SHORT RUN DYNAMICS OF MULTI-CLASS QUEUES

by

Eric J. Friedman and A. S. Landsberg

Memorandum No. UCB/ERL/IGCT M93/76

4 November 1993

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

TITLE PAGE

SHORT RUN DYNAMICS OF MULTI-CLASS QUEUES

by

Eric J. Friedman and A. S. Landsberg

Memorandum No. UCB/ERL/IGCT M93/76

4 November 1993

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Short Run Dynamics of Multi-Class Queues

Eric J. Friedman*
Department of Industrial Engineering and Operations Research,
University of California, Berkeley, CA 94720

A. S. Landsberg
Department of Physics
University of California, Berkeley, CA 94720

November 20, 1992 (Revised: June 4, 1993)

Abstract

We study the dynamics of general queueing systems with multiple classes of customers undergoing self-selection. We prove that if the capacity of a queue is sufficiently large, the equilibrium arrival rate will be globally stable. As the capacity is decreased, the arrival rate typically oscillates near the equilibrium.

Keywords: Queues, Dynamics, Stability, Service Facilities.

1 Introduction

The stability of arrival rates is of great importance in the design and planning of queueing systems such as service facilities, distributed computer networks, and production facilities. Given that the precise nature of the customer population and the detailed properties of the queueing system are typically not known in advance, it is not possible to compute the arrival rate *a priori*. Rather, the arrival rate will vary over time due to the self-selection by customers. If the dynamics of this adjustment process is not stable then the arrival rate may not reach equilibrium and could instead fluctuate.

In a recent paper Stidham [6] considers the problem of stability of an m/m/1 queue with a single class of customers, which is a model of a service facility. For a specific example he showed that if the capacity of this facility is large enough, then the equilibrium arrival rate will be stable. This is crucial for the successful operation of such a facility. However, his model is

*E.J.F. thanks the National Science Foundation Grant IRI-8902813 for support.

very specific, and it is unclear what will occur in more general queues with multiple customer classes.

Our results are an attempt to answer this question. In the next section we present a general model of a queueing system with multiple classes of customers. We then show that Stidham's example is a special case of this more general model. Using techniques from nonlinear dynamics (see Guckenheimer [3]) we can easily rederive Stidham's result, and can in fact compute the analytic solution for his example. We then consider a natural generalization of this model to multiple user classes. We find that similar stability results hold for this class of problems.

We then turn to our general model of queueing systems with many customer classes. Here, we allow the customer classes to have any smooth distribution for their value of service and allow the queue to have an arbitrary delay function, subject to several mild restrictions. We show that if the capacity of the queue is too low then the system typically becomes unstable in a specific manner. In particular, the equilibrium arrival rate undergoes a 'period-doubling bifurcation' [3] and becomes unstable, creating a stable periodic orbit nearby. Thus, even though the equilibrium arrival rate is unstable, the arrival rate always ends up near the equilibrium, and makes small oscillations about it.

2 The Model

We consider a general model of a queueing system with multiple classes of customers. This is a generalization of the model studied in Stidham [6], which was based on work by Dewan and Mendelson [2].

Consider a queue with capacity μ and a delay function $D_\mu(\lambda)$, where $D_\mu(\lambda)$ is the expected delay given the arrival rate λ. A customer is charged p to enter the queue. Here we consider the *short run problem* where the capacity μ is fixed. In the *long run problem* the capacity is varied in order to maximize profit or social benefit. We consider only the short run problem here, and will to discuss the long run problem in a future paper.

We assume that each customer has a linear delay cost and a specific value of service. A class i of customers is a group of customers all having the same delay cost h_i. Within this class different customers may have different values of service. These are represented by a distribution $r_i(v)$, where $r_i(v)$ is the density of customers in class i with value of service v. A customer will enter the queue if her value of service is greater than her total expected cost, which is the sum of the delay cost and the price.

These customers will enter the queue based on self selection. It is easy to see that for a given delay d the arrival rate of customers from class i will be

$$L^i_p(d) = \int_{p+h_i d}^{\infty} r_i(v)\, dv.$$

Note that since we are focusing on preference distinctions among customers, we require that all customers have identical job types, and consider only queueing systems with a single class of service. Thus all customers have the same expected delay, regardless of class.

