Coupling Time in Markovian Queueing Networks

Jill-Jênn Vie
Coordinator: Bruno Gaujal

June 8, 2010

Contents

List of Figures

1 Introduction

Markov chains are central to the understanding of random processes. They are applied in a number of ways in many different fields such as communication networks or performance evaluation, they can even compose music. In the theory of queueing systems, there is a "coveted item": the stationary distribution. It can be calculated for a certain type of queues, for example, those of infinite capacity, but generally it has to be computed. Unfortunately, most techniques are either imprecise or too slow when the state space grows.

There comes "Coupling from the Past" (CFTP). This algorithm produces independent samples of states according to their stationary distribution. Our main interest is setting bounds on the mean coupling time, which represents the complexity of the algorithm, and more precisely to prove that the mean coupling time is $O(\sum_i C_i)$ where C_i denotes the capacity of queue i, for several types of networks.

2 Markovian Queueing Networks

Consider an open queueing network (QN) Q consisting of M queues $Q_1, \ldots, Q_M$. Each queue Q_i ($1 \leqslant i \leqslant M$) has a finite capacity C_i.

Let $\mathbf{C}$ be the vector of capacities : $\mathbf{C} = (C_1, \ldots, C_M)$. So, the state space $\mathcal{S}_\mathbf{C}$ of the network is $\mathcal{S}_\mathbf{C} = \{\mathbf{s} \in \mathbb{Z}^M, \forall i, 0 \leqslant s_i \leqslant C_i\}$.

Jobs join the QN from an external Poissonian source with mean arrival rate λ. The probability that a job joins queue i from outside is p_{0i}. In queue i, each job requires some processing for an exponentially distributed amount of time with mean service rate μ_i. If queue i is not empty, a job is sent to queue j with probability p_{ij}, and is accepted if queue j is not full, otherwise it is lost. Finally, the probability that a job leaves the network after service at queue i is p_{i0}, as if it were sent to a "trash-queue" with capacity zero.

That QN can be seen as a discrete-event system: all possible events are of the form a_{ij} ($i, j \in \{0, \dots, M\}$), corresponding to the service of one job in Q_i joining Q_j, zero representing the outside world. The event space is denoted by $\mathcal{A}$.

The rate of event $a_{ij} \in \mathcal{A}$ is denoted by γ_{ij}. Hence, if $i, j \neq 0$, then $\gamma_{ij} = \mu_i p_{ij}$, $\gamma_{0j} = \lambda p_{0j}$ and $\gamma_{i0} = \mu_i p_{i0}$. The total event rate Γ is finite and $\Gamma = \sum_{i,j} \gamma_{ij}$.

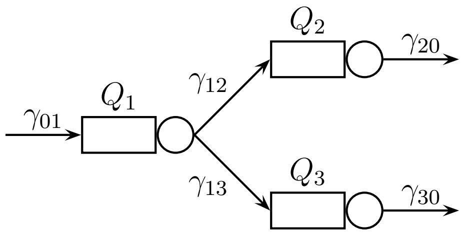

Figure 1: Queueing network.

	Origin	Destination	Enabling condition	Routing policy
a_{01}	*Nowhere*	Q_1	None	Rejection if Q_1 is full
a_{12}	Q_1	Q_2	$s_1 > 0$	Rejection if Q_2 is full
a_{13}	Q_1	Q_3	$s_1 > 0$	Rejection if Q_3 is full
a_{20}	Q_2	*Nowhere*	$s_2 > 0$	None
a_{30}	Q_3	*Nowhere*	$s_3 > 0$	None

We can now define the stochastic process $\{\mathbf{s}(t) \in \mathcal{S}_\mathbf{C}\}_{t \geqslant 0}$, which is a continuous-time Markov chain, where the state of the system $\mathbf{s}(t) = (s_1(t), \dots, s_M(t))$ represents the number of customers in each queue at time t. Besides, we consider the partially ordered set $(\mathcal{S}_\mathbf{C}, \preccurlyeq)$ such that:

$$(s_1, \dots, s_M) \preccurlyeq (s'_1, \dots, s'_M) \Leftrightarrow \forall i \in \{1, \dots, M\}, s_i \leqslant s'_i.$$

2.1 Coupling from the Past

That continuous-time Markov chain can be transformed into (X_n), a discrete-time Markov chain with the same stationary distribution, by uniformization by Γ. It is assumed to be irreducible and aperiodic.

