



UNIVERSITÉ BADJI MOKHTAR-ANNABA, ALGÉRIE

UNIVERSITÉ PARIS 1, PANTHÉON-SORBONNE, FRANCE

ECOLE DOCTORALE SCIENCES MATHÉMATIQUES DE PARIS

CENTRE

LABORATOIRE LAPS ET LANOS, ANNABA

LABORATOIRE SAMM, PARIS 1

THÈSE

présentée en vue d'obtenir le grade de Docteur, spécialité
« Mathématiques Appliquées »

par

ARRAR Nawel Khadidja

PROBLÈMES DE CONVERGENCE, OPTIMISATION D'ALGORITHMES ET ANALYSE STOCHASTIQUE DE SYSTÈMES DE FILES D'ATTENTE AVEC RAPPELS

Thèse soutenue le 10/septembre/2012 devant le jury composé de :

M.	SISSAOUI HOCINE	Université Badji Mokhtar - Annaba	(Président)
M ^{me}	DJELLAB NATALIA	Université Badji Mokhtar - Annaba	(Directeur)
M.	BAILLON JEAN-BERNARD	Université Paris1, Panthéon - Sorbonne	(Directeur)
M.	MOHDEB ZAHER	U. Mentouri, Constantine	(Rapporteur)
M.	AISSANI AMAR	U. S. T. H. B. Alger	(Rapporteur)
M.	BLOT JOËL	Université Paris1, Panthéon - Sorbonne	(Examinateur)
M ^{me}	LASKRI YAMINA	Université Badji Mokhtar - Annaba	(Examinateur)

REMERCIEMENTS

Ces travaux s'inscrivent dans le cadre de l'accord d'une cotutelle signée entre l'université de Paris 1 Panthéon -Sorbonne et l'université Badji Mokhtar-Annaba.

Tout d'abord, Je tiens à exprimer mes sincères remerciement au Professeur Natalia DJELLAB pour m'avoir confié ce sujet de thèse, pour m'avoir encadré durant ces années avec beaucoup de patience, de disponibilité et de professionnalisme, et pour m'avoir transmis sa passion pour le domaine des systèmes de files d'attente,... Ce fut très sincèrement un réel plaisir de travailler à ses côtés pendant ces années et pour m'avoir ainsi permis de réaliser cette thèse.

Je remercie très chaleureusement le Professeur Jean-Bernard BAILLON pour m'avoir accueilli au sein du laboratoire SAMM et pour son encadrement. Je tiens également à le remercier pour sa disponibilité, la confiance qu'il m'a accordé et pour ses conseils et ses commentaires fort utiles qui ont fortement enrichi ma formation.

Je remercie le Professeur Hocine SISSAOUI d'avoir accepter de présider le jury. J' exprimer ma gratitude aux Professeur Amar AÏSSANI et Zaher MOH-DEB qui ont accepté de rapporter cette thèse et pour l'intérêt qu'ils y ont porté. Je remercie également les Professeurs Joël BLOT et Yamina LASKRI pour avoir bien voulu faire partie du jury.

Je voudrais également associer mes meilleurs remerciements à mon collègue et époux pour son soutien moral et pour ses remarques qui ont eu une contribution importante pour ma réussite dans cette thèse.

Je tiens à remercier mes sincères amis Chafia et son époux Foued pour leur accueil chaleureux durant mes stages à Paris, ils ont rendu mes séjours encore plus agréables.

Je remercie mes collègues et amis membres des Laboratoires LaPS et LANOS pour l'amitié qu'ils m'ont témoigné tout au long de ces années de thèse en particulier Docteur Assia Chadli, Docteur Fatima Zohra Benmostefa et Professeur Nacira Seddik-Ameur. Mes remerciements s'adressent également aux membres de l'équipe SAMM pour leur accueil et leur soutien plus particulièrement je remercie Tsirizo Rabenoro et Souheila pour leur gentillesse.

Je remercie tout mes amis pour leur soutien moral : Fahima, Lylia et Nacira ...

Et pour finir, merci à toutes les personnes que j'ai oubliées de citer et qui m'ont permis de mener à bien cette thèse.

TABLE DES MATIÈRES

TABLE DES MATIÈRES	<i>vii</i>
LISTE DES FIGURES	<i>x</i>
LISTE DES TABLEAUX	<i>xi</i>
INTRODUCTION	1
1 MODÉLISATION STOCHASTIQUE	5
1.1 INTRODUCTION	5
1.2 PROPRIÉTÉS DES ALGORITHMES STOCHASTIQUES	6
1.3 ALGORITHMES GÉNÉTIQUES	7
1.3.1 Modélisation par chaîne de Markov	7
1.4 RECUIT SIMULÉ	9
1.4.1 Convergence du recuit simulé	9
1.5 FILES D'ATTENTE	10
1.5.1 Modélisation des systèmes de files d'attente par chaînes de Markov	11
CONCLUSION	11
2 SYSTÈMES DE FILES D'ATTENTE CLASSIQUES	13
2.1 DESCRIPTION DU PHÉNOMÈNE D'ATTENTE	13
2.2 ANALYSE MATHÉMATIQUE D'UN SYSTÈME DE FILES D'ATTENTE	15
2.3 TYPES DE MODÈLES	16
2.3.1 Modèles markoviens	16
2.3.2 Modèles non markoviens	16
2.4 CARACTÉRISTIQUES D'UN SYSTÈME DE FILES D'ATTENTE	17
2.5 MODÈLE D'ATTENTE M/G/1	18
2.5.1 Description du modèle	18
2.5.2 Chaîne de Markov induite	18
2.5.3 Mesures de performance	21
2.6 MODÈLE D'ATTENTE M ^X /G/1	21
2.6.1 Description du modèle	21
2.6.2 Analyse du modèle	21

2.6.3	<i>Chaîne de Markov induite</i>	22
2.6.4	<i>Approche alternative à l'analyse du modèle $M^X/G/1$</i>	22
2.6.5	<i>Mesures de performance</i>	23
	CONCLUSION	24
3	SYSTÈMES DE FILES D'ATTENTE AVEC RAPPELS	25
3.1	INTRODUCTION	25
3.2	MODÈLE D'ATTENTE $M/G/1$ AVEC RAPPELS	26
3.2.1	<i>Description du modèle</i>	26
3.2.2	<i>Chaîne de Markov induite</i>	27
3.2.3	<i>Distribution stationnaire de l'état du système</i>	30
3.2.4	<i>Mesures de performance</i>	34
3.3	MODÉLISATION DE L'IMPATIENCE	35
3.3.1	<i>Description du modèle à un serveur</i>	35
3.3.2	<i>Distribution stationnaire de l'état du système</i>	35
3.3.3	<i>Mesures de performance</i>	38
3.4	SYSTÈMES DE FILES D'ATTENTE $M^X/G/1$ AVEC RAPPELS ET GROUPES IMPATIENTS	38
3.4.1	<i>Description du modèle</i>	39
3.4.2	<i>Chaîne de Markov induite</i>	40
3.4.3	<i>Distribution stationnaire de l'état du système</i>	45
3.4.4	<i>Mesures de performance</i>	49
3.4.5	<i>Exemples d'application</i>	50
	CONCLUSION	51
4	PROPRIÉTÉ DE DÉCOMPOSITION STOCHASTIQUE POUR LE NOMBRE DE CLIENTS DANS LE SYSTÈME	53
4.1	INTRODUCTION	53
4.2	PROPRIÉTÉ DE DÉCOMPOSITION STOCHASTIQUE	54
4.2.1	<i>Modèles d'attente avec vacances</i>	54
4.2.2	<i>Modèles d'attente avec rappels</i>	55
4.3	PROPRIÉTÉ DE DÉCOMPOSITION STOCHASTIQUE DU MODÈLE $M^X/G/1$ AVEC RAPPELS ET IMPATIENCE	58
	CONCLUSION	59
5	COMPORTEMENT ASYMPTOTIQUE DU NOMBRE DE CLIENTS DANS LE MODÈLE $M^X/G/1$ AVEC RAPPELS ET CLIENTS IM- PATIENTS	61
5.1	INTRODUCTION	61

5.2	COMPORTEMENT ASYMPTOTIQUE DU NOMBRE DE CLIENTS EN ORBITE SOUS UN TAUX DE TRAFIC INTENSE	62
5.3	COMPORTEMENT ASYMPTOTIQUE DU NOMBRE DE CLIENTS EN ORBITE SOUS UN TAUX DE RAPPELS FAIBLE	65
5.4	COMPORTEMENT ASYMPTOTIQUE DU SYSTÈME SOUS UN TAUX DE RAPPELS ÉLEVÉ	69
5.5	APPLICATION NUMÉRIQUE	70
	CONCLUSION	74
	CONCLUSION GÉNÉRALE	77
	BIBLIOGRAPHIE	79

LISTE DES FIGURES

2.1	Représentation schématique d'une file d'attente classique . . .	13
3.1	Graphe des transitions du modèle $M/G/1$ avec rappels . . .	32
3.2	Représentation schématique d'un système de files d'attente avec rappels	39
3.3	Graphe des transitions du modèle $M^X/G/1$ avec rappels et impatience	46
5.1	Variation de α en fonction de θ	71
5.2	Distribution du nombre de clients en orbite pour $\theta = 0.2$. . .	72
5.3	Distribution du nombre de clients en orbite pour $\theta = 1$. . .	72
5.4	Distribution du nombre de clients en orbite pour $\theta = 2$. . .	72
5.5	Distribution du nombre de clients en orbite pour $\theta = 10$. . .	72

Liste des tableaux

5.1	Paramètres de la loi Gamma suivant la distribution du temps de service	71
5.2	Moyenne et variance de la loi Normale suivant la distribution du temps de service	73
5.3	Borne inf et borne sup des inégalités vérifiées quand θ tend vers ∞	73
5.4	Longueur des intervalles de convergence de la distribution $\pi_n(\theta)$ vers la distribution $\pi_n(\infty)$	74

INTRODUCTION

*L*A théorie des files d'attente, ou des queues, est un des outils analytiques les plus puissants pour la modélisation de systèmes de logistiques et de communication. Cette théorie a pour objet l'étude de systèmes et réseaux où des entités, appelées clients, cherchent à accéder à des ressources, généralement limitées, afin d'en obtenir un service. La demande concurrente d'une même ressource par plusieurs clients engendre des délais dans la réalisation des services et la formation de file de clients désireux d'accéder à une ressource indisponible. L'analyse théorique de tels systèmes permet d'établir à l'avance les performances de l'ensemble, d'identifier les éléments critiques ou, encore, d'appréhender les effets d'une modification des conditions de fonctionnement.

Ce domaine de recherche, né en 1917, des travaux de l'ingénieur électricien danois Erlang sur la gestion des réseaux téléphoniques de Copenhague entre 1909 et 1920. Par la suite, les files d'attente ont été utilisées dans la modélisation des systèmes de production et des systèmes informatiques. Initialement, leur utilisation était très limitée par le fait qu'elles faisaient vite appel à une puissance de calcul non disponible à cette époque. Ainsi, pendant longtemps, c'est surtout des développements de nature théorique qui ont vu le jour dans le domaine. Le sujet a pourtant inspiré et continue à inspirer, de nombreux chercheurs comme en témoignent les nombreuses publications parues à ce jour dans le domaine.

Avec l'automatisation des systèmes technologiques, notamment l'avènement des systèmes de production flexibles, les convoyeurs automatiques, etc..., les applications réclamant le traitement par une telle théorie se sont également multipliées. Ce n'est pourtant que rarement que les files d'attente apparaissent de manière isolée. Le plus souvent ce sont des réseaux entiers avec un nombre élevé de serveurs interconnectés qu'il s'agit de modéliser. La grande percée dans le domaine est venue avec la découverte en 1957, par Jackson, d'une manière particulièrement élégante et efficace de traiter certains réseaux de files d'attente. C'est grâce aux apports des mathématiciens Khintchine, Palm, Kendall, Pollaczek et Kolmogorov que la théorie s'est vraiment développée.

La théorie des files d'attente classique offre deux possibilités pour résoudre le conflit qui apparaît lorsqu'un client arrive dans le système à serveur unique et trouve le serveur occupé : soit il quitte le système sans recevoir le service (modèle

d'Erlang à demandes refusées), soit il prend place dans une file d'attente. Une possibilité alternative est que le client qui ne peut être servi, libère l'espace du serveur mais après une durée de temps aléatoire revient au système pour répéter sa demande de service. Entre les appels successifs, le client en question se trouve en "orbite". Un tel système est appelé système de files d'attente avec "rappels".

Cette branche de la théorie des files d'attente s'applique dans la modélisation stochastique de différents problèmes de télécommunication, et de réseaux informatiques. Les systèmes de files d'attente avec rappels peuvent être utilisés pour résoudre des problèmes pratiques, tels que l'analyse du comportement des abonnés dans les réseaux téléphoniques, l'évitement de collision dans les réseaux locaux, l'analyse du temps d'attente pour accéder à la mémoire sur les disques magnétiques, ... (20, 112). Par conséquent, de nombreux travaux ont été publiés dans des Journaux spécialisés en probabilités appliquées et modèles stochastiques, statistiques et recherche opérationnelle, télécommunication et ingénierie industrielle, et informatique. Le grand intérêt de ce domaine est confirmé par l'organisation d'une série de workshops sur les systèmes de files d'attente avec rappels : Madrid (1998), Minsk (1999), Amsterdam (2000), Cochin (2002), Seoul (2004), Miraflores de la Sierra (2006), Athens (2008) et Beijing (2010). A cet effet, quelques revues de renommées internationales ont dédié des numéros spéciaux ; c'est le cas du journal *Annals of Operation Research* (23), *European Journal of Operation Research* (32), *Mathematical and Computer Modelling* (22), *Queueing Systems* (110) et *Top* (21).

Parmi les premières contributions sérieuses sur les modèles d'attente avec rappels, on trouve celles de Cohen (1957) (50), de Eldin (1967) (55), de Hashida et Kawashima (1979) (75) et de Lubacz et Roberts (1984) (97). Les progrès récents sont résumés dans les articles de synthèse de Aïssani (1994) (2), Kulkani et Liang (1997) (89), Templeton (1999) (111) et dans les monographies de Falin et Templeton (1997) (65), Artalejo et Gómez-Corral (2008) (31), Gómez-Corral et Ramalhoto (2000) (74), Rodrigo (2006) (106) et dans les travaux bibliographiques de Artalejo (1999 et 2010) (16), (17).

La décomposition stochastique est une des principales approches utilisées ces dernières années pour la résolution de modèles complexes. La propriété de décomposition stochastique est une particularité des modèles d'attente où le serveur est occasionnellement "indisponible" pour certain intervalles de temps (systèmes avec vacances) (54). Des résultats fondamentaux relatifs à cette approche ont été établis par Yang et Templeton (1987) (115), ainsi que par Yang et al. (1994) (117), et ceci pour le modèle avec rappels de type $M/G/1$ dans le cas de distributions exponentielle et générale du temps inter-rappels. Il est prouvé que la fonction génératrice du nombre de clients dans le système $M/G/1$ avec rappels est le produit de deux

fonctions génératrices : la fonction génératrice pour le nombre de clients dans le système $M/G/1$ classique et la fonction génératrice pour le nombre de clients dans le système $M/G/1$ avec rappels étant donnée que le serveur est libre.

Souvent, la fonction génératrice de la distribution stationnaire du nombre de clients en orbite peut être obtenue sous une forme explicite, mais ayant une expression assez lourde, elle ne révèle pas la nature de la distribution en question. Par conséquent, on est amené à l'étude de la convergence asymptotique de ces distributions pour les valeurs limites des différents paramètres : trafic intense, taux de rappels faible et taux de rappels élevé. Une première étude du comportement asymptotique d'un modèle avec rappels de type $M/G/1$ a été réalisée par Falin (1979, 1986) (59, 57) et Falin et Templeton (65). D'autres études ont été établies dans (11, 34, 75, 79, 109). Il existe aussi des travaux sur des modèles classiques sans rappels dans (113).

Le premier chapitre présente un travail de synthèse, avec lequel on entame la thèse, qui a été présenté en 2007 (15). On commence par introduire quelques modèles stochastiques, tels que : le recuit simulé, l'algorithme génétique et les systèmes de files d'attente. La modélisation du premier algorithme est basée sur les mesures invariants, par contre, celle des deux derniers algorithmes est markovienne. Notre intérêt particulier aux applications de ces algorithmes aux différents réseaux de télécommunication et d'Internet, nous a conduit à axer notre travail sur l'étude approfondie des systèmes de files d'attente.

Le deuxième chapitre est un rappel des résultats sur les systèmes de files d'attente classiques de type $M/G/1$, qui permettra d'introduire les chapitres suivants où seront abordés de nouveaux résultats.

Dans le chapitre trois, nous présentons une étude de certains modèles d'attente de type $M/G/1$ avec rappels. Nous commençons par les systèmes d'attente avec rappels et clients persistants, puis ceux avec rappels et clients impatientes. Nous réalisons pour la première fois l'analyse stochastique complète du système de files d'attente du modèle $M^X/G/1$ avec rappels, arrivées par groupes et clients impatientes. Nous donnons aussi quelques exemples d'applications de notre modèle.

Le quatrième chapitre commence par une synthèse bibliographique sur la propriété de décomposition stochastique des systèmes d'attente avec rappels. Par la suite, nous montrons la propriété en question pour le modèle $M^X/G/1$ avec rappels, arrivées par groupes et clients impatientes. Les résultats obtenus sont importants pour comprendre la contribution réelle des rappels pour le nombre de clients dans le système (aux instants de départ). Cette propriété qui est complexe, permet dans le chapitre suivant de réécrire la fonction génératrice du nombre de clients dans le système, sous une forme plus explicite, et ceci dans le but de prouver la convergence asymptotique sous un taux de rappels élevé.

Enfin, dans le cinquième et dernier chapitre, nous étudions le comportement asymptotique du nombre de clients dans le groupe de clients effectuant des tentatives répétées pour accéder au serveur du modèle $M^X/G/1$ avec rappels, arrivées par groupes et clients impatientes, sous un taux de trafic intense, puis sous un taux de rappels faible. Nous effectuons aussi l'étude du comportement asymptotique du nombre de clients dans le système sous un taux de rappels élevé. Ces trois résultats généralisent les résultats du modèle d'attente $M/G/1$ avec rappels, arrivées simples et clients persistants. Nous complétons notre travail par des exemples numériques.

Les résultats issues des trois derniers chapitres ont fait l'objet d'une publication dans le journal *Mathematical and Computer Modelling*, (2012) (12), et de plusieurs communications et publications dans des conférences et proceedings internationaux (13, 14).

MODÉLISATION STOCHASTIQUE



1.1 INTRODUCTION

Pour décrire et optimiser des systèmes techniques et économiques, on est souvent amené à l'étude des systèmes dynamiques déterministes, continus, décrits à l'aide d'équations différentielles. Cependant, il existe deux aspects centraux dans la modélisation de ces systèmes, à savoir le temps et l'incertain. On s'est vite aperçu que l'incertain était un élément incontournable dans toutes les applications, d'où la place importante qu'occupe aujourd'hui les processus stochastiques. Si la trajectoire est totalement décrite par un système dynamique déterministe à partir de conditions initiales, il n'en est rien dans le cas stochastique. En effet, en accord avec les lois probabilistes sous-jacentes, un système dynamique stochastique choisira une trajectoire parmi une infinité, à partir de conditions initiales identiques. Dès lors, plutôt que de s'intéresser aux valeurs prises par une réalisation particulière du processus, on cherche à connaître des probabilités ou d'autres paramètres statistiques associés à l'ensemble de ces trajectoires. Par exemple, au lieu de parler d'un taux de croissance annuel déterministe exprimé en pour cents, on donnera des probabilités ou des intervalles de confiance pour la valeur espérée de ce taux.

Certaines équations aux dérivées partielles (en particulier les équations des ondes) ont été également abordées par une approche probabiliste : la méthode de Glimm (71).

Une idée apparentée à la simulation stochastique, qui s'est avérée particulièrement fertile, a été celle d'introduire le hasard dans la résolution de problèmes purement déterministes, telles que l'évaluation d'intégrales multiples ou l'optimisation combinatoire.

Dans le cadre des algorithmes d'optimisation stochastique, les composantes aléatoires du système permettent d'augmenter la capacité d'exploration et d'adaptation du modèle. En effet, la plupart des algorithmes stochastiques sont fondés sur des mécanismes de recherche mimant des processus d'exploration et d'adaptation biologiques ou physiques (44).

Les chaînes de Markov sont les processus les plus classiques, plus simples et sont à la base de bien des modèles de recherche opérationnelle.

Nous avons choisi d'articuler ce chapitre en présentant trois outils de résolution stochastique :

- les algorithmes génétiques font actuellement partie des techniques de résolution les plus avancées en traitement du signal, en analyse d'événement rares, en mathématiques financières, ainsi qu'en physique et chimie quantique, (46, 72).
- le recuit simulé est un algorithme de recherche aléatoire des minima globaux. Parmi ses nombreux domaines d'application, on cite la conception des circuits électroniques et l'organisation de réseaux informatiques, (80, 90).
- la modélisation markovienne des files d'attente, conçue au début du siècle dernier pour modéliser les phénomènes de congestion dans les réseaux téléphoniques, cette théorie connaît un nouvel essor avec ses applications en informatique (30), en télécommunications (5, 6, 7) et surtout dans le contexte de l'internet (41, 45).

1.2 PROPRIÉTÉS DES ALGORITHMES STOCHASTIQUES

Les algorithmes d'optimisation stochastique vérifient certaines de ces propriétés :

- **Evolutionnaires** : On peut faire la différence entre les algorithmes d'optimisation stochastique "évolutionnaires" qui s'inspirent de phénomènes naturels et qui sont modélisés par des opérateurs spécifiques. Par contre, les algorithmes "non évolutionnaires" sont ceux qui ne s'inspirent pas de phénomènes naturels.
- **Trajectoire et population** : Les algorithmes d'optimisation stochastique qui possèdent "une structure à population", sont ceux qui manipulent un ensemble de solutions, à chaque itération. Par contre, ceux fondés sur la notion de "trajectoire", ne manipulent qu'une seule solution sur l'espace de recherche, à chaque itération.
- **Statiques et dynamiques** : On peut distinguer entre les algorithmes stochastiques selon leur manière d'utiliser la fonction objectif, "statique" (qui demeure inchangée tout au long de l'optimisation) ou "dynamique" (quand la fonction objectif est modifiée au cours de la recherche). Elles dépendent du temps.
- **Structures de voisinages** : La plupart des algorithmes d'optimisation stochastique utilisent "une seule" structure de voisinage. Cependant, il

existe des algorithmes qui permettent de changer de structure de voisinage en cours de recherche.

- **Mémoire à court et à long terme** : Certains algorithmes stochastiques font usage de l'historique de la recherche au cours de l'optimisation, ils sont qualifiés de méthodes à "mémoire à long terme" et à "mémoire à court terme" dans le cas où ils se limitent à considérer l'état de recherche à une itération donnée pour déterminer la prochaine, ces méthodes sont des processus de décision markovienne. Alors que d'autres algorithmes n'ont aucune mémoire du passé.

Les algorithmes génétiques sont considérés comme des algorithmes évolutionnaires possédant une structure à population, statique, utilisant une seule structure de voisinage et à mémoire à court terme. Le recuit simulé est un algorithme non évolutionnaire, à sauf dans des cas particuliers, statique, utilisant une seule structure de voisinage et à mémoire à court terme.

1.3 ALGORITHMES GÉNÉTIQUES

Les algorithmes génétiques sont des algorithmes d'optimisation s'appuyant sur des techniques dérivées de la génétique et de l'évolution naturelle : croisements, mutations, sélection, etc.

1.3.1 Modélisation par chaîne de Markov

Nous travaillerons sur la base d'un codage binaire, p représentant le nombre de bits utilisés pour le codage. La fonction d'évaluation, f est définie sur l'espace $E = \{0, 1\}^p$ à valeurs dans \mathbb{R}_+ . Le problème est donc de localiser l'ensemble des maxima globaux de f , ou à défaut, de trouver rapidement et efficacement des régions de l'espace où se situent ces maxima.

Comme nous l'avons vu, l'algorithme génétique est un algorithme stochastique itératif qui opère sur des ensembles de points et qui est bâti à l'aide de trois opérateurs : mutation, croisement et sélection, que nous présentons plus formellement à présent.

Soit N la taille (fixe) de la population, en terme mathématiques, un algorithme génétique est une chaîne de Markov X_k de la population de la génération k : il s'agit d'une matrice $X_k = (X_k^1, X_k^2, \dots, X_k^N)$ sur un espace produit E^N dont les N éléments sont des chaînes de bits (chromosomes) de taille p . Le passage de la génération k à la génération $k + 1$, c'est à dire de X_k à X_{k+1} se décompose en trois étapes :

$$X_k \xrightarrow{\text{mutation}} Y_k \xrightarrow{\text{croisement}} Z_k \xrightarrow{\text{sélection}} X_{k+1}.$$

Chacune de ces étapes peut être modélisée formellement.

Mutation $X_k \rightarrow Y_k$. L'opérateur considéré est le suivant : pour chaque composante de chaque élément X_k^i , une variable de Bernoulli de paramètre p_m est tirée indépendamment et, suivant le résultat, l'élément binaire examiné est changé ou non (0 est changé en 1 et 1 en 0). La probabilité p_m de mutation doit être préalablement choisie et est généralement faible.

Croisement $Y_k \rightarrow Z_k$. L'opérateur étudié est l'opérateur à un point : la probabilité de croisement p_c est fixée initialement. Pour construire la population Z_k , $N/2$ couples sont formés à partir de la population Y_k (par exemple en appariant les individus consécutifs de Y_k , ou bien en choisissant au hasard et uniformément des individus dans Z_k). Pour chaque couple, une variable de Bernoulli de paramètre p_c est tirée pour décider si le croisement a lieu. Si c'est le cas, un site de coupure est tiré au hasard, et les segments finaux des deux chromosomes sont échangés. Une nouvelle paire d'individus est ainsi obtenue (identique à l'ancienne s'il n'y a pas eu de croisement) et est stockée dans la population Z_k . En général, le paramètre p_c est choisi grand.

Remarquons que les opérateurs de mutation et de croisement ne font pas intervenir la fonction f , ce sont des opérateurs stochastiques d'exploration. C'est le troisième et dernier opérateur, la sélection, qui guide la population vers les valeurs élevées de la fonction f .

Sélection $Z_k \rightarrow X_{k+1}$. Les N individus de la population X_{k+1} sont obtenus après sélection des individus de Z_k . On conserve ainsi les "meilleurs" individus de Z_k , indépendamment à l'aide d'une distribution de probabilité qui favorise les individus de Z_k les mieux adaptés.

Le choix le plus fréquent est l'unique distribution telle que la probabilité de sélection uniforme d'un individu soit proportionnelle à son adaptation, c'est à dire la probabilité de sélection de l'individu Z_k^i est

$$p_i = \mathbb{P} \left(Z_k^i \right) = \frac{f \left(Z_k^i \right)}{\sum_{j=1}^N f \left(Z_k^j \right)}.$$

En tirant les individus dans la population Z_k conformément aux probabilités p_i , on constitue la nouvelle génération X_{k+1} .

Remarque 1.1 *Pour des fonctions d'évaluation pouvant s'annuler sur certaines régions de l'espace, il est possible que tous les individus aient des potentiels nul. Dans cette situation, l'algorithme est stoppé.*

1.4 RECUIT SIMULÉ

Les algorithmes stochastiques sont essentiellement des techniques de simulation de lois de probabilités complexes sur des espaces de grandes dimensions. Ces mesures peuvent être rangées en deux classes : les mesures de Boltzmann-Gibbs, et les mesures de Feynman-Kac (que nous n'utilisons pas dans ce qui suit). Les premières sont définies sur des espaces homogènes E , en terme d'une fonction énergie $U : E \rightarrow [0, \infty)$, un paramètre de température $\beta \in [0, \infty)$, et d'une mesure de référence λ sur E , telle que :

$$\mu_\beta(dx) = \frac{1}{Z_\beta} \exp[-\beta U(x)] \lambda(dx) \text{ avec } Z_\beta = \int \exp[-\beta U] d\lambda.$$

Dans ce contexte, l'espace d'état est souvent difficile à parcourir, et les techniques de simulation élémentaires de type acceptation-rejet sont prises en défaut. En pratique, les stratégies de simulation utilisées sont fondées sur des explorations markoviennes et locales de ces espaces. Le problème consiste alors à trouver une chaîne convergeant le plus rapidement possible vers la mesure d'équilibre recherchée.

