Applications of dynamic games in queues

Eitan Altman

ABSTRACT Queueing phenomena, along with many related decision problems, are well known to all of us from daily situations. We often need to answer questions concerning whether to queue or not, where to queue, how long to queue, etc. In networking applications, both in road traffic as well as in telecommunication networks, individuals or some central controllers are frequently faced with similar questions. These decisions have often to be taken in randomly changing environment, in a decentralized way, and with partial information. This gives rise to many challenging problems in dynamic games. We shall describe in this overview some generic queueing problems (both static and dynamic) that require game theoretic models and solutions.

1.1 Introduction

We shall focus in this overview on some sample of decision problems that arise in queueing, and whose solutions requires game theoretic modeling and solution approaches. The main generic problems are

1. <u>To queue or not to queue?</u> In a context of networks, this question is related to admission control, i.e. to whether a new connection, or an arriving packet are accepted to the network (or to a specific node of a network).

2. <u>When to queue?</u> Individuals who have to queue in order to receive service often face this problem. They may have an idea about the global demand, i.e. the probability distribution of the total number of persons that would arrive and compete for the same service, but the decisions of arrival times of other persons are typically unknown.

3. <u>Where to queue?</u> This is one of the most studied problems in networking games. There are several routes between one or more sources and one or more destinations that involve costs or delays. The latter depend on the congestion at each part of the route (i.e. each node or link). The congestion is of course a function of the routing decisions. Different users have to determine their route choices.

This type of game is known as the traffic assignment problem in the context of road traffic, and as a networking routing game in the context of telecommunications. We shall present two simple examples of stochastic games in which decisions of where to queue have to be taken dynamically (depending on the time varying state of the system).

4. How much to queue? This is perhaps the second most important type of games that occur in telecommunication networks. There are applications that can control the amount or the rate of information sent to the network. For example, file transfers and electronic mail that are sent in the Internet use a dynamic mechanism (called TCP - Transport Control Protocol) to adjust the rate of transmission to the encountered congestion. Real time video and voice applications often use dynamic compression which results not only in controlling the rate of transmission, but also the amount of data to be transmitted. This type of control determines how much data would be queued at different nodes in the network, it influences both the delay (or cost) of that application and contributes to delay (or cost) of other applications.

A key element that appears in many types of networking games and games in queues is the congestion. It has an impact on costs and on delays. In the context of telecommunications, it may further affect other measures of quality of the communications, as perceived by a connection, such as loss probabilities of its packets and the delay variations. Large amount of losses or large delay variations are quite harmful in real time voice and video applications.

In this paper we first present some samples of problems in each of the above categories. Through these problems we shall identify some features of networking games. We shall study in particular some tendencies and properties of equilibrium policies, such as

- **Existence and Uniqueness:** In some games uniqueness of equilibrium will be established, whereas in others cases, counterexamples will be presented to show that multiple equilibria may occur.
- **Structure of equilibria:** Threshold and switching-curves characterize policies that often appear in dynamic games in networking at equilibrium. We shall also present counterexamples where equilibria with such a structure may be expected, but fail to exist.
- **Tendencies between players in equilibria:** We shall encounter games that exhibit the tendency of "avoiding the crowd" as well as games with the opposite tendency of "joining the crowd". Such tendencies are related to dynamic behavior when the system is not at equilibrium. We shall mention, in particular, some recent extensions

due to Yao [38] of submodular games (that were initially introduced by Topkis [35]) along with some queueing examples, which provide tools to identify the two types of tendencies.

- **Efficiency of the equilibrium:** Some of the examples will be used to illustrate how the equilibrium performs in terms of payoff as compared to a globally (cooperative) optimal solution that would be obtained if a there were a single decision maker.

We briefly mention some references on this topic and other closely related ones. The reference [9] presents a survey on zero-sum stochastic games in queueing problems. Such games model a worst case type design for the case that some parameters that are unknown can be modeled as if they were controlled by an opponent player. This reference also presents the methodology for handling unbounded costs which are typical for queueing applications, and for obtaining structural results of saddle point policies. An example in which such techniques have been used in non-zero sum games in queues can be found in [5]. Other recent surveys on applications of games in networks can be found in [6, 13]. Finally, many games occurring in queueing systems are described in details in [22] and references therein.

The structure of the paper is as follows. In the next four sections we address models that fall into the categories of whether to queue or not (Section 1.2), when to queue (Section 1.3), where to queue (Section 1.4) and how much to queue (Section 1.5). All along these sections we address study properties and features of the equilibria. We proceed with two methodological sections that focus on applications of S-modular games in queueing problems, and establish some properties of the values, of policies, and of the convergence to equilibrium.

1.2 To queue or not to queue: The PC-MF game

We summarize in this Section the networking problem from [12] in which individuals that require service have to make the decision whether to be served at some powerful facility that is shared among all the users that are present there, or to be served at a private slower service facility. As illustrated in Fig. 1.1, the slow facility may represent a personal computer (PC), the fast facility may represent a powerful workstation or a mainframe (MF), and an individual may represent a job that has to be processed.

The model.

Requests for processing jobs arrive to the system. Inter-arrival times are general i.i.d. (independent and identically distributed) with mean λ^{-1}. Upon arrival of a request, the user connects to MF and observes the load. Based on this information it decides whether to queue or not to queue there.

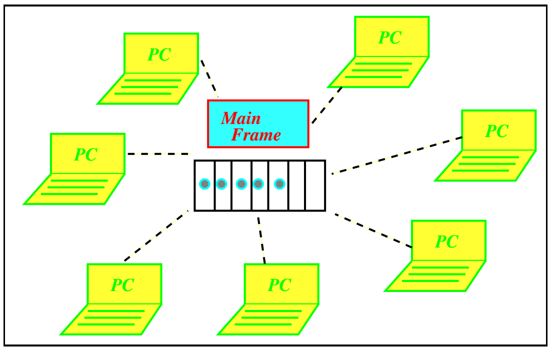

FIGURE 1.1. To queue for a fast service at a mainframe or to receive a slow service at a personal PC?

MF shares its computing capacity between all present users there. This is called **Processor Sharing** discipline. The service at MF is exponentially distributed with rate $\mu(x)$. Let x be the number of jobs queued there. Typically $\mu(x) = \mu$ is a constant and the service intensity per customer is μ/x. But the results below hold more generally provided that $\mu(x)/x$ decreases in x [12].

The PCs offer fixed expected service time of θ^{-1}. (Here, θ is assumed to be the same for all PCs, although one can also solve for different θ's, see [15]).

Denote
$X(t)$:= the number of customers at MF at time t,
$T_k,\ k \geq 0$:= arrival time of job C_k, where $0 = T_0 < T_1 < T_2 < \ldots$ ($t = 0$ is the arrival time of the first customer),

The queue-length process $X(t)$ is defined to be left-continuous (thus $X(T_k)$ is the queue length just *prior* to the possible admission of customer C_k to MF).

The kth arriving individual, C_k, must decide at T_k whether to join MF, or to be served at a PC, after observing $X(T_k)$.

Denote u_k:= the strategy for C_k, is the probability of joining MF if $X(T_k) = x$. Let U be the class of such maps, and let $\pi := (u_0, u_1, \ldots)$ denote a multi-strategy of all customers.

Performance measure.

Define
w_k:= service duration of customer C_k,
$W_k(x,\pi)$:= expected service time of C_k, given that x customers are present at MF at her arrival.
$V_k(x,\pi)$:= expected service time of C_k if it chose to be served at MF, given that x customers are present at MF at her arrival.

Then

$$W_k(x,\pi) = u_k(x)V_k(x,\pi) + (1-u_k(x))\theta^{-1} ,$$

Observe that V_k depends on π through $\{u_l,\, l>k\}$, the decision rules of subsequent customers.

Each customer wishes to minimize her own service time.

To this end, she should evaluate her expected service time at the two queues, namely $V_k(x,\pi)$ and θ^{-1}, and choose the lower one.

Definition 1.1. (**Threshold Policies**) *For any $0 \le q \le 1$ and integer $L \ge 0$, the decision rule u is an $[L,q]$-threshold rule if*

$$u(x) = \begin{cases} 1 & \text{if} \quad x < L \\ q & \text{if} \quad x = L \\ 0 & \text{if} \quad x > L \end{cases} \quad . \tag{1.1}$$

A customer who employs this rule joins MF if the queue length x is smaller than L, while if $x = L$ she does so with probability q. Otherwise she joins PC.
An $[L,q]$ threshold rule will be denoted by $[g]$ where $g = L+q$. Note that $[L,1]$ and $[L+1,0]$ are identical.

Theorem 1.1.
(i) For any equilibrium policy $\pi^* = (u_0, u_1, \ldots)$, each decision rule u_k is a threshold rule.
(ii) A symmetric equilibrium policy $\pi^* = (u^*, u^*, \ldots)$ exists, is unique, and u^* is a threshold rule.

Basic steps of proof:
(i) For every policy π and $k \ge 0$, one shows that $V_k(x,\pi)$ is strictly increasing in x [12].
(ii) Assume that all jobs other than C_k use a threshold policy $[g]$. Then $V_k(x,[g]^\infty)$ can be shown to be:
(1) strictly increasing in g, and
(2) continuous in g.

We note that the best response to a policy g is either a singleton or a whole interval of the form $[i, i+1]$ where i is an integer. The latter occurs when player k is indifferent between a threshold $[i,0]$ and $[i,1]$. It then easily follows that a fixed point exists which provides a symmetric threshold equilibrium [12]. ∎

Numerical Examples
The following numerical example with $\theta = 10, \mu = 100$ is taken from [12]. In that reference, a recursive procedure is given to compute $V(x,[L,q])$. Fig. 1.2 is obtained using that procedure to compute the performance of a threshold x when all other users have threshold $[15,1]$, for various values of input rates. The value of the threshold that defines the best response to

the threshold $[15, 1]$ for a given input rate will then be obtain at the point where $V(x, [15, 1])$ becomes larger than θ^{-1}.

