

Analysis of a two echelon Queueing-service system by considering Speed-Quality tradeoff in service

^{1*}Ata G.Zare, ²Alireza Haji, ³Farshid Jamali

^{1,2}Industrial Engineering Department, Sharif University of Technology, Tehran, Iran

³Aerospace Engineering Department, Sharif University of Technology, Tehran, Iran

Email: ¹a_zare@ie.sharif.edu, ²ahaji@sharif.edu, ³farshid.ae@gmail.com

Abstract. In this paper we study a two-echelon system including a vendor and a service provider in customer-intensive services which introduced by Anand et al. (2011). Customers wish for high quality of service, which leads to longer service time for the service-provider. In the other hand, longer service times can cause to longer waiting time which is not desired for customers. So, tradeoff between service rate and quality of service/or product is critical. The market demand is according to Poisson process. The vendor will immediately send the order to the service-provider which works as a M/M/1 production system. The transportation time between the service-provider and vendor controlled by the service-provider, and is assumed to be deterministic and constant. Each customer has a different waiting cost per unit time and different expected value from a product or service; therefore, they decide whether to join the system or not based on the service value, price and expected waiting cost. We derive the speed quality equilibrium of the system as well as the optimal service rate and maximum demand attraction. Furthermore, we investigate the best policy of the system including optimal transportation time, price and service rate. We show that the optimal transportation time is independent from Degree of customer-intensity (α) of the service. Also, the equilibrium price will cause classifying the market, namely “market base” for small values of α , and “individual base” for high values of α . Unlike in the “market base”, losing the much market share is not undesirable in the “individual base”.

Keywords: speed-quality equilibrium, two-echelon system, service value, demands attraction.

* Corresponding address:

Ata G.Zare,

Industrial Engineering Department, Sharif University of Technology,
Tehran, Iran

Email: a_zare@ie.sharif.edu

1. Introduction

Nowadays, due to the rapid technological improvements and in the presence of the competitive market, many factors are playing different roles in how a customer chooses a product or a service. In the past years, most of the industrial units have experienced a revenue accretion as their production rate increased, assuming that customers receive a constant service value which is independent of the service time. But Anand et al. (2011), for the first time, considered a model in which the service quality and service duration aren't independent. They explored the resulting equilibrium behavior of customers and the service provider's service rate and pricing decisions. In their model, higher revenue and demand attraction are not achieved while the production rate is increased. They investigate the interactions between the service quality (value) and service speed arising from customer-intensity. The product or service value is an important factor which impacts the customer to induce him to decide whether or not to purchase a product. For instance, software comes with various attributes and features; hence people

using this software differently will obtain different values. Software developers consider this fact while developing their software.

Before Anand et al. (2011) many papers explore equilibrium queue joining, pricing and service rate decisions. There are a number of papers in the literature that address the decision-making of customers who decide whether or not to join a queue in order to use the service. Naor (1969) is the earliest paper that studies customer autonomy in service systems; he showed that there is an aggregate equilibrium pattern of behavior, which may not be optimal for the whole society. Many other papers explored equilibrium queue joining in their systems. For example: Cachon and Harker (2002) in a competition between two service-providers with price- and time-sensitive demand; Chen and Wan (2003) in a model of a competition of two make-to-order firms in which each firm tends to take a larger market share; Chen and Frank (2004) in a monopoly system whose customers cannot observe the queue when they must choose whether to place an order (Unobservable queues). They investigate the situations in which the customer preferences are linear and non-linear; Afeche (2006) in the design of price-scheduling mechanisms for a capacity-constrained firm that serves heterogeneous time-sensitive customers with private information on their willingness to pay, time-sensitivity and service requirements; Wang and Debo (2009) in medical diagnostic services (which are call centers that provide guidance to patients about what the most suitable action is according to their symptoms). There are other papers that consider queue joining. For example see Armony and Haviv (2000), Gilbert and Weng (1998), Kalai et al. (1992), Lederer and Li (1997), Li (1992) and Li and Lee (1994). Furthermore, we refer the reader to Hassin and Haviv (2003) who provide a comprehensive review of this literature. There are some papers in the literature that study the speed and quality interaction. Lovejoy and Sethuraman (2000) is among the first papers that studies the tradeoff between quality and speed in a specific manufacturing system using an M/M/1 queuing model. They explain that the quality consequences of complexity can be more significant than the time consequences; Oliva and Sterman (2001) show that working overtime may lead to quality issues. They discover that temporary imbalances between service capacity and demand interact with decision rules for capacity management, over time, and quality aspiration can lead to loss of revenue and permanent erosion of the service standards; Anand et al. (2011) focus on customer-intensive services in a queuing framework. They study the optimal "quality-speed tradeoff" in both single and multiple-server queuing settings; Kostami and Rajagopalan (2009) study the dynamic models in a monopoly system to explore the tradeoff between quality and speed. They allow the price to be a lever in determining demand levels; Lu et al. (2009) consider the tradeoff between quality and speed in a manufacturing rework context. They study how rework routing, wage and piece-rate compensation can strengthen incentives for quality. Kc and Terwiesch (2009) show that overworking is related to quality decline in hospital operations, and has an effect of decreasing the service rate. They point out that personals accelerate the service rate as load increases. Ren and Wang (2009) consider the relationship of the service quality and the patient volume in hospitals. Also, there are some other papers that consider the interactions between quality and speed, for example: De Vericourt and Zhou (2005) in routing unresolved call-backs; Hasija et al. (2009) for an empirical study of call centers; de Vericourt and Sun (2009) in the context of judgment accuracy, and Wang et al. (2010) for medical diagnostic services.

