

وزارة التعليم العالي والبحث العلمي

BADJI MOKHTAR -ANNABA
UNIVERSITY
UNIVERSITE BADJI MOKHTAR
ANNABA



جامعة باجي مختار
- عنابة -

Faculté des Sciences
Département de Mathématiques

THESE

Présentée en vue de l'obtention du diplôme de
DOCTORAT EN MATHÉMATIQUES
Option : Recherche Opérationnelle

Modèles d'attente avec rappels et priorité et leurs applications

Par

Mlle Boutarfa Leila

Sous la direction de **Prof. Natalia Djellab, U.B.M.ANNABA**

Devant le jury

| | | | |
|--------------------|----------------------------|------------|----------------------|
| PRESIDENT | Assia Chadli | MCA | U.B.M. ANNABA |
| DIRECTEUR | Natalia Djellab | Pr | U.B.M. ANNABA |
| EXAMINATEUR | Salim Haddadi | Pr | U. GUELMA |
| EXAMINATEUR | Nacira Seddik-Ameur | Pr | U.B.M. ANNABA |
| EXAMINATEUR | Mohamed Boualem | MCA | U.A.M. BEJAIA |
| EXAMINATEUR | Nadia Oukid | MCA | U.S.D. BLIDA |

Année Universitaire 2014 / 2015

Remerciement

Je remercie vivement mon encadreur Mme DJELLAB pour sa continuelle assistance, sa patience et ses conseils qui m'ont été d'un grand apport et encouragement pendant toute la durée de l'élaboration de ce travail.

Je tiens également à remercier :

- La présidente du jury, d'avoir eu l'amabilité d'accepter la présidence du jury de soutenance*
- Les membres de jury pour leur participation et leur dévouement*

Dédicace

Je dédie ce modeste travail à :

- Mes parents*
- Mes frères et mes sœurs*
- Tous mes proches*

Avec toute mon affection et ma gratitude

Sommaire

| | |
|--|----|
| Introduction générale | 1 |
| Chapitre 1: Systèmes de files d'attente avec rappels | |
| 1.1. Introduction | 11 |
| 1.2. Classification | 11 |
| 1.3. Modèle général | 12 |
| 1.4. Modèle d'attente avec rappels de type M/G/1..... | 12 |
| 1.4.1. Description du modèle | 12 |
| 1.4.2. Chaîne de Markov induite..... | 13 |
| 1.4.3. Distribution conjointe de l'état du serveur et du nombre de clients en orbite..... | 17 |
| 1.4.4. Période d'activité et temps d'attente | 19 |
| 1.5. Les propriétés de monotonie de la chaîne de Markov induite | 21 |
| 1.5.1. Monotonie interne et externe de l'opérateur de transition..... | 23 |
| 1.5.2. Inégalités stochastiques pour la distribution stationnaire du nombre de clients dans le système. Les propriétés de monotonie de la chaîne de Markov induite | 24 |
| 1.5.3. Inégalités stochastiques pour le nombre moyen de clients servis durant la période d'activité | 25 |
| 1.6. Autres modèles | 26 |
| 1.7. Conclusion..... | 26 |
| Chapitre 2: Interruption de service dans les systèmes d'attente avec rappels | |
| 2.1. Introduction | 27 |
| 2.2. Modèles d'attente avec rappels et vacances du serveur | 28 |
| 2.2.1. Description du modèle | 29 |
| 2.2.2. Analyse du modèle..... | 30 |
| 2.2.3. Mesures de performance | 31 |
| 2.3. Modèles d'attente avec rappels et priorité..... | 31 |
| 2.3.1. Modèle d'attente avec rappels de type $M1M2/G1G2/1$ et priorité relative | 32 |
| 2.4. Modèles avec rappels et serveur sujet à des pannes aléatoires | 36 |
| 2.4.1. Modèle d'attente avec rappels de type M/G/1 et serveur non fiable..... | 37 |
| 2.4.2. Modèle M/G/1 avec rappels et deux types de pannes actives..... | 43 |
| 2.5. Conclusion..... | 53 |
| Chapitre 3: Systèmes de files d'attente avec rappels et interruptions de service dans la modélisation des phénomènes réels | |

