

وزارة التعليم العالي والبحث العلمي

MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA
RECHERCHE SCIENTIFIQUE



جامعة باجي مختار - عنابة

BADJI MOKHTAR - ANNABA UNIVERSITY
UNIVERSITÉ BADJI MOKHTAR - ANNABA



كلية العلوم

FACULTY OF SCIENCES
FACULTE DES SCIENCES

Département de Mathématiques
Laboratoire de Probabilités et Statistiques (LaPS)



THÈSE

Présentée en vue de l'obtention du diplôme de
Doctorat en Mathématiques

Option : Modélisation Mathématiques

COMPORTEMENT ASYMPTOTIQUE DE SYSTEME DE FILES D'ATTENTE AVEC RAPPELS ET PRIORITE

Par : Mme DERROUCHE Lamia

Sous la direction de :

DIRECTEUR DE THÈSE : ARRAR REMITA Nawel M.C.A U.B.M. Annaba
CO-DIRECTEUR : DJELLAB Natalia Prof. U.B.M. Annaba

Devant le jury

PRESIDENT : REMITA Med Riad Prof. U.B.M. Annaba
Examineurs : BOUALEM Mohamed Prof. U.A.M. Béjaia
SEDDIK-AMEUR Nacira Prof. U.B.M. Annaba
BENMOSTEFA Fatima Zohra M.C.A U.B.M. Annaba

Année 2018

Dédicaces

À ceux qui m'ont tout donné sans rien demander, qui m'ont offert leur soutien durant toute ma vie, qui m'ont aidé sans restriction, qui m'ont appris les bons principes, à qui tous les mots du monde n'arriveront jamais à les décrire.

*À l'esprit de mon père qui m'a beaucoup soutenu dans ce que je suis aujourd'hui. À ma source de courage, à ce que j'ai de plus cher
ma mère.*

À mon cher mari pour l'aide, le soutien et les encouragements dont il a toujours fait preuve à mon égard.

À mes chères sœurs et mon cher frère. À mon adorable fille Israe à qui je souhaite une longue vie pleine de bonheur et réussite.

À toute ma famille. . .

À toutes mes amies. . .

À toute personne, qui m'a aidé à réaliser ce modeste travail, de près ou de loin.

REMERCIEMENTS

Je tiens à exprimer mes sincères remerciements à Madame Nawel ARRAR REMITA, maître de conférences à l'université de Annaba, 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 de la théorie des systèmes de files d'attente. Elle m'a toujours soutenu moralement et scientifiquement et a pris le temps de répondre à mes nombreuses questions. Ce fut très sincèrement un réel plaisir de travailler à ses côtés pendant ces années et m'avoir ainsi permis de réaliser cette thèse.

Je remercie très chaleureusement Madame Natalia DJELLAB, professeur à l'université de Annaba pour m'avoir confié ce sujet de thèse. Je tiens également à la remercier pour sa disponibilité, la confiance qu'elle m'a accordée et pour ses conseils et ses commentaires fort utiles qui ont fortement enrichi ma formation et guidé dans mon travail de recherche.

J'adresse mes sincères remerciements à Monsieur le Doyen Med Riad REMITA, professeur à l'université de Annaba, pour avoir accepté de présider le jury.

J'exprime ma gratitude à Monsieur Mouhamed BOUALEM, professeur à l'université de Bejaia qui a accepté de rapporter cette thèse et pour l'intérêt qu'il y a porté.

Je remercie également Madame Nacira Seddik-Ameur, professeur à l'université de Annaba et Madame Fatima Zohra Benmostefa, maître de conférences à l'université de Annaba, d'avoir bien voulu examiner ma thèse et faire partie du jury.

Abstract

THE TITLE : On the Asymptotic Behaviour of the queueing systems with Retrials and Priority

In this work, we are interested in M/G/1 retrial queue with priority customers, Bernoulli schedule, FCFS orbit and general retrial times, which stochastic analysis have been performed in the literature. Our contribution consists in the study of the asymptotic behaviour of the number of customers in the retrial group under heavy traffic and also low retrial rate.

We present some numerical illustrations to support the obtained theoretical results.

Key Words : Retrial queue with priority, General retrial times, retrial group, waiting line, Asymptotic Behavior.

Résumé

LE TITRE : Comportement asymptotique des systèmes de files d'attente avec rappels et priorité

Dans ce travail, nous nous sommes intéressés aux modèles $M/G/1$ avec rappels et priorité relative, discipline FCFS en orbite et le temps inter-rappels général, pour lesquelles l'analyse stochastique est réalisée. Notre contribution consiste à l'étude du comportement asymptotique du nombre de clients dans le groupe de rappels sous un trafic intense et également sous un taux de rappels faible et fort.

Nous présentons quelques illustrations numériques pour soutenir les résultats théoriques obtenus.

Mots clés : File d'attente avec rappels et priorité, Temps de rappels général, groupe de rappels, Comportement asymptotiques.

ملخص

العنوان: السلوك المقارب لأنظمة قوائم الانتظار ذات النداء المتكرر والأولوية

في هذا العمل نحن مهتمون بنظام قائمة انتظار $M/G/1$ ذات النداء المتكرر، عملاء نوي الأولوية، جدول برنولي والمدار FCFS ووقت النداء المتكرر عام التي تم تقديم تفكيكها العشوائي

مساهمتنا تتكون في دراسة السلوك المقارب لعدد العملاء في مجموعة النداء المتكرر تحت المرور الكثيف وأيضا انخفاض معدل النداء المتكرر.

نتم عملنا هذا بأمانة عديدة

كلمات البحث: قائمة الانتظار مع النداء المتكرر والأولوية، وقت النداء المتكرر العام، مجموعة من النداءات المتكررة، السلوك المقارب.

Table des matières

0.1	Introduction Générale	4
1	Modèles d'attente avec rappels et priorité	7
1.1	Introduction	7
1.2	Politiques d'accès au serveur	9
1.2.1	Description de la priorité relative	9
1.2.2	Description de la priorité absolue	10
1.3	Synthèse sur les modèles avec rappels et priorité	10
1.3.1	Modèle de Files d'attente $M_1, M_2/G/1$ avec rappels et priorité . . .	10
1.3.2	Modèle de file d'attente $M_1; M_2/G/1$ avec rappels et perte géométrique	14
1.3.3	Modèle de file d'attente $M_1; M_2/G/1/K + 1$ avec rappels et file prioritaire à capacité finie	16
1.3.4	Modèle de file d'attente $M_1; M_2/G/1$ avec rappels	16
1.3.5	Modèle de file d'attente $M_1; M_2/G_1; G_2/1$ avec rappels et feedback . .	16
1.3.6	Modèle de file d'attente $M^{(X_1, X_2)}/G_1; G_2/1$ avec rappels et priorité relative et vacance unique	17
2	Comment simuler un système de files d'attente ?	18
2.1	Introduction à la simulation	18
2.2	Principaux éléments d'un simulateur par événements discrets	19
2.3	Déroulement de la simulation	19
2.4	Simulation à l'aide des logiciels mathématiques	19
2.4.1	Utilisation de MATLAB	23
3	Système de files d'attente $M/G/1$ avec rappels, priorité relative et orbite FCFS	30
3.1	Introduction	30
3.2	Description du modèle	30
3.3	La chaîne de Markov induite	31
3.3.1	Condition d'ergodicité	32
3.3.2	Distribution stationnaire de la chaîne de Markov induite	33
3.4	Distributions stationnaires de l'état du système	35
3.5	Propriété de décomposition stochastique	41

3.5.1	Propriété de décomposition stochastique de la distribution de la taille du système	41
3.5.2	Propriété de décomposition stochastique pour le nombre de clients en orbite	43
3.6	Mesures de performance	44
3.6.1	Cas spéciaux	45
3.7	Conclusion	46
4	Comportement asymptotique du nombre de clients dans le modèle M/G/1 avec rappels et priorité relative et orbite FCFS	47
4.1	Introduction	47
4.2	Trafic intense	47
4.3	Intensité faible des rappels	53
4.4	Intensité forte des rappels	56
4.5	Conclusion	59
5	Illustrations numériques	60
5.1	Cas du trafic intense	60
5.2	Cas de l'intensité faible des rappels	64
5.3	Cas de l'intensité forte des rappels	64
5.4	Effet des paramètres essentiels sur les principales caractéristiques de performance	66
5.5	Effet de la distribution des temps inter-rappels	66
5.5.1	Effet de la politique de priorité	68
5.5.2	Conclusion	68

Table des figures

1.1	Représentation schématique d'un système de files d'attente avec rappels . . .	8
1.2	Graphe des transitions de modèle M/G/1 avec rappels et priorité.	12
5.1	Modèle $M/M/1$ avec rappel. $\theta = 2, p = 0.25$	62
5.2	Modèle $M/M/1$ avec rappel. $\theta = 2, p = 0.5$	62
5.3	Modèle $M/M/1$ avec rappel. $\theta = 2, p = 0.75$	62
5.4	Modèle $M/E_2/1$ avec rappel. $\theta = 2, p = 0.25$	63
5.5	Modèle $M/E_2/1$ avec rappel. $\theta = 2, p = 0.5$	63
5.6	Modèle $M/E_2/1$ avec rappel. $\theta = 2, p = 0.75$	63
5.7	Modèle $M/E_2/1$ avec rappels. $p = 0.5, \lambda = 0.4$	64
5.8	Modèle $M/E_2/1$ avec rappels. $p = 0.5, \lambda = 0.3, \beta_1 = 1$	64
5.9	Modèle $M/E_2/1$ avec rappels. $p = 0.5, \lambda = 0.3, \beta_1 = 1$	65
5.10	Le facteur d'occupation par rapport à θ	67
5.11	Le nombre moyen de clients dans le système par rapport à θ	67
5.12	Le facteur d'occupation par rapport à q	68
5.13	Le nombre moyen du client dans le système par rapport à q	68

0.1 Introduction Générale

La 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, cherchant 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.

Il est possible de classifier les systèmes de files d'attente avec rappels en deux catégories selon

le nombre de clients différents : modèles avec un seul type de clients (population des clients est homogène) et modèles avec plusieurs types de clients où certains jouissent d'une priorité de service. La priorité peut être absolue ou relative. Par priorité absolue, on entend que le service d'un client moins prioritaire sera suspendu lorsqu'un client plus prioritaire se présente devant le serveur. Ce dernier venu commence son service immédiatement. Si la priorité est celle relative, un nouveau client plus prioritaire attend la fin du service avant de commencer le sien. Dans le cas de priorité absolue, deux possibilités se présentent : soit le client suspendu reprend son service là où il a été interrompu, soit il le reprend depuis le début. Le système de files d'attente de type M/G/1 avec rappels et priorité a fait l'objet de plusieurs recherches. Cependant, la plupart des investigations sont portées sur un modèle avec deux types de clients et priorité relative : un client primaire prioritaire (type 1) qui trouve à son arrivée le serveur occupé, prend place dans une file d'attente, tandis qu'un client de type 2 (dans la même situation) entre en orbite et devient source de tentatives répétées. Les clients en orbite peuvent être servis lorsque la file d'attente est vide. La recherche sur les modèles avec rappels et priorité absolue est limitée. Peu de résultats analytiques sont connus sur le sujet.

Dans le but d'étudier mathématiquement un système de file d'attente, en premier lieu, on doit introduire un processus stochastique décrivant l'état du système à un instant donné du temps. En général, on distingue deux catégories de processus stochastique décrivant l'état du système de file d'attente : les processus stochastiques Markoviens et ceux non-Markoviens. Cependant, on dispose de plusieurs méthodes qui permettent de rendre ces derniers Markoviens moyennant certaines transformations (méthode de la chaîne de Markov induite, méthode des variables supplémentaires). Puis, en fonction des quantités qui déterminent la structure du système, on cherche à calculer le régime transitoire. Il est constaté que le calcul explicite de ce dernier est pénible (voir impossible) pour la majorité des modèles. C'est pourquoi, il est souvent préférable de se contenter par la détermination du régime stationnaire. La distribution stationnaire du processus stochastique introduit permet d'obtenir les indices de performance du système : le temps d'attente d'un client, le nombre moyen de client dans le système, le taux d'occupation des dispositifs de service, ... [60].

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) [34]. Des résultats fondamentaux relatifs à cette approche ont été établis par Yang et Templeton [73], ainsi que par Yang et al. [75], 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 [35], [36] et Falin et Templeton [40]. D'autres études ont été établies dans [2], [13], [47], [53], [63]. Il existe aussi des travaux sur des modèles classiques sans rappels dans W. Whitt [70].

Cette thèse est organisée en cinq chapitres. Dans le premier chapitre, nous présentons (ou nous rappelons) les systèmes de files d'attente avec rappels, ainsi que les politiques d'accès au serveur, la propriété de priorité que peut posséder un client ainsi qu'une synthèse sur les systèmes de files d'attente avec rappels et priorité. Dans le deuxième chapitre, un autre objectif est ciblé :

- La simulation de système de files d'attente $M/G/1$ avec rappels en utilisant le MATLAB. La description et l'analyse stochastique du système de files d'attente de type $M/G/1$ avec rappels, priorité relative et orbite FCFS fait l'objet du troisième chapitre. A cet effet, la chaîne de Markov induite est décrite, la condition d'ergodicité est établie à l'aide du critère de Foster et de la condition de Kaplan, la distribution stationnaire de l'état du système est obtenue en appliquant la méthode des variables supplémentaires et les mesures de performance sont déduites. En plus, nous établissons la propriété de décomposition stochastique pour notre modèle, puis, pour certaines valeurs particulières des paramètres de priorité et de rappels, nous présentons des cas spéciaux.

Dans le chapitre quatre, à l'aide de la propriété de décomposition stochastique, nous étudions le comportement asymptotique du nombre de clients dans le groupe de clients réalisant des tentatives répétées pour accéder au serveur du modèle $M/G/1$ avec rappels, priorité relative et orbite FCFS, et cela 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.

Dans le cinquième et dernier chapitre, nous présentons des exemples numériques des résultats théoriques obtenus dans le chapitre précédent.

Chapitre 1

Modèles d'attente avec rappels et priorité

1.1 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, des réseaux de télécommunications Yang et Templeton, [73], des réseaux mobiles cellulaires, des réseaux IP ainsi que des réseaux locaux LAN sous des protocoles de transmission comme CSMA/CD. Une description complète de situations des systèmes de files d'attente avec rappels peut être trouvée dans la monographie de Falin et Templeton, [40]. Une classification bibliographique est donnée dans les articles d'Artalejo [6] et [7]. Le modèle M/G/1 avec rappels et clients non-persistants a été considéré par Falin [37] par Martin et Artalejo [56].

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, 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 fini et si l'orbite est pleine, le client quitte le système pour toujours. Lorsqu'un client est rappelé de

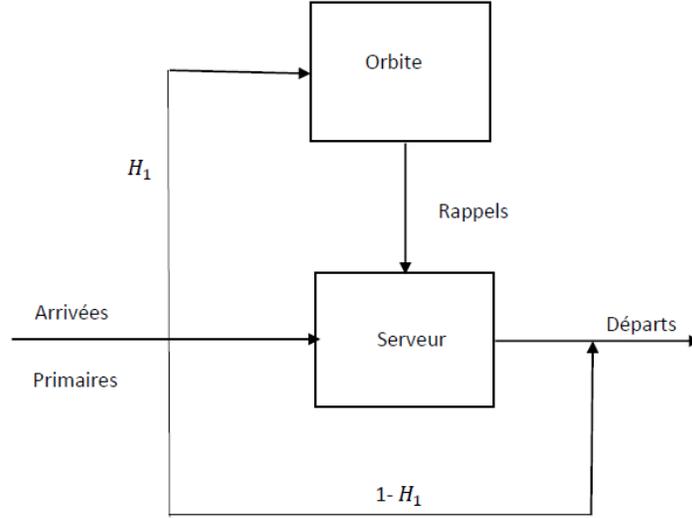


FIG. 1.1 – Représentation schématique d'un système de files d'attente avec rappels

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 $\{\tau_n^e, M_n\}, n \geq 1\}$, 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 $(\Omega, \mathcal{F}, IP)$. 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.

Les systèmes des files d'attente avec rappels peuvent être classifiés en deux catégories selon le nombre de clients différents : modèles avec un seul type de clients et modèles avec plusieurs types de clients (ces derniers peuvent renfermer le phénomène de priorité). Plusieurs auteurs ont étudié les systèmes de files d'attente avec rappels et priorité. Une étude des résultats importantes existe dans l'article (Choi et Chang, [29]). Dans la littérature, il est souvent supposé que les clients prioritaires sont en attente et vont être servis selon une certaine discipline et les clients non-prioritaires (le cas bloqué) entrent en orbite pour retenter sa chance. De plus, les clients prioritaires possèdent une priorité relative (sans préemption) ou

une priorité absolue (avec préemption) sur les clients non-prioritaires (Artalejo, Dudin [14]; Bocharov, Pavlova et Puzikova [20]; Boutarfa et Djellab [21]; Falin et Templeton [40]). Dans les dernières contributions, sur le système des files d'attente avec rappels et le mécanisme de priorité ont analysé par différentes méthodes (Dimitrou [31]; Y.W. Lee [53]; Wu et Lian [71]).

Vu que dans ce travail nous nous intéressons aux modèles avec rappels et priorité, dans le chapitre présent, nous discutons les différentes politiques d'accès au serveur en décrivant la priorité relative et celle absolue, puis nous présentons une synthèse bibliographique sur les modèles en question.

1.2 Politiques d'accès au serveur

Les clients peuvent former une ou plusieurs files d'attente, éventuellement caractérisées par des priorités différentes (la règle de priorité déterminant l'ordre dans lequel les clients vont accéder à la ressource modélisée par le serveur). Les disciplines d'attente classiques, ainsi que leurs acronymes, sont *FCFS* : first come first served "premier arrivé, premier servi", c'est la discipline de service employée le plus souvent et c'est celle qui sera admise par défaut; *LCFS* : last come, first served "dernier arrivé, premier servi"; *SIRO* : service dans un ordre aléatoire; *RR* : round robin "les clients sont servis 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.

La règle de priorité comporte la priorité relative (sans préemption) et la priorité absolue (avec préemption). Le système de files d'attente de type $M/G/1$ avec rappels et priorité a fait l'objet de plusieurs recherches. Cependant, pour décrire la priorité relative et la priorité absolue, on considère un système de file d'attente avec un seul serveur et deux types de clients (type 1 - prioritaire et type 2 - non prioritaire) qui arrivent dans le système selon un processus de Poisson de taux λ_1, λ_2 respectivement. On suppose que la file est régie par une discipline *FCFS* et que le nombre de places d'attente ainsi que celui des clients susceptibles d'accéder au système sont illimités. Le temps de service des deux types de clients peut être différent.

1.2.1 Description de la priorité relative

Dans le cas d'une priorité relative, un nouveau client prioritaire qui arrive et trouve le serveur occupé par le service d'un client (prioritaire ou non prioritaire) prend une place dans la file d'attente, contrairement à un client non prioritaire qui doit (dans la même situation) entrer en orbite.

1.2.2 Description de la priorité absolue

Dans le cas d'une priorité absolue, un client prioritaire qui arrive et trouve le serveur occupé par le service d'un autre client prioritaire, prend une place dans la file d'attente, contrairement au cas où il le trouve occupé par le service d'un client non prioritaire : il commence immédiatement son service. Le client, dont le service a été interrompu par l'arrivée du client prioritaire, quitte le système ou entre en orbite. Toutefois, deux possibilités se présentent : soit le client suspendu reprend son service là où il a été interrompu, soit il le reprend depuis le début.

Dans les deux cas, les clients en orbite ne seront servis sauf s'il n'y a aucun client dans la file d'attente.

1.3 Synthèse sur les modèles avec rappels et priorité

Les modèles d'attente développés ces dernières décennies tentent de prendre en considération des phénomènes de répétition de demandes de service et de priorité à la fois.

Dans ce type de système, il est assez naturel d'envisager une file d'attente avec rappels à deux types de clients où une fois bloqué, les clients de type 1 sont en attente dans l'espace de service et seront servis dès que le serveur soit libre, et les clients de type 2 entrent en orbite pour retenter leur chance, de sorte que les clients de type 1 ont la priorité sur les clients de types 2. L'analyse des systèmes de files d'attente avec rappels avec deux types de clients est essentiellement plus difficile que celle des systèmes de files d'attentes avec rappels à un seul type de client.

1.3.1 Modèle de Files d'attente $M_1, M_2/G/1$ avec rappels et priorité

Dans la plupart des systèmes de files d'attente avec rappels à un seul type de client, il est supposé que les caractéristiques des clients arrivant dans le système, comme le temps d'inter-arrivées, le temps de service et le temps d'inter-rappels sont homogènes. Cependant, dans la pratique, il existe plusieurs types de clients ayant généralement des caractéristiques différentes.

Description du modèle

Deux types de clients (type 1 prioritaire et type 2 non prioritaire) arrivent dans le système selon des processus de Poisson des taux λ_1 et λ_2 , respectivement. Le taux d'arrivée total est $\lambda = \lambda_1 + \lambda_2$. Un client prioritaire qui trouve à son arrivée le serveur occupé, prend place dans une file d'attente, tandis qu'un client type 2 (dans la même situation) entre en orbite et devient source d'appels répétés. Les clients en orbite peuvent être servis lorsque la file d'attente est vide. Par conséquent, les clients type 1 possèdent une priorité relative (sans préemption) sur les clients type 2. Les temps inter-rappels sont mutuellement indépendants et exponentiellement distribués de paramètre $\theta > 0$. Les durées de service suivent une loi

générale de fonction de répartition $B_i(x)$, de Transformée de Laplace-Stieltjes $\tilde{B}_i(x)$, $\text{Re}(s) > 0$ et $i = 1, 2$. Soient les moments d'ordre k $\beta_{i,k} = (-1)^k \tilde{B}_i^{(k)}(0)$. Le taux d'accomplissement du service (l'intensité instantanée du service) est $b_i(x) = \frac{B_i'(x)}{1 - B_i(x)}$, $i = 1, 2$. Toutes les variables définies précédemment sont supposées mutuellement indépendantes.

Soit $\rho = \rho_1 + \rho_2$, où $\rho_i = \lambda_i \beta_{i,1}$, $i = 1, 2$, qui est la charge contribué par des clients type i . On a également,

$$k_{imn} = \int_0^{\infty} \frac{(\lambda_1 x)^m}{m!} e^{-\lambda_1 x} \frac{(\lambda_2 x)^n}{n!} e^{-\lambda_2 x} dB_i(x)$$

la distribution conjointe du nombre de clients primaires (type 1 et type 2) arrivant dans le système durant le temps de service d'un client de type $i = 1, 2$. Sa fonction génératrice est donnée par

$$K_i(z_1, z_2) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} k_{imn} z_1^n z_2^m = \tilde{B}_i(\lambda_1 - \lambda_1 z_1 + \lambda_2 - \lambda_2 z_2).$$

Condition d'ergodicité

Choi et park [26] ont montré que $\rho = \rho_1 + \rho_2 < 1$ est la condition de stabilité d'une file $M_1, M_2/G/1$ avec rappels. La condition d'ergodicité d'une $M_1, M_2/G_1, G_2/1$ a été obtenue dans Falin, Artalejo et Martin [39].

Distribution du nombre de clients dans les deux files

Choi et Park [26] ont obtenu la distribution commune des longueurs des deux files en utilisant la méthode de la variable supplémentaire, où la variable aléatoire supplémentaire est le temps de service restant d'un client ou d'une demande en service.

Soient ;

$N_1(t)$: le nombre d'appels de Type 1 en file d'attente prioritaire (sauf l'appel en service) à l'instant t ,

$N_2(t)$: le nombre d'appels de Type 2 dans l'orbite à l'instant t ,

$$\xi(t) = \begin{cases} 0, & \text{si le serveur est libre à l'instant } t; \\ 1; & \text{si le serveur est occupé par un appel de type 1 à l'instant } t; \\ 2; & \text{si le serveur est occupé par un appel de type 2 à l'instant } t. \end{cases}$$

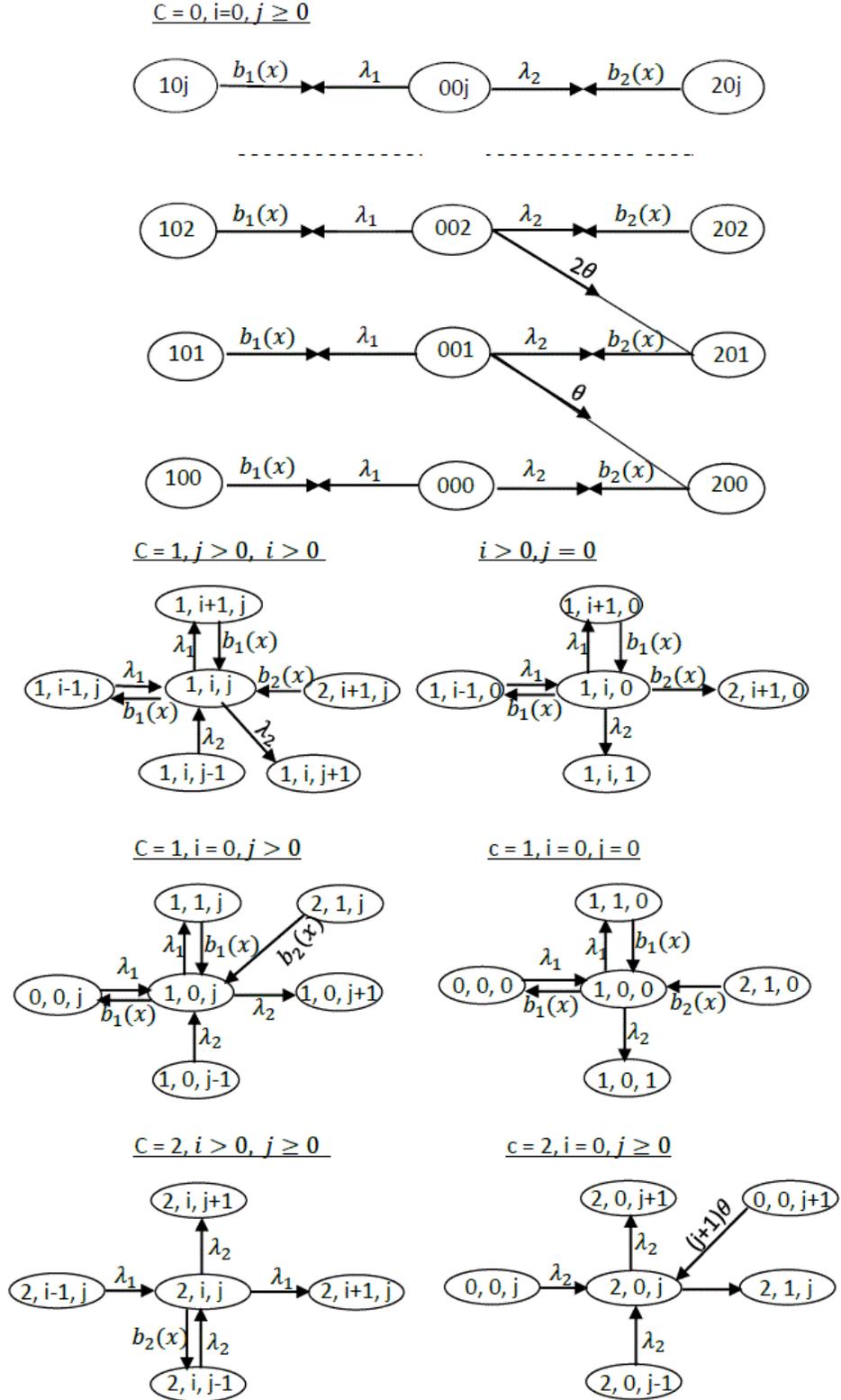


FIG. 1.2 – Graphe des transitions de modèle M/G/1 avec rappels et priorité.