As in Stidham, we break time up into discrete periods. Customers in period $n+1$ decide whether to enter the queue based on the expected delay, which they assume will be the same as the delay in the previous period. Specifically, letting λ^i_n denoting the arrival rate from customer class i in period n, then the arrival rate in period $n+1$ can be computed from $\lambda^i_{n+1} = L^i_p(d_n)$, where $d_n = D_\mu(\sum_i \lambda^i_n)$. From an abstract point of view we can describe the dynamics as a nonlinear mapping $\lambda^i_{n+1} = T_i(\lambda^1_n, \lambda^2_n, \ldots, \lambda^m_n)$ where $T_i(\lambda^1_n, \lambda^2_n, \ldots, \lambda^m_n) = L^i_p(D_\mu(\sum_i \lambda^i))$.

Note that our model is very general. It allows for wide variations of customer classes and queueing systems. In the next section we consider the specific example first analyzed by Stidham [6].

3 Stidham's Example

In [6], Stidham considers a service facilty that is an m/m/1 queue with a single class of customers. Their values of service are distributed uniformly on the interval $[0, a]$ with maximum arrival rate Λ and delay cost h per unit time. In this case $r(v) = \Lambda/a$ for all $v \leq a$ and $r(v) = 0$ for $v > a$. In our notation, $L_p(d) = \Lambda(a - p - hd)/a$. Since the service facility is an m/m/1 queue we see that $D(\lambda) = (\mu - \lambda)^{-1}$ (see, e.g. Kleinrock [5]). The dynamics is therefore described by

$$\lambda_{n+1} = T(\lambda_n) = \Lambda'(1 - \frac{h/a'}{\mu - \lambda_n}), \qquad (1)$$

where following Stidham we have defined $\Lambda' = [(a-p)/a]\Lambda$ and $a' = a - p$. Note that T is non-increasing. (See figure 1.) We must require $a' > h/\mu$ or else no customers will ever enter the queue. [Note that there is an important restriction implicit in this formula, namely $0 \leq \lambda \leq \Lambda$. Thus if $T(0) > T^{-1}(0)$ (which corresponds to $\Lambda' > \mu + h(\Lambda'/\mu - 1)/a'$) then points near the boundary, $\lambda = 0$, will be mapped beyond $T^{-1}(0)$ and then to 0, and the arrival rate will eventually oscillate between 0 and $T(0)$.]

The equilibrium arrival rate satisfies $\lambda^* = T(\lambda^*)$. Thus λ^* is a fixed point of the mapping and is illustrated in figure 1. Since T is non-increasing this fixed point must be unique and is given by

$$\lambda^* = (\Lambda' + \mu - [(\Lambda' + \mu)^2 - 4\Lambda'(\mu - h/a')]^{\frac{1}{2}})/2.$$

A fixed point λ^* is said to be locally stable if there exists a neighborhood S of λ^* such that

$$\lim_{n\to\infty} T^n(\lambda_0) = \lambda^*$$

for any $\lambda_0 \in S$, where we define $T^n(\lambda) = T(T(\cdots T(\lambda)))$ to be the functional composition of the map n times. This implies that if the arrival rate begins in the neighborhood S of the equilibrium, then it will converge to the equilibrium. A fixed point is globally stable if any initial arrival rate converges to the equilibrium, in which case S can be taken to be the entire domain of the function.

In [6], Stidham showed that the equilibrium arrival rate was globally stable if the capacity of the queue was large enough. We now present an easy proof of this result. Our goal here is to develop a simple method which can be extended to more complicated problems.

We observe that the map can be simplified by defining $\rho_n = \mu - \lambda_n$, $\alpha = \mu - \Lambda'$, and $\beta = \Lambda' h/a'$. The simplified map, $\hat{T}(\rho)$, takes the form

$$\hat{T}(\rho) = \alpha + \beta/\rho. \tag{2}$$

Note that $\hat{T}(\rho)$ is defined on the interval $0 \leq \rho \leq \mu$.

