



HAL
open science

Performance et fiabilité des systèmes de grandes tailles par les méthodes de bornes stochastiques

Idriss Ismael-Aouled

► **To cite this version:**

Idriss Ismael-Aouled. Performance et fiabilité des systèmes de grandes tailles par les méthodes de bornes stochastiques. Performance et fiabilité [cs.PF]. PARIS-EST CRÉTEIL, 2012. Français. NNT : . tel-02781116

HAL Id: tel-02781116

<https://theses.hal.science/tel-02781116>

Submitted on 4 Jun 2020

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.

THÈSE

Pour l'obtention du grade de

DOCTEUR

de l'Université **PARIS-EST CRÉTEIL**

Laboratoire d'Algorithmique, Complexité et Logique

École doctorale MSTIC : Mathématiques et STIC

Spécialité : Informatique

Présentée par

Idriss ISMAËL AOULED

Performance et fiabilité des systèmes de grandes tailles par les méthodes de bornes stochastiques

Directeur de thèse : **Prof. Nihal PEKERGIN, UPEC**

Encadrante : **Prof. Hind CASTEL-TALEB, Télécom SudParis**

Soutenue le 18 décembre 2012

Devant la Commission d'Examen

JURY

Prof. Kamel BARKAOUI	CNAM, Cédric	Rapporteur
Prof. Patrice MOREAUX	POLYTECH Annecy-Chambéry, LISTIC	Rapporteur
Prof. Tulin ATMACA	Télécom SudParis	Examineur
Prof. Karim DJOUANI	Université Paris Est (UPEC)	Examineur
Dr. Franck QUESSETTE	Université de Versailles-Saint-Quentin (UVSQ)	Examineur

Remerciement

Table des matières

Table des figures	vii
Liste des tableaux	ix
Introduction générale	1
1 Vers une démarche d'évaluation des performances et de la fiabilité	2
2 Formalismes de modélisation	4
2.1 Réseaux de Petri Stochastique	5
2.2 Réseaux d'Automates Stochastiques	5
2.3 Réseaux de Files d'Attentes	6
3 Méthodes d'analyse quantitative	7
3.1 La mesure	7
3.2 La simulation	7
3.3 Analyse opérationnelle	8
3.4 Méthodes analytiques	9
4 Solutions à forme produit	9
5 Méthodes numériques	10
6 Méthodes d'approximation	10
7 Méthodes des bornes stochastiques	11
8 Apport et organisation de la thèse	12
8.1 Apport de la thèse	12
8.2 Organisation de la thèse	13

I	La comparaison stochastique sur des espaces multidimensionnels	17
1	Introduction	17
2	Les ordres sur des espaces multidimensionnels	19
3	La comparaison stochastique	20
3.1	Comparaisons stochastiques de variables aléatoires	20
3.2	Comparaisons stochastiques de chaînes de Markov	22
4	Le couplage	24
4.1	Exemple	25
5	Méthode des ensembles croissants	27
II	Définition de sous-réseaux bornants dans les réseaux de files d'attente	31
1	Etude d'un réseau de files d'attente $M/M/c_i/k_i$	32
1.1	Description du système	33
1.2	Définition de la projection du processus $X(t)$	34
1.2.1	Projection du processus $X(t)$	35
1.2.2	Définition des taux de transitions de $g(X(t))$	36
1.3	Le couplage pour la construction de la borne supérieure	37
1.4	Calculs des taux de transition des systèmes bornants	39
1.4.1	Borne supérieure	39
1.4.2	Borne inférieure	41
1.5	Analyse numérique des bornes	43
2	Etude d'un réseau G-Networks avec catastrophes	51
2.1	Description du G-Networks avec catastrophes	53
2.2	Définition des bornes du G-Networks avec catastrophes	54
2.2.1	Projection du processus $X(t)$	54
2.2.2	Définition des taux de transition	55
2.3	Construction de la borne supérieure	56
2.4	Construction de la borne inférieure	60
2.5	Comparaison des mesures transitoires	63
2.6	Analyse numérique des indices de performances transitoires	64
3	Conclusion	68
III	Evaluation de la performabilité d'un système de télécommunication	69
1	Modèle composite d'Erlang avec perte	70
1.1	Description du modèle	70

1.2	Résolution approximative par approche MRM	72
2	Système bornant à forme produit	73
2.1	Définition d'un système bornant à forme produit	73
2.2	Comparaison stochastique des processus $X(t)$ et $X^{S1}(t)$	74
3	Systèmes bornants agrégés	77
3.1	Construction des modèles agrégés	77
3.2	Monotonie du processus bornant agrégé	79
3.3	Calculs des taux des systèmes bornants agrégés	80
3.3.1	Définition des taux de la borne supérieure	80
3.3.2	Définition des taux de la borne inférieure	81
4	Résultats	82
5	Conclusion	85
IV Contrôle d'admission et handover forcé dans les réseaux mobiles		87
1	Approche combinée de contrôle d'admission et de handovers forcés dans un réseau monocellulaire	88
1.1	Description du mécanisme proposé	89
1.2	Description de la cellule étudiée	91
1.3	Définitions des transitions des différents mécanismes	92
1.3.1	Les transitions du système MAHO	92
1.3.2	Les transitions du système MG	95
1.3.3	Les transitions du système LMG	97
1.4	Comparaison stochastique des systèmes MG1 et LMG1	100
1.5	Résultats numériques des mécanismes étudiés	102
2	Approche combinée de contrôle d'admission et de handovers forcés dans un réseau multicellulaire	104
2.1	Description du système multicellulaire	104
2.2	Définition des sous-réseaux bornants	108
2.2.1	Projection du processus $X^{LMG2}(t)$	108
2.2.2	Définition des taux de transition	109
2.3	Construction de la borne supérieure	111
2.4	Construction de la borne inférieure	114
3	Resultats numériques sur un réseau multicellulaire	116
3.1	LMG2 vs MG2 : l'influence de la charge	117
3.2	LMG2 et MG2 : l'influence du seuil s	118

3.3	Agrégation bornante : l'impact de la taille des sous-systèmes bornants sur la qualité	120
V	Conclusions générales et perspectives	123
1	Conclusions	123
1.1	Méthode de couplage appliquée à des réseaux de files d'attente	124
1.2	Application de la comparaison stochastique sur des réseaux de communications	126
2	Perspectives	129
2.1	Couplage par fonction de projection : application à des formalismes de haut niveau	129
2.2	Vers un algorithme de couplage de processus	130
2.3	Estimation de l'erreur entre les bornes	130
2.4	Choix de l'ordre sur l'espace d'états	131
2.5	Comparaison de systèmes non-markoviens et semi-markoviens	131
2.6	Ordres stochastiques faibles	131
2.7	Applications futures	132
A	Publications	133
	Références bibliographiques	135

Table des figures

1	Evaluation de performance d'un système	2
2	Analyse opérationnelle sur une boîte noire	8
II.1	Système de 5 files $M/M/c_i/k_i$	34
II.2	Probabilité de blocage de la file 3	44
II.3	Influence de c_1 sur la probabilité de blocage de la file 3	45
II.4	Probabilité de blocage de la file 3 pour $c_2 = 20$	47
II.5	Probabilité de blocage de la file 3 pour $c_2 = 15$	48
II.6	Comportement transitoire de la probabilité de blocage de la file 3 pour différents sous-réseaux bornants pour $c_2 = 15$	48
II.7	Système de 10 files d'attente $M/M/c_i/k_i$	49
II.8	Probabilité de blocage de la file 9 pour différents sous-réseaux bornants, $c_i = 15$ $\forall 1 \leq i \leq 10$	50
II.9	Probabilité de blocage de la file 9 pour différents sous-réseaux bornants, $c_i = 25$ $\forall i \neq 9$ et $c_9 = 15$	51
II.10	Le système G-Network étudié	65
II.11	Temps Tf_7 avant la première infection dans la file 7, selon les différents sous-réseaux.	65
II.12	Encadrement du temps de la première infection de la file 7.	66
II.13	Nombre de clients positifs dans la file 7 pour $\mu_5 = \mu_6 = 2$ et $t \in [0, 10]$	67
III.1	Modèle composite Erlang avec perte pour $n = 3$	71
III.2	Modèle de performance pour $n = 3$	72
III.3	Modèle de disponibilité pour $n = 3$	73

III.4	Modèle bornant à forme produit pour $n = 3$	75
III.5	Agrégation des états avec z identiques	78
III.6	Modèle bornant agrégé de naissances et de morts	78
III.7	Influence de λ sur la probabilité de blocage	83
III.8	Influence du nombre de canaux n	83
III.9	Impact de l'indisponibilité sur les bornes, $1/\mu = 20$ min, $\gamma = 5/hours$, $1/\tau = 15min$, $n = 10$	84
IV.1	La chaîne de Markov du système avec le mécanisme MAHO	94
IV.2	La chaîne de Markov du système avec le mécanisme <i>LMG1</i>	99
IV.3	Probabilité de blocage des nouveaux appels	103
IV.4	Probabilité de dropping des handovers	103
IV.5	Système de réseau mobile avec 7 cellules	117
IV.6	Mesures de performances de la cellule 1 pour un système peu chargé ($\lambda_1 = 10$).	118
IV.7	Mesures de performances de la cellule 1 pour un système chargé ($\lambda_1 = 30$).	118
IV.8	Comparaison de la probabilité de blocage de <i>MG2</i> vs <i>LMG2</i> sur la cellule 1	119
IV.9	Comparaison de la probabilité de dropping de <i>MG2</i> vs <i>LMG2</i> sur la cellule 1	119
IV.10	Influence du seuil s sur la proba. blocage dans la cellule 1 de <i>MG2</i> vs <i>LMG2</i>	119
IV.11	Influence du seuil s sur la proba. dropping dans la cellule 1 de <i>MG2</i> vs <i>LMG2</i>	119
IV.12	Bornes supérieures de la probabilité de blocage	120
IV.13	Bornes supérieures de la probabilité de dropping.	120
IV.14	Bornes inférieures de la probabilité de blocage.	121
IV.15	Bornes inférieures de la probabilité de dropping.	121
V.1	Temps avant la première infection dans la file 7 pour les différents sous-réseaux	125
V.2	Influence de c_1 sur la probabilité de blocage de la file 3	126
V.3	Probabilité de blocage sur un système monocellulaire.	128
V.4	Probabilité de dropping sur un système monocellulaire.	128
V.5	Mesures de performances pour un système peu chargé ($\lambda = 10$) avec une distribution uniforme du trafic.	129
V.6	Probabilité de blocage avec un trafic non-uniforme.	129

Liste des tableaux

II.1	Taux de transition des files du sous-réseau J (borne supérieure)	41
II.2	Taux de transition des files du sous-réseau J (borne inférieure)	42
II.3	Les taux de transition pour la borne supérieure pour $J_1 = \{2, 3\}$	43
II.4	Les taux de transition pour la borne inférieure pour $J_1 = \{2, 3\}$	44
II.5	Valeurs de paramètres	45
II.6	Taux de transitions pour les bornes supérieure et inférieure pour $J_2 = \{3\}$	46
II.7	Description des sous-réseaux bornant	49
II.8	Taux de transition des files du sous-réseau J (borne supérieure)	59
II.9	Résumé des taux de transition de la borne inférieure	62
II.10	Temps de simulation (en seconds) pour les systèmes exacts et bornants	68
III.1	Probabilité de blocage : mesures exactes, bornantes et approximatives de Trivedi	85
IV.1	Paramètres de la simulation	116
IV.2	Subnetworks description	119
V.1	Temps de simulation (seconds)	124
V.2	Probabilités de blocage : valeurs exactes, bornantes et approximatives de Trivedi	127

Table des matières

Introduction générale

Ces dernières années ont connu une évolution fulgurante à la fois des technologies de l'informatique et des réseaux de communication permettant de fournir des services de plus en plus évolués aux utilisateurs. Ainsi, grâce à l'explosion des hauts débits après les années 2000, nous avons vu le développement chez les particuliers de services très attractifs tels que la Vidéo à la Demande, la TV haute définition, et les jeux 3D en réseaux.

Ces évolutions sur les débits mais aussi sur les technologies de l'informatique (précisément du Web2.0) a permis aussi le développement de nouveaux services tels que les réseaux sociaux qui connaissent un très grand succès actuellement.

Le Cloud computing n'en est pas en reste avec le développement sans précédent des offres virtuelles comme les logiciels externes (SaaS, Software as a Service), les plateformes de développement d'applications (PaaS, Platform as a Service) et les infrastructures (IaaS, Infrastructure as a Service).

Enfin, la téléphonie mobile fait complètement partie de notre vie et son développement ne se pose plus en terme de taux de pénétration, l'ARCEP¹ annonçait dès 2011 un taux supérieur à 100% en France, mais de diversité et de qualité de service.

Il est donc clair que devant cette profusion de services, dont nous sommes de plus en plus dépendants, nos exigences sur la qualité des services rendus s'accroissent. Ainsi, le 6 juillet dernier le réseau Orange mobile a connu une panne nationale de 12 heures privant ses abonnés d'appels, de SMS et d'accès à l'Internet mobile sur l'ensemble du territoire. Cet incident a provoqué un grand mécontentement des abonnés.

C'est pour éviter ce genre de problèmes mais aussi parce que les réseaux de communication ainsi que les systèmes informatiques sont de plus en plus complexes en termes de tailles, de

1. <http://www.arcep.fr/index.php?id=35>

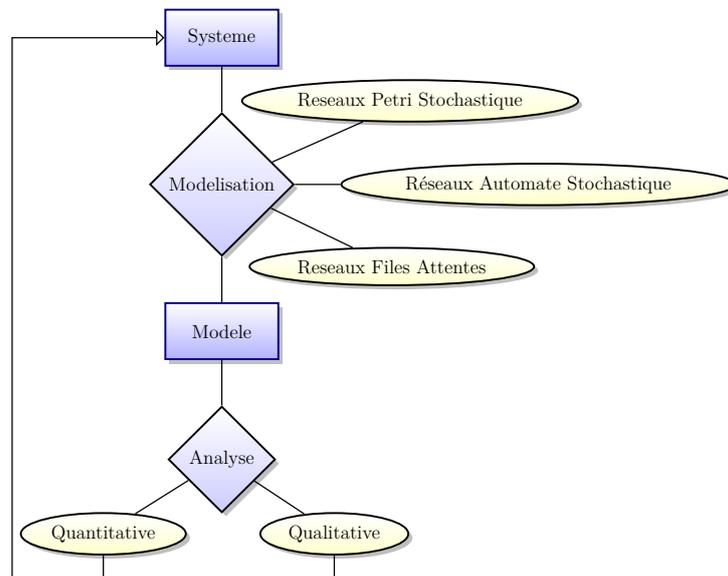


Figure 1 – Evaluation de performance d’un système

diversités techniques et technologiques qu’il faut mettre en œuvre une démarche rigoureuse de prévision des performances et de la fiabilité. Et, aussi importante soit-elle, on ne peut plus se reposer uniquement sur l’intuition et l’expérience.

1 Vers une démarche d’évaluation des performances et de la fiabilité

L’évaluation de performance et de la fiabilité [1] désigne un ensemble d’étapes partant d’un système et aboutissant au calcul des mesures de performances et de fiabilité (Voir Figure1). Comme exemple de mesures de performances, citons le délai de bout en bout dans un réseau ou la probabilité de rejet des appels dans un central téléphonique. En termes de fiabilité, des mesures telles que le taux de pannes d’un routeur sont des paramètres dont les valeurs sont cruciales pour garantir le bon fonctionnement d’un réseau. L’évaluation des performances intervient à deux niveaux dans le cycle de vie d’un système : à sa **conception** et durant son **exploitation**.

- **En conception** : Le système n’existe pas, il s’agit donc de le créer en respectant un cahier des charges. Il est indispensable, vu la complexité des systèmes actuels et de l’exigence des utilisateurs, de mettre en place une démarche claire d’évaluation des performances des systèmes. Ainsi, construire un système sans mener au préalable des études de performance et de fiabilité peut conduire aux problèmes suivants :

- a) Un sous-dimensionnement et donc des performances insuffisantes qui provoquent une dégradation de la qualité de service (QoS). Mais aussi un déficit de fiabilité pouvant engendrer des catastrophes. Nous avons vu l'exemple de l'opérateur Orange mais nous pouvons citer aussi l'explosion de la fusée Ariane V en 96. Ces incidents génèrent une forte dégradation de l'image de la société et des pertes financières importantes.
 - b) Un surdimensionnement entraînant, là aussi, des coûts d'exploitation inutilement élevés et dans certains cas des difficultés de réaliser le système.
- **En exploitation** : Lorsque, sur un système existant, on souhaite réaliser :
- c) Une optimisation : Il s'agit alors d'améliorer les performances du système en évaluant l'impact de tel ou tel changement. Par exemple, un nouveau plan de routage ou une discipline de priorité.
 - d) Des tests pour évaluer les points de ruptures (cas pire) ou la survivabilité du système. Par exemple, un système avionique se doit de résister à une défaillance d'un ou plusieurs éléments du système et continuer à fonctionner même dans des états dégradés. Dans un réseau, on peut s'intéresser à des cas de surcharge du trafic.

Il est donc indispensable de mettre en place une démarche efficace pour évaluer les performances et la fiabilité des systèmes complexes. Cela est d'autant plus important quand il s'agit de systèmes critiques ou sujets à dégradation (pannes). Cette condition de dégradation leur confère une certaine capacité à fonctionner sur plusieurs niveaux de performances qui varient depuis le fonctionnement dans un état parfait (zéro défaut) jusqu'à l'indisponibilité (hors service). Ces systèmes sont définis dans la littérature comme des systèmes tolérants aux fautes ou Multi-states Systems (MSS) [2, 3].

Dans un réseau de communication, la panne d'un lien ou d'un noeud va provoquer une surcharge entraînant une baisse de ses performances. Ainsi, l'analyse des performances du réseau serait trop optimiste si on ignorait les états de panne/réparation. Afin de tenir compte des états de fonctionnement où le système est dégradé, il est nécessaire d'évaluer conjointement la fiabilité et la performance [4] des systèmes. Dans cette thèse, qui s'inscrit dans le thème plus global de la modélisation et de l'évaluation des systèmes, nous nous intéressons entre autre à la prédiction des indicateurs de performabilité qui réfère aux performances du système en présence de pannes et de réparations. Nous précisons que ces indices de performances seront obtenus par des méthodes mathématiques basées essentiellement sur les bornes stochastiques.

D'une manière générale, il existe différentes méthodes pour l'évaluation des performances

des systèmes. Dans le cas où le système existe, le plus simple est peut-être d'effectuer les mesures de performance directement sur le système lui-même, par le biais de sondes matérielles ou logicielles. L'avantage est que les résultats proviennent directement du système réel et non d'un modèle. Toutefois, cette méthode a ses limites d'une part à cause de l'instrumentation qui peut être lourde, et d'autre part pour la collecte des informations et leur interprétation. De plus, elle ne peut s'appliquer que sur un système existant. Or, très souvent le système n'existe pas ou alors (s'il existe) y opérer des manipulations est tout simplement impossible à envisager.

Dans cette thèse, nous nous intéressons à un modèle du système et notre objectif est de développer des méthodes mathématiques d'évaluation des performances.

Dans la section suivante, nous présenterons les différents formalismes de modélisation les plus courants dans l'évaluation de performance avant d'aborder dans les sections suivantes les méthodes permettant de calculer des indices de performances.

2 Formalismes de modélisation

Les systèmes que nous étudions sont des systèmes à événements discrets, c'est-à-dire que l'état de ces systèmes est décrit par des variables d'états discrètes. L'évolution du système est représentée par des changements qui se déclenchent sous l'occurrence d'événements. Par exemple, dans un réseau de communication, l'évolution du système peut-être décrite par le nombre de paquets en attente dans un routeur. Ainsi, les événements tels que l'arrivée des paquets ou leur départ font augmenter ou diminuer la taille de la file d'attente par exemple. Ces systèmes à événements discrets sont présents dans des domaines assez variés tels que les systèmes informatiques, les systèmes de production ou encore les réseaux de communication.

Différents formalismes de modélisation sont utilisés afin de représenter ces systèmes. Certains formalismes sont plus adaptés à une analyse qualitative, et d'autres à une analyse quantitative. Dans [1], l'**analyse qualitative** définit les propriétés structurelles et comportementales du système, telles que la vivacité, les invariants du système ou la stabilité. Alors que l'**analyse quantitative** s'intéresse plus au calcul des mesures de performance. Les travaux qui sont présentés dans cette thèse ne traitent que de l'aspect quantitatif de l'analyse. Pour les aspects qualitatifs, nous laissons le soin au lecteur de se référer à la littérature correspondante [5–7].

L'un des formalismes de modélisation les plus répandus, dans le domaine de l'analyse des

réseaux de communications, est le réseau de files d'attente pour modéliser l'attente sur des ressources partagées. Les réseaux de Pétri et les réseaux d'automates stochastiques sont également utilisés afin de modéliser la synchronisation des processus. Les chaînes de Markov représentent un formalisme de modélisation à un niveau plus bas, permettant d'utiliser des méthodes mathématiques efficaces d'analyse de performance.

Il existe sur le marché des outils logiciels permettant de générer automatiquement une chaîne de Markov à partir des formalismes de spécification de haut niveau tels que les réseaux de Pétri, les réseaux d'automates stochastiques ou les réseaux de files d'attente. Dans cette thèse, nous avons utilisé l'outil SHARPE² [8] développé en 96 par Sahner et Trivedi pour la génération et la résolution de chaînes de Markov et SimEvent³ pour la simulation des systèmes à événements discrets dans un réseau de files d'attente.

Dans la suite, nous présentons différents formalismes de modélisation.

2.1 Réseaux de Petri Stochastique

Dans leur version première, les réseaux de Petri (Petri Nets :PN) ont été développés dès les années 60 [9–11] pour la modélisation et l'analyse qualitative des systèmes complexes présentant des caractéristiques de concurrence et de synchronisation. Une première extension temporelle a été introduite pour adapter les PN à l'analyse quantitative en ajoutant une temporisation sur les transitions [12, 13] ou alors sur les places [14]. Pour tenir compte des comportements probabilistes des systèmes et avoir une équivalence avec les chaînes de Markov, les PN stochastique (SPN) sont proposés dans [15–17] avec différentes sémantiques pour gérer les événements en conflit. Enfin, parmi les extensions des PN, citons la généralisation des SPN (GSPN) qui considèrent à la fois des transitions avec des durées non-nulles, et d'autres avec des durées nulles [18, 19] (transitions immédiates).

2.2 Réseaux d'Automates Stochastiques

Le formalisme des réseaux d'automates stochastique (SAN) est introduit par Plateau [20, 21] pour une représentation à un haut niveau de spécification des chaînes de Markov de grandes tailles. Le système est décrit comme un ensemble de sous-systèmes, exploitant ainsi le fait que les composantes du système peuvent évoluer presque indépendamment rendant ainsi possible le parallélisme (quand les automates n'interagissent pas) et la synchronisation (quand ils interagissent). Chaque sous-système est représenté par un

2. Symbolic Hierarchical Automated Reliability Performance Evaluation

3. <http://www.mathworks.fr/products/simevents/>

automate traditionnel auquel on rajoute des comportements stochastiques et des mécanismes de synchronisation. Ainsi, l'intérêt des réseaux d'automates stochastiques est de pouvoir exprimer la matrice de transition de la chaîne de Markov sous un format de produit tensoriel afin de calculer la distribution de probabilité transitoire et stationnaire. Les premières études ont porté sur le cas des modèles à temps continu [20–22]. Pour représenter les transitions d'un processus stochastique à échelle de temps discrète (distributions géométrique), la notion d'événements compatibles a été introduite pour identifier les événements pouvant avoir lieu simultanément pendant une même unité de temps [23, 24]. Enfin, le formalisme SAN a été enrichi par l'ajout d'un niveau d'abstraction supplémentaire sur l'espace d'états. Une agrégation, directement sur le modèle de haut niveau, est proposée sous certaines conditions. Ces SAN, avec répliquations, sont basés sur le fait qu'un système présente souvent plusieurs entités identiques.

2.3 Réseaux de Files d'Attentes

Le formalisme file d'attente (Queueing Network : QN) est développé pour la modélisation des phénomènes de partage de ressources [19, 25–27]. Ce formalisme est essentiellement utilisé pour l'analyse quantitative des systèmes. Une file d'attente est un système dans lequel les clients **arrivent**, pour recevoir un **service** délivré par un ou plusieurs **serveurs**, puis quittent le système. Les **clients**, accèdent aux ressources de façon prioritaire ou non selon une **discipline de service** et peuvent même subir un temps d'**attente** (bufferisation). L'intérêt des files d'attente est que, dans le cas où les temps des inter-arrivées et les services sont exponentiels, ces systèmes peuvent-être décrits par des chaînes de Markov de type processus de naissance et de mort dont les distributions de probabilité ainsi que des indices de performances peuvent être calculés à partir de formules connues [1]. Ainsi, dans le domaine des télécommunications, la formule d'Erlang B permet de calculer la probabilité de rejet dans un central téléphonique, et l'Erlang C la probabilité d'attente. Dans le cas de lois d'inter-arrivées et de service plus généraux les chaînes de Markov peuvent être déduites, comme par exemple dans le cas de l'Erlang, mais avec des tailles plus importantes que dans le cas de lois exponentielles [1].

Un réseau de files d'attente est un ensemble de files d'attente interconnectées. Ils permettent de représenter par exemple un réseau de routeurs où les paquets sont routés selon une certaine probabilité. Le routage peut être dynamique (selon l'état des files), ou déterministe comme par exemple le routage cyclique. Il existe différents types de réseaux : ouverts ou fermés, mono ou multiclassés qui ont été largement étudiés. Les réseaux de files

d'attente ont été généralisés aux réseaux à clients positifs et négatifs appelés G-Networks (Generalized queueing network ou Gelenbe network) [28]. Les clients négatifs détruisent les clients positifs, ce qui permet de modéliser par exemple des départs synchronisés comme dans le cas de la primitive *Join*. Ils sont également utilisés dans le domaine des réseaux neuronaux.

Dans cette thèse, nous avons essentiellement travaillé sur le formalisme des réseaux de files d'attente et des chaînes de Markov. Nous nous sommes concentrés sur des méthodes mathématiques pour l'analyse quantitative de systèmes dans l'évaluation des performances. Dans la suite, nous présentons les principales méthodes d'analyse quantitative afin de mieux situer celles sur lesquelles nous avons travaillé.

3 Méthodes d'analyse quantitative

3.1 La mesure

La mesure est une méthode empirique de collecte de données issues d'un système en exécution. De ce point de vue, elle impose l'existence d'un système. C'est la façon la plus directe pour évaluer les performances d'un système puisque les mesures sont celles du système lui-même et non ceux d'un modèle.

La mesure peut être aussi employée pour valider les conclusions obtenues à partir d'un modèle d'évaluation de performance. Toutefois, elle n'est pas tout à fait simple à implémenter. En effet, il se pose le problème de la collecte et du stockage des volumes d'informations souvent importants ainsi que leurs interprétations. Mais aussi une instrumentation lourde et des difficultés de positionner les sondes pour la capture des mesures de performances attendues.

Enfin, la mesure est limitée de façon intrinsèque par l'existence d'un système. Ceci réduit considérablement la classe des cas possibles pour son usage.

3.2 La simulation

La simulation à événements discrets consiste à reproduire l'évolution du système dans le temps en étudiant une réalisation particulière du modèle stochastique. L'avantage de la simulation est qu'elle permet d'analyser n'importe quel système ce qui est particulièrement intéressant pour des systèmes où l'on ne peut pas utiliser des méthodes mathématiques. Par contre, son inconvénient est qu'elle est très gourmande en temps de calculs. De plus,

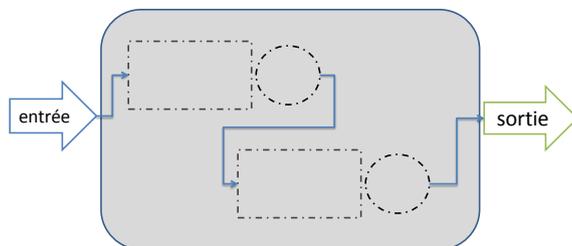


Figure 2 – Analyse opérationnelle sur une boîte noire

il ne s'agit pas d'une technique exacte et les résultats devront être accompagnés d'un intervalle de confiance pour estimer la précision des résultats. De ce point de vue, la simulation nécessite un temps de simulation relativement long. Par exemple, dans le cas des simulations de systèmes à évènements rares, il faut un nombre important de réalisations pour rencontrer au moins une fois ce type d'évènements. Et dans le cas particuliers de systèmes hautement critiques où la panne est un évènement rare, il arrive souvent que ce type d'évènements ne se produise pas sur un ensemble d'exécutions. Donc la simulation peut seulement confirmer l'existence d'un état critique mais ne peut jamais prouver son absence.

3.3 Analyse opérationnelle

L'analyse opérationnelle, voir les travaux de Denning et Buzen [29], consiste à dériver un ensemble de relations à partir d'observations faites sur le système sur une période de temps donnée. Elle permet d'avoir une première approche des mesures de performance d'un système vu comme une boîte noire à partir des paramètres d'entrée et de sortie.

Dans le cas d'une file d'attente simple, l'analyse opérationnelle permet de définir des paramètres de performances en régime transitoire. Dans le cas d'un système stable, on s'intéressera aux valeurs de ces paramètres à l'infini. L'analyse opérationnelle permet d'introduire des critères de performances comme le nombre moyen de clients par unité de temps ou le temps moyen de réponse du système.

Dans le cas de réseaux de files d'attente, on pourra s'intéresser à des paramètres globaux du réseau, ou par station.

3.4 Méthodes analytiques

Les méthodes analytiques permettent une représentation formelle du système sous forme d'équations mathématiques. Il s'agit d'appliquer la théorie des files d'attente afin d'analyser les performances du système. Les mesures de performance peuvent être, dans certains cas, calculées simplement et de façon exacte comme des solutions à forme-produit. Mais le plus souvent, la classe de modèles admissibles aux solutions à forme-produit est très limitée. Dans ce cas, on a recourt à des méthodes de résolutions approximatives ou à des méthodes de bornes.

La section suivante détaille ces méthodes en particulier celles relatives aux encadrements par les techniques de bornes stochastiques qui font l'objet de cette thèse.

4 Solutions à forme produit

Les solutions à forme-produit sont introduites dès les années 60 par Jackson[30] pour les réseaux de files d'attente ouverts, exponentiels avec une discipline PAPS (premier arrivé, premier servi). Dans le cas des réseaux fermés, Gordon et Nowell [31] ont proposé une forme produit.

Cette propriété implique qu'un système multidimensionnel a une solution stationnaire à forme produit lorsque sa solution à l'état stationnaire est le produit de solutions de problèmes à une dimension (les nœuds du système par exemple).

Soit un réseau de Jackson décrit par le processus $(\vec{N}(t))_{t \geq 0}$ où $\vec{N}(t) = (N_1(t), \dots, N_M(t))$ avec $N_i(t)$ nombre de clients présents dans la station i au temps t , le théorème de Jackson nous donne la propriété suivante :

Définition .1. La probabilité stationnaire du réseau possède la forme-produit suivante :

$$\pi(\vec{n}) = \prod_{i=1}^M \pi_i(n_i)$$

où $(\pi_i(n_i))$ est la probabilité stationnaire d'une file $M/M/1$ ayant un taux d'arrivée λ_i et un taux de service μ_i , soit $\pi_i(n_i) = (1 - \rho_i)\rho_i^{n_i}$ où $\rho_i = \frac{\lambda_i}{\mu_i}$.

Les réseaux BCMP [32] généralisent les résultats obtenus, dans le cadre des réseaux de Jackson, aux réseaux multiclassés ouverts, fermés et mixtes et avec différentes disciplines de services et des distributions non-exponentielles.

Les solutions à forme produit sont efficaces pour l'analyse de systèmes multidimensionnels mais les hypothèses sont assez restrictives. Par conséquent, nous devons utiliser d'autres

méthodes pour les systèmes qui n'ont pas de solution à forme produit.

5 Méthodes numériques

Pour les chaînes de Markov n'ayant pas de structures à forme produit, on peut faire appel à des méthodes numériques permettant de résoudre un système d'équations linéaires. Il y a principalement deux types de méthodes numériques : les méthodes directes et les méthodes itératives.

Les méthodes directes modifient la matrice et nécessitent un temps fixe pour calculer la solution, il n'y a pas de problèmes de convergence. Les méthodes itératives consistent à calculer itérativement la distribution de probabilité en utilisant une approche de point fixe. Les méthodes les plus connues sont celle des puissances et de Gauss-Seidel [33].

6 Méthodes d'approximation

Dans un grand nombre de cas, les solutions à forme produit n'existent pas. D'après Van Dijk [34], les phénomènes qui font qu'il n'y ait pas de forme produit sont : les contraintes de capacité, les services bloqués ou les priorités. Il peut, alors, être judicieux de simplifier les hypothèses et de modifier le système de façon à se ramener à un système plus simple représentant une approximation ou un système bornant. Dans [34], on s'intéresse à la mesure de l'écart entre le système modifié et le système original et on cherche à calculer une borne sur l'erreur de la mesure de performance. Par exemple, un réseau de Jackson avec pannes est approximé par un système sans pannes. Cette approximation peut être judicieuse lorsque le système a un taux de pannes faible. Cette idée est reprise pour simplifier un réseau de Jackson à capacité finie par un réseau à capacité infinie.

Les méthodes d'agrégations sont également assez répandues pour approximer les systèmes complexes. Ainsi, dans le cas des réseaux à tandem, on peut regrouper les stations de façon à définir un système simplifié représenté par une seule file [35]. Dans [33, 36], on propose d'agréger des chaînes de Markov ayant des propriétés de "décomposabilité". On identifie au préalable des sous-structures qui peuvent être analysées séparément. Une procédure d'agrégation, qui utilise les résultats calculés de façon indépendante, est ensuite effectuée pour composer les résultats définitifs. L'applicabilité de la méthode doit être vérifiée dans chaque cas. Si la chaîne de Markov a des sous-ensembles d'états tels que ceux-ci sont étroitement couplés les uns par rapport aux autres (taux de transition élevés) à

l'intérieur du sous-ensemble et faiblement couplés avec les états en dehors alors, l'approche peut être appliquée. Les états d'un sous ensemble d'états peuvent être alors agrégés pour former un macro-état. Par conséquent, l'espace d'états est réduit permettant ainsi une simplification de l'analyse. Des études similaires ont également été proposées dans Carasco et Mahevas [37, 38] par analyse de sous-chaînes et/ou agrégation de l'espace d'états. Ces études s'appliquent bien à des systèmes où on peut identifier un sous espace où le processus passe le plus de temps. C'est le cas des systèmes soumis à des pannes car le processus est essentiellement dans des états de fonctionnement ; la panne étant un événement rare. L'intérêt de ces études est de calculer des bornes sur des mesures de performances obtenues à partir des distributions de probabilité stationnaires de taille plus petites.

7 Méthodes des bornes stochastiques

D'une manière générale, l'idée des méthodes de bornes stochastiques consiste à modifier le système initial de façon à définir un système bornant plus simple à étudier. Ainsi, l'approche est plus forte qu'une approximation car le système obtenu fournit une borne sur la mesure de performance initiale. On peut ainsi définir des systèmes bornants inférieurs et supérieurs pour encadrer la mesure de performance exacte. Les méthodes de bornes stochastiques sont basées sur la théorie des ordres stochastiques. Un ordre stochastique se définit comme une relation d'ordre sur des mesures de probabilités, des variables aléatoires, ou encore des processus. L'ordre le plus connu est l'ordre fort "st" [39] qui est équivalent à une comparaison des réalisations. Sur des espaces multidimensionnels des ordres moins contraignants peuvent être aussi définis (dits ordres "wk" et "wk*") [40].

La comparaison de chaînes de Markov, définies sur un même espace d'état, peut se faire selon différentes méthodes. Le couplage de processus [41] revient à coupler les trajectoires des processus de façon à comparer les réalisations après le déclenchement d'un même événement. Cette méthode est assez intuitive et permet de générer l'ordre fort "st". La méthode des ensembles croissants revient à comparer les processus à partir de familles d'ensembles croissants [40]. Cette méthode est assez générale car elle permet de définir l'ordre "st", mais aussi les ordres "wk" et "wk*". Dans [42], Castel-Taleb établit des liens entre ces ordres. Ces liens sont évidents lorsqu'il s'agit de variables aléatoires, mais ce n'est plus le cas pour des chaînes de Markov.

Les méthodes d'agrégations peuvent être également basées sur les bornes stochastiques, en utilisant la comparaison stochastique par fonctions de projection [43]. Certaines études

d'agrégations bornantes sur des espaces multidimensionnels concernent des espaces d'états totalement ordonnés. Ainsi dans [44], la construction d'une matrice bornante "lumpable" (ou agrégeable) a été proposée sur un espace totalement ordonné. Un algorithme basé sur les bornes stochastiques et la propriété d'agrégeabilité (ou lumpability) a été défini dans [45]. Dans [46], on borne la matrice initiale par une matrice "C-lumpable" afin de calculer des mesures de performance bornantes. Contrairement à ces études, j'ai essentiellement travaillé sur des espaces partiellement ordonnés avec, le plus souvent, l'ordre composante par composante qui permet d'analyser nœud par nœud les performances dans un réseau de communication.

8 Apport et organisation de la thèse

8.1 Apport de la thèse

Cette thèse s'inscrit dans le cadre de l'évaluation des performances des systèmes complexes de grandes tailles. Notre objectif est de contribuer au développement d'une méthode d'analyse des performances des grands systèmes multidimensionnels. Pour ce faire, nous utilisons la comparaison stochastique et la méthode du couplage pour analyser des systèmes difficilement analysables en raison de l'explosion combinatoire de l'espace d'états. Le travail présenté ici s'articule autour des deux points suivants :

- Définition des sous-réseaux bornants pour les réseaux G-Nets et les réseaux multi-serveurs.
- Applications :
 - Evaluation de la performabilité sur le modèle d'Erlang avec perte proposé par Trivedi dans [47]. Nous construisons des systèmes bornants, comme décrits dans le point précédent, pour encadrer la mesure de performance recherchée.
 - Introduction d'un nouveau mécanisme d'admission des appels et de contrôle d'occupation des ressources dans les réseaux mobiles. Ce mécanisme est une amélioration de celui proposé dans [3] par Trivedi. Nous tirons partie de l'occupation asymétrique des cellules dans le temps pour provoquer des départs forcés (Forced HandOver) des appels en périphérie de la cellule.

Nous décrivons, dans la suite, l'organisation de la thèse.

8.2 Organisation de la thèse

Après ce premier chapitre consacré à l'introduction de cette thèse, la suite peut être divisée en, principalement, deux parties. Une première partie théorique qui est constituée des chapitres [I](#) et [II](#). Dans le chapitre [I](#), nous donnons les éléments théoriques nécessaires à la compréhension de la comparaison stochastique. Dans le chapitre [II](#), nous développons une méthodologie basée sur le couplage par fonctions de projection pour la définition de sous-réseaux bornants.

La seconde partie, est composée des chapitres [III](#) et [IV](#). On peut la qualifier d'applicative car nous évaluons les performances de réseaux complexes en utilisant la comparaison stochastiques. Ce sont des réseaux de télécommunications mobiles où l'on étudie des problématiques de performabilité et de délestage de charges entre cellules voisines. Le dernier chapitre de la thèse est réservé à la conclusion et aux perspectives de développement ouvert sur le thème, notamment la mise au point d'algorithmes basés sur le couplage pour la génération de processus agrégés bornants.

Chapitre [I](#) : La comparaison stochastique sur des espaces multidimensionnels .

Dans ce chapitre, nous présentons une introduction intuitive des méthodes de bornes stochastiques pour l'évaluation de performances. Nous donnons ensuite les principales notions théoriques des différentes méthodes de la comparaison stochastique illustrées par des exemples.

D'une manière générale, nous avons appliqué la comparaison stochastique de différentes façons.

1. Dans un premier cas, on veut comparer des systèmes qui existent pour montrer que les mesures de performances de l'un sont "inférieures" ou "supérieures" à l'autre. Cela peut être utile quand un système est suffisamment complexe mais que l'on peut comparer avec un autre, plus simple à analyser.
2. Dans un autre cas, seulement un des deux systèmes existe et est beaucoup trop complexe à étudier. On cherche alors à construire, à partir de celui-ci, un système bornant plus simple à analyser (forme produit ou agrégé).

Chapitre [II](#) : Définition de sous-réseaux bornants dans les réseaux de files d'attente .

Généralement, lorsque le réseau est de grande taille, l'analyse exacte est très difficile voire impossible si la distribution des probabilités n'a pas de solution évidente (par exemple une forme produit) à cause de l'explosion de l'espace d'états. Comme nous nous intéressons également au calcul de la distribution des probabilités transitoires, les

calculs sont encore plus complexes car les CMTC (Chaîne de Markov à Temps Continu) sont multidimensionnelles.

Nous utilisons alors la méthode de couplage pour construire des sous-systèmes bornants. L'intérêt de cette méthode est qu'elle est générale et ne suppose pas que l'un des processus soit monotone.

Dans les systèmes de grandes tailles, l'analyse de performance ne porte, en général, que sur une partie du système (un nœud ou un chemin). A partir de cette constatation, nous allons produire des sous-systèmes bornants plus simples à étudier, que le système d'origine. A partir de là, nous encadrons les mesures de performances stationnaires et transitoires.

Nous montrons qu'avec cette approche, nous avons un compromis entre la taille de l'espace d'états et la qualité des bornes obtenues. Nous montrons également que la qualité des bornes est influencée par les paramètres des systèmes en étude.

Le chapitre contient deux illustrations de la comparaison stochastique sur des modèles purement théoriques : les réseaux de files $M/M/c_i/k_i$ et les G-Nets.

Chapitre III : Evaluation de la performabilité d'un système de télécommunication .

Dans ce chapitre, nous considérons un système de commutation téléphonique composé de n canaux avec un nombre infini d'appelants, sujets aux pannes et aux réparations. Pour évaluer ce système complexe, à cause de la prise en compte à la fois des événements arrivées /départs et pannes/réparations, nous proposons d'utiliser les bornes stochastiques pour construire des modèles bornants selon deux approches. L'une consiste à modifier le système initial pour obtenir un système à forme produit et l'autre obtenu en réduisant l'espace d'états par agrégation. Nous comparons les résultats en termes de probabilité de blocage et démontrons l'influence de certains paramètres sur les qualités des bornes obtenues.

Chapitre IV : Mécanisme de CAC et évaluation de la performabilité .

Ce chapitre est consacré à l'évaluation de la performabilité des réseaux mobiles et notamment aux mécanismes d'admission des appels (CAC).

Nous introduisons un nouveau mécanisme de CAC et proposons de le comparer avec celui proposé dans [48] par Madan et Trivedi sur un système monocellulaire dans un premier temps. Nous apportons la preuve et donnons quelques résultats numériques pour justifier de la pertinence de notre mécanisme.

Dans la seconde section, nous étudions un modèle plus réaliste représenté par un en-

semble de cellules afin de modéliser les états de toutes les cellules et l'itinérance (trafic des handovers) entre elles. Nous utilisons l'approche développée dans le Chapitre.II en proposant différents sous-réseaux bornants. Cela permet de contourner le problème de l'explosion de l'espace d'états et de souligner l'importance du compromis entre la taille et la qualité des bornes. Enfin, nous présentons quelques résultats qui illustrent l'influence des paramètres sur la précision des bornes.

Chapitre I

La comparaison stochastique sur des espaces multidimensionnels

Les méthodes de bornes stochastiques s'appliquent essentiellement aux chaînes de Markov et permettent d'apporter des solutions intéressantes pour l'évaluation des performances des systèmes. Nous nous focaliserons sur des chaînes de Markov multidimensionnelles permettant de modéliser des systèmes complexes. L'objectif de ce chapitre est d'expliquer d'abord d'une manière intuitive comment utiliser les méthodes de bornes stochastiques pour l'évaluation des systèmes complexes. Ensuite, on rentrera dans les détails théoriques en présentant différentes méthodes de comparaison stochastique. Ce chapitre est organisé comme suit :

1. Définition d'un ordre sur un espace d'états multidimensionnel.
2. La comparaison stochastique : définitions entre des variables aléatoires et des chaînes de Markov.
3. Méthodes de comparaisons stochastiques : le couplage et les ensembles croissants
4. Comparaison par fonction de projection.

1 Introduction

Un système complexe, comme vu précédemment, peut-être un système de grande taille représentant un réseau avec de multiples nœuds. Il peut être aussi un système où différents types d'événements peuvent se déclencher (arrivées/services, pannes/réparations ou signaux déclencheurs). Ainsi la prise en compte de ces événements peut, de ce fait, entraîner une augmentation du nombre de composantes nécessaires pour représenter un état du système.

