

Bound Computation of Dependability and Performance Measures

Stéphanie Mahévas, Gerardo Rubino

► **To cite this version:**

Stéphanie Mahévas, Gerardo Rubino. Bound Computation of Dependability and Performance Measures. [Research Report] RR-3135, INRIA. 1997. <inria-00073554>

HAL Id: inria-00073554

<https://hal.inria.fr/inria-00073554>

Submitted on 24 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

***Bound computation of dependability and
performance measures***

Stéphanie MAHÉVAS - Gerardo RUBINO

N° 3135

Mars 1997

_____ THÈME 1 _____



***R**apport
de recherche*

Bound computation of dependability and performance measures

Stéphanie MAHÉVAS* - Gerardo RUBINO†

Thème 1 — Réseaux et systèmes
Projet Model

Rapport de recherche n° 3135 — Mars 1997 — 34 pages

Abstract: When evaluating quantitative measures of complex systems using Markov models, a major drawback is the size of the generated state space, due to the intrinsic combinatorics associated with these models. If the number of states is excessively large, the model may be simply untractable. Observe also that in some cases, infinite models are appropriate, in which case, except for particular structures, the numerical procedures are not applicable.

Recently, methods have been proposed to deal with this problem. The idea is to derive bounds of the measures of interest, by replacing a large part of the state space by some compact information. This has been done mainly for asymptotic dependability measures, and the proposed methods are valid under specific and restrictive conditions.

In this paper, we extend these techniques to more general cases. In particular, we show that our approach can also give tight bounds of performance measures. We also show how to handle, in some cases, infinite models. We illustrate the new method with some analytically untractable open queueing networks, as well as with dependability models that cannot be analyzed by previous proposed techniques.

Key-words: dependability, performance, Markovian models, large state spaces, bound computation, equilibrium analysis

(Résumé : tsvp)

* mahevas@irisa.fr

† rubino@irisa.fr

Calcul de bornes de mesures de sûreté de fonctionnement et de performance

Résumé : L'évaluation quantitative de systèmes complexes par des modèles Markoviens a pour inconvénient majeure l'explosion combinatoire de la taille de l'espace d'états généré. Si ce nombre d'états est excessivement grand, l'analyse du modèle peut s'avérer impossible. Observons aussi que dans certains cas, les modèles appropriés sont infinis et donc pour lesquels, exceptées certaines structures particulières, on ne peut pas appliquer les techniques numériques.

Récemment, de nouvelles méthodes ont été proposées pour contrer ce problème. L'idée principale est d'obtenir des bornes des mesures d'intérêt en réduisant une très grande partie de l'espace d'états à l'aide d'information compacte. Cette approche a été essentiellement développée pour des mesures de sûreté de fonctionnement asymptotiques, et les méthodes déjà existantes ne sont valables que sous certaines conditions restrictives.

Dans ce papier, nous élargissons l'applicabilité de ces techniques à des modèles plus généraux. En particulier, nous montrons que ce type de méthodes peut aussi donner des bornes de mesures de performance de bonne qualité. Nous démontrons aussi comment les étendre, dans certains cas, aux modèles infinis. Enfin, nous illustrons notre méthode avec des réseaux de files d'attente ouverts analytiquement intraitables, ainsi qu'avec des modèles en sûreté de fonctionnement qui n'étaient pas solutionnés par les précédentes méthodes.

Mots-clé : sûreté de fonctionnement, performance, modèle Markoviens, grands espaces d'états, calcul de bornes, analyse en équilibre

1 Introduction

To derive performance, dependability or performability measures from a model of a complex system, Markov chains, under different forms, are the most widely used mathematical tools. Sometimes, the user directly builds a Markov chain from system specifications. Most often, the model is described in a higher level language such as queues or networks of queues, stochastic Petri nets, etc., and some tool constructs the stochastic process automatically. The usefulness of Markov chains is due to the power of the theory and to the efficient algorithmic technology associated with. However, such a power has a price. There are two major drawbacks when using Markov models. The first one is the fact that, to be able to represent the more and more complex systems built nowadays, the state spaces are larger and larger. In general, models having, say, hundreds or thousands of millions of states are out of scope for exact numerical analysis in equilibrium. The second one is the so-called “rare events” problem, meaning that in many cases (typically in models of highly available systems), the interesting events (typically, the fact that the system is down) have very low probabilities, making problematic the use of Monte Carlo techniques. This is usually related to high numerical values of the ratios between different transition rates of the chain, which leads to numerical problems with the exact analysis of the models (stiffness). Often, the two problems appear simultaneously. To deal with them, a different approach has been recently developed. It consists of computing bounds of the desired measures instead of exact values or statistical estimations [1], [2], [3]. This is the subject of the paper.

Bounding techniques have been mainly developed for cases in which, at the same time, the model is large and stiff. The second aspect is in fact used to help with dealing with the first one, as we see in the paper. The intuition behind the approach is quite simple. When the model is stiff, the stochastic process spends most of its time in a (small) part of the state space. It is then a natural attempt to try to approximate the interesting measures by working basically with that part of the state space. This is done by replacing, in some way, the rest of the state space by “a few” states. All the difficulty is how to handle that large part of the state space of the model in order to have some control on the accuracy of the method.

The system is represented by a continuous time homogeneous and irreducible Markov chain X over the finite state space S with stationary distribution $\boldsymbol{\pi}$ (row vector). We denote $\pi_i = \Pr(X_\infty = i)$ where X_∞ is a stationary version of X . With each state i we associate a reward $r_i \geq 0$; \mathbf{r} is the (row) vector of rewards. This paper deals with the computation of

$$R = \mathbb{E}(r_{X_\infty}) = \sum_{i \in S} r_i \pi_i = \boldsymbol{\pi} \mathbf{r}^T.$$

For instance, in a dependability context, the states are in general called “operational” when they represent a system delivering its service as expected, and “unoperational” otherwise. If a reward equal to 1 is associated with the formers and equal to 0 to the latters, then R is the asymptotic availability of the system. In a performance context, suppose that the model is a queueing network and that you are interested in the mean number of customers

in queue q . If with each state i we associate a reward equal to the number of customers in station q when the model is in state i , the expectation R is equal to the desired mean number of customers in q . As a second example, if r_i is the number of active processors in some model of a fault-tolerant multi-processor computer when its state is i , then R is a performability measure (the mean *power* of the system).

In this paper we develop an approach that avoids an important drawback of previous work, by extending the class of evaluable pairs (models, measures). The main restriction of previous techniques, which mainly concern dependability models, is that “almost” every state must have at least one transition corresponding to a repair. The approach proposed here works without this condition, but it has a price: in some cases, large linear systems must still be solved. In the paper we illustrate the method with cases where these systems are easy to solve (and such that previous techniques do not apply). In a performance context, it is frequent to work with infinite state spaces. The techniques we propose can, in some situations, deal with these cases as well. This is also developed and illustrated here. The paper is organized as follows. After the introduction in next section, we state some preliminary notation in Section 2 and we develop the main necessary tools to derive our bounding technique in Sections 3, 4, 5 and 6. In Section 6 we also recall the main principles of the method of [3]. Section 7 presents the new approach and Section 8 deals with the infinite state space case. In Section 9 we give examples, both in the dependability and in the performance areas, both in the finite and the infinite cases. Section 10 concludes the paper.

2 Preliminaries

Generalities. We are given a finite and irreducible continuous time homogeneous Markov chain X over the state space S , presumed to model some complex system. We denote by A the infinitesimal generator of X . The asymptotic distribution of X is denoted by π and it will be seen in the paper as a row vector. We then have

$$\pi A = \mathbf{0}.$$

We are given also a vector of positive reals, $\mathbf{r} = (\dots r_i \dots)$ over S . The goal is the evaluation of bounds of $R = \pi \mathbf{r}^T$, *without computing vector π* . To do this, we assume that we know a lower and an upper bound of the rewards, that is, two reals ϱ_1 and ϱ_2 such that for all $i \in S$,

$$0 \leq \varrho_1 \leq r_i \leq \varrho_2 < +\infty.$$

We will see in Section 8, an extension to the case $\varrho_2 = +\infty$. The state space S is assumed to be decomposed (or decomposable) into two disjoint sets, denoted by G and \bar{G} . The first one will be the set of effectively “generated” states (basically, those that are frequently visited by the chain in equilibrium). The techniques discussed here attempt to give bounds of the asymptotic reward R by solving linear systems basically on G . The good situation is then to have $|G| \ll |S|$.

Some notation associated with a subset of states. For any subset of states $C \subseteq S$, we denote

- π_C , the restriction of π to the set C (row vector with size equal to $|C|$), \mathbf{r}_C , the restriction of \mathbf{r} to C , etc.,
- $\pi(C) = \Pr(X_\infty \in C) = \sum_{i \in C} \pi_i = \pi_C \mathbf{1}^T$, where $\mathbf{1}^T$ is the transpose of a row vector having all its entries equal to 1, the dimension being defined by the context,
- $\hat{\pi}_C =$ distribution of X_∞ conditioned to the event $\{X_\infty \in C\}$, (row) vector with size $|C|$, $\hat{\pi}_C = \frac{1}{\pi(C)} \pi_C$,
- $\bar{C} =$ the complement $S - C$ of C ,
- $in(C) = \{j \in C \text{ such that there exists } i \in \bar{C} \text{ with } A_{i,j} > 0\}$,
- $out(C) = \{i \in C \text{ s.t. there exists } j \in \bar{C} \text{ with } A_{i,j} > 0\}$,
- A_C , the block of A corresponding to the transitions inside subset C ,
- $A_{C,C'}$, the block of A corresponding to the transitions from subset C to another subset C' .

3 Forcing the entries in G by a fixed state j

The main idea comes from the basic initial work by Courtois and Semal. It consists of building a family of Markov chains derived from X in the following way. For each state $j \in in(G)$, let us construct the new continuous time homogeneous Markov chain $X^{(j)}$, by forcing the transitions from \bar{G} into G to enter by state j , as illustrated in figure 1. The infinitesimal generator of $X^{(j)}$ is denoted by $A^{(j)}$:

$$A^{(j)} = \begin{pmatrix} A_G & A_{G\bar{G}} \\ A_{\bar{G}G}^{(j)} & A_{\bar{G}} \end{pmatrix}.$$

The transition rate from any state $i \in \bar{G}$ to j is equal to $\sum_{l \in in(G)} A_{i,l}$. In other words, $A_{\bar{G}G}^{(j)} = A_{\bar{G}G} \mathbf{1}^T e_j$, where e_j is the j th row vector of the canonical base in $\mathbb{R}^{|G|}$. The other transition rates of $X^{(j)}$, that is inside G , from G to \bar{G} and inside \bar{G} , are as in X . First, we prove that $X^{(j)}$ has an unique stationary distribution by means of the following lemma.

Lemma 1 *The Markov chain $X^{(j)}$ has an unique recurrent class which includes the state j .*

Proof. Denote by S_j the class of states i such that i is reachable from j and j is reachable from i . The claim implies that S_j can be reached from every state $k \in S$. Assume $S_j \neq S$ and let $k \in S - S_j$. If $k \in \text{out}(\bar{G})$, by definition of $\text{out}()$, k is connected to j . If $k \in \bar{G} - \text{out}(\bar{G})$ then there is necessarily a path from k to some $l \in \text{out}(\bar{G})$ (since X is irreducible), which is completely included in \bar{G} , and by definition of $X^{(j)}$, l is connected to j . It remains the case of a state $k \in G - \{j\}$. If, in X , there is a path from k to j completely included in G , we are done. If not, since X is irreducible, there is at least a path from k to j passing through \bar{G} , thus entering for the first time \bar{G} by some state l and then, we are in the first discussed case. ■

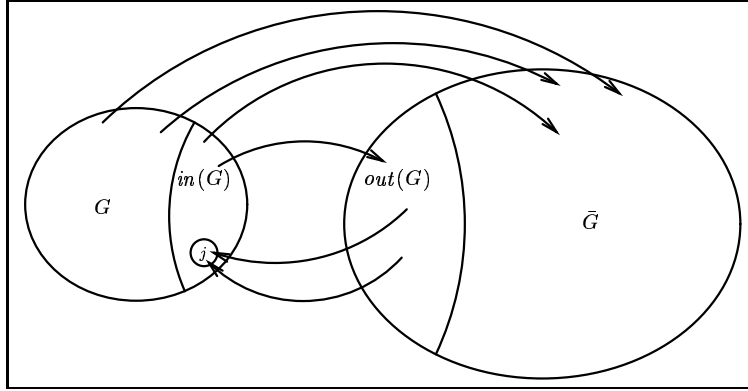


Figure 1: The topology of $X^{(j)}$

It is easy to see that there can be transient states in $X^{(j)}$. For instance, think of this simple situation: some state $i \in G$ has only one transition out to j and one transition in from $k \in \bar{G}$. In $X^{(j)}$, such a state is transient.