We will prove that the fixed point is globally stable by showing that $\hat{T}^2$ is a contraction mapping for $\mu > \Lambda'$. A contraction mapping $\hat{T}^2$ satisfies $|\hat{T}^2(x) - \hat{T}^2(y)| < |x - y|$ for all $x \neq y$ and guarantees that the mapping has a unique globally attracting fixed point. (See, e.g Devaney [1].) It is easy to understand why a contraction mapping must have a unique fixed point. Consider any two points x, y. By the contracting nature of the map the distance between them will decrease monotonically and thus they will converge to the same point. This obviously applies to any collection of points, showing that all initial conditions must converge to a single point.

Lemma 1 *For $\alpha > 0$, $\hat{T}^2$ is a contraction mapping.*

Proof: From equation (2) we find that the second iterate of the map is

$$\hat{T}^2(\rho) = \alpha + \frac{\rho}{1+\rho\alpha/\beta}.$$

Now we compute

$$|\hat{T}^2(x) - \hat{T}^2(y)| = \left|\frac{x}{1+x\frac{\alpha}{\beta}} - \frac{y}{1+y\frac{\alpha}{\beta}}\right| \leq \frac{1}{(1+\frac{\alpha}{\beta}x)(1+\frac{\alpha}{\beta}y)}|x-y|$$

and note that

$$\frac{1}{(1+\frac{\alpha}{\beta}x)(1+\frac{\alpha}{\beta}y)} < 1 \qquad \forall \alpha, \beta, x, y > 0,$$

showing the contracting nature of $\hat{T}^2$. $\diamond$

Theorem 1 *For $\mu > \Lambda'$, λ^* is a globally attracting fixed point of T.*

Proof: Since $\alpha = \mu - \Lambda'$ the theorem requires that $\alpha > 0$. In this case $\hat{T}^2$ is a contraction mapping and therefore $\hat{T}$ must have a unique fixed point by the contraction mapping theorem. Since T is isomorphic to $\hat{T}$ their dynamics must be the same. Note that $\mu > \Lambda'$ implies that $T(0) < T^{-1}(0)$, thus avoiding difficulties with the boundary. $\square$

We point out that for $\mu < \Lambda'$, $\hat{T}^2$ is an expansion mapping which satisfies $|\hat{T}(x) - \hat{T}(y)| > |x-y|$ for all $x \neq y$. This implies that for any initial condition $\lambda_0 \neq \lambda^*$, the arrival rate will eventually run into the boundary.

Finally, we note that the mapping T has a very special structure which allows us to give a complete analytical solution. T is a linear fractional transformation (see, e.g. Hermann [4]) and thus by using its group structure we can write the arrival rate at period n as an explicit function of the initial arrival rate λ_0. We find

$$\lambda_n = T^n(\lambda_0) = \mu - \frac{(\gamma_-^{n+1} - \gamma_+^{n+1})(\mu - \lambda_0) - (\gamma_+\gamma_-^{n+1} - \gamma_-\gamma_+^{n+1})}{(\gamma_-^n - \gamma_+^n)(\mu - \lambda_0) - (\gamma_+\gamma_-^n - \gamma_-\gamma_+^n)}$$

where

$$\gamma_\pm = \frac{\mu - \Lambda'}{2} \pm \sqrt{(\frac{\mu - \Lambda'}{2})^2 + \frac{\Lambda' h}{a'}}.$$

We note, however, that for more general mappings no analytic solution exists. Thus we avoid using analytic solutions in our analysis as we are interested in ideas and techniques which will generalize to complicated situations.

4 Multi-Class Extension of Stidham's Example

Before considering the general multi-class queueing system, we first analyze the dynamics of a particular multi-class extension of Stidham's example. We show that the same techniques we used for analyzing the single class example can be easily applied to this multi-class problem as well.

Consider m classes of customers with delay cost h_i per unit time and uniform distribution parameterized by a_i and Λ_i as in the previous example. We still consider an m/m/1 queue, but now there are multiple customer classes each with its own arrival rate λ^i. For this model the delay d depends on the sum of the individual arrival rates. Thus we write $d_n = D(\lambda_n)$, where $\lambda_n = \sum_i \lambda_n^i$. As before $D(\lambda_n) = (\mu - \lambda_n)^{-1}$.

