Decomposition results for an M/M/k with staggered setup

Anshul Gandhi
Carnegie Mellon University
anshulg@cs.cmu.edu

Mor Harchol-Balter
Carnegie Mellon University
harchol@cs.cmu.edu

Ivo Adan
Eindhoven University of Technology
iadan@win.tue.nl

ABSTRACT

In this paper, we consider an M/M/k queueing system with setup costs. Servers are turned off when there is no work to do, but turning on an off server incurs a setup cost. The setup cost takes the form of a time delay and a power penalty. Setup costs are common in manufacturing systems, data centers and disk farms, where idle servers are turned off to save on operating costs. Since servers in setup mode consume a lot of power, the number of servers that can be in setup at any time is often limited. In the staggered setup model, at most one server can be in setup at any time. While recent literature has analyzed an M/M/k system with staggered setup and exponentially distributed setup times, no closed-form solutions were obtained. We provide the first analytical closed-form expressions for the limiting distribution of the system states, the distribution of response times, and the mean power consumption for the above system. In particular, we prove the following decomposition property: the response time for an M/M/k system with staggered setup is equal, in distribution, to the sum of response time for an M/M/k system without setup, and the setup time.

1. INTRODUCTION

Motivation

Server farms are ubiquitous in manufacturing systems, call centers and service centers. In manufacturing systems, machines are usually turned off when they have no work to do, in order to save on operating costs. Likewise, in call centers and service centers, employees can be dismissed when there are not enough customers to serve. However, there is usually a *setup cost* involved in turning on a machine, or in bringing back an employee. This setup cost is typically in the form of a time delay, which we refer to as the *setup time*.

Server farms are also prevalent in data centers. In data centers, servers consume peak power when they are servicing a job, but still consume about 60% [3] of that peak power, when they are *idle*. Idle servers can be turned off to save power. Again, however, there is a setup cost involved in turning a server back on. This setup cost is in the form of a setup time, *and* a power penalty, since the server consumes peak power during the entire duration of the setup time. Now, if there is a sudden burst of arrivals into the system, then many servers might be turned on simultaneously, resulting in a huge power draw, since servers in *setup* consume peak power. To avoid excessive power draw, data center operators sometime limit the number of servers that can be in *setup* at any point of time. This is referred to as "staggered setup". The idea behind staggered setup is also employed in disk farms, where at most one disk is allowed to spin up at any point of time, to avoid excessive power draw. This is referred to as "staggered spin up" [4, 6]. While staggered setup may help reduce power, its effect on the distribution of response time is not obvious.

Model

Abstractly, we can model a server farm with setup costs using an M/M/k queueing system, with a Poisson arrival process with rate λ, and exponentially distributed job sizes, denoted by the random variable $S \sim Exp(\mu)$. Let $\rho = \frac{\lambda}{\mu}$ denote the system load, where $0 \leq \rho < k$. In this model, a server can be in one of three states: *on* , *off*, or in *setup*. A server is in the *on* state when it is serving jobs. When the server is *on*, it consumes power P_{on}. If there are no jobs to serve, the server turns off instantaneously. While in the *off* state, the server consumes no power. To turn on an *off* server, it must first be put in *setup* mode. However, for the staggered setup model, *at most one server can be in setup at any time*. While in *setup*, a server cannot serve jobs. The time it takes for a server in *setup* mode to turn on is called the *setup time*, and during that entire time, power P_{on} is consumed. We model the setup time as an exponentially distributed random variable, I, with rate $\alpha = \frac{1}{\mathbf{E}[I]}$.

As in an M/M/k queueing system, we assume a First Come First Serve central queue, from which servers pick jobs when they become free. However, setup costs make things more complicated. From the perspective of a job, if a job arrives and finds a server in *setup*, then it simply waits in the queue. However, if the job finds no servers in *setup*, then it randomly picks an *off* server and puts it into the *setup* state. If the job finds no *off* servers, it simply waits in the queue.

