Suraj Khetarpal, Milton Mi

Abstract

In this paper, two queueing models (FIFO, First-In-First Out and OCF, Oil Change First) were developed and studied for the services of automotive repair shops. We controlled and varied system parameters (customer arrival rates, number of technicians, customer tolerance parameters), and analyzed wait times and customer retention. We then analyzed the profitability of a special 30-minute oil change offer. We discovered that an OCF queue has more benefit for shops that are overwhelmed with work, while the special offer is only beneficial to shops with too few customers.

Introduction

Anyone who has ever been to an automotive repair shop to have their car serviced or fixed has likely experienced long wait times. Because shops have a limited number of technicians to work on cars, long queues can form as customers wait to have their cars seen. This creates dissatisfied customers who might take their business elsewhere, never return, spread bad word-of-mouth, and post negative reviews on online websites like Yelp. As a result, it is important that shops prioritize their jobs in such a manner as to minimize wait times for customers. For example, they might do quick jobs before working on longer jobs.

Toyota dealerships take this idea to another level. They sometimes give out special offers which guarantee that if an oil change takes longer than 30 minutes, that oil change is free. Toyota will move these jobs to the front of their queue, because they cannot afford to give away many free oil changes. Prioritizing the oil changes might make sense anyways, since oil changes are the quickest of their services.

In this paper, we assess the effect that such a special offer has on the success of an automotive repair shop. There are three questions to be answered:

1. How does putting oil changes at the front of the queue affect service?
2. How does profit and customer retention change if service times are guaranteed?
3. Does the success of a the special offer system depend upon system parameters, such as the customer arrival rate, the number of technicians working at the shop, or customer tolerance for wait times?

Background Statistics: Automotive Repair Shop

- An average repair job (excluding oil changes) costs roughly $400 *(see ref 1)*
- The cost of an average repair job (excluding oil changes) is roughly 40% labor and 60% parts *(see ref 1)*
- The average cost of labor is roughly $100/hr *(see ref 3)*

- Average oil change costs $30 *(see ref 2)*
- An oil change takes 15 minutes *(see ref 4)*
- The average auto repair shop has 2 working technicians *(see ref 2)*
- An average auto repair shop get 10 repair jobs and 3 oil changes a day. *(see ref 2)*
- Parts are marked up by 25%-50%. *(see ref 5)*

Description of Models

Model Features

- We developed a stochastic queue model that tracks, amongst other outputs, arrivals, wait times, lost customers, and profits. Customers arrive, are enqueued, and then dequeued when they reach the front of the queue and a technician becomes available.
- The model features two queueing systems. The first type is a standard FIFO, or "First-In-First-Out", system where all new customers are put at the end of the queue. The second system is the OCF, or "Oil Changes First", system, in which oil change customers are placed at the front of the queue but behind any pre-existing oil change customers. If the 'reorder' feature is set to 1, this activates "Oil Changes First". If 'reorder' is set to 0, FIFO is activated.
- When the 'special' feature is set to 1, there is a 30-minute time limit to do oil changes before the oil change is free.
- There are two gates at which customers can become 'lost' customers. If 'exp_wait_on' is set to 1, then customers are told their expected waiting time when they arrive at the auto shop. These customers will leave if the expected wait time exceeds their tolerance. If 'wait_limit_on' is set to 1, then enqueued customers will leave if they reach their tolerance.
- Customers have a 'tolerance multiplier' that determines how long they are willing to wait for their service. It is multiplied by the length of the service request time to get their tolerance value. When an oil change special is running, oil change customers are given a second 'tolerance multiplier' to account for their increase in tolerance as they hope for a free oil change.
- If not specified, please assume that all parameters are set to values as described in the 'Model Assumptions' section below.
- The simulations runs over the course of 30 days.
- The time increments are 15 minutes.

Model Simplifications

- We only analyze two categories of services: oil changes, and jobs that are not oil changes, which we refer to as "Repairs" (because many of these jobs involve repairs).