We see that the dynamics is described by

$$\lambda_{n+1}^i = T_i(\vec{\lambda}_n) = \Lambda_i' \left(1 - \frac{h_i/a_i'}{\mu - \lambda_n}\right),$$

where we define $\Lambda_i' = [(a_i - p)/a_i]\Lambda_i$ and $a_i' = a_i - p$. Note that T_i is non-increasing for all classes i.

Again, the map can be simplified by defining $\rho_n^i = \mu/m - \lambda_n^i$, $\alpha_i = \mu/m - \Lambda_i'$, and $\beta_i = \Lambda_i' h_i/a_i'$. We therefore study the map $\hat{T}_i$ given by

$$\rho_{n+1}^i = \hat{T}_i(\vec{\rho_n}) = \alpha_i + \beta_i/\rho_n^{tot}$$

where $\rho_n^{tot} = \sum_i \rho_n^i$.

Define $\alpha^{tot} = \sum_i \alpha_i$, $\beta^{tot} = \sum_i \beta_i$ and $\Lambda' = \sum_i \Lambda_i'$. We now show that there exists a globally stable fixed point of $\hat{T}$ for $\alpha^{tot} > 0$.

Lemma 2 *For $\alpha^{tot} > 0$, $\hat{T}$ has a unique fixed point.*

Proof: The key step is to note that one can construct a map for the quantity ρ_n^{tot} which takes the simple form $\rho_{n+1}^{tot} = \alpha^{tot} + \beta^{tot}/\rho_n^{tot}$. Now using the same argument as in Lemma 1, we see that there is a unique fixed point of this map which we denote by ρ^*. Thus the fixed point of the map $\hat{T}$ must be $\rho^{i*} = \hat{T}_i(\rho^*)$. $\diamond$

Now we prove the global stability of this fixed point for a service facility with sufficient capacity.

Theorem 2 *For $\mu > \Lambda'$, there exists a globally attracting fixed point of T.*

Proof: A simple extension of the proof in Theorem 1 yields the desired result since T and $\hat{T}$ are isomorphic. $\square$

For sake of completeness we mention that, as in the single class case, we can compute the complete analytical solution to this problem. The fixed point of the map T is

$$\lambda^{i*} = \frac{\mu}{m} - (\alpha_i + \beta_i/\rho^*)$$

where

$$\rho^* = \frac{\alpha}{2} + \sqrt{(\frac{\alpha}{2})^2 + \beta}.$$

The arrival rate λ_n^i, expressed in terms of the initial rate λ_0^i, is given by

$$\lambda_n^i = \mu/n - \rho_n^i$$

where

$$\rho_n^i = \frac{[\alpha_i(\gamma_-^n - \gamma_+^n) + \beta_i(\gamma_-^{n-1} - \gamma_+^{n-1})]\rho_0 - \alpha_i(\gamma_+\gamma_-^n - \gamma_-\gamma_+^n) - \beta_i(\gamma_+\gamma_-^{n-1} - \gamma_-\gamma_+^{n-1})}{(\gamma_-^n - \gamma_+^n)\rho_0 - (\gamma_+\gamma_-^n - \gamma_-\gamma_+^n)}$$

and

$$\gamma_\pm = \frac{\alpha^{tot}}{2} \pm \sqrt{(\frac{\alpha^{tot}}{2})^2 + \beta^{tot}}, \qquad \rho_0 = \mu - \sum_i \lambda_0^i.$$

Thus this particular multi-class extension of Stidham's example exhibits many of the same features as the single class case.

We now return to the general model as defined in Section 2.

5 General Theory

The examples in the previous sections are somewhat artificial. Queueing systems do not typically have an m/m/1 delay function and customers' service values do not typically have a uniform distribution. Thus the crucial question is whether the results obtained in these examples apply to more general queues.

In this section we show that subject to a few mild restrictions any queueing system with multiple customer classes has a unique globally stable equilibrium arrival rate if the capacity of the queue is sufficiently large. Since we are not restricting ourselves to a specific queueing model, we must carefully define what we mean by capacity μ. We do this by imposing restrictions on the delay functions $D_\mu(\lambda)$. Thus we will represent the queueing system by a parametrized family of functions $D_\mu(\lambda)$ which must satisfy the following conditions.

