Scheduling with Impatience

Jean-Marie & Hyon

Introduction
The Model
The Question
The Literature
This talk

Dynamic Programming representation

$B = 1$
Results
Propagation
Computation of the threshold

$B \geq 2$

Questions

Optimal Scheduling services in a queueing system with impatience and setup costs

Alain Jean-Marie[1] Emmanuel Hyon[2]

[1]INRIA
LIRMM CNRS/Univ. Montpellier 2

[2]Université Paris Ouest Nanterre la Défense
LIP6

ISCIS 2010, London, 22 September 2010

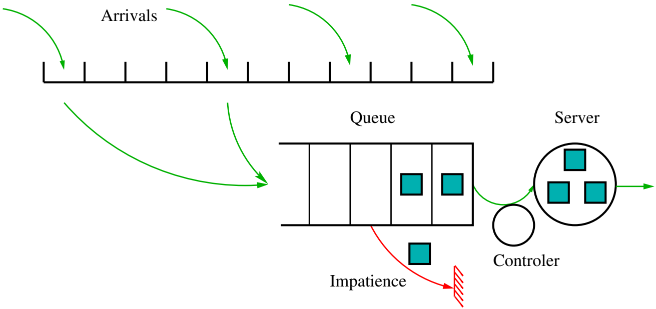
Outline

1. Introduction
   - The Model
   - The Question
   - The Literature
   - This talk
2. Dynamic Programming representation
3. The case $B = 1$
   - Case $B = 1$
   - Results
   - Propagation
   - Computation of the threshold
4. The case $B \geq 2$

Progress

1. Introduction
   - The Model
   - The Question
   - The Literature
   - This talk
2. Dynamic Programming representation
3. The case $B = 1$
   - Case $B = 1$
   - Results
   - Propagation
   - Computation of the threshold
4. The case $B \geq 2$

The Model



Arrival

- Customers arrive to an infinite-buffer queue.
- Time is discrete.
- The distribution of arrivals in each slot A_t, arbitrary with mean λ (customers/slot)

Scheduling with Impatience

Jean-Marie & Hyon

Introduction
The Model
The Question
The Literature
This talk

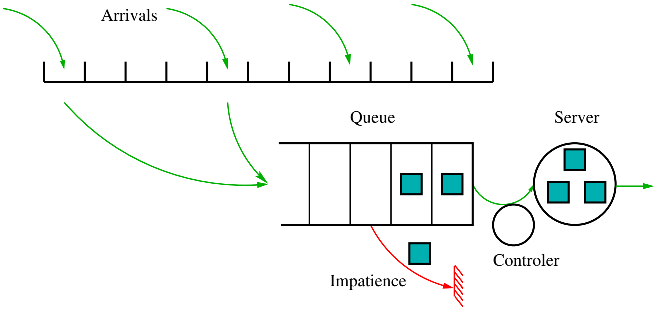
Dynamic Programming representation

$B = 1$
Results
Propagation
Computation of the threshold

$B \geq 2$

Questions

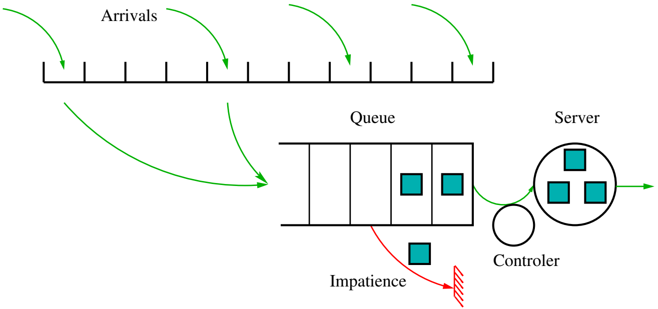
The Model



Services

- Service occurs by *batches* of size B.
- Service time is one slot.

The Model

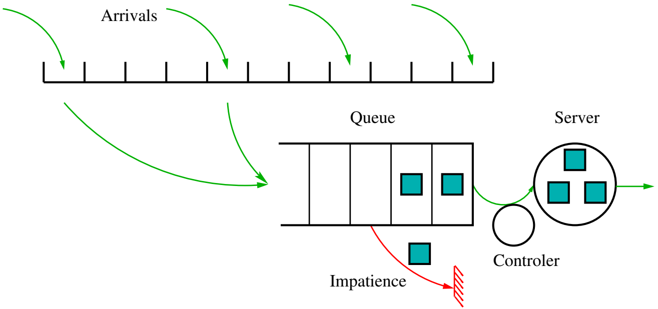


Deadline

Customers are *impatient*: they may leave before service.