1.4.1 Convergence du recuit simulé

L'algorithme de recuit simulé est une méthode de recherche aléatoire des minimas globaux d'une fonction numérique bornée $U : E \rightarrow \mathbb{R}_+$, définie sur un ensemble E . L'exploration aléatoire de l'espace d'état E est définie en terme d'une transition de probabilités $Q(x, dy)$ sur E , réversible par rapport à une mesure λ sur E . C'est à dire,

$$\lambda(dx)Q(x, dy) = \lambda(dy)Q(y, dx)$$

On notera que λ est nécessairement une mesure invariante de Q . L'algorithme de recuit simulé est un algorithme markovien non homogène. Il se présente sous la forme d'une chaîne de Markov dont le noyau de transition à chaque étape $n \geq 1$ dépend d'un paramètre de température $T(n) \in \mathbb{R}_+$.

- Pour $n = 0$, on simule une variable aléatoire X_0 , selon une distribution initiale η_0 .

– A l'étape n , la transition $X_n \rightarrow X_{n+1}$ est décomposée en une étape d'exploration et une étape d'acceptation.

1. L'étape d'exploration consiste à proposer un état Y_n de loi $Q(X_n, \cdot)$.
2. L'étape d'acceptation se décompose à nouveau en deux sous-étapes :
 - Si $U(Y_n) \leq U(X_n)$, on accepte l'état Y_n et on pose

$$X_{n+1} = Y_n.$$

– Si $U(Y_n) > U(X_n)$, alors on effectue le choix aléatoire suivant :

$$X_{n+1} = \begin{cases} Y_n & \text{avec une probabilité } e^{-\frac{1}{T(n)}(U(Y_n)-U(X_n))} \\ X_n & \text{avec une probabilité } 1 - e^{-\frac{1}{T(n)}(U(Y_n)-U(X_n))} \end{cases}$$

Au cours du temps, on fera décroître convenablement la température de sorte que l'algorithme de recherche se "gèle" sur l'un des extrema globaux de la fonction U . Le réglage de la décroissance de $T(n)$, lorsque n tend vers l'infini, sera donc inversement lié aux possibilités de mouvement de l'algorithme. Plus $T(n)$ est faible, plus l'algorithme aura tendance à ne plus changer d'état.

Remarque 1.2 Selon ces quelques remarques, une trop brusque variation de température pourrait conduire et figer l'algorithme dans des états non désirés tels que les extrema locaux de la fonction U . Cette idée naturelle provient de la physique.

D'une manière générale on peut dire que le recuit simulé procure généralement une solution de bonne qualité mais nécessite un grand nombre de paramètres, l'algorithme génétique est très puissant mais difficile à administrer et son efficacité dépend de la qualité du codage.

1.5 FILES D'ATTENTE

La théorie mathématique des files d'attente peut s'appliquer à différentes situations : optimisation des stocks (gestion à flux tendu), gestion des avions au décollage ou à l'atterrissage, attente des clients à un guichet, ou bien encore traitement informatique de données par un serveur.

La modélisation mathématique des files d'attente est un outil de la logistique. Elle relève du calcul des probabilités : les arrivées et départs des clients de la file sont analysés comme un processus stochastique typique d'un processus de naissance et de mort. L'objectif de la modélisation est la recherche des solutions optimales de gestion des files d'attente, ou « queues », telles que la recherche d'une

gestion de priorité (ou discipline) donnant le temps d'attente moyen minimum, ou le temps d'attente au pire des cas minimum, etc., en fonction de la loi de probabilité des arrivées et de la loi donnant le temps de traitement. Elle peut aussi permettre d'évaluer la conséquence de la défaillance d'un serveur sur le temps moyen de résidence dans la file et d'apprécier l'impact de la mise en place d'un serveur supplémentaire, etc.

L'analyse des systèmes de files d'attente s'appuie généralement sur les outils de la théorie des chaînes de Markov à temps continu et à temps discret.

1.5.1 Modélisation des systèmes de files d'attente par chaînes de Markov

Soit $(U_n)_{n \geq 1}$ une suite de v.a. positives et indépendantes de lois respectives $(\mu_n)_{n \geq 1}$. On considère la chaîne de Markov définie de façon récursive par l'équation suivante

$$\begin{cases} X_{n+1} = (X_n - 1)^+ + U_{n+1} \\ X_0 = 0 \end{cases}$$

Ce processus aléatoire peut s'interpréter comme la longueur d'une file d'attente, ou encore le temps d'attente d'un client arrivant à un guichet, servant une personne par unité de temps. Dans ce contexte, la v.a. U_{n+1} représente le nombre de clients arrivant dans la file d'attente au temps $(n + 1)$.

On peut aussi interpréter X_n comme le nombre de paquets (symboles binaires représentant de l'information : voix, vidéo, données, ...) en attente dans la mémoire d'un canal de communication, transmettant un paquet par unité de temps.

On notera que les transitions de cette chaîne sont données pour tout $i \geq 1$, et pour tout $j \geq 0$, par la formule suivante

$$\begin{aligned} \mathbb{P}(X_{n+1} = (i - 1) + j \mid X_n = i) &= \mu_{n+1}(j) \\ &= \mathbb{P}(X_{n+1} = j \mid X_n = 0). \end{aligned}$$

CONCLUSION DU CHAPITRE

On s'est intéressé à la modélisation stochastique comme l'algorithme génétique et le recuit simulé. Les systèmes des files d'attente sont également des outils de la modélisation stochastique c'est pourquoi dans les chapitres suivants nous nous intéressons aux modèles de files d'attente particuliers.

SYSTÈMES DE FILES D'ATTENTE CLASSIQUES

2.1 DESCRIPTION DU PHÉNOMÈNE D'ATTENTE

Un phénomène d'attente peut être décrit comme un système composé d'un certain nombre (fini ou non) de places d'attente d'un ou plusieurs serveurs et de clients arrivant à des instants aléatoires. Les clients attendent, se font servir selon des règles spécifiées et quittent le système. Ils peuvent être des appels téléphoniques, des machines, ..., de même que les serveurs peuvent être un central téléphonique, un processeur,.... Quand les serveurs sont tous occupés, les clients doivent alors patienter dans un espace d'attente (s'il existe) jusqu'à ce qu'un serveur soit disponible.

L'identification des systèmes de files d'attente classiques se base principalement sur trois éléments : le processus stochastique décrivant l'arrivée des clients dans le système, le mécanisme de service (le nombre de serveurs et la loi probabiliste décrivant la durée des services) et la discipline d'attente.

Le processus d'arrivée spécifie les instants auxquels les clients arrivent dans le système. Dans la théorie classique des files d'attente, on fait le plus souvent l'hypothèse que les clients arrivent de manière isolée et indépendamment les uns

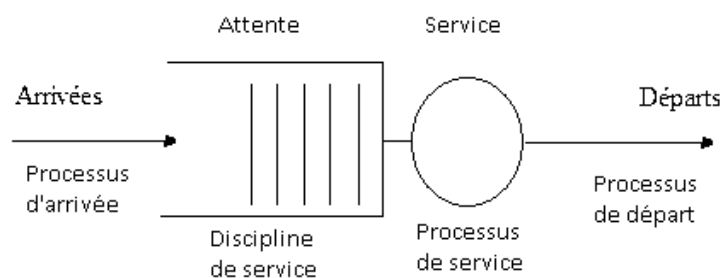


FIGURE 2.1 – Représentation schématique d'une file d'attente classique

des autres. Sous ces hypothèses, les intervalles de temps entre deux arrivées successives forment une suite de variables aléatoires indépendantes et identiquement distribuées.

La liste qui suit résume les lois de probabilité les plus couramment rencontrées dans la modélisation des systèmes de files d'attente ainsi que les symboles associés.

- La lettre M désigne la loi exponentielle.
- La lettre D correspond à une loi déterministe.
- Le symbole E_k désigne un processus où les intervalles de temps entre deux arrivées successives sont des variables aléatoires indépendantes et identiquement distribuées suivant une loi d'Erlang d'ordre k .
- La lettre G est utilisée lorsqu'aucune hypothèse particulière n'est faite sur le processus d'arrivée, ce dernier étant alors un processus de renouvellement quelconque.

Les temps de service nécessaire au traitement des clients sont supposés être des réalisations de variables aléatoires indépendantes et identiquement distribuées. La description du processus de service revient alors à préciser la loi de probabilité de ces variables aléatoires. Les symboles utilisés pour décrire les processus de service sont les mêmes que ceux introduits pour les processus d'arrivée.

Le nombre de serveurs correspond au nombre maximal de clients pouvant être traités simultanément. Tous les serveurs sont supposés identiques, en particulier les temps de service sont indépendants d'un serveur à l'autre et distribués selon une même loi de probabilité.

La capacité d'accueil d'un système de files d'attente correspond au nombre maximal de clients pouvant être présents dans le système à un instant quelconque. Il est égal à la somme du nombre de serveurs et du nombre de places d'attente disponibles. Si un client arrive dans un système ayant atteint sa capacité maximale d'accueil, il est refoulé et doit quitter le système sans avoir été servi.

Le plus souvent, le nombre de clients susceptibles d'accéder au service est supposé illimité et leur fréquence d'arrivée constante. Certaines situations sont, cependant, caractérisées par un nombre fixe et limité de clients. Chaque client présent dans le système diminue alors le nombre d'arrivées potentielles, le taux d'arrivée dans la file n'est donc plus constant mais dépend du nombre de clients présents dans le système. Dans un tel cas, le processus d'arrivée décrit le temps nécessaire à un client entre le moment où il quitte le système et celui où il y revient.

La discipline d'attente, est la règle de priorité déterminant l'ordre dans lequel les clients vont accéder à la ressource modélisé par le serveur. Les disciplines d'attente classiques, ainsi que leurs acronymes, sont

- FIFO : *first in first out "premier arrivé, premier servi"*, c'est la discipline de service employée le plus souvent et c'est celle qui sera admise par défaut ;
- LIFO : *last in, first out "dernier arrivé, premier servi"* ;
- SIRO : *service dans un ordre aléatoire* ;
- RR : *round robin "les clients sont servi à tour de rôle pendant un intervalle de temps fixe, appelé quantum"*
- PS : *processor sharing "cas limite de la discipline RR lorsque le quantum tend vers zéro"*.

Certains modèles de files d'attente font usage de disciplines plus complexes reposant sur des priorités de traitement. Un système peut également accepter la préemption d'un service, c'est-à-dire l'interruption du service d'un client au profit d'un autre. Si un tel mode de fonctionnement est autorisé, le service interrompu peut être simplement terminé plus tard (*preemptive resume*) ou, alors, être recommencé depuis le début, le travail déjà effectué étant perdu (*preemptive repeat*).

Pour la classification des systèmes de files d'attente, on a recours à une notation symbolique introduite par Kendall, comprenant six symboles rangés dans l'ordre $A/B/c/m/P-D$, où

- A décrit la nature du processus d'arrivée,
- B décrit la nature du processus de service,
- c est le nombre de serveurs,
- m est la capacité d'accueil d'un système,
- P est la taille de la population,
- D est la discipline de la file.

Dans sa version courte, seuls les trois premiers symboles $A/B/c$ sont utilisés. Dans un tel cas, on suppose que la file est régie par une discipline FIFO et que le nombre de places d'attente ainsi que celui des clients susceptibles d'accéder au système sont illimités.

2.2 ANALYSE MATHÉMATIQUE D'UN SYSTÈME DE FILES D'ATTENTE

L'étude mathématique d'un système de files d'attente se fait généralement par l'introduction d'un processus stochastique, défini de façon appropriée. On s'intéresse principalement au nombre de clients $N(t)$ se trouvant dans le système à l'instant t ($t \geq 0$).

En fonction des quantités qui définissent le système, on cherche à déterminer :

- Les probabilités d'état $P_n(t) = \mathbb{P}(N(t) = n)$, qui définissent le régime transitoire du processus stochastique $\{N(t) \mid t \geq 0\}$. Il est évident que les fonctions $P_n(t)$ dépendent de l'état initial ou de la distribution initiale du processus.
- Le régime stationnaire du processus stochastique qui est défini par :

$$\pi_n = \lim_{t \rightarrow \infty} P_n(t) = \mathbb{P}(N(\infty) = n) = \mathbb{P}(N = n), (n = 0, 1, 2, \dots).$$

$\{\pi_n\}_{n \geq 0}$ est appelée distribution stationnaire du processus $\{N(t) \mid t \geq 0\}$. Le calcul explicite du régime transitoire s'avère généralement pénible, voire impossible, pour la plupart des modèles donnés. On se contente donc de déterminer le régime stationnaire.

2.3 TYPES DE MODÈLES

2.3.1 Modèles markoviens

Les modèles markoviens caractérisent les systèmes dans lesquels les deux quantités stochastiques principales, qui sont le temps inter-arrivées et la durée de service, sont des variables aléatoires indépendantes et exponentiellement distribuées. La propriété d'absence de mémoire de la loi exponentielle facilite l'étude de ces modèles. L'étude mathématique de tels systèmes se fait par l'introduction d'un processus stochastique approprié. Ce processus est souvent le processus $\{N(t), t \geq 0\}$ défini comme étant le nombre de clients dans le système à l'instant t . L'évolution temporelle du processus markovien est complètement définie grâce à la propriété d'absence de mémoire.

2.3.2 Modèles non markoviens

En l'absence de l'exponentialité ou lorsque l'on s'écarte de l'hypothèse d'exponentialité de l'une des deux quantités stochastiques : le temps des inter-arrivées et la durée de service, ou en prenant en compte certaines spécificités des problèmes par introduction de paramètres supplémentaires, on aboutit à un modèle non markovien. La combinaison de tous ces facteurs rend l'étude mathématique du modèle très délicate, voire impossible- on essaye alors de se ramener à un processus de Markov judicieusement choisi à l'aide de l'une des méthodes d'analyse suivantes :

1. **Méthode des étapes d'Erlang** : Son principe est d'approximer toute loi de probabilité ayant une transformation de Laplace rationnelle par une loi

de Cox (mélange de lois exponentielles), cette dernière possède la propriété d'absence de mémoire par étape.

2. **Méthode de la chaîne de Markov induite** : Elaborée par Kendall, et souvent utilisée, elle consiste à choisir une suite d'instantanés $1, 2, 3, \dots, n$ (déterministes ou aléatoires) tels que la chaîne induite $\{N_n, n \geq 0\}$, où $N_n = N(n)$, soit markovienne et homogène.
3. **Méthode des variables supplémentaires** : Elle consiste à compléter l'information sur le processus $\{N(t), t \geq 0\}$ de telle manière à lui donner le caractère markovien. Ainsi, on se ramène à l'étude du processus $\{N(t), A(t_1), A(t_2), \dots, A(t_n), t \geq 0\}$. Les variables $A(t_k), k \in \{1, 2, \dots, n\}$ sont dites supplémentaires.
4. **Méthode des événements fictifs** : Le principe est d'introduire des événements fictifs qui permettent de donner une interprétation probabiliste aux transformées de Laplace et aux variables aléatoires décrivant le système étudié.
5. **Simulation** : C'est un procédé d'imitation artificielle d'un processus réel effectué sur ordinateur. Elle nous permet d'étudier les systèmes les plus complexes, de prévoir leurs comportements et de calculer leurs caractéristiques. Les résultats obtenus ne sont qu'approximatifs, mais peuvent être utilisés avec une bonne précision. Cette technique se base sur la génération de variables aléatoires suivant les lois gouvernant le système.

2.4 CARACTÉRISTIQUES D'UN SYSTÈME DE FILES D'ATTENTE

On note λ le taux d'arrivée des clients. Cela signifie que l'espérance mathématique de la durée séparant deux arrivées successives est $\mathbb{E}[A] = \frac{1}{\lambda}$.

On note γ le taux de service des clients. Cela signifie que l'espérance de la durée de service est $\mathbb{E}[Se] = \frac{1}{\gamma}$.

L'intensité du trafic s'exprime de la manière suivante

$$\rho = \frac{\lambda}{\gamma} = \frac{\mathbb{E}[Se]}{\mathbb{E}[A]},$$

La distribution stationnaire du processus stochastique introduit permet d'obtenir les caractéristiques d'exploitation du système, telles que : le temps d'attente d'un client (le temps qu'un client passe dans la file d'attente), le temps de séjour d'un client dans le système (composé du temps d'attente et de la durée de service), le taux d'occupation des dispositifs de service, la durée de la période d'activité

(l'intervalle de temps pendant lequel il y a toujours au moins un client dans le système) ; et les mesures de performance suivantes :

- \bar{n} : nombre moyen de client dans le système de files d'attente,
- \bar{n}_f : nombre moyen de clients dans la file d'attente,
- \bar{W}_s : temps moyen de séjour d'un client dans le système,
- \bar{W} : temps moyen d'attente d'un client dans la file d'attente.

Ces valeurs sont liées les unes aux autres par les relations suivantes :

- $\bar{n} = \lambda \bar{W}_s$,
- $\bar{n}_f = \lambda \bar{W}$,
- $\bar{n} = \bar{n}_f + \frac{\lambda}{\gamma}$,
- $\bar{W}_s = \bar{W} + \frac{1}{\gamma}$.

Les deux premières sont appelées "formules de Little". Il est à noter que ces formules sont valables sous la vérification de la condition que le système est en régime stationnaire $\rho = \frac{\lambda}{\gamma} < 1$.

2.5 MODÈLE D'ATTENTE M/G/1

2.5.1 Description du modèle

Les clients arrivent dans le système selon un processus de Poisson de taux $\lambda > 0$. De ce fait, le temps entre deux arrivées successives suit une loi exponentielle de moyenne $\frac{1}{\lambda}$. Le service est assuré par un seul serveur. A l'arrivée d'un client, si le serveur est libre, le client sera pris en charge immédiatement. Dans le cas contraire, il rejoint la file d'attente (de capacité illimitée et discipline FIFO) les durées de service (Se) sont des variables aléatoires indépendantes et identiquement distribuées de loi générale dont la fonction de répartition $B(x)$ et la transformée de Laplace-Stieltjes $\tilde{B}(s)$. Soient $\mathbb{E}[Se] = \frac{1}{\gamma}$ et $\mathbb{E}[Se^2]$.

2.5.2 Chaîne de Markov induite

Nous introduisons le processus stochastique $\{N(t), t \geq 0\}$ qui n'est pas un processus de Markov. Pour le rendre markovien, nous utiliserons la méthode des chaînes de Markov induites.

Soit le processus à temps discret $\{N_n = N(\xi_n), n \geq 1\}$ où ξ_n est l'instant où le $n^{\text{ème}}$ client a fini son service et quitte le système. Vérifions que cette suite de variables définit bien une chaîne de Markov.

Soient les (A_n) des variables aléatoires indépendantes et identiquement distribuées telles que A_n est le nombre de clients arrivants pendant le $n^{\text{ème}}$ service avec la distribution

$$\mathbb{P}(A_n = i) = a_i = \int_0^{\infty} \left(\frac{\lambda t}{i!}\right) i! \exp(-\lambda t) dB(t),$$

où $a_i > 0$ et $i > 0$. Déterminons l'équation fondamentale de la chaîne :

$$N_{n+1} = \begin{cases} N_n - 1 + A_{n+1}, & \text{si } N_n \geq 1 \\ A_{n+1}, & \text{si } N_n = 0 \end{cases}, n \geq 1.$$

Soit la variable aléatoire

$$\delta_n = \begin{cases} 1 & \text{si } N_n > 0 \\ 0 & \text{si } N_n = 0 \end{cases},$$

alors l'équation fondamentale de la chaîne devient :

$$N_{n+1} = N_n - \delta_n + A_{n+1}. \quad (2.1)$$

Il est évident que N_{n+1} dépend de N_n et A_{n+1} seulement et non pas de N_{n-1} , N_{n-2} , D'où la suite $\{N_n, n \geq 1\}$ est une chaîne de Markov induite du processus $\{N(t), t \geq 0\}$ avec les probabilités de transitions $\mathbb{P}(N_{n+1} = j / N_n = i) = p_{ij}$ qui s'expriment de la manière suivante :

$$\begin{cases} p_{0j} = a_j, & \text{si } j \geq 0 \\ p_{ij} = a_{j-i+1}, & \text{si } 0 \leq i \leq j+1 \\ p_{ij} = 0, & \text{ailleurs.} \end{cases}$$

Par conséquent, la matrice de transition M est donnée par

$$\begin{bmatrix} p_{00} & p_{01} & p_{02} & p_{03} & \dots \\ p_{10} & p_{11} & p_{12} & p_{13} & \dots \\ p_{20} & p_{21} & p_{22} & p_{23} & \dots \\ p_{30} & p_{31} & p_{32} & p_{33} & \dots \end{bmatrix} = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & \dots \\ a_0 & a_1 & a_2 & a_3 & \dots \\ 0 & a_0 & a_1 & a_2 & \dots \\ 0 & 0 & a_0 & a_1 & \dots \end{bmatrix}.$$

Puisque nous pouvons passer d'un état à n'importe quel autre, alors la chaîne de Markov est irréductible. En outre la matrice n'est pas décomposable (elle est apériodique), donc la chaîne est ergodique.

La distribution stationnaire existe si $\rho = \frac{\lambda}{\gamma} < 1$. Nous avons la fonction génératrice

$$\begin{aligned}
A(z) &= \sum_{i=0}^{\infty} a_i z^i = \sum_{i=0}^{\infty} z^i \int_0^{\infty} \left(\frac{\lambda t}{i!}\right)^k \exp(-\lambda t) dB(t) \\
&= \int_0^{\infty} \exp\{-(\lambda - \lambda z)t\} dB(t),
\end{aligned}$$

posons $\tilde{B}(s) = \int_0^{\infty} \exp(-st) dB(t)$, alors $A(z) = \tilde{B}(\lambda - \lambda z)$ converge pour $|z| \leq 1$:

1. $|z| < 1$: $0 < a_k < 1 \Rightarrow |a_k z| < |z|^k$
2. $|z| = 1$: $A(1) = \tilde{B}(0) = 1$.

Théorème 2.1 Soit $\rho < 1$. La distribution stationnaire de la chaîne de Markov induite $\{N_n, n \geq 1\}$ possède la fonction génératrice suivante (58)

$$\Pi(z) = \sum_{n=0}^{\infty} z^n \pi_n = \frac{(1-\rho) \tilde{B}(\lambda - \lambda z) (1-z)}{\tilde{B}(\lambda - \lambda z) - z}.$$

Soient les probabilités suivantes :

$$\begin{aligned}
p_j &= \lim_{t \rightarrow \infty} \mathbb{P}(N(t) = j), j \geq 0; \\
\pi_j &= \lim_{n \rightarrow \infty} \mathbb{P}(N(\xi_n) = j), j \geq 0; \\
r_j &= \lim_{n \rightarrow \infty} \mathbb{P}(N(\zeta_n) = j), j \geq 0;
\end{aligned}$$

où ζ_n est l'instant d'arrivée du $n^{\text{ème}}$ client. Comme le processus des arrivées est celui de Poisson de paramètre λ et le nombre de clients dans le système $N(t)$ est discontinu avec un changement de taille ± 1 , alors

$$p_j = r_j = \pi_j.$$

Par conséquent, le processus $\{N(t), t \geq 0\}$ a une distribution stationnaire identique à celle de la chaîne de Markov induite et la fonction génératrice du nombre de clients dans le système est

$$Q(z) = \sum_{j=0}^{\infty} p_j z^j = \Pi(z).$$

2.5.3 Mesures de performance

– Nombre moyen de clients dans le système

$$\bar{n} = \rho + \frac{\rho^2 + \lambda^2 \text{Var}[Se]}{2(1-\rho)}.$$

– Temps moyen de séjour d'un client dans le système

$$\bar{W}_s(s) = \frac{s(1-\rho)}{s-\lambda+\lambda\tilde{B}(s)}\tilde{B}(s).$$

– Temps moyen d'attente d'un client

$$\bar{W}(s) = \frac{s(1-\rho)}{s-\lambda+\lambda\tilde{B}(s)}.$$

2.6 MODÈLE D'ATTENTE $M^X/G/1$

2.6.1 Description du modèle

Considérons un système de files d'attente où le service des clients est assuré par un seul serveur. Les clients arrivent par groupes, ces arrivées des groupes de clients primaires suivent une loi de Poisson de paramètre $\lambda > 0$. Le groupe contient K clients ($1 \leq K \leq \infty$) où K est une variable aléatoire discrète qui est égale à k avec la probabilité c_k et dont la fonction génératrice est $C(z) = \sum_{k=1}^{\infty} c_k z^k$. La taille moyenne des groupes est $\mathbb{E}[K] = C'(1) = \bar{c} = \sum_{k=1}^{\infty} kc_k$.

Le temps de service τ suit une loi générale de fonction de répartition $B(t)$ et de transformée de Laplace-Stieltjes $\tilde{B}(s)$, $\text{Re}(s) > 0$. Soient les moments $\beta_k = (-1)^k \tilde{B}^{(k)}(0)$. Les durées entre deux arrivées consécutives des groupes, la taille des groupes ainsi que les durées de service sont supposées mutuellement indépendantes.

2.6.2 Analyse du modèle

Considérons un intervalle de temps arbitraire de longueur t . Soit $N(t)$ le nombre de clients arrivant dans l'intervalle de temps t et $\nu(t)$ le nombre de groupes arrivant dans cet intervalle de temps. Alors,

$$\mathbb{P}(\nu(t) = n) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}.$$

Soient $\alpha_1, \alpha_2, \dots, \alpha_\nu$ les tailles des groupes, où α_i est indépendante de α_j pour $i \neq j$ et les $\{\alpha_i\}_{1 \leq i \leq \nu}$ sont des variables aléatoires indépendantes et identiquement distribuées. Nous avons alors

$$\begin{aligned}
N(t) &= N = \alpha_1 + \alpha_2 + \cdots + \alpha_\nu \\
\mathbb{E} [z^N | \nu] &= [\mathbb{E} [z^\alpha]]^\nu = C^\nu(z) \\
\mathbb{E} [z^N] &= \sum_{\mu=0}^{\infty} C^\mu(z) \frac{(\lambda t)^\mu e^{-\lambda t}}{\mu!} = e^{-\lambda t[1-C(z)]}. \quad (2.2)
\end{aligned}$$

$\mathbb{E} [z^N]$ est la fonction génératrice du nombre d'arrivées dans un intervalle de temps arbitraire de taille t .

2.6.3 Chaîne de Markov induite

A présent, on considère l'intervalle de temps t comme étant le temps de service actuel, alors la fonction génératrice du nombre d'arrivées sera définie par

$$A(z) = \int_{t=0}^{\infty} \exp(-\lambda(1-C(z))t) B(t) dt = \tilde{B}(\lambda - \lambda C(z)). \quad (2.3)$$

Soit la chaîne de Markov induite aux instants de départs, de la même méthode utilisée dans le cas du modèle $M/G/1$ ordinaire, on effectue une analyse de cette chaîne.

En effet, en remplaçant la valeur $A(z)$ de l'équation (2.3) dans le résultat du modèle $M/G/1$ ordinaire, on obtient

$$\prod(z) = \frac{(1-\rho) \tilde{B}(\lambda - \lambda C(z)) (1-z)}{\tilde{B}(\lambda - \lambda C(z)) - z},$$

$$\text{avec } \rho = A'(1) = \lambda C'(1) \beta_1 = \lambda \bar{c} \beta_1.$$

2.6.4 Approche alternative à l'analyse du modèle $M^X/G/1$

Dans cette approche, on considère les groupes des arrivées comme étant des requêtes avec leur temps de service égal à la somme des temps de services de tous les clients du groupe. La durée de service d'un groupe τ^* suit une loi générale de fonction de répartition $B^*(t)$ et de transformée de Laplace-Stieltjes $\tilde{B}^*(s)$. Alors

$$\tilde{B}^*(s) = \sum_{k=1}^{\infty} c_k (\tilde{B}(s))^k = C(\tilde{B}(s)),$$

avec

$$\begin{aligned}\overline{\tau^*} &= - \left. \frac{d\widetilde{B}^*(s)}{ds} \right|_{s=0} = \left[-\widetilde{B}'(s) C'(\widetilde{B}(s)) \right]_{s=0} \\ &= \beta_1 C'(1) = \bar{c}\beta_1,\end{aligned}$$

et

$$\overline{\tau^{*2}} = \left. \frac{d^2\widetilde{B}^*(s)}{ds^2} \right|_{s=0} = \sigma_\tau^2 \bar{c} + (\bar{c}^2 + \sigma_K^2) \beta_1^2,$$

où σ_τ^2 est la variance de la variable aléatoire τ qui représente la durée de service d'un client et σ_K^2 est la variance de la variable aléatoire K qui représente la taille d'un groupe de clients. Soit $A^*(z) = \widetilde{B}^*(\lambda - \lambda z) = C(\widetilde{B}(\lambda - \lambda z))$ la fonction génératrice du nombre des arrivées des groupes durant le temps de service d'un groupe. Comme le groupe en entier est considéré comme une requête de service, alors on utilise la chaîne de Markov induite aux instants de départs. Soit $Q^*(z)$ la fonction génératrice du nombre de groupes restant après le départ d'un groupe. On peut maintenant remplacer $A(z)$ par $A^*(z)$ avec $\rho = \lambda \bar{c}\beta_1$. D'où la fonction génératrice du nombre de clients dans le système à l'instant de départ d'un groupe

$$Q^*(z) = \frac{(1 - \rho) C(\widetilde{B}(\lambda - \lambda z)) (1 - z)}{C(\widetilde{B}(\lambda - \lambda z)) - z}. \quad (2.4)$$

2.6.5 Mesures de performance

- Temps moyen d'attente d'un groupe \overline{W}_{fg} avant qu'il ne commence son service

$$\overline{W}_{fg} = \frac{\lambda}{2(1 - \rho)} \overline{\tau^{*2}},$$

on peut écrire \overline{W}_{fg} sous la forme

$$\overline{W}_{fg} = \frac{\rho \bar{c}\beta_1}{2(1 - \rho)} \left[1 + \frac{S_\tau^2}{\bar{c}} + S_K^2 \right],$$

où $S_\tau^2 = \frac{\sigma_\tau^2}{\beta_1^2}$ et $S_K^2 = \frac{\sigma_K^2}{\bar{c}^2}$ sont les coefficient de variation quadratique de τ et K .

