Slicing in 5G networks

Update 01/10/2020

Alessandro Spallina

A slice simulator with MDP resolution

Last Call

- Support for finite horizon (mdp)
- Support for immediate action (mdp and simulator)
- Fix bad naming (conservative -> all on, smart conservative -> conservative)
- Fix formulation

A slice simulator with MDP resolution

What's new?

- Batch Manager
- Switch to absolute policy (before was in delta format)
- Support for N allocation policy (as thesis)
- Bugfix "in the timeslot I see the jobs arrived up to the previous instant" (before was "I see the jobs arrived until the last moment of this timeslot")
- Slurm first usage (WIP)
- Support for Bayati's assumptions (WIP)

A slice simulator with MDP resolution

Money Metric - L. M. Bayati's Thesis

"According to research published [MDD10], a single server consumes around something between 238 and 376 Watts. Rajesh et al. [RDSJ08] estimate the cost of one kWh of energy to 0.0897$. These values may vary depending on where the data center is located and how electricity is generated. Using that baseline, one server costs around 300$ per year to run. Every job may generate a profit, and the average profit per job can be computed as a ratio of the total profit over the number of served jobs. For instance, 106 requests (page views) may bring 1000$ of revenue. Thus, it can be said that each job brings $10^{-3}$$ on average. Work in [DM10] suggests that each successfully processed job generates a profit around $6.2\times10^{-6}$$. In this case, a lost job costs $6.2\times10^{-6}$$." [1.3.5]

- Costs for switching on and off?
- Respons Time

A slice simulator with MDP resolution

Timeslot Sizing - Questions

Since the behavior of the system is time-slotted (we see arrivals only at the beginning of a timeslot), the timeslot must be small enough to model what happens in continuous time as closely as possible.

Can we scale down the timeslot after the construction of the arrivals histogram as in L. M. Bayati's Thesis?

A slice simulator with MDP resolution

Multiple timeslot scale - Questions

We have multiple timeslot scale:

- Queue update every X ms
- Server allocation every Y s

Should we support this before multi tenant slices?

A slice simulator with MDP resolution

Next Step

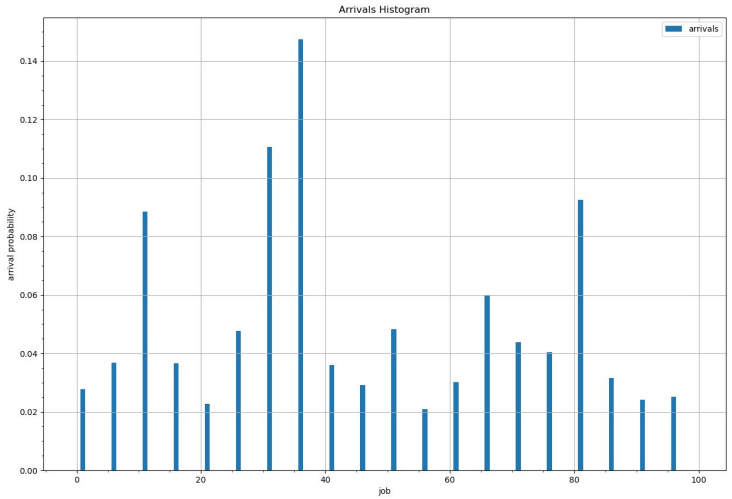
- Support for multi tenant slices
- Multiple timeslot scale support
- Simulations with pseudo-realistic histograms
- Performance optimization

