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*

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

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:

$$\sum_k A_k(t) \leq \alpha N, \quad (1)$$

$$\sum_k B_k(t) \leq \beta N. \quad (2)$$

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 $t = 0, \ldots, T-1$*:*

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

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 $t = 0, \ldots, T-1$:

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 $\sum_k r_{S_k(t)}^{A_k(t)}$;
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 $t = 0, \ldots, T-1$:

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 $\sum_k r^{A_k(t)}_{S_k(t)}$;
4. For every k, the arm k evolves to $S_k(t+1) = s'$ with probability $p^{A_k(t)}_{S_k(t),s'}$.

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

$$\min_{\pi} \quad \mathbb{E}\sum_{t=0}^{T-1}\sum_{s,a} r_s^a Y_{a,s}^{(N)}(t) := V_{opt}^{(N)}(m(0),T) \tag{3a}$$

s.t. Arms follow the Markovian evolution generated by $\Pi_n p^{a_n}_{s_n,s'_n}$, (3b)

$$\sum_{a} Y_{a,s}^{(N)}(t) = M_s^{(N)}(t),\ \forall t \in [[0,T-1]],\ \forall s \in \mathcal{S}, \tag{3c}$$

$$\sum_{s} D_l(s,a) Y_{s,a}^{(N)}(t) \le \alpha_l\ \forall t \in [[0,T-1]],, \tag{3d}$$

$$M_s^{(N)}(0) = m_s(0),\ \forall s \in \mathcal{S}, \tag{3e}$$

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:

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

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.