Since $X^{(j)}$ has an unique recurrent class, it has an unique stationary distribution which we denote by $\pi^{(j)}$. Then, for each state $j \in \text{in}(G)$, we have the following result, which is a slight extension of results in [1]:

Theorem 1 *There exists a positive vector β of size $|\text{in}(G)|$, such that $\beta \mathbf{1}^T = 1$ and*

$$\sum_{j \in \text{in}(G)} \beta_j \pi^{(j)} = \pi.$$

Proof. See Appendix A.

In [1], the authors show this result under the assumption that the chains are irreducible. In the Appendix we show that the only necessary assumption is the existence of an unique stationary distribution for $X^{(j)}$. We also show that the result is still valid in the infinite state space case (see Section 8).

From this theorem, we derive the two following immediate results, put together as a corollary.

Corollary 1 *If we denote $\pi^{(j)}(G) = P(X_\infty^{(j)} \in G)$, we have*

$$\min_{j \in \text{in}(G)} \pi^{(j)}(G) \leq \pi(G), \quad (1)$$

and if $R^{(j)} = \pi^{(j)} \mathbf{r}^T$, we have

$$\min_{j \in \text{in}(G)} R^{(j)} \leq R \leq \max_{j \in \text{in}(G)} R^{(j)}. \quad (2)$$

Proof. Let us denote by \mathbf{u} the row vector equal to $(\mathbf{1}_G, \mathbf{0}_{\bar{G}})$, where $\mathbf{1}_G$ (respectively $\mathbf{0}_{\bar{G}}$) is the row vector having all its entries equal to 1 (respectively 0), the dimension being equal to $|G|$ (respectively to $|\bar{G}|$). From Theorem 1, we have

$$\pi(G) = \pi \mathbf{u}^T = \sum_{j \in \text{in}(G)} \beta_j \pi^{(j)} \mathbf{u}^T = \sum_{j \in \text{in}(G)} \beta_j \pi^{(j)}(G).$$

Since $\beta \mathbf{1}^T = 1$ and $\beta \geq 0$, we have

$$\min_{j \in \text{in}(G)} \pi^{(j)}(G) \leq \sum_{j \in \text{in}(G)} \beta_j \pi^{(j)}(G).$$

In the same way, from $R = \pi \mathbf{r}^T$, we have

$$R = \sum_{j \in \text{in}(G)} \beta_j \pi^{(j)} \mathbf{r}^T \geq \min_{j \in \text{in}(G)} \pi^{(j)} \mathbf{r}^T$$

and the same for the upper bound. ■

Corollary 1 gives the expressions that will be used to derive bounds of R . In the sequel, we will develop a general approach to build a lower bound of $\min_{j \in \text{in}(G)} R^{(j)}$ and an upper bound of $\max_{j \in \text{in}(G)} R^{(j)}$.

4 Partitioning the state space

To go further, we suppose that we are given a partition $\{C_I, I = 0, 1, \dots, M\}$ of S , an integer K with $0 < K < M$, and that

$$G = \bigcup_{I=0}^{K-1} C_I.$$

In a performance context, assuming that we deal with something like a queueing network or a stochastic Petri net, C_I can be, for instance, the set of states where the system, or some

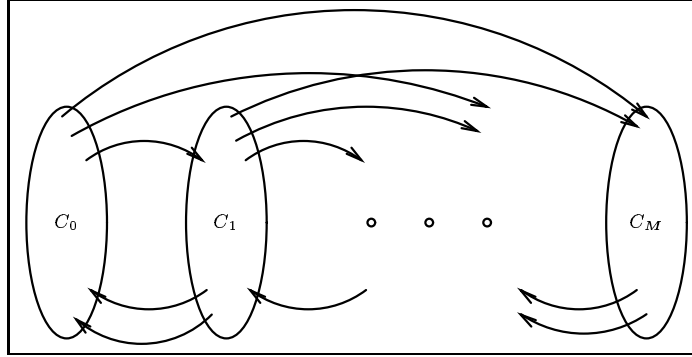


Figure 2: Chain X and the partition $\{C_I, I = 0, 1, \dots, M\}$ of S

of its sub-systems, has I customers or tokens. In a dependability case, if we work with a model of some fault-tolerant multi-component system, C_I can be, for instance, the set of states corresponding to I operational components. The good property for such a partition is that the higher the index I , the lower the probability that X_∞ belongs to the set C_I .

In what follows, we assume that transitions from class C_I to class C_J are not allowed if $J \leq I - 2$, that is, that the following condition holds:

Condition 1 For any two integer indices I and J such that $J \leq I - 2$, for any states $i \in C_I$ and $j \in C_J$ we have $A_{i,j} = 0$.

Observe that, given the irreducibility of X , this implies that for all index $I > 0$ there are at least two states $i \in C_I$ and $j \in C_{I-1}$ such that $A_{i,j} > 0$.

Following with the examples used a few lines before, in the case of a queuing model or a Petri net, this means, for instance, that simultaneous departures are not allowed. In the dependability example, the condition says that simultaneous repairs are not allowed.

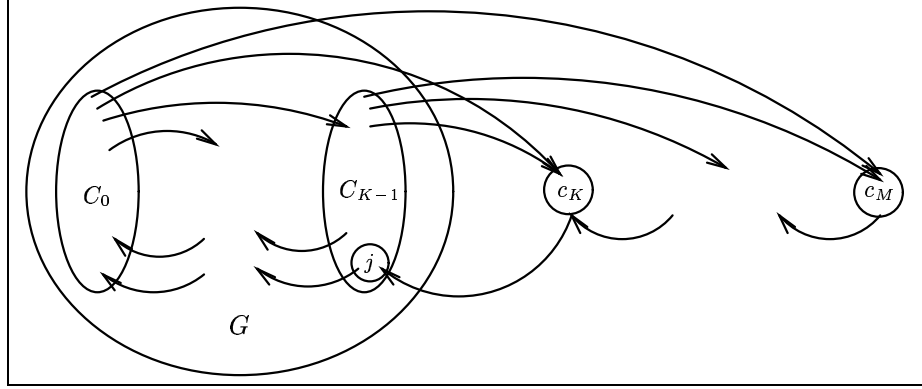
5 Aggregation of states

For each state $j \in in(G)$, we will consider the following aggregation of $X^{(j)}$. We define a continuous time homogeneous Markov chain $X^{(j)agg}$ which is constructed from $X^{(j)}$ by collapsing the classes C_K, C_{K+1}, \dots, C_M of the partition into single states c_K, c_{K+1}, \dots, c_M .

If we denote by $A^{(j)agg}$ the infinitesimal generator of $X^{(j)agg}$, recalling that $\pi^{(j)}(C_I)$ is the probability that $X_\infty^{(j)}$ belongs to C_I , the transition rates of $X^{(j)agg}$ are given by the following expressions:

- for all $h \in G$ (or $h \in out(G)$), and for all $I \geq K$,

$$A_{h,c_I}^{(j)agg} = \sum_{i \in C_I} A_{h,i}, \quad (3)$$

Figure 3: The topology of $X^{(j)\text{agg}}$

- for any $I > K$,

$$A_{c_I, c_{I-1}}^{(j)\text{agg}} = \mu_I^{(j)} = \frac{\sum_{i \in C_I} \pi_i^{(j)} \sum_{l \in C_{I-1}} A_{i,l}}{\pi^{(j)}(C_I)}, \quad (4)$$

-

$$A_{c_K, j}^{(j)\text{agg}} = \mu_K^{(j)} = \frac{\sum_{i \in C_K} \pi_i^{(j)} \sum_{l \in C_{K-1}} A_{i,l}}{\pi^{(j)}(C_K)}, \quad (5)$$

- for any $I \geq K$ and $J > I$,

$$A_{c_I, c_J}^{(j)\text{agg}} = \lambda_{I,J}^{(j)} = \frac{\sum_{i \in C_I} \pi_i^{(j)} \sum_{l \in C_J} A_{i,l}}{\pi^{(j)}(C_I)}. \quad (6)$$

We denote by $S^{\text{agg}} = G \cup \{c_K, c_{K+1}, \dots, c_M\}$ the state space of $X^{(j)\text{agg}}$. Since $X^{(j)}$ has a unique recurrent class (Lemma 1), it is immediate to see that $X^{(j)\text{agg}}$ has also a unique recurrent class including j . Let us denote $\pi_i^{(j)\text{agg}} = \Pr(X_\infty^{(j)\text{agg}} = i)$, $i \in S^{\text{agg}}$, where $X_\infty^{(j)\text{agg}}$ denotes a stationary version of $X^{(j)\text{agg}}$. We then have the following well known result on aggregation:

$$\text{for any } g \in G, \pi_g^{(j)\text{agg}} = \pi_g^{(j)}, \text{ and for any } I \geq K, \pi_{c_I}^{(j)\text{agg}} = \pi^{(j)}(C_I). \quad (7)$$

The chain $X^{(j)\text{agg}}$ is called in [1] “the exact aggregation of $X^{(j)}$ with respect to the given partition”. We adopt here this terminology (in [4], it is called “the pseudo-aggregation of $X^{(j)}$ w.r.t. the given partition”). Of course, to build it we need the stationary distribution $\pi^{(j)}$ of $X^{(j)}$, which is unknown. We define in the next subsection another chain which “bounds”, in some way, chain $X^{(j)\text{agg}}$, and from which the desired bounds of R will be computed.

6 “Bounding” the Markov chain $X^{(j)\text{agg}}$

For each state $j \in \text{in}(G)$, let us define a homogeneous and irreducible Markov chain $Y^{(j)}$ over S^{agg} with the same topology as the aggregated chain $X^{(j)\text{agg}}$, in the following way. We keep the same transition rates inside the subset G and from G to the c_I 's, which are computable without the knowledge of the stationary distribution of $X^{(j)}$. Then, we replace the exact aggregated rates from c_I to c_J , $I < J$, by some $\lambda_{I,J}^+$, the exact aggregated rates from c_I to c_{I-1} , by some μ_I^- , and the exact aggregated rate from c_I to j , by μ_K^- in such a way that

$$\lambda_{I,J}^+ \geq \lambda_{I,J}^{(j)}, \quad (8)$$

$$0 < \mu_I^- \leq \mu_I^{(j)}. \quad (9)$$

Then, between the three chains $X^{(j)}$, $X^{(j)\text{agg}}$ and $Y^{(j)}$, the following relation holds.

Theorem 2 *If we denote by $\mathbf{y}^{(j)}$ the stationary distribution of $Y^{(j)}$, we have*

$$\hat{\pi}_G^{(j)} = \hat{\pi}_G^{(j)\text{agg}} = \hat{\mathbf{y}}_G^{(j)}. \quad (10)$$

Proof. Let us denote by $P^{(j)}$ (respectively by $P'^{(j)}$) the stochastic matrix of $X^{(j)}$ (respectively of $Y^{(j)}$). We have:

$$P^{(j)} = \begin{pmatrix} P_G & P_{G\bar{G}} \\ P_{\bar{G}G}^{(j)} & P_{\bar{G}} \end{pmatrix}, \quad P'^{(j)} = \begin{pmatrix} P_G & P'_{G\bar{G}} \\ P_{\bar{G}G}'^{(j)} & P'_{\bar{G}} \end{pmatrix},$$

where the matrix $P'_{G\bar{G}}$ is equal to $(P_{GC_K} \mathbf{1}^T, \dots, P_{GC_M} \mathbf{1}^T)$.