When a job completes service at a server, j, the job at the head of the queue is moved to server j, without the need for *setup*, since server j is already *on*. Note that even if the job at the head of the queue was already waiting on another server i in *setup*, the job at the head of the queue is still directed to server j; server i is then either turned *off* if the queue is empty, or remains in *setup* if the queue is non-empty. We refer to this model as the M/M/k/STAG model. Fig. 1 shows the Markov chain for an M/M/k/STAG. We use $T_{M/M/k/STAG}$ (respectively, $P_{M/M/k/STAG}$) to denote its response time (respectively, power consumption).

Results

In this paper, we provide the first analysis of M/M/k/STAG, and derive simple closed-form expressions for the limiting distribution of the system states. These in turn yield the mean power consumption and the distribution of response time for the M/M/k/STAG. Interestingly, we prove that the response time for an $M/M/k/STAG$ is equal, in distribution, to the sum of response time for an M/M/k system without setup, and the setup time:

$$T_{M/M/k/STAG} \stackrel{d}{=} I + T_{M/M/k} \qquad (1)$$

Prior work

Prior work on server farms with setup costs has focussed

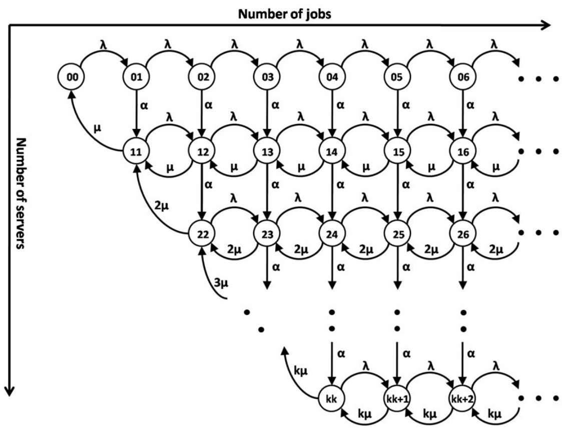

Figure 1: Markov chain for an M/M/k/STAG.

largely on single servers. There is very little work on multi-server systems with setup costs. In particular, no closed-form solutions exist for the M/M/k/STAG system.

In [2], the authors consider an M/M/k/STAG queueing system and solve the steady state equations for the associated Markov chain using a combination of difference equations and matrix analytic methods. The recursive nature of the difference equations does not yield a closed-form solution, but can be solved numerically. The difference equations method used by [2] was previously used in [1], where the authors consider a Markov chain similar to an M/M/k/STAG. Again, the authors provide recursive formulations for various performance measures, which are then numerically solved for various examples.

The above approach differs from ours in that the above papers do not determine closed-form solutions for the limiting probabilities or the distribution of response time. In particular, while the authors of [2] assume the exact same M/M/k/STAG model as ours, they do not derive the decomposition property, Eq. (1), nor do they observe this decomposition property in their graphs.

2. M/M/k/STAG

In this section, we derive the limiting probabilities of the system states (Theorem 1), the distribution of response time (Theorem 2), the mean response time (Corollary 1), and the mean power consumption (Theorem 3) for an M/M/k/STAG. Due to lack of space, we only present proof sketches. Additional details of the proofs can be found in our technical report [5].

Fig. 1 shows the Markov chain for an M/M/k/STAG. The states in the Markov chain are denoted as (a, b), where a represents the number of servers that are *on*, and b represents the number of jobs in the system. The Markov chain consists of $k + 1$ rows. The first row (from the top) consists of states where we have no *on* servers, the second row consists of states where we have exactly one *on* server, and so on. For the setup time, recall that *only one server can be in setup at any time.* Thus, the rate of going from state (i, j) to state $(i + 1, j)$ is α for any $0 \le i < k$ and $i < j$.