In the present model, similar to Anand model Customers exist in the market with an expected value for the product or service and join the queue based on service quality, delay costs and price until their utility function is not negative. There is an inverse relation between service time and product value in the service providers' system. The service rate chosen by the service-provider affects the quality of the service. As the service rate increases, the waiting time in system and also the value for the product decrease, and vice versa. In order to obtain the maximum customer attraction in the market, it is necessary to establish a balance between quality (value) and service speed which Anand calls "Quality-Speed tradeoff". The goal of this paper is investigating the system efficiency in a two-echelon system using Quality-Speed tradeoff and putting optimal production and logistic policies in a balanced condition. Unlike Anand's model which is one-echelon, our model is a two-echelon system: a vendor and a service-provider which serves either service or product. In our system, the transportation time between service-provider and vendor are deterministic and constant. Furthermore, the expected value

for product and waiting cost per unit time are various for each customer and uniformly distributed. We explore the optimal speed-quality equilibrium of the production system; then we investigate the optimal service rate as well as the maximum demand attraction. Assuming that the service-provider controls the transportation system, the optimal transportation time, price, and the service rate are determined. All the results computed by maple 13. It is shown that by improving the transportation system and moderating the price, the service-provider can achieve higher revenue.

In section two the model is explained to its details. In section three, the effective arrival rate and the bounds for transportation time, production rate are obtained based on the customer utility function and market attraction. In section four the service provider revenue function and its decision variable and strategic variable are studied. Section five is devoted to the optimal algorithm and finding the decision variables the strategic analysis on the optimal policy and system equilibrium. Finally in section six the obtained results are reviewed and the future research is proposed.

2. Model

Our model is a two-echelon Queuing-service system which consists of a service-provider and a vendor. Demand arrives to the vendor according to the stationary Poisson process. The vendor will immediately send the order to the Service- provider which works as an $M/M/1$ system. The Service-provider is capable of accelerating the transportation system by undertaking its relevant cost in order to reduce the waiting time. Furthermore, information about market size (η), product or service price (P), service rate (μ), waiting cost per time unit (C) and customer expected value (V) are known. Suppose that both the waiting cost per unit time in the system and the customer expected value have uniform distribution:

$$V_i \in (0, V_u), C_i \in (0, C_u) \quad (1)$$

We define the customer utility function as below:

$$U_i = V_i - C_i W(\lambda, \mu) - P \quad (2)$$

In which $W(\lambda, \mu)$ is the average waiting time for the customer to receive a product or service. That is a function of system arrival rate and service rate. Remember that customers join the queue with the rate of λ . Since the production system is $M/M/1$ and with considering the transportation time, T , we derive:

$$W(\lambda, \mu) = \left(\frac{1}{\mu - \lambda} \right) + T \quad (3)$$

In this case people join the system if their utility function is not negative; otherwise they will refrain. So, considering the Equation (2), we have:

$$V_i \geq C_i W(\lambda, \mu) + P \quad (4)$$

The classical queuing models assume that customers receive a constant service value, V , for every product or service, which is independent of the service time. In those models, increasing the service rate always results in higher revenues, as it allows the firm to lower the expected waiting time and serve more customers. In this paper, like Anand's model, the quality or the service value for customer increase with the spending time for its production. In our model the service value is defined the same as Anand's model:

$$V_0 = V_b + \alpha \mu_b - \alpha \mu \quad (5)$$

In which the parameter $\alpha \geq 0$ captures the customer-intensity of the service provided and is determined according to the market. μ_b and V_b respectively serve as a benchmark service rate and a benchmark service value.