Given that $\pi^{(j)} P^{(j)} = \pi^{(j)}$ and $\mathbf{y}^{(j)} P'^{(j)} = \mathbf{y}^{(j)}$, we have:

$$\begin{cases} \hat{\pi}_G^{(j)} \left(P_G - P_{G\bar{G}} P_{\bar{G}}^{-1} P_{\bar{G}G}^{(j)} \right) = \hat{\pi}_G^{(j)} \\ \hat{\mathbf{y}}_G^{(j)} \left(P_G - P'_{G\bar{G}} (P'_{\bar{G}})^{-1} P_{\bar{G}G}'^{(j)} \right) = \hat{\mathbf{y}}_G^{(j)} \end{cases}$$

From Appendix A, Lemma 8, it follows that $\hat{\pi}_G^{(j)}$ and $\hat{\mathbf{y}}_G^{(j)}$ belong to the polyhedron $\mathcal{P}((I - P_G)^{-1})$.

From Appendix A, Lemma 9, $\hat{\pi}_G^{(j)}$ and $\hat{\mathbf{y}}_G^{(j)}$ are both equal to the j th vertex of $\mathcal{P}((I - P_G)^{-1})$.

From (7), we have directly the first equality between $\hat{\pi}_G^{(j)}$ and $\hat{\pi}_G^{(j)\text{agg}}$ ■

Theorem 3 *Between $\mathbf{y}^{(j)}(G) = \Pr(Y_\infty^{(j)} \in G)$ where $Y_\infty^{(j)}$ is a stationary version of $Y^{(j)}$, and $\pi^{(j)\text{agg}}(G) = \pi^{(j)}(G)$, we have the relation*

$$\mathbf{y}^{(j)}(G) \leq \pi^{(j)}(G). \quad (11)$$

If one of the inequalities (8) or (9) is strict, then $\mathbf{y}^{(j)}(G) < \pi^{(j)}(G)$.

Proof. See Appendix B. ■

Define over S^{agg} the two reward vectors \mathbf{r}_1 and \mathbf{r}_2 obtained by completing vector \mathbf{r}_G with rewards on the aggregated states c_K, c_{K+1}, \dots, c_M equal to ϱ_1 in \mathbf{r}_1 and equal to ϱ_2 in \mathbf{r}_2 .

Theorem 4

$$\min_{j \in \text{in}(G)} \mathbf{y}^{(j)} \mathbf{r}_1^T \leq R, \quad (12)$$

$$\max_{j \in \text{in}(G)} \mathbf{y}^{(j)} \mathbf{r}_2^T \geq R. \quad (13)$$

Proof. Let us consider the expression of $\mathbf{y}^{(j)} \mathbf{r}_1^T$:

$$\begin{aligned} \mathbf{y}^{(j)} \mathbf{r}_1^T &= \mathbf{y}_G^{(j)} \mathbf{r}_G^T + \varrho_1 y^{(j)}(\bar{G}) \\ &= y^{(j)}(G) \left(\hat{\mathbf{y}}_G^{(j)} \mathbf{r}_G^T - \varrho_1 \right) + \varrho_1. \end{aligned}$$

Given that $\pi^{(j)}(G) = \pi^{(j)\text{agg}}(G)$ from (7), that $\hat{\mathbf{y}}_G^{(j)} \mathbf{r}_G^T \geq \varrho_1$ (since $r_i \geq \varrho_1$ for all state i) and using Theorem 2, we have the following inequality:

$$\begin{aligned} \mathbf{y}^{(j)} \mathbf{r}_1^T &\leq \pi^{(j)}(G) \left(\hat{\pi}_G^{(j)} \mathbf{r}_G^T - \varrho_1 \right) + \varrho_1 \\ &= \pi^{(j)} \mathbf{r}_1^T = R_1^{(j)}. \end{aligned}$$

From Corollary 1, we have:

$$\min_{j \in \text{in}(G)} \mathbf{y}^{(j)} \mathbf{r}_1^T \leq \min_{j \in \text{in}(G)} R_1^{(j)} \leq \min_{j \in \text{in}(G)} R^{(j)} \leq R.$$

In the same way, denoting $R_2^{(j)} = \pi^{(j)} \mathbf{r}_2^T$ and writing

$$\mathbf{y}^{(j)} \mathbf{r}_2^T = y^{(j)}(G) \left(\hat{\mathbf{y}}_G^{(j)} \mathbf{r}_G^T - \varrho_2 \right) + \varrho_2$$

and observing that $\hat{\mathbf{y}}_G^{(j)} \mathbf{r}_G^T - \varrho_2 \leq 0$, we obtain

$$\mathbf{y}^{(j)} \mathbf{r}_2^T \geq R_2^{(j)},$$

and thus

$$\max_{j \in \text{in}(G)} \mathbf{y}^{(j)} \mathbf{r}_2^T \geq \max_{j \in \text{in}(G)} R_2^{(j)} \geq \max_{j \in \text{in}(G)} R^{(j)} \geq R.$$

The approach of [3]. It remains the problem of the computation of the bounds $\lambda_{I,J}^+$, μ_I^- . Since we want to compute them *without* the knowledge of $\pi^{(j)}$, the (unknown) stationary distribution of $X^{(j)}$ for each $j \in in(G)$, it would be nice to use

$$\forall I, J \text{ s.t. } K \leq I < J \leq M, \lambda_{I,J}^+ = \max_{i \in C_I} \sum_{l \in C_J} A_{i,l}, \quad (14)$$

$$\forall I \text{ s.t. } K \leq I \leq M, \mu_I^- = \min_{i \in C_I} \sum_{l \in C_{I-1}} A_{i,l}, \quad (15)$$

This is the idea followed in [3]. The use of relation (14) has as an immediate consequence that $\lambda_{I,J}^+ > 0$. Let us examine now the bounds μ_I^- . In order to have $\mu_I^- > 0$ for any value of K , we need a supplementary condition to be satisfied by X :

Condition 2 *For any index $I \neq 0$, for all state $i \in C_I$ there exists at least a state $j \in C_{I-1}$ such that $A_{i,j} > 0$.*

This can be quite restrictive as we will illustrate later, but the interest relies in the fact that it allows to obtain direct lower bounds of the $\mu_I^{(j)}$'s. In Section 7, we develop a new approach that does not need this assumption, allowing to work with much more general models.

Other related works. Before presenting the method that we propose, let us briefly describe other related papers in the area. To improve the bounds obtained by the algorithm of [3], a technique of “duplication of states” is proposed in [5]. The main objective is to reduce the number of linear systems to solve. The technique also allows to improve the bounds by increasing the threshold K without restarting the work from the beginning. That is, the results for level $K + 1$ can use those corresponding to level K . Other line of research is the work by Carrasco in [6] which, improving previous work by the same author, gives better bounds when additional information (“distance to failure”) is available. In that paper the number of linear systems to be solved is also reduced. Let us also mention [7] where an application of [1] and [2] to the particular case of availability analysis with tridiagonal block transition rate matrices is proposed.

7 The proposed method

This is the main part of the paper. Its goal is to derive a method to avoid Condition 2 and still be able to bound the measure R . We need some results of [4], where the authors analyze the sojourn times of a Markov chain X in a part of its state space, and the asymptotic behaviour of these (in general dependent) random variables.

7.1 Sojourn times and aggregation of states

Let us denote by $H_{I,n}^{(j)}$ the length of the n th sojourn of $X_\infty^{(j)}$ in class C_I . The first visited state of C_I during this sojourn is denoted by $V_{I,n}^{(j)}$ ($V_{I,n}^{(j)} \in \text{in}(C_I)$) and after leaving C_I , the next visited state is denoted by $W_{I,n}^{(j)}$ ($W_{I,n}^{(j)} \in \text{in}(\bar{C}_I)$).

The distribution of $V_{I,n}^{(j)}$, as a row vector $\mathbf{v}_I^{(j)}(n)$ defined over C_I , is given by the following expression [4]:

$$\mathbf{v}_I^{(j)}(n) = \mathbf{v}_I^{(j)}(1)(B_{C_I}^{(j)})^{n-1},$$

where $B_{C_I}^{(j)}$ is the stochastic matrix $A_{C_I}^{-1}A_{C_I, \bar{C}_I}^{(j)}A_{\bar{C}_I, C_I}^{-1}$ and $\mathbf{v}_I^{(j)}(1) = \boldsymbol{\pi}_{C_I}^{(j)} - \boldsymbol{\pi}_{\bar{C}_I}^{(j)}A_{\bar{C}_I, C_I}^{-1}A_{C_I, \bar{C}_I}^{(j)}$. Of course for all $n > 1$, $\mathbf{v}_I^{(j)}(n)$ has non-zero entries only on states i belonging to $\text{in}(C_I)$.

The distribution of $H_{I,n}^{(j)}$ is given by [4]

$$\Pr(H_{I,n}^{(j)} > t) = \mathbf{v}_I^{(j)}(n) \exp(-A_{C_I}t) \mathbf{1}^T$$

and its mean is

$$\mathbb{E}(H_{I,n}^{(j)}) = -\mathbf{v}_I^{(j)}(n)A_{C_I}^{-1}\mathbf{1}^T.$$

In [4] it is in particular shown that vector

$$\mathbf{v}_I^{(j)} = \frac{\boldsymbol{\pi}_{C_I}^{(j)}A_{C_I}}{\boldsymbol{\pi}_{C_I}^{(j)}A_{C_I}\mathbf{1}^T} \quad (16)$$

is the stationary distribution of the Markov chain $(V_{I,n}^{(j)})_n$, that is, if $\mathbf{v}_I^{(j)}(1) = \mathbf{v}_I^{(j)}$ then $\mathbf{v}_I^{(j)}(n) = \mathbf{v}_I^{(j)}$ for all $n \geq 1$. We denote by $v_{I,i}^{(j)}$ the component of $\mathbf{v}_I^{(j)}$ corresponding to state $i \in C_I$. We should also note that $\mathbf{v}_G^{(j)} = \mathbf{e}_j$, where \mathbf{e}_j is the j th row vector of the canonical base in $\mathbb{R}^{|\mathcal{G}|}$. Let us consider now some relationships between chains $X^{(j)}$ and $X^{(j)\text{agg}}$ from the sojourn time point of view. We denote by $h_I^{(j)}$ the mean holding time of $X^{(j)\text{agg}}$ in state c_I , $K \leq I \leq M$, that is,

$$\text{for } I \geq K, \quad h_I^{(j)} = \frac{1}{\mu_I^{(j)} + \sum_{J>I} \lambda_{I,J}^{(j)}}, \quad (17)$$

A result needed here is given in the following lemma:

Lemma 2

$$\text{For any } I \geq K, \quad h_I^{(j)} = -\mathbf{v}_I^{(j)}A_{C_I}^{-1}\mathbf{1}^T. \quad (18)$$

For the proof, see [4], basically Corollary 4.6. Lemma 2 says that the mean holding time (i.e. the mean sojourn time) of $X^{(j)\text{agg}}$ in c_I is equal to the mean sojourn time of $X^{(j)}$ in C_I when it enters C_I by state i with probability $v_{I,i}^{(j)}$ (for instance, think of the first sojourn in C_I of a version of $X^{(j)}$ having as initial distribution the vector $\boldsymbol{\alpha}^{(j)}$ such that $\boldsymbol{\alpha}_{C_I}^{(j)} = \mathbf{v}_I^{(j)}$).