Les fonctions génératrices partielles

Les probabilités stationnaires sont obtenues à l'aide des fonctions génératrices :

$$P_0(z_2) = (1 - \rho_1 - \rho_2) \exp \left\{ \frac{1}{\theta} \int_1^{z_2} \frac{\lambda_1 - \lambda_1 h(u) + \lambda_2 - \lambda_2 K_2(h(u), u)}{K_2(h(u), u) - u} du \right\};$$

$$\begin{aligned} P_1(z_1, z_2) &= [\lambda_1 - \lambda_1 h(z_2) + \lambda_2 - \lambda_2 K_2(h(z_2), z_2)] \times K_2(z_1, z_2) \\ &\quad - (\lambda_1 - \lambda_1 z_1 + \lambda_2 - \lambda_2 K_2(z_1, z_2)) \times K_2(h(z_2), z_2) \\ &\quad \times (K_2(h(z_2), z_2) - z_2) (z_1 - K_1(z_1, z_2))^{-1} \\ &\quad \times \frac{1 - K_1(z_1, z_2)}{\lambda_1 - \lambda_1 z_1 + \lambda_2 - \lambda_2 z_2} P_0(z_2); \end{aligned}$$

$$P_2(z_1, z_2) = \frac{\lambda_1 - \lambda_1 h(z_2) + \lambda_2 - \lambda_2 z_2}{K_2(h(z_2), z_2) - z_2} \times \frac{1 - K_1(z_1, z_2)}{\lambda_1 - \lambda_1 z_1 + \lambda_2 - \lambda_2 z_2} P_0(z_2);$$

où $h(z_2)$ est la fonction génératrice du nombre de clients de type 2 arrivant dans le système pendant la période d'activité formée uniquement par des clients de type 1, Encore elle est la solution unique de l'équation $z_1 - \tilde{B}_1(s + \lambda_1 - \lambda_1 z_1) = 0$ où $s = \lambda_2 - \lambda_2 z_2$.

Le nombre moyen de clients dans la file d'attente prioritaire et en orbite sont donnés par :

$$\begin{aligned} E[N_1] &= \frac{\lambda_1 (\lambda_1 \beta_{1,2} + \lambda_2 \beta_{2,2})}{2(1 - \rho_1)}; \\ E[N_2] &= \frac{\lambda_2 (\lambda_1 \beta_{1,2} + \lambda_2 \beta_{2,2})}{2(1 - \rho_1)(1 - \rho_1 - \rho_2)} + \frac{\lambda_2 (\rho_1 + \rho_2)}{\theta(1 - \rho_1 - \rho_2)}. \end{aligned}$$

Probabilité que le serveur est occupé par le service d'un client type 1 (d'un client type 2)

$$P_1 = \rho_1; \quad P_2 = \rho_2.$$

Cas Particuliers

- Si $\lambda_1 = 0$, on a un système de files d'attente $M/G/1$ avec rappels à un seul type d'appels.
- Si $\lambda_2 = 0$, on a un système de files d'attente $M/G/1$ ordinaire.

Encore, pour les clients type 2, la transformée de *Laplace-Stieltjes* de la distribution du temps d'attente est (Choi, Han et Falin, [24]) :

$$\begin{aligned} E[e^{-sW_2}] &= 1 - \rho + \frac{1 - \rho}{s} \int_1^{\Phi_2(s,1)} \left(\frac{\lambda(1-u)[\Phi_1(0, u, 1) - \Phi_1(s, u, 1)]}{[\Phi_1(0, u, 1) - u][\Phi_1(s, u, 1) - u]} \right) \\ &\quad \times \exp \left\{ \int_u^1 \frac{(\theta + s)[\Phi_1(0, x, 1) - x] + \lambda(1-x)[\Phi_1(0, x, 1) - \Phi_1(s, x, 1)]}{\theta[\Phi_1(0, x, 1) - x][\Phi_1(s, x, 1) - x]} du \right\}; \end{aligned}$$

où $\Phi_1(s, u, 1)$ (respectivement, $\Phi_2(s, 1)$) est la solution unique de l'équation

$$\Phi_1(s, u, 1) = \tilde{B}(s + \lambda_1(1 - \Phi_1(s, u, 1)) + \lambda_2(1 - u));$$

(respectivement, $\Phi_2(s, 1) = \Phi_1(s, \Phi_2(s, 1), 1)$) et satisfait l'inégalité

$$0 \leq \Phi_1(s, u, 1) \leq 1;$$

(respectivement, $0 \leq \Phi_2(s, 1) \leq 1$).

– Si $\lambda_1 = 0$, ce résultat coïncide avec le résultat obtenu par Falin et Fricker [38] pour le modèle $M/G/1$ avec rappels.

1.3.2 Modèle de file d'attente $M_1; M_2/G/1$ avec rappels et perte géométrique

Dans la pratique, un client secondaire au bout de quelques tentatives de rappels infructueuses ne peut plus répéter son appel et décider d'abandonner le système. Soit H_j la probabilité qu'un client fera la $(j + 1)^{i\grave{e}me}$ tentative après les $j^{i\grave{e}me}$ échecs. En particulier, nous considérons le système de tirage au sort géométrique telles que $p \equiv H_0$ et $q \equiv H_i (i \geq 1)$. Le cas $q < 1$ et $q = 1$ donnent des solutions fondamentalement différentes du problème.

Choi et Chang [29] ont étudié le modèle de type $M_1; M_2/M/c$ avec rappels, où le temps de service a une distribution exponentielle et l'orbite de capacité finie K , et ont obtenu les résultats suivants : le nombre moyen de client dans le système, temps moyen d'attente, la période moyenne d'occupation et le moyen nombre de rappels par clients.

Considerons une file d'attente $M_1; M_2/G/1$ avec rappels et perte géométrique, on va traiter le cas où $0 \leq p \leq 1, q = 1$ et le cas $p = q, 0 \leq p < 1$ séparément.

Premier cas $0 \leq p \leq 1, q = 1$.

L'analyse de ce modèle est presque identique à celle du modèle principal de B.D. Choi et K.K Park [26] .

Sous la condition $\rho_1 + p\rho_2 < 1$, ce système est stable et la distribution stationnaire de $(N_1(t), N_2(t), \zeta(t))$ possède les fonctions génératrices partielles suivantes :

$$Q(z_2) = \frac{1 - \rho_1 - p\rho_2}{1 + \rho_2(1 - p)} \exp \left[-\frac{\lambda_1 + \lambda_2}{\theta} \int_{z_2}^1 \frac{1 - h(u)}{h(u) - u} du \right];$$

$$P(0, z_1, z_2) = \frac{1 - \rho_1 - p\rho_2}{1 + \rho_2(1 - p)} \frac{\lambda - \lambda \tilde{B}(\lambda_1(1 - z_1) + p\lambda_2(1 - z_2))}{\lambda_1(1 - z_1) + p\lambda_2(1 - z_2)} \times$$

$$\frac{z_1 - h(z_2)}{z_1 - \tilde{B}(\lambda_1(1 - z_1) + p\lambda_2(1 - z_2))} \frac{1 - z_2}{h(z_2) - z_2}$$

$$\times \exp \left[-\frac{\lambda_1 + \lambda_2}{\theta} \int_{z_2}^1 \frac{1 - h(u)}{h(u) - u} du \right];$$

où $h(z_2)$ est la solution unique de l'équation $z_1 - \tilde{B}(\lambda_1(1 - z_1) + p\lambda_2(1 - z_2)) = 0$.

Le nombre moyen de clients dans la file prioritaire et en orbite sont donnés par :

$$E[N_1] = \frac{\lambda_1(\lambda_1\beta_{1,2} + \lambda_2\beta_{2,2})}{2(1 - \rho_1)(1 + \rho_2(1 - p))};$$

$$E[N_2] = \frac{p\lambda_2\lambda}{(1 - \rho_1 - p\rho_2)} \left(\frac{(\beta_{1,1} + \beta_{2,1})}{\theta} + \frac{(\beta_{1,2} + \beta_{2,2})}{2(1 - \rho_1)(1 + \rho_2(1 - p))} \right).$$

La probabilité que le serveur soit occupé par le service d'un client type 1 (d'un client type 2)

$$P_0 = \frac{1 - \rho_1 - p\rho_2}{1 + \rho_2(1 - p)}, \quad P_1 = \frac{\rho_1 + \rho_2}{1 + \rho_2(1 - p)}.$$

Remarque 1.3.1. – Si $\lambda_1 = 0$, ce modèle devient la file d'attente $M/G/1$ avec rappels où un client de type 2 bloqué fait des tentatives répétées avec la probabilité $(1 - p)$ et ne quitte pas le système jusqu'à ce que le serveur devient libre.

– Si $p = 1$, ce modèle devient la file d'attente $M/G/1$ avec rappels et Bernoulli Schedule (Choi et Park, [26]).

Deuxième cas $p = q, 0 \leq p < 1$

Choi et Park [26] ont étudié le modèle $M_1; M_2/G/1$ avec rappels et perte géométrique. Les auteurs ont utilisé la méthode de la variable supplémentaire où la variable supplémentaire était le temps de service écoulé \tilde{S}_k d'un client en service. Sous la condition d'ergodicité $\rho_1 + p\rho_2 < 1$, les mesures de performance suivantes ont été obtenues :

– La probabilité que le serveur est occupé

$$\rho^* = P_1(1, 1, 0)\beta_{1,1} + P_2(1, 1, 0)\beta_{2,1};$$

– Le nombre moyen de clients en orbite quand le serveur est libre

$$E[N_2, \xi = 0] = \frac{[\rho^*(1 - \lambda_2\beta_{2,1}) - (\lambda_1\beta_{1,1} + \lambda_2\beta_{2,1})]}{\theta\beta_{2,1}};$$

– Le nombre moyen de client dans la file prioritaire quand le serveur est occupé avec un client prioritaire.

$$E[N_1, \xi = 1] = \frac{\lambda_1^2 [\beta_{2,1}\beta_{1,2} + (\rho^* - \lambda_1\beta_{1,1})\beta_{1,1}\beta_{2,2}]}{2\beta_{2,2}(1 - \lambda_1\beta_{1,1})};$$

– Le nombre moyen de clients dans la file prioritaire quand le serveur est occupé avec un client de l'orbite.

$$E[N_2, \xi = 2] = \frac{(\rho^* - \lambda_1\beta_{1,1})\beta_{1,1}\beta_{2,2}}{2\beta_{2,2}}.$$

Remarque 1.3.2. – Si $\lambda_1 = 0$, ce modèle devient le modèle de Yang, Posner et Template [74].

– Si $q = 1$ et $B_1(\cdot) = B_2(\cdot)$, ce modèle devient le modèle de Choi et Park [26].

1.3.3 Modèle de file d'attente $M_1; M_2/G/1/K + 1$ avec rappels et file prioritaire à capacité finie

Choi et Lee [28] ont analysé une file d'attente $M_1; M_2/G/1/K + 1$ avec rappels et une file d'attente prioritaire d'une capacité finie K . En général, l'analyse de la file d'attente avec une capacité finie est plus complexe qu'un modèle correspondant à une capacité infinie. Falin [37] a étudié le cas particulier ($K = 0$) et a obtenu la condition de stabilité et la fonction génératrice du nombre de clients dans la file d'attente. Lorsque $K \geq 1$, Choi et Lee [28] ont obtenu la fonction génératrice de la distribution stationnaire du nombre de clients dans la file d'attente en utilisant la méthode de la variable supplémentaire où la variable supplémentaire représente le temps de service restant du client en service (temps résiduels).

1.3.4 Modèle de file d'attente $M_1; M_2/G/1$ avec rappels

Falin, Artalejo et Martin [39] ont étendu le modèle de Choi et Park [26] au modèle où les deux types d'appels ont des temps de service différents. Ils ont utilisé la méthode de la variable supplémentaire où la variable supplémentaire était le temps de service écoulé. Ils ont obtenu les mesures de performance suivantes pour ce système :

- Probabilité que le serveur est occupé par un client de type 1 (respectivement type 2) :

$$\rho_1 \text{ (respectivement, } \rho_2\text{)}.$$

- Le nombre moyen de clients de type 1 dans la file d'attente prioritaire :

$$E[N_1] = \frac{\lambda_1 ((\lambda_1\beta_{1,2} + \lambda_2\beta_{2,2}))}{2(1 - \rho_1)};$$

- Le nombre moyen de clients de type 2 en orbite :

$$E[N_2] = \frac{\lambda_2 ((\lambda_1\beta_{1,2} + \lambda_2\beta_{2,2}))}{2(1 - \rho_1)(1 - \rho_2)} + \frac{\lambda_2\rho_2}{\theta(1 - \rho_2)}.$$

1.3.5 Modèle de file d'attente $M_1; M_2/G_1; G_2/1$ avec rappels et feedback

Choi et Lee [28] ont étudié le système de file d'attente avec rappels et feedback à deux types de clients, où après que le client de type 1(respectivement type 2) soit servi, rejoint la file prioritaire (respectivement l'orbite) ou quitte le système définitivement.

Le phénomène de feedback dans le système de files d'attente avec rappels peut être rencontré dans beaucoup de situations pratiques, par exemple dans les systèmes de télécommunication, où un message d'erreur s'est avéré à destination et est envoyé à nouveau. Dans le système de file d'attente avec rappels et feedback le client de type 1 servi quitte le système avec une probabilité $1 - \delta_1$ ou retourne à la file d'attente prioritaire pour être servi à nouveau avec

une probabilité δ_1 . Le client de type 2 qui a servi quitte le système avec probabilité $1 - \delta_2$ ou rejoint l'orbite avec probabilité δ_2 .

La fonction génératrice de la distribution stationnaire de la taille de la file d'attente est obtenue en utilisant la méthode de la variable supplémentaire.

1.3.6 Modèle de file d'attente $M^{(X_1, X_2)}/G_1; G_2/1$ avec rappels et priorité relative et vacance unique

Langaris et Moutzoukis [52] ont analysé la file d'attente $M^{(X_1, X_2)}/G_1; G_2/1$ avec rappels et priorité relative et vacances uniques du serveur. Le serveur prend une seule vacance durant que la file prioritaire est vide. Les clients arrivent par groupe dans le système selon un processus de Poisson avec un taux λ . Chaque groupe arrivé peut contenir des clients type 1 et des clients type 2. Les clients type 1 ont toujours une priorité relative pour être servis par rapport aux clients type 2. A l'arrivée d'un groupe si le serveur est libre, alors un des clients type 1 commence son service s'il n'y a aucun client dans la file prioritaire et le reste des clients type 1 rejoint la file prioritaire et attendent le service suivant la discipline FIFO. Le reste du groupe (le cas des clients de type 2) rejoint l'orbite. Si dans le groupe arrivé il n'est y a aucun client type 1, alors un des clients type 2 du groupe arrivé commence son service et le reste des clients type 2 du groupe entre en orbite. Si à l'arrivée d'un groupe, le serveur est en vacances ou occupé par client type 1, alors les clients type 1 et les clients type 2 du groupe rejoignent la file prioritaire et l'orbite, respectivement. Si à l'arrivée d'un groupe contenant au moins un client type 1, le serveur est occupé avec un client type 2, alors le service du client type 2 dans service est interrompu et le serveur commence le service d'un client type 1 du groupe. Quand la file prioritaire devient libre et avant que le serveur part en vacance, le dernier client type 2 interrompu reprend son service à partir du point interrompusion.

Langaris et Moutzoukis [52] ont obtenus la fonction génératrice conjointe de la taille de file d'attente dans l'état transitoire et dans l'état stationnaire. Le temps d'attente virtuel d'un client en orbite a été analysé de la même manière que Falin et Fricker [38] et Choi, Han et Falin [24].

Chapitre 2

Comment simuler un système de files d'attente ?

2.1 Introduction à la simulation

Dans la simulation des systèmes stochastiques, le paradigme le plus répandu est l'approche par événements discrets. En réalité, il ne s'agit de rien d'autre que de la simulation d'un processus markovien à temps continu, dont l'espace d'états et la loi des transitions pourraient être donnés de manière explicite. Du côté informatique, il est intéressant de relever que l'approche de simulation par événements discrets a donné lieu au développement de nombreux langages de programmation spécialisés, dont notamment : GPSS, SIMSCRIPT, SIMULA (première génération) ; et que ce dernier a inspiré la programmation orientée objets, devenue standard de nos jours. Parmi les plus modernes, on trouve OP-NET, qui contient les fonctions graphiques ainsi que l'animation, et également QNAP, qui est un langage spécifique à l'application (réseaux de files d'attente).

Dans un modèle de simulation par événements discrets, le système considéré se trouve à tout instant dans un certain état. Celui-ci ne peut changer que lorsqu'il se produit un événement. En dehors de provoquer un changement d'état du système, un événement peut également provoquer le déclenchement d'autres événements qui vont se produire dans le futur, ou encore modifier le cours des événements qui avaient déjà été prévus auparavant. Les événements ne se produisent qu'à des instants discrets, d'où le nom de l'approche. Dans la modélisation, où un processus composé de certaines activités, telles que par exemple les opérations successives dans une chaîne de montage, on ne conserve de ces activités que les événements qui marquent leur début et leur fin, ainsi que leur effet final. Par contre, tant qu'elle est en cours, l'activité en question n'a aucun effet visible sur le système, autre que d'en bloquer des ressources, le cas échéant, l'état du système n'est donc pas modifié par cette activité pendant qu'elle est en train de se dérouler.

2.2 Principaux éléments d'un simulateur par événements discrets

- Un descripteur des éléments de l'état du système.
- Des compteurs et des places où seront stockés des résultats.
- Une liste chronologique pour placer les événements futurs, avec les algorithmes de mise à jour permettant de gérer l'insertion d'un événement à la bonne place et la recherche du prochain événement et son exclusion de la liste.
- Les différents types d'événements, chacun avec sa description (celle-ci comporte l'action de l'événement sur l'état du système, le déclenchement conditionnel et la mise dans la liste chronologique d'autres événements futurs, ainsi que la mise à jour des différents compteurs et statistiques).
- Une horloge indiquant l'instant courant de la simulation.

2.3 Déroulement de la simulation

Début : Au départ, le système se trouve dans son état initial. Un premier événement DEBUT, placé d'office dans la liste chronologique au temps $t=0$, initialise le processus, inscrit d'autres événements dans la liste chronologique et prévoit, en particulier, l'événement FIN marquant la fin de la simulation.

Boucle : Par la suite, on traite les événements successifs dans l'ordre chronologique de leur apparition dans la liste chronologique : on déplace donc le temps (horloge) jusqu'à l'instant du premier événement agendé, on enlève cet événement de la liste chronologique et on l'exécute, c'est à dire on effectue le changement d'état requis, on inscrit les événements subséquents aux bons endroits de la liste, on met à jour les compteurs et les statistiques, et enfin, on remet l'heure à l'échéance du prochain événement.

Fin : On procède de la sorte jusqu'à ce qu'arrive le temps (l'horloge indique l'heure) de l'événement spécial FIN qui marque l'arrêt de la simulation. Ce dernier peut avoir été déclenché par d'autres événements, ou bien être tout simplement l'événement agendé au départ pour indiquer que la durée prévue de la simulation s'est écoulée. Cet événement déclenchera l'exécution des algorithmes de présentation des résultats.

2.4 Simulation à l'aide des logiciels mathématiques

Lorsque les systèmes sont modélisés comme des processus stochastiques ou des systèmes de files d'attente, ils deviennent complexes et dynamiques, les solutions analytiques ou numériques peuvent devenir intraitables. Dans de tels cas, un programme informatique qui imite le comportement du système (ou du moins les comportements d'intérêt) peut être utilisé. Le programme informatique (ou simulation) est exécuté avec plusieurs valeurs aléatoires et les comportements ainsi modélisés sont enregistrés pour analyse.

La clé d'une bonne simulation est la qualité des générateurs de nombres aléatoires qui y sont utilisés. Les nombres aléatoires constituent la base même du traitement informatique du hasard : toute simulation de phénomènes aléatoires quelque soit leur nature fait par le biais de ce type de nombres. Il est important de disposer de méthodes fiables et efficaces pour les générer. On s'intéresse donc à la source capable de produire une suite de nombres aléatoires $\{u_i\}$ indépendants et uniformément distribués sur l'intervalle $[0, 1]$. On peut à partir d'une telle suite générer toute autre suite de variables aléatoires de loi arbitraire. Par conséquent la loi de densité uniforme $U [0, 1]$ est la base des générateurs de nombres aléatoires.

En pratique, on utilise des procédés physiques ou algorithmiques permettant de produire les nombres aléatoires. Ces derniers sont pseudo aléatoires, c'est à dire ils possèdent des propriétés statistiques identiques à celles des nombres réellement aléatoires.

Un générateur algorithmique doit posséder une bonne forme mathématique, simple et facilement programmable. L'algorithme doit être récursif,

$$u_{i+1} = f(u_i, u_{i-1}, \dots, u_{i-k}) \text{ avec } k < i \text{ et } i \geq 1,$$

ou

$$u_{i+1} = f(u_i) \text{ avec } i \geq 1.$$

Ici f est une fonction donnée à valeurs dans $[0, 1]$. Un bon générateur de nombres aléatoires devrait avoir une longue période avant que les nombres aléatoires ne se recyclent. La corrélation entre les nombres successifs d'une séquence devrait également être faible. La méthode linéaire congruentielle (LC) est une technique largement utilisée pour générer des nombres aléatoires. Dans cette méthode, le nombre aléatoire suivant r_i est généré en utilisant le nombre aléatoire actuel r_{i-1} en utilisant l'équation suivante :

$$r_i = (a \times r_{i-1} + b) \text{ modulo } m,$$

où a et b sont des constantes non négatives. Pour produire m nombres différents, les conditions suivantes doivent être vérifiées :

- Les réels m et b sont premiers entre eux.
- Tout diviseur premier de m divise $(a - 1)$.
- Les nombres $\{u_i\}$ uniformes sur $[0, 1]$ seront obtenus par normalisation $u_i = \frac{r_i}{m}$, $i \geq 0$.

Pour démarrer le générateur, on donne la racine r_0 . En général m est une puissance de 2 ou 100.

Pour augmenter la gamme des nombres générés et réduire la corrélation entre les nombres successifs, plusieurs variantes de la méthode LC ont été proposées.

Ceux-ci incluent la LC multiplicative (où $b = 0$) et la LC additive (où $r_i = (r_{i-1} + r_{i-k}) \text{ modulo } m$). En raison de l'intérêt croissant pour la sécurité informatique utilisant la cryptographie, qui nécessite la génération de clés aléatoires, il existe plusieurs nouvelles techniques pour générer de longues séquences de nombres aléatoires. Pour la plupart des simulations, on peut utiliser un générateur de nombres aléatoires qui a été testé pour sa qualité (par exemple, ceux fournis par (MATLAB)).

En utilisant un générateur de nombres aléatoires qui représente une distribution de probabilité uniforme sur l'intervalle $[0, 1]$, d'autres distributions de probabilité peuvent être générées. Par exemple, la fonction suivante génère les résultats d'une distribution de Poisson avec un taux d'arrivée λ , et un intervalle de temps T fixé.

Algorithme 2.4.1. *int poisson (float lambda, T)*

```

{
float r, temp;
int n;
n = 0;
temp = -1 / (lambda * ln(random_number(seed)));
while ( temp < T)
{ n = n+1;
temp = temp - 1 / (lambda * ln(random_number(seed)));
}
return n;
}

```

La précision d'une simulation dépend également d'une compréhension claire du système modélisé, y compris des interactions entre les différents sous-systèmes ainsi que de la qualité du logiciel développé. Puisque les comportements complexes mènent à des modèles complexes et à des programmes complexes, il est difficile de les valider correctement. Une bonne simulation devrait permettre une variance dans les données (ou les paramètres de simulation) pour étudier les systèmes modélisés dans différentes conditions.

Puisque les simulations de systèmes stochastiques utilisent des nombres aléatoires, elles sont appelées simulations de Monte Carlo. Typiquement, les simulateurs d'ordinateur simulent seulement des événements spécifiques à des temps discrets et sont donc également connus sous le nom de simulateurs d'événements discrets. Un événement peut être considéré comme un moment où le modèle change d'état. Des exemples d'événements incluent : l'arrivée d'un nouveau client (ou d'un nouveau travail), le début d'un service ou la fin d'un service. Les variables d'état définies par le programme sont utilisées pour suivre l'état du système. Les exemples de variables d'état incluent le nombre de tâches en attente sur chaque serveur (ou dans chaque file d'attente lorsque plusieurs classes de travaux sont modélisées). D'autres variables sont utilisées pour définir les paramètres du système, notamment les taux d'arrivée, les débits de service et les tailles maximales des files d'attente. Le programme va simuler les événements en changeant les valeurs des variables du système et en changeant l'heure (ou l'horloge simulée) à l'heure des événements.

Ce qui suit décrit une structure générique de simulateurs typiques.

Algorithme 2.4.2. *Initialize ;*

```

//Initialize termination conditions
//Initialize system state variables, clocks
//Schedule an initial event

```

```

while (termination is false)
{ set_clock; //move clock to next event time
simulate_next_event; //execute procedures to simulate the event
//remove the simulated event
update_statistics;
}
Analyse_results; //produce statistical reports

```

Pour développer un simulateur de système de file d'attente (par exemple, $M/M/1$), nous pouvons sélectionner l'une des deux variantes possibles. Nous pouvons créer tous les événements d'arrivée de traitements au tout début de la simulation. Nous utilisons un générateur de nombres aléatoires pour générer l'heure d'arrivée de chaque traitement (en ajoutant l'heure d'arrivée au moment où le dernier travail est arrivé). Alternativement, nous pouvons générer un traitement à la fois. Dans ce cas, nous générons aléatoirement un nouvel événement qui peut être une arrivée ou un service. Nous recommandons le premier choix parce qu'il sera plus facile de contrôler la simulation, et cette approche permet également la reproduction de la population de telle sorte que différentes disciplines de file d'attente (telles que l'ordonancement prioritaire (priority scheduling), l'anticipation des délais (earliest-deadline-first) (EDF), l'emploi le plus court en premier (shortest-job-first) (SJF)) peut être appliqué à la même population.