- We roll the charge for labor and the charge for parts into one rate, as if they are both proportional to the time that the job takes. We do this because we assume that longer jobs will likely involve more parts.
- We estimate customer tolerance for queue wait time based on the length of work they request. We ignore the fact that some auto shop customers drop their cars off and come back later, while others wait in the shop. In the real world, these two categories of customer would have very different queue tolerances. However, some shops, including Toyota, have comfortable waiting areas with wifi, making it possible for customers who stay in the shop to wait for hours.
- We treat consecutive days as if they are continuous in time. In the real world, auto shop days are semi-continuous. Many repair jobs are allowed to sit overnight and then "roll-over" to the next day. However, this may not be true if a customer is waiting for his car, and it is certainly not true for oil change customers.
- In our simulation, we will track the revenue minus the cost of parts for each customer. At the end of the simulation, we calculate and deduct the total cost of labor for the entire simulation to get the profit.

Model Assumptions

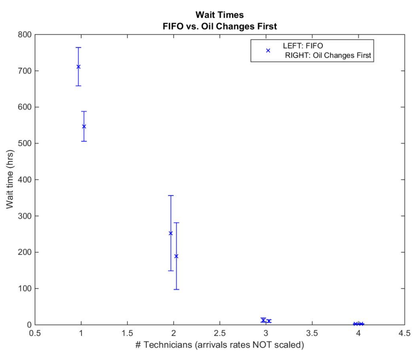
- Days are 12 hours.
- Poisson customers arrivals. An average shop receives 13 per day (industry average). The arrival rate scales with the shop size (or number of technicians divided by 2).
- A average shop has 2 working technicians (industry average).
- Each customer has a 3/13 probability (industry average) of needing an oil change and 10/13 probability of needing a repair job.
- Oil change customers require a 15-minute service request.
- Repair customers require a Poisson service request with a mean of 2.5 hours (industry average). We calculated this using statistics about an average repair:
$$400\frac{\$}{average\ job} * 60\%\frac{labor\ \$}{\$}/100\frac{\$}{labor\ hour} \approx 2.5\frac{labor\ hour}{average\ job}$$
- Revenue from repairs is \$160/hr. We calculated this as follows:
$$400\frac{\$}{average\ job}/2.5\frac{hours}{average\ job} = 160\frac{\$}{hour}$$
- We assume that labor is \$100/hr, leaving \$60/hr of revenue from parts. We assume a 50% markup, which means that the shop pays \$40/hr for parts during repairs. So the revenue minus the cost of parts is \$120/hr.
- All oil change provides \$30 in revenue whether or not a special offer is used.
- Customer tolerance for queue wait time is $(service\ request)\ x\ (tolerance\ multiplier)$. We assume that people will wait between 1 and 3 times the length of the requested work.

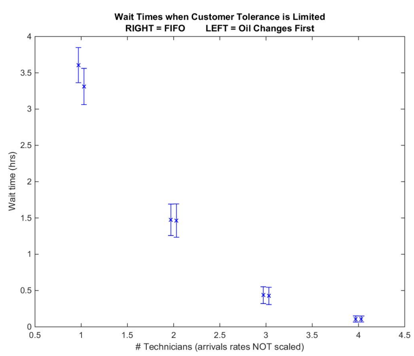
Analysis

In the first portion of our analysis, we will explore the effect of putting oil change customers at the front of the queue. We do not yet include the special offer of a free oil change after 30 minutes.

Note that many of the following figures show two layers of results, meaning that at each x axis value, two data points are shown. The first data point is slightly offset to the left, and it represents the result from when a FIFO ("First-In-First-Out") queue system is employed. The second result, slightly offset to the right, is from when an OCF ("Oil Changes First") queue system is employed.

We started by looking at what happens if customers are never lost, but instead wait indefinitely for their service.

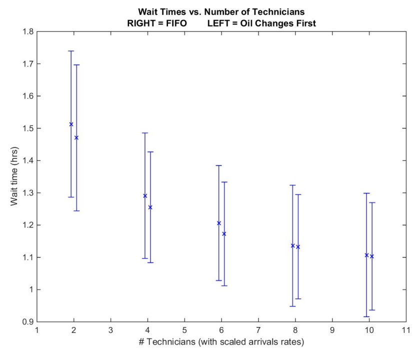




The upper graph shows the average customer waiting time as a function of the number of technicians working at the auto shop. The customer arrival rate is held constant (meaning that it does not scale with the number of technicians), so that we can see the effect of having different numbers of technicians trying to deal with the same customer load. As expected, wait times approach zero when there are more technicians working. However, when there are only one or two technicians, wait times are many hundreds of hours. This does not bode well for an average shop with only 2 technicians. The good news is that by putting oil change customers first, the wait time is reduced by almost 25% for an average sized shop.