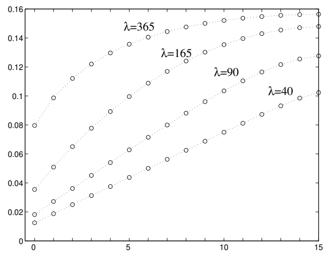

FIGURE 1.2. $V(x, [15, 1])$ **as a function of** x **for various** λ**'s**

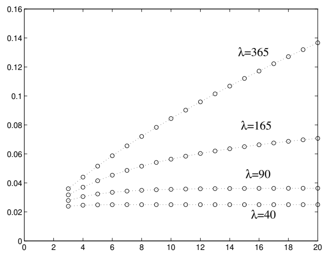

FIGURE 1.3. $V(3, [L, 1])$ **as a function of** L **for various** λ**'s**

Fig. 4 shows the equilibrium symmetrical threshold that is obtained for this problem. It also compares it to the socially optimal problem (that would be obtained if a single decision maker tried to minimize the expected average waiting time of all customers). We see that the equilibrium threshold is always larger than the one corresponding to the socially optimal threshold. In particular, as the input rate λ becomes very large, (overload conditions) the equilibrium threshold will converge to 10 and the socially optimal threshold will converge to 1. The equilibrium threshold of 10 in overload provides an expected waiting time at the MF that equals to θ^{-1}. It is ten times larger that the expected waiting time at MF for the socially optimal policy.

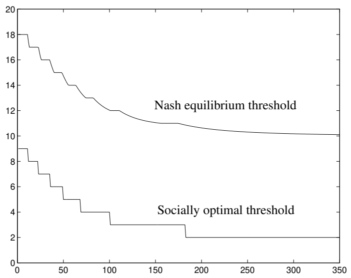

FIGURE 1.4. $g^* = L^* + q^*$ **as a function of** λ

Conclusions We have observed in this first example that when all other users have a threshold g, the threshold of the best response policy is decreasing in that g. Thus the behavior of a user here goes in the opposite direction as the behavior of the rest of the users. Following the terminology of Hassin and Haviv [20], we call this property "Avoiding the Crowd". It is a similar kind of behavior than we would expect in road traffic, where drivers try to avoid rush hours. In future sections we shall also encounter the opposite type of behavior, of "Joining the Crowd". In our example, this behavior implied the uniqueness of the equilibrium.

1.3 When to queue?

We present two games that fall into this categories. The first is the problem of when to arrive at a bank. The second is of when to retry to make a phone call. We then briefly mention a third related game that also involves retransmission decisions and that is related to access to a radio channel.

1.3.1 When to arrive at the bank?

The aim of this example is to illustrate that a very common situation that we frequently face is in fact a queueing game, where the decision is of when to queue.

We consider the following scenario. A bank opens between 9h00 to 12h00. All customers that arrive before 12h00 are served that day. A random number X of customers wish to get a service at a given day. Service time are i.i.d. with exponential distribution. The order of service is FCFS (First Come First Served). Each customer wishes to minimize her own expected waiting time, i.e. the expected time elapsed from her arrival at the bank till she gets served.

This problem has been studied in [17]. The authors show that there exists a symmetric equilibrium distribution F of the arrival time with support $[T_0, 12h00]$ for all players, with $T_0 < 9h00$, such that if all customers follow F, then no individual has an incentive to deviate from using the distribution F.

We finally note that if we eliminate the FCFS regime among those who arrive before the bank opens, then $T_0 = 9h00$. This could reduce average waiting time!

1.3.2 When to retry to make a phone call?

1.3.3 games in retrial queues

We are all used to obtaining occasionally a busy signal when attempting to make phone calls. This signal is either due to the fact that the destination is currently busy with another phone call or to congestion that may cause calls to be blocked. A person whose call is blocked may typically retry calling. The goal of the example, taken from [16, 21], is to analyze the individual's choice of time between retrials.

The model:

Calls arrive according to a Poisson process with average rate λ. Service rates are i.i.d. with mean τ and finite variance σ^2. Let $S^2 := \tau^2 + \sigma^2$ and $\rho := \lambda\tau$. Between retrials calls are said to be "in orbit". Times between retrials of the ith call in orbit are exponentially distributed with expected value of $1/\theta_i$.

We note that exponentially distributed times are frequently used in models in telecommunications, and in particular in telephony models due to much accumulated experimental results. In particular, it is known that the duration of phone calls is well modeled by the exponential distribution. The exponential distribution is also attractive mathematically since it leads to simple Markovian models due to its memoryless property (i.e. the property $P(X > t + s | X \geq t) = P(X > s)$ for any $t, s > 0$).

We assume that each retrial costs c, and the waiting time costs w per time unit.

We begin by presenting the performance of the socially optimal policy [27].

The socially optimal policy.

The expected time in orbit when all retrials use the same parameter θ is

$$W = \frac{\rho}{1-\rho}\left(\frac{1}{\theta} + \frac{S^2}{2\tau}\right),$$

so the average cost per call is

$$(w + c\theta)W =$$
$$\frac{\rho}{1-\rho}\left(\frac{cS^2}{2\tau}\theta + \frac{w}{\theta}\right) + \frac{\rho}{1-\rho}\left(\frac{wS^2}{2\tau} + c\right).$$

This is minimized at

$$\theta^* = \frac{\sqrt{2w\tau/c}}{S}.$$

A remarkable observation is that this rate is independent of the arrival rate!

If θ^* is used, the two terms that depend on θ turn out to be equal: the <u>waiting</u> cost and the <u>retrial</u> costs coincide.

The game

In the game case, the authors of [27] compute $g(\theta, \gamma)$, which is the expected waiting time of an individual who retries at rate γ while all the others use retrial rate θ. This allows them to obtain the equilibrium rate:

$$\theta_e = \frac{w\rho + \sqrt{w^2\rho^2 + 16w\tau c(1-\rho)(2-\rho)/S^2}}{4c(1-\rho)}.$$

We observe that θ_e monotone increases to infinity as λ increases (to $1/\sigma$).

Conclusions.

Here are some conclusions from this example.

1. The ratio between the cost at equilibrium and the globally optimal cost tends to infinity. Thus, the noncooperative nature of the game leads to high inefficiency.

2. The equilibrium retrial rate is larger than the globally optimal retrial rate. They tend to coincide as $\rho \to 0$.

3. Both equilibrium and optimal retrial rates are monotone decreasing in the variance of the service times.

4. There is a unique equilibrium.

1.3.4 Access to a radio channel: when to retransmit

Phone calls are not the only example where we find retransmissions. Another example, typical to cellular communications as well as satellite communications, is in the access of packets to a common radio channel. If more than one terminal attempt to transmit a packet at the same time then a collision occurs and all terminals involved in the collision have to retransmit their packet later. A terminal wishes typically to maximize the throughput (i.e. the expected average number of successful transmissions). A further cost may be incurred per each transmission attempt (it could represent the energy consumption). The problem has various formulations as stochastic games, in which each terminal controls the time till it attempts a retransmission of a packet that has collided, or the probability of its retransmission at a given time slot [8, 24, 28]. In [8], a comparison between a cooperative solution and a noncooperative solution is presented. It is shown that as the arrival rate of packets increases, the cooperative solution tends to reduce the retransmission probabilities; retransmissions are thus delayed so as to avoid more collisions. Due to this behavior, the throughput was shown to increase with the arrival rate of packets. In the noncooperative case, a single symmetric equilibrium is obtained in which, in contrast, the retransmission probabilities tend to increase with the arrival rate of packets. Thus, as the arrival rate increases, terminals become more aggressive, and the system throughput decreases. For large arrival rates it decreases to zero.

1.4 Where to queue?

We briefly mention some background on the well studied traffic assignment problem in a general network, and then present two examples in more details of stochastic games occurring in simple queueing problems.

1.4.1 The traffic assignment problem

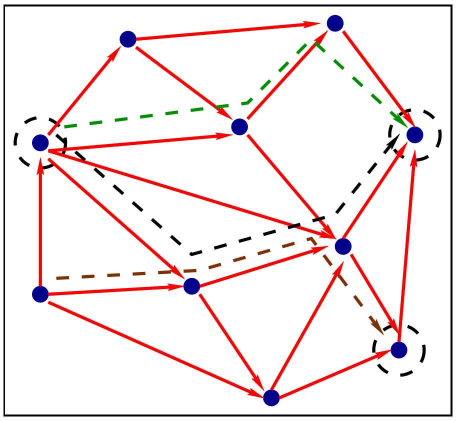

FIGURE 1.5. The assignment problem and competitive routing

One of the most studied game problem that arises in networks is the traffic assignment problem [29, 30]. The generic problem is to determine which path to take in a network, that may be shared by many types of users having different sources and destinations. The costs, or delays (or other performance measures, such as loss probabilities) for each user may be related to the congestion it suffers between the source and destination, and is a function of choices of routes of other users as well. This problem has been studied already in the fifties by Wardrop [36] in the context of road traffic, where a user corresponds to a single driver. The game has been modeled using a continuum number of players where the influence of a single player (driver) on the cost of others can be neglected. Due to this property, such a player is often called atomless. Existence and uniqueness of the equilibrium (known as Wardrop equilibrium) have been established under quite general conditions, and many efficient computational methods have been developed (see [30] for an overview). This type of equilibrium has also been discovered in the context of mobile telecommunications in [18] and in the context of Potential games in [33] (see also [32]). Wardrop principles have also been obtained independently around thirty years before Wardrop in an economical context [31].

