A UNIFYING COMPUTATION OF WHITTLE'S INDEX FOR MARKOVIAN BANDITS

Manu K. Gupta[2]

Joint work with

U. Ayesta[1,2] & I.M. Verloop[1,2]

[1]Centre National de la Recherche Scientifique (CNRS),
[2]Institut de Recherche en Informatique de Toulouse (IRIT), Toulouse

Outline

Background and overview

- A particular case of constrained Markov Decision Process.
  - Stochastic resource allocation problem.

Background and overview

- A particular case of constrained Markov Decision Process.
  - Stochastic resource allocation problem.
- A generalization of multi-armed bandit problem (MABP).

Background and overview

- A particular case of constrained Markov Decision Process.
  - Stochastic resource allocation problem.
- A generalization of multi-armed bandit problem (MABP).
- Powerful modeling technique for diverse applications:
  - Routing in clusters (Niño-Mora, 2012a), sensor scheduling (Niño-Mora and Villar, 2011).
  - Machine repairman problem (Glazebrook et al., 2005), content delivery problem (Larrañaga et al., 2015)
  - Minimum job loss routing (Niño-Mora, 2012b), inventory routing (Archibald et al., 2009), processor sharing queues (Borkar and Pattathil, 2017), congestion control in TCP (Avrachenkov et al., 2013) etc.

Background and overview

- A particular case of constrained Markov Decision Process.
  - Stochastic resource allocation problem.
- A generalization of multi-armed bandit problem (MABP).
- Powerful modeling technique for diverse applications:
  - Routing in clusters (Niño-Mora, 2012a), sensor scheduling (Niño-Mora and Villar, 2011).
  - Machine repairman problem (Glazebrook et al., 2005), content delivery problem (Larrañaga et al., 2015)
  - Minimum job loss routing (Niño-Mora, 2012b), inventory routing (Archibald et al., 2009), processor sharing queues (Borkar and Pattathil, 2017), congestion control in TCP (Avrachenkov et al., 2013) etc.

Major challenges

- Establishing indexability and computations of Whittle's index.

Multi-armed bandit problem (MABP)

- A particular case of MDP.
- At each decision epoch, scheduler selects one *bandit*.

Multi-armed bandit problem (MABP)

- A particular case of MDP.
- At each decision epoch, scheduler selects one *bandit*.
- Selected bandit evolves *stochastically*, while the remaining bandits are *frozen*.

Multi-armed bandit problem (MABP)

- A particular case of MDP.
- At each decision epoch, scheduler selects one *bandit*.
- Selected bandit evolves *stochastically*, while the remaining bandits are *frozen*.
- States, rewards and transition probabilities are known.
- Objective is to maximize the total average reward.

Multi-armed bandit problem (MABP)

- A particular case of MDP.
- At each decision epoch, scheduler selects one *bandit*.
- Selected bandit evolves *stochastically*, while the remaining bandits are *frozen*.
- States, rewards and transition probabilities are known.
- Objective is to maximize the total average reward.
- In general, optimal policy depends on all the input parameters.

Multi-armed bandit problem (MABP)

- A particular case of MDP.
- At each decision epoch, scheduler selects one *bandit*.
- Selected bandit evolves *stochastically*, while the remaining bandits are *frozen*.
- States, rewards and transition probabilities are known.
- Objective is to maximize the total average reward.
- In general, optimal policy depends on all the input parameters.

Gittin's index

- For MABP, optimal policy is an index rule (Gittins et al., 2011).
- For example, $c\mu$ rule in multi-class queues.

Restless Bandit Problem (RBP)

- RBP is a generalization of MABP.
  - Any number of bandits (more than 1) can be made active.
  - All bandits might evolve *stochastically*.

Restless Bandit Problem (RBP)

- RBP is a generalization of MABP.
  - Any number of bandits (more than 1) can be made active.
  - All bandits might evolve *stochastically*.
- Objective is to optimize the average performance criterion.
- Computing optimal policy is typically out of reach.

Restless Bandit Problem (RBP)

- RBP is a generalization of MABP.
  - Any number of bandits (more than 1) can be made active.
  - All bandits might evolve *stochastically*.
- Objective is to optimize the average performance criterion.
- Computing optimal policy is typically out of reach.
  - RBPs are *PSPACE-complete* (Papadimitriou and Tsitsiklis, 1999).
  - Much more convincing evidence of intractability than NP-hardness.

Restless Bandit Problem (RBP)

- RBP is a generalization of MABP.
  - Any number of bandits (more than 1) can be made active.
  - All bandits might evolve *stochastically*.
- Objective is to optimize the average performance criterion.
- Computing optimal policy is typically out of reach.
  - RBPs are *PSPACE-complete* (Papadimitriou and Tsitsiklis, 1999).
  - Much more convincing evidence of intractability than NP-hardness.

Whittle's relaxation (Whittle, 1988)

Restriction on number of active bandits to be respected on *average* only.

- Optimal solution to the relaxed problem is of index type.
- The Whittle's index recovers Gittin's index for non-restless case.

Whittle's index policy

- A heuristic for the original problem.
  - A bandit with the highest Whittle's index is made active.

Whittle's index policy

- A heuristic for the original problem.
  - A bandit with the highest Whittle's index is made active.
- Whittle's index policy performs strikingly well (Niño-Mora, 2007).

Whittle's index policy

- A heuristic for the original problem.
  - A bandit with the highest Whittle's index is made active.
- Whittle's index policy performs strikingly well (Niño-Mora, 2007).
- Asymptotically optimal under certain conditions (Weber and Weiss, 1990, 1991).
  - A generalization to several classes of bandits, arrivals of new bandits and multiple actions (Verloop, 2016).

Whittle's index policy

- A heuristic for the original problem.
  - A bandit with the highest Whittle's index is made active.
- Whittle's index policy performs strikingly well (Niño-Mora, 2007).
- Asymptotically optimal under certain conditions (Weber and Weiss, 1990, 1991).
  - A generalization to several classes of bandits, arrivals of new bandits and multiple actions (Verloop, 2016).

Results

- A unifying framework for obtaining Whittle's index.
- Retrieve many available Whittle's indices in literature including machine repairman problem, content delivery problem etc.

Model description and notations

$$\begin{aligned}
K &: \text{Number of ongoing projects or bandits.}\\
a &: \text{Binary action to make the bandit active or passive.}\\
\phi &: \text{The policy to make a bandit active or passive.}\\
N_k^{\phi}(t) &: \text{State of bandit } k \text{ at time } t \text{ under policy } \phi.\\
S_k^{\phi}(\vec{N}^{\phi}(t)) \in \{0,1\} &: \text{Whether or not bandit } k \text{ is made active at time } t.\\
C_k(n,a) &: \text{Cost per unit of time when bandit } k \text{ is in state } n.\\
C_k^{\infty}(x,y,a) &: \text{The lump cost for bandit } k \text{ when state instanteneously}\\
&\quad \text{changes form } x \text{ to } y \text{ under action } a.
\end{aligned}$$

- Each bandit is modeled as continuous time Markov chain.
- Both *finite* and *infinite* transition rates are allowed.