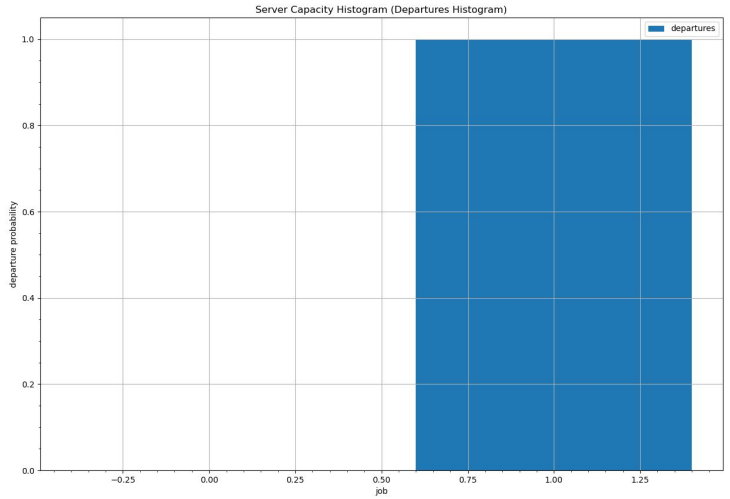
Simulation Results

Common Parameters

A slice simulator with MDP resolution

Simulation Results - Common Parameters





A slice simulator with MDP resolution

Simulation Results - Common Parameters

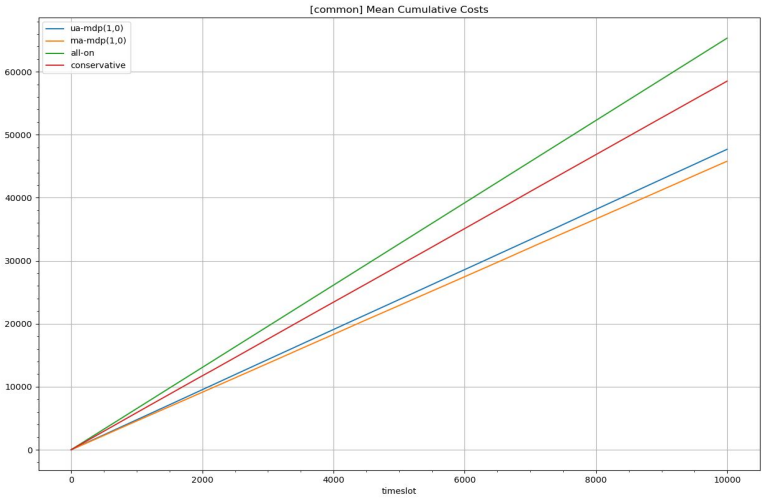
- Queue size: 20
- Max allocated servers: 20
- C_j: 1; alpha: 1
- C_s: 1; beta: 1
- C_l: 1; gamma: 1
- Number of simulations: 10
- Simulation Time: 10k time slots
- MDP discount value: 0.99

