One-step policy improvements
to reduce the waiting time at traffic lights

René Haijema and Jan van der Wal

workshop on Optimal Control in Stochastic Systems
EURANDOM, Eindhoven, November 27, 2010

Outline

1. One-step policy improvement examples
2. Control of traffic lights
3. Markov Decision Problem
4. One-step policy improvement over FC
    - FC – decomposition,
    - FC – relative values.
5. Results
6. Conclusions

1. One-step policy improvement

To approximately solve an MDP:

1. Choose initial policy π

2. Value determination step

   $\forall x : v^\pi(x) + g^\pi = C(x, \pi(x)) + \sum_y P_{xy}(\pi(x))v^\pi(y).$

   relative values $v^\pi(x)$ and gain g^π

3. Policy improvement step

   $$\forall x : \pi'(x) = \arg\min_a \left[C(x,a) + \sum_y P_{xy}(a)v^\pi(y) \right].$$

Complication

- the number of states can be too large to determine v^π

  $\forall x : v^\pi(x) + g^\pi = C(x, \pi(x)) + \sum_y P_{xy}(\pi(x)) v^\pi(y).$

- even when v^π is approximate iteratively
  e.g. by successive approximations (SA)

  $\forall x : v_n^\pi(x) = C(x, \pi(x)) + \sum_y P_{xy}(\pi(x)) v_{n-1}^\pi(y).$

- especially true when x is multi-dimensional

Approximate solution

Break down the big MC into sub MCs

by decomposition of the state space (Norman, 1972)

1. Slightly modify the problem assumptions to relax the dependencies

   and/or

2. Choose a well structured initial policy that allows decomposition

Examples – Decomposition and One-step policy improvement

Routing telephone calls in a network or call center

- Krishnan and Ott (1987), Sassen, Tijms and Nobel (1997), Bhulai (2008)

Producing multiple items on a single machine

- Wijngaard (1979), De Bruin and Van der Wal (2010)

Traffic lights

- Haijema and Van der Wal (2008)

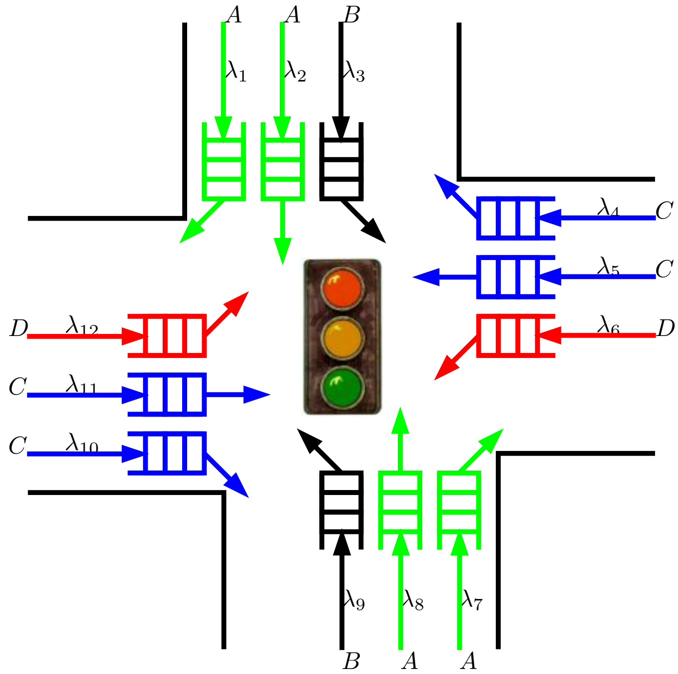
Outline

1. One-step policy improvement examples
2. Control of traffic lights
3. Markov Decision Problem
4. One-step policy improvement over FC
   - FC – decomposition,
   - FC – relative values.
5. Results
6. Conclusions

2. Problem of interest

Minimize the overall mean waiting time per car!

- F streams in
  C (disjoint) combinations,
- Green to ≤ 1 combination,
- Grouping and sequencing
  is given,
- Lights:
  green $\longrightarrow$ yellow $\longrightarrow$ red $\longrightarrow$ green.