Objective

To minimize the long-run average cost:

$$\mathcal{C}^{\phi} := \limsup_{T\to\infty} \sum_{k=1}^{K} \frac{1}{T}\mathbb{E}\left(\int_0^T C_k(N_k^{\phi}(t), S_k^{\phi}(\vec{N}^{\phi}(t))) \right.$$

$$\left. + \sum_{\tilde{N}_k} C_k^{\infty}(\tilde{N}_k, y, S_k^{\phi}(\vec{N}_{-k}^{\phi}(t)))q_k^{S_k^{\phi}(\vec{N}_k^{\phi}(t))}(N_k^{\phi}(t), \tilde{N}_k)\mathcal{I}_k(\tilde{N}_k, S_k^{\phi}(\vec{N}_{-k}^{\phi}(t))dt \right) \tag{1}$$

where $q_k^a(n,\tilde{n})$ be the transition rate of going from state n to $\tilde{n}$ under action a and

$$\mathcal{I}_k(n,a) = \begin{cases} 1 \text{ if bandit } k \text{ results in infinite transition rates} \\ 0 \text{ otherwise} \end{cases}$$

Hard constraint

$$\sum_{k=1}^{K} f_k(N_k^{\phi}, S_k^{\phi}(\vec{N})) \leq M. \tag{2}$$

Hard constraint

$$\sum_{k=1}^{K} f_k(N_k^{\phi}, S_k^{\phi}(\vec{N})) \leq M. \tag{2}$$

- If $f_k(N_k^{\phi}, S_k^{\phi}(\vec{N})) = S_k^{\phi}(\vec{N})$, constraint (2) implies $\sum_{k=1}^{K} S_k^{\phi}(\vec{N}) \leq M$.
  - Standard restless bandit constraint.

Hard constraint

$$\sum_{k=1}^{K} f_k(N_k^{\phi}, S_k^{\phi}(\vec{N})) \leq M. \tag{2}$$

- If $f_k(N_k^{\phi}, S_k^{\phi}(\vec{N})) = S_k^{\phi}(\vec{N})$, constraint (2) implies $\sum_{k=1}^{K} S_k^{\phi}(\vec{N}) \leq M$.
  - Standard restless bandit constraint.
- If $f_k(N_k^{\phi}, S_k^{\phi}(\vec{N})) = N_k^{\phi} S_k^{\phi}(\vec{N})$, constraint (2) implies $\sum_{k=1}^{K} N_k^{\phi} S_k^{\phi}(\vec{N}) \leq M$.
  - Buffer constraint for TCP (Avrachenkov et al., 2013).

Hard constraint

$$\sum_{k=1}^{K} f_k(N_k^{\phi}, S_k^{\phi}(\vec{N})) \leq M. \tag{2}$$

- If $f_k(N_k^{\phi}, S_k^{\phi}(\vec{N})) = S_k^{\phi}(\vec{N})$, constraint (2) implies $\sum_{k=1}^{K} S_k^{\phi}(\vec{N}) \leq M$.
  - Standard restless bandit constraint.
- If $f_k(N_k^{\phi}, S_k^{\phi}(\vec{N})) = N_k^{\phi} S_k^{\phi}(\vec{N})$, constraint (2) implies $\sum_{k=1}^{K} N_k^{\phi} S_k^{\phi}(\vec{N}) \leq M$.
  - Buffer constraint for TCP (Avrachenkov et al., 2013).
- $f_k(N_k^{\phi}, S_k^{\phi}(\vec{N}))$ represents the capacity occupation (volume) in state N_k^{ϕ} under action $S_k^{\phi}(\vec{N})$.
  - Family of sample path knapsack capacity allocation constraint (Jacko, 2016; Graczová and Jacko, 2014).

Finite transition rates

The transitions rates of vector $\vec{N} = (N_1, N_2, ..., N_K)$ are:

$$\begin{cases} \vec{N} \to \vec{N} + \vec{e}_k & \text{with transition rate } b_k^a(N_k) \\ \vec{N} \to \vec{N} - \vec{e}_k & \text{with transition rate } d_k^a(N_k) \\ \vec{N} \to \vec{N} + \alpha^a(n_k)\vec{e}_k & \text{with transition rate } h_k^a(N_k) \\ \vec{N} \to \vec{N} - \beta^a(n_k)\vec{e}_k & \text{with transition rate } l_k^a(N_k), \end{cases}$$

The long run average cost:

$$\mathcal{C}^{\phi} = \limsup_{T\to\infty} \sum_{k=1}^{K} \frac{1}{T}\mathbb{E}\left(\int_0^T C_k(N_k^{\phi}(t), S_k^{\phi}(\vec{N}^{\phi}(t)))dt\right) \tag{3}$$

Finite transition rates

The transitions rates of vector $\vec{N} = (N_1, N_2, ..., N_K)$ are:

$$\begin{cases} \vec{N} \to \vec{N} + \vec{e}_k & \text{with transition rate } b_k^a(N_k) \\ \vec{N} \to \vec{N} - \vec{e}_k & \text{with transition rate } d_k^a(N_k) \\ \vec{N} \to \vec{N} + \alpha^a(n_k)\vec{e}_k & \text{with transition rate } h_k^a(N_k) \\ \vec{N} \to \vec{N} - \beta^a(n_k)\vec{e}_k & \text{with transition rate } l_k^a(N_k), \end{cases}$$

The long run average cost:

$$\mathcal{C}^{\phi} = \limsup_{T\to\infty} \sum_{k=1}^{K} \frac{1}{T}\mathbb{E}\left(\int_0^T C_k(N_k^{\phi}(t), S_k^{\phi}(\vec{N}^{\phi}(t)))dt\right) \tag{3}$$

Machine repairman problem, class selection problem, load balancing problem.

Examples of finite transition rates

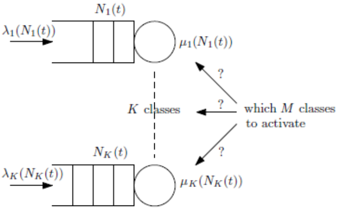

Figure: Class selection problem

Examples of finite transition rates

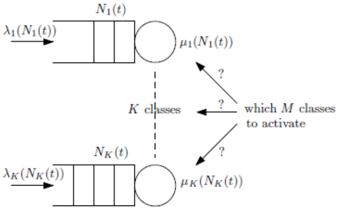

Figure: Class selection problem

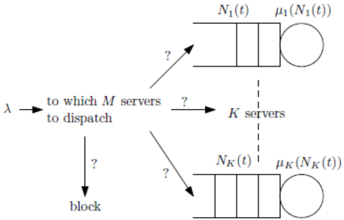

Figure: Load balancing problem

Examples of finite transition rates

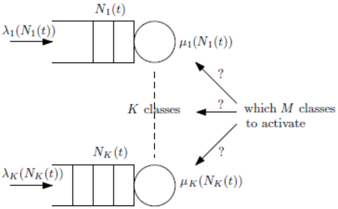

Figure: Class selection problem

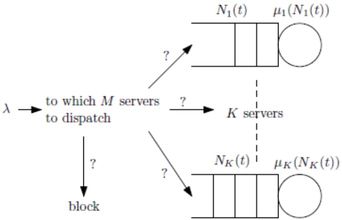

Figure: Load balancing problem

- Machine repairman problem (Glazebrook et al., 2005)
  - M machines to be repaired by R repairmen.