An alternative game formulation arises when decisions are not taken by each individual, but instead, by a finite number of big organizations (the players) such as service providers, in the context of telecommunications, or of transportation companies, in the context of road traffic. Such players are often called atomic players. The decision taken by a player is what fraction of its flow to ship over each possible route. Nash equilibrium is then used as the solution concept. Some references in this type of modeling are [29, 23], as well as [30] and references therein.

Till now we have not mentioned the relation between the traffic assignment problem and queueing. It turns out that there is a large class of queueing networks for which the average expected delays or the expected queue size satisfy the assumptions of the traffic assignment models studied in the literature. These are known as BCMP queueing networks (known to have a property called "product form", which makes link costs additive). We refer in particular to the paper by Kameda and Zhang [25, 11] that describes these kinds of networks in the context of the traffic assignment problem.

Interesting extensions to dynamic and differential games context exist for the assignment model. We shall mention [37] and references therein. In this reference, the players have to ship a given amount of flow within a certain period and can decide dynamically at what rate to ship at each instant. A dynamic mixed equilibrium (that combines both atomless and atomic users) is computed using the maximum principle.

1.4.2 The gas station game

The example we present here illustrates the dynamic routing choices between two paths. When a routing decision is made, the decision maker knows the congestion state of only one of the routes; the congestion state in the second route is unknown to the decision maker. The problem originates from the context of two gas stations on a highway [19]. A driver arriving at the first station sees the amount of other cars already queued there and has to decide whether to join that queue, or to proceed to the next gas station. The state of the next gas station (i.e. the number of cars there) is not available when making the decision. The situation is illustrated in Fig. 1.4.2. The exact mathematical solution of the model was obtained in [10] and we describe it below.

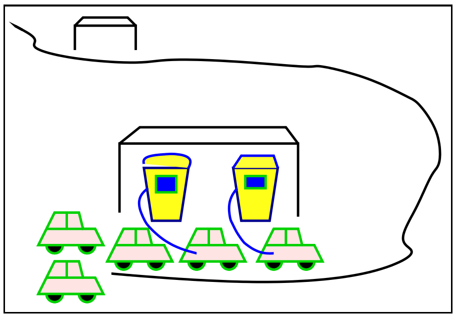

FIGURE 1.6. The gas station problem

This problem has natural applications in telecommunication networks:

when making routing decisions for packets in a network, the state in a down stream node may become available after a considerable delay, which makes that information irrelevant when taking the routing decisions.

Although the precise congestion state of the second route is unknown, its probability distribution, which depends on the routing policy, can be computed by the router.

We assume that the times that corresponds to the arrivals instants of individuals is a Poisson process with rate λ. Each arrival is a player, so there is a countable number of players, each of whom takes one routing decision, of whether to join the first or the second service station. A player eventually leaves the system and does not affect it anymore, once it receives service.

To obtain an equilibrium, we need to compute the joint distribution of the congestion state in both routes as a function of the routing policy.

We restrict to random threshold policies (n, r):

- if the the number of packets in the first path is less than or equal to $n-1$ at the instance of an arrival, the arriving packet is sent to path 1.
- If the number is n then it is routed to path 1 with probability r.
- If the number of packets is greater than n then it is routed to path 2.

The delay in each path is modeled by a state dependent queue:

- Service time at queue i is exponentially distributed with parameter μ_i
- Global inter-arrival times are exponential i.i.d. with parameter λ.

When all arrivals use policy (n, r), the steady state distribution is obtained by solving the steady state probabilities of the continuous time Markov chain [10].

If an arrival finds i customers at queue 1, it computes

$$E_i[X_2] = E\left[X_2 | X_1 = i\right]$$

and takes a routing decision according to whether

$$T^{n,r}(i,1) := \frac{i+1}{\mu_1} \leq_? \frac{E_i[X_2]+1}{\mu_2} =: T^{n,r}(i,2).$$

To compute it, the arrival should know the policy (n, r) used by all previous arrivals.

If the decisions of the arrival as a function of i coincide with (n, r) then (n, r) is a Nash equilibrium.

The optimal response against $[g] = (n, r)$ is monotone decreasing in g. This is the **Avoid The Crowd** behavior.

Computing the conditional distributions, one can show [10] that there are parameters $(\mu_1, \mu_2, \lambda, n, r)$ for which the optimal response to (n, r) is indeed a threshold policy.

Denote

$$\rho := \frac{\lambda}{\mu_1}, \qquad s := \frac{\mu_2}{\mu_1}$$

There are other parameters for which the optimal response to (n, r) is a two-threshold policy characterized by $t^-(n, \rho, s)$ and $t^+(n, \rho, s)$ as follows.

It is optimal to route a packet to queue 2 **if** $t^-(n, \rho, s) \leq X_1 \leq t^+(n, \rho, s)$ **and to queue** 1 **otherwise.**

At the boundaries t^- and t^+ routing to queue 1 or randomizing is also optimal if $T^{n,r}(i, 1) = T^{n,r}(i, 2)$. For parameters in which the best response does not have a single threshold, we cannot conclude anymore what is the structure of a Nash equilibrium.

Example [10]. Consider $n = 3,\ r = 1,\ \rho = \lambda/\mu_1 = 1$ and $s = \mu_2/\mu_1 = 0.56$. We plot in Fig. 1.7 $T^{n,r}(i, 1)$ and $T^{n,r}(i, 2)$ for $i = 0, 1, \ldots, 4$.

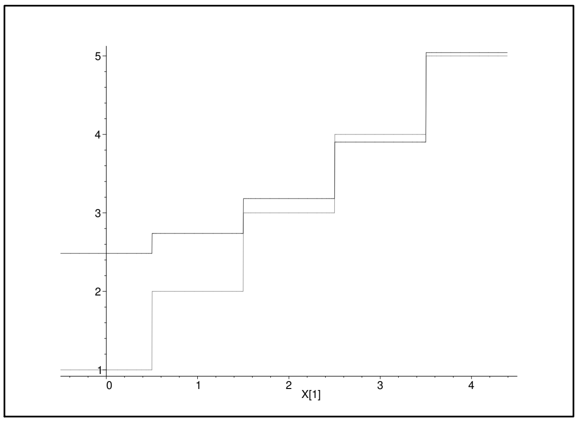

FIGURE 1.7. $T^{n,r}(i, 1)$ **and** $T^{n,r}(i, 2)$

Conclusions: As opposed to the example of the choice of players between a PC and a MF in Sec. 1.2, we saw in this section an example where for some parameters there may be no threshold type (n, r) equilibria. The form of the Nash equilibrium in these cases remains an open problem. Moreover, even the question of existence of a Nash equilibrium is then an open question.

Yet, in the case of equal service rates in both stations, a threshold equilibrium does exist, and it turns out to have the same type of behavior as in the PC-MF game, i.e. it is unique, and the best response has a tendency of "Avoiding the Crowd" [19].

By actually analyzing (numerically) the equilibrium for equal service rates, it was noted in [19] that when players use the equilibrium strategy, then the revenue of the first station is higher than the second one. Thus the additional information that the users have on the state of the first station produces an extra profit to that station. An interesting open problem is whether the second station can increase its profits by using a different pricing than the first station, so that users will have an extra incentive to go to that station. Determining an optimal pricing is also an interesting problem.

1.4.3 Where to queue: queues with priority

We present a second queueing problem modeled as a stochastic game with infinitely many individual players due to [2, 20]. We assume again that players arrive at the system according to a Poisson process with intensity λ, and have to take a decision of whether to join a low priority (second class) or a high priority (first class) queue, as illustrated in Fig. 1.4.3. There is a single server that serves both queues but gives strict priority to first class customers. Thus a customer in the second class queue gets served only when the first class queue is empty. We assume exponentially distributed service time with parameter μ and define $\rho := \lambda/\mu$.

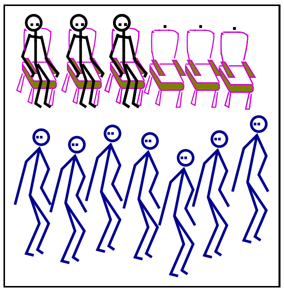

FIGURE 1.8. Choice between first and second class priorities

The game model is then as follows:

The actions Upon arrival, a customer (player) observes the two queues and may purchase the high priority for a payment of an amount θ, or join the low priority queue.

The state: The state is a pair of integers (i, j) corresponding to the number of customers in each queue; i is the number of high priority customers and j, the number of low priority ones.

The analysis of this problem can be considerably simplified by using the following **monotonicity property**, identified in [2]: If for some strategy adopted by everybody, it is optimal for an individual to purchase priority at (i, j), then he must purchase priority at (r, j) for $r > i$.

This implies that the problem has an effective **lower dimensional state space:** It follows that starting at $(0, 0)$ and playing optimally, there is some

n such that the only reachable states are

$$(0, j),\ j \leq n, \quad \text{and } (i, n),\ i \geq 1.$$

Indeed, due to monotonicity, if at some state $(0, m)$ it is optimal not to purchase priority, it is also optimal at states $(0, i)$, for $i \leq m$. Let $n-1$ be the largest such state. Then starting from $(0,0)$ we go through states $(0, i)$, $i < n$, until $(0, n-1)$ is reached. At $(0, n)$ it is optimal to purchase priority. We then move to state $(1, n)$.

The low priority queue does not decrease as long as there are high-priority customers. Due to monotonicity, it also does not increase as long as there are high-priority customers since at (i, n), $i \geq 1$ arrivals purchase priority! Therefore we remain at (i, n), as long as $i \geq 1$.

The Equilibrium. Suppose that the customers in the population, except for a given individual, adopt a common threshold policy $[g]$. Then the optimal threshold for the individual is non-decreasing in g.

This property is called **"Follow The Crowd" Behavior**

This property clearly implies *Existence* of an equilibrium, that can be obtained by a monotone best response argument.