3. Queue Joining Decision

In order to receive a service or purchase a product the customer considers his utility function which is the combination of the costs and the gains. In fact the customer wants the obtained value of the service or product to be greater than the costs such as waiting cost and purchase cost. In the following section the customer behavior in this equilibrium is investigated.

3.1. Determining system arrival rate:

Now suppose that the service-provider serves its customers with the value of V_0 , considering the customer expected value and waiting cost per unit time for them. Then for the service-provider to be able to cover a part of market, it is essential that V_0 satisfies the condition below because the customers' utility function should be non negative:

$$P + C_i \left(\frac{1}{\mu} + T \right) \leq V_0 \leq P + C_i \left(\frac{1}{\mu - \eta} + T \right) \quad (6)$$

We can simply determine the effective arrival rate, λ , that is the fraction of the total market size that join the queue. We know that the portion of the total market join the system if their expected value is not greater than V_0 and satisfies the Equation (4). So the effective arrival rate can be easily obtained from below equation:

$$\lambda = \eta \frac{\int_0^{C_u} \int_{C_i W(\lambda, \mu) + P}^{V_0} dV dC}{C_u V_u} \quad (7)$$

Solving the above equation, two roots are obtained for λ , but considering $0 < \lambda < \eta$, the feasible arrival rate is given by:

$$\lambda_e = \frac{1}{4V_u} \left\{ -(C_u T - 2V_0 + 2P)^2 \eta^2 + 4V_u \sqrt{(T\mu + 2)\eta C_u + 2\mu\eta(P - V_0) + 4V_u^2 \mu^2} \right. \\ \left. + (-C_u T + 2V_0 - 2P)\eta + 2V_u \mu \right\} \quad (8)$$

It is necessary to V_0 to be greater than or equal to $P + C_u W(\lambda, \mu)$. For satisfying this Strategic condition, considering the Equations (3) and (5), we solve this equation with respect to T . Then we have:

$$T \leq \frac{2V_u [\alpha \mu (\mu_b - \mu) - \mu(P - V_b) - C_u] - \eta [(\mu - \mu_b)\alpha + P - V_b]^2}{C_u \{ \eta [(\mu - \mu_b)\alpha + P - V_b] + 2\mu V_u \}} \quad (9)$$

Lemma 1: no customer will join the system if the price is greater than $V_0 - \frac{C_u}{2\mu} - \frac{C_u T}{2}$

Proof: Clearly, the arrival rate is decreasing with respect to the price, so we could determine the price in which the arrival rate would be equal to zero and after which the arrival rate would be negative. By setting the right-hand-side of the equation (8) equal to zero and solving it with respect to P , the maximum price could be obtained.

Lemma2: with constant service rate, λ_e is a strongly decreasing function with respect to p .

Proof: By derivating the Equation (8) with respect to P , and setting it equal to zero:

$$\frac{-2\eta V_0 + \eta(TC_u + 2P) + 2\mu V_u}{\sqrt{4\eta^2 V_0^2 - 4\eta V_0[\eta(TC_u + 2P) + 2\mu V_u] + \eta^2(TC_u + 2P)^2 + 4\eta V_u C_u(T\mu + 2) + 2\mu P + (2\mu V_u)^2}} = -1 \quad (10)$$

Consider that the denominator and nominator of the left-hand-side of the above expression are always positive, so, the derivative doesn't cross zero at any point. On the other hand, λ_e has positive and zero values, and also, $\lim_{P \rightarrow \infty} \lambda_e(\mu, P, T) = -\infty$ then it would be strongly decreasing function with respect to P.

Lemma 3: Supposing the constant service rate, λ_e is decreasing with respect to T.

Proof: By derivating the Equation (8) with respect to T, and set it equal to zero, we get the Equation (10). Similarly, considering that the derivative has no root, and $\lim_{T \rightarrow \infty} \lambda_e(\mu, P, T) = -\infty$. Then it would be strongly decreasing function with respect to T.