Examples of finite transition rates

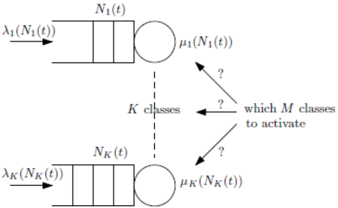

Figure: Class selection problem

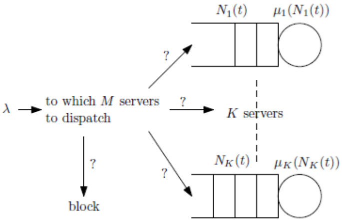

Figure: Load balancing problem

- Machine repairman problem (Glazebrook et al., 2005)
  - M machines to be repaired by R repairmen.
- Load balancing problem (Argon et al., 2009)
  - With dedicated arrivals to each queues.

Infinite transition rates

The transition rates of vector $\vec{N} = (N_1, N_2, ..., N_K)$ for this case are:

$$\begin{cases} \vec{N} \rightarrow \vec{N} + \vec{e}_k & \text{with transition rate } b_k^a(N_k) \\ \vec{N} \rightarrow \vec{N} - \vec{e}_k & \text{with transition rate } d_k^a(N_k) \\ \vec{N} \rightarrow \vec{N} + \alpha^a(n_k)\vec{e}_k & \text{with transition rate } h_k^a(N_k) \\ \vec{N} \rightarrow \vec{N} - \beta^a(n_k)\vec{e}_k & \text{with transition rate } l_k^a(N_k) \\ \vec{N} \rightarrow \vec{N} + \gamma^a(n_k)\vec{e}_k & \text{with impulse rate } \tilde{h}_k^a(N_k), \\ \vec{N} \rightarrow \vec{N} - \delta^a(n_k)\vec{e}_k & \text{with impulse rate } \tilde{l}_k^a(N_k), \end{cases}$$

Infinite transition rates

The transition rates of vector $\vec{N} = (N_1, N_2, ..., N_K)$ for this case are:

$$\begin{cases} \vec{N} \to \vec{N} + \vec{e}_k & \text{with transition rate } b_k^a(N_k) \\ \vec{N} \to \vec{N} - \vec{e}_k & \text{with transition rate } d_k^a(N_k) \\ \vec{N} \to \vec{N} + \alpha^a(n_k)\vec{e}_k & \text{with transition rate } h_k^a(N_k) \\ \vec{N} \to \vec{N} - \beta^a(n_k)\vec{e}_k & \text{with transition rate } l_k^a(N_k) \\ \vec{N} \to \vec{N} + \gamma^a(n_k)\vec{e}_k & \text{with impulse rate } \tilde{h}_k^a(N_k), \\ \vec{N} \to \vec{N} - \delta^a(n_k)\vec{e}_k & \text{with impulse rate } \tilde{l}_k^a(N_k), \end{cases}$$

- Content delivery problem (Larrañaga et al., 2015).
- bottleneck router in TCP etc (Avrachenkov et al., 2013).

Infinite transition rates

The transition rates of vector $\vec{N} = (N_1, N_2, ..., N_K)$ for this case are:

$$\begin{cases} \vec{N} \to \vec{N} + \vec{e}_k & \text{with transition rate } b_k^a(N_k) \\ \vec{N} \to \vec{N} - \vec{e}_k & \text{with transition rate } d_k^a(N_k) \\ \vec{N} \to \vec{N} + \alpha^a(n_k)\vec{e}_k & \text{with transition rate } h_k^a(N_k) \\ \vec{N} \to \vec{N} - \beta^a(n_k)\vec{e}_k & \text{with transition rate } l_k^a(N_k) \\ \vec{N} \to \vec{N} + \gamma^a(n_k)\vec{e}_k & \text{with impulse rate } \tilde{h}_k^a(N_k), \\ \vec{N} \to \vec{N} - \delta^a(n_k)\vec{e}_k & \text{with impulse rate } \tilde{l}_k^a(N_k), \end{cases}$$

- Content delivery problem (Larrañaga et al., 2015).
- bottleneck router in TCP etc (Avrachenkov et al., 2013).
- Instantaneous change in state.

Lagrangian Relaxation

$$\limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left(\int_0^T \sum_{k=1}^{K} f_k(N_k^{\phi}(t), S_k^{\phi}(\vec{N}^{\phi}(t)))dt\right) \leq M \text{ (On average)} \quad (4)$$

The unconstrained problem is to find a policy ϕ that minimizes

$$C^{\phi}(W) = \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left(\int_0^T \left(\sum_{k=1}^{K} C_k(N_k^{\phi}(t), S_k^{\phi}(\vec{N}^{\phi}(t)))\right.\right.$$

$$+\sum_{\tilde{N}_k} C_k^{\infty}(\tilde{N}_k, y, S_k^{\phi}(\vec{N}_{-k}^{\phi}(t)))q_k^{S_k^{\phi}(\vec{N}_k^{\phi}(t))}(N_k^{\phi}(t), \tilde{N}_k)\mathcal{I}_k(\tilde{N}_k, S_k^{\phi}(\vec{N}_{-k}^{\phi}(t))$$

$$\left.\left.-W\left(\sum_{k=1}^{K} f_k(N_k^{\phi}(t), S_k^{\phi}(\vec{N}^{\phi}(t))) - M\right) dt\right)\right), \quad (5)$$

Decomposition

The problem can be decomposed (key observation in Whittle (1988)):

$$\mathbb{E}(C_k(N_k^\phi, S_k^\phi(N_k^\phi))) + \sum_{\tilde{N}_k} \mathbb{E}\left(C_k^\infty(\tilde{N}_k, y, S_k^\phi(N_k^\phi)) q_k^{S_k^\phi(N_k^\phi)}(N_k^\phi, \tilde{N}_k) \mathcal{I}_k(\tilde{N}_k, S_k^\phi(\tilde{N}_k)) \right) - W\mathbb{E}(f_k(N_k^\phi, S_k^\phi(N_k^\phi)) \tag{6}$$

- The solution to the relaxed problem:
  - Combining the solution of K separate problems.

Decomposition

The problem can be decomposed (key observation in Whittle (1988)):

$$\mathbb{E}(C_k(N_k^\phi, S_k^\phi(N_k^\phi))) +$$
$$\sum_{\tilde{N}_k} \mathbb{E}\left(C_k^\infty(\tilde{N}_k, y, S_k^\phi(N_k^\phi)) q_k^{S_k^\phi(N_k^\phi)}(N_k^\phi, \tilde{N}_k) \mathcal{I}_k(\tilde{N}_k, S_k^\phi(\tilde{N}_k)) \right) - W\mathbb{E}(f_k(N_k^\phi, S_k^\phi(N_k^\phi)) \tag{6}$$

- The solution to the relaxed problem:
  - Combining the solution of K separate problems.
- The decomposed problem is an MDP.
  - The optimal policy is the solution of the dynamic programming equations.