However, it turns out that there is no uniqueness of the equilibrium! Indeed, Hassin and Haviv have shown in [20] that there may be up to

$$\left\lfloor \frac{1}{1-\rho} \right\rfloor$$

pure threshold Nash equilibria, as well as other mixed equilibria! They further present numerical examples of multiple equilibria.

We conclude that in this problem we have definitely a different behavior of the equilibria than in the previous stochastic games in which we had threshold equilibria (the PC-MF game and the gas station game).

1.5 How much to queue?

Two main approaches exist for controlling the congestion at the network in high speed communication systems. The first, in which sources control their *transmission rate*, has been adopted by the ATM-forum [1] (an international standardization organism for high speed telecommunication networks) for applications that have flexibility in their requirement for bandwidth (and which then use the so called "Available Bit Rate" service). Although the general approach and signaling mechanisms have been standardized, the actual way to control the sources has been left open and is still an open area of research and development. An alternative approach is the so called "*window based*" flow control which is used at the Internet. An example of a stochastic game based on the latter approach is given in [26].

We present in this section a linear quadratic differential game model for flow control based on the first approach, which has been introduced and analyzed in [7], in which M users control their transmission rate into a single bottleneck queue. Thus, each user has to answer dynamically the question of "how much to queue", see Fig. 1.9.

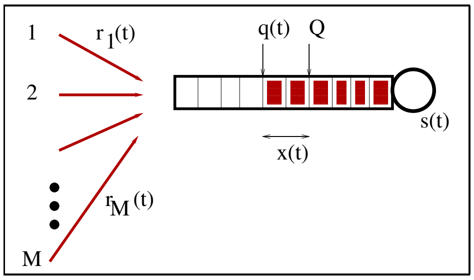

FIGURE 1.9. A competitive flow control model

The output rate of the queue (i.e. the server's rate) is given by $s(t)$. The *controlled* input rate of user m is denoted by $r_m(t)$. Let $q(t)$ denote the instantaneous queue length.

It is desirable for telecommunication applications to keep the instantaneous size of the queue close to some target: Q. Indeed, if the queue is too large then information packets might be lost due to overflow, and moreover, packets suffer large delays. If the queue is empty, on the other hand, then the output rate is constrained by the input rate, which might be lower than $s(t)$. We thus may have underutilization of the available service rate and we loose in throughput.

Define $x(t) := q(t) - Q$ to be the shifted instantaneous queue length. It will serve as the state of the queue. We assume that there is some desirable share of bandwidth available for user m, given by $a_m s(t)$, where $\sum_m a_{m=1}^M = 1$.

Define $u_m(t) := r_m(t) - a_m s(t)$ to be the shifted control.

Ignoring the nonlinearity in the dynamics that are due to boundary effects (at an empty queue or at a full queue), we obtain the following idealized dynamics:

$$\frac{dx}{dt} = \sum_{m=1}^{M} (r_m - a_m s) = \sum_{m=1}^{M} u_m \,, \tag{1.2}$$

Justification for using the linearized model are presented in [7].

Policies and information: We consider the following class of history dependent policies for all users:

$$u_m(t) = \mu_m(t, x_t), \qquad t \in [0, \infty) \,.$$

μ_m is piecewise continuous in its first argument, piecewise Lipschitz continuous in its second argument. The class of all such policies for user m is $\mathcal{U}_m$.

Objectives. The cost per user is a linear combination of the cost related to deviation from the desirable queue length value Q and deviating from the desirable share of the bandwidth. We consider two versions that differ by some scaling factor:

- N1: the individual cost to be minimized by user m $(m \in \mathcal{M} = \{1, ..., M\})$ is

$$J_m^{N1}(u) = \int_0^\infty \left(|x(t)|^2 + \frac{1}{c_m} |u_m(t)|^2 \right) dt\,. \tag{1.3}$$

- N2: the individual cost to be minimized by controller m $(m \in \mathcal{M})$ is

$$J_m^{N2}(u) = \int_0^\infty \left(\frac{1}{M} |x(t)|^2 + \frac{1}{c_m} |u_m(t)|^2 \right) dt\,. \tag{1.4}$$

In case N2 the "effort" for keeping the deviations of the queue length from the desired value is split equally between the users.

Nash equilibria. We seek a multi-policy $\mu^* := (\mu_1^*, ..., \mu_M^*)$ such that no user has an incentive to deviate from, i.e.

$$J_m^{N1}(\mu^*) = \inf_{\mu_m \in \mathcal{U}_m} J_m^{N1}([\mu_m | \mu_{-m}^*]) \tag{1.5}$$

where $[\mu_m|\mu_{-m}^*]$ is the policy obtained when for each $j \neq m$, player j uses policy μ_j^*, and player m uses μ_m. We define similarly the problem with the cost J^{N_2}.

Main results We shall show that the flow control game has a simple computable equilibrium and value. We further show a uniqueness result.
<u>The equilibrium:</u> For case Ni $(i = 1, 2)$, there exists an equilibrium given by

$$\mu_{Ni,m}^*(x) = -\beta_m^{Ni} x\,, \ m = 1, ..., M,$$

where β_m^{Ni} is given by

$$\beta_m^{N1} = \overline{\beta}^{(N1)} - \sqrt{\overline{\beta}^{(N1)^2} - c_m}$$

where $\overline{\beta}^{(N1)} := \sum_{m=1}^M \beta_m^{N1}$, $i = 1, 2$, are the unique solutions of

$$\overline{\beta}^{(N1)} = \frac{1}{M-1} \sum_{m=1}^M \sqrt{(\overline{\beta}^{(N1)})^2 - c_m}$$

and for the case N2:

$$\beta_m^{N2} = \overline{\beta}^{(N2)} - \sqrt{\overline{\beta}^{(N2)^2} - \frac{c_m}{M}},$$

where $\overline{\beta}^{(N2)} := \sum_{m=1}^{M} \beta_m^{N2}$, $i = 1, 2$, are the unique solutions of

$$\overline{\beta}^{(N2)} = \frac{1}{M-1} \sum_{m=1}^{M} \sqrt{(\overline{\beta}^{(N2)})^2 - \frac{c_m}{M}} = \frac{\overline{\beta}^{(N1)}}{\sqrt{M}} .$$

Moreover,

$$\beta_m^{N1} = \beta_m^{N2}\sqrt{M}.$$

For each case, this is the unique equilibrium among stationary policies and is time-consistent.

The value. The costs accruing to user m, under the two Nash equilibria above, are given by

$$J_m^{N1}(\mu_{N1}^*) = \frac{\beta_m^{N1}}{c_m} x^2$$

and

$$J_m^{N2}(\mu_{N2}^*) = \frac{\beta_m^{N2}}{c_m} x^2 = \frac{1}{\sqrt{M}} J_m^{N1}(u_{N1}^*).$$

The symmetric case. In the case $c_m = c_j =: c$ for all $m, j \in \mathcal{M}$ we obtain:

$$\beta_m^{N1} = \sqrt{\frac{c}{2M-1}}, \qquad \text{and} \qquad \beta_m^{N2} = \sqrt{\frac{c}{M(2M-1)}}, \quad \forall m \in \mathcal{M} \, ;$$

The case of $M = 2$. We assume general values of c_m's. we have for $m = 1, 2, \ j \neq m$,

$$\beta_m^{N1} = \left[-\frac{2c_j - c_m}{3} + 2\frac{\sqrt{c_1^2 - c_1 c_2 + c_2^2}}{3} \right]^{1/2}, \qquad \beta_m^{N2} = \frac{\beta_m^{N1}}{\sqrt{2}}.$$

If moreover, $c_1 = c_2 = c$ then $\beta_m^{N1} = \sqrt{c/3}$, $\beta_m^{N2} = \sqrt{c/6}$.

We now sketch the proof for N1. The one for N2 is similar.

Proof for (N1): Choose a candidate solution

$$u_m^*(x) = -\beta_m x \, , \; m = 1, ..., M, \quad \text{where } \beta_m = \overline{\beta} - \sqrt{\overline{\beta}^2 - c_m}$$

where $\overline{\beta} := \sum_{m=1}^{M} \beta_m$, are the unique solution of

$$\overline{\beta} = \frac{1}{M-1} \sum_{m=1}^{M} \sqrt{\overline{\beta}^2 - c_m}.$$

Fix u_j for $j \neq m$. Player m is faced with a linear quadratic optimal control problem with the dynamics

$$dx/dt = u_m - \beta_{-m} x \, , \ \beta_{-m} = \sum_{j \neq m} \beta_j$$

and cost $J_m^{N1}(u)$ that is strictly convex in u_m. Her optimal response is $u_m = -c_m P_m x$, where P_m is the unique positive solution of the Riccati equation

$$-2\beta_{-m} P_m - P_m^2 c_m + 1 = 0. \tag{1.6}$$

Denoting $\beta'_m = c_m P_m$, we obtain from (1.6)

$$\beta'_m = f_m(\beta_{-m}) := -\beta_{-m} + \sqrt{\beta_{-m}^2 + c_m}.$$

u is in equilibrium if and only if $\beta' = \beta$, or

$$\overline{\beta}^2 = \beta_{-m}^2 + c_m. \tag{1.7}$$

Hence

$$\beta_m = \overline{\beta} - \sqrt{\overline{\beta}^2 - c_m}.$$

Summing over $m \in \mathcal{M}$ we obtain

$$\Delta := \overline{\beta} - \frac{1}{M-1} \sum_{m=1}^{M} \sqrt{\overline{\beta}^2 - c_m} = 0$$

Uniqueness follows since

- Δ is strictly decreasing in $\overline{\beta}$ over the interval $[\max_m \sqrt{c_m}, \infty)$,
- it is positive at $\overline{\beta} = \max_m \sqrt{c_m}$ and
- it tends to $-\infty$ as $\beta \to \infty$.