Theorem 1: the maximum customer attraction from the market is as below:

$$\lambda_{max} = -\eta \left(\frac{\sqrt{2C_u \alpha} - V_b - \alpha \mu_b}{V_u + \alpha \eta} \right) \quad (11)$$

Proof: considering the Equation (8), it can be easily shown that $\lim_{\mu \rightarrow \infty} \lambda_e(\mu, P, T) = -\infty$ and $\lim_{\mu \rightarrow -\infty} \lambda_e(\mu, P, T) = -\infty$. Because λ_e is a continuous function, so, the joining rate has a maximum point with respect to μ . Using simple calculus, we can have μ^1 that maximize the joining rate as follows:

$$\mu^1 = \frac{1}{2} \left(\frac{\sqrt{2\alpha C_u (V_u - \alpha \eta)^2} + 2\eta \mu_b \alpha^2 + \alpha \eta (2V_b - C_u T - 2P)}{\alpha (V_u + \alpha \eta)} \right) \quad (12)$$

Introducing the above equation in the Equation (8) we get:

$$\lambda_e(\mu^1, P, T) = -\frac{\eta}{2} \left(\frac{2\sqrt{2\alpha C_u} - 2\alpha \mu_b + 2P + C_u T - 2V_b}{V_u + \alpha \eta} \right) \quad (13)$$

Regarding the market conditions (C_u, V_u, α) , in the extreme condition, the service-provider sets the values of the price and transportation time to zero, so, Equation (11) can be obtained.

This means that considering the market conditions, for any service providing policy, the value of the maximum attractable demand would not be greater than λ_{max} . In other words, the market imposes these conditions on the service-provider at the beginning of establishing the service providing system. With studying the behavior of λ_{max} for small and large values of α , we can infer its relations to the benchmark parameters. For very small values of α , we have $\lim_{\alpha \rightarrow 0} \lambda_{max} = \eta \left(\frac{V_b}{V_u} \right)$. This means that with the insensitive α in the market, the maximum value of attraction rate highly depends on the benchmark service value. Actually, in the insensitive market, the proportion of the maximum attraction rate to the potential rate is equal to the proportion of benchmark value to the maximum expected value of customers. Similarly for large values of α with sensitive market conditions, we get $\lim_{\alpha \rightarrow \infty} \lambda_{max} = \mu_b$, the maximum value of attraction rate highly depends on the benchmark production rate, and is almost equal to it.

3.2. Determining feasible production region

In this section we want to determine the feasible region for the service rate. We consider boundary conditions, which justify the existence of the provider and service system. In this condition, the service rate should be determined in a way that arrival utility function is positive

with the price of zero and without market coverage. In other words, $C_u(T + 1/\mu) \leq V_0$. In which $V_0 = V_b + \alpha\mu_b - \alpha\mu$. Now according to this equation, we can find the feasible region for the production rate.

Lemma4: For having a valid service rate, it should be in the below ranges:

$$A_1(T, \alpha) \leq \mu \leq A_2(T, \alpha) \quad (14)$$

Proof: By solving $C_u \left(\frac{1}{\mu} + T \right) \leq V_0$, we can determine a band for the service rate like below:

$$A_1(T, \alpha) = \frac{1}{2\alpha} \left[V_b + \alpha\mu_b - C_u T - \sqrt{\alpha^2 \mu_b^2 + (-2\mu_b C_u T + 2V_b \mu_b - 4C_u) \alpha + (C_u T - V_b)^2} \right] \quad (15)$$

$$A_2(T, \alpha) = \frac{1}{2\alpha} \left[V_b + \alpha\mu_b - C_u T + \sqrt{\alpha^2 \mu_b^2 + (-2\mu_b C_u T + 2V_b \mu_b - 4C_u) \alpha + (C_u T - V_b)^2} \right] \quad (16)$$

We denote this region by $F(\alpha, T)$. Figure 1 shows the variations of these two bands with respect to α . The red line shows the upper band variations while the green line shows the lower one. This figure shows that, as α increases, these two bands converge at one point. For greater value than this point (with respect to α) there is no feasible region for μ . In other words, the service-provider cannot provide the service to customers with their expected values. In order to analyze the effect of variations of T on feasible region consider that for the service rate to be positive two conditions should hold: first, the upper band should be a real number; secondly, the upper band should be positive. With these conditions, we define lemma 5 as follows. We will use these bands in the next sections in order to find the optimal solutions.

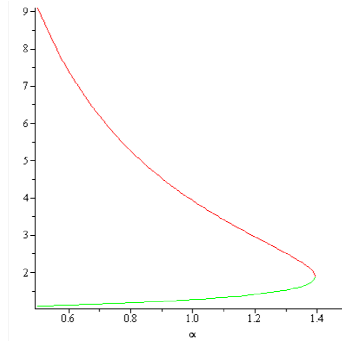


Figure 1. The behaviour of relation (14) versus α .

Lemma5: For planning the transportation time in the system, it is necessary to satisfy the below condition:

$$T \leq \frac{\alpha\mu_b + V_b - 2\sqrt{\alpha C_u}}{C_u} \quad (17)$$

Proof: According to Equation (14), the service rate must be lower than $A_2(T, \alpha)$, furthermore, the service rate should be positive. Consequently, it is necessary to the right-hand-side of the Equation (16) to be positive, So, by simplifying the $A_2(T, \alpha) \geq 0$, the Equation (17) can be obtained.

