Applications of Markov Decision Processes in Communication Networks : a Survey

Eitan Altman

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Applications of Markov Decision Processes in Communication Networks: a Survey

Eitan Altman

N° 3984

Août 2000

THÈME 1

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--3984--FR+ENG

Applications of Markov Decision Processes in Communication Networks: a Survey

Eitan Altman

Thème 1 — Réseaux et systèmes
Projet Mistral

Rapport de recherche n° 3984 — Août 2000 — 51 pages

Abstract: We present in this research report a survey on applications of MDPs to communication networks. We survey both the different applications areas in communication networks as well as the theoretical tools that have been developed to model and to solve the resulting control problems.

Key-words: MDPs, Communication networks

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, B.P. 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : 04 92 38 77 77 - International : +33 4 92 38 77 77 — Fax : 04 92 38 77 65 - International : +33 4 92 38 77 65

Applications de processus de décision markoviens aux réseaux de communication

Résumé : Dans ce rapport de recherche, nous présentons un état de l'art des applications de processus de contrôle markovien aux problèmes dans les réseaux de télécommunications. Nous présentons d'abord les différents champs d'applications dans les réseaux, puis les différents outils qui ont été développés pour modéliser et pour résoudre ces problèmes.

Mots-clés : MDPs, Réseaux de communication

Contents

1 Introduction

Various traditional communication networks have long coexisted providing disjoint specific services: telephony, data networks and cable TV. Their operation has involved decision making that can be modelled within the stochastic control framework. Their decisions include the choice of routes (for example, if a direct route is not available then a decision has to be taken which alternative route can be taken) and call admission control; if a direct route is not available, it might be wise at some situations not to admit a call even if some alternative route exists.

In contrast to these traditional networks, dedicated to a single application, today's networks are designed to integrate heterogeneous traffic types (voice, video, data) into one single network. As a result, new challenging control problems arise, such as congestion and flow control and dynamic bandwidth allocation. Moreover, control problems that had already appeared in traditional networks reappear here with a higher complexity. For example, calls corresponding to different applications require typically different amount of network resources (e.g., bandwidth) and different performance bounds (delays, loss probabilities, throughputs). Admission control then becomes much more complex than it was in telephony, in which all calls required the same performance characteristics and the same type of resources (same throughput, bounds on loss rates and on delay variation).

We do not aim at a complete survey of the area, since several other surveys [86, 90, 144, 114, 145, 147, 245, 242, 244, 247] on related issues already exist, see also [215]. Other references that focus on the methodology that allows to use MDPs to communication networks can be found in the following books [4, 51, 108, 233, 263, 275] as well as in the survey paper [187].

We have two goals in this survey. First, we wish to present to researchers who specialise in MDPs a central application area which provides a vast field of challenging problems. We would like to familiarise these researchers with special complex features in control problems that arise in communications: complex information structure, problems with multiobjective and multiagents. A second objective is to familiarise researchers in communications with tools that have been developed for modelling and for solving control problems in networks.

Problems that are described in this survey are MDPs in a general sense: they are described as Markov chains whose transition probabilities are controlled. This control can be done by a single or several controllers, having the same or having different objectives. It is often tools other than the standard dynamic programming that are used to solve these control problems. For completeness, we occasionally mention approaches that have been used for control of communication systems which are not based on MDPs, and then relate or compare these to MDPs.

2 The Control Theoretical Framework

The most popular telecommunication network architectures today are the Internet and the ATM (Asynchronous Transfer Mode) networks. The Internet offers today a "best effort" type service, i.e., the resources in the network are shared among all users, and when the number of users increases, the quality of service (in terms of delay, throughput, losses) per user decreases. In contrast, ATM networks provide mostly guaranteed services: if a session needs a given Quality Of Services (QOS) it may establish a contract with the network that guarantees that all along the duration of the session, some required bounds would hold on given performance measures (loss probabilities, delays, delay variation, throughput), as long as the source respects the terms of the contract (in terms of the average and the maximum throughput it sends, as well as some constraints on its bursty behavior). Two guaranteed service classes are defined in ATM: the CBR (Constant Bit Rate) and VBR (Variable Bit Rate) service classes. ATM contains also two best effort type services: the Available Bit Rate (ABR) service and the Unspecified Bit Rate (UBR) service. In the ABR service, the network determines the allowed transmission rate of each source by sending to them periodically appropriate control signals. As long as a source abides to those commands, the network guarantees some given bounds on its loss rates. In the UBR service no guarantees are given on performance measures.

From a control theoretical point of view, an important classification of the control in networks is according to who is controlling and what the objectives are. Three general frameworks are possible:

1. Centralised control, or a single controller. This is the case in problems such as call admission control: a request for a new connection arrives and the network has to decide whether to accept it or not.

2. Team theory [45]: several controllers but a single objective. This is the case when all the control decisions are taken by different elements (or agents) in the network rather than by the users. The common objective(s) might be the efficient use of the network and providing a good service to the network users.

3. Game theory [45]: there is more than one selfish decision makers (players) and each has its own objective. The goal of each player is to maximise its own performance, and the decision of each player has an impact on the performance of other players. This framework models the case of several users who can control their own flow, or the routes of their own traffic.

From a control theoretical point of view, ATM networks can be viewed generally as *team control problems*: there are several agents within the network, typically situated in the routers or in the access points to the network, that take decisions. Their objectives is to guarantee the performances that they wish to provide to the various applications and, if possible, to optimise them. In addition, an important objective is to use efficiently network resources (such as memory, bandwidth). (Often different controllers have different information, as will be discussed in Section 3.) The literature on problems in telecommunications (or related models) that fall into this category is very rich, see [31, 33, 109, 232, 278].

The appropriate theoretical framework to consider the Internet is more involved. It could be considered as in a *game* framework since essential control decisions (such as flow control) are taken by the users which are a-priori non-cooperative. The sources may have different objectives and their dynamic behavior may have an influence on other users. In practice, however, the actual design of controllers of data transfer (TCP/IP [136]) is frequently done (through the software that comes with computers handling Internet connections) in an unselfish way. The controllers can still be changed by the users or by applications to improve their individual performance, but this is seldom done. On the other hand, rate controllers are frequently applied in the Internet to interactive voice and video (by changing the compression rate and thus the quality of service). In these applications there is much more variety of controllers and these are often designed in a selfish way.

Two optimality concepts are used in the non-cooperative games arising in networks. The first is the Nash equilibrium: it arises when the performances are determined by a finite number of controllers and it constitutes of a strategy set for each of these controllers such that no user can benefit from deviating from its own strategy as long as the other controllers stick to their strategy. A second type of equilibrium concept arises when the number of users is infinite (which may model in practice a finite but large number of users) and the influence of a single user on the performances of other users is negligible. This is called an atomic game. In the context of routing problems in networks, the corresponding equilibrium is called a *Wardrop equilibrium* [264]. This concept turns out to be useful also in cases in which both the flow and the routing are controlled [221].

If we consider a very large number of sessions in a network, the routing decisions for a single session will typically have a negligible influence on the performance of other sessions. This can be viewed as an atomic game. If however, the routing decisions are not made by the application that initiated the session but, rather, by the service provider to which the user is subscribed, then the analysis will be in terms of a Nash equilibrium, given that the routing decisions of any service provider can have a nonnegligible impact on the performance of users of other service providers as well. Another example of where the Nash equilibrium is the proper theoretical concept is flow control on the Internet initiated by Web connections. A single request to access a Web page that contains many pictures can result in opening *simultaneously* several connections each of which is used to fetch the data of another pictures. In such a case, a single application can have an important impact on many other applications.

Most work that studied telecommunication networks within the the game theoretic framework did not use a dynamic setting and restricted to determining the size and the routes of average flows. We present some examples that did use stochastic dynamic models (Markov games – which are the extension of MDPs to a game situation of selfish controllers).

Korilis and Lazar established in [163] the existence of an equilibrium in a distributed flow control problem in a network in a setting of Nash equilibrium. They further characterised its solution using coupled linear programs.

The references [42, 34, 50, 74] treat a problem related to the choice of connection between a best-effort type service and a dedicated guaranteed service. Consider a request that arrives for a connection for a data transfer. The user that initiates the connection can either use a guaranteed service and request a fixed dedicated amount of resources (such as bandwidth), or it could use a typically cheaper best-effort service in which the resources are shared between all best-effort ongoing connections. The information available for taking the decision is the number of ongoing best-effort connections. The user wishes to minimise the expected time it takes for the data transfer. Note that this time depends also on the decisions that will be made by future users that will wish to connect: if more users in the future will decide to use the best-effort service too, the expected transfer time will become longer, since the available resources are shared between all ongoing connections. However, the decisions of future arrivals are unknown. This explains the need for the Markov game approach and the equilibrium concept that is used in these references. The existence and uniqueness of an equilibrium is established in [34] and it is further computed in the case of a Poisson arrival process of sessions.

Other examples of Markov games which study models related to telecommunication applications appear in the references [3, 28, 118].

Before ending the discussion on game models, we mention zero-sum Markov games in which there are only two players with opposite objectives. Several models within this special framework have been used for the study of problems in telecommunications with a single *central controller* in which some parameters are unknown. The unknown parameters were assumed to vary in time in a way which is unpredictable to the central controller (called player 1). The goal of player 1 is to guarantee the best performance under the worst possible (unknown) dynamic choice of the unknown parameters. To solve such problems, one models the unknown parameters as if they were chosen by a second player with opposite objective. This gives rise to a zero-sum game. Admission control, routing, flow control and service assignment problems were considered in the references [1, 2, 5, 15, 20, 14].

In all three theoretical frameworks that we discussed above, the objective of a controller may be in fact a vector. The objective is said to be *multicriteria*. In the case of a central controller one may wish to minimise delays, minimise losses, minimise rejection rate of sessions, minimise waste of resources. One could add up the objectives and consider minimise instead a weighted sum of objectives instead. Alternative formulations are

1. Minimising the vector of objective in the Pareto sense; a Pareto solution is a policy such that no strict improvement in the performance of one component can be achieved without strictly decreasing another component.

2. Minimising one of the components of the objective subject to constraints on other components.

We state some example of the multicriteria approach in the modelling of telecommunication problems:

(1) *The maximisation of the throughput of some traffic, subject to constraints on its delays* (this problem has been studied along with its dual problem: the minimisation of expected delay subject to a lower bound constraint on the throughput). A rich research literature in this direction was started up by Lazar [181] and has been pursued and developed together with other researchers; some examples are the references [67, 131, 132, 163, 164, 258]. In all these cases, limit-type optimal policies were obtained (known as window flow control).

Koole [157] and Hordijk and Spieksma [130] considered the problem introduced in [181] as well as other admission control problems within the framework of MDPs, and discovered that for some problems, optimal policies are not of a limit-type (the so called "thinning policies" were shown to be optimal under some conditions).

In [4], the author considers a discrete time model that extends the framework of the above problems and also includes service control. The latter control can model bandwidth assignment or control of quality of service. The flow control has the form of the control of the probability of arrivals at a time slot. The control of service is modelled by choosing the service rate, or more precisely, by assigning the probability of service completion within a time slot. A tradeoff exists between achieving high throughput, on the one hand, and low expected delays on the other. It is further assumed that there are costs on the service rates. The problem is formulated as a constrained MDP, where we wish to minimise the costs related to the delay subject to constraints on the throughputs and on the costs for service.

(2) *Dynamic control of access of different traffic types.* A pioneering work by Nain and Ross considered in [207] the problem where several different traffic types compete for some resource; some weighted sum of average delays of some traffic types is to be minimised, whereas for some other traffic types, a weighted sum of average delays should be bounded by some given limit. This research stimulated further investigations; for example, Altman and Shwartz [35] who considered several constraints and Ross and Chen [43] who analyzed the control of a whole network. The typical optimal policies for these types of models requires some randomisation or some time-sharing between several fixed priority policies.

(3) *Controls of admission and routing in networks.* Feinberg and Reiman have solved in [95] the problem of optimal admission of calls of two types into a multichannel system with finite capacity. They established the optimality of a randomised trunk reservation policy.

Other problems in telecommunications which have been solved by constrained MDPs are reported in [75, 57, 194]. A study of a constrained MDP in a queueing model with a removable server, with possible applications in telecommunications or in production, was done in [94].

3 Information issues and action delays

When attempting to model control protocols in telecommunication networks within the optimal stochastic control framework, we note that many non standard features arise in the information structure.

We recall that a "full information" framework is a setting in which each controller knows at each time instant all previous states as well as the previous actions taken by all controllers, plus the current state. In addition, it is assumed that all controllers know the initial state (or the initial distribution over the initial state). We list below some complex aspects of the information available to the decision maker.

3.1 MDPs with incomplete information and MDPs with partial information

The decision maker does not have full information about the network state, but only some observations of the history. For example, the flow control in ABR services in ATM is performed by routers within the networks which have information only on their own congestion state (the number of cells queued at the router) and may have some rough information on congestion experienced by connections that use this router; but they have no information on that of other connections. There are various models that handle MDPs without full state information. By "MDP with incomplete information" we refer to a a Baysian-type framework in which the decision maker(s) can take actions according to the observed history, and knows the initial state, or an initial distribution over the state.

In contrast, in the framework called "MDPs with partial information", the initial distribution or the initial state need not be available (see [128, 129, 169, 190].) To illustrate this framework, we mention the problem of two service stations in tandem. Customers of two different types arrive at a single server facility. The control decisions of the server are which type to of customer to serve at each time. Customers that are have arrived and are have not yet been served are queued and wait their turn. Once a customer is served in the first station it is routed to a second service station that has the same structure, in which a second server has to decide which customer to serve at each time. There again queueing may occur. The decision of each server may depend only on the number of customers of each type queued in that centre. This is a very realistic assumption in practice. This problem is solved in [129].

Another example of MDPs with partial information is window-based flow control in networks. We consider a network with several sources and destinations. Each source of packets has to take decisions concerning the transmission of new packets. When a packet reaches the destination the source receives an acknowledgement. The only information that a source has on the state of the network is through these acknowledgements, from which it can infer how many packets have not yet reached the destination (or more precisely, for how many packets acknowledgements have not yet come back). Thus each controller has some different local information. This problem has been fully solved in [131, 132] by exploiting the so called *Norton Equivalence*, which allows each source to consider the rest of the network as an equivalent single queue (see also [258]).