1. Positivity and Convexity: $D_\mu(\lambda)$ is a non-negative, monotonically increasing, differentiable convex function.

2. Capacity: $D_\mu(\mu) = \infty$.

3. Delay : $D_\mu(\lambda) < D_{\mu'}(\lambda)$ if $\mu > \mu'$.

4. Marginal increase in delay: $D'_\mu(\lambda) < D'_{\mu'}(\lambda)$ if $\mu > \mu'$. (Here $D'_\mu(\lambda) = \frac{\partial D_\mu(\lambda)}{\partial \lambda}$.)

5. Sufficient marginal capacity: $\forall \epsilon, \lambda > 0$, there exists a μ such that $D'_\mu(\lambda) \leq \epsilon$.

These restrictions are quite mild and are satisfied by any reasonable queue, e.g. m/m/c, GI/GI/c [5]. We discuss them briefly here. Conditions (1) and (2) state that the delay is positive, convex, smooth, and becomes infinite as the capacity is reached. Conditions (3) and (4) imply that larger capacity is always more desirable. Condition (5) states that any service demand could be satisfied with sufficient capacity. Note that in this model μ can be interpreted as the number of servers, as well as the speed of the server.

Our restriction on the customer classes is that the value of service $r_i(v)$ is continuous, bounded and non-zero over a finite interval. Also, the total arrival rate is finite, which is equivalent to $L_i(0) < \infty$. [Note that if $r_i(v)$ is not bounded then various pathologies can occur. For example, if $r_i(d)$ has support on a Cantor set then $L_i(d)$ could be a 'devil's staircase' [1]. Then even for a queueing system with a very high capacity the equilibrium arrival rate may not be attracting, and could lead to sustained, highly complex fluctuations.]

Our main result is that for any queueing system and customer classes satisfying the above assumptions, the equilibrium arrival rate will be globally attracting if the capacity is large enough. We prove this using an argument similar to that used for the examples.

Theorem 3 *There exists a $\mu_g > 0$ such that if $\mu > \mu_g$, λ^* is a globally attracting fixed point of T.*

Proof: Define $\lambda_n = \sum_i \lambda_n^i$. Then we see that $\lambda_{n+1} = \tilde{T}(\lambda_n) = \sum_i T_i(\lambda_n)$. Defining $L(d) = \sum_i L_i(d)$, it follows that L is non-increasing and differentiable, with bounded derivative. Define $\hat{l} = L(0)$ and $k = \max_d [mh_i r_i(d)]$. Thus we can write $\tilde{T}$ as $\tilde{T}(\lambda) = L(D_\mu(\lambda))$. Note that for all $n > 0$, $0 \leq \lambda_n \leq \hat{l}$. Now by condition (5) of the definition for D_μ there exists a μ_g such that $D'_{\mu_g}(\hat{l}) < 1/k$ and $\tilde{T}_{\mu_g}(0) \leq \tilde{T}^{-1}_{\mu_g}(0)$.

Now for $0 < \lambda < \hat{l}$ we see that

$$|\tilde{T}'(\lambda)| = |L'(D_{\mu_g}(\lambda))| \cdot |D'_{\mu_g}(\lambda)| < 1$$

since $|L'(D_{\mu_g}(\lambda))| \leq k$. This implies that

$$|\tilde{T}(x) - \tilde{T}(y)| < |x - y|$$

showing that $\tilde{T}$ is a contraction mapping. Therefore $\tilde{T}$ must have a unique fixed point λ^* which is globally attracting. Noting that $\lambda^{i*} = T_i(\lambda^*)$ completes the proof. □

Thus, as in the specific examples, the general case has a globally attracting equilibrium for a queue with large enough capacity $\mu > \mu_g$. However, below the critical point ($\mu < \mu_g$) the behavior is in general quite different. In the previous examples the equilibrium arrival rate is globally repelling below the critical point, since the arrival rate eventually hits the boundary.

However, for most models, when the equilibrium arrival rate loses global stability it would either become locally stable, or it would become locally unstable and give rise to small, but stable, oscillations. We describe this process in the next section.