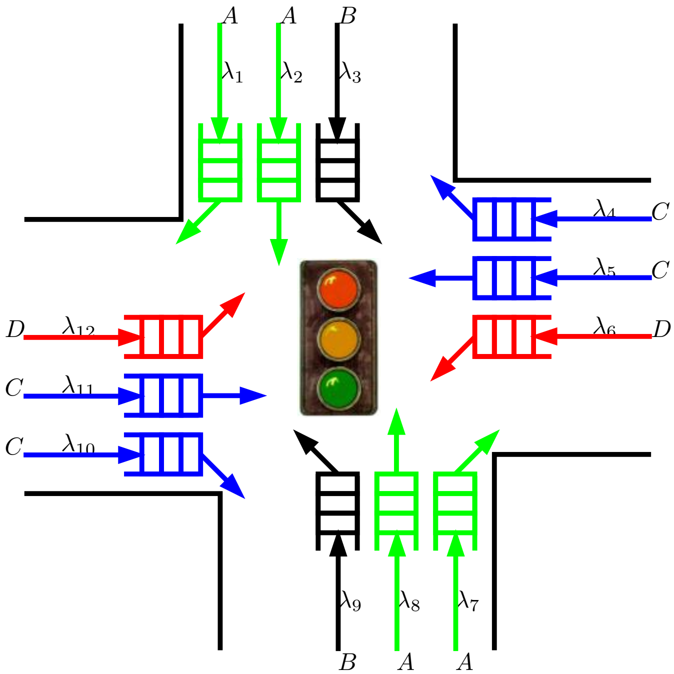


Clean problem

- Cars only,
- Known # cars at queue f,
- Under-saturated cases only,
- Identical clearance times.

Generalization possible!!

When to switch?



3. Markov Decision Problem (MDP)

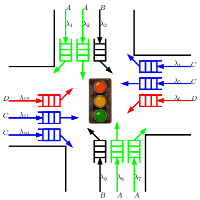
3.1. Discrete time

- 1 slot = 2 sec.

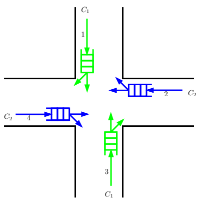
- Switching to another combination takes 3 slots:
  - 2 slots yellow
  - 1 slot all red

- Events within 1 slot:
  - observe state,
  - change lights,
  - 0 or 1 arrival/queue (1 w.p. p_f),
  - $\leq$ 1 departure per "G-or-Y" queue.

- State = $(q_1, \ldots, q_F; l)$,
  q_f = # cars at queue $f < Q$
  l = state of light
- Total # states =
  $Q^F \cdot [1 + C(1 + 2)]$
- Examples — $q_f \le 20$ cars



$F = 12,\ C = 4:$ #states $\approx 10^{17}$ → intractable.



$F = 4,\ C = 2:$ # states $\approx 10^{6}$ → (just) tractable

MDP: curse of dimensionality

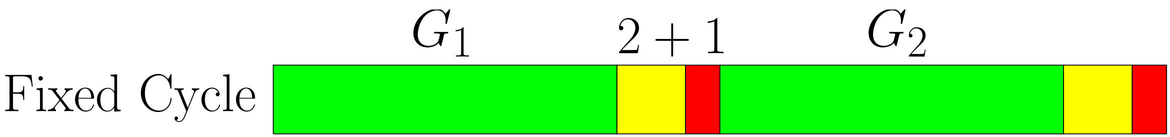
Heuristics:

- Exhaustive control (XC)
- Fixed cycle (FC)
- ...