Decomposition

The problem can be decomposed (key observation in Whittle (1988)):

$$\mathbb{E}(C_k(N_k^\phi, S_k^\phi(N_k^\phi))) +$$
$$\sum_{\tilde{N}_k} \mathbb{E}\left(C_k^\infty(\tilde{N}_k, y, S_k^\phi(N_k^\phi)) q_k^{S_k^\phi(N_k^\phi)}(N_k^\phi, \tilde{N}_k) \mathcal{I}_k(\tilde{N}_k, S_k^\phi(\tilde{N}_k)) \right) - W\mathbb{E}(f_k(N_k^\phi, S_k^\phi(N_k^\phi)) \tag{6}$$

- The solution to the relaxed problem:
  - Combining the solution of K separate problems.
- The decomposed problem is an MDP.
  - The optimal policy is the solution of the dynamic programming equations.
- Indexability and Whittle's index.

Indexability[1]

Definition 1.

A bandit is indexable if the set of states in which passive is an optimal action in (6) (denoted by $D_k(W)$) increases in W, that is,
$W' < W \Rightarrow D_k(W') \subseteq D_k(W)$.

[1] Peter Whittle. Restless bandits: Activity allocation in a changing world. *Journal of applied probability*, 25 (A):287-298, 1988.

Indexability[1]

Definition 1.

A bandit is indexable if the set of states in which passive is an optimal action in (6) (denoted by $D_k(W)$) increases in W, that is, $W' < W \Rightarrow D_k(W') \subseteq D_k(W)$.

- A natural definition.
- Usually difficult to establish and sometimes doesn't hold.

[1] Peter Whittle. Restless bandits: Activity allocation in a changing world. *Journal of applied probability*, 25 (A):287-298, 1988.

Whittle's index[2]

If indexability is satisfied, Whittle's index in state N_k is defined as follows:

Definition 2.

The smallest value of W such that an optimal policy for (6) is indifferent of the action in state n_k. The Whittle's index is denoted by $W_k(n_k)$.

[2] Peter Whittle. Restless bandits: Activity allocation in a changing world. *Journal of applied probability*, 25 (A):287-298, 1988.

Whittle's index[2]

If indexability is satisfied, Whittle's index in state N_k is defined as follows:

Definition 2.

The smallest value of W such that an optimal policy for (6) is indifferent of the action in state n_k. The Whittle's index is denoted by $W_k(n_k)$.

- Optimal solution to relaxed problem
  - To activate all bandits that are in a state n_k such that $W_k(n_k) > W$.

[2] Peter Whittle. Restless bandits: Activity allocation in a changing world. *Journal of applied probability*, 25 (A):287-298, 1988.

Whittle's index[2]

If indexability is satisfied, Whittle's index in state N_k is defined as follows:

Definition 2.

The smallest value of W such that an optimal policy for (6) is indifferent of the action in state n_k. The Whittle's index is denoted by $W_k(n_k)$.

- Optimal solution to relaxed problem
  - To activate all bandits that are in a state n_k such that $W_k(n_k) > W$.
- Optimality of Whittle's index policy for single armed restless bandits.
  - Relaxation and original constraint will give the same set of policies.

[2] Peter Whittle. Restless bandits: Activity allocation in a changing world. *Journal of applied probability*, 25 (A):287-298, 1988.

Monotone policies

Definition 3.

There is a threshold $n_k(W)$ such that when bandit k is in a state $m_k \leq n_k(W)$, then action a is optimal, and otherwise action a' is optimal, $a,\ a' \in \{0, 1\}$ and $a \neq a'$.

Monotone policies

Definition 3.

There is a threshold $n_k(W)$ such that when bandit k is in a state $m_k \leq n_k(W)$, then action a is optimal, and otherwise action a' is optimal, $a,\ a' \in \{0, 1\}$ and $a \neq a'$.

- A policy $\phi = n$ denotes a threshold policy with threshold n,
  - 0-1 type if $a = 0$ and $a' = 1$
  - 1-0 type if $a = 1$ and $a' = 0$
- For certain problems, optimal solution of problem (6) is of threshold type.

Closed form expression for Whittle's index

Theorem 1.

Assume an optimal solution of (6) is of threshold type, and $\mathbb{E}(f_k(N_k^n, S_k^n(N_k^n)))$ is strictly increasing in n. Then, bandit k is indexable.
If the structure of an optimal solution of problem (6) is of 0-1 type, then, in case

$$\frac{F_k^n(N_k^n, S_k^n(N_k^n)) - F_k^{n-1}(N_k^{n-1}, S_k^{n-1}(N_k^{n-1}))}{\mathbb{E}(f_k(N_k^n, S_k^n(N_k^n))) - \mathbb{E}(f_k(N_k^{n-1}, S_k^{n-1}(N_k^{n-1})))} \tag{7}$$

is non-decreasing in n, Whittles index $W_k(n_k)$ is given by (7) and is hence non-decreasing. Similarly, if the structure of an optimal solution of problem (6) is of 1-0 type, then, in case (7) is non-decreasing in n, $-W_k(n_k)$ is given by (7) and hence Whittles index is non-increasing.

$F_k^n(N_k^n, S_k^n(N_k^n))$ is the expected cost under the threshold policy n for bandit k.

Optimality of threshold policies

Proposition 1.

Consider the finite transition rates and assume

$$\begin{aligned} b_k^a(N_k) &= \lambda_k^0(n_k)(1-a) \\ d_k^a(N_k) &= \mu_k^1(n_k)a + \mu_k^0(n_k)(1-a) \\ h_k^a(N_k) &= 0 \\ l_k^a(N_k) &= l_k^1(n_k)a + l_k^0(n_k)(1-a) \end{aligned}$$

Then there exists an $n_k \in \{-1, 0, 1, \ldots\}$ such that a 0-1 type of threshold policy, with threshold n_k, optimally solves problem (6).

Details of the proof

1-0 type policies

If instead,

$$
\begin{aligned}
b_k^a(N_k) &= \lambda_k^1(n_k)a + \lambda_k^0(n_k)(1-a) \\
d_k^a(N_k) &= \mu_k^0(n_k)(1-a) \\
h_k^a(N_k) &= h_k^1(n_k)a + h_k^0(n_k)(1-a) \\
l_k^a(N_k) &= 0
\end{aligned}
$$

Then there exists an $n_k \in \{-1, 0, 1, ...\}$ such that a 1-0 type of threshold policy, with threshold n_k, optimally solves problem (6).

Infinite transition rates

Proposition 2.

