Towards the Control of Markov Chains with Constraints: Theory and Applications

B. Miller
School of Mathematical Sciences
Monash University, Melbourne Australia
and
A. A. Kharkevish Institute for Infromation Transmission Problems
Russain Academy of Sciences, Moscow Russia
boris.miller@monash.edu

Abstract

A problem of stochastic optimal control for Markov chains with a finite number of states in continuous time is considered. The problem statement is assumed a non-stationary finite horizon problem with constraints given as a set of inequalities. It is shown that the optimal control within the class of Markov policies, if it exists, may be found with the aid of maximum principle as a solution of some deterministic optimal control problem. Meanwhile it can be shown that under simple assumptions of convexity, which are common for optimal control problems, this solution is optimal also within more general class of predictive controls. An approach to the numerical solution of the optimal control problem has been suggested and its effectiveness is illustrated by examples from the areas of the INTERNET congestion avoidance and dams' management.

Key words: Markov chains, optimal control, problem with constraints, maximum principle

1 Introduction

Controlled Markov Chains (CMC) constitute a universal mathematical tool applicable to a wide class of applied problems including, but not limited to: stock and production management (Davis 1993), Internet congestion control (Low ea. 2002), large dams management (B. Miller and D. McInnes 2011), and many others.

Solutions to optimal control problems with feedback and active users have recently been developed and applied to Internet congestion control (A. Miller 2010). Even more recent is the extension to the class of connected controlled Markov chains, which is important for control of networks, such as wireless Internet, gas and water allocation systems (A. Miller and B. Miller 2011).

However, the existing theory covers mainly systems with complete observations, where the control depends on the current state of the Markov chains (MC). Even if this class of controls provides the optimal solution to the problem with constraints (B. Miller ea. 2011), there is an important class of systems where the state is not observable and have to be estimated on the basis of observations. These problems are usually referred to as partially observable Markov decision processes (POMDP). Typical problems arising in the development of transmission control protocols (TCP) in the Internet (Srikant 2004) also belong to this class.

Recent years have also seen a significant growth in the interest in the study of POMDPs with the emergence of important results on the existence of an optimal policy based on the observation of another jump process (Ceci ea. 2002) (with applications to life testing (Ceci and Mazliak 2004)). These results lead to the separation principle in joint estimation and stochastic control problems (Ceci and Mazliak 2004), (Davis 1993) and to the development of numerical algorithms (Hao Zhang 2010). Usually, these problems may be solved with the aid of a two-step procedure: estimation of the unobservable state on the basis of filtering (Aggoun-Elliott 2004), (Liptser and Shiryaev 2005) and reduction of the problem to completely observable MDP (Markov decision problems) on the basis of the separation principle.

However, in order to obtain the optimal control one needs first to solve the dynamic programming equation, which belongs to a class of functional-differential equations and which is difficult to solve even numerically and even for small dimensions.

Meanwhile, problems related to TCP usually invoke stochastic control with incomplete observations. The typical situation is that on the customer side the state of router is unknown, and the customer sends data packets without any guaranty of acceptance or rejection. In some very popular active queue management algorithms, like Random Early Detection (RED) (Floyd and Jacobson 1993), the number of rejected packets serves as a proxy to the router's state and serves as feedback for customers to control their transmission rate and prevent congestion. The control of the transmission rate based on the estimation of the router's state is suggested in (B. Miller ea. 2005-2). In some special cases it has been proven that the separation principle is valid and new algorithms for control of transmission rate have been proposed. It was shown that the optimal law of the transmission rate is different from the widely known additive increase/multiplicative-decrease (AIMD) algorithm and should be realized with the aid of a concave window curve which depends on the dynamic of the conditional probabilities for the non-observable state of the router. This algorithm provides much less variability (up to five times (B. Miller ea. 2005-2)) of the transmission rate and therefore shows more stable performance particularly for high speed, long distance and wireless communication networks. This is why these algorithms were used in some newly developed TCPs like TCP-Illinois (Liu ea. 2008) and Westwood+ (Westwood+ 2008).

In summary, the state-of-the-art in control of MC with a finite number of states is as follows.