We'll now solve the Markov chain shown in Fig. 1 for the limiting probabilities of being in any state. We first find the limiting probabilities for the states in the 1st row, in terms of $\pi_{0,0}$. Next, we solve for the limiting probabilities of being in the states of the 2nd row, in terms of the solution for the 1st row, which in turn is expressed in terms of $\pi_{0,0}$. Continuing in this way, we can solve for the limiting probabilities of all the states of the Markov chain in terms of $\pi_{0,0}$. We'll then solve for $\pi_{0,0}$ using the equation $\sum_{i,j} \pi_{i,j} = 1$.

Theorem 1. *The limiting probabilities for an M/M/k/STAG are given by:*

$$\pi_{i,j} = \frac{\pi_{0,0} \cdot \gamma^i}{i!} \beta^j \quad \text{for } 0 \le i < k \text{ and } j \ge i$$

$$\pi_{k,j} = \frac{\pi_{0,0} \gamma^k k\mu}{k!(k\mu - (\lambda + \alpha))} \beta^j - \frac{\pi_{0,0} k^k (\lambda + \alpha)}{k!(k\mu - (\lambda + \alpha))} \left(\frac{\rho}{k}\right)^j \quad \text{for } j \ge k$$

$$\pi_{0,0} = (1 - \beta) \cdot \left\{ \sum_{0 \le i < k} \frac{\rho^i}{i!} + \frac{\rho^k \mu}{(k-1)! \cdot (k\mu - \lambda)} \right\}^{-1}$$

where $\alpha = \frac{1}{\mathbb{E}[I]}$, $\beta = \frac{\lambda}{\lambda + \alpha}$ *and* $\gamma = \frac{\lambda + \alpha}{\mu}$.

Proof. The relevant balance equations for the 1st row are given by:

$$\pi_{0,j} \cdot (\lambda + \alpha) = \pi_{0,j-1} \cdot \lambda \quad \text{for } j > 0$$

$$\Longrightarrow \pi_{0,j} = \pi_{0,0} \cdot \beta^j \text{ for } j \ge 0, \text{ where } \beta = \frac{\lambda}{\lambda + \alpha} \quad (2)$$

The relevant balance equations for the 2nd row are given by, when $j > 1$:

$$\pi_{1,j} \cdot (\lambda + \alpha + \mu) = \pi_{1,j-1} \cdot \lambda + \pi_{0,j} \cdot \alpha + \pi_{1,j+1} \cdot \mu \quad (3)$$

The RHS of Eq. (3) above consists of states of the 2nd row as well as states of the 1st row. Thus, we use difference equations (see [1] for more information on difference equations) to solve for $\pi_{1,j}$.

$$\pi_{1,j} = A_{1,1} x^j + A_{1,2} \beta^j \text{ for } j > 1, \quad (4)$$

where x is a solution of the homogeneous equation:

$$x \cdot (\lambda + \alpha + \mu) = \lambda + x^2 \cdot \mu \quad (5)$$

We now solve for $A_{1,2}$ by plugging in Eqs. (2) and (4) into Eq. (3) for $j > 2$. This gives us $A_{1,2} = \pi_{0,0} \cdot \gamma$, where $\gamma = \frac{\lambda + \alpha}{\mu}$.

To get $A_{1,1}$, we use the balance equation for $\pi_{1,2}$, which will contain $\pi_{1,1}$. Using the balance equation for $\pi_{0,0}$:

$$\pi_{0,0} \cdot \lambda = \pi_{1,1} \cdot \mu$$

$$\Longrightarrow \pi_{1,1} = \pi_{0,0} \cdot \rho \quad (6)$$

We now use Eqs. (2) and (4) in Eq. (3) for $j = 2$. This gives us $A_{1,1} = 0$.