- Temps moyen d'attente \overline{W}_2 pour un appel sachant que le service de son groupe a commencé

$$\bar{W}_2 = \widetilde{W}_2(s) \Big|_{s=1} = \frac{\beta_1 [C''(1) - \bar{c}]}{2\bar{c}} = \beta_1 \left[\frac{\bar{c}}{2} (1 + S_K^2) - \frac{1}{2} \right].$$

– Temps moyen d'attente global \bar{W}_f pour un appel (dans le groupe) est donné par

$$\begin{aligned} \bar{W}_f &= \bar{W}_{fg} + \bar{W}_2 \\ &= \frac{\rho \bar{c} \beta_1}{2(1-\rho)} \left[1 + \frac{S_\tau^2}{\bar{c}} + S_K^2 \right] + \frac{\beta_1}{2} [\bar{c} (1 + S_K^2) - 1]. \end{aligned}$$

CONCLUSION DU CHAPITRE

Nous avons rappelé et présenté les notions et techniques de base sur les systèmes de files d'attente classiques de type $M/G/1$, et de type $M^X/G/1$ étudiés aussi dans (62, 78) qui seront utilisées dans la modélisation des systèmes de files d'attente avec rappels et arrivées par groupes.

SYSTÈMES DE FILES D'ATTENTE

AVEC RAPPELS

3.1 INTRODUCTION

Les systèmes de files d'attente avec rappels sont caractérisés par la propriété qu'un client qui trouve à son arrivée tous les serveurs occupés quitte l'espace de service et rappelle ultérieurement à des instants aléatoires. Entre deux rappels successifs, le client est dit "en orbite". Ces systèmes de files d'attente sont largement utilisés dans la modélisation des systèmes informatiques et des réseaux de télécommunications (26, 27, 38, 39, 43, 82, 83, 98, 104, 107). Une description complète de situations où les systèmes de files d'attente avec rappels peut être trouvée dans la monographie de Falin et Templeton (1997) (65) et dans (9, 10, 29, 103, 104). Une classification bibliographique est donnée dans les articles de Artalejo (1999) (16) et (2010) (17). Le modèle M/G/1 avec rappels et clients non-persistants a été considéré par Falin (1990) (60), par Martin et Artalejo (1995) (99) et Martin et Gomez-Corral (1995) (100).

Pour identifier un système de files d'attente avec rappels, on a besoin des spécifications suivantes : la nature stochastique du processus des arrivées, la distribution du temps de service, le nombre de serveurs qui composent l'espace de service, la capacité et la discipline d'attente ainsi que la spécification concernant le processus de répétition d'appels.

Le modèle général d'un système de files d'attente avec répétition d'appels, étudiés dans (37, 40, 53, 64, 73, 81, 88, 105), peut être décrit comme suit : le système est composé de $c \geq 1$ dispositifs de service et de $m - c$ ($m \geq c$) positions d'attente. Les clients arrivent dans le système selon un processus aléatoire avec une loi de probabilité donnée, et forment un flux d'appels primaires. A l'arrivée d'un client, s'il y a une position d'attente libre, le client rejoint la file d'attente. Dans le cas contraire, il quitte l'espace de service temporairement avec une probabilité H_0 pour tenter sa chance après une durée de temps aléatoire, ou il quitte le système

définitivement avec une probabilité $1 - H_0$. Entre les tentatives, le client est "en orbite" et devient source d'appels répétés ou d'appels secondaires. La capacité O de l'orbite peut être finie ou infinie. Dans le cas où O est finie et si l'orbite est pleine, le client quitte le système pour toujours. Lorsqu'un client est rappelé de l'orbite, il est traité de la même manière qu'un client primaire avec une probabilité H_k (s'il s'agit de la $k^{\text{ième}}$ tentative échouée).

La notation de Kendall est $A/B/c/m/O/H$, où A et B décrivent respectivement la distribution du temps inter-arrivées et la distribution du temps de service, c est le nombre de serveurs identiques et indépendants, $m - c$ est la capacité du tampon, O est la capacité de l'orbite, H est la fonction de persistance $H = \{H_k, k \geq 0\}$. Si m, O et H sont absents dans la notation de Kendall, alors $m = c, O = \infty$ et $H_k = 1$ pour tout $k \geq 0$. La distribution du temps inter-rappels n'est pas indiquée.

On décrit l'entrée dans le système par une suite $\{(\tau_n^e, M_n), n \geq 1\}$ (60), où τ_n^e est l'intervalle de temps entre les arrivées des $n^{\text{ème}}$ et $(n + 1)^{\text{ème}}$ clients primaires, M_n est une marque associée au $n^{\text{ème}}$ client primaire. Cette marque comprend : τ_n^s la durée de service, w_n le nombre maximal de rappels autorisés (on suppose $w_n \rightarrow \infty$), $\tau_n^r = \{\tau_{n1}^r, \tau_{n2}^r, \dots, \tau_{nw_n}^r\}$ une suite d'intervalles de temps entre deux rappels successifs. Les variables aléatoires τ_n^e, τ_n^s et τ_n^r sont indépendantes et définies sur l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. Dans ce qui suit, on suppose que les suites $\{\tau_n^e\}, \{\tau_n^s\}$ et $\{\tau_n^r\}$ sont des suites indépendantes de variables aléatoires indépendantes et identiquement distribuées.

Dans ce chapitre, nous présentons une étude de certains modèles avec rappels. Nous commençons par les systèmes de files d'attente avec rappels de type $M/G/1$ avec clients persistants et ceux avec rappels et clients impatientes. Puis nous réalisons pour la première fois l'analyse stochastique complète du modèle $M^X/G/1$ avec rappels, arrivées par groupes et clients impatientes.

3.2 MODÈLE D'ATTENTE $M/G/1$ AVEC RAPPELS

Le modèle $M/G/1$ avec rappels est le modèle le plus étudié par les spécialistes. Il existe une littérature abondante sur ses diverses propriétés (1, 8, 18, 25, 35, 77, 95, 29, 36, 52, 84, 96, 101, 102).

3.2.1 Description du modèle

Les clients arrivent dans le système selon un processus de Poisson de taux $\lambda > 0$: $P(\tau_n^e \leq x) = 1 - e^{-\lambda x}$. Le service des clients est assuré par un seul serveur. La durée de service τ est de loi générale $P(\tau_n^s \leq x) = B(x)$ et de transformée

de Laplace-Stieltjes $\tilde{B}(s)$, $\text{Re}(s) > 0$. Soient les moments $\beta_k = (-1)^k \tilde{B}^{(k)}(0)$, l'intensité du trafic $\rho = \lambda\beta_1$ et $\gamma = \frac{1}{\beta_1}$. La durée entre deux rappels successifs d'une même source secondaire est exponentiellement distribuée de paramètre $\theta > 0$: $T(x) = P(\tau_n^r \leq x) = 1 - e^{-\theta x}$.

Le système évolue de la manière suivante : On suppose que le $(n-1)^{\text{ème}}$ client termine son service à l'instant ξ_{n-1} (les clients sont numérotés dans l'ordre de service) et le serveur devient libre ; même s'il y a des clients dans le système, ils ne peuvent pas occuper le serveur immédiatement à cause de leur ignorance de l'état de ce dernier. Donc il existe un intervalle de temps R_n durant lequel le serveur reste libre avant que le $n^{\text{ème}}$ client n'entre en service. A l'instant $\xi_n = \eta_n + R_n$ le $n^{\text{ème}}$ client débute son service durant un temps τ_n^s . Les rappels qui arrivent durant ce temps de service n'influent pas sur ce processus. A l'instant $\xi_n = \eta_n + \tau_n^s$ le $n^{\text{ème}}$ client achève son service, le serveur devient libre et ainsi de suite.

3.2.2 Chaîne de Markov induite

Considérons le processus $\{C(t); N_o(t); t \geq 0\}$, où $C(t)$ représente l'état du serveur

$$C(t) = \begin{cases} 0 & \text{si le serveur est libre} \\ 1 & \text{si le serveur est occupé} \end{cases}$$

et $N_o(t)$ est le nombre de clients en orbite à la date t . En général, ce processus n'est pas un processus de Markov, mais il possède une chaîne de Markov induite. Cette chaîne a été décrite pour la première fois par Choo et Conolly (1979) (48).

Soit (q_n) la chaîne de Markov induite aux instants de départs, où $q_n = N_o(\xi_n)$ représente le nombre de clients en orbite après le $n^{\text{ème}}$ départ, dont l'équation fondamentale est :

$$q_{n+1} = q_n - \delta_{q_n} + v_{n+1},$$

où v_{n+1} est le nombre d'clients primaires arrivant dans le système durant le service du $(n+1)^{\text{ème}}$ client. Elle ne dépend pas des événements qui se sont produits avant l'instant η_{n+1} (où l'instant 0 en faisant une translation) du début de service du $(n+1)^{\text{ème}}$ client. La distribution de v_{n+1} est la suivante :

$$\mathbb{P}(v_n = i) = a_i = \int_0^\infty \frac{(\lambda x)^i}{i!} \exp(-\lambda x) dB(x),$$

où $a_i > 0, i \geq 0$. On a les résultats suivants

$$\text{si } v = \lim_{n \rightarrow \infty} v_n, E[v] = \rho; \text{ alors } A(z) = \sum_{i=0}^{\infty} a_i z^i = \tilde{B}(\lambda - \lambda z).$$

La variable aléatoire δ_{q_n} est une variable de Bernoulli définie par

$$\delta_{q_n} = \begin{cases} 1 & \text{si le } (n+1)^{\text{ième}} \text{ client servi provient de l'orbite} \\ 0 & \text{si le } (n+1)^{\text{ième}} \text{ client servi est primaire} \end{cases}.$$

Elle dépend de q_n et sa distribution est

$$\begin{aligned} \mathbb{P}(\delta_{q_n} = 1/q_n = i) &= \frac{i\theta}{\lambda + i\theta}, \\ \mathbb{P}(\delta_{q_n} = 0/q_n = i) &= \frac{i}{\lambda + i\theta}. \end{aligned}$$

Les probabilités de transition de l'état i à l'état j ($\forall j \geq 0$ et $0 \leq i \leq j$) sont

$$r_{ij} = \mathbb{P}(q_{n+1} = j/q_n = i) = a_{j-i} \frac{\lambda}{\lambda + i\theta} + a_{j-i+1} \frac{i\theta}{\lambda + i\theta}.$$

La condition d'existence du régime stationnaire peut être obtenue comme suit :
L'accroissement moyen de la chaîne vaut

$$\begin{aligned} E[q_{n+1} - q_n/q_n = i] &= E[v_{n+1}] - E[\delta_{q_n} = 1/q_n = i] \\ &= \rho - \frac{i\theta}{\lambda + i\theta}. \end{aligned}$$

Si $\rho < 1$, alors $\lim_{i \rightarrow \infty} E[q_{n+1} - q_n/q_n = i] = \rho - 1 < 0$ et la chaîne est donc ergodique. Par contre, si $\rho \geq 1$, alors $\lim_{i \rightarrow \infty} E[q_{n+1} - q_n/q_n = i] = \rho - \frac{i\theta}{\lambda + i\theta} \geq 1 - \frac{i\theta}{\lambda + i\theta} = \frac{\lambda}{\lambda + i\theta} > 0$. Puisque la chaîne est bornée inférieurement par la chaîne induite du système M/G/1 classique, donc la chaîne n'est pas ergodique (elle est transitoire). Soit $\pi_n = \lim \mathbb{P}(N_o(\xi_i) = n)$. Les équations de Kolmogorov se présentent de la manière suivante :

$$\pi_n = \sum_{m=0}^n \pi_m \frac{\lambda}{\lambda + m\theta} a_{n-m} + \sum_{m=1}^{n+1} \pi_m \frac{m\theta}{\lambda + m\theta} a_{n-m+1} \text{ et } n = 0, 1, \dots$$

Vu la présence de convolution, cette équation peut être transformée, à l'aide des fonctions génératrices $\varphi(z) = \sum_{n=0}^{\infty} z^n \pi_n$ et $\psi(z) = \sum_{n=0}^{\infty} z^n \frac{\pi_n}{\lambda + n\theta}$,

$$\varphi(z) = A(z) (\lambda \psi(z) + \theta \psi'(z)).$$

D'un autre côté,

$$\begin{aligned}
 \varphi(z) &= \sum_{n=0}^{\infty} z^n \pi_n = \sum_{n=0}^{\infty} z^n \pi_n \frac{\lambda + n\theta}{\lambda + n\theta} \\
 &= \lambda \sum_{n=0}^{\infty} z^n \pi_n \frac{\pi_n}{\lambda + n\theta} + \theta \sum_{n=0}^{\infty} n z^n \frac{\pi_n}{\lambda + n\theta} \\
 &= \lambda \psi(z) + \theta \psi'(z).
 \end{aligned} \tag{3.1}$$

Par conséquent

$$\begin{aligned}
 \lambda \psi(z) + \theta \psi'(z) &= A(z) (\lambda \psi(z) + \theta \psi'(z)), \\
 \theta \psi'(z) [A(z) - z] &= \lambda \psi(z) [1 - A(z)].
 \end{aligned} \tag{3.2}$$

Lemme 3.1 La fonction analytique $f(z) = A(z) - z$ est positive, croissante et pour $z \in [0, 1]$, $\rho < 1 : z < A(z) < 1$.

Démonstration. Soit

$$f(z) = \tilde{B}(\lambda - \lambda z) - z, f(1) = \tilde{B}(0) - 1 = 0.$$

En plus

$$f'(z) = -\lambda \tilde{B}'(\lambda - \lambda z) - 1, \text{ et } f'(1) = \rho - 1 < 0,$$

alors 1 est le seul zéro de f . En outre,

$$f''(z) = -\lambda \tilde{B}''(\lambda - \lambda z) + \lambda^2 \tilde{B}''(\lambda - \lambda z) \geq 0.$$

Alors $f(z)$ est décroissante sur $[0, 1]$, positive pour $\rho = \frac{\lambda}{\gamma} < 1$ et pour $z \in [0, 1]$:

$$z < f(z) < 1.$$

□

Notons aussi que

$$\lim_{z \rightarrow 1^-} \frac{1 - \tilde{B}(\lambda - \lambda z)}{\tilde{B}(\lambda - \lambda z) - z} = \frac{\rho - 1}{1 - \rho} < \infty.$$

Théorème 3.1 Soit $\rho < 1$. La distribution stationnaire de la chaîne de Markov induite possède la fonction génératrice suivante (58)

$$\varphi(z) = \sum_{n=0}^{\infty} z^n \pi_n = \frac{(1 - \rho)(1 - z)A(z)}{A(z) - z} \exp \left\{ \frac{\lambda}{\theta} \int_1^z \frac{1 - A(u)}{A(u) - u} du \right\},$$

où $A(z) = \tilde{B}(\lambda - \lambda z)$.

Démonstration. Le lemme 3.1 nous permet de réécrire l'équation (3.2) pour tout $z \in [0, 1]$ comme suit

$$\psi'(z) = \frac{\lambda}{\theta} \left[\frac{1 - A(z)}{A(z) - z} \right] \psi(z),$$

qui a pour solution

$$\psi(z) = \psi(1) \exp \left[\frac{\lambda}{\theta} \int_1^z \frac{1 - A(u)}{A(u) - u} du \right]$$

Si $\rho < 1$

$$\psi(z) = \psi(1) \exp \left\{ \frac{\lambda}{\theta} \int_1^z \frac{1 - A(u)}{A(u) - u} du \right\}.$$

De (3.1), On a

$$\begin{aligned} \varphi(z) &= \lambda \psi(z) + \theta \psi'(z) \\ &= \lambda \psi(z) + \theta \frac{\lambda}{\theta} z \frac{1 - A(z)}{A(z) - z} \psi(z) \\ &= \lambda \psi(z) A(z) \frac{1 - z}{A(z) - z}. \end{aligned}$$

Puisque $\varphi(1) = 1$, on a $\psi(1) = \sum_{n=0}^{\infty} z^n \frac{\pi_n}{\lambda + n\theta} = \frac{1-\rho}{\lambda}$. Enfin, on obtient la fonction génératrice

$$\varphi(z) = \sum_{n=0}^{\infty} z^n \pi_n = \frac{(1-\rho)(1-z)A(z)}{A(z)-z} \exp \left\{ \frac{\lambda}{\theta} \int_1^z \frac{1 - A(u)}{A(u) - u} du \right\}.$$

□

3.2.3 Distribution stationnaire de l'état du système

Le premier résultat sur le système $M/G/1$ avec rappels a été obtenu par Keilson et al. (1968) (76), basé sur la méthode des variables supplémentaires. Une des approches permettant de trouver la distribution stationnaire jointe de l'état du serveur et de la taille de l'orbite a été introduite par De Kok (1984) (51). Elle consiste à décrire le processus des arrivées comme processus de Markov avec dépendance de l'état de paramètre λ_{in} quand $\{C(t), N_o(t)\}$ est dans l'état (i, n) et à appliquer les schémas récursifs. L'état du système peut être décrit par

le processus

$$X(t) = \begin{cases} N_o(t) & \text{si } C(t) = 0 \\ \{C(t); N_o(t); \xi(t)\} & \text{si } C(t) = 1 \end{cases} ,$$

où $\xi(t)$ est une variable aléatoire supplémentaire à valeurs dans \mathbb{R}^+ , et désignant la durée de service écoulé à la date t . Notons par

$$\begin{aligned} p_{0n} &= \lim_{t \rightarrow \infty} P(C(t) = 0, N_o(t) = n); \\ p_{1n}(x) &= \lim_{t \rightarrow \infty} \frac{d}{dx} P(C(t) = 1, \xi(t) \leq x, N_o(t) = n). \end{aligned}$$

A partir du graphes des transitions 3.1, les probabilités p_{0n} et $p_{1n}(x)$ vérifient le système d'équations de balance :

$$\begin{aligned} (\lambda + n\theta) p_{0n} &= \int_0^{\infty} p_{1n}(x) b(x) dx; \\ p'_{1n}(x) &= -(\lambda + b(x)) p_{1n}(x) + \lambda p_{1n-1}(x); \\ p_{1n}(0) &= \lambda p_{0n} + (n+1) \theta p_{0n+1}; \end{aligned}$$

où $b(x) = B'(x) / (1 - B(x))$ est l'intensité instantanée du service étant donné que la durée écoulée est égale à x .

Soient les fonctions génératrices, telles que $P_0(z) = \sum_{n=0}^{\infty} z^n p_{0n}$ et $P_1(z, x) = \sum_{n=0}^{\infty} z^n p_{1n}(x)$. Le système d'équations de balance devient

$$\left\{ \begin{aligned} \lambda \sum_{n=0}^{\infty} z^n p_{0n} + \theta \sum_{n=0}^{\infty} z^n n p_{0n} &= \int_0^{\infty} \sum_{n=0}^{\infty} z^n p_{1n}(x) b(x) dx; \\ \sum_{n=0}^{\infty} z^n p'_{1n}(x) &= -(\lambda + b(x)) \sum_{n=0}^{\infty} z^n p_{1n}(x) + \lambda \sum_{n=0}^{\infty} z^n p_{1n-1}(x); \\ \sum_{n=0}^{\infty} z^n p_{1n}(0) &= \lambda \sum_{n=0}^{\infty} z^n p_{0n} + \theta \sum_{n=0}^{\infty} z^n (n+1) p_{0n+1} \end{aligned} \right.$$

D'où

$$\left\{ \begin{aligned} \lambda P_0(z) + \theta z P'_0(z) &= \int_0^{\infty} P_1(z, x) b(x) dx; \\ P'_1(z, x) &= (\lambda z - \lambda - b(x)) P_1(z, x); \\ P_1(z, 0) &= \lambda P_0(z) + \theta P'_0(z). \end{aligned} \right. \quad (3.3)$$

De la deuxième équation de (3.3), on a

$$P_1(z, x) = P_1(z, 0) [1 - B(x)] \exp(-(\lambda - \lambda z)x).$$

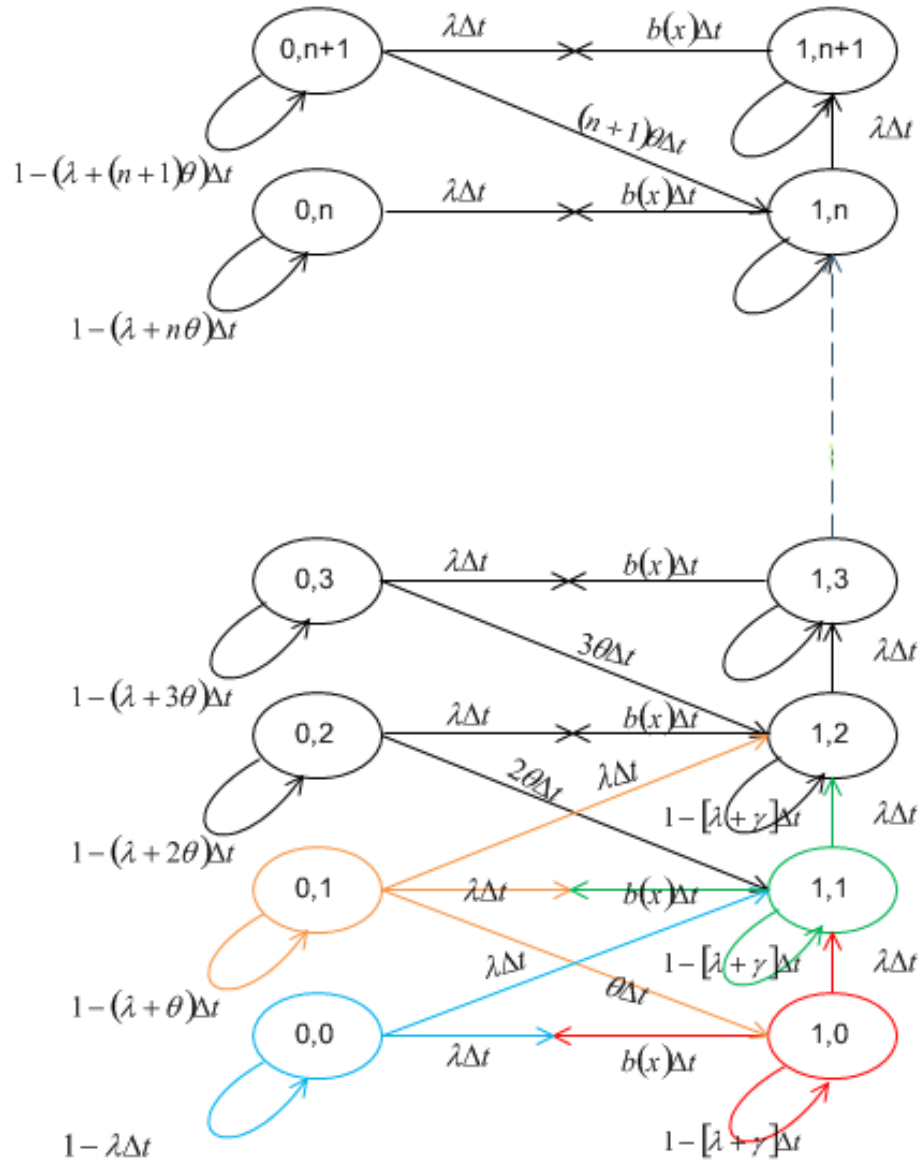


FIGURE 3.1 – Graphe des transitions du modèle M/G/1 avec rappels

Donc, la première équation de (3.3) devient

$$\begin{aligned}\lambda P_0(z) + \theta z P_0'(z) &= \int_0^{\infty} P_1(z, 0) [1 - B(x)] \exp(-(\lambda - \lambda z)x) b(x) dx \\ &= P_1(z, 0) \tilde{B}(\lambda - \lambda z) = P_1(z, 0) A(z).\end{aligned}\quad (3.4)$$

A partir des équations (3.3) et (3.4), on a

$$\begin{aligned}P_1(z, 0) f(z) &= \lambda P_0(z) + \theta z \left(\frac{P_1(z, 0)}{\theta} - \frac{\lambda}{\theta} P_0(z) \right); \\ P_1(z, 0) &= \frac{\lambda - \lambda z}{A(z) - z} P_0(z) [1 - B(x)] \exp(-(\lambda - \lambda z)x).\end{aligned}$$

En intégrant cette équation, et en utilisant la formule $\int_0^{\infty} \exp(-sx) [1 - B(x)] dx = (1 - \tilde{B}(s)) / s$, on obtient

$$P_1(z) = \int_0^{\infty} P_1(z, x) dx = P_0(z) \frac{1 - A(z)}{A(z) - z}.$$

De (3.3) et (3.4), on peut obtenir $P_0(z)$,

$$\lambda P_0(z) + \theta z P_0'(z) = A(z) [\lambda P_0(z) + \theta P_0'(z)]; \quad (3.5)$$

$$\theta [A(z) - z] P_0'(z) = \lambda [1 - A(z)] P_0(z). \quad (3.6)$$

Considérons $f(z) = A(z) - z$. Du lemme 3.1, $f(z)$ est une fonction décroissante sur $[0, 1]$, positive et pour $\rho < 1$ et $z \in [0, 1]$: $z < A(z) < 1$. En plus, $\lim_{z \rightarrow 1^-} \frac{1 - A(z)}{A(z) - z} = \frac{A'(1)}{A(1) - 1} = \frac{\rho}{1 - \rho} < \infty$. De ce fait, pour $z = 1$, la fonction $\frac{1 - A(z)}{A(z) - z} = \frac{\rho}{1 - \rho}$.

