Dynamic resource allocation problems in communication networks:

Weakly Coupled Markov decision processes

Alexandre Reiffers-Masson

Equipe Maths&Net, IMT Atlantique, CS department
LabSTICC (UMR CNRS 6285)

July 11, 2023

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Lab-STICC

Example: Load balancing and service rate planning in parallel queue networks

- **Scenario:** N queues are processing jobs. The rate of each queue can be controlled. Moreover, the scheduler can also decide to which queue a job can be sent.

Example: Load balancing and service rate planning in parallel queue networks

- **Scenario:** N queues are processing jobs. The rate of each queue can be controlled. Moreover, the scheduler can also decide to which queue a job can be sent.
- **Challenge:** In order to minimize the total load on the system:
  - which queue should a job be *allocated* at each instant t?
  - which queue should see its *service rate* increased or decreased at each instant t ?

Arrival Process

We consider that at every time instant αN new jobs arrives in the system with probability $p \in (0, 1)$. Let $T_n \in \mathbb{N}_+$ be the arrival time of the n-th batch new jobs.

Note that:

$$\mathbb{P}(T_n - T_{n-1} = \tau) = (1 - p)^{\tau - 1} p, \ \forall \tau \geq 1, \ \forall n \in \mathbb{N}_+.$$

Dynamic of the queue

The length of the k-th queue, denoted by $S_k(T_{n+1})$ at instant T_{n+1} is given by:

$$S_k(T_{n+1}) = S_k(T_n) - D_k(T_{n+1} - T_n) + I\{S_k(T_n) < K\}A_k(T_n)$$

where:

- K is the finite buffer size of a queue;
- $D_k(T_n)$ the number of process jobs between T_n and T_{n+1}. We assume that the probability that a job is processed during one-time unit is equal to $B_k(t) \in \{\underline{b}, \bar{b}\}$. We assume that between two arrivals $B_k(t)$ is constant for all t and k;
- $A_k(T_n) \in \{0, 1\}$ is equal to one if one job from n-th batch is sent to the queue k.

Transition Probability

From that fact that the arrival are i.d.d. and the departure only depends on the inter-arrival time, we can rewrite the dynamic of the queue:

$$S_k(t+1) = S_k(t) - D_k(\tau) + I\{S_k(t) < K\}A_k(t).$$

We have the following lemma:

Lemma

For $s + a < k$, we have that the probability $\mathbb{P}(S_k(t+1) = s' | S_k(t+1) = s,\ A_k(t) = a,\ B_k(t) = b)$ is equal to

$$\sum_{\tau=1}^{+\infty} (1-p)^{\tau-1} p I_{s' < \min\{\tau, s+a\}} \binom{\tau}{s'} b^{\tau - s'} (1-b)^{s'}.$$

Cost functions and constraints

Costs: We will assume that there are two instantaneous costs:

- *Energy cost:* $\sum_k C_s(S_k(t)) + \sum_k C_q(B_k(t))$, where $C_s(\cdot)$ and $C_q(\cdot)$ are convex increasing.
- *Job rejection cost:* $-\gamma \sum_k A_k(t)$, with $\gamma > 0$. This cost implies that we prefer to send jobs.

Constraints: We will also assume that there are two instantaneous constraints:

$$\sum_k A_k(t) \leq \alpha N, \tag{1}$$

$$\sum_k B_k(t) \leq \beta N. \tag{2}$$

Motivation

Weakly Coupled Markov decision processes

Construction of LP-Admissible Policy

Weakly Coupled Markov decision processes

A **Weakly Coupled Markov decision processes** is composed of N statistically equivalent MDPs where:

Weakly Coupled Markov decision processes

A **Weakly Coupled Markov decision processes** is composed of N statistically equivalent MDPs where:

- $S_k(t) \in \mathcal{S}$ is the state of the arm k at the discrete decision time $t \in \{0, \cdots, T\}$,

Weakly Coupled Markov decision processes

A **Weakly Coupled Markov decision processes** is composed of N statistically equivalent MDPs where:

- $S_k(t) \in \mathcal{S}$ is the state of the arm k at the discrete decision time $t \in \{0, \cdots, T\}$,
- $A_k(t) \in \mathcal{A}$ is the action taken by the decision maker at the discrete decision time $t \in \{0, \cdots, T\}$.