Let us denote by $\widehat{h}_{i,I}$ the mean sojourn time in C_I conditioned to the fact that the process enters the set C_I by state i . Observe that, for all $j \in in(G)$,

$$\widehat{h}_{i,I} = \mathbb{E}(H_{I,n}^{(j)} \mid V_{I,n}^{(j)} = i) \text{ for all } n \geq 1. \quad (19)$$

From Relations (18) and (17), we can write

$$h_I^{(j)} = \sum_{i \in in(C_I)} v_{I,i}^{(j)} \widehat{h}_{i,I}. \quad (20)$$

Now, for the purposes of this paper, we have to consider the event “when the n th sojourn of $X^{(j)}$ in C_I ends, the next visited state belongs to C_{I-1} ”, that is, $\{W_{I,n}^{(j)} \in C_{I-1}\}$. It is straightforward to verify that the probability of this event is

$$\text{for } I > K, \quad \Pr(W_{I,n}^{(j)} \in C_{I-1}) = -\mathbf{v}_I^{(j)}(n) A_{C_I}^{-1} A_{C_I, C_{I-1}} \mathbf{1}^T.$$

When $I = K$, we also have

$$\Pr(W_{K,n}^{(j)} = j \in in(G)) = -\mathbf{v}_K^{(j)}(n) A_{C_K}^{-1} A_{C_K, C_{K-1}} \mathbf{1}^T,$$

where $A_{C_K, C_{K-1}}^{(j)}$ is the matrix equal to $A_{C_K, C_{K-1}} \mathbf{1}^T \mathbf{e}_j$, and \mathbf{e}_j is the j th row vector of the canonical base in $\mathbb{R}^{|C_K|}$.

The event similar to $\{W_{I,n}^{(j)} \in C_{I-1}\}$ in the aggregated chain $X^{(j)\text{agg}}$ is “when leaving state c_I , the chain jumps to c_{I-1} ”. Its probability is

$$p_I^{(j)} = \frac{\mu_I^{(j)}}{\mu_I^{(j)} + \sum_{J>I} \lambda_{I,J}^{(j)}}.$$

The following result (similar to Lemma 2) holds:

Lemma 3

$$\text{For any } I > K, \quad p_I^{(j)} = -\mathbf{v}_I^{(j)} A_{C_I}^{-1} A_{C_I, C_{I-1}} \mathbf{1}^T \quad (21)$$

and

$$p_K^{(j)} = -\mathbf{v}_K^{(j)} A_{C_K}^{-1} A_{C_K, C_{K-1}} \mathbf{1}^T. \quad (22)$$

Proof.

Let $I > K$. From (6) and from (4), we can write

$$\lambda_{I,J}^{(j)} = \frac{\pi_{C_I}^{(j)} A_{C_I, C_J} \mathbf{1}^T}{\pi_{C_I}^{(j)} \mathbf{1}^T}, \quad \mu_I^{(j)} = \frac{\pi_{C_I}^{(j)} A_{C_I, C_{I-1}} \mathbf{1}^T}{\pi_{C_I}^{(j)} \mathbf{1}^T}.$$

Then,

$$\begin{aligned}\mu_I^{(j)} + \sum_{J>I} \lambda_{I,J}^{(j)} &= \frac{\pi_{C_I}^{(j)} A_{C_I, \bar{C}_I} \mathbf{1}^T}{\pi_{C_I}^{(j)} \mathbf{1}^T} \\ &= -\frac{\pi_{C_I}^{(j)} A_{C_I} \mathbf{1}^T}{\pi_{C_I}^{(j)} \mathbf{1}^T} \\ &\quad (\text{since } A_{C_I, \bar{C}_I} \mathbf{1}^T = -A_{C_I} \mathbf{1}^T).\end{aligned}$$

This leads to

$$\begin{aligned}p_I^{(j)} &= \frac{\mu_I^{(j)}}{\mu_I^{(j)} + \sum_{J>I} \lambda_{I,J}^{(j)}} \\ &= \frac{\pi_{C_I}^{(j)} A_{C_I, C_{I-1}} \mathbf{1}^T}{\pi_{C_I}^{(j)} \mathbf{1}^T} \left(-\frac{\pi_{C_I}^{(j)} \mathbf{1}^T}{\pi_{C_I}^{(j)} A_{C_I} \mathbf{1}^T} \right) \\ &= -\frac{\pi_{C_I}^{(j)} A_{C_I, C_{I-1}} \mathbf{1}^T}{\pi_{C_I}^{(j)} A_{C_I} \mathbf{1}^T}.\end{aligned}$$

It remains to check that the last expression is equal to $-\mathbf{v}_I^{(j)} A_{C_I}^{-1} A_{C_I, C_{I-1}} \mathbf{1}^T$. The case of class C_K is proven in the same way. ■

As for Lemma 2, Lemma 3 says that the probability that $X^{(j)\text{agg}}$ will jump to C_{I-1} when leaving C_I is the same as the conditional probability for $X^{(j)}$ to jump to C_{I-1} when leaving C_I , given that $X^{(j)}$ enters C_I by state i with probability $v_{I,i}^{(j)}$. This implies that, if we denote by $\hat{p}_{i,I}$ the conditional probability that $X^{(j)}$ jumps to C_{I-1} when leaving C_I , given that the sojourn started in state $i \in C_I$, we have

$$p_I^{(j)} = \sum_{i \in \text{in}(C_I)} v_{I,i}^{(j)} \hat{p}_{i,I}. \quad (23)$$

7.2 The bounding algorithm

Let us assume that Condition 2 is not satisfied. To obtain the bounds of R given in (12) and (13), we proceed as in [3]. The problem is the computation of lower bounds of the $\mu_I^{(j)}$'s.

First let us consider a new subset of states of C_I , $\text{in}(C_I)^*$, which is the set of entry points i of C_I such that if $X^{(j)}$ enters C_I by i , there is a non-null probability that the next visited class is C_{I-1} :

$$\text{in}(C_I)^* = \{i \in \text{in}(C_I) \mid \hat{p}_{i,I} > 0\}.$$

Then we distinguish between the case where $\text{in}(C_I)^* = \text{in}(C_I)$, and the case where there exists at least a state i of $\text{in}(C_I)$ such that the probability to jump from C_I to C_{I-1} when the sojourn in C_I starts in i , is null. Let us define a new condition.

Condition 3 For all $I > 1$ and all $i \in \text{in}(C_I)$, the probability to jump from C_I to C_{I-1} when the sojourn in C_I starts in i , is not null (that is, $\text{in}(C_I)^* = \text{in}(C_I)$).

Under this new condition, lower bounds of the $\mu_I^{(j)}$'s are given in the following result.

Theorem 5 For all $I \geq K$, for all $j \in \text{in}(G)$,

$$\mu_I^* = \min_{i \in \text{in}(C_I)} \frac{\widehat{p}_{i,I}}{\widehat{h}_{i,I}} \leq \mu_I^{(j)}. \quad (24)$$

Proof. The result simply follows from relations (20) and (23), writing that, for any $I \geq K$,

$$\mu_I^{(j)} = \frac{p_I^{(j)}}{h_I^{(j)}} = \frac{\sum_{i \in \text{in}(C_I)} v_{I,i}^{(j)} \widehat{p}_{i,I}}{\sum_{i \in \text{in}(C_I)} v_{I,i}^{(j)} \widehat{h}_{i,I}},$$

and then using the fact that $\sum_{i \in C_I} v_{I,i}^{(j)} = 1$. ■

Let us resume the algorithm. The input data are the partition and a given level K . The steps to be followed are the following:

- First, compute the starred bounds given by (24).
- Then, for any $j \in \text{in}(G)$, find the stationary distribution $\mathbf{y}^{(j)}$ of the chain $Y^{(j)}$ with the choice $\mu_I^- = \mu_I^*$.
- Compute the lower and upper bounds of R using (12) and (13).

The main drawback of this algorithm is that the computation of the $\widehat{p}_{i,I}$'s and the $\widehat{h}_{i,I}$'s may be numerically intractable due to the possible size of class C_I . Moreover, even if they can be calculated, the induced cost may be too high for the user. A possible way to handle this problem is to try to obtain new bounds on these numbers. We are following this direction in our current research work. In this paper, we want only to illustrate the use of our approach in cases where deriving the bounds can be done analytically. However we can note that if Condition 2 holds, the bounds obtained by the new algorithm are better than those of [3], as stated in next result.

Lemma 4 If Condition 2 holds (which implies that Condition 3 holds as well), then for all $I \geq K$, $\mu_I^- \leq \mu_I^* \leq \mu_I$.

Proof. Let us denote by \mathbf{e}_i the vector whose entries are 0 except the i th one, which is equal to 1. From Lemmas 2 and 3, we have:

$$\widehat{h}_{i,I} = -\mathbf{e}_i A_{C_I}^{-1} \mathbf{1}^T, \quad (25)$$

$$\widehat{p}_{i,I} = -\mathbf{e}_i A_{C_I}^{-1} A_{C_I, C_{I-1}} \mathbf{1}^T. \quad (26)$$

Let us consider the vector $A_{C_I, C_{I-1}} \mathbf{1}^T$. Each one of its components are greater than $\min_{i \in C_I} \sum_{j \in C_J} A_{i,j}$, that is, greater than μ_I^- . From the relations above, we have

$$\widehat{p}_{i,I} \geq \mu_I^- \widehat{h}_{i,I}$$

which ends the proof. ■

8 Bounding R in infinite models

In this section, we adapt the method described before to the case of an infinite state space S and finite classes c_I 's (implying that $M = +\infty$). We assume that X is ergodic. We also assume, of course, that $R < +\infty$. Given that each class is finite, the cardinality of $\text{in}(G)$ is also finite. Observe first that when $|S| = +\infty$, the arguments used in Lemma 1 remain valid, proving here that from any state i there is a finite path to j in $X^{(j)}$. Now, since the infinitesimal generator of $X^{(j)}$ is the same as the infinitesimal generator of X except for a finite part of S , necessary $X^{(j)}$ is also ergodic, and therefore, with a single positive recurrent class containing j . In Appendix C, we show that the conclusions of Theorem 1 are still valid in the infinite state space case. Moreover, since $R < +\infty$, we necessarily have $R^{(j)} < +\infty$ as well. To see this, observe that if for some $j_0 \in \text{in}(G)$ we have $R^{(j_0)} = \boldsymbol{\pi}^{(j_0)} \mathbf{r}^T = +\infty$, then from the expression

$$R = \boldsymbol{\pi} \mathbf{r}^T = \sum_{j \in \text{in}(G)} \boldsymbol{\pi}^{(j)} \mathbf{r}^T$$

we obtain $R = +\infty$, in contradiction with our starting assumption.

To simplify the analysis, we add the assumption that the transitions from C_I to C_J are null if $I - J > 1$. If this condition is not verified, the corresponding relations are more complex, but the method still applies.

Assume that Conditions 1 and 3 hold. We can compute the starred bounds of previous subsection.

