Queues with a Dynamic Schedule

John Gilbertson

October 2016

Outline

Background

Queues with Scheduled Arrivals

- System where customers queue for service
- Instead of arriving randomly, customer arrival times are scheduled in advance
- Common example is appointments for doctor
- Scheduler determines arrival times to minimize expected cost
- Expected cost is a linear combination of expected total customers' waiting time and expected server availability time

Assumptions

- Single server
- Service times are independent exponential RVs with mean μ
- All customers arrive punctually
- Queue operates on a first in, first out (FIFO) basis
- Total number of customers is fixed

Static vs. Dynamic

- Static schedules:
  - Fixed for the duration of service
- Dynamic schedules:
  - Chosen progressively during service
  - On a customer's arrival, scheduler chooses the next customer's arrival time
  - Reflect ability to reschedule customer arrival times
- Aim to examine the differences between the two schedules

Literature Review

- Bailey (1952) was the first to study scheduled arrivals
- Pegden and Rosenshine (1990) propose a method for finding the optimal static schedule
- Mendel (2006) extends this model to allow for no-shows
- Fiems, Koole, and Nain (2007) include emergency requests that immediately halt the server
- Wang (1993) considers the problem of adding a new customer to a fixed static schedule

Static Schedules

Objective Function

- Denote the expected waiting time of customer i by w_i

$$\mathbb{E}\Big[\text{total customers' waiting time}\Big] = \sum_{i=1}^{n} w_i$$

- Denote the interarrival time between customer i and customer $i+1$ by x_i

$$\mathbb{E}\Big[\text{server availability time}\Big] = \sum_{i=1}^{n-1} x_i + w_n + \mu$$

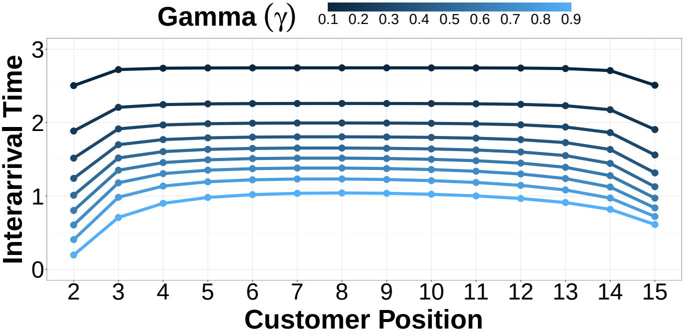
- Objective function is a linear combination of these times

$$\phi(\mathbf{x}_n) := (1-\gamma)\sum_{i=1}^{n} w_i + \gamma\left(\sum_{i=1}^{n-1} x_i + w_n + \mu\right)$$

- Objective is find $\mathbf{x}_n^* = (x_1^*, \ldots, x_{n-1}^*)$ to minimise $\phi(\mathbf{x}_n)$

$$\mathbf{x}_n^* = \arg\min_{\mathbf{x}_n} \phi(\mathbf{x}_n)$$

Model for 15 Customers



- Dome-shape: increase for first customers, remain constant, then decrease for last few customers
- As relative cost of server availability time (γ) increases, customers arrive earlier

Dynamic Schedules

Dynamic Schedules

- Static schedules are fixed for the duration of service
- Could be advantageous to allow schedule to vary during service
- Dynamic schedule is chosen progressively during service
- Arrival of customer i is scheduled on arrival of customer $i - 1$

Markov Decision Process

- Consider the problem of scheduling N customers
- Denote the number of customers in the system on arrival of customer i by k_i
- $\{k_1, \ldots, k_N\}$ is a discrete-time Markov chain
- On each customer's arrival, scheduler needs to schedule the arrival time of the next customer denoted by a
- Set of possible times is $\mathcal{A} = [0, \infty)$
- Naturally modeled as Markov decision process

State Transitions

- Denote the current state of n customers remaining to be scheduled and k customers currently in the system by (n, k)
- Initial state: $(N, 0)$
- State on arrival of first customer: $(N - 1, k_1)$
- State on arrival of last customer: $(0, k_N)$
- From state (n, k), transition to a state $(n - 1, j)$ where

$$j \in \left\{1, 2, \ldots, k + 1\right\}$$

- State transition occurs over time interval a

Expected Cost of Schedule

- Denote the expected cost of state (n, k) assuming the next customer is scheduled to arrive in a time units by $C_n(a, k)$
- Optimal policy a^* minimises the expected cost

$$C_n^*(k) = C_n(a^*, k) = \min_{a \geq 0} C_n(a, k)$$

- Probability of each transition denoted by $p_a(k, j)$
- Expected cost of each transition denoted by $R_a(k, j)$
- Expected cost of state (n, k) by Bellman equation:

$$C_n^*(k) := \min_{a \geq 0} C_n(a, k) = \min_{a \geq 0} \left[\sum_{j=1}^{k+1} p_a(k, j) \Big(R_a(k, j) + C_{n-1}^*(j) \Big) \right]$$