Théorème 3.2 Si $\rho = \lambda \beta_1 < 1$, le système est en régime stationnaire et les fonctions génératrices de la distribution conjointe de l'état du serveur et de la taille de l'orbite sont données par

$$\begin{aligned}P_0(z) &= \sum_{n=0}^{\infty} z^n p_{0n} = (1 - \rho) \exp \left[\frac{\lambda}{\theta} \int_1^z \frac{1 - A(u)}{A(u) - u} du \right] \\ P_1(z) &= \sum_{n=0}^{\infty} z^n p_{1n} = \frac{1 - A(z)}{A(z) - z} P_0(z).\end{aligned}$$

Démonstration. L'équation (3.6) devient

$$P_0'(z) = \frac{\lambda}{\theta} \frac{1 - A(z)}{A(z) - z} P_0(z).$$

La résolution de cette équation nous donne

$$\begin{aligned} P_0(z) &= (1 - \rho) \exp \left[\frac{\lambda}{\theta} \int_1^z \frac{1 - A(u)}{A(u) - u} du \right] \\ P_1(z) &= \frac{1 - A(z)}{A(z) - z} P_0(z). \end{aligned}$$

De plus, $P_1(1) = \frac{\rho}{1-\rho} P_0(1)$. Et, vu que $P_0(1) + P_1(1) = 1$, on obtient $P_1(1) = \rho$ et $P_0(1) = 1 - \rho$. \square

Par conséquent, la distribution marginale du nombre de serveurs occupés s'exprime de la manière suivante

$$P_0 = \lim_{t \rightarrow \infty} P(C(t) = 0) = (1 - \rho) P_1 = \lim_{t \rightarrow \infty} P(C(t) = 1) = \rho.$$

La fonction génératrice de la distribution marginale de la taille de l'orbite est définie par :

$$\begin{aligned} P(z) &= P_0(z) + P_1(z) = \frac{(1 - z)}{A(z) - z} P_0(z) \\ &= \frac{(1 - \rho)(1 - z)}{A(z) - z} \exp \left[\frac{\lambda}{\theta} \int_1^z \frac{1 - A(u)}{A(u) - u} du \right], \end{aligned}$$

et la fonction génératrice de la distribution de l'état stationnaire du nombre de clients dans le système est

$$\begin{aligned} Q(z) &= P_0(z) + zP_1(z) \\ &= \frac{(1 - \rho)(1 - z)A(z)}{A(z) - z} \exp \left[\frac{\lambda}{\theta} \int_1^z \frac{1 - A(u)}{A(u) - u} du \right]. \end{aligned}$$

3.2.4 Mesures de performance

Les caractéristiques du modèle sont (116) :

– Nombre moyen de clients dans le système

$$\bar{n} = Q'(1) = \rho + \frac{\lambda^2 \beta_2}{2(1 - \rho)} + \frac{\lambda \rho}{\theta(1 - \rho)};$$

– Nombre moyen de clients en orbite

$$\bar{n}_o = P'(1) = \bar{n} - \rho = \frac{\lambda^2 \beta_2}{2(1 - \rho)} + \frac{\lambda \rho}{\theta(1 - \rho)};$$

– Temps moyen d'attente d'un client

$$\bar{W} = \frac{\bar{n}_o}{\lambda} = \frac{\lambda \beta_2}{2(1 - \rho)} + \frac{\rho}{\theta(1 - \rho)};$$

- Nombre moyen de rappels par client (d'après la formule de Little)

$$\bar{R} = \bar{W}\theta = \frac{\lambda\theta\beta_2}{2(1-\rho)} + \frac{\rho}{1-\rho}.$$

3.3 MODÉLISATION DE L'IMPATIENCE

Dans les systèmes de téléphonie, on rencontre souvent le phénomène suivant : l'abonné quitte le système définitivement (sans recevoir le service) après un certain nombre de tentatives infructueuses d'établir la communication. Dans les systèmes de files d'attente la prise en compte de cet effet se fait à l'aide de la fonction de persistance $H = \{H_k, k \geq 1\}$, où H_k est la probabilité qu'après une $k^{\text{ème}}$ tentative échouée, l'abonné décide d'effectuer une $(k+1)^{\text{ème}}$ tentative. On admet que chaque décision est indépendante du nombre de tentatives précédentes, c'est à dire $H_2 = H_3 = \dots = H_k, \forall k \geq 2$. Les mesures statistiques dans les réseaux de téléphone montrent que ceci est une hypothèse tout à fait réaliste. Les modèles de files d'attente avec clients impatientes ont été étudiés dans (39, 33, 66, 87, 99, 114).

3.3.1 Description du modèle à un serveur

Les clients primaires arrivent dans le système à serveur unique selon un processus de Poisson de paramètre $\lambda > 0$. A l'arrivée d'un client primaire, si le serveur est libre, le client en question est pris en charge. Dans le cas contraire, le client quitte le système définitivement avec une probabilité $1 - H_1$, ou entre en orbite avec une probabilité $H_1 < 1$ et devient source d'appels secondaires. Nous supposons que la probabilité $H_2 = 1$. Les temps de service suivent une loi générale de fonction de répartition $B(x)$, de transformée de Laplace-Stieltjes $\tilde{B}(s), \text{Re}(s) > 0$. Soient les moments $\beta_k = (-1)^k \tilde{B}^{(k)}(0)$, et le taux de service $\gamma = \frac{1}{\beta_1}$. La durée entre deux rappels successifs d'une même source secondaire est exponentiellement distribuée de paramètre $\theta > 0$ qui dépend du nombre de clients de cette source, $T(x) = 1 - e^{-\theta x}$. Le flux de clients primaires, les durées entre deux tentatives successives, ainsi que les temps de service sont supposés mutuellement indépendants.

3.3.2 Distribution stationnaire de l'état du système

L'état du système peut être décrit par le processus

$$X(t) = \begin{cases} N_o(t) & \text{si } C(t) = 0 \\ \{C(t), N_o(t), \xi(t)\} & \text{si } C(t) = 1 \end{cases}$$

$C(t)$ est 0 ou 1 selon le fait que le serveur soit libre ou occupé, $N_o(t)$ est le nombre de clients en orbite, $\xi(t)$ est une variable aléatoire supplémentaire à valeurs dans \mathbb{R}^+ , désignant la durée de service écoulée à la date t . Supposons que le système est en régime stationnaire c'est-à-dire lorsque $\rho = \lambda H_1 \beta_1 < 1$. Notons,

$$\begin{aligned} p_{0n} &= \lim_{t \rightarrow \infty} P(C(t) = 0, N_o(t) = n); \\ p_{1n}(x) &= \lim_{t \rightarrow \infty} \frac{d}{dx} P(C(t) = 1, \xi(t) \leq x, N_o(t) = n). \end{aligned}$$

Les probabilités p_{0n} et $p_{1n}(x)$ vérifient le système d'équations de balance

$$\left\{ \begin{array}{l} (\lambda + n\theta) p_{0n} = \int_0^{\infty} p_{1n}(x) b(x) dx; \\ p'_{1n}(x) = -(\lambda H_1 + b(x)) p_{1n}(x) + \lambda H_1 p_{1n-1}(x); \\ p_{1n}(0) = \lambda p_{0n} + (n+1)\theta p_{0n+1}; \end{array} \right.$$

où $b(x) = B'(x) / (1 - B(x))$ est l'intensité instantanée du service étant donné que la durée écoulée est égale à x . On introduit les fonctions génératrices, telles que $P_0(z) = \sum_{n=0}^{\infty} z^n p_{0n}$ et $P_1(z, x) = \sum_{n=0}^{\infty} z^n p_{1n}(x)$. Le système d'équation de balance devient :

$$\left\{ \begin{array}{l} \lambda P_0(z) + \theta P'_0(z) = \int_0^{\infty} P_1(z, x) b(x) dx; \\ P'_1(z, x) = -(\lambda H_1 (1-z) + b(x)) P_1(z, x); \\ P_1(z, 0) = \lambda P_0(z) + \theta P'_0(z). \end{array} \right. \quad (3.7)$$

De l'équation (3.7), on a

$$P_1(z, x) = P_1(z, 0) [1 - B(x)] \exp(-\lambda H_1 (1-z)x). \quad (3.8)$$

A l'aide de l'équation (3.8), l'équation (3.7) peut être réécrite comme suit

$$\lambda P_0(z) + \theta P'_0(z) = P_1(z, 0) \tilde{B}(\lambda H_1 (1-z)). \quad (3.9)$$

En éliminant $P_1(z, 0)$ des équations (3.7) et (3.9), on obtient

$$\theta [\tilde{B}(\lambda H_1 (1-z)) - z] P'_0(z) = \lambda [1 - \tilde{B}(\lambda H_1 (1-z))] P_0(z). \quad (3.10)$$

Lemme 3.2 Soit $f(z) = A(z) - z$, où $A(z) = \tilde{B}(\lambda H_1 (1-z))$, alors la fonction $f(z)$ est décroissante sur $[0, 1]$ et positive pour $\rho < 1$ et $z \in [0, 1]$, $z < A(z) < 1$.

En plus on a

$$\lim_{z \rightarrow 1^-} \frac{1 - A(z)}{A(z) - z} = \frac{\rho H_1}{1 - \rho H_1} < \infty.$$

Théorème 3.3 *Si $\rho < 1$, le système est en régime stationnaire et les fonctions génératrices de la distribution jointe de l'état du serveur et de la taille de l'orbite sont données par*

$$P_0(z) = \frac{1 - \rho H_1}{1 + \rho(1 - H_1)} \exp \left\{ \frac{\lambda}{\theta} \int_1^z \frac{1 - A(u)}{A(u) - u} du \right\}$$

$$P_1(z) = \int_0^\infty P_1(z, x) dx = \frac{1 - A(z)}{A(z) - z} P_0(z),$$

où $A(z) = \tilde{B}(\lambda H_1(1 - z))$.

Démonstration. Du lemme 3.2, l'équation (3.10) devient

$$P_0'(z) = \frac{\lambda}{\theta} \frac{1 - A(z)}{A(z) - z} P_0(z).$$

Ceci donne

$$P_0(z) = P_0(1) \exp \left\{ \frac{\lambda}{\theta} \int_1^z \frac{1 - A(u)}{A(u) - u} du \right\}.$$

A partir de l'équation (3.7), on déduit que

$$P_1(z, 0) = \lambda \frac{(1 - z)}{A(z) - z} P_0(z).$$

Par conséquent,

$$P_1(z, x) = \lambda \frac{(1 - z)}{A(z) - z} P_0(z) \times [1 - B(x)] \exp(-\lambda H_1(1 - z)x); \quad (3.11)$$

$$P_1(z) = \int_0^\infty P_1(z, x) dx = \frac{1 - A(z)}{A(z) - z} P_0(z). \quad (3.12)$$

La constante $P_0(1)$ peut être trouvée à l'aide de l'équation de normalisation $P_0(1) + P_1(1) = 1$. A partir de l'équation (3.12), nous obtenons $P_1(1) = P_0(1) \frac{\rho}{1 - \rho H_1}$. D'où $P_0(1) = \frac{1 - \rho H_1}{1 + \rho(1 - H_1)}$. \square

Aussi, la fonction génératrice de la distribution marginale du nombre de

clients en orbite est donnée par :

$$\begin{aligned} P(z) &= P_0(z) + P_1(z) \\ &= \frac{1 - \rho H_1}{1 + \rho(1 - H_1)} \frac{(1 - z)}{A(z) - z} \\ &\quad \times \exp \left\{ \frac{\lambda}{\theta} \int_1^z \frac{1 - A(u)}{A(u) - u} du \right\}. \end{aligned}$$

3.3.3 Mesures de performance

– Nombre moyen de clients en orbite

$$\begin{aligned} \bar{n}_0 &= \lim_{t \rightarrow \infty} E[N_0(t)] = P'(1) \\ &= \frac{\lambda^2 H_1}{1 - \rho H_1} \left(\frac{\beta_1}{\theta} + \frac{\beta_2}{2(1 + \rho(1 - H_1))} \right); \end{aligned}$$

– Probabilité de groupage

$$P_1 = P_1(1) = \frac{\rho}{1 + \rho(1 - H_1)}.$$

Les modèles de files d'attente avec clients impatientes ont été étudiés dans (66) et (99).

3.4 SYSTÈMES DE FILES D'ATTENTE $M^X/G/1$ AVEC RAPPELS ET GROUPES IMPATIENTS

La première étude des systèmes de files d'attente avec rappels et arrivées par groupe était faite par Falin (56), qui a supposé la règle suivante : " Si le serveur est occupé à une arrivée, alors la totalité du groupe rejoint l'orbite, et si le serveur est libre, alors un des arrivants commence son service et le reste rejoint l'orbite. A l'arrivée d'un groupe, si le serveur est occupé, la totalité du groupe rejoint l'orbite ; dans le cas contraire, l'un des clients arrivant commence son service et le reste rejoint l'orbite." . Ce modèle peut être utilisé pour évaluer la performance des réseaux locaux à bus opérant sous des protocoles comme le CSMA/CD (Carrier Sense Multiple Access with Collision Detection) (47). Certaines situations de files d'attente avec rappels et arrivées par groupe sont étudiées dans (3, 85, 91, 92, 93). Des contributions récentes sur ce problème incluent les travaux de Aissani (4), Krishna Kumar et Pavai Madheswari (86), Artalejo et Atencia (24) et aussi Atencia et al. (42).

La plus part des travaux sur les files d'attente avec rappels considéraient le

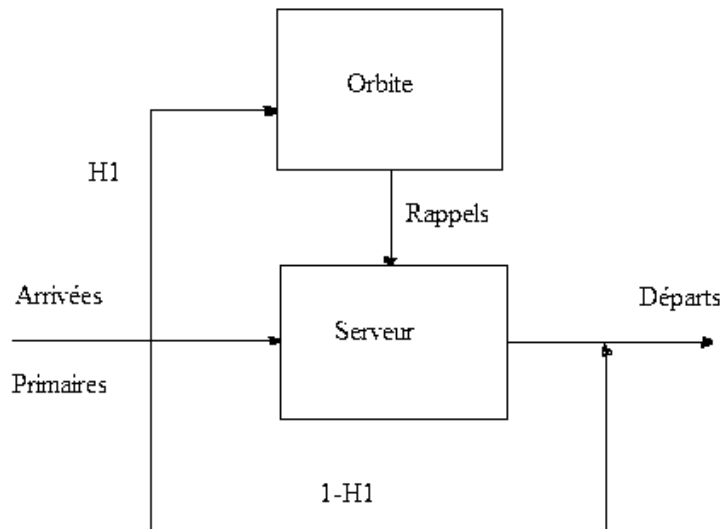


FIGURE 3.2 – Représentation schématique d'un système de files d'attente avec rappels

temps d'attente comme une alternative au modèle classique du réseau téléphonique. Dans ce contexte, chaque client bloqué génère des appels répétés indépendamment du reste des clients en orbite. Alors, dans cette situation, les intervalles entre les essais successifs sont exponentiellement distribués de paramètre θ , quand le nombre de clients en orbite est j . Ce type de discipline de rappels est connu comme une politique de rappels classique étudiée dans (20, 22).

3.4.1 Description du modèle

Les clients primaires arrivent dans le système selon un processus de Poisson de taux $\lambda > 0$. Les clients arrivent par groupes de taille K qui est une variable aléatoire, posons $P(K = k) = c_k(t)$, $k \geq 0$, (la probabilité que la taille du groupe soit égale à k à la date t). Le service est assuré par un seul serveur. A l'arrivée d'un groupe primaire, si le serveur est occupé, le groupe entre en orbite avec une probabilité H_1 , sinon il quitte le système avec une probabilité $(1 - H_1)$, ($H_2 = 1$). Par contre si le serveur est libre, l'un des clients sera pris en charge par le serveur et le reste du groupe entre en orbite. Les clients en orbite répètent les appels jusqu'à ce que le serveur soit libre, et ceci avec un taux de rappel $\theta > 0$, qui dépend du nombre de clients en orbite. Les durées inter-rappels sont exponentiellement distribuées; $T(x) = 1 - e^{-\theta x}$, $x > 0$.

Les durées de service suivent une loi générale $\mathbb{P}(\tau_n^s \leq x) = B(x)$ de transformée de Laplace-Stieltjes $\tilde{B}(s)$, $\text{Re}(s) > 0$. Soient les moments $\beta_k = (-1)^k \tilde{B}^{(k)}(0)$, le taux de service $\gamma = \frac{1}{\beta_1}$.

Soit la fonction génératrice de la distribution stationnaire de la taille des

groupes $C(z) = \sum_{k=1}^{\infty} c_k z^k$ et $\bar{c} = C'(1)$ est la taille moyenne des groupes. L'intensité du trafic est $\rho = \lambda \bar{c} H_1 \beta_1$. Enfin, nous admettons que toutes les variables définies précédemment soient mutuellement indépendantes.

L'état du système est décrit par le processus

$$X(t) = \begin{cases} N_o(t) & \text{si } C(t) = 0 \text{ (serveur libre)} \\ \{C(t), N_o(t), \xi(t)\} & \text{si } C(t) = 1 \text{ (serveur occupé)}, \end{cases}$$

où, $N_o(t)$ est le nombre de clients en orbite, $C(t)$ est l'état du serveur à l'instant t . Nous avons $C(t)$ est égale à 0 ou 1 selon le fait que le serveur est libre ou occupé. Si $C(t) = 1$, $\xi(t)$ représente le temps de service écoulé à la date t .

$$c_k = \lim_{t \rightarrow \infty} c_k(t), k \geq 1.$$

Le phénomène de l'impatience est représentée par la fonction de persistance

$$H = \{H_k, k \geq 1\} \text{ telle que } H_1 < 1 \text{ et } H_2 = H_3 = \dots = 1.$$

Le processus ci-dessus peut être étudié à l'aide de deux manières, soit en utilisant la méthode de la chaîne de Markov induite, soit avec la méthode des variables supplémentaires.

3.4.2 Chaîne de Markov induite

Considérons le processus $\{C(t), N_o(t), t \geq 0\}$ qui n'est pas en général markovien, mais possède un chaîne de Markov induite. Soit $q_n = N_o(\xi_n)$ le nombre de clients en orbite après le $n^{\text{ème}}$ départ. La suite des variables aléatoires $\{q_n, n \geq 1\}$ forme une chaîne de Markov induite, dont l'équation fondamentale est

$$q_{n+1} = q_n - \delta_{q_n} + v_{n+1}. \quad (3.13)$$

La variable aléatoire v_{n+1} représente le nombre de clients primaires arrivant dans le système durant le service du $(n+1)^{\text{ème}}$ client. Elle ne dépend pas des événements qui se sont produits avant l'instant du début de service du $(n+1)^{\text{ème}}$ client.

Théorème 3.4 La distribution du nombre de clients primaires arrivant dans le système durant un service est donnée par

$$\mathbb{P}(v_n = i) = a_i = \int_0^{\infty} \sum_j \left(\frac{\lambda H_1 x}{j!} \right)^j \exp(-\lambda H_1 x) c_i^{(j)} dB(x)$$

où $c_i^{(j)}$ est le $j^{\text{ème}}$ produit de convolution de la suite c_i et $a_i > 0, \forall i \geq 0$. Sa fonction génératrice est définie par

$$A(z) = \sum_{n=0}^{\infty} a_n z^n = \tilde{B}(\lambda H_1 (1 - C(z))).$$

En outre,

$$\mathbb{E}[v_n] = \sum_{i=0}^{\infty} i a_i = \rho.$$

Démonstration. Si on a l clients qui arrivent, alors seulement j d'entre eux peuvent entrer en orbite (à cause de l'impatience). De ce fait, la probabilité du nombre de clients primaires qui arrivent durant le $(n+1)^{\text{ème}}$ service est :

$$\begin{aligned} \mathbb{P}(v_n = i) &= \int_0^{\infty} \sum_{l=0}^{\infty} \sum_{j=0}^l C_l^j H_1^j (1 - H_1)^{l-j} c_i^{(j)} \frac{(\lambda \xi)^l}{l!} \exp(-\lambda \xi) dB(\xi) \\ &= \int_0^{\infty} \sum_{j=0}^{\infty} \sum_{l=0}^j \frac{l!}{j! (l-j)!} H_1^j (1 - H_1)^{l-j} c_i^{(j)} \frac{(\lambda \xi)^l}{l!} \exp(-\lambda \xi) dB(\xi). \end{aligned}$$

Posons : $l' = l - j$, alors

$$\mathbb{P}(v_n = i) = \int_0^{\infty} \sum_{l'=0}^{\infty} \left[\frac{1}{l'!} (1 - H_1)^{l'} (\lambda \xi)^{l'} \right] \frac{1}{j!} H_1^j (\lambda \xi)^j \exp(-\lambda \xi) . c_i^{(j)} dB(\xi).$$

D'où :

$$\begin{aligned} \mathbb{P}(v_n = i) &= \int_0^{\infty} \sum_{j=0}^{\infty} \exp((1 - H_1) \lambda \xi) \frac{1}{j!} H_1^j (\lambda \xi)^j \exp(-\lambda \xi) . c_i^{(j)} dB(\xi) \\ &= \int_0^{\infty} \sum_{j=0}^{\infty} \frac{(H_1 \lambda \xi)^j}{j!} \exp(-\lambda H_1 \xi) . c_i^{(j)} dB(\xi). \end{aligned}$$

□

La variable aléatoire δ_{q_n} est définie par

$$\delta_{q_n} = \begin{cases} +1 & \text{si le } (n+1)^{\text{ème}} \text{ client servi provient de l'orbite} \\ -(l-1) & \text{si le } (n+1)^{\text{ème}} \text{ client servi provient d'une arrivée} \\ & \text{primaire avec la taille du groupe égale à } l. \end{cases}$$

Elle dépend seulement de q_n et sa probabilité conditionnelle est donnée par

$$\mathbb{P}(\delta_{q_n} = +1 / q_n = i) = \frac{i\theta}{\lambda + i\theta}, \mathbb{P}(\delta_{q_n} = -(l-1) / q_n = i) = \frac{i}{\lambda + i\theta}.$$

Les probabilités de transition de i à j ($\forall j \geq 0$ et $0 \leq i \leq j$) de notre chaîne sont

$$\begin{aligned} r_{ij} &= \mathbb{P}(q_{n+1} = j/q_n = i) = \mathbb{P}(q_n - \delta_{q_n} + v_{n+1} = j/q_n = i) \\ &= \mathbb{P}(v_{n+1} = j - i + \delta_{q_n}/q_n = i) \\ &= c_l \mathbb{P}(v_{n+1} = j - i - (l - 1)/q_n = i, \delta_{q_n} = -(l - 1)) \mathbb{P}(\delta_{q_n} = -(l - 1)/q_n = i) \\ &\quad + \mathbb{P}(v_{n+1} = j - i - (l - 1)/q_n = i, \delta_{q_n} = +1) \mathbb{P}(\delta_{q_n} = +1/q_n = i). \end{aligned}$$

$$r_{ij} = \mathbb{P}(q_{n+1} = j/q_n = i) = \sum_{l=1}^{j-i+1} c_l a_{j-i-l+1} \frac{\lambda}{\lambda + i\theta} + a_{j-i+1} \frac{i\theta}{\lambda + i\theta}. \quad (3.14)$$

Notons que $r_{ij} \neq 0, \forall i = 0, 1, 2, \dots, j + 1$. La question qui se pose est l'ergodicité de la chaîne. De (3.13), $\{q_n, n \geq 1\}$ est une chaîne irréductible et apériodique. Nous utilisons le critère de Foster (49) pour trouver la condition suffisante. Selon le critère en question, il faut montrer l'existence d'une fonction non négative $f(i), i \in S$, (dans notre cas $f(q_n) = i$) et d'un réel $\varepsilon > 0$ tels que l'accroissement moyen de la chaîne de Markov induite $x_i = E[q_{n+1} - q_n/q_n = i] < -\varepsilon$ pour tout $i \in S$, à l'exception peut être d'un ensemble fini. On a :

$$\begin{aligned} x_i &= E[q_{n+1} - q_n/q_n = i] \\ &= E[-\delta_{q_n}/q_n = i] + E[v_{n+1}/q_n = i] \\ &= -E[\delta_{q_n} = 1/q_n = i] + E[v_{n+1}] \\ x_n &= -\frac{i\theta}{\lambda + i\theta} + \rho. \end{aligned}$$

Soit $x = \lim_{i \rightarrow \infty} x_i$, alors $x = \rho - 1 < 0$. Ainsi, $\rho < 1$ est la condition suffisante pour l'ergodicité de notre chaîne. Pour montrer qu'elle est aussi une condition nécessaire, nous appliquons la condition de Kaplan : Une chaîne de Markov n'est pas ergodique si $x_i < \infty$ pour tout $i \geq 0$ et il existe $i_0 \in \mathbb{N}$ tel que $x_i \geq 0$ pour $i \geq i_0$ (108). Dans notre cas, cette condition est vérifiée car $r_{ij} = 0$ pour $j < i - 1$ et $i > 0$ (voir(3.14)). Par conséquent, si $\rho > 1$, la chaîne de Markov induite $\{q_n, n \geq 1\}$ n'est pas ergodique. Enfin, $\{q_n, n \geq 1\}$ est ergodique si et seulement si $\rho < 1$. Notre second but est de trouver π_n la distribution stationnaire de la chaîne de Markov induite $\{q_n\}$. Soit

$$\pi_n = \lim_{i \rightarrow \infty} \mathbb{P}(N_o(\xi_i) = n).$$

De (49), les probabilités limites $\pi_n = \lim_{i \rightarrow \infty} \mathbb{P}(q_i = n)$ existent et sont positives si la chaîne de Markov induite $\{q_n, n \geq 1\}$ est ergodique. Pour $\rho < 1$ les équations

de Kolmogorov π_n se présentent de la manière suivante

$$\pi_n = \sum_{m=0}^n \pi_m \frac{\lambda}{\lambda + m\theta} \sum_{i=1}^{n-m+1} c_i a_{n-m-i+1} + \sum_{m=1}^{n+1} \pi_m \frac{m\theta}{\lambda + m\theta} a_{n-m+1}, \forall n \geq 0. \quad (3.15)$$

Vu la présence de convolution, nous introduisons les fonctions génératrices

$$\varphi(z) = \sum_{n=0}^{\infty} z^n \pi_n, \text{ et } \psi(z) = \sum_{n=0}^{\infty} z^n \frac{\pi_n}{\lambda + n\theta}.$$

L'équation (3.15) devient

$$\begin{aligned} \varphi(z) &= \sum_{n=0}^{\infty} \sum_{m=0}^n z^n \pi_m \frac{\lambda}{\lambda + m\theta} \sum_{i=1}^{n-m+1} c_i a_{n-m-i+1} + \sum_{n=0}^{\infty} \sum_{m=0}^{n+1} z^n \pi_m \frac{m\theta}{\lambda + m\theta} a_{n-m+1} \\ &= \lambda \sum_{m=0}^{\infty} \frac{z^m \pi_m}{\lambda + m\theta} \sum_{i=1}^{\infty} c_i z^{i-1} \sum_{i=n-m+1}^{\infty} z^{n-m-i+1} a_{n-m-i+1} \\ &\quad + \theta \sum_{m=1}^{\infty} \frac{z^{m-1} \pi_m}{\lambda + m\theta} \sum_{n=m-1}^{\infty} a_{n-m+1} z^{n-m+1}; \end{aligned}$$

Enfin,

$$\varphi(z) = \tilde{B}(\lambda H_1(1 - C(z))) \left[\lambda \frac{C(z)}{z} \psi(z) + \theta \psi'(z) \right].$$

D'un autre côté, nous avons

$$\begin{aligned} \varphi(z) &= \sum_{n=0}^{\infty} z^n \pi_n = \sum_{n=0}^{\infty} z^n \pi_n \frac{\lambda + n\theta}{\lambda + n\theta} \\ &= \lambda \sum_{n=0}^{\infty} z^n \frac{\pi_n}{\lambda + n\theta} + \theta z \sum_{n=0}^{\infty} n z^{n-1} \frac{\pi_n}{\lambda + n\theta}; \end{aligned}$$

et par suite,

$$\varphi(z) = \lambda \psi(z) + \theta z \psi'(z). \quad (3.16)$$

D'où

$$\varphi(z) = \tilde{B}(\lambda H_1(1 - C(z))) \left[\lambda \frac{C(z)}{z} \psi(z) + \theta \psi'(z) \right]$$

et

$$\psi'(z) \left[\tilde{B}(\lambda H_1(1 - C(z))) - z \right] \theta = \psi(z) \lambda \left[1 - \tilde{B}(\lambda H_1(1 - C(z))) \frac{C(z)}{z} \right]. \quad (3.17)$$

Lemme 3.3 La fonction $f(z) = A(z) - z$ est décroissante, positive; et pour $\rho < 1$ et $z \in [0, 1]$, on a $z \leq \tilde{B}(\lambda H_1(1 - C(z))) \leq 1$, où $A(z) = \tilde{B}(\lambda H_1(1 - C(z)))$.

Démonstration. Nous avons $f(1) = \tilde{B}(0) - 1 = 0$, et

$$f'(z) = -\lambda H_1 C'(z) \tilde{B}'(\lambda H_1 (1 - C(z))) - 1.$$

Donc $f'(1) = \rho - 1 < 0$, et

$$f''(z) = -\lambda H_1 C''(z) \tilde{B}'(\lambda H_1 (1 - C(z))) + \lambda^2 H_1^2 (C'(z))^2 \tilde{B}''(\lambda H_1 (1 - C(z))) \geq 0.$$

Alors $f(z)$ est décroissante sur $[0, 1]$, positive pour $\rho = \frac{\lambda H_1 \bar{c}}{\gamma} < 1$ et pour $z \in [0, 1] : z < A(z) < 1$. \square

En plus on a

$$\lim_{z \rightarrow 1^-} \frac{1 - A(z) \frac{C(z)}{z}}{A(z) - z} = \frac{\rho + \bar{c} - 1}{1 - \rho} < \infty.$$

Théorème 3.5 Soit $\rho < 1$. La distribution stationnaire de la chaîne de Markov induite possède la fonction génératrice suivante

$$\varphi(z) = \frac{1 - \rho}{\bar{c}\rho} \frac{A(z)(1 - C(z))}{A(z) - z} \exp \left[\frac{\lambda}{\theta} \int_1^z \frac{1 - A(u) \frac{C(u)}{u}}{A(u) - u} du \right],$$

où $A(z) = \tilde{B}(\lambda H_1 (1 - C(z)))$.

