2, we will show that one of the three conditions $P_i(Z^{\gamma_1}(t), Z^{\gamma_2}(t))$ for $i = 1, 2, 3$ is true at all times $t \in \mathbb{N}$.

3 Main results

In this section, we characterize the optimal policies for the queueing system with delayed observations. We first obtain the policies for simple cases of $D = 1$ and $D = 2$. The proof of optimality for these two cases are obtained by modifying the arguments used in [3] to handle the observation delays.

The optimal policies for these two simple cases provide some intuition about the optimal policy for the case of an arbitrary delay $D > 0$. In Sec.3.3, under some mild assumptions on the parameters D and N, we characterize the optimal policy for the case of general $D > 0$.

3.1 One step delay ($D = 1$)

Let us first consider the simplest case of one step delay. Here the observations are given by $Z(t) = \{X(t-1), U(t-1), C(t-1)\}$. Let $V_i(t) := (X_i(t-1) - U_i(t-1)C_i(t-1))^+$ denote the *deterministic queue* length as defined in [1]. Then we claim that the policy π_1 of choosing the queue with the largest value of $V_i(t)$ is optimal among the class of all admissible policies

Proposition 1. *For the one step delay case, the policy π_1 dominates (see Def. 1) any arbitrary admissible policy γ.*

The proof of this result proceeds quite similarly to the proof of the optimality of the Longest Connected First (LCF) policy given in [3]. As in [3], we proceed in the following two steps:

- First we show that for any given policy γ, there exists a policy γ_1 which selects the first action according to the policy π_1, and which dominates γ.

- Then by using forward induction in time, we show that the policy π_1 dominates all other admissible policies.

Since we are considering symmetric arrivals it suffices for us to show that $\sum_{i=1}^{N} V_i^{\pi_1}(t) \leq \sum_{i=1}^{N} V_i^{\gamma}(t)$ for all $t \in \mathbb{N}$. The update rule for $V^{\pi_1}(t)$ is given by

$$V_i^{\pi_1}(t+1) = \left(V_i^{\pi_1}(t) + a_i(t) - u_i(t)C_i(t)\right)^+$$

$U_i(t-2).C_i(t-2)$	$U_i(t-1).C_i(t-1)$	X(t)
0	0	$a_i(t-2)+a_i(t-1)$
1	0	$a_i(t-2)+a_i(t-1)$
0	1	$a_i(t-1)$
1	1	$a_i(t-1)$

Table 1: The table shows the different possible forms of $X_i(t)$ if the observed queue length $X_i(t-2)=0$ at time t.

Thus it is equivalent to a zero delay observation model in which the arrival occurs before the control action is applied, and the proof of this result is very similar to that in [3]. The details of the proof are given in Appendix- A for completion.

3.2 Two steps delay ($D = 2$)

Now we proceed to the case of two steps observation delay. As in the $D = 1$ case, we can again write the current queue length $X(t)$ as a sum of a deterministic part $V(t)$ and a stochastic part. However, unlike the $D = 1$ case, the stochastic part of the current queue length random variables are not iid.

$$X_i(t) = \big((X_i(t-2) - u_i(t-2).C_i(t-2))^+ + a_i(t-2) - u_i(t-1).C_i(t-1)\big)^+ + a_i(t-1) \quad (3)$$

From Eq. 3, we can see that if $X_i(t-2) - d_i(t) \geq 0$, then we have $V_i(t) = X_i(t-2) - d_i(t)$ and $X_i(t) = V_i(t) + A_2^{(i)}(t)$. However, if $X_i(t) - d_i(t) < 0$, then $V_i(t) = 0$, and the stochastic part will either be $a_i(t-1)$ or $A_2^{(i)}(t)$ depending upon the previous actions, as shown in Table 1.

Thus, for $D = 2$, it is not sufficient to consider just the deterministic part of the queue length when $X_i(t-2) - d_i(t) < 0$; we also have to consider the stochastic part. Let us define a policy π_2 which takes the following control actions:

- If the set $S_t = \{i : X_i(t-2) - d_i(t) \geq 0\}$ is not empty, then it chooses $\arg\max_{i \in S_t}\big(X_i(t-2) - d_i(t)\big)$.
- Otherwise, it selects a queue i such that $X_i(t) = a_i(t-2) + a_i(t-1)$. Note that if $N \geq 2$, then such a queue always exists.

First we state the following result, which allows us to prove that the policy π_2 dominates all other admissible policies.

Lemma 1. *Let γ be any admissible policy for the $D = 2$ case. Then there exists a policy γ_1 dominating γ (see Def. 1), which at $t = 1$ selects its actions according to the policy π_2.*

Proof. The proof of this result is based on the ideas used in [3], i.e., we proceed by constructing a coupling of the observation processes under the two policies:

- **Step-1:** First we show that at the end of the first time slot, the observations under the two policies will satisfy one of three properties (P1-P3), all of which imply that the conditional distribution of the sum of current queue lengths under γ_1 is stochastically dominated by that of γ.
- **Step-2:** Then we show that if at any time t, the two observation processes satisfy the properties P1,P2 or P3, then we can choose control actions for the policy γ_1 at time t so that they continue to satisfy these properties at time $t+1$ as well, to complete the proof.