Below we review information structures that are, on one hand, suitable for modelling telecommunication systems and on the other hand, lead to solutions to the stochastic control problem (within a reasonable complexity).

3.2 Quantised information

This is a special case of incomplete information; due to the discrete nature of data networks, information that are originally represented by real numbers have to be transformed into numerical date using a finite number of bits. This causes loss of precision in the information. For example, congestion information in networks is often transmitted only by one or two information bits, so that the transmitted information takes a small number of possible values. This is the case in the TCP flow control on the Internet, where the way to infer that congestion exists is by a binary information on whether a packet is lost or not.

3.3 Delayed information

The information available to decision makers on the state often suffers from delays. A one way propagation delay can exceed 250ms in a network that contains a Geostationary satellite link. Around 20 ms of propagation delay is incurred in a communication between the west and east coast of the USA. In addition to propagation delays, large random time varying delays are often incurred due to queueing. These components depend on the congestion state of the network and are sometime unknown to the controllers. In a network with several controllers, the delays further vary from one controller to another. The delays could be neglected in analysis of networks in which throughputs are small and transmission delays large; in such cases the time scales related to events in the networks are larger than those involving the delays. Today high speed networks achieve very large throughputs, and the time between the transmission of two consecutive information bits can be smaller than 10^{-9}sec (when considering Gbit/s networks). Hence information delays become a crucial practical (and theoretical) problem.

We briefly review work on control problems with delayed information in telecommunications. Flow control with delayed information has been studied in [25, 36, 171] by transforming the problem into an equivalent MDP with full information. The first paper has been extended to noisy delayed information in [21]. Two types of flow control have been studied. The first type is a rate-base flow control, in which the rate of transmission of packets is directly controlled. The second type is a window-based flow control, in which the controller adjusts its window dynamically; a window stands for the number of packets that can be sent before acknowledgements to the source arrive from the destination. Work on rate-based flow control with delay in the framework of linear quadratic control (linear dynamics and quadratic cost) has appeared in [27, 29, 33]. The impact of delay on window-based flow control in the framework of Jackson network is analyzed in [67]. Routing with delayed information has been investigated in [40, 170, 253]. Finally, a problem of optimal priority assignment for access to a single channel with delay has been investigated in [19].

3.4 Sampled Information

Sampled information means that the decision maker (either within the network or at the source) does not get the information on the network state continuously, but it get some occasional updates. For example, the information used for flow control in the Internet are the acknowledgement that return from the destination to indicate the receipt of a packet. Acknowledgements return however at discrete times, may be lost, and their rate further depends on the transmission rate of the original transmitted data packets. To our knowledge there has been almost no work on MDPs with sampled information in the context of telecommunication networks. We should mention that modelling the flow control within the linear quadratic control framework (in particular - when considering Gaussian noise or the

H^∞ approach, see [26, 29]) allows one to handle sampled information, using the theory in [44, Sec. 5.3]. But this has not yet been done. In [10] an alternative framework is used to handle sampled information; properties of optimal policies are established whenever the value function is multimodular in the controls.

3.5 Asynchronous Information

This is a special case of sampled information in which several controllers in a network receive information at different times, possibly independent of each other; a controller may not know when another controller receives an information update. For example, routing information for establishing the shortest path in the network are typically gathered in different nodes (routers) in which local routing decisions are taken. Such routers exchange occasionally information to determine shortest paths in an asynchronous way.

3.6 Delayed Sharing Information

This is a special case of information structure that arises in team or game problems, see [133, 172, 173, 227, 260]. The state is given as a product of several local components where each component corresponds to one controller (or player). Time is discrete. All controllers have a one step delayed information about the global state of the system. However, each controller gets immediate information about the local component of the state that corresponds to it.

We present some examples of applications of this type of information structure in telecommunications. In [232], the authors have studied decentralised control in packet switched satellite communication and a decentralised control problem for multiaccess broadcast networks have been studied in [109]. In both examples, each controller has to decide whether to transmit or not, without knowing if packets have arrived in the current time unit to other nodes. If they did, then packets from other nodes could be scheduled for transmission at the same time and collisions could occur.

3.7 No information

When transferring short files, the information about the state of the network may come back to the source after the whole file is transferred. This situation occurs when transmission delays are negligible with respect to information delays. This illustrates the fact that controllers often have to make decisions with no available state information. There have been several approaches in solving MDPs with no state information. Within the partial information framework, the algorithm of Hordijk and Loeve [190] has been used in a problem of routing into parallel queues [128]. An alternative approach has been proved to be useful in solving routing control, admission control and polling [135, 6, 152]. This approach applies to cases in which instead of keeping the whole history of previous actions, only some finite (bounded) number of events are recorded. In all above papers, the events which have to be recorded when considering routing to several queues are simply how long ago a packet was routed to each one of the queues. This allows one to transform problems with no information to equivalent MDPs with full information, to establish the optimality of periodic policies, and to obtain other characteristic of optimal policies. It is in particular remarkable that models which are not Markovian (arrival of customers may be general stationary ergodic processes) can be handled by MDPs with finite spaces of states and actions [6]. A third approach consists of using special structure of queueing systems (queueing systems are often used to model telecommunication networks) which leads to the optimality of policies that are regular in some sense. In particular, whenever the value function can be shown to be multimodular in the controls, a rich theory exists for obtaining optimal policies with no state information. We refer to [113, 11] for the definition and properties of multimodular functions, and refer to [113, 6, 13, 11, 8, 9, 12] for applications in routing, admission control and polling with no state information. Weaker notions of regularity of policies are presented in [13, 7, 69, 140, 122, 135]. In particular, the Golden-ratio approach is used in [122, 135] for an optimal channel assignment problem with no state information to obtain simple policies which are close to optimal. This approach is adapted to a context of optimal scheduling of search engines on the Web in [140] and to optimal polling in [69].

3.8 Nested information

In the case of several controllers, say N, we say that information is nested if we can order the controllers in a way that information available to a controller i is a subset of the information available to controller $i+1$ for all $1 \leq i < N$. This structure is again a special case of the incomplete information setting. Note that the case where several controllers receive all information after some fixed delay (that may depend on the controller) is a special case of the nested information structure. The controllers are then ordered with decreasing delays. An example of a flow control problem that gives rise to this information structure appears in [33].

3.9 On the tractability of complex information structure

There is one appealing feature in the nested information as well as in the one-delayed sharing pattern. In both cases it is possible to transform the problem into an MDP with full information with somewhat larger state and action spaces. The way to transform an MDP with incomplete information to an equivalent one with full information can be found in [121, 133, 151, 260] and references therein. For the case of several controllers, we cite a general condition that allows one to transform an incomplete information MDP with several controllers into an equivalent MDP that has a tractable solution the following [45, p. 369]: "An agent's information at a particular stage n can depend on the control of some other agent at some stage $k < n$ only if he also has access to his information available to that agent at that stage k". This condition includes the one-delayed sharing pattern as well as the nested information. It is called in [45] a "classical information pattern".

We note that some complex information structures seem natural for modelling telecommunication systems but have not been actually used in the literature since their solution is either hard, or unknown, or requires policies which are complex to realize (require many computation and/or memory). An example is team or game problems with no state information, with full information on the actions of the agents, but in which the different agents have different knowledge on the initial states. (This problem is related to the one in [98].)

An alternative conservative approach for handling various information structures is to consider noisy information, where the noise is allowed to be a general, possibly state dependent disturbance. A worst case design can then be used, see [26, 29]. This approach may be used for example to handle the case of quantised information.

3.10 Action delays

Another important implication of the delays in high speed networks is the so called *action delay*: even in absence of information delay, a large delay may elapse between the moment that a decision is taken by a controller till this decision has an impact on the network. For example, in ABR service of ATM networks, routers issue commands on flow control that are forwarded to the transmitting sources through special information cells. A router has immediate information on the local state of the queues in that router. Based on this information it sends commands to the sources. But by the time the sources react to these commands, and by the time the reaction influences the queues at the router, a large *action delay* elapses.

It has been shown in [33] that MDPs with both information and action delays can be transformed into a equivalent MDPs with only information delay.

In the following sections we summarise in this section some central control issues that arise in telecommunications.

4 Call admission control

The decision to accept another call to the system may influence both the performance of that call as well as that of ongoing call in the system. This decision changes the state of the system and thus has also an impact on whether future calls will be accepted. Call admission control (CAC) is not applied today on the Internet, but is implemented in ATM networks. Whereas many protocols and control policies have been standardised in the ATM, the implementation of CAC in ATM network is left to the constructor and will probably not be standardised. We describe below several type of admission control problems that arise in telecommunication. Admission control is often combined with routing; we discuss this in the subsection on routing. Other issues related to admission control in wireless communication will be discussed in Section 10.

4.1 Admission control for CBR sources

If the network guarantees a fixed bandwidth per accepted call (e.g., in telephony or in CBR connections in ATM), then the performances of accepted calls is no more influenced by future decisions. In particular, in an ATM environment, delays and throughputs are guaranteed along the entire duration of the call once it is admitted. The main performance measure of interest in that context is then the average rejection rate of a session. Dynamic decisions are required in that framework if there are different classes of calls, each with its own requirement of network resources. This type of admission control problems is frequently modelled at a session level: the state is taken to be the number of sessions of each class, and the MDP is formulated by using the arrival rate of sessions, there average bandwidth requirement and the probability distribution of the duration of a session.

In the case of a single node, a given bandwidth has to be shared between sessions. This problem is known as the *stochastic knapsack problem*, see [16, 185, 220, 230] and references therein. In this setting, there may be cases in which if we accept a call that requires a small amount of bandwidth then there may not be sufficient room for accepting a large call. This type of dilemma explains the fact that optimal acceptance policies are quite complex and are often non-monotone, see [16, 220].

In special cases, the acceptance policies turn out to be monotone and the acceptance region is convex [95, 185, 220, 230]. In the case of two classes with the same bandwidth requirements per class, the optimal policy is known as a trunk reservation policy, which means that we reserve some bandwidth for a high priority class and calls of the other classes are rejected when the remaining bandwidth for the priority class goes below that level. We mention here that in [95] the problem is posed as a constrained MDP and the trunk reservation policy requires randomisation.

In the case of jobs with different requirements for bandwidth, an elegant solution of the optimal call admission control is presented in [104, 229] (for two classes) when restricting to coordinate convex policies, see also [206, 230]. The solution is based on the observation that for such policies, the steady state probabilities have a product form. A fluid approximation approach is used in [16] when relaxing the above restriction. The solution shows, in fact, that in the appropriate scaling, the limit fluid model has a trunk reservation solution which is a special case of a coordinate convex policy.

In the above discussion it is assumed that if a call is not accepted then it disappears. There are, however, cases in which a requirement for establishing can be put into some finite waiting queue instead of being completely rejected. An MDP is used in [217] to solve this problem and its performance is compared to other methods.

4.2 Admission control of sources with variable throughput

A very rich literature exists on admission control of calls with variable transmission rate (VBR sources). Assume that a single node with a given bandwidth is to be shared between several sources of this type. A conservative approach would be to consider the sum of peak rates of all existing connections and to check whether the overall bandwidth would be exceeded if we further added to that the peak rate of the new connection. If it is then the call is rejected. Instead, one tries to make use of *statistical multiplexing*, i.e., of the fact that rarely all sources transmit at peak rate. The typical question that is posed then is: how many calls can we accept so that the probability that a packet is lost is below some threshold? note that a packet is lost when the sum of the instantaneous transmission rates exceeds the available bandwidth. A popular approach is to characterise sources by a parameter called the *effective bandwidth* (see [88, 143, 188] and references therein) which is, roughly speaking, the amount of bandwidth the source needs so that the average loss rate is below some threshold. It is typically larger than the average transmission rate, but is still smaller than the peak rate. There are methods to estimate or to compute effective bandwidth [88, 143, 188].

The effective bandwidth approach itself is not related to MDPs, but it can be combined with MDPs. Indeed, once we know the effective bandwidth of a source with a variable transmission rate, we can use methods from the two previous sections for admitting calls and consider that the bandwidth they require is the effective bandwidth (instead of the peak rate bandwidth). This allows one to accept more calls and thus use the network more efficiently at a price of some additional loss probability of packets in each connection. If the effective bandwidth of existing calls is unknown to the network, adaptive mechanisms can be used to estimate it and combining that with call admission control. This approach is known as *measurement based admission control*, see [110].

The effective bandwidth is not the only approach that can be used to combine session level considerations with packet level phenomena (i.e., the actual process of transmission of sources). In [204] an MDP is formulated to obtain a Call admission control that takes both into account, and a solution is obtained through linear programming.

4.3 Call admission in an integrated service environment

The control problems that we mentioned in the previous subsections were related to "non-elastic" calls, i.e., calls that have some given bandwidth requirements and cannot adjust their transmission rate to the available bandwidth. On the other hand, there are applications that can adapt their transmission rate to the available bandwidth (for example, data transfer). In this case it may seem that call admission is control is not required; the ATM forum (a standardisation institution for ATM) has decided [84] that ABR sessions will not be subject to CAC, unless they require from the network a guarantee on minimum cell rate. (An ABR with minimum cell rate can adapt its transmission to the available bandwidth: it will use more bandwidth if available, but it still requests a minimum guaranteed bandwidth.)

The following questions then arise:

- How to control acceptance of ABR sessions when they do have a minimum cell rate requirement?
- How to handle call acceptance of CBR or VBR traffic in the presence of ABR traffic?

In [23], the authors study the second problem using diffusion approximations under general distribution of interarrival times. They obtain numerically policies that have a monotone switching curve structure and show that a substantial improvement in the performance of ABR traffic classes can be obtained at the price of a slight increase in the rejection rate of CBR and VBR traffic classes. In [137] the monotone switching curve structure is proved to be optimal in the case of exponential distributions.

In [215, 214, 218], routing and call admission of combined CBR, VBR and ABR is considered. Here ABR traffic is also subject to admission control as it is assumed to have a minimum cell rate requirement.

5 Buffer management and packet admission control

Admission control may occur not only at a session level, but also at a packet level. In absence of packet admission control, packets that arrive to routers queue in dedicated buffers; if buffer overflow then the packets are lost. In both ATM as well as in the Internet environment [168, 97] it has been recognised that performance may be improved if the network rejects arriving packets even before the buffer overflows. This is especially the case when we wish to provide different performances to packets of different priorities. Yet, even in the case of a single priority, if real time interactive applications are used (such as video or voice) then it may be useful to discard packets before buffers overflow in order to avoid large queueing delays.