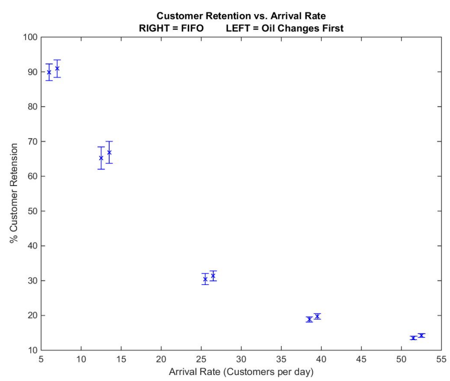
But will the benefits of an OCF queue system persist if the model allows for lost customers? The lower graph is similar to the one above except for that customers have a queue tolerance. We see that the proportional difference in wait times between a FIFO queue and an OCF queue is not very impressive.

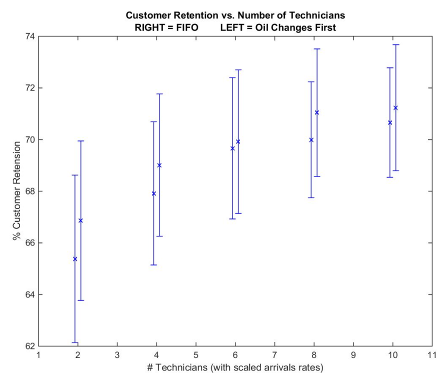
From here on out, we will only consider the case where customers can be lost at each of the two loss points; the first is when they enter the queue and are given an expected wait time, and the second is when they wait too long in the queue.



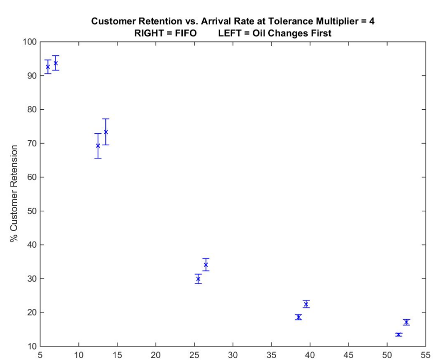
We next look at the role that the size of auto shop has on the wait times. A larger shop has not only more technicians, but also more customers, and so we scale the arrival rate. Surprisingly, even though arrivals rates are scaled with shop size, the wait times still come down as the shop size increases. Again, we see that the difference in wait times is very little between the FIFO and OCF queue systems. However, OCF does slightly better across the board, and seems to have more impact when there are fewer technicians.

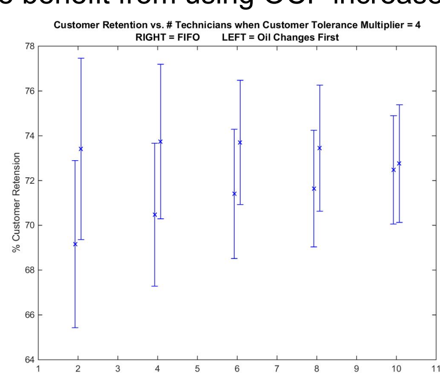
We next examine customer retention (the percentage of customers who are not lost) in both queue systems. Two parameters, the customer arrival rate and the number of technicians, have opposite effects.





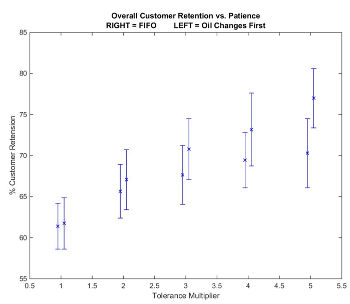
Notice that the figures have different y-axis scales. In both graphs, the customer retention rate is very similar for FIFO and OCF queue systems. However, OCF consistently performs 1% to 2% better. These graphs were made with a tolerance multiplier of 2. If we increase the tolerance multiplier to a less realistic 4, the benefit from using OCF increases. See below.