Thus, we have from Eqs. (4):

$$\pi_{1,j} = \pi_{0,0} \cdot \beta^j \cdot \gamma \quad \text{for } j > 0 \quad (7)$$

For row i, $3 \le i \le k$, we again use difference equations, and derive the values of $A_{1,1}$ and $A_{1,2}$ as above. We find that $A_{1,1} = 0$ and $A_{1,2} = \frac{\pi_{0,0} \cdot \gamma^i}{i!}$. Thus:

$$\pi_{i,j} = \frac{\pi_{0,0} \cdot \gamma^i}{i!} \beta^j \text{ for } j \ge i \text{ and } i \le k \quad (8)$$

For row $(k + 1)$, the difference equations suggest:

$$\pi_{k,j} = A_{k,1} x^j + A_{k,2} \beta^j \quad \text{for } j > k\ , \quad (9)$$

where x is a solution of the homogeneous equation:

$$x \cdot (\lambda + k\mu) = \lambda + x^2 \cdot k\mu \quad (10)$$

This time, $A_{k,1}$ will not be zero. Thus, we need to solve the above equation for x. Solving Eq. (10) for x, we find $x = \frac{\rho}{k}$ (the other solution $x = 1$ is trivially discarded). Thus, we have:

$$\pi_{k,j} = A_{k,1}\left(\frac{\rho}{k}\right)^j + A_{k,2}\beta^j \quad \text{for } j > k\ , \qquad (11)$$

We now solve for $A_{1,1}$ and $A_{1,2}$ as we did for row 2, and find that, for $j \geq k$:

$$\pi_{k,j} = \frac{\pi_{0,0}\gamma^k k\mu}{k!(k\mu - (\lambda+\alpha))}\beta^j - \frac{\pi_{0,0}k^k(\lambda+\alpha)}{k!(k\mu-(\lambda+\alpha))}\left(\frac{\rho}{k}\right)^j$$

Finally, we derive $\pi_{0,0}$ by setting $\sum_{i,j}\pi_{i,j} = 1$. □

Interestingly, we have $\pi_{0,0} = (1-\beta)\cdot\pi'_0$, where π'_0 is the limiting probability of having 0 jobs in an M/M/k system without setup.

We now use the limiting probabilities to derive the distribution of response times for an M/M/k/STAG.

THEOREM 2. *For an M/M/k/STAG, we have:*

$$T_{M/M/k/STAG} \stackrel{d}{=} I + T_{M/M/k}$$

where $T_{M/M/k}$ is the random variable representing the response time for an M/M/k system, which is independent of the setup time, I.

PROOF. In order to derive the distribution of response times for an M/M/k/STAG, we'll first derive the z-transform of the number of jobs in queue, $\hat{N}_Q(z)$. Then, we'll use this to obtain $\tilde{T}_Q(s)$, the Laplace-Stieltjes transform for the time in queue of an M/M/k/STAG.

Using Theorem 1, the limiting probabilities for the number of jobs in queue, N_Q, for an M/M/k/STAG can be expressed as:

$$\begin{aligned}
Pr[N_Q = i] &= \pi_{0,i} + \pi_{1,1+i} + \pi_{2,2+i} + \ldots + \pi_{k,k+i} \\
&= \pi_{0,0}\left(\sum_{j=0}^{k}\frac{\rho^j}{j!}\right)\beta^i + \frac{\pi_{0,0}\rho^k(\lambda+\alpha)\left(\beta^i - \left(\frac{\rho}{k}\right)^i\right)}{k!(k\mu-\lambda-\alpha)} \\
\hat{N}_Q(z) &= \sum_{i=0}^{\infty} Pr[N_Q = i]\cdot z^i \\
&= \frac{\pi_{0,0}\left(\sum_{j=0}^{k}\frac{\rho^j}{j!}\right)}{1-\beta z} + \frac{\pi_{0,0}\rho^k(\lambda+\alpha)\lambda z}{k!(k\mu-\lambda z)(\lambda+\alpha-\lambda z)}
\end{aligned}$$

To convert $\hat{N}_Q(z)$ to $\tilde{T}_Q(s)$, observe that, by PASTA, an arrival sees the steady state number in the queue, which is the same (in distribution) as the number of jobs seen by a departure in the queue. However, the jobs left behind by a departure are exactly the ones that arrived during the job's time spent in the queue. Thus we have $\hat{N}_Q(z) = \tilde{T}_Q(\lambda(1-z))$, or equivalently, $\tilde{T}_Q(s) = \hat{N}_Q(1-\frac{s}{\lambda})$. This gives us:

$$\tilde{T}_Q(s) = \frac{\pi_{0,0}(\lambda+\alpha)}{s+\alpha}\left\{\frac{\rho^k}{k!}\left(\frac{\lambda-s}{k\mu-\lambda+s} - \frac{\lambda}{k\mu-\lambda}\right) + \frac{\alpha\pi_{0,0}^{-1}}{(\lambda+\alpha)}\right\}$$