The evolution of (X_n) can be written under the form $X_{n+1} = \varphi(X_n, u_n)$, where u_n is a random variable over the event space $\mathcal{A}$ that takes value a_{ij} with probability γ_{ij}/Γ.

Let $\mathbf{e_i}$ be the unit vector in direction i of M.
The transition function $\varphi : S_{\mathbf{C}} \times \mathcal{A} \to S_{\mathbf{C}}$ is defined as follows:

- If $i, j \neq 0$, $\varphi(\mathbf{s}, a_{ij}) = \begin{cases} \mathbf{s} - \mathbf{e_i} + \mathbf{e_j} & \text{if } \mathbf{0} \leqslant \mathbf{s} - \mathbf{e_i} + \mathbf{e_j} \leqslant \mathbf{C} \\ \mathbf{s} - \mathbf{e_i} & \text{else, and if } \mathbf{s} - \mathbf{e_i} \geqslant \mathbf{0} \\ \mathbf{s} & \text{else.} \end{cases}$

- If $i = 0$, $\varphi(\mathbf{s}, a_{0j}) = \begin{cases} \mathbf{s} + \mathbf{e_j} & \text{if } \mathbf{s} + \mathbf{e_j} \leqslant \mathbf{C} \\ \mathbf{s} & \text{else.} \end{cases}$

- If $j = 0$, $\varphi(\mathbf{s}, a_{i0}) = \begin{cases} \mathbf{s} - \mathbf{e_i} & \text{if } \mathbf{s} - \mathbf{e_i} \geqslant \mathbf{0} \\ \mathbf{s} & \text{else.} \end{cases}$

For example, in figure 1, for event a_{12} we get:

$$\varphi(\square, a_{12}) : (s_1, s_2, s_3) \mapsto \begin{cases} (s_1 - 1, s_2 + 1, s_3) & \text{if } s_1 \geqslant 1 \text{ and } s_2 < C_2 \\ (s_1 - 1, s_2, s_3) & \text{if } s_1 \geqslant 1 \text{ and } s_2 = C_2 \ (Q_2 \text{ full}) \\ (s_1, s_2, s_3) & \text{if } s_1 = 0 \ (Q_1 \text{ empty}). \end{cases}$$

Let $\varphi^{(n)} : \mathcal{S}_{\mathbf{C}} \times \mathcal{A}^n \to \mathcal{S}_{\mathbf{C}}$ denote the function which output is the state of the chain after n iterations starting in state $\mathbf{s} \in \mathcal{S}_{\mathbf{C}}$.

$$\varphi^{(n)}(\mathbf{s}, u_{1\to n}) \triangleq \varphi(\ldots \varphi(\varphi(\mathbf{s}, u_1), u_2), \ldots, u_n)$$

This notation can be extended to sets of states:

$$\forall E \subset \mathcal{S}_{\mathbf{C}}, \varphi^{(n)}(E, u_{1\to n}) \triangleq \{\varphi^{(n)}(\mathbf{s}, u_{1\to n}), \mathbf{s} \in E\}.$$

In the following, $|E|$ denotes the cardinality of set E.
Note that φ is monotone for each event of $\mathcal{A}$:

$$\forall n \in \mathbb{N}, \forall u_1, \ldots, u_n \in \mathcal{A}, X_0 \preccurlyeq X'_0 \Rightarrow \varphi^{(n)}(X_0, u_{1\to n}) \preccurlyeq \varphi^{(n)}(X'_0, u_{1\to n}).$$

Theorem ([3]).

$$\lim_{n\to+\infty} \left|\varphi^{(n)}(\mathcal{S}_{\mathbf{C}}, u_{-n+1\to 0})\right| = 1 \textit{ almost surely.}$$

Furthermore, the value of $\varphi^{(n)}(\mathcal{S}_{\mathbf{C}}, u_{-n+1\to 0})$ *is steady-state distributed.*

As $\mathcal{S}_{\mathbf{C}}$ is partially ordered by $\preccurlyeq$ and $\varphi^{(n)}$ is monotone, we only have to simulate trajectories from the minimal state $\mathbf{0}$ and the maximal state $\mathbf{C}$, as illustrated in figure 2. We hence obtain the algorithm 1.