Les solutions apportées par les méthodes de bornes stochastiques consistent à construire des systèmes bornants plus faciles à étudier. A partir de ces derniers, on pourra générer des mesures de performances bornantes (supérieures et inférieures) de la mesure exacte. Nous soulignons que les bornes ainsi obtenues sont valables aussi bien en **stationnaire** qu'en **transitoire**. La comparaison stochastique est basée sur la théorie des ordres stochastiques qui est compliquée lorsque l'espace d'état est multidimensionnel [39, 40, 49] car plusieurs ordres stochastiques peuvent être définis et les méthodes de comparaison sont plus difficiles à utiliser que sur un espace monodimensionnel. L'ordre le plus connu est l'ordre fort noté \preceq_{st} (dit ordre "strong"), équivalent à une comparaison des trajectoires ("sample path ordering"). Dans cette thèse, nous ne travaillons que sur cet ordre et donc, nous ne présenterons que celui-ci.

Nous supposons que le système peut être représenté par une chaîne de Markov à temps continu (CMTC), notée $\{X(t), t \geq 0\}$, et définie sur un espace d'état A (représenté en général par \mathbb{N}^n). On note par $\Pi(x, t)$ la probabilité à l'instant t que le processus soit à l'état x . Nous supposons que l'espace d'état A est muni d'un préordre¹ (exemple : l'ordre composante par composante). Nous voulons calculer la mesure de performance suivante :

$$R(t) = \sum_{x \in A} \Pi(x, t) f(x) \quad (\text{I.1})$$

où f est une fonction croissante $A \rightarrow \mathbb{R}^+$. Quand $t \rightarrow \infty$, si le processus a un comportement stationnaire, on note par $\Pi(x)$ la probabilité stationnaire d'être à l'état x . R représente alors la mesure calculée à partir de la distribution stationnaire. Si $\Pi(t)$ (ou Π) n'a pas de solution simple alors, le calcul de $R(t)$ (ou R) devient difficile quand la taille de l'espace d'état A est importante.

Nous proposons d'appliquer la comparaison stochastique. C'est-à-dire, borner $\{X(t), t \geq 0\}$ par $\{X^S(t), t \geq 0\}$ (au sens borne supérieure) et $\{X^I(t), t \geq 0\}$ (au sens borne inférieure) dans le sens de l'ordre stochastique \preceq_{st} : tel que $\{X^S(t), t \geq 0\}$ et $\{X^I(t), t \geq 0\}$ soient plus faciles à analyser car les distributions sont calculables plus facilement ou définies sur un espace d'état plus petit (voir Fourneau et al. [45]). Ainsi, on peut déduire la mesure bornante $R^S(t)$, calculée à partir de la distribution $\Pi^S(t)$ et tel que :

$$R(t) \leq R^S(t) \quad (\text{I.2})$$

1. voir la section suivante pour la définition du préordre

ou $R^I(t)$, calculée à partir de la distribution $\Pi^I(t)$ et tel que :

$$R^I(t) \leq R(t) \tag{I.3}$$

2 Les ordres sur des espaces multidimensionnels

Dans cette section, nous donnons les bases théoriques essentielles à l'utilisation des ordres stochastiques sur des espaces multidimensionnels. Dans la plupart des cas, je travaillerai sur l'espace $A = \mathbb{N}^n$ qui est discret, dénombrable, muni d'au moins un préordre \preceq (relation binaire au moins réflexive et transitive :

1. Réflexivité : $\forall x \in A, x \preceq x$
2. Transitivité : $\forall x, y, z \in A$, si $x \preceq y$ et $y \preceq z$ alors, $x \preceq z$.

Ainsi, un ordre partiel sur un ensemble A est un préordre vérifiant en plus l'antisymétrie :

- 3 Antisymétrie : $\forall x, y \in A$, si $x \preceq y$ et $y \preceq x$ alors, $x = y$.

Un ordre total est un ordre partiel qui satisfait à la propriété de comparabilité :

- 4 Comparabilité : $\forall x, y \in A, x \preceq y$ ou $y \preceq x$

Par exemple, sur $A = \mathbb{N}^n$, l'ordre lexicographique est un ordre total alors que l'ordre composante par composante (noté \leq) est partiel. Nous utilisons souvent l'ordre composante par composante, \preceq , pour la comparaison des processus multidimensionnels :

$$\forall x, y \in \mathbb{N}^n, x \preceq y \Leftrightarrow x_i \leq y_i, \forall 1 \leq i \leq n$$

Ainsi, l'ordre composante par composante permet de comparer file par file les réseaux de communication, ce qui génère des inégalités sur les mesures de performances telles que les temps de réponses, ou les probabilités de blocage d'une file. On peut remarquer que cet ordre partiel peut être plus intéressant que l'ordre total lexicographique car cela génère moins d'inégalités à vérifier pour la comparaison de chaînes de Markov et donc, une meilleure qualité des bornes. Dans la suite, nous définissons la comparaison stochastique de variables aléatoires et des processus markoviens sur un espace A muni d'au moins un préordre \preceq .

3 La comparaison stochastique

La comparaison stochastique est fondée sur les ordres stochastiques. Un ordre stochastique est une relation d'ordre qui permet de comparer des variables aléatoires, des processus stochastiques ou des distributions de probabilité. Il existe deux grandes familles d'ordres : les ordres intégraux et les ordres ensemblistes. Nous orientons le lecteur intéressé à l'article de Massey [40] pour les ordres ensemblistes et aux ouvrages de Stoyan [50] et de Shaked et al. [51] pour les ordres intégraux. La thèse de Taleb [52] propose une couverture des deux approches et regroupe les principaux résultats et définitions sur les ordres stochastiques ensemblistes et intégraux ainsi que, les liens d'équivalences qui peuvent exister entre eux. Nous présentons dans la suite les définitions de base dans le cas de l'ordre fort st .

3.1 Comparaisons stochastiques de variables aléatoires

Soit donc, un espace d'état A muni d'un préordre \preceq . Considérons deux variables aléatoires X et Y définies sur A avec des mesures de probabilités respectives p et q telles que $p[i] = Prob(X = i)$, $\forall i \in A$ (resp. $q[i] = Prob(Y = i)$, $\forall i \in A$). Leur comparaison suivant l'ordre fort \preceq_{st} , est définie comme suit [39] :

Définition .2. $X \preceq_{st} Y$ si et seulement si $E[f(X)] \leq E[f(Y)]$, $\forall f : A \rightarrow \mathbb{R}^+$, \preceq *-croissante*

Sur les espaces multidimensionnels d'autres ordres, plus faibles en termes de contraintes, peuvent être définis. Par exemple, dans [40], l'ordre "weak" (noté \preceq_{wk}) permet de comparer les queues de distribution et l'ordre "weak*" (noté \preceq_{wk*}), les fonctions de répartition. Il existe plusieurs formalismes pour définir un ordre stochastique : fonctions croissantes ou ensembles croissants. Ces deux formalismes sont équivalents car les fonctions croissantes peuvent être générées à partir des combinaisons linéaires des fonctions indicatrices des ensembles croissants. Nous proposons d'utiliser celui des ensembles croissants afin de nous ramener à des inégalités entre matrices (matrices des générateurs ou matrices des probabilités de transitions) pour la comparaison de chaînes de Markov.

Soit $\Gamma \subset A$, Γ est un ensemble croissant si et seulement s'il est équivalent à une séquence d'éléments croissants de A . On note par :

$$\Gamma \uparrow = \{y \in A \mid y \succeq x, x \in \Gamma\}$$

et donc la définition formelle d'un ensemble croissant est la suivante :

Définition .3. Γ est un ensemble croissant si et seulement si $\Gamma = \Gamma \uparrow$

Exemples :

- Soit $A = \{0, 1, 2, 3, 4\}$, muni de l'ordre total \leq . $\Gamma_1 = \{2, 3, 4\}$ est un ensemble croissant contrairement à $\Gamma_2 = \{1, 4\}$. En effet, nous avons $1 \leq 2$ et $1 \leq 3$ or 2 et 3 ne sont pas dans Γ_2 :
- Considérons l'ensemble $A = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ pour lequel nous définissons l'ordre composante par composante \preceq . Nous remarquons que :

$$\Gamma_1 = \{(1, 0), (1, 1)\}$$

est un ensemble croissant alors que :

$$\Gamma_2 = \{(0, 0), (0, 1)\}$$

n'en est pas un. En effet, l'état $(1, 1)$ n'appartient pas à Γ_2 alors que, $(1, 1) \succeq (0, 1)$. Trois ordres stochastiques sont définis à partir des familles d'ensembles croissants [40]. Le plus fort, l'ordre \preceq_{st} , est défini à partir de la famille $\Phi_{st}(A)$ qui contient tous les ensembles croissants de A :

$$\Phi_{st}(A) = \{\Gamma \subset A \mid \Gamma = \Gamma \uparrow\}.$$

Bien entendu, il existe d'autres familles d'ensembles croissants, définies pour les ordres stochastiques faibles [40], mais elles ne sont pas présentées dans cette thèse. Donc pour la suite, même si certaines définitions que nous présentons restent valables pour une famille quelconque d'ordres croissants, nous ne travaillerons que sur l'ordre \preceq_{st} (et la famille $\Phi_{st}(A)$).

L'ordre stochastique \preceq_{st} est défini, à partir de la famille $\Phi_{st}(A)$ [53], comme suit :

Définition .4.

$$\forall \Gamma \in \Phi_{st}(A), X \preceq_{st} Y \Leftrightarrow \sum_{x \in \Gamma} p[x] \leq \sum_{x \in \Gamma} q[x] \quad (\text{I.4})$$

Exemple : Soit $A = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$, l'ensemble muni de l'ordre composante par composante \preceq .

$$Q_{st}(A) = \{\{(1, 1)\}, \{(1, 0), (1, 1)\}, \{(0, 1), (1, 1)\}, \{(0, 1), (1, 0), (1, 1)\}, A\}$$

On suppose que :

$$p = [0.4, 0.2, 0.2, 0.2] \text{ et } q = [0.3, 0.2, 0.2, 0.3]$$

On voit bien que pour

$$\Gamma = \{(1, 1)\}, \sum_{x \in \Gamma} p[x] = 0.2 < \sum_{x \in \Gamma} q[x] = 0.3$$

$$\Gamma = \{(1, 0), (1, 1)\}, \sum_{x \in \Gamma} p[x] = 0.4 < \sum_{x \in \Gamma} q[x] = 0.5$$

En étudiant tous les ensembles croissants on vérifie la définition .4 et donc $X \preceq_{st} Y$.

Dans la suite, nous nous intéressons à la comparaison stochastique de chaînes de Markov.

3.2 Comparaisons stochastiques de chaînes de Markov

Soit $\{X(t), t \geq 0\}$ (resp. $\{X^S(t), t \geq 0\}$) des CMTCs (Chaînes de Markov à Temps Continu) définies sur A . La comparaison stochastique " \preceq_{st} " des chaînes est définie comme une comparaison à chaque instant t [39] :

Définition .5.

$$\{X(t), t \geq 0\} \preceq_{st} \{X^S(t), t \geq 0\}$$

si :

$$X(0) \preceq_{st} X^S(0) \implies X(t) \preceq_{st} X^S(t), \forall t > 0$$

Supposons que la CMTC $\{X(t), t \geq 0\}$ (resp. $\{X^S(t), t \geq 0\}$) soit homogène dans le temps, de générateur Q (resp. Q^S). Si p (resp. q) est un vecteur de probabilité sur A représentant le vecteur des probabilités initiales $P(X(0))$ (resp. $P(X^S(0))$), alors le vecteur des probabilités à l'instant t est $P(X(t)) = p \exp(tQ)$ (resp. $P(X^S(t)) = q \exp(tQ^S)$). Nous avons alors la définition suivante de la comparaison stochastique des CMTCs [40] :

Définition .6.

$$\{X(t), t \geq 0\} \preceq_{st} \{X^S(t), t \geq 0\}$$

si, pour tous les vecteurs de probabilités p et q sur A , nous avons :

$$p \preceq_{st} q \implies p \exp(tQ) \preceq_{st} q \exp(tQ^S), \forall t > 0 \tag{I.5}$$

La comparaison des CMTDs (Chaînes de Markov en Temps Discret) se définit par la comparaison à l'instant n (où $n \in \mathbb{N}$) des chaînes. Ainsi, si $\{X(n), n \geq 0\}$ et $\{X^S(n), n \geq 0\}$ sont deux CMTDs de matrices de probabilités de transition P et P^S , alors nous avons la définition suivante :

Définition .7. $\{X(n), n \geq 0\} \preceq_{st} \{X^S(n), n \geq 0\}$ si, pour tous vecteurs de probabilité p et q sur A , nous avons :

$$p \preceq_{st} q \implies pP \preceq_{st} qP^S \quad (\text{I.6})$$

Lorsque les chaînes sont définies sur des espaces d'états différents, elles peuvent être comparées sur un espace commun [40, 43]. Supposons que $\{X(t), t \geq 0\}$ soit définie sur un espace A , et $\{X^S(t), t \geq 0\}$ sur F . Nous définissons une fonction de projection $g : A \rightarrow F$ afin de les comparer sur l'espace commun F , où le préordre \preceq est défini.

Définition .8.

$$\{g(X(t)), t \geq 0\} \preceq_{\Phi_{st}} \{X^S(t), t \geq 0\}$$

si :

$$g(X(0)) \preceq_{st} X^S(0) \implies g(X(t)) \preceq_{st} X^S(t), \forall t > 0 \quad (\text{I.7})$$

Cette définition peut également s'écrire en fonction des générateurs infinitésimaux [40]. Comme nous verrons plus loin, la comparaison stochastique utilise la monotonie stochastique qui est une propriété de croissance (ou de décroissance) du processus avec le temps. Dans le sens croissant, cette propriété est très utilisée pour la comparaison stochastique. Elle se définit comme suit [39] :

Définition .9. $\{X(t), t \geq 0\}$ est \preceq_{st} -monotone, si :

$$X(t) \preceq_{st} \forall \tau \geq 0, X(t+\tau), \forall t \geq 0. \quad (\text{I.8})$$

Dans le cas discret, la monotonie d'une chaîne de Markov $\{X(n), n \geq 0\}$ se définit comme une croissance en fonction de n . Si P représente la matrice des probabilités de transition, la monotonie se définit comme suit [40] :

Définition .10. $\{X(n), n \geq 0\}$ est \preceq_{st} -monotone si, pour tous les vecteurs de probabilités p et q sur A , nous avons :

$$p \preceq_{st} q \implies pP \preceq_{st} qP \quad (\text{I.9})$$

Pour les CMTCs, la monotonie s'exprime en fonction du générateur infinitésimal [40]. Dans la suite, je présenterai les deux méthodes de comparaison stochastique des chaînes de Markov multidimensionnelles : le couplage et les ensembles croissants.

4 Le couplage

Le couplage, pour la comparaison stochastique \preceq_{st} de variables aléatoires, est équivalent à une comparaison de leurs réalisations :

Théorème .11. *Pour les variables X et Y , définies sur A et de fonctions de répartition F_X et F_Y , les points suivants sont équivalents :*

1. $X \preceq_{st} Y$
2. *Il existe un espace de probabilités $(\Omega, \mathcal{A}, \mathcal{P})$ et des variables aléatoires \widehat{X} et \widehat{Y} , de fonctions de répartition F_X et F_Y , tel que :*

$$\widehat{X}[\omega] \preceq \widehat{Y}[\omega], \forall \omega \in \Omega$$

Dans [54], le couplage est défini entre les mesures de probabilités P (resp. P') de X (resp. Y). Nous avons $P \preceq_{st} P'$ si et seulement s'il existe une mesure de probabilité couplée sur $A \times A$ à valeurs dans $K = \{(x, y) \in A \times A \mid x \preceq y\}$ et dont la première distribution marginale est P et la seconde P' . C'est un cas spécial du théorème de Strassen [39].

Le couplage des chaînes est basé sur la construction d'un couple de trajectoires dépendant des mêmes événements afin de les comparer [55]. Ce formalisme est intéressant car il est intuitif et basé sur la description des systèmes à événements discrets. Pour le couplage de $\{X(t), t \geq 0\}$ et $\{X^S(t), t \geq 0\}$, nous définissons les deux CMTCs sur A :

- $\{\widehat{X}(t), t \geq 0\}$ a le même générateur infinitesimal que $\{X(t), t \geq 0\}$
- $\{\widehat{X}^S(t), t \geq 0\}$ a le même générateur infinitesimal que $\{X^S(t), t \geq 0\}$

Le couplage des processus vise à définir un processus couplé à partir des processus à comparer et dont les composantes représentent les réalisations des processus à comparer [55] :

Théorème .12. *Les propositions suivantes sont équivalentes :*

1. $\{X(t), t \geq 0\} \preceq_{st} \{X^S(t), t \geq 0\}$
2. *il existe la chaîne couplée :*

$$\{(\widehat{X}(t), \widehat{X}^S(t)), t \geq 0\}$$

tel que :

$$\widehat{X}(0)[\omega] \preceq \widehat{X}^S(0)[\omega] \Rightarrow \widehat{X}(t)[\omega] \preceq \widehat{X}^S(t)[\omega], \forall t > 0, \forall \omega \in \Omega \quad (\text{I.10})$$

3. il existe la chaîne $\{Z(t) = (\widehat{X}(t), \widehat{X^S}(t)), t \geq 0\}$ à valeurs dans

$$K = \{(x, y) \in A \times A, x \preceq y\}$$

Le couplage est, bien sûr, applicable à la comparaison des CMTD en comparant les trajectoires à chaque étape $n \in \mathbb{N}$. La monotonie stochastique peut se vérifier également par le couplage en comparant la chaîne avec elle même et en prenant des conditions initiales comparables [56].

Dans le cas où les processus ne sont pas définis sur le même espace d'états, on peut appliquer le couplage par fonctions de projections [43]. Supposons que $\{X(t), t \geq 0\}$ soit défini sur un espace A , et $\{X^S(t), t \geq 0\}$ sur F . On définit une fonction de projection $g : A \rightarrow F$ afin de les comparer.

Théorème .13. *Les propositions suivantes sont équivalentes :*

1. $\{g(X(t)), t \geq 0\} \preceq_{st} \{X^S(t), t \geq 0\}$
2. Il existe la chaîne $\{(\widehat{X}(t), \widehat{X^S}(t)), t \geq 0\}$ tel que $\forall \omega \in \Omega :$

$$g(\widehat{X}(0)(\omega)) \preceq \widehat{X^S}(0)(\omega) \Rightarrow g(\widehat{X}(t)(\omega)) \preceq \widehat{X^S}(t)(\omega), \forall t > 0$$

3. Il existe la chaîne $\{Z(t) = (\widehat{X}(t), \widehat{X^S}(t)), t \geq 0\}$ à valeurs dans

$$K = \{(x, y) \in A \times F, g(x) \preceq y\}$$

Notons que ce théorème peut également être utilisé pour comparer $\{g(X(t)), t \geq 0\}$ avec un processus représentant une borne inférieure, il suffira alors d'inverser les inégalités.

Nous donnons un exemple de couplage par fonction de projection pour définir un processus bornant supérieur.

4.1 Exemple

Pour illustrer l'application du Théorème .13, prenons l'exemple d'un réseau de n files d'attente. Au niveau de chaque file $1 \leq i \leq n$, les paramètres sont les suivants :

- Arrivées de type Poisson et de taux λ_i
- Services exponentiels et de paramètres μ_i .
- Avec la probabilité $p_{i,j}$, les paquets transitent de la file i vers la file j
- Avec la probabilité d_i , ils quittent le système à partir de la file i .

Chapitre I. La comparaison stochastique sur des espaces multidimensionnels

Ce système est représenté par une CMTC $\{X(t), t \geq 0\}$, avec le générateur infinitésimal Q . Considérons la CMTC $\{X^S(t), t \geq 0\}$ avec le générateur infinitésimal Q^S , telle que $g(X(t)) \leq_{st} X^S(t)$. g est la fonction surjective de $\mathbb{N}^n \rightarrow \mathbb{N}$, telle que $g(x_1, \dots, x_n) = \sum_{i=1}^n x_i$, représente le nombre total de clients dans le système.

Définissons Q^S , tel que $(g(X(t)), X^S(t))$ reste dans l'ensemble K :

$$K = \{(x, y) \in \mathbb{N}^n \times \mathbb{N} \mid g(x) \leq y\}.$$

De toute évidence, nous avons deux cas où $g(x)$ peut être modifiée :

1. la première modification est une augmentation correspondant à une arrivée dans une file i .
2. et la seconde, à une diminution traduisant un départ à partir de n'importe quelle file i .

On notera, donc, qu'un transit entre deux files du système n'affecte en rien le nombre total de clients dans le système.

Etudions maintenant les sauts qui font que le processus quitte l'ensemble K défini précédemment. Et définissons les inégalités entre les taux de transition pour garantir que le processus couplé reste dans l'ensemble K .

1. Pour toute transition de la première composante de x vers x' tel que $g(x') = g(x) + 1$, nous pouvons avoir : $g(x) + 1 > y$ (si $\sum_{i=1}^n x_i = y$). Alors, la seconde composante y doit compenser en réalisant un saut de y vers $y + 1$ afin d'avoir $g(x) + 1 \leq y + 1$. Ainsi, la première composante doit augmenter moins vite que la seconde, et nous en déduisons l'inégalité suivante :

$$\sum_{g(x')=g(x)+1} Q(x, x') \leq Q^S(y, y').$$

Comme $\sum_{g(x')=g(x)+1} Q(x, x') = \sum_{i=1}^n \lambda_i$, alors nous avons $Q^S(y, y') = \sum_{i=1}^n \lambda_i$.

2. Si y a une transition vers y' tel que nous ayons $y' = y - 1$, alors nous devons avoir une transition de x vers x' tel que $g(x') = g(x) - 1$. Par conséquent, la première composante doit diminuer plus vite que la seconde.

$$\sum_{g(x')=g(x)-1} Q(x, x') \geq Q^S(y, y').$$

Comme $\sum_{g(x')=g(x)-1} Q(x, x') = \sum_{i=1}^n \mu_i d_i 1_{x_i > 0}$, donc nous avons $Q^S(y, y') = \min_{1 \leq i \leq n} \mu_i d_i$.

Ainsi avec ces taux de transition, $\{X^S(t), t \geq 0\}$ fournit une borne supérieure du processus image de $\{X(t), t \geq 0\}$, par la fonction g :

$$\{g(X(t)), t \geq 0\} \leq_{st} \{X^S(t), t \geq 0\}$$

Dans cet exemple, on notera que le système original est constitué d'un réseau de n files d'attente là où le modèle bornant ne possède qu'une file d'attente. Les taux d'arrivée et de départ ont été définis pour avoir une borne supérieure sur le nombre total de clients dans le système.

Abordons maintenant la comparaison stochastique en utilisant les ensembles croissants. L'intérêt de ce formalisme est qu'il est basé sur l'utilisation de matrices, permettant ainsi de définir des algorithmes. De plus, il permet à la fois la définition des ordres forts et faibles. Cependant comme il a été dit plus tôt, nous n'introduisons que l'ordre \preceq_{st} .

5 Méthode des ensembles croissants

Dans le cas de CMTDs homogènes, la comparaison stochastique peut être démontrée à partir d'inégalités sur les lignes des matrices de probabilités de transitions calculées à partir de familles d'ensembles croissants. Par exemple, si nous considérons l'ordre \preceq_{st} , nous avons des inégalités sur les lignes x et y comparables ($x \preceq y$) de la façon suivante [40] :

Théorème .14.

$$\{X(n), n \geq 0\} \preceq_{st} \{X^S(n), n \geq 0\}$$

si et seulement si, $\forall \Gamma \in \Phi_{st}(A), \forall x \preceq y$:

$$\sum_{z \in \Gamma} P(x, z) \leq \sum_{z \in \Gamma} P^S(y, z)$$

Exemple : Soit $A = \{0, 1, 2, 3\}$ muni de l'ordre \leq

$$P = \begin{pmatrix} 0.3 & 0.3 & 0.2 & 0.2 \\ 0.4 & 0.3 & 0.1 & 0.3 \\ 0.2 & 0.3 & 0.1 & 0.4 \\ 0.1 & 0.2 & 0.2 & 0.5 \end{pmatrix} \quad P^S = \begin{pmatrix} 0.2 & 0.2 & 0.1 & 0.3 \\ 0.3 & 0.2 & 0.1 & 0.4 \\ 0.2 & 0.2 & 0.1 & 0.5 \\ 0.1 & 0.1 & 0.1 & 0.7 \end{pmatrix}$$

$$Q_{st}(A) = \{\{3\}, \{2, 3\}, \{1, 2, 3\}, A\}$$

- On compare la ligne 0 de P avec les lignes 0, 1, 2 et 3 de P^S .
- On vérifie les inégalités sur les autres lignes en utilisant le théorème .14.

On peut déduire donc d'après le théorème .14, que : $\{X(n), n \geq 0\} \preceq_{st} \{X^S(n), n \geq 0\}$. La comparaison stochastique de chaîne de Markov peut se faire en utilisant les chaînes associées, en en utilisant donc le théorème .14. D'autres théorèmes sont basés sur la comparaison des lignes de générateurs. Il faut considérer que l'utilisation du générateur infinitésimal implique qu'il faut considérer l'élément de la diagonale comme appartenant ou non aux ensembles croissants pour éviter le problème des éléments négatifs [39, 53].

Théorème .15.

$$\{X(t), t \geq 0\} \preceq_{st} \{X^S(t), t \geq 0\}$$

si et seulement si, $\forall \Gamma \in \Phi_{st}(A), \forall x \preceq y \mid x, y \in \Gamma, x, y \notin \Gamma :$

$$\sum_{z \in \Gamma} Q(x, z) \leq \sum_{z \in \Gamma} Q^S(y, z)$$

Exemple : Soit $A = \{0, 1, 2, 3\}$ muni de l'ordre \leq

$$Q = \begin{pmatrix} -0.3 & 0 & 0.1 & 0.2 \\ 0.1 & -0.5 & 0.3 & 0.1 \\ 0.2 & 0.3 & -0.7 & 0.2 \\ 0 & 0.2 & 0.2 & -0.4 \end{pmatrix} \quad Q^S = \begin{pmatrix} -0.4 & 0 & 0.2 & 0.2 \\ 0.2 & -0.6 & 0.1 & 0.3 \\ 0.2 & 0.2 & -0.8 & 0.4 \\ 0 & 0.1 & 0.2 & -0.3 \end{pmatrix}$$

- On compare la ligne 0 de Q avec la ligne 0 de Q^S .
- en utilisant le théorème .15, on compare les autres lignes des générateurs.

D'après le théorème .15 on a donc que $\{X(t), t \geq 0\} \preceq_{st} \{X^S(t), t \geq 0\}$.

Si l'un des processus à comparer est monotone, alors nous avons le théorème suivant pour la comparaison stochastique [40] :

Théorème .16. $\{X(t), t \geq 0\} \preceq_{st} \{X^S(t), t \geq 0\}$, si les conditions suivantes sont vérifiées :

1. $X(0) \preceq_{st} X^S(0)$
2. $\{X(t), t \geq 0\}$ ou $\{X^S(t), t \geq 0\}$ est \preceq_{st} -monotone
3. Comparaison des générateurs infinitésimaux :

$$\forall \Gamma \in \Phi_{st}(A), \forall x \in A, \sum_{z \in \Gamma} Q(x, z) \leq \sum_{z \in \Gamma} Q^S(x, z).$$

Dans le domaine des réseaux de files d'attente, la monotonie a été vérifiée dans [56] pour les réseaux de Jackson. Dans [55], elle est également vérifiée pour des réseaux avec des hypothèses similaires, et multi-serveurs. Nous pouvons remarquer que les G-Networks ne sont pas monotones à cause des départs synchronisés.

La monotonie peut se vérifier par des inégalités sur les matrices. Par exemple, pour la monotonie \preceq_{st} d'une CMTD, on utilisera le théorème .14 avec la même matrice de transition P . Nous avons le théorème suivant [39, 53] :

Théorème .17. $\{X(n), n \geq 0\}$ est \preceq_{st} -monotone si et seulement si, $\forall \Gamma \in \Phi_{st}(A), \forall x \preceq y$:

$$\sum_{z \in \Gamma} P(x, z) \leq \sum_{z \in \Gamma} P(y, z)$$

D'une manière équivalente, on peut définir la monotonie stochastique d'une CMTC en utilisant le théorème .15 avec le générateur Q .

Théorème .18. $\{X(t), t \geq 0\}$ est \preceq_{st} -monotone si et seulement si, $\forall \Gamma \in \Phi_{st}(A), \forall x \preceq y \mid x, y \in \Gamma, x, y \notin \Gamma$:

$$\sum_{z \in \Gamma} Q(x, z) \leq \sum_{z \in \Gamma} Q(y, z)$$

Lorsque les processus sont définis sur des espaces d'états différents, ils peuvent être comparés sur un espace commun.

Supposons que $\{X(t), t \geq 0\}$ soit défini sur un espace A , et $\{X^S(t), t \geq 0\}$ sur F . Définissons la fonction de projection $g : A \rightarrow F$ afin de les comparer.

Théorème .19. $\{g(X(t)), t \geq 0\} \preceq_{st} \{X^S(t), t \geq 0\}$, si les conditions suivantes sont vérifiées :

1. $g(X(0)) \preceq_{st} X^S(0)$
2. $\{X(t), t \geq 0\}$ ou $\{X^S(t), t \geq 0\}$ est \preceq_{st} -monotone
3. Comparaison des générateurs infinitésimaux Q et Q^S :

$$\forall \Gamma \in \Phi_{st}(F), \forall x \in A, y \in F \mid g(x) = y, \sum_{g(z) \in \Gamma} Q(x, z) \leq \sum_{z \in \Gamma} Q^S(y, z).$$

Ce théorème peut aussi s'appliquer à des chaînes de Markov à temps discret (CTMD) en remplaçant les générateurs par des matrices de probabilités de transitions.

Exemple : Soient $A = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ muni de l'ordre composante par composante \preceq , et $F = \{0, 1, 2\}$ muni de l'ordre total \leq . On définit la matrice de probabilité

Chapitre I. La comparaison stochastique sur des espaces multidimensionnels

de transition P de la chaîne $\{X(n), n \geq 0\}$, et P^S de la chaîne $\{X^S(n), n \geq 0\}$, comme suit :

$$P = \begin{pmatrix} 0.3 & 0.3 & 0.2 & 0.2 \\ 0.4 & 0.2 & 0.1 & 0.3 \\ 0.2 & 0.3 & 0.1 & 0.4 \\ 0.1 & 0.2 & 0.2 & 0.5 \end{pmatrix} \quad P^S = \begin{pmatrix} 0.3 & 0.5 & 0.2 \\ 0.2 & 0.4 & 0.4 \\ 0.1 & 0.4 & 0.5 \end{pmatrix}$$

La famille des ensembles croissants sur F est :

$$Q_{st}(F) = \{A, \{2\}, \{2, 1\}\}$$

On peut remarquer que $\{X^S(n), n \geq 0\}$ est \leq_{st} -monotone (en appliquant le théorème .17).

On peut remarquer que :

$$\forall \Gamma \in \Phi_{st}(F), \forall x \in A, y \in F \mid \sum_{g(z) \in \Gamma} P(x, z) \leq \sum_{z \in \Gamma} P^S(y, z)$$

et donc les CMTD sont comparables. Comme nous le verrons plus tard, la comparaison par fonction de projection peut-être utile dans le cas de construction de chaînes agrégées bornantes, quand g est une fonction de projection dans un espace d'état plus petit. Il peut, aussi, être utile à la comparaison de processus non markoviens [40]. Nous reviendrons sur ce point dans les perspectives.

Dans la suite de la thèse, nous présentons des exemples d'applications de la comparaison stochastique. D'une manière générale, nous avons appliqué la comparaison stochastique de différentes façons. Dans le premier cas, les systèmes à comparer existent et nous cherchons à montrer qu'un système est inférieur ou supérieur à l'autre. Dans le second cas, le système à étudier est difficilement analysable et nous cherchons à construire, à partir de celui-ci, un système bornant plus simple à analyser par ce qu'il présente une forme produit ou représente une forme agrégée.

Chapitre II

Définition de sous-réseaux bornants dans les réseaux de files d'attente

Dans l'évaluation des performances des réseaux de communication, on s'intéresse souvent qu'à une partie du réseau : délai de bout en bout d'un chemin, probabilité de blocage ou de panne d'un nœud. Lorsque le réseau est de grande taille, l'analyse exacte est très difficile si la distribution de probabilité du système n'a pas de solution évidente (une forme produit par exemple). De plus, la résolution du système par des méthodes numériques ou sa simulation est quasi impossible à cause de l'explosion de l'espace d'états. Nous nous intéressons également au calcul de la distribution de probabilité transitoire qui est très complexe pour des CMTC multidimensionnelles. A partir de cette constatation, nous allons définir des sous-systèmes bornants plus simples à étudier que le système d'origine, et qui nous permettent d'encadrer les mesures de performance stationnaires et transitoires en utilisant la comparaison stochastique.

Avec cette approche, nous proposons un compromis entre la taille de l'espace d'états et la qualité des bornes obtenues. En ce sens, elle est une solution intéressante pour l'évaluation des systèmes complexes. Nous verrons également, comment la qualité des bornes est influencée par les paramètres des systèmes étudiés. Nous appliquons l'approche en étudiant deux types de réseaux :

- Un réseau de files d'attente multi-serveurs pour calculer des bornes sur des probabilités de blocage transitoires. Les résultats de ce travail ont fait l'objet d'une publication dans la revue *Computer journal* en avril 2011.
- Un réseau G-Network avec catastrophe qui modélise par exemple un réseau de communication infecté de virus. On cherchera à calculer des bornes sur les temps avant l'arrivée dans une station d'un premier virus ou le nombre de fois qu'une station a été infectée. Ce travail a été publié dans la conférence *ValueTools* en mai 2011.

Ce chapitre comporte deux grandes sections, une pour chacun des deux réseaux ci-dessus. Dans chacune d'elles, les systèmes sont décrits et la construction des systèmes bornants est détaillée en utilisant la comparaison stochastique basée sur le couplage par fonction de projection dans un espace d'état plus petit. Nous donnons, en fin de chacune des sections, des résultats numériques avant de conclure par une discussion autour des paramètres qui peuvent influencer sur les systèmes bornants.

Comme nous l'avons dit, nous utilisons la méthode du couplage par fonction de projection. L'intérêt de cette méthode est qu'elle est assez générale car elle ne suppose pas que l'un des processus soit monotone. Ainsi, les réseaux multi-serveur sont monotones (voir [56]) mais les G-Networks ne le sont pas à cause du départ synchronisé de clients. Nous proposons d'utiliser cette méthode qui est intuitive car elle est basée sur les équations d'évolution des systèmes et la comparaison des états du système après le déclenchement des événements.

1 Etude d'un réseau de files d'attente $M/M/c_i/k_i$

Nous considérons un système représenté par un réseau de files d'attente $M/M/c_i/k_i$. Avec des hypothèses markoviennes, les modèles de bas niveau sous-jacents sont des CMTMC multidimensionnelles. Pour certains réseaux, avec certaines conditions [1, 35], les solutions à forme produit sont connues à l'état d'équilibre. Par contre, s'il n'y a pas de solution à forme produit, le calcul de la distribution de probabilité stationnaire s'avère très dur voire insoluble en raison de l'explosion de l'espace d'états. L'analyse transitoire de ces systèmes étant, bien évidemment, encore plus difficile [57] car il est impossible de dériver une expression analytique de la distribution transitoire pour le système que nous étudions. De plus, l'utilisation de méthodes numériques pour calculer les distributions stationnaire et transitoire, est difficile à cause de l'explosion combinatoire de l'espace d'états. Dans ce cas, la comparaison stochastique permet d'apporter des solutions intéressantes. Massey, [53], a utilisé la comparaison par les ensembles croissants afin de définir l'ordre faible "weak" entre un réseau de files d'attente du type Jackson et un réseau de files indépendantes $M/M/1$. Il a ainsi pu borner la queue de la distribution transitoire d'un réseau de Jackson par le produit des queues de distributions transitoires des files $M/M/1$. L'intérêt est que les distributions de probabilités transitoires d'une $M/M/1$ sont calculables.

L'approche de Massey consiste à construire un système bornant en coupant les liens entre toutes les files, pour se ramener à un système dont la distribution de probabilité transitoire est calculable. L'approche que nous proposons est différente car nous coupons les liens

entre un sous-réseau et le reste du réseau de façon à définir des sous-réseaux bornants. Ainsi l'analyse transitoire par des méthodes numériques ou la simulation de réseaux de taille plus petite est plus facile. De plus on garde la dynamique du sous-réseau et donc nos bornes sont plus proches du système exact.

Un autre avantage de notre approche est donc, de proposer une famille de systèmes bornants obtenus par découpage, plus ou moins important, des liens entre les files d'attente afin de proposer un compromis entre la précision des bornes et la complexité des calculs. Le point intéressant de ces sous-systèmes bornants est la capacité de maintenir inchangée la dynamique interne puisque nous conservons le transit entre les files d'attente du sous-réseau.

Les taux de transition des systèmes bornants sont calculés en utilisant le couplage par fonction de projection [43] dans l'espace du sous-réseau. La technique est similaire à l'agrégation [58] mais dans notre cas, l'espace d'états obtenu n'est pas un sous-ensemble de l'espace d'état général mais un autre espace d'état qui est plus petit que celui d'origine. De plus, certains algorithmes comme celui de Truffet [44] ne peut pas être utilisé car il nécessite un espace totalement ordonné. Ces algorithmes sont basés sur un formalisme matriciel, alors que la méthode du couplage est basée sur l'étude des équations d'évolution et la comparaison des processus après le déclenchement des événements. Des relations d'inégalités sont obtenues sur les taux de transitions des systèmes. Les résultats obtenus sont assez intuitifs. Par exemple, pour la construction de la borne supérieure, le taux d'arrivée dans une file du sous-réseau est la somme du taux d'arrivée des clients venant de l'extérieur et le taux de transit provenant des files inter connectées et extérieures au sous-réseau. Ainsi pour les files extérieures au sous-réseau, on considère le cas limite où les serveurs sont toujours occupés.

Pour la borne inférieure, le taux d'arrivée dans une file du sous-réseau est égal au taux d'arrivée des clients de l'extérieur. On ne tient pas compte des transits provenant des files extérieures au sous-réseau. Ainsi pour les files extérieures au sous-réseau, on considère le cas limite où elles sont toujours vides.

Dans la suite, nous allons décrire le système sur lequel nous basons cette première étude.

1.1 Description du système

Nous considérons un système général de files d'attente multi-serveurs qui contient n files d'attente à capacités finies et illustré dans la Figure.II.1. Chaque file i est représentée par c_i serveurs et une capacité finie k_i . Ainsi, le système représente l'interconnexion de n files

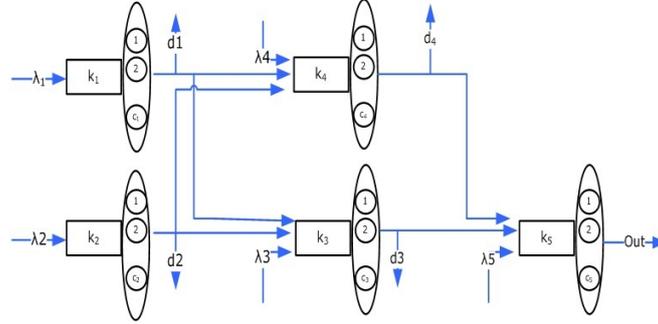


Figure II.1 – Système de 5 files $M/M/c_i/k_i$

$M/M/c_i/k_i$ où chaque file i est définie par les paramètres suivants :

- Arrivées : Les inter-arrivées sont exponentielles de taux λ_i
- Services : Chaque serveur a un temps de service exponentiel de taux μ_i
- Transits : Après qu'un client ait reçu son service, il y a deux possibilités :
 1. avec la probabilité $p_{i,j}$, le client transite de la file i vers la file j si cette dernière n'est pas pleine. Si la file j est pleine, alors le client est perdu.
 2. avec la probabilité d_i , le client quitte le réseau.

En outre, nous supposons qu'un client ne peut revenir sur la file qu'il vient de quitter soit $p_{i,i} = 0$. Alors, nous avons : $\sum_{j \neq i} p_{i,j} + d_i = 1, \forall i = 1 \dots n$.

Soit $\vec{e}_i \in \mathbb{N}^n$, un vecteur tel que tous ses composantes sont nulles à l'exception de la composante i qui est égale à 1. A partir de l'état $\vec{x} = (x_1, \dots, x_n) \in \mathbb{N}^n$, les transitions sont :

- $\vec{x} \rightarrow \vec{x} + \vec{e}_i$ avec le taux λ_i , si $x_i < k_i, 1 \leq i \leq n$.
- $\vec{x} \rightarrow \vec{x} - \vec{e}_i$ avec le taux $\min(x_i, c_i)\mu_i d_i$, si $x_i > 0, 1 \leq i \leq n$.
- $\vec{x} \rightarrow \vec{x} - \vec{e}_i + \vec{e}_j$ avec le taux $\min(x_i, c_i)\mu_i p_{i,j}$, si $(x_i > 0, x_j < k_j), 1 \leq i, j \leq n$, avec $i \neq j$.

1.2 Définition de la projection du processus $\mathbf{X}(t)$

Nous notons par $E = \{1, \dots, n\}$, l'ensemble des indices des nœuds du système de files d'attente considéré. Nous partitionnons cet ensemble en sous-ensembles d'indices disjoints J et H . De sorte que J contienne les nœuds du sous-réseau que nous prenons en compte pour l'étude de la mesure de performance, alors que H est l'ensemble complémentaire. Pour simplifier la notation, supposons que les nœuds du réseau ont une numérotation qui débute par ceux du sous-réseau formé par l'ensemble J . Ainsi, si le sous-réseau J possède k files $k \leq n$, alors elles sont numérotées de $\{1, \dots, k\}$ tandis que pour H , elles

sont numérotées de $\{k+1, \dots, n\}$. Le système que nous considérons est représenté par la CMTC multidimensionnelle $\{X(t), t \geq 0\}$ définie sur l'espace d'états $A = \mathbb{N}^n$. Il est clair que la distribution de probabilités n'a pas de forme produit car les files sont de capacités finies. Nous allons donc nous intéresser à une partie du système en réduisant l'espace d'états grâce à la fonction de projection $g : \mathbb{N}^n \rightarrow \mathbb{N}^k$. Le comportement du système par la fonction de projection g est décrit par $\{g(X(t)), t \geq 0\}$.

Supposons que nous nous intéressons à la probabilité de blocage de la file 3 du système décrit dans la Figure II.1. Alors, le sous-système J , que nous utilisons pour encadrer la mesure de performance, doit contenir la file 3 obligatoirement. Nous pourrions, par exemple, étudier le sous-système $J = \{2, 3\}$ défini en coupant les liens entre J et le reste du système. A savoir, les liens entre 1 et 3 (noté 1-3), 2-4 et 3-5 et dont leurs trafics seront considérés comme externes.

1.2.1 Projection du processus $X(t)$

Soit $\{g(X(t)), t \geq 0\}$, le processus image de $\{X(t), t \geq 0\}$ par la fonction de projection g . Si nous désignons par \vec{z} les états du sous-réseau J , alors :

$$\forall x \in \mathbb{N}^n, g(\vec{x}) = \vec{z} = (x_1, \dots, x_k) \in \mathbb{N}^k \quad (\text{II.1})$$

Nous définissons $\vec{e}_i \in \mathbb{N}^n$ et $\vec{v}_i \in \mathbb{N}^k$, deux vecteurs dont toutes les composantes sont nulles exceptée la composante i qui vaut 1.

Suite à un événement (une arrivée dans une file d'attente, un service, un transit entre les files d'attente), le modèle original $\{X(t), t \geq 0\}$ évolue de l'état $\vec{x} \in \mathbb{N}^n$ vers un état $\vec{x}' \in \mathbb{N}^n$.

Nous distinguons deux cas, selon que l'événement modifie le sous-réseau J ou non. Si l'événement n'a pas d'incidence sur les files d'attente dans J , alors $g(\vec{x}) = g(\vec{x}')$. Sinon, pour $g(\vec{x}')$, nous devons prendre en compte les modifications des files d'attente de J en raison de ces événements. Par exemple, considérons les deux cas suivants à partir de l'état \vec{x} :

- si nous avons un transit de la file i vers la file j , n'appartenant pas au sous-réseau J , alors il existe une transition de l'état \vec{x} vers l'état $\vec{x} - \vec{e}_i + \vec{e}_j$ (où $i, j > k$). Pour ce cas, la projection donne $g(\vec{x} - \vec{e}_i + \vec{e}_j) = g(\vec{x})$ et cet événement n'a pas d'impact sur $g(X(t))$.
- si nous avons une arrivée dans la file $i \in J$, alors il y a une transition de l'état \vec{x} vers l'état $\vec{x} + \vec{e}_i$ (où $1 \leq i \leq k$). La projection de l'état $\vec{x} + \vec{e}_i$ est $g(\vec{x} + \vec{e}_i) = g(\vec{x}) + \vec{v}_i$ (où

$1 \leq i \leq k$), cet événement influe donc sur $g(X(t))$ car $g(\vec{x} + \vec{e}_i) \neq g(\vec{x})$.