Démonstration. Du lemme 3.3 nous pouvons réécrire l'équation (3.17) pour tout $z \in [0, 1]$ comme suit

$$\psi'(z) = \frac{\lambda}{\theta} \left[\frac{1 - A(z) \frac{C(z)}{z}}{A(z) - z} \right] \psi(z),$$

qui a pour solution

$$\psi(z) = \psi(1) \exp \left[\frac{\lambda}{\theta} \int_1^z \frac{1 - A(u) \frac{C(u)}{u}}{A(u) - u} du \right].$$

De (3.16), nous avons

$$\begin{aligned} \varphi(z) &= \lambda \psi(z) + \theta z \psi'(z) \\ &= \lambda \psi(z) + \lambda z \left[\frac{1 - A(z) \frac{C(z)}{z}}{A(z) - z} \right] \psi(z) \\ &= \frac{\lambda A(z)(1 - C(z))}{A(z) - z} \psi(z). \end{aligned}$$

De l'équation de normalisation $\varphi(1) = 1$, nous obtenons

$$\begin{aligned} 1 &= \lim_{z \rightarrow 1} \frac{-\lambda^2 H_1 A'(z) C(z) - C'(z) \lambda A(z)}{-\lambda H_1 C'(z) A'(z) - 1} \psi(1); \\ 1 &= \frac{-\lambda^2 \beta_1 H_1 \bar{c}^2}{\lambda \beta_1 H_1 \bar{c} - 1} \psi(1). \end{aligned}$$

D'où $\psi(1) = \frac{1-\rho}{\lambda \bar{c}}$. Enfin, nous obtenons la fonction génératrice $\varphi(z) = \sum_{n=0}^{\infty} \pi_n z^n$ de la distribution stationnaire de la chaîne de Markov induite $\{q_n, n \geq 1\}$ aux moments de départs :

$$\varphi(z) = \frac{1-\rho}{\bar{c}\rho} \frac{A(z)(1-C(z))}{A(z)-z} \exp \left[\frac{\lambda}{\theta} \int_1^z \frac{1-A(u) \frac{C(u)}{u}}{A(u)-u} du \right]. \quad (3.18)$$

□

3.4.3 Distribution stationnaire de l'état du système

Nous obtenons la distribution stationnaire du processus $\{C(t), N_o(t), \xi(t), t \geq 0\}$ à l'aide de la méthode des variables supplémentaires. Posons $\rho < 1$ et introduisons

$$\begin{aligned} p_{0,n} &= \lim_{t \rightarrow \infty} P(C(t) = 0, N_o(t) = n); \\ p_{1,n}(x) &= \lim_{t \rightarrow \infty} \frac{d}{dx} P(C(t) = 1, \xi(t) \leq x, N_o(t) = n); \\ p_{1,n} &= \lim_{t \rightarrow \infty} P(C(t) = 1, N_o(t) = n) = \int_0^{\infty} p_{1,n}(x) dx. \end{aligned}$$

A partir du graphe de transition (3.3), nous obtenons le système d'équations de balance suivant :

$$(I) \begin{cases} (\lambda + n\theta) p_{0n} = \int_0^{\infty} p_{1n}(x) b(x) dx; \\ p'_{1,n}(x) = -(\lambda H_1 + b(x)) p_{1,n}(x) + \lambda H_1 \sum_{k=0}^n c_k p_{1,n-k}(x) \\ p_{1,n}(0) = \lambda \sum_{k=1}^{n+1} c_k p_{0,n-k+1} + (n+1) \theta p_{0,n+1}. \end{cases}$$

où $b(x) = \frac{B'(x)}{1-B(x)}$ est l'intensité instantanée du service étant donnée que la durée de service écoulé est égale à x . Pour la résolution de ce système on introduit les fonctions génératrices telles que : $P_0(z) = \sum_{n=0}^{\infty} p_{0,n} z^n$ et $P_1(z, x) = \sum_{n=0}^{\infty} p_{1,n}(x) z^n$. Le système (I) devient

$$\begin{cases} \lambda \sum_{n=0}^{\infty} p_{0,n} z^n + \theta \sum_{n=0}^{\infty} n p_{0,n} z^n = \int_0^{\infty} \sum_{n=0}^{\infty} z^n p_{1,n}(x) b(x) dx; \\ \sum_{n=0}^{\infty} z^n p'_{1,n}(x) = -(\lambda H_1 + b(x)) \sum_{n=0}^{\infty} z^n p_{1,n}(x) + \lambda H_1 \sum_{n=0}^{\infty} z^n \sum_{k=0}^n c_k p_{1,n-k}(x) \\ \sum_{n=0}^{\infty} z^n p_{1,n}(0) = \lambda \sum_{n=0}^{\infty} z^n \sum_{k=1}^{n+1} c_k p_{0,n-k+1} + \theta \sum_{n=0}^{\infty} z^n (n+1) p_{0,n+1}. \end{cases}$$

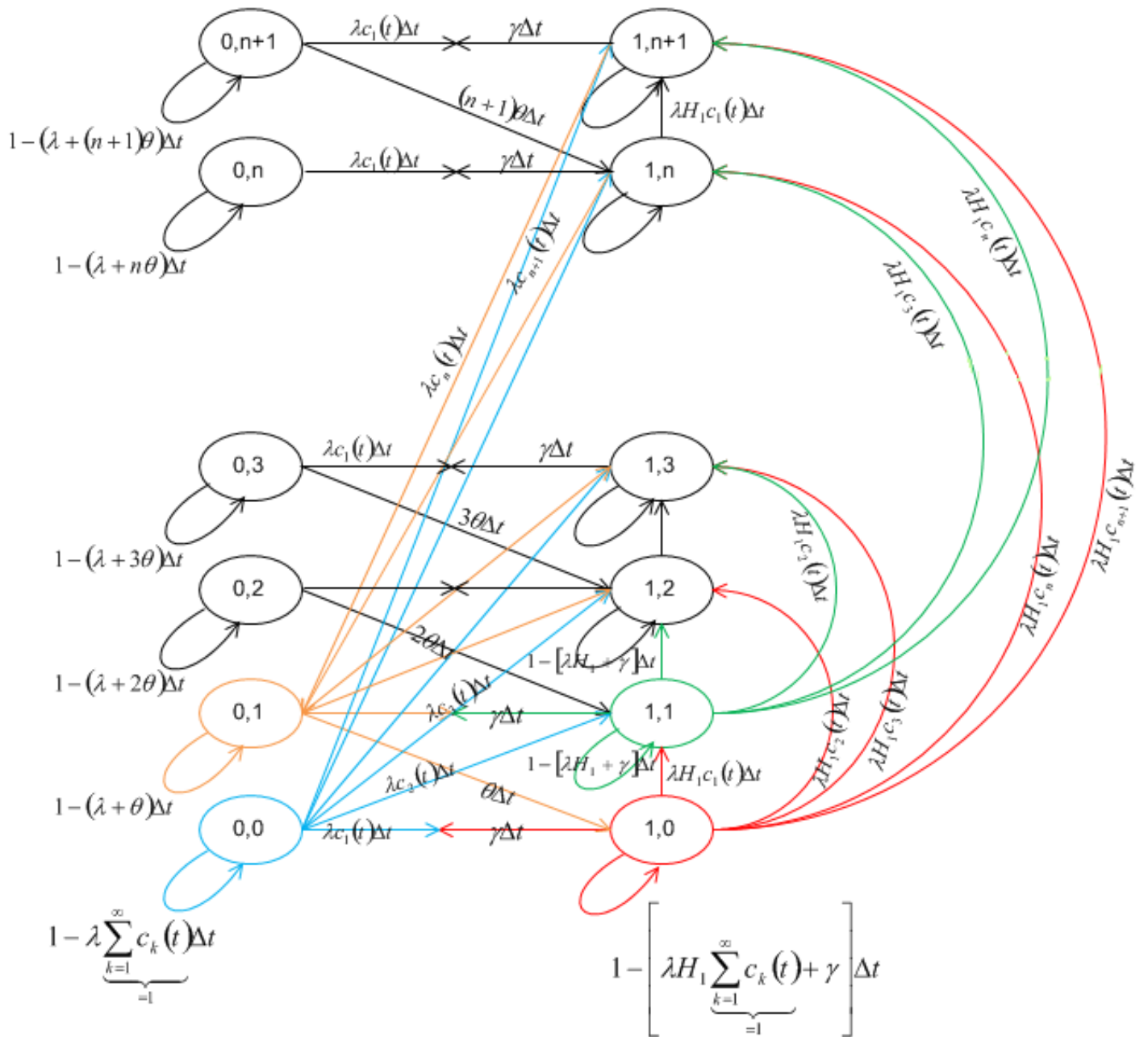


FIGURE 3.3 – Graphe des transitions du modèle $M^X/G/1$ avec rappels et impatience

D'où

$$\begin{cases} \lambda P_0(z) + \theta z P_0'(z) = \int_0^\infty P_1(z, x) b(x) dx; \\ P_1'(z, x) = -(\lambda H_1 + b(x)) P_1(z, x) + \lambda H_1 C(z) P_1(z, x) \\ P_1(z, 0) = \frac{\lambda}{z} C(z) P_0(z) + \theta P_0'(z). \end{cases} \quad (3.19)$$

De la deuxième équation du système (3.19), on a

$$P_1'(z, x) = -[\lambda H_1 (1 - C(z)) + b(x)] P_1(z, x),$$

d'où

$$\frac{P_1'(z, x)}{P_1(z, x)} = -\lambda H_1 (1 - C(z)) - \frac{B'(x)}{1 - B(x)}.$$

Alors,

$$P_1(z, x) = P_1(z, 0) (1 - B(x)) \exp[-\lambda H_1 (1 - C(z)) x]. \quad (3.20)$$

Donc la première équation du système (3.19) devient

$$\begin{aligned} \lambda P_0(z) + \theta z P_0'(z) &= \int_0^\infty P_1(z, 0) (1 - B(x)) \exp[-\lambda H_1 (1 - C(z)) x] b(x) dx \\ &= P_1(z, 0) \int_0^\infty B'(x) \exp[-\lambda H_1 (1 - C(z)) x] dx \\ &= P_1(z, 0) \int_0^\infty \exp[-\lambda H_1 (1 - C(z)) x] dB(x); \\ \lambda P_0(z) + \theta z P_0'(z) &= P_1(z, 0) A(z), \end{aligned} \quad (3.21)$$

où $A(z) = \tilde{B}(\lambda H_1 (1 - C(z)))$. De la troisième équation du système (3.19), l'équation (3.21) donne

$$\begin{aligned} \lambda P_0(z) + \theta z P_0'(z) &= \left[\frac{\lambda}{z} C(z) P_0(z) + \theta P_0'(z) \right] A(z) \\ \theta [A(z) - z] P_0'(z) &= \lambda \left[1 - A(z) \frac{C(z)}{z} \right] P_0(z). \end{aligned} \quad (3.22)$$

Théorème 3.6 Si $\rho = \lambda \bar{c} H_1 \beta_1 < 1$, et le système est en régime stationnaire alors les fonctions génératrices de la distribution jointe de l'état du serveur et de la taille de l'orbite sont données par

$$\begin{aligned} P_0(z) &= \sum_{n=0}^{\infty} z^n p_{0n} = \frac{H_1 (1 - \rho)}{\rho + H_1 (1 - \rho)} \exp \left[\frac{\lambda}{\theta} \int_1^z \frac{1 - A(u) \frac{C(u)}{u}}{A(u) - u} du \right] \\ P_1(z) &= \sum_{n=0}^{\infty} z^n p_{1n} = \frac{1 - A(z)}{(A(z) - z) H_1} P_0(z). \end{aligned}$$

Démonstration. Du lemme 3.3, l'équation (3.22) devient

$$\frac{P_0'(z)}{P_0(z)} = \frac{\lambda}{\theta} \left[\frac{1 - A(z) \frac{C(z)}{z}}{A(z) - z} \right] \quad (3.23)$$

et

$$\lim_{z \rightarrow 1^-} \frac{1 - A(z) \frac{C(z)}{z}}{A(z) - z} = \frac{\rho + \bar{c} - 1}{1 - \rho} < \infty.$$

De (3.23), on a la première fonction génératrice partielle

$$\begin{aligned} P_0(z) &= P_0(1) \exp \left[\frac{\lambda}{\theta} \int_1^z \frac{1 - A(u) \frac{C(u)}{u}}{A(u) - u} du \right] \\ &= \frac{H_1(1 - \rho)}{\rho + H_1(1 - \rho)} \exp \left[\frac{\lambda}{\theta} \int_1^z \frac{1 - A(u) \frac{C(u)}{u}}{A(u) - u} du \right]. \end{aligned}$$

De plus, la troisième équation du système (3.19) peut être réécrite comme suit

$$\begin{aligned} P_1(z, 0) &= \frac{\lambda}{z} C(z) P_0(z) + \lambda \left[\frac{1 - A(z) \frac{C(z)}{z}}{A(z) - z} \right] P_0(z) \\ &= \lambda \left[\frac{1 - C(z)}{A(z) - z} \right] P_0(z) \end{aligned}$$

et l'équation (3.20) donne

$$P_1(z, x) = \lambda \frac{[1 - C(z)][1 - B(x)]}{A(z) - z} P_0(z) \exp[-\lambda H_1(1 - C(z))x].$$

D'où

$$\begin{aligned} P_1(z) &= \sum_{n=0}^{\infty} z^n P_{1,n} = \int_0^{\infty} P_1(z, x) dx \\ &= \lambda \frac{(1 - C(z)) P_0(z)}{A(z) - z} \int_0^{\infty} [1 - B(x)] \exp[-\lambda H_1(1 - C(z))x] dx. \end{aligned}$$

Alors, la deuxième fonction génératrice partielle est obtenue :

$$\begin{aligned} P_1(z) &= \lambda \frac{(1 - C(z)) P_0(z)}{A(z) - z} \frac{1 - A(z)}{\lambda H_1(1 - C(z))} \\ &= \frac{1 - A(z)}{(A(z) - z) H_1} P_0(z). \end{aligned} \quad (3.24)$$

□

A présent, il est possible de calculer la fonction génératrice de la distribution

stationnaire marginale du nombre de clients en orbite

$$\begin{aligned} P(z) &= P_0(z) + P_1(z) \\ &= \frac{(1 - H_1 z) - A(z)(1 - H_1)}{(A(z) - z)} \\ &\quad \times \frac{(1 - \rho)}{\rho + H_1(1 - \rho)} \exp \left[\frac{\lambda}{\theta} \int_1^z \frac{1 - A(u) \frac{C(u)}{u}}{A(u) - u} du \right], \quad (3.25) \end{aligned}$$

ainsi que la fonction génératrice de la distribution de stationnaire du nombre de clients dans le système

$$\begin{aligned} Q(z) &= P_0(z) + zP_1(z) \\ &= \frac{H_1(A(z) - z) + z(1 - A(z))}{(A(z) - z)} \\ &\quad \times \frac{(1 - \rho)}{\rho + H_1(1 - \rho)} \exp \left[\frac{\lambda}{\theta} \int_1^z \frac{1 - A(u) \frac{C(u)}{u}}{A(u) - u} du \right] \end{aligned}$$

Il faut noter que cette fonction génératrice n'est pas identique à celle obtenue pour la distribution stationnaire de la chaîne de Markov induite (3.18). Nous avons aussi la distribution marginale du nombre de serveurs occupés

$$\begin{aligned} P_0 &= \lim_{t \rightarrow \infty} P(C(t) = 0) = P_0(1) = \frac{H_1(1 - \rho)}{\rho + H_1(1 - \rho)}; \\ P_1 &= \lim_{t \rightarrow \infty} P(C(t) = 1) = P_1(1) = \frac{\rho}{\rho + H_1(1 - \rho)}. \end{aligned}$$

3.4.4 Mesures de performance

– Nombre moyen de clients dans le système

$$\begin{aligned} \bar{n} &= \lim_{t \rightarrow \infty} E[C(t) + N_O(t)] = Q'(1) \\ &= \rho + \frac{\lambda^2 H_1^2 (C'(1))^2 \beta_2 + \rho C''(1) / C'(1)}{2(1 - \rho)} + \frac{\lambda \rho + \bar{c} - 1}{\theta (1 - \rho)}; \end{aligned}$$

– Nombre moyen de clients en orbite

$$\begin{aligned} \bar{n}_O &= \lim_{t \rightarrow \infty} E[N_O(t)] = P'(1) \\ &= \frac{\lambda^2 H_1^2 \bar{c}^2 \beta_2 + \rho C''(1) / \bar{c}}{2(1 - \rho)} + \frac{\lambda \rho + \bar{c} - 1}{\theta (1 - \rho)}; \end{aligned}$$

– Temps moyen d'attente d'un client

$$\bar{W} = \frac{\bar{n}_O}{\lambda H_1 \bar{c}} = \frac{\lambda H_1 \bar{c} \beta_2 + \rho C''(1) / \lambda H_1 \bar{c}^2}{2(1 - \rho)} + \frac{1}{\theta H_1 \bar{c}} \frac{\rho + \bar{c} - 1}{1 - \rho};$$

– Nombre moyen de rappels par client

$$\bar{R} = \theta \bar{W} = \theta \frac{\lambda H_1 \bar{c} \beta_2 + \rho C''(1) / \lambda H_1 \bar{c}^2}{2(1 - \rho)} + \frac{1}{H_1 \bar{c}} \frac{\rho + \bar{c} - 1}{1 - \rho}.$$

3.4.5 Exemples d'application

Exemple 3.1 Dans les réseaux locaux (LAN), l'un des protocoles de communication les plus utilisés est CSMA (Carrier-Sence Multiple Acces) non-persistant. Supposons qu'un réseau local est composé de n stations connectées par un seul bus. La communication entre les stations est réalisée au moyen de ce bus. Les messages de longueur variables arrivent aux stations du monde extérieur. En recevant le message, la station le découpe en un nombre fini de paquets de longueur fixe et consulte le bus pour voir s'il est occupé. Si le bus est libre, l'un des paquets est transmis via ce bus à la station de destination, et les autres paquets sont stockés dans les tampons pour transmission ultérieure. Autrement, tous les paquets sont stockés dans le tampon et la station peut consulter le bus après une certaine durée aléatoire. Les questions concernant ce problème sont : Quel est le temps moyen d'attente d'un paquet ? Quel est le nombre moyen de messages (paquets) dans le tampon d'une station ? Si les messages arrivent selon un processus de Poisson, le système peut être modélisé comme un système M/G/1 avec rappels et arrivées par groupes. Le serveur est le bus et les tampons des stations représentent l'orbite. Si la capacité des tampons est très grande, on a un système de files d'attente avec rappels, arrivées par groupes et capacité de l'orbite infinie.

Exemple 3.2 Dans le modèle opérationnel du serveur WWW, les requêtes HTTP arrivent au serveur selon un processus de Poisson, il faut noter que ces requêtes doivent être définies par un utilisateur avant l'arrivée au serveur. Quand les requêtes arrivent au serveur, une requête est prise en charge par le serveur et le reste entre dans un tampon qui est à l'intérieur du serveur. Dans le tampon, chaque requête attend une certaine durée de temps puis de nouveau rappelle le serveur. Il existe un programme implémenté dans le serveur WWW pour gérer les requêtes du tampon. Après chaque rappel échoué, il fait une nouvelle tentative pour tenter sa chance après une durée de temps aléatoire. Si la page Web ciblée se trouve dans le même serveur WWW, la requête peut rester dans le serveur. Pour garder le serveur WWW fonctionnel, des maintenances telles que des anti-virus peuvent être utilisés quand le serveur est libre. Ce type de maintenance peut être programmé pour s'exécuter régulièrement. Néanmoins, ces maintenance ne se répètent pas continuellement. Quand ces activités sont terminées, le serveur WWW devient libre et attend l'arrivée de nouvelles requêtes. Dans la terminologie des files d'attente, le tampon, le serveur WWW, la retransmission et les activités de maintenance

(quand le serveur est libre) correspondent respectivement à l'orbite, le serveur, la discipline de rappel et à la vacation.

CONCLUSION DU CHAPITRE

Nous avons établi l'analyse stochastique complète du modèle $M^X/G/1$ avec rappels, arrivées par groupes et clients impatientes. Nous avons mis en évidence la chaîne de Markov induite associée et donné la distribution du nombre de clients primaires arrivant dans le système durant un service donné. Par la méthode des variables supplémentaires et à l'aide du graphe des transitions nous avons obtenu les fonctions génératrices de la distribution conjointe de l'état du serveur et de la taille de l'orbite, ainsi que la fonction génératrice de la distribution stationnaire du nombre de clients dans le système et la distribution marginale du nombre de serveurs occupés. Enfin, nous avons établi aussi les mesures de performances de ce modèle.

PROPRIÉTÉ DE DÉCOMPOSITION STOCHASTIQUE POUR LE NOMBRE DE CLIENTS DANS LE SYSTÈME

4.1 INTRODUCTION

Parmi les approches permettant d'étudier les systèmes de files d'attente avec rappels, on rencontre celles basées sur la propriété de décomposition stochastique que peut posséder un modèle. Le concept général de la propriété de décomposition stochastique (PDS) d'un système d'attente $M/G/1$ est défini de la manière suivante : le nombre de clients se trouvant dans le système à une date aléatoire est distribué comme la somme de deux variables aléatoires indépendantes ou plus ; l'une de ces variables représente le nombre de clients se trouvant dans le système $M/G/1$ ordinaire à une date aléatoire (le serveur est toujours disponible). Les systèmes évoqués sont en régime stationnaire. Ce type de décomposition a été observé auparavant pour les systèmes d'attente avec vacances. Ces derniers sont caractérisés par le fait que le temps inoccupé du serveur peut être utilisé pour des tâches extérieures (par exemple : tâches prioritaires ou de maintenance) (54). Dans un certain sens, les modèles d'attente avec rappels peuvent être considérés comme un type particulier des modèles avec vacances, où les vacances commencent après chaque service, et leur durée dépend du processus des arrivées et de l'état du système.

Dans ce chapitre, nous réalisons une synthèse bibliographique sur la propriété de décomposition stochastique des systèmes d'attente avec rappels et établissons la propriété en question pour le modèle $M^X/G/1$ avec rappels et clients impatientes.

4.2 PROPRIÉTÉ DE DÉCOMPOSITION STOCHASTIQUE

4.2.1 Modèles d'attente avec vacances

On distingue les vacances du serveur dans le cas d'un service exhaustif et celles dans le cas d'un service non-exhaustif. Dans la première situation, le serveur prend ces vacances lorsque le système est vide. En outre, le serveur est en vacances jusqu'à ce qu'il trouve (en revenant des vacances) au moins un client dans le système (vacances multiples), ou bien il prend une seule vacation après chaque période d'activité (vacation unique). Dans la seconde situation, le serveur est autorisé à prendre des vacances en présence des clients dans le système. De plus, la distribution des vacances peut être dépendante ou indépendante du processus de service. Il est également possible que les durées des vacances soient mutuellement indépendantes, ou dépendantes des vacances précédentes. Ajoutons une suite de vacances $\{V_k\}$ à la description du modèle $M/G/1$, et supposons qu'elle soit indépendante du processus des arrivées et du processus de service. Il s'agit d'une suite composée de variables aléatoires positives, indépendantes et identiquement distribuées. La décomposition stochastique pour le nombre de clients dans le système $M/G/1$ avec vacances dans le cas d'un service exhaustif (les clients, qui arrivent dans le système, jouissent donc d'une priorité sans préemption sur les vacances) a été observée par Doshi (1986) (54). Une première étude approfondie sur les modèles avec vacances dans le cas d'un service non-exhaustif a été réalisée par Gaver (1962) (69). En introduisant la mesure de performance (période d'accomplissement du service), l'auteur a explicitement établi la validité de la propriété de décomposition stochastique. Il a également justifié la relation entre les modèles avec vacances et ceux avec priorité. Cependant, ce sont Fuhrmann et Cooper (1985) (68) qui ont défini une série d'hypothèses caractérisant les systèmes de files d'attente vérifiant la propriété de décomposition stochastique, particulièrement pour les systèmes d'attente avec vacances généralisées. Ces hypothèses sont :

- Les clients arrivent dans le système selon un processus de Poisson.
- La durée de service suit une loi générale. Les temps de service des différents clients sont indépendants les uns des autres et sont indépendants du processus des arrivées. Chaque durée de service est aussi indépendante de la suite de période des vacances qui la précède.
- Tous les clients arrivant dans le système sont éventuellement servis, et $\rho < 1$.
- Les clients sont servis dans un ordre qui est indépendants de leur durée de service.

- Les règles qui déterminent quand le serveur débute et termine les vacances n'ont aucune influence sur le processus des arrivées.
- Le nombre de clients arrivant dans le système durant les vacances est indépendant du nombre de clients présents dans le système au début des vacances.

Dans le cas où le système est en régime stationnaire on a la décomposition suivante pour la fonction génératrice de la distribution stationnaire du nombre de clients dans le système à un instant arbitraire d'accomplissement du service

$$\varphi(z) = \pi(z) \chi(z),$$

où $\pi(z)$ est la fonction génératrice pour le nombre de clients dans le système $M/G/1$ ordinaire sans vacances (formule de Pollczek-Khintchine, $\chi(z)$ est la fonction génératrice de la distribution stationnaire du nombre de clients dans le système étant donné que le serveur est en vacances. Ce résultat est valable pour tout système d'attente avec vacances. Il rend possible de se concentrer uniquement sur l'étude des effets des vacances sur le nombre de clients dans le système étant donné que le serveur est en vacances.