Il est également nécessaire de décider d'un test de terminaison basé soit sur un nombre total de tâches traitées par la simulation, soit sur une période maximale sur laquelle le système est simulé. Dans le premier cas, tous les travaux entrant dans le système seront traités, alors que dans le second cas, tous les travaux entrants ne peuvent pas être traités au moment où la simulation est terminée.

Il est nécessaire de décider des informations à associer à chaque travail. Dans un système $M/M/1$ simple utilisant la discipline du premier entré, premier sorti (FCFS), il est seulement nécessaire de conserver l'heure d'arrivée, l'heure à laquelle un service est initié et le temps de service pour chaque traitement. À partir de ces informations, il est possible de calculer les temps d'attente et les temps de réponse pour chaque travail, ainsi que les temps d'attente moyens et les temps de réponse pour le système. Pour les systèmes en temps réel, il est nécessaire de respecter les délais d'exécution d'un travail. La date limite peut être basée sur les heures de service ou créée de manière aléatoire.

Les changements dans le traitement des listes de tâches en attente peuvent simuler des variations dans les disciplines de file d'attente FCFS. Pour mettre en œuvre la planification EDF, il est nécessaire de trier la liste d'attente des travaux selon leurs échéances. Pour implémenter la planification SJF, la liste est triée en fonction des heures de service des tâches en attente. Les files d'attente de priorité peuvent être simulées en maintenant des listes séparées pour chaque priorité.

Pour simuler $M/M/1$, le temps de simulation est défini sur l'heure d'arrivée du travail suivant dans la liste d'attente. Si le serveur est occupé, l'heure de simulation est définie sur l'heure d'achèvement du service en cours de traitement (qui est égale à l'heure de démarrage du service plus le temps de service). À ce moment, le prochain travail d'attente est programmé

pour le service. Ce processus est répété jusqu'à ce que la condition de terminaison soit remplie. Des langages de programmation et des bibliothèques de logiciels sont disponibles pour simplifier la conception des programmes de simulation. Ils fournissent des générateurs de nombres aléatoires prêts à l'emploi, des fonctions pour générer diverses distributions de probabilité, des structures de données pour mettre en file d'attente des événements, gérer le temps, enregistrer les résultats et produire des analyses statistiques communes. L'une des premières langues est SIMULA, datant des années 1960. Des versions plus récentes de SIMULA basées sur C++ et JAVA ont été développées dans diverses universités, souvent en freeware. Un autre exemple est le langage de balisage de provisionnement de services (SMPL), développé par McDougall au MIT, qui contient un ensemble de fonctions de langage C pouvant être utilisées pour simuler des systèmes de files d'attente. D'autres langages et outils commerciaux sont disponibles à l'achat. Dans ce chapitre, nous allons nous concentrer sur le développement de systèmes de simulation utilisant MATLAB.

2.4.1 Utilisation de MATLAB

Le programme MATLAB ci-dessous permet d'effectuer une simulation d'événement discret d'une file d'attente $M/M/1$ avec un taux d'arrivée $\lambda = 0,5$ et un taux de service $\mu = 1$. La variable prochaine arrivée (**nextarrival**) donne les heures d'arrivée du prochain client. De même, le prochain départ (**nextdeparture**) indique l'heure de départ du client en cours de traitement (défini sur l'infini si la file d'attente est actuellement vide). L'instruction clé est si **nextarrival** < **nextdeparture**, qui détermine si l'événement suivant doit se produire sera une arrivée ou une sortie. Pour une arrivée, on déplace la variable maintenant vers l'heure d'arrivée, on augmente la longueur de la file d'attente (**currentlength**) de 1, on annonce l'arrivée avec une instruction **disp** et on planifie l'arrivée suivante (après celle-ci) en réinitialisant **nextarrival**. Rappelons que $(-1/\lambda) * \log(\mathbf{rand})$ génère un temps d'inter-arrivée exponentiel de taux λ . Si le client nouvellement arrivé est le seul présent (c'est-à-dire si **currentlength** == 1), le client peut aller directement en service, ainsi nous décidons également combien de temps le service prendra en générant un temps de service aléatoire $(-1/\mu) * \log(\mathbf{rand})$ avec la distribution Exponentielle de taux μ , et fixer le prochain départ en conséquence.

Pour gérer un départ, nous diminuons la longueur de la file d'attente actuelle de 1 et annonçons le départ avec une autre instruction **disp**. Cela laisse la file d'attente vide, auquel cas **nextdeparture** doit être infini, ou prend un autre client en service, auquel cas **nextdeparture** doit être défini en générant un temps de service pour ce client.

Le traitement complet est inclu dans une boucle **while** qui maintient la simulation jusqu'à **targettime**, qui est le moment où la simulation doit se terminer.

```

Algorithme 2.4.3. lambda=0.5;
mu=1.0;
targettime=50;
nextarrivale= (-1/lambda)*log(rand);
now=0;

```

```

nextdeparture= inf; % infinity
currentlength= 0;
while now < targettime,
if nextarrival < nextdeparture,
now= nextarrival;
currentlength= currentlength +1;
disp (sprintf('Arrival at : %f (current length %d)', now,currentlength));
nextarrival= now +(-1/lambda)*log(rand);
if currentlength==1,
nextdeparture= now + (-1/mu)*log(rand);
end
else
now=nextdeparture;
currentlength= currentlength - 1;
disp(sprintf('Departure at : %f (currentlength %d)', now,currentlength));
if currentlength > 0,
nextdeparture=now+(-1/mu)*log(rand);
else
nextdeparture= inf;
end
end
end

```

Lorsque le programme est exécuté, la sortie peut être comme suit :

Arrivée à	0.102314	durée actuelle 1
Départ à	0.601800	durée actuelle 0
Arrivée à	3.031791	la longueur actuelle 1
Départ à	3.146866	durée actuelle 0
Arrivée à	4.474956	la longueur actuelle 1
Arrivée à	5.018319	la longueur actuelle 2
Départ à	5.259194	durée actuelle 1

Chaque fois que le programme passe par la boucle principale, le programme génère une ligne de sortie, correspondant à une arrivée ou à un départ.

Voici un autre exemple d'une simple simulation de file d'attente $M/M/1$ qui représente graphiquement le nombre moyen de clients dans le système, le temps moyen de séjour.

Algorithme 2.4.4. %% Implementation of a simple $M/M/1$

```

queue_lim=200000; %system limit
arrival_mean_time(1 :65)=0.01;
service_mean_time=0.01;
sim_packets=750; %number of clients to be simulated
util(1 :65) = 0;

```

```

avg_num_in_queue(1 :65) = 0;
avg_delay(1 :65) = 0;
P(1 :65) = 1;
for j=1 :64 %loop for increasing the mean arrival time
arrival_mean_time(j+1)=arrival_mean_time(j) + 0.001;
num_events=2;
% initialization
sim_time = 0.0;
server_status=0;
queue_size=0;
time_last_event=0.0;
num_puck_insys=0;
total_delays=0.0;
time_in_queue=0.0;
time_in_server=0.0;
delay = 0.0;
time_next_event(1)= sim_time + exprnd(arrival_mean_time(j+1));
time_next_events(2)= exp(30);
disp(['Launching Simulation...',num2str(j)]);
while(num_pack\insys < sim_packets)
min_time_next_event = exp(29);
type_of_event=0;
for i=1 :num_events
if(time_next_event(i)<min_time_next_event)
min_time_next_event = time_next_event (i);
type_of_event = i;
end;
end
if(type_of_event==0)
disp(['no event in time ',num2str(sim_time)]);
end
sim_time = min_time_next_event;
time_since_last_event = sim_time-time_last_event;
time_last_event = sim_time;
time_in_queue = time_in_queue + queue_size
* time_since_last_event;
time_in_server = time_in_server + server_status
* time_since_last_event;
if (type_of_event==1)
%disp(['packet arrived']);
%-----arrival-----
time_next_event(1) = sim_time + exprnd(arrival_mean_time(j+1));

```

```

%epomenos xronos afiksis
if(server_status == 1)
num_pack_insys = num_pack_insys + 1;
queue_size = queue_size + 1;
if(queue_size > queue_lim)
disp(['queue size = ', num2str(queue_size)]);
disp(['System Crach at ', num2str(sim_time)]);
pause
end
arr_time(queue_size) = sim_time;
else
server_status = 1;
time_next_events(2) = sim_time + exprnd(service_mean_time);
end
elseif (type_of_event==2)
% -----service and departure-----
if(queue_size == 0)
server_status = 0;
time_next_event(2) = exp(30);
else
queue_size = queue_size - 1;
delay = sim_time - arr_time(1);
total_delays = total_delays + delay;
time_next_event(2) = sim_time + exprnd(service_mean_time);
for i = 1 :queue_size
arr_time(i)=arr_time(i+1);
end
end
end
end
%results output
util(j+1) = time_in_server/sim_time;
avg_num_in_queue(j+1) = time_in_queue/sim_time;
avg_delay(j+1) = total_delays/num_pack_insys;
P(j+1) = service_mean_time./arrival_mean_time(j+1);
end
%-----graphs-----
figure('name','mean numberof clients in system diagram(simulated)');
plot(P,avg_num_in_queue,'r');
Xlabel('P');
Ylabel('mean number of clients');
axis([0 0.92 0 15]);

```

```

figure('name','mean delay in system diagram (simulated)');
plot(p,avg_delay,'m');
Xlabel('P');
Ylabel('mean delay (hrs)');
axis([0 0.92 0 15]);
figure('name','UTILISATION DIAGRAM');
plot(P,util,'b');
Xlabel('P');
Ylabel('UTILISATION');
axis([0 0.92 0 1]);

```

Algorithm 2.4.5. %% *Routines to simulate M/G/1 and M/G/∞*

```

function [jumptimes, systsize, systtime] = simmg1(tmax, lambda)
% SIMMG1 simulate a M/G/1 queueing system. Poisson arrivals
% of intensity lambda, uniform service times.
%
% [jumptimes, systsize, systtime] = simmg1(tmax, lambda)
%
% Inputs : tmax - simulation interval
% lambda - arrival intensity
%
% Outputs : jumptimes - time points of arrivals or departures
% systsize - system size in M/G/1 queue
% systtime - system times
% set default parameter values if ommited
if (nargin==0)
tmax=1500; % simulation interval
lambda=0.99; % arrival intensity
end
arrtime=-log(rand)/lambda; % Poisson arrivals
i=1;
while (min(arrtime(i,:))<=tmax)
arrtime = [arrtime; arrtime(i,:)-log(rand)/lambda];
i=i+1;
end
n=length(arrtime); % arrival times t_1,...,t_n
service=2.*rand(1,n); % service times s_1,...,s_k
cumservertime=cumsum(servtime);
arrsubtr=arrtime-[0 cumservertime(:,1:n-1)]'; % t_k-(k-1)
arrmatrix=arrsubtr*ones(1,n);
deptime=cumservertime+max(triu(arrmatrix)); % departure times
% u_k=k+max(t_1,...,t_k-k+1)
% Output is system size process N and waiting

```

```

% times W.
B=[ones(n,1) arrtime; -ones(n,1) deptime'];
Bsort=sortrows(B,2);
jumps=Bsort(:,1);
jumptimes=[0;Bsort(:,2)];
systsize=[0;cumsum(jumps)]; % size of M/G/1 queue
systime=deptime-arrtime'; % system times
figure(1)
stairs(jumptimes,systsize);
xmax=max(systsize)+5;
axis([0 tmax 0 xmax]);
grid
figure(2)
hist(systtime,20);
function [jumptimes, systsize] = simmginfy(tmax, lambda)
% SIMMGINFY simulate a M/G/infinity queueing system. Arrivals are
% a homogeneous Poisson process of intensity lambda. Service times
% Pareto distributed (can be modified).
%
% [jumptimes, systsize] = simmginfy(tmax, lambda)
%
% Inputs : tmax - simulation interval
%          systsize - number of customers in system
%
% set default parameter values if omitted
if ( nargin==0)
tmax=1500;
lambda=1;
end
% generate Poisson arrivals
% the number of points is Poisson-distributed
npoints = poissrnd(lambda*tmax);
% conditioned that number of points is N,
% the points are uniformly distributed
if (npoints>0)
arrt = sort(rand(npoints, 1)*tmax);
else
arrt = [];
end
% uncomment if not available POISSONRD
% generate Poisson arrivals
% arrt=-log(rand)/lambda;

```

```

% i=1;
% while (min(arrt(i, :)-log(rand)/lambda);
% arrt = [arrt; arrt(i, :)-log(rand)/lambda];
% i=i+1;
% end
% npoints=length(arrt); % arrival times t_1,...,t_n
% servt=50.*rand(n,1); % uniform service times s_1,...,s_k
alpha = 1.5; % Pareto service times
servt = rand.^(-1/(alpha-1))-1; % stationnary renewal process
servt = [servt; rand(npoints-1,1).^(-1/alpha)-1];
servt = 10.*servt; % arbitrary choice of mean
dept = arrt+servt; % departure times
% Output is system size process N.
B = [ones(npoints, 1) arrt; -ones(npoints, 1) dept];
Bsort = sortrows(B, 2); % sort jumps in order
jumps = Bsort(:, 1);
jumptimes = [0; Bsort(:, 2)];
systsize = [0; cumsum(jumps)]; % M/G/infinity system size
% process
stairs(jumptimes, systsize);
xmax = max(systsize)+5;
axis([0 tmax 0 xmax]);
grid
%
```

Chapitre 3

Systeme de files d'attente $M/G/1$ avec rappels, priorité relative et orbite FCFS

3.1 Introduction

La grande partie de recherche sur les systèmes des files d'attente avec rappels est basée sur le fait que les rappels fonctionnent sous une politique de rappels classique, tels que les intervalles de temps entre deux rappels successifs d'une même source secondaire sont supposés exponentiellement distribués de taux total $j\theta$ (j est le nombre de clients dans le groupe de rappels). Il y a une autre discipline, appelée politique de rappels constant, où le taux de rappels total ne dépend pas du nombre de clients en orbite (groupe de rappels) : les clients forment une file d'attente et uniquement le client en tête de la file peut demander le service. La discipline en question est introduite par Fayolle [43]. La file d'attente avec un seul serveur, où le temps inter-rappels ainsi que le temps de service suivent une loi générale, a été étudié par Gomez-Corral [46]. Des contributions récentes sur ce sujet incluent les travaux de Atencia, Fortes et Sánchez [18]; Choudhury et Ke [30]; Moreno [57]; Sumitha et Chandrika [67]; Wu, Wang et Liu [72]. La stabilité d'un système de files d'attente avec rappels sous une distribution générale pour le temps inter-rappel est étudié Kernane [48].

Dans ce chapitre, nous considérons un modèle où les phénomènes de priorité relative, de l'orbite FCFS ainsi que des temps inter-rappels généraux sont présents. Le modèle en question a été étudié dans [17]. Nous présentons les résultats essentiels obtenus ainsi que nos propres développements.

3.2 Description du modèle

On considère un système de file d'attente avec rappels et un seul serveur où les clients primaires arrivent dans le système selon un processus de Poisson de taux $\lambda > 0$. À l'arrivée d'un client si le serveur est libre, il entre immédiatement en service, sinon il décide d'entrer

en orbite avec une probabilité p ou de rejoindre la file d'attente (d'une capacité infinie) avec une probabilité $(1 - p)$. On suppose que seulement le client en tête de l'orbite est autorisé à accéder au serveur. En orbite le client en question répète les tentatives jusqu'à ce que le serveur devient libre et commence son service. Ce type de modèle est appelé système de file d'attente avec rappels et orbite FCFS. Le temps inter-rappel suit une distribution générale de fonction de répartition $A(x)$ et de transformé de Laplace-Stieltejes $\tilde{A}(s) = \int_0^{\infty} e^{-sx} dA(x)$, $\text{Re}(s) > 0$ et soient les moments d'ordre k , $\alpha_k = (-1)^k \tilde{A}^{(k)}(0)$. Il est clair qu'un client de l'orbite ne peut être admis pour service seulement si l'espace d'attente est vide. Ainsi, les clients de la file d'attente (qui sont servis selon une discipline FCFS) possèdent une priorité relative (sans préemption) sur les clients en orbite. À l'achèvement d'un service, le serveur devient libre seulement si la file prioritaire est vide et une compétition entre une distribution Exponentielle de taux λ et une distribution générale des temps inter-rappels de taux $\theta = \frac{1}{\alpha_1}$ détermine le prochain client qui sera servi. Le temps de service suit une loi général de fonction de répartition $B(x)$ et de transformée de Laplace-Stieltejes $\tilde{B}(s) = \int_0^{\infty} e^{-sx} dB(x)$, $\text{Re}(s) > 0$. Soit $\beta_k = (-1)^k \tilde{B}^{(k)}(0)$ les moments d'ordre k . L'intensité du trafic $\rho = \lambda\beta_1$.

$$k_{ij} = \int_0^{\infty} \frac{(\lambda(1-p)x)^i}{i!} \frac{(\lambda px)^j}{j!} e^{-\lambda x} dB(x)$$

est la distribution conjointe du nombre de client arrivant dans la file prioritaire et en orbite étant donné que le serveur est occupé, sa fonction génératrice est

$$K(z_1, z_2) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} k_{ij} z_1^i z_2^j = \tilde{B}(\lambda - \lambda(1-p)z_1 - \lambda pz_2).$$

Finalement, on admet l'hypothèse de l'indépendance mutuelle entre toutes les variables aléatoires définies ci-dessus.

3.3 La chaîne de Markov induite

L'état de système à la date t peut être décrit à l'aide du processus

$$\{C(t), N_f(t), N_o(t), \xi_1(t), \xi_2(t), t \geq 0\}$$

où $N_f(t)$ et $N_o(t)$ sont respectivement les nombres de clients dans la file prioritaire et en orbite, $C(t)$ représente l'état du serveur à l'instant t :

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

Si $C(t) = 0$, $N_f(t) = 0$ et $N_o(t) > 0$, alors $\xi_1(t)$ représente le temps écoulé après le rappel du client en tête de l'orbite (à la date t). Si $C(t) = 1$, $\xi_2(t)$ est le temps de service écoulé du client dans le serveur à la date t .

Considérons la chaîne de Markov induite associée à ce modèle. Soit, ζ_n l'instant de départ du $n^{\text{ième}}$ client, et soient $N_{f,n} = N_f(\zeta_n)$ et $N_{o,n} = N_o(\zeta_n)$ le nombre de clients dans la file d'attente prioritaire et le nombre de clients en orbite respectivement, juste avant le temps ζ_n . Les équations fondamentales récursives sont définies comme suit :

$$N_{f,n} = \begin{cases} N_{f,n-1} - 1 + v_{f,n} & \text{si } N_{f,n-1} > 0 \\ v_{f,n} & \text{si } N_{f,n-1} = 0 \end{cases} \quad (3.3.1)$$

$$N_{o,n} = \begin{cases} N_{o,n-1} + v_{o,n} & \text{si } N_{f,n-1} \geq 1 \\ N_{o,n-1} - \delta_n + v_{o,n} & \text{si } N_{f,n-1} = 0 \text{ et } N_{o,n-1} \geq 1 \\ v_{o,n} & \text{si } N_{f,n-1} = 0 \text{ et } N_{o,n-1} = 0 \end{cases} \quad (3.3.2)$$

où $v_{f,n}$ et $v_{o,n}$ sont les nombres de clients entrant dans la file d'attente prioritaire et en orbite respectivement pendant le $n^{\text{ème}}$ temps de service; $\delta_{f,n}$ est égale à 0 ou 1 selon le $n^{\text{ième}}$ client servi est primaire ou secondaire.

La suite des vecteurs aléatoires $X_n = (N_{f,n}, N_{o,n})$, $n \in \mathbb{N}$ forme une chaîne de Markov, qui est la chaîne de Markov induite de notre système de file d'attente avec l'espace d'états \mathbb{N}^2 . $\{X_n, n \in \mathbb{N}\}$ est irréductible et apériodique.

3.3.1 Condition d'érgodicité

Notre premier objectif sera l'étude de l'érgodicité de la chaîne de Markov. En raison de la structure récursive des équations 3.3.1 et 3.3.2, nous utiliserons le critère de Foster et la condition de Kaplan, qui indique qu'une chaîne de Markov irréductible et apériodique X_n avec l'espace des états \mathbb{N}^2 est ergodique s'il existe une fonction non-négative $F(i, j)$, $(i, j) \in \mathbb{N}^2$, appelée fonction teste, et $\epsilon > 0$ tel que l'accroissement moyen

$$x_{i,j} = E [F(X_n) - F(X_{n-1}) / X_{n-1} = (i, j)]$$

soit fini pour tout $(i, j) \in \mathbb{N}^2$ et $x_{(i,j)} \leq -\epsilon$ pour tout $(i, j) \in \mathbb{N}^2$ sauf peut-être un nombre fini. Dans notre cas, on considère la fonction test linéaire (Fonction de Lyapunov) suivante

$$F(i, j) = \gamma i + j$$

où γ est un paramètre positif qui sera déterminé plus tard. Alors l'accroissement moyen $x_{i,j}$ est donné par

$$x_{i,j} = \begin{cases} \rho q \gamma + \rho p & \text{si } i = 0 \text{ et } j = 0, \\ \rho q \gamma - \tilde{A}(\lambda) + \rho p & \text{si } i = 0 \text{ et } j \geq 1, \\ \rho p - (1 - \rho q) \gamma & \text{si } i \geq 1 \text{ et } j \geq 0. \end{cases}$$

La chaîne de Markov $\{N_{f,n}, N_{o,n}, n \geq 1\}$ est ergodique si le paramètre γ satisfait les inégalités suivantes

$$\rho q \gamma - \tilde{A}(\lambda) + \rho p < 0,$$

et

$$\rho p - (1 - \rho q) \gamma < 0.$$

Cette condition signifie que γ appartient à l'intervalle $\left(\frac{\rho p}{1 - \rho q}, \frac{\tilde{A}(\lambda) - \rho p}{\rho q} \right)$. Tel que γ peut être déterminé si cet intervalle n'est pas vide, c'est à dire

$$\rho < \frac{\tilde{A}(\lambda)}{p + (1 - p)\tilde{A}(\lambda)}. \quad (3.3.3)$$

Ainsi $\rho < \frac{\tilde{A}(\lambda)}{p + (1 - p)\tilde{A}(\lambda)}$ est une condition suffisante pour l'ergodicité de $\{N_{f,n}, N_{o,n}, n \geq 1\}$,

Comme on verra dans ce qui suit, que cette condition est aussi nécessaire pour l'ergodicité.

3.3.2 Distribution stationnaire de la chaîne de Markov induite

Notre second objectif est de trouver la distribution stationnaire de la chaîne de Markov induite X_n

$$\pi_{i,j} = \lim_{n \rightarrow \infty} P[N_{f,n} = i, N_{o,n} = j].$$

En première étape les probabilités de transition sont données par les formules

$$p_{(0,0)(i,j)} = k_{i,j};$$

$$p_{(0,j)(i,j)} = \left[1 - \tilde{A}(\lambda) \right] k_{i,j-m} + \tilde{A}(\lambda) k_{i,j-m+1}; \quad m = 1, \dots, j;$$

$$p_{(0,j+1)(i,j)} = \tilde{A}(\lambda) k_{i,0};$$

$$p_{(h,m)(i,j)} = k_{i-h+1,j-m}; \quad h = 1, \dots, i+1; \quad m = 0, \dots, j.$$

D'où, l'équation de Kolmogorov pour la distribution $\pi_{i,j}$ est

$$\begin{aligned} \pi_{i,j} &= \pi_{0,0} k_{i,j} + (1 - \delta_{0j}) \left[1 - \tilde{A}(\lambda) \right] \sum_{m=1}^j \pi_{0,m} k_{i,j-m} \\ &\quad + \tilde{A}(\lambda) \sum_{m=1}^{j+1} \pi_{0,m} k_{i,j-m+1} \\ &\quad + \sum_{h=1}^{i+1} \sum_{m=0}^j \pi_{h,m} k_{i-h+1,j-m}; \end{aligned}$$

pour $i \geq 0, j \geq 0$.

Afin de résoudre les équations ci-dessus, nous introduisons la fonction génératrice de la distribution stationnaire de la chaîne de Markov induite

$$\varphi(z_1, z_2) = \sum_{i=1}^{\infty} \sum_{j=0}^{\infty} \pi_{i,j} z_1^i z_2^j$$

et la fonction génératrice auxiliaire

$$\psi(z_2) = \sum_{j=0}^{\infty} \pi_{0,j} z_2^j.$$

Prenant en compte les fonctions génératrices précédentes, l'équation de Kolmogorov devient :

$$\frac{z_1 - K(z_1, z_2)}{K(z_1, z_2)} \varphi(z_1, z_2) = \frac{\tilde{A}(\lambda) z_1 (1 - z_2) - z_2 (1 - z_1)}{z_2} \psi(z_2) - (1 - z_1) \pi_{0,0} \quad (3.3.4)$$

Maintenant on considère la fonction

$$f(z_1, z_2) = z_1 - K(z_1, z_2) = z_1 - \tilde{B}(\lambda - \lambda(1 - p)z_1 - \lambda p z_2)$$

pour chaque z_2 fixé avec $|z_2| < 1$, on considère $f(z_1, z_2)$ comme une fonction de z_1 . Si $|z_1| = 1$ on a que $\operatorname{Re}(\lambda - \lambda(1 - p)z_1 - \lambda p z_2) > 0$. On sait que $|\tilde{B}(s)| < 1$ si $\operatorname{Re}(s) > 0$.