- the individual probability of being impatient in each slot: α
- memoryless, geometrically distributed patience

The Model



Control

Service is *controlled*.

- The controller knows the number of customers but not their amount of patience: just the distribution.
- It decides whether to serve a batch or not.

The Question

What is the optimal *policy* π^* of the controller, so as to minimize the θ*-discounted* global cost:

$$v_\theta^\pi(x) \; = \; \mathbb{E}_x^\pi \left[\sum_{n=0}^{\infty} \theta^n \, c(x_n, q_n) \right] \; ,$$

where:

- x_n: number of customers at step n;
- q_n: decision taken at step n;

and $c(x, q)$ is the cost incurred, involving:

- c_B: cost for serving a batch (*setup cost*)
- c_H: per capita *holding cost* of customers
- c_L: per capita *loss cost* of impatient customers.

Related Literature

Control of queues and/or impatience (or reneging, abandonment) has a long history.

Optimal, deadline-based scheduling:

- Bhattacharya & Ephremides, 1989
- Towsley & Panwar, 1990

Optimal admission/service control (without impatience)

- Deb & Serfozo, 1973
- Altman & Koole, 1998 (admission)
- Papadaki & Powell, 2002 (service)

Optimal *routing* control with impatience

- Kocaga & Ward, 2009
- Movaghar, 2005

No optimal control of batch service in presence of stochastic impatience, so far.

Purpose of this talk

In the talk, we:

- give the solution to this problem for $B = 1$
- explain what goes wrong when $B \geq 2$

Progress

1 Introduction
- The Model
- The Question
- The Literature
- This talk

2 Dynamic Programming representation

3 The case $B = 1$
- Case $B = 1$
- Results
- Propagation
- Computation of the threshold

4 The case $B \geq 2$

State dynamics

x_n: number of customers in the queue at time n.
$q_n = 1$ is service occurs, $q_n = 0$ if not, at time n.

Sequence of events (at each slot)

1. Begining of the slot
2. Admission in service
3. Impatience on remaining customers
4. Arrivals

State dynamics (ctd.)

The sequence of events leads to :

$$x_{n+1} \;=\; S\left([x_n - q_n B]^+\right) \;+\; A_{n+1} \;.$$

$S(x)$: the (random) number of "survivors" after impatience, out of x customers initially present.

$I(x)$: the number of impatient customers.
$\Longrightarrow$ binomially distributed random variables

Costs

The cost at step n is:

$$c_B q_n \; + \; c_L I([x_n - q_n B]^+) \; + \; c_H [x_n - q_n B]^+$$

Average Cost

$$c(x,q) \; = \; q\, c_B + (c_L\, \alpha + c_H)\, (x - qB)^+ \; = \; q\, c_B + c_Q\, (x - qB)^+ \, .$$

Optimization criterion:

$$v_\theta^\pi(x) \; = \; \mathbb{E}_x^\pi \left[\sum_{n=0}^{\infty} \theta^n \, c(x_n, q_n) \right] \; .$$

Dynamic programming equation

The optimal *value function* $V(x)$ is solution to:

The dynamic programming equation

$$V(x) = \min_{q \in \{0,1\}} \{c_B q + c_Q [x - Bq]^+ + \theta \mathbb{E} \left(V(S([x - Bq]^+) + A) \right) \}.$$

The optimal policy is *Markovian* and feedback: there exists a function of the state x, $d(x)$, such that

$$\pi^* = (d, d, \ldots, d, \ldots)$$

and $d(x)$ is given by:

The optimal policy

$$d(x) = \arg \min_{q \in \{0,1\}} \{...\}.$$

Progress

Introduction
The Model
The Question
The Literature
This talk

Dynamic Programming representation

$B = 1$
Results
Propagation
Computation of the threshold

$B \geq 2$

Questions

1 Introduction
- The Model
- The Question
- The Literature
- This talk

2 Dynamic Programming representation

3 The case $B = 1$
- Case $B = 1$
- Results
- Propagation
- Computation of the threshold

4 The case $B \geq 2$

Optimality Results

Scheduling with Impatience

Jean-Marie & Hyon

Introduction
The Model
The Question
The Literature
This talk

Dynamic Programming representation

$B = 1$
Results
Propagation
Computation of the threshold

$B \geq 2$

Questions

Theorem

The optimal policy is of threshold type: there exists a ν such that $d(x) = 1_{\{x \geq \nu\}}$.