After a few steps of algebra, we get:

$$\begin{aligned}
\tilde{T}_Q(s) &= \left(\frac{\alpha}{s+\alpha}\right)\left\{(1-P_Q) + P_Q\frac{k\mu-\lambda}{k\mu-\lambda+s}\right\} \\
&= \tilde{I}(s)\cdot\tilde{T}_{Q_{M/M/k}}(s) \qquad (12)
\end{aligned}$$

where P_Q is the probability of queueing in an M/M/k. Thus:

$$\begin{aligned}
T_{Q_{M/M/k/STAG}} &\stackrel{d}{=} I + T_{Q_{M/M/k}} \\
\Longrightarrow T_{M/M/k/STAG} &\stackrel{d}{=} I + T_{M/M/k}
\end{aligned}$$

□

COROLLARY 1.

$$\mathbb{E}[T_{M/M/k/STAG}] = \frac{1}{\alpha} + \mathbb{E}[T_{M/M/k}] \qquad (13)$$

THEOREM 3. *For an M/M/k/STAG, the mean power consumption is given by:*

$$\mathbb{E}[P_{M/M/k/STAG}] = P_{on}\left(\beta + \rho - \frac{\pi_{0,0}\rho^k\beta}{k!(1-\beta)(1-\frac{\rho}{k})}\right)$$

where $\alpha = \frac{1}{\mathbb{E}[I]}$, $\beta = \frac{\lambda}{\lambda+\alpha}$ and $\pi_{0,0}$ is given by Theorem 1.

PROOF. From Section 1, we know that a server can be in any of the following three states: (i) *off*, (ii) *on* or (iii) *setup*. The server consumes zero power in the *off* state and P_{on} power in the *on* or *setup* states. Thus, we can compute $\mathbb{E}[P_{M/M/k/STAG}]$ by conditioning on the power consumed in state (i,j). □

3. CONCLUSION

In this paper, we derive closed-form expressions for the limiting probabilities in an M/M/k with staggered setup. From these, an interesting *decomposition property* is illuminated: the response time for an M/M/k with staggered setup is distributed as the sum of two independent random variables, one corresponding to the response time for an M/M/k system without setup and the other to the setup time. The simplicity of this result makes it extremely appealing to system designers who can immediately ascertain the effect of setup costs on response time in data centers. Further, using the limiting probabilities of the system states, one can easily derive other relevant performance measures such as mean power consumption (included herein), as well as the variance of response time and the variance of power consumption.

4. REFERENCES

[1] I.J.B.F Adan and J. van der Wal. Combining make to order and make to stock. *OR Spektrum*, 20:73–81, 1998.

[2] J. R. Artalejo, A. Economou, and M. J. Lopez-Herrero. Analysis of a multiserver queue with setup times. *Queueing Syst. Theory Appl.*, 51(1-2):53–76, 2005.

[3] Luiz André Barroso and Urs Hölzle. The case for energy-proportional computing. *IEEE Computer*, 40(12):33–37, 2007.

[4] Intel Corporation. Serial ATA Staggered Spin-Up (White paper), September 2004.

[5] Anshul Gandhi and Mor Harchol-Balter. M/G/k with Exponential Setup. Technical Report CMU-CS-09-166, Carnegie Mellon University, 2009.

[6] Mark W. Storer, Kevin M. Greenan, Ethan L. Miller, and Kaladhar Voruganti. Pergamum: replacing tape with energy efficient, reliable, disk-based archival storage. In *FAST'08*, pages 1–16, Berkeley, CA, USA, 2008. USENIX Association.