- The theory for completely observable systems with constraints guarantees the existence of an optimal control and its explicit representation with the help of the maximum principle. Numerical implementation can then be achieved by first reducing the dynamic programming equation to a system of ordinary differential equations (ODE) and second by solving the associated maxmin problem (B. Miller ea. 2010)and (B. Miller ea. 2011). In this case effective numerical algorithms have been developed – in comparison with popular TCP

RED (Srikant 2004), the application of optimized algorithms permits to improve the performance characteristics, by up to 50% for the average waiting time and by up to 20% for the reduction of the number of rejected demands (A. Miller and B. Miller 2012), (A. Miller and B. Miller 2013).

- While both existence and characterisation of an optimal control are also guaranteed for partially observable systems (Ceci ea. 2002), (Ceci and Mazliak 2004), a solution, whether explicit or numerical, is not available.

- Unlike for completely observable systems, when only partial observations are available, one has to estimate the states (usually via filtering). However, this effectively results in preventing the reduction of the dynamic programming equation to a system of ODEs (curse of dimensionality coupled with a differential-difference form of the equation) and the use of the maximum principle for a numerical implementation of the solution.

  On the other hand the direct use of the separation principle – a substitution of the MC state with its conditional expectation within the exact control (that for the corresponding completely observable case), is only sub-optimal. Furthermore, no estimation of the difference between optimal and sub-optimal solutions exists.

This article aims to demonastrate some promissing approach to the optimization of POCMC (partially observable) CMCs which permits:

- develop a solution for the stochastic control problem for POCMCs, both unconstrained and constrained,

- develop effective numerical algorithms for solution of optimal control problem of POCMCs, both unconstrained and constrained,

- develop approaches to new TCP for establishing the connection in abruptly changing environment based on the available observation and the estimation of unobservable MC states.

2 Towards the optimization of POCMC

2.1 Overview

2.1.1 Theory

The first step is to develop a solution of the unconstrained stochastic control problem for particular classes of POCMCs. So we shall initially consider the case of control-independent generators. The general case will require the novel implementation of a widely used technique (for example in financial mathematics), namely the change of measure. The final stage of the theoretical aspect of the approach is the formulation of the maxmin problem for solution of the problems with constraints.

It should be stressed that this is a non-classical problem that encompasses the control of both process and observations, and therefore needs the development of new filtering algorithms for discrete-continuous (hybrid) observations. These must handle both counting processes such as the acknowledgment or rejection notifications in TCP (which are discrete in nature), and continuous range measurements such as round-trip time (RTT).

The next step is the extension of this approach to the class of constrained problems. This can be achieved by judiciously exploiting the convex properties of the criterions' attainable set and by using a Lagrange multipliers approach - as in the the case of completely observable MDPs.

2.1.2 Numerical Algorithms

The aim is to develop effective numerical algorithms for solution of constrained optimal control problem of POCMCs. This part is based on the synthesis of stochastic modelling for controlled MCs and filtering results for observation of the MC states. Since the optimal values of criterions may be evaluated only statistically, one needs to join together Monte-Carlo statistical modelling and strong results for the description of optimal feedback controls.

2.1.3 Applications

The aim is to mathematically model and develop approaches to new TCP for establishing the connection in abruptly changing environment based on the available observation and the estimation of unobservable transmission lines states.

2.2 The control-independent case without constraints – an explicit solution

Here we use the standard problem statement (Elliott 1995, 2008), Chapter 12, and previous results obtained in (B. Miller ea. 2005-1), (B. Miller ea. 2005-2). We describe our approach in general terms to highlight the main ideas. Suppose that the state of a MC is described by a vector $X_t \in \mathcal{S} = \{e_1, .., e_N\}$ where N is the number of states and e_i is the vector with all null entries, except of i-th which is 1. Vector X_t satisfies the stochastic differential equation (Elliott 1995, 2008)

$$X_t = X_0 + \int_0^t A(s)X_s ds + W_t, \quad W_t \text{ – square integrable martingal} \tag{1}$$

with control independent generator $A(t)$. We assume that the system (1) is completed by a set of stochastic differential equations describing the discrete-continuous observations Y_t, generating $\mathcal{Y}_t = \sigma\{Y_s, 0 \leq s \leq t\}$ (for details see (Elliott 1995, 2008)). The aim of the control is to minimize the cost function

$$J_0[u(\cdot)] = E\left\{\int_0^T \langle g(s.u(s)), X_s\rangle ds\right\} \tag{2}$$

over all $\mathcal{Y}_t$ - predictable controls $u(t) \in U$, where U is the set of admissible control values.