Let us assume that the observations under γ and γ_1 at time t are represented by $Z^{\gamma}(t)$ and $Z^{\gamma_1}(t)$ respectively, and let $X(t)$ and $Y(t)$ denote the queue length vectors under these two policies. We assume that the two policies have the same observations at $t = 1$, i.e., $Z^{\gamma}(1) = Z^{\gamma_1}(1) =$

$(X^{(0)}(-1), u(-1), u(0), C(-1), C(0))$, i.e.,($X(-1) = Y(-1) = X^{(0)}(-1)$). To prove Step-1, let us suppose at $t = 1$, policy γ chooses queue j and π_2 chooses queue i. Let us denote by v_l the quantity $X_l^{(0)}(-1) - d_l(1)$ for all $l \in [N]$. In the ensuing discussion, we will match the connectivity random variables at the queues being served under the two policies.

- If $v_j = v_i$, then we can match the arrivals at queue i under γ_1 with the arrivals at queue j under γ to ensure that $P2(Z^{\gamma_1}(2), Z^{\gamma}(2))$ holds.
- If $v_j = 0$ and $v_i > 0$, then we allow the same arrivals for same queues under the two policies. If $v_j = v_i - 1$ and $(a_j(-1), a_i(-1)) = (0, 1)$, then $P2(Z^{\gamma_1}(2), Z^{\gamma}(2))$ holds. In all other cases, $P1(Z^{\gamma_1}(2), Z^{\gamma}(2))$ holds.
- If $v_j < 0$ and $v_j \leq v_i \leq 0$, then let us allow the same arrivals at same queues under the two policies. In this case, we have $\bar{Y}_i(1) = A_2^{(i)}(1)$ and $\bar{X}_j(1) = a_j(0)$, and if $a_i(-1) = 1$ then $P1(Z_2(2), Z_1(2))$ holds, otherwise $P2(Z^{\gamma_1}(2), Z^{\gamma}(2))$ is true.

Thus, we see that at the end of time $t = 1$, one of the properties $P1(Z^{\gamma_1}(2), Z^{\gamma}(2))$ or $P2(Z^{\gamma_1}(2), Z^{\gamma}(2))$ are true. This concludes the proof of step-1.

To complete the proof, we need to show Step-2, i.e., we can choose appropriate actions for the policy γ_1 to ensure that either of $P1, P2$, or $P3$ continue to be true.

- If $P3(Z^{\gamma_1}(t), Z^{\gamma}(t))$ is true, matching all arrivals and control actions under γ and γ_1 will ensure that $P3$ continues to be true at $t + 1$.
- If $P2(Z^{\gamma_1}(t), Z^{\gamma}(t))$ holds, then we match the arrivals and control actions at queue i under γ with queue $\sigma(i)$ under γ_1 (where σ is the permutation used in the definition of property $P2$) to ensure that $P2(Z^{\gamma_1}(t + 1), Z^{\gamma}(t + 1))$ holds.
- Suppose $P1(Z^{\gamma_1}(t), Z^{\gamma}(t))$ is true. If γ selects the queue $l \in [N] \setminus \{i, j\}$ at time t, then we select the queue $\sigma(l)$ under γ_1 which ensures that $P1(Z^{\gamma_1}(t + 1), Z^{\gamma}(t + 1))$ holds. If γ chooses one of i or j, then depending on the values of arrivals at time $t - 2$, either $P1, P2$ or $P3$ holds at $t + 1$. The details are in Appendix B.

Thus we have completed Step-2, and this concludes the proof of Lemma 1.

□

We can now apply the result of Lemma 1 along with induction, to prove the next result which says that the policy π_2 dominates all other admissible policies.

Proposition 2. *For the two step delay case, the policy π_2 dominates all other admissible policies π.*

Proof. It suffices to show that for any $n \in \mathbb{N}$, the conditional distribution of the queue length vector, $\bar{X}^{\pi_D}(n)$ stochastically dominates $\bar{X}^{\gamma}(n)$ for any admissible policy γ. But for any admissible γ, from Lemma 1 and induction, there exists a policy γ_n which for the first n time steps acts according to π_D, and which dominates γ. Thus we have

$$\bar{X}^{\pi_D}(n) = \bar{X}^{\gamma_n}(n) \preccurlyeq_{st} \bar{X}^{\gamma}(n)$$

as required. Also, since the policy γ_n converges to π_D pointwise as $n \to \infty$ and the single step costs are all non-negative, we have the following

$$J^{\pi_D}(Z(0)) \leq \liminf_{n \to \infty} J^{\gamma_n}(Z(0)) \leq J^{\gamma}(Z(0))$$

□

3.3 Arbitrary delay ($D > 0$)

Now we consider the case of arbitrary observation delay. In the two simple cases that we considered earlier, the random variables $(\bar{X}_i(t))_{i=1}^N$ were all comparable to each other, and the optimal policy was to select the queue with the *largest* $\bar{X}_i(t)$ in the sense of usual stochastic ordering.