Selon les événements se déclenchant dans $X(t)$, $g(X(t))$ évolue comme suit :

1. si nous avons une arrivée dans une file $i \notin J$: si $x_i < k_i$ alors $\vec{x}' = \vec{x} + \vec{e}_i$. Et comme la file i n'est pas dans le sous-réseau alors, $g(\vec{x}') = g(\vec{x})$.
2. si nous avons une arrivée dans la file $i \in J$: si $x_i < k_i$ alors $\vec{x}' = \vec{x} + \vec{e}_i$. La file i n'étant pas dans le sous-réseau alors, $g(\vec{x}') = g(\vec{x}) + \vec{v}_i$.
3. si nous avons un transit de la file $i \notin J$ vers la file $j \in J$: si $x_i > 0$ et $x_j < k_j$ alors, $\vec{x}' = \vec{x} - \vec{e}_i + \vec{e}_j$ et $g(\vec{x}') = g(\vec{x}) + \vec{v}_j$. Autrement, si $x_j = k_j$ alors $\vec{x}' = \vec{x} - \vec{e}_i$. Et comme nous avons supposé la perte du client dans ce cas de figure alors, $g(\vec{x}') = g(\vec{x})$.
4. si nous avons un transit de la file $i \in J$ vers la file $j \in J$: si $x_i > 0$ et $x_j < k_j$ alors, $\vec{x}' = \vec{x} - \vec{e}_i + \vec{e}_j$ et $g(\vec{x}') = g(\vec{x}) - \vec{v}_i + \vec{v}_j$. Et si $x_j = k_j$ alors, $\vec{x}' = \vec{x} - \vec{e}_i$ et $g(\vec{x}') = g(\vec{x}) - \vec{v}_i$.
5. si nous avons un transit de la file $i \notin J$ vers la file $j \notin J$: si $x_i > 0$ et $x_j < k_j$ alors, $\vec{x}' = \vec{x} - \vec{e}_i + \vec{e}_j$ et $g(\vec{x}') = g(\vec{x})$. Et si $x_j = k_j$ alors, $\vec{x}' = \vec{x} - \vec{e}_i$ et $g(\vec{x}') = g(\vec{x})$.
6. si nous avons un transit de la file $i \in J$ vers la file $j \notin J$: si $x_i > 0$ et $x_j < k_j$ alors, $\vec{x}' = \vec{x} - \vec{e}_i + \vec{e}_j$, et $g(\vec{x}') = g(\vec{x}) - \vec{v}_i$. Et si $x_j = k_j$ alors, $\vec{x}' = \vec{x} - \vec{e}_i$ et $g(\vec{x}') = g(\vec{x}) - \vec{v}_i$.
7. si nous avons un service dans la file $i \in J$ pour quitter le réseau : si $x_i > 0$ alors, $\vec{x}' = \vec{x} - \vec{e}_i$ et $g(\vec{x}') = g(\vec{x}) - \vec{v}_i$.
8. si nous avons un service dans la file $i \notin J$ pour quitter le réseau : si $x_i > 0$ alors $\vec{x}' = \vec{x} - \vec{e}_i$, et $g(\vec{x}') = g(\vec{x})$.

A ce stade, nous pouvons déduire l'évolution générale du processus $g(X(t))$. Pour les cas 1, 5 et 8, les événements n'ont pas d'influence sur le sous-réseau et $g(\vec{x}') = g(\vec{x})$. Pour les autres événements, on voit bien que le processus $g(X(t))$ est modifié.

Dans la suite, nous allons définir les taux de transition de $g(X(t))$

1.2.2 Définition des taux de transitions de $g(X(t))$

- Si nous avons une arrivée dans la file $i \in J$ (cas 2) ou une transition de la file $j \notin J$ vers la file $i \in J$ (cas 3¹) alors, $g(\vec{x}') = g(\vec{x}) + \vec{v}_i$. Le taux de transition équivalent est

1. on notera que le sens du trafic est inversé entre les files i et j

alors la somme des taux des cas 2 et 3 :

$$\sum_{\vec{x}' | g(\vec{x}') = g(\vec{x}) + \vec{v}_i} Q(\vec{x}, \vec{x}') = \lambda_i + \sum_{j \notin J} \min(x_j, c_j) \mu_j p_{j,i} \quad (\text{II.2})$$

- Si nous avons un transit de la file $i \in J$ vers la file $j \in J$ (cas 4) si $x_j < k_j$, alors, $g(\vec{x}') = g(\vec{x}) - \vec{v}_i + \vec{v}_j$. Le taux de transition est alors :

$$\sum_{\vec{x}' | g(\vec{x}') = g(\vec{x}) - \vec{v}_i + \vec{v}_j} Q(\vec{x}, \vec{x}') = \min(x_i, c_i) \mu_i p_{i,j} \quad (\text{II.3})$$

- Si nous avons un transit de la file $i \in J$ vers l'extérieur (cas 7) ou un transit de la file $i \in J$ vers la file $j \notin J$ (cas 6) alors, $g(\vec{x}') = g(\vec{x}) - \vec{v}_i$. Le taux de transition est alors la somme de taux des cas 6 et 7 :

$$\sum_{\vec{x}' | g(\vec{x}') = g(\vec{x}) - \vec{v}_i} Q(\vec{x}, \vec{x}') = \min(x_i, c_i) \mu_i [d_i + \sum_{j \notin J} p_{i,j}] \quad (\text{II.4})$$

Encadrons maintenant le processus $\{g(X(t)), t \geq 0\}$ en utilisant l'ordre partiel composante par composante noté \preceq :

$$\forall \vec{x}, \vec{y} \in \mathbb{N}^n, \vec{x} \preceq \vec{y} \Leftrightarrow x_i \leq y_i, \forall i = 1, \dots, n$$

Nous utilisons le couplage pour définir les processus bornants $\{X^S(t), t \geq 0\}$ pour la borne supérieure et $\{X^I(t), t \geq 0\}$ pour la borne inférieure :

$$\forall t, X^I(t) \preceq_{st} g(X(t)) \preceq_{st} X^S(t) \quad (\text{II.5})$$

où $\{X^S(t), t \geq 0\}$ (resp. $\{X^I(t), t \geq 0\}$) est définie dans l'espace d'états \mathbb{N}^k avec le générateur infinitésimal Q^S (resp. Q^I).

1.3 Le couplage pour la construction de la borne supérieure

Soit $Z(t) = (\widehat{X}(t), \widehat{X}^S(t))$, le processus de couplage associé à K au sens du théorème .13 de la section 4.

Dans la suite, nous étudions $Z(t)$ de telle sorte que ses réalisations restent dans :

$$K = \{(\vec{x}, \vec{y}) \in \mathbb{N}^n \times \mathbb{N}^k \mid g(\vec{x}) \preceq \vec{y}\}$$

1. si nous sortons de K par un saut de la première composante, alors nous devons compenser par un saut de la seconde composante pour rester dans K . Pour les états $(\vec{x}, \vec{y}) \in K$, si \vec{x} transite vers \vec{x}' tel que $(\vec{x}', \vec{y}) \notin K$ (si $g(\vec{x}') \not\leq \vec{y}$), alors la seconde composante \vec{y} devra compenser ce saut, par une transition vers \vec{y}' de façon à ce que $(\vec{x}', \vec{y}') \in K$ ($g(\vec{x}') \leq \vec{y}'$).
2. si nous quittons K par un saut de la seconde composante, alors nous devrons compenser ce saut par un saut de la première composante. Pour les états $(\vec{x}, \vec{y}) \in K$, si \vec{y} transite vers \vec{y}' tel que $(\vec{x}, \vec{y}') \notin K$ (si $g(\vec{x}) \not\leq \vec{y}'$) alors, \vec{x} devra compenser par une transition vers \vec{x}' telle que $(\vec{x}', \vec{y}') \in K$ ($g(\vec{x}') \leq \vec{y}'$).

A partir des points (1) et (2), nous définissons les états frontières de K

Remarque : les états frontières sont les états $(\vec{x}, \vec{y}) \in K$ où se produisent les transitions suivantes :

$$\begin{aligned} \vec{x} \rightarrow \vec{x}' \quad \text{tel que} \quad (\vec{x}', \vec{y}) \notin K \quad (\text{cas1}) \\ \vec{y} \rightarrow \vec{y}' \quad \text{de sorte que} \quad (\vec{x}, \vec{y}') \notin K \quad (\text{cas2}). \end{aligned}$$

Par exemple, si nous considérons la transition de \vec{x} vers $\vec{x} + \vec{e}_i$, alors $(\vec{x} + \vec{e}_i, \vec{y}) \notin K$ si $x_i = y_i$. Alors, les états frontières sont les états $(\vec{x}, \vec{y}) \in K$ pour lesquels $x_i = y_i$. D'une manière générale, les états frontières sont définis en fonction du type de sauts :

1. Pour une arrivée ou un service, (saut qui modifie uniquement une file i de -1 ou $+1$), les états frontières sont $(\vec{x}, \vec{y}) \in K$ tels que $x_i = y_i$.
2. Pour des transits entre des files i et j , les états frontières sont les états $(\vec{x}, \vec{y}) \in K$ tels que $x_i = y_i$ ou $x_j = y_j$

De plus, pour les états $(\vec{x}, \vec{y}) \in K$ considérés, nous supposons que pour que les événements puissent se déclencher, nous avons les conditions suivantes :

- pour les services, $x_i > 0$ et donc $y_i > 0$ puisque $x \preceq y$
- pour les transits de $i \rightarrow j$, $(x_i > 0, x_j < k_j)$ et donc $(y_i > 0, y_j < k_j)$ puisque $x \preceq y$

Avec ces conditions, quand une composante effectue un saut, l'autre peut toujours en effectuer un autre pour compenser.

Pour définir les systèmes bornants, nous allons étudier les états $(\vec{x}, \vec{y}) \in K$ qui sont des états frontières, et avec les conditions citées précédemment. Pour définir les taux de transition des systèmes bornants, nous étudierons les sauts qui font sortir le processus couplé de K , c'est à dire dans les cas suivants :

- un saut de la composante \vec{x} après le déclenchement d'une arrivée.
- un saut de la composante \vec{y} après le déclenchement d'un service.

- un saut de la composante \vec{x} ou de la composante \vec{y} après le déclenchement d'un transit entre des files.

Nous considérons, par la suite, les différents événements qui se produisent sur la file i pour analyser les sauts des composantes. Nous verrons que la compensation des sauts va nous permettre d'obtenir des conditions sur les taux de transitions pour définir les systèmes bornants.

1.4 Calculs des taux de transition des systèmes bornants

Nous commençons par les taux de la borne supérieure.

1.4.1 Borne supérieure

1. Si nous avons une arrivée dans la file $i \in J$ qui génère une transition de l'état \vec{x} vers \vec{x}' telle que $g(\vec{x}') = g(\vec{x}) + \vec{v}_i$ alors, nous devons avoir une transition de \vec{y} vers \vec{y}' telle que $\vec{y}' = \vec{y} + \vec{v}_i$ pour avoir $g(\vec{x}') \preceq \vec{y}'$. Ainsi, les taux de transition doivent satisfaire la condition suivante :

$$Q^S(\vec{y}, \vec{y} + \vec{v}_i) \geq \sum_{\vec{x}' | g(\vec{x}') = g(\vec{x}) + \vec{v}_i} Q(\vec{x}, \vec{x}') \quad (\text{II.6})$$

Puisque l'inégalité (II.6) doit être satisfaite pour tous les états \vec{x} tels que $(\vec{x}, \vec{y}) \in K$, alors nous prenons le maximum de la partie droite pour $\vec{x} \in \mathbb{N}^n$:

$$\max_{\vec{x} | g(\vec{x}) \preceq \vec{y}} \left\{ \sum_{\vec{x}' | g(\vec{x}') = g(\vec{x}) + \vec{v}_i} Q(\vec{x}, \vec{x}') \right\}$$

L'évolution de $\{g(X(t)), t \geq 0\}$, en raison de cet événement, est donnée dans l'équation (II.2) et le maximum du taux est atteint pour les états \vec{x} tels que $x_j = c_j, \forall j \notin J$. Ainsi, nous choisissons :

$$Q^S(\vec{x}, \vec{x} + \vec{v}_i) = \lambda_i + \sum_{j \notin J} c_j \mu_j p_{j,i}, \forall i \in J.$$

2. Si nous avons un service dans une file d'attente $i \in J$ pour \vec{y} alors, nous devons avoir aussi un service pour \vec{x} . L'évolution de $X(t)$ est donnée par l'équation (II.4). Etant donné que, dans les états frontaliers, $y_i = x_i$ nous prenons le même taux pour

$g(X(t))$. Ainsi, nous avons :

$$Q^S(\vec{y}, \vec{y} - \vec{v}_i) = \min(y_i, c_i) \mu_i [d_i + \sum_{j \notin J} p_{i,j}], \forall i \in J.$$

3. le transit de la file i vers la file j peut générer un saut en dehors de K pour la première ou la seconde composante :

- (a) Pour la première composante \vec{x} , si nous avons un transit entre les files d'attente i et j alors, cet événement peut être compensé par une hausse similaire $(-\vec{v}_i + \vec{v}_j)$ ou par une arrivée dans la file d'attente j $(+ \vec{v}_j)$ pour la seconde composante \vec{y} . Comme le taux d'arrivée dans la file d'attente j a déjà été utilisé pour la compensation des arrivées dans la file d'attente j (cas 1²), seul le taux de transit sera pris en considération. L'évolution de $g(X(t))$ est donnée par l'équation (II.3) alors, on obtient l'inégalité suivante :

$$\min(x_i, c_i) \mu_i p_{i,j} \leq Q^S(\vec{y}, \vec{y} - \vec{v}_i + \vec{v}_j) \quad (\text{II.7})$$

- (b) Pour la composante \vec{y} , la transition entre les files d'attente i et j peut être compensée par le même saut $(-\vec{v}_i + \vec{v}_j)$ ou par un service de la file d'attente i $(-\vec{v}_i)$ sur la première composante \vec{x} .

Ici aussi, le dernier événement a déjà été utilisé pour compenser le saut en raison d'un service sur la deuxième composante dans le cas 2, seul la transition sera pris en compte :

$$Q^S(\vec{y}, \vec{y} - \vec{v}_i + \vec{v}_j) \leq \min(x_i, c_i) \mu_i p_{i,j} \quad (\text{II.8})$$

A partir des équations (II.7) et (II.8), nous prenons $Q^S(\vec{y}, \vec{y} - \vec{v}_i + \vec{v}_j) = \min(x_i, c_i) \mu_i p_{i,j}$. Et comme $x_i = y_i$ alors, il est possible de remplacer x_i par y_i dans les équations (II.7) et (II.8) et nous obtenons :

$$\forall (i \neq j) \in J : Q^S(\vec{y}, \vec{y} - \vec{v}_i + \vec{v}_j) = \min(y_i, c_i) \mu_i p_{i,j}.$$

Par conséquent, nous pouvons définir un système bornant supérieur. Dans le tableau (II.1), nous avons récapitulé les taux de transition de ce système.

Notez que, si nous voulons comparer le sous-réseau et le système bornant, les inégalités données par le théorème 2 dans [59] pour les réseaux de Jackson avec les taux de services

2. en prenant j au lieu de i

généraux sont satisfaits.

Les taux d'arrivée de la borne supérieure sont plus grands, tandis que les taux de service et les probabilités de routage sont identiques à ceux du sous-réseau.

Enfin, le sous-réseau et le système bornant évoluent, tous les deux, de la même manière quand des transitions se produisent.

Intuitivement, pour définir les taux de transition de la borne supérieure :

- nous ajoutons le taux de transit des files extérieures à J au taux d'arrivée des files de J . Comme le trafic provenant des files extérieures à J dépend du nombre de serveurs occupés, on prend le cas limite où tous les serveurs sont occupés
- nous ajoutons le taux de transit sortant des files J et allant vers les files extérieures à J au taux de service vers l'extérieur.
- nous considérons que les transits à l'intérieur d'un sous-réseau ne sont pas modifiés, nous conservons donc la dynamique du sous-réseau.

Définissons maintenant les taux de transitions de la borne inférieure.

1.4.2 Borne inférieure

Nous appliquons le Théorème .13 (cas : $\forall t, X^I(t) \preceq_{st} g(X(t))$) afin de définir $\{X^I(t), t \geq 0\}$. Dans ce cas :

$$K = \{(\vec{x}, \vec{y}) \in \mathbb{N}^k \times \mathbb{N}^n \mid \vec{x} \preceq g(\vec{y})\}$$

Les conditions sur les états frontières sont les mêmes que pour la borne supérieure. Nous appliquons la technique du couplage général sur les états $(\vec{x}, \vec{y}) \in K$ et nous considérons différents événements qui se produisent dans la file d'attente i dans le but d'analyser les sauts des composantes :

1. si nous avons une arrivée dans la file $i \in J$ qui génère une transition de \vec{x} vers \vec{x}' telle que $\vec{x}' = \vec{x} + \vec{v}_i$ alors, nous devons avoir également une transition de \vec{y} vers \vec{y}'

Tableau II.1 – Taux de transition des files du sous-réseau J (borne supérieure)

Taux	Expression
arrivée dans la file $i \in J$	$\lambda_i + \sum_{j \notin J} c_j \mu_j p_{j,i}$, si $y_i < k_i$.
service dans la file $i \in J$	$\min(y_i, c_i) \mu_i [d_i + \sum_{j \notin J} p_{i,j}]$, si $y_i > 0$.
transit de i vers j ($i, j \in J, i \neq j$)	$\min(y_i, c_i) \mu_i p_{i,j}$, si $y_i > 0$, et $y_j < k_j$.

de sorte que, $g(\vec{y}') = g(\vec{y}) + \vec{v}_i$:

$$Q^I(\vec{x}, \vec{x} + \vec{v}_i) \leq \sum_{\vec{y}' | g(\vec{y}') = g(\vec{y}) + \vec{v}_i} Q(\vec{y}, \vec{y}') \quad (\text{II.9})$$

En considérant le taux minimal dans le second terme de l'équation (II.2) ($x_j = 0, \forall j \notin J$), nous obtenons :

$$Q^I(\vec{x}, \vec{x} + \vec{v}_i) = \lambda_i, \forall i \in J.$$

2. si nous avons un service dans la file $i \in J$ pour \vec{y} alors, nous devons avoir aussi un service dans la file i pour \vec{x} . Il résulte de l'équation.(II.4) :

$$Q^I(\vec{x}, \vec{x} - \vec{v}_i) \geq \min(y_i, c_i) \mu_i [d_i + \sum_{j \notin J} p_{i,j}] \quad (\text{II.10})$$

Comme cette inégalité doit être vérifiée pour tout $\vec{y} \in \mathbb{N}^n$ tel que $\vec{x} \preceq g(\vec{y})$ et que $x_i = y_i$ alors, pour obtenir l'inégalité.(II.10) nous prenons :

$$Q^I(\vec{x}, \vec{x} - \vec{v}_i) = \min(x_i, c_i) \mu_i [d_i + \sum_{j \notin J} p_{i,j}], \forall i \in J,$$

3. si nous avons un transit entre les files d'attente i et j pour état \vec{y} tel que $i, j \in J$ alors, nous devons avoir aussi un transit pour l'état \vec{x} (la preuve est similaire au cas de la borne supérieure). Donc :

$$Q^I(\vec{x}, \vec{x} - \vec{v}_i + \vec{v}_j) = \min(x_i, c_i) \mu_i p_{i,j}, \forall i \neq j \in J.$$

Le Tableau.II.2 résume les taux de transition pour une file $i \in J$.

Pour la borne inférieure, le taux d'arrivée est calculé en considérant le cas limite où les files d'attente du reste du réseau sont vides, ce qui correspond à un trafic nul. Le taux de

Tableau II.2 – Taux de transition des files du sous-réseau J (borne inférieure)

Taux	Expression
arrivée dans la file $i \in J$	λ_i , si $x_i < k_i$.
service dans la file $i \in J$	$\min(x_i, c_i) \mu_i [d_i + \sum_{j \notin J} p_{i,j}]$, si $x_i > 0$.
transit de i vers $j \neq i \in J$	$\min(x_i, c_i) \mu_i p_{i,j}$, si $x_i > 0$ et $x_j < k_j$.

service est semblable à celui de la borne supérieure car un client sortant du sous-réseau est considéré comme quittant le système. Les transits entre les files du sous-réseau J ne sont pas modifiés.

Dans la suite, nous effectuons une analyse numérique des systèmes bornants.

1.5 Analyse numérique des bornes

Nous considérons à nouveau le réseau de files d'attente de la Figure.II.1. Nous nous intéressons à la probabilité de blocage de la file 3. Ainsi, les sous-réseaux considérés pour calculer les bornes sur les mesures de performances contiennent la file d'attente 3. Nous fournissons des valeurs numériques par le biais du simulateur *SimEvent* de MathWorks puis discutons de la qualité des bornes proposées en fonction de la taille des systèmes bornants. Nous considérons un intervalle de confiance de 95% avec un écart type de 0,01. En d'autres termes, si la valeur estimée est \hat{X} , la valeur exacte X est comprise entre $\hat{X} - 0.01\hat{X} \leq X \leq \hat{X} + 0.01\hat{X}$ avec une probabilité de 0.95.

Nous proposons d'étudier le sous-réseau de files d'attente $J_1 = \{2, 3\}$. Dans le Tableau.II.3, nous donnons les taux de transition pour la borne supérieure. La borne inférieure a des taux similaires à l'exception du taux d'arrivée dans la file d'attente 3 qui vaut λ_3 (voir Tableau.II.4). Et nous récapitulons, dans le Tableau.II.5, les valeurs des paramètres d'entrée utilisés dans les simulations.

1. Influence des paramètres d'entrée

Le lien entre les files 1 et 3 étant coupé dans le système bornant J_1 , nous étudions l'influence de la probabilité de routage $p_{1,3}$ sur la probabilité de blocage stationnaire de la file 3. Dans la Figure.II.2, nous donnons les probabilités de blocage de la file 3 pour les sous-réseaux bornants et le système exact par analyse de l'ensemble du réseau. Nous notons, dans la figure, par I la borne inférieure et par S la borne supérieure.

Tableau II.3 – Les taux de transition pour la borne supérieure pour $J_1 = \{2, 3\}$

	File 2	File 3
Taux d'arrivée	λ_2	$\lambda_3 + c_1\mu_1 p_{1,3}$
Taux de transit	$2 \rightarrow 3$ $\min(c_2, x_2)\mu_2 p_{2,3}$	pas de transit
Taux de service	$\min(c_2, x_2)\mu_2 d_2$ $+ \min(c_2, x_2)\mu_2 p_{2,4}$	$\min(c_3, x_3)\mu_3 d_3$ $+ \min(c_3, x_3)\mu_3 p_{3,5}$

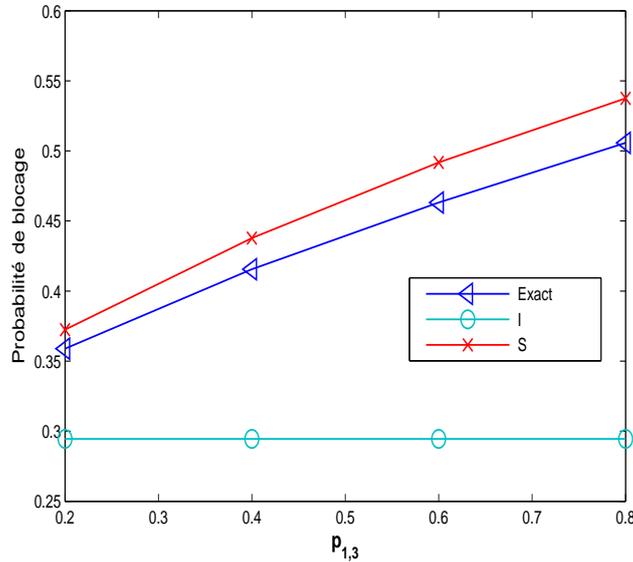


Figure II.2 – Probabilité de blocage de la file 3

Nous pouvons facilement constater que lorsque $p_{1,3}$ augmente alors, la probabilité de blocage augmente pour le système exact et la borne supérieure. Par contre, $p_{1,3}$ n'influe pas ou très peu sur la qualité de la borne inférieure. En effet, $p_{1,3}$ n'est pas prise en compte dans le calcul des taux de transition de la borne inférieure.

Ensuite, nous fixons $p_{1,3} = 0,2$ et nous faisons varier le nombre de serveurs c_1 de la file 1. Nous pouvons voir dans la Figure.II.3 que la qualité de la borne supérieure baisse lorsque le nombre de serveurs c_1 augmente. Ce résultat est évident car le taux d'arrivée dans la file d'attente 3 augmente avec c_1 et la probabilité de blocage de la borne supérieure augmente avec ce taux d'arrivée (voir Tableau.II.3). Par contre, dans le système exact nous atteignons l'état d'équilibre. On peut remarquer aussi que c_1 n'a aucune influence sur la borne inférieure puisque ce paramètre n'est pas pris en compte dans les taux de transition.

Tableau II.4 – Les taux de transition pour la borne inférieure pour $J_1 = \{2, 3\}$

	File 2	File 3
Taux d'arrivée	λ_2	λ_3
Taux de transit	$2 \rightarrow 3$ $\min(c_2, x_2)\mu_2 p_{2,3}$	pas de transit
Taux de service	$\min(c_2, x_2)\mu_2 d_2$ $+ \min(c_2, x_2)\mu_2 p_{2,4}$	$\min(c_3, x_3)\mu_3 d_3$ $+ \min(c_3, x_3)\mu_3 p_{3,5}$

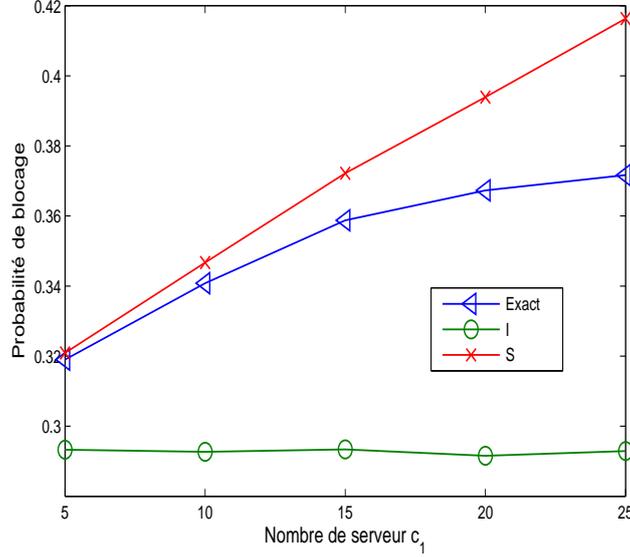


Figure II.3 – Influence de c_1 sur la probabilité de blocage de la file 3

Dans la suite, nous considérons les probabilités de blocage stationnaires et transitoires de la file 3 pour différents sous réseaux de façon à étudier l'impact de la taille du sous-réseau sur la qualité des bornes.

2. Impact de la topologie du sous-réseau bornant

Soit $J_2 = \{3\}$, le sous-réseau bornant obtenu en coupant tous les liens vers le nœud 3. Les taux de transition sont donnés dans le Tableau II.6. A noter que les sous-réseaux bornant J_1 et J_2 fournissent, tous deux, des bornes sur la probabilité de blocage de la file d'attente 3.

A noter également que $J_2 \subset J_1$. En effet, J_2 est obtenu à partir de J_1 en coupant le lien entre les files d'attente 2 et 3.

Nous cherchons à mettre en évidence l'influence de la coupure du lien entre les files

Tableau II.5 – Valeurs de paramètres

Taux d'arrivée	$\lambda_i = 30, i = 1, \dots, 5$
Taux de service	$\mu_i = 2, i = 1, \dots, 5$
Taille de la file	$k_i = 100, i = 1, \dots, 5$
Nombre de serveurs	$c_1 = 20, c_2 = c_3 = c_4 = c_5 = 15$
Probabilités de routage	$p_{1,4} = 0.4, p_{1,3} = 0.2, p_{2,3} = 0.3$ $p_{2,4} = 0.3, p_{3,5} = 0.6, p_{4,5} = 0.6$ $d_i = 0.4, i = 1, \dots, 5$

2 et 3 dans le sous réseau J_2 sur la qualité des bornes.

Dans la Figure.II.4, nous supposons que $c_2 = 20$ et conservons, pour les autres paramètres, les valeurs du Tableau.II.5. Nous notons par I_1 (resp. I_2) la borne inférieure de J_1 (resp. J_2), et par S_1 (resp. S_2) la borne supérieure de J_1 (resp. J_2).

Nous constatons aisément l'influence de la définition des systèmes bornants sur la qualité des bornes supérieures. S_1 fournit une borne de meilleure qualité que S_2 . Ceci est normal car la transition entre les files 2 et 3 est prise en considération dans S_1 tandis que pour S_2 , la même transition est considérée comme un flux externe additionnel sur la file 3. Par contre, la complexité de S_2 est moindre en comparaison de S_1 (si la taille des files est de 100, alors la taille de l'espace d'états de S_2 est divisée par 100). Nous remarquons également que la coupure du lien impacte aussi la borne inférieure puisque I_1 donne une meilleure borne que I_2 .

Dans la Figure.II.5, nous diminuons le nombre de serveurs de la file 2 à $c_2 = 15$. Nous constatons que les probabilités de blocages S_1 et S_2 de la file 3 sont moins élevées que dans le cas précédent (voir Figure.II.4). On note que les bornes supérieures sont plus proches entre elles.

Les résultats obtenus sont logiques : le taux d'arrivée de la borne supérieure S_2 est de $\lambda_3 + c_1\mu_1p_{13} + c_2\mu_2p_{23}$ et le flux d'arrivée venant de la file 2 considéré dans cette somme est $c_2\mu_2p_{23}$. Cela suppose que tous les serveurs de la file 2 sont occupés. Quand on diminue c_2 , la file 2 est plus chargée et donc le flux $c_2\mu_2p_{23}$ que l'on rajoute à la borne supérieure est proche du comportement exact. Dans ce cas, il est intéressant de prendre le sous-réseau bornant J_2 puisqu'il est défini uniquement sur la file 3 et donc les valeurs numériques sont obtenues plus rapidement.

La borne inférieure obtenue I_2 n'a pas changé car elle ne dépend pas de c_2 . Par contre, I_1 diminue légèrement quand c_2 diminue. Donc l'écart entre les deux bornes inférieures est plus petit quand $c_2 = 15$ que lorsque $c_2 = 20$.

Nous pouvons conclure sur cet exemple, que la valeur de c_2 a une influence sur la

Tableau II.6 – Taux de transitions pour les bornes supérieure et inférieure pour $J_2 = \{3\}$

	Borne supérieure 2 File 3	Borne inférieure 2 File 3
Taux d'arrivée	$\lambda_3 + c_1\mu_1p_{1,3}$ $+ c_2\mu_2p_{2,3}$	λ_3
Taux de départ	$\min(x_3, c_3)\mu_3d_3$ $+ \min(x_3, c_3)\mu_3p_{3,5}$	$\min(x_3, c_3)\mu_3d_3$ $+ \min(x_3, c_3)\mu_3p_{3,5}$

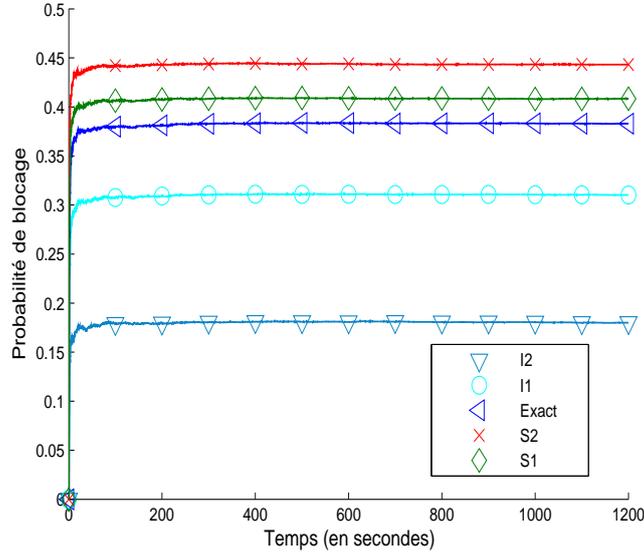


Figure II.4 – Probabilité de blocage de la file 3 pour $c_2 = 20$

qualité des bornes. Ainsi, on peut également remarquer que la topologie des sous réseaux bornants (et donc leur taille) a une influence sur la qualité des bornes, et la valeur des paramètres peut augmenter ou diminuer l'écart entre les bornes.

La Figure.II.6 illustre le comportement transitoire des différents systèmes dans l'intervalle de temps $[0, 100]$. Nous pouvons voir la pertinence de notre approche en fournissant des bornes pour des comportements à la fois stationnaires et transitoires.

3. *Etude de systèmes de grande taille*

Nous proposons maintenant d'étudier un système plus complexe de 10 files d'attente représenté dans la Figure.II.7 et pour lequel nous considérons les paramètres suivants :

$$\left[\begin{array}{l} \lambda_i = 30 \text{ et } \mu_i = 2, \forall i \in [1, \dots, 10] \\ p_{1,2} = p_{1,3} = p_{3,5} = p_{3,6} = 0, 3 \\ p_{1,4} = p_{3,7} = 0, 4 \\ p_{2,5} = p_{2,9} = p_{5,8} = p_{5,9} = 0, 5 \\ p_{6,9} = p_{8,10} = p_{9,10} = d_{10} = 1 \end{array} \right]$$

L'objectif est de calculer la probabilité de blocage de la file 9 pour les sous-réseaux bornants décrits dans le Tableau.II.7. Les systèmes de bornes sont construits en coupant progressivement les liens de façon à obtenir $J_4 \subset J_3 \subset \dots \subset J_1$. Nous obtenons ainsi des systèmes plus faciles à analyser et qui permettent d'encadrer l'indice de

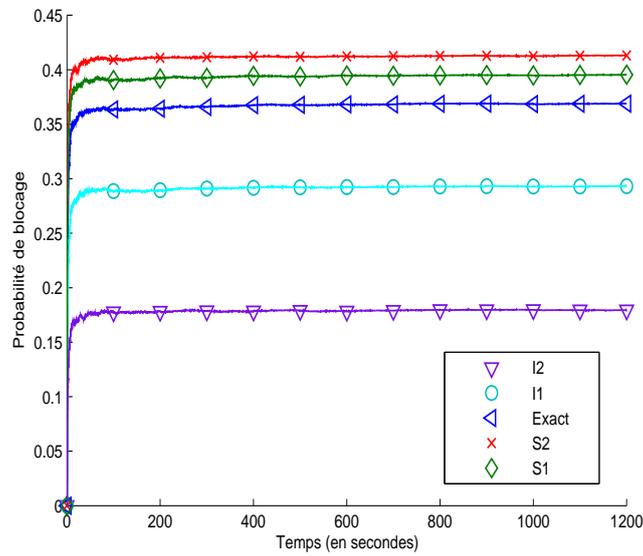


Figure II.5 – Probabilité de blocage de la file 3 pour $c_2 = 15$

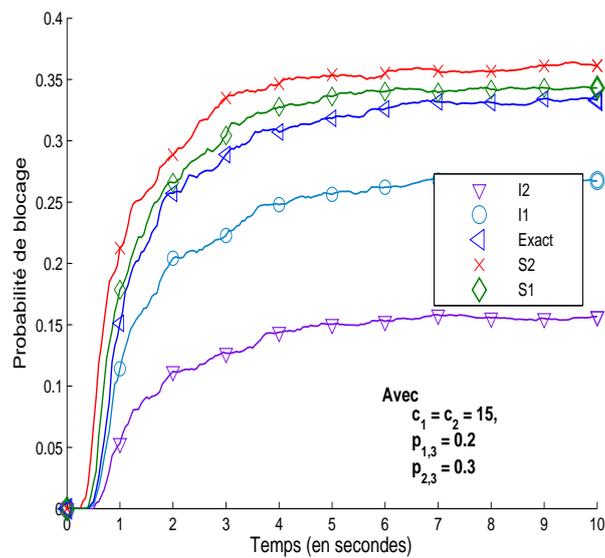


Figure II.6 – Comportement transitoire de la probabilité de blocage de la file 3 pour différents sous-réseaux bornants pour $c_2 = 15$

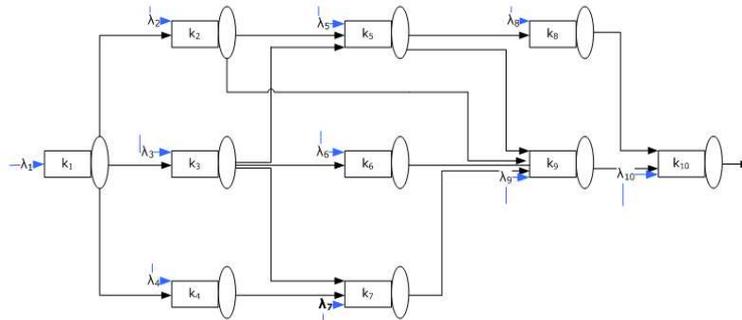


Figure II.7 – Système de 10 files d'attente $M/M/c_i/k_i$

performance recherché. Nous pouvons étudier l'impact de la relation d'inclusion et l'influence des paramètres sur la qualité des bornes.

Dans la Figure.II.8, nous considérons que toutes les files ont le même nombre de serveurs ($c_i = 15, \forall i \in [1, \dots, 10]$). Ensuite, nous augmentons ce nombre dans toutes les files sauf pour la file 9 ($c_i = 25, \forall i \in [1, \dots, 10]$ et $c_9 = 15$) dans le cas de la Figure.II.9.

On peut remarquer, dans la Figure.II.8, que la qualité des bornes supérieures dépend peu du sous-réseau considéré, mais on peut noter que les bornes S_1 et S_2 fournissent des valeurs qui sont, par ailleurs, très proches de la valeur exacte. Dans la Figure.II.9, nous pouvons voir que les bornes supérieures sont aussi très proches, il y a cependant un écart un peu plus important.

Ces résultats sont compatibles avec les courbes précédentes. Lorsque le nombre de serveurs, pour les files d'attente interconnectées à la file 9, est élevé alors, la topologie du sous-réseau a un impact sur la qualité de la borne supérieure de la probabilité de blocage.

Nous savons que lorsque nous coupons un lien, le débit correspondant est ajouté comme un flux supplémentaire à l'entrée de la file d'attente 9. Et comme le débit augmente avec le nombre de serveurs, lorsque ce nombre est élevé la rupture du lien

Tableau II.7 – Description des sous-réseaux bornant

J_1	$\{1, 2, 3, 5, 6, 9\}$
J_2	$\{1, 2, 3, 6, 9\}$
J_3	$\{1, 3, 6, 9\}$
J_4	$\{9\}$

dégrade la qualité de la borne supérieure. Ainsi, la définition du sous-réseau bornant a un impact certain sur la qualité des bornes.

Par contre, lorsque le nombre de serveurs diminue, comme le flux supplémentaire est moins important, la topologie n'est pas très importante pour la précision des bornes.

On constate que pour les bornes inférieures, la topologie a une influence sur la qualité des bornes. Ainsi, I_1 donne la meilleure qualité des bornes car ce système tient compte des transits entre les files. Contrairement à I_4 où tous les liens sont coupés.

Par contre, par rapport à la borne supérieure, on constate que le nombre de serveurs pour les files extérieures aux sous-réseaux a peu d'influence sur la qualité des bornes. Ainsi on remarque, entre les Figures. II.8 et II.9, que la qualité des bornes inférieures est similaires. C'est logique car contrairement à la borne supérieure, le taux d'arrivée n'est pas modifié par les différents sous-réseaux bornants ; il n'y a que les transits entre les files qui est considéré.

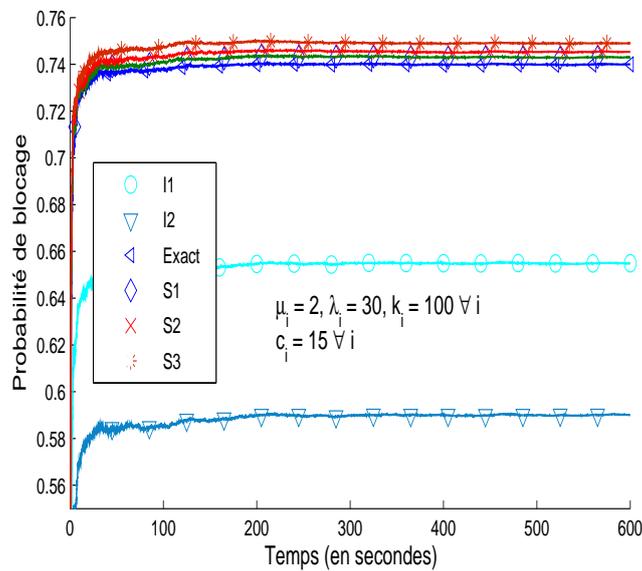


Figure II.8 – Probabilité de blocage de la file 9 pour différents sous-réseaux bornants, $c_i = 15 \forall 1 \leq i \leq 10$.

Dans la suite, nous étudions un autre type de réseau de files d'attente. Il s'agit de réseau de clients positifs et négatifs dit G-Networks ou G-Nets avec catastrophes.

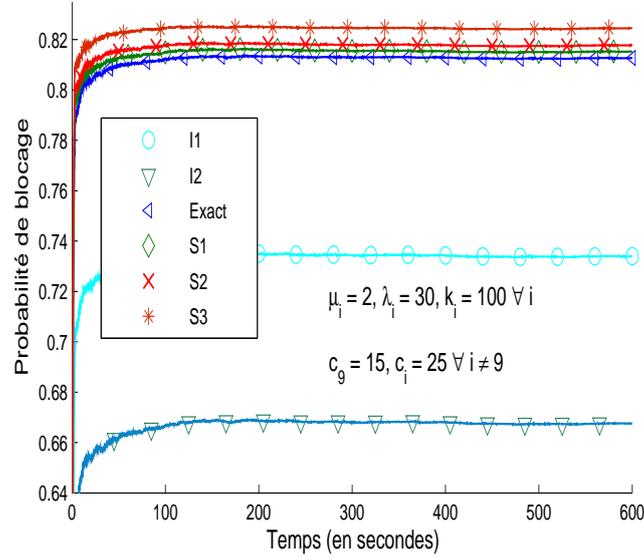


Figure II.9 – Probabilité de blocage de la file 9 pour différents sous-réseaux bornants, $c_i = 25 \forall i \neq 9$ et $c_9 = 15$.

2 Etude d'un réseau G-Networks avec catastrophes

Depuis la série inaugurale de Gelenbe dans le début des années 90 [60–62], les réseaux généralisés de files d'attente ont reçu une attention considérable. Actuellement, il y a plusieurs centaines de références consacrées à ce sujet et un livre [63] donne un aperçu de quelques-unes des orientations de la recherche, du développement et des applications dans le domaine des réseaux de files d'attente avec des clients positifs et négatifs. Les clients positifs sont les clients en attente dans le réseau et les clients négatifs sont souvent appelés "Signaux", et leur rôle est de détruire les clients positifs. Dans les modèles des réseaux de files d'attente (tels que Jackson ou BCMP), il n'existe pas de mécanismes pour que certains clients éliminent d'autres clients, ou les réorientent parmi les files d'attente. En d'autres termes, les clients de réseaux de files d'attente traditionnelles ne peuvent pas exercer un contrôle direct sur les autres clients. Les modèles des réseaux généralisés (G-Networks ou G-Net) permettent de surmonter certaines des limites des modèles classiques de réseaux de files d'attente tout en conservant la propriété de forme produit pour la distribution de probabilité stationnaire. Ils contiennent des clients inhabituels pouvant avoir les comportements suivants :

- les clients négatifs qui éliminent les clients normaux,

- les catastrophes qui vident tous les clients dans une file d'attente [62]
- ou les déclencheurs qui déplacent les clients d'une file d'attente vers une autre [61].

Les réseaux G-Networks ont été introduits pour modéliser les réseaux de neurones dans les échanges des signaux inhibiteurs et excitateurs [64]. Ils sont également utilisés pour modéliser des opérations complexes telles que la suppression des travaux [65] ou des infections virales des logiciels [66].

Comme vu au début de ce chapitre, l'analyse transitoire de ces réseaux de files d'attente est très difficile et la comparaison stochastique offre des solutions intéressantes pour résoudre ce problème en permettant la construction de bornes stationnaires et transitoires des distributions des probabilités. Dans [67, 68], les études de Massey [53] ont été reprises, et donc le formalisme des ensembles croissants est développé en fonction des événements pour définir l'ordre "weak" dans le cas des G-Networks et des G-Networks avec catastrophes. Les systèmes bornants sont des files d'attente $M/M/1$ dont les paramètres sont définis. Cela permet de borner les queues des distributions de probabilités transitoires des G-Networks par le produit des queues des distributions de probabilité transitoires de $M/M/1$. Dans le cas des G-Networks avec catastrophes, on peut borner par des files $M/M/1$ avec retour à 0 dont la distribution transitoire a été calculée.