As was observed in (B. Miller ea. 2005-1), if the generator of CMC does not depend on the control, then the optimal control may be found by application of the maximum principle Thm. 2.9 (Elliott 1995, 2008) or Remark 7, Thm. 4 (B. Miller ea. 2005-1). This simplification does not need the stipulation of the adjoint variables and the optimal control reduces to

$$u^*(t, q_t^*) = \underset{U}{argmax}\langle g(t,u), q_t^*\rangle = \underset{U}{argmax}\langle g(t,u), \pi_t^*\rangle, \tag{3}$$

where $\pi_t^* = E(X_t^*|\mathcal{Y}_t)$ and q_t^* is the conditional expectation of X_t^* with respect to some reference probability measure (provided by the change of measure) and $*$ refers to the optimal path. Since the conditional probabilities may be determined for any predictable control, the substitution of (3) into the system dynamics gives the optimal values of the running cost (2).

2.3 The control-independent case with constraints – a synthesis of the explicit solution and a Monte-Carlo approach

Unfortunately, there is no direct way to evaluate the optimal cost without solution of the dynamic programming equation. However, the value of the optimal cost corresponding to a given initial condition may be evaluated by Monte-Carlo.

Confidence levels may be estimated and for a given optimal control u^* one can obtain a confidence interval for the value function $\bar{J}_0 = E\{J_0[u^*]\}$.

As we have to solve the problem with constraints, which are represented in the form of inequalities $\bar{J}_k \leq 0, k = 1, ..., M$ for J_k having the same form as J_0 (2), the problem may be reduced to the unconstrained one by using the following observation.

An important characterization of the attainable set of constrained criteria is that under rather mild conditions its closure is convex (B. Miller ea. 2010) which permits to use the dual approach and to reduce the problem to a *maxmin* one (B. Miller ea. 2011), such as

$$E\left\{\mathcal{L}(\bar{\lambda}, u)\right\} = E\left\{J_0[u] + \sum_{k=1}^{M} \lambda_k J_k[u]\right\} \to \max_{\lambda_k \geq 0} \min_{u \in U},$$

where the inner minimization is made over the observation measurable control. The calculation of the current value of $\mathcal{L}(\bar{\lambda}, u)$ is made by Monte-Carlo modelling and via the maximization, by stochastic search, over Lagrange multipliers.

2.4 The general optimal control with constraints – a measure transformation approach

The difficulty in Monte-Carlo modelling resides in the dependence of the transition rates on controls; the solution of the stochastic equations needs the evaluation of very small transition probabilities if we need to approximate the problem with small time-steps. This automatically creates numerical challenges.

However, earlier we suggested another approach to modelling of CMC which is based on the change of measure (Aggoun-Elliott 2004) and which permits to transform the counting processes responsible for transitions into standard Poisson flows with constant transition intensities. This approach was used successfully for modelling of the optimal control of data transmission flow (B. Miller ea. 2005-1).

This change of measure approach is effective if the original CMC has a control-dependent generator and is described by the equation

$$X_t = X_0 + \int_0^t A(s, u(s)) X_s ds + W_t, \tag{4}$$

where $X_t \in \mathcal{S} = \{e_1, ..., e_N\}$ and X_0 is the initial state of the MC. The $(N \times N)$ matrix $A(t,u)$ is the generator of the MC, with entries $\lambda^{j,k}(t,u)$ and is assumed to be continuous on $(t,u) \in [0,T] \times U$, where $T < \infty$ and U are the time horizon and the control set respectively. Here the process W_t is a square integrable martingale. We also assume that the control is based on the observation of some vector discrete-continuous process $Y(t)$, which joins together discrete (counting processes) and continuous observations (diffusions) - for details see (Elliott 1995, 2008), Chapter 12. The difficulties of the optimal control search for this class of problems are described above. However, one can transform this problem to the equivalent problem with the control-independent generator $\bar{A}(t)$ with the aid of the following measure transformation (Aggoun-Elliott 2004), p. 227.

We assume that the control-dependent intensities $\lambda^{i,j}(t,u)$ (entries of $A(t,u)$) are either nil (transition from i to j is not possible) or uniformly-in-u bounded away from 0: $\lambda^{i,j}(t,u) \geq \delta^{i,j}(t) > 0$. Let $\mathcal{G}_t = \sigma\{X_s, y_s | 0 \leq s \leq t\}$. Define the conditional density

$$\Lambda_t = \mathbf{E}\left\{\frac{d\bar{P}}{dP}|\mathcal{G}_t\right\} =$$

$$\prod_{\substack{i \neq j \\ i \hookrightarrow j}} \exp\left\{\int_0^t \log\frac{\bar{\lambda}^{i,j}(s)}{\lambda^{i,j}(s,u(s))} d\mathcal{J}^{i,j}(s) - \int_0^t (\bar{\lambda}^{i,j}(s) - \lambda^{i,j}(s,u(s))ds\right\}, \tag{5}$$

where $\mathcal{J}^{i,j}(t)$ is the counting process of jumps from state i to j and $i \hookrightarrow j$ indicates that the transition is possible.