Weakly Coupled Markov decision processes

A **Weakly Coupled Markov decision processes** is composed of N statistically equivalent MDPs where:

- $S_k(t) \in \mathcal{S}$ is the state of the arm k at the discrete decision time $t \in \{0, \cdots, T\}$,
- $A_k(t) \in \mathcal{A}$ is the action taken by the decision maker at the discrete decision time $t \in \{0, \cdots, T\}$.
- We assume that the decision-maker has to respect the following resource allocation constraints:

$$\sum_k D_l(S_k(t), A_k(t)) \leq N\alpha_l, \; \forall l = 1, \ldots, L$$

Weakly Coupled Markov decision processes

For each time-step $t = 0, \ldots, T-1$*:*

Weakly Coupled Markov decision processes

For each time-step $t = 0, \dots, T-1$*:*

1. The decision-maker gets full knowledge of the current system state $S(t) := [S_1(t), \dots, S_N(t)] \in \mathcal{S}^N$;

Weakly Coupled Markov decision processes

For each time-step $t = 0, \dots, T-1$*:*

1. The decision-maker gets full knowledge of the current system state $S(t) := [S_1(t), \dots, S_N(t)] \in \mathcal{S}^N$;
2. Once $S(t)$ has been observed, the decision-maker chooses a control $A(t) := [A_1(t), \dots, A_N(t)] \in \mathcal{A}^N$, such that:

$$\sum_k D_l(S_k(t), A_k(t)) \leq N\alpha_l, \ \forall l = 1, \dots, L$$

Weakly Coupled Markov decision processes

For each time-step $t = 0, \ldots, T-1$:

1. The decision-maker gets full knowledge of the current system state $S(t) := [S_1(t), \ldots, S_N(t)] \in \mathcal{S}^N$;
2. Once $S(t)$ has been observed, the decision-maker chooses a control $A(t) := [A_1(t), \ldots, A_N(t)] \in \mathcal{A}^N$, such that:

$$\sum_k D_l(S_k(t), A_k(t)) \leq N\alpha_l, \ \forall l = 1, \ldots, L$$

3. The decision-maker collects the reward $\sum_k r_{S_k(t)}^{A_k(t)}$;

Weakly Coupled Markov decision processes

For each time-step $t = 0, \dots, T-1$:

1. The decision-maker gets full knowledge of the current system state $S(t) := [S_1(t), \dots, S_N(t)] \in \mathcal{S}^N$;
2. Once $S(t)$ has been observed, the decision-maker chooses a control $A(t) := [A_1(t), \dots, A_N(t)] \in \mathcal{A}^N$, such that:

$$\sum_k D_l(S_k(t), A_k(t)) \leq N\alpha_l, \ \forall l = 1, \dots, L$$

3. The decision-maker collects the reward $\sum_k r_{S_k(t)}^{A_k(t)}$;
4. For every k, the arm k evolves to $S_k(t+1) = s'$ with probability $p_{S_k(t),s'}^{A_k(t)}$.

Weakly Coupled Markov decision processes

For each time-step $t = 0, \dots, T-1$:

1. The decision-maker gets full knowledge of the current system state $S(t) := [S_1(t), \dots, S_N(t)] \in \mathcal{S}^N$;
2. Once $S(t)$ has been observed, the decision-maker chooses a control $A(t) := [A_1(t), \dots, A_N(t)] \in \mathcal{A}^N$, such that:

$$\sum_k D_l(S_k(t), A_k(t)) \leq N\alpha_l, \ \forall l = 1, \dots, L$$

3. The decision-maker collects the reward $\sum_k r_{S_k(t)}^{A_k(t)}$;
4. For every k, the arm k evolves to $S_k(t+1) = s'$ with probability $p_{S_k(t),s'}^{A_k(t)}$.