3 Coupling Time

Almost surely, there exists a finite time τ, the *coupling time*, defined by:

$$\tau = \min\{n \in \mathbb{N}, \left|\varphi^{(n)}(\mathcal{S}, e_{-n+1\to 0})\right| = 1\}.$$

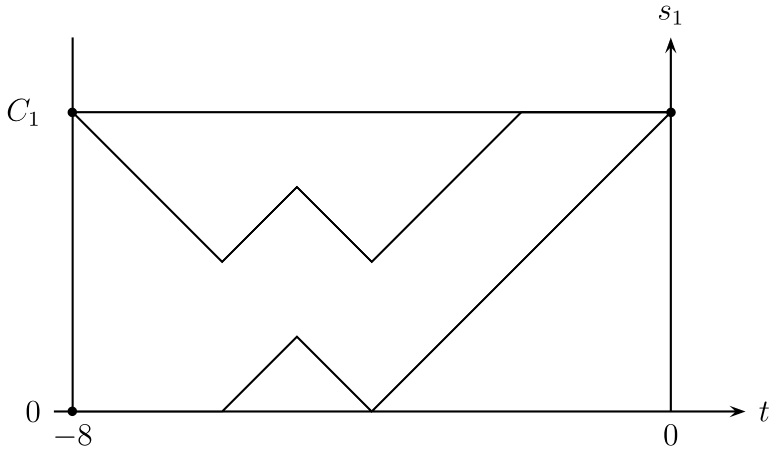

Figure 2: Coupling queue. The sequence of events is $(a_{10}^{(2)}, a_{01}, a_{10}, a_{01}^{(4)})$ ($a_{ij}^{(n)}$ means a_{ij} for n times).

Algorithm 1 Perfect simulation algorithm (PSA)

Data: φ, $(u_{-n})_{n\in\mathbb{N}}$.
Result: $\mathbf{s}^* \in \mathcal{S}_{\mathbf{C}}$, generated according to the stationary distribution of the system.

```
  n ← 1, M ← C, m ← o
  repeat
    for i = n − 1 downto 0 do
      M ← φ(M, u_{−i})
      m ← φ(m, u_{−i})
    end for
    n ← 2n
  until M = m
  s* ← M
  return s*
```

Proposition ([3]). *The average time and space complexity of PSA is $O(\mathbb{E}\tau)$.*

3.1 Case of one queue

Proposition ([1]). *The mean coupling time $\mathbb{E}\tau$ of one queue with capacity C, arrival rate λ and service rate μ is bounded using $p = \frac{\lambda}{\lambda+\mu}$ and $q = 1-p$.*

Critical bound	$\forall p \in [0,1]$,	$\mathbb{E}\tau \leqslant \frac{C^2+C}{2}$.
Heavy traffic bound	if $p > \frac{1}{2}$,	$\mathbb{E}\tau \leqslant \frac{C}{p-q} - \frac{q\left(1-\left(\frac{q}{p}\right)^C\right)}{(p-q)^2}$.
Light traffic bound	if $p < \frac{1}{2}$,	$\mathbb{E}\tau \leqslant \frac{C}{q-p} - \frac{p\left(1-\left(\frac{p}{q}\right)^C\right)}{(q-p)^2}$.

3.2 Ψ^2, the Perfect Simulator

Ψ^2 is a simulator developed by INRIA that implements PSA.

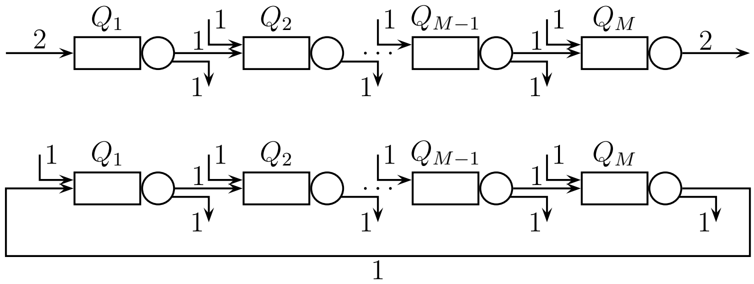

Figure 3: Queueing networks (acyclic/cyclic) in critical case (as many arrivals than departures) simulated with Ψ^2. Results are in the appendix (A).

With respect to the results given in the appendix (A), we can conjecture that the mean coupling time is $O(\sum_i C_i)$ (and $O(\sum_i C_i^2)$ for the critical case), whether the network is acyclic or not (see also figure 4).