Ainsi nous obtenons

$$|[z_1 - K(z_1, z_2)] - z_1| = |K(z_1, z_2)| < 1 = |z_1|.$$

Selon le théorème de Rouché, pour chaque z_2 avec $|z_2| < 1$, il y a une solution unique $z_1 = g(z_2)$ de l'équation $f(z_1, z_2) = 0$ dans le disque unité, telle que

$$f(g(z_2), z_2) = g(z_2) - \tilde{B}(\lambda - \lambda(1 - p)g(z_2) - \lambda p z_2) = 0.$$

Il est clair que

$$\begin{cases} g(1) = 1; \\ g'(1) = \frac{\rho p}{1 - \rho q}; \\ g''(1) = \frac{\lambda^2 p^2 \beta_2}{(1 - \rho q)^3}. \end{cases}$$

Posons $z_1 = g(z_2)$ dans l'équation 3.3.4, $\psi(z_2)$ devient

$$\psi(z_2) = \frac{z_2 [1 - g(z_2)]}{\tilde{A}(\lambda) g(z_2) (1 - z_2) - z_2 (1 - g(z_2))} \pi_{0,0}. \quad (3.3.5)$$

En remplaçant $\psi(z_2)$ dans $\varphi(z_1, z_2)$ donne

$$\varphi(z_1, z_2) = \frac{z_1 - g(z_2)}{z_1 - K(z_1, z_2)} \frac{\tilde{A}(\lambda)(1 - z_2)K(z_1, z_2)}{\tilde{A}(\lambda)g(z_2)(1 - z_2) - z_2(1 - g(z_2))} \pi_{0,0}.$$

Utilisant la condition de normalisation $\varphi(1, 1) = 1$, On peut trouver la constante inconnue p_0

$$\pi_{0,0} = p_0 = 1 - \frac{p + (1 - p)\tilde{A}(\lambda)}{\tilde{A}(\lambda)}\rho. \quad (3.3.6)$$

De la formule de 3.3.6 et comme $p_0 > 0$, on obtient

$$\rho < \frac{\tilde{A}(\lambda)}{p + (1 - p)\tilde{A}(\lambda)}$$

qui est une condition nécessaire pour l'érgodicité de la chaîne de Markov.

On résume les résultats précédents dans le théorème suivant :

- La chaîne de Markov induite aux instants de départ est ergodique si seulement si $\rho < \frac{\tilde{A}(\lambda)}{p + (1 - p)\tilde{A}(\lambda)}$.

Théorème 3.3.1. - *La fonction génératrice de la distribution stationnaire de la chaîne de Markov induite est donnée par :*

$$\begin{aligned} \varphi(z_1, z_2) &= \left\{ 1 - \frac{p + (1 - p)\tilde{A}(\lambda)}{\tilde{A}(\lambda)}\rho \right\} \times \frac{z_1 - g(z_2)}{z_1 - K(z_1, z_2)} \\ &\times \frac{\tilde{A}(\lambda)(1 - z_2)K(z_1, z_2)}{\tilde{A}(\lambda)g(z_2)(1 - z_2) - z_2(1 - g(z_2))} \end{aligned}$$

- où $g(z_2)$ est l'unique racine de z_1 de l'équation $z_1 - K(z_1, z_2) = z_1 - \tilde{B}(\lambda - \lambda(1 - p)z_1 - \lambda pz_2) = 0$.

Remarque 3.3.1. Si $\rho < \frac{\tilde{A}(\lambda)}{p + (1 - p)\tilde{A}(\lambda)}$, la chaîne de Markov induite est érgodique pour toutes les valeurs de $p; q \in [0, 1]$.

3.4 Distributions stationnaires de l'état du système

Nous définissons la probabilité limite

$$P_0 = \lim_{t \rightarrow \infty} P(C(t), N_f(t) = 0, N_o(t) = 0);$$

et les densités de probabilités limites

$$P_{0,0,j} = \int_0^{\infty} \lim_{t \rightarrow \infty} \frac{d}{dx} P(C(t) = 0, N_f(t) = 0, N_o(t) = j, \xi_1(t) \leq x); j \geq 1, x \geq 0);$$

$$P_{1,i,j} = \int_0^{\infty} \lim_{t \rightarrow \infty} \frac{d}{dx} P(C(t) = 1, N_f(t) = i, N_o(t) = j, \xi_2(t) \leq x; i \geq 0, j \geq 0, x \geq 0);$$

Par la méthode des variables supplémentaires, on obtient les équations d'équilibre

$$\lambda P_0 = \int_0^{\infty} P_{1,0,j}(x) b(x) dx; \quad (3.4.1)$$

$$\frac{d}{dx} P_{0,0,j}(x) = -(\lambda - a(x)) P_{0,0,j}(x); j \geq 1; \quad (3.4.2)$$

$$\begin{aligned} \frac{d}{dx} P_{1,i,j}(x) &= -(\lambda - b(x)) P_{1,i,j}(x) + (1 - \delta_{0i}) \lambda q P_{1,i-1,j}(x) \\ &+ (1 - \delta_{0j}) \lambda p P_{1,i,j-1}(x), i \geq 0, j \geq 0. \end{aligned} \quad (3.4.3)$$

où les fonctions $a(x)$ et $b(x)$ sont les taux de conditions d'achèvement (à l'instant x) pour les rappels successifs et pour le service respectivement.

$$a(x) = \frac{A'(x)}{1 - A(x)} \text{ et } b(x) = \frac{B'(x)}{1 - B(x)}.$$

Les conditions limites sont

$$P_{0,0,j}(0) = \int_0^{\infty} P_{1,0,j}(x) b(x) dx, j \geq 1; \quad (3.4.4)$$

$$\begin{aligned} P_{1,i,j}(0) &= \delta_{0i} \delta_{0j} \lambda P_0 + \delta_{0i} (1 - \delta_{0j}) \lambda \int_0^{\infty} P_{0,0,j}(x) dx \\ &+ \delta_{0i} \int_0^{\infty} P_{0,0,j+1}(x) a(x) dx \\ &+ \int_0^{\infty} P_{1,i+1,j}(x) b(x) dx, \end{aligned} \quad (3.4.5)$$

pour $i \geq 0, j \geq 0$, et la condition de normalisation est

$$p_0 + \sum_{j=1}^{\infty} \int_0^{\infty} P_{0,0,j}(x) dx + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \int_0^{\infty} P_{1,i,j}(x) dx = 1. \quad (3.4.6)$$

Pour résoudre les équations. 3.4.1 - 3.4.6, nous définissons les fonctions génératrices de probabilités

$$P_0(x, z_2) = \sum_{j=1}^{\infty} P_{0,0,j}(x) z_2^j,$$

$$P_1(x, z_1, z_2) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} P_{1,i,j}(x) z_1^i z_2^j.$$

En utilisant les fonctions génératrices précédentes, les équations. 3.4.2 et 3.4.3 deviennent les équations différentielles partielles suivantes :

$$\frac{\partial}{\partial x} P_0(x, z_2) = -(\lambda - a(x)) P_0(x, z_2),$$

$$\frac{\partial}{\partial x} P_1(x, \lambda, z_2) = -(\lambda - \lambda q z_1 - \lambda p z_2 + b(x)) P_1(x, z_1, z_2)$$

dont les solutions sont données par

$$P_0(x, z_2) = P_0(0, z_2) [1 - A(x)] e^{-\lambda x}, \quad (3.4.7)$$

$$P_1(x, z_1, z_2) = P_1(0, z_1, z_2) [1 - B(x)] e^{-(\lambda - \lambda q z_1 - \lambda p z_2)x}. \quad (3.4.8)$$

En termes de fonctions génératrices ci-dessus, les équations. 3.4.4 et 3.4.5 peuvent être écrites comme suit

$$P_0(0, z_2) = \int_0^{\infty} P_1(x, 0, z_2) b(x) dx - \int_0^{\infty} P_{1,0,0}(x) b(x) dx, \quad (3.4.9)$$

$$P_1(0, z_1, z_2) = \lambda p_0 + \lambda \int_0^{\infty} P_0(x, z_2) dx + \frac{1}{z_2} \int_0^{\infty} P_0(x, z_2) a(x) dx$$

$$+ \frac{1}{z_1} \int_0^{\infty} [P_1(x, z_1, z_2) - P_1(x, 0, z_2)] b(x) dx. \quad (3.4.10)$$

En utilisant 3.4.1 après avoir remplacé 3.4.7 et 3.4.8 dans 3.4.9 et 3.4.10, on obtient

$$P_0(0, z_2) = \tilde{B}(\lambda - \lambda p z_2) P_1(0, 0, z_2) - \lambda p_0, \quad (3.4.11)$$

$$\begin{aligned}
& \left[z_1 - \tilde{B}(\lambda - \lambda(1-p)z_1 - \lambda pz_2) \right] P_1(0, z_1, z_2) \\
&= \frac{z_2 + (1-z_2)\tilde{A}(\lambda)}{z_2} z_1 P_0(0, z_2) \\
& \quad - \tilde{B}(\lambda - \lambda pz_2) P_1(0, 0, z_2) + \lambda z_1 p_0.
\end{aligned} \tag{3.4.12}$$

Posons $z_1 = g(z_2)$ alors, l'équation 3.4.12 devient

$$\tilde{B}(\lambda - \lambda pz_2) P_1(0, 0, z_2) = \frac{z_2 + (1-z_2)\tilde{A}(\lambda)}{z_2} g(z_2) P_0(0, z_2) + \lambda g(z_2) p_0$$

et en remplaçant l'équation ci-dessus dans 3.4.11, nous obtenons

$$P_0(0, z_2) = \frac{\lambda z_2 [1 - g(z_2)]}{\tilde{A}(\lambda) (1 - z_2) g(z_2) - z_2 [1 - g(z_2)]} p_0. \tag{3.4.13}$$

Alors, de 3.4.13, l'équation 3.4.11 donne

$$P_1(0, 0, z_2) = \frac{\lambda g(z_2) \tilde{A}(\lambda)}{\tilde{B}(\lambda - \lambda pz_2) \tilde{A}(\lambda) (1 - z_2) g(z_2) - z_2 [1 - g(z_2)]} (1 - z_2) p_0. \tag{3.4.14}$$

En remplaçant les équations, 3.4.13 et 3.4.14 dans l'équation 3.4.12 on obtient

$$\begin{aligned}
P_1(0, z_1, z_2) &= \frac{z_1 - g(z_2)}{z_1 - \tilde{B}(\lambda - \lambda(1-p)z_1 - \lambda pz_2)} \\
&\quad \times \frac{\tilde{A}(\lambda) (1 - z_2) p_0}{\tilde{A}(\lambda) (1 - z_2) g(z_2) - z_2 [1 - g(z_2)]}
\end{aligned}$$

Maintenant nous allons définir

$$P_{0,0,j} = \int_0^{\infty} P_{0,0,j}(x) b(x) dx, j \geq 1;$$

comme la probabilité qu'il y est j clients en orbite et que le serveur soit libre. Introduisons la fonction génératrice correspondante

$$P_0(z_2) = \sum_{j=1}^{\infty} P_{0,0,j} z_2^j = \int_0^{\infty} P_0(x, z_2) dx = \frac{1 - \tilde{A}(\lambda)}{\lambda} P_0(0, z_2).$$

D'autre part, nous allons définir

$$P_{1,i,j} = \int_0^{\infty} P_{1,i,j}(x) dx, i \geq 0, j \geq 0.$$

comme la probabilité qu'il y est i clients dans la file d'attente prioritaire, j clients en orbite et que le serveur est occupé. Et introduisons la fonction génératrice correspondante

$$\begin{aligned} P_1(z_1, z_2) &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} P_{1,i,j} z_1^i z_2^j = \int_0^{\infty} P_1(x, z_1, z_2) dx \\ &= \frac{1 - \tilde{B}(\lambda - \lambda(1-p)z_1 - \lambda pz_2)}{\lambda - \lambda(1-p)z_1 - \lambda pz_2} P_1(0, z_1, z_2) \end{aligned}$$

Alors à partir de la condition de normalisation $p_0 + P_0(1) + P_1(1, 1) = 1$, il s'ensuit que

$$p_0 = 1 - \frac{p + (1-p)\tilde{A}(\lambda)}{\tilde{A}(\lambda)} \rho. \quad (3.4.15)$$

Evidemment à partir de 3.4.15, puisque $p_0 > 0$, on obtient que $\rho < \frac{\tilde{A}(\lambda)}{p + (1-p)\tilde{A}(\lambda)}$ est une condition nécessaire à la stabilité de notre système.

En résumé, nous avons le théorème suivant.

Théorème 3.4.1. *Si $\rho < \frac{\tilde{A}(\lambda)}{p + (1-p)\tilde{A}(\lambda)}$, alors l'état du système est en régime stationnaire la fonction génératrice de la distribution stationnaire de nombre de clients en orbite, étant donné que le serveur est*

- libre, est donnée par

$$\begin{aligned} P_0(x, z_2) &= \left\{ 1 - \frac{p + (1-p)\tilde{A}(\lambda)}{\tilde{A}(\lambda)} \rho \right\} \\ &\quad \times \frac{z_2(1-g(z_2))}{\tilde{A}(\lambda)(1-z_2)g(z_2) - z_2(1-g(z_2))} [1 - A(x)] \lambda e^{-\lambda x}; \end{aligned}$$

- occupé, est

$$\begin{aligned} P_1(x, z_1, z_2) &= \left\{ 1 - \frac{p + (1-p)\tilde{A}(\lambda)}{\tilde{A}(\lambda)} \rho \right\} \times \frac{\tilde{A}(\lambda)(1-z_2)}{\tilde{A}(\lambda)(1-z_2)g(z_2) - z_2(1-g(z_2))} \\ &\quad \times \frac{z_1 - g(z_2)}{z_1 - \tilde{B}(\lambda - \lambda(1-p)z_1 - \lambda pz_2)} \\ &\quad \times [1 - A(x)] \lambda e^{-(\lambda - \lambda(1-p)z_1 - \lambda pz_2)x}; \end{aligned}$$

où $g(z_2)$ est la racine unique de z_1 de l'équation $z_1 - \tilde{B}(\lambda - \lambda(1-p)z_1 - \lambda pz_2) = 0$.

Si nous négligeons le temps inter-rappel écoulé depuis $\xi_0(t)$ et le temps de service écoulé $\xi_1(t)$, nous obtenons les fonctions génératrices correspondantes.

Après plusieurs calculs, simplifications et remplacements, les fonctions génératrices correspondantes sont

$$P_0(z_2) = \left\{ 1 - \frac{p + (1-p)\tilde{A}(\lambda)}{\tilde{A}(\lambda)}\rho \right\} \frac{z_2(1-g(z_2))(1-\tilde{A}(\lambda))}{\tilde{A}(\lambda)(1-z_2)g(z_2) - z_2(1-g(z_2))}$$

et

$$\begin{aligned} P_1(z_1, z_2) &= \left\{ 1 - \frac{p + (1-p)\tilde{A}(\lambda)}{\tilde{A}(\lambda)}\rho \right\} \frac{\tilde{A}(\lambda)(1-z_2)}{\tilde{A}(\lambda)(1-z_2)g(z_2) - z_2(1-g(z_2))} \\ &\times \frac{z_1 - g(z_2)}{z_1 - \tilde{B}(\lambda - \lambda(1-p)z_1 - \lambda pz_2)} \\ &\times \frac{1 - \tilde{B}(\lambda - \lambda(1-p)z_1 - \lambda pz_2)}{1 - (1-p)z_1 - pz_2}. \end{aligned}$$

A partir du théorème (3.4.1), nous obtenons une généralisation de la Formule de Pollaczek-Khinchin.

Théorème 3.4.2 (Atencia [15]). [*formule généralisée de Pollaczek-Khinchin*]

– La fonction génératrice de la distribution stationnaire du nombre de clients dans le système N est donnée par

$$\begin{aligned} Q(z) &= p_0 + P_0(z) + zP_1(z, z) \\ &= \left\{ 1 - \frac{p + (1-p)\tilde{A}(\lambda)}{\tilde{A}(\lambda)}\rho \right\} \\ &\times \frac{(1-z)\tilde{B}(\lambda - \lambda(1-p)z_1 - \lambda pz_2)}{\tilde{B}(\lambda - \lambda(1-p)z_1 - \lambda pz_2) - z} \times \frac{\tilde{A}(\lambda)[z - g(z)]}{\tilde{A}(\lambda)(1-z)g(z) - z(1-g(z))} \end{aligned}$$

qui sera appelée *formule généralisée de Pollaczek-Khinchin*

– La fonction génératrice de la distribution stationnaire du nombre de clients dans la file prioritaire N_f est donnée par

$$R(z) = p_0 + P_0(1) + P_1(z, 1) = 1 - \rho + \frac{1 - \rho q}{q} \frac{1 - \tilde{B}(\lambda q(1-z))}{\tilde{B}(\lambda q(1-z)) - z}$$

– La fonction génératrice de la distribution stationnaire du nombre de clients en orbite N_o est donnée par

$$\begin{aligned} P(z) &= p_0 + P_0(z) + P_1(1, z) \\ &= \left\{ 1 - \frac{p + (1-p)\tilde{A}(\lambda)}{\tilde{A}(\lambda)}\rho \right\} \\ &\times \frac{\tilde{A}(\lambda)}{p} \frac{1 - qg(z) - pz}{\tilde{A}(\lambda)(1-z)g(z) - z[1-g(z)]}; \end{aligned}$$

ou bien

$$\begin{aligned}
P(z) = & \frac{(1 - \lambda\beta_1)(1 - qg(z) - pz)}{(g(z) - z)p} \times \frac{1}{(1 - \lambda\beta_1)} \times \left[1 - \frac{p + q\tilde{A}(\lambda)}{\tilde{A}(\lambda)}\rho \right] \times \tilde{A}(\lambda) \\
& \times \frac{g(z) - z}{\tilde{A}(\lambda)(1 - z)g(z) - z(1 - g(z))} \tag{3.4.16}
\end{aligned}$$

3.5 Propriété de décomposition stochastique

Parmi les approches développées ces dernières années permettant d'analyser les systèmes de files d'attente avec rappels, on trouve celle basée sur la propriété de décomposition stochastique que peut posséder un modèle. Elle offre les avantages de simplification de résolution de modèles complexes. Le concept général de la propriété de décomposition stochastique 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).

3.5.1 Propriété de décomposition stochastique de la distribution de la taille du système

La littérature sur les modèles de vacation reconnaît cette propriété comme l'une des caractéristiques les plus intéressantes sur ce sujet. Le premier résultat sur la décomposition stochastique est le résultat de Fuhrmann et Cooper [44], qui a été généralisé par Shanthikumar dans [64]. Yang et Templeton étudié dans [73], la décomposition stochastique des systèmes de files d'attente avec rappel, dont les applications ont été discutés plus tard par Artalejo et Falin [12], Yang et al. [75]. L'interprétation classique des propriétés de décomposition stochastique montre que la distribution de la taille du système se décompose en deux variables aléatoires l'un des ce qui correspond à la taille du système de la file d'attente ordinaire sans vacation. L'interprétation de l'autre variable aléatoire est généralement liée à la taille de système étant donné que le serveur est en vacation (voir proposition 5 dans [44]). En particulier, dans le contexte de notre système, nous observons les relations suivantes entre les fonctions génératrices

$$\lim_{\tilde{A}(\lambda) \rightarrow 1} Q(z) = Q^\infty(z),$$

où

$$Q^\infty(z) = (1 - \rho) \frac{(1 - z)\tilde{B}(\lambda(1 - z))}{\tilde{B}(\lambda q(1 - z)) - z}$$

est la fonction génératrice de la distribution stationnaire du nombre de clients dans le système $M/G/1/\infty$, c'est la formule de Pollaczek-Khinchin. Par conséquent, comme $\tilde{A}(\lambda) \rightarrow 1$ (ce qui équivaut à considérer la distribution exponentielle de rappels avec taux croissance illimité), dans ce cas notre modèle peut être considéré comme un système de file d'attente $M/G/1/\infty$ ordinaire. Nous notons maintenant que la formule généralisée de Pollaczek-Khinchin peut être écrite comme suit

$$Q(z) = Q^\infty(z) \frac{p_0 + P_0(z)}{p_0 + P_0(1)}$$

où le second terme sur la droite représente la fonction génératrice de la distribution stationnaire du nombre de clients en orbite étant donné que le serveur est libre. En fait, c'est la propriété de décomposition stochastique pour notre système de file d'attente ; alors, le nombre total de clients dans notre système de files d'attente peut se présenter comme la somme de deux variables aléatoires indépendantes : l'une est le nombre total de clients dans le système ordinaire correspondant et l'autre est le nombre de clients dans l'orbite étant donné que le serveur est libre. Ce résultat peut être résumé dans le théorème suivant.

Remarque 3.5.1. Soient les notations suivantes :

- (M_∞) : le modèle classique $M/G/1/\infty$ avec deux type de clients
- $(M_{\tilde{A}(\lambda)})$: le modèle $M/G/1$ avec rappels, deux type de clients et FCFS orbite.

Théorème 3.5.1. *Le nombre de clients dans le système N peut être écrit comme la somme de deux variables aléatoires indépendantes, la première représente le nombre de clients dans le système associé au modèle (M_∞) et la deuxième est le nombre de clients répétés (en orbite) associé au modèle $(M_{\tilde{A}(\lambda)})$ sachant que le serveur est libre.*

$$N = N_\infty + R.$$

Il n'est pas surprenant que le système étudié satisfait cette propriété, puisque notre système peut être considéré comme une file d'attente prioritaire à deux niveaux avec vacances du serveur. Dans ce modèle de vacation, le serveur commence ses vacances quand un service se termine et il n'y a pas de clients prioritaires dans la file d'attente. La durée des vacances dépend du processus d'arrivée et de la distribution qui régit le temps inter-rappels.

Les vacances se terminent lorsque le serveur est activé selon une compétition entre une loi exponentielle de taux λ (correspondant aux arrivées primaires) et la distribution générale du temps de rappels. Avec cette définition, la propriété de décomposition stochastique observée précédemment pour notre système est cohérente avec celui rapporté par Fuhrmann et Cooper [44] pour le modèle $M/G/1$ avec vacances.

Remarque 3.5.2. Puisque nous avons prouvé cette propriété de décomposition stochastique, l'égalité suivante est évidente

$$E[N] = E[N_\infty] + \frac{1}{1-\rho} P'_0(1),$$

qui représente une extension de la formule Pollaczek-Khinchin.

3.5.2 Propriété de décomposition stochastique pour le nombre de clients en orbite

En ce qui concerne la décomposition stochastique pour le nombre de clients en orbite, nous pouvons déduire (de manière analogue) le résultat suivant :

De l'équation 3.4.16, la fonction génératrice de la distribution stationnaire du nombre de clients en orbite $P(z)$ peut s'écrire sous la forme suivante :

$$P(z) = X(z) \times Y(z),$$

où

$$X(z) = \frac{(1 - \lambda\beta_1)(1 - qg(z) - pz)}{(g(z) - z)p}$$

qui est la fonction génératrice du nombre de client dans la file d'attente aux instants de départ associée à (M_∞) modèle d'attente classique $M/G/1$ avec deux types de clients.

Le deuxième facteur

$$Y(z) = \frac{1}{(1 - \lambda\beta_1)} \times \left[1 - \frac{p + q\tilde{A}(\lambda)}{\tilde{A}(\lambda)}\rho \right] \times \tilde{A}(\lambda) \times \frac{g(z) - z}{\tilde{A}(\lambda)(1 - z)g(z) - z(1 - g(z))}$$

représente la fonction génératrice du nombre de client en orbite associée à $(M_{\tilde{A}(\lambda)})$ système $M/G/1$ avec rappels, priorité relative et orbite FCFS sachant que le serveur est libre.

Alors, la décomposition stochastique peut être exprimée comme suit :

$$\left\{ 0, N_{o,\tilde{A}(\lambda)}(t), t \geq 0 \right\} = \left\{ 0, N_{f,\infty}(t), t \geq 0 \right\} + \left\{ 0, N_{\tilde{A}(\lambda)}(t), t \geq 0 \right\}.$$

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

Théorème 3.5.2. *Le nombre de clients en orbite $N_{o,\tilde{A}(\lambda)}$ associé au modèle $(M_{\tilde{A}(\lambda)})$ peut être écrit comme la somme de deux variables aléatoires indépendantes, dont la première $N_{f,\infty}$ est le nombre de clients dans la file d'attente associé au modèle classique (M_∞) et la deuxième $N_{\tilde{A}(\lambda)}$ est le nombre de clients en orbite associé à $(M_{\tilde{A}(\lambda)})$ étant donné que le serveur est libre.*

$$N_{o,\tilde{A}(\lambda)} = N_{f,\infty} + N_{\tilde{A}(\lambda)}.$$

De plus on a,

$$E \left[N_{o,\tilde{A}(\lambda)} \right] = E \left[N_{f,\infty} \right] + \frac{1}{1 - \rho} P'_0(1).$$

3.6 Mesures de performance

Dans le corollaire suivant, nous présentons quelques mesures de performance pour le système dans un régime stationnaire (Atencia) [17].

- Le système est libre avec une probabilité

$$p_0 = 1 - \frac{p + (1-p)\tilde{A}(\lambda)}{\tilde{A}(\lambda)}\rho;$$

- Le système est occupé avec une probabilité

$$P_0(1) + P_1(1,1) = \frac{p + (1-p)\tilde{A}(\lambda)}{\tilde{A}(\lambda)}\rho;$$

- Le serveur est libre avec une probabilité

$$p_0 + P_0(1) = 1 - \rho;$$

- Le serveur est occupé avec une probabilité

$$P_1(1,1) = \rho;$$

- Le nombre moyen de clients dans la file prioritaire est

$$E[N_f] = \frac{\lambda^2(1-p)\beta_2}{2(1-\rho q)};$$

- Le nombre moyen de clients en orbite est

$$E[N_o] = \frac{\lambda p}{2(1-\rho(1-p))} \frac{2\beta_1(1-\rho(1-p)) \left[1 - \tilde{A}(\lambda)\right] + \lambda\beta_2 \left[p + (1-p)\tilde{A}(\lambda)\right]}{\tilde{A}(\lambda) - \left[p + (1-p)\tilde{A}(\lambda)\right]\rho};$$

- Le nombre moyen de clients dans le système est

$$E[N] = \rho + \frac{\lambda^2\beta_2}{2(1-\rho)} + \frac{p \left[1 - \tilde{A}(\lambda)\right]}{\tilde{A}(\lambda) - \left[p + (1-p)\tilde{A}(\lambda)\right]\rho} \\ \times \left\{ \rho + \frac{\lambda^2 p \beta_2}{2(1-\rho)(1-\rho(1-p))} \right\};$$

- Un client qui se fait servir est considéré comme membre de la file d'attente prioritaire avec la probabilité $(1-p)$ et est considéré provenant de l'orbite avec la probabilité p .
- Le temps moyen d'attente pour chaque type de client est donné par

$$\bar{W}_f = \frac{E[N_f]}{\lambda(1-p)} = \frac{\lambda\beta_2}{2(1-\rho(1-p))};$$

$$\begin{aligned}\bar{W}_o &= \frac{E[N_o]}{\lambda p} \\ &= \frac{2\beta_1(1-\rho(1-p))\left[1-\tilde{A}(\lambda)\right] + \lambda\beta_2\left[p+(1-p)\tilde{A}(\lambda)\right]}{2(1-\rho(1-p))\left\{\tilde{A}(\lambda) - \left[p+(1-p)\tilde{A}(\lambda)\right]\rho\right\}},\end{aligned}$$

– Le temps moyen de séjour (y compris le temps de service) est donné par

$$\begin{aligned}\bar{W}_s &= \frac{E[N]}{\lambda} = \beta_1 + \frac{\lambda\beta_2}{2(1-\rho)} \\ &\quad + \frac{p\left[1-\tilde{A}(\lambda)\right]}{\tilde{A}(\lambda) - \left[p+(1-p)\tilde{A}(\lambda)\right]\rho} \left\{ \beta_1 + \frac{\lambda p\beta_2}{2(1-\rho)(1-\rho(1-p))} \right\}.\end{aligned}$$

Remarque 3.6.1. Nous observons que la distribution stationnaire du nombre de clients dans la file d’attente prioritaire ne dépend pas de la distribution des temps inter-rappel (c’est-à-dire qu’elle est indépendante de la loi qui régit l’accès depuis l’orbite au serveur).