1.5.1 Greedy decentralized algorithms

The Nash equilibrium requires some coordination, i.e. a (non-binding) agreement according to which all user follow the Nash equilibrium. Moreover, to compute the equilibrium, a player needs to have the knowledge of other individual utilities (c_m). Both the coordination as well as the knowledge of others' utilities are restrictive and non-realistic assumptions. This motivated the authors in [7] to propose several greedy decentralized "best response" algorithms.

A greedy "best response" algorithm is defined by the following four conditions [14]:

- (i) Each user updates from time to time its policy by computing the best response against the most recently announced policies of the other users.
- (ii) The time between updates is sufficiently large, so that the control problem faced by a user when it updates its policy is well approximated by the original infinite horizon problem.
- (iii) The order of updates is arbitrary, but each user performs updates infinitely often.
- (iv) When the nth update occurs, a subset $K_n \subset \{1, \ldots, M\}$ of users simultaneously update their policies.

Proposed algorithms

- **Parallel update algorithm (PUA):** $K_n = \{1, ..., M\}$ for all n.
- **Round robin algorithm (RRA):** K_n is a singleton for all n and equals $(n+k)\text{mod}M + 1$, where k is an arbitrary integer.
- **Asynchronous algorithm (AA):** K_n is a singleton for all n and is chosen at random.

The initial policy used by each user is linear.
$\beta^{(n)}$:= value of the linear coefficient defining the policies corresponding to the end of the nth iteration.
The optimal response at each step n:

$$\beta_m^{(n)} = \begin{cases} f_m(\beta_{-m}^{(n-1)}) & \text{if } m \in K_n \\ \\ \beta_m^{(n-1)} & \text{otherwise} \ , \end{cases} \tag{1.8}$$

where

$$f_m(\beta_{-m}) := -\beta_{-m} + \sqrt{\beta_{-m}^2 + c_m}. \tag{1.9}$$

Convergence results We briefly mention the convergence results obtained in [7].

Theorem 1.2. *Consider PUA.*
(i.a) Let $\beta_k^{(1)} = 0$ for all k. Then $\beta_k^{(2n)}$ monotonically decrease in n and $\beta_k^{(2n+1)}$ monotonically increase in n, for every player k, and thus, the following limits exist: $\hat{\beta}_k := \lim_{n\to\infty} \beta_k^{(2n)}$, $\tilde{\beta}_k := \lim_{n\to\infty} \beta_k^{(2n+1)}$.
(i.b) Assume that $\hat{\beta}_k = \tilde{\beta}_k$ (defined as above, with $\beta_k^{(1)} = 0$ for all k). Consider now a different initial condition satisfying either $\beta_k^{(1)} \leq \beta_k$ for all k, (where β_k is the unique Nash) or $\beta_k^{(1)} \geq \beta_k$ for all k. Then for all k, $\lim_{n\to\infty} \beta_k^{(n)} = \beta_k$.

Global convergence. A global convergence result is obtained for $M = 2$:

Theorem 1.3. *If*
– (ii.a) $M = 2$, and either $\beta_k^{(1)} \leq \beta_k$ for all k, or $\beta_k^{(1)} \geq \beta_k$ for all k; or if
– (ii.b) $\beta_k^{(1)}$ and $c := c_k$ are the same for all k,
then $\beta^{(n)}$ converges to the unique equilibrium β^*.

Local convergence.

Theorem 1.4. *For arbitrary c_k, there exists some neighborhood V of the unique equilibrium β^* such that if $\beta_k^{(1)} \in V$ then $\beta^{(n)}$ converges to the unique equilibrium β^*.*

Numerical examples Some examples from [7] on the convergence of greedy algorithm are presented below. The parallel updates are seen to converge slowest with an oscillating behavior.

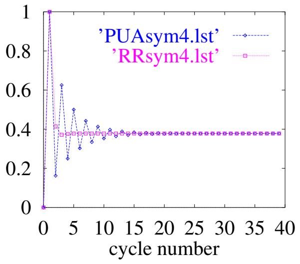

FIGURE 1.10. PUA versus RRA for $M = 4$

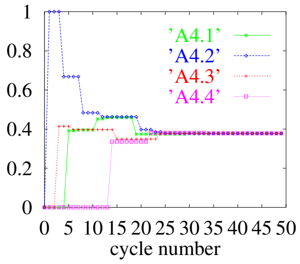

FIGURE 1.11. AA for $M = 4$

1.6 S-modular games

We have seen in several examples in previous Sections that equilibria best response strategies have either the "Join the Crowd" property or the the "Avoid the Crowd" property, which typically leads to threshold equilibria policies. These properties turn out to be also useful when we seek to obtain convergence to equilibria from a non-equilibrium initial point. These issues will be presented in this section within the framework of S-modular games due to Yao [38] who extends the notion of submodular games introduced by Topkis [35].

1.6.1 Model, definitions and assumptions

General model are developed in [35, 38] for games where the strategy space S_i of player i is a compact sublattice of $R^{\overline{m}}$. By sublattice we mean that it has the property that for any two elements x, y that are contained in S_i, also $\min(x, y)$ (denoted by $x \wedge y$) and $\max(x, y)$ (denoted by $x \vee y$) are contained there (by $max(x, y)$ we mean the componentwise max, and similarly with the min). We describe below the main results for the case that $\overline{m} = 1$.

Definition 1.2. *The utility f_i for player i is supermodular if and only if*

$$f_i(x \wedge y) + f_i(x \vee y) \geq f_i(x) + f_i(y).$$

It is submodular if the opposite inequality holds.

If f_i is twice differentiable then supermodularity is equivalent to

$$\frac{\partial^2 f_i(x)}{\partial x_1 \partial x_2} \geq 0.$$

Monotonicity of maximizers. The following important property was shown to hold in [35]. Let f be a supermodular function. Then the maximizer with respect to x_i is increasing in $x_j, j \neq i$.

More precisely, define the best response

$$BR_1^*(x_2) = argmax_{x_1} f(x_1, x_2);$$

if there are more than one argmax above we shall always limit ourselves to the smallest one (or always limit ourselves to the largest one). Then $x_2 \leq x_2'$ implies $BR_1^*(x_2) \leq BR_1^*(x_2')$. This monotonicity property holds also for non-independent policy sets such as (1.10), provided that they satisfy the ascending property (defined below).

Definition 1.3. (**Monotonicity of sublattices**) *Let A and B be sublattices. We say that $A \prec B$ if for any $a \in A$ and $b \in B$, $a \wedge b \in A$ and $a \vee b \in B$.*

Next, we introduce some properties on the policy spaces.

Consider two players. We allow S_i to depend on x_j

$$S_i = S_i(x_j), \quad i, j = 1, 2, \quad i \neq j. \tag{1.10}$$

<u>Monotonicity of policy sets</u> We assume

$$x_j \leq x_j' \Longrightarrow S_i(x_j) \prec S_i(x_j').$$

This is called the **Ascending Property**. We define similarly the **Descending Property**.

<u>Lower semi continuity of policies</u> We say that the point to set map $S_i(\cdot)$ is lower semi continuous if for any $x_j^k \to x_j^*$ and $x_i^* \in S_i(x_j^*)$ $(j \neq i)$, there exist $\{x_i^k\}$ s.t. $x_i^k \in S_i(x_j^k)$ for each k, and $x_i^k \to x_i^*$.

1.6.2 Existence of Equilibria and Round Robin algorithms

Consider an n-player game. Yao [38, Algorithm 1] and Topkis [35, algorithm I] consider a greedy round robin scheme where at some infinite strictly increasing sequence of time instants T_k, players update their strategies using each the best response to the strategies of the others. Player l updates at times T_k with $k = mn + l$, $m = 1, 2, 3, \ldots$.

Assume lower semi-continuity and compactness of the strategy sets. Under these conditions, Supermodularity together with the ascending property imply monotone convergence of the payoffs to an equilibrium [38]. The monotonicity is in the same direction for all players: the sequences of strategies for each player either all increase or all decrease.

The same type of result is also obtained in [38, Thm. 2.3] with submodularity instead of supermodularity for the case of two players, where the ascending property is replaced by the descending property. The monotone convergence of the round robin policies still holds but it is in opposite directions: the sequence of responses of one player increases to his equilibrium strategy, while the ones of the other player decreases.

In both cases, there need not be a unique equilibrium.

Yao [38] further extends these results to cases of costs (or utilities) that are submodular in some components and supermodular in others. The notion of s-modularity is used to describe either submodularity or supermodularity. Another extension in [38] is to vector policies (i.e. a strategy of a player is in a sublattice of R^k).

Next we present several examples for games in queues where s-modularity can be used. The first two examples are due to Yao [38].

1.6.3 Example of supermodularity: queues in tandem

Consider a set of queues in tandem. Each queue has a server whose speed is controlled. The utility of each server rewards the throughput and penalizes the delay. Under appropriate conditions, it is then shown in [38] that the players have compatible incentives: if one speeds up, the other also want to speed up.

More precisely, consider two queues in tandem with i.i.d. exponentially distributed service times with parameters μ_i, $i = 1, 2$. Let $\mu_i \leq u$ for some constant u. Server one has an infinite source of input jobs There is an infinite buffer between server 1 and 2. **The throughput** is given by $\mu_1 \wedge \mu_2$.

The expected number of jobs in the buffer is given [38] by

$$\frac{\mu_1}{\mu_2 - \mu_1}$$

when $\mu_1 < \mu_2$, and is otherwize infinite.

Let

- $p_i(\mu_1 \wedge \mu_2)$ be the profit of server i,
- $c_i(\mu_i)$ be the operating cost,
- $g(\cdot)$ be the inventory cost.