4. One-step policy improvement

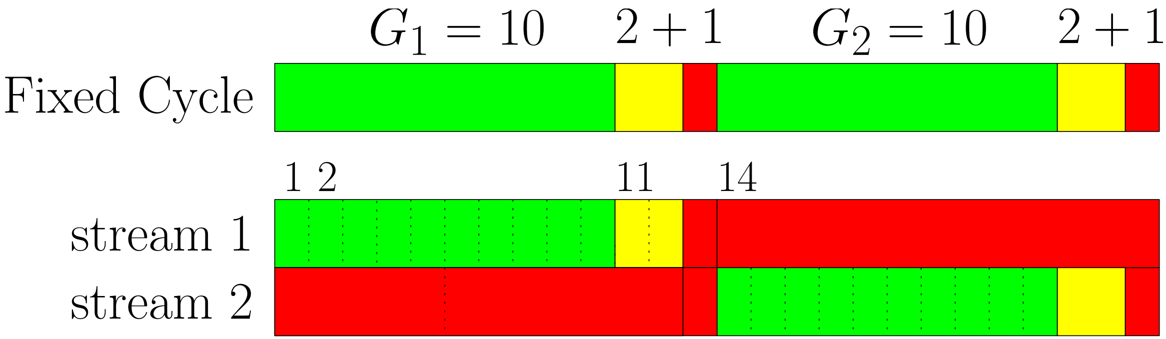
4.1. Initial policy Fixed Cycle (FC)

- Fixed cyclic order
- Fixed green periods: $G_1, \ldots, G_C$
  $\longrightarrow$ Fixed cycle length $= D$ slots



Example

- $G_1 = G_2 = 10 \rightarrow D = 26$ slots, labeled $1, 2, \ldots\ldots, 26$



- Remaining green time depends on slot number only

- Decomposition $\rightarrow$ evaluate each stream in isolation.

- Relative value of state $(t, q_1, \ldots, q_F) = \sum_{f=1}^{F} v_f^{FC}(t, q_f)$.

- where under FC for stream f:

$v_f^{FC}(t, q_f)$ = relative value (bias term) of state (t, q_f)

$$v_f^{FC}(t, q_f) = \lim_{n \to \infty} \frac{1}{D} \sum_{j=n}^{n+D-1} \left(v_j^f(t, q_f) - v_j^f(t, 0) \right)$$

- and

$v_n^f(t, q_f) =$ total expected costs at stream f over a planning horizon of n slots when starting in state (t, q_f).

FC – Value iteration

$$v^f_{n+1}(t, q_f) = q_f \quad + \quad p_f \cdot v^f_n(t+1, q_f+1) \quad + (1-p_f) \cdot v^f_n(t+1, q_f)$$

, if t implies red to f

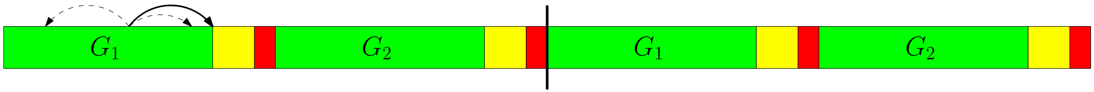
$$v^f_{n+1}(t, q_f) = q_f \quad + \quad p_f \cdot v^f_n(t+1, q_f) \quad + (1-p_f) \cdot v^f_n(t+1, q_f-1)$$

, if t implies green/yellow to f.

4.4. Policy improvement step – break FC

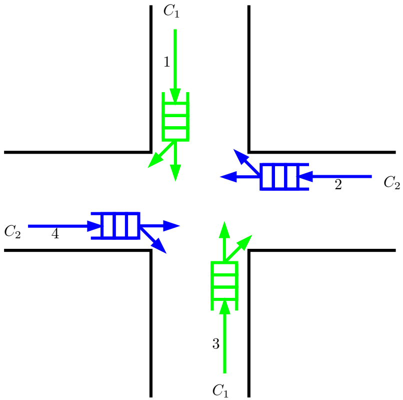
Every decision epoch:

1. observe $(t, q_1, \ldots, q_F)$
2. adjust state of traffic light



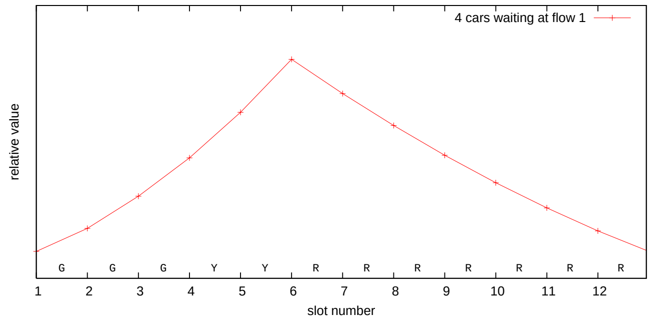
Lengthen or shorten green periods

Choose a best 'time-jump' from slot t to slot a: relative values

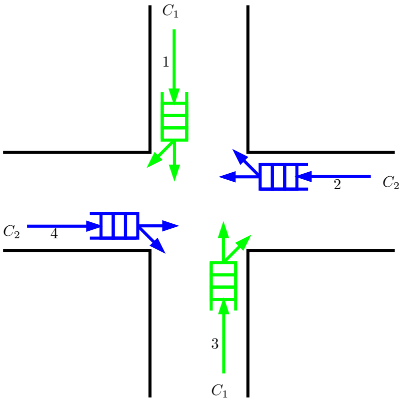


4 identical streams
in 2 combinations
$G_1 = G_2 = 3 \to D = 12$

Best slot a given t and $\mathbf{q} = (4, 2, 2, 1)$?

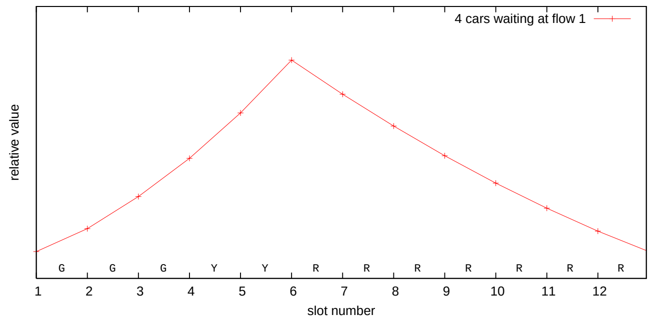


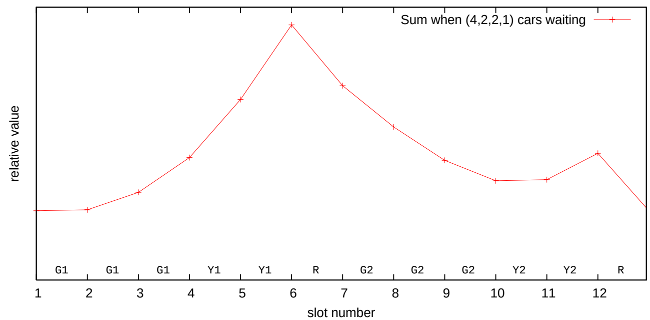
$v_1^{FC}(a, 4)$.



4 identical streams
in 2 combinations
$G_1 = G_2 = 3 \rightarrow D = 12$

Best slot a given t and $\mathbf{q} = (4, 2, 2, 1)$?





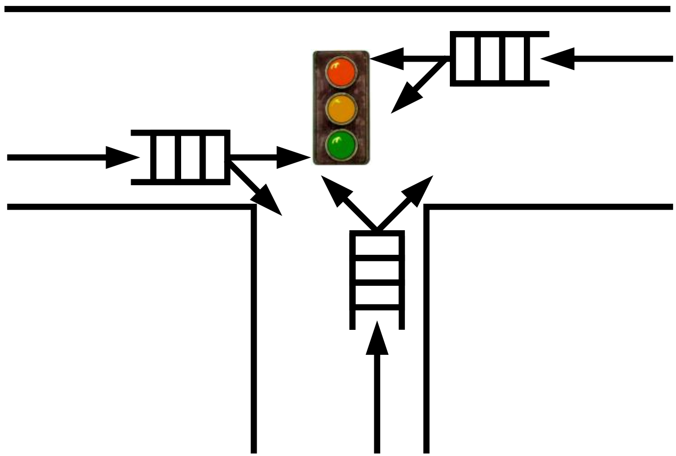
$$\sum_{f=1}^{F} v_f^{FC}(a, q_f).$$

5. Simulation results

Compare some cyclic policies:

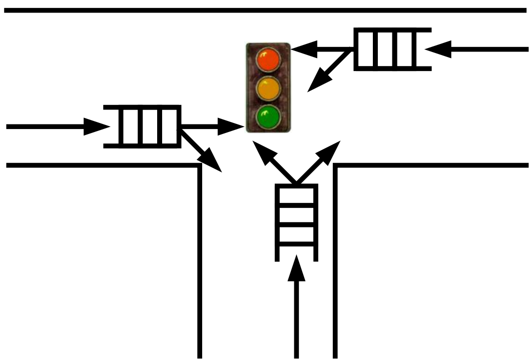
RV	:	one-step policy improvement over FC
MDP	:	optimal cyclic policy
FC	:	fixed cycle
XC	:	switch when $q_f = 0$
XC-2	:	switch when $q_f \leq 2$

At most 1 out of 3 has green:

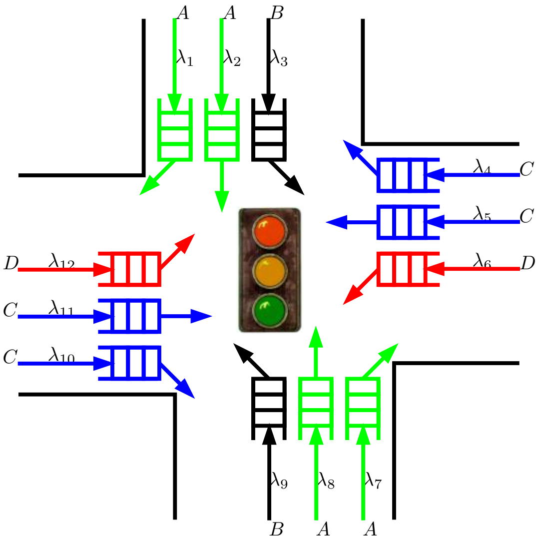


At medium and high arrival intensities

Overall average waiting time per car in seconds.



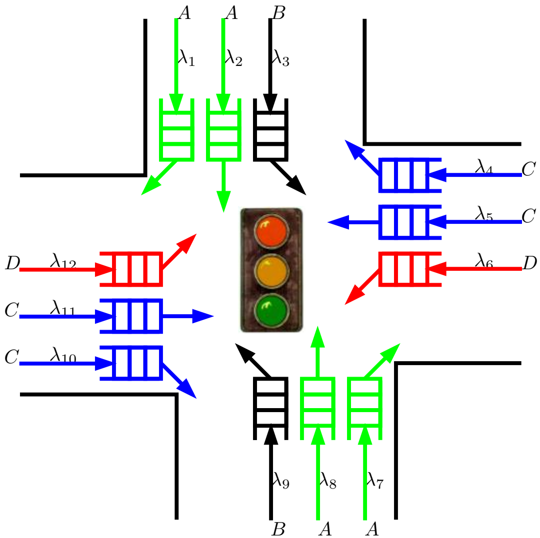
load	Medium	High
RV	10.7	20.3
MDP	10.6	18.2
FC	16.0	33.8
XC	14.0	27.8
XC-2	10.7	18.4



- Dilemma:
  - A or $C \rightarrow \leq 4$ dep/slot
  - B or $D \rightarrow \leq 2$ dep/slot
- Definition of XC?

A more complex infrastructure

Overall average waiting time per car in seconds.



load	High	% above RV
RV	41.8	
FC	51.2	+22%
XC	90.0	+115%
XC-2	53.6	+28%

WAGENINGEN UNIVERSITY
WAGENINGEN UR

Tail distributions

Rule	*EW* overall		*EW* C_1, C_3	*EW* C_2, C_4
RV1	41.8		37.4	50.6
FC	50.5	+21%	50.5	50.4
XC	89.8	+115%	88.5	92.4
XC-1	70.1	+68%	68.9	72.4
XC-2	53.3	+28%	52.1	55.8

- RV improves FC:
  - 50-65% waiting time reduction for ‘simple’ cases,
  - 22% at complex intersection.