Theorem

Let ψ be the number defined by

$$\psi \; = \; c_B \; - \; \frac{c_Q}{1 - \overline{\alpha}\theta} \; .$$

Then,

1. *If $\psi > 0$, the optimal threshold is $\nu = +\infty$.*
2. *If $\psi < 0$, the optimal threshold is $\nu = 1$.*
3. *If $\psi = 0$, any threshold $\nu \geq 1$ gives the same value.*

Method of Proof

Introduction
The Model
The Question
The Literature
This talk
Dynamic Programming representation
$B = 1$
Results
Propagation
Computation of the threshold
$B \geq 2$
Questions

Framework: propagation of properties through the dynamic programming operator (Puterman, Glasserman & Yao).

Requirement 1

$$\exists w(\cdot) \geq 0, \qquad \sup_{(x,q)} \frac{|c(x,q)|}{w(x)} \; < \; +\infty \; ,$$

$$\sup_{(x,q)} \frac{1}{w(x)} \sum_y \mathbb{P}(y|x,q) w(y) \; < \; +\infty \; ,$$

and $\forall \mu$, $0 \leq \mu < 1$, $\exists \eta$, $0 \leq \eta < 1$, $\exists J$, such that: $\forall$ J-uple of Markov Deterministic decision rules $\pi = (d_1, \ldots, d_J)$, and $\forall x$,

$$\mu^J \sum_y P_\pi(y|x) w(y) \; \leq \; \eta w(x) \; .$$

$\longrightarrow$ works with $w(x) = C + c_Q x$

Method of Proof

Framework: propagation of properties through the dynamic programming operator (Puterman, Glasserman & Yao).

Requirement 2

$\exists V^\sigma, \mathcal{D}^\sigma$

1. $v \in V^\sigma$ implies $Lv \in V^\sigma$,
2. $v \in V^\sigma$ implies there exists a decision d such that $d \in \mathcal{D}^\sigma \cap \arg\min_d L_d v$,
3. V^σ is a closed by simple convergence.

$\longrightarrow$ works with:

V^σ = { increasing and convex } and

$\mathcal{D}^\sigma$ = { monotone controls }

Propagation of structure

Theorem

Let, for any function v, $\tilde{v}(x) = \min_q Tv(x,q)$. Then:

1. *If v increasing, then $\tilde{v}$ increasing*
2. *If v increasing and convex, then $\tilde{v}$ increasing convex*

Theorem

If v is increasing and convex, then $Tv(x,q)$ is submodular over $\mathbb{N} \times \mathcal{Q}$. As a consequence, $x \mapsto \arg\min_q Tv(x,q)$ is increasing.

Submodularity (Topkis, Glasserman & Yao, Puterman)

g submodular if, for any $\overline{x} \geq \underline{x} \in \mathcal{X}$ and any $\overline{q} \geq \underline{q} \in \mathcal{Q}$:

$$g(\overline{x},\overline{q}) - g(\underline{x},\overline{q}) \leq g(\overline{x},\underline{q}) - g(\underline{x},\underline{q}).$$

Optimal Threshold / 1

The system under threshold ν evolves as:

$$x_{n+1} \;=\; R_\nu(x_n) \;:=\; S\left([x_n - 1_{\{x\geq\nu\}}]^+\right) \;+\; A_{n+1}\ .$$

A direct computation gives:

$$V_\nu(x) = \frac{c_Q}{1-\theta\overline{\alpha}}\left(x \;+\; \frac{\theta\lambda}{1-\theta}\right) \;+\; \psi\ \Phi(\nu,x)$$

$$\Phi(\nu,x) = \sum_{n=0}^{\infty} \theta^n \mathbb{P}(R_\nu^{(n)}(x) \geq \nu)$$

$$\psi = c_B \;-\; \frac{c_Q}{1-\overline{\alpha}\theta}\ .$$

Optimal Threshold / 2

Lemma

The function $\Phi(\nu, x)$ is decreasing in $\nu \geq 1$, for every x.

Proof by a coupling argument. If $O_\nu^{(n)}$ = set of customers present at time n under threshold ν, starting from $x_0 = x$:

Lemma

For every trajectory, we have

$$O_{\nu+1}^{(n)} = \begin{cases} \text{either} & O_\nu^{(n)} \\ \text{or} & O_\nu^{(n)} \dot\cup \{j_n\} \end{cases}$$

where j_n is the customer of smaller index in $O_{\nu+1}^{(n)}$.

$$\Longrightarrow \left\{ R_{\nu+1}^{(n)}(x) \geq \nu + 1 \right\} \subset \left\{ R_\nu^{(n)}(x) \geq \nu \right\} .$$