For each $j \in \text{in}(G)$, let us define a new chain $Z^{(j)}$ over the state space $G \cup \{c\}$ with the same transition rates than X inside G . From any $g \in G$ to c the transition rate is equal to $A_{g,c_K}^{(j)\text{agg}}$ as in $X^{(j)\text{agg}}$. From c to G , there exists a single non-null transition rate which is from c to j , denoted by ν_j , and defined by

$$\nu_j = \frac{\mu_K^{(j)}}{\sum_{I=K}^{+\infty} \theta_I^{(j)}} \quad (27)$$

where

$$\theta_I^{(j)} = \frac{\lambda_{K,K+1}^{(j)} \cdots \lambda_{I-1,I}^{(j)}}{\mu_{K+1}^{(j)} \cdots \mu_I^{(j)}}, \quad \theta_K^{(j)} = 1. \quad (28)$$

Then denoting by $\mathbf{z}^{(j)}$ the stationary distribution of $Z^{(j)}$, we have

Lemma 5

$$z_c^{(j)} = \sum_{I=K}^{+\infty} \pi_{c_I}^{(j)\text{agg}}$$

and for all $g \in G$,

$$z_g^{(j)} = \pi_g^{(j)\text{agg}}.$$

Proof. The proof is immediate by writing the equilibrium equations for $X_\infty^{(j)\text{agg}}$

$$\pi_{c_I}^{(j)\text{agg}} \lambda_{I,I+1}^{(j)} = \pi_{c_{I+1}}^{(j)\text{agg}} \mu_{I+1}^{(j)},$$

which imply that

$$\pi_{c_I}^{(j)\text{agg}} = \theta_I^{(j)} \pi_{c_K}^{(j)\text{agg}}.$$

In order to have the equality $z_c^{(j)} = \sum_{I=K}^{+\infty} \pi_{c_I}^{(j)\text{agg}}$ we need

$$\nu_j = \frac{\pi_{c_K}^{(j)\text{agg}} \mu_K^{(j)}}{\sum_{I=K}^{+\infty} \pi_{c_I}^{(j)\text{agg}}}.$$

The result follows by writing $\pi_{c_I}^{(j)\text{agg}}$ as a function of $\pi_{c_K}^{(j)\text{agg}}$ in this last relation. — ■

In the same way, let us define another irreducible Markov chain $Z'^{(j)}$ and its stationary distribution $z'^{(j)}$ over $G \cup \{c\}$, with the same rates than for $Z^{(j)}$ inside G and from any $g \in G$ to c . The only entry in G from c is j and the transition rate from c to j , denoted by ν'_j , is defined as follows:

$$\nu'_j = \frac{\mu_K^*}{\sum_{I=K}^{+\infty} \theta_I^*} \quad (29)$$

where

$$\forall I > K, \theta_I^* = \frac{\lambda_{K,K+1}^+ \cdots \lambda_{I-1,I}^+}{\mu_{K+1}^* \cdots \mu_I^*}, \theta_K^* = 1. \quad (30)$$

Then, we have the following results:

Lemma 6

$$z_c'^{(j)} = \sum_{I=K}^{+\infty} y_{c_I}^{(j)}$$

and for all $g \in G$,

$$z_g'^{(j)} = y_g^{(j)}.$$

Proof. From the balance equations

$$y_{c_I}^{(j)} \lambda_{I,I+1}^+ = y_{c_{I+1}}^{(j)} \mu_{I+1}^*,$$

we have

$$y_{c_I}^{(j)} = \theta_I^* y_{c_K}^{(j)}.$$

In order to have the equality $z_c'^{(j)} = \sum_{I=K}^{+\infty} y_{c_I}^{(j)}$ we need

$$\nu_j' = \frac{y_{c_K}^{(j)} \mu_K^*}{\sum_{I=K}^{+\infty} y_{c_I}^{(j)}}.$$

The result follows by writing $y_{c_I}^{(j)}$ as a function of $y_{c_K}^{(j)}$ in this last relation. ■

At this point, you should note that $\nu_j' \leq \nu_j$, and that ν_j' is in fact independent of j .

Let us come back to our bounds. Let us define over $G \cup \{c\}$, the two reward vectors \mathbf{r}'_1 and \mathbf{r}'_2 obtained by completing vector \mathbf{r}_G with a reward on the aggregated state c equal to ϱ_1 in \mathbf{r}'_1 and denoted by r_c in \mathbf{r}'_2 . Since $0 \leq \varrho_1 < +\infty$, to obtain a lower bound on R we proceed exactly as in the previous section. The problem can arise in the case of $\varrho_2 = +\infty$. First, let us consider the case when ϱ_2 is finite.

Theorem 6 *If $\varrho_2 < \infty$, let r_c be equal to ϱ_2 . Then, we have the following bounds of R :*

$$\min_{j \in \text{in}(G)} z'^{(j)} \mathbf{r}'_1{}^T \leq R, \quad (31)$$

$$\max_{j \in \text{in}(G)} z'^{(j)} \mathbf{r}'_2{}^T \geq R. \quad (32)$$

Proof. From Lemma 6, we know that

$$z'^{(j)} \mathbf{r}'_1{}^T = \mathbf{y}_G^{(j)} \mathbf{r}_G^T + (1 - y^{(j)}(G)) \varrho_1,$$

and

$$z'^{(j)} \mathbf{r}'_2{}^T = \mathbf{y}_G^{(j)} \mathbf{r}_G^T + (1 - y^{(j)}(G)) \varrho_2.$$

Then, the proof is identical to the proof of Theorem 4. ■

Assume now that $\varrho_2 = +\infty$ and denote by r_{c_I} an upper bound of the rewards on C_I . Then we have the following preliminary result:

Lemma 7 *If $\varrho_2 = +\infty$, under the condition*

$$\sum_{I=K}^{+\infty} \theta_I^{(j)} r_{c_I} < +\infty,$$

letting

$$r_c = \max \left\{ \max_{j \in \text{in}(G)} \frac{\sum_{I=K}^{+\infty} \theta_I^{(j)} r_{c_I}}{\sum_{I=K}^{+\infty} \theta_I^{(j)}}, \max_{i \in G} r_i \right\}, \quad (33)$$

an upper bound of the expected reward R is

$$\max_{j \in \text{in}(G)} \left(\sum_{i \in G} z_i^{(j)} r_i + r_c z_c^{(j)} \right) \geq R. \quad (34)$$

Proof. For any $j \in \text{in}(G)$,

$$\begin{aligned} \boldsymbol{\pi}^{(j)} \mathbf{r}^T &= \boldsymbol{\pi}_G^{(j)} \mathbf{r}_G^T + \boldsymbol{\pi}_G^{(j)} \mathbf{r}_G^T \\ &= \boldsymbol{\pi}_G^{(j)\text{agg}} \mathbf{r}_G^T + \boldsymbol{\pi}_G^{(j)} \mathbf{r}_G^T \\ &\leq \boldsymbol{\pi}_G^{(j)\text{agg}} \mathbf{r}_G^T + \sum_{I=K}^{+\infty} \pi_{c_I}^{(j)\text{agg}} r_{c_I}. \end{aligned}$$

From the proof of Lemma 5, $\pi_{c_I}^{(j)\text{agg}} = \theta_I^{(j)} \pi_{c_K}^{(j)\text{agg}}$, so, given that $\sum_{I=K}^{+\infty} \theta_I^{(j)} r_{c_I} < \infty$,

$$\boldsymbol{\pi}^{(j)} \mathbf{r}^T \leq \mathbf{z}_G^{(j)} \mathbf{r}_G^T + \pi_{c_K}^{(j)\text{agg}} \sum_{I=K}^{+\infty} \theta_I^{(j)} r_{c_I}.$$

From the definition of r_c , we have

$$\boldsymbol{\pi}^{(j)} \mathbf{r}^T \leq \mathbf{z}_G^{(j)} \mathbf{r}_G^T + z_c^{(j)} r_c. \quad \blacksquare$$

Thus, obtaining an upper bound of r_c allows us to derive an upper bound of R , as stated in the next theorem:

Theorem 7 *If $r_c^* \leq r_c$, then an upper bound of the expected reward R is*

$$\max_{j \in \text{in}(G)} \left(\sum_{i \in G} z_i'^{(j)} r_i + r_c^* z_c'^{(j)} \right) \geq R. \quad (35)$$

Proof. Consider the expression of $\mathbf{z}'^{(j)} \mathbf{r}_2'^T$

$$\mathbf{z}'^{(j)} \mathbf{r}_2'^T = \mathbf{z}'^{(j)} \mathbf{r}_G^T + z_c'^{(j)} r_c.$$

From Theorem 2, we have

$$\hat{\mathbf{z}}_G'^{(j)} = \hat{\mathbf{z}}_G^{(j)}.$$

Using the remark above, $\nu'_j \leq \nu_j$, and Theorem 3, we also have $z'^{(j)}(G) \leq z^{(j)}(G)$, and so,

$$z_c'^{(j)} \geq z_c^{(j)}.$$

Then given that $r_c \geq r_i$, $\forall i \in in(G)$,

$$z^{(j)} \mathbf{r}_2'^T \leq z'^{(j)} \mathbf{r}_2'^T.$$

It follows that

$$\sum_{i \in G} z_i^{(j)} r_i + r_c z_c^{(j)} \leq \sum_{i \in G} z_i'^{(j)} r_i + r_c z_c'^{(j)}.$$

Then if $r_c^* \geq r_c$, we obtain (35). ■

This tells us that if we can compute an upper bound of r_c , we have in the right hand side of (35) an upper bound of R .

A particular case

Let us analyse what happens when $\lambda_{I,J}^+$ and μ_I^* are constant and respectively equal to λ and μ . A condition for the stability of the model is given by $\lambda < \mu$. Denoting $\varrho = \lambda/\mu$, we have

$$\theta_I^* = \varrho^{I-K}, \quad \nu_j' = \mu(1 - \varrho).$$

In the examples that follow (9.2 and 9.3), we are in this particular case. We just have to bound r_c with

$$r_c^* = \max \left(\max_{i \in G} r_i, \frac{\sum_{I=K}^{+\infty} \theta_I^* r_{cI}}{\sum_{I=K}^{+\infty} \theta_I^*} \right) = \max \left(\max_{i \in G} r_i, (1 - \varrho) \sum_{I=K}^{+\infty} \varrho^{I-K} r_{cI} \right).$$

This is the technique that we will use to obtain bounds in the queuing models of next section.

9 Illustrations

This section illustrates the efficiency of the bounding method proposed in the paper. First, we use a standard dependability model, a ‘‘Machine Repair Model’’ which leads to a large finite Markov chain that can not be handled by the technique published in [3]. The second example is an open queueing network composed by two queues in series, leading to an infinite Markov chain with no known analytical solution. In this case, we bound the mean number of customers in each node. We can observe that this model can not be handled by matrix-geometric techniques. The third example is another open tandem of queues. Here, there are blocking mechanisms since we consider finite buffers in all the nodes except the first one, and we bound blocking probabilities. These two open examples can be transformed into closed versions by limiting the total allowed number of customers and in this case, the conditions necessary to use the method of [3] do not hold, as in the first example.

9.1 Bounding the asymptotic availability of a MRM

Our first example is a standard multi-component system subject to failures and repairs. There are two types of machines. The number of machines of type k is denoted by N_k and the time to failure of type k machines is exponentially distributed with parameter λ_k , $k = 1, 2$. After a failure, the machines enter a repair system with one server. The repair time is distributed according to a d_k -stage Erlangian distribution for type k machines, with mean d_k/ν . Machines of type 1 are served with higher priority than type 2 machines, and the priority is non preemptive. We assume that type 2 machines are put immediately in operation when repaired, but that type 1 machines need a delay exponentially distributed (with parameter μ) to come back to operation. Thus, in the model, type 1 customers go to a second infinite server queue. This allows us to illustrate the method when the repair subsystem is more complex than a single queue.

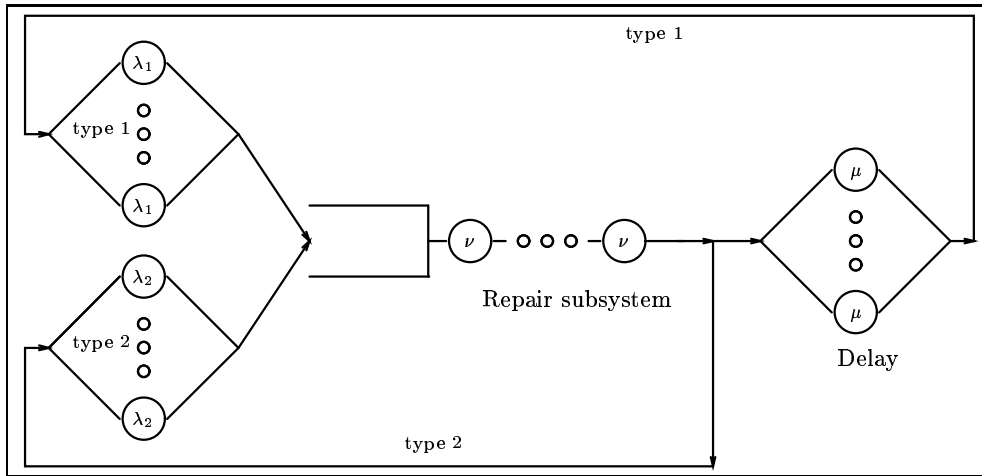


Figure 4: A Machine Repair Model

Let n_1 (respectively n_2) be the number of machines of type 1 (respectively 2) in the repair queue, denote by k the type of the machine being repaired (with value 0, say, if the repair station is empty), let d be the number of the phase of the current service with $1 \leq d \leq d_k$ (0, say, if the repair station is empty), and denote by n_3 the number of machines of type 1 in transit (that is, in the delay station). We are interesting in bounding the asymptotic availability of the system. It is said to be available as soon as $nMin_1$ (respectively $nMin_2$) machines of type 1 (respectively of type 2) are up.