4. The Service-provider revenue function

The revenue function of the service-provider consists of three parts. The first part is the amount of incomes which are related to the customers' arrival and the service costs. The second part is the service providing cost, which depends on the service providing rate and finally the third part is the transportation costs which is a function of the arrival rate and transportation time. The service-provider should spend more cost in order for rapid sending and reducing the waiting time; hence, this cost has an inverse relation with the transportation time. So, the revenue function is determined as follows:

$$E(P, \mu, T) = P\lambda_e(P, \mu, T) - k\mu - \frac{\beta}{T}\lambda_e(P, \mu, T) \quad (18)$$

In which k and β are service providing cost per unit time and transportation fee respectively. Our numerical results and behaviour of the revenue function show that this function is concave in μ for a given price. An example of this function behaviour is shown in figure 2. Thus there exists a service rate $\mu^*(P)$ that maximizes the service provider's revenues.

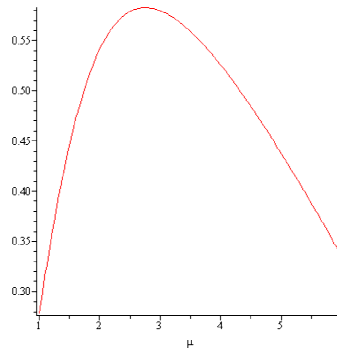


Figure 2. Concavity of revenue function versus μ .

Now considering the effective arrival rate (Equation 8), and by introducing it in the profit function (Equation 18), the optimum value of μ could be obtained as below by derivation with respect to μ :

$$\mu^*(P) = \frac{1}{2} \left\{ \frac{-2\sqrt{2}\gamma}{(kT - PT + \beta)[\alpha\eta(P - \beta) + kTV_u](\alpha\eta + V_u)} \right\} \quad (19)$$

Where γ is:

$$\gamma = \sqrt{-\eta C_u \left[\left(\frac{1}{2}\eta\beta P + kV_u - \frac{1}{2}PV_u \right) T + \frac{1}{2}\beta(V_u - \alpha\eta) \right]^2 (kT - PT + \beta) C_u (kTV_u + \eta\alpha PT - r_i)} - \eta(C_u T - 2\alpha\mu_b + 2P - 2V_b)(kT - PT + \beta)[\alpha\eta(P - \beta) + kTV_u] \quad (20)$$

It is proved, with the obtained results, by substituting $\mu^*(P)$, the corresponding equilibrium revenues, $E(P, \mu^*(P))$ is concave with respect to price and has an optimal point.

5. Optimal policy for the system

It is assumed that the service-provider can reduce the transportation time by incurring some costs, but it is desired to find out the beneficial transportation policy. Numerical results show that the equilibrium revenue function, $E(P, \mu^*(P))$, is concave in both T and P , and has a maximum value (see Figure 3); so, it can be optimized due to T or P . The results computed by MAPLE13, show that for any given P , the optimal T can be easily achieved through the numerical methods, and vice versa. Considering the concavity of equilibrium revenue function with respect to P and T , the optimal values of T and P are simultaneously obtained by using a iterative search algorithm.

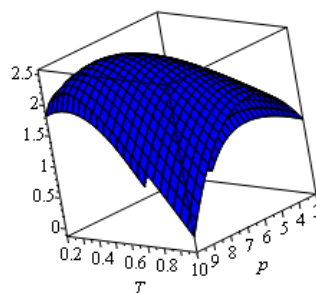


Figure 3. The behaviour of equilibrium revenue function in T and p .

It is obvious that T can not possess any arbitrary values in the real world problems. In fact T varies as a scale of a constant amount which is defined as the variation coefficient (scale) in the iterative algorithm. Starting with the minimum acceptable T , the optimal P and revenue function are obtained. By discretely increasing T based on the scale, the new optimal P and revenue are calculated in each step. The algorithm goes on until the first decline is observed in the revenue comparing to the previous step revenue. Table 1 shows the behavior of revenue function with respect to P and T using this algorithm. As shown in the table, the optimal T is 0.5.