Consider the infinite transition rates and assume

$$
\begin{aligned}
b_k^a(N_k) &= \lambda_k^0(n_k)(1-a)\\
d_k^a(N_k) &= \mu_k^0(n_k)(1-a)\\
h_k^a(N_k) &= 0\\
l_k^a(N_k) &= l_k^0(n_k)(1-a)\\
\tilde{h}_k^a(N_k) &= 0\\
\tilde{l}_k^a(N_k) &= \infty \textit{ for } a = 1 \textit{ (and 0 otherwise)}
\end{aligned}
$$

Then there exists an $n_k \in \{-1, 0, 1, \ldots\}$ *such that a 0-1 type of threshold policy, with threshold* n_k*, optimally solves problem (6).*

Applications

- Machine repairman problem
- Content delivery problem
- Congestion control in TCP flows

Machine Repairman problem

$$\begin{aligned} M &: \text{ Non-identical Machines} \\ R &: \text{Number of repairmans}, R \leq M \\ X_k(t) &: \text{The state of machine } k \end{aligned}$$

Machine Repairman problem

$$
\begin{aligned}
M &: \text{ Non-identical Machines} \\
R &: \text{Number of repairmans}, R \leq M \\
X_k(t) &: \text{The state of machine } k
\end{aligned}
$$

- States of the machine are the degree of deterioration.

Machine Repairman problem

$$\begin{aligned} M &: \text{ Non-identical Machines} \\ R &: \text{Number of repairmans}, R \leq M \\ X_k(t) &: \text{The state of machine } k \end{aligned}$$

- States of the machine are the degree of deterioration.
- Action $a = 1$ (use the repairman)
  - State improves.
  - Machine is returned to pristine state 0.

Machine Repairman problem

$$\begin{aligned} M &: \text{ Non-identical Machines} \\ R &: \text{Number of repairmans}, R \leq M \\ X_k(t) &: \text{The state of machine } k \end{aligned}$$

- States of the machine are the degree of deterioration.
- Action $a = 1$ (use the repairman)
  - State improves.
  - Machine is returned to pristine state 0.
- Action $a = 0$
  - State further deteriorates.
  - Machine spends a random amount of time in its current damage state before deteriorating to the next one.

Machine repairman problem

- Possibility of a catastrophic breakdown with rate $\psi_k(n_k)$
- Repair rates be $r_k(n_k)$ from state n_k.
- Deterioration rates be $\lambda_k(n_k)$.

Machine repairman problem

- Possibility of a catastrophic breakdown with rate $\psi_k(n_k)$
- Repair rates be $r_k(n_k)$ from state n_k.
- Deterioration rates be $\lambda_k(n_k)$.

$$C_k^b(n_k,0) : \text{Huge lump cost for breakdown.}$$
$$C_k^r(n_k,1) : \text{Cost of using the repairman.}$$
$$C_k^{pd}(n_k,0) : \text{Per unit cost of deterioration.}$$

$$C_k^b(n_k,0) >> C_k^r(n_k,1)$$

Machine repairman problem

- Possibility of a catastrophic breakdown with rate $\psi_k(n_k)$
- Repair rates be $r_k(n_k)$ from state n_k.
- Deterioration rates be $\lambda_k(n_k)$.

$$C_k^b(n_k, 0) : \text{Huge lump cost for breakdown.}$$
$$C_k^r(n_k, 1) : \text{Cost of using the repairman.}$$
$$C_k^{pd}(n_k, 0) : \text{Per unit cost of deterioration.}$$

$$C_k^b(n_k, 0) >> C_k^r(n_k, 1)$$

Objective

To deploy the repairmen to minimize the total expected cost.

The Markov decision process is characterized by the following transition rates:

$$
\begin{aligned}
b_k^a(N_k) &= \lambda_k(n_k)(1-a) \\
d_k^a(N_k) &= 0 \\
h_k^a(N_k) &= 0 \\
l_k^a(N_k) &= r_k(n_k)a + \psi_k(n_k)(1-a)
\end{aligned}
$$

$$
f_k(N_k^\phi, S_k^\phi(\vec{N})) = S_k^\phi(\vec{N})
$$

The Markov decision process is characterized by the following transition rates:

$$\begin{aligned} b_k^a(N_k) &= \lambda_k(n_k)(1-a) \\ d_k^a(N_k) &= 0 \\ h_k^a(N_k) &= 0 \\ l_k^a(N_k) &= r_k(n_k)a + \psi_k(n_k)(1-a) \end{aligned}$$

$$f_k(N_k^{\phi}, S_k^{\phi}(\vec{N})) = S_k^{\phi}(\vec{N})$$

Threshold optimality

0-1 type of threshold policy is optimal.

Dynamics of a bandit in machine repairman problem

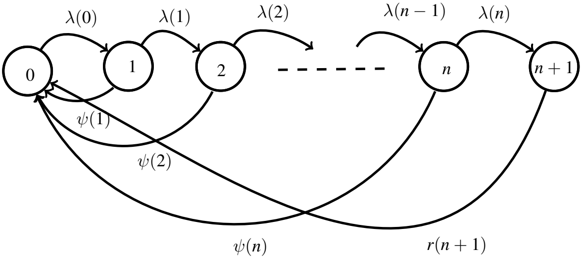

Figure: Transition diagram for threshold policy 'n' of machine repairman problem

Indexabiliity

Lemma 1.

Machine k is indexable if repair rates are non-decreasing in its state, i.e., $r_k(n_k) \leq r_k(n_k+1)\ \forall\ n_k$. In particular, all machines are indexable for state independent repair rates.

Indexabiliity

Lemma 1.

Machine k is indexable if repair rates are non-decreasing in its state, i.e., $r_k(n_k) \leq r_k(n_k+1)\ \forall\ n_k$. In particular, all machines are indexable for state independent repair rates.

- Follows from Theorem 1.
  - $\mathbb{E}(f_k(N_k^{n_k}, S_k^{n_k}(N_k^{n_k})))$ is strictly increasing in n_k.
- Equivalently, $\sum_{m=0}^{n_k} \pi_k^{n_k}(m)$ is strictly increasing in n_k for machine repairman problem.

Details of the proof

Whittle's index

Lemma 2.

The Whittle's index, $W_k(n)$, for machine k is given by

$$\frac{\left(C_{Sum}(n)+C_k^r(n+1,1)P_n\right)\left(P_{Sum}(n-1)+\frac{P_{n-1}}{r_k(n)}\right)-\left(C_{Sum}(n-1)+C_k^r(n,1)P_{n-1}\right)\left(P_{Sum}(n)+\frac{P_n}{r_k(n+1)}\right)}{\frac{P_{n-1}}{r_k(n)}\sum_{i=0}^{n}\frac{P_i}{\lambda_k(i)}-\frac{P_n}{r_k(n+1)}\sum_{i=0}^{n-1}\frac{P_i}{\lambda_k(i)}} \quad (8)$$

where $C_{Sum}(n)=\sum_{i=1}^{n}\left[(P_{i-1}-P_i)C_k^b(i,0)+\frac{P_iC_k^{pd}(i,0)}{\lambda_k(i)}\right]$, $P_{Sum}(n)=\sum_{i=0}^{n}\frac{P_i}{\lambda_k(i)}$,

$P_i=\prod_{j=1}^{i}p_k(j)$, $p_k(j)=\frac{\lambda_k(j)}{\lambda_k(j)+\psi_k(j)}$ *and* $P_0=1$*, if (8) is non-decreasing in n.*

- Follows from Theorem 1.