On states $s = (n_1, n_2, d, k, n_3)$ we have a Markov chain on which we consider the subsets of states C_I defined by

$$C_I = \{s \mid n_1 + n_2 + n_3 = I\}$$

with $0 \leq I \leq N_1 + N_2$. They define a partition of the state space. Observe that Condition 2 is not satisfied but Condition 3 is. Thus the method in [3] can not be applied. The size of the state space S is

$$\begin{aligned}
|S| &= 1 + N_1 + \frac{1}{2}(d_1 + d_2) \left(\frac{1}{6}N_1(N_1 + 1)(2N_1 + 1) + \frac{1}{2}N_1(N_1 + 1) \right) \\
&\quad + (N_2 - N_1)(N_1 + 1) \left(d_2 + \frac{1}{2}N_1(d_1 + d_2) \right) \\
&\quad + N_1 \left(d_2(N_1 + N_2 + 1) + \frac{1}{2}(d_1 + d_2)N_1(N_1 + 1) \right) \\
&\quad - d_2 \left(\frac{1}{2}(N_1 + N_2)(N_1 + N_2 + 1) - \frac{1}{2}N_2(N_2 + 1) \right) \\
&\quad - \frac{1}{2}(d_1 + d_2) \left(\frac{1}{6}N_1(N_1 + 1)(2N_1 + 1) - \frac{1}{2}N_1(N_1 + 1) \right),
\end{aligned}$$

and if we aggregate each C_I for $I \geq K$ in a single state, the size of the aggregated space S^{agg} is

- if $K < N_1$,

$$|S^{\text{agg}}| = K + \frac{1}{2}(d_1 + d_2) \left(\frac{1}{6}(K - 1)K(K - 1) + \frac{1}{2}(K - 1)K + 1 \right),$$

- if $N_1 \leq K < N_2$,

$$\begin{aligned}
|S^{\text{agg}}| &= K + \frac{1}{2}(d_1 + d_2) \left(\frac{1}{6}(K - 1)K(2K - 1) + \frac{1}{2}(K - 1)K \right) \\
&\quad + (K - N_1 + 1) \left((N_1 + 1)d_2 + \frac{1}{2}(d_1 + d_2)N_1(N_1 + 1) \right) + 1,
\end{aligned}$$

- if $N_2 \leq K < N_2 + N_1$,

$$\begin{aligned}
|S^{\text{agg}}| &= K + \frac{1}{2}(d_1 + d_2) \left(\frac{1}{6}(K - 1)K(2N_1 + 1) + \frac{1}{2}(K - 1)K \right) \\
&\quad + (K - N_1 + 1) \left((N_1 + 1)d_2 + \frac{1}{2}N_1(N_1 + 1)(d_1 + d_2) \right) \\
&\quad + (K - 1 - N_2) \left(d_2(N_1 + N_2 + 1) + \frac{1}{2}N_1(N_1 + 1)(d_1 + d_2) \right) \\
&\quad - d_2 \left(\frac{1}{2}(K - 1)K - \frac{1}{2}N_2(N_2 + 1) \right) \\
&\quad - \frac{1}{2}(d_1 + d_2) \left(\frac{1}{6}(K - 1 - N_2)(K - N_2)(2K - 1 - 2N_2) \right. \\
&\quad \left. - \frac{1}{2}(K - 1 - N_2)(K - N_2) \right) + 1.
\end{aligned}$$

Let us consider the following parameter values:

- $N_1 = 80$, $N_2 = 120$, $nMin_1 = 77$, $nMin_2 = 115$, $\lambda_1 = 0.00004$, $\lambda_2 = 0.00003$, $\nu = 1.0$, $d_1 = 6$, $d_2 = 5$ and $\mu = 3.0$. The size of the whole state space is $|S| = 4344921$. Using two small values of K , we obtain the following numerical results:

K	$ S^{agg} $	Lower bound	Upper bound
5	226	0.999999698	0.999999852
10	1826	0.999999841	0.999999841

- $N_1 = 80$, $N_2 = 120$, $nMin_1 = 79$, $nMin_2 = 115$, $\lambda_1 = 0.00004$, $\lambda_2 = 0.00003$, $\nu = 1.0$, $d_1 = 6$, $d_2 = 5$ and $\mu = 3.0$. The size of the whole state space is $|S| = 4344921$. We obtain

K	$ S^{agg} $	Lower bound	Upper bound
5	226	0.9997597121	0.9997597466
10	1826	0.9997597349	0.9997597349

9.2 Bounding the mean number of customers in a two-node tandem

Consider the following simple open queueing network with 2 FIFO nodes (Figure 5). In this example, customers arrive from outside according to a Poisson process with rate λ . Each queue has an infinite capacity, and service time at each node are 2-stages Erlangian distributed, with average equal to $2/\nu$.

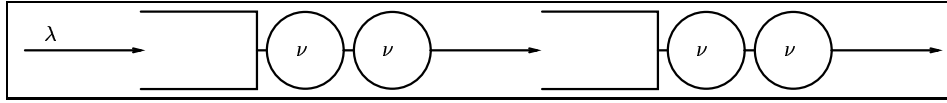


Figure 5: An two-node open queueing network

We consider the standard Markov representation of this queueing system with states $s = (n_1, d_1, n_2, d_2)$, with n_i customers in node i and phase d_i in the server of node i , $i = 1, 2$, with the convention that $d_i = 0$ if $n_i = 0$. Let us define the partition $(C_I)_{I \geq 0}$ of the state space with

$$C_I = \{(n_1, d_1, n_2, d_2) \mid n_1 + n_2 = I\}.$$

This model does not possess a known closed form solution and it is not possible to solve it directly due to its infinite state space cardinality. Moreover given that Condition 2 is not satisfied, the method in [3] can not be applied. Observe also that the natural partition considered makes that matrix-geometric [8] approach can not be used neither, since the state space has a 2-dimensional structure and both dimensions are unbounded.

Let us apply our method to bound the mean number of customers in each node. In both cases, the a priori bound ϱ_2 on the rewards is infinite. First, we should note that λ is an upper bound of $\lambda_{I, I+1}^+$ for all I . Concerning the needed lower bound on the μ_I 's, we use the regular structure of the C_I 's. In Annex D we show that for each C_I , the value of μ_I^* is

obtained for state $(1, 2, I - 1, 1)$, if $I \geq 2$, and for state $(1, 2, 0, 0)$ if $I = 1$, and that its value is independent of I . Thus, if we let ϱ be equal to λ/μ_K^* , we have

$$\theta_I^* = \left(\frac{\lambda}{\mu_K^*} \right)^{I-K} = \varrho^{I-K}, \quad \text{and } \nu_j' = (1 - \varrho)\mu_K^*.$$

The fact that μ_I^* is independent of I , give directly a sufficient condition of ergodicity of the aggregated process, which is $\lambda < \mu_I^*$.

Let us set $\nu = 1.0$. To be able to apply the method, we need $\lambda < 0.42$. If we choose $\lambda = 0.35$, the mean number of customers in node 1 belongs to $]1.6784474634, 1.6784474659[$ and the mean number of customers in node 2 belongs to $]1.9250000000, 1.9250000006[$. When $\lambda = 0.4$, the mean number of customers in node 1 belongs to $]3.2000, 3.2020[$ and in node 2 belongs to $]2.7101, 2.7139[$. These values are obtained after generating 19802 states.

9.3 Bounding blocking probabilities in a three-node tandem

Let us consider the three-node open queueing network shown in Figure 6. Excepting the first node which has an infinite capacity, the two other nodes have a finite capacity of respective sizes H_1 and H_2 . Customers arrive at the first node according to a Poisson process with rate λ . All the services are exponentially distributed and the service rates are respectively μ_1 , μ_2 and μ_3 in the first, the second and the third nodes. We also assume that all the nodes implement a FIFO service discipline, and that there is a blocking-after-service behaviour in the first and second nodes. This means that when the second node is saturated, the first one blocks its server after the end of the current service and until the departure of the customer being serviced in the second one (“blocking-after-service”, B.A.S.). Then, if the latter is not saturated, simultaneously a customer leaves the second node to the third node and another customer passes from node 1 to node 2. On the other hand, if the last node is saturated, the second node as the first one blocks its server after the end of the current service and until the customer serviced in the third node leaves the system.

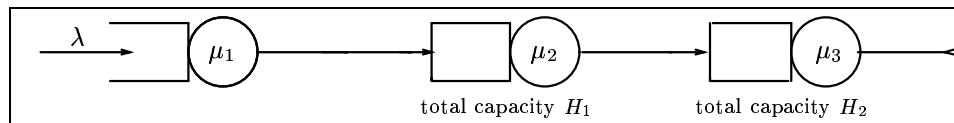


Figure 6: A three-node open queueing network with blocking-after-service

Let n_i be the number of customers in the i th node, $i = 1, 2, 3$, and let $s = (n_1, n_2, n_3)$ be a state of the model. We use $C_I = \{s \mid n_1 + n_2 + n_3 = I\}$ with $I \geq 0$. As in the previous example, there is no closed form solution [9] and the state space cardinality is infinite. Given that $n_2 \leq H_1$ and $n_3 \leq H_2$, it is easily verified that for $I \geq H_1 + H_2 + 1$, $|C_I|$ is constant. This allows us to simplify the analysis by choosing any $K > H_1 + H_2 + 1$ (note that if we had chosen some $K \leq H_1 + H_2 + 1$, we still could have done the same by collapsing in c all the aggregated states c_I with $I > H_1 + H_2 + 1$ and keeping $c_K, \dots, c_{H_1+H_2+1}$). The condition for

stability due to the Markov chain used to bound the asymptotic measures is $\sum_{j=1}^{\infty} \theta_j^* < \infty$. Given that $K > H_1 + H_2 + 1$, for all $I \geq K$, $\mu_I^* = \mu_K^*$ and we have $\lambda_{I,I+1}^+ = \lambda$. So, letting $\varrho = \lambda/\mu_K^*$, we have

$$\theta_I^* = \left(\frac{\lambda}{\mu_K^*}\right)^{I-K} = \varrho^{I-K}, \quad \nu_j' = (1 - \varrho)\mu_K^*.$$

We are going to bound the mean number of customers in the first node (case ϱ_2 infinite), and blocking probabilities of the first and second nodes (case ϱ_2 finite).

Let us consider the following parameter values: $\lambda = 0.2$, $\mu_1 = 0.7$, $\mu_2 = 1.5$, $\mu_3 = 0.2$, $H_1 = 18$ and $H_2 = 10$. After generating only 3839 states, we give measure value with an absolute error less than 10^{-10} , that is, the difference between the computed upper and lower bound is less than 10^{-10} . The mean number of customers in node 1 is 0.4000000000.

In the same way, we show respectively in Figure 7 and in Figure 8 the probability of blocking of nodes 1 and 2. As above, given that we are interesting in performance measures with an absolute error less than 10^{-10} , we only plot the average between both computed bounds. The probability of blocking of nodes 1 is plotted as a function of the service rate of the corresponding server, μ_2 , with $(H_1, H_2, \lambda, \mu_1, \mu_3) = (10, 8, 0.1224, 1.5, 0.5)$ and the probability of blocking of nodes 2 as a function of μ_3 , with $(H_1, H_2, \lambda, \mu_1, \mu_2) = (18, 10, 0.1224, 1.5, 0.7)$.