Table 1. Obtaining Optimal Policy

T	μ	P	λ	Optimal Revenue
0.1	1.67588	6.89856	0.302716	1.16392
0.2	1.77111	6.23196	0.3438	1.53569
0.3	1.79603	5.97558	0.355088	1.64634
0.4	1.80444	5.82205	0.358947	1.68502
0.5	1.80638	5.70969	0.359838	1.69401
0.6	1.80509	5.61793	0.359244	1.68801
0.7	1.80195	5.53793	0.357802	1.6735
0.8	1.79765	5.46528	0.35583	1.65375
0.9	1.79257	5.39753	0.353504	1.6306
1	1.78691	5.33321	0.35093	1.60516

The parameters are: $C_u = 2$, $V_u = 14$, $V_b = 12$, $\mu_b = 1.5$, $K = 0.1$, $B = 0.25$, $\text{scale} = 0.1$

Consider that for small transportation time, the service-provider incurs high costs, thus finding appropriate transportation time for the service-provider is considered as a problem. The service-provider wants to maximize his profit, and, on the other hand, the customers want to maximize their utility. Hence both of these groups impose their conditions on each other. If the service-provider sets the transportation time too small, the waiting time of customers would be shorter and their satisfaction and consequently the arrival to the system rise. But the service-provider would face many costs. In order to achieve the profit, he would increase his prices and this will worsen the customers' satisfaction. Finally under these mutual interactions arrival to system and consequently the profit will be decreased as shown in Table 1.

Now with increasing the transportation time, moderating the price and service speed, arrival to system and consequently the profit would increase. On the other hand, if the service-provider sets the transportation time too high, the waiting cost for the customer would be so much that even moderating the price or speed wouldn't cause to better customer's utility. Therefore, this action has no good result for the system. For these reasons, finding the appropriate transportation time is one of the strategic variables for the service-provider in order to achieving high demand attraction in the market and more profit. With utilizing the repetitive algorithm discussed before, one could easily obtain the optimal transportation time. Based on the obtained results for 200 random problems, similar to table 1, it has been proved that the service-provider can achieve a better profit with investing in its transportation system and moderating the price, rather than the situation where there is no control over the transportation time.

Proposition 1: the optimal transportation time for the service-provider is independent from density coefficient α .

Proof: Considering our observation and analyses, α doesn't have any effect on choosing the optimal transportation time. As shown in Table 2, the optimal transportation time does not change with variation of α . Variations of α affect the service value (V_0). This would naturally change the arrival rate of the system. So, the price and service rate would be moderated in order to achieve the maximum profit. Hence α doesn't have any direct considerable effect on determining T.

Table 2. Behaviour of decision parameters with respect to α

A	T	μ	P	λ	Optimal Revenue
0	0.25	2.70386	5.24186	0.364759	1.49573
0.4	0.25	1.79582	4.95728	0.332633	1.33632
0.8	0.25	1.47411	4.85175	0.315344	1.25642
1.2	0.25	1.29393	4.81098	0.30359	1.20974
1.6	0.25	1.17447	4.80615	0.294942	1.18211
2	0.25	1.0878	4.82491	0.288327	1.16704

The parameters are: $\eta = 1$, $C_u = 3$, $V_u = 13$, $V_b = 11$, $\mu_b = 1$, $k = 0.1$, $\beta = 0.1$, scale = 0.05

Proposition 2: the joining (Effective arrival) rate and the service rate are decreasing with respect to α .

Proof: Keep in mind that the relation between the system and the market is very complex, because the reactions are mutual and simultaneous. All the decision variables of the service-provider have mutual effect on each other and on the demand attraction. Here we take their equilibrium behaviour into consideration. Remember that the sensitivity coefficient α has a direct effect on the service value and on moderating the service rate. Now with α increasing, the service-provider would decrease the service rate in order to achieve the desired equilibrium service value. With this mutual reaction and the customer-intensity increase, the arrival rate would decrease. Then, the service rate and arrival rate have similar equilibrium behaviour towards the increase in α . Figures 4 and 5 show this behaviour for Effective arrival rate and service rate, respectively.

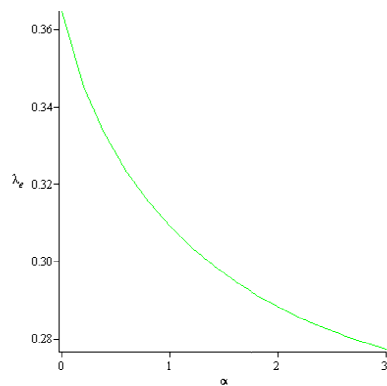


Figure 4. Effective arrival rate and customer-intensity value

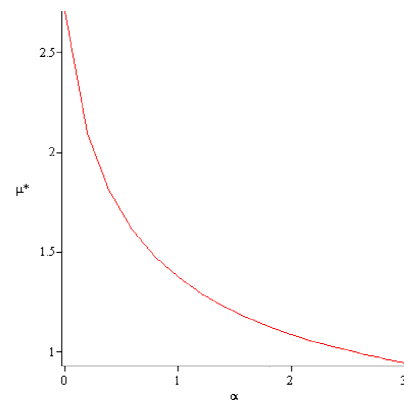


Figure 5. Optimal service rate and customer-intensity value

Proposition3: the equilibrium price and the profit function have minima with respect to the α .