Scenario 1

Multiple action vs Unitary action

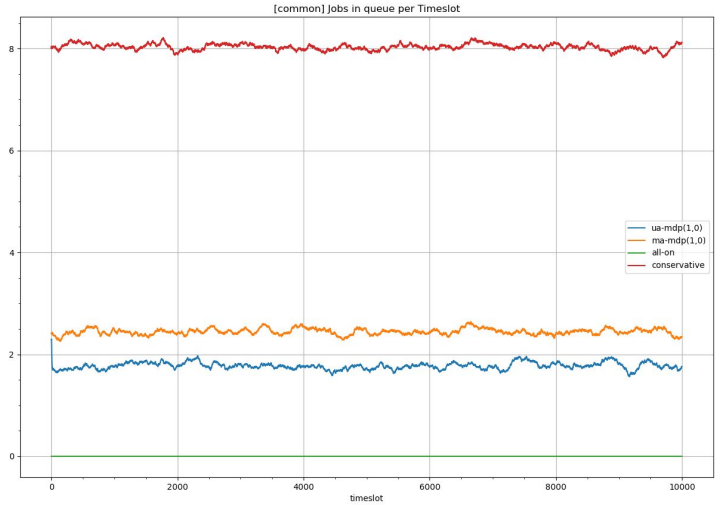
A slice simulator with MDP resolution

Scenario 1: Multiple action vs Unitary action



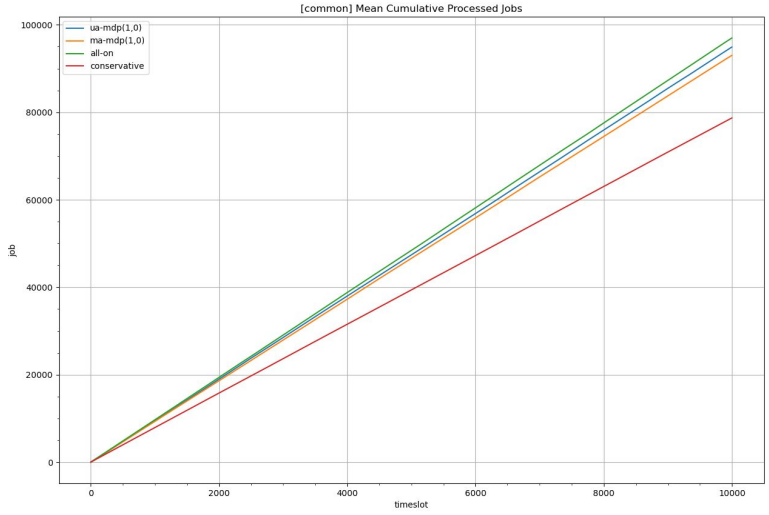
A slice simulator with MDP resolution

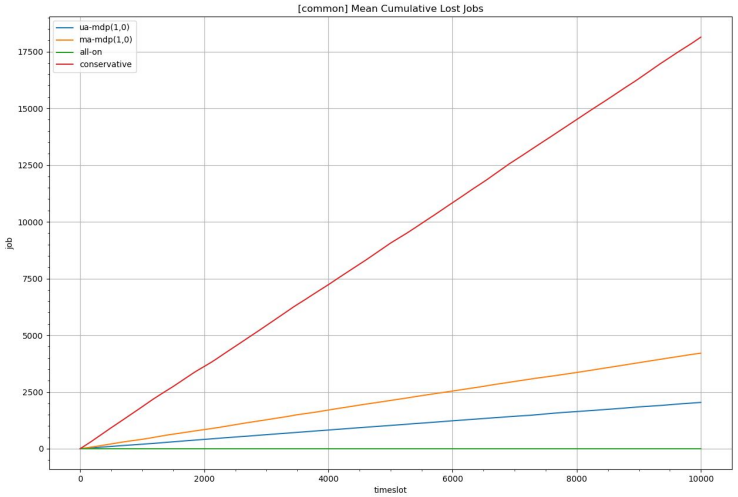
Scenario 1: Multiple action vs Unitary action



A slice simulator with MDP resolution

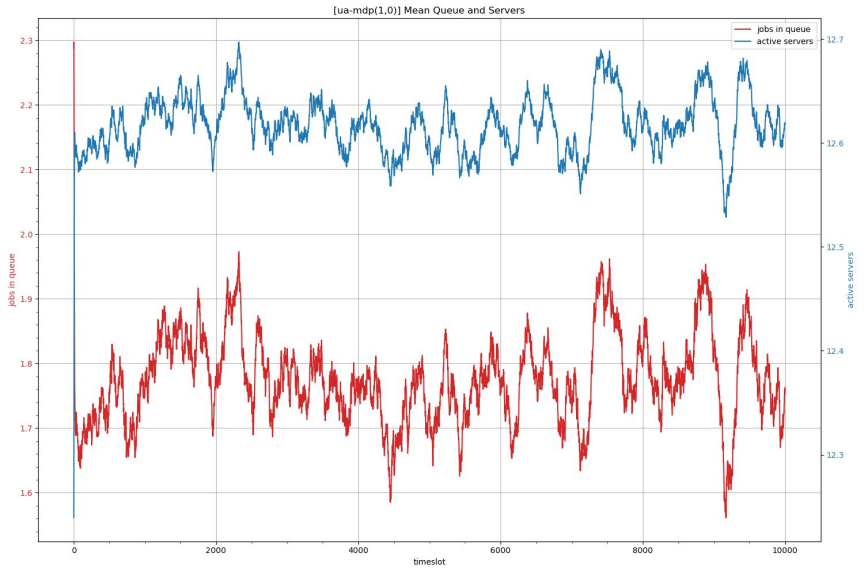
Scenario 1: Multiple action vs Unitary action