- RV and XC-2 are nearly optimal for ‘simple’ cases,

- RV is superior for complex intersection.

6. Conclusions and remarks

- Optimal control of traffic lights is complicated,
- Approximation by One-step policy improvement
  - improves FC,
  - is fast even for very large intersections,
  - is often better than other dynamic control rules.

Not in this talk

1. Modified RV rules

   - Acyclic control,

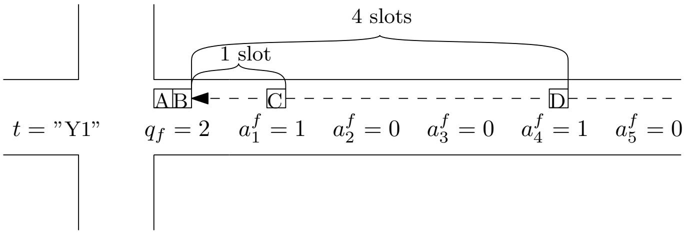
   - "2-step" policy improvement rule.

2. Information on near-future arrival,

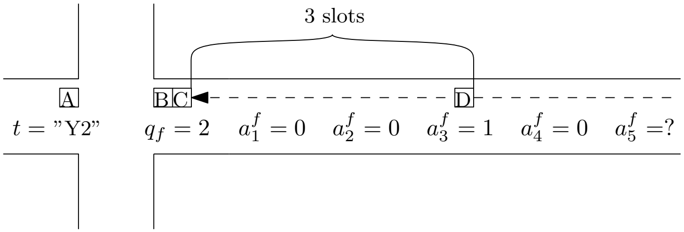
3. Arterials and Networks.

Extension – Arrival information

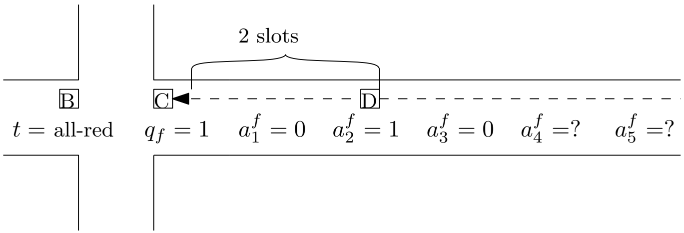
$$\text{State} = (t, q_f, a_1^f, a_2^f, a_3^f, a_4^f, a_5^f)$$



After one slot:



After two slots:



Extension – Arrival information

Relative values of states:

$$\lim_{N\to\infty} \sum_{f=1}^{F} \frac{1}{D} \sum_{d=0}^{D-1} \left[V_{N+d}\left(t, q_f, \mathbf{a}^f\right) - (n+d) \cdot g^{(f)} \right] \tag{1}$$

where $g^{(f)}$ and $V_{N+d}\left(t, q_f, \mathbf{a}^f\right)$ by value iteration algorithm.

$V_{n+1}^{f}\left(t, q_f, a_1^f, a_2^f, a_3^f, a_4^f, a_5^f\right) = q_f +$

$+(1-\lambda_f) \cdot V_n^f\left(t, (q_f + a_1^f - \Delta_t^f)^+, a_2^f, a_3^f, a_4^f, a_5^f, 0\right)$

$+\lambda_f \cdot V_n^f\left(t, (q_f + a_1^f - \Delta_t^f)^+, a_2^f, a_3^f, a_4^f, a_5^f, 1\right)$