La distribution stationnaire de l’état du serveur $p_0 + P_0(1) = 1 - \rho$, $P_1(1, 1) = \rho$, dépend de la distribution des temps de service $B(x)$ seulement à travers sa moyenne β_1 et ne dépend pas de la distribution des temps inter-rappels.

Comme prévu, $E\left[N_{\tilde{A}(\lambda)}\right] = E\left[N_{o,\tilde{A}(\lambda)}\right] + E[N_f] + \rho$ et par conséquent $\bar{W}_s = (1-p)\bar{W}_f + p\bar{W}_o + \beta_1$.

3.6.1 Cas spéciaux

Dans ce paragraphe, nous observons que les résultats présentés ci-dessus sont compatibles avec ceux connus dans la littérature.

– Lorsque $p = 0$, le modèle considéré devient le système de files d’attente classique $M/G/1/\infty$. Dans ce cas, $N_2 = 0$. L’équation du **Théorème 3.4.2** sera réduite à

$$Q(z) = (1-\rho) \frac{(1-z)\tilde{B}(\lambda(1-z))}{\tilde{B}(\lambda(1-z)) - z};$$

qui est en accord avec la formule de Pollaczek-Khinchin bien connu.

– Quand $p = 1$, le modèle à l’étude devient une file d’attente $M/G/1$ avec un temps de rappel général, qui a été étudié dans [46].

Dans ce cas, $N_1 = 0$ et $g(z_2) = \tilde{B}(\lambda - \lambda z)$. Les équations de **théorème 3.4.1** sont réduites à

$$p_0 = 1 - \frac{\rho}{\tilde{A}(\lambda)};$$

$$P_0(x, z_2) = \left\{ 1 - \frac{\rho}{\tilde{A}(\lambda)} \right\} \frac{z_2 [1 - g(z_2)]}{\tilde{A}(\lambda) (1 - z_2) g(z_2) - z_2 [1 - g(z_2)]} \\ \times [1 - A(x)] \lambda e^{-\lambda x};$$

$$P_1(x, 1, z_2) = \left\{ 1 - \frac{\rho}{\tilde{A}(\lambda)} \right\} \frac{\tilde{A}(\lambda) (1 - z_2)}{\tilde{A}(\lambda) (1 - z_2) g(z_2) - z_2 [1 - g(z_2)]} \\ \times [1 - B(x)] \lambda e^{-(\lambda - \lambda z_2)x};$$

qui sont en accord avec le **théorème 2** dans [46].

– Quand $A(x) = 1 - e^{-\nu x}$, nous avons un système de files d'attente $M/G/1$ avec rappels et Bernoulli Schedule, qui a été étudié dans [16]. Les fonctions génératrices principales du modèle sont réduites à

$$p_0 + P_0(z_2) = [\nu - (\lambda p + \nu) \rho] \frac{g(z_2) - z_2}{\lambda z_2 [g(z_2) - 1] + \nu [g(z_2) - z_2]};$$

$$P_1(z_1, z_2) = [\nu - (\lambda p + \nu) \rho] \frac{1 - z_2}{\lambda z_2 [g(z_2) - 1] + \nu [g(z_2) - z_2]} \\ \times \frac{z_1 - g(z_2)}{z_1 - \tilde{B}(\lambda - \lambda q z_1 - \lambda p z_2)} \frac{1 - \tilde{B}(\lambda - \lambda q z_1 - \lambda p z_2)}{1 - q z_1 - p z_2};$$

qui sont en accord avec le **théorème 4.1** dans [16] les caractéristiques importantes deviennent :

$$E[N_f] = \frac{\lambda^2 q \beta_2}{2(1 - \rho q)};$$

$$E[N_o] = \frac{\lambda \rho p (1 - \rho q) + (\lambda p + \nu) \frac{\lambda^2 p \beta_2}{2}}{(1 - \rho q) [\nu - (\lambda p + \nu) \rho]}.$$

3.7 Conclusion

Nous avons présenté l'analyse mathématique d'un système de files d'attente avec rappels, priorité relative, orbite FCFS et distribution générale du temps inter-rappels. Les résultats obtenus sont d'une complexité particulière et de cette manière leur utilisation pratique est restreinte. C'est pourquoi, dans le chapitre suivant, nous étudions le comportement asymptotique du modèle considéré.

Chapitre 4

Comportement asymptotique du nombre de clients dans le modèle $M/G/1$ avec rappels et priorité relative et orbite FCFS

4.1 Introduction

Une première étude du comportement asymptotique d'un modèle avec rappels a été réalisée par Falin [35], [36] et ceci pour le système avec rappels de type $M/G/1$. Une autre étude plus récente a été faite par J. Kim et B. Kim [51], 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).

Bien que les caractéristiques de performance du système avec rappels soient disponibles sous forme explicite, elles sont complexes et d'une utilisation restreinte du point de vue pratique. Prenons l'expression de la fonction génératrice de la distribution stationnaire du nombre de clients en orbite obtenue dans le chapitre 3. Elle ne permet pas d'étudier le caractère de la distribution de $N_o(t)$.

Dans ce chapitre, nous étudions le comportement asymptotique du modèle $M/G/1$ avec rappels, priorité relative et orbite FCFS, et ceci dans le cas des valeurs extrêmes de certains paramètres : régime intense, taux de rappels faible et taux de rappels fort.

4.2 Trafic intense

Nous nous intéressons dans ce paragraphe au cas du trafic intense, c'est-à-dire quand le taux d'arrivées croît de telle manière que l'intensité du trafic $\rho^* = \frac{p + q\hat{A}(\lambda)}{\hat{A}(\lambda)}\rho$ tend vers 1. Nous avons obtenu le résultat suivant :

Théorème 4.2.1. *Si la file d'attente $M/G/1$ avec rappels, deux types de clients, priorité relative, l'orbite FCFS et temps des rappels non Exponentielle est en état stationnaire, et si le deuxième moment de la distribution du temps de service $\beta_2 < \infty$, alors*

$$\lim_{z \rightarrow 1} P(z) = \frac{1}{1 + s \frac{\beta_2}{2\beta_1^2}} \times \frac{1}{1 + s \left[\frac{p(1-\tilde{A}(\lambda))\tilde{A}(\lambda)}{(p+q\tilde{A}(\lambda))^2} + \frac{(\tilde{A}(\lambda))^2 \beta_2}{2(p+q\tilde{A}(\lambda))\beta_1^2} \right]}. \quad (4.2.1)$$

C'est à dire, dans le cas d'un trafic intense, le nombre de clients en orbite $N_o(t)$ suit asymptotiquement une distribution d'Erlang généralisée d'ordre deux.

Démonstration. Soit

$$P(z) = \left[1 - \frac{p + q\tilde{A}(\lambda)}{\tilde{A}(\lambda)} \rho \right] \times \frac{\tilde{A}(\lambda)}{p} \times \frac{1 - qg(z) - pz}{\tilde{A}(\lambda)(1-z)g(z) - z(1-g(z))}$$

la fonction génératrice du nombre de clients en orbite réécrite pour appliquer la propriété de décomposition stochastique. En utilisant la propriété de décomposition stochastique du modèle [11], nous pouvons présenter $P(z)$ de la manière suivante :

$$P(z) = X(z) \times Y(z), \quad (4.2.2)$$

Où

$$X(z) = \frac{(1 - \lambda\beta_1)(1 - qg(z) - pz)}{(g(z) - z)p}$$

est la fonction génératrice de la distribution stationnaire du nombre de clients dans l'espace d'attente de la file d'attente ordinaire avec deux types de clients (sans rappels),
et

$$Y(z) = \frac{1}{(1 - \lambda\beta_1)} \times \left[1 - \frac{p + q\tilde{A}(\lambda)}{\tilde{A}(\lambda)} \rho \right] \times \tilde{A}(\lambda) \times \frac{g(z) - z}{\tilde{A}(\lambda)(1-z)g(z) - z(1-g(z))}$$

est la fonction génératrice de la distribution stationnaire du nombre de clients en orbite étant donné que le serveur est libre.

Alors

$$\lim_{z \rightarrow 1} P(z) = \lim_{z \rightarrow 1} X(z) \times \lim_{z \rightarrow 1} Y(z).$$

Dans un premier temps, calculons le terme $\lim_{z \rightarrow 1} X(z)$. En effet,

$$\lim_{z \rightarrow 1} X(z) = \frac{(1 - \lambda\beta_1)(1 - qg(z) - pz)}{(g(z) - z)p}.$$

Supposons que $\varepsilon = (1 - \rho) \rightarrow 0$ et $z = e^{-\varepsilon s}$. Ainsi $\varepsilon = (1 - \rho) = 1 - \lambda\beta_1$ et $\lambda = \frac{1}{\beta_1}(1 - \varepsilon)$.

Sous cette hypothèse, on obtient

$$\lim_{\varepsilon \rightarrow 0} X(e^{-\varepsilon s}) = \lim_{\varepsilon \rightarrow 0} \frac{\varepsilon (1 - qg(e^{-\varepsilon s}) - pe^{-\varepsilon s})}{(g(e^{-\varepsilon s}) - e^{-\varepsilon s})p}.$$

Soit $g(e^{-\varepsilon s}) = g(z)$, où $z = e^{-\varepsilon s}$. On peut développer $g(z)$ de la manière suivante :

$$e^{-\varepsilon s} = 1 - \varepsilon s + \frac{\varepsilon^2 s^2}{2} + o(\varepsilon).$$

$$g(u) = g(1 + (u - 1)) = g(1) + (u - 1)g'(1) + \frac{1}{2}(u - 1)^2 g''(1).$$

Ceci donne

$$\begin{aligned} g(e^{-\varepsilon s}) &= g(1) + (-\varepsilon s + \frac{\varepsilon^2 s^2}{2} + o(\varepsilon))g'(1) + \frac{1}{2}(-\varepsilon s + o(\varepsilon))^2 g''(1) \\ &= 1 - \varepsilon s g'(1) + \frac{\varepsilon^2 s^2}{2} g'(1) + \frac{\varepsilon^2 s^2}{2} g''(1) + o(\varepsilon^2). \end{aligned}$$

Nous avons également que $g(1) = 1$, $g'(1) = \frac{\rho p}{1 - \rho q}$, $g''(1) = \frac{\lambda^2 p^2 \beta_2}{(1 - \rho q)^3}$.

Donc,

$$\begin{aligned} g(e^{-\varepsilon s}) &= 1 - \frac{\varepsilon s p}{(1 - (1 - \varepsilon)q)} + \frac{\varepsilon^2 s p}{(1 - (1 - \varepsilon)q)} \\ &+ \frac{\varepsilon^2 s^2}{2} \frac{p}{(1 - (1 - \varepsilon)q)} + \frac{\varepsilon^2 s^2}{2} \frac{(1 - \varepsilon)^2 \frac{\beta_2}{\beta_1^2} p^2}{(1 - (1 - \varepsilon)q)^3} + o(\varepsilon^2) \\ &= 1 + \frac{1}{(1 - (1 - \varepsilon)q)^3} \left[-\varepsilon s p^3 + \varepsilon^2 s p^2 (p - 2q) + \frac{\varepsilon^2 s^2}{2} p^2 (p + \frac{\beta_2}{\beta_1^2}) \right] + o(\varepsilon^2). \end{aligned}$$

Prenant en compte que $\beta_2 < \infty$, nous obtenons

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} X(e^{-\varepsilon s}) &= \lim_{\varepsilon \rightarrow 0} \frac{\varepsilon (1 - qg(e^{-\varepsilon s}) - pe^{-\varepsilon s})}{(g(e^{-\varepsilon s}) - e^{-\varepsilon s})p} \\ &= \lim_{\varepsilon \rightarrow 0} \frac{\varepsilon \left(\frac{1}{(1 - (1 - \varepsilon)q)^3} \left[\varepsilon s q p^3 - \varepsilon^2 s q p^2 (p - 2q) - \frac{\varepsilon^2 s^2}{2} q p^2 (p + \frac{\beta_2}{\beta_1^2}) \right] - p \left(-\varepsilon s + \frac{\varepsilon^2 s^2}{2} \right) \right)}{\left(\frac{1}{(1 - (1 - \varepsilon)q)^3} \left[-\varepsilon s p^3 + \varepsilon^2 s p^2 (p - 2q) + \frac{\varepsilon^2 s^2}{2} p^2 (p + \frac{\beta_2}{\beta_1^2}) \right] + \varepsilon s - \frac{\varepsilon^2 s^2}{2} \right) p} \\ &= \lim_{\varepsilon \rightarrow 0} \frac{\varepsilon^2 s q p^3 + \varepsilon^2 s p^4}{\varepsilon^2 s p^2 (p - 2q + 3q) + \frac{\varepsilon^2 s^2}{2} p^3 (p + \frac{\beta_2}{\beta_1^2} - p)} \\ &= \lim_{\varepsilon \rightarrow 0} \frac{\varepsilon^2 s p^3}{\varepsilon^2 s p^3 + \frac{\varepsilon^2 s^2}{2} p^3 \frac{\beta_2}{\beta_1^2}} = \frac{1}{1 + \frac{s \beta_2}{2 \beta_1^2}}. \end{aligned}$$

A présent, calculons le deuxième terme $\lim_{z \rightarrow 1} Y(z)$:

$$Y(z) = \frac{1}{(1 - \lambda\beta_1)} \times \left[1 - \frac{p + q\tilde{A}(\lambda)}{\tilde{A}(\lambda)} \rho \right] \times \tilde{A}(\lambda) \times \frac{g(z) - z}{\tilde{A}(\lambda) (1 - z) g(z) - z (1 - g(z))}.$$

Dans ce cas $\rho = \lambda\beta_1 < \frac{\tilde{A}(\lambda)}{p + q\tilde{A}(\lambda)}$.

Posons $\varepsilon = \frac{\tilde{A}(\lambda)}{p + q\tilde{A}(\lambda)} - \rho \rightarrow 0$ et $z = e^{-\varepsilon s}$, ainsi $\varepsilon = \frac{\tilde{A}(\lambda)}{p + q\tilde{A}(\lambda)} - \lambda\beta_1$ et $\lambda = \frac{1}{\beta_1} \left(\frac{\tilde{A}(\lambda)}{p + q\tilde{A}(\lambda)} - \varepsilon \right)$.

Posons $\eta_1 = \frac{\tilde{A}(\lambda)}{p + q\tilde{A}(\lambda)}$. Alors $\lambda = \frac{1}{\beta_1} (\eta_1 - \varepsilon)$.

Sous ces hypothèses, $\lim_{z \rightarrow 1} Y(z)$ devient :

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} Y(e^{-\varepsilon s}) &= \lim_{\varepsilon \rightarrow 0} \frac{1}{((1 - \eta_1) + \varepsilon)} \times \left[1 - \frac{p + q\tilde{A}(\lambda)}{\tilde{A}(\lambda)} (\eta_1 - \varepsilon) \right] \\ &\quad \times \tilde{A}(\lambda) \times \frac{g(e^{-\varepsilon s}) - e^{-\varepsilon s}}{\tilde{A}(\lambda) (1 - e^{-\varepsilon s}) g(e^{-\varepsilon s}) - e^{-\varepsilon s} (1 - g(e^{-\varepsilon s}))}, \end{aligned}$$

et $g(e^{-\varepsilon s})$ peut être obtenu sous la forme suivante :

$$g(e^{-\varepsilon s}) = 1 - \varepsilon s g'(1) + \frac{\varepsilon^2 s^2}{2} g'(1) + \frac{\varepsilon^2 s^2}{2} g''(1) + o(\varepsilon^2).$$

D'où

$$\begin{aligned} g(e^{-\varepsilon s}) &= 1 - \frac{\varepsilon s (\eta_1 - \varepsilon) p}{(1 - (\eta_1 - \varepsilon) q)} + \frac{\varepsilon^2 s^2}{2} \frac{(\eta_1 - \varepsilon) p}{(1 - (\eta_1 - \varepsilon) q)} + \frac{\varepsilon^2 s^2}{2} \frac{(\eta_1 - \varepsilon)^2 \frac{\beta_2}{\beta_1^2} p^2}{(1 - (\eta_1 - \varepsilon) q)^3} + o(\varepsilon^2) \\ &= 1 - \frac{\varepsilon s \eta_1 p}{(1 - (\eta_1 - \varepsilon) q)} + \frac{\varepsilon^2 s p}{(1 - (\eta_1 - \varepsilon) q)} + \frac{\varepsilon^2 s^2}{2} \frac{\eta_1 p}{(1 - (\eta_1 - \varepsilon) q)} \\ &\quad + \frac{\varepsilon^2 s^2}{2} \frac{\eta_1^2 \frac{\beta_2}{\beta_1^2} p^2}{(1 - (\eta_1 - \varepsilon) q)^3} + o(\varepsilon^2) \\ &= 1 + \frac{1}{(1 - (\eta_1 - \varepsilon) q)^3} (-\varepsilon s \eta_1 p (1 - q\eta_1)^2 + \varepsilon^2 s p (1 - q\eta_1) (1 - 3q\eta_1) \\ &\quad + \frac{\varepsilon^2 s^2}{2} \eta_1 p ((1 - q\eta_1)^2 + \eta_1 \frac{\beta_2}{\beta_1^2} p)) + o(\varepsilon^2). \end{aligned}$$

D'où

$$\lim_{\varepsilon \rightarrow 0} Y(e^{-\varepsilon s}) = \lim_{\varepsilon \rightarrow 0} \frac{1}{((1 - \eta_1) + \varepsilon)} \times \frac{(p + q\tilde{A}(\lambda)) \varepsilon (g(e^{-\varepsilon s}) - e^{-\varepsilon s})}{\tilde{A}(\lambda) (1 - e^{-\varepsilon s}) g(e^{-\varepsilon s}) - e^{-\varepsilon s} (1 - g(e^{-\varepsilon s}))}.$$

Prenant en compte que $\beta_2 < \infty$, on a

$$\lim_{\varepsilon \rightarrow 0} Y(e^{-\varepsilon s}) = \lim_{\varepsilon \rightarrow 0} \frac{\left(p + q\tilde{A}(\lambda)\right) \times \varepsilon \times (g(e^{-\varepsilon s}) - e^{-\varepsilon s})}{((1 - \eta_1) + \varepsilon) \tilde{A}(\lambda) (1 - e^{-\varepsilon s}) g(e^{-\varepsilon s}) - ((1 - \eta_1) + \varepsilon) e^{-\varepsilon s} (1 - g(e^{-\varepsilon s}))}$$

Posons

$$Y_1(e^{-\varepsilon s}) = \left(p + q\tilde{A}(\lambda)\right) \times \varepsilon \times (g(e^{-\varepsilon s}) - e^{-\varepsilon s})$$

et

$$Y_2(e^{-\varepsilon s}) = ((1 - \eta_1) + \varepsilon) \tilde{A}(\lambda) (1 - e^{-\varepsilon s}) g(e^{-\varepsilon s}) - ((1 - \eta_1) + \varepsilon) e^{-\varepsilon s} (1 - g(e^{-\varepsilon s})).$$

La fonction $Y_1(e^{-\varepsilon s})$ peut être exprimée de la manière suivante :

$$Y_1(e^{-\varepsilon s}) = \left(p + q\tilde{A}(\lambda)\right) \times \varepsilon \times (g(e^{-\varepsilon s}) - e^{-\varepsilon s}),$$

où

$$\begin{aligned} Y_1(e^{-\varepsilon s}) &= \left(p + q\tilde{A}(\lambda)\right) \times \varepsilon \times \frac{1}{(1 - (\eta_1 - \varepsilon)q)^3} \left((-\varepsilon s \eta_1 p (1 - q\eta_1)^2 + \varepsilon^2 s p (1 - q\eta_1) (1 - 3q\eta_1)) \right. \\ &\quad \left. + \frac{\varepsilon^2 s^2}{2} \eta_1 p ((1 - q\eta_1)^2 + \eta_1 \frac{\beta_2}{\beta_1^2} p) + \varepsilon s - \frac{\varepsilon^2 s^2}{2} \right) + o(\varepsilon^2) \\ &= \left(p + q\tilde{A}(\lambda)\right) \times \frac{1}{(1 - (\eta_1 - \varepsilon)q)^3} \left[-\varepsilon^2 s \eta_1 p (1 - q\eta_1)^2 + \varepsilon^2 s (1 - q\eta_1)^3 \right] + o(\varepsilon^2) \\ &= \frac{\left(p + q\tilde{A}(\lambda)\right)}{(1 - (\eta_1 - \varepsilon)q)^3} \left[\varepsilon^2 s (1 - q\eta_1)^2 (1 - q\eta_1 - \eta_1 p) \right]. \end{aligned}$$

En effectuant quelques simplifications et remplacements nous obtenons

$$Y_1(e^{-\varepsilon s}) = \frac{1}{(1 - (\eta_1 - \varepsilon)q)^3} \times \varepsilon^2 s \frac{p^3 (1 - \tilde{A}(\lambda))}{\left(p + q\tilde{A}(\lambda)\right)^2}.$$

Maintenant considérons $Y_2(e^{-\varepsilon s})$:

$$Y_2(e^{-\varepsilon s}) = ((1 - \eta_1) + \varepsilon) \tilde{A}(\lambda) (1 - e^{-\varepsilon s}) g(e^{-\varepsilon s}) - ((1 - \eta_1) + \varepsilon) e^{-\varepsilon s} (1 - g(e^{-\varepsilon s})).$$

Posons $Y_2(e^{-\varepsilon s}) = l_1(e^{-\varepsilon s}) + l_2(e^{-\varepsilon s})$,

où

$$l_1(e^{-\varepsilon s}) = ((1 - \eta_1) + \varepsilon) \tilde{A}(\lambda) (1 - e^{-\varepsilon s}) g(e^{-\varepsilon s})$$

et

$$l_2(e^{-\varepsilon s}) = -((1 - \eta_1) + \varepsilon) e^{-\varepsilon s} (1 - g(e^{-\varepsilon s})).$$

On peut écrire $l_1(\varepsilon)$ sous la forme suivante :

$$\begin{aligned} l_1(e^{-\varepsilon s}) &= ((1 - \eta_1) + \varepsilon) \tilde{A}(\lambda) (1 - e^{-\varepsilon s}) g(e^{-\varepsilon s}) + O(\varepsilon^2) \\ &= \frac{1}{(1 - (\eta_1 - \varepsilon)q)^3} (\varepsilon s \tilde{A}(\lambda) (1 - \eta_1) (1 - q\eta_1)^3 + \varepsilon^2 s \tilde{A}(\lambda) (1 - q\eta_1)^2 (3q(1 - \eta_1) \\ &\quad + (1 - q\eta_1)) - \frac{\varepsilon^2 s^2}{2} \tilde{A}(\lambda) (1 - \eta_1) (1 - q\eta_1)^2 (1 - q\eta_1 + 2\eta_1 p)), \end{aligned}$$

et $l_2(e^{-\varepsilon s})$ sous la forme

$$\begin{aligned} l_2(e^{-\varepsilon s}) &= \frac{1}{(1 - (\eta_1 - \varepsilon)q)^3} \times (-\varepsilon s p \eta_1 (1 - \eta_1) (1 - q\eta_1)^2 \\ &\quad + \varepsilon^2 s p (1 - q\eta_1) [(1 - \eta_1) (1 - 3q\eta_1) - \eta_1 (1 - q\eta_1)] \\ &\quad + \frac{\varepsilon^2 s^2}{2} \eta_1 p (1 - \eta_1) \left[3(1 - q\eta_1)^2 + \eta_1 p \frac{\beta_2}{\beta_1^2} \right]). \end{aligned}$$

Par conséquent, nous pouvons obtenir $Y_2(e^{-\varepsilon s}) = l_1(e^{-\varepsilon s}) + l_2(e^{-\varepsilon s})$. En effet,

$$\begin{aligned} Y_2(e^{-\varepsilon s}) &= \frac{1}{(1 - (\eta_1 - \varepsilon)q)^3} \\ &\times \left\{ \varepsilon^2 s \frac{p^3 (1 - \tilde{A}(\lambda))}{(p + q\tilde{A}(\lambda))} + \frac{\varepsilon^2 s^2}{2} \left(\frac{2p^4 \tilde{A}(\lambda) (1 - \tilde{A}(\lambda))^2}{(p + q\tilde{A}(\lambda))^4} + \frac{p^3 (\tilde{A}(\lambda))^2 (1 - \tilde{A}(\lambda)) \frac{\beta_2}{\beta_1^2}}{(p + q\tilde{A}(\lambda))^3} \right) \right\}. \end{aligned}$$

Alors

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} Y(e^{-\varepsilon s}) &= \lim_{\varepsilon \rightarrow 0} \frac{Y_1(e^{-\varepsilon s})}{Y_2(e^{-\varepsilon s})} \\ &= \lim_{\varepsilon \rightarrow 0} \frac{\varepsilon^2 s \frac{p^3 (1 - \tilde{A}(\lambda))}{(p + q\tilde{A}(\lambda))}}{\varepsilon^2 s \frac{p^3 (1 - \tilde{A}(\lambda))}{(p + q\tilde{A}(\lambda))} + \frac{\varepsilon^2 s^2}{2} \left(\frac{2p^4 \tilde{A}(\lambda) (1 - \tilde{A}(\lambda))^2}{(p + q\tilde{A}(\lambda))^4} + \frac{p^3 (\tilde{A}(\lambda))^2 (1 - \tilde{A}(\lambda)) \frac{\beta_2}{\beta_1^2}}{(p + q\tilde{A}(\lambda))^3} \right)} \\ &= \frac{1}{1 + s \left[\frac{p (1 - \tilde{A}(\lambda)) \tilde{A}(\lambda)}{(p + q\tilde{A}(\lambda))^2} + \frac{(\tilde{A}(\lambda))^2 \beta_2}{2 (p + q\tilde{A}(\lambda)) \beta_1^2} \right]}. \end{aligned}$$

Finalement, nous obtenons

$$\lim_{\varepsilon \rightarrow 0} X(e^{-\varepsilon s}) \times \lim_{\varepsilon \rightarrow 0} Y(e^{-\varepsilon s}) = \frac{1}{1 + \frac{s\beta_2}{2\beta_1^2}} \times \frac{1}{1 + s \left[\frac{p(1 - \tilde{A}(\lambda))\tilde{A}(\lambda)}{(p + q\tilde{A}(\lambda))^2} + \frac{(\tilde{A}(\lambda))^2\beta_2}{2(p + q\tilde{A}(\lambda))\beta_1^2} \right]}$$

ou

$$\lim_{z \rightarrow 1} P(z) = \frac{1}{1 + \frac{s\beta_2}{2\beta_1^2}} \times \frac{1}{1 + s \left[\frac{p(1 - \tilde{A}(\lambda))\tilde{A}(\lambda)}{(p + q\tilde{A}(\lambda))^2} + \frac{(\tilde{A}(\lambda))^2\beta_2}{2(p + q\tilde{A}(\lambda))\beta_1^2} \right]}.$$

Ce dernier résultat est le produit de deux transformées de *Laplace-Stieltjes* de deux variables exponentielles avec des taux différents $\lambda_1 = \frac{\beta_2}{2\beta_1^2}$ et $\lambda_2 = \frac{p(1 - \tilde{A}(\lambda))\tilde{A}(\lambda)}{(p + q\tilde{A}(\lambda))^2} +$

$\frac{(\tilde{A}(\lambda))^2\beta_2}{2(p + q\tilde{A}(\lambda))\beta_1^2}$. Ce qui correspond à une distribution d'Erlang généralisée de degré 2 avec moyenne $\frac{1}{\lambda_1} + \frac{1}{\lambda_2}$. □

4.3 Intensité faible des rappels

A présent, nous nous intéressons à la loi de probabilité (asymptotique) du nombre de clients en orbite lorsque le taux des rappels est faible ($\theta \rightarrow 0$). Le résultat obtenu est résumé dans le théorème suivant.