4.2.2 Modèles d'attente avec rappels

On peut constater que les vacances du serveur débutent après chaque service accompli, et le serveur se met à nouveau selon la compétition entre deux flux indépendants. L'un de ces flux est poissonnien, et correspond aux clients primaires ; l'autre (correspond aux clients en orbite) possède une structure complexe et son intensité dépend du nombre de clients en orbite. Par conséquent, le modèle sans vacances est le système d'attente classique sans rappels, et les vacances sont occasionnées par les tentatives répétées. Pour les modèles avec vacances, la propriété de décomposition stochastique a lieu aussi bien pour la distribution stationnaire de la taille du système que pour le temps d'attente (66, 70). Pour les modèles avec rappels, la validité de la PDS pour le temps d'attente est une conjoncture. Cependant, cette propriété pour le nombre de clients dans le système a été prouvée pour certains modèles avec rappels. On s'intéresse au processus stochastique

$$\{N(t) = C(t) + N_o(t), t \geq 0\},$$

où $N(t)$ est le nombre de clients dans le système, $C(t)$ est 0 ou 1 selon le fait que le serveur est libre ou occupé, $N_o(t)$ est le nombre de clients en orbite à la date t . L'étude de ce processus peut être ramenée à celle d'une chaîne de Markov induite aux instants ξ_n où le serveur devient libre pour la $n^{\text{ème}}$ fois $\{N(\xi_n)\}$. Supposons

que le système est en régime stationnaire. Ce dernier existe si et seulement si $\rho = \frac{\lambda}{\gamma} < 1$ (60). De plus, on a que la solution ergodique pour $\{N(\xi_n)\}$ est également la solution ergodique pour $N(t)$, (voir section 2 chapitre 2). La décomposition stochastique pour le nombre de clients dans le système $M/G/1$ avec rappels a été observée par Yang et Templeton (1987) (115). En supposant que le temps inter-rappels suit une loi exponentielle : les auteurs ont obtenu les résultats suivants sur la décomposition stochastique pour la fonction génératrice $\varphi(z)$ de la distribution stationnaire de la chaîne de Markov induite $\{N(\xi_n)\}$ lorsque n tend vers ∞ :

$$\varphi(z) = \frac{(1-\rho)(1-z)\tilde{B}(\lambda-\lambda z)\Phi(z)}{\tilde{B}(\lambda-\lambda z)-z\Phi(1)},$$

où

$$\Phi(z) = \exp\left\{-\frac{\lambda}{\theta}\int_0^z\frac{1-\tilde{B}(\lambda-\lambda x)}{\tilde{B}(\lambda-\lambda x)-x}dx\right\}.$$

Le facteur $\frac{(1-\rho)(1-z)\tilde{B}(\lambda-\lambda z)}{\tilde{B}(\lambda-\lambda z)-z}$ est la formule de Pollaczek-Khintchine pour le nombre de clients dans le système $M/G/1$ classique. Il est indépendant du temps inter-rappels. Le facteur $\frac{\Phi(z)}{\Phi(1)}$ est la fonction génératrice de la distribution stationnaire du nombre de clients dans le système $M/G/1$ avec rappels étant donné que le serveur est libre. Yang et al. (1994) (117) ont exploré la propriété de décomposition du système $M/G/1$ avec rappels et ont prouvé (à l'aide de la méthode de la chaîne de Markov induite) que cette propriété est toujours vraie pour la distribution générale du temps inter-rappels. La propriété de décomposition stochastique présente diverses applications pratiques dans le modèle $M/G/1$ avec rappels (28)

- obtention de relations explicites pour les moments factoriels de la taille de l'orbite et du nombre de clients en attente dans le système $M/G/1$ classique ;
- estimation de la vitesse de convergence vers le système $M/G/1$ classique :

$$\theta \text{ tend vers } \infty, D = \sum_{i=0}^1 \sum_{n=0}^{\infty} |p_{in}(\theta) - p_{in}(\infty)| = o\left(\frac{1}{\theta}\right);$$

- estimation de la distance entre les distributions stationnaires du modèle $M/G/1$ avec rappels et le modèle $M/G/1$ classique correspondant :

$$2(1-\rho)(1-R_o(\theta)) < D < (1-R_o(\theta)),$$

où

$$R_o(\theta) = \exp\left\{-\frac{\lambda}{\theta}\int_0^1\frac{1-\tilde{B}(\lambda-\lambda x)}{\tilde{B}(\lambda-\lambda x)-x}dx\right\}.$$

La validité de la PDS a été étendue aux modèles avec rappels et arrivées par groupes (115). On observe aussi cette propriété dans les systèmes avec rappels et priorité (61, 94). Dans (94), les auteurs considèrent un système $M_2/G_2/1$ avec rappels et deux types de clients : prioritaires et non-prioritaires. Dans le cas où le service est refusé, les clients prioritaires forment une file d'attente, tandis que les clients non-prioritaires rejoignent l'orbite. Les temps inter-rappels suivent une loi exponentielle. Le service est sans "préemption". L'état du système à la date t a été décrit par $\{C(t), N_o(t), A(t), \xi(t)\}$, où $C(t)$ est 0, 1 ou 2 selon le fait que le serveur est libre, un client prioritaire est en service ou un client non-prioritaire se trouve dans le serveur; $N_o(t)$ est le nombre de clients en orbite; $A(t)$ est le nombre de clients prioritaires en attente; $\xi(t) > 0$ est la durée de service écoulée à la date t . Lorsque le système est en régime stationnaire, la décomposition stochastique se présente de la manière suivante :

$$\{C_\theta(t), N_{o,\theta}(t), A_\theta(t)\} = \{C_\infty(t), N_\infty(t), A_\infty(t)\} + \{0, R_\theta(t), 0\}.$$

Les processus $\{C_\theta(t), N_{o,\theta}(t), A_\theta(t)\}$ et $\{0, R_\theta(t), 0\}$ sont associés au modèle $M_2/G_2/1$ avec rappels de taux de rappels $\theta > 0$, où $R_\theta(t)$ représente le nombre de clients en orbite à la date t étant donné que le serveur est libre et aucun client prioritaire n'est en attente. Le processus $\{C_\infty(t), N_\infty(t), A_\infty(t)\}$ est associé au modèle $M_2/G_2/1$ ordinaire, avec priorité et service sans "préemption". Soit un système de files d'attente où le phénomène de rappels ainsi que les vacances du serveur sont présents. A l'aide de l'approche régénératrice, Artalejo (1997) (19) a étudié un système d'attente de type $M/G/1$ avec rappels constants et vacances dans le cas d'un service exhaustif. L'état du système à la date t a été décrit par le processus $\{C(t), N_o(t), \xi(t)\}$, où $C(t)$ est 0, 1 ou 2 selon le fait que le serveur est libre, occupé ou en vacances; $\xi(t)$ est une variable aléatoire supplémentaire à valeurs dans R^+ , et désignant la durée de service écoulé (si $C(t) = 1$) ou la durée des vacances écoulée (si $C(t) = 2$) à la date t . Le serveur est donc libre lorsqu'il a débuté ses propres vacances, ou à cause des rappels. En supposant que le système est en régime stationnaire et à l'aide d'un système auxiliaire $M/G/1$ avec rappels et sans vacances, l'auteur a obtenu la décomposition stochastique pour le nombre de clients dans le système. Cette décomposition fournit trois composantes :

$$P(z) = \frac{(1-\rho)(1-z)\tilde{B}(\lambda-\lambda z)}{\tilde{B}(\lambda-\lambda z)-z} \frac{P_2(z)}{P_2} \frac{\bar{P}_0(z)}{\bar{P}_0},$$

où la première est liée au système $M/G/1$ ordinaire, la seconde aux rappels et la troisième aux vacances (propres). Enfin, dans Langaris et Moutzoukis (1995)

(92), la validité de la propriété de décomposition stochastique a été prouvée pour un système $M/G/1$ avec rappels, arrivées par groupes, priorité et vacances.

4.3 PROPRIÉTÉ DE DÉCOMPOSITION STOCHASTIQUE DU MODÈLE $M^X/G/1$ AVEC RAPPELS ET IMPATIENCE

Soit le processus stochastique $\{C(t), N_o(t), \xi(t), t \geq 0\}$. Nous nous intéressons au nombre de clients dans le système aux instants de départ.

Notations :

M_∞ : Modèle $M^X/G/1$ classique avec arrivées par groupes et groupes impatients ;

M_θ : Modèle $M^X/G/1$ avec rappels, arrivées par groupes et groupes impatients ;

N_θ : nombre de clients dans M_θ (régime stationnaire) ;

$N_{o,\theta}$: nombre de clients en orbite dans M_θ (régime stationnaire).

N_∞ : nombre de clients dans M_∞ (régime stationnaire) ;

$N_{f,\infty}$: nombre de clients dans la file d'attente dans M_∞ (régime stationnaire).

Considérons l'équation (3.18), nous pouvons voir que la fonction génératrice $\varphi(z)$ peut être décomposée en deux facteurs

$$\varphi(z) = \Omega(z) \cdot \Psi(z),$$

où $\Omega(z) = \frac{1-\rho}{\bar{c}\rho} \frac{\tilde{B}(\lambda H_1(1-C(z)))(1-C(z))}{\tilde{B}(\lambda H_1(1-C(z)))^{-z}}$ est la fonction génératrice du nombre de clients aux instants de départ associée au système de file d'attente classique $M^X/G/1$ avec clients impatients et arrivées par groupes (modèle M_∞). C'est un système d'attente avec un seul serveur où les clients arrivent par groupes de taille k (avec la probabilité $c_k, k \geq 1$) selon le processus de Poisson de paramètre $\lambda > 0$. Si à l'arrivée d'un groupe, le serveur est occupé, alors avec une probabilité $1 - H_1 > 0$, tous les clients du groupe quittent le système sans recevoir de service et avec la probabilité H_1 tout le groupe rejoint la file d'attente ; par contre, si le serveur est libre, alors un des clients du groupe commence son service et le reste du groupe rejoint la file d'attente ; à la fin de chaque service le client quitte le système.

Le deuxième facteur $\Psi(z) = \frac{P_0(z)}{P_0(1)} = \exp \left\{ \frac{\lambda}{\theta} \int_1^z \frac{1-\tilde{B}(\lambda H_1(1-C(z))) \frac{C(u)}{u}}{\tilde{B}(\lambda H_1(1-C(z)))^{-u}} du \right\}$ représente la fonction génératrice du nombre de clients aux instants de départ associée au système d'attente $M^X/G/1$ avec rappels et impatience (modèle M_θ) sachant que le serveur est libre. Alors, la décomposition stochastique peut être exprimée comme suit :

$$\{0, N_{o,\theta}(t), t \geq 0\} = \{0, N_{f,\infty}(t), t \geq 0\} + \{0, N_\theta(t), t \geq 0\}. \quad (4.1)$$

Les processus $\{0, N_{O,\theta}(t), t \geq 0\}$ et $\{0, N_\theta(t), \xi_n(t) \geq 0\}$ sont associés au modèle M_θ , où $N_\theta(\xi_n(t))$ représente le nombre de clients en orbite à l'instant $\xi_n(t)$ sachant que le serveur est libre. Le processus $\{0, N_{f,\infty}(t), t \geq 0\}$ est associé au modèle M_∞ , où $N_{f,\infty}(t)$ est le nombre de clients dans la file d'attente à l'instant t .

CONCLUSION DU CHAPITRE

Par conséquent, nous avons obtenu les résultats suivants : Le nombre de clients aux instants de départ dans le système M_θ considéré est la somme de deux variables aléatoires indépendantes ; le nombre de clients aux instants de départ dans le système M_∞ et le nombre de clients dans le système correspondant mais avec rappels sachant que le serveur est libre.

$$N_\theta = N_\infty + (N_\theta / \text{le serveur est libre}). \quad (4.2)$$

De même, le nombre de clients en orbite aux instants de départ dans M_θ est la somme de deux variables aléatoires indépendantes ; la première est le nombre de clients en orbite aux instants de départ dans le système M_∞ et la seconde est le nombre de clients en orbite dans le système correspondant mais avec rappels sachant que le serveur est libre.

$$N_o = N_{f,\infty} + (N_o / \text{le serveur est libre}). \quad (4.3)$$

Les résultats obtenus sont importants pour comprendre la contribution réelle des rappels pour le nombre de clients dans le système (aux instants de départ).

COMPORTEMENT ASYMPTOTIQUE DU NOMBRE DE CLIENTS DANS LE MODÈLE $M^X/G/1$ AVEC RAPPELS ET CLIENTS IMPATIENTS

5.1 INTRODUCTION

Dans ce chapitre, nous étudions le comportement asymptotique du modèle $M^X/G/1$ avec rappels et clients impatients.

Une première étude du comportement asymptotique d'un modèle avec rappels a été réalisée par Falin (1979, 1986) (59), (57) et ceci pour le système avec rappels de type $M/G/1$. Il a été établi que :

- si λ tend vers $\frac{1}{\beta_1} - 0$ et $\beta_2 < \infty$, la distribution de la variable aléatoire du nombre de clients en orbite $(1 - \rho)N_o(t)$ converge vers la distribution Gamma de paramètres $\alpha^* = \frac{1}{\theta\beta_1} + \frac{\beta_2}{2\beta_1^2}$ et $\mu^* = \frac{\beta_2}{2\beta_1^2} \left(\frac{1}{\theta\beta_1} + \frac{\beta_2}{2\beta_1^2} \right)$.*
- si θ tend vers 0, $N_o(t)$ le nombre de clients en orbite est asymptotiquement Gaussienne de moyenne $\frac{\lambda\rho}{(1 - \rho)\theta}$ et de variance $\frac{\lambda^3\beta_2 + 2\lambda\rho - 2\lambda\rho^2}{2(1 - \rho)^2\theta}$.*

Une autre étude plus récente a été faite par J. Kim et B. Kim (2007) (79), où les auteurs ont démontré que la distribution de la taille de l'orbite converge asymptotiquement vers une fonction géométrique multipliée par une fonction puissance (un polynôme). Dans le chapitre 3, nous avons obtenu la fonction génératrice de la distribution stationnaire du nombre de clients en orbite. L'expression établie (3.25) ne permet pas d'étudier le caractère de la distribution en question dans certaines applications importantes : régime chargé, intensité faible des rappels. En outre, nous voulons avoir des informations sur le comportement du système étudié lorsque le taux des rappels θ tend vers ∞ . Nous obtenons des résultats

qui peuvent être considérés plus généraux que ceux établis dans la littérature par exemple par (57, 59, 65).

5.2 COMPORTEMENT ASYMPTOTIQUE DU NOMBRE DE CLIENTS EN ORBITE SOUS UN TAUX DE TRAFIC INTENSE

Nous nous intéressons dans cette partie au cas du trafic intense, c'est-à-dire quand le taux d'arrivées λ croît de telle manière que ρ tend vers 1, nous citons le travail de Falin et Falin (63). Nous avons le théorème suivant :

Théorème 5.1 Si $\rho = \lambda \bar{c} H_1 \beta_1 < 1$ et $\beta_2 < \infty$, alors

$$\lim_{\lambda \rightarrow \frac{1}{\bar{c} H_1 \beta_1} - 0} P(z) = \left(1 + \left(\frac{\beta_2}{2\beta_1^2} + \frac{C''(1)}{2\bar{c}} \right) s \right)^{-1 - \frac{\frac{1}{\beta_1 H_1}}{\theta \left(\frac{\beta_2}{2\beta_1^2} + \frac{C''(1)}{2\bar{c}} \right)}} \quad (5.1)$$

la variable aléatoire $N_o = \lim_{t \rightarrow \infty} N_o(t)$, représentant le nombre de clients en orbite, est asymptotiquement distribuée selon une loi Gamma.

Démonstration. Considérons la fonction génératrice

$$\begin{aligned} P(z) &= P_0(z) + P_1(z) = \frac{(1 - H_1 z) - \tilde{B}(\lambda H_1 (1 - C(z))) (1 - H_1)}{(\tilde{B}(\lambda H_1 (1 - C(z))) - z)} \\ &\quad \times \frac{(1 - \rho)}{\rho + H_1 (1 - \rho)} \exp \left[\frac{\lambda}{\theta} \int_1^z \frac{1 - \tilde{B}(\lambda H_1 (1 - C(u))) \frac{C(u)}{u}}{\tilde{B}(\lambda H_1 (1 - C(u))) - u} du \right]. \end{aligned} \quad (5.2)$$

Supposons que $\varepsilon = 1 - \rho$ tend vers 0 quand λ tend vers $\frac{1}{\bar{c} H_1 \beta_1} - 0$ et posons $z = e^{-\varepsilon s}$. Sous cette hypothèse le premier terme de (5.2) devient

$$\lim_{\varepsilon \rightarrow 0} \frac{(1 - H_1 e^{-\varepsilon s}) - \tilde{B}(\lambda H_1 (1 - C(e^{-\varepsilon s}))) (1 - H_1)}{(\tilde{B}(\lambda H_1 (1 - C(e^{-\varepsilon s}))) - e^{-\varepsilon s})} \times \frac{\varepsilon}{1 - \varepsilon + H_1 \varepsilon}.$$

Soit $A(e^{-\varepsilon s}) = \tilde{B}(t)$ où $t = \lambda H_1 (1 - C(e^{-\varepsilon s}))$. Nous pouvons développer $\tilde{B}(t)$ de la manière suivante :

$$x = \lambda H_1 (1 - C(e^{-\varepsilon s})) = \lambda H_1 \left[1 - C \left(1 - s\varepsilon + \frac{s^2}{2} \varepsilon^2 + \varepsilon^2 o(\varepsilon) \right) \right].$$

Nous avons

$$\begin{aligned} C(u) &= C(1 + (u - 1)) \\ &= C(1) + (u - 1) C'(1) + \frac{1}{2} (u - 1)^2 C''(1) + (u - 1)^2 o(u - 1). \end{aligned}$$

D'où

$$\begin{aligned} C(e^{-\varepsilon s}) &= C(1) + \left(-\varepsilon s + \frac{s^2}{2}\varepsilon^2\right) C'(1) + \frac{1}{2}(-\varepsilon s)^2 C''(1) + \varepsilon^2 o(\varepsilon) \\ &= 1 - s\bar{c}\varepsilon + \frac{s^2}{2}(C'(1) + C''(1))\varepsilon^2 + \varepsilon^2 o(\varepsilon), \end{aligned}$$

et x devient

$$x = \lambda H_1 \left[\bar{c}s\varepsilon - \frac{s^2}{2}(C'(1) + C''(1))\varepsilon^2 + \varepsilon^2 o(\varepsilon) \right].$$

Comme $\varepsilon = 1 - \lambda\bar{c}\beta_1 H_1$, $\lambda = \frac{1-\varepsilon}{\beta_1\bar{c}H_1}$, alors

$$\begin{aligned} x &= \frac{1-\varepsilon}{\beta_1\bar{c}} \left[\bar{c}s\varepsilon - \frac{s^2}{2}(C'(1) + C''(1))\varepsilon^2 + \varepsilon^2 o(\varepsilon) \right] \\ &= \frac{s}{\beta_1}\varepsilon - \eta_1(s)\varepsilon^2 + \varepsilon^2 o(\varepsilon), \end{aligned}$$

où $\eta_1(s) = \left[\frac{(C'(1)+C''(1))s^2}{2\bar{c}\beta_1} - \frac{1}{\beta_1}s \right]$. Nous avons aussi

$$\begin{aligned} \tilde{B}(x) &= \tilde{B}(0) + x\tilde{B}'(0) + \frac{x^2}{2}\tilde{B}''(0) \\ &= \beta_0 - \beta_1 x + \frac{x^2}{2}\beta_2 + x^2 o(x) \\ &= 1 - \beta_1 \left[\frac{1}{\beta_1}\varepsilon s - \eta_1(s)\varepsilon^2 \right] + \frac{1}{2\beta_1^2}s^2\beta_2\varepsilon^2 + \varepsilon^2 o(\varepsilon). \end{aligned}$$

Donc

$$\tilde{B}(x) = 1 - s\varepsilon + \eta_2(s)\varepsilon^2 + \varepsilon^2 o(\varepsilon),$$

où $\eta_2(s) = \left(\frac{\beta_2}{2\beta_1^2} + \frac{(C'(1)+C''(1))}{2\bar{c}} \right) s^2 + s$. En notant que $\beta_2 < \infty$, nous obtenons

$$\begin{aligned} &\lim_{\varepsilon \rightarrow 0} \frac{(1 - H_1 e^{-\varepsilon s}) - \tilde{B}(\lambda H_1 (1 - C(e^{-\varepsilon s}))) (1 - H_1)}{\left(\tilde{B}(\lambda H_1 (1 - C(e^{-\varepsilon s}))) - e^{-\varepsilon s} \right)} \times \frac{\varepsilon}{1 - \varepsilon + H_1 \varepsilon} \\ &= \lim_{\varepsilon \rightarrow 0} \frac{(1 - H_1 e^{-\varepsilon s}) - \tilde{B}(t) (1 - H_1)}{\left(\tilde{B}(t) - e^{-\varepsilon s} \right)} \times \frac{\varepsilon}{1 - \varepsilon + H_1 \varepsilon} \\ &= \lim_{\varepsilon \rightarrow 0} \frac{\left[1 - H_1 \left(1 - s\varepsilon + \frac{s^2}{2}\varepsilon^2 \right) \right] - (1 - s\varepsilon + \eta_2(s)\varepsilon^2) (1 - H_1) + \varepsilon^2 o(\varepsilon)}{(1 - s\varepsilon + \eta_2(s)\varepsilon^2) - \left(1 - s\varepsilon + \frac{s^2}{2}\varepsilon^2 \right) + \varepsilon^2 o(\varepsilon)} \times \frac{\varepsilon}{1 - \varepsilon + H_1 \varepsilon} \\ &= \frac{s}{\eta_2(s) - \frac{s^2}{2}}. \end{aligned}$$

Maintenant, nous calculons le second terme du membre droit de (5.2)

$$\lim_{\lambda \rightarrow \frac{1}{\bar{c}\beta_1 H_1} - 0} \exp \left[\frac{\lambda}{\theta} \int_1^z \frac{1 - \tilde{B}(\lambda H_1 (1 - C(u))) \frac{C(u)}{u}}{\tilde{B}(\lambda H_1 (1 - C(u))) - u} du \right].$$

Introduisons la nouvelle variable $u = e^{-\varepsilon\sigma}$. Alors $du = -\varepsilon e^{-\varepsilon\sigma} d\sigma$. Nous avons donc

$$C(e^{-\varepsilon\sigma}) = 1 - \bar{c}\sigma\varepsilon + \frac{(C'(1) + C''(1))}{2} \sigma^2 \varepsilon^2 + \varepsilon^2 o(\varepsilon)$$

et

$$\tilde{B}(t) = 1 - \sigma\varepsilon + \eta_2(\sigma) \varepsilon^2 + \varepsilon^2 o(\varepsilon).$$

Par conséquent,

$$\begin{aligned} & \frac{1 - \tilde{B}(\lambda H_1 (1 - C(e^{-\varepsilon\sigma}))) \frac{C(e^{-\varepsilon\sigma})}{e^{-\varepsilon\sigma}}}{\tilde{B}(\lambda H_1 (1 - C(e^{-\varepsilon\sigma}))) - e^{-\varepsilon\sigma}} e^{-\varepsilon\sigma} \\ = & \frac{\frac{\varepsilon^2 \sigma^2}{2} \left[1 - \frac{\beta_2}{2\beta_1^2} - \frac{(\bar{c} + C''(1))}{\bar{c}} - 2\bar{c} - C''(1) \right] - \varepsilon^2 \sigma + \varepsilon \sigma \bar{c} + \varepsilon^2 o(\varepsilon)}{\varepsilon^2 \left[\left(\frac{\beta_2}{2\beta_1^2} + \frac{(\bar{c} + C''(1))}{2\bar{c}} \right) \sigma^2 + \sigma \right] - \frac{\varepsilon^2 \sigma^2}{2} + \varepsilon^2 o(\varepsilon)} \\ = & \frac{\bar{c} - \varepsilon + \varepsilon \left(-\frac{\beta_2}{2\beta_1^2} - \frac{(\bar{c} + 1)C''(1)}{\bar{c}} - \bar{c} \right) \sigma}{\varepsilon + \varepsilon \left(\frac{\beta_2}{2\beta_1^2} + \frac{C''(1)}{2\bar{c}} \right) \sigma}. \end{aligned}$$

Alors

$$\begin{aligned} I_\varepsilon &= \varepsilon \int_0^s \frac{1 - \tilde{B}(\lambda H_1 (1 - C(e^{-\varepsilon\sigma}))) \frac{C(e^{-\varepsilon\sigma})}{e^{-\varepsilon\sigma}}}{\tilde{B}(\lambda H_1 (1 - C(e^{-\varepsilon\sigma}))) - e^{-\varepsilon\sigma}} e^{-\varepsilon\sigma} d\sigma \\ &= \int_0^s \frac{\bar{c} - \varepsilon + \varepsilon \left(-\frac{\beta_2}{2\beta_1^2} - \frac{(\bar{c} + 1)C''(1)}{\bar{c}} - \bar{c} \right) \sigma}{1 + \sigma \left(\frac{\beta_2}{2\beta_1^2} + \frac{C''(1)}{2\bar{c}} \right)} d\sigma. \end{aligned}$$

Soit $A = \varepsilon \left(-\frac{\beta_2}{2\beta_1^2} - \frac{(\bar{c} + 1)C''(1)}{\bar{c}} - \bar{c} \right)$; $B = \varepsilon - \bar{c}$ et $D = \left(\frac{\beta_2}{2\beta_1^2} + \frac{C''(1)}{2\bar{c}} \right)$.

Alors

$$\begin{aligned} I_\varepsilon &= \int_0^s \frac{A\sigma - B}{D\sigma + 1} d\sigma \\ &= \frac{A}{D}s - \frac{A + BD}{D^2} \left(\ln \left| s + \frac{1}{D} \right| - \ln \left| \frac{1}{D} \right| \right) \\ &= \frac{A}{D}s - \frac{A + BD}{D^2} \ln |Ds + 1|. \end{aligned}$$

Nous avons

$$\begin{aligned} \frac{A}{D} &= \frac{\epsilon \left(-\frac{\beta_2}{2\beta_1^2} - \frac{(\bar{c}+1)}{\bar{c}} C''(1) - \bar{c} \right)}{\frac{\beta_2}{2\beta_1^2} + \frac{C''(1)}{2\bar{c}}} \xrightarrow{\epsilon \rightarrow 0} 0 \\ \frac{A + BC}{D^2} &= \frac{\epsilon \left(-\frac{\beta_2}{2\beta_1^2} - \frac{(\bar{c}+1)}{\bar{c}} C''(1) - \bar{c} \right) + (\epsilon - \bar{c}) \left(\frac{\beta_2}{2\beta_1^2} + \frac{C''(1)}{2\bar{c}} \right)}{\left(\frac{\beta_2}{2\beta_1^2} + \frac{C''(1)}{2\bar{c}} \right)^2} \\ &\xrightarrow{\epsilon \rightarrow 0} \frac{-\bar{c}}{\left(\frac{\beta_2}{2\beta_1^2} + \frac{C''(1)}{2\bar{c}} \right)}. \end{aligned}$$

Alors

$$I_\epsilon \xrightarrow{\epsilon \rightarrow 0} \frac{-\bar{c}}{\left(\frac{\beta_2}{2\beta_1^2} + \frac{C''(1)}{2\bar{c}} \right)} \ln \left| \left(\frac{\beta_2}{2\beta_1^2} + \frac{C''(1)}{2\bar{c}} \right) s + 1 \right|,$$

d'où

$$\begin{aligned} &\exp \left[\frac{\lambda}{\theta} \int_1^z \frac{1 - \tilde{B}(\lambda H_1(1 - C(u))) \frac{C(u)}{u}}{\tilde{B}(\lambda H_1(1 - C(u))) - u} du \right] \\ &= \exp \left[\frac{-\lambda \bar{c}}{\theta \left(\frac{\beta_2}{2\beta_1^2} + \frac{C''(1)}{2\bar{c}} \right)} \ln \left| \left(\frac{\beta_2}{2\beta_1^2} + \frac{C''(1)}{2\bar{c}} \right) s + 1 \right| \right] \\ &= \left(\left(\frac{\beta_2}{2\beta_1^2} + \frac{C''(1)}{2\bar{c}} \right) s + 1 \right)^{\frac{-\lambda \bar{c}}{\theta \left(\frac{\beta_2}{2\beta_1^2} + \frac{C''(1)}{2\bar{c}} \right)}}. \end{aligned}$$

Finalement,

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} P(z) &= \frac{1}{1 + s \left(\frac{\beta_2}{2\beta_1^2} + \frac{C''(1)}{2\bar{c}} \right)} \left[\left(\frac{\beta_2}{2\beta_1^2} + \frac{C''(1)}{2\bar{c}} \right) s + 1 \right]^{\frac{-\lambda \bar{c}}{\theta \left(\frac{\beta_2}{2\beta_1^2} + \frac{C''(1)}{2\bar{c}} \right)}} \\ &= \left(1 + \left(\frac{\beta_2}{2\beta_1^2} + \frac{C''(1)}{2\bar{c}} \right) s \right)^{-1 - \frac{1}{\theta \left(\frac{\beta_2}{2\beta_1^2} + \frac{C''(1)}{2\bar{c}} \right)}}. \end{aligned}$$

□

5.3 COMPORTEMENT ASYMPTOTIQUE DU NOMBRE DE CLIENTS EN ORBITE SOUS UN TAUX DE RAPPELS FAIBLE

Quand le taux de rappel est faible, θ tend vers 0, le nombre de clients en orbite se comporte asymptotiquement selon le théorème que nous présentons ci dessous.

Théorème 5.2 Si $\beta_2 < \infty$, alors lorsque θ tend vers 0 le nombre de clients en orbite $N_o(t)$ est

asymptotiquement Gaussien de moyenne $\frac{\lambda}{\theta} \frac{\bar{c} + \rho - 1}{1 - \rho}$ et de variance $\frac{\lambda}{\theta} \frac{[(1 - \rho)d_0 + \bar{c}\rho d_1]}{2(1 - \rho)^2}$,
 où $d_0 = \frac{\rho^2}{\beta_1} + C''(1)$ et $d_1 = 1 - \rho + \rho \frac{\beta_2}{\beta_1} + \frac{C''(1)}{\bar{c}}$.