| | | |
|--|---|------------|
| 3.1. | Introduction | 54 |
| 3.2. | Systèmes réels modélisables comme systèmes d'attente avec rappels et priorité relative .. | 55 |
| 3.2.1. | Modèle 1 | 55 |
| 3.2.2. | Modèle 2 | 56 |
| 3.2.3. | Exemples d'application..... | 56 |
| 3.3. | Systèmes réels modélisables comme systèmes d'attente avec rappels et priorité absolue .. | 58 |
| 3.3.1. | Bureaux d'information..... | 58 |
| 3.3.2. | Réseaux de communication | 59 |
| 3.4. | Systèmes réels modélisables comme systèmes de files d'attente avec rappels et vacances du serveur | 59 |
| 3.4.1. | Exemples d'application..... | 61 |
| 3.5. | Systèmes réels modélisables comme un système de files d'attente avec rappels et serveur sujet à des pannes | 64 |
| 3.5.1. | Transmission des messages à travers des réseaux fac-similes | 64 |
| 3.5.2. | Réseaux informatiques avec infection virale..... | 64 |
| 3.5.3. | Interaction hôte-parasite (application biologique)..... | 67 |
| 3.6. | Conclusion..... | 67 |
| Chapitre 4: Modèle d'attente avec rappels et priorité absolue | | |
| 4.1. | Introduction | 68 |
| 4.2. | Description mathématique du modèle..... | 69 |
| 4.3. | Condition d'ergodicité | 71 |
| 4.4. | Distribution stationnaire de l'état du système | 74 |
| 4.5. | Mesures de performances | 82 |
| 4.6. | Illustration numérique..... | 84 |
| 4.7. | Conclusion..... | 86 |
| Chapitre 5: Monotonie de la chaîne de Markov induite d'un modèle M/G/1 avec rappels et priorité absolue | | |
| 5.1. | Introduction | 87 |
| 5.2. | Les propriétés de monotonie de la chaîne de Markov induite | 88 |
| 5.3. | Inégalités stochastiques pour la distribution stationnaire | 99 |
| 5.4. | Applications numériques | 101 |
| 5.5. | Conclusion..... | 104 |
| Conclusion Générale | | 105 |

| | |
|----------------------------|-----|
| Bibliographie | 107 |
|----------------------------|-----|

Introduction générale

La théorie des files d'attente, est une technique de la Recherche Opérationnelle relevant du domaine des probabilités. Elle a été développée pour fournir des modèles mathématiques pour prédire le comportement des systèmes admettant un phénomène d'attente.

On peut schématiser un système (ou phénomène) d'attente comme suit : un ensemble d'individus, qu'on appelle clients, arrivent suivant un processus quelconque (le plus souvent aléatoire) pour acquérir un service auprès d'un autre individu dit serveur. La constitution de la file d'attente commence à se manifester dès que le taux des arrivées excède le taux de service (par taux, on entend le nombre moyen de clients arrivants ou servis par unité de temps). La file d'attente peut ne pas se manifester de façon personnifiée : on parle alors d'une file de machines en panne qui attendent la réparation dans un atelier, ou d'un ensemble de programmes qui attendent l'acquisition d'une composante de la machine,...etc. Un système de files d'attente comprend donc un espace de service avec un ou plusieurs dispositifs de service (serveurs), et un espace d'attente dans lequel se forme une éventuelle file d'attente.

Une abondante littérature couvre le sujet des systèmes de files d'attente [20] [7].... Les premiers travaux sur les files d'attente sont dus au mathématicien danois Erlang (1878-1929), alors qu'il s'intéressait à la modélisation du réseau téléphonique de Copenhague, afin d'en réduire les phénomènes d'attente. La théorie mathématique des files d'attente s'est développée par la suite grâce notamment aux travaux de Palm, de Kolmogorov et de Khintchine.

La théorie de 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 le client quitte le système sans recevoir le service, soit il prend place dans une file d'attente. Une possibilité alternative est de permettre au client de répéter sa demande de service après une durée de temps aléatoire. Entre deux tentatives successives (rappels), le client en question est en orbite. Un tel système est appelé système de files

d'attente avec rappels. L'étude de cette catégorie de modèles est motivée par l'avènement de nouvelles technologies, notamment dans les systèmes de télécommunication.