Théorème 4.3.1. *Si $\beta_2 < \infty$, et $\tilde{A}(\lambda) \rightarrow 0$, alors le nombre de clients en orbite est asymptotiquement Gaussien de moyenne*

$$\frac{p(1 - \tilde{A}(\lambda))}{\tilde{A}(\lambda) - [p + q\tilde{A}(\lambda)]\rho} \left(\rho + \frac{\lambda^2 p \beta_2}{2(1 - \rho)(1 - \rho q)} \right) \text{ et de va-}$$

$$\text{riance } \frac{p\lambda}{(1 - \rho^*)} \left(\beta_1 + \frac{\lambda p \beta_2}{2(1 - \rho)(1 - \rho q)} \right),$$

$$\text{où } \rho^* = \frac{p + q\tilde{A}(\lambda)}{\tilde{A}(\lambda)}\rho \text{ et } \rho = \lambda\beta_1.$$

Démonstration. Considérons $\bar{n}_o = P'(1)$, le nombre moyen de clients en orbite. Selon la propriété de décomposition stochastique du modèle considéré [17], $\bar{n}_o = \bar{n}_{q\infty} + \bar{n}_o^*$, où $\bar{n}_{q\infty} = \rho + \frac{\lambda^2\beta_2}{2(1 - \rho)}$ est le nombre moyen de clients en attente dans la file d'attente ordinaire avec

deux types de clients (prioritaires et moins prioritaires) et

$$\bar{n}_o^* = \frac{P'_0(1)}{1-\rho} = \frac{p(1-\tilde{A}(\lambda))}{\tilde{A}(\lambda) - [p+q\tilde{A}(\lambda)]\rho} \left(\rho + \frac{\lambda^2 p \beta_2}{2(1-\rho)(1-\rho q)} \right)$$

représente le nombre moyen de clients en orbite étant donné que le serveur est libre.

Supposons que $\nu = p\rho + \frac{\lambda^2 p^2 \beta_2}{2(1-\rho)(1-\rho q)}$ et $\rho^* = \frac{p+q\tilde{A}(\lambda)}{\tilde{A}(\lambda)}\rho$.

À présent, considérons

$$\begin{aligned} \bar{n}^* &= \frac{\bar{n}_o - \bar{n}_o^*}{\frac{1}{\sqrt{\tilde{A}(\lambda)}}} = \frac{\bar{n}_o - \frac{1-\tilde{A}(\lambda)}{\tilde{A}(\lambda) - [p+q\tilde{A}(\lambda)]\rho} \nu}{\frac{1}{\sqrt{\tilde{A}(\lambda)}}} \\ &= \sqrt{\tilde{A}(\lambda)} \times \bar{n}_o - \sqrt{\tilde{A}(\lambda)} \times \frac{1-\tilde{A}(\lambda)}{\tilde{A}(\lambda) - [p+q\tilde{A}(\lambda)]\rho} \nu. \end{aligned}$$

La fonction caractéristique $\mathbf{E}[\exp(it\bar{n}^*)]$ peut être donnée en termes de la fonction génératrice $P(z)$ comme suit :

$$\mathbf{E}[\exp(it\bar{n}^*)] = P\left(e^{it\sqrt{\tilde{A}(\lambda)}}\right) \exp\left\{-it\sqrt{\tilde{A}(\lambda)} \times \frac{1-\tilde{A}(\lambda)}{\tilde{A}(\lambda)(1-\rho^*)} \nu\right\},$$

où

$$P\left(e^{it\sqrt{\tilde{A}(\lambda)}}\right) = X\left(e^{it\sqrt{\tilde{A}(\lambda)}}\right) \times Y\left(e^{it\sqrt{\tilde{A}(\lambda)}}\right)$$

avec

$$X\left(e^{it\sqrt{\tilde{A}(\lambda)}}\right) = \frac{(1-\rho)\left(1 - qg\left(e^{it\sqrt{\tilde{A}(\lambda)}}\right) - pe^{it\sqrt{\tilde{A}(\lambda)}}\right)}{\left(g\left(e^{it\sqrt{\tilde{A}(\lambda)}}\right) - e^{it\sqrt{\tilde{A}(\lambda)}}\right)p}$$

et

$$Y\left(e^{it\sqrt{\tilde{A}(\lambda)}}\right) = \frac{1}{(1-\rho)} \times \frac{(1-\rho^*) \times \tilde{A}(\lambda) \times \left(g\left(e^{it\sqrt{\tilde{A}(\lambda)}}\right) - e^{it\sqrt{\tilde{A}(\lambda)}}\right)}{\tilde{A}(\lambda) \left(1 - e^{it\sqrt{\tilde{A}(\lambda)}}\right) g\left(e^{it\sqrt{\tilde{A}(\lambda)}}\right) - e^{it\sqrt{\tilde{A}(\lambda)}} \left(1 - g\left(e^{it\sqrt{\tilde{A}(\lambda)}}\right)\right)}.$$

Alors

$$\begin{aligned} \mathbf{E}[\exp(it\bar{n}^*)] &= \lim_{\theta \rightarrow 0} P\left(e^{it\sqrt{\tilde{A}(\lambda)}}\right) \exp\left\{-it\sqrt{\tilde{A}(\lambda)} \times \frac{1-\tilde{A}(\lambda)}{\tilde{A}(\lambda)(1-\rho^*)} \nu\right\} \\ &= \lim_{\theta \rightarrow 0} X\left(e^{it\sqrt{\tilde{A}(\lambda)}}\right) \times \lim_{\theta \rightarrow 0} Y\left(e^{it\sqrt{\tilde{A}(\lambda)}}\right) \exp\left\{-it\sqrt{\tilde{A}(\lambda)} \times \frac{1-\tilde{A}(\lambda)}{\tilde{A}(\lambda)(1-\rho^*)} \nu\right\}. \end{aligned}$$

On obtient

$$\lim_{\theta \rightarrow 0} X(e^{it\sqrt{\tilde{A}(\lambda)}}) = \frac{(1 - \rho) \left(1 - qg(e^{it\sqrt{\tilde{A}(\lambda)}}) - pe^{it\sqrt{\tilde{A}(\lambda)}} \right)}{\left(g(e^{it\sqrt{\tilde{A}(\lambda)}}) - e^{it\sqrt{\tilde{A}(\lambda)}} \right) p} \rightarrow 1,$$

et

$$\begin{aligned} & \lim_{\theta \rightarrow 0} Y(e^{it\sqrt{\tilde{A}(\lambda)}}) \exp \left\{ -it\sqrt{\tilde{A}(\lambda)} \times \frac{1 - \tilde{A}(\lambda)}{\tilde{A}(\lambda)(1 - \rho^*)} \nu \right\} \\ = & \lim_{\theta \rightarrow 0} \frac{1}{(1 - \rho)} \times \frac{(1 - \rho^*) \times \tilde{A}(\lambda) \times \left(g(e^{it\sqrt{\tilde{A}(\lambda)}}) - e^{it\sqrt{\tilde{A}(\lambda)}} \right)}{\tilde{A}(\lambda) \left(1 - e^{it\sqrt{\tilde{A}(\lambda)}} \right) g \left(e^{it\sqrt{\tilde{A}(\lambda)}} \right) - e^{it\sqrt{\tilde{A}(\lambda)}} \left(1 - g(e^{it\sqrt{\tilde{A}(\lambda)}}) \right)} \\ & \times \exp \left\{ -it\sqrt{\tilde{A}(\lambda)} \times \frac{1 - \tilde{A}(\lambda)}{\tilde{A}(\lambda)(1 - \rho^*)} \nu \right\}. \end{aligned}$$

Calculons

$$\begin{aligned} & \lim_{\theta \rightarrow 0} Y(e^{it\sqrt{\tilde{A}(\lambda)}}) \exp \left\{ -it\sqrt{\tilde{A}(\lambda)} \times \frac{1 - \tilde{A}(\lambda)}{\tilde{A}(\lambda)(1 - \rho^*)} \nu \right\} \\ & \times \exp \left\{ \left(e^{it\sqrt{\tilde{A}(\lambda)}} - 1 \right) \times \frac{1 - \tilde{A}(\lambda)}{\tilde{A}(\lambda)(1 - \rho^*)} \nu \right\} \\ & \times \exp \left\{ - \left(e^{it\sqrt{\tilde{A}(\lambda)}} - 1 \right) \times \frac{1 - \tilde{A}(\lambda)}{\tilde{A}(\lambda)(1 - \rho^*)} \nu \right\} \\ = & \lim_{\theta \rightarrow 0} Y(e^{it\sqrt{\tilde{A}(\lambda)}}) \times \exp \left\{ - \left(e^{it\sqrt{\tilde{A}(\lambda)}} - 1 \right) \times \frac{1 - \tilde{A}(\lambda)}{\tilde{A}(\lambda)(1 - \rho^*)} \nu \right\} \\ & \times \exp \left\{ \left(\left(e^{it\sqrt{\tilde{A}(\lambda)}} - 1 \right) - it\sqrt{\tilde{A}(\lambda)} \right) \times \frac{1 - \tilde{A}(\lambda)}{\tilde{A}(\lambda)(1 - \rho^*)} \nu \right\}. \end{aligned}$$

Posons

$$f_1 \left(\tilde{A}(\lambda) \right) = \lim_{\theta \rightarrow 0} \exp \left\{ \left(\left(e^{it\sqrt{\tilde{A}(\lambda)}} - 1 \right) - it\sqrt{\tilde{A}(\lambda)} \right) \times \frac{1 - \tilde{A}(\lambda)}{\tilde{A}(\lambda)(1 - \rho^*)} \nu \right\}$$

et

$$f_2 \left(\tilde{A}(\lambda) \right) = \lim_{\theta \rightarrow 0} Y(e^{it\sqrt{\tilde{A}(\lambda)}}) \times \exp \left\{ - \left(e^{it\sqrt{\tilde{A}(\lambda)}} - 1 \right) \times \frac{1 - \tilde{A}(\lambda)}{\tilde{A}(\lambda)(1 - \rho^*)} \nu \right\}.$$

On a $e^{it\sqrt{\tilde{A}(\lambda)}} = 1 + it\sqrt{\tilde{A}(\lambda)} - \frac{t^2}{2}\tilde{A}(\lambda)$. Alors

$$\begin{aligned} f_1(\tilde{A}(\lambda)) &= \lim_{\theta \rightarrow 0} \exp \left\{ \left(1 + it\sqrt{\tilde{A}(\lambda)} - \frac{t^2}{2}\tilde{A}(\lambda) - 1 - it\sqrt{\tilde{A}(\lambda)} \right) \times \frac{1 - \tilde{A}(\lambda)}{\tilde{A}(\lambda)(1 - \rho^*)} \nu \right\} \\ &= \exp \left\{ -\frac{t^2}{2} \frac{\nu}{(1 - \rho^*)} \right\} \\ &= \exp \left\{ -\frac{t^2}{2} \frac{1}{(1 - \rho^*)} \left(p\rho + \frac{\lambda^2 p^2 \beta_2}{2(1 - \rho)(1 - \rho q)} \right) \right\} \end{aligned}$$

et

$$\begin{aligned} f_2(\tilde{A}(\lambda)) &= \lim_{\theta \rightarrow 0} Y(e^{it\sqrt{\tilde{A}(\lambda)}}) \times \exp \left\{ -\left(e^{it\sqrt{\tilde{A}(\lambda)}} - 1 \right) \times \frac{1 - \tilde{A}(\lambda)}{\tilde{A}(\lambda)(1 - \rho^*)} \nu \right\} \\ &= \lim_{\theta \rightarrow 0} \frac{1}{(1 - \rho)} \times \frac{(1 - \rho^*) \times \tilde{A}(\lambda) \times \left(g(e^{it\sqrt{\tilde{A}(\lambda)}}) - e^{it\sqrt{\tilde{A}(\lambda)}} \right)}{\tilde{A}(\lambda) \left(1 - e^{it\sqrt{\tilde{A}(\lambda)}} \right) g(e^{it\sqrt{\tilde{A}(\lambda)}}) - e^{it\sqrt{\tilde{A}(\lambda)}} \left(1 - g(e^{it\sqrt{\tilde{A}(\lambda)}}) \right)} \\ &\quad \times \exp \left\{ -\left(e^{it\sqrt{\tilde{A}(\lambda)}} - 1 \right) \times \frac{1 - \tilde{A}(\lambda)}{\tilde{A}(\lambda)(1 - \rho^*)} \nu \right\}. \end{aligned}$$

En appliquant la règle de l'hôpital, on obtient

$$f_2(\tilde{A}(\lambda)) = \lim_{\theta \rightarrow 0} \frac{1}{(1 - \rho)} \times \frac{-p\rho \times (-g'(1) + 1)}{-g'(1)},$$

où $g'(1) = \frac{p\rho}{1 - q\rho}$. Alors

$$f_2(\tilde{A}(\lambda)) = \frac{1}{(1 - \rho)} \times \frac{-p\rho \times (-g'(1) + 1)}{-g'(1)} = \frac{1}{(1 - \rho)} \times \frac{p\rho \times (-p\rho + 1 - \rho + p\rho)}{1 - q\rho} \times \frac{1 - q\rho}{p\rho} = 1.$$

En fin,

$$\mathbf{E}[\exp(it\tilde{n}^*)] = \exp \left\{ -\frac{t^2}{2} \frac{1}{(1 - \rho^*)} \left(p\rho + \frac{\lambda^2 p^2 \beta_2}{2(1 - \rho)(1 - \rho q)} \right) \right\} \quad (4.3.1)$$

Il s'agit donc de la fonction caractéristique de la variable aléatoire Gaussienne de moyenne égale à 0 et variance $\frac{1}{(1 - \rho^*)} \left(p\rho + \frac{\lambda^2 p^2 \beta_2}{2(1 - \rho)(1 - \rho q)} \right)$. \square

4.4 Intensité forte des rappels

Nous nous intéressons au comportement asymptotique du système considéré lorsque ce dernier est soumis sous l'intensité des rappels forte.

Intuitivement, quand $\tilde{A}(\lambda) \rightarrow 1$ la distribution stationnaire de notre système converge vers un système (limite), qui coïncide avec la distribution stationnaire d'une file d'attente ordinaire avec deux types de clients. Cependant, en considérant le nombre de clients dans le système, la distribution stationnaire du modèle $(M_{\tilde{A}(\lambda)})$ est en accord avec la distribution stationnaire du système classique $M/G/1/\infty$. Pour démontrer cet argument nous utilisons la propriété de décomposition stochastique, posons :

$$\begin{aligned}\pi_j(\tilde{A}(\lambda)) &= \lim_{k \rightarrow \infty} P(N_o(\xi_k) = j) \\ \pi_j(\infty) &= \lim_{k \rightarrow \infty} P(N_{f,\infty}(\xi_k) = j) \\ \hat{\pi}_j(\tilde{A}(\lambda)) &= \lim_{k \rightarrow \infty} P(N_{\tilde{A}(\lambda)}(\xi_k) = j / C(t) = 0)\end{aligned}$$

où $\hat{\pi}_j(\tilde{A}(\lambda)) = \frac{P_0(0, z_2)}{p_0 + P_0(1)}$ et sa fonction génératrice est $Y(z)$ correspondant au nombre de clients en orbite associée au modèle $(M_{\tilde{A}(\lambda)})$ sachant que le serveur est libre. $\pi_j(\infty) = (1 - \rho) \frac{(1 - z) \tilde{B}(\lambda(1 - z))}{\tilde{B}(\lambda q(1 - z)) - z}$ est la fonction génératrice de la distribution stationnaire du nombre de clients dans le système (M_∞) (la formule de Pollaczek-Khinchin).

Théorème 4.4.1. *Les inégalités suivantes sont vérifiées*

$$2\rho p \frac{1 - \tilde{A}(\lambda)}{\tilde{A}(\lambda)} \leq \sum_{j=0}^{\infty} \left| P[N_{o,\tilde{A}(\lambda)} = j] - P[N_{f,\infty} = j] \right| \leq \frac{2\rho p}{1 - \rho} \frac{1 - \tilde{A}(\lambda)}{\tilde{A}(\lambda)}. \quad (4.4.1)$$

Par conséquent, nous avons dérivé une mesure de la proximité entre la distribution d'état stationnaire pour le système de file d'attente classique $M/G/1/\infty$ et notre système de file d'attente. L'importance de ces limites est de fournir des estimations supérieures et inférieures pour la distance entre les deux distributions.

Démonstration. De (3.5.1), nous remarquons que $\pi_j(\tilde{A}(\lambda))$ est une convolution de deux distributions $\pi_j(\infty)$ et $\hat{\pi}_j(\tilde{A}(\lambda))$, et que

$$\pi_j(\tilde{A}(\lambda)) = \sum_{k=0}^j \pi_k(\infty) \hat{\pi}_{j-k}(\tilde{A}(\lambda)). \quad (4.4.2)$$

L'expression 4.4.2 peut être réécrite sous la forme

$$\pi_j(\tilde{A}(\lambda)) - \pi_j(\infty) = \pi_j(\infty) \hat{\pi}_0(\tilde{A}(\lambda)) - \pi_j(\infty) + \sum_{k=0}^{j-1} \pi_k(\infty) \hat{\pi}_{j-k}(\tilde{A}(\lambda)).$$

Comme,

$$\begin{aligned}
\left| \pi_j \left(\tilde{A}(\lambda) \right) - \pi_j(\infty) \right| &\leq \left| \pi_j(\infty) \hat{\pi}_0 \left(\tilde{A}(\lambda) \right) - \pi_j(\infty) \right| + \sum_{k=0}^{j-1} \pi_k(\infty) \hat{\pi}_{j-k} \left(\tilde{A}(\lambda) \right). \\
&\leq \pi_j(\infty) \left(1 - \hat{\pi}_0 \left(\tilde{A}(\lambda) \right) \right) + \pi_j \left(\tilde{A}(\lambda) \right) - \pi_j(\infty) \hat{\pi}_0 \left(\tilde{A}(\lambda) \right) \\
&\leq \pi_j(\infty) \left(1 - 2\hat{\pi}_0 \left(\tilde{A}(\lambda) \right) \right) + \pi_j \left(\tilde{A}(\lambda) \right).
\end{aligned}$$

Par conséquent,

$$\sum_{j=0}^{\infty} \left| \pi_j \left(\tilde{A}(\lambda) \right) - \pi_j(\infty) \right| \leq \sum_{j=0}^{\infty} \pi_j(\infty) \left(1 - 2\hat{\pi}_0 \left(\tilde{A}(\lambda) \right) \right) + \sum_{j=0}^{\infty} \pi_j \left(\tilde{A}(\lambda) \right).$$

Vu que $\sum_{j=0}^{\infty} \pi_j(\infty) = 1$ et $\sum_{j=0}^{\infty} \pi_j \left(\tilde{A}(\lambda) \right) = 1$,

$$\sum_{j=0}^{\infty} \left| \pi_j \left(\tilde{A}(\lambda) \right) - \pi_j(\infty) \right| \leq \left(1 - 2\hat{\pi}_0 \left(\tilde{A}(\lambda) \right) \right) + 1,$$

$$\sum_{j=0}^{\infty} \left| \pi_j \left(\tilde{A}(\lambda) \right) - \pi_j(\infty) \right| \leq 2 \left(1 - \hat{\pi}_0 \left(\tilde{A}(\lambda) \right) \right),$$

où

$$\hat{\pi}_0 \left(\tilde{A}(\lambda) \right) = \frac{1}{(1-\rho)} \times \left[1 - \frac{p+q\tilde{A}(\lambda)}{\tilde{A}(\lambda)} \rho \right],$$

d'où la deuxième inégalité. En utilisant l'inégalité $|a-b| \geq a-b$, nous obtenons

$$\sum_{j=0}^{\infty} \left| \pi_j \left(\tilde{A}(\lambda) \right) - \pi_j(\infty) \right| \geq \left| \pi_0 \left(\tilde{A}(\lambda) \right) - \pi_0(\infty) \right| + \sum_{j=1}^{\infty} \left| \pi_j \left(\tilde{A}(\lambda) \right) - \pi_j(\infty) \right|$$

$$\sum_{j=0}^{\infty} \left| \pi_j \left(\tilde{A}(\lambda) \right) - \pi_j(\infty) \right| \geq \left| \pi_0 \left(\tilde{A}(\lambda) \right) - \pi_0(\infty) \right| + \pi_0(\infty) - \pi_0 \left(\tilde{A}(\lambda) \right). \quad (4.4.3)$$

De 4.4.2, nous pouvons voir que $\pi_0 \left(\tilde{A}(\lambda) \right) = \pi_0(\infty) \hat{\pi}_0 \left(\tilde{A}(\lambda) \right) < \pi_0(\infty)$. Alors l'expression 4.4.3 devient

$$\pi_0(\infty) = 1 - \rho,$$

$$\left| \pi_0 \left(\tilde{A}(\lambda) \right) - \pi_0(\infty) \right| = \pi_0(\infty) - \pi_0 \left(\tilde{A}(\lambda) \right) = \pi_0(\infty) \left(1 - \pi_0 \left(\tilde{A}(\lambda) \right) \right).$$

$$\sum_{j=0}^{\infty} \left| \pi_j \left(\tilde{A}(\lambda) \right) - \pi_j(\infty) \right| \geq 2\pi_0(\infty) - \pi_0 \left(\tilde{A}(\lambda) \right).$$

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

$$\sum_{j=0}^{\infty} \left| \pi_j \left(\tilde{A}(\lambda) \right) - \pi_j(\infty) \right| \geq 2(1 - \rho) ((1 - \pi_0 \left(\tilde{A}(\lambda) \right))).$$

□

Remarque 4.4.1. La distance $\sum_{j=0}^{\infty} \left| P \left[N_{o,\tilde{A}(\lambda)} = j \right] - P \left[N_{f,\infty} = j \right] \right|$ entre les distributions des variables aléatoires $N_{o,\tilde{A}(\lambda)}$ et $N_{f,\infty}$ diminue lorsque p se rapproche de 0 ou $\tilde{A}(\lambda)$ proche de 1.

4.5 Conclusion

Dans ce chapitre, les investigations ont été portées sur le comportement asymptotique de la variable aléatoire représentant le nombre de clients en orbite. Il a été établi que la variable en question se comporte comme une variable de loi Erlang généralisée d'ordre deux lorsque l'intensité du trafic $\rho \rightarrow 1$ et comme une variable normale lorsque le taux de rappel $\theta \rightarrow 0$. Les estimateurs des bornes supérieures et inférieures entre le modèle avec rappels considéré et celui équivalent mais sans rappels sont également obtenus.

Chapitre 5

Illustrations numériques

Ce chapitre est concerné par des illustrations numériques pour discuter des résultats obtenus dans les théorèmes établis dans le chapitre précédent.

5.1 Cas du trafic intense

Dans un premier temps, considérons 3.4.16. Il est facile de voir que $\lim_{z \rightarrow 1} P(z)$ est le produit de deux transformées de Laplace-Stieltjes de deux distributions correspondant à la somme de deux variables indépendants Exponentielles :

$$\tilde{F}(s) = \frac{1}{1 + \frac{s\beta_2}{2\beta_1^2}} \times \frac{1}{1 + s \left[\frac{p(1 - \tilde{A}(\lambda))\tilde{A}(\lambda)}{(p + q\tilde{A}(\lambda))^2} + \frac{(\tilde{A}(\lambda))^2\beta_2}{2(p + q\tilde{A}(\lambda))\beta_1^2} \right]}.$$

La moyenne de la variable aléatoire ayant la transformée de Laplace-Stieltjes $\tilde{F}(s)$ est $m_1 = (-1)\tilde{F}'(s=0)$. D'où,

$$m_1 = \frac{\beta_2}{2\beta_1^2} \left[1 + \frac{(\tilde{A}(\lambda))^2}{(p + q\tilde{A}(\lambda))} \right] + \frac{p(1 - \tilde{A}(\lambda))\tilde{A}(\lambda)}{(p + q\tilde{A}(\lambda))^2}.$$

Pour illustrer le comportement du nombre moyen de clients en orbite, sous un trafic lourd $\rho^* \rightarrow 1$, nous supposons que le taux de service $\gamma = \frac{1}{\beta_1} = 1$ et tenons en compte des distributions de temps de service suivantes :

- Exponentielle (E) : $B_E(x) = 1 - e^{-\gamma x}$, $x \geq 0$, $\beta_1 = 1$ et $\beta_2 = 2$;
- Erlang d'ordre 2 (E_2) : $B_{E_2}(x) = 1 - e^{-2\gamma x} - 2\gamma e^{-2\gamma x}$, $x \geq 0$, $\beta_1 = 1$ et $\beta_2 = 1.5$.

Nous avons donc les modèles $M/M/1$ et $M/E_2/1$ avec rappels. Pour les durées inter-rappels, nous choisissons :

- Exponentielle (E) : $A_E(x) = 1 - e^{-\theta x}$, $x \geq 0$;

- Erlang d'ordre 2 (E_2) : $A_{E_2}(x) = 1 - e^{-2\theta x} - 2\theta x e^{-2\theta x}$, $x \geq 0$;
- Hyperexponentielle d'ordre 2 (H_2) : $A_{H_2}(x) = 1 - \zeta_1 e^{-\theta_1 x} - \zeta_2 e^{-\theta_2 x}$, $x \geq 0$, $\frac{\zeta_1}{\theta_1} + \frac{\zeta_2}{\theta_2} = \frac{1}{\theta}$, $\zeta_1 + \zeta_2 = 1$, $\theta_1 = 2\zeta_1\theta$, $\theta_2 = 2\zeta_2\theta$.