The utilities of the players are defined as

$$f_1(\mu_1, \mu_2) := p_1(\mu_1 \wedge \mu_2) - c_1(\mu_1) - g\left(\frac{\mu_1}{\mu_2 - \mu_1}\right)$$
$$f_2(\mu_1, \mu_2) := p_2(\mu_1 \wedge \mu_2) - c_2(\mu_2) - g\left(\frac{\mu_1}{\mu_2 - \mu_1}\right).$$

The strategy spaces are given by

$$S_1(\mu_2) = \{\mu_1 : 0 \leq \mu_1 \leq \mu_2\},$$
$$S_2(\mu_1) = \{\mu_2 : \mu_1 \leq \mu_2 \leq u\}.$$

It is shown in [38] that if g is convex increasing then f_i are supermodular. So we can apply the results of the previous subsection, and obtain (1) the property of "joining the crowd" of the best response policies, (2) existence of an equilibrium, (3) convergence to equilibrium of some round robin dynamic update schemes.

1.6.4 *Example of submodularity: flow control*

We consider now an example for submodularity. There is a single queueing centre with two input streams with Poisson arrivals with rates λ_1 and λ_2. The rates of the streams are controlled by 2 players.

The queueing center consists of c servers and no buffers. Each server has one unit of service rate.

When all servers are occupied, an arrival is blocked and lost.

The blocking probability is given by the Erlang loss formula:

$$B(\lambda) = \frac{\lambda^c}{c!}\left[\sum_{k=0}^{c} \frac{\lambda^k}{k!}\right]^{-1}$$

where $\lambda = \lambda_1 + \lambda_2$.

Suppose user i maximizes

$$f_i = r_i(\lambda_i) - c_i(\lambda B(\lambda)).$$

c_i is assumed to be convex increasing. $\lambda B(\lambda)$ is the total loss rate.

Then it is easy to check [38] that f_i are submodular.

Two different settings can be assumed for the strategy sets. In the first, the available set for player i consists of $\lambda_i \leq \overline{\lambda}$. Alternatively, we may consider that the strategy sets of the players depend on each other and the sum of input rates has to be bounded: $\lambda \leq \overline{\lambda}$. Then S_i satisfy the descending property. We can thus apply again the results of the subsection 1.6.2.

1.6.5 A flow versus service control

Exponentially inter-arrivals as we used in previous examples are quite appealing to handle mathematically, and they can model sporadic arrivals, or alternatively, information packets that arrive one after the other but whose size can be approximated by an exponential random variable. In this example we consider, in contrast, a constant time T between arrivals of packets, which can be used for modeling ATM (Asynchronous Transfer Mode) networks in which information packet have a fixed size.

We consider a single node with a periodic arrival process, in which the first player controls the constant time period $T \in [\underline{T}, \overline{T}]$ between any two consecutive arrivals. We consider a single server with no buffer. The service time distribution is exponential with a parameter $\mu \in [\underline{\mu}, \overline{\mu}]$, which is controlled by the second player. If an arrival finds the server busy then it is lost. The loss probability is given by

$$P_l = \exp(-\mu T),$$

which is simply the probability that the random service time of (the previous) customer is greater than the constant T.

The transmission rate of packets is T^{-1}, but since a fraction P_l is lost then the goodput (the actual rate of packets that are transmitted successfully) is

$$G = \frac{1}{T}(1 - \exp(-\mu T)).$$

We assume that the utility of the first player is the goodput plus some function of the input rate T^{-1}. The server earns a reward that is also proportional to the goodput, and has some extra operation costs which is a function of the service rate μ. In other words,

$$J_1(T, \mu) = \frac{1}{T}(1 - \exp(-\mu T)) + f(T^{-1}), \qquad J_2(T, \mu) = \frac{1}{T}(1 - \exp(-\mu T)) + g(\mu).$$

We then have for $i = 1, 2$

$$\frac{\partial J_i^2}{\partial T \partial \mu} = -\mu \exp(-\mu T) \leq 0.$$

We conclude that the cost is submodular.

1.7 Applications of S-modularity to stochastic games

In the previous examples of S-modularity, we used that notion in static games and applied it directly to the policies. An alternative use of (a stronger version of) S-modularity is in applying it to dynamic programming operators, which allows us to obtain structural results for equilibrium policies. We shall restrict in this section to zero-sum games in which player one maximizes some cost and player two seeks to minimize it.

Before introducing the stochastic game setting, we first introduce the structural results in a setting of a set of matrix games. We shall then explain in Subsection 1.7.2 how this has been used to solve in [3, 4] a stochastic zero-sum game in which both flow and service are controlled. We then solve in the subsequent subsection a more involved stochastic game in which there are several sources of flow. We conclude with a discussion.

1.7.1 Monotone policies is matrix games

Let $\mathbf{X} = \{0, 1, 2, ..., L\}$ where L may be finite or infinite. Consider a set of matrix games $\{S(x, a, b)\}_{a,b}$ indexed by $x \in \mathbf{X}$ where $a \in \mathbf{A}(x)$ and $b \in \mathbf{B}(x)$ are the finite set of actions of player 1 and 2, respectively. We assume that $\mathbf{A}(x)$ is monotone increasing in x and that $\mathbf{B}(x)$ is monotone decreasing in x. If game x is played and the players choose actions a and b, then player 2 pays to player 1 $S(x, a, b)$.

We introduce the following assumptions:

J1 (Strong submodularity): For any $b^1 > b^2$ and any a^1 and a^2,

$$S(x, a^1, b^2) - S(x, a^1, b^1) - [S(x-1, a^2, b^2) - S(x-1, a^2, b^1)] > 0, \quad 0 < x \leq L.$$

J2 (strong supermodularity): For any $a^1 > a^2$ and any b^1 and b^2,

$$S(x, a^2, b^1) - S(x, a^1, b^1) - [S(x-1, a^2, b^2) - S(x-1, a^1, b^2)] < 0, \quad 0 < x \leq L.$$

u_x (resp. v_x) will denote a randomized strategy for player 1 and of player 2, respectively, at game x. A policy $u = \{u_x\}, x \in \mathbf{X}$ for player 1 is a sequence of randomized strategies for all possible games. A policy for player 2 is defined similarly.

We now describe the type of monotonicity of optimal policies that will occur in the above problems. Let $v : \mathbf{X} \to \mathcal{M}(\mathbf{B})$. Denote $b_x^{sup}(v) :=$ the greatest b in the support of u_x, i.e. the greatest $b \in \mathbf{B}$ that is chosen by u with positive probability in state x. Denote $b_x^{inf}(u) :=$ the smallest b in the support of u_x. We say that a decision rule u_t at time t is **strongly monotone decreasing** if for any $x \in \mathbf{X}$ and y with $y < x$, $b_y^{inf}(u_t) \geq b_x^{sup}(u_t)$. The analogous definitions hold naturally for player 1.

Note that as a direct consequence from the definition of strongly monotone policies we have

Lemma 1.1. *If u is strongly monotone then it randomizes in at most$|\mathbf{A}| - 1$ games. If v is strongly monotone then it randomizes in at most $|\mathbf{B}| - 1$ games.*

Denote val$S(x)$:= the value of the matrix game x, and $S(x, u, v)$:= the expected cost that player 2 pays to player 1 at game x when mixed strategies u and v are used.

We now present the main structural results for the saddle point policies for the games $x \in \mathbf{X}$.

Theorem 1.5. *Choose any saddle point policies u and v for player 1 and 2, respectively.*
(i) Assume J1. Then u is strongly monotone.
(ii) Assume J2. Then v is strongly monotone.

Proof. Assume J1. Let v be any saddle point policy for player 2. Then for any $b^2 > b^1$, and any policy u, we have by J1

$$\Delta(u,x) := S(x,u_x,b^2) - S(x,u_x,b^1) - [S(x-1,u_{x-1},b^2) - S(x-1,u_{x-1},b^1)] > 0.$$

We show that this implies that $b_x^{sup}(v) \le b_{x-1}^{inf}(v)$. Suppose $\Delta(x,u)$ is positive for all u, but the latter does not hold, i.e. $b_x^{sup}(v) > b_{x-1}^{inf}(v)$. Set $b^1 = b_{x-1}^{inf}(v)$ and $b_2 = b_x^{sup}(v)$. Let u be any saddle point policy for player 1. It then immediately follows (for more details see e.g. Lemma 3.1 (i) in [3]) that

$$\mathrm{val} S(x) = S(x,u_x,b_2) \le S(x,u_x,b_1).$$

Hence since $\Delta(x,u)$ is strictly positive, we have

$$S(x-1,u_{x-1},b_2) < S(x-1,u_{x-1},b_1) = \mathrm{val} S(x-1)$$

(the last inequality is again straight forward and more details can be found in Lemma 3.1 (i) in [3]). But this contradicts the definition of the value of the game $S(x-1)$. This establishes (1) by contradiction. The proof of (ii) is symmetric. ∎

Corollary 1.1. *(i) Assume that J1 holds. Assume that $\mathbf{A} = \{a^1, a^2\}$ where $a^1 < a^2$. Any saddle point policy u^* for player 1 satisfies the following. There exists $m_u \in \mathbf{X}$ such that*

$$u_x^* = \begin{cases} (1,0) & \text{if } x > m_u \\ (q_u, \bar{q}_u) & \text{if } x = m_u \\ (0,1) & \text{if } x < m_u \end{cases} \tag{1.11}$$

where $q_u \in [0,1]$ is some constant. Moreover, any saddle point v_x^* for player 2 needs no randomizations in any game except for (perhaps) $x = m_u$.
(ii) Assume that $\mathbf{B} = \{b^1, b^2\}$ where $b^2 > b^1$. Any saddle point policy v^* for player 2 satisfies the following. Then there exists $m_v \in \mathbf{X}$ such that

$$v_x^* = \begin{cases} (1,0) & \text{if } x > m_v \\ (q_v, \bar{q}_v) & \text{if } x = m_v \\ (0,1) & \text{if } x < m_v \end{cases} \tag{1.12}$$

where $q_v \in [0,1]$ is some constant.u Moreover, u_x^* needs no randomizations in any game except for (perhaps) $x = m_v$.