There has been a rich literature on optimal acceptance of customers into a single queue (or a network of queues) using MDPs, either directly related or with potential application to the above queueing model, starting from the late 1960s and beginning of the 1970s [211, 240, 280]. Later references are [64, 49, 138, 241, 243, 247], see also [234, 235].

In general, the type of results obtained in these papers is the existence of threshold optimal policies that reject packets if and only if the queue size exceeds some level.

Surprising structural results have are obtained in [130] when considering multiobjective control problems. They consider the problem of minimising the expected average delay under some lower bound constraint on the expected throughput, or alternatively, the problem of maximising the expected throughput subject to an upper bound constraint on the expected average delay. Optimal policies are shown to be of threshold type, but in addition to that, there is one state in which randomisation (between acceptance and rejection) is required. Surprisingly, this randomisation is performed not necessarily at the threshold, so the optimal policy is not monotone.

Admission control with *no state information* has been studied in [113] using the concept of multimodularity. This has been extended to a framework of a network in [11, 8]. The case of delayed (and other) information was studied in [10, 36, 171].

Remark 1 In flow control problem the source has to decide at which rate to send packets. The admission control problem can thus be viewed as a special case of a flow control problem, in which only two actions exist: that of transmitting at rate zero (corresponding to rejection) and that of transmitting at maximum rate (corresponding to acceptance). Thus much of the literature on flow control can be of potential use for admission control.

We would finally like to mention [30] in which combined rate-based flow control and admission control are handled, and the reference [3] in which the control of admission and of service (in a non-zero sum stochastic game setting) are considered.

Other issues in buffer management in which MDPs have been used are given in [83, 237]. These concern optimisation of policies for sharing a buffer between several classes of arrival streams.

6 Flow and congestion control

In many applications, the users have to adapt the transmission rate of the packets they transmit to the instantaneous state of the network. There are two main approaches to do so: the rate-based flow control, in which the rate is directly controlled, and the window-based control, in which the window size is to be determined dynamically. More precisely, when a packet reaches the destination the source receives an acknowledgement. The window is defined to be the number of packets that are allowed to be transmitted before being acknowledged. If this number is fixed to, say, one hundred, then the source can first send one hundred packets, and then later send one packet per each acknowledgement that returns from the destination. Depending on some parameters that can be measured (as delays, or losses), the source can dynamically change the window size. The rate-based flow controlled is used in ATM networks [84], whereas the window-based flow control is used in the Internet [136].

Many papers have been devoted to flow control into a single queue. The network has often been modelled as a single bottleneck queue in which congestion occurs, and the rest of the network just adds to additional propagation delay. This has been justified by both experimentation and theoretical analysis [65, 68, 131, 132, 258]. In the last three references, optimal window-based flow control is obtained in the case of several sources and in a whole network. The problem is solved by reducing it to an equivalent single queue models.

A key research issue has been to establish the optimality of flow control policies that are monotone in the state, under different information structure. The monotonicity in the case of a single queue has been studied in [10, 21, 25, 36, 234, 235, 248, 247]. In the case of two actions, these monotone policies become threshold policies. The monotonicity in the case of a whole network is treated in [107, 108, 265] (see also [22, 150] for extensions), based on the submodularity concept [256] (see also [272]). In general, the proof of the monotone structure uses value iteration and it requires that the cost be convexity nondecreasing. The convexity is, however, not necessary in the case of expected average cost, see [248] and [3].

Note that the flow control in a network often has the form of the control of service rate of a queue since by controlling it we determine the flow rate into the downstream nodes. Typical objectives are to minimise expected delays or other monotone functions of the workload or number of queued packets. In the case of finite queues, one often considers the minimisation of losses as an objective. In that context, a tandem queueing system with finite buffer capacities is analyzed in [162], where both queues are fed by exogenous arrivals. The first server can be stopped so as to avoid congestion at the downstream queue. It is shown that the optimal control policy for the minimisation of the total loss rate (in both queues) has a monotone switching curve. The authors then compare the optimal policy to other policies that are simpler to implement.

Monotone structure for flow control has also been obtained in the frameworks of constrained MDPs and in stochastic games. In the setting of constrained MDPs, the expected delay is to be minimised subject to a lower bound on the throughput, or vice versa: the throughput is to be maximised subject to an upper bound on the expected delay [4, 66, 67, 131, 132, 181]. Adaptive implementations of this type of constrained problems can be found in [191, 192]. In the framework of zero-sum stochastic games, the service rate may change in time in a way which is unpredictable to the flow controller [1, 2, 14]. The server is then modelled as an adversary player and the flow controller seeks to guarantee the best performance under the worst case behavior of the server. (The opposite case, in which the flow rate is controlled by an adversary player, is studied in [15].) Finally, a model that combines non-zero sum stochastic games with constraints has been analyzed in [163].

An alternative framework for the study of flow control is the LQ (linear dynamics, quadratic cost) model. A plausible objective is to keep queue lengths around some desired level. This objective is derived from the fact that when queues are large then the risk of overflow increases, which results in undesirable loss of packets. On the other hand, when queues become empty then the output rate of the queue cannot exceed the input rate (which is required to be, on the average, lower than the service rate, in order to avoid instabilities); an empty queue thus results in loss of throughput. By letting the queues' size track in an appropriate way the desirable level, optimal performance (in terms of throughput and losses) can be achieved. When the queue seldom becomes empty, the dynamics of the queue can be well approximated as being linear in the control (the input rate). The well known linear quadratic framework is obtained by setting the immediate cost to be quadratic in the deviation from the desired queue level. Other objectives in terms of (undesirable) transmission rate variation, or in terms of (desirable) good tracking of the available bandwidth can be included in that framework, see [26, 29]. The available bandwidth may be described as an ARMA model, which is also suitable for the description of additional noise in the information available to the controller.

One advantage of this setting is that delayed information is not difficult to handle [27, 29]. Moreover, it allows one to handle the case of several controllers (modelled as a team problem) with different action as well as information delay [31, 32, 33]. Finally, this setting allows one to obtain an explicit solution of the problem: both the optimal policy as well as the optimal value can be explicitly computed.

7 Routing

Routing consists of determining the route of each packet from a given source to one or more destinations. There are networks in which each packet may use a different route (this is frequently the case in the Internet), such networks are called packed-switched networks. In such cases dynamic routing control is performed on a very granular scale. In contrast, in networks based on a circuit-switched architecture (such as traditional telephony or ATM networks), routing decisions are made for each connection, and all packets of the connection use the same path.

Among the mostly used routing algorithms are the Bellman-Ford algorithm, the Dijkstra's Algorithm and the Floyd-Marshall Algorithm [52, Sec. 5.2]. In all these algorithms, the dynamic programming principle plays a central role. The objective of routing algorithms is to find the shortest path between nodes, either in terms of number of hops, or in terms of the link length (delay). The latter may change in time either due to link failures and repairs, or due to changing traffic conditions in the network. Routing algorithms should therefore adapt to such changes. Routing is further decentralised: finding shortest paths involves computations that occur in parallel in various nodes of the network. Finally, it is typically asynchronous, in the sense that updates in different nodes do not occur at the same time. Some examples of such algorithms are the one used in the original 1969 ARPANET [52, p. 327], and the RIP (Routing Information Protocol) [195] (the latter reference can be considered to be the standard of the Internet routing).

The above characteristics of routing algorithms, namely adaptivity to time changes, asynchronicity and decentralisation may cause oscillations and stability problems. Discussions on these problems can be found in [52, Sec. 5.2.5]. Therefore an important research issue on these asynchronous dynamic programming algorithms is the study of their convergence and correctness. An example of such a proof can be found in [52, Sec. 5.2.4].

The general approach of routing along shortest paths is well suited to routing of packets, since the routing decision for one individual packet has a negligible impact on the delay along that path and the computed delay will be approximately the one to be experienced by that packet. Thus from a theoretical point of view, the routing decisions result in solutions that related to the concept of Wardrop equilibrium (Section 2). An important feature in Wardrop equilibrium is that all routes from a source to a destination that are actually used have the same delay. This property is appealing in applications in telecommunications, since it reduces the overhead of resequencing in case different packets take different routes (see discussion in [111]).

Below we further describe research on routing in packet-switched and circuit-switched networks. We mention additional issues that arise in routing in mobile networks in Section 10.

7.1 Routing and admission control in circuit-switched networks

When routing whole sessions rather than individual packets along fixed paths (as is typically the case of telephony or of ATM networks), the routing decisions will have a nonnegligible impact on the delay on that path, and thus on the delay experienced by the session to be routed as well as by future sessions. Other types of dynamic programming formulations have frequently been used in those cases.

In case that all packets of a call have the same route, the Call Admission Control is coupled with the routing problem: the question is whether there exists a route along which we should accept a connection. This is the *call setup problem.* Sometimes there is the possibility of taking alternative long routes in case that a direct route is occupied by other calls, and the call admission controller has to decide whether or not to use the alternative route. To illustrate this, consider three nodes: A, B and C, and assume that between each two nodes there is a direct link with a capacity to handle one hundred calls. If all the capacity between A and B is used, one can still attempt to route a call from A to B through the point C. This path is called an *alternative route*.

In networks that use alternative routing whenever a direct route is not available, it has been observed that bi-equilibria behavior occur: the network spends long time in an uncongested mode and a long time in a congested mode. In the uncongested equilibrium, many direct links are available. In the congested mode, many ongoing calls in the network use alternative routes and the blocking probability of a new call is high. To understand this, note that an alternative route uses more resources than a call established on a direct link, and it can increase congestion in the network: the danger with using alternative routes is that if a large number of connections are actually routed through alternative routes then the amount of resources used is probably high, and the chances that a new call will find a direct route is small. If accepted, it would probably also require an alternative route which will further increase congestion and would further decrease the chances that the next call will be accepted or will find a free direct link.

To avoid the congested mode, trunk reservation is often used: it is a policy that does not allow a connection to use an alternative route on a link if the free capacity on the link is below some threshold. The remaining capacity of the link (the trunk) is reserved for direct calls. This means that new calls can be rejected even if there are available resources to handle them (see e.g., [39, 146, 134]).

In the 1980s, dynamic and adaptive routing has been introduced into telephony, which resulted in substantial improvement in network performance and reliability [87]. A number of methods have been proposed for adaptive routing: methods based on decentralised adaptive schemes [212], centralised time-variable schemes and adaptive methods based on the least loaded path (for more details and references, see [87]). Several papers use MDPs to compute off line state-dependent routing and admission control (see e.g., [87, 166, 167] and references therein). Since the state space is huge, it is impossible to obtain an optimal policy except for small networks. However, policies with good performance are obtained as follows. First, some independence assumption is used which allows us to decompose the network problem into a set of MDPs each related to another link. The optimal policy for the decomposed problem is easy to compute, but it is not optimal for the original problem for which the decomposition assumption does not hold. But based upon this policy, one obtains an improved policy using a one step policy improvement iteration [182, 165, 166, 167, 231] or several steps of policy improvement [87, 223]. The low complexity of this approach allows us to obtain good policies based on real-time measurements, see [87]. Another method that uses a similar decomposition approach as a starting point is given in [39].

In some cases of regular topologies (for example, a fully connected networks), it becomes easier to obtain good policies as the network becomes very large. In [134] the policy that uses the least loaded alternative route with trunk reservation is shown to be asymptotically optimal as the number of nodes grow to infinity. For more references on routing in circuit-switched networks using MDPs, see [119, 115].

Diffusion approximations in routing of calls has been studied in [178].

7.2 Routing and Admission in packet-switched networks

In addition to the question of decentralised asynchronous dynamic routing in networks, other theoretical questions have attracted much attention. In particular, many researchers have considered the case of routing into parallel queues. Two types of structural results have been obtained: the optimality of monotone switching-curve policies, and the optimality of policies that send a packet to the shorter queue, if it has a faster server. These characterisations of optimal policies may seem quite trivial (although it is not at all straight-forward to prove their optimality). Yet, they turn out to hold under quite specific conditions. In fact, several counterexamples have been presented in which this structure does not hold once we deviate from these conditions, see e.g., [273] and [41].

A routing policy into N queues is described by partitioning the space state into disjoint sets S_i, such that in set i it is optimal to route to queue i, $i = 1, ..., N$.

Monotone switching curves In [112, 101, 257, 277], an optimal policy is shown to have the following structure under various assumptions, for routing into two queues: the two sets S_i are separated by a monotone curve; for each given state x_i of queue i there is a threshold L^i such that it is optimal to route packet to queue $j \neq i$ at state (x_i, x_j) if the number of packets x_j at queue j is smaller than $L^i(x_i)$.

This type of results is extended to different information structure and to more than one controller. The delayed information is treated in [10, 40, 170], and other information patterns (including the sampled information and the case of no information) are handled in [10].

Joining the shortest queue with fastest server In [91, 125, 139, 197, 141, 142, 276], an optimal policy is shown to have the following structure under various assumptions when routing to into N queues. If the available information is the number of packets in the queues then a packet should be routed to a queue if it is has the smallest number of customers. If the workload in the queue is known, then the routing is done to the queue with the shortest workload.

This type of results is extended to different information structure and to more than one controller. The delayed information is treated in [170].

An extension to a game setting is given in [3, 14] and references therein.

Diffusion approximations for routing have been used in [99, 100, 144] and references therein.

7.3 Routing with no state information

An active area of research in packet-switched network is the routing to N queues, or N servers, or N networks with no state information.

In the particular case of symmetric queues, the optimality of a round robin policy has been established in [12, 189, 263] (for the case of average costs). The case of finite horizon with no state information, but with a given prior distribution has been studied in [202, 203].

The case of routing to N networks (not necessarily identical) that are linear in the so called "$max+$" algebra, is studied in [12]. Based on the theory of multimodularity policies with some regular properties are shown to be optimal in many cases, under very general assumptions on the service and arrival distributions. The objective is to minimise expected average waiting times or workloads (or convex functions of the latter). This framework includes as special cases the routing into N parallel queues.