$\theta = 2$	λ	Temps de rappels E		Temps de rappels E_2		Temps de rappels H_2	
		$\tilde{A}_E(\lambda)$	ρ^*	$\tilde{A}_{E_2}(\lambda)$	ρ^*	$\tilde{A}_{H_2}(\lambda)$	ρ^*
$p = 0.25$	0.1	0.952 38	0.101 25	0.951 81	0.101 27	0.953 07	0.101 23
	0.2	0.909 09	0.205 00	0.907 03	0.205 13	0.912 79	0.204 78
	0.3	0.869 57	0.311 25	0.865 33	0.311 67	0.877 09	0.310 51
	0.4	0.833 33	0.42000	0.826 45	0.42100	0.845 02	0.418 34
	0.5	0.80000	0.531 25	0.790 12	0.533 2	0.815 89	0.528 21
	0.6	0.769 23	0.64500	0.756 14	0.648 38	0.789 22	0.640 06
	0.7	0.740 74	0.761 25	0.724 31	0.766 61	0.764 62	0.753 87
	0.8	0.714 29	0.88000	0.694 44	0.888 00	0.741 81	0.869 61
$p = 0.5$	0.1	0.952 38	0.102 5	0.951 81	0.102 53	0.953 07	0.102 46
	0.2	0.909 09	0.2100	0.907 03	0.210 25	0.912 79	0.209 55
	0.3	0.869 57	0.322 5	0.865 33	0.323 34	0.877 09	0.321 02
	0.4	0.833 33	0.4400	0.826 45	0.442 00	0.845 02	0.436 68
	0.5	0.80000	0.562 5	0.790 12	0.566 41	0.815 89	0.556 41
	0.6	0.769 23	0.6900	0.756 14	0.696 75	0.789 22	0.680 12
	0.7	0.740 74	0.822 5	0.724 31	0.833 22	0.764 62	0.807 74
	0.8	0.714 29	0.9600	0.694 44	0.976 00	0.741 81	0.939 22
$p = 0.75$	0.1	0.952 38	0.103 75	0.951 81	0.103 80	0.953 07	0.103 69
	0.2	0.909 09	0.21500	0.907 03	0.215 38	0.912 79	0.214 33
	0.3	0.869 57	0.333 75	0.865 33	0.335 02	0.877 09	0.331 53
	0.4	0.833 33	0.46000	0.826 45	0.463 00	0.845 02	0.455 02
	0.5	0.80000	0.593 75	0.790 12	0.599 61	0.815 89	0.584 62
	0.6	0.769 23	0.735 00	0.756 14	0.745 13	0.789 22	0.720 19
	0.7	0.740 74	0.883 75	0.724 31	0.899 83	0.764 62	0.861 62

Tableau 1 : L'intensité de trafic ρ^* pour les modèles $M/M/1$ et $M/E_2/1$ avec rappels

Le tableau 1 présente les valeurs numériques de l'intensité du trafic ρ^* calculé selon

$$\rho^* = \frac{(p + (1-p)\tilde{A}(\lambda))}{\tilde{A}(\lambda)} \times \lambda\beta_1, \text{ pour différentes valeur des taux d'arrivée } \lambda, \text{ des probabilités de priorité } p \text{ et}$$

ainsi que pour les différentes distributions du temps de rappels. Nous avons supposé que l'intensité des rappels $\theta = 2$. Pour taux de service $\gamma = \frac{1}{\beta_1} = 1$, nous obtenons des valeurs très proches de ρ^* pour les deux modèles considérés.

Les figures (5.1 à 5.3) montrent le comportement asymptotique de m_1 du nombre moyen de clients en orbite, par rapport au taux d'arrivée λ pour le modèle $M/M/1$ avec rappels; tandis que les figures (5.4 à 5.6) sont concernées par le modèle $M/E_2/1$ avec rappels. Les résultats numériques sont obtenus pour les différentes valeurs de la probabilité p ($p = 0.25$, $p = 0.5$ and $p = 0.75$) ainsi que pour les distributions de temps de rappels mentionnées ci-dessus : Exponentielle (courbe bleue), Erlang d'ordre 2 (courbe rouge) et Hyperexponentielle d'ordre 2 (courbe verte). Il est facile de voir que la probabilité p a peu d'effet sur les valeurs numériques de m_1 .

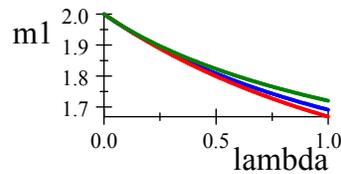


FIG. 5.1 – Modèle $M/M/1$ avec rappel. $\theta = 2$, $p = 0.25$.

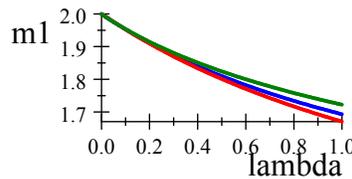


FIG. 5.2 – Modèle $M/M/1$ avec rappel. $\theta = 2$, $p = 0.5$.

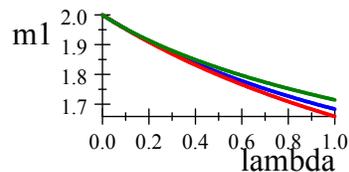


FIG. 5.3 – Modèle $M/M/1$ avec rappel. $\theta = 2$, $p = 0.75$.

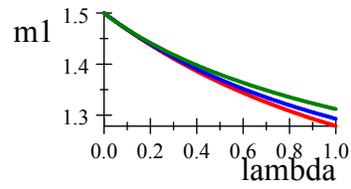


FIG. 5.4 – Modèle $M/E_2/1$ avec rappel. $\theta = 2$, $p = 0.25$.

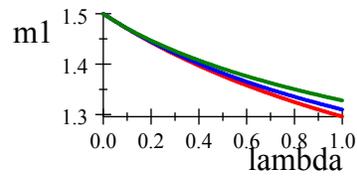


FIG. 5.5 – Modèle $M/E_2/1$ avec rappel. $\theta = 2$, $p = 0.5$.

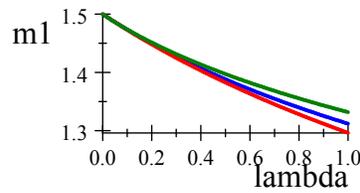


FIG. 5.6 – Modèle $M/E_2/1$ avec rappel. $\theta = 2$, $p = 0.75$.

Nous avons également examiné les effets de l'intensité des rappels θ sur le comportement asymptotique du nombre moyen de clients en orbite quand $\rho^* \rightarrow 1$.

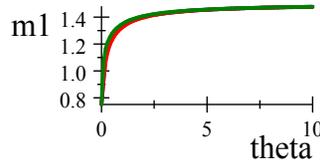


FIG. 5.7 – Modèle $M/E_2/1$ avec rappels. $p = 0.5, \lambda = 0.4$.

A partir de la figure 5.7, nous pouvons voir que l'augmentation de l'intensité des rappels θ entraîne une augmentation importante de la mesure de performance considérée, et ceci pour les trois distributions des temps de rappels considérées : Exponentielle (courbe bleue), Erlang d'ordre 2 (courbe rouge) et Hyperexponentielle de degré 2 (courbe verte).

5.2 Cas de l'intensité faible des rappels

Dans cette application, nous nous intéressons au résultat établi dans le théorème 4.3.1. Nous étudions le comportement de la moyenne et de la variance du nombre de clients en orbite (lorsque l'intensité des rappels ($\theta \rightarrow 0$) et $\lambda = 0.3; \beta_1 = 1$, le modèle $M/E_2/1$ avec rappel est pris en considération. Les figures (5.8 à 5.9) présente les valeurs numériques des paramètres gaussiens quand le temps de rappel est distribué selon la loi exponentielle (courbe bleue), loi d'Erlang d'ordre 2 (courbe rouge) et hyperexponentielle d'ordre 2 (vert courbe)

La convergence de la variance vers une valeur minimale est plus rapide dans le cas d'une distribution hyperexponentielle d'ordre deux pour le temps des rappels

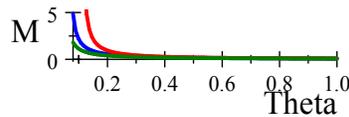


FIG. 5.8 – Modèle $M/E_2/1$ avec rappels. $p = 0.5, \lambda = 0.3, \beta_1 = 1$.

5.3 Cas de l'intensité forte des rappels

Dans les tableaux 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 ∞ avec $\lambda = 0.3; \beta_1 = 1$, et $p = 0.5$. Nous

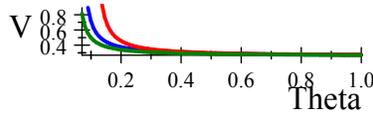


FIG. 5.9 – Modèle $M/E_2/1$ avec rappels. $p = 0.5$, $\lambda = 0.3$, $\beta_1 = 1$.

observons que dans le cas où le temps de service suit la loi Exponentielle, la distribution du nombre de clients dans le système $M/G/1$ avec rappels, priorité relative et orbite FCFS converge plus rapidement vers la distribution du nombre de clients dans le système classique $M/G/1/\infty$.

M/E/1			
θ	$\tilde{A}_E(\lambda)$	BI	BS
1	0.909 09	0.03	$4.285\ 8 \times 10^{-2}$
5	0.943 40	$1.799\ 9 \times 10^{-2}$	$2.571\ 2 \times 10^{-2}$
10	0.970 87	$9.001\ 2 \times 10^{-3}$	$1.285\ 9 \times 10^{-2}$
10^2	0.997 01	$8.996\ 9 \times 10^{-4}$	$1.285\ 3 \times 10^{-3}$
10^3	0.999 7	$9.002\ 7 \times 10^{-5}$	$1.286\ 1 \times 10^{-4}$
10^4	0.999 97	$9.000\ 3 \times 10^{-6}$	$1.285\ 8 \times 10^{-5}$
10^5	1.000	0	0
M/H/1			
θ	$\tilde{A}_H(\lambda)$	BI	BS
1	0.769 23	0.09	0.128 57
5	0.943 40	$1.799\ 9 \times 10^{-2}$	$2.571\ 2 \times 10^{-2}$
10	0.970 87	$9.001\ 2 \times 10^{-3}$	$1.285\ 9 \times 10^{-2}$
10^2	0.997 01	$8.996\ 9 \times 10^{-4}$	$1.285\ 3 \times 10^{-3}$
10^3	0.999 7	$9.002\ 7 \times 10^{-5}$	$1.286\ 1 \times 10^{-4}$
10^4	0.999 7	$9.000\ 3 \times 10^{-6}$	$1.285\ 8 \times 10^{-5}$
10^5	1	0	0
M/E ₂ /1			
θ	$\tilde{A}_{E_2}(\lambda)$	BI	BS
1	0.756 14	$9.675\ 2 \times 10^{-2}$	0.138 22
5	0.942 60	$1.826\ 9 \times 10^{-2}$	$2.609\ 8 \times 10^{-2}$
10	0.970 66	$9.068\ 1 \times 10^{-3}$	$1.295\ 4 \times 10^{-2}$
10^2	0.997 01	$8.996\ 9 \times 10^{-4}$	$1.285\ 3 \times 10^{-3}$
10^3	0.999 7	$9.002\ 7 \times 10^{-5}$	$1.286\ 1 \times 10^{-4}$
10^4	0.999 97	$9.000\ 3 \times 10^{-6}$	$1.285\ 8 \times 10^{-5}$
10^5	1.000	0	0

Tableaux 2 – Borne inf et borne sup des inégalités vérifiées quand $\tilde{A}(\lambda)$ tend vers 1 (ou quand θ tend vers ∞).

Quand $\tilde{A}(\lambda) \rightarrow 1$ la distribution stationnaire de notre système converge vers un système limite, qui coïncide avec la distribution stationnaire d'un file d'attente de priorité à deux niveaux. Cependant, en ce qui concerne le nombre total de clients dans le système, sa distribution stationnaire converge la distribution stationnaire du système classique M /G / 1/∞. Alors nous obtenons le tableau suivant donnant les longueurs des intervalles pour chacune des distributions du temps de service.

	M/E/1	M/H/1	M/E ₂ /1
θ	BS-BI	BS-BI	BS-BI
1	1.2858×10^{-2}	0.03857	4.1468×10^{-2}
5	7.713×10^{-3}	7.713×10^{-3}	7.829×10^{-3}
10	3.8578×10^{-3}	3.8578×10^{-3}	3.8859×10^{-3}
10^2	3.8561×10^{-4}	3.8561×10^{-4}	3.8561×10^{-4}
10^3	3.8583×10^{-5}	3.8583×10^{-5}	3.8583×10^{-5}
10^4	3.8577×10^{-6}	3.8577×10^{-6}	3.8577×10^{-6}
10^5	0	0	0

Tableau 3 – Longueur des intervalles de convergence de la distribution $\pi_j(\tilde{A}(\lambda))$ vers la distribution $\pi_j(\infty)$

Nous déduisons que la loi Exponentielle du temps de service est la mieux adaptée à notre modèle quand le taux de rappels augmente.

5.4 Effet des paramètres essentiels sur les principales caractéristiques de performance

Pour illustrer l'effet des paramètres, tels que la distribution du temps inter-rappels et la politique de priorité, sur les principales caractéristiques de performance, nous considérons dans ce paragraphe, le modèle avec $\lambda = 0.3$ et le temps de service exponentiellement distribué avec la moyenne $\beta_1 = 1$. Les valeurs des paramètres sont choisies de manière à ce que la condition de stabilité soit satisfaite.

5.5 Effet de la distribution des temps inter-rappels

Nous supposons que q est égal à 0,4 et les temps de rappels sont distribués selon une loi d'Erlang d'ordre k de paramètre θ , telle que $\tilde{A}(s) = \left(\frac{\theta}{\theta + s}\right)^k$. Rappelons que la variable d'Erlang peut être exprimée comme la somme de k variables indépendantes exponentiellement distribuées de taux θ . Nous présentons trois courbes qui correspondent respectivement à $k = 1, 5$ et 10.

Dans la figure 5.10, la probabilité que le système soit occupé est tracée par rapport au taux des rappels θ . En effet, le facteur d'occupation du système diminue avec l'accroissement du taux de rappels θ et le décroissement de k . Pratiquement les mêmes effets sont visibles sur la figure 5.11, qui illustre le comportement du nombre moyen de clients dans le système $E[N]$ en fonction de θ .

De plus, nous observons que, lorsque θ se rapproche de la condition de stabilité, la moyenne de la taille du système tend vers l'infini (à cause de l'instabilité du système) par conséquent, la probabilité que le système soit occupé converge vers 1.

Remarquons que le facteur d'occupation et le nombre moyen de clients dans le système sont asymptotiquement cohérents (quand $\theta \rightarrow \infty$) avec les caractéristiques du modèle de files d'attente classique $M/G/1/\infty$ telles que $\rho = 0.3$ et $E[N] = 0.428571$, d'où les résultats limites

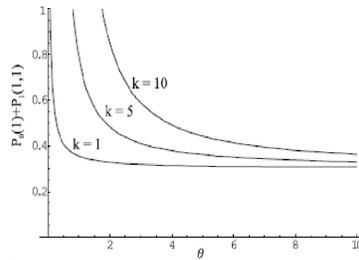


FIG. 5.10 – Le facteur d'occupation par rapport à θ .

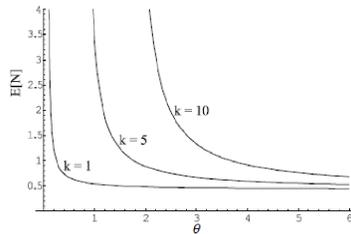


FIG. 5.11 – Le nombre moyen de clients dans le système par rapport à θ .

$$\lim_{\tilde{A}(\lambda) \rightarrow 1} = [P_0(1) + P_1(1, 1)] = \rho$$

et

$$\lim_{\tilde{A}(\lambda) \rightarrow 1} E[N] = \rho + \frac{\lambda^2 \beta_2}{2(1 - \rho)}$$

En outre, nous notons que le facteur d'occupation et la taille moyenne du système dans le modèle de files d'attente $M/G/1/\infty$ classique sont une borne inférieure pour les caractéristiques correspondantes de notre système de files d'attente.

5.5.1 Effet de la politique de priorité

Supposons que les temps inter-rappels suivent une loi Exponentielle de taux $\theta > 0$ (politique des rappels constante). L'influence du paramètre q sur le facteur d'occupation est représentée dans la figure 5.12. La courbe la plus élevée dans la figure 5.13 correspond à la valeur la plus faible de θ .

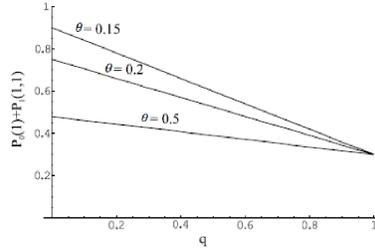


FIG. 5.12 – Le facteur d'occupation par rapport à q

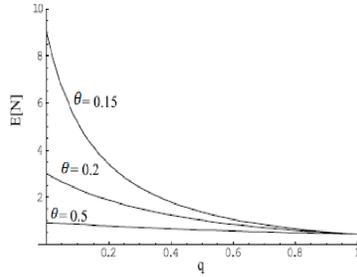


FIG. 5.13 – Le nombre moyen du client dans le système par rapport à q .

Par intuition la probabilité que le système soit occupé augmente avec l'accroissement de θ . A partir de ce graphique, nous voyons que le facteur d'occupation diminue avec l'augmentation des valeurs de q , ce qui correspond également aux attentes intuitives. La même discussion s'applique à la tendance des valeurs de $E[N]$. (Voir figure 5.13).

5.5.2 Conclusion

Pour le modèle $M/G/1$ avec rappels, priorité relative, orbite FCFS et temps de rappels général, l'analyse du comportement asymptotique de la distribution (en régime stationnaire) du nombre de clients en orbite, qui n'a pas révélé la nature de la variable étudiée, nous a permis d'approximer cette dernière par des distributions dont les caractéristiques sont connues. Nous avons obtenu les résultats suivants :

- Dans le cas d'un trafic intense, la distribution du nombre de clients en orbite converge vers la distribution généralisée d'Erlang d'ordre deux, alors le nombre moyen de clients en orbite est :

$$\bar{n}_o = \frac{2\beta_1^2}{\beta_2} + \left[\frac{p(1 - \tilde{A}(\lambda))\tilde{A}(\lambda)}{(p + q\tilde{A}(\lambda))^2} + \frac{(\tilde{A}(\lambda))^2\beta_2}{2(p + q\tilde{A}(\lambda))\beta_1^2} \right]^{-1}.$$

- Dans le cas d'un taux des rappels faible, la distribution du nombre de clients en orbite converge vers une distribution Gaussienne. Alors le nombre moyen de clients en orbite est le suivant :

$$\bar{n}_o = \frac{p(1 - \tilde{A}(\lambda)) \left(\rho + \frac{\lambda^2 p \beta_2}{2(1 - \rho)(1 - \rho q)} \right)}{\tilde{A}(\lambda) - [p + q\tilde{A}(\lambda)] \rho}.$$

- Dans le cas d'une intensité des rappels forte, la distance $\sum_{j=0}^{\infty} |P[N_{o,\tilde{A}(\lambda)} = j] - P[N_{f,\infty} = j]|$ entre les distributions des variables aléatoires N_o et $N_{f,\infty}$ diminue lorsque p se rapproche de 0 ou $\tilde{A}(\lambda)$ est proche de 1.

Conclusion Générale et Perspectives

Dans ce travail nous nous sommes intéressés aux systèmes de files d'attente de type M/G/1 avec rappels de loi générale, priorité relative et orbite FCFS :

Dans un premier temps, nous avons passé en revue certains résultats consacrés aux systèmes de files d'attente de type M/G/1 avec temps inter-rappels générale et priorité relative ainsi que sur les approches utilisées par les chercheurs dans leurs investigations. Une attention particulière a été accordée à la définition des politiques d'accès au serveur.

Dans un deuxième temps, nous avons présenté une analyse mathématique du système de files d'attente M/G/1 avec rappels de loi générale, priorité relative et orbite FCFS. A cet effet, la chaîne de Markov induite est décrite, la condition d'ergodicité est établie à l'aide du critère de Foster et de la condition de Kaplan, la distribution stationnaire de l'état du système est obtenue en appliquant la méthode des variables supplémentaires ainsi que les mesures de performance sont déduites. Enfin, le comportement asymptotique du nombre de clients en orbite et celui dans le système a été étudié.

A cet effet, les lois de probabilité 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 établies. Enfin, le comportement asymptotique du nombre de clients dans le système a été étudié. Nos investigations théoriques sont appuyées par des illustrations numériques.

Par conséquent les conclusions suivantes sont tirées :

1. Sous un régime chargé, la distribution stationnaire du nombre de clients en orbite converge vers la loi Erlang généralisée d'ordre deux.
2. Sous un taux de rappels faible, la distribution stationnaire du nombre de clients en orbite converge vers la loi Gaussienne.
3. Sous un taux de rappels élevé, la proximité avec le système M/G/1 avec rappels de loi générale, priorité relative et orbite FCFS a été estimée.

De ce fait, nous avons établi que quel que soit le modèle avec rappels de type M/G/1 gouverné par la politique FCFS et distribution générale du temps inter-rappels, les conclusions ci-dessus s'appliquent avec succès.

Dans la continuation de ce travail, les résultats obtenus permettent d'envisager de nouvelles perspectives de recherche :

Nous nous intéressons aux modèles avec recherche en orbite, où le serveur après avoir terminer un service peut aller chercher un client en orbite (sous-entendu que la file d'attente prioritaire soit vide).

Considérer des modèles plus complexes par exemple, en rajoutant le phénomène de l'impatience, ces systèmes peuvent être appliquer en actuariat.

Annexe A

Cette annexe présente des propriétés supplémentaires et des distributions connexes qui sont souvent utilisées dans les applications. Des autres distributions que celles mentionnées ici, mais qui sont parfois utilisées dans des applications, peuvent être trouvées dans des textes standard en statistique.

A.1 Distribution Hyperexponentielle (HE)

Soient les variables aléatoires $\{Z_1, Z_2, \dots\}$ être distribuées comme suit

$$F(x) = 1 - \sum_{i=1}^K p_i e^{-\lambda_i x}; \quad 0 \leq x < \infty$$

$\lambda_i > 0$, pour tout i pour lequel ;

$$0 \leq p_i \leq 1; \quad \sum_{i=1}^K p_i = 1. \quad (\text{A.1.1})$$

On a

$$E(Z_n) = \sum_{i=1}^K \frac{p_i}{\lambda}$$

et sa transformation de Laplace

$$\psi(\theta) = \sum_{i=1}^K p_i \left(\frac{\lambda_i}{\theta + \lambda_i} \right) \quad (\text{A.1.2})$$

aussi

$$E[Z_n^2] = \sum_{i=1}^K \frac{2p_i}{\lambda_i^2} \quad \text{et} \quad CV(Z_n) = \left[\frac{2 \sum_{i=1}^K p_i / \lambda_i^2}{\left(\sum_{i=1}^K p_i / \lambda_i \right)^2} - 1 \right]^{\frac{1}{2}}.$$

Cette distribution est générée si les événements tombent dans des classes identifiables $(1, 2, \dots, k)$ et qu'un événement appartenant à la classe i arrive avec probabilité p_i avec un temps d'inter occurrence qui est exponentiel avec une moyenne de $\frac{1}{\lambda_i}$. En fonction des valeurs de p_i, λ_i et des valeurs possibles de i , une grande variété de distributions peut être générée.

Afin de conserver la même moyenne $\frac{1}{\lambda}$, la forme suivante de la distribution HE peut être utilisée.

$$F(x) = 1 - \sum_{i=1}^K p_i e^{-K p_i \lambda x} \quad (x \geq 0) \quad \text{A.3.3} \quad (\text{A.1.3})$$

$\lambda > 0$, $0 \leq p_i \leq 1$ et $\sum_{i=1}^K p_i = 1$.

Alors

$$E[Z_n] = \frac{1}{\lambda}$$

$$E[Z_n^2] = \frac{2}{(K\lambda)^2} \sum_{i=1}^K \left(\frac{1}{p_i}\right)$$

$$CV[Z_n] = \left[\left(\sum_{i=1}^K \frac{1}{p_i} \right) \frac{2}{K^2} - 1 \right] \frac{1}{2}$$

La valeur de K couramment utilisée dans les applications est 2.

A.2 Distribution Erlang (E_k)

Soient les variables aléatoires $\{Z_1, Z_2, \dots\}$ être distribuées comme suit

$$\begin{aligned} F(x) &= \int_0^x e^{-\lambda y} \frac{\lambda^k y^{k-1}}{(k-1)!} dy \quad 0 \leq x \leq \infty; \quad \lambda > 0 \\ &= 1 - \sum_{r=0}^{k-1} e^{-\lambda x} \frac{(\lambda x)^r}{r!}. \end{aligned} \quad (\text{A.2.1})$$

On a

$$E[Z_n] = \frac{k}{\lambda}$$

$$\psi(\theta) = \left(\frac{\lambda}{\theta + \lambda} \right)^k$$

et

$$E[Z_n^2] = \left(1 + \frac{1}{k}\right) \frac{k^2}{\lambda^2} \quad \text{et} \quad CV[Z_n] = \frac{1}{\sqrt{k}}.$$

La distribution $F(x)$ est une distribution à deux paramètres, et est communément connue sous le nom de distribution Erlang (A.K. Erlang a démontré son utilisation dans l'analyse de la congestion du système téléphonique), ou la distribution gamma, ou la distribution Poisson type 3 avec des valeurs intégrales pour le paramètre k . (C'est aussi un cas particulier de la distribution χ^2).

A.3 Distributions mixtes d'Erlang

La distribution HE de la section A.3 ci-dessus est obtenue en utilisant un mélange fini de distributions exponentielles. De manière similaire, afin de fournir une polyvalence, nous pouvons obtenir des distributions mixtes Erlang.

(a) Constante λ ; variant k (k_1, k_2, \dots, k_n). Soit

$$F(x) = \int_0^x \sum_{i=1}^N p_i e^{-k_i \lambda y} \frac{(k_i \lambda)^{k_i} y^{k_i-1}}{(k_i-1)!} dy \quad 0 \leq x \leq \infty; \quad \lambda > 0. \quad \text{A.51} \quad (\text{A.3.1})$$

On a

$$E[Z_n] = \frac{1}{\lambda}$$

$$\psi(\theta) = \sum_{i=1}^N p_i \left(\frac{k_i \lambda}{\theta + k_i \lambda} \right)^{k_i} \quad (\text{A.3.2})$$

et

$$E[Z_n^2] = \frac{1}{\lambda^2} \sum_{i=1}^N p_i \left(1 + \frac{1}{k_i} \right) \quad \text{et} \quad CV[Z_n] = \left[\sum_{i=1}^N \frac{p_i}{k_i} \right]^{\frac{1}{2}}.$$

Ceci ajoute une autre dimension de généralité à la distribution d'Erlang. Il a été démontré par plusieurs auteurs que cette distribution se rapproche très bien de presque toutes les distributions d'intérêt pratique. Une limite finie pour les valeurs de N a également été jugée satisfaisante. (Voir Luchak (1956)).