Proof: The Service-provider determines the price based on his profit function. This price determines which level of customers in the market will join the system considering their utility function. Hence, the price value will cause the market classification. Considering what has been seen in the results, at first with increasing of α , the service-provider will decrease his price in order not to lose much market share. We name this behaviour of the system “market base”. Because in these occasions, the system consider the whole market and decrease the price in order not to lose the profit and demand. But the system increase the price from a special level of α because there are customers in the market with high value (V), and losing much of the market share is not undesirable. Because the system has changed its strategy and wants to attract special people with high expected service value.

In this occasion, it is desired that the system attracts fewer people but with more expected value that are able to pay more prices. We name this behaviour of the system “individual base”. With adopting this policy, increasing the price, and also attracting special customers, the system is able to achieve much profit in high α 's, hence the profit behaviour curve increase from a special level of α and continue this behaviour. Figures 6 and 7 illustrate the behaviour of the optimal price and optimal revenue with respect to α respectively.

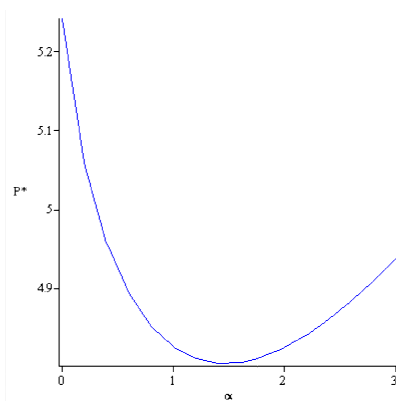


Figure 6. Optimal price and customer-intensity value

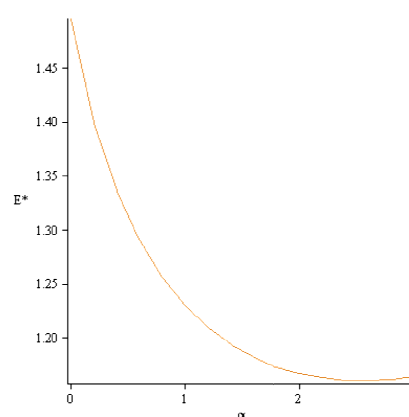


Figure 7. Optimal revenue and customer-intensity value

6. Conclusion

The paper studied the customer-intensity of a service in a queuing framework. We analyzed a two-echelon system consists of a vendor and a single service-provider. The vendor will immediately send the order to the service-provider which works as a $M/M/1$ production system. We assume that each customer has a different waiting cost per unit time and different expected value from a product or service; therefore, they decide whether to join the system or not based on the service value, price and expected waiting cost.

For the service-provider, there is an inverse relation between the service time and product value. So the service-provider decides to adjust the optimal service rate to obtain the maximum revenue. We derived the speed-quality equilibrium of the system as well as the optimal service rate and Maximum demand attraction. Also, we determined the maximum price in which the market coverage still exists. We showed that the feasible region for the production rate is sensitive to the transportation time (T) and customer-intensity (α). We investigated the optimal service rate for a given price and, system behaviour in optimal speed-quality equilibrium condition. Assuming that the service-provider controls the transportation system, we studied the best policy of the system including optimal transportation time, price and service rate. It was shown that by improving the transportation system and moderating the price, the service provider can achieve higher revenue.

It is shown that the optimal transportation time is independent from Degree of customer-intensity (α) of the service. Also, the equilibrium price will cause classifying the market, namely "market base" for small values of α , and "individual base" for high values of α . Unlike in the "market base", losing the much market share is not undesirable in the "individual base". In "market based", the system consider the whole market and decrease the price with increasing of α , in order not to lose much market share. However, in the "individual based", the system has changed its strategy and wants to attract fewer people but with more expected values that are able to pay more prices. With adopting this policy, increasing the price, and also attracting special customers, the system is able to achieve much profit in high α 's, hence the profit behaviour curve increase from a special level of α and continue this behaviour.

Our research could be extended in several possible directions. First, we assumed that both the waiting cost per unit time in the system and the customer expected value have uniform distribution. These distributions can be assumed generally. Second, we could extend the current model to multi-vendor system in which holding the inventory can be allowed. Third, a multi-product model can make the solutions become much more experimental and closer to the real world problems. Also many different vendors at different locations can be considered, for example geographically speaking, in different locations and population specification expected value and waiting cost and transportation time and service receiving time. Therefore to dominate the market in different places, more complex policies must be taken which can be studied in future studies.