In [6, 9, 152] similar structural results have also been obtained in the case of routing to N parallel servers *with no buffering*, with exponential service times, where the throughput is to be maximised. It has been shown in [6] that even for *non Markovian inter-arrival times*, the control problem can be transformed into an equivalent MDP with full information where all but a finite number of states are recurrent. For practical purposes, this means that an MDP with a finite number of states and actions can be used to solve this problem. Moreover, it is shown that there exist optimal deterministic *periodic* policies. In the particular case of two queues, the optimal period is of the form $(1, 2, 2, ..., 2)$, where 2 corresponds to the faster server.

7.4 Routing after queueing

A special routing problem that is somewhat different than the previous one is that of routing after queueing: there are several servers with different speeds. In general it is optimal to send a packet to the faster server if it is not busy; the question is then whether a a packet should be sent to a slow server (or should we wait until the fast server is free). In the case of two servers it has been shown in [184] that there is some threshold on the number of queued packets: an arriving packet should be routed to the slow server if and only if the number of packets in the is below the threshold. The original lengthy proof is based on a policy iteration argument. Since then a simpler sample-path proof has been presented in [262], and an even simpler proof based on dynamic programming has appeared in [159].

Surprisingly, this intuitive structure of the optimal policy does not extend to more than two queues, see [41].

8 Scheduling of service

Optimal service scheduling models many scenarios in telecommunications: access control to a communication channel, dynamic priority assignment between different traffic types and dynamic bandwidth allocation in ATM networks.

8.1 Infinite queues and linear costs: the $c\mu$ rule

One of the mostly studied problem in stochastic optimal scheduling of service is the one in which there are K parallel infinite buffer queues and the average weighted expected sum of queue lengths (or of work load) in the queue is to be minimised. The weight factors are given by some positive constants c_i, $i = 1, ..., K$. The service times in queue i is assumed to be exponentially distributed with parameter μ_i whereas inter-arrival times are generally distributed. It has been known already from the early sixties that the optimal policy has is a fixed priority policy; it is the so called "$c\mu$" rule [85, p. 84-85]: the different queues are ordered according to the decreasing order of the product of the weight c_i times the service rate μ_i, and a queue is served only if those queues with a higher product of $\mu_i c_i$ are empty.

These results have been adapted to other framework, in particular to the discrete time setting, see [48, 47, 76].

Other extensions and generalisations both in the model as well as in the proof techniques can be found in [79, 80, 126, 158, 161, 205, 208, 259, 263] (many related references are presented in [263]). In particular, the case in which packets can be rerouted and change class after they terminate their service is analyzed in [148] see also [56, 268]. We should mention that the first proofs have used the theory of Bandits for which it was known that optimal index (Gittens index) rule exists, see [106, 261, 268, 274, 275].

This problem has received a particular attention in the setting of MDPs with additional constraints. As an example, consider the problem where different interactive and non-interactive traffic compete for the access to a single channel. One may wish to minimise the expected delay of the non-interactive traffic, but yet to impose bounds on the average delays of interactive traffic. The case of a single constraint has been studied in [207], who proposed a randomised mixture at each step between two fixed-priority policies; they show the existence of an optimal policy within this class. The case of several constraints has been solved in [35] who use a finite linear programming to determine an optimal policy within the class of so called "time sharing policies": the controller switches between different fixed priority policies when the queues all empty. The linear programming determines the relative frequency at which each one of the finite number of fixed priority policies should be used. The solutions in both references [35] and [207] use a Lagrangian argument that transforms the control problem into a nonconstrained one, for which the $c\mu$ rule is known to be optimal.

As a side result of the structural results of optimal policies in [35], the authors characterise the achievable set of performance measures attained by *any* policy. When restricting to policies that do not idle when the system is non-empty, this region is shown to be a polytope whose extreme points correspond to fixed priority policies.

Later, a reverse approach can be used to solve the constrained problem: first the achievable region is characterised, see [92, 93], which then allows one to obtain the solution of constrained control problem, see [56, 53, 236]. This approach turned out to be quite powerful in more complex control problems as well, see [54, 55, 213]. The achievable region is often referred to as *conservation laws*, see e.g., [263, p. 258].

The type of results for the constrained control problem has been extended to a network in [43]. Yet one should be careful when considering networks: there is a counterexample that shows that the $c\mu$ rule is not optimal in the second node in a tandem network, see [127].

The $c\mu$ rule has been shown in [20] to be optimal for a problem of scheduling service in parallel queues in a setting of a stochastic zero-sum game (that corresponds to a worst case control) where the arrival are MADP (Markov Decision Arrival Processes), an extension of

MAP (Markov Arrival Processes) that allows the arrival process to be controlled and to depend on the state of the system.

8.2 Other scheduling problems

We briefly mention other scheduling problems. Scheduling with deadline constraints has been studied in [58].

In the case of finite queues, it has been observed in [239] that the control model is equivalent to that of optimal routing. This observation has been used in [6, 13] in which stochastic scheduling of service with no state information is studied. Other references on the case of no information are [122, 135].

A special class of scheduling is polling problems, which we describe in the next section.

9 Polling

The terms polling is used when a single server moves between several queues. Upon arrival to a queue, some customers in the queue are served and then the server moves to another queue. The number to be served may depend on the number of customers in the queue. An important feature of polling systems is that when the server moves from one queue to another then switching times or switching costs are incurred.

Polling systems are used to model LANs – local area networks (these are the networks that interconnect computers and printers within a department or a university) or MANs – metropolitan area networks (these are deployed over an area of up to 100km^2 and serve to interconnect local area networks). Examples of LANs that can be modelled using polling systems are the IEEE 802.4 token bus and the IEEE 802.5 token ring (Chapters 4.5.3–4.5.4 in [52]). Examples of MANs modelled as polling systems are given in [228].

Polling systems can also be used to model the scheduling of transmission of packets in the output of an ATM switch.

Interesting control problems arise in the context of polling; the *schedule problem:* in what order should queues be visited and the *service regime:* how many packets are to be served upon a visit to a queue.

In [71, 72, 281], a polling system with infinite queues is considered. The problem of which queue to visit next (for a fixed given service regime) is formulated as an MDP. The objective is to minimise the expected weighted waiting time accruing to the system per unit time. Another control problem is then formulated: the choice of a permutation of the set of queues $\{1, ..., N\}$. The server then visits these queues according to that order (using a fixed service regime in each queue). Once all the queues are visited (we call this a cycle) a new choice is made for the next cycle. The objective is to minimise the expected cycle time. This problem is completely solved in [71] leading to a simple rule which is optimal for both gated[1] as well as exhaustive service regimes[2]: the visit is according to the increasing order of the values of n_i/λ_i, where n_i is the number of packets at queue i in the beginning of the cycle, and λ_i is the arrival rate to that queue. The optimality of the same criterion under other service disciplines is established in [281].

A similar formulation of an MDP is derived in [73] for polling systems with a single buffer, and some optimality results are obtained under particular assumptions.

Several papers study the optimal dynamic polling under various information structure. The maximisation of the throughputs for polling system with a single buffer is studied in [13] in the case of no state information, whereas the corresponding infinite buffer case (in which the weighted sum of expected workload is minimised) is studied in [12]. Other information patterns are handles in [210, 186].

Other papers that consider the objective of minimisation of expected (weighted) workload or weighted average queue lengths in polling systems are [18, 24, 38, 70, 69, 103, 123, 155, 183, 180, 282] and references therein. In general, the problem of minimising the expected (weighted) workload is a difficult one and standard dynamic programming techniques have been seldom used in the references above, see [123]. Characterisation and structure of optimal policies have been derived using sophisticated sample path approaches [183, 186], techniques based on multimodularity [12], fluid approximations [180] and diffusion approximations [24]. Some papers restrict to a given rich class of policies and consider the optimisation among this class [37, 60, 61, 62, 63].

10 Wireless and satellite communications

In wireless and mobile satellite networks there are other dynamic control problems that are due to the mobility.

10.1 Control of handover and admission

If a call is accepted and established then the mobile terminal transmits and receives information from a local base station that covers the area in which it is located. This area is called a cell. If a mobile moves from one cell to another, then a handover of the communication to the new base station has to take place in order for the call to continue. Usually one puts more importance to maintain an existing call than to accept a new one. Thus when accepting a call at a cell, one should have in mind that this decreases the number of calls that can be handed-over to that cell from neighbouring ones. Several papers have used MDPs to solve the call admission control for mobile networks, see [120, 224].

Another related problem is how to decide whether and when to perform a handover. The decision may take into account the movements of the mobile and the strength of the signal. A solution of this problem using an MDP can be found in [226].

[1] in the gated regime, a packet is served in queue i if and only if it is present there upon the arrival of the service to that queue

[2] the server stays in a queue till it becomes empty

10.2 Routing in wireless networks

More complex routing problems occur in the case where users are mobile. Moreover, the nodes themselves may be mobile, as is the case in satellite communications using Low Earth Orbit satellites [102]), or in so called "Ad-hoc Networks" [111, 222]. In Ad-hoc networks, adaptive decentralised routing is again used, but it should be very sensitive to changes in the topology. In satellite communications one can often use simpler routing schemes since the movement of the routers (the satellites) is predictable. Routing in such networks is an active area of research. To the best of our knowledge, dynamic programming or MDP tools have not yet been applied to routing through moving satellites, whereas in Ad-hoc networks, dynamic programming is still the basic tool for determining shortest paths.

10.3 Scheduling transmission opportunities

The scheduling of transmission opportunities between the mobile terminals is determined by the base station. An important factor that is specific to wireless communication is that there may be disconnectivity problems due to the fact that the communication link is more vulnerable to physical obstacles, fading and noise. Some papers that have addressed this control problem are [24, 78, 153, 252, 254].

Another related problem is that of random access to a common channel in a distributed setting, i.e. in the absence of a base station. In that case, users send their packets independently of each other; simultaneous transmissions by more than a single user result in collisions and in the need of retransmission. This type of random access to a common wireless channel was first used in 1970 to interconnect the islands of Hawaii and is still quite common today in satellite communications. Control is needed to determine the retransmission probabilities so as to avoid a large number of collisions during that phase. Finding an optimal retransmission probability is slotted ALOHA systems was modelled in [96] as a finite state MDP with compact action spaces. It was shown that this MDP is unichained and several properties such as monotonicity and estimates for optimal policies were derived. By using natural initial approximations for an optimal policy, a policy iteration algorithm was implemented. The algorithm computed optimal retransmission probabilities after few iterations.

10.4 Other control problems

Some other control problems in both wireless and satellite networks are mobility tracking [46, 193], and energy control [279]; for the latter problem we do not know of an MDP solution approaches, but we believe that it could be used.

11 MDPs in applications of the World Wide Web

Most of today's traffic on the Internet is World Wide Web traffic, which makes it an important field of application of optimal control techniques. The World Wide Web offers search engines, such as Altavista, Lycos, Infoseek and Yahoo, that serve as a database that allows one to search information on the Web. These search engines often use robots that periodically traverse a part of the Web structure in order to keep the database up-to-date by copying the Web pages from around the world into the database. An important control problem is the efficient design of these search engines. In particular, the question that arises is how often should pages be fetched in order for the information in the database to be updated. It is required to minimise the probability that a request for an information on a page finds that page in the database out-of-date, i.e., that the page has since been modified but the new version has not yet been updated in the database. Fetching pages is a costly operation, and efficient updates (taking into account the frequency that a page is requested) is crucial.

Several papers have studied this problem using MDPs. In [140], the authors consider a problem where there is a fixed number of M Web-pages. The contents of page i is modified at time epochs that follow a Poisson process with parameter μ_i. The time a page is considered up-to-date by the search engine is the time since the last visit by the robot until the next time instant of modification; at this point the Web-page is considered out-of-date until the next time it is visited by the robot. The times between updates by the robot are given by a sequence τ_n, assumed to be i.i.d. A simple policy based on the golden-ratio approach is shown in [122, 135] to perform close to optimum. In [6] the i.i.d. assumption is relaxed and the τ_n sequence is only assumed to be stationary. The problem is then solved using a finite MDP, and the existence of periodic optimal policies is established. The solution, using MDPs of some related problems can be found in http://www.path.berkeley.edu/~guptar/webtp/index.html.

Other applications of MDPs to the control of search engines in the Web can be found in [249, 250].

12 Solution methodologies

We shall survey in this section some solution methodologies that were successful in problems in telecommunications. Some of these are classical and are related directly to general solution methods of dynamic programming equations. Some other solution methods, however, make use of special properties of queueing networks which are useful in modelling problems in telecommunications.

We have already mentioned in the previous sections the use of Bandits together with the Gittens index for scheduling problems, and the use of conservation laws (which are especially used in scheduling as well). We further discussed in the part on flow control the advantage of linear quadratic framework, which could be also used in other contexts (e.g., in dynamic bandwidth allocation). Below we present several other techniques and methodologies which are very useful in telecommunication applications.

12.1 Structural characteristics of optimal policies

Due to the curse of dimensionality of dynamic programming, researchers have been interested in inferring the structure of optimal policies and/or of the value function. In some cases, when one knows the structure of optimal policies, the original optimal control problem can be reduced to that of an optimisation problem over a small subclass of policies that possess the required properties.

A very popular method for obtaining the structure of optimal policies goes through value iteration as follows. Under quite general conditions, if one knows the value function then one can compute the optimal policies as the policy that chooses the argument of the optimisation (maximisation or minimisation) of the corresponding dynamic programming. The question is then how does the argument of the optimisation behave as a function of the state. This of course depends on the properties of the immediate rewards (or costs), the transition probabilities, and the value of the dynamic programming. For a given set of transition probabilities, one can often show that if the immediate rewards and value have some properties (such as convexity or concavity or submodularity) then this implies the required structure of the argument of the optimisation and thus of the optimal policy (see e.g., [149]).

For example, in many one-dimensional queueing systems, if the value function is convex then the policy that minimises the costs is of threshold type [108, 150]. In two dimensional problems, submodularity often implies the optimality of monotone switching curve policies [150]. In routing problems, weak Schur convexity of the value function implies typically the optimality of the policy that routes a customer to the shortest queue [125].

In order to obtain the required properties of the value function, one proceeds by value iteration. One first checks that the terminating cost has this property; in case of infinite horizon (discounted cost) one can choose some arbitrary terminating cost (typically a cost that is everywhere zero) that has the required property (which will not have an influence on the value for the infinite horizon). Then one checks by iterating the dynamic programming operator that the value for the n-step horizon also possesses the required structure for any integer n. For the infinite horizon case, one then has to establish that the property also holds for the limit (as n tends to infinity), which coincides with the value for the infinite horizon (under fairly general conditions).