To present the results for the general delay case, we will make the following two assumptions:

- **A1:** First we assume that the delay D and number of queues N satisfy the following:

$$D + 1 \leq N$$

- **A2:** We assume that the current queue connectivity, i.e., the vector $C(t)$, is not available to the controller. This is equivalent to assuming that the queues are always connected, and the server serves one packet in a time slot with probability p.

For arbitrary delay $D > 0$, it is not clear whether the queue lengths are comparable to each other for all points in the observation space. However, under assumption $A1$ there always exists a dominating queue as shown in Claim. 1 below. Assumption $A2$ allows us to sidestep the case where the dominating queue guaranteed by Claim 1 is not connected, and the controller is forced to choose among queues which may not be comparable.

Claim 1. *Under assumption A1, for any value of the observation $Z(t)$, there always exists a queue i^* such that $\bar{X}_j \preccurlyeq_{st} \bar{X}_{i^*}$ for all $j \neq i^*$.*

Proof. This result follows directly from the fact that for $N \geq D + 1$, there always exists a queue which has not been allotted the server in the last D time slots, and so there is at least one queue for which $\bar{X}_i(t) \stackrel{d}{=} q_i + A_D^{(i)}(t)$ for some $q_i \in \mathbb{N}$. Among such queues, the queue with the largest value of q_i stochastically dominates all other queues. □

Let us represent by π_D the stationary policy that chooses at time t, the queue i^* for which $\bar{X}_{i^*}(t)$ stochastically dominates $\bar{X}_j(t)$ of all other queues j. The above result tells us that in the regime $N \geq D+1$, the policy π_D is well defined. The main result of this paper is that this policy π_D is the best policy among the class of all admissible policies. We will prove this result with the help of the following lemmas.

Lemma 2. *Let us fix an $n \in \mathbb{N}$. Suppose γ is any admissible policy, and $\gamma^{(n-1)}$ is a policy which takes the same actions as γ for $1 \leq t < n-1$ and acts according to π_D at $t = n-1$. Then we have*

$$\bar{X}^{\gamma^{(n-1)}}(n) \preccurlyeq_{st} \bar{X}^{\gamma}(n)$$

Proof. Let $X(t)$ and $Y(t)$ denote the queue length vectors under the policies γ and $\gamma^{(n-1)}$ respectively. Then we have $\bar{X}(t) \stackrel{d}{=} \bar{Y}(t)$ for all $t \leq n-1$. Suppose at $t = n-1$, γ chooses queue j and $\gamma^{(n-1)}$ selects queue i. Then clearly $\bar{X}_j \preccurlyeq_{st} \bar{Y}_i(n-1) = \bar{X}_i(n-1)$. Let $\bar{Y}_i(n-1) \stackrel{d}{=} q_i + A_D^{(i)}(n-1)$, and let $q_j = X_j(n-1-D) - d_j(n-1)$. Let us consider the different possible cases (In all these scenarios we match the connectivity random variable at the queues i and j in the two systems):

- Case-1: $\underline{q_i > q_j > 0}$
  If the connectivity is zero, then $\bar{X}_l(n) = \bar{Y}_l(n)$ for all $l \in [N]$, and the result holds trivially. If the connectivity is 1, then we have

$$\begin{aligned}
\bar{Y}_i(n) &= q_i - 1 + a_i(n-1-D) + A_D^{(i)}(n) \\
Y_{\bar{j}}(n) &= q_j + a_j(n-1-D) + A_D^{(j)}(n) \\
\bar{X}_i(n) &= q_i + a_i(n-1-D) + A_D^{(i)}(n) \\
\bar{X}_j(n) &= q_j - 1 + a_j(n-1-D) + A_D^{(j)}(n)
\end{aligned}$$

  If $q_j = q_i - 1$ and $(a_i(n-1-D), a_j(n-1-D)) = (0, 1)$, then $P2(Z_2(n), Z_1(n)$ is true. In all other cases $P1(Z_2(n), Z_1(n))$ holds.

- Case-2: $q_i > 0$, $q_j \le 0$
  Again assuming that the common connectivity random variable is 1 and setting $l = -q_j$, we have the following updated states:

$$\begin{aligned}
\bar{Y}_i(n) &= q_i - 1 + a_i(n-1-D) + A_D^{(i)}(n)\\
\bar{Y}_j(n) &= a_j(n-1-D) - l + I_{k_1}^{(1)}(n-1) + \dots I_{k_1,k_2,\dots,k_l}^{(l)}(n-1) + A_D^{(j)}(n)\\
\bar{X}_i(n) &= q_i + a_i(n-1-D) + A_D^{(i)}(n)\\
\bar{X}_j(n) &= \bar{Y}_j(n) - 1 + I_{k_1,k_2,\dots,k_{D-k_1-k_2-\dots-k_l}}^{(l+1)}(n-1)
\end{aligned}$$

  Here $P1(Z_2(n), Z_1(n))$ is true because $0 \preccurlyeq_{st} I_{k_1,k_2,\dots,k_{D-k_1-k_2-\dots-k_l}}^{(l+1)}(n-1)$.