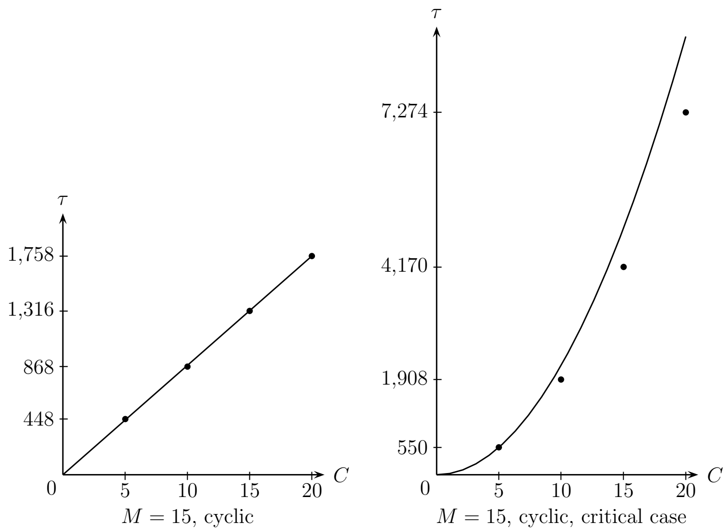

Figure 4: Evolution of τ in function of the capacities.

3.3 Acyclic networks

If the network is acyclic, we can number the queues according to the topological order. Therefore, no event occuring in Q_j has any effect on queue Q_i if $j > i$: once Q_1 couples, it will stay coupled until the end of the simulation. We can then consider Q_2, that couples almost surely, etc.

In [1], it is hence proven that $\mathbb{E}\tau = O(\sum_i C_i^2)$.

3.4 Cyclic networks

It is possible to prove that the mean coupling time of QNs is $\Theta(\sum_i C_i)$ for networks having arbitrary topology.

The main idea consists in simulating two QNs, one of capacities $\mathbf{C}$, the other of capacities $\mathbf{C} - \mathbf{e_i}$ for a certain i, with the same events.

For example, with $M = 2$ queues, the trajectories starting from $\mathbf{0}$ and $\mathbf{C} - \mathbf{e_2}$ couple in $(0, C_2 - 1)$ and then reach the bottom line at point $\mathbf{s}$ (see figure 5). At that time, the dashed trajectory starting from $\mathbf{C}$ reaches $\mathbf{s} + \mathbf{e_2}$. Both trajectories then couple in constant time ([2]).

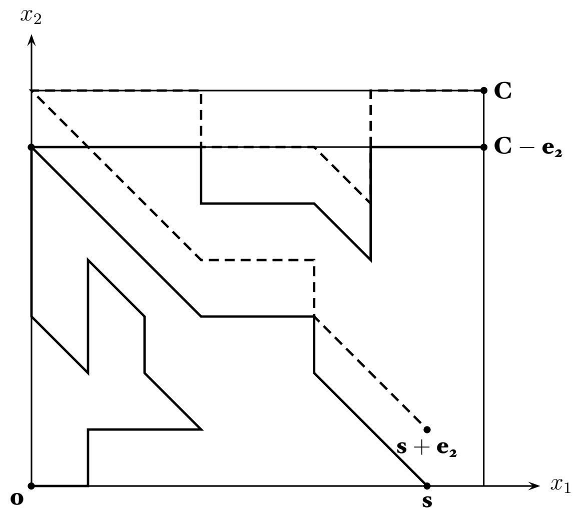

Figure 5: Coupling of the trajectories starting from $\mathbf{0}$ and $\mathbf{C} - \mathbf{e_2}$ in a queueing network of arbitrary topology.
Events: $(a_{01}, a_{02}, a_{01}^{(2)}, a_{12}, a_{02}, a_{12}, a_{20}^{(2)}, a_{12}, a_{10}^{(2)}, a_{02}^{(3)}, a_{10}^{(3)}, a_{21}^{(3)}, a_{01}^{(2)}, a_{20}, a_{21}^{(2)})$.

Note that the difference between trajectories can change: a queue can be decoupled. When it happens, another queue couples.

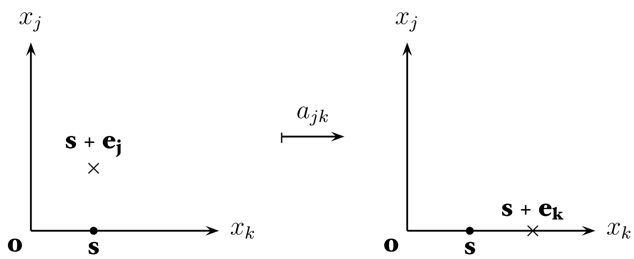

Figure 6: Decoupling of Q_k, coupling of Q_j.