Ici, nous appliquons la méthode des bornes avec une approche complètement différente de [68], afin de définir des systèmes bornants plus évolués. En effet, dans [68] tous les liens entre les files ont été coupés et les bornes sont définies à partir d'une file $M/M/1$ en isolation. Les bornes ainsi obtenues sont peu précises puisqu'aucune dynamique interne entre les files d'attente n'est conservée. Dans notre travail, nous conservons les transits entre les files d'attente à l'intérieur du sous-réseau. Nous cherchons donc, à généraliser cette approche en bornant le sous-réseau à l'étude par un sous-réseau de même taille mais avec des taux obtenus en utilisant la comparaison stochastique.

Nous appliquons la comparaison stochastique par fonction de projection [43] sur un espace d'états plus petit afin de définir les systèmes bornants. Comme les G-Networks ne sont pas monotones il n'est pas possible d'appliquer l'algorithme présenté dans [58]. Et l'algorithme de Truffet [44] ne peut pas être utilisé car il nécessite un espace totalement ordonné.

Dans cette étude, nous utilisons l'ordre partiel composante par composante sur l'espace d'état afin de comparer file par file le processus.

2.1 Description du G-Networks avec catastrophes

Considérons un réseau G-Networks avec catastrophes avec n files d'attente interconnectées comme illustré dans la Figure.II.10. Chaque file i est définie par les paramètres suivants :

- (a) la file i a un serveur avec un taux de service μ_i
- (b) les arrivées sont de type Poisson avec un taux λ_i^+ pour les clients positifs et un taux λ_i^- pour les clients négatifs
- (c) à la fin d'un service, le client transite de la file i vers la file j comme un client positif avec une probabilité $p_{i,j}^+$, ou comme un client négatif avec une probabilité $p_{i,j}^-$ ou alors quitte le réseau avec une probabilité d_i

A son arrivée dans une file d'attente, si le serveur est occupé alors le client positif attend son service et augmente la taille de la file de +1. Tandis qu'à son arrivée dans une file d'attente, le client négatif va détruire tous les clients positifs qui sont présents dans la file d'attente, s'ils existent, et va ensuite disparaître.

Nous admettons qu'il n'y a pas de rebouclage dans les matrices de routage : c'est-à-dire que $\forall i, p_{i,i}^+ = p_{i,i}^- = 0$. Nous avons la relation suivante pour toute file i :

$$\sum_j p_{i,j}^+ + \sum_j p_{i,j}^- + d_i = 1$$

Soit $\{X(t), t \geq 0\}$, la CMTCC avec le générateur infinitésimal Q représentant l'évolution du G-Net avec catastrophes. A partir d'un état $\vec{x} \in \mathbb{N}^n$, les transitions qui décrivent ce système sont :

- $\vec{x} \rightarrow \vec{x} + \vec{e}_i$ avec le taux λ_i^+ , $1 \leq i \leq n$.
- $\vec{x} \rightarrow \vec{x} - x_i \vec{e}_i$ avec le taux λ_i^- , $x_i > 0, 1 \leq i \leq n$.
- $\vec{x} \rightarrow \vec{x} - \vec{e}_i$ avec le taux $\mu_i d_i$, si $x_i > 0, 1 \leq i \leq n$.
- $\vec{x} \rightarrow \vec{x} - \vec{e}_j + \vec{e}_i$ avec le taux $\mu_j p_{j,i}^+$, si $x_j > 0, 1 \leq i, j \leq n, i \neq j$.
- $\vec{x} \rightarrow \vec{x} - \vec{e}_j - x_i \vec{e}_i$ avec le taux $\mu_j p_{j,i}^-$, si $x_j > 0, x_i > 0, 1 \leq i, j \leq n, i \neq j$.

Nous rappelons que \vec{e}_i (resp. \vec{e}_j) $\in \mathbb{N}^n$ est un vecteur tel que toutes ses composantes sont nulles à l'exception de la composante i (resp. j) qui est égale à 1.

Avec ces hypothèses, il est prouvé dans [62] que les G-Networks ont une forme produit pour les distributions stationnaires.

Le comportement transitoire de ce système étant très difficile à étudier, nous proposons la comparaison stochastique afin de définir des systèmes bornants de sorte qu'il soit possible de calculer la distribution des probabilités transitoires.

2.2 Définition des bornes du G-Networks avec catastrophes

Soit $E = \{1, \dots, n\}$, l'ensemble des indices des files d'attente considéré et J et H les sous-ensembles d'indices disjoints tels que $E = J \cup H$ (voir section 1.2) et la fonction de projection $g : \mathbb{N}^n \rightarrow \mathbb{N}^k$. Décrivons maintenant l'impact d'un événement à la fois sur notre processus original $\{X(t), t \geq 0\}$ mais aussi sur son image par la fonction de projection g :

2.2.1 Projection du processus $X(t)$

- (a) si nous avons une arrivée d'un client positif sur la file i alors, $\vec{x}' = \vec{x} + \vec{e}_i$ et nous avons les cas suivants pour $g(\vec{x}')$:
- i. si la file $i \notin J$ alors, $g(\vec{x}') = g(\vec{x})$.
 - ii. si la file $i \in J$ alors, $g(\vec{x}') = g(\vec{x}) + \vec{v}_i$ (avec $1 \leq i \leq k$). Cette fois, i est dans le sous-réseau alors cet événement impacte $g(X(t))$ car $g(\vec{x} + \vec{e}_i) \neq g(\vec{x})$.
- (b) si nous avons un transit d'un client positif de la file i vers j : si $x_i > 0$ alors, $\vec{x}' = \vec{x} - \vec{e}_i + \vec{e}_j$ et nous avons les cas suivants pour $g(\vec{x}')$:
- i. si $i \notin J$ et $j \notin J$ alors, $g(\vec{x}') = g(\vec{x})$.
 - ii. si $i \notin J$ et $j \in J$ alors, $g(\vec{x}') = g(\vec{x}) + \vec{v}_j$.
 - iii. si $i \in J$ et $j \notin J$ alors, $g(\vec{x}') = g(\vec{x}) - \vec{v}_i$.
 - iv. si $i \in J$ et $j \in J$ alors, $g(\vec{x}') = g(\vec{x}) - \vec{v}_i + \vec{v}_j$.
- (c) si nous avons un service vers l'extérieur, à partir de la file i : si $x_i > 0$ alors, $\vec{x}' = \vec{x} - \vec{e}_i$ et nous avons les cas suivants pour $g(\vec{x}')$:
- i. si $i \notin J$ alors, $g(\vec{x}') = g(\vec{x})$.
 - ii. si $i \in J$ alors, $g(\vec{x}') = g(\vec{x}) - \vec{v}_i$.
- (d) si nous avons une arrivée d'un client négatif sur la file i : si $x_i > 0$ alors, $\vec{x}' = \vec{x} - x_i \vec{e}_i$ et nous avons pour $g(\vec{x}')$:
- i. si $i \notin J$ alors, $g(\vec{x}') = g(\vec{x})$.
 - ii. si $i \in J$ alors, $g(\vec{x}') = g(\vec{x}) - x_i \vec{v}_i$.
- (e) enfin, si nous avons un transit d'un client négatif de la file i vers j : si $x_i > 0$ alors, $\vec{x}' = \vec{x} - \vec{e}_i - x_j \vec{e}_j$ si $x_j > 0$ sinon, $\vec{x}' = \vec{x} - \vec{e}_i$ et nous avons pour $g(\vec{x}')$:
- i. si $i \notin J$ et $j \notin J$ alors, $g(\vec{x}') = g(\vec{x})$

II.2 Etude d'un réseau G-Networks avec catastrophes

ii. si $i \notin J$ et $j \in J$ alors, $g(\vec{x}') = g(\vec{x}) - x_j \vec{v}_j$ si $x_j > 0$.

iii. si $i \in J$ et $j \notin J$ alors, $g(\vec{x}') = g(\vec{x}) - \vec{v}_i$.

iv. si $i \in J$ et $j \in J$ alors, $g(\vec{x}') = g(\vec{x}) - \vec{v}_i - x_j \vec{v}_j$ si $x_j > 0$, sinon, $g(\vec{x}') = g(\vec{x}) - \vec{v}_i$.

Nous en déduisons les évolutions du processus $g(X(t))$: dans les cas (a)-i, (b)-i, (c)-i, (d)-i et (e)-i, les événements n'ont pas d'influence sur le sous-réseau et donc $g(\vec{x}') = g(\vec{x})$.

Pour les autres cas, les taux de transitions sont définis dans la section suivante.

2.2.2 Définition des taux de transition

– si nous avons une arrivée dans la file $i \in J$ (cas (a)-ii) ou un transit de la file $j \notin J$ vers $i \in J$ d'un client positif (cas (b)-ii) alors, $g(\vec{x}') = g(\vec{x}) + \vec{v}_i$. Et le nouveau taux de transition est la somme des cas (a)-ii et (b)-ii :

$$\sum_{\vec{x}' | g(\vec{x}') = g(\vec{x}) + \vec{v}_i} Q(\vec{x}, \vec{x}') = \lambda_i^+ + \sum_{j \notin J} \mu_j p_{j,i}^+ 1_{x_j > 0} \quad (\text{II.11})$$

– si nous avons un transit d'un client positif de la file $i \in J$ vers la file $j \in J$ (cas (b)-iv) alors, $g(\vec{x}') = g(\vec{x}) - \vec{v}_i + \vec{v}_j$. Et le taux de transition est :

$$\sum_{\vec{x}' | g(\vec{x}') = g(\vec{x}) - \vec{v}_i + \vec{v}_j} Q(\vec{x}, \vec{x}') = \mu_i p_{i,j}^+ \quad (\text{II.12})$$

– si nous avons un départ d'un client positif à partir de $i \in J$ vers l'extérieur (cas (c)-ii) ou un transit d'un client positif ou négatif de la file $i \in J$ vers $j \notin J$ (cas (b)-iii et (e)-iii) alors, $g(\vec{x}') = g(\vec{x}) - \vec{v}_i$. Le taux de transition correspondant est :

$$\begin{aligned} \sum_{\vec{x}' | g(\vec{x}') = g(\vec{x}) - \vec{v}_i} Q(\vec{x}, \vec{x}') &= \mu_i d_i + \sum_{j \notin J} \mu_i p_{i,j}^+ \\ &+ \sum_{j \notin J} \mu_i p_{i,j}^- \end{aligned} \quad (\text{II.13})$$

– si nous avons une arrivée d'un client négatif de l'extérieur dans la file $i \in J$ (cas (d)-ii) ou un transit de la file $j \notin J$ vers i (cas (e)-ii) alors, nous avons

$g(\vec{x}') = g(\vec{x}) - x_i \vec{v}_i$. Et le taux de transition est :

$$\sum_{\vec{x}' | g(\vec{x}') = g(\vec{x}) - x_i \vec{v}_i} Q(\vec{x}, \vec{x}') = \lambda_i^- + \sum_{j \notin J} \mu_j p_{j,i}^- 1_{x_j > 0} \quad (\text{II.14})$$

– si nous avons un transit d'un client négatif de la file $j \in J$ vers $i \in J$ (cas (e)-iv) alors, $g(\vec{x}') = g(\vec{x}) - \vec{v}_j - x_i \vec{v}_i$ si $x_i \geq 0$. Le taux de transition est dans ce cas :

$$\sum_{\vec{x}' | g(\vec{x}') = g(\vec{x}) - \vec{v}_j - x_i \vec{v}_i} Q(\vec{x}, \vec{x}') = \mu_j p_{j,i}^- \quad (\text{II.15})$$

Nous allons maintenant définir les systèmes bornants en utilisant la fonction de projection g sur le processus $X(t)$. Nous utilisons l'ordre partiel noté \preceq sur l'espace d'états \mathbb{N}^n :

$$\forall (\vec{x}, \vec{y}) \in \mathbb{N}^n, \vec{x} \preceq \vec{y} \Leftrightarrow x_i \leq y_i, \forall i = 1, \dots, n$$

Nous appliquons le couplage par fonction de projection et le théorème .13 afin de définir le processus de $\{X^S(t), t \geq 0\}$ comme étant une borne supérieure pour le sous-réseau J :

$$\{g(X(t)), t \geq 0\} \preceq_{st} \{X^S(t), t \geq 0\} \quad (\text{II.16})$$

ainsi que le processus $\{X^I(t), t \geq 0\}$ définie comme étant une borne inférieure :

$$\{X^I(t), t \geq 0\} \preceq_{st} \{g(X(t)), t \geq 0\} \quad (\text{II.17})$$

Nous allons par la suite définir les taux de transitions pour la construction de la borne supérieure.

2.3 Construction de la borne supérieure

Soient $\widehat{X}(t)$ et $\widehat{X}^S(t)$, deux processus markoviens avec les générateurs infinitésimaux Q et Q^S . Nous définissons le processus couplé $Z(t) = (\widehat{X}(t), \widehat{X}^S(t))$ associé à K et telles que les réalisations restent dans :

$$K = \{(\vec{x}, \vec{y}) \in \mathbb{N}^n \times \mathbb{N}^k \mid g(\vec{x}) \preceq \vec{y}\}.$$

Examinons maintenant, en fonction des événements, comment un saut d'une composante doit être compensé par le saut de l'autre composante, afin d'obtenir les

II.2 Etude d'un réseau G-Networks avec catastrophes

conditions sur les taux de transition. On notera qu'une arrivée occasionne un saut sur la composante \vec{x} . Le processus couplé peut alors sortir de K et ce saut devra donc être compensé par un saut équivalent sur \vec{y} . Les sauts sur \vec{y} , qui font que le processus couplé quitte K , sont un service ou l'arrivée d'un signal dans une file d'attente. Il faudra aussi les compenser par des sauts similaires de la composante \vec{x} .

- (a) si nous avons une arrivée dans la file $i \in J$ d'un client positif qui génère une transition de l'état \vec{x} vers \vec{x}' telle que $g(\vec{x}') = g(\vec{x}) + \vec{v}_i$ alors, nous devons avoir une transition de \vec{y} vers \vec{y}' telle que $\vec{y}' = \vec{y} + \vec{v}_i$ pour avoir $g(\vec{x}') \preceq \vec{y}'$. Ainsi les taux de transition doivent satisfaire la condition suivante :

$$Q^S(\vec{y}, \vec{y} + \vec{v}_i) \geq \sum_{\vec{x}' | g(\vec{x}') = g(\vec{x}) + \vec{v}_i} Q(\vec{x}, \vec{x}') \quad (\text{II.18})$$

Puisque l'inégalité (II.18) doit être satisfaite pour tous les états \vec{x} tels que $(\vec{x}, \vec{y}) \in K$, alors nous prenons le maximum de la partie droite pour $\vec{x} \in N^n$:

$$Q^S(\vec{y}, \vec{y} + \vec{v}_i) = \max_{\vec{x} | g(\vec{x}) \preceq \vec{y}} \left\{ \sum_{\vec{x}' | g(\vec{x}') = g(\vec{x}) + \vec{v}_i} Q(\vec{x}, \vec{x}') \right\}$$

L'évolution de $\{g(X(t)), t \geq 0\}$, à cause de cet événement, est donnée dans l'équation (II.11). Et le maximum pour ce taux est atteint pour les états \vec{x} tels que $x_j > 0 \forall j \notin J$. Nous choisissons donc :

$$Q^S(\vec{y}, \vec{y} + \vec{v}_i) = \lambda_i^+ + \sum_{j \notin J} \mu_j p_{j,i}^+$$

- (b) un transit de la file i vers la file j d'un client positif peut générer un saut en dehors de K pour la première ou la seconde composante.

- i. pour la première composante, si nous avons une transition entre \vec{x} et \vec{x}' telle que $g(\vec{x}') = g(\vec{x}) - \vec{v}_i + \vec{v}_j$ alors, nous devons avoir une transition de \vec{y} vers \vec{y}' telle que nous avons :

$$\vec{y}' = \vec{y} - \vec{v}_i + \vec{v}_j \text{ ou } \vec{y}' = \vec{y} + \vec{v}_j.$$

- ii. pour la seconde composante, si nous avons une transition de \vec{y} vers \vec{y}' telle que $\vec{y}' = \vec{y} - \vec{v}_i + \vec{v}_j$ alors, nous devons avoir une transition de \vec{x} vers \vec{x}' telle

que :

$$g(\vec{x}') = g(\vec{x}) - \vec{v}_i + \vec{v}_j \text{ ou } g(\vec{x}') = g(\vec{x}) - \vec{v}_i.$$

Comme les sauts pour les services et les arrivées sont utilisés pour compenser ces événements, nous ne prenons que la transition entre les files d'attente i et j . Il en résulte le même taux de transition, donc à partir de l'équation (II.12) :

$$Q^S(\vec{y}, \vec{y} - \vec{v}_i + \vec{v}_j) = \mu_i p_{i,j}^+.$$

- (c) si nous avons un transit d'un client négatif de la file i vers la file j qui déclenche une transition de \vec{y} vers \vec{y}' telle que $\vec{y}' = \vec{y} - \vec{v}_i - y_j \vec{v}_j$ alors, nous devons avoir une transition de \vec{x} vers \vec{x}' telle que $g(\vec{x}') = g(\vec{x}) - \vec{v}_i - x_j \vec{v}_j$ si $x_i > 0$. Comme cet événement ne se produit dans le système exact que si $x_i > 0$, alors nous prenons le taux minimum :

$$Q^S(\vec{y}, \vec{y} - \vec{v}_i - y_j \vec{v}_j) = 0.$$

- (d) si nous avons un service dans la file $i \in J$ pour lequel nous avons une transition de \vec{y} vers \vec{y}' telle que $\vec{y}' = \vec{y} - \vec{v}_i$ alors, nous devons avoir une transition de \vec{x} vers \vec{x}' telle que $g(\vec{x}') = g(\vec{x}) - \vec{v}_i$, ainsi nous avons l'inégalité suivante :

$$Q^S(\vec{y}, \vec{y} - \vec{v}_i) \leq \sum_{\vec{x}' | g(\vec{x}') = g(\vec{x}) - \vec{v}_i} Q(\vec{x}, \vec{x}')$$

Comme elle doit être vérifiée pour tous les $(\vec{x}, \vec{y}) \in K$ alors, nous prenons :

$$Q^S(\vec{y}, \vec{y} - \vec{v}_i) = \min_{x | g(\vec{x}) \leq \vec{y}} \left\{ \sum_{\vec{x}' | g(\vec{x}') = g(\vec{x}) - \vec{v}_i} Q(\vec{x}, \vec{x}') \right\}$$

et à partir de l'équation (II.13), nous obtenons :

$$Q^S(\vec{y}, \vec{y} - \vec{v}_i) = \mu_i d_i + \sum_{j \notin J} \mu_i p_{i,j}^+ + \sum_{j \notin J} \mu_i p_{i,j}^-.$$

Comme dans le cas précédent, supposons que $Q^S(\vec{y}, \vec{y} - \vec{v}_i - y_j \vec{v}_j) = 0$ ($i, j \in J$). Si nous considérons aussi que, pour la borne supérieure, le taux $\mu_i p_{i,j}^-$ n'a pas d'impact sur la composante j et contribue seulement à diminuer la composante

II.2 Etude d'un réseau G-Networks avec catastrophes

i , nous obtenons alors :

$$Q^S(\vec{y}, \vec{y} - \vec{v}_i) = \mu_i d_i + \sum_{j \notin J} \mu_i p_{i,j}^+ + \sum_{j \notin J} \mu_i p_{i,j}^- + \sum_{j \in J} \mu_i p_{i,j}^-.$$

- (e) si nous avons une arrivée, de l'extérieur, d'un client négatif sur la file $i \in J$ qui génère une transition de \vec{y} vers \vec{y}' telle que $\vec{y}' = \vec{y} - y_i \vec{v}_i$ alors, nous devons avoir une transition de \vec{x} vers \vec{x}' telle que $g(\vec{x}') = g(\vec{x}) - x_i \vec{v}_i$ si $x_j = 0$ ou une transition de \vec{x} vers \vec{x}' telle que $g(\vec{x}') = g(\vec{x}) - \vec{v}_j - x_i \vec{v}_i$ (pour $j \neq i$ et $j \in J$) si $x_j > 0$. Dans les deux cas, nous prenons les taux minimaux obtenus lorsque $x_j = 0 \forall j \neq i$. Donc à partir de l'équation (II.14), nous obtenons :

$$Q^S(\vec{y}, \vec{y} - \vec{v}_i) = \lambda_i^-$$

Le Tableau II.8 résume les taux de transition de la borne supérieure pour la file $i \in J$.

Les taux de $\{X^S(t), t \geq 0\}$ sont définis de sorte que le processus couplé reste dans l'ensemble K et donc, l'équation II.16 est vérifiée.

Pour la borne supérieure, les taux de transition $\mu_j p_{j,i}^+$, des clients positifs en provenance du reste du réseau, sont ajoutés aux taux d'arrivée externes λ_i^+ de clients positifs. Le taux de départ correspond au taux de départ du sous-réseau. Pour les clients négatifs, seuls les taux d'arrivée externes sont considérés. À l'intérieur du sous-réseau, les transitions des clients positifs sont maintenues mais pas celles des clients négatifs qui sont prises en compte dans le taux de départ uniquement.

Intuitivement, une file d'attente de la borne supérieure a plus de clients positifs et moins de clients négatifs car elle subit moins de remises à zéro des files que dans le système exact.

Tableau II.8 – Taux de transition des files du sous-réseau J (borne supérieure)

Taux	Expression
arrivée client positif	$\lambda_i^+ + \sum_{j \notin J} \mu_j p_{j,i}^+$
arrivée client négatif	λ_i^-
service	$\mu_i d_i + \sum_{j \notin J} \mu_i p_{i,j}^+ + \sum_{j \neq i} \mu_i p_{i,j}^-$
transit vers $j \in J$	$\mu_i p_{i,j}^+$

On peut remarquer aussi que la borne supérieure est la plus précise que nous pourrions obtenir des contraintes d'inégalités. En outre, comme les transitions se produisent quels que soient les états, alors la borne est monotone (voir le théorème .13).

Dans la suite, nous allons définir les taux des transitions de la borne inférieure.

2.4 Construction de la borne inférieure

Nous cherchons à définir le processus bornant $\{X^I(t), t \geq 0\}$ de sorte que, le processus couplé $Z(t) = ((X^I(t), X(t)))$ puisse avoir ses réalisations dans

$$K = \{(\vec{x}, \vec{y}) \in \mathbb{N}^k \times \mathbb{N}^n \mid \vec{x} \preceq g(\vec{y})\}$$

- (a) Si nous avons une arrivée dans la file $i \in J$ d'un client positif qui génère une transition de \vec{x} vers \vec{x}' telle que $\vec{x}' = \vec{x} + \vec{v}_i$ alors, nous devons avoir aussi une transition de \vec{y} vers \vec{y}' telle que $g(\vec{y}') = g(\vec{y}) + \vec{v}_i$. On obtient donc la condition suivante pour le taux de transition :

$$Q^I(\vec{x}, \vec{x} + \vec{v}_i) \leq \sum_{\vec{y}' | g(\vec{y}') = g(\vec{y}) + \vec{v}_i} Q(\vec{y}, \vec{y}')$$

En considérons le taux minimum dans l'équation (II.11) ($x_j = 0, \forall j \notin J$), nous prenons :

$$Q^I(\vec{x}, \vec{x} + \vec{v}_i) = \lambda_i^+.$$

- (b) un transit de la file i vers la file j d'un client positif peut générer un saut en dehors de K pour la première ou la seconde composante. Ce cas est identique au cas (b) de la borne supérieure. Nous prenons la même valeur pour le taux de transition :

$$Q^I(\vec{x}, \vec{x} - \vec{v}_i + \vec{v}_j) = \mu_i p_{i,j}^+.$$

- (c) Si nous avons un service dans la file $i \in J$ sur la composante \vec{y} alors, nous devons aussi avoir une diminution du nombre de clients dans la file i sur \vec{x} soit par un service dans la même file i de \vec{x} soit par un transit d'un client positif ou négatif de la file i vers tout autre file $j \notin J$. Donc, à partir de l'équation (II.13), nous avons :

$$Q^I(\vec{x}, \vec{x} - \vec{v}_i) = \mu_i d_i + \sum_{j \notin J} \mu_i p_{i,j}^+ + \sum_{j \notin J} \mu_i p_{i,j}^-.$$

II.2 Etude d'un réseau G-Networks avec catastrophes

- (d) Si nous avons un transit de la file i vers la file j à l'intérieur du sous-réseau J , d'un client négatif qui provoque une transition de \vec{y} vers \vec{y}' telle que $g(\vec{y}') = g(\vec{y}) - \vec{v}_i - y_j \vec{v}_j$ alors, nous pouvons avoir une transition de \vec{x} vers \vec{x}' telle que $\vec{x}' = \vec{x} - \vec{v}_i - x_j \vec{v}_j$ si $x_i > 0$ ou $\vec{x}' = \vec{x} - x_j \vec{v}_j$ si $x_i = 0$.

si $x_i > 0$, nous avons :

$$\sum_{\vec{y}' | g(\vec{y}') = g(\vec{y}) - \vec{v}_i - y_j \vec{v}_j} Q(\vec{y}, \vec{y}') \leq Q^I(\vec{x}, \vec{x} - \vec{v}_i - x_j \vec{v}_j)$$

en remplaçant la première partie de l'inégalité par le taux donné dans l'équation (II.15), alors nous obtenons, si $x_i > 0$:

$$Q^I(\vec{x}, \vec{x} - \vec{v}_i - x_j \vec{v}_j) = \mu_i p_{i,j}^-$$

si $x_i = 0$, alors nous obtenons :

$$\sum_{\vec{y}' | g(\vec{y}') = g(\vec{y}) - \vec{v}_i - y_j \vec{v}_j} Q(\vec{y}, \vec{y}') \leq Q^I(\vec{x}, \vec{x} - x_j \vec{v}_j) \quad (\text{II.19})$$

ce qui revient à écrire :

$$Q^I(\vec{x}, \vec{x} - x_j \vec{v}_j) \geq \mu_i p_{i,j}^- \quad (\text{II.20})$$

Et donc on obtient :

$$Q^I(\vec{x}, \vec{x} - x_j \vec{v}_j) \geq \sum_{i \in I} \mu_i p_{i,j}^- 1_{x_i=0} \quad (\text{II.21})$$

ici, seul le 1er cas nous intéresse ($x_i > 0$), où l'on obtient le résultat suivant :

$$Q^I(\vec{x}, \vec{x} - \vec{v}_i - x_j \vec{v}_j) = \mu_i p_{i,j}^-$$

Le cas où $x_i = 0$ est utilisé par la suite.

- (e) Si nous avons une arrivée d'un client négatif dans la file $i \in J$ qui génère une transition de \vec{y} vers \vec{y}' telle que $g(\vec{y}') = g(\vec{y}) - y_i \vec{v}_i$, nous pouvons avoir une transition de \vec{x} vers \vec{x}' telle que :

i. $\vec{x}' = \vec{x} - x_i \vec{v}_i$

Chapitre II. Définition de sous-réseaux bornants dans les réseaux de files d'attente

- ii. ou bien $\vec{x}' = \vec{x} - x_j - x_i \vec{v}_i \forall j \in J$ si $x_j > 0$. Cette transition est déjà utilisée pour compenser la transition de \vec{y} vers \vec{y}' telle que $g(\vec{y}') = g(\vec{y}) - \vec{v}_i - y_j \vec{v}_j$ (dans le cas précédent).

A partir de l'équation (II.14), nous prenons la valeur maximale en considérant $x_j > 0 \forall j \neq J$.

En considérant en plus l'inégalité donnée par II.19, on obtient ainsi le taux externe des clients négatifs de la file d'attente i :

$$Q^I(\vec{x}, \vec{x} - x_i \vec{v}_i) = \lambda_i^- + \sum_{j \notin J} \mu_j p_{j,i}^- + \sum_{j \in J, j \neq i} \mu_j p_{j,i}^- 1_{x_j=0}.$$

Ainsi pour la file $i \in J$, les taux de transition de la borne inférieure sont résumés dans le Tableau II.9.

Comme les taux de $\{X^I(t), t \geq 0\}$ sont définis de telle sorte que le processus couplé reste dans l'ensemble K , alors l'équation II.17 est vérifiée.

Pour la borne inférieure, on remarque que les taux de transition $\mu_j p_{j,i}^-$, des flux de clients négatifs en provenance du reste du réseau, sont ajoutés aux flux externes de clients négatifs. Le taux de départ est celui de départ vers l'extérieur du sous-réseau. Pour les arrivées de clients positifs, seuls les taux d'arrivée des clients externes positifs sont pris en compte. À l'intérieur du sous-réseau, les transitions de clients positifs et négatifs sont maintenues.

En outre, le taux $\mu_i p_{i,j}^-$, des clients négatifs entre les files d'attente, est pris en compte dans la borne inférieure $\forall x_i$, si $x_i > 0$. Et si $x_i = 0$, ce taux est ajouté au taux d'arrivée de signal externe. Donc, avec ce taux une file d'attente j est toujours

Tableau II.9 – Résumé des taux de transition de la borne inférieure

Taux	Expression
arrivée client positif	λ_i^+ .
arrivée client négatif service	$\lambda_i^- + \sum_{j \notin J} \mu_j p_{j,i}^- + \sum_{j \in J, j \neq i} \mu_j p_{j,i}^- 1_{x_j=0}$.
transit d'un client positif vers $j \in J$	$\mu_i d_i + \sum_{j \notin J} \mu_i p_{i,j}^+ + \sum_{j \in J} \mu_i p_{i,j}^-$.
transit d'un client négatif vers j	$\mu_i p_{i,j}^-$.

vidée quel que soit l'état des autres files d'attente.

On peut remarquer également que ce système est la borne inférieure la plus précise pour le sous-réseau. Et comme la transition se produit quel que soit l'état, alors la borne inférieure est monotone (voir le théorème .13).

Intuitivement, pour la borne inférieure, chaque file d'attente a moins de clients positifs car elle subit plus d'arrivées de signaux. que dans le processus exact.

2.5 Comparaison des mesures transitoires

Dans cette partie, nous identifions les mesures qui pourraient être comparées entre le G-Network avec catastrophes et les systèmes bornants. Vu l'ordre à définir sur l'espace d'états et la comparaison stochastique des processus, de toute évidence, nous avons la comparaison du nombre moyen de clients dans la file d'attente $i \in J$ à tout instant t .

Les G-Networks sont des modèles très efficaces pour étudier les infections virales des logiciels [66]. Dans ces réseaux, les remises à zéros (reset) des files d'attente peuvent représenter les infections par des programmes malveillants qui causent la suppression d'informations (représentées par les clients positifs) dans les stations. Certaines mesures de performance intéressantes peuvent être calculées en utilisant les méthodes de bornes stochastiques. Soit $N_i(t)$, le nombre de fois qu'une station i du sous-réseau J a été infectée dans l'intervalle de temps $[0, T]$. Nous désignons par $N_i^S(t)$ (resp. $N_i^I(t)$) la mesure correspondante pour $\{X^S(t), t \geq 0\}$ (resp. $\{X^I(t), t \geq 0\}$). Nous avons alors la comparaison suivante :

Proposition .20. $\forall t \geq 0, \forall i \in J :$

$$N_i(t) \leq_{st} N_i^I(t), \text{ et } N_i^S(t) \leq_{st} N_i(t)$$

Preuve. Nous utilisons le couplage des processus avec l'événement de type signal car il est le seul événement qui modifie le comportement de ces processus. On définit le processus $\{\widehat{N}_i(t), t \geq 0\}$ (resp. $\{\widehat{N}_i^I(t), t \geq 0\}$) avec les mêmes taux de transitions que $\{N_i(t), t \geq 0\}$ (resp. $\{N_i^I(t), t \geq 0\}$), avec la condition initiale suivante : $\widehat{N}_i(0)[\omega] \leq \widehat{N}_i^I(0)[\omega]$. Pour la preuve, nous supposons que :

$$\widehat{N}_i(t)[\omega] \leq \widehat{N}_i^I(t)[\omega] \tag{II.22}$$

apportons maintenant la preuve que l'inégalité reste vraie à $t + dt$.

Nous avons montré que les processus $\{g(X(t)), t \geq 0\}$ et $\{X^I(t), t \geq 0\}$ sont \preceq_{st} -comparable. Nous avons plus d'occurrences des signaux (vidages) sur le processus $\{X^I(t), t \geq 0\}$ que dans le réseau correspondant au système exact. Donc, si $\widehat{N}_i(t)$ augmente alors, $\widehat{N}_i^I(t)$ augmente aussi et l'équation (II.22) est vérifiée au temps $t + dt$, donc : $N_i(t) \leq_{st} N_i^I(t)$. De même, nous pouvons prouver que $N_i^S(t) \leq_{st} N_i(t)$. \square

Une autre mesure intéressante, qui peut être calculée, est le temps pour qu'une première infection survienne dans la file d'attente i du sous-réseau J . Si Tf_i désigne le temps de la première infection, on en déduit l'inégalité suivante :

Proposition .21.

$$\forall i \in J, Tf_i \leq_{st} Tf_i^S \text{ et } Tf_i^I \leq_{st} Tf_i$$

Preuve. Soit \widehat{Tf}_i (resp \widehat{Tf}_i^S) une variable aléatoire avec la même distribution de probabilité que Tf_i (resp. Tf_i^S). Nous supposons le cas où aucune infection n'a encore eu lieu dans l'intervalle de temps $[0, t]$ pour les deux systèmes. Il est clair que lorsqu'une infection se produit dans la file i de $X^S(t)$, alors elle a déjà eu lieu dans $X(t)$. Donc, nous en déduisons que : $\widehat{Tf}_i[\omega] \preceq \widehat{Tf}_i^S[\omega], \forall \omega \in \Omega$. De même, nous pouvons prouver que $Tf_i^I \leq_{st} Tf_i$, et la proposition .21 est prouvée. \square

2.6 Analyse numérique des indices de performances transitoires

Nous présentons quelques résultats des mesures de performances transitoires présentées précédemment. Nous étudions un G-Net, avec 7 files d'attente, représenté dans la figure II.10. Nous nous concentrons sur la file d'attente 7 pour le calcul des mesures de performances transitoires.

Nous considérons les trois sous-systèmes bornants $J_3 = \{7\}$, $J_2 = \{5, 7\}$ et $J_1 = \{5, 6, 7\}$ (avec $J_3 \subset J_2 \subset J_1$) pour étudier la précision des bornes face à la simplification du système d'origine. Les résultats sont obtenus avec l'outil Matlab/Simevent en considérant un intervalle de confiance de 95% et un écart type de 0,01.

Pour la simulation, nous considérons les valeurs suivantes des paramètres :

II.2 Etude d'un réseau G-Networks avec catastrophes

$$\left[\begin{array}{l} \lambda_7^+ = 3 \text{ et } d_7 = 0.2 \\ \lambda_i^+ = 2 \text{ et } d_i = 0.2 \forall i \in [1, \dots, 6] \\ \mu_i = 2 \text{ et } \lambda_i^- = \frac{1}{25} \forall i \in [1, \dots, 7] \\ P_{1,3}^+ = P_{1,4}^+ = 0.3 \text{ et } P_{1,3}^- = P_{1,4}^- = 0.1 \\ P_{2,4}^+ = 0.6 \text{ et } P_{2,4}^- = 0.2 \\ P_{3,5}^+ = P_{3,6}^+ = 0.3 \text{ et } P_{3,5}^- = P_{3,6}^- = 0.1 \\ P_{4,6}^+ = 0.4 \text{ et } P_{4,6}^- = 0.4 \\ P_{5,7}^+ = 0.6 \text{ et } P_{5,7}^- = 0.2 \\ P_{6,7}^+ = 0.3 \text{ et } P_{6,7}^- = 0.5 \end{array} \right]$$

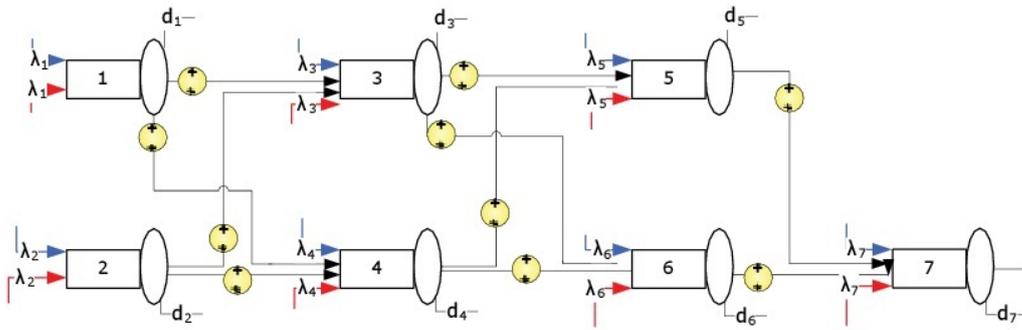


Figure II.10 – Le système G-Network étudié

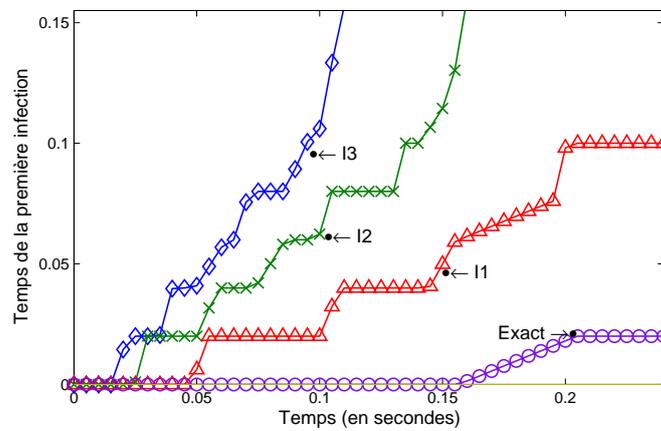


Figure II.11 – Temps T_{f7} avant la première infection dans la file 7, selon les différents sous-réseaux.

Pour les sous-réseaux J_i , pour tout $1 \leq i \leq 3$, nous désignons par

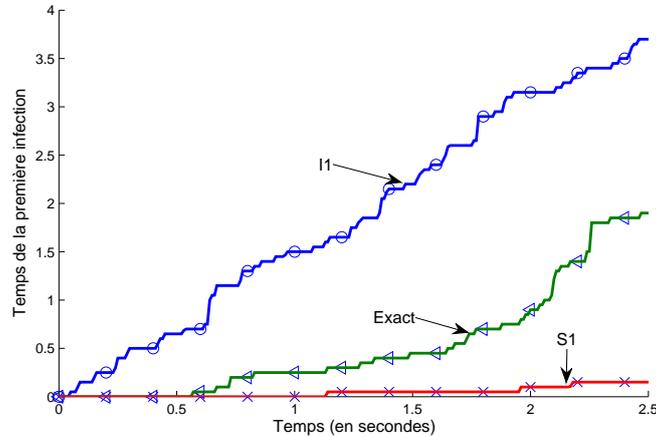


Figure II.12 – Encadrement du temps de la première infection de la file 7.

- $I1$, $I2$ et $I3$, les bornes inférieures
- et $S1$, $S2$ et $S3$, les bornes supérieures.

Dans la suite, nous donnons les résultats obtenus pour le système exact avec 7 files et les différents sous-réseaux J_1 , J_2 et J_3 . Et les résultats obtenus confirment la pertinence de la proposition des sous-réseaux bornants pour l'encadrement du réseau exact.

(a) **Temps de la première arrivée d'une panne Tf_7**

Nous étudions le temps nécessaire pour qu'une première panne survienne Tf_7 au niveau de la file 7 dans l'intervalle de temps $[0, t]$. Dans la Figure.II.11, nous constatons l'influence de la définition des sous-réseaux bornants. Et c'est le sous-réseau J_1 qui fournit la meilleure borne inférieure ($I1$) tandis que $I3$ est la moins efficace des bornes inférieures.

Ce résultat est normal car pour le sous-réseau bornant J_3 , nous avons pris en compte les taux de transit des clients négatifs des files 5 et 6 qui n'appartiennent pas à J_3 . Par contre, pour J_1 nous conservons la dynamique des transitions avec la file 7 comme pour le système exact.

Nous remarquons également que sur l'intervalle de temps $[0, 0.2]$, nous n'observons toujours pas de déclenchement pour les bornes supérieures. Cela s'explique par le fait que, dans la définition des taux de transition des clients négatifs, les taux des files connectées à la file 7 ne sont pas pris en considération.

La Figure.II.12 met en évidence la pertinence de l'approche en proposant un encadrement de Tf_7 comme mesure de performance du système exact où on

prend en considération l'ensemble des nœuds.

(b) *Influence des sous-réseaux sur le nombre de clients en attente dans la file 7*

Dans la Figure II.13, nous représentons le nombre de clients dans la file 7 dans le temps.

Comme on pouvait s'attendre, moins la file est soumise à l'arrivée de clients négatifs et plus sa "population" de clients positifs augmente. Ainsi, la borne S_3 est la plus mauvaise puisqu'elle n'est constituée que de la file 7 et donc qu'elle reçoit peu de clients négatifs qui détruisent les clients positifs. D'une manière générale, les comportements des bornes supérieures sont éloignés du système exact, car nous avons très peu pris en compte les clients négatifs dans les systèmes bornants. Cela est dû au fait que la comparaison stochastique engendre des contraintes un peu trop fortes dans certains cas.

Par contre, les bornes inférieures sont proches du système exact car nous avons conservé les arrivées des clients négatifs.

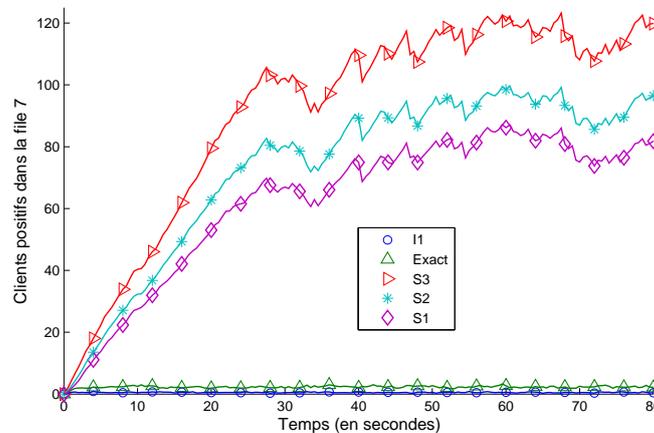


Figure II.13 – Nombre de clients positifs dans la file 7 pour $\mu_5 = \mu_6 = 2$ et $t \in [0, 10]$.

(c) *Influence de la complexité sur le temps de simulation*

Nous cherchons à mettre en évidence l'impact de la taille du système, donc sa complexité, sur le temps de simulation.

Les systèmes que nous simulons ici sont relativement petits par rapport à des systèmes avec plusieurs milliers de nœuds. Toutefois, nous pouvons remarquer dans le Tableau II.10 que le temps de simulation est 7 fois plus important pour le modèle exact que pour la borne supérieure S_3 . Nous constatons donc, la perti-

nence de notre approche et la nécessité de décider, dans certaines circonstances, d'un compromis entre le temps de simulation et la qualité des bornes donc de la complexité du système étudié.

3 Conclusion

Dans ce chapitre, nous avons vu que l'application de la comparaison stochastique pour l'analyse transitoire des grands réseaux de files d'attente fournit une solution intéressante car on obtient un compromis entre la qualité des bornes, et les temps de simulations qui dépendent de la complexité de calcul.

Les bornes sont construites dans le sens de l'ordre \preceq_{st} par la méthode du couplage par fonction de projection de l'espace d'états du réseau avec n files d'attente vers l'espace d'états du sous-réseau de k files d'attente (où $k < n$).

Nous donnons des études de cas pour montrer la précision des bornes en fonction des paramètres du modèle. De toute évidence, avec un sous-réseau bornant de plus grande taille il est logique que l'on obtienne des résultats plus précis. Mais cette précision à un coût notamment en temps de simulation ou en capacité de stockage. La comparaison stochastique est une solution intéressante pour étudier les grands réseaux de files d'attente en fournissant des bornes sur des mesures de performances stationnaires et transitoires. Nous avons pu voir par ailleurs, que la qualité de ces bornes dépendaient aussi de la taille des sous-réseaux mais aussi des paramètres du système. Nous aurons l'occasion d'y revenir dans le chapitre IV.

Dans le chapitre suivant, nous étudions l'évaluation de la performabilité sur le modèle d'Erlang avec perte proposé par Trivedi dans [47].

Sous-système bornant	Temps de simulation
Exact	363 s
S_1	149 s
S_2	100 s
S_3	53 s

Tableau II.10 – Temps de simulation (en seconds) pour les systèmes exacts et bornants

Chapitre III

Evaluation de la performabilité d'un système de télécommunication

Dans les domaines de l'informatique (Cloud computing, systèmes distribués) basés sur le WEB et des télécommunications (réseaux optiques, mobiles), les systèmes deviennent physiquement et logiquement plus complexes et la performance de ces systèmes est souvent dégradée à cause des défauts internes et externes qui réduisent la qualité du service rendu. La prise en compte de cette condition de dégradation leur confère une certaine capacité à fonctionner sur plusieurs niveaux de performances qui varient depuis le fonctionnement d'un état parfait (zéro défaut) jusqu'à l'indisponibilité (hors service). Ces systèmes sont définis dans la littérature comme des systèmes à tolérance aux fautes ou Multi-states systems (MSS) [2, 3]. Dans [4], on s'intéresse à la modélisation de ces systèmes, et une approche hiérarchique est proposée afin de prendre en compte à la fois l'aspect performance et fiabilité. Il est donc nécessaire dans l'évaluation des performances, d'évaluer l'impact de la fiabilité sur les performances des systèmes car les modèles générés sont souvent de grandes tailles.