Under the probability measure $\bar{P}$ the counting processes $\mathcal{J}^{i,j}(t)$ have control-independent intensities $\bar{\lambda}^{i,j}(t)$. The problem of control of MC (4) can now be replaced by the equivalent problem for the MC with control-independent generator $\bar{A}$. The cost function must be replaced by

$$J[u] = \mathbf{E}\left\{\langle\phi, X_T\rangle + \int_0^T \langle g(t,u(t)), X_t\rangle\right\} = \bar{\mathbf{E}}\left\{\langle\phi, q_T^u\rangle + \int_0^T \langle g(t,u(t)), q_t^u\rangle\right\},$$

where $q_t^u = \bar{E}[\Lambda_t^u X_t^u | \mathcal{Y}_t]$ is the new phase variable which has to be substituted into relation (3) for determining the optimal control.

3 Model of the TCP descrided by Markov chains with incomplete information

Some recent results on controlled Markov chains with incomplete information (B. Miller ea. 2005-1) permit to analyse the flow rate control scheme with unknown probability of access which assumed to be changing according to the Hidden Markov Model driven by a Gilbert-type path model (Gilbert 1960) with two states. Transitions between two states are governed by the generator

$$A = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}.$$

Losses or ECN marks arrive according to a counting process N_t with intensities c_1 and c_2 ($c_1 < c_2$); that is, state 1 corresponds to a "good" state and state 2 corresponds to a "bad" of the transmission line.

The equation for the conditional probability $\pi_t^1 = E(I\{X_t = 1\}|\mathcal{F}_t^N)$ has the form (Liptser and Shiryaev 2005, Chapter 19, Example 2)

$$d\pi_t^1 = [-\lambda\pi_t^1 + (1-\pi_t^1)\mu + u_t\pi_t^1(1-\pi_t^1)(c_2 - c_1)]\,dt - \frac{\pi_t^1(1-\pi_t^1)(c_2-c_1)}{\pi_t^1 c_1 + \pi_t^2 c_2}\,dN_t. \tag{6}$$

Therefore, between jumps, control $u_t > 0$ acts to increase the conditional probability π^1, but at the arrival instants of jumps of N_t this probability abruptly decreases.

To calculate the "optimal" control, we take the cost function

$$f_0(u,x) = -\frac{a^2}{RTT^2(x)u},$$

which is frequently used to select the control in the TCP. This is a standard utility function of TCP New Reno (Srikant 2004), and it is rather convenient for calculations since the control maximizing the expression $\langle f_0(u) - kcu, \pi_t\rangle$ is uniformly bounded, strictly positive, and can be calculated explicitly.

Here the parameter $RTT(x)$ is the so-called *round trip time*, i.e., the sum of the time between sending a packet and its arrival to a recipient and the time necessary to obtain an acknowledge receipt from the recipient to the sender. This parameter depends on an unobservable state of the Markov chain as well. The optimal control, which maximizes $\langle f_0(u) - kcu, \pi_t\rangle$, has the form

$$u_{\text{opt}}(\pi_t) = \frac{a}{\sqrt{k}}\sqrt{\frac{\dfrac{\pi_t^1}{RTT^2(1)} + \dfrac{1-\pi_t^1}{RTT^2(2)}}{c^1\pi_t^1 + c^2(1-\pi_t^1)}}. \tag{7}$$

Between two consecutive losses, the conditional probability for the "good" state increases. Consequently, the "optimal" sending rate increases as well. When a packet is lost or a congestion notification arrives, the conditional probability for the "good" state, as well as the "optimal" sending rate, decreases abruptly by a jump.

In (B. Miller ea. 2005-1) we developed the optimal stochastic control scheme which demonstrates more stable behavior of the transmission rate and much less variability even in the case of abruptly changing probability of access (see Fig. 3). The difference in operating of the optimal and the standard AIMD TCP is shown in the Fig. 1 These results show that the optimal control is more effective than existing TCP Reno basing on the same admissible information. Standard TCPs use typically AIMD (Additive increase/muptiple decrease) scheme which is one of suboptimal stochastic controls scheme where the flow of rejected packages serves as the observations and the control (transmission rate) is a function of these observations.