- Case-3: $q_i = 0$ and $q_j < 0$
  For this case we have the following updated queue distributions:

$$\begin{aligned}
\bar{Y}_i(n) &= a_i(n-1-D) - 1 + I_D^{(1)}(n-1) + A_D^{(i)}(n-1)\\
\bar{Y}_j(n) &= a_j(n-1-D) - l + I_{k_1}^{(1)}(n-1) + \dots I_{k_1,k_2,\dots,k_l}^{(l)}(n-1) + A_D^{(j)}(n)\\
\bar{X}_i(n) &= a_i(n-1-D) + A_D^{(i)}(n)\\
\bar{X}_j(n) &= \bar{Y}_j(n) - 1 + I_{k_1,k_2,\dots,k_{D-k_1-k_2-\dots-k_l}}^{(l+1)}(n-1)
\end{aligned}$$

  to complete this step, we require that $I_D^{(1)}(n-1) + A_D^{(i)}(n-1) \preccurlyeq_{st} I_{k_1,k_2,\dots,k_{D-k_1-k_2-\dots-k_l}}^{(l+1)}(n-1)$ occurs, which is true because $\mathbb{P}\big(I_D^{(1)}(n-1) = 1\big) = (1-\lambda)^D \le \mathbb{P}\big(I_{k_1,k_2,\dots,k_{D-k_1-k_2-\dots-k_l}}^{(l+1)}(n-1)) = 1\big)$

Thus in all cases the two queue observations at time n satisfy either $P1$ or $P2$ and hence $\sum_{i=1}^{N} \bar{X}^{\gamma_{n-1}}(n) \preccurlyeq_{st} \sum_{i=1}^{N} \bar{X}^{\gamma}(n)$ as required. □

Lemma 3. *Let $k \in N$, and let γ be any admissible policy. Suppose $\gamma^{(k+1)}$ is a policy which acts according to γ for $1 \le t \le k$, and according to π_D for $t \ge k+1$. Then there exists an admissible policy γ_k which selects its actions according to γ for $1 \le t \le k-1$, and according to π_D at $t = k$, and this policy dominates $\gamma^{(k+1)}$.*

Proof. Let $Z_1(t)$ and $Z_2(t)$ represent the observation process under $\gamma^{(k+1)}$ and γ_k respectively. We will construct a coupling between these two process for showing this result as follows:

- For $t \le k-1$, let us match the arrivals and connectivity processes for the two policies so that $Z_1(t) = Z_2(t)$ for $t \le k-1$.
- At $t = k$, γ_k acts according to π_D while $\gamma^{(k+1)}$ acts according to γ. From the proof of Lemma 2, we know that either $P1(Z_2(k+1), Z_1(k+1))$ or $P2(Z_2(k+1), Z_1(k+1))$ is true.
- From $t = k+1$ onward, $\gamma^{(k+1)}$ acts according to π_D. So to complete the proof we will show that we can choose actions for the policy γ_k so that if one of $P_i(Z_2(t), Z_1(t))$ holds, then one of them is also true at $t+1$.

Let us consider the different cases:

- If $P2(Z_2(t), Z_1(t))$ or $P3(Z_2(t), Z_1(t))$ are true, then matching the arrival, connectivity and control actions for queue i with $\sigma(i)$ ensures that these properties are true at $t+1$ as well.
- Suppose the property $P1(Z_2(t), Z_1(t))$ is true. Then, if $\gamma^{(k+1)}$ selects the queue $l \in [N] \setminus \{i, j\}$, then we select the queue $\sigma(i)$ under the policy γ_k to ensure that $P1(Z_2(t+1), Z_1(t+1))$ holds. Since $\gamma^{(k+1)}$ acts according to π_D for $t \ge k+1$, it will not select the queue j, as $\bar{X}_j(t) \preccurlyeq \bar{X}_i(t)$. So the only remaining case to check is if $\gamma^{(k+1)}$ selects queue i. We have the following cases:

– If $\bar{X}_i(t) = x_i + A_D^{(i)}(t)$ and $\bar{Y}_{\sigma(i)}(t) = y_i + A_D^{(i)}(t)$ with $0 < y_i < x_i$, then we choose the queue $\sigma(i)$ under the policy γ_k and $P1(Z_2(t+1), Z_1(t+1))$ holds.

– If $\bar{X}_i(t) = 1 + A_D^{(i)}(t)$ and $\bar{Y}_{\sigma(i)}(t) = A_D^{(i)}(t)$, then selecting the queue $\sigma(i)$ ensure that $P1(Z_2(t+1), Z_1(t+1))$ is true. (The details of this step are in Appendix C)