Details of the proof

Model 1: Deterioration cost per unit

- No breakdowns: $\psi_k(n_k) = 0$ and $C_k^b(n_k, 0) = 0$

Model 1: Deterioration cost per unit

- No breakdowns: $\psi_k(n_k) = 0$ and $C_k^b(n_k, 0) = 0$
- Repair cost and repair rates are state independent: $C_k^{pr}(n, 1) = C_k^{pr}(n+1, 1) = C_k^{pr}, r_k(n) = r_k(n+1) = r_k$

Model 1: Deterioration cost per unit

- No breakdowns: $\psi_k(n_k) = 0$ and $C_k^b(n_k, 0) = 0$
- Repair cost and repair rates are state independent: $C_k^{pr}(n,1) = C_k^{pr}(n+1,1) = C_k^{pr}, r_k(n) = r_k(n+1) = r_k$

Corollary 1.

If $C_k^{pd}(n,0)$, deterioration cost, is an increasing sequence, then, all machines are indexable and Whittle's index is given by

$$W_k(n) = r_k \left[\sum_{i=0}^{n-1} \frac{C_k^{pd}(n,0) - C_k^{pd}(i,0)}{\lambda_k(i)} + \frac{C_k^{pd}(n,0) - C_k^{pr}}{r_k} \right] \tag{9}$$

Model 1: Deterioration cost per unit

Discrete time analogy

- For $1/r_k = 1$, we recover the index for average cost criterion in discrete time (see Equation (19) in Glazebrook et al. (2005)).

Model 1: Deterioration cost per unit

Discrete time analogy

- For $1/r_k = 1$, we recover the index for average cost criterion in discrete time (see Equation (19) in Glazebrook et al. (2005)).
- For $1/r_k = 1$, $C_k^{pr} = 0$ and $C_k^{pd}(n, 0) = C_k n$, we get the index which is consistent with the result of Whittle's approximate evaluation (See ch. 14.6 in Whittle (1996)).

Model 2: Lump cost for breakdown

- No deterioration cost: $C_k^{pd}(n_k, 0) = 0$
- Repair cost, breakdown cost and repair rates are state independent: $C_k^r(n, 1) = R_k, C_k^b(n, 0) = B_k, r_k(n) = r_k(n+1) = r_k$

Model 2: Lump cost for breakdown

- No deterioration cost: $C_k^{pd}(n_k, 0) = 0$
- Repair cost, breakdown cost and repair rates are state independent: $C_k^r(n, 1) = R_k, C_k^b(n, 0) = B_k, r_k(n) = r_k(n+1) = r_k$

Corollary 2.

If $\psi_k(n)$ is an increasing sequence, then, all machines are indexable and Whittle's index is given by

$$W_k(n) = \frac{B_k \left(\frac{1-p_k(n)}{r} - \frac{p_k(n)}{\lambda_k(n)} + \sum_{i=0}^{n} \frac{P_i}{\lambda_k(i)} - p_k(n) \sum_{i=0}^{n-1} \frac{P_i}{\lambda_k(i)} \right)}{\frac{1}{r_k} \left(\sum_{i=0}^{n} \frac{P_i}{\lambda_k(i)} - p_k(n) \sum_{i=0}^{n-1} \frac{P_i}{\lambda_k(i)} \right)} - R_k \qquad (10)$$

Discrete time analogy

Expected time to change the state is 1. Mathematically,

$$\frac{1}{r_k} = 1 \text{ and } \frac{1}{\lambda_k(i) + \psi_k(i)} = 1 \ \forall\ i,$$

We recover the index for average cost criterion in discrete time (see Equation (48) in Glazebrook et al. (2005)).

Discrete time analogy

Expected time to change the state is 1. Mathematically,

$$\frac{1}{r_k} = 1 \text{ and } \frac{1}{\lambda_k(i) + \psi_k(i)} = 1 \ \forall\, i,$$

We recover the index for average cost criterion in discrete time (see Equation (48) in Glazebrook et al. (2005)).

- Cost structure
  - Constant in Glazebrook et al. (2005) for model 2.
  - Linear in Whittle (1996) for model 1.

Discrete time analogy

Expected time to change the state is 1. Mathematically,

$$\frac{1}{r_k} = 1 \text{ and } \frac{1}{\lambda_k(i) + \psi_k(i)} = 1 \; \forall \, i,$$

We recover the index for average cost criterion in discrete time (see Equation (48) in Glazebrook et al. (2005)).

- Cost structure
  - Constant in Glazebrook et al. (2005) for model 2.
  - Linear in Whittle (1996) for model 1.
- Disadvantages of constant or linear cost (Van Mieghem (1995), Ansell et al. (1999)).
  - Index for a state dependent cost structure in continuous time.

Content delivery problem

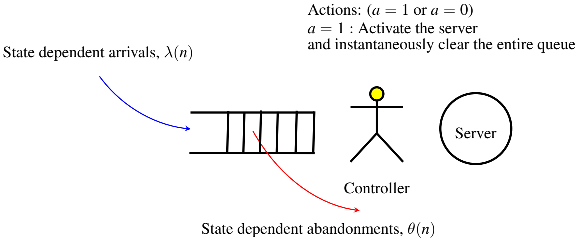

Figure: Optimal clearing framework as single-armed restless bandit

Efficient content delivery (Larrañaga et al., 2015)

- Bulk of traffic is delay tolerant (software updates, video content etc.).
- Requests can be delayed and grouped.

Content delivery problem

$C^h(n)$: State dependent holding cost per unit of time jobs held in the queue.

$C^a(n)$: State dependent abandonment cost for the jobs abandoning the queue.

$C_s^\infty(n)$: Set-up (lump) cost of clearing the batch of size n.

Content delivery problem

$C^h(n)$: State dependent holding cost per unit of time jobs held in the queue.

$C^a(n)$: State dependent abandonment cost for the jobs abandoning the queue.

$C_s^\infty(n)$: Set-up (lump) cost of clearing the batch of size n.

Objective

- To minimize the average cost.
  - To balance the gains against the risk of not meeting the deadline.

Content delivery problem

This Markov decision process is characterized by the following transitions:

$$
\begin{aligned}
b^a(N) &= \lambda(n) \\
d^a(N) &= \theta(n) \\
h^a(N) &= 0 \\
l^a(N) &= 0 \\
\tilde{h}^a(N) &= 0 \\
\tilde{l}^a(N) &= \infty \text{ for } a = 1 \text{ (and 0 otherwise)}
\end{aligned}
$$

$$f(N^\phi, S^\phi(\vec{N})) = S^\phi(\vec{N})$$

where $a = 1$ $(a = 0)$ stands for serving (not serving) the jobs.

Content delivery problem

This Markov decision process is characterized by the following transitions:

$$\begin{aligned}
b^a(N) &= \lambda(n) \\
d^a(N) &= \theta(n) \\
h^a(N) &= 0 \\
l^a(N) &= 0 \\
\tilde{h}^a(N) &= 0 \\
\tilde{l}^a(N) &= \infty \text{ for } a = 1 \text{ (and 0 otherwise)}
\end{aligned}$$

$$f(N^\phi, S^\phi(\vec{N})) = S^\phi(\vec{N})$$

where $a = 1$ $(a = 0)$ stands for serving (not serving) the jobs.

Threshold optimality

0-1 type of threshold policy is optimal.