Remark 1.1. *It follows from Corollary 1.1 that if $\mathbf{A} = \{a^1, a^2\}$ and $\mathbf{B} = \{b^1, b^2\}$ and if $m_u \neq m_v$ then no randomizations needed at any state by both u^* and v^*. If $m_u = m_v$ then randomization may be needed only at $x = m_u = m_v$.*

1.7.2 Stochastic game between the control of flow and the control of service

We now briefly present the stochastic game in [3, 4] consisting in a queue with a single input flow controlled by one player, whose adversary controls the service. We shall emphasize here the type of results and the solution approach. In the next subsection we shall then present in more details a model with several input flows and a controlled server.

Considered is a discrete-time single-server queue with a buffer of size $L \leq \infty$. We shall use below x to denote the state, which refers to the number of customers at the queue. We assume that at most one customer may join the system in a time slot. This possible arrival is assumed to occur immediately after the beginning of the time slot. The state corresponds to the number of customers in the queue at the beginning of a time slot. Let a_{min} and a_{max} be two real numbers satisfying $0 < a_{min} \leq a_{max} < 1$. At the end of the slot, if the queue is non-empty and if the action of the server is a, then a service of a customer is successfully completed with probability $a \in \mathbf{A}$ where $\mathbf{A}$ is a finite subset of $[a_{min}, a_{max}]$. If the service fails the customer remains in the queue, and if it succeeds then the customer leaves the system.

Let b_{min}, b_{max} be two real numbers satisfying $0 \leq b_{min} \leq b_{max} < 1$. At the beginning of each time slot, if the state is x then the flow controller chooses an action b from a finite set $\mathbf{B}(x) \subset [b_{min}, b_{max}]$. In this case, the probability of having one arrival during this time slot is equal to b. If the buffer is finite ($L < \infty$) we assume that $0 \in \mathbf{B}(x)$ for all x; moreover, when the buffer is full, no arrivals are possible ($\mathbf{B}(L) = \{0\}$). In all states other than L we assume that the available actions are the same, and we denote them by $\mathbf{B}(x) = \mathbf{B}$.

We assume that a customer that enters an empty system may leave the system (with probability a, when action a is used) at the end of this same time slot. The transition law P is:

$$P_{xaby} := \begin{cases} \bar{b}a, & \text{if } L \geq x \geq 1,\ y = x-1; \\ ba + \bar{b}\bar{a}, & \text{if } L \geq x \geq 1,\ y = x; \\ b\bar{a}, & \text{if } L > x \geq 0,\ y = x+1; \\ 1 - b\bar{a}, & \text{if } y = x = 0; \end{cases}$$

(for any number $\chi \in [0,1]$, $\bar{\chi} := 1 - \chi$). We define an immediate payoff

$$c(x, a, b) := h(x) + \theta(a) + \zeta(b) \tag{1.13}$$

for all $x \in \mathbf{X}$, $a \in \mathbf{A}$ and $b \in \mathbf{B}$. We assume that $h(x)$ is a real-valued *increasing convex* function on $\mathbf{X}$ which is polynomially bounded, θ is a real function on $\mathbf{A}$ and ζ is a real function on $\mathbf{B}$. It is natural (although not necessary for the structural results) to assume that θ is increasing in a and $\theta \geq 0$ whereas ζ is decreasing in b and $\zeta \leq 0$. h can be interpreted as a function which gives the holding cost rates, ζ as a reward function related to the acceptance of incoming customer, and θ as a cost function per quality of service.

We thus consider a flow control problem, where the real controller is the one that chooses actions b, and it is playing in an unknown environment of service conditions, so that the first player, which we call "nature" represents an imaginary service controller.

Let U (V) be the class of profiles of player 1 (player 2 resp.). A profiles $u \in U$ ($v \in V$) is a sequence $u = (u_0, u_1, ...)$ ($v = (v_0, v_1, ...)$ resp.) where u_n (resp. v_n) is a conditional probability on $\mathbf{A}$ (resp. $\mathbf{B}$) given the history of all states and actions of both players as well as the current state. Thus each player is assumed to have the information of all last actions of both players as well as the current and past states of the system. Both players know the action sets, the immediate cost c, the initial state and the transition probabilities P. A special class of profiles are the stationary policies, which are those we defined in the previous subsection: the choice of a mixed action is then only a function of the current state.

Let u be a policy of player 1 and v a policy of player 2. Let ξ be a fixed number in $[0, 1)$. Define the discounted cost:

$$V_\xi(x, u, v) := E^{u,v}\left[\sum_{t=0}^{\infty} \xi^t\, c(X_t, A_t, B_t) \mid X_0 = x\right], \tag{1.14}$$

where X_t, A_t and B_t are the stochastic processes describing the state and actions at time t. Define the following problem ($\mathbf{Q}_\xi$): Find u, v that achieve

$$V_\xi(x) := \sup_{u\in U} \inf_{v \in V} V_\xi(x, u, v), \qquad \forall x \in \mathbf{X}. \tag{1.15}$$

It is known [3, 4] that there exist stationary saddle point policies (u^*, v^*) for the players, i.e. stationary policies that satisfy

$$\begin{aligned} \sup_{u\in U} \inf_{v\in V} V_\xi(x, u, v) &= \inf_{v\in V} \sup_{u\in U} V_\xi(x, u, v) = \sup_{u \in U} V_\xi(x, u, v^*) \\ &= \inf_{v\in V} V_\xi(x, u^*, v) = V_\xi(x, u^*, v^*). \end{aligned}$$

$V_\xi(x)$ is called the ξ-discounted *value* of the stochastic game. This value is further known to be finite. (The existence of stationary saddle point policies and the finiteness of the value also holds in the expected average cost, yet for case of infinite buffer, some stability conditions have to be imposed (see [3, 4].)

In the discounted stochastic game, the value function is known to satisfy the dynamic programming equation (due to Shapely [34]), which says that for every state x, the value function $V_\xi(x)$ evaluated at that state equals to the value of some *matrix game* $S(x,a,b)$ (whose entries depend on x, on the discount factor, the immediate costs, the transition probabilities and on the value function at some other states). Moreover, the stationary policies u and v for which u_x and v_x are saddle points of the game $S(x)$ are optimal policies for the original stochastic games. This representation also holds for the case of infinite state space (see [4, 9]). It is shown in [3] using dynamic programming arguments that the set $S(x)$ of matrix games satisfies properties J1 and J2 of the previous subsection. This allows us to conclude that any stationary policies obtained from the saddle point of the matrix games $S(x)$ are strongly monotone, as described in the previous subsection, and are saddle point strategies for the stochastic game. This structure also carries on to the expected average cost criterion (see [9]). More details will be given in the next subsection for a more general problem.

1.7.3 *The case of several input flows*

We now extend our previous model to allow in addition to the controlled flow and the controlled service, additional non-controlled flows that join the same queue. They can be viewed as higher priority flows since they are not subject to flow control and packets of these flows cannot be rejected. The probability that the total number of packets brought by all the non-controlled flows together equals i is given by p_i, $i \geq 0$. We assume that $p_0 > 0$ (this is a necessary condition for the queue to be stable). All other parameters are the same as in the previous subsection. We shall only consider here the case of infinite buffer.

Define for $x \in \mathbf{X}$, $f : \mathbf{X} \cup \{-1\} \to R$,

$$\begin{aligned} S(x,a,b,f) &:= c(x,a,b) + \xi\Big(\bar{a}bf(x+1) + (ab+\bar{a}\bar{b})f(x) + a\bar{b}f(x-1)\Big), \\ g(f,x) &:= \sum_{i=0}^{\infty} p_i f(x+i), \qquad x \in \mathbf{X} \cup \{-1\}. \end{aligned}$$

The dynamic programming equation for this problem for the discounted cost is given by

$$V_\xi(x) = \mathrm{val}_{a,b} S(x,a,b,g(V_\xi)), \quad x \in \mathbf{X}$$

where we define (with some abuse of notation) $V_\xi(-1) = V_\xi(0)$.

We shall say that $f : \mathbf{X} \to R$ satisfies assumption:

WC (weakly convex) if for all $x \in \mathbf{X}$,

$$f(x+2) - f(x+1) \geq f(x+1) - f(x). \tag{1.16}$$

SC(x) (strongly convex) if for x given,

$$f(x+2) - f(x+1) > f(x+1) - f(x). \qquad (1.17)$$

MI if $f(x)$ is monotone increasing in x, i.e. for any $x \in \mathbf{X}$,

$$f(x+1) \geq f(x) \qquad (1.18)$$

We further use the above properties (with some abuse of notation) for the class of policies $f : \mathbf{X} \cup \{-1\} \to R$ that satisfy $f(-1) = f(0)$.

Let U^* be the set of stationary policies for the service controller such that $u \in U^*$ if and only if for any $x \in \mathbf{X}$, u_x is optimal for player 1 in the matrix game $S(x, V_\xi)$. Let V^* be the set of stationary policies for the flow controller such that $v \in V^*$ if and only if for any $x \in \mathbf{X}$, v_x is optimal for player 2 in the matrix game $S(x, V_\xi)$. Any pair (u, v) such that $u \in U^*$ and $v \in V^*$ is optimal for problem $\mathbf{Q}_\xi$ (see e.g. [9]).

We now present the main results.

Theorem 1.6. *Consider the discounted cost. If the holding cost h satisfies* **MI**, **WC** *and either $h(1) > h(0)$ or* **SC(0)** *then*
(i) The value function V_ξ is strictly convex and increasing,
(ii) V_ξ satisfies assumptions J1 and J2,
(iii) Any of the optimal policies $u \in U^*$ and $v \in V^*$ are strongly monotone.