– Finally if $\gamma^{(k+1)}$ selects the queue i and $\bar{X}_i(t) = A_D^{(i)}(t)$. Since $\bar{Y}_{\sigma(i)} \preccurlyeq_{st} \bar{X}_i(t)$, the queue $\sigma(i)$ must have been served at least once in the last D time steps. IN this case we argue that there exists another queue with index $\sigma(m) \notin \{\sigma(i), \sigma(j)\}$ such that $\bar{Y}_{\sigma(m)} = A_D^{(\sigma(m))}(t)$. This is because of the following two facts, (1) since we have imposed the condition that $D + 1 \leq N$, there must be at least one queue under γ_k which has not been allocated the server in last D time steps, and (2) since $\gamma^{(k+1)}$ selected the queue i for which $\bar{X}_i(t) = A_D^{(i)}(t)$, there exists no queue for which $\bar{Y}_s(t) = y_s + A_D^{(s)}(t)$ with $y_s > 0$. So under the policy γ_k we select this queue $\sigma(m)$. Then it follows that $P1(Z_2(t+1), Z_1(t+1))$ is true with the permutation σ_1 which is defined as

$$\sigma_1(l) = \begin{cases} \sigma(l) & \text{if } l \notin \{i, m\} \\ \sigma(m) & \text{if } l = i \\ \sigma(i) & \text{if } l = m \end{cases}$$

Thus we have shown that if either one of three properties $P1, P2$ or $P3$ are always true, and so the policy γ_k dominates $\gamma^{(k+1)}$.

□

We are now in a position to state and prove the main result of this paper.

Theorem 1. *For the case of arbitrary observation delay D satisfying $N \geq D + 1$, the policy π_D dominates (see Def. 1) all other admissible policies γ.*

Proof. We need to show that for any $n \in \mathbb{N}$ and for any admissible policy γ, $\bar{X}^{\pi_D}(n) \preccurlyeq_{st} \bar{X}^{\gamma}(n)$. We will proceed by backward induction, i.e., for a fixed $n \in \mathbb{N}$, we will show that for any $1 \leq k \leq n-1$, there exists a policy $\gamma^{(k)}$ such that

$$\gamma^{(k)}(t) = \begin{cases} \gamma(t) & \text{if } t \leq k-1 \\ \pi_D(t) & \text{if } t \geq k \end{cases}$$

This is sufficient for our purpose as the policy $\gamma^{(1)}$ takes the same actions as π_D for $1 \leq t \leq n-1$, and hence $\bar{X}^{\pi_D}(n) = \bar{X}^{\gamma^{(1)}}(n) \preccurlyeq_{st} \bar{X}^{\gamma}(n)$, which is what we need to prove.

It suffices to show that for a fixed $n \in \mathbb{N}$, for all $1 \leq k < n-1$, and for any admissible policy γ, there exists a policy $\gamma^{(k)}$ which selects actions according to π for $t \leq k-1$ and according to the policy π_D for $k \leq t \leq n-1$, and this policy $\pi^{(i)}$ dominates π. We note that for $k = 0$, $\pi^{(k)} = \pi_D$ which proves the result.

We now proceed to prove the result by induction:

Base Case: (k=n-1) For any admissible policy, the existence of the policy $\gamma^{(n-1)}$ follows from Lemma 2

Induction Step:

Let us assume that for any admissible policy γ, there exists a policy $\gamma^{(k+1)}$ satisfying the above definition. We need to show that the policy $\gamma^{(k)}$ also dominates γ. This follows from the following:

- From the induction hypothesis, $\gamma^{(k+1)}$ dominates γ.
- From Lemma 3, we know that there exists an admissible policy γ_k which dominates $\gamma^{(k+1)}$ (and hence γ) and satisfies the following conditions $\gamma_k(t) = \begin{cases} \gamma^{(k+1)}(t) & \text{if } t \leq k-1 \\ \pi_D(t) & \text{if } t = k \end{cases}$.
- Now, again from the induction hypothesis, the policy $\gamma_k^{(k+1)}$ defined as $\gamma_k^{(k+1)}(t) = \begin{cases} \gamma_k(t) & \text{if } t \leq k \\ \pi_D(t) & \text{if } t \geq k+1 \end{cases}$ dominates the policy γ_k.

- However, from the construction of γ_k we note that $\gamma_k^{(k+1)}$ is the same as the required policy $\gamma^{(k)}$ which dominates γ_k and hence γ.

This completes the induction step, and hence we have shown that for all $l, k \in [n-1]$ with $l < k$, the policy $\gamma^{(l)}$ dominates $\gamma^{(k)}$.

$\square$

It follows from the definition of stochastic dominance that the policy π_D at any time t selects the queue which has got he largest expected current queue length, and also the queue with the smallest probability of idling.

4 Numerical Experiments

In this section, we study through simulations queueing systems with observation delays and non-symmetric arrivals. For the case of symmetric arrivals, we have shown that the optimal policy (π_D) is to choose the queue for which the conditional distribution of the current queue length given the current observations, stochastically dominates the corresponding distribution for all other queues.With heterogeneous arrival processes however, the conditional distributions of the different queues may not be comparable in the usual stochastic ordering. We study the following heuristic policies which are motivated by the policy π_D for the symmetric case:

- Policy γ_1 chooses the queue with the largest expected queue length.
- Policy γ_2 chooses the queue with the smallest probability of idling. If the probability of idling for all the queues is zero, then it chooses the queue with the largest expected queue length.

In the symmetric model, since π_D chooses the queue which is maximal in the sense of stochastic ordering, the chosen queue also has the largest mean and the smallest probability of idling.