Dynamics of a bandit in content delivery problem

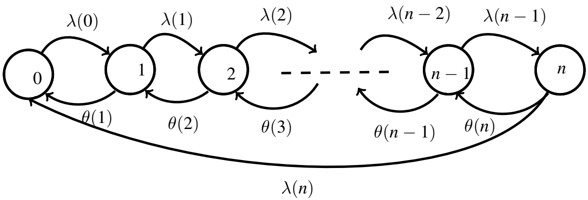

Figure: Transition diagram for threshold policy 'n' in content delivery network

Dynamics of a bandit in content delivery problem

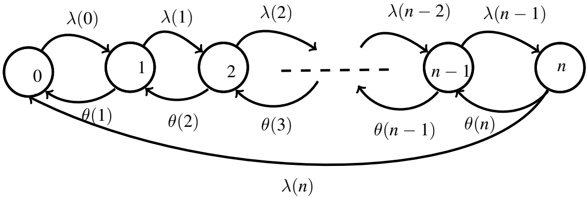

Figure: Transition diagram for threshold policy 'n' in content delivery network

Indexability

- Follows from Theorem 1.
- $\pi^n(n)$ is strictly decreasing in n.
- Index for state dependent cost and rates.

Whittle's index

Corollary 3.

If the rates and costs are state independent, i.e., $\lambda(i) = \lambda$, $\theta(i) = i\theta$, $C^h(i) = C^h$, $C^a(i) = C^a$ *and* $C_s^\infty(i) = C_s^\infty$ $\forall$ *i, then, the Whittle's index is given by*

$$W(n) = \tilde{C}\frac{\mathbb{E}(N^n) - \mathbb{E}(N^{n-1})}{\pi^{n-1}(n-1) - \pi^n(n)} - \lambda C_s^\infty \tag{11}$$

if (11) is non-decreasing in n, where $\tilde{C} = C^h + \theta C^a$ *and* $\mathbb{E}(N^n)$ *is the expected number of jobs under threshold policy n.*

Whittle's index

Corollary 3.

If the rates and costs are state independent, i.e., $\lambda(i) = \lambda$, $\theta(i) = i\theta$, $C^h(i) = C^h$, $C^a(i) = C^a$ *and* $C_s^\infty(i) = C_s^\infty$ $\forall$ *i, then, the Whittle's index is given by*

$$W(n) = \tilde{C}\frac{\mathbb{E}(N^n) - \mathbb{E}(N^{n-1})}{\pi^{n-1}(n-1) - \pi^n(n)} - \lambda C_s^\infty \tag{11}$$

if (11) is non-decreasing in n, where $\tilde{C} = C^h + \theta C^a$ *and* $\mathbb{E}(N^n)$ *is the expected number of jobs under threshold policy n.*

- The index recovers the optimal policy of propostion 3 in Larrañaga et al. (2015).

Congestion control of TCP flows

- K flows trying to deliver packets via a bottleneck router.
- Congestion window is adapted according to received acknowledgement.
    - For ACK, window is increased by 1.
    - For NACK, window is decreased.

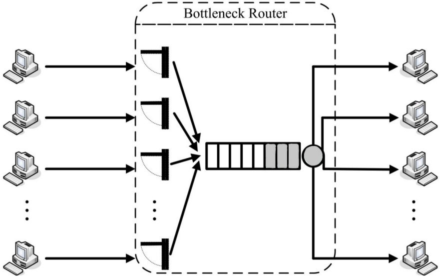

Figure: A bottleneck router in TCP with multiple flows (Avrachenkov et al., 2013)

Congestion control problem

Let $R_k(n,a)$ be the generalized $\alpha-$fairness or reward earned by flow k in state n under action a, $R_k(n,0)=0$ and

$$R_k(n,1)=\begin{cases}\frac{(1+n)^{(1-\alpha)}-1}{1-\alpha} & \text{if } \alpha\neq 1,\\ \log(n+1) & \text{if } \alpha=1;\end{cases}$$

- $C_k(n,a)=-R_k(n,a)$

Congestion control problem

Let $R_k(n,a)$ be the generalized $\alpha-$fairness or reward earned by flow k in state n under action a, $R_k(n,0) = 0$ and

$$R_k(n,1) = \begin{cases} \frac{(1+n)^{(1-\alpha)}-1}{1-\alpha} & \text{if } \alpha \neq 1, \\ \log(n+1) & \text{if } \alpha = 1; \end{cases}$$

- $C_k(n,a) = -R_k(n,a)$
- Objective is to minimize the total average cost.
    - Constraint on bottleneck router.

The Markov decision process is characterized by the following transitions:

$$\begin{aligned} b_k^a(N_k) &= \lambda_k(1-a) \\ d_k^a(N_k) &= 0 \\ h_k^a(N_k) &= 0 \\ l_k^a(N_k) &= 0 \\ \tilde{h}_k^a(N_k) &= 0 \\ \tilde{l}_k^a(N_k) &= \infty \text{ for } a=1 \text{ (and 0 otherwise)} \end{aligned}$$

$$f_k(N_k^\phi, S_k^\phi(\vec{N})) = N_k^\phi(1 - S_k^\phi(\vec{N}))$$

where action $a = 1$ (or $a = 0$) stands for sending NACK (or ACK). Jump parameter $\delta_k^1(n_k) = n_k - \max\{\lfloor \gamma_k . n_k \rfloor, 1\}$.

The Markov decision process is characterized by the following transitions:

$$\begin{aligned}
b_k^a(N_k) &= \lambda_k(1-a)\\
d_k^a(N_k) &= 0\\
h_k^a(N_k) &= 0\\
l_k^a(N_k) &= 0\\
\tilde{h}_k^a(N_k) &= 0\\
\tilde{l}_k^a(N_k) &= \infty \text{ for } a=1 \text{ (and 0 otherwise)}
\end{aligned}$$

$$f_k(N_k^{\phi}, S_k^{\phi}(\vec{N})) = N_k^{\phi}(1 - S_k^{\phi}(\vec{N}))$$

where action $a = 1$ (or $a = 0$) stands for sending NACK (or ACK). Jump parameter $\delta_k^1(n_k) = n_k - \max\{\lfloor \gamma_k . n_k \rfloor, 1\}$.

Threshold optimality

0-1 type of threshold policy is optimal.

Dynamics of a bandit in congestion control problem

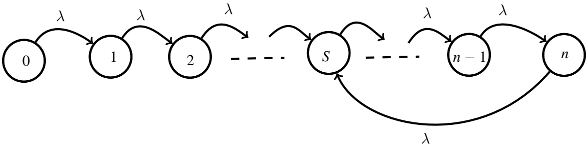

Figure: Transition diagram for TCP congestion control problem.

Dynamics of a bandit in congestion control problem

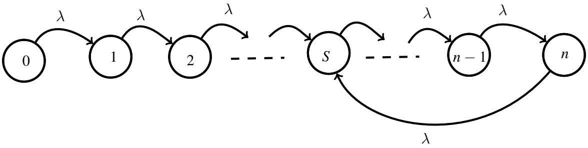

Figure: Transition diagram for TCP congestion control problem.