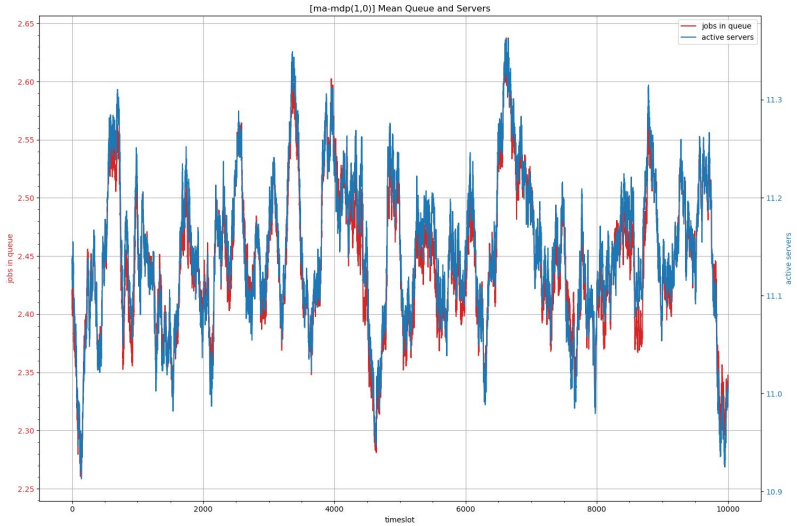




A slice simulator with MDP resolution

Scenario 1: Multiple action vs Unitary action





A slice simulator with MDP resolution

Scenario 1: Multiple action vs Unitary action

ua-mdp(1,0)

	0 servers	1 servers	2 servers	3 servers	4 servers	5 servers	6 servers	7 servers	8 servers	9 servers	10 servers	11 servers	12 servers	13 servers	14 servers	15 servers	16 servers	17 servers	18 servers	19 servers	20 servers
0 jobs	1	2	3	4	5	6	7	8	9	10	11	12	12	12	13	14	15	16	17	18	19
1 jobs	1	2	3	4	5	6	7	8	9	10	11	12	13	13	13	14	15	16	17	18	19
2 jobs	1	2	3	4	5	6	7	8	9	10	11	12	13	13	13	14	15	16	17	18	19
3 jobs	1	2	3	4	5	6	7	8	9	10	11	12	13	13	13	14	15	16	17	18	19
4 jobs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	14	14	15	16	17	18	19
5 jobs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	14	14	15	16	17	18	19
6 jobs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	14	14	15	16	17	18	19
7 jobs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	15	15	16	17	18	19
8 jobs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	16	16	17	18	19
9 jobs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	16	16	17	18	19
10 jobs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	17	17	18	19
11 jobs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	17	17	18	19
12 jobs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	17	17	18	19
13 jobs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	17	17	18	19
14 jobs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	17	17	18	19
15 jobs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	17	17	18	19
16 jobs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	17	17	18	19
17 jobs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	17	17	18	19
18 jobs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	17	17	18	19
19 jobs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	17	17	18	19
20 jobs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	17	17	18	19

A slice simulator with MDP resolution

Scenario 1: Multiple action vs Unitary action

ma-mdp(1,0)