For implementing the two policies, we estimate the expected queue lengths and the idling probabilities by drawing a large number of samples based on the arrival distributions. We fixed the connectivity probability for all the queues at $p = 0.8$, and considered seven different values of the total input arrival rates from 0.1 to 0.7 in steps of 0.1. For each arrival rate, we generated 20 arrival rate vectors (of size $N \times 1$) randomly and ran the two policies for every choice of parameters for $T = 5000$ time steps. Finally, to study the performance of the two policies we compare the average cost under the two policies at different values of the arrival rates.

The performance of the two policies for $N = 5$ and $D = 5, 10$ and 15 are shown in Table. 2. The values in the table suggest that it is preferable to choose the queue which minimizes the probability of idling. This is expected because the policy γ_2 is in some sense an extension of the *work preserving policies* to the delayed observation case, and it is known that any work preserving policy is optimal for the case of fixed connectivity [3].

In Figure 1, we show the behavior of the two policies for the case of $(N, D) = (5, 20)$. This y-coordinate of this plot is the quantity $100 \times \left(\frac{J^{\gamma_2} - J^{\gamma_1}}{J^{\gamma_2}}\right)$, where J^{γ_i} for $i = 1, 2$ is the average cost of policy γ_i. The figure shows that it is much more beneficial to use policy γ_2 when the load of the system is small, as the queues are more likely to idle under this condition. As the system load increases, this effect becomes less prominent and the performance of the two policies become similar.

References

[1] N. Ehsan and M. Liu, "Server allocation with delayed state observation: Sufficient conditions for the optimality of an index policy," *IEEE Transactions on Wireless Communications*, vol. 8, no. 4, pp. 1693–1705, 2009.

$(N, D) = (10, 5)$	Bernoulli Arrivals		Poisson Arrivals	
$(\sum_{i=1}^{N} \lambda_i)$	Policy γ_1	Policy γ_2	Policy γ_1	Policy γ_2
0.1	0.5864	0.4374		
0.2	1.0024	0.8118		
0.3	1.6055	1.2852		
0.4	2.2724	1.8345		
0.5	3.0208	2.4472		
0.6	4.5137	3.5889		
0.7	7.7304	6.4267		
$(N, D) = (10, 10)$				
0.1	0.6674	0.4321		
0.2	1.2506	0.9032		
0.3	1.8544	1.4511		
0.4	2.6615	2.1499		
0.5	3.7642	3.1051		
0.6	5.3100	4.4114		
0.7	8.5873	7.0866		
$(N, D) = (10, 15)$				
0.1	0.6813	0.4028		
0.2	1.2597	0.9464		
0.3	1.8598	1.4830		
0.4	2.8481	2.3051		
0.5	4.1027	3.4071		
0.6	5.9605	5.0038		
0.7	9.0979	7.7003		

Table 2: Simulation results are shown for a queueing system with $N = 10$ and three different values of observation delay (D). The connectivity probability for each queue is fixed at $p = 0.8$ and the total arrival rates are varied from 0.1 to 0.7. As a measure of performance of the two policies, the average cost averaged over 5000 time steps, and 20 random choices of arrival rate vectors is used. Policy γ_1 selects the queue with the largest expected current queue length, and policy γ_2 selects the queue with minimum probability of idling. The values of the simulations suggest that γ_2 is preferable to policy γ_1 uniformly for all loads.

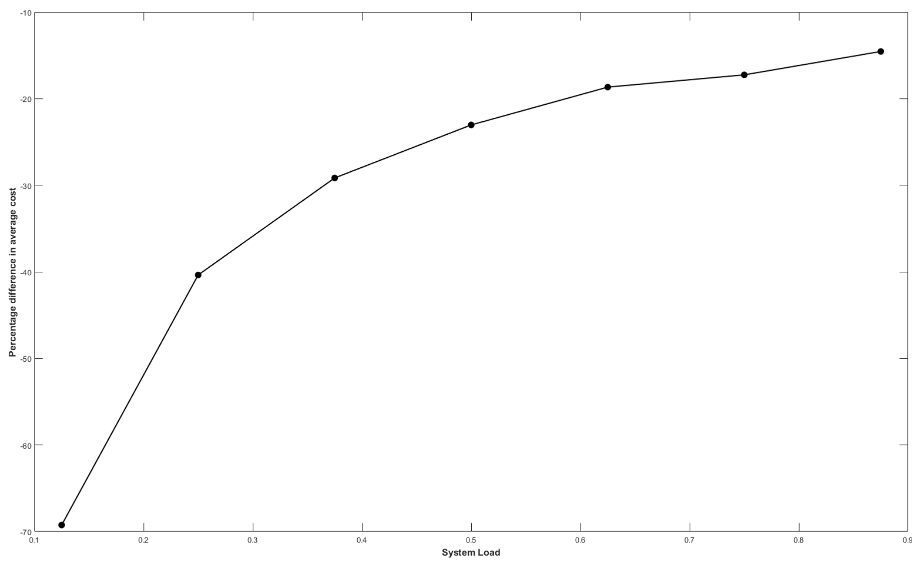

Figure 1: The figure shows the change in

[2] Z. Liu, P. Nain, and D. Towsley, “Sample path methods in the control of queues,” *Queueing Systems*, vol. 21, no. 3-4, pp. 293–335, 1995.