Erlang Distribution

- Waiting time of customer in position $r+1$ is sum of r independent Exponential RVs with mean μ

$$X = \sum_{i=1}^{r} S_i \sim \text{Erlang}(r, \mu)$$

- Distribution function:

$$F(a; r) := \mathbb{P}(X \leq x) = \begin{cases} 0 & \text{where } x = 0 \\ 1 - \sum_{i=0}^{r-1} \frac{1}{i!} \left(\frac{x}{\mu}\right)^i \mathrm{e}^{\frac{-x}{\mu}} & \text{where } x > 0 \end{cases}$$

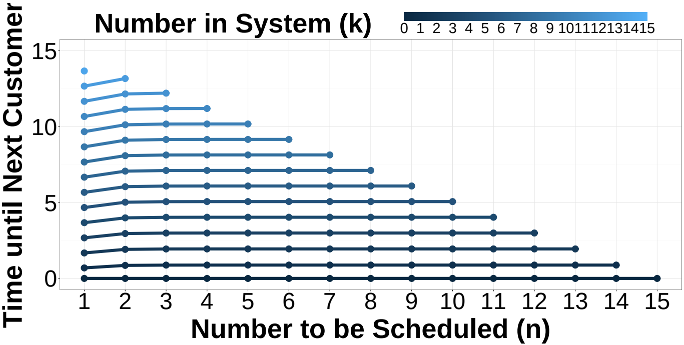
- Conditional expectation:

$$\mathbb{E}\Big[X \mid X \leq a\Big] = \mu r \times \frac{F(a; r+1)}{F(a; r)}$$

- Suppose $Y \sim \text{Exp}(\mu)$ independent of X

$$\mathbb{E}\Big[X \mid X \leq a, X + Y > a\Big] = \frac{ar}{r+1}$$

Model for 15 Customers



- Plot of optimal interarrival times a^* for each possible state
- $a^* = 0$ for initial state $(15, 0)$, and $a^* = 10.17$ for state $(3, 10)$
- a^* is independent of n for $n \geq 2$

Schedule Comparison

Expected Cost Comparison

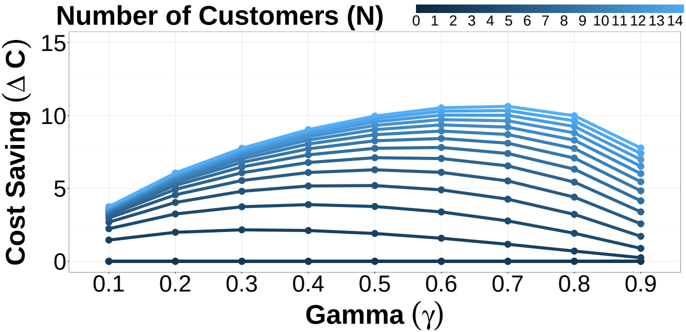
- Consider the problem of scheduling N customers
- Expected cost of optimal static schedule is $\phi(\mathbf{x}_N^*)$
- Expected cost of optimal dynamic schedule is $C_N^*(0)$
- Dynamic schedule is never worse (in expectation):

$$C_N^*(0) \leq \phi(\mathbf{x}_N^*)$$

- Equality holds for $N \in \{0, 1, 2\}$
- Define the expected percentage cost saving as

$$\Delta C := 100 \times \frac{\phi(\mathbf{x}_N^*) - C_N^*(0)}{\phi(\mathbf{x}_N^*)}$$

Expected Percentage Cost Saving



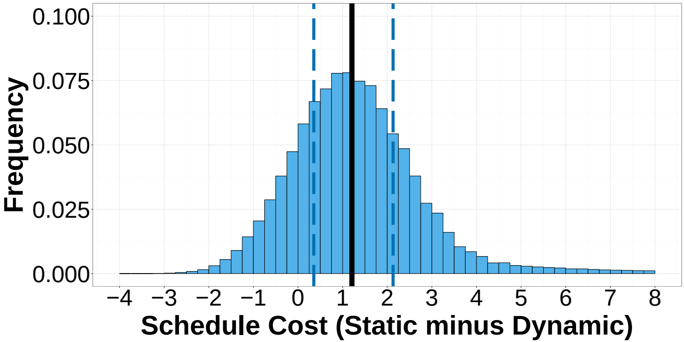
- For fixed γ, ΔC increases as N increases (at a decreasing rate)
- ΔC is at a minimum for the extreme values of γ where one of the costs is heavily prioritised