6 Period-Doubling Bifurcation

In Stidham's example and in our specific multi-class generalization of it, we saw that as soon as the equilibrium arrival rate lost stability, the system became totally unstable since all initial arrival rates would rapidly diverge from the equilibrium, until they hit the boundaries. However, this is not the typical behavior for general queues and customer classes. Instead we find that the arrival rate begins to oscillate in the vicinity of the equilibrium. Consider the following example.

Example: Consider a m/m/1 queue with a single class of service with linear delay cost $b = 1$. Now let $r_i(v) = 3 - v$ for $0 \leq v \leq 3$ and 0 otherwise and set $p = 1$. It is easy to compute that $L(d) = 2 - 2d + d^2/2$ and

$$T(\lambda; \mu) = 2 - \frac{2}{\mu - \lambda} + \frac{1}{2(\mu - \lambda)^2}$$

The equilibrium arrival rate of T can be found by setting $\lambda^*(\mu) = T(\lambda^*(\mu); \mu)$ and solving a cubic equation. For $\mu > 3/2$ we see that the fixed point is $\lambda^*(3/2) = 1/2$, which is locally stable since

$$0 > \frac{\partial T(1/2; 3/2)}{\partial \lambda} > -1.$$

A more careful analysis reveals that for all $\mu > 3/2$ the fixed point is in fact globally stable.

However, when $\mu < 3/2$ the equilibrium becomes unstable and small but stable oscillations near the equilibrium arise. This occurs due to a 'period-doubling bifurcation' [3]. In order to understand this we consider the second iterate of the map, $T^2(\lambda; \mu)$, and approximate it near the 'bifurcation point' $\lambda^* = 1/2$, $\mu^* = 3/2$ using a Taylor expansion.

Define $G(\lambda; \mu) = T^2(\lambda, \mu)$. Keeping only the 'important' terms in the expansion [3], we write

$$G(\lambda; \mu) \approx G + \frac{\partial G}{\partial \lambda}(\lambda - \lambda^*) + \frac{\partial G}{\partial \mu}(\mu - \mu^*) + \frac{\partial^2 G}{\partial \lambda \, \partial \mu}(\lambda - \lambda^*)(\mu - \mu^*) + \frac{\partial^3 G}{\partial \lambda^3}(\lambda - \lambda^*)^3/6$$

where all derivatives are evaluated at (λ^*, μ^*).

We find that G takes the form

$$G(\lambda;\mu) \approx \frac{1}{2} + (\lambda - \lambda^*) - (\lambda - \lambda^*)(\mu - \mu^*) - \frac{(\lambda - \lambda^*)^3}{2}.$$

Solving for the fixed points ($G(\lambda;\mu) = \lambda$) we find three solutions: $\lambda^* = 1/2, \lambda^+ = 1/2 + \sqrt{2(\mu^* - \mu)}$, and $\lambda^- = 1/2 - \sqrt{2(\mu^* - \mu)}$.

The solution λ^* simply corresponds to the equilibrium arrival rate. However, the fixed points λ^+, λ^- of the map $G(\lambda;\mu)$ represent oscillatory solutions of the map $T(\lambda;\mu)$. In particular, the arrival rate will jump back and forth between λ^+ and λ^- with each iteration. A '2-cycle' is said to have formed. Note that this 2-cycle (λ^+, λ^-) grows out of the equilibrium arrival rate λ^*, since λ^+, λ^- coincide with λ^* at $\mu = \mu^*$, and begin to move outward as μ is decreased below μ^*. Since this 2-cycle is stable, the arrival rate will converge to this 2-cycle from any initial value. (See figure 2.)

Even though the arrival rate is unstable, it is still very well behaved and remains close to the equilibrium. Thus, instability of the equilibrium arrival does not completely destabilize the queueing system. However, as the capacity of the queue is decreased further the periodic orbit may move farther away from the equilibrium arrival rate, thus making the queueing system unstable in a practical sense.

The emergence of a stable 2-cycle ("period-doubling") is not just a special feature of this specific example. Rather, we can show that this phenomenon is what can typically be expected to occur in general models of multi-class queues. Formally, we say that this behavior is generic. Thus the absence of this period-doubling bifurcation in Stidham's example should be viewed as an anomaly resulting from the special properties of his mapping.