(b) Les deux paramètres λ et k variant

$$F(x) = \int_0^x \sum_{i=1}^N p_i e^{-k_i \lambda_i y} \frac{(k_i \lambda_i)^{k_i} y^{k_i-1}}{(k_i-1)!} dy. \quad (\text{A.3.3})$$

Cette forme générale admet à la fois la distribution hyperexponentielle et Erlang comme cas spéciaux.

HE : $k_i = 1$ pour $i = 1, 2, \dots, N$;

Erlang Mixte: $k_i = \lambda$ pour $i = 1, 2, \dots, N$.

En supposant que le coefficient de variation soit une mesure fournissant une représentation adéquate du modèle, les distributions d'*Erlang* (avec $CV \leq 1$) et d'*HE* (avec $CV \geq 1$) offrent un spectre vide de choix aux fins de la sélection du modèle. Dans le modèle d'*Erlang*, le CV est diminué en augmentant la valeur du paramètre k et dans le modèle *HE* avec $N = 2$, CV est augmenté en éloignant p_1 et p_2 de $1/2$.

A.4 Une Distribution Générale

Soit $F(x)$ une fonction de distribution continue, avec fonction de densité de probabilité $f(x)$. Nous avons

$$f(x) = -\frac{d}{dx} [1 - F(x)]. \quad (\text{A.4.1})$$

En utilisant le concept de fonction de risque introduit en a11, nous avons

$$\begin{aligned} h(x) &= \frac{1}{1 - F(x)} \left\{ -\frac{d}{dx} [1 - F(x)] \right\} \\ &= -\frac{d}{dx} \ln [1 - F(x)] \end{aligned} \quad (\text{A.4.2})$$

Intégrer $h(x)$, nous trouvons

$$\begin{aligned} \int_0^x h(y) dy &= -\ln [1 - F(x)] \\ F(x) &= 1 - e^{-\int_0^x h(y) dy} \end{aligned}$$

et

$$f(x) = h(x) e^{-\int_0^x h(y) dy} \quad (\text{A.4.3})$$

Annexe B

Processus de Markov

B.1 Equations de Kolmogorov

Soit $\{X(t), t \in T\}$ un processus de Markov dans un temps homogène

$$P_{i,j}(t) = P[X(t) = j / X(0) = i]. \quad (\text{B.1.1})$$

Il existe deux types d'équations différentielles pour la détermination de $P_{ij}(t)$ dans les processus de Markov. Ce sont des équations de Kolmogorov *forward* et des équations de Kolmogorov *backward*. Les équations de Kolmogorov *forward* sont celles couramment utilisées dans les applications en raison de leur structure pratique, même si les équations *backward* sont considérées comme plus fondamentales en raison de la nature des propriétés limitantes utilisées dans leur dérivation. Afin de dériver ces équations, nous procédons comme suit. Dans un processus de Markov dans un temps homogène

$$P_{ij}(s, t) = \sum_{k \in S} P_{ik}(s, u) P_{kj}(u, t); \quad s < u < t.$$

représentant la relation Chapman-Kolmogorov, peut être écrit comme

$$P_{ij}(s + t) = \sum_{k \in S} P_{ik}(t) P_{kj}(s)$$

met $s = \Delta t$, alors

$$P_{ij}(t + \Delta t) = \sum_{k \in S} P_{ik}(t) P_{kj}(\Delta t).$$

En soustrayant $P_{ij}(t)$ des deux côtés de l'équation et en divisant par Δt ,

$$\frac{P_{ij}(t + \Delta t) - P_{ij}(t)}{\Delta t} = \sum_{k \neq j} \frac{P_{ik}(t) P_{kj}(\Delta t)}{\Delta t} + P_{ij}(t) \frac{P_{jj}(\Delta t)}{\Delta t}.$$

Soit $\Delta t \rightarrow 0$; on a

$$P'_{ij}(t) = -\lambda_{jj} P_{ij}(t) + \sum_{k \neq j} \lambda_{kj} P_{ik}(t). \quad (\text{B.1.2})$$

L'équation (B.1.2) pour $i, j \in S$ est connue sous le nom d'équation de *Kolmogorov forward*.

$$P'(t) = P(t)A \quad (\text{B.1.3})$$

où

$$A = \begin{bmatrix} -\lambda_{00} & \lambda_{01} & \lambda_{02} & \cdots \\ \lambda_{10} & -\lambda_{11} & \lambda_{12} & \cdots \\ \lambda_{20} & \lambda_{20} & -\lambda_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

La probabilité de transition $P_{ij}(t)$ peut être déterminée en résolvant ces équations différentielles avec la condition limite $P(0) = I$.

Les équations de Kolmogorov *Backward* peuvent être obtenues de manière similaire, en commençant par la relation

$$P_{ij}(t + \Delta t) = \sum_{k \in S} P_{ik}(t) P_{kj}(\Delta t).$$

L'équation matricielle correspondante peut être donnée comme

$$P'(t) = AP(t).$$

Formellement, la solution pour les deux ensembles d'équations peut être donnée comme

$$P(t) = e^{At} = I + \sum_{n=1}^{\infty} A^n \frac{t^n}{n!}. \quad (\text{B.1.4})$$

B.2 Le Processus de Poisson

Nous montrons ici comment les équations de Kolmogorov peuvent être utilisées pour déterminer la distribution de probabilité de transition du processus de Poisson. Dans la section A.1 de l'annexe A, nous avons listé les propriétés suivantes :

1. Les événements se produisant dans des intervalles de temps sans chevauchement sont indépendants les uns des autres.
2. Il y a une constante λ telle que les probabilités d'apparition d'événements dans un petit intervalle de longueur Δt sont données comme suit :

$$P \{ \text{Nombre d'événements survenus dans } (t, t + \Delta t] = 0 \} = 1 - \lambda \Delta t + o(\Delta t).$$

$$P \{ \text{Nombre d'événements survenus dans } (t, t + \Delta t] = 1 \} = \lambda \Delta t + o(\Delta t).$$

$$P \{ \text{Nombre d'événements survenus dans } (t, t + \Delta t] > 1 \} = o(\Delta t).$$

où $o(\Delta t)$ est telle que $o(\Delta t) / \Delta t \rightarrow 0$ comme $\Delta t \rightarrow 0$.

En utilisant les notations et les équations développées pour les processus de Markov, dans ce contexte, nous avons

$$P'_{ij}(0) = \lambda; \quad P'_{ii}(0) = -\lambda$$

résultant dans la matrice du générateur

$$A = \begin{bmatrix} -\lambda & \lambda & 0 & 0 & 0 & \dots \\ 0 & -\lambda & \lambda & 0 & 0 & \dots \\ 0 & 0 & -\lambda & \lambda & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (\text{B.2.1})$$

Le processus de Poisson est un processus de comptage dont la valeur initiale est 0, c'est-à-dire $X(0) = 0$. Ecrire $P_{0n}(t) = P_n(t)$ pour plus de commodité et noter que $\mathbf{P}(t) = (P_0(t), P_1(t), \dots)$ et $\mathbf{P}'(t) = (P'_0(t), P'_1(t), \dots)$ les équations individuelles dans (B.1.2) peuvent être écrites comme

$$\begin{aligned} P'_0(t) &= -\lambda P_0(t) \\ P'_n(t) &= -\lambda P_n(t) + \lambda P_{n-1}(t) \quad n > 0 \end{aligned}$$

avec $P_0(0) = 1$ et $P_n(0) = 0$ pour $n > 0$.

Résoudre ces équations différentielles nous obtenons

$$P_n(t) = e^{-\lambda t} \frac{(\lambda t)^n}{n!} \quad n = 0, 1, 2, \dots \quad (\text{B.2.2})$$

La solution aux équations en (B.1.2) n'est pas toujours facile à déterminer. Dans le cas du processus de Poisson, à cause de la structure bidiagonale de A et de l'élément constant λ , les équations différentielles pourraient être résolues à l'aide des méthodes standard. Lorsque de telles simplifications ne sont pas disponibles, dans des cas plus simples, nous pouvons utiliser des transformées de Laplace et des PGF dans leurs solutions. Lorsque A est fini et diagonalisable, la méthode des valeurs propres peut être utilisée pour déterminer la solution sous la forme (B.1.4). De plus, il existe des méthodes de calcul pour obtenir des solutions à partir des équations différentielles directement. (Voir Bailey (1964), Stewart (1994) et Bhat et Miller (2002)).

$\lambda_{ij}, j \neq i$ est le taux instantané pour la transition $i \rightarrow j$. Nous savons aussi que $\sum_{j \neq i} \lambda_{ij} = \lambda_{ii}$

Cela signifie que λ_{ii} est également la somme de tous les taux de transition instantanée hors de l'état i .

Cela nous permet d'interpréter $\frac{1}{\lambda_{ii}}$ comme la durée moyenne du processus dans l'état i pendant une visite. La durée pendant laquelle le processus reste dans un état au cours d'une visite est connue sous le nom de temps de séjour dans cet état. Ce temps de séjour du processus de Markov dans l'état i a été montré pour avoir une distribution Exponentielle avec moyenne $\frac{1}{\lambda_{ii}}$.

B.3 Classifications des états

Afin de décrire un processus stochastique, nous devons spécifier l'espace d'état et l'espace des paramètres. L'espace de paramètre est facilement catégorisé comme étant discret ou continu. L'espace d'état, cependant, en plus d'être discret ou continu, peut inclure des états ou des groupes d'états ayant des propriétés spéciales.

1. En commençant de l'état i , si la chaîne de Markov est certaine de revenir à i , on dit que l'état est récurrent. Puisque tous les états de la classe d'équivalence communiquent entre eux, la classe elle-même est récurrente. Une classification plus poussée est effectuée en fonction de la durée de récurrence, qui est le temps moyen que prend le processus pour revenir au même état pour la première fois. Si le temps de récurrence est fini, l'état (et la classe à laquelle il appartient) est connu comme positif récurrent. Si elle est infinie, l'état et la classe sont connus comme récurrents nuls. Notez qu'un état absorbant est récurrent.
2. En commençant de l'état i , si le retour de la chaîne de Markov dans cet état n'est pas certain, on dit qu'il est transitoire. Puisque tous les états de la classe d'équivalence communiquent entre eux, la classe elle-même est transitoire.

La classification des états d'un processus stochastique comme la longueur de la file d'attente (nombre de clients dans le système) joue un rôle majeur dans la compréhension de son comportement. Nous donnons ci-dessous quelques-unes des propriétés qui peuvent être déduites de la nature des états du processus.

1. S'il y a des états transitoires dans l'espace d'état du processus, à long terme $n \rightarrow \infty$, le processus ne sera pas trouvé dans ces états. Ainsi, s'il y a des états transitoires et récurrents dans l'espace d'états, le processus se terminera toujours dans les états récurrents.
2. Un processus commençant dans un état récurrent i restera toujours dans la classe d'équivalence récurrente à laquelle appartient l'état i .
3. En raison des propriétés 1 et 2 ci-dessus, seul le processus avec des modèles de processus de Markov irréductibles doit être pris en compte pour comprendre le comportement à long terme du système. Nous pouvons établir les conditions dans lesquelles des distributions limitées existent pour de tels processus.
4. Lorsque l'espace d'état comprend à la fois des états transitoires et récurrents, l'une des caractéristiques d'intérêt est la transition des états transitoires vers un état de la classe récurrente. Par exemple, les propriétés de distribution de la période d'occupation dans un système de file d'attente peuvent être déterminées en considérant 0 comme un état absorbant pour le processus de longueur de file d'attente, alors que tous les autres états sont transitoires.

Pour une élaboration sur la classification des états et leur utilité dans la modélisation stochastique.

Annexe C

C.1 Fonction génératrice

Soit X une variable aléatoire discrète non négative telle que $P(X = n) = p(n), n \in \mathbb{N}$.
La fonction génératrice $P_X(z)$ de X est définie par :

$$P_X(z) = E[z^X] = \sum_{n=0}^{\infty} p(n)z^n.$$

Du fait que $p(n) \geq 0$ et que $\sum_{n=0}^{\infty} p(n) = 1$, la fonction $P_X(z)$ est définie pour Z tel que $|Z| \leq 1$ (Z une variable complexe).

$$P_X(0) = p(0), P_X(1) = 1, P'_X(1) = E[X].$$

Et d'une manière générale,

$$P_X^{(k)}(1) = E(X(X-1)\dots(X-k+1));$$

où l'exposant (k) désigne la dérivée $k^{\text{ième}}$.

Soit $Z = X + Y$ la somme de deux variables aléatoires indépendantes discrètes X et Y .
Alors,

$$P_Z(z) = P_X(z) \cdot P_Y(z).$$

Lorsque Z a une probabilité q d'être égale à X et une probabilité $1 - q$ d'être égale à Y ,
alors

$$P_Z(z) = qP_X(z) + (1 - q)P_Y(z).$$

C.2 Transformée de Laplace-Stieltjes

La transformée de Laplace-Stieltjes $\tilde{X}(s)$ d'une variable aléatoire non négative X de fonction de répartition $F(\cdot)$, est définie par :

$$\tilde{X}(s) = E(e^{-sX}) = \int_{x=0}^{\infty} e^{-sx} dF(x), s \geq 0.$$

Lorsque la variable aléatoire admet une densité de probabilité $f(\cdot)$, la transformée se simplifie à :

$$\tilde{X}(s) = \int_{x=0}^{\infty} e^{-sX} f dx.$$

Notifions que $|\tilde{X}(s)| \leq 1$ pour tout $s = a + ib$, tel que $a \geq 0$.

$$\tilde{X}(0) = 1, \tilde{X}'(0) = -E(X), \tilde{X}^{(k)}(0) = (-1)^k E(X^k).$$

Où l'expression (k) désigne la dérivée $k^{i\grave{e}me}$.

Soit $Z = X + Y$ la somme de deux variables aléatoires indépendantes discrètes X et Y .

Alors,

$$\tilde{Z}(s) = \tilde{X}(s) \cdot \tilde{Y}(s).$$

Lorsque Z a une probabilité q d'être égale à X et une probabilité $1 - q$ d'être égale à Y , alors

$$\tilde{Z}(s) = q\tilde{X}(s) + (1 - q)\tilde{Y}(s).$$

Liste de communications

- Nawel Arrar, Lamia Derrouiche and Natalia Djellab. On the Asymptotic Behaviour of the M/G/1 Retrial Queue With Priority Customers, Bernoulli Schedule and General Retrial Times, *IAENG International Journal of Applied Mathematics*, 48, 207-213, 2018.
- L. Derrouiche, N. K. Arrar, and N. Djellab. Système de files d'attente M/G/1 avec rappels et priorité, orbite FCFS, Deuxième workshop sur les Equations aux Dérivées Partielles WEDP' , 25-26 Novembre, 2018 Sétif, Algérie.
- N. K. Arrar, L. Derrouiche and N. Djellab. M/G/1 retrial queue with priority customers and general retrial times. *30th International Conference of the Jangjeon Mathematical Society (ICJMS-2017)*, Alger, 12-15 Juillet 2017.
- L. Derrouiche, N. K. Arrar, and N. Djellab. Sur le comportement asymptotique d'un modèle d'attente avec priorité, Journées Jeunes Chercheurs JJC'14, 30 Septembre-01 Octobre, 2014, Annaba, Algérie.

Bibliographie

- [1] A. Aissani. Modélisation et simulation . OPU Office des Publications Universitaire, 02, 2011
- [2] V.V. Anisimov and M. Kurtulush. Some Markovian queuing retrial system under light-traffic conditions , *Cybernetics and Systems Analysis*, 37, 876-887, (2001).
- [3] N.K. Arrar, N.V. Djellab and J-B. Baillon. On the asymptotic behavior of $M/G/1$ retrial queues with batch arrivals and impatience phenomenon. *Mathematical and Computer Modelling* 55, 654-665, 2012.
- [4] N. K. Arrar, N. Djellab and J-B. Baillon. On stochastic decomposition property of single server retrial queueing systems. *Turkish Journal of Mathematics* 41, 918-931, 2017.
- [5] N. K. Arrar, L. Derrouiche and N. Djellab. $M/G/1$ retrial queue with priority customers and general retrial times. *30th International Conference of the Jangjeon Mathematical Society (ICJMS-2017)*, Alger, 12-15 Juillet 2017.
- [6] J.R. Artalejo. A classified bibliography of research on retrial queues : progress in 1990-1999. *TOP* 7(2), 187-211, 1999.
- [7] J.R. Artalejo. A classified bibliography of research on retrial queues : progress in 2000-2009. *Mathematical and Computer Modelling* 51, 1071-1081, 2010.
- [8] 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).
- [9] J. R. Artalejo. Retrial queues with a finite number of sources. *Journal of the Korean Mathematical Society*, 35(3) : 503-525, (1998).
- [10] J. R. Artalejo, 1st International Workshop on Retrial Queues, *Top*, 7, No. 2, 169-353, (1999).
- [11] J. R. Artalejo. Retrial Queueing systems, *Mathematical and Computer Modelling* 30, No. 3-4, 1-228, (1999).
- [12] J.R. Artalejo, G.I. Falin, Stochastic decomposition for retrial queues, *Top* 2 (2) (1994) 329–342.
- [13] J. R. Artalejo and M. J. Lòpez-Herrero. On the busy period of the $M/G/1$ retrial queue. *Naval Research Logistics*, 47, 115-127, (2000). (Cité page 3.)
- [14] 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.

- [15] I. Atencia, A queueing system under LCFS PR discipline with general retrial times, *In International Conference Modern Mathematical Methods of Investigating of the Information Net works*, vol. 16, Minsk, 2001, pp. 30–34.
- [16] I. Atencia, G. Bouza, R. Rico, A queueing system with constant repeated attempts and Bernoulli schedule, in : Fifth International Conference on Operations Research, La Habana, 2002.
- [17] I. Atencia and P. Moreno. A single-server retrial queue with general retrial times and Bernoulli schedule. *Applied Mathematics and Computation* 162, 855-880, 2005.
- [18] I. Atencia, I. Fortes and S. Sánchez. A discrete-time retrial queueing system with starting failures, Bernoulli feedback and general retrial times. *Computers and Industrial Engineering* 57, 1291-1299, 2009.
- [19] N. Bartoli and P. Del Moral. Simulation et Algorithmes Stochastiques. Cépaduès éditions, (2001).
- [20] P.P. Bocharov, O.I. Pavlova and D.A. Puzikova. M/G/1/r retrial queueing systems with priority of primary customers. *Mathematical and Computer Modelling* 30, 89-98, 1999.
- [21] L. Boutarfa and N. Djellab. On the performance of the $M_1, M_2/G_1, G_2/1$ retrial queue with pre-emptive resume policy. *YUJOR* 25, 153-164, 2015.
- [22] P. Chen, Y. Zhou and C. Li. Discrete-time retrial queue with Bernoulli vacation, preemptive resume and feedback customers. *IJEM*, Vol. 8, N4, 1236–1250, 2015.
- [23] B.D. Choi, Y. Chang, Single server retrial queues with priority calls, *Mathematical and Computer Modelling* 30 (1999) 7–32.
- [24] B.D. Choi, D.B. Han and G.I. Falin, On the virtual waiting time for an M/G/1 retrial queue with two types of calls, *J. Appl. Math. Stoch. Anal* 6, 11-24 (1993).
- [25] B.D. Choi, K.K. Park, The M/G/1 retrial queue with Bernoulli schedule, *Queueing Systems* 7 (1990) 219–227.
- [26] B.D. Choi, K.K. Park, C.E.M. Pearce, An M/M/1 retrial queue with control policy and general retrial times, *Queueing Systems* 14 (1993) 275–292.
- [27] B.D. Choi, K.H. Rhee, K.K. Park, The M/G/1 retrial queue with retrial rate control policy, *Probability in the Engineering and Informational Sciences* 7 (1993) 29–46.
- [28] B.D. Choi, K.B. Choi and Y.W. Lee, M/G/1 retrial queueing system with two types of calls and finite capacity, *Queueing Systems* 19, 215-229 (1995).
- [29] B.D. Choi and Y.Chang. Single Server Retrial Queues with Priority calls. *Mathematical and Computer Modelling* 30, 7-32, 1999.
- [30] G. Choudhury and J-C. Ke. A batch arrival retrial queue with general retrial times under Bernoulli vacation schedule for unreliable server and delaying repair. *Applied Mathematical Modelling* 36, 255-269, 2012.
- [31] I. Dimitrou. A mixed priority retrial queue with negative arrivals, unreliable server and multiple vacations. *Applied Mathematical Modelling* 37, 1295-1309, 2003.

- [32] N. V. Djellab. On the M/G/1 Retrial Queue to Breakdowns. *RAIRO : Operations Research*, 36 :299-310, (2002).
- [33] N.V. Djellab. On the single-server retrial queue. *YUJOR* 16(1), 45-53, (2006).
- [34] B. T. Doshi. Queueing systems with vacations. A survey. *Queueing systems*, 1 :29-66, (1986) .
- [35] G.I. Falin. A Single-line System with Secondary Orders. *Engineering Cybernetics Review*, 17(2), 76-83, (1979).
- [36] G.I. Falin. Single-line Repeated Orders Queueing Systems. *Mathematische Operationsforschung und Statistik, Optimization*, 5, 649-667, (1986).
- [37] G.I. Falin, A survey of retrial queues, *Queueing Systems* 7, 127-168 (1990).
- [38] G. I. Falin and C. Fricker, On the virtual waiting time in an M/G/1 retrial queue, *Journal of Applied Probability*, 28 (1991), 446-460.
- [39] G.I. Falin, J.R. Artalejo, M. Martin, On the single server retrial queue with priority customers, *Queueing Systems* 14 (1993) 439–455.
- [40] G.I. Falin and J.G.C. Templeton. Retrial queues. *Chapman&Hall*, London, (1997).
- [41] G. I. Falin and J. R. Artalejo, A finite source retrial queue, *European Journal of Operational Research*, 108 (1998), pp. 409-424.
- [42] K. Farahmand, Single line queue with repeated demands, *Queueing Systems* 6 (1990) 223–228.
- [43] G. Fayolle. A simple telephone exchange with delayed feedback. *Teletraffic Analysis and Computer Performance Evaluation*, 1986.
- [44] S.W. Fuhrmann, R.B. Cooper, Stochastic decompositions in the M/G/1 queue with generalized vacations, *Operations Research* 33 (5) (1985) 1117–1129.
- [45] D. P. GAVER, A waiting line with interrupted service including priorities, *J. Roy. Stat. Soc*, B24 (1962), 73-90.
- [46] A. Gomez-Corral. Stochastic analysis of a single server retrial queue with general retrial times. *Naval Research Logistics* 46, 561-581, 1999.
- [47] O. Hashida and K. Kawashima. Buffer Behavior with Repeated Calls. *Electronics and Communication in Japan*, 62-B, 27, (1979).
- [48] T. Kernane. Conditions for stability and instability of retrial queueing systems with general retrial times. *Statistics and Probability Letters* 78, 3244-3248, 2008.
- [49] Z. Khalil and G. Falin, Stochastic inequalities for M/G/1 retrial queues, *Operations Research Letters*, 16 (1994), pp. 285-290.
- [50] 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).
- [51] B. Kim and J. Kim. Waiting time distributions in an M/G/1 retrial queue with two classes of customers. *Annals of Operations Research*, Vol. 252, Issue 1, 121–134, May 2017.

- [52] C. Langaris and E. Moutzoukis. A retrial queue with structured batch arrivals, priorities and server vacations. *Queueing Systems* 20, 341-368, (1995).
- [53] Y.W. Lee. The M/G/1 feedback retrial queue with two types of customers. *Bulletin of Korean Mathematical Society* 42, 875-887, 2005.
- [54] 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).
- [55] 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).
- [56] M. Martin, J.R. Artalejo, Analysis of an M/G/1 queue with two types of impatient units, *Advances in Applied Probability* 27 (3) (1995) 840–861.
- [57] P. Moreno. An M/G/1 retrial queue with recurrent customers and general retrial times. *Applied Mathematics and Computation* 159, 651-666, 2004.
- [58] U. B. Narayan. An introduction to queueing theory. Statistics for industry and technology, USA, (2010).
- [59] S. M. Ross, Stochastic Processes, Willey, New York, 1983.
- [60] A. Ruegg, Processus Stochastiques. Presses polytechniques romandes, Lausannes, 1989.
- [61] L. I. Senott, P. A. Humblet and R. I. Tweedie, Mean drifts and non ergodicity of Markov chains, *Operations Research*, 31 (1983), 783-789.
- [62] M. Shaked and J. G. Shanthikumar, Stochastic Orders and Their Applications, AcademicPress, 1994.
- [63] 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) .
- [64] J.G. Shanthikumar, On stochastic decomposition in M/G/1 type queues with generalized server vacations, *Operations Research* 36 (4) (1988) 566–569.
- [65] C. Shekhar, A. A. Raina and A. Kumar. A brief review on retrial queue : Progress in 2010–2015. *IJASER*, vol. 5, Issue 4, 324-336, 2016.
- [66] D. Stoyan, Comparison Methods for Queues and Other Stochastic Models, New York, 1983.
- [67] D. Sumitha and K.U. Chandrika. $M^X/G/1$ retrial queue with second optional service, feedback, admission control and vacation. *IJIRSET*, vol. 3, Issue 1, 8320-8329, January 2014.
- [68] J. G. C. Templeton, Retrial Queues, *Queueing Systems*, 7, No. 2, 125-227, (1990).
- [69] J. G. C. Templeton, Retrial Queues, *Top*, 7 : 351-353, (1999).
- [70] W. Whitt. On Approximations for Queues, I : Extremal Distributions. *AT & T Bell Laboratoire Technical Journal*, 63, 115-138, (1984).
- [71] J. Wu and Z. Lian. A single-server retrial G-queue with priority and unreliable server under Bernoulli vacation schedule. *Computers and Industrial Engineering*, 64, 84-93, 2013.

- [72] J. Wu, J. Wang and Z. Liu. A discrete-time Geo/G/1 retrial queue with preferred and impatient customers. *Applied Mathematical Modelling* , 37, 2552-2561, 2013.
- [73] T. Yang and J.G.C. Templeton. A survey on retrial queues. *Queueing Systems* 2, 201-233, (1987).
- [74] T. Yang, M.J.M. Posner, J.G.C. Templeton, H. Li, An approximation method for the M/G/1 retrial queue with general retrial times, *European Journal of Operational Research* 76 (1994)552–562.
- [75] 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).