Indexability

- Follows from Theorem 1.
- $\mathbb{E}(f_k(N_k^{n_k}, S_k^{n_k}(N_k^{n_k})))$ is strictly increasing in n.
  - $\mathbb{E}(f_k(N_k^{n_k}, S_k^{n_k}(N_k^{n_k}))) = \sum_{m=0}^{n_k} m\pi_k^{n_k}(m)$

Whittle's index

Lemma 3.

The Whittle's index for flow k is given by,

$$W_k(n) = \begin{cases} \dfrac{2\lambda_k(n-S)(1-(1+n)^{1-\alpha}) - \sum\limits_{m=S}^{n-1}(1-(1+m)^{1-\alpha})}{(n-S)(n-S+1)(1-\alpha)} & \text{if } \alpha \neq 1, \\ \dfrac{2\lambda_k\left(\sum\limits_{m=S}^{n-1}\log(1+m) - (n-S)\log(1+n)\right)}{(n-S)(n-S+1)} & \text{if } \alpha = 1; \end{cases}$$

if $W_k(n)$ *is non-decreasing in n with* $S = \max\{\lfloor \gamma_k.(n+1) \rfloor, 1\}$.

Load balancing with heterogeneous schedulers

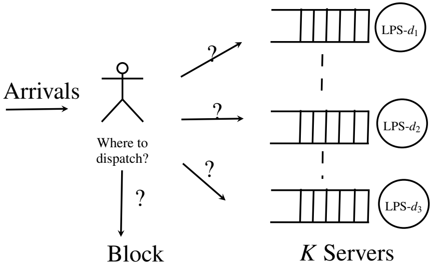

Figure: Abstraction of load balancing problem in a multi-server system with heterogeneous service disciplines.

- Significant improvement over the standard dispatching rules (JSEW).
- Index policy is close to optimal.

Summary and future directions

- A general framework for obtaining Whittle's index for the average cost criterion.
  - Machine repairman problem
  - Content delivery problem
  - Congestion control in TCP.
- Easy way to find the optimal policy for single armed restless bandit.
- Extensions to Partially observable MDPs.
- Other two dimensional control problems of interest such as batch service, polling systems etc.
- Constraint with certain probability.

References I

PS Ansell, KD Glazebrook, I Mitrani, and J Nino-Mora. A semidefinite programming approach to the optimal control of a single server queueing system with imposed second moment constraints. *Journal of the Operational Research Society*, 50(7):765–773, 1999.

Thomas W Archibald, DP Black, and Kevin D Glazebrook. Indexability and index heuristics for a simple class of inventory routing problems. *Operations research*, 57(2):314–326, 2009.

Nilay Tanik Argon, Li Ding, Kevin D Glazebrook, and Serhan Ziya. Dynamic routing of customers with general delay costs in a multiserver queuing system. *Probability in the Engineering and Informational Sciences*, 23(2):175–203, 2009.

Konstantin Avrachenkov, Urtzi Ayesta, Josu Doncel, and Peter Jacko. Congestion control of TCP flows in internet routers by means of index policy. *Computer Networks*, 57(17):3463–3478, 2013.

Vivek S Borkar and Sarath Pattathil. Whittle indexability in egalitarian processor sharing systems. *Annals of Operations Research*, pages 1–21, 2017.

John Gittins, Kevin Glazebrook, and Richard Weber. *Multi-armed bandit allocation indices*. John Wiley & Sons, 2011.

Kevin D Glazebrook, HM Mitchell, and PS Ansell. Index policies for the maintenance of a collection of machines by a set of repairmen. *European Journal of Operational Research*, 165(1):267–284, 2005.

References II

Darina Graczová and Peter Jacko. Generalized restless bandits and the knapsack problem for perishable inventories. *Operations Research*, 62(3):696–711, 2014.

Peter Jacko. Resource capacity allocation to stochastic dynamic competitors: knapsack problem for perishable items and index-knapsack heuristic. *Annals of Operations Research*, 241(1-2):83–107, 2016.

Maialen Larrañaga, Onno J Boxma, Rudesindo Núñez-Queija, and Mark S Squillante. Efficient content delivery in the presence of impatient jobs. In *Teletraffic Congress (ITC 27), 2015 27th International*, pages 73–81. IEEE, 2015.

José Niño-Mora. Dynamic priority allocation via restless bandit marginal productivity indices. *Top*, 15(2): 161–198, 2007.

José Niño-Mora. Admission and routing of soft real-time jobs to multiclusters: Design and comparison of index policies. *Computers & Operations Research*, 39(12):3431–3444, 2012a.

José Niño-Mora. Towards minimum loss job routing to parallel heterogeneous multiserver queues via index policies. *European Journal of Operational Research*, 220(3):705–715, 2012b.

José Niño-Mora and Sofía S Villar. Sensor scheduling for hunting elusive hiding targets via whittle's restless bandit index policy. In *Network Games, Control and Optimization (NetGCooP), 2011 5th International Conference on*, pages 1–8. IEEE, 2011.

Christos H Papadimitriou and John N Tsitsiklis. The complexity of optimal queuing network control. *Mathematics of Operations Research*, 24(2):293–305, 1999.

References III

Jan A Van Mieghem. Dynamic scheduling with convex delay costs: The generalized c/mu rule. *The Annals of Applied Probability*, pages 809–833, 1995.

Ina Maria Verloop. Asymptotically optimal priority policies for indexable and nonindexable restless bandits. *The Annals of Applied Probability*, 26(4):1947–1995, 2016.

Richard R Weber and Gideon Weiss. On an index policy for restless bandits. *Journal of Applied Probability*, 27(3):637–648, 1990.

Richard R Weber and Gideon Weiss. Addendum to on an index policy for restless bandits'. *Advances in Applied probability*, 23(2):429–430, 1991.

Peter Whittle. Restless bandits: Activity allocation in a changing world. *Journal of applied probability*, 25 (A):287–298, 1988.

Peter Whittle. *Optimal Control: Basics and Beyond*. Wiley Online Library, 1996.

Thank You!!!

Stationary distribution in Machine Repairman Model

$$\pi_k^{n_k}(m_k) = \frac{P_{m_k}}{\lambda_k(m_k)\left(\sum_{i=0}^{n_k}\frac{P_i}{\lambda_k(i)} + \frac{P_{n_k}}{r_k(n_k+1)}\right)} \;\forall\, m_k = 0, 1, 2, \ldots n_k, \quad (12)$$

$$\pi_k^{n_k}(n_k+1) = \frac{P_{n_k}}{r_k(n_k+1)\left(\sum_{i=0}^{n_k}\frac{P_i}{\lambda_k(i)} + \frac{P_{n_k}}{r_k(n_k+1)}\right)} \quad (13)$$

$$\pi_k^{n_k}(m_k) = 0 \;\forall\, m_k = n_k+2, \ldots \quad (14)$$

Back to Machine Repairman Problem

Proof of threshold optimality

Define $n^* = \min\{m \in \{0, 1, \cdots\} : S^{\phi^*}(m) = 1\}$

- From the definition of transition rates, all states $m > n^*$ are transient.
- This implies that the optimal average cost is same as the cost under the 0-1 type threshold policy with threshold n^*.

Back to Threshold Optimality Result