The following theorem asserts that a globally stable equilibrium rate will generically undergo a period-doubling bifurcation when it becomes unstable. The only way that this will not occur is if several derivatives of the mapping T happen to cancel, a rare event. (This is precisely what happens in Stidham's example.)

Theorem 4 *Let μ_g be the parameter at which the equilibrium arrival rate $\lambda^*(\mu_g)$ first becomes* *unstable and assume that for $\mu > \mu_g$ the equilibrium is globally attracting. Then if*

$$\frac{\partial T_\mu(\lambda)}{\partial \mu}\frac{\partial^2 T_\mu(\lambda)}{\partial \lambda^2} + 2\frac{\partial^2 T_\mu(\lambda)}{\partial \lambda\, \partial \mu} \neq 0$$

and

$$\frac{1}{2}\left(\frac{\partial^2 T_\mu(\lambda)}{\partial \lambda^2}\right)^2 + \frac{1}{3}\frac{\partial^3 T_\mu(\lambda)}{\partial \lambda^3} \neq 0$$

a period-doubling bifurcation will occur as μ is decreased below μ_g, creating a stable 2-cycle for $\mu < \mu_g$. This 2-cycle will be close to λ^* for μ close to μ_g. (Note that all derivatives in the above expressions are evaluated at (λ^*, μ^*).)

Proof: This is proved by combining the monotonicity of T_μ, with the theorem in Guckenheimer and Holmes [3, p. 158], and noting that the bifurcation must be supercritical since the equilibrium is globally attracting for $\mu > \mu_g$. □

The significance of this result lies in the fact that it does not depend on the precise details of the model. We believe that the consideration of generic properties of queueing systems is of fundamental importance in understanding actual systems, which lack an exact mathematical description.

7 Acknowledgements

E.J.F. would like to thank Shmuel Oren for useful conversations.

References

[1] R.L. Devaney. *An Introduction to Chaotic Dynamical Systems.* Benjamin/Cummings, Menlo Park, Calif, 1986.

[2] S. Dewan and H. Mendholson. User delay costs and internal pricing for a service facility. *Management Science*, 36(12):1502–17, December 1989.

[3] J. Guckenheimer and P. Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields.* Springer-Verlag, New York, 1986.

[4] R. Hermann. *Lie Groups for Physicists.* W. A. Benjamin, New York, 1966.

[5] L. Kleinrock. *Queueing Systems, Vol I.* John Wiley, New York, 1975.

[6] S. Stidham. Pricing and capacity decisions for a service facility: Stability and multiple local optima. *Management Science*, 38(8):1121–1139, 1992.