Démonstration. Soit la variable aléatoire n^* définie par

$$n^* = \frac{n_0 - \frac{\lambda}{\theta} \left(\frac{\bar{c} + \rho - 1}{1 - \rho} \right)}{\frac{1}{\sqrt{\theta}}} = \sqrt{\theta} n_0 - \frac{\lambda}{\sqrt{\theta}} \left(\frac{\bar{c} + \rho - 1}{1 - \rho} \right).$$

La fonction caractéristique $\mathbb{E}[\exp(itn^*)]$ peut être exprimée en fonction de la fonction génératrice $P(z)$ de la manière suivante

$$\begin{aligned} \mathbb{E}[\exp(itn^*)] &= P\left(e^{it\sqrt{\theta}}\right) e^{-it \frac{\lambda}{\sqrt{\theta}} \left(\frac{\bar{c} + \rho - 1}{1 - \rho} \right)} \\ &= \frac{\left(1 - H_1 e^{it\sqrt{\theta}}\right) - G(t)(1 - H_1)}{\left(G(t) - e^{it\sqrt{\theta}}\right)} \frac{(1 - \rho)}{\rho + H_1(1 - \rho)} \\ &\quad \exp\left\{ \frac{\lambda}{\theta} \int_1^{e^{it\sqrt{\theta}}} \frac{1 - A(u) \frac{C(u)}{u}}{A(u) - u} du - it \frac{\lambda}{\sqrt{\theta}} \left(\frac{\bar{c} + \rho - 1}{1 - \rho} \right) \right\} \end{aligned} \quad (5.3)$$

où $G(t) = A\left(e^{it\sqrt{\theta}}\right)$ et $A(z) = \tilde{B}(\lambda H_1(1 - C(z)))$ Considérons le premier terme (5.3). Quand θ tend vers 0 et $z = e^{it\sqrt{\theta}}$ tend vers 1, alors

$$\frac{\left(1 - H_1 e^{it\sqrt{\theta}}\right) - G(t)(1 - H_1)}{H_1 \left(G(t) - e^{it\sqrt{\theta}}\right)} \frac{H_1(1 - \rho)}{\rho + H_1(1 - \rho)} \rightarrow 1.$$

Nous transformons le second terme de (5.3) comme suit

$$\begin{aligned} &\exp\left\{ \frac{\lambda}{\theta} \int_1^{e^{it\sqrt{\theta}}} \frac{1 - A(u) \frac{C(u)}{u}}{A(u) - u} du - it \frac{\lambda}{\sqrt{\theta}} \left(\frac{\bar{c} + \rho - 1}{1 - \rho} \right) \right\} \\ &= \exp\left\{ \frac{\lambda}{\theta} \int_1^{e^{it\sqrt{\theta}}} \frac{1 - A(u) \frac{C(u)}{u}}{A(u) - u} du - \frac{\lambda}{\theta} it \sqrt{\theta} \left(\frac{\bar{c} + \rho - 1}{1 - \rho} \right) \right\} \\ &= \exp\left\{ \frac{\lambda}{\theta} \int_1^{e^{it\sqrt{\theta}}} \left(\frac{1 - A(u) \frac{C(u)}{u}}{A(u) - u} - \frac{\bar{c} + \rho - 1}{1 - \rho} \right) du \right. \\ &\quad \left. + \frac{\lambda \left(e^{it\sqrt{\theta}} - 1 \right) - it \lambda \sqrt{\theta} \left(\frac{\bar{c} + \rho - 1}{1 - \rho} \right)}{\theta} \right\}. \end{aligned} \quad (5.4)$$

Pour calculer la limite du premier terme du membre de droite de (5.4), nous introduisons la fonction

$$f(\theta) = \exp\left\{ \frac{\lambda}{\theta} \int_1^{e^{it\sqrt{\theta}}} \left(\frac{1 - A(u) \frac{C(u)}{u}}{A(u) - u} - \frac{\bar{c} + \rho - 1}{1 - \rho} \right) du \right\}.$$

Nous pouvons réécrire (5.4) de la manière suivante

$$\frac{\lambda}{\theta} \int_1^{e^{it\sqrt{\theta}}} \frac{1 - A(u) \frac{C(u)}{u}}{A(u) - u} du - it \frac{\lambda}{\sqrt{\theta}} \left(\frac{\bar{c} + \rho - 1}{1 - \rho} \right) = \frac{\lambda}{\theta} f_1(\theta) + f_2(\theta),$$

où

$$f_1(\theta) = \int_1^{e^{it\sqrt{\theta}}} \left(\frac{1 - A(u) \frac{C(u)}{u}}{A(u) - u} - \frac{\bar{c} + \rho - 1}{1 - \rho} \right) du,$$

et

$$f_2(\theta) = \frac{\lambda \left(e^{it\sqrt{\theta}} - 1 \right) - it\lambda\sqrt{\theta} \left(\frac{\bar{c} + \rho - 1}{1 - \rho} \right)}{\theta}.$$

Remarquons que $f_1(0) = 0$, et

$$f_1'(\theta) = -\frac{t^2}{2} \left[\frac{\left(e^{it\sqrt{\theta}} - e^{2it\sqrt{\theta}} + e^{it\sqrt{\theta}} G(t) - G(t) C(e^{it\sqrt{\theta}}) \right)}{it\sqrt{\theta} \left(G(t) - e^{it\sqrt{\theta}} \right)} - \frac{\bar{c}}{(1-\rho)} \frac{e^{it\sqrt{\theta}} \left(G(t) - e^{it\sqrt{\theta}} \right)}{it\sqrt{\theta} \left(G(t) - e^{it\sqrt{\theta}} \right)} \right].$$

Alors

$$f_1'(0) = \lim_{\theta \rightarrow 0} f_1'(\theta) = -\frac{t^2}{4(\rho-1)} \left[-2 + 2\rho - \frac{\rho^2}{\beta_1} - \bar{c} - \rho\bar{c} - C''(1) - \frac{\bar{c}}{(1-\rho)} \left(3(\rho-1) + \left(\rho^2 \frac{\beta_2}{\beta_1^2} + \rho \frac{C''(1)}{\bar{c}} \right) \right) \right],$$

comme θ tend vers 0 et $\beta_2 < \infty$, nous avons

$$\begin{aligned} f_1(\theta) &= f_1(0) + \theta f_1'(0) + o(\theta) \\ &= -\theta \frac{t^2}{4(\rho-1)} \left[-2 + 2\rho - \frac{\rho^2}{\beta_1} - \bar{c} - \rho\bar{c} - C''(1) - \frac{\bar{c}}{(1-\rho)} \left(3(\rho-1) + \left(\rho^2 \frac{\beta_2}{\beta_1^2} + \rho \frac{C''(1)}{\bar{c}} \right) \right) \right] + o(\theta). \end{aligned}$$

Par conséquent,

$$\begin{aligned} \frac{\lambda}{\theta} f_1(\theta) &= -\frac{\lambda t^2}{4(\rho-1)} \left[-2 + 2\rho - \frac{\rho^2}{\beta_1} - \bar{c} - \rho\bar{c} - C''(1) - \frac{\bar{c}}{(1-\rho)} \left(3(\rho-1) + \left(\rho^2 \frac{\beta_2}{\beta_1^2} + \rho \frac{C''(1)}{\bar{c}} \right) \right) \right] + o(1). \end{aligned}$$

D'où

$$\lim_{\theta \rightarrow 0} f_2(\theta) = -\frac{\lambda t^2}{2} \left(\frac{\bar{c} + \rho - 1}{1 - \rho} \right).$$

Donc

$$\lim_{\theta \rightarrow 0} \mathbb{E} [\exp \{itn^*\}] = \exp \left\{ -\frac{t^2}{2} \cdot \lambda \frac{(1-\rho)d_0 + \bar{c}\rho d_1}{2(1-\rho)^2} \right\},$$

où $d_0 = \frac{\rho^2}{\beta_1} + C''(1)$ et $d_1 = 1 - \rho + \rho \frac{\beta_2}{\beta_1^2} + \frac{C''(1)}{\bar{c}}$. C'est la fonction caractéristique d'une variable aléatoire Gaussienne de moyenne $\frac{\lambda}{\theta} \cdot \frac{\bar{c} + \rho - 1}{1 - \rho}$ et de variance $\frac{\lambda}{\theta} \frac{[(1-\rho)d_0 + \bar{c}\rho d_1]}{2(1-\rho)^2}$. \square

Nous récapitulons nos deux résultats précédents, dans ce théorème :

Théorème 5.3 *Supposons que nous avons deux systèmes de files d'attente : $M/G/1$ avec rappels, arrivées par groupes et clients impatientes (modèle M_θ^X) ayant $P(z) = \sum_{n=0}^{\infty} z^n p_n$ pour fonction génératrice de la distribution stationnaire p_n du nombre de clients en orbite $N_o = \lim_{t \rightarrow \infty} N_o(t)$. Et $M/G/1$ avec rappels, arrivées simples et clients persistants (modèle M_θ) ayant $P^*(z) = \sum_{n=0}^{\infty} z^n p_n^*$ pour fonction génératrice de la distribution stationnaire p_n^* du nombre de clients en orbite $N_o^* = \lim_{t \rightarrow \infty} N_o^*(t)$. Alors, $N_o(t)$ et $N_o^*(t)$ ont asymptotiquement la même distribution et possèdent les mêmes propriétés.*

Démonstration. 1. Considérons le cas du trafic dense tel que : ρ tend vers 1.

Pour $M/G/1$ avec rappels, arrivées par groupes et impatience, nous avons montré que la distribution du nombre de clients en orbite $N^*(t)$, converge asymptotiquement vers la loi Gamma avec

$$\lim_{\lambda \rightarrow \frac{1}{\beta_1 \bar{c} H_1} - 0} P(z) = \left(1 + \left(\frac{\beta_2}{2\beta_1^2} + \frac{C''(1)}{2\bar{c}} \right) s \right)^{-1 - \frac{\frac{1}{\beta_1 H_1}}{\theta \left(\frac{\beta_2}{2\beta_1^2} + \frac{C''(1)}{2\bar{c}} \right)}}. \quad (5.5)$$

2. Considérons maintenant, le cas où le taux de rappels est faible tel que θ tend vers 0.

Nous avons montré que la distribution du nombre de clients en orbite N_o pour le modèle M_θ est asymptotiquement Gaussien avec $E[N_o] = \frac{\lambda}{\theta} \cdot \frac{\bar{c} + \rho - 1}{1 - \rho}$ et $Var[N_o] = \frac{\lambda}{\theta} \frac{[(1-\rho)d_0 + \bar{c}\rho d_1]}{2(1-\rho)^2}$, où $d_0 = \frac{\rho^2}{\beta_1} + C''(1)$ et $d_1 = 1 - \rho + \rho \frac{\beta_2}{\beta_1^2} + \frac{C''(1)}{\bar{c}}$. Dans (5.6), pour le modèle M_θ , il est montré que la distribution de la variable aléatoire N_o^* est aussi asymptotiquement Gaussienne avec $E[N_o^*] = \frac{\lambda}{\theta} \cdot \frac{\rho}{1 - \rho}$ et $Var[N_o^*] = \frac{\lambda}{\theta} \frac{\lambda^2 \beta_2 + 2\rho - 2\rho^2}{2(1-\rho)^2}$ (ici $\rho = \lambda\beta_1$). Posons $\bar{c} = 1$ et $H_1 = 1$. Alors l'expression (5.5) devient

$$\lim_{\lambda \rightarrow \frac{1}{\beta_1}} P(z) = \left(1 + \frac{\beta_2}{2\beta_1^2} s \right)^{-1 - \frac{2\beta_1}{\theta\beta_2}}.$$

Cette limite est une loi Gamma vers laquelle converge la distribution de la variable aléatoire N_o^* pour le modèle M_θ (56).

Remplaçons : $\bar{c} = 1$ et $H_1 = 1$ dans $E[N_o]$ et $Var[N_o]$, nous obtenons : $E[N_o] = E[N_o^*]$. \square

5.4 COMPORTEMENT ASYMPTOTIQUE DU SYSTÈME SOUS UN TAUX DE RAPPELS ÉLEVÉ

Maintenant, nous étudions le comportement asymptotique de notre système quand le taux des rappels est élevé. En général, quand θ tend vers ∞ la distribution stationnaire d'un système avec rappels converge vers une limite, qui est généralement la distribution stationnaire d'un certain système limite. Dans notre cas, il est intuitif que c'est le modèle M_∞ (voir chapitre 4). Pour prouver cet argument heuristique, nous utilisons la propriété de décomposition stochastique. Soit

$$\begin{aligned}\pi_n(\theta) &= \lim_{k \rightarrow \infty} P(N_{o,\theta}(\xi_k) = n), \\ \pi_n(\infty) &= \lim_{k \rightarrow \infty} P(N_{f,\infty}(\xi_k) = n), \\ \hat{\pi}_n(\theta) &= \lim_{k \rightarrow \infty} P(N_\theta(\xi_k) = n / C(t) = 0).\end{aligned}$$

Notons que $\hat{\pi}_n(\theta) = \frac{P_{0,n}}{\binom{H_1(1-\rho)}{\rho+H_1(1-\rho)}}$ et sa fonction génératrice est $\Psi(z)$. Avec l'aide de la méthode présentée dans (65), nous pouvons annoncer le résultat suivant pour le taux de convergence de la distribution $\pi_n(\theta)$ vers la distribution $\pi_n(\infty)$.

Théorème 5.4 Les inégalités suivantes sont vérifiées

$$2 \frac{1-\rho}{\bar{c}\rho\tilde{B}(\lambda H_1)} (1 - \hat{\pi}_0(\theta)) \leq \sum_{n=0}^{\infty} |\pi_n(\theta) - \pi_n(\infty)| \leq 2(1 - \hat{\pi}_0(\theta)),$$

où

$$\hat{\pi}_0(\theta) = \exp \left\{ -\frac{\lambda}{\theta} \int_0^1 \frac{1 - A(u) \frac{C(u)}{u}}{A(u) - u} du \right\}.$$

Nous avons aussi

$$\sum_{n=0}^{\infty} |\pi_n(\theta) - \pi_n(\infty)| = o\left(\frac{1}{\theta}\right).$$

Démonstration. De (4.1), nous remarquons que $\pi_n(\theta)$ est une convolution de deux distributions $\pi_n(\infty)$ et $\hat{\pi}_n(\theta)$, et que

$$\pi_n(\theta) = \sum_{k=0}^n \pi_k(\infty) \hat{\pi}_{n-k}(\theta). \quad (5.6)$$

L'expression (5.6) peut être réécrite sous la forme

$$\pi_n(\theta) - \pi_n(\infty) = \pi_n(\infty)\hat{\pi}_0(\theta) - \pi_n(\infty) + \sum_{k=0}^n \pi_k(\infty)\hat{\pi}_{n-k}(\theta).$$

Comme

$$\begin{aligned} |\pi_n(\theta) - \pi_n(\infty)| &\leq |\pi_n(\infty)\hat{\pi}_0(\theta) - \pi_n(\infty)| + \sum_{k=0}^{n-1} \pi_k(\infty)\hat{\pi}_{n-k}(\theta) \\ &\leq \pi_n(\infty)(1 - 2\hat{\pi}_0(\theta)) + \pi_n(\theta), \end{aligned}$$

$$\sum_{n=0}^{\infty} |\pi_n(\theta) - \pi_n(\infty)| \leq 2(1 - \hat{\pi}_0(\theta)),$$

où

$$\hat{\pi}_0(\theta) = \Psi(0) = \exp \left\{ -\frac{\lambda}{\theta} \int_0^1 \frac{1 - A(u) \frac{C(u)}{u}}{A(u) - u} du \right\}.$$

D'où la deuxième inégalité. En utilisant l'inégalité $|x - y| \geq x - y$, nous obtenons

$$\sum_{n=0}^{\infty} |\pi_n(\theta) - \pi_n(\infty)| \geq |\pi_0(\theta) - \pi_0(\infty)| + \pi_0(\infty) - \pi_0(\theta). \quad (5.7)$$

De (5.6), nous pouvons voir que $\pi_0(\theta) = \pi_0(\infty)\hat{\pi}_0(\theta) < \pi_0(\infty)$. Alors l'expression (5.7) devient

$$\sum_{n=0}^{\infty} |\pi_n(\theta) - \pi_n(\infty)| \geq 2\pi_0(\infty)(1 - \hat{\pi}_0(\theta)).$$

En Posant $z = 0$, on obtient la probabilité $\pi_0(\infty) = \frac{1-\rho}{\bar{c}\rho\tilde{B}(\lambda H_1)}$ de la fonction génératrice $\frac{1-\rho}{\bar{c}\rho} \frac{(1-C(z))}{g(z)-z}$ de la variable aléatoire $N_{f,\infty}$. Alors

$$\sum_{n=0}^{\infty} |\pi_n(\theta) - \pi_n(\infty)| \geq 2 \frac{1-\rho}{\bar{c}\rho\tilde{B}(\lambda H_1)} (1 - \hat{\pi}_0(\theta)).$$

□

5.5 APPLICATION NUMÉRIQUE

Exemple 5.1 Dans le tableau ci-dessous nous donnons les valeurs des paramètres de la distribution Gamma $\Gamma(\alpha, \mu)$ obtenues pour le système de files d'attente $M^X/G/1$ avec

rappels et clients impatientes, où

$$\alpha = 1 + \frac{\frac{1}{\beta_1 H_1}}{\theta \left(\frac{\beta_2}{2\beta_1^2} + \frac{C''(1)}{2\bar{c}} \right)} \text{ et } \mu = \frac{1}{\left(\frac{\beta_2}{2\beta_1^2} + \frac{C''(1)}{2\bar{c}} \right)}.$$

Considérons que la taille des groupes est distribuée suivant une loi géométrique de paramètre $p = 0.2$ et moyenne $\bar{c} = 5$ (alors $C''(1) = 4$). Les paramètres du système d'attente sont définis comme suit : $\gamma = 1, H_1 = 0.8$. La distribution du temps de service est

1. Exponentielle \mathcal{E} : $B(x) = 1 - e^{-\gamma x}, x \geq 0$, avec coefficient de variation $CV = 1$;
2. Erlang d'ordre deux \mathcal{E}_2 : $B(x) = 1 - e^{-2\gamma x} - 2\gamma x e^{-2\gamma x}, x \geq 0$, avec coefficient de variation $CV \simeq 0.7$;
3. Hyperexponentielle d'ordre deux : \mathcal{H}_2 : $B(x) = 1 - p_1 e^{-\gamma_1 x} - p_2 e^{-\gamma_2 x}, x \geq 0$, où $p_1 + p_2 = 1$ et $\frac{p_1}{\gamma_1} + \frac{p_2}{\gamma_2} = \frac{1}{\gamma}$. Nous choisissons le coefficient de variation $CV = 1.5$, alors $p_1 \simeq 0.19, p_2 \simeq 0.81$ et $\gamma_1 \simeq 0.38, \gamma_2 \simeq 2$.

TABLE 5.1 – Paramètres de la loi Gamma suivant la distribution du temps de service

	Exponentielle	Erlang d'ordre deux	Hyperexponentielle d'ordre deux
β_2	2	0.75	3.25
α	$1 + \frac{1}{1.12 \cdot \theta}$	$1 + \frac{1}{0.62 \cdot \theta}$	$1 + \frac{1}{1.62 \cdot \theta}$
μ	0.7143	1.2903	0.4938

D'après les résultats obtenus et pour chacune des distributions du temps de service, nous faisons la représentation graphique de α où $\alpha = 1 + \frac{\frac{1}{\beta_1 H_1}}{\theta \left(\frac{\beta_2}{2\beta_1^2} + \frac{C''(1)}{2\bar{c}} \right)}$ (le premier paramètre de la loi Gamma vers laquelle converge la distribution du nombre de clients en orbite), en fonction de θ , le taux de rappels. Pour différentes

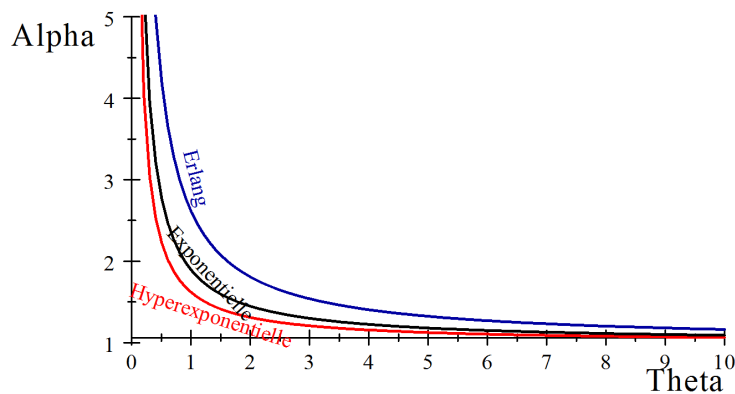


FIGURE 5.1 – Variation de α en fonction de θ

valeurs du taux des rappels θ , la distribution du nombre de clients en orbite a la forme suivante

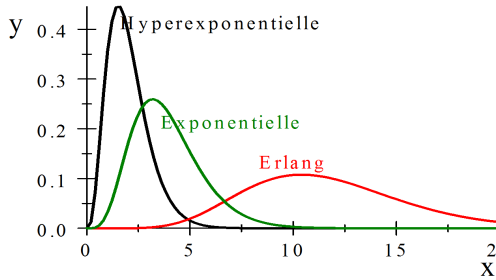


FIGURE 5.2 – Distribution du nombre de clients en orbite pour $\theta = 0.2$

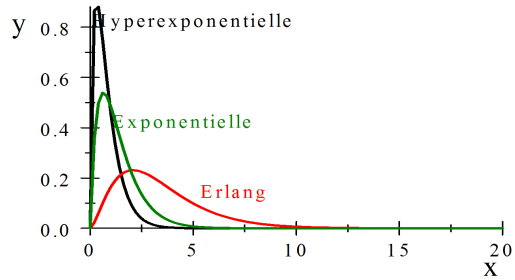


FIGURE 5.3 – Distribution du nombre de clients en orbite pour $\theta = 1$

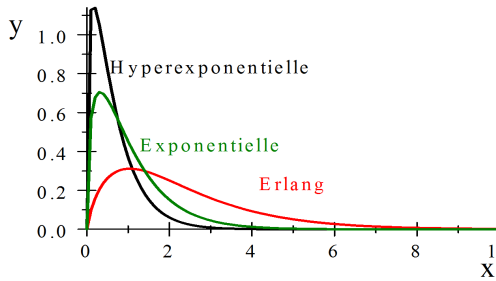


FIGURE 5.4 – Distribution du nombre de clients en orbite pour $\theta = 2$

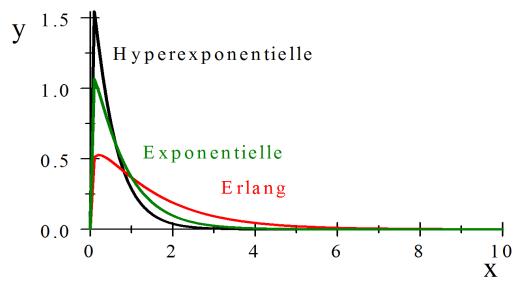


FIGURE 5.5 – Distribution du nombre de clients en orbite pour $\theta = 10$

A partir des résultats présentés dans les figures ci-dessus, nous pouvons conclure que la distribution du nombre de clients en orbite converge vers une loi exponentielle de paramètre μ où $\mu = \frac{1}{\left(\frac{\beta_2}{2\beta_1^2} + \frac{C''(1)}{2c}\right)}$ (le deuxième paramètre de la loi Gamma vers laquelle converge la distribution du nombre de clients en orbite). Cette convergence est plus rapide lorsque le temps de service suit une loi Hyperexponentielle d'ordre deux.

Exemple 5.2 Nous affectons les valeurs suivantes aux différents paramètres du système de files d'attente $M^X/G/1$ avec rappels, arrivées par groupe et clients impatientes : $\lambda = 0.2, \gamma = 1, H_1 = 0.8$ (tel que $\rho = 0.8$). Supposons que la distribution du temps de service suit la loi :

1. Exponentielle E : $B(x) = 1 - e^{-\gamma x}, x \geq 0$, avec coefficient de variation $CV = 1$;
2. Erlang d'ordre deux E_2 : $B(x) = 1 - e^{-2\gamma x} - 2\gamma x e^{-2\gamma x}, x \geq 0$, avec coefficient de variation $CV \simeq 0.7$;
3. Hyperexponentielle d'ordre deux H_2 : $B(x) = 1 - p_1 e^{-\gamma_1 x} - p_2 e^{-\gamma_2 x}, x \geq 0$, où $p_1 + p_2 = 1$ et $\frac{p_1}{\gamma_1} + \frac{p_2}{\gamma_2} = \frac{1}{\gamma}$. Le coefficient de variation $CV = 1.5$, alors $p_1 \simeq 0.19, p_2 \simeq 0.81$ et $\gamma_1 \simeq 0.38, \gamma_2 \simeq 2$.

Et que la taille des groupes est distribuée selon une loi géométrique de paramètre $p = 0.2$ et moyenne $\bar{c} = 5$ (alors $C''(1) = 4$). Dans le tableau ci-dessous nous présentons quelques valeurs numériques de la moyenne et la variance de la loi Gaussienne vers laquelle converge la distribution du nombre de clients en orbite sous un taux de rappel faible.

TABLE 5.2 – Moyenne et variance de la loi Normale suivant la distribution du temps de service

	Exponentielle		Erlang d'ordre deux		Hyperexponentielle d'ordre deux	
β_2	2		0.75		3.25	
θ	moyenne	variance	moyenne	variance	moyenne	variance
0.9	7.56	31.47	7.56	20.36	7.56	42.58
0.5	13.60	56.64	13.60	36.64	13.60	76.64
0.1	68.00	283.20	68.00	183.20	68.00	383.20
0.01	680.00	2832.00	680.00	1832.00	680.00	3832.00

Dans le cas des trois distributions considérées que suit le temps de service, nous observons que les moyennes de la loi Gaussienne sont égales et elles augmentent quand le taux de rappels est proche de 0. Les variances augmentent aussi quand le taux des rappels est proche de 0, mais la dispersion est plus faible dans le cas où le temps de service suit la loi d'Erlang d'ordre deux. Nous déduisons que cette dernière distribution est la mieux adaptée à notre modèle quand le taux de rappels θ est faible. Dans le tableau ci-dessous nous présentons quelques valeurs numériques des Bornes inférieures et des bornes supérieures des inégalités vérifiées par le nombre de clients dans le système quand θ le taux de rappel tend vers ∞ . Nous observons que dans le cas où le temps de service suit la loi Exponentielle, la

TABLE 5.3 – Borne inf et borne sup des inégalités vérifiées quand θ tend vers ∞

	Exponentielle		Erlang d'ordre deux		Hyperexponentielle d'ordre deux	
θ	Borne inf (BI)	Borne sup (BS)	Borne inf (BI)	Borne sup (BS)	Borne inf (BI)	Borne sup (BS)
1	0.00298004	0.05138	0.05586706	0.95794	0.19439570	0.93146
5	0.00060204	0.01038	0.01425924	0.24450	0.04918224	0.23566
10	0.00030160	0.00520	0.00736232	0.12624	0.02536122	0.12152
100	0.00003016	0.00052	0.00075816	0.01300	0.00260875	0.01250
10^3	0.00000348	0.00006	0.00007582	0.00130	0.00026296	0.00126
10^4	0	0	0.00000816	0.00014	0.00002504	0.00012
10^5	0	0	0.00000117	0.00002	0.00000417	0.00002
10^6	0	0	0	0	0	0

distribution du nombre de clients dans le système $M^X/G/1$ avec rappels, arrivées par groupe et clients impatient converge plus rapidement vers la distribution du nombre de clients dans le système classique $M^X/G/1$ avec arrivées par groupe et clients impatient.

Alors nous obtenons le tableau suivant donnant les longueurs des intervalles pour chacune des distributions du temps de service.

TABLE 5.4 – Longueur des intervalles de convergence de la distribution $\pi_n(\theta)$ vers la distribution $\pi_n(\infty)$

	Exponentielle	Erlang d'ordre deux	Hyperexponentielle d'ordre deux
θ	$(BS) - (BI)$	$(BS) - (BI)$	$(BS) - (BI)$
1	0.04839996	0.90207294	0.73706430
5	0.00977796	0.23024076	0.18647776
10	0.00489840	0.11887768	0.09615878
100	0.00048984	0.01224184	0.00989125
10^3	0.00005652	0.00122418	0.00099704
10^4	0	0.00013184	0.00009496
10^5	0	0.00001883	0.00001583
10^6	0	0	0

Nous observons que dans le cas où le temps de service suit la loi Exponentielle la vitesse de convergence est d'ordre 10^4 , par contre pour le cas des deux autres distributions du temps de service, la loi d'Erlang d'ordre deux et la loi Hyperexponentielle d'ordre deux, la vitesse de convergence est d'ordre 10^6 . Nous déduisons que la loi Exponentielle du temps de service est la mieux adaptée à notre modèle quand le taux de rappels θ augmente.

CONCLUSION DU CHAPITRE

Pour les différentes valeurs limites des paramètres du système d'attente $M^X/G/1$ avec rappels, arrivées par groupe et clients impatientes, nous avons établie les résultats suivants

- Sous un taux de trafic intense, la distribution du nombre de clients en orbite converge vers la loi Gamma et selon les résultats numériques la convergence est plus rapide lorsque le temps de service suit une loi Hyperexponentielle d'ordre deux.
- Sous un taux de rappels faible, la distribution du nombre de clients en orbite converge vers la loi Gaussienne et selon les résultats numériques la convergence est plus rapide lorsque le temps de service suit une loi d'Erlang d'ordre deux.
- Sous un taux de rappels élevé, la distribution du nombre de clients dans le système converge vers la distribution du nombre de clients dans le système classique $M/G/1$ avec arrivées par groupes et clients impatientes et selon les résultats numériques la convergence est plus rapide lorsque le temps de service suit une loi Exponentielle.

Des résultats similaires sont obtenus, pour les systèmes d'attente M/G/1 avec rappels, arrivées simples et clients persistants, dans (65). Nous avons réalisé une généralisation de ces résultats obtenus .