Ainsi, l'évaluation de la performabilité induit une complexité dans la mesure où la chaîne de Markov représentant le système possède un espace d'états très grand en raison du nombre d'événements (occupation/libération et panne/réparation) à prendre en compte. Ce qui rend difficile voire impossible une résolution directe même par des méthodes numériques à cause de la taille de l'espace d'états. Il existe des méthodes approximatives telles que celle de Courtois [33] qui permettent de décomposer le système global en sous-systèmes représentant des macros-états. Cela permet d'obtenir un système agrégé à partir duquel le calcul de la distribution de probabilité stationnaire est plus facile à analyser. Nous proposons de reprendre l'étude menée

par Trivedi [47] qui consiste à calculer la probabilité de blocage dans un système de télécommunications soumis à des pannes. Le système étant multidimensionnel et donc difficile à évaluer, son approche consiste à décomposer le système initial en deux systèmes indépendants : un premier pour représenter l'occupation et un second pour la disponibilité. Une solution approximative de la probabilité de blocage, basée sur les modèles de Markov avec récompense (MRM), est obtenue.

Nous proposons, dans ce chapitre, une approche basée sur les bornes stochastiques pour calculer à la fois une borne inférieure et une borne supérieure afin d'encadrer la probabilité de blocage. Notre travail a fait l'objet d'une publication à la conférence internationale ASMTA10 [69]. Ce chapitre est organisé comme suit :

- (a) Présentation du système et de l'approche de Trivedi
- (b) Construction d'un système bornant à forme produit
- (c) Construction de systèmes bornants agrégés
- (d) Résultats numériques des probabilités de blocage en fonction de différentes valeurs des paramètres d'entrée.

1 Modèle composite d'Erlang avec perte

Les systèmes dégradables sont conçus pour poursuivre leur exploitation, même en présence de défaillances de composants bien qu'à un niveau de performance réduite. Leur performance ne peut pas être évaluée avec précision sans tenir compte de l'impact des changements structurels (pannes et réparations). Le modèle Erlang avec perte composite est un modèle de performabilité car il combine les analyses de performance et de fiabilité.

1.1 Description du modèle

Nous considérons un système de commutation téléphonique composée de n canaux avec un nombre infini d'appelants [47, 70]. Le modèle composite pour l'analyse de la performabilité peut être représenté par une CTMC $\{X(t), t \geq 0\}$ de générateur infinitésimal Q . $\{X(t), t \geq 0\}$ est définie sur l'espace d'états $A \subset E = \{0, \dots, n\} \times \{0, \dots, n\}$, où n est le nombre de canaux disponibles. Soit $\vec{x} = (x_1, x_2) \in A$, un état de $X(t)$, où x_1 est le nombre de canaux disponibles (qui ne sont en pannes) et x_2

III.1 Modèle composite d'Erlang avec perte

ceux qui sont occupés. On a les relations suivantes : $x_1 \geq x_2$, et $x_1 \leq n$, $x_2 \leq n$. Les événements qui se déclenchent dans ce système sont :

– **Arrivées et services**

Un nouvel appel arrive avec un processus poissonnien de taux λ et n'est admis que s'il y a une disponibilité (au moins un canal libre et opérationnel). La durée d'un appel (ou temps de service) suit une distribution exponentielle de moyenne $1/\mu$.

– **Pannes et réparations**

Les pannes et réparations sont distribués exponentiellement de taux γ pour les pannes et τ pour les réparations. Les canaux occupés ne bénéficient pas de protection particulière. Donc si un canal occupé tombe en panne alors l'appel est perdu. Enfin, nous supposons que tous les canaux partagent un serveur unique pour la réparation.

Le diagramme d'états de la figure III.1 donne une illustration pour $n = 3$ canaux. La taille de $X(t)$ est $(n+1)(n+2)/2$ car $x_1 \geq x_2$. Cependant, on comprend aisément que l'augmentation de n entraîne celle de l'espace d'états et par conséquent, rend difficile le calcul de la distribution de la probabilité stationnaire.

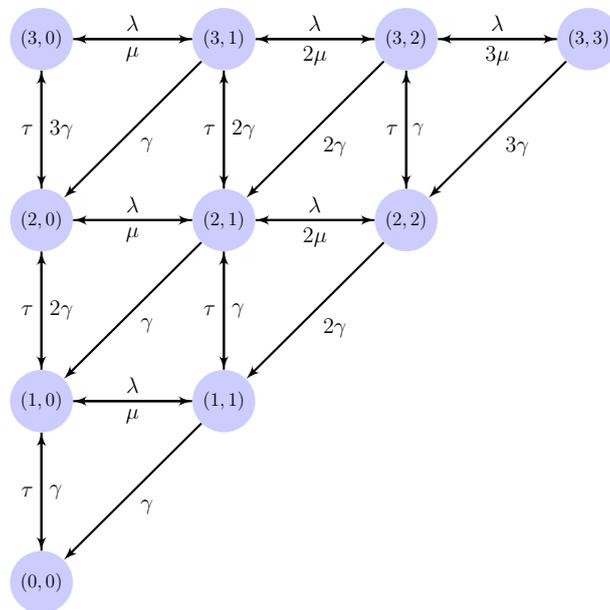


Figure III.1 – Modèle composite Erlang avec perte pour $n = 3$

Les différents événements (arrivée, service, panne ou réparation) qui font évoluer le processus $X(t)$, à partir de l'état \vec{x} , sont décrits par les équations d'évolutions suivantes :

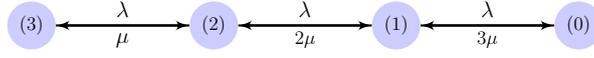


Figure III.2 – Modèle de performance pour $n = 3$

$$\begin{aligned}
 \vec{x} &\rightarrow (x_1, \min\{n, x_2 + 1\}), \text{ avec le taux } \lambda \\
 &\rightarrow (x_1, \max\{0, x_2 - 1\}), \text{ avec le taux } x_2\mu \\
 &\rightarrow (\max\{0, x_1 - 1\}, x_2), \text{ avec le taux } (x_1 - x_2)\gamma \\
 &\rightarrow (\max\{0, x_1 - 1\}, \max\{0, x_2 - 1\}), \text{ avec le taux } x_2\gamma \\
 &\rightarrow (\min\{n, x_1 + 1\}, x_2), \text{ avec le taux } \tau
 \end{aligned}
 \tag{III.1}$$

En notant par Π , la probabilité de distribution stationnaire de $X(t)$, l'objectif est donc de calculer la probabilité totale de blocage T_b définie comme la probabilité pour que le système n'accepte plus un nouvel appel :

$$T_b = \sum_{x_1=0}^n \Pi[x_1, x_1]
 \tag{III.2}$$

Nous remarquons qu'il n'existe pas de formule exacte pour T_b , et la résolution du système par des méthodes numériques peut-être difficile quand n croit. Nous présentons d'abord la solution approximative de Trivedi basée sur les modèles de Markov avec récompenses *MRM*.

1.2 Résolution approximative par approche MRM

Dans [47], une approximation de la probabilité de blocage est obtenue avec une approche markovienne avec récompenses (Markov Reward Model : MRM). Un modèle, haut niveau, de fiabilité est évalué avec pour récompenses les résultats du modèle de performance.

Soit la récompense $r_i = p_b(i)$ pour $i \geq 1$, définie comme la probabilité de blocage pour i canaux disponibles et calculée sur le modèle de performance de la Figure.III.2 :

$$p_b(i) = \frac{(\lambda/\mu)^i / i!}{\sum_{j=0}^i (\lambda/\mu)^j / j!}
 \tag{III.3}$$

L'équation III.3 est connue sous le nom de formule d'Erlang'B avec perte. La probabilité $p_b(i)$ est utilisée comme une fonction de récompense dans le modèle de disponibilité de la figure III.3 en régime stationnaire :

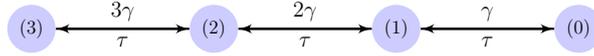


Figure III.3 – Modèle de disponibilité pour $n = 3$

$$\pi_i = \frac{1}{i!} (\tau/\gamma)^i \pi_0$$

avec π_0 qui représente la probabilité stationnaire d'indisponibilité et qui est donnée par $\pi_0 = [\sum_{i=0}^n \frac{1}{i!} (\tau/\gamma)^i]^{-1}$.

Soit, pour la probabilité de blocage totale approximative, l'expression suivante :

$$T_b^* = \sum_{i=0}^n r_i \pi_i = \sum_{i=1}^n p_b(i) \pi_i + \pi_0 \quad (\text{III.4})$$

Dans l'étude de Trivedi, la probabilité exacte de blocage T_b a été comparée avec l'approximation T_b^* . Dans la suite, à partir du système exact, nous définissons différents systèmes bornants afin de calculer des bornes (supérieures et inférieures) sur la probabilité de blocage.

Commençons par présenter le système bornant ayant une forme produit.

2 Système bornant à forme produit

Le principe est de construire un système multidimensionnel constitué de deux composantes (l'une pour le nombre de canaux disponible, et l'autre pour le nombre de canaux occupés), évoluant d'une manière indépendante, afin d'arriver à un système à forme produit.

2.1 Définition d'un système bornant à forme produit

Considérons le processus $X^{S1}(t)$, de la Figure.III.4, où chacun des états \vec{y} est représenté par le couple (y_1, y_2) . La première composante y_1 , désigne le nombre de canaux disponibles et la seconde y_2 , le nombre de canaux occupés. Le processus $X^{S1}(t)$ possède un comportement similaire à $X(t)$ excepté les points suivants :

- (1) L'occurrence d'une panne sur un canal occupé, dans le processus $X(t)$, génère une transition d'un état (x_1, x_2) vers un état $(x_1 - 1, x_2 - 1)$ avec un taux $x_2\gamma$. Tandis que, pour $X^{S1}(t)$, cet événement génère une transition de \vec{x} vers $(x_1 - 1, x_2)$ avec le même taux. Intuitivement, cela signifie que lorsqu'un canal occupé par un appel tombe en panne alors, cet appel bascule vers un autre canal disponible. Il n'est donc pas perdu.
- (2) Il est possible d'avoir, pour le processus $X^{S1}(t)$, $y_1 \leq y_2$ ou $y_1 \geq y_2$. Parce que nous considérons que les composantes évoluent d'une manière indépendantes. Ceci fait que, $X^{S1}(t)$ ne représente pas le système exact.

Intuitivement, nous pouvons déduire que $X^{S1}(t)$ représente une borne supérieure. En effet, avec la modification des taux de transition décrite dans (1), il est clair que nous augmentons les taux de transition vers les états (y_1, y_1) pour $y_1 = \{0, \dots, n\}$, sur lesquels nous calculons la probabilité de blocage T_b .

En outre, dans $X^{S1}(t)$, une seule composante varie à chaque transition de sorte que l'on retrouve l'indépendance des composantes. Et donc, la distribution de probabilité stationnaire Π^{S1} a une forme produit.

Les équations ci-dessous décrivent les différentes transitions de $X^{S1}(t)$ à partir d'un état \vec{y} :

$$\begin{aligned}
 \vec{y} &\rightarrow (y_1, \min\{n, y_2 + 1\}) \text{ avec le taux } \lambda \\
 &\rightarrow (y_1, \max\{0, y_2 - 1\}) \text{ avec le taux } y_2\mu \\
 &\rightarrow (\max\{0, y_1 - 1\}, x_2) \text{ avec le taux } y_1\gamma \\
 &\rightarrow (\min\{n, y_1 + 1\}, y_2) \text{ avec le taux } \tau
 \end{aligned}$$

(III.5)

2.2 Comparaison stochastique des processus $X(t)$ et $X^{S1}(t)$

Nous proposons de comparer $X(t)$ et $X^{S1}(t)$ afin d'obtenir des inégalités sur les probabilités de blocage.

Les comparaisons stochastiques de processus de Markov peuvent être effectuées uniquement sur les espaces d'états où un préordre est défini. Sur les espaces d'états multidimensionnels, l'ordre composante par composante est largement utilisé dans l'évaluation des performances des systèmes informatiques car il génère la comparai-

III.2 Système bornant à forme produit

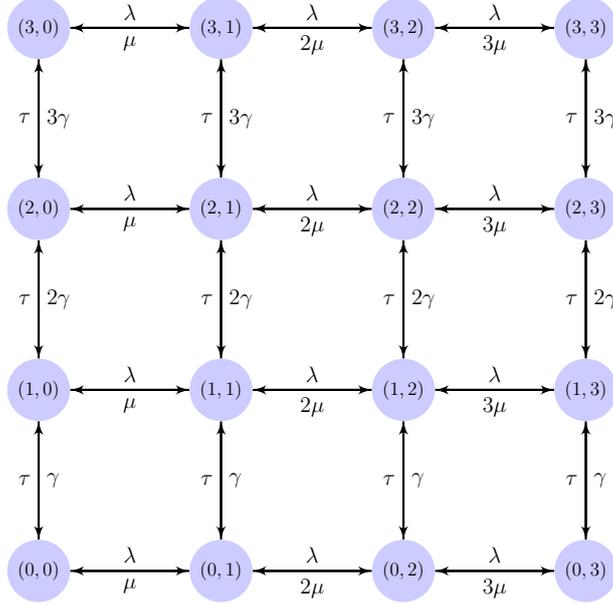


Figure III.4 – Modèle bornant à forme produit pour $n = 3$

son de certaines mesures de performance importantes comme : les probabilités de blocage, les temps de réponse, etc.

Dans le cas présent, il n'est pas possible d'utiliser cet ordre. En effet, pour T_b la récompense vaut, par exemple, 1 pour l'état $(0, 0)$ et 0 pour $(1, 0)$. De ce fait, pour l'ordre composante par composante, la récompense n'est pas une fonction croissante. Par contre, nous pouvons remarquer, d'après l'équation III.2, que T_b est calculée sur les états \vec{x} tels que $x_1 - x_2 = 0$. C'est pourquoi, nous proposons que ces états représentent les états supérieurs de l'espace d'états. Ce qui nous permet de définir sur A , le préordre \preceq suivant :

$$\vec{x} \preceq \vec{y} \iff x_1 - x_2 \geq y_1 - y_2 \quad (\text{III.6})$$

Appliquer la technique de couplage pour comparer les processus $X(t)$ et $X^{S^1}(t)$ requière qu'ils soient définis sur le même espace d'états. Or, nous remarquons que le processus $X(t)$ est défini sur A tandis que $X^{S^1}(t)$ est défini sur E . Comme $A \subset E$, nous définissons la fonction de projection $g : E \rightarrow A$, afin de comparer les processus sur l'espace d'états commun A . Donc g est définie comme suit :

$$g(x_1, x_2) = \begin{cases} (x_1, x_1) & \text{si } x_1 \leq x_2 \\ (x_1, x_2) & \text{sinon} \end{cases} \quad (\text{III.7})$$

Nous pouvons, à présent, établir la preuve que le processus $X^{S1}(t)$ est pour $X(t)$ une borne supérieure aussi bien en régime stationnaire que transitoire.

Proposition .22.

$$\{X(t), t \geq 0\} \preceq_{st} \{g(X^{S1}(t)), t \geq 0\} \quad (\text{III.8})$$

Preuve. Nous utilisons le théorème de couplage (voir Théorème .13) pour prouver l'existence de deux processus $\{\widehat{X}(t), t \geq 0\}$ et $\{\widehat{X}^{S1}(t), t \geq 0\}$ avec les mêmes matrices de générateurs infinitésimaux que, respectivement, $\{X(t), t \geq 0\}$ et $\{X^{S1}(t), t \geq 0\}$, représentant deux réalisations différentes et nous vérifions si :

$$\widehat{X}(0) \preceq g(\widehat{X}^{S1}(0)) \Rightarrow \widehat{X}(t) \preceq g(\widehat{X}^{S1}(t)), t > 0 \quad (\text{III.9})$$

Supposons que : $\widehat{X}(t) \preceq g(\widehat{X}^{S1}(t))$. Et comme $\widehat{X}(t) = \vec{x}$ et $\widehat{X}^{S1}(t) = \vec{y}$, alors nous obtenons $(x_1, x_2) \preceq g(y_1, y_2)$. Vérifions maintenant si :

$$\widehat{X}(t + \Delta t) \preceq g(\widehat{X}^{S1}(t + \Delta t)) \quad (\text{III.10})$$

Si nous considérons l'évolution des événements survenus au cours de l'intervalle de temps Δt , on constate que les deux processus ont la même évolution. Leurs comportements diffèrent à la survenue d'une panne sur un canal occupé par une communication.

(a) **Arrivée des appels :**

$$\widehat{X}(t + dt) = (x_1, \min\{n, x_2 + 1\}) \text{ et } \widehat{X}^{S1}(t + dt) = (y_1, \min\{n, y_2 + 1\}).$$

Comme $g(\widehat{X}^{S1}(t + dt)) = (y_1, y_2 + 1)$ ou (y_1, y_1) et comme nous avons supposé que $(x_1, x_2) \preceq g(y_1, y_2)$ alors, $\widehat{X}(t + dt) \preceq g(\widehat{X}^{S1}(t + dt))$.

(b) **Fin de service :**

$$\widehat{X}(t + dt) = (x_1, \max\{0, x_2 - 1\}) \text{ et } \widehat{X}^{S1}(t + dt) = (y_1, \max\{0, y_2 - 1\}).$$

Comme $g(\widehat{X}^{S1}(t + dt)) = (y_1, y_2 - 1)$ ou (y_1, y_1) alors, $\widehat{X}(t + dt) \preceq g(\widehat{X}^{S1}(t + dt))$

(c) **Panne sur un canal :**

$$\widehat{X}(t + dt) = (\max\{0, x_1 - 1\}, x_2) \text{ ou } (\max\{0, x_1 - 1\}, \max\{0, x_2 - 1\}) \text{ et } \widehat{X}^{S1}(t + dt) = (\max\{0, y_1 - 1\}, y_2).$$

Comme $g(\widehat{X}^{S1}(t + dt)) = (y_1 - 1, y_2)$ ou $(y_1 - 1, y_1 - 1)$ donc, quelque soit le cas : $\widehat{X}(t + dt) \preceq g(\widehat{X}^{S1}(t + dt))$.

(d) **Réparation d'un canal :**

$$\widehat{X}(t + dt) = (\min\{n, x_1 + 1\}, x_2) \text{ et } \widehat{X}^{S^1}(t + dt) = (\min\{n, y_1 + 1\}, y_2).$$

Comme $g(\widehat{X}^{S^1}(t + dt)) = (y_1 + 1, y_2)$ ou $(y_1 + 1, y_1 + 1)$ alors, $\widehat{X}(t + dt) \preceq g(\widehat{X}^{S^1}(t + dt))$.

Par conséquent, nous pouvons déduire que l'inégalité III.10 est vérifiée et donc, que la proposition .22 est prouvée. \square

D'après la comparaison des processus, nous pouvons déduire que :

$$\sum_{\vec{z} \in \Gamma} \Pi(\vec{x}) \leq \sum_{g(\vec{x}) \in \Gamma} \Pi^{S^1}(\vec{x}), \forall \Gamma \in \Phi_{st}(A) \quad (\text{III.11})$$

$\Gamma = \{(0, 0), (1, 1), \dots, (n, n)\}$, étant un ensemble croissant de $\Phi_{st}(A)$ alors, à partir de l'équation III.11 nous avons :

$$\sum_{\vec{x} \mid x_1 = x_2} \Pi(\vec{x}) \leq \sum_{\vec{x} \in E \mid x_1 \leq x_2} \Pi^{S^1}(\vec{x})$$

Donc, nous obtenons :

$$T_b \leq T_b^{S^1} = \sum_{\vec{x} \in E \mid x_1 \leq x_2} \Pi^{S^1}(\vec{x}). \quad (\text{III.12})$$

L'intérêt de ce résultat est que $T_b^{S^1}$ peut-être calculée facilement car la distribution de probabilité Π^{S^1} a une forme produit.

Jusqu'ici nous avons une borne supérieure mais pas d'encadrement. Alors, nous proposons maintenant d'utiliser une autre approche pour définir des systèmes bornants par agrégation des états du système initial. Nous cherchons à construire des systèmes de tailles plus petites.

3 Systèmes bornants agrégés

L'objectif est de construire des systèmes agrégés en tenant compte de la dynamique du système initial et de la mesure de performance à calculer.

3.1 Construction des modèles agrégés

Pour mémoire, la probabilité de blocage T_b est calculée comme le produit d'une fonction de récompense par la distribution de probabilité Π (voir équation III.2). Cette

fonction vaut 1 pour $x_1 - x_2 = 0$ et 0 sinon. De plus, on peut remarquer que les états ayant la même valeur de $x_1 - x_2$ ont une évolution similaire (voir Figure.III.5). Nous exploitons cette caractéristique pour construire des processus markoviens agrégés et bornants sur l'espace d'états $F = \{0, \dots, n\}$. Pour cela, nous réalisons une projection des états qui ont la même différence $x_1 - x_2$ vers un état $z = x_1 - x_2 \in F$ (voir Figure.III.5). Avec cette représentation, l'évolution est seulement soit de +1 (pour une fin de service ou une réparation) ou soit de -1 (pour une arrivée ou une panne). Par conséquent, le processus de Markov agrégé est équivalent à un processus de naissances et de morts, comme illustré dans la Figure.III.6, et avec des taux de transition dépendants de l'état pour être plus généraliste. Le processus est main-

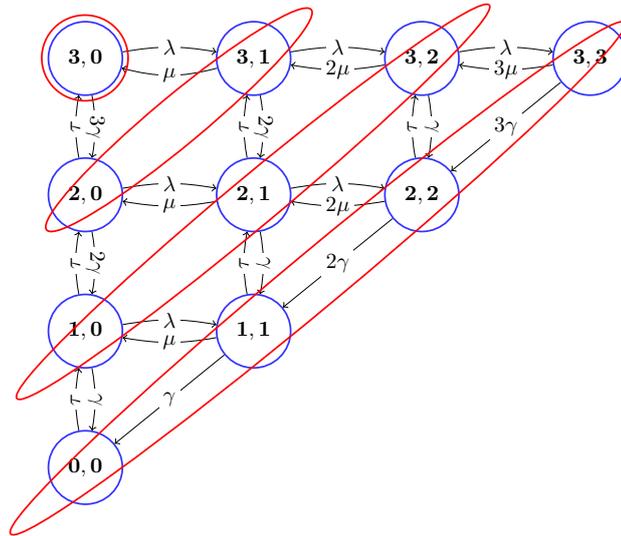


Figure III.5 – Agrégation des états avec z identiques

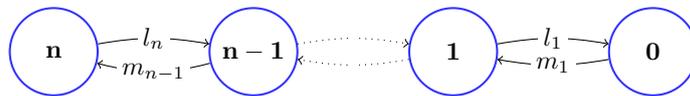


Figure III.6 – Modèle bornant agrégé de naissances et de morts

tenant représenté par une composante unique, il est donc plus facile de calculer les probabilités de distribution stationnaires et transitoires. Nous définissons sur F , le préordre \preceq suivant :

$$\forall x_1, x_2 \in F, x_1 \preceq x_2 \iff x_1 \geq x_2 \tag{III.13}$$

L'ordre est défini de sorte que l'état 0 de F soit le plus grand et l'état n le plus petit. Ainsi, les états tels que $x_1 - x_2 = 0$ sont les plus grands car ils sont utilisés pour le calcul de T_b . Avec cet ordre, tout système bornant fournira une valeur bornante pour cette mesure de performance.

A partir d'un état $n \preceq x \preceq 0$, un saut de -1 correspond à une transition de x vers $x - 1$, soit une naissance ou une transition vers un état plus grand avec le taux l_x . Un saut de $+1$ correspond, quand à lui, à une transition de x vers $x + 1$, équivalent à une mort ou un transit vers un état petit, avec le taux m_x .

Pour la comparaison des processus, nous utilisons la comparaison stochastique par fonction de projection en appliquant le théorème .19, basé sur la monotonie de l'un des processus. Nous allons montrer que notre processus, de naissances et de morts, est \preceq_{st} -monotone suivant la relation d'ordre établie.

3.2 Monotonie du processus bornant agrégé

Nous désignons par $Y(t)$ le processus de Markov de générateur Q_Y , défini sur l'espace d'états F avec le préordre \preceq et, équivalent à un processus de naissances et de morts comme décrit précédemment. Nous avons la proposition suivante :

Proposition .23. $Y(t)$ est \preceq_{st} -monotone.

Preuve. Apportons maintenant la preuve que $Y(t)$ est \preceq_{st} -monotone. Soit Q_Y le générateur infinitésimal, nous appliquons le théorème .18 et nous vérifions que :

$$\sum_{z \in \Gamma} Q_Y(x, z) \leq \sum_{z \in \Gamma} Q_Y(y, z) \quad (\text{III.14})$$

$\forall x, y \in F$, et $\forall \Gamma \in \Phi_{st}(F)$ tel que :

$$\forall (x, y) \preceq F \mid x, y \in \Gamma \text{ ou } x, y \notin \Gamma \quad (\text{III.15})$$

.

Nous considérons les deux états x et y , tel que $x \preceq y$, et l'ensemble croissant Γ tel que la condition (III.15) est vérifiée. Un ensemble croissant est une séquence croissante d'états (voir définition .3). Soit $\Gamma_x = \{x, x - 1, \dots, 1, 0\}$, l'ensemble croissant qui représente les états supérieurs à x . Si nous considérons les états x et $y = x - 1$ dans l'équation (III.15) alors, nous prenons l'ensemble croissant $\Gamma_x = \{x, x - 1, \dots, 0\}$ et

$\Gamma_{x-2} = \{x-2, \dots, 0\}$ pour la comparaison. En effet, cet ensemble vérifie la condition (III.15) et au moins, un des taux de transitions $\sum_{z \in \Gamma} Q_Y(x, z)$ ou $\sum_{z \in \Gamma} Q_Y(y, z)$ est non-nul dans l'inégalité (III.14). Pour les ensembles croissants Γ_x et Γ_{x-2} nous avons :

$$\begin{aligned} \sum_{z \in \Gamma_x} Q_Y(x, z) &= -m_x \quad \text{et} \quad \sum_{z \in \Gamma_x} Q_Y(x-1, z) = 0 \\ \sum_{z \in \Gamma_{x-2}} Q_Y(x, z) &= 0 \quad \text{et} \quad \sum_{z \in \Gamma_{x-2}} Q_Y(x-1, z) = l_{x-1} \end{aligned}$$

D'après ces résultats, on remarque bien que l'inégalité III.14 est vérifiée. Dans le cas où $x = y$, il est évident que les taux de transition sont égaux. Nous allons généraliser maintenant ces résultats pour n'importe quel couple d'états (x, y) , tel que $x \prec y$. Dans le cas général, nous considérons les ensembles croissants Γ_x et Γ_{y-1} . Pour Γ_x , nous obtenons $\sum_{z \in \Gamma_x} Q_Y(x, z) = -m_x$, et $\sum_{z \in \Gamma_x} Q_Y(y, z) = 0$. Et pour Γ_{y-1} , nous avons $\sum_{z \in \Gamma_{y-1}} Q_Y(x, z) = 0$ et $\sum_{z \in \Gamma_{y-1}} Q_Y(y, z) = l_{y-1}$.

Ainsi l'inégalité (III.14) est vérifiée, et la proposition .23 est prouvée. \square

Dans la suite, nous allons calculer les taux de transition du processus de naissance et de mort de façon à définir des systèmes bornants.

3.3 Calculs des taux des systèmes bornants agrégés

Soient $X^{S^2}(t)$ et $X^{I^2}(t)$, deux processus de naissances et de morts similaires à $Y(t)$ et définis respectivement pour être une borne supérieure et une borne inférieure pour $X(t)$. Pour chaque état x , nous désignons par :

- l_x , le taux de transition de x vers $x-1$,
- et par m_x celui de x vers $x+1$.
- $h : A \rightarrow F$, la fonction de projection surjective telle que pour un état $x = (x_1, x_2) \in A$, $h(x) = x_1 - x_2$.

3.3.1 Définition des taux de la borne supérieure

Appliquons le théorème .19 pour définir le processus $X^{S^2}(t)$ tel que :

$$h(X(t)) \preceq_{st} X^{S^2}(t)$$

Comme $X^{S^2}(t)$ est \preceq_{st} -monotone alors, nous pouvons appliquer l'inégalité (3) du théorème .19.

Nous prenons l'ensemble croissant pour lequel les taux de transition ne sont pas nuls.

A partir d'un état $x \in F$, du processus $X^{S2}(t)$, le taux de transition sont différents de 0 pour les ensembles croissants Γ_{x-1} , et Γ_x .

Ainsi, si nous prenons Γ_{x-1} alors, nous obtenons pour chaque état $x \in F$:

$$\sum_{h(z) \in \Gamma_{x-1}} Q(y, z) \leq l_x^{S2}, \forall y \in A \mid h(y) = x \quad (\text{III.16})$$

et donc,

$$l_x^{S2} = \max_{y \mid h(y)=x} \sum_{h(z) \in \Gamma_{x-1}} Q(y, z) = \lambda + x\gamma \quad (\text{III.17})$$

Et si nous prenons l'ensemble croissant Γ_x alors, nous avons :

$$\sum_{h(z) \in \Gamma_x} Q(y, z) \geq m_x^{S2}, \forall y \in A \mid h(y) = x \quad (\text{III.18})$$

Donc :

$$m_x^{S2} = \min_{y \mid h(y)=x} \sum_{h(z) \in \Gamma_{x-1}} Q(y, z) \quad (\text{III.19})$$

Nous choisissons la valeur suivante :

$$m_x^{S2} = \min(\mu, \tau).$$

3.3.2 Définition des taux de la borne inférieure

Pour la borne inférieure $X^{I2}(t)$, l'inégalité sur les taux de transitions est inversée. Ce qui nous donne les équations suivantes :

$$l_x^{I2} = \min_{y \mid h(y)=x} \sum_{h(z) \in \Gamma_{x-1}} Q(y, z) = \lambda + x\gamma. \quad (\text{III.20})$$

$$m_x^{I2} = \max_{y \mid h(y)=x} \sum_{h(z) \in \Gamma_x} Q(y, z) = \tau + (n - x)\mu \quad (\text{III.21})$$

Nous effectuons ensuite le calcul de la distribution de probabilité stationnaire Π^{S2} (respectivement Π^{I2}) de $X^{S2}(t)$ (respectivement $X^{I2}(t)$). Nous en déduisons les probabilités de blocage $T_b^{S2} = \Pi^{S2}(0)$ pour la borne supérieure et $T_b^{I2} = \Pi^{I2}(0)$ pour la borne inférieure. D'après la comparaison stochastique, si nous prenons l'ensemble croissant $\Gamma = \{0\}$ alors, nous obtenons la relation suivante :

$$T_b = \sum_{x_1=0}^n \Pi[x_1, x_1] \leq \Pi^{S2}(0) = T_b^{S2} \quad (\text{III.22})$$

et de la même manière ;

$$T_b \geq T_b^{I2}. \quad (\text{III.23})$$

4 Résultats

Dans cette section, nous cherchons d'une part à vérifier numériquement les relations d'inégalités entre les probabilités de blocage obtenues à partir des différents systèmes et d'autre part, à évaluer l'influence des paramètres sur la qualité des bornes. Les probabilités de blocage des différents systèmes bornants, sont calculées à partir des distributions de probabilité obtenues avec l'outil logiciel SHARPE¹.

Nous considérons que le système a les valeurs suivantes de paramètres :

- Le nombre de canaux est $n = 10$
- La durée d'un appel est $1/\mu = 2$ minutes.
- Le un taux de panne $\gamma = 1$ par heure
- Le temps de réparation $1/\tau = 30$ minutes.

Enfin, les appels arrivent avec une fréquence λ qui varie de 2 à 50 appels par minute. Dans les figures, nous désignons par $S1$, la borne supérieure fournie par le système à forme produit et par $S2$ et $I2$, les bornes supérieure et inférieure calculées à partir des systèmes agrégés.

a) Influence du taux d'arrivée des appels λ

La Figure III.7 illustre ce point et nous constatons :

- La borne inférieure $I2$ est de très bonne qualité car ses taux sont assez proches des taux du système exact.
- Le système à forme produit ($S1$) fourni une meilleure borne en comparaison du système agrégé ($S2$). Parce que les taux de transition des systèmes agrégés sont un peu loin du système exact En particulier, le $\min(\mu, \tau)$ qui est loin des valeurs des taux du système exact

b) Influence de n

Afin, d'évaluer l'influence de n sur la probabilité de blocage mais aussi sur la qualité des bornes, nous augmentons le nombre de canaux de communications dans le

1. SHARPE (Symbolic Hierarchical Automated Reliability and Performance Evaluator).
<http://www.sharpe.pratt.duke.edu/>

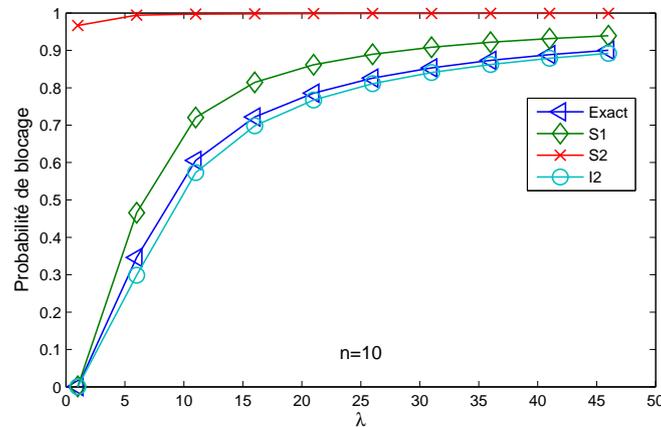


Figure III.7 – Influence de λ sur la probabilité de blocage

système à $n = 27$.

Nous remarquons sur la Figure III.8 que, l'ordre sur les bornes n'est pas modifié et que $S1$ fourni toujours une meilleure borne par rapport à $S2$. Par contre, la qualité des bornes s'est dégradée pour les bornes supérieures. En effet, ceci s'explique pour le système à forme produit que le calcul de la probabilité de blocage est une somme de probabilité sur plus d'états. Pour la borne supérieure du système agrégé, comme les taux dépendent des états alors l'écart avec les taux exacts est plus important quand n augmente. Par contre, pour la borne inférieure, la qualité demeure intéressante car les taux de transition du système agrégé demeurent plus proches de ceux du système exact.

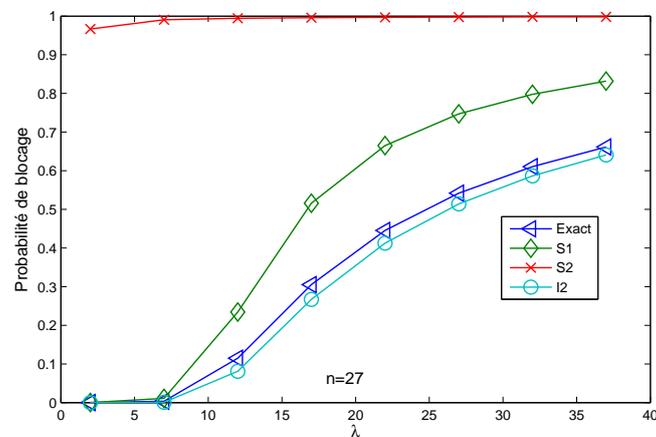


Figure III.8 – Influence du nombre de canaux n

c) Influence de la disponibilité sur les systèmes bornants

Nous décidons de dégrader fortement le système en augmentant la fréquence des pannes, γ , qui est maintenant de 5 par heure. Le temps de réparation est de $1/\tau = 15$ minutes et nous supposons que la durée des appels est plus longue : $1/\mu = 20$ minutes. Nous nous intéressons maintenant aux bornes supérieures uniquement.

Dans ces conditions, la Figure.III.9 montre que le système agrégé ($S2$) possède une meilleure qualité de borne par rapport au système à forme produit ($S1$). En effet, l'augmentation du taux de panne γ doit certainement augmenter les probabilités des états que l'on prend en plus dans la probabilité de blocage (états en dessous de la diagonale) ce qui dégrade la borne du système à forme produit. Pour le système agrégé, la plus grande proximité de μ et de τ améliore la qualité des bornes.

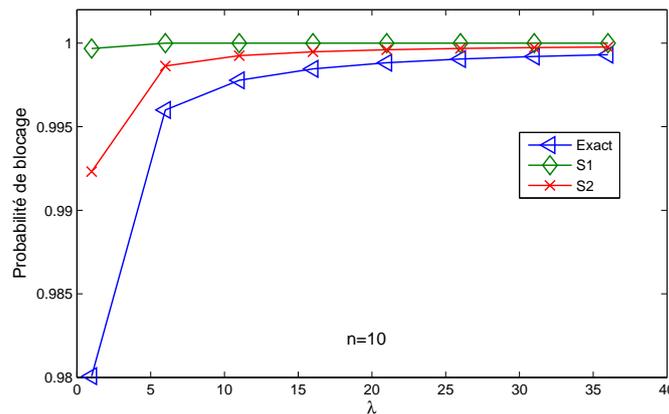


Figure III.9 – Impact de l'indisponibilité sur les bornes, $1/\mu = 20$ min, $\gamma = 5/hours$, $1/\tau = 15min$, $n = 10$

Dans la suite, nous comparons les résultats des bornes avec ceux de l'approximation de Trivedi.

d) Comparaison des bornes avec l'approximation de Trivedi

Dans le tableau III.1, nous donnons les probabilités de blocage pour les systèmes bornants, le système exact et l'approximation de Trivedi. Nous prenons les valeurs de paramètres utilisés pour obtenir la figure III.7. Nous remarquons que l'approximation de Trivedi donne des résultats bien meilleurs que ceux obtenus avec les systèmes bornants. Dans cet exemple, la borne supérieure S1 donne de meilleurs résultats que pour S2. On peut remarquer que lorsque $\lambda \geq 6$, nous sommes capables de proposer une borne supérieure (S1) de qualité intéressante ainsi qu'une borne inférieure assez précise. Ce résultat est important car bien que l'approximation de Trivedi soit de

λ	Exact	S1	S2	I2	Tb*
1	8,95E-04	1,27E-03	9,67E-01	3,65E-05	9,26E-04
6	3,46E-01	4,66E-01	9,94E-01	2,98E-01	3,48E-01
11	6,06E-01	7,20E-01	9,97E-01	5,73E-01	6,07E-01
16	7,22E-01	8,14E-01	9,98E-01	6,98E-01	7,23E-01
21	7,86E-01	8,62E-01	9,98E-01	7,67E-01	7,86E-01
26	8,26E-01	8,90E-01	9,99E-01	8,11E-01	8,26E-01
31	8,53E-01	9,08E-01	9,99E-01	8,41E-01	8,54E-01
36	8,73E-01	9,22E-01	9,99E-01	8,62E-01	8,74E-01
41	8,88E-01	9,32E-01	9,99E-01	8,79E-01	8,89E-01
46	9,00E-01	9,39E-01	9,99E-01	8,92E-01	9,01E-01

Tableau III.1 – Probabilité de blocage : mesures exactes, bornantes et approximatives de Trivedi

qualité, nous n'avons pas une preuve théorique sur le fait qu'elle soit supérieure ou inférieure à la valeur exacte. L'intérêt de notre solution est de proposer un encadrement de la valeur exacte, d'une qualité intéressante.

5 Conclusion

Dans ce chapitre, nous avons vu que la comparaison stochastique est une technique efficace dans les évaluations des systèmes complexes en proposant des bornes sur les mesures de performabilité. Ces bornes sont très souvent suffisantes pour quantifier un système dont la complexité rend difficile une résolution exacte.

Pour ce, nous avons défini des systèmes bornants en simplifiant le système exact. L'objectif étant d'obtenir d'autres systèmes plus simples à analyser. La mesure de performabilité, ici la probabilité de blocage, est encadrée par les valeurs obtenues à partir des systèmes bornants.

Enfin, nous avons montré qu'il n'y a pas un seul système bornant mais plusieurs dont les qualités dépendent des paramètres du système à évaluer. Ainsi l'intérêt de ce travail est d'utiliser différentes approches basées sur la comparaison stochastique afin de construire des systèmes bornants plus simples à analyser. On peut calculer différentes valeurs bornantes, supérieures et inférieures afin d'encadrer la valeur exacte, ce qui peut être un résultat plus important qu'une valeur approximative dont il n'est pas facile d'évaluer théoriquement la qualité.

Chapitre III. Evaluation de la performabilité d'un système de télécommunication

Chapitre IV

Contrôle d'admission et handover forcé dans les réseaux mobiles

Depuis déjà plusieurs années, le succès de la téléphonie mobile ne cesse de se confirmer grâce à la diversité des services offerts (téléphonie, internet, GPS, réseaux sociaux, etc...). Les opérateurs doivent donc répondre à une explosion des volumes des trafics engendrés par ces nouveaux usages de la part des utilisateurs. C'est pourquoi, il est indispensable de concevoir des réseaux mobiles capables de garantir des contraintes de qualité de service (QoS). Parmi les indicateurs de la QoS qui nous intéressent dans ce chapitre, nous étudierons la probabilité de blocage des nouveaux appels notée P_b (qui est la probabilité pour que le réseau rejette les nouveaux appels) et la probabilité de dropping des handovers (notée P_d ou dropping probability, qui est la probabilité pour qu'un appel en cours soit rejeté dans une situation de mobilité lors du transfert d'une cellule à une autre).

Il est clair que, du point de vue de l'utilisateur, la coupure d'un appel en cours est plus ennuyeuse que la non-allocation de ressource pour un nouvel appel entrant. Par conséquent, des mécanismes de contrôle des admissions des appels (CAC) [71–73] sont mis en œuvre. Parmi lesquels, le mécanisme GC (Guard Channel) qui est basé sur la réservation d'un certain nombre de canaux pour les appels en handovers. Cela permet de donner la priorité aux appels handovers par rapport aux nouveaux appels, afin de réduire la probabilité de dropping des handovers. Malheureusement, ces mécanismes ont un impact négatif sur la capacité du réseau à satisfaire la demande des nouveaux appels entrants et donc augmentent la probabilité de blocage des nouveaux appels.

Pour limiter cette dégradation, nous introduisons un nouveau mécanisme qui combine à la fois les techniques de CAC et le partage de charge entre les cellules avoi-

sinantes. Ce mécanisme, que nous désignons par *LMG* (pour Load sharing Mobile assisted handover with Guard channel), est une amélioration du mécanisme *MG* (MAHO with GC) introduit dans [48] par Madan et Trivedi. Nous reviendrons sur ces différents mécanismes plus tard dans ce chapitre.

Nous décomposons ce chapitre en deux grandes parties. Dans la première, nous étudierons un réseau monocellulaire et dans la seconde, un réseau multicellulaire.

(a) La première partie est organisée comme suit :

- Nous présentons le réseau étudié et les différents mécanismes impliqués dans l'élaboration du mécanisme *LMG*
- Ensuite, nous validons sur un réseau monocellulaire, la pertinence de *LMG* par rapport au mécanisme *MG* de Trivedi. Dans un premier temps, nous apportons la preuve par la comparaison stochastique et ensuite par l'analyse numérique en utilisant l'environnement de résolution SHARPE¹.

Ce travail a fait l'objet d'une publication dans la conférence internationale ASMTA12 [74].

(b) La seconde partie est, quand à elle, organisée ainsi :

- Nous décrivons le nouveau réseau sur lequel nous travaillons. Ce réseau multicellulaire nous permet, pour un ensemble de cellules adjacentes, de tenir compte de leurs états respectifs et également de l'itinérance (trafic de handover) entre elles à tout instant. En ce sens, ce réseau est plus réaliste que celui de la section précédente.
- Enfin, le réseau étant multidimensionnel, il se pose le problème de l'explosion de l'espace d'états. Pour contourner cette difficulté, nous utilisons l'approche développée dans le Chapitre.II pour proposer différents sous-réseaux bornants pour encadrer la mesure de performance recherchée.

Nous envisageons de publier les résultats de cette partie prochainement.

1 Approche combinée de contrôle d'admission et de handovers forcés dans un réseau monocellulaire

Dans cette partie, nous introduisons un mécanisme qui propose à la fois un contrôle sur les admissions pour les appels téléphoniques (nouveaux appels et handovers), et

1. Symbolic Hierarchical Automated Reliability Performance Evaluation

une gestion de l'état de la cellule en effectuant des handovers forcés lorsque celle-ci est surchargée.

1.1 Description du mécanisme proposé

Dans les réseaux cellulaires, maintenir les appels en cours durant les phases de mobilité (HandOver) relève de la nature même du service fourni aux usagers. De nombreux travaux ont été conduits dans le domaine du contrôle des admissions [71, 75] pour réduire la probabilité de dropping des handovers.