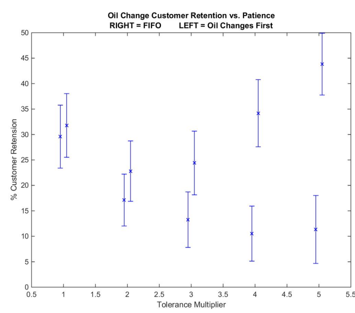


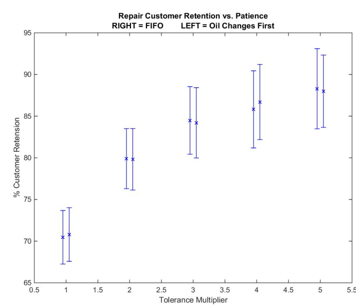


We see that at a high customer tolerance, the OCF queue system improves customer retention by an impressive 5% for most arrival rates. This improvement is proportionally very large at the higher arrival rates, where they more than double the retention rate. OCF also improves retention at all measured tolerance levels, but the effect shrinks to near zero as the number of technicians increases.

We now look in detail at the effect that the tolerance multiplier has upon customer retention. The leftmost graph shows overall customer retention, while the middle and right graphs show retention of just oil change customers and just repair customers.

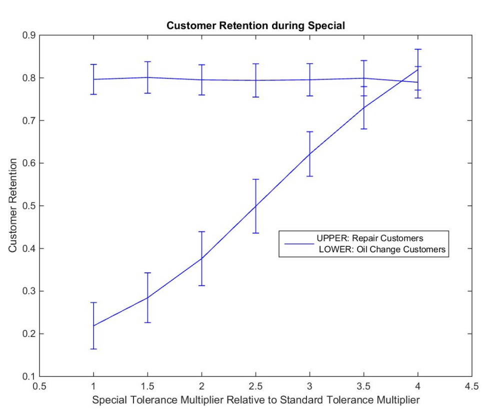




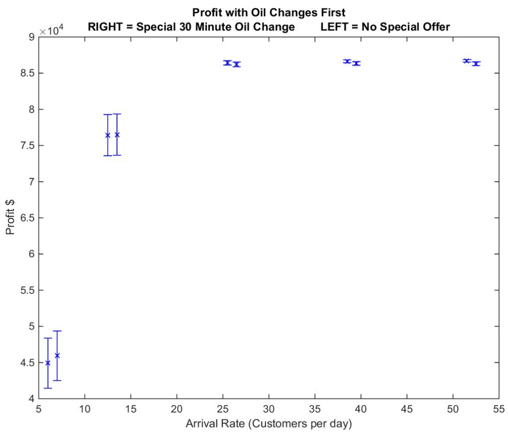


As expected, overall customer retention goes up as queue tolerance increases. However, the benefit from using an OCF queue increases as customers become more tolerant. From the middle and right graphs, we see that the benefit from OCF is not coming from repair customers, but rather from oil change customers. The middle graph is surprising because the CFO customer retention dips down before it starts its ascent. It is also surprising that repair customer retention is virtually unaffected by switching to a OCF queue. One would expect that more would leave after being bumped back by oil change customers.

Finally, we address the “special offer” of a free oil change after 30 minutes. Now that there is the risk of free services, it makes a lot of sense to judge the system based off of both customer retention and profitability.



In this graph, we gained insights into how customer retention breaks down between oil change customers and repair customers. Since the special drives up the queue tolerance for only oil change customers, we use a secondary tolerance multiplier that applies only to them. As the secondary tolerance multiplier increases, oil change customer retention increases dramatically. However, repair customer retention is slightly depressed.



This graph shows us the how profit is affected as a function of the arrival rate. We see that arrival rates drive up profit until the shop reaches its capacity, at which point profit becomes flat. Interestingly, even though oil changes might be given away for free, profits are relatively unaffected. This is because oil changes make up such a small proportion of the entire business. We also see that for shops with lower arrival rates, a small increase in arrival rate will drive up profit.