Objective: Maximize the expected total sum of rewards over the T time-steps.

Weakly Coupled Markov decision processes

For each time-step $t = 0, \dots, T-1$:

1. The decision-maker gets full knowledge of the current system state $S(t) := [S_1(t), \dots, S_N(t)] \in \mathcal{S}^N$;
2. Once $S(t)$ has been observed, the decision-maker chooses a control $A(t) := [A_1(t), \dots, A_N(t)] \in \mathcal{A}^N$, such that:

$$\sum_k D_l(S_k(t), A_k(t)) \leq N\alpha_l, \ \forall l = 1, \dots, L$$

3. The decision-maker collects the reward $\sum_k r_{S_k(t)}^{A_k(t)}$;
4. For every k, the arm k evolves to $S_k(t+1) = s'$ with probability $p_{S_k(t),s'}^{A_k(t)}$.

Objective: Maximize the expected total sum of rewards over the T time-steps.

Discussion with respect to the constraints

We assume that all terms in $D_l(s, a)$ and α_l are non-negative numbers, and that $D(s, 0) = 0$.

This is a natural assumption under the resource allocation context in which $a = 0$ corresponds to a passive action that consumes no resources.

Implication: The later also implies that our resource constraint problem has at least a feasible solution by always choosing the passive action.

Mathematical Formulation

$$
\begin{aligned}
\min_{\pi} \quad & \mathbb{E}\sum_{t=0}^{T-1}\sum_{s,a} r_s^a Y_{a,s}^{(N)}(t) := V_{opt}^{(N)}(m(0),T) && \text{(3a)}\\
\text{s.t.} \quad & \text{Arms follow the Markovian evolution generated by } \Pi_n p_{s_n,s'_n}^{a_n}, && \text{(3b)}\\
& \sum_a Y_{a,s}^{(N)}(t) = M_s^{(N)}(t),\ \forall t\in[[0,T-1]],\ \forall s\in\mathcal{S}, && \text{(3c)}\\
& \sum_s D_l(s,a) Y_{s,a}^{(N)}(t) \leq \alpha_l\ \forall t\in[[0,T-1]],, && \text{(3d)}\\
& M_s^{(N)}(0) = m_s(0),\ \forall s\in\mathcal{S}, && \text{(3e)}
\end{aligned}
$$

where $m_s(0) = \frac{1}{N}\sum_{k=1}^{N} I\{S_k(0)=s\}$, for all $s\in\mathcal{S}$.

Difficulty

The key difficulty of Weakly Coupled Markov decision processes is coming from:

$$\sum_s D_l(s,a) Y_{s,a}^{(N)}(t) \leq \alpha_l \; \forall t \in [[0, T-1]],$$

which couples all the arms together.

Challenge of the day:

How to design an efficient heuristic to solve such problem?

A different one that the projection policy.

Outline of the approach

1. **Relaxation:** Classical approach is to relax this constraint and consider a problem where this constraint has to be satisfied only in expectation:

Outline of the approach

1. **Relaxation:** Classical approach is to relax this constraint and consider a problem where this constraint has to be satisfied only in expectation:

$$\sum_s D_l(s,a)\mathbb{E}[Y_{s,a}^{(N)}(t)] \leq \alpha_l \ \forall t \in [[0, T-1]],$$

Outline of the approach

1. **Relaxation:** Classical approach is to relax this constraint and consider a problem where this constraint has to be satisfied only in expectation:

$$\sum_s D_l(s,a)\mathbb{E}[Y_{s,a}^{(N)}(t)] \leq \alpha_l \ \forall t \in [[0, T-1]],$$

2. **Interpolation:** Construct a sequence of decision rules $\pi_t : \Delta^d \to \Delta^{2d}$ which is optimal for the relaxed problem.