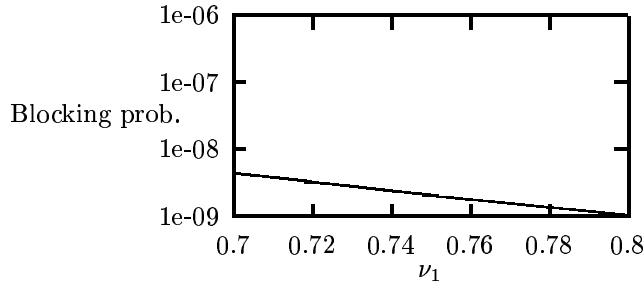


Figure 7: Probability of blocking of node 1 as a function of μ_2

10 Conclusions

This paper proposes a new way of obtaining upper and lower bounds of asymptotic performance measures, from finite or infinite Markov models. The asymptotic performance includes as particular cases the asymptotic availability in a dependability context, or standard asymptotic performance measures such as mean number of customers, blocking probabilities, loss probabilities, etc. To be applied, the method, as presented here, needs enough knowledge of the structure of the model in order to be able to derive analytically or

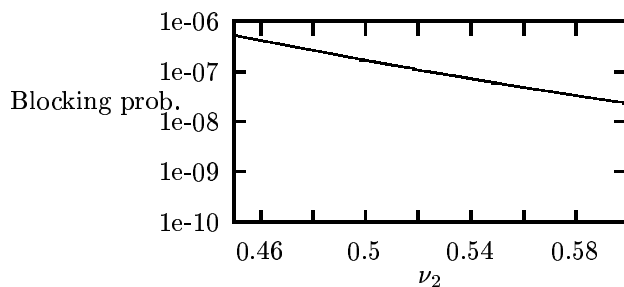


Figure 8: Probability of blocking of node 2 as a function of μ_3

numerically certain parameters necessary to obtain the bounds. This is not always possible and current research aims to deal with this situation. In any case, there are many models such as the type of infinite queueing networks used in this paper to illustrate the method, with no known closed solution and where, to the best of our knowledge, any other available bounding techniques do not apply.

Acknowledgements

The authors would like to thank J. Carrasco and B. Sericola for helpful discussions on different technical points.

References

- [1] P. Semal P.-J. Courtois. – Bounds for the positive eigenvectors of nonnegative matrices and for their approximations by decomposition. – *Journal of the association for computing machinery*, 31(4):804–825, october 1984.
- [2] P. Semal P.-J. Courtois. – Computable bounds for conditional steady-state probabilities in large markov chains and queueing models. – *IEEE on selected areas in communications*, 4(6):926–936, september 1986.
- [3] A. Goyal R.R. Muntz, E Silva E. De Souza. – Bounding availability of repairable computer systems. – *IEEE journal on selected areas in communication*, 38(12), december 1989.
- [4] Rubino G. and Sericola B. – Sojourn times in finite markov process. – *J.Appl.Prob*, 27:744–756, July 1989.
- [5] J.C.S. Lui R.R. Muntz. – Evaluating bounds on steady-state availability of repairable systems from markov models. – In *Numerical solution of Markov chains*, pages 435–454, 1991.

- [6] J. Carrasco. – Improving availability bounds using the failure distance concept. – In *Procs. of the DCCA-4 Int. Conf.*, janary 1994.
- [7] P. Semal. – Refinable bounds for large markov chains. – *IEEE transactions on computers*, 44(10):1216–1222, october 1995.
- [8] Marcel F. Neuts. – *Matrix-Geometric Solutions in Stochastic Models. An Algorithmic Approach*. – The Johns Hopkins University Press, 1981.
- [9] Harry G. Perros. – *Queueing Networks with Blocking*. – Oxford University Press, 1994.
- [10] Alexander Graham. – *Nonnegatige matrices and applicable topics in linear Algebra*. – Ellis Horwood Limited, 1987.
- [11] Erhan Cinlar. – *Introduction to Stochastic Processes*. – Prentice-Hall, Inc., 1975.

A Proof of Theorem 1

First, let us consider the irreducible and aperiodic stochastic matrix P , obtained by uniformization of A with respect to the uniformization rate $\Lambda \geq \sup_i |A_{ii}|$, that is, matrix $P = I + A/\Lambda$. We have $\pi = \pi P$. Then, let us construct the matrix Q with the same size as P , such that $Q_{i,j} = 0$ for all $i \in C_K$ and $j \in in(G)$, and $Q_{i,j} = P_{i,j}$ in the other cases. Given that for all $i, j \in S$ we have $Q_{i,j} \leq P_{i,j}$, Q is a lower bound of P , that is a sub-stochastic matrix. We should note that Q is a strict lower bound because of the irreducibility of P . Thus the matrix $(I - Q)$ is invertible [10, Chapter 1, p. 66].

Let us denote by $\mathcal{P}(M)$ the polyhedron given by the set of convex combinations of the normalized rows of the square matrix M . Observe that if P is a stochastic matrix and if $\pi = \pi P$, then $\pi \in \mathcal{P}(P)$. Consider

$$\mathcal{P}((I - Q)^{-1}) = \{v \in \mathbb{R}_{1 \times |S|} \mid \exists \beta \in \mathbb{R}_{1 \times |S|}, \beta \mathbf{1}^T = 1, v = \beta \Sigma^{-1} (I - Q)^{-1}\}, \quad (36)$$

where $\Sigma^{-1} = \text{Diag}((I - Q)^{-1} \mathbf{1}^T)^{-1} = \text{Diag}(\sigma_k)$ is the normalization matrix. We will denote by $z^{(1)}, \dots, z^{(|S|)}$ the vertices of $\mathcal{P}((I - Q)^{-1})$.

Proof of Theorem 1 As for P , we consider the uniformization of $A^{(j)}$, denoted by $P^{(j)}$. Because of the irreducibility of A , there exists an unique normalized vector $\pi^{(j)}$ (Lemma 1), such that $\pi^{(j)} A^{(j)} = 0$ and $\pi^{(j)} P^{(j)} = \pi^{(j)}$.

First we show that $\pi^{(j)}$ belongs to $\mathcal{P}((I - Q)^{-1})$, and then that $\pi^{(j)}$ is equal to the j th vertex of $\mathcal{P}((I - Q)^{-1})$.

Lemma 8 $\pi^{(j)} \in \mathcal{P}((I - Q)^{-1})$.

Proof. Matrix $(I - Q)^{-1}$ is invertible; this implies that there always exists a vector β such that

$$\pi^{(j)} = \beta \Sigma^{-1} (I - Q)^{-1}. \quad (37)$$

Since

$$\Sigma^{-1}(I - Q)^{-1}\mathbf{1}^T = \mathbf{1}^T,$$

then

$$\mathbf{1} = \boldsymbol{\pi}^{(j)}\mathbf{1}^T = \boldsymbol{\beta}\Sigma^{-1}(I - Q)^{-1}\mathbf{1}^T = \boldsymbol{\beta}\mathbf{1}^T. \quad (38)$$

Moreover, from (37), and given that $\boldsymbol{\pi}^{(j)}(I - P^{(j)}) = 0$, it follows

$$\boldsymbol{\beta} = \boldsymbol{\pi}^{(j)}(I - Q)\Sigma = \boldsymbol{\pi}^{(j)}(I - Q)\Sigma - \boldsymbol{\pi}^{(j)}(I - P^{(j)})\Sigma = \boldsymbol{\pi}^{(j)}(P^{(j)} - Q) \geq 0. \quad (39)$$

Equations (37), (38), and (39), end the proof of this lemma. ■

Lemma 9 *The j th vertex of $\mathcal{P}((I - Q)^{-1})$ is $\boldsymbol{\pi}^{(j)}$.*

Proof. We prove first the existence of a stochastic matrix such that the j th vertex of $\mathcal{P}((I - Q)^{-1})$, denoted by $\boldsymbol{z}^{(j)}$, is its stationary distribution. Then, we show that this stochastic matrix is $P^{(j)}$. To start, let us introduce a new matrix $\Omega = \text{Diag}(\mathbf{1}(I - Q)^{-1})$. By definition of $\boldsymbol{z}^{(j)}$, there exists a normalized and positive vector $\boldsymbol{\beta}^{(j)}$ such that $\boldsymbol{z}^{(j)} = \boldsymbol{\beta}^{(j)}\Sigma^{-1}(I - Q)^{-1}$. Vector $\boldsymbol{\beta}^{(j)}$ is defined by $\beta_i^{(j)} = 1$ if $i = j$ and $\beta_i^{(j)} = 0$ in the other cases. Indeed $(I - Q)^{-1}$ has full rank. That means that each row of the matrix belongs to the base of the polyhedron. Let us consider the matrix $C^{(j)}$ equal to $\frac{1}{c^{(j)}}\Omega^{-1}\boldsymbol{\nu}^T\boldsymbol{\beta}^{(j)}\Sigma^{-1}$, where $c^{(j)}$ is a constant equal to $\boldsymbol{\beta}^{(j)}\Sigma^{-1}(I - Q)^{-1}\Omega^{-1}\boldsymbol{\nu}^T$ and $\boldsymbol{\nu}$ is a positive and normalized row vector.

Then we have the following relations:

$$\boldsymbol{z}^{(j)}(Q + C^{(j)}) = \boldsymbol{\beta}^{(j)}\Sigma^{-1}(I - Q)^{-1}(Q + C^{(j)}) = \boldsymbol{z}^{(j)}, \quad (40)$$

and if we denote $\boldsymbol{w}^T = (I - Q)^{-1}\Omega^{-1}\boldsymbol{\nu}^T$,

$$(Q + C^{(j)})\boldsymbol{w}^T = (Q + C^{(j)})(I - Q)^{-1}\Omega^{-1}\boldsymbol{\nu}^T = \boldsymbol{w}^T. \quad (41)$$

Note that the matrix $(Q + C^{(j)})$ is a stochastic matrix. Indeed from (40) and because of the equality $\boldsymbol{z}^{(j)}\mathbf{1}^T = 1$, we have $(Q + C^{(j)})\mathbf{1}^T = \mathbf{1}^T$.

Let us prove now that $Q + C^{(j)} = P^{(j)}$. First consider the expression of the constant $c^{(j)}$:

$$c^{(j)} = \boldsymbol{\beta}^{(j)}\Sigma^{-1}(I - Q)^{-1}\Omega^{-1}\boldsymbol{\nu}^T.$$

Given that \boldsymbol{w}^T is a right eigenvector of the stochastic matrix $(Q + C^{(j)})$ (relation (41)), there exists an invertible matrix Δ such that $\boldsymbol{w}^T = \Delta^{-1}\mathbf{1}^T$. And thus,

$$c^{(j)} = \boldsymbol{\beta}^{(j)}\Sigma^{-1}\Delta^{-1} = \Delta^{-1} \sum_k \frac{\beta_k^{(j)}}{\sigma_k} = \Delta^{-1} \frac{\beta_j^{(j)}}{\sigma_j}.$$

It follows that

$$C^{(j)} = \left(\frac{\beta_j^{(j)}}{\sigma_j} \right)^{-1} (I - Q)\mathbf{1}^T\boldsymbol{\beta}^{(j)}\Sigma^{-1}.$$

That means that the only non null elements of $C^{(j)}$ are on its j th column, which is equal to $[(I - Q)\mathbf{1}^T]_j$. And so we have, $Q + C^{(j)} = P^{(j)}$. ■

At this point, from lemmas above we know that $\pi^{(j)} = z^{(j)} \in \mathcal{P}((I - Q)^{-1})$. To prove Theorem 1, we only have to show that π belongs to $\mathcal{P}((I - Q)^{-1})$. As for $\pi^{(j)}$ and given that $\pi(I - P) = 0$ and $(P - Q) \geq 0$, there exists a normalized and positive vector β such that $\pi = \beta \Sigma^{-1}(I - Q)^{-1}$. It follows that π belongs to $\mathcal{P}((I - Q)^{-1})$ and $\pi = \sum_{j \in S} \beta^{(j)} z^{(j)}$. Moreover from the expression of π above, we have the following equality for β : $\beta = \pi(P - Q)\Sigma$. Matrix $(P - Q)$ is equal to a matrix whose columns are null excepted those associated with the entry points in G . This means that the only non null elements of β are β_i for $i \in in(G)$. Then from Lemma 9, it follows that $\pi = \sum_{j \in in(G)} \beta^{(j)} \pi^{(j)}$. ■

B Proof of Theorem 3

Let us first consider the exact aggregated Markov chain constructed from X , by collapsing the subset G into a single state g , and each subset C_I for $I \geq K$ into a single state c_I . We denote the aggregated transition rate from g to any c_I , by $\lambda_{g,I}$, from any c_I to c_J (for $J > I$) by $\lambda_{I,J}$, from c_I to c_{I-1} by μ_I ($I > K$) and from c_K to g by μ_K . Now consider a second Markov chain having the same topology the first one, but such that the transition rates from c_I to c_J when $J \geq I$, denoted by $\lambda'_{I,J}$, are upper bounds of the corresponding rates in the first aggregation, and lower bounds in case of transitions from c_I to c_{I-1} or from c_K to g (the respective values are denoted by μ'_I , $I \geq K$).