	0 servers	1 servers	2 servers	3 servers	4 servers	5 servers	6 servers	7 servers	8 servers	9 servers	10 servers	11 servers	12 servers	13 servers	14 servers	15 servers	16 servers	17 servers	18 servers	19 servers	20 servers
0 jobs	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
1 jobs	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
2 jobs	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
3 jobs	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
4 jobs	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12
5 jobs	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17
6 jobs	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17
7 jobs	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18
8 jobs	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19
9 jobs	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
10 jobs	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
11 jobs	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
12 jobs	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
13 jobs	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
14 jobs	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
15 jobs	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
16 jobs	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
17 jobs	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
18 jobs	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
19 jobs	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
20 jobs	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20

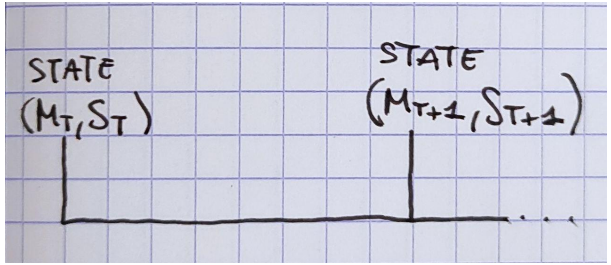
Backup

A slice simulator with MDP resolution

Formulation - Assumption

Delayed Action (timeslot view):

1. State
2. Action chosen according to the state
3. Arrival phase (losses)
4. Processing phase
5. Execution of the action chosen in (2)



Immediate Action (timeslot view):

1. State
2. Action chosen according to the state
3. Execution of the action chosen in (2)
4. Arrival phase (losses)
5. Processing phase

A slice simulator with MDP resolution

Formulation - Assumptions of L. M. Bayati's Thesis

"We begin by serving the waiting jobs of the buffer, next we fill the free operational servers by the new jobs, then we fill the buffer." [1.3.1]

"At the beginning of each slot, and based on the current state of the system, an action $\alpha j \in A$ will be made to determine how many servers will be operational during the current slot." [4.1.2.1]

Immediate Action + phases exchange (timeslot view):

1. State
2. Action chosen according to the state
3. Execution of the action chosen in (2)
4. Processing phase
5. Arrival phase (losses)

A slice simulator with MDP resolution

Formulation - Transition Probability

This can be generalized as follows:

$$Q(m, s \to m', s') = \sum_{a=[m'-m]^+}^{\text{qsize}-m} P(\text{arr} = a) \cdot P(\text{proc} = m + a - m' | a + m) \quad (2)$$

$$+ \sum_{a=\text{qsize}-m+1}^{\infty} P(\text{arr} = a) P(\text{proc} = \text{qsize} - m' | \text{qsize}) \quad (3)$$

Where $s' = s + action$ and

$$action = \begin{cases} 0 & \text{do nothing} \\ +1 & \text{allocate 1 server} \\ -1 & \text{deallocate 1 server} \end{cases}$$

(2) non full queue
(3) full queue but we have missing probabilities due the histograms

Where $P(\text{proc} = x|y)$ is the probability of processing x jobs given that y jobs are found in the queue the instant when the processor starts to pick jobs from the queue. Observe that

Delayed Action

$$P(\text{proc} = x|y) = \begin{cases} H^s_{\text{departures}}(x) & \text{if } x < y \\ \sum_{x=y}^{\infty} H^s_{\text{departures}}(x) & if x \geq y \end{cases}$$

Immediate Action

$$P(\text{proc} = x|y) = \begin{cases} H^{s'}_{\text{departures}}(x) & \text{if } x < y \\ \sum_{x=y}^{\infty} H^{s'}_{\text{departures}}(x) & if x \geq y \end{cases}$$

Notice that is the number of current servers s is equal to 0, then the departure histogram will be just $\Delta_1([1., 0., ..., 0.])$

A slice simulator with MDP resolution

Formulation - Transition Probability

The transition probability is then:

$$Q^{\text{action}}(m, s \to m', s') = \begin{cases} Q(m, s \to m', s') & \text{if } s' = s + \text{action} \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

A slice simulator with MDP resolution

Example - Timeslot Sizing

Assuming:

- QueueSize = 3
- 1 Server always On
- 1 job/sec