To perform numerical experiments, one needs to substitute (7) into (6) and solve this differential equation numerically. It is important to note that one should solve the differential equation and generate the counting process N_t^u simultaneously since N_t^u depends on the sending rate u and the state of the path. Some examples of numerical solutions are presented in (B. Miller ea. 2005-1).

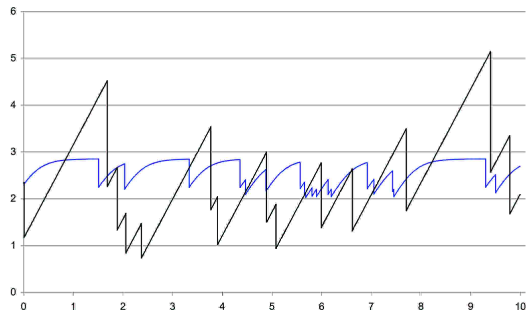

Figure 1: The time evolution of the sending rate for standard TCP Reno - (black) using AIMD scheme and for Optimal sending rate scheme - (blue)

The modeling was based on the standard driving Poisson process that does not depend on the state and control (Elliott 1995, 2008). In statistical modeling, this method is more efficient if the state of the Markov chain does not depend on the control since in this case one only needs to generate one standard state-and-control-independent Poisson process. Modeling results for the average control, its standard deviation, and the corresponding performance criteria obtained with the aid of this method are given below. All calculations were made with the same parameters for the router with two states, driven by the control independent Markov chain, where the control is the user sending rate, who is trying to maximize the volume of transmitted information "minus" the cost of traffic and the average number of lost packages (see the utility function corresponding to the case of so-called *minimum potential delay fairness* (Srikant 2004) p. 14). Note that the average values of the "optimal" and standard TCP controls appeared to be almost identical (see Fig. 1). We proved this by statistical modeling of 100 paths. The average values of the performance criterion averaged also by time are

$$\bar{J}[u_{\mathrm{OPT}}] = -2.86, \qquad \bar{J}[u_{AIMD}] = -3.06,$$

and the average values of the control are

$$\bar{u}_{\mathrm{OPT}} = 2.43, \qquad \bar{u}_{AIMD} = 2.32.$$

Even though the difference between the respective average criteria is not so substantial, the difference between the standard deviations is quite impressive:

$$\sigma(u_{\mathrm{OPT}}) = 0.763, \qquad \sigma(u_{AIMD}) = 4.452.$$

In other words, modeling results show that, under the same average control level, variations of the "optimal" control are much less than for the standard TCP. The average values and standard deviations obtained with the aid of 100 paths are shown in Figs. 2 and 3.

Monte-Carlo modelling of these two scheme shows that under the same at the average transmission rate (see Fig. 2), the variance, which characterized the stability under abruptly changing access probability, of the Optimal sending transmission rate scheme is 6 - times less (see Fig. 3).

The advantages of this Optimal sending rate scheme were successfully used in two new TCP, namely, in TCP-Illinois (Liu ea. 2008) and TCP-Westwood+ (Westwood+ 2008).

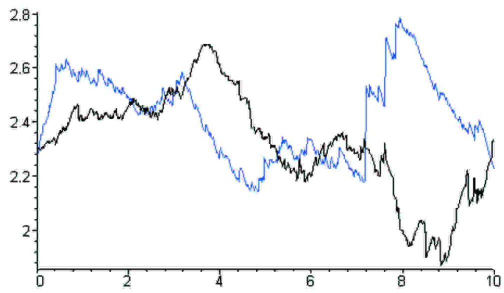

Figure 2: Average sending rate of TCP Reno - (black) using AIMD scheme and for Optimal sending rate scheme - (blue)

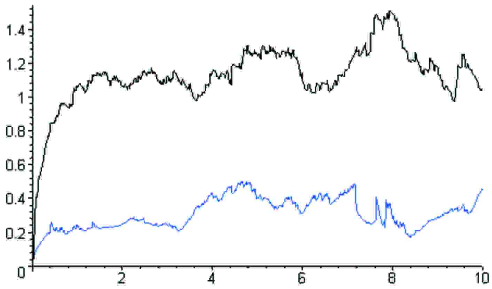

Figure 3: Variance of the sending rate for TCP Reno - (black) using AIMD scheme and for Optimal sending rate scheme - (blue)

References

Aggoun L. and Elliott R. J. 2004. “Measure Theory and Filtering: Theory and Applications.” *Cambridge University Press* April 2004.