Let us denote by \mathbf{v} (respectively by \mathbf{v}') the stationary distribution of the first chain (respectively of the second). Then we have:

Lemma 10 $\mathbf{v}'_g \leq \mathbf{v}_g$.

Proof. Suppose that $\mathbf{v}'_g > \mathbf{v}_g$. From the equilibrium equations, we can write that:

$$\mathbf{v}_{c_K} = \frac{(\sum_{J \geq K} \lambda_{g,J}) \mathbf{v}_g}{\mu_K},$$

$$\mathbf{v}'_{c_K} = \frac{(\sum_{J \geq K} \lambda'_{g,J}) \mathbf{v}'_g}{\mu'_K}.$$

By definition, we know that $\lambda_{g,J} \leq \lambda'_{g,J}$ and $\mu_K \leq \mu'_K$. Thus it follows:

$$\mathbf{v}'_{c_K} > \mathbf{v}_{c_K}.$$

Recursively, if we write for each $I > K$ the equilibrium equations, similar results are obtained. That is:

$$\forall I > K, \mathbf{v}'_{c_I} > \mathbf{v}_{c_I}.$$

Then given that \mathbf{v} is a stationary distribution, we have:

$$\mathbf{v}'_g + \sum_{I \geq K} \mathbf{v}'_{c_I} > 1$$

This means the first assumption is false. ■

In the same way, let us consider the two exact aggregated Markov chains constructed, as above, from $X^{(j)\text{agg}}$ and from $Y^{(j)}$, by collapsing the subset G in a single state g . Given that there is a single point to enter in G from \bar{G} , the aggregated rate from c_K to g in former chain (respectively in the latter chain), is equal to the transition rate from c_k to j in $X^{(j)\text{agg}}$ (respectively in $Y^{(j)}$). The transition rates from g to any c_I in the two new chains are given by the following expressions:

$$\begin{aligned} \lambda_{g,I} &= \sum_{i \in G} \hat{\pi}_{G,i}^{(j)\text{agg}} A_{i,c_I}^{(j)\text{agg}}, \\ \lambda'_{g,I} &= \sum_{i \in G} \hat{y}_{G,i}^{(j)} A_{i,c_I}^{(j)\text{agg}}. \end{aligned}$$

By construction of $X^{(j)\text{agg}}$ and $Y^{(j)}$, the restriction of their respective infinitesimal generators to the subset G are identical. From Theorem 2, we have the equality between the two conditional vectors $\hat{\pi}_G^{(j)\text{agg}}$, $\hat{y}_G^{(j)}$. That means that

$$\lambda_{g,I} = \lambda'_{g,I}.$$

Moreover the transition rates between the other aggregated states are the same as in $X^{(j)\text{agg}}$ and $Y^{(j)}$. Let us denote by $\pi_g^{(j)\text{agg}}$ (respectively by $y_g^{(j)}$), the stationary probability of being in state g for the exact aggregated matrix obtained from $X^{(j)\text{agg}}$ (respectively from $Y^{(j)}$). Thus, the conditions required to apply Lemma 10 are verified. It follows:

$$\pi_g^{(j)\text{agg}} \geq y_g^{(j)}.$$

The fact that $\pi_g^{(j)\text{agg}} = \pi^{(j)\text{agg}}(G)$ and $y_g^{(j)} = y^{(j)}(G)$ (same results as (7)), ends the proof.

C The existence of $(I - Q)^{-1}$

To apply Theorem 1 in the infinite state space case, it remains to show that matrix $(I - Q)$ is invertible. First let us recall the block structure of Q :

$$Q = \begin{pmatrix} P_G & P_{G\bar{G}} \\ 0 & P_{\bar{G}} \end{pmatrix}.$$

One way to obtain the existence of inverse of $(I - Q)$ is to consider each block separately. Given that the only infinite blocks in Q are $P_{G\bar{G}}$ and $P_{\bar{G}}$, we just have to show that $I - P_{\bar{G}}$

is invertible. This can be done by means of standard results as presented in [11]. Since $P_{\bar{G}}$ is a sub-stochastic matrix, the sum $I + P_{\bar{G}} + P_{\bar{G}}^2 + \dots$ is finite and sub-stochastic. Using [11, 6.4.5], if \mathbf{f}^T denotes the vector defined by

$$f_i = \lim_{n \rightarrow \infty} \sum_{j \in \bar{G}} (P_{\bar{G}}^n)_{i,j},$$

we have that \mathbf{f}^T is the maximal solution of the linear system $\mathbf{x}^T = P_{\bar{G}} \mathbf{x}^T$, with $0 \leq x_i \leq 1$, and either $f_i = 0$ for all i , or $\sup_{i \in \bar{G}} f_i = 1$. Since X is irreducible and have only positive recurrent states, the system of linear equations above, have a unique solution verifying $f_i = 0$ for all i [11, 5.3.29]. This means that the matrix $P_{\bar{G}}$ verifies:

$$\lim_{n \rightarrow \infty} P_{\bar{G}}^n \mathbf{1}^T = \mathbf{0}^T.$$

Given that all the elements of $P_{\bar{G}}$ are positive, it follows that

$$\lim_{n \rightarrow \infty} P_{\bar{G}}^n = 0,$$

which is a sufficient condition for the convergence of the series.

D Proof of a result concerning Example 9.2

In this section, we prove that for each C_I , the value of μ_I^* is obtained for the entry-state in C_I $(1, 2, I - 1, 1)$, if $I \geq 2$ and $(1, 2, 0, 0)$ if $I = 1$, and is independent of I . Note that we are only interested in computing the value of μ_I^* for $I \geq K$, that is, in the worst case for $I \geq 1$. We will consider later the particular case $I = 1$.

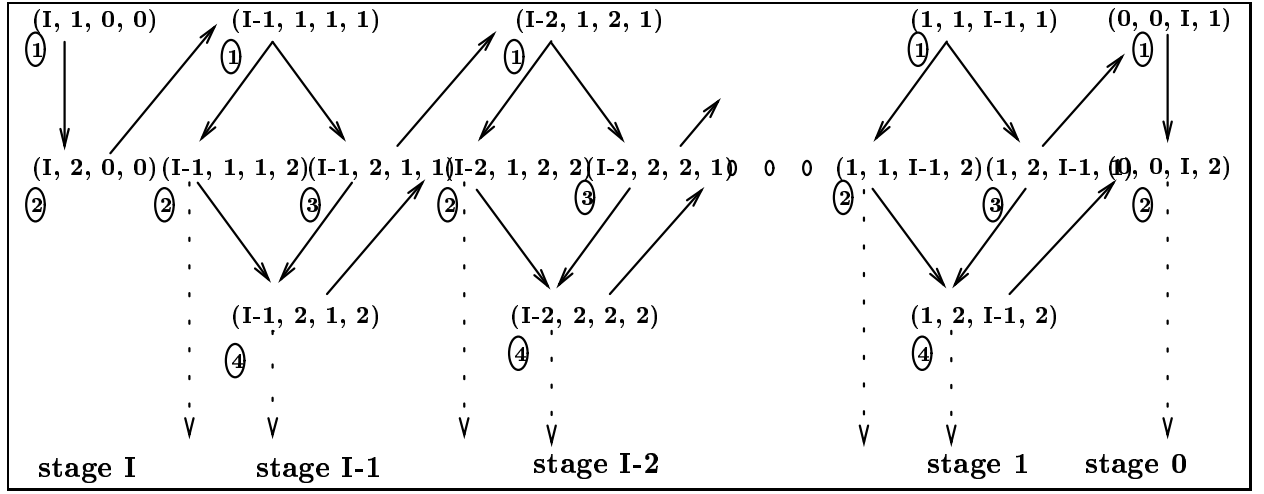
Consider the structure of each class C_I for $I \geq 2$ and the transitions from each state of C_I represented in Figure 9. The full (respectively dotted) arrows stand for the transition rate equal to ν inside the class C_I (respectively from the class C_I to C_{I-1}). From each state in C_I , there exists a transition, with rate λ , to a state in C_{I+1} , deliberately omitted to simplify the figure.

Let us cut C_I into $I + 1$ stages and denumerate the states from 1 to 4 in each stage, excepted in stages 0 and I , where they are denumerated from 1 to 2, as it is shown in Figure 9. Let us denote by p_i^n (respectively by h_i^n) the probability to jump from C_I to C_{I-1} (respectively the mean sojourn time in C_I) when the sojourn in C_I starts in i th state of the n th stage. You should note that all the states of C_I excepted states $(0, 0, I, 2)$ and $(1, 2, I - 1, 2)$, belong to $in(C_I)$.

Consider now four reals a, b, c, d defined by

$$a = \frac{1}{\nu + \lambda}, \quad b = \frac{1}{2\nu + \lambda}, \quad c = \frac{\nu}{\nu + \lambda}, \quad d = \frac{\nu}{2\nu + \lambda}.$$

We have the following relations:


 Figure 9: Topological aspect of a class C_I for $I \geq K$

$$\begin{aligned}
 \text{At stage 0 : } & \begin{cases} p_1^0 = c, & h_1^0 = a, \\ p_2^0 = c, p_1^0 & h_2^0 = a + ch_1^0. \end{cases} \\
 \text{At stage n : } & \begin{cases} p_4^n = d + dp_2^{n-1}, & h_4^n = b + dh_2^{n-1}, \\ p_3^n = d(p_4^n + p_1^{n-1}), & h_3^n = b + d(h_4^n + h_1^{n-1}), \\ p_2^n = d + dp_4^n, & h_2^n = b + dh_4^n, \\ p_1^n = d(p_3^n + p_2^n), & h_1^n = b + d(h_3^n + h_2^n). \end{cases} \\
 \text{At stage I : } & \begin{cases} p_2^I = c + cp_1^{I-1}, & h_2^I = a + ch_1^{I-1}, \\ p_1^I = cp_2^I, & h_1^I = a + ch_2^I. \end{cases}
 \end{aligned}$$

Then it is easy to see that

$$\frac{p_1^1}{h_1^1} = \frac{p_3^1}{h_3^1} = \frac{p_1^0}{h_1^0} = \frac{\nu^2}{2\nu + \lambda},$$

and

$$\frac{p_2^1}{h_2^1} = \frac{p_4^1}{h_4^1} = \frac{p_2^0}{h_2^0} = \nu.$$

And then by recurrence, we have:

$$\text{for all odd } n, \frac{p_1^n}{h_1^n} = \frac{p_1^0}{h_1^0}, \quad \text{for all even } n, \frac{p_1^n}{h_1^n} = \frac{p_4^1}{h_4^1}.$$

But, for any value of ν and λ , $\nu > \frac{\nu^2}{2\nu + \lambda}$; it follows that for any $I \geq 2$,

$$\mu_I^* = \frac{\nu^2}{2\nu + \lambda}.$$

For the particular case $I = 1$, there are only two stages and all the states of C_1 , excepted state $(0, 0, 1, 2)$, belong to $in(C_1)$. Then we obtain

$$\mu_1^* = \frac{(\nu^2 + 2\nu\lambda + \lambda^2)\nu^2}{(3\nu^2 + 3\nu\lambda + \lambda^2)(2\nu + \lambda)}.$$



Unit e de recherche INRIA Lorraine, Technop le de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS L ES NANCY
Unit e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit e de recherche INRIA Rh ne-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unit e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unit e de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

 diteur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
<http://www.inria.fr>
ISSN 0249-6399