Relaxed problem

$$\min_{\pi} \quad \mathbb{E}[\sum_{t=0}^{T-1} \sum_{s,a} r_s^a Y_{a,s}^{(N)}(t)] =: V_{rel}^{(N)}(m(0), T) \tag{4a}$$

$$\text{s.t.} \quad \text{Arms follow the Markovian evolution generated by } \Pi_n p_{s_n, s'_n}^{a_n}, \tag{4b}$$

$$\sum_a Y_{a,s}^{(N)}(t) = M_s^{(N)}(t),\ \forall t \in [[0, T-1]],\ \forall s \in \mathcal{S}, \tag{4c}$$

$$\sum_s D_l(s,a)\mathbb{E}[Y_{s,a}^{(N)}(t)] \leq \alpha_l\ \forall t \in [[0, T-1]],\ \forall l, \tag{4d}$$

$$M_s^{(N)}(0) = m_s(0),\ \forall s \in \mathcal{S}, \tag{4e}$$

LP formulation

Let us define the following LP problem:

$$
\begin{aligned}
\min_{y \geq 0} \quad & \sum_{t=0}^{T-1} \sum_{s,a} r_s^a y_{s,a}(t) =: V_{LP}(m(0), T) \\
\text{s.t.} \quad & \sum_{a} y_{s,a}(t) = m_s(t),\ \forall t \in [[0, T-1]],\ \forall s \in \mathcal{S}, \\
& m_s(t) = \sum_{s'} \sum_{a} y_{s',a}(t-1) p_{s',s}^a \ \forall t \in [[1, T-1]],\ \forall s \in \mathcal{S}, \\
& \sum_{s} D_l(s,a) y_{s,a}(t) \leq \alpha_l \ \forall t \in [[0, T-1]],\ \forall l, \\
& m_s(0) = m^0,\ \forall s \in \mathcal{S}
\end{aligned}
\tag{5}
$$

LP formulation

Let us define the following LP problem:

$$
\begin{aligned}
\min_{y \geq 0} \quad & \sum_{t=0}^{T-1} \sum_{s,a} r_s^a y_{s,a}(t) =: V_{LP}(m(0), T) \\
\text{s.t.} \quad & \sum_a y_{s,a}(t) = m_s(t),\ \forall t \in [[0, T-1]],\ \forall s \in \mathcal{S}, \\
& m_s(t) = \sum_{s'} \sum_a y_{s',a}(t-1) p_{s',s}^a\ \forall t \in [[1, T-1]],\ \forall s \in \mathcal{S}, \\
& \sum_s D_l(s,a) y_{s,a}(t) \leq \alpha_l\ \forall t \in [[0, T-1]],\ \forall l, \\
& m_s(0) = m^0,\ \forall s \in \mathcal{S}
\end{aligned}
\tag{5}
$$

We denote by $y^* := [[[y_{s,a}^*(t)]]]_{s,a,t}$ the optimal solution of (6) and we also define $m^* := [[m_s(t) := \sum_a y_{s,a}^*(t)]]_{s,t}$.

Motivation

Weakly Coupled Markov decision processes

Construction of LP-Admissible Policy

Feasible control

We define the set of feasible control at time t by:

$$\mathcal{Y}(M^{(N)}(t)) := \Big\{ y \in \mathbb{R}_+^{2S} | \sum_a y_{s,a} = M_s^{(N)}(t) \ \forall s \in \mathcal{S};$$

$$\sum_s \sum_a D_l(s,a) y_{s,a} \leq \alpha_l \Big\}$$

Feasible control

We define the set of feasible control at time t by:

$$\mathcal{Y}(M^{(N)}(t)) := \Big\{ y \in \mathbb{R}_+^{2S} | \ \sum_a y_{s,a} = M_s^{(N)}(t) \ \forall s \in \mathcal{S};$$

$$\sum_s \sum_a D_l(s,a) y_{s,a} \leq \alpha_l \Big\}$$

Some observations:

Feasible control

We define the set of feasible control at time t by:

$$\mathcal{Y}(M^{(N)}(t)) := \left\{ y \in \mathbb{R}_+^{2S} | \sum_a y_{s,a} = M_s^{(N)}(t) \ \forall s \in \mathcal{S}; \right.$$

$$\left. \sum_s \sum_a D_l(s,a) y_{s,a} \leq \alpha_l \right\}$$

Some observations:

1. In general, note that $y^*(t)$ are not an integers;
2. In general $y^*(t) \notin \mathcal{Y}(M^{(N)}(t))$;

Feasible control

We define the set of feasible control at time t by:

$$\mathcal{Y}(M^{(N)}(t)) := \Big\{ y \in \mathbb{R}_+^{2S} | \sum_a y_{s,a} = M_s^{(N)}(t) \; \forall s \in \mathcal{S};$$

$$\sum_s \sum_a D_l(s,a) y_{s,a} \leq \alpha_l \Big\}$$

Some observations:

1. In general, note that $y^*(t)$ are not an integers;
2. In general $y^*(t) \notin \mathcal{Y}(M^{(N)}(t))$;
3. In general $y^*(t) \in \mathcal{Y}(m^*(t))$.

Feasible control

We define the set of feasible control at time t by:

$$\mathcal{Y}(M^{(N)}(t)) := \Big\{ y \in \mathbb{R}_+^{2S} | \sum_a y_{s,a} = M_s^{(N)}(t) \ \forall s \in \mathcal{S};$$
$$\sum_s \sum_a D_l(s,a) y_{s,a} \leq \alpha_l \Big\}$$

Some observations:

1. In general, note that $y^*(t)$ are not an integers;
2. In general $y^*(t) \notin \mathcal{Y}(M^{(N)}(t))$;
3. In general $y^*(t) \in \mathcal{Y}(m^*(t))$.

Resolving policy

We redefine the following LP:

$$
\begin{aligned}
\min_{y \geq 0} \quad & \sum_{t=0}^{T-t-1} \sum_{s,a} r_s^a y_{s,a}(t) =: V_{LP}(m(0), \textbf{T-t}) \\
\text{s.t.} \quad & \sum_a y_{s,a}(t) = m_s(t),\ \forall t \in [[0, \textbf{T-t-1}]],\ \forall s \in \mathcal{S}, \\
& m_s(t) = \sum_{s'} \sum_a y_{s',a}(t-1) p_{s',s}^a \ \forall t \in [[1, \textbf{T-t-1}]],\ \forall s \in \mathcal{S}, \\
& \sum_s y_{s,1}(t) \leq \alpha,\ \forall t \in [[0, \textbf{T-t-1}]],, \\
& m_s(0) = m^0,\ \forall s \in \mathcal{S}
\end{aligned} \tag{6}
$$

Resolving policy

We redefine the following LP:

$$
\begin{aligned}
\min_{y \geq 0} \quad & \sum_{t=0}^{T-t-1} \sum_{s,a} r_s^a y_{s,a}(t) =: V_{LP}(m(0), \textbf{T-t}) \\
\text{s.t.} \quad & \sum_a y_{s,a}(t) = m_s(t), \ \forall t \in [[0, \textbf{T-t-1}]], \ \forall s \in \mathcal{S}, \\
& m_s(t) = \sum_{s'} \sum_a y_{s',a}(t-1) p_{s',s}^a \ \forall t \in [[1, \textbf{T-t-1}]], \ \forall s \in \mathcal{S}, \\
& \sum_s y_{s,1}(t) \leq \alpha, \ \forall t \in [[0, \textbf{T-t-1}]], , \\
& m_s(0) = m^0, \ \forall s \in \mathcal{S}
\end{aligned}
\tag{6}
$$

The solution of this LP is denoted by

$$
y^{Res}(m(0), T-t) = [y_{t'}^{Res}(m(0), T-t)]_{0 \leq t' \leq T-t-1}.
$$

Algorithm to solve the LP

What could be a possible algorithm to solve this LP?