The above approach can be used not only to establish properties of an optimal policy or the optimal value, but also to compare the values of different policies or the value corresponding to different statistical assumptions. For example, this method is used in [154] to show the advantage of stochastic multiplexing of many small sources, by comparing the performance of one big source to that of several small sources.

We finally note that other techniques can be used to establish structural properties of optimal policies, and in particular sample path methods.

12.2 Sample path methods

Several sample path techniques have been used for solving MDPs in queueing applications, see e.g., [205, 209, 208, 238, 263]. A thorough survey on these methodologies can be found in [187]. The most frequently used are interchange arguments, in which one can show that by interchanging the order in which actions are taken, the policy can be improved. This technique has been used for both routing as well as scheduling problems, see e.g., [263]. This approach is based on coupling: one constructs on the same probability space the evolution of a two stochastic systems that differ only in the control, but not in the driving sequence (such as interarrival times or service times). An opposite approach is, on the contrary, to change the probability space and solve the control problem in the new space instead, which if chosen appropriately is simpler to solve. In some cases, although the probability spaces are different, the costs depend only on some marginal (rather than joint) distributions and is thus the same for the two models. An example is routing to parallel symmetric queues with no state information (for appropriate costs). One constructs a new model for which (potential) service times are the same in all queues, which allows one to establish the optimality of a round robin policy for the new model. it then turns out that this policy is also optimal for the original model [263, p. 264].

12.3 Stability analysis

There is a whole class of MDPs that arise in telecommunications for which the solution goes through stability analysis, and typically through Lyapunov function techniques.

As an example, consider the assignment or polling problem when the connections between each queue and the server is broken at random times and for random durations. The possibility of such interruptions complicates the control problem considerably, since the possibility that any queue might not be available to the server at any future time needs to be accounted for in choosing the current server allocation. This problem is typical for wireless communications in which noise can cause disconnectivity problems. The goal is to maximise the throughput. The solution approach is to first find stability conditions that are *necessarily* for an arbitrary policy, and then show that a the same stability condition is a sufficient condition under some given candidate policy. This approach has been used in [252, 254] where it was shown that the policy that serves the longest connected queue is optimal. Lyapunov functions that are quadratic in the state were used to obtain the stability condition. A related problem in satellite communications was solved with a similar technique in [78]. The same problem as in [252] was later solved under more refined criteria [24] using diffusion approximations.)

In [251, 253] state dependent routing and flow control is considered in a queueing network with arbitrary topology. The routing is based on local state information. In addition, the rate of a server is controlled based on local information (which means that the outflow is controlled). A distributed policy is shown to achieve maximum throughput in the case of delayed state information [253].

Optimal scheduling in another type of network topology is considered in [105], namely ring networks. Again, a scheduling policy with maximal stability region is obtained.

12.4 Fluid limits and diffusion approximations

Telecommunication systems are usually described as discrete event systems, where discrete units of information (bits, cells, packets or sessions) are transmitted at at discrete times. Due to the curse of dimensionality of dynamic programming it is often impossible to solve optimal control problems in networks within this framework. Two less granular approaches are the fluid limits and the diffusion approximations, in which the transmission of discrete information units is approximated by the transmission of a deterministic fluid and of a stochastic fluid, respectively. In both approximations, the fluid represents the expected workload or the number of customers that arrive, that are present, and that leave the system. Using functional laws of large numbers for the fluid approximation and functional central limit theorems for the diffusion approximation, one can typically show that the approximating processes become tight at high loads. The limiting processes allows us to approximate the value and to obtain policies that are almost optimal for the original system. Public domain software exists which is specially adapted to the use of diffusion approximations for telecommunication problems, see http://www.dam.brown.edu/lcds/software.html. Fluid and diffusion approximations have often the advantage of a collapse of the dimension of the state space: in many cases, a high dimensional problem reduces to a lower dimensional one in these approximations, see e.g., [16] for the case of fluid approximations and [23] for the case of diffusion approximations.

Some other papers using fluid limits in control of queues in this line are [16, 39, 82, 81, 89, 180, 200, 198, 199, 201, 269, 270, 271]. A public domain animations of fluid control corresponding to the papers [269, 270, 271] is available on the Web in the home page http://rstat.haifa.ac.il/ gweiss of Weiss. Some references on diffusion models are [116, 117, 144, 174, 175, 176, 177, 178, 179, 219, 225, 266, 267].

In yet another type of fluid models, one consider two time scales. The distribution of the arrival and/or service in a network may very in some stochastic way (whose distribution is possibly controlled) at a time scale which is much slower than the transmission time of packets. In that case one uses some environment state to describe the slow variation of the distributions of the parameters of the network, and then use a fluid approximation only to replace the granular arrivals and service in each given environment by a fluid. In that context, recent work [17] shows that fluid limits are not only approximations that become asymptotically tight (in some appropriate scaling), but also that they give in fact *optimistic bounds* on the performance.

Using this approach, the authors in [30] consider combined admission and flow control. The state of the network depends on the number of sessions of different traffic types, which varies at a time scale much slower than the controlled transmission of packets. Termination of sessions (and thus the decrease in their number) occurs according to an exponential distribution. The arrival of new sessions of different types occur according to a Poisson process, and the admission control can decide to accept or reject an arrival of a session.

Once in the system, the rate of transmission of packets (modelled as fluid) of each session, is controlled. Another example of this approach is the reference [246] where the authors consider a fluid model for the optimal flow control.

12.5 Power series algorithm

The power series algorithm, developed originally in [124] and further in [59, 160] in the context of non-controlled Markov chains, allows one to obtain the performance measures by a recursive numerical method that is based in the expansion of the value function in the load parameter. Recently this method has been extended to MDPs and optimisation of Markov chains in [60, 156] and applied to several problems in optimisation and control of queueing systems, which can be used to model scenarios in telecommunications [62, 63].

12.6 Neuro-dynamic programming

Neuro-dynamic programming (NDP), also known as reinforcement learning, is a recent class of methods that can be used to solve very large and complex dynamic optimisation problems [51]. NDP combines simulation, learning, neural networks or other approximation architectures and heuristics, with the central ideas in dynamic programming. It provides a rigorous framework for addressing challenging and often intractable problems from a broad variety of fields. As such, it is a promising tool for for solving large scale control problems in telecommunications. Applications of this methodology in telecommunications can be found in [51, Sec. 8.5], in [77, 196, 216, 255] and in references therein.

13 Concluding remarks

We have provided a survey of the areas in telecommunications in which MDPs have been applied and have potential application. In addition, we have surveyed some modelling issues (multiagent and information issues) as well as solution techniques (other than the standard dynamic programming) that are special to telecommunication applications of MDPs.

In preparing this survey we have interviewed around forty researchers working on MDPs and on telecommunications. Our general impression from these interviews and from the preparation of the survey are the following.

(i) Communication networks is a very rich area of application that has an impact on MDPs, including the development of theoretical tools that seem adapted to problems encountered in telecommunications.

(ii) Optimal control and applied mathematics are not central in the development of communications and network technology, as opposed perhaps to areas such as aeronautics, robotics. Some people interviewed regretted that persons that venture into today's communication problems do not have a solid background of the fundamentals issues addressed by control theory.

(iii) Dynamic programming techniques have an important impact in some areas, and in particular on routing (Bellman-Ford and other algorithms). Theoretical work using MDPs that were mostly cited in the interviews as having an impact in telecommunications is that of [87, 167] on admission control and routing. Dziong and Mason's work [87] was implemented in the Bell Canada network. Moreover for the work they received the First Stentor award for Industry-University Collaboration (along with A. Girard, J. Regnier and H. Cameron from NORTEL) in Telecoms, which is one of the most prestigious awards in Canada.
(iv) We believe that MDPs has an important potential for applications in telecommunications, in particular in the areas of scheduling in ATM switches, in buffer management schemes (admission of packets of different priorities to buffers) and in flow control.

References

[1] E. Altman. Flow control using the theory of zero-sum Markov games. *IEEE Transactions on Automatic Control*, 39:814–818, 1994.

[2] E. Altman. Monotonicity of optimal policies in a zero sum game: A flow control model. *Advances of Dynamic Games and Applications*, 1:269–286, 1994.

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

[4] E. Altman. *Constrained Markov Decision processes.* Chapman and Hall/CRC, 1999.

[5] E. Altman. A Markov game approach for optimal routing into a queueing network. *Anals of Dynamic Games Vol 5: Stochastic and Differential Games, Theory and Numerical Methods, M. Bardi, T.E.S. Raghavan and T. Parthasarathy (Editors), Birkhauser Boston, Basel, Berlin*, pages 359–376, 1999.

[6] E. Altman, S. Bhulai, B. Gaujal, and A. Hordijk. Optimal routing to M parallel servers with no buffers. *Journal of Applied Probability*, 37(3), 2000.

[7] E. Altman, B. Gaujal, and A. Hordijk. Regularity for admission control comparisons. In *Proceedings of the 37th IEEE Conference on Decision and Control*, Dec. 1998.

[8] E. Altman, B. Gaujal, and A. Hordijk. Admission control in stochastic event graphs. *to appear in JACM*, 2000.

[9] E. Altman, B. Gaujal, and A. Hordijk. Balanced sequences and optimal routing. *to appear in IEEE Trans. Automatic Control*, 2000.

[10] E. Altman, B. Gaujal, and A. Hordijk. Multimodular value functions: monotonicity of feedback control. *in preparation*, 2000.

[11] E. Altman, B. Gaujal, and A. Hordijk. Multimodularity, convexity and optimization properties. *Mathematics of Operations Research*, 25:324–347, 2000.

[12] E. Altman, B. Gaujal, and A. Hordijk. Optimal open-loop control of vacations, polling and service assignment. *to appear in Qeueuing Systems*, 2000.

[13] E. Altman, B. Gaujal, A. Hordijk, and G. Koole. Optimal admission, routing and service assignment control: the case of single buffer queues. In *Proceedings of the 37th IEEE Conference on Decision and Control*, Tampa, Florida, USA, Dec. 1998.

[14] E. Altman and A. Hordijk. Zero-sum Markov games and worst-case optimal control of queueing systems. *Queueing Systems*, 21:415–447, 1995.

[15] E. Altman, A. Hordijk, and F.M. Spieksma. Contraction conditions for average and α-discount optimality in countable state markov games with unbounded rewards. *Mathematics of Operations Research*, 22 No. 3:588–618, 1997.

[16] E. Altman, T. Jimenez, and G. Koole. On optimal call admission control. In *Proceedings of the 37th IEEE Conference on Decision and Control*, pages 569–574, Tampa, Florida, USA, Dec. 1998.

[17] E. Altman, T. Jimenez, and G. Koole. Comparing tandem queueing systems and their fluid limits. *Probability in the Engineering and Informational Sciences*, 2000.

[18] E. Altman, A. Khamisy, and U. Yechiali. On elevator polling with globally gated regime. *Queueing Systems*, 11:85–90, 1992.

[19] E. Altman, D. Kofman, and U. Yechiali. Discrete time queues with delayed information. *Queueing Systems*, 19:361–376, 1995.

[20] E. Altman and G. M. Koole. Stochastic scheduling games and Markov decision arrival processes. *Computers and Mathematics with Applications*, 26(6):141–148, 1993.

[21] E. Altman and G.M. Koole. Control of a random walk with noisy delayed information. *Systems and Control Letters*, 24:207–213, 1995.

[22] E. Altman and G.M. Koole. On submodular value functions and complex dynamic programming. *Stochastic Models*, 14:1051–1072, 1998.

[23] E. Altman and H. Kushner. Admission control for combined guaranteed performance and best effort communications systems under heavy traffic. *SIAM J. Control and Optimization*, 37(6):1780–1807, 1999.

[24] E. Altman and H. Kushner. Control of polling in presence of vacations in heavy traffic with applications to satellite and mobile radio systems. In *Proceedings of the 37rd Allerton Conference on Communication, Control, and Computing*, Illinois, USA, Sept. 1999.

[25] E. Altman and P. Nain. Closed-loop control with delayed information. *Performance Evaluation Review*, 20:193–204, 1992.

[26] E. Altman and T. Başar. Optimal rate control for high speed telecommunication networks. In *Proc. of the 34th IEEE Conference on Decision and Control*, New Orleans, Louisiana, USA, Dec. 1995.

[27] E. Altman and T. Başar. Optimal rate control for high speed telecommunication networks: the case of delayed information. In *First Workshop on ATM Traffic Management, WATM, IFIP, WG.6.2 Broadband Communication, Paris*, pages 115–122, Dec 1995.

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

[29] E. Altman, T. Başar, and N. Hovakimian. Worst-case rate-based flow control with an arma model of the available bandwidth. *Anals of Dynamic Games*, 6, 1999.

[30] E. Altman, T. Başar, and Z. Pan. Piecewise-deterministic differential games and dynamic teams with hybrid controls. *Anals of Dynamic Games*, 6, 1999.

[31] E. Altman, T. Başar, and R. Srikant. Multi-user rate-based flow control with action delays: a team-theoretic approach. In *Proc. of the 36th IEEE Conference on Decision and Control*, San Diego, California, Dec. 1997.

[32] E. Altman, T. Başar, and R. Srikant. Robust rate control for abr sources. In *IEEE INFOCOM, San-Fransisco, California, USA*, 1998.

[33] E. Altman, T. Başar, and R. Srikant. Congestion control as a stochastic control problem with action delays. *Automatica*, 1999.

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

[35] E. Altman and A. Shwartz. Optimal priority assignment: a time sharing approach. *IEEE Transactions on Automatic Control*, AC-34:1089–1102, 1989.

[36] E. Altman and S. Stidham, Jr. Optimality of monotonic policies for two-action Markovian decision processes, including information and action delays. *Queueing Systems*, 12 No. 2:307–328, 1996.

[37] E. Altman and U. Yechiali. Cyclic Bernoulli polling. *ZOR - Mathematical Methods of Operations Research*, 38:55–76, 1993.

[38] E. Altman and U. Yechiali. Polling in a closed network. *Probability in the Engineering and Informational Sciences*, 8:327–343, 1994.

[39] V. Anantharam and M. Benchekroun. Trunk reservation based control of circuit switched networks with dynamic routing. In *Proceedings of the 29th Conference on Decsion and Control*, pages 2102–2105, Honolulu, Hawaii, Dec. 1990.