Lorsqu'un appel est en cours dans une cellule, certains mécanismes se focalisent sur la qualité du signal pour décider du transfert (handover) vers une cellule voisine. C'est le cas du mécanisme MAHO (Mobile Assisted handover) [76] qui est basé sur une décision collective entre le terminal mobile et la station de base à laquelle il est rattaché. Le terminal mobile rend compte périodiquement des valeurs de son couple RSSI/BER (force du signal reçu/ taux d'erreur binaire) pour aider la station de base à prendre la décision du transfert (handover) vers une cellule voisine qui offre de meilleures conditions radios. Ce mécanisme assure au terminal un bon niveau de signal avant son transfert vers une cellule cible. Par contre, il ne fournit pas d'information sur la disponibilité des canaux au niveau de cette même cellule. Le risque de voir cet appel rejeté existe donc malgré la qualité du signal offert.

D'autres mécanismes sont plus orientés vers l'accès aux ressources pour limiter la perte des appels en itinérance entre les cellules. Dans [72], les auteurs réservent un certain nombre de canaux (désignés comme Canaux de Garde - GC) pour les appels en handover qui sont transférés des cellules voisines. Le mécanisme GC donne une plus grande priorité aux appels en handover par rapport aux nouveaux appels. Ces derniers ne sont admis que si le nombre de canaux disponibles est supérieur à un seuil (g) défini de façon statique ou dynamique. Par exemple, dans [73, 77, 78], le nombre de canaux de garde n'est pas fixé, mais s'adapte en fonction de certains paramètres de trafic et améliore donc le mécanisme GC initial. Dans [79], les auteurs utilisent une évaluation conjointe de la disponibilité des ressources et la priorité des appels pour gérer l'accès aux canaux de garde. Lorsque cette ressource ne suffit pas, le CAC supprime les appels non prioritaires (préemption) ou réduit la ressource allouée (privation). Il est évident que ce mécanisme présente un avantage certain mais, il se pose deux problèmes. Le premier est que ce mécanisme ne garantit pas la qualité du signal radio offerte à l'appel en handover. Le second problème, c'est que

la mise en œuvre de ce mécanisme dégrade la probabilité de blocage des nouveaux appels.

Pour faire face à une mauvaise qualité radio durant la mobilité intercellulaire, Trivedi et al combinent les mécanismes MAHO et GC dans [47]. Cette approche, désignée sous le nom de MAHO avec GC (noté *MG*), est proposée comme une amélioration du mécanisme GC.

Contrairement à GC, qui réserve $g \subset n$ canaux pour les appels de handovers mais qui reste aveugle sur la qualité radio offerte, *MG* s'assure que l'appel reçoit un canal libre mais aussi un signal de bonne qualité afin de limiter la probabilité de dropping. Cette approche apporte effectivement une amélioration au mécanisme GC mais ne résout pas le problème de l'augmentation de la probabilité de blocage des nouveaux appels. En effet, réserver exclusivement un certain nombre de canaux pour protéger les appels handovers améliore la probabilité de dropping des handovers P_d certes, mais dégrade la probabilité de rejet des nouveaux appels P_b .

D'un autre côté, les expériences montrent que les cellules ne sont pas toujours occupées de la même manière. Une cellule peut être densément chargée dans la journée (quartier d'affaires par exemple) et moins chargée dans la soirée et vice-versa pour une cellule correspondant à une zone résidentielle. Nous proposons donc de profiter de la dynamique liée à l'asymétrie des charges entre les cellules voisines en répartissant les charges entre elles. *LMG* agit donc comme un mécanisme préventif contre la congestion au niveau de la cellule et améliore ainsi, l'efficacité du réseau. Donc, si une cellule venait à entrer en surcharge alors, *LMG* déclenche un "transfert forcé" d'un appel en périphérie de la cellule vers une cellule voisine à condition que celle-ci soit en mesure d'offrir un canal de libre et de bonnes conditions radios. Si au moins l'une de ces conditions n'est pas remplie alors, l'appel est conservé dans sa cellule d'origine. Le mécanisme *LMG* fonctionne en deux étapes parallèles :

- Il agit comme un mécanisme de CAC pour contrôler l'admission aux ressources avec une priorité pour les appels handovers sur les nouveaux appels entrants. Pour cela, il est identique au mécanisme *MG*.
- Il exerce une gestion dynamique des ressources. Si une cellule dépasse le seuil de surcharge, *LMG* force le transfert d'un appel en périphérie de la cellule surchargée vers une cellule voisine.

Dans la suite, nous décrivons le réseau monocellulaire étudié et définissons les paramètres qui vont être utilisés.

1.2 Description de la cellule étudiée

Nous considérons un réseau où les cellules ont un nombre fini n de canaux qui sont sujets à des pannes et des réparations. Nous admettons que toutes les cellules voisines sont statistiquement identiques. Dans ce cas, une seule cellule en isolation est représentative du système et toutes les interactions entre les cellules adjacentes seront capturées grâce à un processus d'arrivée des appels de handovers [72]. Nous admettons également que chaque cellule dispose de deux régions distinctes, le centre et la périphérie, et qu'il n'y a pas de mobilité entre ces deux régions. La mobilité intracellulaire peut être désactivée [80] si on applique la technique de saut de fréquence car, dans ce cas, tous les canaux ont le même niveau d'interférence.

Ainsi, une cellule de notre système, est définie par les paramètres suivants

- Arrivées : Les inter-arrivées sont exponentielles avec les taux suivants :
 - λ_c (resp. λ_e) pour les nouveaux appels arrivant au centre (resp. en périphérie) de la cellule.
 - λ_h pour les handovers arrivant en périphérie de la cellule.
- Services : Les temps de services sont exponentiels avec le paramètre μ_t pour un appel qui se termine (au centre ou en périphérie de la cellule) .
- Handover vers une cellule voisine : Le temps de séjour d'un appel dans la cellule avant un handover est exponentiel de paramètre μ_r .
- Panne : Un canal peut tomber en panne et, on suppose que, les temps entre deux pannes sont exponentiels de paramètre γ . A noter que si un canal occupé tombe en panne alors, l'appel en cours est transféré sur un canal disponible et perdu sinon [47].
- Réparation : Les réparations sont effectuées avec un seul serveur et les durées sont exponentielles de moyenne $\frac{1}{\tau}$.

Sachant que n désigne le nombre de canaux de la cellule, nous notons par $N = \{0, \dots, n\}$. Chacun des états de la cellule est défini par le tuple $\vec{x} = (x, x_c, x_e)$, où $x \in N$ désigne le nombre de canaux disponibles (en fonctionnement) et $x_c \in N$ (resp. $x_e \in N$) est le nombre de canaux occupés au centre (resp. en périphérie). Nous avons logiquement la relation suivante $(x_e + x_c) \leq x$ car le nombre total de canaux occupés ne peut pas dépasser le nombre de canaux en fonctionnement.

Ainsi, le modèle de performabilité décrit ci-dessus est une CMTC (Chaîne de Markov à Temps Continu) homogène et irréductible prenant ses valeurs dans l'espace d'états fini $A = N^3$.

Comme nous l'avons dit précédemment, le système *LMG* que nous proposons est basé sur un ensemble de systèmes : le système MAHO, le système *MG* puis les handovers forcés entre cellules voisines. Avant de présenter le système *LMG*, nous présentons ces différents systèmes.

1.3 Définitions des transitions des différents mécanismes

1.3.1 Les transitions du système MAHO

Appliquons le mécanisme MAHO à la cellule décrite plus haut, qui nous rappelons est un modèle qui combine la performance et la fiabilité. Comme le mécanisme MAHO est basé uniquement sur la qualité du signal radio, l'accès aux ressources est équitable entre les différents types d'appels (nouveaux et handovers). Les différents cas de transitions sont alors les suivants :

- le taux d'arrivée des nouveaux appels au centre est λ_c si $x_c + x_e < x$, sinon il est nul. Donc, l'équation d'évolution du système est telle que x_c augmente de un (voir transition IV.1).
- le taux d'arrivée des nouveaux appels en périphérie est λ_e si $x_c + x_e < x$ et nul sinon (voir transition IV.2).
- le taux d'arrivée pour les appels en handovers en périphérie est λ_h si $x_c + x_e < x$, sinon ce taux est nul (voir transition IV.2).
- le taux de service μ_c des appels au centre est le taux pour qu'un canal se libère au centre de la cellule (lorsqu'un appel se termine). Comme nous avons x_c canaux occupés alors $\mu_c = x_c \mu_t$ (voir transition IV.3).
- le taux de service μ_e des appels en périphérie est la somme des taux pour que les appels se terminent dans la cellule ou qu'ils partent vers les cellules voisines (handovers). Donc, nous avons $\mu_e = x_e(\mu_t + \mu_r)$ (voir transition IV.4).
- le taux de panne est $x_c \gamma$ pour les canaux occupés dans le centre si tous les canaux sont occupés (c'est à dire $x = x_c + x_e$) et dans ce cas, x et x_c diminuent de 1 (voir transition IV.5). Nous avons une évolution similaire pour les appels en périphérie de la cellule (voir transition IV.6).
- le taux de panne est de $(x - (x_c + x_e))\gamma + (x_c + x_e)\gamma$ s'il y a au moins un canal de disponible ($x - (x_c + x_e) > 0$). Si la panne survient sur un canal occupé (au centre ou en périphérie) alors la communication passe sur un autre canal, qui est disponible. Dans ce cas, seul le nombre x de canaux en fonctionnement décroît de

IV.1 Approche combinée de contrôle d'admission et de handovers forcés dans un réseau monocellulaire

1. De même, si la panne survient sur un canal non occupé alors, seul x décroît de 1 (voir transition IV.7).
- le taux de réparation est de τ et donc, x croît de un (voir transition IV.8).

Nous résumons ci-dessous les équations des différentes transitions du système avec le mécanisme MAHO. Donc, à partir d'un état \vec{x} du système nous avons les équations suivantes :

$$\vec{x} \rightarrow (x, \min\{n, x_c + 1\}, x_e), \text{ avec le taux } \lambda_c, \quad (\text{IV.1})$$

si $x_c + x_e < x$.

$$\rightarrow (x, x_c, \min\{n, x_e + 1\}), \text{ avec le taux } \lambda_e + \lambda_h, \quad (\text{IV.2})$$

si $x_c + x_e < x$.

$$\rightarrow (x, \max\{0, x_c - 1\}, x_e), \text{ avec le taux } \mu_c = x_c \mu_t. \quad (\text{IV.3})$$

$$\rightarrow (x, x_c, \max\{0, x_e - 1\}), \text{ avec le taux } \mu_e = x_e (\mu_t + \mu_r). \quad (\text{IV.4})$$

$$\rightarrow (\max\{0, x - 1\}, \max\{0, x_c - 1\}, x_e), \quad (\text{IV.5})$$

avec le taux $x_c \gamma$ *si* $x = x_c + x_e$.

$$\rightarrow (\max\{0, x - 1\}, x_c, \max\{0, x_e - 1\}), \quad (\text{IV.6})$$

avec le taux $x_e \gamma$ *si* $x = x_c + x_e$.

$$\rightarrow (\max\{0, x - 1\}, x_c, x_e), \quad (\text{IV.7})$$

avec le taux $(x_e + x_c) \gamma + (x - (x_c + x_e)) \gamma$,
si $x > (x_c + x_e)$.

$$\rightarrow (\min\{n, x + 1\}, x_c, x_e), \quad (\text{IV.8})$$

avec le taux τ .

La Figure IV.1 représente la CMTC correspondant au système que l'on vient de décrire et nous pouvons aisément remarquer la complexité du système à partir du nombre d'états et les différents événements qui se produisent. Les mesures de performance auxquels nous nous intéressons sont P_d et P_b . Elles sont calculées sur les mêmes états, précisément, sur les états \vec{x} tels que $x = x_c + x_e$. Nous désignons par $\pi_{i,j,k}^{MAHO}$, la probabilité pour que l'état $\vec{x} = (x, x_c, x_e)$ ait la valeur (i, j, k) . Ainsi, les probabilités P_d^{MAHO} et P_b^{MAHO} sont exprimées comme suit :

$$P_d^{MAHO} = P_b^{MAHO} = \sum_{j=0}^n \sum_{i=j}^n \pi_{i,j,i-j}^{MAHO} \quad (\text{IV.9})$$

1.3.2 Les transitions du système MG

Le mécanisme *MG* considère à la fois la qualité du signal radio, avec *MAHO*, et le mécanisme de *CAC* qui est basé sur un système de canaux de garde (*GC*). *GC* réserve un nombre g de canaux pour les handovers entre les cellules. Chaque fois que le nombre de canaux de libre, dans la cellule, est inférieur ou égal au seuil g , le mécanisme *GC* donne la priorité aux appels handovers en rejetant les nouveaux appels. Tous les types d'appels sont à nouveau acceptés dès que le nombre de canaux disponibles remonte au-dessus du seuil. Dans le cas où tous les canaux sont occupés, l'ensemble des appels qui se présentent sont rejetés. De plus, pour toujours garantir une qualité de signal acceptable à tout appel en handover, le mécanisme *MG* utilise la technique *MAHO*. Cela signifie que $\alpha = 1$, avec α qui exprime la probabilité pour que le handover soit un succès [48]. En d'autres termes, *MG* s'assure que l'appel qui sera transféré va obtenir à la fois un canal de libre mais aussi un bon niveau de signal radio.

Nous donnons maintenant les différentes transitions du système *MG*, à partir d'un état $\vec{x} = (x, x_c, x_e)$:

- (a) Le taux total des arrivées des nouveaux appels au centre de la cellule est λ_c si $x - (x_c + x_e) > g$ et nul sinon (voir transition IV.10).
- (b) Le taux total des arrivées en périphérie de la cellule est la somme du taux des nouveaux appels et du taux des appels en handover. Ce taux est $\lambda_e + \lambda_h$ si le nombre de canaux disponibles est supérieur à g ($x - (x_c + x_e) > g$) (voir transition IV.11). Autrement, seul les appels handover sont admis et le taux est λ_h (voir transition IV.12).
- (c) Le taux total de service pour les appels au centre est $\mu_c = x_c \mu_t$, (voir transition IV.13).
- (d) Le taux total de service pour les appels en périphérie de la cellule est $\mu_e = x_e(\mu_t + \mu_r)$, (voir transition IV.14).
- (e) Le taux de panne est $x_c \gamma$ pour les canaux occupés au centre de la cellule et quand, tous les canaux sont occupés ($x = x_c + x_e$) on a la transition IV.15. L'équation d'évolution est similaire concernant la panne d'un canal en périphérie et le taux est $x_e \gamma$ (voir transition IV.16).
- (f) Le taux de panne pour les canaux qui se trouvent au centre ou en périphérie de la cellule, qu'ils soient occupés ou pas, est $(x - (x_c + x_e))\gamma + (x_c + x_e)\gamma$ s'il y

a au moins un canal de disponible ($x - (x_c + x_e) > 0$). A noter que si un canal occupé tombe en panne alors, la communication passe sur un nouveau canal de disponible. Donc, dans tous les cas de figure seul le nombre x de canaux qui fonctionnent diminue de un (voir transition IV.17).

(g) Le taux de réparation est τ et x augmente de un (voir transition IV.18).

Donnons maintenant les équations d'évolution du système avec le mécanisme *MG*. A partir d'un état \vec{x} nous avons :

$$\vec{x} \rightarrow (x, \min\{n, x_c + 1\}, x_e), \text{ avec le taux } \lambda_c, \quad (\text{IV.10})$$

$$\text{si } x - (x_c + x_e) > g.$$

$$\rightarrow (x, x_c, \min\{n, x_e + 1\}), \text{ avec le taux } \lambda_e + \lambda_h, \quad (\text{IV.11})$$

$$\text{si } x - (x_c + x_e) > g.$$

$$\rightarrow (x, x_c, \min\{n, x_e + 1\}), \text{ avec le taux } \lambda_h, \quad (\text{IV.12})$$

$$\text{si } x - (x_c + x_e) \leq g.$$

$$\rightarrow (x, \max\{0, x_c - 1\}, x_e), \text{ avec le taux } \mu_c = x_c \mu_t. \quad (\text{IV.13})$$

$$\rightarrow (x, x_c, \max\{0, x_e - 1\}), \text{ avec le taux } \mu_e = x_e(\mu_t + \mu_r). \quad (\text{IV.14})$$

$$\rightarrow (\max\{0, x - 1\}, x_c, \max\{0, x_e - 1\}), \quad (\text{IV.15})$$

$$\text{avec le taux } x_e \gamma \text{ si } x = x_c + x_e.$$

$$\rightarrow (\max\{0, x - 1\}, \max\{0, x_c - 1\}, x_e), \quad (\text{IV.16})$$

$$\text{avec le taux } x_c \gamma \text{ si } x = x_c + x_e.$$

$$\rightarrow (\max\{0, x - 1\}, x_c, x_e), \quad (\text{IV.17})$$

$$\text{avec le taux } (x_e + x_c) \gamma + (x - (x_c + x_e)) \gamma,$$

$$\text{si } x > (x_c + x_e).$$

$$\rightarrow (\min\{n, x + 1\}, x_c, x_e), \quad (\text{IV.18})$$

$$\text{avec le taux } \tau.$$

On peut remarquer que le système *MG* ressemble au système *MAHO*, sauf pour les équations IV.10, IV.11 et IV.12 qui correspondent à la mise en place du CAC.

Nous nous intéressons maintenant aux mesures de performances du système *MG*. Nous désignons par $\pi^{MG}(i, j, k)$, la probabilité pour que l'état $\vec{x} = (x, x_c, x_e)$ du système soit égal à (i, j, k) . Pour le modèle *MG*, P_d^{MG} est calculée sur les états \vec{x} tels que $x = x_c + x_e$. P_d^{MG} est donc la somme des probabilités des états (i, j, k) telles

IV.1 Approche combinée de contrôle d'admission et de handovers forcés dans un réseau monocellulaire

que $k = i - j$ en considérant que $0 \leq j \leq n - g$ à cause de la réservation de canaux pour les handovers. Elle est exprimée comme suit :

$$P_d^{MG} = \sum_{j=0}^{n-g} \sum_{i=j}^n \pi_{i,j,i-j}^{MG}. \quad (\text{IV.19})$$

P_b^{MG} est, quand à elle, calculée en sommant les probabilités $\pi_{i,j,k}^{MG}$ telles que $i - (j + k) \leq g$ et est exprimée comme suit :

$$P_b^{MG} = \sum_{j=0}^{n-g} \sum_{k=0}^g \sum_{i=j+k}^n \pi_{i,j,i-j-k}^{MG}. \quad (\text{IV.20})$$

Il apparait logique que le mécanisme MG améliore la mesure P_d^{MG} à cause du mécanisme de réservation des canaux pour le handover. A contrario, il est normal que P_b^{MG} augmente pour la même cause. Nous verrons plus tard, à travers les résultats numériques les changements qu'à introduit MG .

Dans la suite, nous introduisons notre mécanisme, LMG , afin de limiter la dégradation de la probabilité de blocage des nouveaux appels.

1.3.3 Les transitions du système LMG

LMG est construit sur le mécanisme MG auquel on ajoute une fonction de contrôle de l'état de congestion de la cellule. Donc, en plus du seuil g qui définit le nombre de canaux de garde, nous utilisons un autre seuil, noté s , pour gérer l'état de la cellule en temps réel. Ainsi, s représente le nombre minimum de canaux de disponibles et en dessous duquel LMG déclenche un départ forcé d'un appel en périphérie de la cellule (qui est rentrée en congestion) vers les cellules avoisinantes susceptibles d'offrir un canal de libre et de bonnes conditions radio.

Le mécanisme de handover forcé modifie le temps de résidence des appels en périphérie de la cellule lorsque celle-ci est surchargée. Nous supposons que le temps de séjour des handovers forcés suit une loi exponentielle de paramètre μ_{r0} . En reprenant les mêmes paramètres que ceux du système MG , nous avons les cas suivants pour les taux de service des appels en périphérie. Si $x - (x_c + x_e) > s$, alors le taux de service total est la somme du taux pour que les appels se terminent dans la cellule et le taux correspondant au temps de séjour dans la cellule (channel residency) d'un appel avant d'être transféré vers une autre cellule (handover). Ce taux est $\mu_e = x_e(\mu_t + \mu_r)$

(voir transition IV.25).

Dans le cas où le nombre de canaux non occupés est inférieur à s , le taux de service total est $\mu_{e0} = x_e(\mu_t + \mu_r + \mu_{r0})$ (voir transition IV.26).

Dans la suite du chapitre, nous adoptons la notation suivante :

- *LMG1* et *MG1* font référence respectivement aux mécanismes *LMG* et *MG* dans le cas de l'étude du réseau à une cellule.
- *LMG2* et *MG2*, quand à eux, font référence respectivement aux mécanismes *LMG* et *MG* dans le cas de l'étude du réseau multicellulaire.

A partir des hypothèses sur les arrivées des appels et les services, nous en déduisons que le système *LMG1* est représenté par la chaîne de Markov $\{X^{LMG1}(t), t \geq 0\}$ et les différentes transitions qui ont lieu à partir d'un état $\vec{x} = (x, x_c, x_e)$ sont :

$$\vec{x} \rightarrow (x, \min\{n, x_c + 1\}, x_e), \text{ avec le taux } \lambda_c, \quad (\text{IV.21})$$

si $x - (x_c + x_e) > g$.

$$\rightarrow (x, x_c, \min\{n, x_e + 1\}), \text{ avec le taux } \lambda_e + \lambda_h, \quad (\text{IV.22})$$

si $x - (x_c + x_e) > g$.

$$\rightarrow (x, x_c, \min\{n, x_e + 1\}), \text{ avec le taux } \lambda_h, \quad (\text{IV.23})$$

si $x - (x_c + x_e) \leq g$.

$$\rightarrow (x, \max\{0, x_c - 1\}, x_e), \text{ avec le taux } \mu_c = x_c \mu_t. \quad (\text{IV.24})$$

$$\rightarrow (x, x_c, \max\{0, x_e - 1\}), \text{ avec le taux } \mu_e = x_e(\mu_t + \mu_r), \quad (\text{IV.25})$$

si $x - (x_c + x_e) > s$.

$$\rightarrow (x, x_c, \max\{0, x_e - 1\}), \text{ avec le taux } \mu_{e0} = x_e(\mu_t + \mu_r + \mu_{r0}), \quad (\text{IV.26})$$

si $x - (x_c + x_e) \leq s$.

$$\rightarrow (\max\{0, x - 1\}, x_c, \max\{0, x_e - 1\}), \quad (\text{IV.27})$$

avec le taux $x_e \gamma$ *si* $x = x_c + x_e$.

$$\rightarrow (\max\{0, x - 1\}, \max\{0, x_c - 1\}, x_e), \quad (\text{IV.28})$$

avec le taux $x_c \gamma$ *si* $x = x_c + x_e$.

$$\rightarrow (\max\{0, x - 1\}, x_c, x_e), \quad (\text{IV.29})$$

avec le taux $(x_e + x_c)\gamma + (x - (x_c + x_e))\gamma$,
si $x > (x_c + x_e)$.

$$\rightarrow (\min\{n, x + 1\}, x_c, x_e), \text{ avec le taux } \tau. \quad (\text{IV.30})$$

IV.1 Approche combinée de contrôle d'admission et de handovers forcés dans un réseau monocellulaire

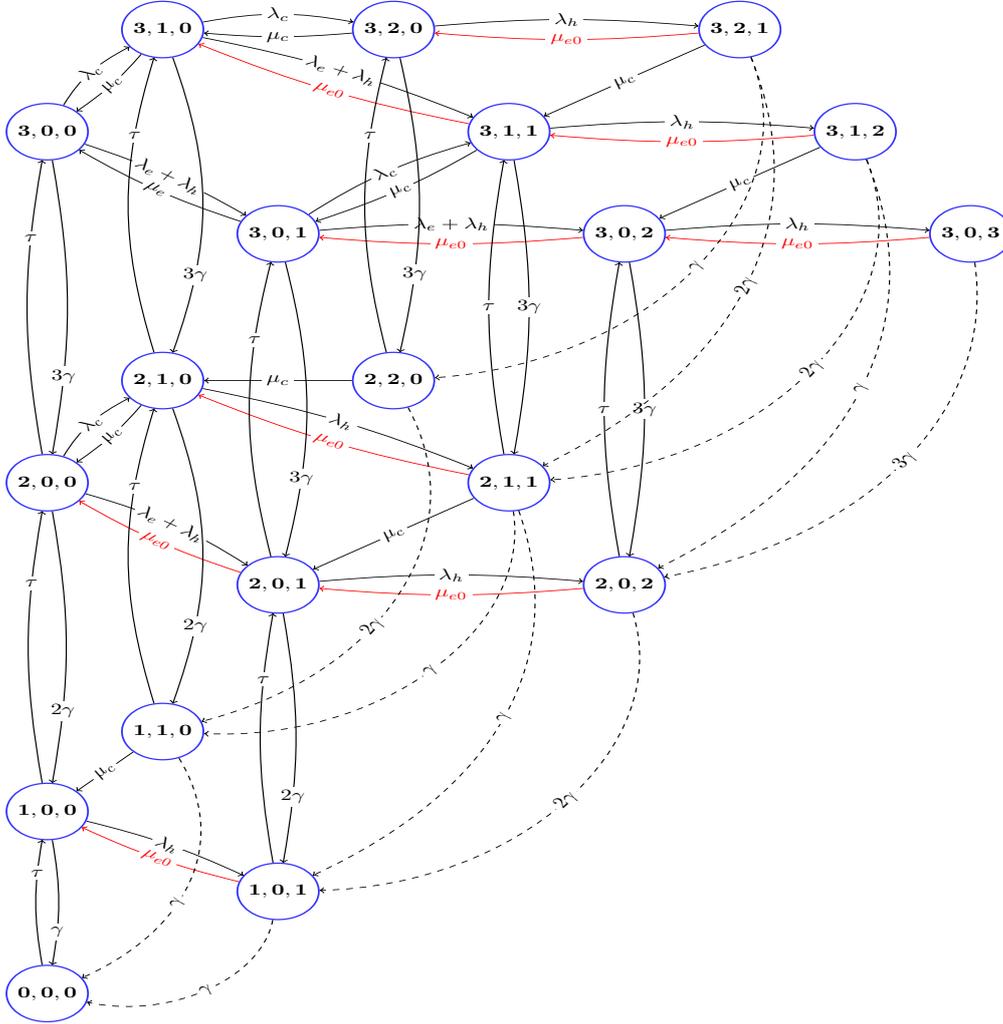


Figure IV.2 – La chaîne de Markov du système avec le mécanisme $LMG1$

Nous représentons dans la Figure IV.2 la chaîne de Markov du système avec le mécanisme $LMG1$ pour $g = s = 1$ et $n = 3$. Pour ce modèle, le calcul des mesures de performances P_d^{LMG1} et P_b^{LMG1} est identique (en utilise les mêmes états) au cas précédent du mécanisme $MG1$. Si nous désignons par $\pi^{LMG1}(i, j, k)$, la probabilité pour que l'état $\vec{x} = (x, x_c, x_e)$ du système $LMG1$ soit égal à (i, j, k) alors, P_d^{LMG1} et P_b^{LMG1} sont exprimées comme suit :

$$P_d^{LMG1} = \sum_{j=0}^{n-g} \sum_{i=j}^n \pi_{i,j,i-j}^{LMG1}. \quad (IV.31)$$

$$P_b^{LMG1} = \sum_{j=0}^{n-g} \sum_{k=0}^g \sum_{i=j+k}^n \pi_{i,j,i-j-k}^{LMG1}. \quad (\text{IV.32})$$

Dans la suite, nous allons comparer stochastiquement les systèmes $MG1$ et $LMG1$ de façon à vérifier les inégalités sur les mesures de performances.

1.4 Comparaison stochastique des systèmes $MG1$ et $LMG1$

Dans ce paragraphe, nous cherchons à prouver que le mécanisme $LMG1$ permet d'améliorer la probabilité de dropping des handovers mais aussi la probabilité de blocage des nouveaux appels par rapport au mécanisme $MG1$. La comparaison avec $MAHO$ est inutile car $MG1$ est déjà une amélioration de $MAHO$. Nous allons effectuer une comparaison entre les mesures de performances grâce à la comparaison stochastique qui est basée sur la théorie des ordres stochastiques.

Nous désignons par $\{X^{MG1}(t), t \geq 0\}$, le processus de Markov représentant le système avec le mécanisme $MG1$ et par $\{X^{LMG1}(t), t \geq 0\}$, celui représentant le système avec le mécanisme $LMG1$. Ces deux processus sont définis dans l'espace d'états $A = N^3$.

Comme P_d est calculée en sommant les états $\vec{x} = (x, x_c, x_e) \in A$ tels que nous avons $x = x_c + c_e$ alors, nous pouvons l'écrire comme une fonction de récompense sur la distribution de probabilité avec une récompense qui vaut 1 pour les états tels que $x = x_c + c_e$ et 0 pour tous les autres. P_b est, elle aussi, écrite comme une fonction de récompense avec une récompense qui vaut 1 pour les états tels que $x - (x_c + c_e) \leq g$ et 0 dans les autres cas.

Comme la comparaison des systèmes ne peut être faite que si la fonction de récompense est une fonction croissante selon le préordre défini sur l'espace d'états A alors, celui-ci est tel que : quand le nombre de canaux libres diminue alors, l'état est plus grand.

Pour tout $\vec{x}, \vec{y} \in A$, nous définissons le préordre \preceq . Donc $\vec{x} \prec \vec{y}$ si :

$$x - (x_e + x_c) > y - (y_e + y_c)$$

ou

$$x - (x_e + x_c) = y - (y_e + y_c) \text{ et } x_e > y_e, x_c > y_c$$

IV.1 Approche combinée de contrôle d'admission et de handovers forcés dans un réseau monocellulaire

et

$$\vec{x} = \vec{y} \text{ si } x = y, \text{ et } x_c = y_c, x_e = y_e$$

Nous allons apporter la preuve que :

$$\{X^{LMG1}(t), t \geq 0\} \preceq_{st} \{X^{MG1}(t), t \geq 0\} \quad (\text{IV.33})$$

Nous appliquons le Théorème .12, donc nous définissons $\{\widehat{X}^{LMG1}(t), t \geq 0\}$ (resp. $\{\widehat{X}^{MG1}(t), t \geq 0\}$) avec le même générateur infinitésimal que $\{X^{LMG1}(t), t \geq 0\}$ (resp. $\{X^{MG1}(t), t \geq 0\}$).

Nous supposons que : $\widehat{X}^{LMG1}(t) \preceq \widehat{X}^{MG1}(t)$ et nous allons prouver que :

$$\widehat{X}^{LMG1}(t + dt) \preceq \widehat{X}^{MG1}(t + dt) \quad (\text{IV.34})$$

Considérons alors les différents événements qui se produisent dans les systèmes :

- si nous avons une arrivée dans la cellule : comme nous avons le même taux d'arrivée (λ_c) dans les deux systèmes alors, si nous avons une arrivée pour $X^{LMG1}(t)$ nous avons aussi une arrivée pour $X^{MG1}(t)$. Donc, l'inégalité (IV.34) est vérifiée. Pour les autres types d'arrivées (en périphérie de la cellule pour des nouveaux appels ou des handovers), l'inégalité est aussi vérifiée car les deux processus ont un fonctionnement équivalent.
- pour la fin des communications dans le centre de la cellule, si nous avons une fin de communication dans $X^{MG1}(t)$ alors, elle a lieu aussi dans $X^{LMG1}(t)$ car les deux processus fonctionnent de la même façon. Donc, l'inégalité (IV.34) est vérifiée.
- si la fin de communications a lieu en périphérie de la cellule, nous avons deux cas possibles :
 - (a) dans le cas où $x - (x_e + x_c) > s$, les processus fonctionnent de la même façon et avec le même taux : donc si une fin de communication se produit dans $X^{MG1}(t)$ alors, elle se produit aussi dans $X^{LMG1}(t)$
 - (b) dans le cas où $x - (x_e + x_c) \leq s$, en raison des handovers forcés, le taux de service total pour libérer un canal en périphérie de la cellule est $x_e(\mu_t + \mu_r + \mu_{r0})$ pour le processus $X^{LMG1}(t)$. Ce taux est plus grand que $x_e(\mu_t + \mu_r)$, qui est le taux du processus $X^{MG1}(t)$ pour réaliser le même événement. Donc, l'inégalité IV.34 est là aussi vérifiée.
- pour les pannes et les réparations, les processus ont des comportements similaires donc, l'inégalité (IV.34) est vérifiée.

Nous avons montré à travers les différents événements que l'inégalité IV.34 est toujours vérifiée. Ce résultat est logique car vu le fonctionnement des systèmes, $X^{LMG1}(t)$ a plus de canaux disponibles que $X^{MG1}(t)$ à cause du mécanisme actif de libération des cellules s'il y a congestion. Par conséquent, nous en déduisons que l'équation (IV.33) est vérifiée. Et pour la distribution de probabilité stationnaire nous avons d'après la théorie des ordres stochastiques : $\pi^{LMG1} \preceq_{st} \pi^{MG1}$.

Et comme les probabilités de dropping des handovers et les probabilité de blocage des nouveaux appels de $X^{MG1}(t)$ et $X^{LMG1}(t)$ ont la même fonction de récompense qui est aussi une fonction croissante selon le préordre \preceq alors, nous en déduisons les inégalités suivantes entre les mesures de performance :

$$P_d^{LMG1} \leq P_d^{MG1} \text{ et } P_b^{LMG1} \leq P_b^{MG1} \quad (\text{IV.35})$$

Nous pouvons remarquer que la comparaison stochastique implique la comparaison des distributions de probabilités transitoires, et donc des mesures de performances transitoires. Dans cette partie, nous nous sommes intéressés qu'au comportement stationnaire. Par la suite, nous allons donner quelques résultats numériques pour comparer les mesures de performances des différents systèmes.

1.5 Résultats numériques des mécanismes étudiés

Nous voulons mettre en évidence l'intérêt du mécanisme LMG1 que nous avons proposé en donnant les valeurs numériques des mesures de performances correspondant aux différents systèmes : MAHO, MG1, et LMG1. Les valeurs ont été obtenues par analyse numérique en utilisant SHARPE. Nous considérons les valeurs suivantes pour les paramètres considérés dans la simulation :

- le nombre de canaux dans la cellule étudiée est de $n = 10$
- $\lambda \in [10, 40]$, désigne le taux total des arrivées dans la cellule. On considère que 60 % arrivent au centre ($\lambda_c = 0.6\lambda$) et 40% en périphérie ($\lambda_e = 0.4\lambda$)
- les handovers arrivent avec un taux $\lambda_h = 0.1\lambda$
- les communications durent en moyenne $1/\mu_t = 3 \text{ minutes}$
- le temps de séjour dans la cellule est d'en moyenne $1/\mu_r = 1.5 \text{ minutes}$
- le taux des pannes est $\gamma = 1/\text{heure}$ et le taux de réparation est $1/\tau = 30 \text{ minutes}$
- on prendra $g = s = 2$.

IV.1 Approche combinée de contrôle d'admission et de handovers forcés dans un réseau monocellulaire

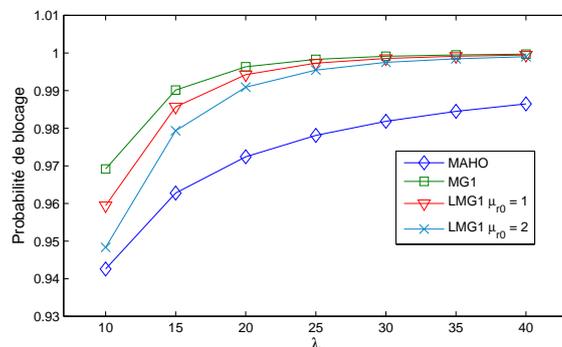


Figure IV.3 – Probabilité de blocage des nouveaux appels

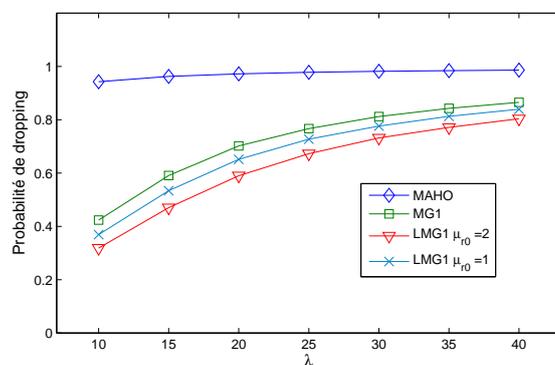


Figure IV.4 – Probabilité de dropping des handovers

Nous voulons observer l'apport de *LMG1* en comparant les mesures de performances du système avec les deux autres mécanismes : *MAHO* et *MG1*. Nous allons montrer l'intérêt du taux de départ forcé μ_{r0} et pour cela nous considérerons deux valeurs pour μ_{r0} : 1 et 2 par minute pour voir l'influence de ce paramètre. Dans les figures IV.3 et IV.4, nous donnons les probabilités de blocage des nouveaux appels et les probabilités de dropping des handovers pour les trois mécanismes : *MAHO*, *MG1* et *LMG1*.

Dans un premier temps, d'après la figure IV.4, nous pouvons voir que le mécanisme *MG1* améliore la probabilité de dropping des handovers par rapport au mécanisme *MAHO* mais que la probabilité de blocage des nouveaux appels dans *MG1* est dégradée d'après la figure IV.3. Ces résultats sont conformes à ce qui a été dit au début de ce chapitre. En effet, la réservation de canaux de garde pour les handovers permet de diminuer la probabilité de dropping mais en contre partie, cela dégrade la probabilité de blocage des nouveaux appels car il y a moins de ressources (de

canaux) pour eux.

D'après ces figures, nous remarquons aussi que le mécanisme *LMG1* améliore à la fois la probabilité de blocage mais aussi la probabilité de dropping par rapport au mécanisme *MG1*. Ces résultats sont conformes à la comparaison stochastique effectuée dans la section précédente. Cela est dû au fait que le mécanisme *LMG1* libère des ressources en forçant des appels en périphérie de la cellule à transiter vers les cellules adjacentes à condition qu'elles offrent des canaux disponibles et de bonnes conditions radios.

Enfin, nous pouvons remarquer l'impact du taux μ_{r0} . En effet, quand ce taux augmente alors, les mesures de performances sont améliorées. Ceci est logique car augmenter μ_{r0} revient à déclencher le mécanisme *LMG1* plus souvent et donc éviter la congestion de la cellule.

Maintenant que nous avons apporté la preuve de l'efficacité de *LMG1* sur une cellule, nous voulons le valider sur un système plus réaliste composé de plusieurs cellules et ainsi tenir compte de l'interaction qui existe entre elles.

2 Approche combinée de contrôle d'admission et de handovers forcés dans un réseau multicellulaire

Dans la section précédente, nous avons mis en évidence l'apport du mécanisme *LMG1* sur une cellule isolée. Nous avons supposé que le transfert forcé peut être toujours accepté dans des cellules voisines ce qui peut être trop optimiste. Nous considérons maintenant un système plus réaliste représenté par un ensemble de cellules afin de modéliser les états de toutes les cellules ainsi que les handovers entre elles.

Nous allons dans la suite décrire le mécanisme dans un système multicellulaire. Il sera noté *LMG2* dans la suite.

2.1 Description du système multicellulaire

Nous considérons un réseau, formé d'un ensemble de m cellules adjacentes. Chaque cellule de ce réseau est définie comme dans la section précédente. Nous considérons maintenant que des appels peuvent être routés entre une cellule et une autre voisine

IV.2 Approche combinée de contrôle d'admission et de handovers forcés dans un réseau multicellulaire

(par des handovers) selon des probabilités de routage.

Pour chaque cellule i , nous avons :

- pour les arrivées : λ_i^c est le taux d'arrivée des nouveaux appels au centre et λ_i^e , le taux d'arrivée des nouveaux appels à la périphérie.
- pour les services : μ_i^t est le taux de service au centre et μ_i^r , le taux de service à la périphérie de la cellule.
- pour les handovers : on considère que les handovers forcés, en cas de surcharge de la cellule, se font à la périphérie avec un taux μ_i^{r0} .

Pour l'ensemble du réseau de cellules, on note par $p_{i,j}$, la probabilité qu'un appel en périphérie de la cellule i soit transféré vers la cellule j en handover normal, et d_i la probabilité pour qu'un appel se termine dans la cellule i . En supposant que $p_{i,i} = 0$ alors, nous avons $\sum_{j \neq i} p_{i,j} + d_i = 1$. Nous notons par $p_{i,j}^0$ la probabilité de routage forcée entre une cellule i et une cellule j , et nous avons pour toute cellule i , $\sum_{j \neq i} p_{i,j}^0 = 1$. On supposera également qu'un handover forcé ne pourra se faire d'une cellule vers une autre cellule, que si la cellule cible n'est pas en surcharge. Si au contraire, elle est surchargée alors la communication continue dans la cellule d'origine.

On suppose que les temps d'inter-arrivées et de services suivent une loi exponentielle dont les paramètres sont donnés plus haut. Dans ce cas, le système que nous étudions peut être représenté par la chaîne de Markov $\{X^{LMG^2}(t), t \geq 0\}$ dont les états sont décrits par $\vec{x} = (x_1, x_i, \dots, x_m)$.

Pour chaque cellule $i = 1 \dots m$, nous définissons son état comme suit : $\vec{x}_i = (x_i^t, x_i^c, x_i^e)$, où x_i^t désigne le nombre total de canaux disponibles (en fonctionnement), x_i^c (resp. x_i^e), le nombre de canaux occupés dans le centre (resp. périphérie). Nous avons l'inégalité suivante : $x_i^t \geq (x_i^c + x_i^e)$, car le nombre de canaux occupés dans la cellule i ne dépasse pas le nombre de canaux en fonctionnement.

Décrivons maintenant, les évolutions du système à partir d'un état $\vec{x} \in \mathbb{N}^{3m}$:

$$\rightarrow (\vec{x}_1, \dots, (x_i^t, \min\{n, x_i^c + 1\}, x_i^e), \dots, \vec{x}_m), \quad (\text{IV.36})$$

avec le taux λ_i^c , si $x_i^t - (x_i^c + x_i^e) > g$.

$$\rightarrow (\vec{x}_1, \dots, (x_i^t, x_i^c, \min\{n, x_i^e + 1\}), \dots, \vec{x}_m), \quad (\text{IV.37})$$

avec le taux λ_i^e , si $x_i^t - (x_i^c + x_i^e) > g$.

$$\rightarrow (\vec{x}_1, \dots, (x_i^t, \max\{0, x_i^c - 1\}, x_i^e), \dots, \vec{x}_m), \quad (\text{IV.38})$$

avec le taux $x_i^c \mu_i^t$, si $x_i^c > 0$.

$$\rightarrow (\vec{x}_1, \dots, (x_i^t, x_i^c, \max\{0, x_i^e - 1\}), \dots, \vec{x}_m), \quad (\text{IV.39})$$

avec le taux $x_i^e \mu_i^r d_i$, si $x_i^e > 0$.

$$\rightarrow (\vec{x}_1, \dots, (x_i^t, x_i^c, \max\{0, x_i^e - 1\}), \dots, (x_j^t, x_j^c, \min\{n, x_j^e + 1\}), \dots, \vec{x}_m),$$

avec le taux $x_i^e \mu_i^r p_{i,j}$, si $(x_i^e > 0 \text{ et } x_j^t - (x_j^c + x_j^e) > 0)$. (IV.40)

$$\rightarrow (\vec{x}_1, \dots, (x_i^t, x_i^c, \max\{0, x_i^e - 1\}), \dots, (x_j^t, x_j^c, \min\{n, x_j^e + 1\}), \dots, \vec{x}_m),$$

avec le taux $x_i^e \mu_i^{r0} p_{i,j}^0$, (IV.41)

si $(x_i^e > 0, x_i^t - (x_i^c + x_i^e) \leq s)$, et $(x_j^t - (x_j^c + x_j^e) > s)$

$$\rightarrow (\vec{x}_1, \dots, (\max\{0, x_i^t - 1\}, x_i^c, x_i^e), \dots, \vec{x}_n), \quad (\text{IV.42})$$

avec le taux $x_i^t \gamma_i$, si $x_i^t > 0$, et $x_i^t > (x_i^c + x_i^e)$

$$\rightarrow (\vec{x}_1, \dots, (\max\{0, x_i^t - 1\}, \max\{0, x_i^c - 1\}, x_i^e), \dots, \vec{x}_n), \quad (\text{IV.43})$$

avec le taux $x_i^c \gamma_i$, si $x_i^c > 0$, et $x_i^t = (x_i^c + x_i^e)$

$$\rightarrow (\vec{x}_1, \dots, (\max\{0, x_i^t - 1\}, x_i^c, \max\{0, x_i^e - 1\}), \dots, \vec{x}_n), \quad (\text{IV.44})$$

avec le taux $x_i^e \gamma_i$, si $x_i^e > 0$, et $x_i^t = (x_i^c + x_i^e)$

$$\rightarrow (\vec{x}_1, \dots, (\min\{n, x_i^t + 1\}, x_i^c, x_i^e), \dots, \vec{x}_n), \quad (\text{IV.45})$$

avec le taux τ_i .