Solution 1: Simplex or Convex optimisation?

Solution 2: Dynamic programming. Observe that:

$$V_{LP}(m, T-t) = \min_{y \in \mathcal{Y}(m)} \sum_{s,a} r_s^a y_{s,a} + V_{LP}(\phi(m,y), T-t-1),$$

where $\phi_s(m,y) = \sum_{s'} \sum_a y_{s',a} p_{s',s}^a$ for all s.

Resolving operator

We define the following operator:

$$\pi_t^{Res}(M^{(N)}) := y_0^{Res}(M^{(N)}, T-t). \tag{7}$$

Resolving operator

We define the following operator:

$$\pi_t^{Res}(M^{(N)}) := y_0^{Res}(M^{(N)}, T-t). \tag{7}$$

Note that:

Resolving operator

We define the following operator:

$$\pi_t^{Res}(M^{(N)}) := y_0^{Res}(M^{(N)}, T-t). \qquad (7)$$

Note that:

1. In general, note that $\pi_t^{Res}(M^{(N)})$ are not an integers;

Resolving operator

We define the following operator:

$$\pi_t^{Res}(M^{(N)}) := y_0^{Res}(M^{(N)}, T - t). \tag{7}$$

Note that:

1. In general, note that $\pi_t^{Res}(M^{(N)})$ are not an integers;
2. $\pi_t^{Res}(M^{(N)}) \in \mathcal{Y}(M^{(N)}(t))$;

Resolving operator

We define the following operator:

$$\pi_t^{Res}(M^{(N)}) := y_0^{Res}(M^{(N)}, T-t). \tag{7}$$

Note that:

1. In general, note that $\pi_t^{Res}(M^{(N)})$ are not an integers;
2. $\pi_t^{Res}(M^{(N)}) \in \mathcal{Y}(M^{(N)}(t))$;
3. $y^*(t) = pi_t^{Res}(m^*(t))$. (P-Admissible Policy)

Algorithm

Resolving Policy

- **Input:** Initial system configuration vector $m(0)$ and time horizon T.
- **Set** $\hat{M}$:= m(0);
- **For** $t = 0, 2, \ldots, T-1$ **do:**
  1. **Compute** $y^{Res}(\hat{M}, T-t)$; Set $\hat{y}(t) = y_0^{Res}(\hat{M}, T-t)$
  2. *Rounding step:* For all $s \in \mathcal{S}$, set:

$$\hat{Y}_{s,a}^{(N)}(t) = \begin{cases} N^{-1}\lfloor N\hat{y}_{s,1}(t)\rfloor & \text{if } a = 1, \\ \hat{M}_s - N^{-1}\lfloor N\hat{y}_{s,1}(t)\rfloor & \text{otherwise.} \end{cases}$$

  3. Use control $\hat{Y}^{(N)}$ to advance to the next time-step ;
  4. Set $\hat{M}$:= current empirical distribution;

Certainty equivalent control

Our policy is inspired from the *certainty equivalent control* (CEC).

Certainty equivalent control

Our policy is inspired from the *certainty equivalent control* (CEC).

Principle of the CEC

Sub-optimal control that applies at each stage the control that would be optimal if some or all of the uncertain quantities were fixed at their expected values.

Bibliography

- The proof of the main theorem and more advance theorem can be found here: Gast, Nicolas, Bruno Gaujal, and Chen Yan. "The LP-update policy for weakly coupled Markov decision processes." arXiv preprint arXiv:2211.01961 (2022).
- If you want to have a quick introduction to dynamic programming, please have a look to the lecture note of Nahum Shimkin: `https://webee.technion.ac.il/shimkin/LCS11/LCS11index.html`