[40] D. Artiges. Optimal routing into two heterogeneous service stations with delayed information. *IEEE Transactions on Automatic Control*, 40(7):1234–1236, 1995.

[41] D. Artiges. *Controle et evaluation des reseaux de telecommunication (in french)*. PhD thesis, INRIA, Sophia Antipolis, France, 1996.

[42] D. Assaf and M. Haviv. Reneeging from time sharing and random queues. *Mathematics of Operations Research*, 15:129–138, 1990.

[43] K. Ross B. and Chen. Optimal scheduling of interactive and non-interactive traffic in telecommunications systems. *IEEE Transactions on Automatic Control*, 33:261–267, 1988.

[44] T. Başar and P. Bernhard. *H^∞-Optimal Control and Relaxed Minimax Design Problems: A Dynamic Game Approach*. Birkhauser, Boston, MA, USA, 1991 (2nd edition, 1995).

[45] T. Başar and J. B. Cruz. Concepts and methods in multi-person coordination and control. In S. G. Tzafestas, editor, *Optimization and Control of Dynamic Operational Research Methods*, pages 351–394. North-Holland Publishing Company, 1982.

[46] A. Bar-Noy, I. Kessler, and M. Sidi. Mobile users: To update or not to update? *Wireless Networks journal*, 1:175–186, 1995.

[47] J. S. Baras, A. J. Dorsey, and A. M. Makowski. K competing queues with geometric service requirements and linear costs: the μc rule is always optimal. *Systems and Control Letters*, 6:173–180, 1985.

[48] J. S. Baras, A. J. Dorsey, and A. M. Makowski. Two competing queues with linear costs and geometric service requirements : the μc rule is often optimal. *Advances in Applied Probability*, 17:186–209, 1985.

[49] M. Bartroli. On the structure of optimal control policies for networks of queues. Ph.D. dissertation, Department of Operations Research, University of North Carolina at Chapel Hill, 1989.

[50] I. Ben-Shahar, A. Orda, and Nahum Shimkin. Dynamic service sharing with heterogeneous preferences. *submitted to QUESTA*, 1999.

[51] D. P. Bertsekas and J. N. Tsitsiklis. *Neuro-Dynamic Programming*. Athena Scientific, Belmont, MA, 1996.

[52] D.P. Bertsekas and R.G. Gallager. *Data Networks*. Prentice-Hall, 1987.

[53] D. Bertsimas and Jose Nino-Mora. Conservation laws, extended polymatroids and multi-armed bandit problems; a unified polyhedral approach. *Mathematics of Operations Research*, 21:257–306, 1996.

[54] D. Bertsimas and Jose Nino-Mora. Optimization of multiclass queueing networks with changeover times via the achievable region method: Part i, the single-station case. *To appear in Mathematics of Operations Research*, 1999.

[55] D. Bertsimas and Jose Nino-Mora. Optimization of multiclass queueing networks with changeover times via the achievable region method: Part ii, the multi-station case. *To appear in Mathematics of Operations Research*, 1999.

[56] D. Bertsimas, I. Paschalidis, and J. N. Tsitsiklis. Branching bandits and klimov's problem: Achievable region and side constraints. *IEEE Transactions on Automatic Control*, 40:2063–2075, 1995.

[57] F.J. Beutler and K.W. Ross. Time-average optimal constrained semi-markov decision processes. *Advances of Applied Probability*, 18:341–359, 1986.

[58] P.P. Bhattacharya, L. Tassiulas, and A. Ephremides. Optimal scheduling with deadline constraints in tree networks. *IEEE Transactions on Automatic Control*, 42(12):1703–1705, 1997.

[59] J.P.C. Blanc. On a numerical method for calculating state probabilities for queueing systems with more than one waiting line. *Journal of Computational and Applied Mathematics*, 20:119–125, 1987.

[60] J.P.C. Blanc. Performance analysis and optimization with the power-series algorithm. In L. Donatiello and R. Nelson, editors, *Performance Evaluation of Computer and Communication Systems*, pages 53–80. Springer-Verlag, 1993. Lecture Notes in Computer Science 729.

[61] J.P.C. Blanc and R. D. van der Mei. The power series algorithm applied to polling systems with a dormant server. (9346), 1993.

[62] J.P.C. Blanc and R. D. van der Mei. Optimization of polling systems with Bernoulli schedules. *Performance Evaluation*, 22:139–158, 1995.

[63] J.P.C. Blanc and R. D. van der Mei. Computation of derivatives by means of the power-series algorithm. *INFORMS J. Comput.*, 2:45–54, 1996.

[64] J.P.C. Blanc, P.R. de Waal, P. Nain, and D. Towsley. Optimal control of admission to a multiserver queue with two arrival streams. *IEEE Transactions on Automatic Control*, pages 785–797, 1992.

[65] J.-C. Bolot. End-to-end delay and loss behavior in the internet. In *Proceedings of ACM Sigcomm '93, San Francisco, CA, USA*, pages 289–298, Sept. 1993.

[66] A. D. Bovopoulos and A. A. Lazar. Optimal load balancing algorithms for Jacksonian networks with acknowledgement delays. *IEEE Transactions on Communications*, pages 144–151, 1988.

[67] A. D. Bovopoulos and A. A. Lazar. The effect of delayed feedback information on network performance. *Annals of Operations Res.*, pages 581–588, 1991.

[68] O.J. Boxma. Sojourn times in cyclic queues – the influence of the slowest server. In G. Iazeolla, P. J. Courtois, and O. J. Boxma, editors, *Computer Performance and Reliability*, pages 84–88. Elsevier Science Publishers B.V. (North-Holland), 1988.

[69] O.J. Boxma, H. Levy, and J.A. Weststrate. Efficient visit frequencies for polling tables: Minimization of waiting costs. *Queueing Systems*, 9:133–162, 1991.

[70] O.J. Boxma, H. Levy, and U. Yechiali. Cyclic reservation schemes for efficient operation of multiple-queue single-server systems. *Annals of Operations Research*, 31:187–208, 1991.

[71] S. Browne and U. Yechiali. Dynamic priority rules for cyclic-type queues. *Advances in Applied Probability*, 21:432–450, 1989.

[72] S. Browne and U. Yechiali. Dynamic routing in polling systems. In M. Bonati, editor, *Teletraffic Science*, pages 1455–1466. Elsevier Science Pub. (North-Holland), 1989.

[73] S. Browne and U. Yechiali. Dynamic scheduling in single-server multiclass service systems with unit buffers. *Naval Research Logistics*, 38:383–396, 1991.

[74] R. Buche and H. J. Kushner. Stochastic approximation and user adaptation in a competitive resource sharing system. In *Proceedings of the 37th Conference on Decision and Control*, Tampa, Florida, USA, Dec. 1998.

[75] E. B. N. Bui. Contrôle de l'allocation dynamique de trame dans un multiplexeur integrant voix et données (in french). Master Thesis 89 E 005, TELECOM, Département Réseaux, Paris, June 1989.

[76] C. Buyukkoc, P. Varaiya, and J. Walrand. The $c\mu$ rule revisited. *Advances in Applied Probability*, 17:237–238, 1985.

[77] J. Carlstrom and E. Nordstrom. Control of self-similar ATM call traffic by a reinforcement learning. In J. Alspector, R. Goodman, and T. X. Brown, editors, *Proceedings of the International Workshop on Applications of Neural Networks to Telecommunication 3, IWANNT'97*, pages 54–62, Melbourne, Australia, 1997. Lawrence Erlbaum.

[78] M. Carr and B. Hajek. Scheduling with asynchronous service opportunities with applications to multiple sattelite systems. *IEEE Transactions on Automatic Control*, 38(12):1820–1833, 1993.

[79] C.S. Chang, X. Chao, M. Pinedo, and R.R. Weber. On the optimality of LEPT and $c\mu$-rules for machines in parallel. *Journal of Applied Probability*, 29:667–681, 1992.

[80] C.S. Chang and R. Righter. The optimality of LEPT in parallel machine scheduling. Working paper, 1993.

[81] H. Chen and D. Yao. Dynamic scheduling of a multi-class fluid network. *Operations Research*, 41(6):1104–1115, 1993.

[82] R. R. Chen and S. P. Meyn. Value iteration and optimization of multiclass queueing network, invited paper. 32:65–97, 1999.

[83] I. Cidon, L. Georgiadis, R. Guerin, and A. Khamisy. Optimal buffer sharing. In *IEEE INFOCOM*, Boston, MA, USA, Apr. 1995.

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

[85] D.R. Cox and W.L. Smith. *Queues.* John Wiley, New York, 1961.

[86] T. Crabill, D. Gross, and M. Magazine. A classified bibliography of research on optimal design and control of queues. *Operations Research*, 25:219–232, 1977.

[87] Z. Dziong and L. G. Mason. Call admission and routing in multi-service loss networks. *IEEE Transactions on Communications*, 42:2011–2022, 1994.

[88] A. I. Elwalid and D. Mitra. Effective bandwidth of general markovian traffic sources and admission control of high speed networks. *IEEE/ACM Trans. on Networking*, 1:329–343, 1993.

[89] D. Eng, J. Humphrey, and S. Meyn. Fluid network models: Linear programs for control and performance bounds. In J. Cruz J. Gertler and Eds. M. Peshkin, editors, *Proceedings of the 13th World Congress of International Federation of Automatic Control*, volume B, pages 19–24, San Francisco, California, June 30 to July 5 1996.

[90] A. Ephremides and S. Verdu. Control and optimization methods in communication network problems. *IEEE Transactions on Automatic Control*, 34:930–942, 1989.

[91] T.M. Farrar. Generalisations of the join the shortest queue rule for symmetric queues - a sample-path proof. 1993.

[92] A. Federgruen and H. Groenevelt. Characterization and control of achievable performance in general queueing systems. *Operations Research*, 36:733–741, 1988.

[93] A. Federgruen and H. Groenevelt. M/g/c queueing systems with multiple customer classes: character- ization and control of achievable performance. *Management Science*, 34, 1988.

[94] E.A. Feinberg and D.J. Kim. Bicriterion optimization of an $M|G|1$ queue with a removeable server. *Probability in the Engineering and Informational Sciences*, 10:57–73, 1996.

[95] E.A. Feinberg and M.I. Reiman. Optimality of randomized trunk reservation. *Probability in the Engineering and Informational Sciences*, 8:463–489, 1994.

[96] E.A. Feinberg, Y.A.Kogan, and A.N. Smirnov. Optimal control by the retransmission probability in slotted ALOHA systems. *Performance Evaluation*, 5:85–96, 1985.

[97] S. Floyd and V. Jacobson. Random early detection gateways for congestion avoidance. *IEEE/ACM Transactions on Networking*, 1:397–413, 1993.

[98] F. Forges. Repeated games of incomplete information: Non-zero-sum. In R. J. Aumann and S. Hart, editors, *Handbook of Game Theory*, volume 1. Elsevier, North-Holland.

[99] G. Foschini. On heavy traffic diffusion analysis and dynamic routing in packet-switched networks. *Computer Performance*, pages 499–513, 1977. Chandy, K.M. and Reiser, M. (eds.), North-Holland.

[100] G. Foschini and J. Salz. A basic dynamic routing problem and diffusion. *IEEE Transactions on Communication*, 26:320–327, 1978.

[101] R. Hariharan V. G., Kulkarni, and S. Stidham. Optimal control of admission and routing to two parallel infinite- server queues. *Proceedings of 29th IEEE Conference on Decision and Control*, 1990.

[102] J. Galtier. Geographical reservation for guaranteed handover and routing in low earth orbit constellations. In *WCSF'99*, 1999.

[103] A.S. Gandhi and C.G. Cassandras. Optimal control of polling models for transportation applications. *Journal of Mathematical and Computer Modeling*, 23(11-12):1–23, 1996.

[104] M. Gavious and Z. Rosberg. A restricted complete sharing policy for a stochastic knapsack problem in B-ISDN. *IEEE Transactions on Communications*, 42, No. 7:2375–2379, 1994.

[105] L. Georgiadis, W. Szpankowski, and L. Tassiulas. A scheduling policy with maximal stability region for ring networks with spatial reuse. *Queueing Systems*, 19:131–148, October 1995.

[106] J. C. Gittins. Bandit processes and dynamic allocation indices. *Journal Royal Statistical Society*, B41:148–164, 1979.

[107] P. Glasserman and D. Yao. Monotone optimal control of permutable GSMPs. *Mathematics of Operations Research*, 19:449–476, 1994.

[108] P. Glasserman and D. Yao. *Monotone Structure in Discrete-Event Systems.* Wiley, New York, 1994.

[109] J. W. Grizzle, S. I. Marcus, and K. Hsu. A decentralized control strategy for multiaccess broadcast networks. *Large Scale Systems*, 3:75–88, 1982.

[110] M. Grossglauser and D. Tse. A framework for robust measurement-based admission control. *IEEE/ACM Trans. on Networking*, 7:293–309, 1999.

[111] P. Gupta and P. R. Kumar. A system and traffic dependent adaptive routing algorithm for Ad Hoc networks. In *Proceedings of the 37th IEEE Conference on Decision and Control*, Tampa, Florida, USA, Dec. 1998.

[112] B. Hajek. Optimal control of two interacting service stations. *IEEE Transactions on Automatic Control*, 29:491–499, 1984.

[113] B. Hajek. Extremal splittings of point processes. *Mathematics of Operations Research*, 10:543–556, 1985.

[114] R. Hariharan, V. G. Kulkarni, and S. Stidham. A survey of research relevant to virtual-circuit routing in telecommunication networks. preprint, Department of Operations Research, University of North Carolina at Chapel Hill, 1990.

[115] R. Hariharan, V.G. Kulkarni, and S. Stidham, Jr. A survey of research relevant to virtual-circuit routing in telecommunication networks. Technical Report UNC/OR/TR90-13, University of N.C. at Chapel Hill, 1990.

[116] J. M. Harrison and L. M. Wein. Scheduling networks of queues: heavy traffic analysis of a simple open network. *Queueing Systems*, 5(4):265–279, 1989.

[117] J. M. Harrison and L. M. Wein. Scheduling networks of queues: heavy traffic analysis of a two-station closed network. *Operations Research*, 38(6):1052–1064, 1990.