4 Extension to Negative Customers

In this section, we extend the approach to QNs having "negative customers": we add events a^-_{ij} $(i, j \in \{1, \dots, M\})$ to $\mathcal{A}$ corresponding to the service of one job in Q_i that joins Q_j, kills a customer then commits suicide (this may happen everyday and everywhere in daily life, so please be careful when you are waiting in a queue).

4.1 Simulation

Ψ^2 handles negative customers.

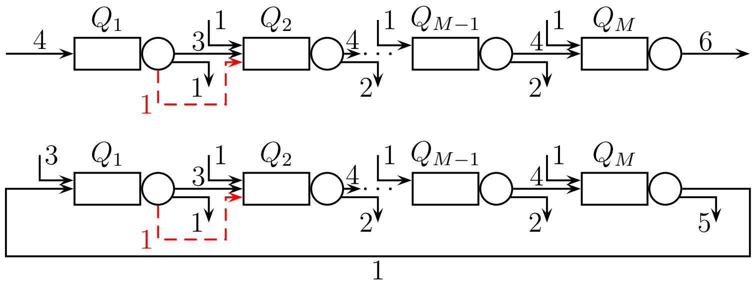

Figure 7: Queueing networks (acyclic/cyclic) with negative customers (in red, dashed) simulated with Ψ^2. Results are in the appendix (A).

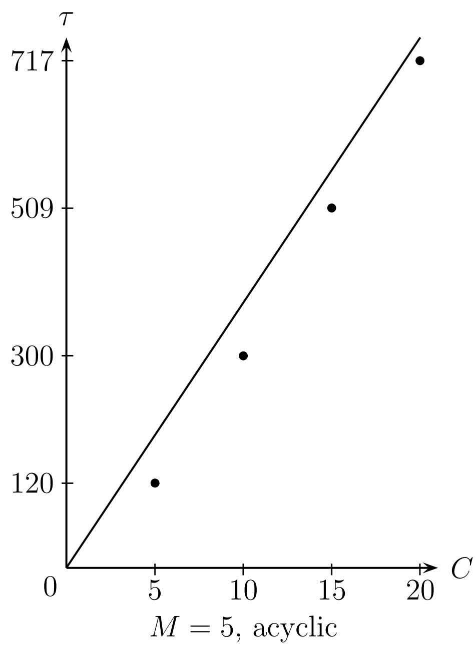

Figure 8: Evolution of τ in function of the capacities, with negative customers.

Once more, the mean coupling time seems to be $O(\sum_i C_i)$.

4.2 Non-monotonicity

The resulting system is not monotone. Indeed, for $M = 2$, $(0,1) \preccurlyeq (1,1)$, but $\varphi((0,1), a_{12}^-) = (0,1) \not\succcurlyeq (0,0) = \varphi((1,1), a_{12}^-)$.

4.3 Envelopes

Therefore, the only lower bound we can have for $E = \varphi^{(n)}(\mathcal{S}_\mathbf{C}, u_{-n+1\to 0})$ is $\mathbf{b} = (b_1, \dots, b_M)$ where $\forall i, b_i = \min_{\mathbf{s}\in E} s_i$. Similarly, the upper bound is defined by $\mathbf{B} = (B_1, \dots, B_M)$ where $\forall i, B_i = \max_{\mathbf{s}\in E} s_i$. These bounds, called ***envelopes***, can be computed starting with $\mathbf{b}(1) = \mathbf{o}$, $\mathbf{B}(1) = \mathbf{C}$, associated to $\mathcal{S}_\mathbf{C}$, then computing $\mathbf{b}(k)$ and $\mathbf{B}(k)$ for $k = 2, \dots, n$, associated to $E_k = \varphi^{(k)}(\mathcal{S}_\mathbf{C}, u_{-n+1\to -n+k})$.