Dans la suite, nous expliquons les équations d'évolution.

- (a) Dans la transition (IV.36), le nombre de canaux occupés dans le centre de la cellule i augmente de un avec le taux λ_i^c si $x_i^t - (x_i^c + x_i^e) > g$ sinon, il reste inchangé.
- (b) Dans la transition (IV.37), le nombre de canaux occupés dans la périphérie de la cellule i augmente de un avec le taux λ_i^e , si le nombre de canaux de libre est supérieur au nombre de canaux de garde $(x_i^t - (x_i^c + x_i^e) > g)$.

IV.2 Approche combinée de contrôle d'admission et de handovers forcés dans un réseau multicellulaire

- (c) Dans la transition (IV.38), si un appel se termine dans le centre de la cellule i alors, le nombre de canaux occupés dans le centre diminue de un avec le taux $x_i^c \mu_i^t$.
- (d) Dans la transition (IV.39), le nombre de canaux occupés en périphérie de la cellule i diminue de un si l'appel se termine avec le taux $x_i^e \mu_i^r d_i$.
- (e) Dans la transition (IV.40), nous pouvons voir que lorsqu'un appel en périphérie de la cellule i transite vers la périphérie de la cellule j à cause du handover alors, l'appel est accepté si un canal est libre dans la cellule j , c'est à dire ($x_j^t > (x_j^c + x_j^e)$) sinon, l'appel est perdu. Donc le nombre de canaux occupés en périphérie de la cellule i diminue de un et le nombre de canaux occupés en périphérie de la cellule j augmente de un avec le taux $x_i^e \mu_i^r p_{i,j}$.
- (f) Dans la transition (IV.41), nous pouvons voir que lorsque le nombre de canaux non occupés dans la cellule i est inférieur à s alors, nous forçons l'appel à transiter vers la cellule j avec le taux $\mu_i^{r0} p_{i,j}^0$, si la cellule j n'est pas en surcharge. Dans le cas où la cellule cible est surchargée alors, l'appel se termine dans la cellule d'origine et donc, il n'y a pas de handover forcé.
- (g) Dans la transition (IV.42), comme $x_i^t - (x_i^c + x_i^e) > 0$ alors, nous pouvons avoir des pannes sur des canaux de libre ou occupés. Dans le cas où un canal occupé tombe en panne, la communication est commutée sur un autre canal de libre donc, seul le nombre de canaux total x_i^t de la cellule i diminue de un. Dans le cas où c'est un canal non occupé qui tombe en panne, là aussi seul x_i^t est modifié. Le taux de panne est alors de $(x_i^t - (x_i^c + x_i^e))\gamma + (x_i^c + x_i^e)\gamma = x_i^t \gamma$.
- (h) Dans les transitions (IV.43) et (IV.44), il n'existe aucun canal de disponible : $x_i^t - (x_i^c + x_i^e) = 0$. Alors, quand un canal tombe en panne la communication est perdue. Donc, dans la transition (IV.43) les deux composantes x_i^t et x_i^c diminuent et dans (IV.44), ce sont x_i^t et x_i^e qui vont diminuer.
- (i) Dans la transition (IV.45), le canal est réparé avec le taux τ donc, x_i^t augmente de un.

Nous considérons par la suite des mesures de performance transitoires. De ce qui précède, nous pouvons déduire les expressions de $Pd_i^{LMG2}(t)$ et de $Pb_i^{LMG2}(t)$, à

l'instant t et pour toute cellule i :

$$Pd_i^{LMG2}(t) = \sum_{\vec{x} \mid x_i^c=0}^{n-g} \sum_{x_i^t=x_i^c}^n \pi_{\vec{x}_1, \dots, (x_i^t, x_i^c, x_i^t-x_i^c), \dots, x_n}^{LMG2}(t) \quad (\text{IV.46})$$

$$Pb_i^{LMG2} = \sum_{\vec{x} \mid x_i^c=0}^{n-g} \sum_{x_i^e=0}^g \sum_{x_i^t=x_i^c+x_i^e}^n \pi_{\vec{x}_1, \dots, (x_i^t, x_i^c, x_i^t-x_i^c-x_i^e), \dots, x_n}^{LMG2}(t) \quad (\text{IV.47})$$

Il est clair que ce système est complexe à analyser. Nous proposons alors d'utiliser la comparaison stochastique par fonctions de projections et plus précisément l'approche proposée au chapitre.?? afin de construire des sous réseaux bornants.

2.2 Définition des sous-réseaux bornants

Soit $E = \{1, \dots, m\}$, l'ensemble des indices des cellules du réseau mobile et $J = \{1, \dots, k\}$, $k \leq m$, le sous-ensemble de cellules à partir desquelles nous voulons calculer les mesures de probabilités. On rappelle que J doit contenir la cellule i si cette dernière est utilisée pour calculer les mesures de probabilités. Ainsi, le sous-réseau bornant J est obtenu en coupant les liens entre les cellules de J et le reste du réseau, noté H . Nous avons bien-sûr $E = J \cup H$.

Intuitivement, pour la définition de la borne supérieure, le taux de transfert des clients (appels handovers) d'une cellule $j \notin J$ vers une cellule $i \in J$ est ajouté au taux d'arrivée en périphérie de la cellule i . En revanche, les taux de transfert de clients à partir d'une cellule $i \in J$ vers une cellule $j \notin J$ est considéré comme un taux de départ. Nous soulignons que la pertinence de ces bornes est de maintenir inchangée la dynamique interne du sous-réseau puisque nous gardons les transferts entre les cellules du sous-réseau.

Dans la suite, nous allons d'abord expliquer l'évolution du sous-réseau J .

2.2.1 Projection du processus $X^{LMG2}(t)$

Afin de voir comment le sous-réseau J évolue, nous utilisons les fonctions de projections sur les processus.

Définissons $\{g(X^{LMG2}(t)), t \geq 0\}$, le processus image par g de la CMTC $\{X^{LMG2}(t), t \geq 0\}$, où g est la fonction de projection $\mathbb{N}^{3m} \rightarrow \mathbb{N}^{3k}$ tel que nous soyons en mesure de décrire les évolutions de l'image du processus $\{X^{LMG2}(t), t \geq 0\}$ dans le but de définir des sous-systèmes bornants.

Ainsi, $\forall \vec{x} \in \mathbb{N}^{3m}$, $g(\vec{x}) = \vec{z}$, où \vec{z} représente un état du sous-réseau J , donc $g(\vec{x}) = \vec{z} = (x_1, \dots, x_k) \in \mathbb{N}^{3k}$.

Afin de calculer les taux de transitions, nous introduisons les vecteurs suivants :

- $v_i^c \in \mathbb{N}^{3k}$, un vecteur où toutes ses composantes sont nulles sauf $v_i^c = 1$,
- $v_i^e \in \mathbb{N}^{3k}$, un vecteur où toutes ses composantes sont nulles sauf $v_i^e = 1$.
- $v_i^t \in \mathbb{N}^{3k}$, un vecteur où toutes ses composantes sont nulles sauf $v_i^t = 1$.

Parmi les événements (une arrivée dans une cellule, un service, un transfert d'une cellule à une autre cellule) qui déclenchent sur le modèle original $\{X^{LMG2}(t), t \geq 0\}$ une transition de l'état \vec{x} vers l'état \vec{x}' , on distingue deux cas selon que l'événement modifie le sous-réseau J ou non. Si l'événement n'a pas d'incidence sur les cellules dans J alors, $g(\vec{x}) = g(\vec{x}')$. Sinon, pour $g(\vec{x}')$, nous devons prendre en compte la modification des cellules dans J en raison de certains événements qui se déclenchent. On obtient les taux de transition suivants :

2.2.2 Définition des taux de transition

- **Arrivée au centre** : si nous avons une arrivée au centre de la cellule $i \in J$ d'un nouvel appel alors, dans le cas où le nombre de canaux libres est supérieur à g nous avons $g(\vec{x}') = g(\vec{x}) + v_i^c$ et le taux de transition est :

$$\sum_{\vec{x}' | g(\vec{x}') = g(\vec{x}) + v_i^c} Q(\vec{x}, \vec{x}') = \lambda_i^c 1_{x_i^t - (x_i^e + x_i^c) > g} \quad (\text{IV.48})$$

- **Arrivée en périphérie** : si nous avons une arrivée d'un nouvel appel, quand le nombre de canaux libres est supérieur à g , ou un transfert (handover) provenant des cellules $j \notin J$, ou enfin un départ forcé (handover forcé) d'un appel des cellules $j \notin J$ quand les cellules j sont surchargées alors, $g(\vec{x}') = g(\vec{x}) + v_i^e$ avec le taux suivant :

$$\begin{aligned} \sum_{\vec{x}' | g(\vec{x}') = g(\vec{x}) + v_i^e} Q(\vec{x}, \vec{x}') = & \lambda_i^e 1_{x_i^t - (x_i^e + x_i^c) > g} \quad (\text{IV.49}) \\ & + \sum_{j \notin J} \min(x_j^e, x_j^t) \mu_j^r p_{j,i} \\ & + \sum_{j \notin J} \min(x_j^e, x_j^t) \mu_j^{r0} p_{j,i}^0 1_{x_j^t - (x_j^e + x_j^c) \leq s} 1_{x_i^t - (x_i^e + x_i^c) > s} \end{aligned}$$

- **Transferts (handovers) entre cellules du sous-réseau J** : si nous avons un transfert entre la cellule $i \in J$ vers la cellule $j \in J$ alors, $g(\vec{x}') = g(\vec{x}) + \vec{v}_j^e - \vec{v}_i^e$ et le taux de transition est :

$$\sum_{\vec{x}' | g(\vec{x}') = g(\vec{x}) + \vec{v}_j^e - \vec{v}_i^e} Q(\vec{x}, \vec{x}') = \min(x_i^e, x_i^t) \mu_i^r p_{i,j} \quad (\text{IV.50})$$

- **Transferts (handovers) forcés entre cellules du sous-réseau J** : si la cellule $i \in J$ est surchargée ($x_i^t - (x_i^e + x_i^c) \leq s$) alors, les appels en périphérie sont forcés à transiter vers une cellule $j \in J$ si celle-ci n'est pas surchargée, donc : $g(\vec{x}') = g(\vec{x}) + \vec{v}_j^e - \vec{v}_i^e$, et le taux de transition est le suivant :

$$\sum_{\vec{x}' | g(\vec{x}') = g(\vec{x}) + \vec{v}_j^e - \vec{v}_i^e} Q(\vec{x}, \vec{x}') = \min(x_i^e, x_i^t) \mu_i^{r0} p_{i,j}^0 \mathbf{1}_{x_i^t - (x_i^e + x_i^c) \leq s} \mathbf{1}_{x_j^t - (x_j^e + x_j^c) > s} \quad (\text{IV.51})$$

- **Départs d'appels de la périphérie de la cellule** : si un appel se termine dans la cellule i ou est transféré dans la cellule $j \notin J$ (normal ou forcé) alors, nous avons une transition vers $g(\vec{x}') = g(\vec{x}) - \vec{v}_i^e$ et le taux de transition est :

$$\begin{aligned} \sum_{\vec{x}' | g(\vec{x}') = g(\vec{x}) - \vec{v}_i^e} Q(\vec{x}, \vec{x}') = & \min(x_i^e, x_i^t) \mu_i^r [d_i + \sum_{j \notin J} p_{i,j}] \quad (\text{IV.52}) \\ & + \min(x_i^e, x_i^t) \mu_i^{r0} [\sum_{j \notin J} p_{i,j}^0] \mathbf{1}_{x_i^t - (x_i^e + x_i^c) \leq s} \mathbf{1}_{x_j^t - (x_j^e + x_j^c) > s} \end{aligned}$$

- **Pannes des canaux** : nous avons différents cas de figures. S'il existe de canaux de libres ($x_i^t - (x_i^e + x_i^c) > 0$) alors, pour tout type d'appels au centre ou en périphérie nous avons $g(\vec{x}') = g(\vec{x}) - \vec{v}_i^t$ et le taux de transition est :

$$\sum_{\vec{x}' | g(\vec{x}') = g(\vec{x}) - \vec{v}_i^t} Q(\vec{x}, \vec{x}') = x_i^t \gamma_i \quad (\text{IV.53})$$

s'il y a pas de canal de libre ($x_i^t - (x_i^e + x_i^c) = 0$) alors, si un canal au centre tombe en panne nous avons $g(\vec{x}') = g(\vec{x}) - \vec{v}_i^t - \vec{v}_i^e$ et le taux de transition est :

$$\sum_{\vec{x}' | g(\vec{x}') = g(\vec{x}) - \vec{v}_i^t - \vec{v}_i^e} Q(\vec{x}, \vec{x}') = x_i^c \gamma_i \quad (\text{IV.54})$$

s'il y a pas de canal de libre et si un canal en périphérie tombe en panne alors,

nous avons $g(\vec{x}') = g(\vec{x}) - v_i^t - v_i^e$ et le taux de transition est :

$$\sum_{\vec{x}' | g(\vec{x}') = g(\vec{x}) - v_i^t - v_i^e} Q(\vec{x}, \vec{x}') = x_i^e \gamma_i \quad (\text{IV.55})$$

- **Réparations des canaux** : le nombre de canaux disponibles augmente de 1 si $x_i^t < n$ alors, $g(\vec{x}') = g(\vec{x}) + v_i^t$ et le taux de transition est :

$$\sum_{\vec{x}' | g(\vec{x}') = g(\vec{x}) + v_i^t} Q(\vec{x}, \vec{x}') = \tau_i \quad (\text{IV.56})$$

Nous proposons maintenant de définir des sous-systèmes bornants supérieurs pour le processus $\{g(X^{LMG2}(t)), t \geq 0\}$. L'objectif est de définir un sous-système de cellules plus petit et qui représente une borne supérieure pour la probabilité de dropping des handovers, et la probabilité de rejet des nouveaux appels de chaque cellule du sous-système.

2.3 Construction de la borne supérieure

La définition de l'ordre sur l'espace d'états est très importante car elle doit permettre la comparaison des différentes mesures de performances. Intuitivement, comme le système bornant supérieur a des probabilités de dropping et de blocage des nouveaux appels plus grandes alors, il a moins de canaux libres que le système exact. Donc, nous disons que l'état \vec{x} est plus petit que l'état \vec{y} , si \vec{x} a plus de canaux libres que \vec{y} . Si nous désignons par \preceq cet ordre alors, nous avons :

$$\forall \vec{x}, \vec{y} \in \mathbb{N}^{3m}, \vec{x} \preceq \vec{y} \Leftrightarrow \forall 1 \leq i \leq m, x_i^t - (x_i^c + x_i^e) \geq y_i^t - (y_i^c + y_i^e),$$

Les systèmes bornants sont des réseaux de k cellules où chacun d'eux fournit des bornes sur le sous-réseau J . Nous appliquons le couplage par fonction de projection afin de définir la CMTC $\{Y^S(t), t \geq 0\}$, représentant une borne supérieure :

$$\forall t, g(X^{LMG2}(t)) \preceq_{st} Y^S(t)$$

où $\{Y^S(t), t \geq 0\}$ est définie sur l'espace d'états \mathbb{N}^k avec le générateur infinitésimal Q^S .

Nous suivons la même démarche que celle adoptée dans le chapitre II. Nous allons étudier le processus couplé $Z(t) = (X^{\hat{L}MG^2}(t), Y^{\hat{S}}(t))$ telles que ses réalisations restent dans :

$$K = \{(\vec{x}, \vec{y}) \in \mathbb{N}^{3m} \times \mathbb{N}^{3k} | g(\vec{x}) \preceq \vec{y}\}$$

où $X^{\hat{L}MG^2}(t)$ et $Y^{\hat{S}}(t)$ sont aussi des chaînes de Markov avec les générateurs infinitésimaux Q^{LMG^2} et Q^S . En dehors de K , les deux coordonnées peuvent être modifiées indépendamment mais dans K , les sauts des deux coordonnées ne peuvent être indépendants si l'on veut que le processus couplé reste dans K . Les étapes suivies pour la construction de la borne supérieure ressemblent à celles utilisées dans le chapitre II donc, nous ne donnerons que les principales étapes.

En considérant les événements qui se produisent dans $\{X^{LMG^2}(t), t \geq 0\}$ et qui déclenchent une transition à partir de \vec{x} vers \vec{x}' , on calcule les taux de transition de $\{Y^S(t), t \geq 0\}$ telles que les transitions du processus couplé $\{Z(t), t \geq 0\}$ restent dans K . Ci-après, nous décrivons les différents événements qui se produisent afin de définir les taux de transition de la borne supérieure.

(a) *Arrivée au centre de la cellule :*

Si nous avons une arrivée dans la cellule $i \in J$ si $x_i^t - (x_i^c + x_i^e) > g$, avec le taux λ_i^c , qui génère une transition de l'état \vec{x} vers \vec{x}' telle que $g(\vec{x}') = g(\vec{x}) + \vec{v}_i^c$ alors, nous devons avoir une transition de \vec{y} vers \vec{y}' telle que $\vec{y}' = \vec{y} + \vec{v}_i^c$, afin d'obtenir $g(\vec{x}') \preceq \vec{y}'$. Ce taux de transition doit satisfaire la condition suivante :

$$Q^S(\vec{y}, \vec{y} + \vec{v}_i^c) \geq \sum_{\vec{x}' | g(\vec{x}') = g(\vec{x}) + \vec{v}_i^c} Q(\vec{x}, \vec{x}') \quad (\text{IV.57})$$

Comme l'inégalité IV.57 doit être satisfaite pour tous les états \vec{x} tels que $(\vec{x}, \vec{y}) \in K$, nous prenons le maximum de la partie droite, $\vec{x} \in \mathbb{N}^{3m}$:

$$Q^S(\vec{y}, \vec{y} + \vec{v}_i^c) = \max_{\vec{x} | g(\vec{x}) \preceq \vec{y}} \left\{ \sum_{\vec{x}' | g(\vec{x}') = g(\vec{x}) + \vec{v}_i^c} Q(\vec{x}, \vec{x}') \right\}$$

Et donc, nous obtenons :

$$Q^S(\vec{y}, \vec{y} + \vec{v}_i^c) = \lambda_i^c 1_{x_i^t - (x_i^c + x_i^e) > g} \quad (\text{IV.58})$$

(b) *Arrivée en périphérie de la cellule :*

IV.2 Approche combinée de contrôle d'admission et de handovers forcés dans un réseau multicellulaire

Le taux de transition doit satisfaire la condition suivante :

$$Q^S(\vec{\mathbf{y}}, \vec{\mathbf{y}} + \vec{v}_i^e) \geq \sum_{\vec{\mathbf{x}}' | g(\vec{\mathbf{x}}') = g(\vec{\mathbf{x}}) + \vec{v}_i^e} Q(\vec{\mathbf{x}}, \vec{\mathbf{x}}') \quad (\text{IV.59})$$

D'après l'équation IV.49, nous obtenons :

$$Q^S(\vec{\mathbf{y}}, \vec{\mathbf{y}} + \vec{v}_i^e) \geq \lambda_i^e 1_{x_i^t - (x_i^e + x_i^c) > g} + \sum_{j \notin J} x_j^t \mu_j^r p_{j,i} + \sum_{j \notin J} x_j^t \mu_j^{r0} p_{j,i} 1_{x_j^t - (x_j^e + x_j^c) \leq s} 1_{x_i^t - (x_i^e + x_i^c) > s}$$

en prenant le maximum du second terme, nous obtenons :

$$Q^S(\vec{\mathbf{y}}, \vec{\mathbf{y}} + \vec{v}_i^e) = \lambda_i^e 1_{x_i^t - (x_i^e + x_i^c) > g} + \sum_{j \notin J} x_j^t \mu_j^r p_{j,i} + \sum_{j \notin J} x_j^t \mu_j^{r0} p_{j,i} 1_{x_i^t - (x_i^e + x_i^c) > s}$$

(c) *Départs du centre de la cellule (fin d'appel)*

Nous avons la condition suivante :

$$Q^S(\vec{\mathbf{y}}, \vec{\mathbf{y}} - \vec{v}_i^c) \leq \sum_{\vec{\mathbf{x}}' | g(\vec{\mathbf{x}}') = g(\vec{\mathbf{x}}) - \vec{v}_i^c} Q(\vec{\mathbf{x}}, \vec{\mathbf{x}}') \quad (\text{IV.60})$$

Comme l'évolution de $X^{LMG2}(t)$ est donnée par $\min(x_i^c, x_i^t) \mu_i^t$, alors :

$$Q^S(\vec{\mathbf{y}}, \vec{\mathbf{y}} - \vec{v}_i^c) = \min(x_i^c, x_i^t) \mu_i^t$$

(d) *Départs d'appels de la périphérie de la cellule :*

Dans ce cas là, nous avons l'inégalité suivante :

$$\sum_{\vec{\mathbf{x}} | g(\vec{\mathbf{x}}) = g(\vec{\mathbf{x}}) - \vec{v}_i^c} Q(\vec{\mathbf{x}}, \vec{\mathbf{x}}') \geq Q^S(\vec{\mathbf{y}}, \vec{\mathbf{y}} - \vec{v}_i^c)$$

En prenant le minimum de l'équation de gauche, et d'après l'équation IV.52 on obtient :

$$\sum_{\vec{\mathbf{x}} | g(\vec{\mathbf{x}}) = g(\vec{\mathbf{x}}) - \vec{v}_i^c} Q(\vec{\mathbf{x}}, \vec{\mathbf{x}}') = \min(x_i^e, x_i^t) \mu_i^r [d_i + \sum_{j \notin J} p_{i,j}] \quad (\text{IV.61})$$

Pour les autres transitions qui impactent seulement les cellules de J comme les transferts, les transferts forcés, les pannes ou les réparations, nous prenons pour la borne supérieure les mêmes taux que pour le système exact.

Ainsi, la borne supérieure est caractérisée par les taux de transition suivants pour une cellule $i \in J$:

- Taux des arrivées dans le centre de la cellule :

$$Q^S(\vec{\mathbf{y}}, \vec{\mathbf{y}} + \vec{v}_i^c) = \lambda_i^c 1_{x_i^t - (x_i^e + x_i^c) > g}$$

- Taux des arrivées dans la périphérie de la cellule :

$$Q^S(\vec{\mathbf{y}}, \vec{\mathbf{y}} + \vec{v}_i^e) = \lambda_i^e 1_{x_i^t - (x_i^e + x_i^c) > g} + \sum_{j \notin J} x_j^t \mu_j^r p_{j,i} + \sum_{j \notin J} x_j^t \mu_j^{r0} p_{j,i}^0 1_{x_i^t - (x_i^e + x_i^c) > s}$$

- Taux de départs des appels du centre de la cellule :

$$Q^S(\vec{\mathbf{y}}, \vec{\mathbf{y}} - \vec{v}_i^c) = \min(x_i^c, x_i^t) \mu_i^t$$

- Taux des départs des appels de la périphérie de la cellule :

$$Q^S(\vec{\mathbf{y}}, \vec{\mathbf{y}} - \vec{v}_i^e) = \min(x_i^e, x_i^t) \mu_i^r [d_i + \sum_{j \notin J} p_{i,j}]$$

- Taux de transition des handovers entre les cellules $i \in J$ et $j \in J$ pour les appels en périphérie, à l'intérieur du sous-réseau J :

$$Q^S(\vec{\mathbf{y}}, \vec{\mathbf{y}} - \vec{v}_i^e + \vec{v}_j^e) = \min(x_i^e, x_i^t) \mu_i^r p_{i,j}$$

- Taux de transition des handovers forcés entre la cellule i et la cellule j du sous-réseau, si $x_i^t - (x_i^e + x_i^c) \leq s$:

$$Q^S(\vec{\mathbf{y}}, \vec{\mathbf{y}} - \vec{v}_i^e + \vec{v}_j^e) = \min(x_i^e, x_i^t) \mu_i^{e0} p_{i,j}^0 1_{x_j^t - (x_j^e + x_j^c) > s}$$

Dans la suite, nous définissons les taux de la borne inférieure.

2.4 Construction de la borne inférieure

Nous appliquons le couplage par fonction de projection afin de définir le processus $\{Y^I(t), t \geq 0\}$, représentant une borne inférieure :

$$Y^I(t) \preceq_{st} g(X^{LMG^2}(t)), \forall t \geq 0,$$

IV.2 Approche combinée de contrôle d'admission et de handovers forcés dans un réseau multicellulaire

où $\{Y^I(t), t \geq 0\}$ est une CMTC définie sur l'espace d'états \mathbb{N}^k avec le générateur infinitésimal Q^I .

Nous ne détaillerons pas le calcul des taux de transition, l'approche est similaire que pour la borne supérieure, sauf que les inégalités sont inversées. Cela revient à minimiser les taux d'arrivée, à maximiser les taux de départs et à garder la dynamique à l'intérieur du sous-réseau.

– Taux des arrivées dans le centre de la cellule :

$$Q^I(\vec{x}, \vec{x} + \vec{v}_i^c) = \lambda_i^c 1_{x_i^t - (x_i^e + x_i^c) > g}$$

– Taux des arrivées en périphérie de la cellule :

$$Q^I(\vec{x}, \vec{x} + \vec{v}_i^e) = \lambda_i^e 1_{x_i^t - (x_i^e + x_i^c) > g}$$

– Taux des départs des appels du centre de la cellule : si $x_i^c > 0$,

$$Q^I(\vec{x}, \vec{x} - \vec{v}_i^c) = \min(x_i^c, x_i^t) \mu_i^t$$

– Taux des départs des appels de la périphérie de la cellule :

$$Q^I(\vec{x}, \vec{x} - \vec{v}_i^e) = \min(x_i^e, x_i^t) \mu_i^r [d_i + \sum_{j \in I} p_{i,j}] + \min(x_i^e, x_i^t) \mu_i^{r0} \sum_{j \notin I} p_{i,j}^0 1_{x_i^t - (x_i^e + x_i^c) \leq s}$$

– Taux de transition des handovers entre les cellules $i \in I$ et $j \in I$ pour les appels en périphérie de la cellule :

$$Q^I(\vec{x}, \vec{x} - \vec{v}_i^e + \vec{v}_j^e) = \min(x_i^e, x_i^t) \mu_i^r p_{i,j}$$

– Taux de transition des handovers forcés entre une cellule i surchargée et une autre cellule j à l'intérieur d'un sous-réseau :

$$Q^I(\vec{x}, \vec{x} - \vec{v}_i^e + \vec{v}_j^e) = \min(x_i^e, x_i^t) \mu_i^{r0} p_{i,j}^0 1_{x_i^t - (x_i^e + x_i^c) \leq s} 1_{x_j^t - (x_j^e + x_j^c) > s}$$

Pour résumer, les bornes agrégées obtenues sont assez intuitives. Comme l'idée générale consiste à couper les liens entre le sous réseau J et le reste du réseau, on considère différemment l'impact de cette rupture selon qu'il s'agit une borne supérieure ou inférieure. Ainsi, pour la

borne supérieure, les taux de handovers normaux et forcés sont rajoutés aux taux d'arrivée des appels en périphérie. Alors que pour la borne inférieure, ces taux ne sont pas rajoutés.

3 Resultats numériques sur un réseau multicellulaire

Nous étudions un ensemble de 7 cellules donné par la Figure.IV.5 avec un nombre $n = 10$ de canaux dans chaque cellule. Nous nous sommes intéressés aux probabilités de dropping des handovers et de blocage des nouveaux appels de la cellule 1. Pour évaluer la pertinence de *LMG2*, nous considérons que les cellules environnantes sont moins chargées.

En effet, nous sommes partis de l'hypothèse que le mécanisme peut-être utile lors d'une occupation asymétrique des cellules dans le temps. Pour cela, elles reçoivent un trafic équivalent à 50% du total du trafic qui arrive dans la cellule 1. Si nous avons un trafic de taux $\lambda_1 = \lambda_1^e + \lambda_1^c$ sur la cellule 1 alors, les autres cellules reçoivent $0.5\lambda_1$. Nous supposons que le trafic $0.5\lambda_1$ se répartit d'une manière équilibrée sur les cellules avoisinantes de la cellule 1. Les résultats sont comparés avec ceux que nous obtenons avec le système *MG2*.

La comparaison stochastique des systèmes *MG1* et *LMG1* est évidente et a été prouvée. Par contre, dans le cas multicellulaire, la comparaison stochastique est impossible car dans le système *LMG2*, on a des handovers forcés entre les files, alors que l'on n'en a pas dans le *MG2*. Donc, vu que l'on a défini un ordre cellule par cellule et que les transits forcés existent dans un système et pas dans un autre alors, l'ordre n'est plus vérifié entre les états des systèmes.

Les mesures de performance sont obtenues avec l'outil SimEvent. Les résultats sont obtenus avec un intervalle de confiance de 95% et un écart type de moins de 0,01. Nous reprenons les valeurs des paramètres utilisés dans la section précédente et pour le reste des paramètres, ils sont résumés dans le Tableau.IV.1.

Tableau IV.1 – Paramètres de la simulation

Seuils	$g = 2,$ $s = \{2, 6\},$
Probabilités de fin d'appel	$d_1 = 0.7,$ $d_i = 1 - p_{i,1}, \forall i \in m, i \neq 1$
Probabilités de routage avec la cellule 1 non surchargée	$p_{1,i} = 0.05, \forall i \in m, i \neq 1$ $p_{2,1} = 0.15, p_{3,1} = 0.1, p_{5,1} = 0.2, p_{4,1} = p_{6,1} = p_{7,1} = 0.3$
Probabilités de routage avec la cellule 1 surchargée	$p_{1,2}^0 = p_{1,5}^0 = 0.2, p_{1,3}^0 = p_{1,4}^0 = p_{1,6}^0 = p_{1,7}^0 = 0.15$

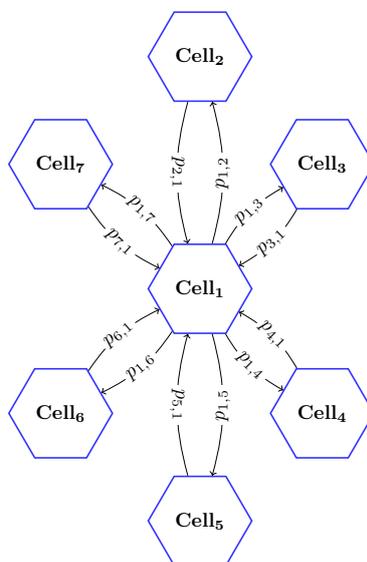


Figure IV.5 – Système de réseau mobile avec 7 cellules

3.1 LMG2 vs MG2 : l'influence de la charge

Nous varions λ_1 et comparons, dans les figures IV.6 et IV.7, les probabilités de blocage et de dropping dans la cellule 1 obtenues avec les mécanismes MG2 et LMG2. Différents scénarios ont été étudiés. Par exemple pour un système peu ou fortement chargé. Par ailleurs, dans un premier temps nous considérons que le trafic dans la cellule 1 est uniformément réparti entre le centre et la périphérie. Nous avons donc, 50% du trafic au centre et 50% en périphérie. Nous précisons que les résultats obtenus ne concernent que la cellule 1.

Nous observons que le mécanisme LMG2, dans les deux cas, donne de meilleurs résultats à la fois pour les probabilités de dropping des handovers que pour les probabilités de blocage pour les nouveaux appels. C'est conforme aux attentes puisque le mécanisme LMG2 agit contre la congestion des cellules en forçant les appels en périphérie de la cellule à s'associer avec les cellules adjacentes sous certaines conditions.

On note aussi que, l'amélioration est plus importante pour le système légèrement chargé que le système fortement chargé. Ceci est normal car les cellules voisines offrent moins de canaux pour les transferts forcés.

Dans les figures IV.8 et IV.9, nous étudions l'influence de la distribution des appels dans les différentes régions de la cellule. Nous considérons alors, les cas suivants : le centre reçoit 70% et la périphérie 30% du trafic (noté 70/30 dans la figure) ensuite, nous inversons. Donc, 30% du trafic arrivent au centre quand la périphérie reçoit 70% (noté 30/70 dans la figure). Ces cas seront désignés par la suite comme une distribution non-uniforme de trafic en opposition

aux cas où nous avons une distribution uniforme du trafic. Nous observons alors, de meilleurs résultats lorsque la majorité des appels sont en bordure de cellules. Ces résultats sont conformes à nos attentes car le mécanisme *LMG2* agit sur les appels en périphérie de la cellule alors qu'il n'a aucune influence sur ceux au centre.

Nous remarquons, par ailleurs, que l'impact est plus marqué sur la probabilité de dropping (Figure.IV.9) que sur la probabilité de blocage (Figure.IV.8). En effet, les premiers canaux libérés par le mécanisme *LMG2* vont bénéficier d'abord aux appels handovers à cause du mécanisme d'admission des appels *GC*.

3.2 LMG2 et MG2 : l'influence du seuil s

Dans les figures IV.10 et IV.11, nous voulons évaluer l'influence du seuil s sur le mécanisme *LMG2*. Nous décidons de déclencher le mécanisme *LMG2* à différents niveaux de fonctionnement du système (i.e $s = 2$ and $s = 6$). De plus, nous considérons à nouveau le cas où la distribution du trafic est uniforme entre le centre et la périphérie.

Nous en déduisons que le déclenchement précoce des handovers forcés dans le mécanisme *LMG2*, c'est à dire en augmentant le seuil s , donne de meilleures performances. Ceci s'explique par le fait que le mécanisme de répartition de la charge entre en action plus tôt et par conséquent, permet de réduire les mesures de performances.

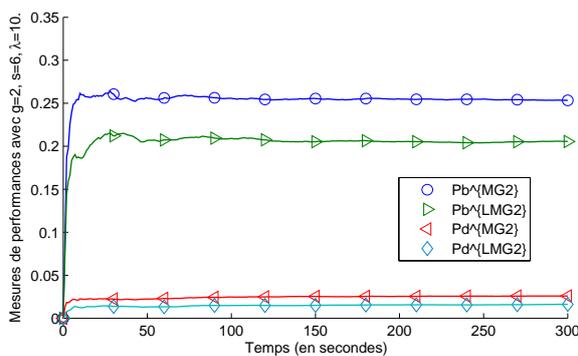


Figure IV.6 – Mesures de performances de la cellule 1 pour un système peu chargé ($\lambda_1 = 10$).

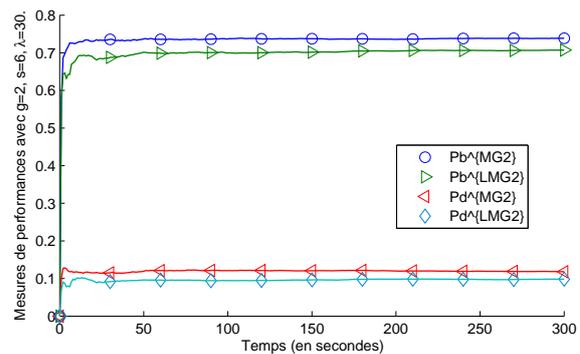


Figure IV.7 – Mesures de performances de la cellule 1 pour un système chargé ($\lambda_1 = 30$).

Encore une fois, nous précisons que les résultats obtenus ne concernent que la cellule 1. Il est clair que les performances des autres cellules vont être dégradées. Nous supposons qu'elles sont suffisamment sous-chargées (hypothèse de départ) pour ne pas avoir une augmentation importantes des mesures sur les performances. On mise sur le fait que, globalement, les performances du système sont meilleures grâce au partage de charge.

IV.3 Resultats numériques sur un réseau multicellulaire

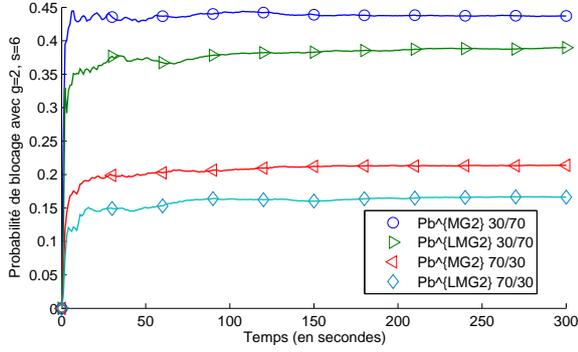


Figure IV.8 – Comparaison de la probabilité de blocage de *MG2* vs *LMG2* sur la cellule 1

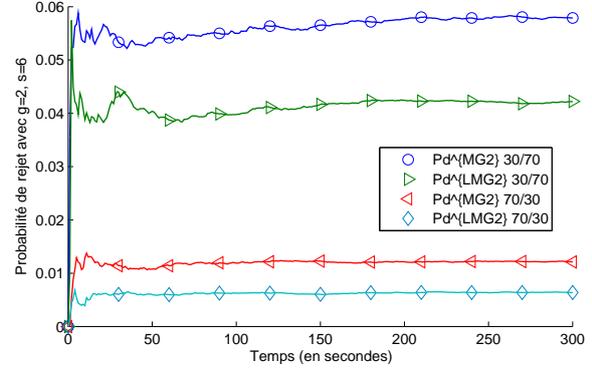


Figure IV.9 – Comparaison de la probabilité de dropping de *MG2* vs *LMG2* sur la cellule 1

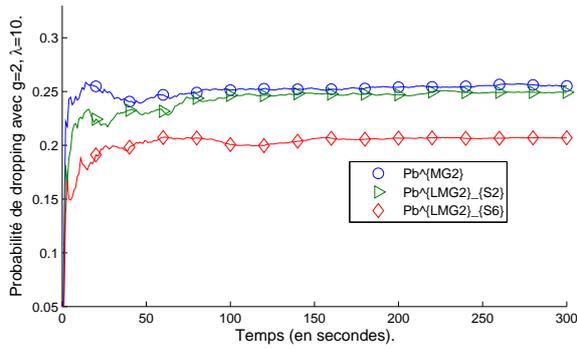


Figure IV.10 – Influence du seuil s sur la proba. blocage dans la cellule 1 de *MG2* vs *LMG2*

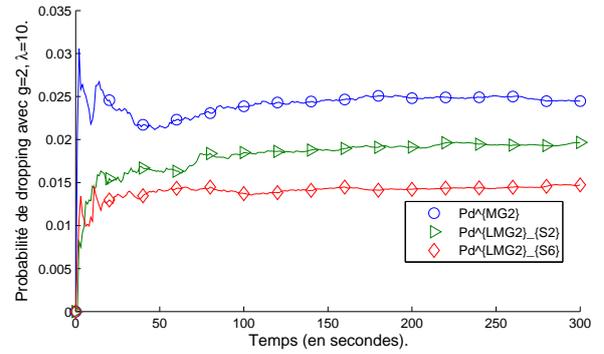


Figure IV.11 – Influence du seuil s sur la proba. dropping dans la cellule 1 de *MG2* vs *LMG2*

Nous remarquons que les simulations requièrent des ressources importantes et prennent donc beaucoup de temps. Par exemple, le temps nécessaire pour la simulation de la figure IV.11 est 470 secondes en moyenne. Comme nous cherchons à calculer les probabilités de blocage et de rejet de la cellule 1. Nous proposons d'étudier les différents sous-réseaux bornants décrits dans le Tableau.IV.2 avec les relations d'inclusion suivantes $J3 \subset J2 \subset J1$. Les sous-systèmes bornants sont définis par la suppression progressive des liens du réseau de façon à obtenir un sous-système plus facile à analyser et générant des bornes sur les mesures de performances.

Nous allons étudier l'impact des relations d'inclusion sur la précision des bornes.

Tableau IV.2 – Subnetworks description

$J1$	$\{1, 2, 3, 4, 5\}$
$J2$	$\{1, 2, 3\}$
$J3$	$\{1\}$

3.3 Agrégation bornante : l'impact de la taille des sous-systèmes bornants sur la qualité

Dans la suite, nous considérons les mesures de performances stationnaires et transitoires de la cellule 1 afin de voir l'impact de la taille du sous-réseau bornant sur la qualité des bornes.

Dans les figures IV.12 et IV.13, nous observons l'impact de la définition des sous-systèmes bornants sur la qualité des bornes supérieures. Nous notons par $S1$, $S2$ et $S3$, les bornes supérieures des sous-systèmes $J1$, $J2$ et $J3$. De toute évidence, $S1$ est plus précise que $S2$ et $S3$. Nous expliquons ceci par le fait que, contrairement au sous-réseau $J2$ où les routages entre les cellules 4 et 1 d'une part, et les cellules 5 et 1 d'autre part sont considérés comme des flux additionnels, pour le sous-réseau $J1$, ces routages sont pris en compte.

En revanche, $J1$ reste plus complexe que les sous-réseaux $J2$ et $J3$. Ainsi, comme la taille de l'espace d'états d'une cellule est $(n + 1)^3$ alors, la taille de l'espace d'états de $J1$ est $(n + 1)^{15}$ et celle de $J2$ est $(n + 1)^9$, ainsi on a divisé la taille de l'espace d'états par n^6 (où $n=10$). Par conséquent, il est évident qu'il y a un compromis à faire entre la taille et la qualité des sous-systèmes bornants.

De même la façon, dans les figures IV.14 et IV.15, nous avons les bornes inférieures. Et en couplant les deux résultats (bornes inférieures et supérieures), nous sommes en mesure de proposer des encadrements des mesures de performances recherchées.

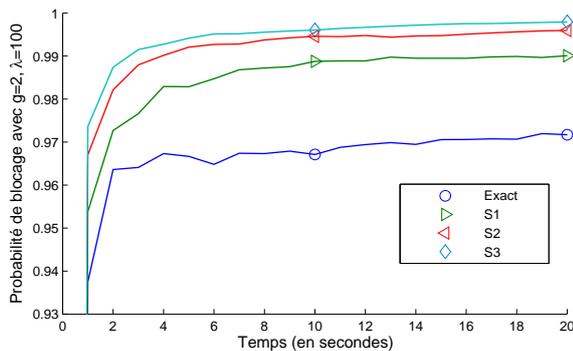


Figure IV.12 – Bornes supérieures de la probabilité de blocage

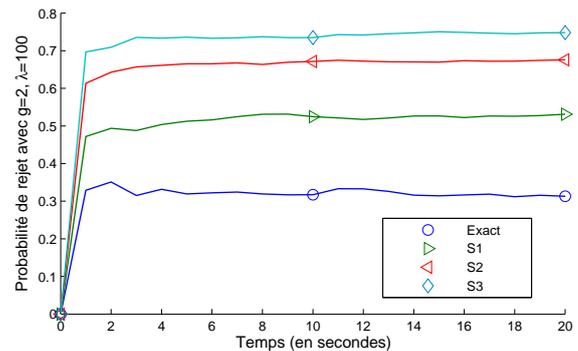


Figure IV.13 – Bornes supérieures de la probabilité de dropping.

IV.3 Resultats numériques sur un réseau multicellulaire

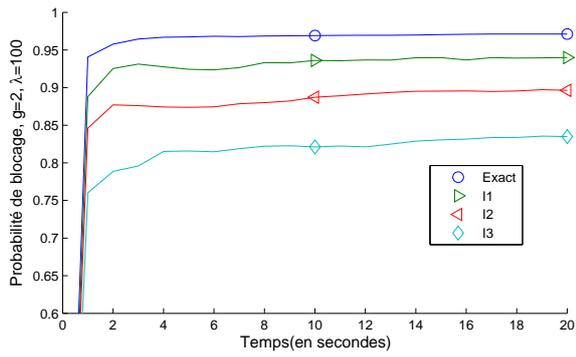


Figure IV.14 – Bornes inférieures de la probabilité de blocage.

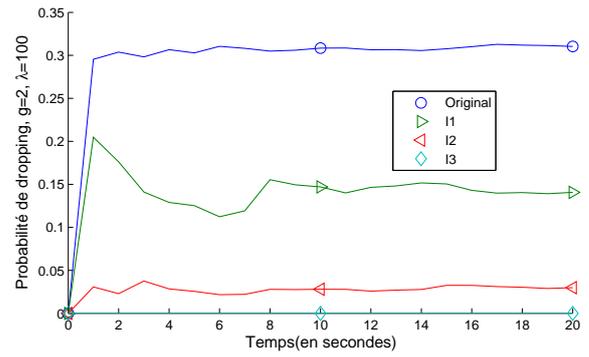


Figure IV.15 – Bornes inférieures de la probabilité de dropping.

Chapitre V

Conclusions générales et perspectives

Le travail présenté dans ce mémoire concerne l'évaluation des mesures de performance stationnaires et transitoires des systèmes complexes de grandes tailles en utilisant la comparaison stochastique, et plus précisément le couplage des chaînes de Markov. Dans ce chapitre, dans un premier temps, nous résumons les travaux effectués au cours de cette thèse en soulignant les contributions à travers les résultats obtenus. Pour mieux montrer l'intérêt des travaux, nous avons essayé, pour certains cas, de comparer nos résultats avec ceux de la communauté scientifique. Nous évoquons ensuite, les perspectives ouvertes sur le sujet et les axes de recherche en cours.