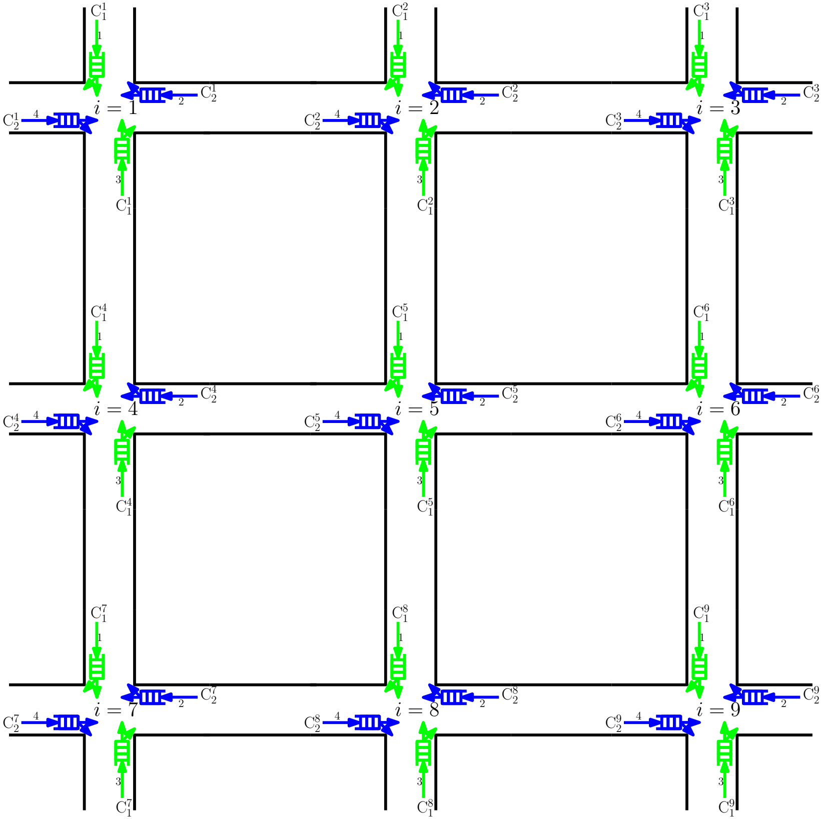
or discrete simulation over m slots + $\sum_{f=1}^{N} R_f^{FC}(t', q'_f)$ as terminal cost.

Results – F12C4 with 5 slots arrival info

Mean waiting times (in sec.) for partly-asymmetric F12C4
(Identical arrival rates: 0.1, 0.15, and 0.2)

Rule	$\rho = 0.4$		$\rho = 0.6$		$\rho = 0.8$	
Cyclic policies:						
RV1	13.9		20		44	
RV1(5)	13.1	-6%	19	-5%	43	-3%
FC	15.5	+12%	24	+24%	53	+21%
XC	20.0	+44%	35	+76%	96	+118%
XC-2	13.9	0%	21	+2%	58	+32%
FC cycle length	32		40		88	
FC departure times	(6, 6, 6, 6)		(8, 8, 8, 8)		(20, 20, 20, 20)	

Extension – Network of intersections (I3x3F4C2)



Results – Network of 9 intersections (I3x3F4C2)

Mean waiting times (in sec.) for skew I3x3F4C2, $\rho^i = 0.8$:

$(\lambda_1^i = 0.32, \lambda_2^i = \lambda_3^i = 0.16$ and $\lambda_4^i = 0.48)$

Policy	No cars turn right		17% turns right		34% turns right	
RV1(5)[a]	8.9		9.6		10.0	
RV1[a]	12.9	+44%	12.7	+33%	12.6	+27%
FC, coordinated	8.8[b]	-0.9%	12.3[c]	+29%	14.4[d]	+44%
XC-2	11.6	+30%	11.6	+21%	11.6	+16%

[a]Based on FC with $D_i = 44$, $d_{i,1} = 16$ and $d_{i,2} = 24$ seconds.
[b]FC with $D_i = 42$, $\psi = 42$, $d_{i,1} = 16$ and $d_{i,2} = 22$ seconds.
[c]FC with $D_i = 44$, $\psi = 42$, $d_{i,1} = 16$ and $d_{i,2} = 24$ seconds.
[d]FC with $D_i = 48$, $\psi = 40$, $d_{i,1} = 18$ and $d_{i,2} = 26$ seconds.