[118] R. Hassin and M. Haviv. Equilibrium strategies and the value of information in a two line queueing system with threshold jockeying. *Stochastic Models*, 10:415–435, 1994.

[119] M. Herzberg. An optimal decision process for routing circuit-switched calls originated by users of a private distribution network. In A. Jensen and V. B. Iversen, editors, *Teletraffic and Datatraffic in a period of change, ITC-13*, pages 453–458. Elsevier Science Publisher B. V. (North-Holland), 1991.

[120] M. Herzberg and D. McMillan. State-dependent control of call arrivals in layered cellular mobile networks. *Telecommunication Systems*, 1:365–378, 1993.

[121] K. F. Hinderer. *Foundation of Non-stationary Dynamic Programming with Discrete Time Parameter, Lectur Notes in Operations Research and Mathematical Systems No. 33*. Springer-Verlag, Berlin, Heidelberg, New York, 1970.

[122] M. Hofri and Z. Rosberg. packet delay under the golden ratio weighed tdm policy in a multiple-access channel. *IEEE Trans. Inform. Theory*, 33:341–349, 1987.

[123] M. Hofri and K.W. Ross. On the optimal control of two queues with server setup times and its analysis. *SIAM Journal on Computing*, 16:399–420, 1987.

[124] G. Hooghiemstra, M. Keane, and S. van de Ree. Power series for stationary distributions of coupled processor models. *SIAM J. Appl. Math.*, 48(5):1159–1166, 1988.

[125] A. Hordijk and G. M. Koole. On the assignment of customers to parallel queues. *Probability in the Engineering and Informational Sciences*, 6:495–511, 1992.

[126] A. Hordijk and G. M. Koole. On the optimality of LEPT and μc rules for parallel processors and dependent arrival processes. *Advances in Applied Probability*, 25:979–996, 1993.

[127] A. Hordijk and G.M. Koole. The μc-rule is not optimal in the second node of the tandem queue: A counterexample. *Advances in Applied Probability*, 24:234–237, 1992.

[128] A. Hordijk, G.M. Koole, and J.A. Loeve. Analysis of a customer assignment model with no state information. *Probability in the Engineering and Informational Sciences*, 8:419–429, 1994.

[129] A. Hordijk and J.A. Loeve. Undiscounted Markov decision chains with partial information; an algorithm for computing a locally optimal periodic policy. *ZOR - Mathematical Methods of Operations Research*, 40:163–181, 1994.

[130] A. Hordijk and F. Spieksma. Constrained admission control to a queueing system. *Advances in Applied Probability*, 21:409–431, 1989.

[131] M. T. Hsiao and A. A. Lazar. Optimal flow control of multiclass queueing networks with partial information. *IEEE Transactions on Automatic Control*, 35 No. 7:855–860, 1990.

[132] M. T. Hsiao and A. A. Lazar. Optimal decentralized flow control of Markovian queueing networks with multiple controllers. *Performance Evaluation*, 13:181–204, 1991.

[133] K. Hsu and S.I. Marcus. Decentralized control of finite state Markov processes. *IEEE Transactions on Automatic Control*, 27 No. 2:426–431, 1982.

[134] P. J. Hunt and C. N. Laws. Asymptotically optimal loss network control. *Mathematics of Operations Research*, 18(4):880–900, 1993.

[135] A. Itai and Z. Rosberg. A golden ratio control policy for a multiple-access channel. *IEEE Transactions on Automatic Control*, 29:712–718, 1984.

[136] V. Jacobson. Congestion avoidance and control. In *ACM SIGCOMM 88*, pages 273–288, 1988.

[137] T. Jimenez. Optimal admission control for high-speed networks: A dynamic programming approach. In *Proceedings of the 39th IEEE Conference on Decision and Control, Sidney, Australia*, Dec. 2000.

[138] S. G. Johansen and S. Stidham. Control of arrivals to a stochastic input-output system. *Advances in Applied Probability*, 12:972–999, 1980.

[139] P. K. Johri. Optimality of the shortest line discipline with state-dependent service times. *European Journal of Operational Research*, 41:157–161, 1990.

[140] E.G. Coffman Jr, Z. Liu, and R.R. Weber. Optimal robot scheduling for web search engines. *Journal of Scheduling*, 1, 1994.

[141] T. Kämpke. On the optimality of static priority policies in stochastic scheduling on parallel machines. *Journal of Applied Probability*, 24:430–448, 1987.

[142] T. Kämpke. Optimal scheduling of jobs with exponential service times on identical parallel processors. *Operations Research*, 37:126–133, 1989.

[143] F. P. Kelly. Effective bandwidth at multi-class queues. *Queueing Systems*, 9:5–16, 1991.

[144] F. P. Kelly and C. N. Laws. Dynamic routing in open queueing networks: Brownian models, cust constraints and resource pooling. *Queueing Systems*, 13:47–86, 1993.

[145] P. B. Key. Some control issues in telecommunications networks. In F. P. Kelly, editor, *Probability, Statistics and Optimisation A tribute to Peter Whittle*, pages 383–395. Wiley, 1994.

[146] P.B. Key. Optimal control and trunk reservation in loss networks. *Probability in the Engineering and Informational Sciences*, 4:203–242, 1990.

[147] M. Kitaev and V. Rykov. *Controlled Queueing Systems*. CRC Press, 1995.

[148] G. P. Klimov. Time sharing systems. *Theory of Probability and Applications*, 9:532–551, 1974.

[149] G. M. Koole. Dynamic programming tools for control of telecommunication systems. In *Proceedings of the 35th IEEE CDC*, Dec. 1996.

[150] G. M. Koole. Structural results for the control of queueing systems using event-based dynamic programming. *Queueing Systems*, 20:323–339, 1998.

[151] G. M. Koole. A transformation method for stochastic control problems with partial observations. *Systems and Control Letters*, 35:301–308, 1998.

[152] G. M. Koole. On the static assignment to parallel servers. *IEEE Transactions on Automatic Control*, 44:1588–1592, 1999.

[153] G. M. Koole. Optimal transmission policies for noisy channels. Research Report WS-515, Vrije Universiteit, Amsterdam, 1999.

[154] G. M. Koole and Z. Liu. Stochastic bounds for queueing systems with multiple on-off sources. *Probability in the Engineering and Informational Sciences*, 12:25–48, 1998.

[155] G. M. Koole and P. Nain. On the value function of a priority queue with an application to a controlled polling model. *to appear in QUESTA*, 1999.

[156] G. M. Koole and O. Passchier. Optimal control in light traffic Markov decision processes. *ZOR - Mathematical Methods of Operations Research*, 45:63–79, 1997.

[157] G.M. Koole. Stochastische dynamische programmering met bijvoorwaarden (translation: Stochastic dynamic programming with additional constraints). Master's thesis, Leiden University, 1990.

[158] G.M. Koole. Optimal server assignment in the case of service times with monotone failure rates. *Systems and Control Letters*, 20:233–238, 1993.

[159] G.M. Koole. A simple proof of the optimality of a threshold policy in a two-server queueing system. *Systems and Control Letters*, 26:301–303, 1995.

[160] G.M. Koole. On the use of the power series algorithm for general Markov processes, with an application to a Petri net. *INFORMS Journal on Computing*, 9:51–56, 1997.

[161] G.M. Koole. Stochastic scheduling with event-based dynamic programming. *ZOR - Mathematical Methods of Operations Research*, 51, 2000.

[162] G.M. Koole and Z. Liu. Nonconservative service for minimizing cell loss in ATM networks. In *Proceedings of the 33rd Allerton Conference on Communication, Control, and Computing*, pages 736–745, Illinois, USA, 1995.

[163] 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.

[164] Y.A. Korilis and A. lazar. Why is flow control hard: optimality, fairness, partial and delayed information. *preprint*, 1995.

[165] K. R. Krishnan and F. Hubner-Szabo de Bucs. Admission control and state-dependent routing for multirate circuit-switched traffic. In *Proceedings of the 15th ITC*, pages 1043–1055. Elsevier Science B. V., 1997.

[166] K.R. Krishnan and T.J. Ott. State-dependent routing for telephone-traffic: Theory and results. In *Proceedings of the 25th IEEE Conference on Decision and Control*, pages 2124–2128, 1986.

[167] K.R. Krishnan and T.J. Ott. Separable routing: A scheme for state-dependent routing of circuit switched telephone traffic. *Annals of Operations Research*, 35:43–68, 1992.

[168] H. Kroner, G. Hebuterne, P. Boyer, and A. Gravey. Priority management in ATM switching nodes. *IEEE J. Selected Areas in Communications*, pages 418–427, Apr. 1991.

[169] V.G. Kulkarni and Y. Serin. Optimal implementable policies: Discounted cost case. In W. J. Stewart, editor, *Proceeding of the International Meeeting on Computations with Markov Chains*, pages 283–306, Raleigh, NC, USA, 1995. Kluwer Acaedmic Publishers.

[170] J. Kuri and A. Kumar. Optimal control of arrivals to queues with delayed queue length information. In *Proceedings of the 31th IEEE Conference on Decision and Control*, 1992.

[171] J. Kuri and A. Kumar. On the optimal control of arrivals to a single queue with arbitrary feedback delay. *Queueing Systems*, 27(1-2):1–16, 1997.

[172] B. Kurtaran. Decentralized stochastic control with delayed sharing information pattern. *IEEE Transactions on Automatic Control*, 24:656–657, 1976.

[173] B. Kurtaran. Corrections and extensions to "Decentralized stochastic control with delayed sharing information pattern". *IEEE Transactions on Automatic Control*, 24:656–657, 1979.

[174] H. J. Kushner. Control of trunk line systems in heavy traffic. *SIAM J. Control Optim.*, 33:765–803, 1995.

[175] H. J. Kushner. Heavy traffic analysis of controlled multiplexing systems. *Queueing Systems*, 28:79–107, 1998.

[176] H.J. Kushner and L.F. Martins. Heavy traffic analysis of a data transmission system with many independent sources. *SIAM J. Appl. Math.*, 53:1095–1122, 1993.

[177] H.J. Kushner and L.F. Martins. Heavy traffic analysis of a controlled multi class queueing network via weak convergence theory. *SIAM J. on Contral and Optimization*, 34:1781–1797, 1996.

[178] H.J. Kushner and J. Yang. Numerical methods for controlled routing in large trunk line systems via stochastic control theory. *ORSA J. Computing*, 6:300–316, 1994.

[179] H.J. Kushner, J. Yang, and D. Jarvis. Controlled and optimally controlled multiplexing systems: A numerical exploration. *Queueing Systems*, 20(3-4):255–291, 1995.

[180] W.M. Lan and T. Lennon Olsen. A lower bound for dynamic scheduling of single machine multi-product systems with setups. manuscript, 1999.

[181] A. A. Lazar. Optimal flow control of a class of queuing networks in equilibrium. *IEEE Transactions on Automatic Control*, 28:1001–1007, 1983.

[182] W. G. Lazarev and S. M. Starobinets. The use of dynamic programming for optimization of control in networks of communications of channels. *Engineering Cybernetics (Academy of Sciences, USSR)*, (3), 1997.

[183] H. Levy, M. Sidi, and O.J.Boxma. Dominance relations in polling systems. *Queueing Systems*, 6:155–172, 1990.

[184] W. Lin and P.R. Kumar. Optimal control of a queueing system with two heterogeneous servers. *IEEE Transactions on Automatic Control*, 29:696–703, 1984.

[185] S.A. Lippman. Applying a new device in the optimization of exponential queueing systems. *Operations Research*, 23:687–710, 1975.

[186] Z. Liu, P. Nain, and D. Towsley. On optimal polling policies. *Queueing Systems*, 11:59–83, 1992.

[187] Z. Liu, P. Nain, and D. Towsley. Sample path methods in the control of queues. *Queueing Systems*, 21:293–336, 1995.

[188] Z. Liu, P. Nain, and D. Towsley. Exponential bounds with application to call admission. *Journal of the ACM*, 44:366–394, 1997.

[189] Z. Liu and D. Towsley. Optimality of the round-robin routing policy. *Journal of Applied Probability*, 31:466–475, 1994.

[190] J.A. Loeve. *Markov Decision Chains with Partial Information*. PhD thesis, Leiden University, 1995.

[191] D.-J. Ma and A. M. Makowski. A class of steering policies under a recurrence condition. In *Proceedings of the 27th IEEE Conference on Decision and Control*, pages 1192–1197, Austin, TX, USA, Dec. 1988.

[192] D.-J. Ma and A. M. Makowski. A class of two-dimensional stochastic approximations and steering policies for markov decision processes. In *Proceedings of the 31st IEEE Conference on Decision and Control*, pages 3344–3349, Tucson, Arizona, USA, Dec. 1992.

[193] U. Madhow, M.L. Honing, and K. Steiglitz. Optimization of wireless resources for personal communications mobility tracking. In *Proceedings of IEEE Infocom '94*, pages 577–584, 1994.

[194] B. Maglaris and M. Schwartz. Optimal fixed frame multiplexing in integrated line- and packet-switched communication networks. *IEEE Transactions on Information Theory*, 28:263–273, 1982.

[195] G. Malkin. Rip version 2 carying additional information. In *IETF RFC 1388*, 1994.

[196] P. Marbach, O. Mihatsch, and J. N. Tsitsiklis. Call admission control and routing in integrated service networks using neuro-dynamic programming. *IEEE Journal on Selected Areas in Communications*, 18(2):197–208, 2000.

[197] R. Menich and R. Serfozo. Optimality of routing and servicing in dependent parallel processing systems. *Queueing Systems*, 9(4):403–418, 1991.

[198] S. P. Meyn. The policy improvement algorithm: General theory with applications to queueing networks and their fluid models. In *Proceedings of the 35th IEEE Conference on Decision and Control*, Kobe, Japan, Dec. 1996.

[199] S. P. Meyn. The policy improvement algorithm for Markov Decision Processes with general state space. *IEEE Transactions on Automatic Control*, 42:191–196, 1997.

[200] S. P. Meyn. Stability and optimization of queueing networks and their fluid models. In *Proceedings of the Summer Seminar on The Mathematics of Stochastic Manufacturing Systems, VA, June 17–21, 1996, Williamsburg, VA. In Lectures in Applied Mathematics,* ***33****, American Mathematical Society*, volume 33, 1997.