Ceci C., Gerardi A., and Tardelli P. 2002. “Existence of Optimal Controls for Partially Observed Jump Processes.” *Acta Applicandae Mathematicae* Vol. 74, pp. 155–175.

Ceci C. and Mazliak L. 2004. “Optimal Design in Nonparametric Life Testing.” *Statistical Inference for Stochastic Processes* Vol. 7, pp. 305–325.

Davis M.H.A. 1993. “Markov Models and Optimization.” *London: Chapman and Hall.*

Elliott R.J., Aggoun L. and Moore J.B. 1995, 2008. “Hidden Markov Models. Estimation and Control.” *New York: Springer Verlag.*

Floyd S. and Jacobson V. 1993. “Random Early Detection Gateways for Congestion Avoidance.” *IEEE/ACM Transactions on networking* Vol. 1, No 4, pp. 393–413.

Gilbert, E. N. 1960. “Capacity of a Burst-Noise Channel.” *Bell Syst. Tech. J.* Vol. 1, No 5, pp. 1253–1265.

Kliazovich D. ea. 2008. “Logarithmic window increase for TCP Westwood+ for improvement in high speed, long distance networks.” *Computer Networks* Vol. 52, pp. 2395–2410.

Liptser R. Sh. and Shiryaev, A. N. 2005. "Statistics of Random Processes." *New York: Springer.*

Liu S., Basar T., and Shrikant R. 2008. "TCP-Illinois: A loss-and delays-based congestion control algorithm for high-speed networks." *Performance Evaluation* Vol. 65, pp. 417–440.

Low S. H., Paganini F. and Doyle J.C. 2002. "Internet Congestion Control" *IEEE Control Systems Magazine*, Vol. 22(1), pp. 28–43.

Miller A. B. 2010. "Using Methods of Stochastic Control to Prevent Overloads in Data Transmission Networks" *Automation and Remote Control*, Vol. 71, No. 9, pp. 1804–1815.

Miller A. and Miller B. 2012. "Towards the Analysis of TCP with the aid of Stochastic Control" *Proceedings of the IX International Conference Systems Identification and Control Problems 2012, Moscow, Russia, 30/01/2012 02/02/2012* pp. 326–346.

Miller A. and Miller B. 2013. "Application of stochastic control to analysis and optimization of TCP" *Proceedings of the 2013 Australian Control Conference, Perth, Australia, 04/11/2013 05/11/2013* pp. 238–243.

Miller A. B. and Miller B. M. 2011. "Control of connected Markov chains. Application to congestion avoidance in the Internet" *Proceedings of 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*, Orlando, FL, USA, December 12-15, pp. 7242–7248.

Miller B., Avrachenkov K., Stepanyan K., and Miller G. 2005. "The problem of the optimal stochastic control of a data flow with incomplete information" *Problems of Information Transmission* Vol. 41 No 2, pp. 150–170.

Miller B., Avrachenkov K., Stepanyan K., and Miller G. 2005. "Flow Control as Stochastic Optimal Control Problem with Incomplete Information" *Proceedings of 24th IEEE INFOCOM*, Miami, FL, USA, March 13-17, pp. 1328–1337.

Miller B. M. and McInnes D. J. 2011. "Management of a Large Dam via Optimal Price Control" *Proceedings of the 18th IFAC World Congress, August 28 - September 2, 2011, Milan, Italy* pp. 12432–12438.

Miller B. M., Miller G. B., and Semenikhin K. V. 2010. "Torwards the optimal control of Markov chains with constraints" *Automatica* Vol. 46, pp. 1495–1502.

Miller B. M., Miller G. B., and Semenikhin K. V. 2011. "Methods to Design Optimal Control of Markov Process with Finite State Set in the Presence of Constraints" *Automation and Remote Control*, Vol. 72, No. 2, pp. 323–341.

Srikant R. 2004. "The mathematics of Internet congestion control" *Birkhuser*, Boston.

Hao Zhang 2010. "Partially Observable Markov Decision Processes: A Geometric Technique and Analysis" *Operations Research*, Vol. 58, No. 1, pp. 214–228.