Les systèmes de files d'attente satisfaisant les hypothèses suivantes :

- Les clients arrivent par groupes dans le système selon un processus de Poisson. Le temps de service suit une loi générale. Les temps de services des différents clients sont indépendants les uns des autres et aussi indépendants du processus des arrivées.*
- Les clients primaires arrivants dans le système peuvent le quitter si le serveur est occupé.*
- Tous les clients arrivants dans le système seront éventuellement servis, $\rho < 1$.*
- Les clients sont servis dans un ordre indépendant du temps de leur service.*
- La durée entre deux rappels successifs d'une même source secondaire est exponentiellement distribuée.*

vérifient les résultats du théorème 5.3 et aussi le théorème 5.4.

CONCLUSION GÉNÉRALE

Dans ce travail nous nous sommes intéressés aux systèmes de files d'attente de type $M^X/G/1$ avec rappels, arrivées par groupes et clients impatientes :

- Les chapitres un et deux sont une introduction pour le système de files d'attente de type $M^X/G/1$ avec rappels, arrivées par groupe et clients impatientes. Nous avons ainsi présenté les propriétés des algorithmes d'optimisation stochastique et la modélisation de certains d'entre eux dont les files d'attente.*
- Dans les trois derniers chapitres, nous avons réalisé l'analyse stochastique complète du modèle $M^X/G/1$ avec rappels, arrivées par groupes et clients impatientes, à cet effet la fonction génératrice de la chaîne de Markov induite a été obtenue, la propriété de décomposition stochastique a été établie, le comportement asymptotique du nombre de clients en orbite dans le cas d'un régime chargé et dans le cas d'un taux de rappels faible ont été étudiés. Enfin, le comportement asymptotique du nombre de clients dans le système a été étudié.*

Par conséquent les conclusions suivantes sont tirées :

- ☛ sous un régime chargé, la distribution stationnaire du nombre de clients en orbite converge vers la loi Gamma,*
- ☛ sous un taux de rappels faible, la distribution stationnaire du nombre de clients en orbite converge vers la loi Gaussienne,*
- ☛ sous un taux de rappels élevé, la proximité avec le système $M^X/G/1$ avec arrivées par groupe et clients impatientes classique a été estimée.*

BIBLIOGRAPHIE

- [1] A. Aïssani. *Influence des pannes des serveurs sur la distribution de nombre de clients en orbite et dans un système M/G/1/o avec rappels*. *Technologies Avancées*, 2, 23-38, (1991). (Cité page 26.)
- [2] A. Aïssani. *A Survey on Retrial Queueing Models*. *Actes des Journées Statistiques Appliquées, U.S.T.H.B., Alger*, 1-11, (1994). (Cité page 2.)
- [3] A. Aïssani. *An $M^X/G/1$ retrial queue with exhaustive vacations*. *Journal of Statistics and Management Systems* 3, 269-286, (2000). (Cité page 38.)
- [4] A. Aïssani. *An $M^X/G/1$ retrial queue with unreliable server and vacations*. *Proceedings of the 10th International Conference on Analytical and Stochastic Modelling Techniques and Applications, ASMTA'03*, pp. 175-180. SCS- European Publishing House, Nottingham, (2003). (Cité page 38.)
- [5] M. Ajmone Marsan, G. De Carolis, E. Leonardi, R. Lo Cigno and M. Meo. *An approximate model for the computation of blocking probabilities in cellular networks with repeated calls*, *Telecommunication Systems*, 15, 53-62, (2000). (Cité page 6.)
- [6] M. Ajmone Marsan, G. De Carolis, E. Leonardi, R. Lo Cigno and M. Meo. *Efficient estimation of call blocking probabilities in cellular mobile telephony networks with customer retrials*, *IEEE Journal on Selected Areas in Communications*, 19, 332-346, (2001). (Cité page 6.)
- [7] A.S. Alfa and W. Li. *PCS networks with correlated arrival process and retrial phenomenon*. *IEEE Transactions on Wireless Communications*, 1, 630-637, (2002). (Cité page 6.)
- [8] J. Amador and J. R. Artalejo. *The M/G/1 retrial queue : New descriptors of the customer's behavior*. *Journal of computational and Applied Mathematics*, 223, 15-26, (2009). (Cité page 26.)
- [9] J. Amador and J. R. Artalejo. *Transient analysis of the successful and blocked events in retrial queues*. *Telecommunication Systems*, 41, 255-265, (2009). (Cité page 25.)

- [10] V.V. Anisimov. *Switching stochastic models and applications in retrieval queues*, *Top*, 7, 169-186, (1999). (Cité page 25.)
- [11] V.V. Anisimov and M. Kurtulush. *Some Markovian queuing retrieval system under light-traffic conditions*, *Cybernetics and Systems Analysis*, 37, 876-887, (2001). (Cité page 3.)
- [12] N. K. Arrar, N. Djellab and J-B. Baillon. *On the asymptotic behaviour of M/G/1 retrieval queues with batch arrivals and impatience phenomenon*. *Mathematical and computer modelling* 55. 654-665, (2012). (Cité page 4.)
- [13] N. K. Arrar, N. Djellab and J-B. Baillon. *On the $M^X/G/1$ retrieval queue with impatient customers*. *Proceeding of the International Conference in Modern Probabilistic Methods for Analysis and Optimization of Information and Telecommunication Networks*, Minsk, Belarussie, 49-51, (2011). (Cité page 4.)
- [14] N. K. Arrar, N. Djellab and J-B. Baillon. *M/G/1 retrieval queue with impatient batches under low and high intensities of retrials*. *Verification and evaluation of computer and communication systems, Tunisie, VECOS11, 2011* (Cité page 4.)
- [15] N. K. Arrar and N. Djellab. *Classification of some stochastic algorithms*, *Proceeding of the 9th Symposium of Operationnal Research, Slovénia SOR07*, 429-432, (2007). (Cité page 3.)
- [16] J. R. Artalejo. *Accessible bibliography on retrieval queues*. *Mathematical and computer modelling* 30. 1-6, (1999). (Cité pages 2 et 25.)
- [17] J. R. Artalejo. *Accessible bibliography on retrieval queues : Progress in 2000-2009*. *Mathematical and computer modelling* 51, 1071-1081, (2010). (Cité pages 2 et 25.)
- [18] J. R. Artalejo. *Aunified cost function for M/G/1 queueing systems with removable server*. *Trabajo de Investigacion Operativa* 7, 95-104, (1992). (Cité page 26.)
- [19] J. R. Artalejo. *Analysis of an M/G/1 queue with constant repeated attempts and server vacations*. *Computers and Operations Research*, 24(6) :493-504, (1997). (Cité page 57.)
- [20] J. R. Artalejo. *Retrial queues with a finite number of sources*. *Journal of the Korean Mathematical Society*, 35(3) :503-525, (1998). (Cité pages 2 et 39.)

- [21] J. R. Artalejo, 1st International Workshop on Retrial Queues, *Top*, 7, No. 2, 169-353, (1999). (Cité page 2.)
- [22] J. R. Artalejo. Retrial Queueing systems, *Mathematical and Computer Modelling* 30, No. 3-4, 1-228, (1999). (Cité pages 2 et 39.)
- [23] J. R. Artalejo. Algorithmic Methods in Retrial Queues. *Annals of Operation Research*, 141,1-301, (2006). (Cité page 2.)
- [24] J.R. Artalejo and I. Atencia. On the single server retrial queue with batch arrivals. *Sankhya* 66, 140-158,(2004). (Cité page 38.)
- [25] J. R. Artalejo and G. Choudhoury. Steady state analysis of an M/G/1 queue with repeated attempts and two phase service. *Quality Technology and Quantitative Management*, 1(2), 189-199, (2004). (Cité page 26.)
- [26] J. R. Artalejo, A.N. Dudin and V.I. Klimenok. Stationary analysis of a retrial queue with preemptive repeated attempts. *Operations Research Letters*, 28, 173-180, (2001). (Cité page 25.)
- [27] J.R. Artalejo, A. Economou and A. Gomez-Corral. Applications of maximum queue lengths to call center management. *Computers & Operations Research*, 34, 983-996, (2007). (Cité page 25.)
- [28] J. R. Artalejo and G. I. Falin. Stochastic decomposition for retrial queue. *TOP*, 2, 329-342, (1994). (Cité page 56.)
- [29] J. R. Artalejo, G. I. Falin and M. J. López-Herrero. A second order analysis of the waiting time in the M/G/1 retrial queue. *Asia-Pacific Journal of Operational Research*, 19, 131-148, (2002). (Cité pages 25 et 26.)
- [30] J.R. Artalejo and A. Gomez-Corral. Channel idle periods in computer and telecommunication systems with customer retrials. *Telecommunication Systems*, 24, 29-46, (2003). (Cité page 6.)
- [31] J.R. Artalejo and A. Gomez-Corral. Retrial Queueing Systems. A Computational Approach. Springer, (2008). (Cité page 2.)
- [32] J. R. Artalejo and A. Gómez-Corral, *Advances in Retrial Queues*, *European Journal of Operation Research*, (2008). (Cité page 2.)
- [33] J. R. Artalejo and M. J. López-Herrero. On the single server retrial queue with balking. *Infor*, 38, 33-50, (2000). (Cité page 35.)
- [34] J. R. Artalejo and M. J. López-Herrero. On the busy period of the M/G/1 retrial queue. *Naval Research Logistics*, 47, 115-127, (2000). (Cité page 3.)

- [35] J. R. Artalejo and M. J. Lòpez-Herrero. On the M/G/1 queue with quadratic repeated attempts. *Statistical Methods*, 3, 60-78, (2001). (Cit  page 26.)
- [36] J. R. Artalejo and M. J. Lòpez-Herrero. The M/G/1 retrial queue : An information theoretic approach. *Statistics and Operations Association*, 2, 1-17, (2005). (Cit  page 26.)
- [37] J. R. Artalejo and M. J. Lòpez-Herrero. On the distribution of the number of retrials. *Applied Mathematical Modelling*, 31, 478-489, (2007). (Cit  page 25.)
- [38] J. R. Artalejo and M. J. Lòpez-Herrero. Cellular mobile networks with repeated calls operating in random environment. *Computers & Mathematics with Applications*, 37, 1158-1166, (2010). (Cit  page 25.)
- [39] J. R. Artalejo and V. Pla. On the impact of customer balking, impatience and retrials in telecommunication systems. *Computers & Mathematics with Applications*, 57, 217-229, (2009). (Cit  pages 25 et 35.)
- [40] J. R. Artalejo, V. Rajagopalan and R. Sivasamy. On finite Markovian queues with repeated attempts. *Investigacion Operativa*, 9, 83-94, (2000). (Cit  page 25.)
- [41] I. Atencia. A Queueing System under LCFS PR Discipline with General Retrial Times. *Proceeding of the Int. Conference "Modern Mathematical Methods of Investigating of the Information Networks"*. Minsk, pp 30-34, (2001). (Cit  page 6.)
- [42] I. Atencia, G. Bouza and P. Moreno. An $M^{[X]}/G/1$ retrial queue with server breakdowns and constant rate of repeated attempts. *Annals of Operations Research* 157, 225-243, (2008). (Cit  page 38.)
- [43] K. Avrachenkov and U. Yechiali. Retrial networks with finite buffers and their application to internet data traffic. *Probability in the Engineering and Informational Sciences*, 22, 519-536, (2008). (Cit  page 25.)
- [44] N. Bartoli and P. Del Moral. *Simulation et Algorithmes Stochastiques*. C epadu s  ditions, (2001). (Cit  page 5.)
- [45] L. Breuer, V.I. Klimenok, A. Birukov, A.N. Dudin and U.R. Krieger. Mobile networks modeling the access to a wireless network at hot spots. *European Transactions on Telecommunications*, 16, 309-316, (2005). (Cit  page 6.)
- [46] R. Cerf. *Une Th orie Asymptotique des Algorithmes G n tiques*. Th se de Doctorat, Universit  de Montpellier II (France), (1994). (Cit  page 6.)

- [47] B.D. Choi, Y.W. Shin and W.C. Ahn. Retrial queues with collision arising from unslotted CSMA/CD protocol. *Queueing Systems* 11, 335-356, (1992). (Cité page 38.)
- [48] Q. H. Choo and B. Conolly. New results in the theory of repeated orders queueing systems. *Journal of applied Probability*, 16 :631-640, (1979). (Cité page 27.)
- [49] E. Cinlar. *Introduction to Stochastic Processes*. Prentice-Hall, New York, (1975). (Cité page 42.)
- [50] J. W. Cohen. *Basic Problems of Telephone Traffic Theory and Influence of Repeated Calls*. *Philips Telecom. Review*, 18(2) :49-100, (1957). (Cité page 2.)
- [51] A. G. De Kok. *Algorithmic methods for single server systems with repeated attempts*. *Statistica Neerlandica*, 38 :23-32, (1984). (Cité page 30.)
- [52] N. V. Djellab. *On the M/G/1 Retrial Queue to Breakdowns*. *RAIRO : Operations Research*, 36 :299-310, (2002). (Cité page 26.)
- [53] N.V. Djellab. *On the single-server retrial queue*. *YUJOR* 16(1), 45-53, (2006). (Cité page 25.)
- [54] B. T. Doshi. *Queueing systems with vacations. A survey*. *Queueing systems*, 1 :29-66, (1986) . (Cité pages 2, 53 et 54.)
- [55] A. Eldin. *Approach of Theoretical Description of Repeated Call Attempts*. *Ericsson Thecnics*, 23(3) : 346-407, (1967). (Cité page 2.)
- [56] G.I. Falin. *Aggregate arrival of customers in one line systems with repeated calls*. *Ukrainian Mathematical Journal* 28, 337-340, (1976). (Cité pages 38, 68 et 69.)
- [57] G.I. Falin. *A Single-line System with Secondary Orders*. *Engineering Cybernetics Review*, 17(2), 76-83, (1979). (Cité pages 3, 61 et 62.)
- [58] G. I. Falin. *Single-line repeated orders queueing systems*. *Engineering Cybernetics Review*, 21 (6), 21-25, (1983). (Cité pages 20 et 29.)
- [59] G.I. Falin. *Single-line Repeated Orders Queueing Systems*. *Mathematische Operationsforschung und Statistik, Optimization*, 5, 649-667, (1986). (Cité pages 3, 61 et 62.)
- [60] G. I. Falin. *Asurvey of retrial queues*. *Queueing systems*, 7 :127-168, (1990) . (Cité pages 25, 26 et 56.)

- [61] G. I. Falin, J. R. Artalejo and M. Martin. On the single server retrial queue with priority cutomers. *Queueing Systems*, 14, 439-455, (1993). (Cité page 57.)
- [62] G. I. Falin. A single-server batch arrival queue with returning customers. *European Journal of Operational Research*. 201, 786-790, (2010). (Cité page 24.)
- [63] G. I. Falin and A.I. Falin. Heavy traffic analysis of M/G/1 type queueing systems with Markov-modulated arrivals. *Top*, 7, 279-291, (1999). (Cité page 62.)
- [64] G. I. Falin and A. Gomez-Corral. On a bivariate Markov process arising in the theory of single-server retrial queues. *Statistica Neerlandica*, 54, 67-78, (2000). (Cité page 25.)
- [65] G.I. Falin and J.G.C. Templeton. *Retrial Queues*. Chapman and Hall, (1997). (Cité pages 2, 3, 25, 62, 69 et 75.)
- [66] G. Fayolle and M.A. Brun. On a system with impatience and repeated calls. In : *Queueing theory and applications. Liber Amicorum for J.W. Cohen*, pp. 283-305. North Holland, Amsterdam, (1988). (Cité pages 35, 38 et 55.)
- [67] S. Fuhrmann and R. B. Cooper. A note on the M/G/1 queue with server vacations. *Operations research*, 31 :1368, (1981).
- [68] S. Fuhrmann and R. B. Cooper. Stochastic décompositions in the M/G/1 queue with generalized vacations. *Operations research*, 33, 1117-1129, (1985). (Cité page 54.)
- [69] D.P.Gaver. A waiting line with interrupted service including priorities. *J. Roy. Stat. Soc.*, B25, 73-90, (1962). (Cité page 54.)
- [70] E. Gelenbe and R. Iasnogorodski. A queue with server of walking type. *Annales de l'institut Henri Poincaré (B)*, 16(1) :63-73, (1980). (Cité page 55.)
- [71] J. Glimm. Solutions in the Large for Nonlinear Systems of Conservation Laws. *Comm. Pure Appl. Math.*, vol. 18, 1965, p. 695-715. (Cité page 5.)
- [72] D. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, (1989). (Cité page 6.)
- [73] A. Gòmez-Corral. Stochastic Analysis of a Single Server Retrial Queue with General Retrial Times. *Naval Research Logistics*, 46 :561-581, (1999). (Cité page 25.)

- [74] A. Gómez-Corral and M.F. Ramalhoto. *On the waiting time distribution and the busy period of a retrial queue with constant retrial rate*. *Stochastic Modelling and Applications*, 3, 37-47, (2000). (Cité page 2.)
- [75] O. Hashida and K. Kawashima. *Buffer Behavior with Repeated Calls*. *Electronics and Communication in Japan*, 62-B, 27, (1979). (Cité pages 2 et 3.)
- [76] J. Keilson, J. Cozzolino and H. Young. *A service system with unfilled requests repeated*. *Operation research*, 16 :1126-1137, (1968). (Cité page 30.)
- [77] Z. Khalil and G. Fallin. *Stochastic Inequalities for M/G/1 Retrial Queues*. *Operations Research Letters*, 16, 285-290, (1994). (Cité page 26.)
- [78] J. Kim, A. Kanechny and A.N. Dudin. *Threshold control by a single-server retrial queue with batch arrivals and group services*. *Operations Research Letters*, 34, 548-556, (2006). (Cité page 24.)
- [79] J. Kim, B. Kim and S.S. Ko. *Tail Asymptotic for the Queue Size Distribution in an M/G/1 Retrial Queue*. *Journal Applied Probability*, 44, 1111-1118, (2007). (Cité pages 3 et 61.)
- [80] S. Kirkpatrick, C.D. Gelatt Jr, and M.P. Vecchi, *Optimization by Simulated Annealing*, *Science*, V. 220, No. 4598, pp. 671 - 680, (1983). (Cité page 6.)
- [81] E.V. Koba. *Stability conditions for some typical retrial queues*. *Cybernetics and Systems Analysis*, 41, 100-103, (2005). (Cité page 25.)
- [82] E.V. Koba and S.V. Pustovaya. *Call center as retrial queueing system*. *Journal of Automation and Information Sciences*, 39, 37-47, (2007). (Cité page 25.)
- [83] G. Koole and A. Mandelbaum. *Queueing models of call centers : An introduction*. *Annals of Operations Research*, 113, 41-59, (2002). (Cité page 25.)
- [84] B. Krishna Kumar and D. Arivudainambi. *The M/G/1 retrial queue with Bernoulli schedule and general retrial times*. *Computers and Mathematics with Applications*, 43, 251-264, (2002). (Cité page 26.)
- [85] B. Krishna Kumar, D. Arivudainambi and A. Vijayakumar. *On the $M^X/G/1$ retrial queue with Bernoulli schedule and general retrial times*. *Asia Pacific Journal of Operational Research* 19, 177-194, (2002). (Cité page 38.)

- [86] B. Krishna Kumar and S. Pavaai Madheswari. *M^X/G/1 retrial queue with multiple vacations and starting failures*. *Operational Research* 40, 115-137, (2003). (Cité page 38.)
- [87] A. Krishnamoorthy, T.G. Deepak and V.C. Joshua. *An M/G/1 retrial queue with nonpersistent customers and orbital search*. *Stochastic Analysis and Applications*, 23, 975-997, (2005) . (Cité page 35.)
- [88] A. Krishnamoorthy, S.V. Nair and V.C. Narayanan. *An inventory model with retrial and orbital search*. *Bulletin of Kerala Mathematics Association, Special issue*, 47-65, (2009) . (Cité page 25.)
- [89] V. G. Kulkarni and H. M. Liang. *Retrial Queues Revisited*. *Frontiers in Queueing* (J.H. Dshalov, ed.) CRC Press Boca Raton, pp 19-34, (1997). (Cité page 2.)
- [90] P.J.M. Laarhoven and E.H.L. Aarts. *Simulated annealing : theory and applications*, Kluwer Academic Publishers, (1989). (Cité page 6.)
- [91] C. Langaris. *Gated polling models with customers in orbit*. *Mathematical and Computer Modelling* 30, 171-187, (1999). (Cité page 38.)
- [92] C. Langaris and E. Moutzoukis. *A retrial queue with structured batch arrivals, priorities and server vacations*. *Queueing Systems* 20, 341-368, (1995). (Cité pages 38 et 58.)
- [93] C. Langaris and E. Moutzoukis. *A batch arrival reader-writer queue with retrial writers*. *Stochastic Models* 13, 523-545, (1997). (Cité page 38.)
- [94] H. Li and T. Yang. *Geo/G/1 discrete time retrial queue with Bernoulli schedule*. *European Journal of Operation Research*, 111 :629-649, (1998) . (Cité page 57.)
- [95] M.J. Lopez-Herrero. *On the number of customers served in M/G/1 Retrial Queue : first moments and maximum entropy approach*. *Computers Operations Research*, 29, 1739-1757, (2002). (Cité page 26.)
- [96] M.J. Lopez-Herrero. *Distribution of the number of customers served in an M/G/1 retrial queue*. *Journal of applied Probability*, 39, 407-412, (2002). (Cité page 26.)
- [97] J. Lubacz and J. Roberts. *A new approach to the single-server repeated attempt system with balking*. *Proceedings of the Third Int. Seminar on Teletraffic Theory. Moscow*, 290-293, (1984). (Cité page 2.)

- [98] F. Machihara and M. Saitoh. *Mobile customer model with retrials*. *European Journal of Operational Research*, 189, 1073-1087, (2008). (Cité page 25.)
- [99] M. Martin and J.R. Artalejo. *Analysis of an M/G/1 queue with two types of impatient units*. *Advances in Applied Probability* 27, 840-861, (1995). (Cité pages 25, 35 et 38.)
- [100] M. Martin and A. Gomez-Corral. *On the M/G/1 retrial queueing system with linear control policy*. *TOP*, 3, 285-305, (1995). (Cité page 25.)
- [101] P. Moreno. *An M/G/1 retrial queue with recurrent customers and general server times*. *Applied Mathematics and Computation*, 159, 651-666, (2004). (Cité page 26.)
- [102] R.D. Nobel and H.C. Tijms. *Waiting-time probabilities in the M/G/1 retrial queue*. *Statistica Neerlandica*, 60, 73-78, (2006). (Cité page 26.)
- [103] R.D. Nobel and P. Moreno. *A discrete-time retrial queueing model with one server*. *European Journal of Operational Research*, 189, 1088-1103, (2008). (Cité page 25.)
- [104] H. Ohmura and Y. Takahashi. *An Analysis of Repeated Call Model with Finite Number of Sources*. *Electronics and Communication in Japan*, 68 :112-121, 27, (1985). (Cité page 25.)
- [105] A. Rodrigo and M. V Vazquez. *Large Sample Inference in Retrial Queues*. *Mathematical and Computer Modelling*, 30, 197-206, (1999). (Cité page 25.)
- [106] A. Rodrigo. *Estimators of the retrial rate in M/G/1 retrial queues*. *Asia-Pacific Journal of Operational Research*, 23, 193-213, (2006). (Cité page 2.)
- [107] J. Roszik, J. Sztrik and C.S. Kim. *Retrial queues in the performance modelling of cellular mobile networks using MOSEL*. *International Journal of Simulation* 6, 38-47, (2005). (Cité page 25.)
- [108] L.I. Sennott, P.A. Humblet and R.I. Tweedie. *Mean drifts and non ergodicity of Markov chains*. *Operations Research* 31, 783-789, (1983). (Cité page 42.)
- [109] W. Shang, L. Liu and Q.L. Li. *Tail asymptotics for the queue length in an M/G/1 retrial queue*. *Queueing Systems*, 52, 193-198, (2006). (Cité page 3.)

- [110] J. G. C. Templeton, *Retrial Queues, Queueing Systems*, 7, No. 2, 125-227, (1990). (Cité page 2.)
- [111] J. G. C. Templeton, *Retrial Queues, Top*, 7 : 351-353, (1999). (Cité page 2.)
- [112] P. Tran-Gia and M. Mandjes. *Modelling of customer retrial phenomenon in cellular mobile networks. IEEE Journal on Selected Areas in Communication*, 15 :1406-1414, (1997) . (Cité page 2.)
- [113] W. Whitt. *On Approximations for Queues, I : Extremal Distributions. AT &T Bell Laboratoire Technical Journal*, 63, 115-138, (1984). (Cité page 3.)
- [114] X. Wu, P. Brill, M. Hlynka and J. Wang. *An M/G/1 retrial queue with balking and retrials during service. Internationam Journal of Operational Research*, 1, 30-51, (2005) . (Cité page 35.)
- [115] T.Yang and J.G.C. Templeton. *A survey on retrial queues. Queueing Systems* 2, 201-233, (1987). (Cité pages 2, 56 et 57.)
- [116] T.Yang, J.G.C. Templeton. *The $C_a/M/s/m$ retrial queue : a computational approach. ORSA Journal on Computing*, 4(2) :182-191, (1992). (Cité page 34.)
- [117] T. Yang et al. *An approximation method for the M/G/1 retrial queue with general retrial times. European Journal of Operation Research*, 76 :552-562, (1994). (Cité pages 2 et 56.)

Titre *Problèmes de convergence, optimisation d'algorithmes et analyse stochastique de systèmes de files d'attente avec rappels*

Résumé *Pour optimiser la gestion des réseaux de télécommunication, nous considérons le système de file d'attente $M^X/G/1$ avec rappels et clients impatientes. En utilisant la méthode des variables supplémentaires, nous obtenons les fonctions génératrices partielles de l'état stationnaire conjointe de l'état du serveur et du nombre de clients dans le groupe de rappels. Pour compléter l'analyse du modèle considéré, nous calculons la distribution stationnaire de la chaîne de Markov induite, grâce à laquelle nous présentons la propriété de la décomposition stochastique. Cependant, la fonction génératrice de la distribution stationnaire du nombre de clients dans le groupe de rappels, est obtenu sous une forme explicite, très complexe et ne révèle pas la nature de la distribution en question. Alors, nous étudions le comportement asymptotique de la variable aléatoire représentant le nombre de clients en orbite et dans le système pour des valeurs limites des différents paramètres. Nous complétons notre travail par des exemples numériques.*

Mots-clés *Rappels, arrivées par groupe, impatience, décomposition stochastique, comportement asymptotique, optimisation stochastique.*

Title *Convergence problems, optimization of algorithms and stochastic analysis of retrial queueing systems*

Abstract *To optimize the networks control in telecommunication, we consider an $M^X/G/1$ retrial queue with impatient customers. By using the method of supplementary variables, we obtain the partial generating functions of the steady state joint distribution of the server state and the number of customers in the retrial group are obtained. To complete the analysis of the considered model, we find the steady state distribution of the embedded Markov chain. We investigate the stochastic decomposition property. Although the generating function of the steady state distribution of the number of customers in the retrial group can be obtained in explicit form, it is cumbersome and does not reveal the nature of the distribution in question. Therefore, we investigate the asymptotic behaviour of the random variable representing the number of customers in the retrial group and in the system under limit values of various parameters. We complete this work by numerical examples.*

Keywords *Retrial queues, batches customer, impatience, stochastic decomposition, asymptotic behaviour, stochastic optimization.*

لعنوان مسألة التقارب، تحسين الخوارزميات والتحليل العشوائي لأنظمة قوائم الانتظار ذات النداء المتكرر.

ملخص لتحسين إدارة شبكات الاتصالات السلكية واللاسلكية، نعتبر نظام قائمة انتظار $M^X / G / 1$ ذات النداء المتكرر، الوصول في أفواج و العملاء نافذة الصبر. باستخدام أسلوب المتغيرات الإضافية، نحصل على الدوال المولدة الجزئية للتوزيع الثابت المشترك لوضع الموزع وعدد العملاء في قائم ة الانتظار ذات النداء المتكرر. لإتمام تحليل هذا النموذج، حسبنا التوزيع الثابت لسلسلة ماركوف المستخلصة، التي من خلالها نقدم خاصية التفكير العشوائي. إن الدوال المولدة للتوزيع الثابت لعدد العملاء في الطابور ذات النداء المتكرر و أيضا لعدد العملاء في كامل النظام، تحصلنا عليها في شكل واضح ، لكن محقد للغاية ولا يكشف طبيعة توزيعها. لذا ندرس سلوكها التقريبية وذلك حسب قيم قصوى لمختلف العوامل. نتم عملنا هذا بأمتلة عددية.

كلمات البحث قوائم الانتظار ذات النداء المتكرر ، الوصول في أفواج ، نقاد الصبر، التفكير العشوائي، السلوك التقريبية ، التحسين العشوائي.