Progress

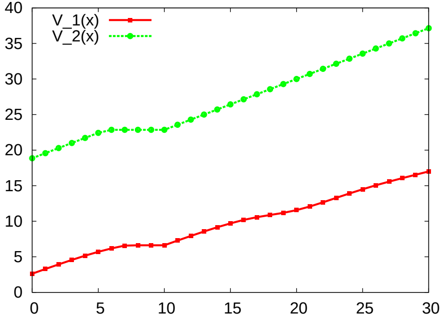
1. Introduction
   - The Model
   - The Question
   - The Literature
   - This talk
2. Dynamic Programming representation
3. The case $B = 1$
   - Case $B = 1$
   - Results
   - Propagation
   - Computation of the threshold
4. The case $B \geq 2$

What goes wrong when $B \geq 2$

Numerical experiments and exact results in special cases reveal that:

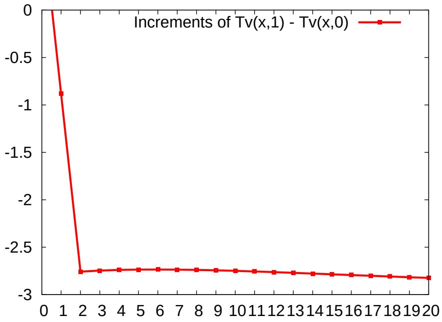
- The value function $V(x)$ is not convex in general
- The function $TV(x, q)$ is not submodular in general

Examples with $B = 10$, $\alpha = 1/10$, $\theta = 8/10$: V not convex



What goes wrong when $B \geq 2$, ctd.

Submodularity: if $Tv(x, 1)$ is submodular, then $x \mapsto Tv(x, 1) - Tv(x, 0)$ is decreasing.
A counterexample with $B = 2$, $\lambda = 1/10$, $\alpha = 9/10$, $\theta = 9/10$.



What goes wrong when $B \geq 2$, end.

Papadaki & Powell study the same problem without impatience.

Dynamics without impatience

$$x_{n+1} \;=\; [x_n - q_n B]^+ \;+\; A_{n+1} \;.$$

They show that the following “K-convexity” propagates:

K-convexity

$$V(x+K) - V(x) \;\geq\; V(x-1+K) - V(x-1) \;.$$

Also used in Altman & Koole for batch arrivals.

$\Longrightarrow$ does not work here.

Koole (2006) and Koçağa & Ward (2009) mention the incompatibility of impatience with structure theorems.

Questions?

Bibliography

The complete paper

E. Hyon and A. Jean-Marie.
Scheduling in a queuing system with impatience and setup costs.
Technical Report RR-6881, version 2, INRIA, Feb. 2010. In revision for *The Computer Journal*.

References on optimal Markovian control theory

M. Puterman.
Markov Decision Processes Discrete Stochastic Dynamic Programming.
Wiley, 2005.

P. Glasserman and D. Yao.
Monotone Structure in Discrete-Event Systems.
Wiley, 1994.

G. Koole.
Monotonicity in markov reward and decision chains: Theory and applications.
Foundation and Trends in Stochastic Systems, 1(1), 2006.

Bibliography (ctd)

References on the control of queues

R. K. Deb and R. F. Serfozo.
Optimal control of batch service queues.
Advances in Applied Probability, 5(2):340–361, 1973.

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

K. P. Papadaki and W. B. Powell.
Exploiting structure in adaptative dynamic programming algorithms for a stochastic batch service problem.
European Journal of Operational Research, 142:108–127, 2002.

Y. L. Koçağa and A. R. Ward.
Admission control for a multi-server queue with abandonment.
Forthcoming, *Queuing Systems*, July 2009.

Bibliography (end)

Control of queues with deadlines

P. P. Bhattacharya and A. Ephremides.
Optimal scheduling with strict deadlines.
IEEE Trans. Automatic Control, 34(7):721–728, July 1989.

D. Towsley and S. S. Panwar.
On the optimality of minimum laxity and earliest deadline scheduling for real-time multiprocessors.
In *Proceedings of IEEE EUROMICRO-90 Real Time Workshop*, pages 17–24, June 1990.

A. Movaghar.
Optimal control of parallel queues with impatient customers.
Performance Evaluation, 60:327–343, 2005.

Extensions to the model

Average case / no discount: $\theta = 1$.
$\implies$ should work as long as $\alpha \neq 0$ ($\overline{\alpha} \neq 1$)

Critical value:

$$\psi = c_B - c_Q \frac{1}{\alpha} = c_B - c_L - \frac{c_H}{\alpha} .$$

Branching processes: at each step, each customer is replaced by X customers. $\overline{\alpha} = \mathbb{E}X$, must be $\overline{\alpha} < \theta^{-1}$.
$\implies$ same formula for the optimal policy

Critical value:

$$\psi = c_B - \frac{c_Q}{1 - \overline{\alpha}\theta} .$$