Proof. We proceed using value iteration to show that value function V_ξ is weakly convex and increasing. We note that the function which is zero for all x satisfies these properties. Assume that a function $f : \mathbf{X} \to R$ satisfies these properties and so it also satisfies them when extending the range of f to $\mathbf{X} \cup \{-1\}$ and setting $f(-1) := f(0)$. It clearly follows that $g(f)$ also satisfies these properties. One can then use exactly the same proof as in Lemma 4.3 in [3] (replacing f by $g(f)$) to show that

$$T_\xi(f, x) := \mathrm{val}(S(x, g(f))$$

satisfies these properties. It then follows by induction that for all integers n, $T_\xi^n(0, x)$ satisfies the properties. These properties pass to the limit as $n \to \infty$, and this limit equals the value of the game (see e.g. [9]): $V_\xi(x) = \lim_{n\to\infty} T_\xi^n(0, x)$. Hence V_ξ is increasing and weakly convex. We can now proceed as in the proof of Lemma 4.3 in [3] to conclude that V_ξ is in fact strictly convex so (i) is established.

(ii) The proof of this part follows as in the proof of Theorem 4.1 in [3].

(iii) follows from (ii) by applying Theorem 1.5. ∎

1.7.4 Discussion

We obtained structural results for a zero-sum stochastic game using a stronger version of S-modularity notions, and applying them to the matrix games that appear in the dynamic programming equation. This allowed us to show that both players in the stochastic game have strongly monotone saddle point policies. In the case that each player has only two actions, the saddle point policies are of a threshold type. Such threshold policies can be parameterized by a single real number, as we saw in Section 1.2. One could then ask the question whether S-modularity can be applied directly to these strategies. In other words, if both players use threshold policies, is the corresponding discounted or average cost S-modular? We have not been able to answer that question. First, even when restricting to threshold policies, the corresponding costs do not have a simple expression as in the examples in the previous section, so one cannot check easily the S-modularity. In fact, the power of the approach we use in this section is that we do not need to compute the cost corresponding to each pair of threshold policies.

An alternative way to show that there exist equilibria within threshold policies in an N-player game would be to show that the optimal response of a player when all other players restrict to some threshold policy, is also a threshold policy, and then use some simple fixed point arguments. We used a similar approach in previous Sections, such as Section 1.2. Yet the setting there was quite different, there were infinitely many players, each taking one action, then staying for some time in the system without taking further decisions, and then disappearing from the system. In this section, in contrast, we consider a finite number of "permanent players" that control continuously the flow or the service. In this context of two permanent players, we have not been able to show that a best response to a threshold policy is also a threshold policy, and whether it is true remains an open problem. This problem made it in particular hard to extend our problem to non zero-sum games (for example, in which there are two players that send flow to the system and their flow is controlled by two non cooperative players). Yet, there are some examples of stochastic non-zero sum games with a finite number of permanent players in which it is possible to follow that approach and show that the best response to threshold policies of other players is also a threshold policy, and that there exist equilibria among threshold policies (see [5]).

References

[1] The ATM Forum Technical Committee, *Traffic Management Specification*, Version 4.0, af-tm-0056, April 1996.

[2] I. Adiri and U. Yechiali, "Optimal priority purchasing and pricing decisions in nonmonopoly and monopoly queues", *Operations Research*, 1974.

[3] E. Altman, "Monotonicity of optimal policies in a zero sum game: a flow control model", *Advances of dynamic games and applications*, **1**, pp. 269-286, 1994.

[4] E. Altman, "Flow control using the theory of zero-sum Markov games", *IEEE Trans. Automatic Control*, **39**, No. 4, pp. 814-818, 1994.

[5] E. Altman, "Non zero-sum stochastic games in admission, service and routing control in queueing systems", *Queueing Systems*, **23**, pp. 259-279, 1996.

[6] E. Altman, T. Boulogne, R. El Azouzi, T. Jiménez and L. Wynter, "A survey on networking games", submitted to Telecommunication Systems, Nov. 2000. Available on http://www-sop.inria.fr/mistral/personnel/Eitan.Altman/ntkgame.html

[7] E. Altman and T. Başar, "Multi-user rate-based flow control", *IEEE Trans. on Communications*, pp. 940-949, 1998.

[8] E. Altman, R. El-Azouzi, and T. Jiménez, "Slotted Aloha as a stochastic game with partial information", In *Proceedings of WiOpt'03*, Sophia-Antipolis, France, 3-5, March 2003.

[9] E. Altman and A. Hordijk, "Zero-sum Markov games and worst-case optimal control of queueing systems", invited paper, *Queueing Systems* **21**, special issue on optimization of queueing systems, Ed. S. Stidham, 415-447, 1995.

[10] E. Altman, T. Jiménez, R. Nunez-Queija and U. Yechiali, Queueing analysis for optimal routing with partial information, manuscript, 2001.

[11] E. Altman and H. Kameda, "Equilibria for Multiclass Routing in Multi-Agent Networks", *40th IEEE Conference on Decision and Control*, Orlando, Florida, U.S.A., Dec. 2001.

[12] E. Altman and N. Shimkin, "Individual equilibrium and learning in processor sharing systems", *Operations Research*, vol. 46, pp. 776–784, 1998.

[13] E. Altman and L. Wynter, "Equilibrium, games, and pricing in transportation and telecommunication networks", to appear in the special issue of "Networks and Spatial Economics, Crossovers between Transportation Planning and Telecommunications", 2003. Available as INRIA research report, RR-4632, Nov. 2002.

[14] T. Başar and G.J. Olsder, *Dynamic Noncooperative Game Theory*, SIAM Classics in Applied Mathematics, 1999.

[15] I. Ben-Shahar and A. Orda and N. Shimkin, "Dynamic service sharing with heterogeneous preferences", *QUESTA*, **35**(1-4), 83–103, 2000.

[16] A. Elcan, "Optimal customer return rate for an M/M/1 queueing system with retrials", *Probability in the Engineering and Informational Sciences*, **8**, 521-539, 1994.

[17] A. Glazer and R. Hassin, "?/M/1: On the equilibrium distribution of customer arrivals", European J. of OR, 13 146-150, 1983.

[18] P. Gupta and P. R. Kumar, "A system and traffic dependent adaptive routing algorithm for Ad Hoc networks", In *Proceedings of the 36th IEEE Conference on Decision and Control*, 2375–2380, San Diego, USA, Dec. 1997.

[19] R. Hassin, "On the advantage of being the first server", *Management Science*, Vol. 42, 618–623, 1996.

[20] R. Hassin and M. Haviv, "Equilibrium threshold strategies: the case of queues with priorities", *Operations Research*, **45**, 966-973, 1997.

[21] R. Hassin and M. Haviv, "On optimal and equilibrium retrial rates in a busy system", *Prob. in the Eng. and Informational Sciences*, Vol. 10, 223-227, 1996.

[22] R. Hassin and M. Haviv, *To queue or not to queue: equilibrium behavior in queueing systems*, International series in Operations Research and Management Science, Volume 59, Kluwer Academic Publishers, Boston, 2002.

[23] A. Haurie and P. Marcotte, "On the relationship between Nash-Cournot and Wardrop equilibria", *Networks*, 15:295–308, 1985.

[24] Y. Jin and G. Kesidis, "Equilibria of a noncooperative game for heterogeneous users of an ALOHA network", *IEEE Comm. Letters* Vol. 6, No. 7, pp. 282-284, 2002.

[25] H. Kameda and Y. Zhang, "Uniqueness of the solution for optimal static routing in open BCMP queueing networks" *Mathematical and Computer Modeling*, 22:119–130, 1995.

[26] Y.A. Korilis and A. Lazar, "On the existence of equilibria in noncooperative optimal flow control", *Journal of the ACM*, 42 No. 3:584–613, 1995.

[27] V. G. Kulkarni, "On queueing systems with retrials", *Journal of Applied Probability*, **20**, pp. 380-389, 1983.

[28] A. B. MacKenzie and S. B. Wicker, "Selfish users in Aloha: A game theoretic approach", *Proceedings of the Fall 2001 IEEE Vehicular Technology Conference*, 2001.

[29] A. Orda, R. Rom and N. Shimkin, Competitive routing in multi-user communication networks, *IEEE/ACM Transactions on Networking*, **1**(5):510-520, 1993.

[30] M. Patriksson, *The Traffic Assignment Problem: Models and Methods*, VSP BV, P.O. Box 346, 3700 AH Zeist, The Netherlands, 1994.

[31] A. C. Pigou, *The Economics of Welfare*, McMillan & Co., London, 1920.

[32] R. W. Rosenthal, "A class of games possessing pure strategy Nash equilibria", *Int. J. Game Theory*, 2:65–67, 1973.

[33] W. H. Sandholm, "Potential games with continuous player sets", *Journal of Economic Theory* Vol. 97, 81-108, 2001.

[34] L. S. Shapely,"Stochastic games", *Proceeding of the National Academy of Sciences USA* **39**, pp. 1095-1100, 1953.

[35] D. Topkis, "Equilibrium points in nonzero-sum n-person submodular games", *SIAM J. Control and Optimization*, 17:773–787, Nov. 1979.

[36] J. G. Wardrop. "Some theoretical aspects of road traffic research communication networks", *Proc. Inst. Civ. Eng.*, Part 2, 1:325–378, 1952.

[37] B. W. Wie. "A differential game approach to the dynamic mixed behavior traffic network equilibrium problem", *European Journal of Operational Research*, 83:117–136, 1995.

[38] D. D. Yao, "S-Modular Games, with Queuing Applications", *Queuing Systems: Theory and Applications 21*, pp.449-475, 1995.