[201] S. P. Meyn. Feedback regulation for sequencing and routing in multiclass queueing networks. In *2000 IEEE International Symposium on Information Theory*, Sorrento, Italy, June 25 - June 30 2000.

[202] R. Milito and E. Fernández-Gaucherand. Open-loop routing of n arrivals to m parallel queues. *IEEE Transactions on Automatic Control*, 40:2108–2114, 1995.

[203] R. Milito and E. Fernández-Gaucherand. Routing arrivals to queues in parallel. In *Proc. 34th IEEE Conference on Decision and Control*, pages 1415–1420, New Orleans, LA, 1995.

[204] D. Mitra, M.I. Reiman, and J. Wang. Robust dynamic admission control for unified cell and call qos in statistical multiplexers. *IEEE J. Sel. Areas in Commun.*, 16, No.5:692–707, June 1998.

[205] P. Nain. Interchange arguments for classical scheduling problems in queues. *Systems and Control Letters*, 12:177–184, 1989.

[206] P. Nain. Qualitative properties of the Erlang blocking model with heterogeneous user requirements. *Queueing Systems*, 6:189–206, 1990.

[207] P. Nain and K. W. Ross. Optimal priority assignment with hard constraint. *IEEE Transactions on Automatic Control*, 31:883–888, 1989.

[208] P. Nain, P. Tsoucas, and J. Walrand. Interchange arguments in stochastic scheduling. *Journal of Applied Probability*, 27:815–826, 1989.

[209] P. Nain, P. Tsoucas, and J. Walrand. Interchange arguments in stochastic scheduling. *Journal of Applied Probability*, 27:815–826, 1989.

[210] K. Nakade, M. Ohnishi, T. Ibaraki, and T. Ohno. On the average optimality of circular assignment policy. *Queueing Systems*, 11:241–254, 1992.

[211] P. Naor. On the regulation of queueing size by levying tolls. *Econometrica*, 37:15–24, 1969.

[212] K. S. Narendra, E. A. Wright, and L. G. Mason. Application of learning automata to telephone traffic routing and control. *IEEE Trans. Systems, Man and Cybernetics*, 7(11), 1977.

[213] J. Nino-Mora. On the throughput-wip trade-off in queueing systems, diminishing returns and the threshold property: A linear programming approach. *Submitted to Mathematics of Operations Research*, 1998.

[214] E. Nordstrom. Call admission control for preemptive and partially blocking service integration schemes in ATM networks. submitted, 1998.

[215] E. Nordstrom. *Markov Decision Problems in ATM Traffic Control.* PhD thesis, Department of Computer Systems, Upsala University, available as report DoCS 98/100, 1998.

[216] E. Nordstrom and J. Carlstrom. A reinforcement learning scheme for adaptive link allocation in ATM networks. In J. Alspector, R. Goodman, and T. X. Brown, editors, *Proceedings of the International Workshop on Applications of Neural Networks to Telecommunication 2, IWANNT'95*, pages 88–95, Stockholm, Sweedan, 1995. Lawrence Erlbaum.

[217] E. Nordstrom and J. Carlstrom. Near-optimal link allocation of blokable narrow-band and queueable wide-band call traffic in ATM networks. In V. Ramasawami and P. E. Wirth, editors, *Proceedings of the 15th International Teletraffic Congress, ITC'15*, pages 987–996, Washington D. C, USA, 1997. Elseivier Science.

[218] E. Nordstrom and J. Carlstrom. Call admission control and routing for integrated CBR/VBR and ABR services: A Markov decision approach. In *Proceedings of the ATM99 workshop*, Kochi, Japan, May 1999.

[219] J. Ou and L. M. Wein. On the improvement from scheduling a two-station queueing network in heavy traffic. *Operations Research Letters*, 11(4):225–232, 1992.

[220] J. D. Papastavrou, S. Rajagopalan, and A. J. Kleywegt. The dynamic and stochastic knapsack problem with deadlines. *Management Science*, 42, No. 12:1706–1718, 1996.

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

[222] C. E. Perkins and E. M. Royer. Ad-hoc on-demand distance vector routing. In *Proceedings of the 2nd IEEE Workshop on Mobile computer systems and Applications*, pages 90–100, 1999.

[223] D. Towsley R.-H. Hwang, J.F. Kurose. MDP routing for multirate loss networks. *Computer Networks and ISDN*, 2000.

[224] R. Ramjee, R. Nagarajan, , and D. Towsley. On optimal call admission control in cellular networks. In *Proceedings of IEEE INFOCOM 96*, pages 43–50, 1996.

[225] M.I. Reiman and L.M. Wein. Dynamic scheduling of a two-class queue with setups. *Operations Research*, 46(4):532–547, 1998.

[226] R. Rezaiifar, A. M. Makowski, and S. P. Kumar. Stochastic control of handoffs in cellular networks. *IEEE Journal of Selected Areas in Communications*, 13(7):1348–13162, September 1995.

[227] U. Rieder. Decentralized markov decision problems with delayed information. In *Abstracts of the 10th INFORMS Applied Probability Conference*, University of Ulm, Germany, July 1999.

[228] F. E. Ross. An overview of fddi: the fiber distributed data interface. *IEEE J. Select. Areas Commun.*, 7(7):1043–1051, Sept. 1989.

[229] K. Ross and D. H. K. Tsang. The stochastic knapsack problem. *IEEE Transactions on Communications*, 37:740–747, 1989.

[230] K. Ross and D. D. Yao. Monotonicity properties for the stochastic knapsack. *IEEE Transactions on Information Theory*, 36:1173–1179, 1990.

[231] H. Rummukainen and J. Virtamo. Polynomial cost approximations in Markov decision theory based least cost routing. *submitted, available in http://www.tct.hut.fi/tutkimus/com2/*, 2000.

[232] F.C. Schoute. Decentralized control in packet switched satellite communication. *IEEE Transactions on Automatic Control*, 23:362–371, 1978.

[233] L. I. Sennott. *Stochastic dynamic programming and the control of queueing systems.* Wiley, New York, 1999.

[234] R. Serfozo. Monotone optimal policies for Markov decision processes. In R. Wets, editor, *Stochastic Systems, II: Optimization*, volume 6, pages 202–215, Amsterdam, New York, 1976. North-Holland. Mathematical Programming Studies.

[235] R. Serfozo. Optimal control of random walks, birth and death processes, and queues. *Advances in Applied Probability*, 13:61–83, 1981.

[236] J.G. Shanthikumar and D.D. Yao. Multiclass queueing systems: Polymatroidal structure and optimal scheduling control. *Operations Research*, 40:S293–S299, 1992.

[237] S. Sharma and Y. Viniotis. Optimal buffer management policies for shared-buffer ATM switches. *IEEE/ACM transactions on networking*, 7(4):575–587, August 1999.

[238] P.D. Sparaggis. Routing and scheduling in heterogeneous systems: A sample path approach. *IEEE Transactions on Automatic Control*, 40:156–161, 1995.

[239] P.D. Sparaggis, C.G. Cassandras, and D.F. Towsley. On the duality between routing and scheduling systems with finite buffer spaces. *IEEE Transactions on Automatic Control*, 38:1440–1446, 1993.

[240] S. Stidham. Socially and individually optimal control of arrivals to a $GI|M|1$ queue. *Management Science*, 24:1598–1610, 1970.

[241] S. Stidham. Optimal control of arrivals to queues and networks of queues. In *Proceedings of the 21th IEEE Conference on Decision and Control*, 1982.

[242] S. Stidham. Optimal control of admission, routing, and service in queues and networks of queues: a tutorial review. *Proceedings ARO Workshop: Analytic and Computational Issues in Logistics R and D*, pages 330–377, 1984. George Washington University.

[243] S. Stidham. Optimal control of admission to a queueing system. *IEEE Transactions on Automatic Control*, 30:705–713, 1985.

[244] S. Stidham. Scheduling, routing, and flow control in stochastic networks. *Stochastic Differential Systems, Stochastic Control Theory and Applications*, IMA-10:529–561, 1988. W. Fleming and P.L. Lions, eds.

[245] S. Stidham and N. U. Prabhu. Optimal control of queueing systems. In A.B. Clarke, editor, *Mathematical Methods in Queueing Theory*, volume 98, pages 263–294, Berlin, 1974. Springer-Verlag. Lecture Notes in Economics and Mathematical Systems.

[246] S. Stidham, S. Rajagopal, and V. G. Kulkarni. Optimal flow control of a stochastic fluid-flow system. *IEEE Journal on Selected Areas in Communications*, 13:1219–1228, 1995.

[247] S. Stidham and R. Weber. A survey of Markov decision models for control of networks of queues. *Queueing Systems*, 13:291–314, 1993.

[248] S. Stidham and R.R. Weber. Monotonic and insensitive optimal policies for control of queues with undiscounted costs. *Operations Research*, 37:611–625, 1989.

[249] J. Talim, Z. Liu, P. Nain, and E. G. Coffman. Controlling robots in web search engines. *Submitted to Performance Evaluation*, 1999.

[250] J. Talim, Z. Liu, P. Nain, and E. G. Coffman. Optimizing the number of robots for web search engines. *To appear in Telecommunication Systems*, 1999.

[251] L. Tassiulas and A. Ephremides. Jointly optimal routing and scheduling in packet radio networks. *IEEE Transactions on Information Theory*, 38(1):165–168, January 1992.

[252] L. Tassiulas and A. Ephremides. Dynamic server allocation to parallel queues with randomly varying connectivity. *IEEE Transactions on Information Theory*, 39(2):466–478, March 1993.

[253] L. Tassiulas and A. Ephremides. Throughput properties of a queueing network with distributed dynamic routing and flow control. *Advances in Applied Probability*, 28:285–307, March 1996.

[254] L. Tassiulas and S. Papavassiliou. Optimal anticipative scheduling with asynchronous transmission opportunities. *IEEE Transactions on Automatic Control*, 40(12):2052–2062, Dec. 1995.

[255] H. Tong and T. X. Brown. Adaptive call admission control under quality of service constraints: A reinforcement learning solution. *IEEE Journal on Selected Areas in Communications*, 18(2):209–221, Feb. 2000.

[256] D. Topkis. Minimizing a submodular function on a lattice. *Operations Research*, 26:305–321, 1978.

[257] D. Towsley, P.D. Sparaggis, and C.G. Cassandras. Optimal routing and buffer allocation for a class of finite capacity queueing systems. *IEEE Transactions on Automatic Control*, 37:1446–1451, 1992.

[258] F. Vakil and A. A. Lazar. Flow control protocols for integrated networks with partially observed voice traffic. *IEEE Transactions on Automatic Control*, 32:2–14, 1987.

[259] M. P. van Oyen, D.G. Pandelis, and D. Teneketzis. Optimality of index policies for stochastic scheduling with switching costs. *Journal of Applied Probability*, 29:957–966, 1992.

[260] P. Varaiya and J. Walrand. On delayed sharing patterns. *IEEE Transactions on Automatic Control*, 23:443–445, 1978.

[261] P. Varaiya, J. Walrand, and C. Buyukkoc. Extensions of the multiarmed bandit problem: the discounted case. *IEEE Transactions on Automatic Control*, 30:426–439, 1985.

[262] J. Walrand. A note on 'Optimal control of a queueing system with heterogeneous servers'. *System Control Letters*, 4:131–134, 1984.

[263] J. Walrand. *An Introduction to Queueing Networks.* Prentice Hall, Englewood Cliffs, NJ, 1988.

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

[265] R. Weber and S. Stidham. Control of service rates in networks of queues. *Advances in Applied Probability*, 24:202–218, 1987.

[266] L. M. Wein. Optimal control of a two-station brownian network. *Mathematics of Operations Research*, 5(2):215–242, 1990.

[267] L. M. Wein. Scheduling networks of queues: heavy traffic analysis of a two-station network with controllable inputs. *Operations Research*, 38(6):1065–1078, 1990.

[268] G. Weiss. Branching bandit processes. *Probability in the Engineering and Informational Sciences*, 2:269–278, 1988.

[269] G. Weiss. Optimal draining of a fluid re-entrant line. In Frank Kelly and Ruth Williams, editors, *Stochastic Networks*, volume 71 of IMA volumes in Mathematics and its Applications, pages 91–103. Springer-Verlag, New York, 1995.

[270] G. Weiss. Optimal draining of fluid re-entrant lines: Some solved examples. In *Stochastic Networks: Theory and Applications*, volume 4 of Royal Statistical Society Lecture Notes Series, pages 19–34. Oxford University Press, Oxford, 1996.

[271] G. Weiss. Scheduling and control of manufacturing systems – a fluid approach. In *Proceedings of the 37rd Allerton Conference on Communication, Control, and Computing*, Illinonis, USA, Sept. 1999.

[272] C. C. White. Monotone control laws for noisy, countable-state Markov chains. *European Journal of Operational Research*, 5:124–132, 1980.

[273] W. Whitt. Deciding which queue to join; some counterexamples. *Operations Research*, 34:55–62, 1986.

[274] P. Whittle. Multi-armed bandits and the gittins index. *Journal Royal Statistical Society*, B42:143–149, 1980.

[275] P. Whittle. *Optimal control: basics and beyond.* John Wiley and sons, 1996.

[276] W. Winston. Optimality of the shortest-line discipline. *Journal of Applied Probability*, 14:181–189, 1977.

[277] S.H. Xu and H. Chen. A note on the optimal control of two interacting service stations. Working paper, 1991.

[278] D. D. Yao and Z. Schechner. Decentralized control of service rates in a closed jackson network. *IEEE Transactions on Automatic Control*, 34 No. 2:236–240, 1989.

[279] R. D. Yates. A framework for uplink power control in cellular radio systems. *IEEE Journal on Selected Areas in Communications*, 13(7):1341–1347, September 1995.

[280] U. Yechiali. On optimal balking rules and toll charges in a $GI|M|1$ queueing process. *Operations Research*, 19:349–370, 1971.

[281] U. Yechiali. Optimal dynamic control of polling systems. In J.W. Cohen and C.D. Pack, editors, *Queueing, Performance and Control in ATM*, pages 205–217. North-Holland, 1991.

[282] U. Yechiali. Analysis and control of polling systems. In L. Donatiello & R. Nelson, editor, *Performance Evaluation of Computer and Communication Systems*, pages 630–650. Springer-Verlag, 1993.

Index

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - B.P. 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Lorraine : Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 Villers lès Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot St Martin (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, B.P. 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399