[3] L. Tassiulas and A. Ephremides, “Dynamic server allocation to parallel queues with randomly varying connectivity,” *IEEE Transactions on Information Theory*, vol. 39, no. 2, pp. 466–478, 1993.

A Proof of Proposition 1

Let the deterministic queue observations obtained under policies γ and γ_1 be $V(t)$ and $W(t)$ respectively, and the actual queue length processes under the two policies be $X(t)$ and $Y(t)$ respectively. Then we will couple the two processes, $V(t)$ and $W(t)$ in such a way that for every $t \in \mathbb{N}$, $\sum_{i=1}^{N} \bar{Y}_i(t) \preccurlyeq_{st} \sum_{i=1}^{N} \bar{X}_i(t)$, which means that $g(Z^{\gamma_1}(t) \leq g(Z^{\gamma}(t)$ a.s. Thus from the monotone coupling theorem, we know that the total discounted cost under γ_1 is stochastically dominated by the total discounted cost under the policy γ.

Let us first define the following properties:

- **Q1:** If at any time $V(t)$ and $W(t)$ satisfy:

$$\begin{aligned} &W_i(t) = V_i(t) - 1 \\ &W_j(t) = V_j(t) + 1; \quad W_j(t) \leq W_i(t) \\ &W_l(t) = V_l(t) \qquad \forall l \in [N] \setminus \{i, j\} \end{aligned}$$

- **Q2:** If the following holds

$$\begin{aligned} &W_i(t) = V_j(t) \\ &W_j(t) = V_i(t) \\ &W_l(t) = V_l(t) \qquad \forall l \in [N] \setminus \{i, j\} \end{aligned}$$

- **Q3:** If the following holds:

$$W_l(t) \leq V_l(t) \qquad \forall l \in [N]$$

At time $t = 1$, $V(1) = W(1)$. Suppose policy γ chooses queue j and γ_1 chooses i. If $V_j(1) = W_i(1)$ then we can swap the connectivity, arrival and service of the queues i and j in the two processes to get identical cost functions at all t. So we can assume that $V_j(1) < W_i(1)$. Let us match the connectivity at time $t = 1$ at the queue j under γ with queue i under γ_1. We first have the following result:

Lemma 4. *If $W_i(1) = V_i(1) > V_j(1) = W_j(1)$ and the policy γ and γ_1 select queues j and i respectively at $t = 1$, then the observations $W(2)$ and $V(2)$ satisfy one of the properties $Q1, Q2$ or $Q3$ almost surely.*

Proof. Then we can have the following cases.

- Case-1: $0 < V_j(1) < W_i(1)$: Same arrivals are allotted to same queues in the two systems. The observations are:

$$\begin{aligned} V_i(2) &= V_i(1) + a_i(0) \qquad & W_i(2) &= W_i(1) - 1 + a_i(0) \\ V_j(2) &= V_j(1) - 1 + a_j(0) \qquad & W_j(2) &= w_J(1) + a_j(0) \end{aligned}$$

  Thus if $V_i(1) = V_j(1) + 1$ and $(a_i(0), a_j(0)) = (0, 1)$ the the observations at $t = 2$ satisfy $Q2$. In all other cases the observations at $t = 2$ satisfy $Q1$.

- Case-2: $V_j(1) = 0 < V_i(1)$: Again, same arrivals are allotted to same queues in the two systems. Then we have the following:

$$\begin{aligned} V_i(2) &= V_i(1) + a_i(0) \qquad & W_i(2) &= W_i(1) - 1 + a_i(0) \\ V_j(2) &= (a_j(0) - 1)^+ = 0 \qquad & W_j(2) &= a_j(0) \end{aligned}$$

  In this case, if $a_j(0) = 0$, then at $t = 2$, property $Q3$ is satisfied. If $(a_i(0), a_j(0)) = (0, 1)$ and $V_i(1) = 1$, then at $t = 2$ property $Q2$ is satisfied. Otherwise $Q1$ is satisfied at $t = 2$.

Thus the observations $W(2)$ and $V(2)$ satisfy one of the properties $Q1, Q2$ and $Q3$ almost surely. $\square$

If at a given time t, the two observations satisfy one of the three properties, $Q_i, i = 1, 2, 3$, then we have $\sum_{i=1}^{N} W_i(t) \leq \sum_{i=1}^{N} V_i(t)$, which means that $\sum_{i=1}^{N} \mathbb{E}[\bar{Y}_i(t)] \leq \sum_{i=1}^{N} \mathbb{E}[\bar{X}_i(t)]$. Thus, to complete the proof we will show the following result:

Lemma 5. *If the observations $V(t)$ and $W(t)$ satisfy one of the properties $Q_i, i = 1, 2, 3$ at time t, then we can assign the control actions to $W(t)$ such that they also satisfy one of these properties at time $t+1$ almost surely.*

Proof. We consider the following cases:

- $W(t)$ and $V(t)$ satisfy $P1$: As in [3], we define $\alpha_1(t) = \arg\max\{V_i(t), V_j(t)\}$ and $\alpha_2(t) = \arg\min\{V_i(t), V_j(t)\}$, and similarly define $\hat{\alpha}_1(t)$ and $\hat{\alpha}_2(t)$ for $W(t)$. We have the following cases:
  - If the control policy γ allots the server to a queue $l \in [N] \setminus \{i, j\}$, then serve the same queue under γ_1.
  - Match the arrivals, connectivity and the control variables for the queues α_1 with $\hat{\alpha}_1$; and similarly match α_2 with $\hat{\alpha}_2$. Then in all the cases, $W(t+1)$ and $V(t+1)$ will satisfy either $Q1$, $Q2$ or $Q3$.
- W(t) and V(t) satisfy $Q2$ or $Q3$: If $V(t)$ and $W(t)$ satisfy $P2$ (resp.$P3$) at time t, then matching all the variables for the queues α_k with $\hat{\alpha}_k$ for $k = 1, 2$, ensures that $Q2$ (resp. $Q3$) is satisfied at time $t+1$.

$\square$

Finally, we note that if the observations $W(t)$ and $V(t)$ satisfy either one of Q_i, for $i = 1, 2, 3$, then $\sum_{i=1}^{N} \bar{Y}_i(t) \preccurlyeq_{st} \sum_{i=1}^{N} \bar{X}_i(t)$ as required. Thus we have constructed a policy γ_1 which dominates the policy π, and which chooses the actions according to the policy π_1 at $t = 1$.

Since the above construction of γ_1 works for any starting observations, by repeated applications for any $n \in \mathbb{N}$, we can construct a policy γ_n dominating π, which selects the first n actions according to π_1.

B Remaining step in the proof of Lemma 1

In this section, we provide the missing step in the proof of Lemma. 1. We are given that $P1(Z_2(t), Z_1(t))$ is true, and under the policy γ, the controller selects either the queue i or j. Let $X(t)$ and $Y(t)$ represent the queue length processes under the policies γ and γ_1 respectively. Let us also define the terms $L_i(t) = X_i(t-D) - d_i^{\gamma}(t)$ and $M_i(t) = Y_i(t) - d_i^{\gamma_1}(t)$.

Suppose the policy γ selects queue i. In all the cases considered below we will match the connectivity for the queues being allotted under the two policies.

- $L_i(t) > M_i(t) > 0$: In this case we select the queue $\sigma(i)$ for γ_1 and also match the arrivals at queues l with $\sigma(l)$ for all $l \in [N]$. If $L_i(t) = M_i(t) + 1$ and $L_j(t) = M_j(t) - 1$, and $(a_j(t-D), a_i(t-D)) = (1, 0)$ then $P2(Z_2(t+1), Z_1(t+1))$ holds. In all other cases the property $P1$ is satisfied.
- If $M_i(t) \leq 0$, then we match the arrivals at queue i under γ with queue j under γ_1 (i.e., let a_1 and a_2 be two *i.i.d.* Bernoulli arrivals. Then we set $a_i^{\gamma}(t-D) = a_j^{\gamma_1}(t-2) = a_1$ and $a_j^{\gamma}(t-2) = a_i^{\gamma_1}(t-2) = a_2$). Depending upon the value of $L_i(t)$ and the arrivals at $t-2$ either of the two properties $P1$ or $P3$ is true at $t+1$.

C Remaining step in the proof of Lemma 3

Here we provide the details of the step used in the proof of Lemma 3. Suppose $\bar{X}_i(t) = 1 + A_D^{(i)}(i)$ and $\bar{Y}_{\sigma(i)}(t) = A_D^{(\sigma(i))}(t)$. Since $P1(Z_2(t), Z_1(t))$ is true, and because $\gamma^{(k+1)}$ selects its actions according to π_D for $t \geq k+1$, $\bar{X}_j(t)$ and $\bar{Y}_{\sigma(j)}(t)$ will have the following form (see Sec. 2 for notations used):

$$\bar{Y}_{\sigma(j)}(t) = -l + I_{k_1}^{(1)}(t) + \dots I_{k_1,\dots,k_l}^{(l)}(t) + A_D^{(\sigma(j))}(t)$$

$$\bar{X}_j(t) = -(l+1) + I_{k_1}^{(1)}(t) + \dots I_{k_1,\dots,k_l}^{(l)}(t) + I_{k_1,\dots,k_{l+1}}^{(l+1)} + A_D^{(j)}(t)$$

Thus for $\bar{Y}_{\sigma(i)}(t+1) + \bar{Y}_{\sigma(j)}(t+1) \preccurlyeq_{st} \bar{X}_i(t+1) + \bar{X}_j(t+1)$, we need that

$$I_{D-1}^{(1)} \preccurlyeq_{st} I_{k_1,\dots,k_{l+1}}^{(l+1)}$$

which is true because $P(I_{D-1}^{(1)} = 1) = (1-\lambda)^{D-1} \leq P(I_{k_1,\dots,k_{l+1}}^{(l+1)}) = 1)$. It can be checked that if $l \geq 1$, then $P1(Z_2(t+1), Z_1(t+1))$ is true with the same permutation σ. If $l = 0$, it is true again but with a new permutation σ_1 which swaps $\sigma(i)$ and $\sigma(j)$.