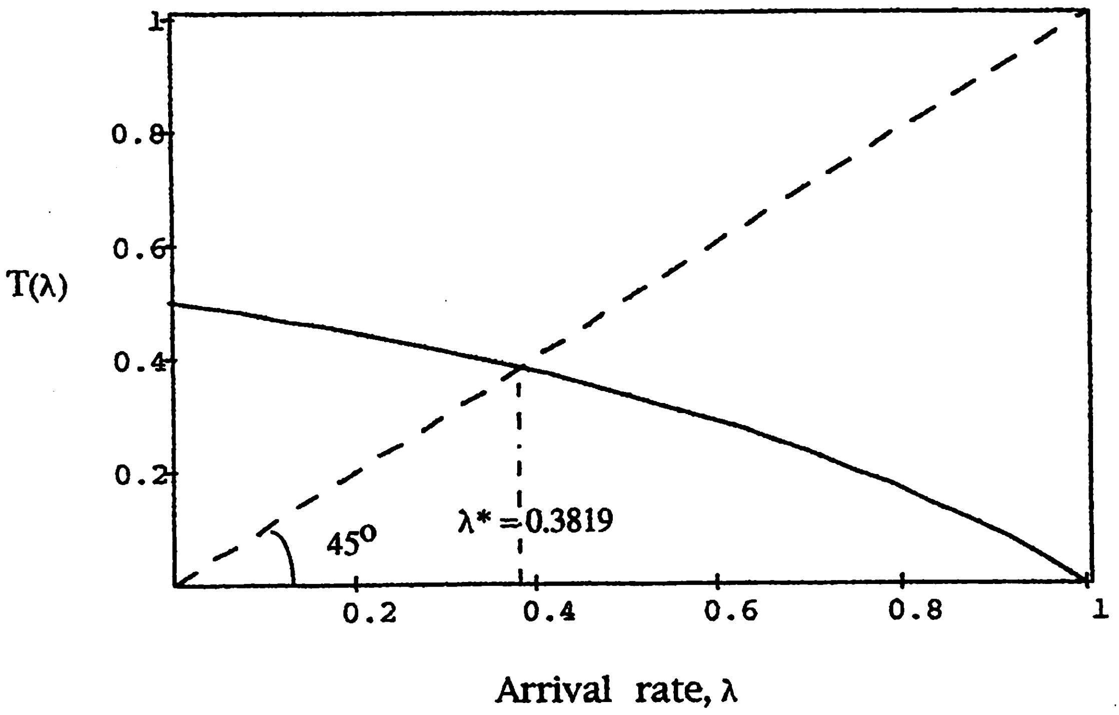

Figure 1: The mapping T(λ) versus λ for a single class of uniformly distributed customers with Λ'=1, h=a'=1, and μ=2. The fixed point is λ*=.3819.

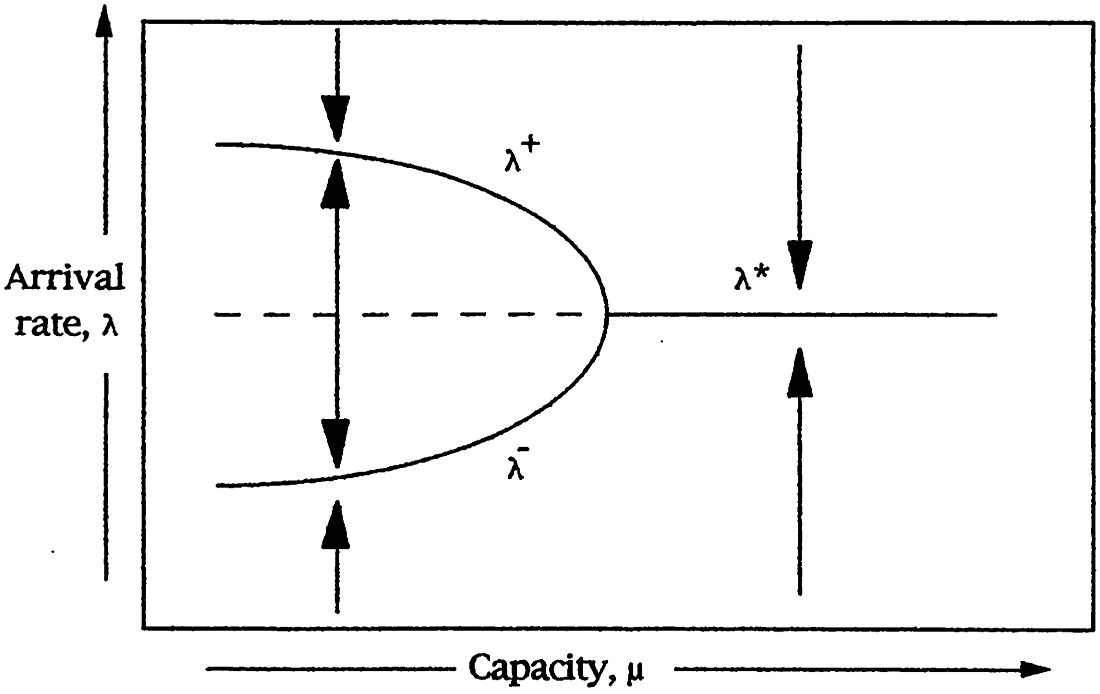

Figure 2: The period doubling bifurcation. For $\mu > \mu_g$ the equilibrium arrival rate λ* is globally stable (solid line). For $\mu < \mu_g$, λ* becomes unstable (dashed line) and the 2-cycle (λ^+, λ^-) is stable and attracting.