Simulation Studies

Simulation

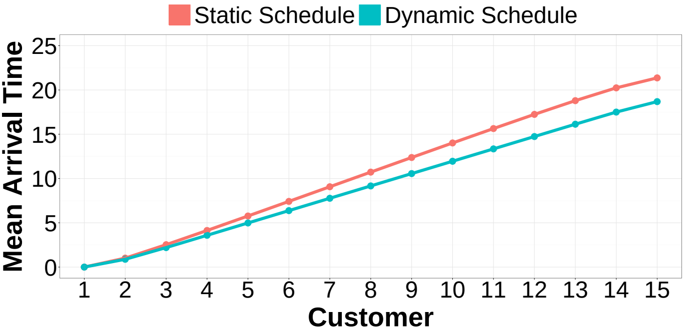
- Consider the problem of scheduling 15 customers
- Assume $\mu = 1$ and $\gamma = 0.5$
- Expected cost of static schedule is 15.05
- Expected cost of dynamic schedule is 13.55
- Desire broader understanding of difference between schedules
- Simulate a million runs to compare schedule performance

Schedule Cost



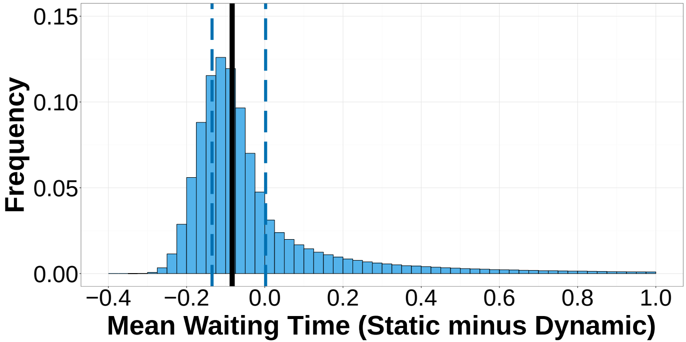
- Static schedule outperforms the dynamic schedule for a proportion of the simulation runs
- For the majority of the runs, the static schedule has a considerably greater cost than the dynamic schedule

Customer Arrival Times



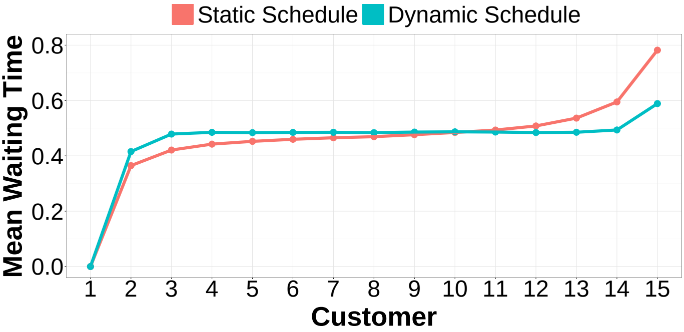
- Mean arrival time is similar for first four or five customers
- Later customers arrive significantly earlier in dynamic schedule

Waiting Time Per Customer



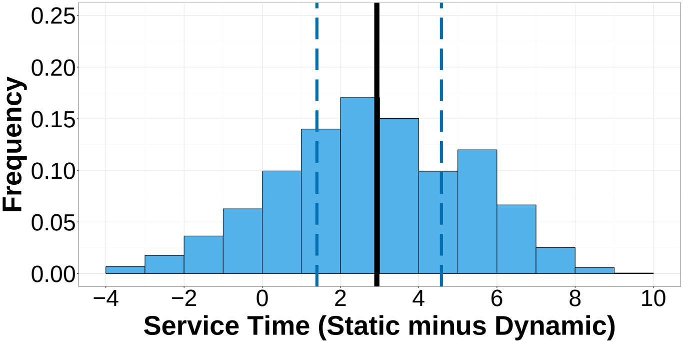
- For the majority of the runs, the customers in the static schedule have shorter waiting time
- Dynamic schedule is less prone to runs with extremely long waiting times

Customer Waiting Times



- First few customers wait longer in the dynamic schedule, but last few customers wait longer in the static schedule
- Dynamic schedule is fairer

Server Availability Time



- Dynamic schedule has generally lower server availability time
- Static schedule is limited by fixed arrival time of last customer

Conclusion

Conclusion

- Dynamic schedules often significantly outperform static schedules
- Dynamic schedules are able to adapt during service, thus less prone to runs with extremely high cost
- Possible extensions:
  - Minimum notice period
  - Server's idle time in objective function
  - Possibility of customer arriving late or not at all

References

References i

- Bailey, Norman TJ (1952). "A study of queues and appointment systems in hospital out-patient departments, with special reference to waiting-times". *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 185–199.
- Fiems, Dieter, Ger Koole, and Philippe Nain (2007). "Waiting times of scheduled patients in the presence of emergency requests". *Technisch Rapport*.
- Mendel, Sharon (2006). "Scheduling arrivals to queues: A model with no-shows". MA thesis. Tel-Aviv University.
- Pegden, Claude Dennis and Matthew Rosenshine (1990). "Scheduling arrivals to queues". *Computers & Operations Research* 17.4, pp. 343–348.

References ii

Wang, P Patrick (1993). “Static and dynamic scheduling of customer arrivals to a single-server system”. *Naval Research Logistics (NRL)* 40.3, pp. 345–360.