Discussion

Our early analysis revealed that switching from a FIFO to an OCF queue system has the most efficacy when wait times are allowed to stretch to infinity. In an infinite queue, prioritizing an oil change customer would reduce his wait times far more than it would increase the wait times of repair customers. In fact, the net reduction in overall wait time would roughly scale with the length of the queue. However, this advantage is diminished when wait times are capped by customer tolerance. Customer loss keeps queue lengths short, so moving up oil change customers makes less difference to their wait times.

It was interesting to see that wait times come down in larger shops, even when arrivals are scaled to shop size. This makes sense since a system with many technicians is less likely to grind to a halt when they receive a string of long service request jobs.

As one would expect, increasing the arrival rate in our simulated auto shop drove down customer retention, while increasing the number of technicians pushed it back up. These two parameters have opposite effects and need to be matched against each other in order to create a successful shop. However, customer tolerance was the parameter that made the most difference in determining whether or not an OCF queue system would improve customer retention. At low tolerance values, the benefits of an OCF queue are mediocre because repair customers are more likely to leave when they are bumped by oil change customers. At higher tolerance values, an OCF queue becomes advantageous at all arrival rates, but especially for shops with very high arrival rates. These shops are overwhelmed with work and are losing many customers. OCF serves to retain many of the oil change customers who would otherwise be lost. OCF also has a major impact on retention when there are few technicians, because this situation is again a representation of an overwhelmed work force that is losing customers.

In an OCF queue, oil change customer retention improved greatly as tolerance increased. When tolerance is low, oil change customers are likely to leave if a technician is not immediately available. But when tolerance increases, they are very likely to stick around long enough to be served in the OCF system. Alternatively, in a FIFO system where oil change customers have to wait behind repair jobs, increasing tolerance actually hurts oil change customer retention. Since tolerance is proportional to service request length, increasing the tolerance multiplier drives up repair customer tolerances faster than oil change customer tolerances. This pushes up wait times beyond what oil change customers can handle.

The 30-minute special offer had unsurprising results. It drove up oil change customer retention while slightly suppressing repair customer retention. In terms of profit, these two effects somewhat cancelled each other out because oil change customers make up a smaller portion of revenue than repair customers. However, the special could have a positive impact on profit if the shop is not at customer capacity, since pushing up the arrival rate would drive up profit. On the other hand, if the shop is at capacity, then increasing the arrival rate would have no benefit. In fact, the special would probably reduce profits slightly, making it very pointless.

Conclusion

We found that an OCF queueing system is most beneficial to auto shops when customers are at least reasonably tolerant of queues, and when shops have too few technicians relative to their customer arrival rate. There is some risk of decreasing repair customer retention, but this is balanced by the strong improvements in oil change customer retention. Running the special 30-minute offer is a good idea for shops that have the potential to take on a larger customer load, but is harmful otherwise.

Going forward with this model, it would be helpful to have some sort of feedback loop that allows customer satisfaction to adjusts arrival rates. Also, it would be more realistic if some customers waited at the shop for their car while other dropped off their car. Customers who wait are more likely to be served first.

References

1. CarMD. (2015). The 2015 CarMD Vehicle Health Index Report. Available at: https://www.carmd.com/wp/vehicle-health-index-introduction/2015-carmd-vehicle-health-index/ (Accessed 18 July 2016)

2. The University of Georgia, BOS/SBDC. (2001). Industry Fact Sheets—Mechanical Automobile Repair (NAICS 8111). Available at: https://www.georgiasbdc.org/pdfs/automotive.pdf (Accessed 18 July 2016)
3. ABC News. (2011). Why do Mechanics Cost So Much? Available at: http://www.wcti12.com/lifestyle/automotive/2014-autos/Why-do-mechanics-cost-so-much/4011326 (Accessed 18 July 2016)
4. Valvoline. (2015). How Long Will an Oil Change Take. Available at: http://www.vioc.com/services/faq/full-oil-change-service/how-long-will-an-oil-change-take (Acessed 18 July 2016)
5. Mechanic Advisor. (2014). How Much Does a Mechanic Mark Up Parts? Available at: https://www.mechanicadvisor.com/articles/how-much-does-a-mechanic-mark-up-parts (Accessed 18 July 2016)