Because of these definitions, the evolution of $\mathbf{b}$ depends on both $\mathbf{b}$ and $\mathbf{B}$, such as the evolution of $\mathbf{B}$. For instance, with $M = 2$ queues, if $\mathbf{b}(k) = (b_1, b_2)$, $\mathbf{B}(k) = (B_1, B_2)$ and the event $u_{-n+k+1} = a_{ij}^-$, then if we denote $\mathbf{B}(k+1) = (B_1', B_2')$ we have two possibilities for B_2': if $b_1 = 0$, then $(0, B_2)$ ***may*** be in E_k. As $\varphi((0, B_2), a_{ij}^-) = (0, B_2)$, we have $B_2' = B_2$. Else, we know for sure that $B_2' = B_2 - 1$. Thus, we must define a new transition function $\varphi_- : \mathcal{S}_\mathbf{C} \times \mathcal{S}_\mathbf{C} \times \mathcal{A} \to \mathcal{S}_\mathbf{C}^2$ such that:

$$\forall k \in \{1, \dots n-1\}, (\mathbf{b}(k+1), \mathbf{B}(k+1)) = \varphi_-(\mathbf{b}(k), \mathbf{B}(k), u_{-n+k+1})$$

($\varphi_-^{(n)}$ is defined as previously).

- $\forall i, j, \varphi_-(\mathbf{b}, \mathbf{B}, a_{ij}) = (\varphi(\mathbf{b}, a_{ij}), \varphi(\mathbf{B}, a_{ij}))$ (non-trivial but true when $i, j \neq 0$)
- $\forall i, j \in \{1, \dots, M\}, \varphi_-(\mathbf{b}, \mathbf{B}, a_{ij}^-) = (\mathbf{b}', \mathbf{B}')$

$$\text{where } \mathbf{b}' = \begin{cases} \mathbf{b} - \mathbf{e_i} - \mathbf{e_j} & \text{if } \mathbf{b} - \mathbf{e_i} - \mathbf{e_j} \geqslant \mathbf{o} \\ \mathbf{b} - \mathbf{e_i} & \text{else, and if } \mathbf{b} - \mathbf{e_i} \geqslant \mathbf{o} \text{ } (Q_j \text{ may be empty}) \\ \mathbf{b} - \mathbf{e_j} & \text{else, and if } \mathbf{b} - \mathbf{e_j} \geqslant \mathbf{o} \text{ and } B_i > 0 \\ & (Q_i \text{ is not necessarily empty}) \\ \mathbf{b} & \text{else } (Q_i \text{ is empty or } b_i = b_j = 0). \end{cases}$$

$$\text{and } \mathbf{B}' = \begin{cases} \mathbf{B} - \mathbf{e_i} - \mathbf{e_j} & \text{if } \mathbf{B} - \mathbf{e_i} - \mathbf{e_j} \geqslant \mathbf{o} \text{ and } b_i > 0 \text{ } (Q_i \text{ is not empty}) \\ \mathbf{B} - \mathbf{e_i} & \text{else, and if } \mathbf{B} - \mathbf{e_i} \geqslant \mathbf{o} \\ & (Q_j \text{ is empty or } Q_i \text{ may be empty}) \\ \mathbf{B} & \text{else } (Q_i \text{ is empty}). \end{cases}$$

So, if the trajectories of the envelopes $\mathbf{b}$ and $\mathbf{B}$ couple, we know the system has coupled.

4.4 Main result

As in 3.4, we simulate two QNs. One of capacity $\mathbf{C}$ which envelopes are $(\mathbf{b}, \mathbf{B})$, another of capacity $\mathbf{C} - \mathbf{e_i}$ for a certain i, associated to $(\mathbf{b}', \mathbf{B}')$.

$\mathbf{b} = \mathbf{b}'$ until $\mathbf{b}$ and $\mathbf{B}$ couple. At this time, $\mathbf{b} = \mathbf{b}' = \mathbf{B}$, and $\exists j, \mathbf{B}' = \mathbf{b}' + \mathbf{e_j}$.

Again, some queues can decouple while others couple (at the same time). This is illustrated by figure 9.