Marketing strategy can be applied to this model for example in hi-tech aspects, skimming pricing strategy is suitable attract the demand, where in this case more factors are playing in decision making. Furthermore different competitors can be considered and their counteraction can be studied.

References

- [1] Afeche, Philipp, "Incentive-compatible Revenue Management in Queueing systems: Optimal Strategic delay and other Delay Tactics", University of Toronto working paper, 2006.
- [2] Anand, K., Fazil Pac, M., Veeraraghavan, S., "Quality-Speed Conundrum: Tradeoffs in Customer-Intensive Services", *Management Science*, 57:40-56, 2011.
- [3] Armony M., M. Haviv, "Price and Delay Competition between Two Service Providers", *European Journal of Operational Research*, 147(1), 32-50, 2000.

- [4] Cachon, G., P. Harker, "Competition and Outsourcing with Scale Economies", *Management Science*, 48(10), 1314-1333, 2002.
- [5] Chen, H., M. Frank, "Monopoly Pricing When Customers Queue", *IIE Transactions*, 36(6), 569-581, 2004.
- [6] Chen, H., Y. Wan, "Price competition of make-to-order firms", *IIE Trans.*, 35(9), 871-832, 2003.
- [7] De Vericourt, F., P. Sun, "Judgement Accuracy under Congestion in Service Systems", Duke University working paper, 2009.
- [8] De Vericourt, F., Y. Zhou, "Managing Response Time in a Call-Routing Problem with Service Failure", *Operations Research*, 53(6), pp. 968-981, 2005.
- [9] Gilbert, S. M., Z. K. Weng, "Incentive Effects Favor Non-Consolidating Queues in a Service System: The Principal Agent Perspective", *Management Science*, 44(12), 1662-1669, 1998.
- [10] Hasija, S., E. Pinker, R.A. Shumsky, "Work Expands to Fill the Time Available: Capacity Estimation and Staffing under Parkinson's Law", *M&SOM*, 12(1), 1-18, 2009.
- [11] Hassin, R., M. Haviv, "To Queue or not to Queue: Equilibrium behaviour in queuing systems", Kluwer Academic Publishers, Norwell, MA, 2003.
- [12] Kalai, E., M. Kamien, M. Rubinovitch, "Optimal Service Speeds in a Competitive Environment", *Management Science*, 38(8), 1154-1163, 1992.
- [13] Kc, D. S., Terwiesch, C., "Impact of Workload on Service Time and Patient Safety: An Econometric Analysis of Hospital Operations", *Management Science*, 55 (9), 1486-1498, 2009.
- [14] Kostami, V., S. Rajagopalan, "Speed Quality Tradeoffs in a Dynamic Model", University of Southern California working paper, 2009.
- [15] Lederer, P. J., L. Li, "Pricing, production, scheduling, and delivery-time competition", *Operations Research*, 45(3), 407-420, 1997.
- [16] Li, L., "The role of inventory in delivery time-competition", *Management Science*, 38(2) 182|197, 1992.
- [17] Li, L., Y. S. Lee, "Pricing and Delivery-Time Performance in a Competitive Environment", *Management Science*, 40(5), 633-646, 1994.
- [18] Lovejoy, W., K. Sethuraman, "Congestion and Complexity Costs in a Plant with Fixed Resources that Strives to Make Schedule", *M & SOM* 2(3), 221|239, 2000.
- [19] Lu, L. X., Miegheem, J. A. V., Savaskan, R. C., "Incentives for Quality Through Endogenous Routing", *Manufacturing and Service Operations Management* 11 (2), 254-273, 2009.
- [20] Naor, P., "The Regulation of Queue Sizes by Levying Tolls", *Econometrica*, 37(1), 15-24, 1969.
- [21] Oliva, R., R. J. Sterman, "Cutting Corners and Working Overtime: Quality Erosion in the Service Industry", *Management Science*, 47(7), 894-914, 2001.
- [22] Ren, Z. J., Wang, X., "Should Patients be Steered to High Volume Hospitals? An Empirical Investigation of Hospital Volume and Operations Service Quality", Working Paper, 2009.
- [23] Wang, X., L. Debo, A. Scheller-Wolf, S. Smith, "Design and Analysis of Diagnostic Service Centers", *Management Science*, 2010, 56(11), 1873-1890 (lead article), 2010.



Free download this article
and more information