1 Conclusions

Prédire les performances d'un système dès sa conception permet de s'assurer que le système atteint les objectifs pour lesquels il a été conçu. Cependant, nous avons vu que les systèmes à évaluer deviennent rapidement complexes notamment lorsque que l'on prend en compte un certain nombre d'événements à la fois comme les pannes/réparation, et les arrivées/services dans des systèmes de performabilité. La prise en compte de plusieurs événements à la fois nécessite parfois de générer des modèles multidimensionnels dont l'analyse peut être complexe. C'est le cas également des réseaux de files d'attente, où l'analyse nœud par nœud nécessite de générer un système multidimensionnel souvent difficile à analyser par des méthodes mathématiques.

L'application des méthodes de bornes stochastiques présente un intérêt certain car l'approche permet d'encadrer des mesures de performances stationnaires mais aussi transitoires. Par exemple, on pourra déduire une probabilité de panne ou de non disponibilité sur un intervalle de temps, ou le temps moyen avant le déclenchement d'une panne.

D'une manière générale, nous avons utilisé les méthodes de bornes stochastiques de deux façons dans cette thèse :

1. Soit nous n'avons qu'un système, de grande taille et donc difficile à analyser, et nous construisons un système bornant plus simple à analyser.
2. Soit nous avons deux systèmes et nous voulons les comparer pour montrer que les mesures de performance qui nous intéressent sont comparables.

Pour notre part, nous avons utilisé essentiellement l'agrégation bornante basée sur le couplage par fonction de projection sur des espaces d'états plus petits. La méthode du couplage a l'avantage d'être intuitive car elle est basée sur la comparaison des trajectoires des processus en fonction des événements qui se déclenchent dans les systèmes.

1.1 Méthode de couplage appliquée à des réseaux de files d'attente

Dans le chapitre II, nous expliquons concrètement comment appliquer les méthodes de bornes et la comparaison stochastique par fonction de projection afin d'analyser plus facilement les performances d'un réseau de files $M/M/c_i/k_i$ et d'un G-Networks avec catastrophes. Sur ces systèmes, souvent nous n'avons besoin que d'évaluer les indices de performances qui ne dépendent que d'un nombre limité de nœuds (probabilité de blocage, disponibilité d'un système). Nous avons donc construit des chaînes de Markov de tailles réduites à partir desquelles nous calculons des mesures de performances bornantes (supérieures ou inférieures). L'intérêt de ces techniques est de proposer un compromis entre la qualité de la borne et la complexité de calcul. Pour illustrer nos propos, considérons par exemple notre système informatique sujet aux infections de virus du chapitre II section.2.

Considérons le G-Net de la figure II.10. Nous remarquons que la précision du temps moyen pour qu'un virus affecte la file 7 (Figure.V.1) est en rapport avec le type de sous-réseau (Tableau.V.1).

Nous avons également montré à travers les résultats numériques que les valeurs des paramètres intrinsèques au système original jouent également un rôle. C'est le cas des bornes construites dans le Chapitre II pour le réseaux de files $M/M/c_i/k_i$. Ainsi dans le système de la figure II.1, nous nous sommes intéressés à la probabilité de blocage de la file 3. On étudie

Sous-système bornant	Temps de simulation
Exact= $\{1, \dots, 7\}$	363 secondes
$J1 = \{5, 6, 7\}$	149 secondes
$J2 = \{5, 7\}$	100 secondes
$J3 = \{7\}$	53 secondes

Tableau V.1 – Temps de simulation (seconds)

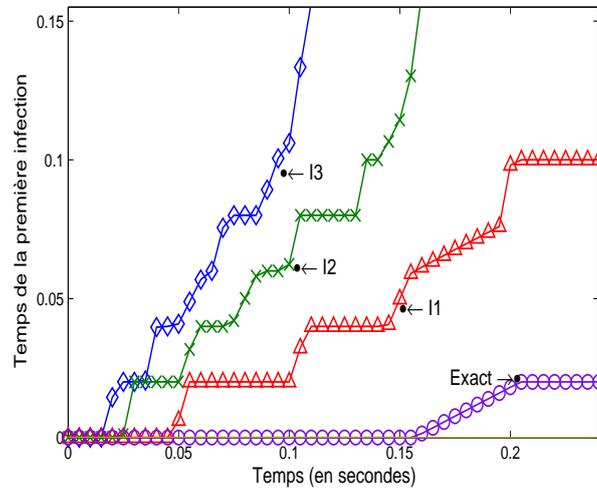


Figure V.1 – Temps avant la première infection dans la file 7 pour les différents sous-réseaux

le sous réseau $J_1 = \{2, 3\}$, et un système bornant supérieur est défini. Une comparaison de la borne supérieure de la probabilité de blocage de la file 3 et de la valeur exacte a été menée en fonction des différents paramètres.

La Figure V.2 montre que, lorsque le nombre de serveurs augmente alors, la qualité de la borne supérieure se dégrade. Ceci parce que, le taux d'arrivée dans la file d'attente 3 augmente avec c_1 et la probabilité de blocage de la borne supérieure augmente avec ce taux d'arrivée (voir Tableau.3.3). Par contre, dans le système exact nous atteignons un état d'équilibre.

Ce type d'étude a été repris dans des réseaux de taille plus grande, avec différents sous-réseaux, et la conclusion est similaire. Lorsque le nombre de serveurs, des files connectées à la file dont on veut calculer la mesure de performance, augmente alors, couper le lien dégrade nettement la qualité de la borne.

Donc à travers les exemples, nous avons montré deux tendances :

- Premièrement, plus le sous-réseau a des files (structure proche du réseau exact), meilleure est la qualité de la borne.
- Deuxièmement, des paramètres tels que le nombres de serveurs connectés aux files du sous-réseau ont un impact sur la qualité de la borne.

Ainsi, lorsque le nombre de serveurs diminue alors, la coupure du lien a peu d'impact sur la qualité de la borne, ainsi on peut trouver des sous-réseaux avec peu de files qui sont pratiquement de qualité similaire à des grands sous-réseaux.

Nous constatons donc, l'intérêt de ces techniques qui est de proposer un compromis entre la qualité de la borne et la complexité de calcul. Par contre, on notera que l'approche proposée

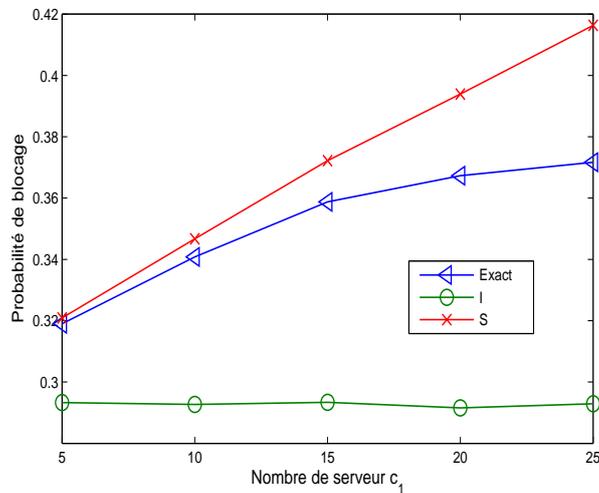


Figure V.2 – Influence de c_1 sur la probabilité de blocage de la file 3

est différente des autres types d'agrégations classiques telles que celles de Courtois [36] et aussi celles de [37, 38, 81] qui obtiennent que des bornes sur des mesures stationnaires, alors que les bornes stochastiques permettent d'obtenir à la fois des bornes sur les stationnaires et les transitoires.

Dans la suite, nous expliquons comment nous avons appliqué la comparaison stochastique à des cas de réseaux mobiles. L'objectif était essentiellement de comparer deux systèmes existants, pour comparer leurs mesures de performance.

1.2 Application de la comparaison stochastique sur des réseaux de communications

Dans le chapitre III, nous nous sommes intéressés à un problème de performabilité dans les réseaux de télécommunication. On étudie le modèle de perte d'Erlang (Erlang loss model) et on suppose que chaque canal peut tomber en panne pour être ensuite réparé au bout d'un certain temps. Le fait de prendre en compte à la fois les problèmes d'occupation et de fiabilité des ressources font que le système devient vite complexe. Trivedi a proposé, dans [47], une approximation de la probabilité de blocage calculée sur un modèle composite de performance et de fiabilité. Certes, son approximation est proche de la valeur exacte mais il n'est pas possible de dire, à tout instant, si c'est une borne supérieure ou inférieure.

Nous avons proposé de construire différents systèmes bornants (supérieurs et inférieurs). Ainsi, dans le tableau V.2, nous donnons différentes probabilités de blocage : valeurs exactes, bornantes

λ	Exact	S1	S2	I2	Tb*
1	8,95E-04	1,27E-03	9,67E-01	3,65E-05	9,26E-04
6	3,46E-01	4,66E-01	9,94E-01	2,98E-01	3,48E-01
11	6,06E-01	7,20E-01	9,97E-01	5,73E-01	6,07E-01
16	7,22E-01	8,14E-01	9,98E-01	6,98E-01	7,23E-01
21	7,86E-01	8,62E-01	9,98E-01	7,67E-01	7,86E-01
26	8,26E-01	8,90E-01	9,99E-01	8,11E-01	8,26E-01
31	8,53E-01	9,08E-01	9,99E-01	8,41E-01	8,54E-01
36	8,73E-01	9,22E-01	9,99E-01	8,62E-01	8,74E-01
41	8,88E-01	9,32E-01	9,99E-01	8,79E-01	8,89E-01
46	9,00E-01	9,39E-01	9,99E-01	8,92E-01	9,01E-01

Tableau V.2 – Probabilités de blocage : valeurs exactes, bornantes et approximatives de Trivedi

et approximatives de Trivedi. Nous remarquons que l’approximation de Trivedi donne de très bons résultats. L’avantage de notre approche est de proposer un encadrement (voir figure.V.3) de la mesure de performance à étudier. Ce qui est important car on peut situer la mesure de performance dans un intervalle.

Dans le chapitre IV, nous avons repris l’étude de Trivedi ([48]) qui a porté sur l’amélioration de la probabilité de dropping des appels en itinérance (handovers). Le mécanisme, MG, proposé dans cette étude a permis de diminuer la probabilité de dropping des appels handovers. Mais il ne propose rien contre la dégradation de la probabilité de blocage des nouveaux appels en raison de la réservation des canaux de gardes pour les appels en handovers.

Nous avons donc proposé un nouveau mécanisme, LMG, pour limiter à la fois la probabilité de dropping des appels handovers mais aussi la probabilité de blocage des nouveaux appels en tirant profit de l’occupation asymétrique des cellules dans le temps pour provoquer des départs forcés (handovers forcés) d’une cellule chargée vers des cellules moins chargées sous certaines conditions.

Nous avons apporté la preuve que les processus de Markov représentant les systèmes *MG* et *LMG* sont tels que $\{X^{LMG1}(t), t \geq 0\} \preceq_{st} \{X^{MG1}(t), t \geq 0\}$, dans l’hypothèse d’une seule cellule en utilisant la comparaison stochastique. Puis, nous avons validé l’analyse numérique par la simulation (voir Figures V.3 et V.4) afin de montrer que le système *LMG*, que nous avons proposé, améliore à la fois la probabilité de dropping des handovers et la probabilité de blocage des nouveaux appels.

Ensuite, nous avons considéré plusieurs cellules (Chapitre ?? section.2) afin de tenir compte des états de toutes les cellules et l’itinérance (trafic de handover) entre elles. Le système devient

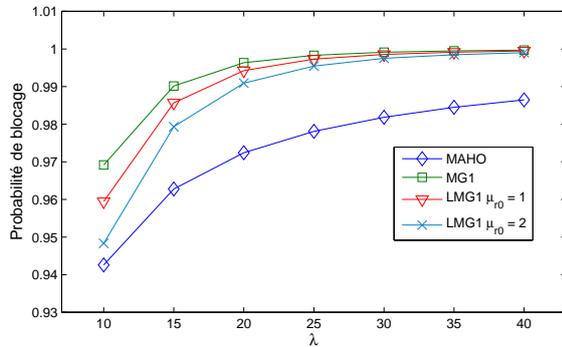


Figure V.3 – Probabilité de blocage sur un système monocellulaire.

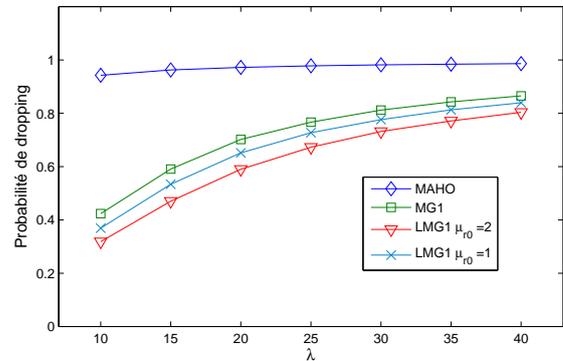


Figure V.4 – Probabilité de dropping sur un système monocellulaire.

donc, rapidement complexe à cause du nombre d'états et des événements (arrivées de nouveau appels voix, handovers, services, pannes et réparations).

Contrairement au cas où nous avons considéré une cellule unique, ici il est difficile de réaliser une comparaison stochastique des processus à cause du handover forcé qui ne se produit que dans le mécanisme *LMG*. Par contre, les résultats numériques confirment la pertinence de *LMG* puisqu'il permet d'obtenir des mesures de performances meilleures que dans le cas du mécanisme *MG* (voir Figure.V.5). Cependant, ces résultats ont été obtenus pour une configuration particulière de la charge du réseau. D'autres systèmes devraient être étudiés pour tirer des conclusions plus générales.

Enfin, ici aussi, nous avons pu mettre en évidence la nécessité du compromis entre la qualité des bornes, obtenues avec des sous-systèmes bornants plus ou moins grands, et la complexité (voir Figure.V.6).

Dans cette partie, nous avons vu qu'il est possible d'étudier des applications complexes par les méthodes d'analyse mathématiques telles que les méthodes de bornes. L'avantage d'un encadrement de mesures à la fois stationnaires et transitoires permet de déduire des mesures de performances dont les calculs sont autrement beaucoup trop difficiles à envisager.

A partir des résultats obtenus dans cette thèse, nous allons voir maintenant dans quelles voies nous aimerions diriger nos futures activités de recherche.

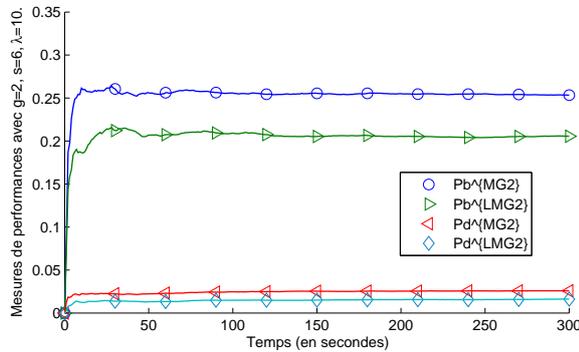


Figure V.5 – Mesures de performances pour un système peu chargé ($\lambda = 10$) avec une distribution uniforme du trafic.

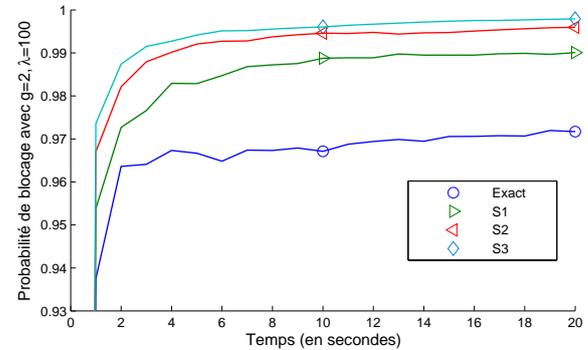


Figure V.6 – Probabilité de blocage avec un trafic non-uniforme.

2 Perspectives

Le formalisme du couplage est intéressant car en utilisant les événements, on se rapproche d'un formalisme de spécification de haut niveau. Il est donc naturel d'envisager l'application du couplage à des formalismes de haut niveau tels que les réseaux de Pétri et les automates stochastiques.

2.1 Couplage par fonction de projection : application à des formalismes de haut niveau

Comme nous l'avons vu, le couplage par fonction de projection appliqué à des modèles de réseaux de files d'attente permet de proposer des bornes sur des sous-réseaux. Il existe également des travaux appliqués aux réseaux d'automates stochastiques. Ainsi, dans [82], on applique les méthodes de bornes à des automates stochastiques représentant des chaînes de Markov de grande tailles. Les bornes sont obtenues en utilisant les informations relatives à un automate en isolation, puis elles sont raffinées en incluant d'autres automates. La qualité des bornes dépend du degré de dépendance entre les automates. Dans [83], on étudie des réseaux d'automates stochastiques "lumpables" (ou agrégeables). On suppose que les chaînes de Markov générées par ces automates ont cette propriété, ce qui simplifie le calcul de la distribution de probabilité du système. Dans [84], on étudie des réseaux de Pétri stochastiques dont le graphe de marquage (représentant des chaînes de Markov) ont des structures particulières de "lumpability" afin de réduire la taille de l'espace d'états. Par rapport à ces différentes études, les agrégations que nous avons proposées ne dépendent pas de la structure de la chaîne de Markov mais sont plus générales car elles permettent de construire des chaînes définies dans un espace d'états différent

de l'espace d'états initial et pas forcément un sous-espace d'états. C'est pourquoi, il nous paraît intéressant, et pas impossible, d'appliquer ce type d'agrégation directement à un niveau de spécification plus élevé tels que les réseaux d'automates ou de Pétri stochastiques.

Le couplage étant basé sur la comparaison des équations d'évolution en fonction des événements, nous l'avons appliquée dans cette thèse à différents systèmes. L'approche étant similaire à chaque fois, il serait intéressant de développer un algorithme permettant d'automatiser ces traitements.

2.2 Vers un algorithme de couplage de processus

On considère deux processus avec des événements se déclenchant dans chacun d'eux. Ainsi dans [85], nous avons développé une première partie de cet algorithme. On définit d'abord le processus couplé et l'ensemble K de ses valeurs. L'idée est de définir les points frontières et de considérer les événements qui font sortir le processus couplé de K par la première composante. On vérifie que la deuxième composante peut compenser par des événements de façon à ce que le processus couplé reste dans K . L'objectif est de considérer ces différents sauts et de comparer les taux de transition correspondants, de façon à vérifier que les taux qui compensent sont toujours supérieurs à ceux qui font sortir. Nous n'avons développé que la première partie de l'algorithme, c'est à dire la sortie par la première composante, il faudrait donc le faire également pour la seconde composante. De plus, nous avons considéré certaines hypothèses pour simplifier l'algorithme, il faudrait proposer un algorithme plus complet et plus général.

2.3 Estimation de l'erreur entre les bornes

Nous avons vu que la comparaison stochastique a permis de fournir des encadrements des mesures de performances exactes. Ce résultat est très intéressant par rapport à des approximations car nous avons comme résultat un intervalle de valeurs possibles pour la valeur exacte. Les valeurs numériques nous permettent de voir l'écart entre la borne supérieure et la borne inférieure. Par contre, nous n'avons pas obtenu pour l'instant de résultat théorique concernant cet écart. Ainsi, si P^S est une borne supérieure d'une mesure de performance, et P^I une borne inférieure, l'objectif est de trouver la valeur b tel que $|P^S - P^I| < b$. Il existe un certain nombre de travaux que nous aimerions exploiter.

Dans [86], Van Dijk propose une méthode d'estimation de l'erreur pour un réseau avec un nombre fini de stations avec pertes. Dans [87], Van Dijk propose des bornes sur les erreurs dans le cas des réseaux de Jackson finis. L'encadrement de l'erreur est essentiellement basé sur la théorie des modèles de Markov avec récompense (Markov Reward Model : MRM) et

l'estimation de l'erreur. Ainsi, les travaux qui existent dans ce domaine concernent soit des réseaux particuliers comme les réseaux de Jackson, ou soit des méthodes approximatives. Il est donc important de pouvoir proposer une solution générale pour l'encadrement de l'erreur dans le cas de bornes.

2.4 Choix de l'ordre sur l'espace d'états

Dans certains cas, le choix de l'ordre sur l'espace d'états n'est pas évident. Dans le cas de réseaux de files d'attente, on utilise souvent l'ordre composante par composante. Dans le cas général d'un système modélisé par une chaîne de Markov, il n'est pas toujours facile, à partir des états utilisés pour calculer l'indice de performance, de choisir cet ordre. Plus précisément, on peut définir l'ordre de telle façon que l'ensemble de ces états soit un ensemble croissant. Ainsi, dans [88] la construction d'un ordre sur l'espace d'états est déduit de la fonction de coût et des propriétés de monotonie de la matrice. Ainsi, la recherche d'un ordre sur l'espace d'états permettant de comparer les mesures de performance est une étape importante. Il serait intéressant de poursuivre pour voir l'impact de l'ordre choisi sur la qualité des bornes.

2.5 Comparaison de systèmes non-markoviens et semi-markoviens

La comparaison stochastique a été peu appliquée à des processus semi-markoviens ou non-markoviens. Toutefois, il existe un certain nombre d'articles concernant ces sujets là. Dans [89], la comparaison stochastique est appliquée à des processus semi-markoviens. Des conditions suffisantes sur les taux de transitions sont établies pour définir l'ordre fort entre les processus. De plus, il existe un certain nombre de travaux [90] consistant à appliquer la comparaison stochastique par fonctions de projections afin de comparer un processus non-markovien avec un processus markovien.

2.6 Ordres stochastiques faibles

La comparaison stochastique basée sur des ordres faibles (weak ou weak*) [53] peut être utile lorsque l'ordre fort (strong) n'existe pas entre les processus. Il serait important de comprendre quand on peut les utiliser. C'est à dire, de façon intuitive pourquoi dans certains cas un ordre existe et pas un autre. Il serait important (sans passer par le formalisme matriciel) d'arriver à deviner, à partir du comportement des systèmes, le type l'ordre pouvant être défini. Il en est de même pour la monotonie stochastique.

2.7 Applications futures

Les bornes stochastiques permettront certainement d'apporter des solutions à des applications actuelles et futures dans le domaine des réseaux et de l'informatique. Ainsi, dans le cloud computing la complexité du système, du point de vue de la taille mais aussi de la dynamique, rend difficile l'analyse par des méthodes classiques. En effet, l'informatique répartie pose un défi en termes de variation de la charge (demande de services, allocation de ressources) et de la disponibilité des capacités (machine physique par groupe d'utilisateurs, machines virtuelles par machine physique). Dans [91], une approche consistant à la décomposition du système en sous modèles (sous chaîne de Markov) est proposée afin d'analyser les performances.

Nous pouvons également recourir à la comparaison stochastique dans les technologies du Green Networking où des ressources peuvent être mises en veille quand la charge du réseau diminue [92]. La prise en compte de la mise en veille des ressources engendre une complexité du système. De plus, une analyse transitoire du système est importante pour évaluer la variation de la charge dans le temps en raison de la mise en veille des ressources. Et donc, les méthodes de bornes stochastiques peuvent apporter, dans ce contexte, des solutions intéressantes.

Annexe A

Publications

Journal :

- Hind Castel-Taleb, Idriss Ismael-Aouled et Nihal Pekergin
"Bounding aggregations for transient and stationary performance analysis of subnetworks", The Computer Journal 2011, Oxford Journals, Oxford University Press.

Articles :

- Hind Castel-Taleb, Idriss Ismael-Aouled et Nihal Pekergin
"Stochastic comparisons for performability of telecommunication systems", 17th International Conference on Analytical and Stochastic Modelling Techniques and Applications, ASMTA2010, 14-16 juin 2010, Cardiff, UK.
- Hind Castel-Taleb, Idriss Ismael-Aouled et Nihal Pekergin
"Bounding techniques for transient analysis of G-Networks with catastrophes", 5th International ICST Conference on Performance Evaluation Methodologies and Tools, VALUE-TOOLS 2011, 16-20 mai 2011, Cachan, France.
- Idriss Ismael-Aouled et Hind Castel-Taleb
"CAC and Forced Handoff for Mobile Network Performability", 19th International Conference on Analytical and Stochastic Modelling Techniques and Applications, ASMTA'12, 4-6 juin 2012, Grenoble, France.

Fast abstract :

- Hind Castel-Taleb et Idriss Ismael-Aouled
"Stochastic comparison of Markov processes through coupling by mapping functions", 8th International Conference on Quantitative Evaluation of SysTems, QEST11, 5-8 September 2011, Aachen, Germany.

Références bibliographiques

- [1] B. BAYNAT. *Théorie des files d'attente : Des chaînes de Markov aux réseaux à forme produit*. Réseaux et télécommunications. Hermes Science Publications (2000). 2, 4, 6, 32
- [2] KISHOR S. TRIVEDI, JOGESH K. MUPPALA, STEVEN P. WOOLET, AND BOUDEWIJN R. HAVERKORT. *Composite performance and dependability analysis*. Perform. Eval. **14**(3-4), 197–215 (1992). 3, 69
- [3] KISHOR S. TRIVEDI, XIAOMIN MA, AND SELVAMUTHU DHARMARAJA. *Performability modelling of wireless communication systems*. Int. J. Communication Systems **16**(6), 561–577 (2003). 3, 12, 69
- [4] JOHN F. MEYER. *On evaluating the performability of degradable computing systems*. IEEE Trans. Computers **29**(8), 720–731 (1980). 3, 69
- [5] J.C.M. BAETEN. *Applications of process algebra*. Cambridge University Press (1990). 4
- [6] K. JENSEN. *Coloured Petri Nets : Basic Concepts, Analysis Methods and Practical Use*. Monographs in Theoretical Computer Science. Springer (1997). 4
- [7] TADAO MURATA. *Petri nets : Properties, analysis and applications*. pages 541–580 April (1989). NewsletterInfo : 33Published as Proceedings of the IEEE, volume 77, number 4. 4
- [8] ROBIN SAHNER, K.S. TRIVEDI, AND A. PULIAFITO. *Performance and Reliability Analysis of Computer Systems : An Example-Based Approach Using the Sharpe Software Package*. Kluwer Academic Publishers (1996). 5
- [9] CARL ADAM PETRI. *Communication with automata*. Thèse de Doctorat, Universität Hamburg (1966). 5
- [10] JAMES LYLE PETERSON. *Petri Net Theory and the Modeling of Systems*. Prentice Hall PTR, Upper Saddle River, NJ, USA (1981). 5

- [11] WOLFGANG REISIG. *Petri nets : an introduction*. Springer-Verlag New York, Inc., New York, NY, USA (1985). 5
- [12] M.K. MOLLOY. *On the integration of delay and throughput measures in distributed processing models*. Tesis doctorales. University of California, Los Angeles (1981). 5
- [13] C. RAMCHANDANI. *Analysis of asynchronous concurrent systems bu timed Petri Nets*. PhD Thesis. Massachusetts Institute of Technology, Cambridge, MA, USA (1974). 5
- [14] JOSEPH SIFAKIS. *Use of petri nets for performance evaluation*. pages 75–93 (1977). 5
- [15] S.O. NATKIN. *Les réseaux de PETRI stochastiques et leur application à l'évaluation des systèmes informatiques*. Tesis doctorales. CNAM (1980). 5
- [16] MICHAEL K. MOLLOY. *Discrete time stochastic petri nets*. IEEE Trans. Softw. Eng. **11**(4), 417–423 April (1985). 5
- [17] GIANFRANCO CIARDO. *Discrete-time markovian stochastic petri nets*. Technical report (1995). 5
- [18] MARCO AJMONE MARSAN, GIANFRANCO BALBO, AND GIANNI CONTE. *A class of generalised stochastic petri nets for the performance evaluation of multiprocessor systems*. pages 198–199 (1983). 5
- [19] M.A. MARSAN, G. BALBO, AND G. CONTE. *Performance models of multiprocessor systems*. (1986). 5, 6
- [20] B. PLATEAU. *De l'évaluation du parallélisme et de la synchronisation*. Thèse doctorat. (1984). 5, 6
- [21] BRIGITTE PLATEAU. *On the stochastic structure of parallelism and synchronization models for distributed algorithms*. pages 147–154 (1985). 5, 6
- [22] P.H.L. FERNANDES. *Méthodes numériques pour la solution de systèmes Markoviens à grand espace d'États*. Thèse doctorat. Institut national polytechnique de Grenoble (1998). 6
- [23] K. ATIF. *Modélisation du parallélisme et de la synchronisation*. Thèse doctorat. INPG (1992). 6
- [24] O. GUSAK, T. DAYAR, AND J.-M. FOURNEAU. *Discrete-time stochastic automata networks and their efficient analysis*. Performance Evaluation. **53**(1), 43–69 (2003). 6
- [25] J D C. LITTLE. *A proof for the queuing formula : lequil;/?w*. (1961). 6
- [26] E. GELENBE AND I. MITRANI. *Analysis and synthesis of computer systems*. (1980). 6
- [27] S.S. LAVENBERG. *Computer performance modeling handbook*. Notes and reports in computer science and applied mathematics. Academic Press (1983). 6

- [28] EROL GELENBE. *Random neural networks with negative and positive signals and product form solution*. Neural Comput. **1**(4), 502–510 December (1989). [7](#)
- [29] PETER J. DENNING AND JEFFREY P. BUZEN. *The operational analysis of queueing network models*. ACM Comput. Surv. **10**(3), 225–261 September (1978). [8](#)
- [30] J.R. JACKSON. *Jobshop-like queueing systems*. Management Science **10**, 131–142 (1963). [9](#)
- [31] W. J. GORDON AND G. F. NEWELL. *Closed queueing systems with exponential servers*. (1967). [9](#)
- [32] FOREST BASKETT, K. MANI CHANDY, RICHARD R. MUNTZ, AND FERNANDO G. PALACIOS. *Open, closed, and mixed networks of queues with different classes of customers*. J. ACM **22**(2), 248–260 (1975). [9](#)
- [33] G. BOLCH, S. GREINER, H. DE MEER, AND K.S. TRIVEDI. *Queueing Networks and Markov Chains : Modeling and Performance Evaluation with Computer Science Applications*. Wiley-Interscience publication. John Wiley & Sons (2006). [10](#), [69](#)
- [34] NICO M. VAN DIJK. *Analytic comparison results for communication networks*. Computer Communications **21**(17), 1495–1508 (1998). [10](#)
- [35] S. FDIDA AND G. PUJOLLE. *Modèles de systèmes et de réseaux*. Number vol. 1 in Modèles de systèmes et de réseaux. Eyrolles (1989). [10](#), [32](#)
- [36] P. J. COURTOIS. *Decomposability : queueing and computer system applications*. (1977). [10](#), [126](#)
- [37] JUAN A. CARRASCO. *Bounding steady-state availability models with group repair and phase type repair distributions*. Perform. Eval. **35**(3-4), 193–214 May (1999). [11](#), [126](#)
- [38] STEPHANIE MAHÉVAS AND GERARDO RUBINO. *Bound computation of dependability and performance measures*. IEEE Trans. Comput. **50**(5), 399–413 May (2001). [11](#), [126](#)
- [39] A. MÜLLER AND P.D. STOYAN. *Comparison Methods for Stochastic Models and Risks*. Wiley Series in Probability and Statistics. John Wiley & Sons (2002). [11](#), [18](#), [20](#), [22](#), [23](#), [24](#), [28](#), [29](#)
- [40] WILLIAM A. MASSEY. *Stochastic ordering for markov processes on partially ordered spaces*. Math. Oper. Res. **12**(2) May (1987). [11](#), [18](#), [20](#), [21](#), [22](#), [23](#), [27](#), [28](#), [30](#)
- [41] T. LINDVALL. *Lectures on the Coupling Method*. Dover Books on Mathematics Series. Dover Publications (2002). [11](#)
- [42] H. CASTEL-TALEB. *Agrégation de trafic et de systèmes markoviens pour garantir la QoS des réseaux complexes*. Habilitation à Diriger Les Recherches de l'Université Pierre et Marie Curie-Paris VI. (Mai 2011). [11](#)

- [43] M. DOISY. *A coupling technique for stochastic comparison of functions of markov processes*. JAMDS **4**(1), 39–64 (2000). [11](#), [23](#), [25](#), [33](#), [52](#)
- [44] LAURENT TRUFFET. *Reduction techniques for discrete-time markov chains on totally ordered state space using stochastic comparisons*. Journal of Applied Probability **37**(3), pp. 795–806 (2000). [12](#), [33](#), [52](#)
- [45] JEAN-MICHEL FOURNEAU AND NIHAL PEKERGIN. An algorithmic approach to stochastic bounds. In *Performance Evaluation of Complex Systems : Techniques and Tools, Performance 2002, Tutorial Lectures*, pages 64–88, London, UK, UK (2002). Springer-Verlag. [12](#), [18](#)
- [46] JAMES LEDOUX AND LAURENT TRUFFET. *Markovian bounds on functions of finite markov chains*. Advances in Applied Probability **33**(2) (2001). [12](#)
- [47] KISHOR S. TRIVEDI. *Probability and statistics with reliability, queuing and computer science applications*. John Wiley and Sons Ltd., Chichester, UK, 2nd edition edition (2002). [12](#), [68](#), [70](#), [72](#), [90](#), [91](#), [126](#)
- [48] B. B. MADAN, S. DHARMARAJA, AND K. S. TRIVEDI. *Combined Guard Channel and Mobile-Assisted Handoff for Cellular Networks*. IEEE Transactions on Vehicular Technology **57**, 502–510 (2008). [14](#), [88](#), [95](#), [127](#)
- [49] M. SHAKED AND J. G. SHANTHIKUMAR. *Stochastic Orders and Their Applications*. Academic Press, Boston (1994). [18](#)
- [50] A. MULLER AND D. STOYAN. *Comparison methods for Stochastic Models and Risks*. J. Wiley and son in Probability and Statistics (2002). [20](#)
- [51] M. SHAKED AND J.G. SHANTHIKUMAR. *Stochastic orders and their applications*. Probability and mathematical statistics. Academic Press (1994). [20](#)
- [52] H. CASTEL-TALEB. *Bornes stochastiques pour l'évaluation des réseaux informatiques*. Thèse de doctorat de l'Université de Paris 6. (Janvier 1996). [20](#)
- [53] WILLIAM A. MASSEY. *A family of bounds for the transient behavior of a jackson network*. Journal of Applied Probability **23**(2), 543–549 (1986). [21](#), [28](#), [29](#), [32](#), [52](#), [131](#)
- [54] T. KAMAE AND U. KRENGEL. *Stochastic inequalities on partially ordered spaces*. Annals of Probability **5**, 899–912 (1977). [24](#)
- [55] TORGNY LINDVALL. *Lectures on the Coupling Method*. Dover, New York (2002). [24](#), [29](#)
- [56] TORGNY LINDVALL. *Stochastic monotonicities in jackson queueing networks*. Probability in the Engineering and Informational Sciences **11**(01), 1–9 (1997). [25](#), [29](#), [32](#)
- [57] WILLIAM J. STEWART. *Introduction to the numerical solution of Markov Chains*. Princeton University Press (1994). [32](#)

- [58] HIND CASTEL-TALEB, LYNDA MOKDAD, AND NIHAL PEKERGIN. *Aggregated bounding markov processes applied to the analysis of tandem queues*. page 43 (2007). 33, 52
- [59] ANTONIS ECONOMOU. *Necessary and sufficient conditions for the stochastic comparison of jackson networks*. *Probab. Eng. Inf. Sci.* **17**(1), 143–151 January (2003). 40
- [60] E. GELENBE. *Product-form queueing networks with negative and positive customers*. *Journal of Applied Probability* pages 656–663 (1991). 51
- [61] E. GELENBE. *G-networks with instantaneous customer movement*. *Journal of Applied Probability* **3**(20), 742–748 (1993). 51, 52
- [62] E. GELENBE. *G-networks with signals and batch removal*. *Probability in the Engineering and Informational Sciences* **7**(03), 335–342 (1993). 51, 52, 53
- [63] X. CHAO, M. MIYAZAWA, AND M. PINEDO. *Queueing Networks : Customers, Signals, and Product Form Solutions*. Wiley-Interscience series in systems and optimization. Wiley (1999). 51
- [64] E. GELENBE. *G-networks : a unifying model for neural and queueing networks*. *Annals of Operations Research* **48**(5), 433–461 (1994). 52
- [65] JESUS R. ARTALEJO. *G-networks : A versatile approach for work removal in queueing networks*. *European Journal of Operational Research* **126**(2), 233–249 (2000). 52
- [66] E. GELENBE. *Dealing with software viruses : a biological paradigm*. information security technical report **12**(4), 242–250 (2007). 52, 63
- [67] HIND CASTEL-TALEB AND NIHAL PEKERGIN. *Stochastic bounds on the transient behaviors of the g-network*. page 23 (2002). 52
- [68] HIND CASTEL-TALEB, JEAN-MICHEL FOURNEAU, AND NIHAL PEKERGIN. *Stochastic comparisons applied to g-networks with catastrophes*. pages 57–60 (2010). 52
- [69] HIND CASTEL-TALEB, IDRIS ISMAEL-AOULED, AND NIHAL PEKERGIN. *Stochastic Comparisons for Performability of Telecommunication Systems*. pages 189–203 (2010). 70
- [70] BOUDEWIJN R. HAVERKORT AND RICK HARPER. *Performance and dependability techniques and tools*. *Perform. Eval.* **44**(1-4), 1–4 (2001). 70
- [71] MOHAMED HOSSAM AHMED. *Call admission control in wireless networks : A comprehensive survey*. *IEEE Communications Surveys Tutorials* **7**, 50–69 (2005). 87, 89
- [72] DAEHYOUNG HONG AND STEPHEN S. RAPPAPORT. *Traffic Model and Performance Analysis for Cellular Mobile Radio Telephone Systems with Prioritized and Non-Prioritized Handoff Procedures*. In *IEEE International Conference on Communications*, pages 1146–1150 (1986). 87, 89, 91

- [73] WHA SOOK JEON AND DONG GUEN JEONG. *Call admission control for mobile multimedia communications with traffic asymmetry between uplink and downlink*. IEEE Transactions on Vehicular Technology **50**, 59–66 (2001). [87](#), [89](#)
- [74] IDRIS ISMAEL AOULED AND HIND CASTEL-TALEB. *Combined cac and forced handoff for mobile network performability*. pages 76–90 (2012). [88](#)
- [75] MAJID GHADERI AND RAOUF BOUTABA. *Call admission control in mobile cellular networks : a comprehensive survey : Research articles*. Wirel. Commun. Mob. Comput. **6**(1), 69–93 February (2006). [89](#)
- [76] MICHEL MOULY AND MARIE-BERNADETTE PAUTET. *The GSM System for Mobile Communications*. Telecom Publishing (1992). [89](#)
- [77] OLIVER T. W. YU AND VICTOR C. M. LEUNG. *Adaptive Resource Allocation for Prioritized Call Admission over an ATM-Based Wireless PCN*. IEEE Journal on Selected Areas in Communications **15**, 1208–1225 (1997). [89](#)
- [78] NOVELLA BARTOLINI AND IMRICH CHLAMTAC. *Improving call admission control procedures by using hand-off rate information*. Wireless Communications and Mobile Computing **1**, 257–268 (2001). [89](#)
- [79] S.H. CHO, S.J. YANG, S. HEU, AND S.H. PARK. QoS Oriented Bandwidth Management Scheme for Stable Multimedia Services on the IMT2000 Networks. In *International Symposium on Information Technology*, pages 289–294 (2000). [89](#)
- [80] T.T. NIELSEN AND J. WIGARD. *Performance Enhancements in a Frequency Hopping GSM Network*. Springer (2000). [91](#)
- [81] P. J. COURTOIS AND P. SEMAL. *Computable bounds for conditional steady-state probabilities in large markov chains and queueing models*. IEEE Journal on Selected Area in Communication **4**(6), 926–937 September (1986). [126](#)
- [82] PETER BUCHHOLZ. *An iterative bounding method for stochastic automata networks*. Perform. Eval. **49**(1/4), 211–226 (2002). [129](#)
- [83] OLEG GUSAK, TUGRUL DAYAR, AND JEAN MICHEL FOURNEAU. *Iterative disaggregation for a class of lumpable discrete-time stochastic automata networks*. Performance Evaluation **53**, 43–69 (2003). [129](#)
- [84] SOUHEIB BAARIR, MARCO BECCUTI, CLAUDE DUTHEILLET, GIULIANA FRANCESCHINIS, AND SERGE HADDAD. *Lumping partially symmetrical stochastic models*. Performance Evaluation **68**, 21–44 (2011). [129](#)
- [85] HIND CASTEL-TALEB AND I. ISMAEL-AOULED. *Stochastic comparison of markov processes through coupling by mapping functions*. 8th International Conference on Quantitative Evaluation of Systems (Fast Abstract,QEST 2011, Aachen Germany) pages 5–6 September (2011). [130](#)

- [86] NICO M. VAN DIJK AND HAYRI KOREZLIOGLU. *On product form approximations for communication networks with losses : error bounds*. Ann. Oper. Res. **35**(1-4), 69–94 April (1992). [130](#)
- [87] NICO M. VAN DIJK. *Error bounds for state space truncation of finite jackson networks*. European Journal of Operational Research **186**(1), 164–181 (2008). [130](#)
- [88] DAVID DALY, PETER BUCHHOLZ, AND WILLIAM H. SANDERS. *A preorder relation for markov reward processes*. Statistics and Probability Letters **77**(11), 1148 – 1157 (2007). [131](#)
- [89] D. SONDERMAN. *Comparing semi-markov processes*. Mathematics of Operations Research **1** (Feb. 1980). [131](#)
- [90] W. WITT. *Stochastic comparisons for non-markov processes*. Mathematics of Operations Research **11**(4) (Nov. 1986). [131](#)
- [91] RAHUL GHOSH, FRANCESCO LONGO, VIJAY K. NAIK, AND KISHOR S. TRIVEDI. *Modeling and performance analysis of large scale iaas clouds*. Future Generation Computer Systems (0) (2012). [132](#)
- [92] G. SHEN, J. DENG, AND P. HO. *Green backbone optical networks : The way forward*. In *Information, Communications and Signal Processing (ICICSP) 2011 8th International Conference on*, pages 1–5. IEEE (2011). [132](#)

Résumé

Notre travail s'inscrit dans le thème plus global de la modélisation et de l'évaluation des systèmes multidimensionnels et adresse plus particulièrement la prédiction des indicateurs de performabilité qui réfère aux performances des systèmes en présence de pannes et de réparations. L'analyse exacte, de ces derniers, est très difficile voire impossible si la distribution des probabilités n'a pas de solution évidente (par exemple une forme produit) à cause de l'explosion de l'espace d'états. Pour contourner ce problème, et parce que souvent on ne s'intéresse qu'à une partie du système (nœud ou un chemin), nous utilisons l'agrégation bornante basée sur le couplage par fonction de projection sur des espaces d'états plus petits. La méthode du couplage, basée sur la comparaison des trajectoires des processus en fonction des événements qui se déclenchent dans les systèmes, nous permet de construire à partir du système original des sous-systèmes bornants plus simple à analyser. Nous pouvons, alors, proposer des mesures de performances bornantes (supérieures et inférieures) de la mesure exacte. Ces bornes, ainsi obtenues, sont valables pour les régimes stationnaires mais aussi transitoires. Nous mettons également en évidence l'influence des paramètres des systèmes, comme le nombre de serveurs ou la charge, sur la qualité des bornes.

Enfin, nous proposons un nouveau mécanisme de contrôle d'admission (CAC) dans les réseaux mobiles. Nous l'avons comparé à d'autres mécanismes et apporté la preuve stochastique que celui-ci améliore les probabilités de dropping des handovers et de blocage des nouveaux appels dans le cas d'un réseau monocellulaire. De plus, les résultats numériques nous ont permis de valider ce mécanisme sur un réseau multicellulaire.

Mots clés : Performabilité, comparaison stochastique, couplage, réseaux de files d'attente, G-Net, $M/M/c_i/k_i$, CAC, espaces d'états

Abstract

Our work fits into the broader topic of modeling and evaluation of multidimensional systems and more specifically address the prediction of performability indicators that refer to system performance in the presence of failures and repairs. The exact analysis, of such systems, is very difficult or impossible if the probability distribution has no obvious solution (eg a product form) because of the explosion of the state space. To circumvent this problem, and because we often only interested in one part of the system (node or a path), we use bounds aggregation based on the coupling by mapping function on smaller states space. Coupling methodology, based on the comparison of the trajectories of processes based on events that trigger systems, allows us to construct, from the original, sub-systems easier to analyze. We can, then, propose bounds (upper and lower) on the performances measures that we are looking for. Thus bounds remain valid for the steady-states but also for transient regimes. We highlight also the influence of the parameters of the systems, as the number of servers or load, on the precision of the bounds.

Finally, we propose a new mechanism of call admission control (CAC) in mobile networks. We compared it with other mechanisms and brought stochastic proof that it improves the handed calls dropping probability and new call blocking probability in the case of a single network. And the numerical results have allowed us to validate this mechanism on a network with cluster of cells.

Keywords : Performability, stochastic comparison, coupling, queuing networks, G-Net, $M/M/c_i/k_i$, CAC, state space