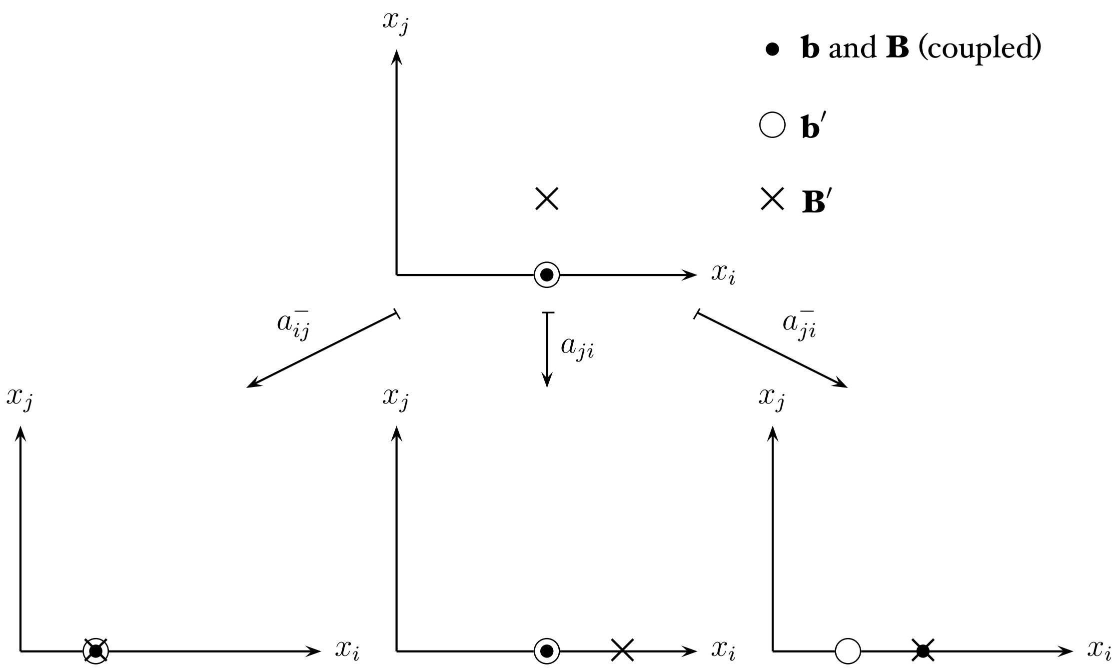

Figure 9: Effect of negative customers on trajectories.

The following table lists the events that may occur:

Before	• was with	Event	• will be with	After
$\mathbf{B}' = \mathbf{b}' + \mathbf{e_j}$ $b'_j = 0$	○	a_{ji}^-	×	$\mathbf{B}' = \mathbf{b}' + \mathbf{e_i}$
	○	a_{ji}	○	
	×	a_{ji}^-	○	
	×	a_{ij}^-	×	
	○ or ×	a_{ji}	○ and ×	Q_i and Q_j are coupled

○ and × stay close to •, therefore we can adapt [2] to prove that the mean coupling time for QN including negative customers is $O(\sum_i C_i)$.

5 Conclusion

We are now able, for any monotone QN, to generate a state according to the stationary distribution of the system in linear time in the size of the capacities of all queues (which is far better than $O(\prod_i C_i)$), and even with certain non-monotone QNs, which is surprising because the envelopes seemed to be relatively crude bounds at first sight.

A Values returned by Ψ^2

The algorithm used to generate these results is not exactly PSA (algorithm 1): instead of doubling τ at each loop, it increments it. Else, we would not be able to get accurate mean coupling times.

All entries are of the form τ_c/τ_a, where τ_c is the mean coupling time (over 1000 simulations) of the cyclic network, τ_a the mean coupling time of the acyclic network associated to it.

In bold, the entries represented in the graphs.

A.1 Positive customers

M \ C	5	10	15	20
5	140/127	277/274	423/421	568/573
10	290/266	572/547	861/836	1145/1133
15	**448**/412	**868**/830	**1316**/1280	**1758**/1720
20	609/565	1166/1111	1762/1711	2359/2299

A.2 Positive customers with critical case

See also figures 3 and 4.

M \ C	5	10	15	20
5	135/126	460/408	967/822	1624/1331
10	332/320	1135/1098	2433/2381	4288/4137
15	**550**/540	**1908**/1863	**4170**/4021	**7274**/7007
20	784/768	2795/2682	5971/5905	10396/10068

A.3 Positive and negative customers

See also figures 7 and 8.

M \ C	5	10	15	20
5	129/**120**	329/**300**	510/**509**	688/**717**
15	599/590	1416/1382	2116/2058	2735/2742

References

[1] Jantien Dopper, Bruno Gaujal, and Jean-Marc Vincent. Bounds for the Coupling Time in Queueing Networks Perfect Simulation. Research Report RR-5828, INRIA, 2006.

[2] Bruno Gaujal. Coupling Time in Markovian Queueing Networks. Technical report, INRIA, 2010.

[3] J. G. Propp and D. B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 9(1-2):223–252, 1996.