Habituellement, lors de l'étude des problèmes de la théorie des files d'attente classique ou celle avec rappels, on supposait que le service s'effectue d'une manière continue et bien déterminée (sans interruptions). Cependant en pratique, on rencontre souvent des cas où les serveurs sont sujets à des pannes aléatoires, ou le service d'un client est interrompu à cause d'arrivée d'un client prioritaire. Les interruptions de service dans les système de files d'attente (classique ou avec rappels) peuvent être dues aussi à la possibilité de vacance du serveur [26], [17]. Dans un système d'attente avec vacances, le serveur prend occasionnellement une vacance d'une durée aléatoire, qui peut être utilisée pour accomplir une ou plusieurs tâches secondaires (prioritaires ou de maintenance).

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 (voire 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,...[37].

Il est constaté que ces indices de performance sont difficiles à obtenir sous formes explicites pour la plupart des systèmes de files d'attente ou bien leur forme explicite est complexe (elles contiennent des transformées de Laplace, des expressions intégrales,...) difficile à interpréter en pratique. Pour surmonter cette difficulté, il est possible de faire recours à une des méthodes d'approximation qui est la méthode de comparaison stochastique (autrement dit : méthode de bornes stochastiques).

La méthode de bornes stochastiques [40] s'applique aux chaînes de Markov multidimensionnelles et permet ainsi d'apporter des solutions intéressantes pour l'évolution des performances des systèmes complexes. Un système complexe peut-être un système de grande taille représentant un réseau avec beaucoup de nœuds. Il peut être aussi un système où différents types d'événements peuvent se déclencher (arrivées/service, pannes physiques/ réparation, arrivée prioritaire, vacance du serveur,...). Car la prise en compte de plusieurs événements peut augmenter le nombre de composantes de l'état du système, les solutions apportées par ces méthodes consisteraient à construire des systèmes bornants plus faciles à étudier. A partir de ces systèmes bornants, on pourra générer des mesures de performances bornantes (supérieures et inférieures) de la mesure exacte.

La comparaison stochastique est basée sur la théorie des ordres stochastiques. Un ordre stochastique est simplement une relation d'ordre définie sure un espace des fonctions de répartition. Pour l'étude de systèmes de files d'attente, les ordres : stochastique fort (\leq_{st}), convexe (\leq_v) et Laplacien (\leq_L), s'imposent bien car ils permettent de comparer différents indices de performances comme par exemple, le nombre moyen de clients dans un système, le taux de perte, la probabilité de panne,...etc.

La monotonie est une propriété importante en termes de borne, c'est-à-dire que la meilleure chaîne bornante (plus proche du système exacte) est monotone.

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 ce travail, nous réalisons l'analyse d'un système de files d'attente M/G/1 avec rappels et priorité absolue. Nous considérons deux types de clients (plus prioritaire type 1 et moins prioritaire type 2). Un client type 1, arrivant dans le système et trouvant un autre client type 1 dans le serveur, rejoint l'orbite et devient source de tentatives répétées. Un client type 1, arrivant dans le système et trouvant un client type 2 dans le serveur, interrompt le service en cours et commence le sien. Le client dont le service a été interrompu, reste auprès du serveur pour terminer son service ultérieurement. En outre, les clients type 2, arrivant dans le système et trouvant le serveur occupé, quittent le système sans entrer dans l'espace de service. Notre analyse inclut : description de la chaîne de Markov induite, condition d'ergodicité, distribution stationnaire de l'état du système, obtention de mesures de performance, étude des propriétés de monotonie de la chaîne de Markov induite. Des inégalités stochastiques pour la distribution stationnaire du nombre de clients dans le système considéré sont également introduites. Notre modèle peut être utilisé pour modéliser une situation (fréquemment observée dans quelques bureaux d'information), où un agent seul répond aux appels téléphoniques et sert le client présent. Dans ce contexte, les appels téléphoniques ont une priorité absolue par rapport au client présent. Il faut noter que notre modèle n'est pas étudié dans la littérature. Son analyse mathématique demande un grand effort analytique. Les résultats obtenus sont appuyés par des illustrations numériques.

Cette thèse est organisée en cinq chapitres. Dans le premier chapitre, nous présentons quelques rappels sur les systèmes de files d'attente avec rappel (sans interruption de service).

La description du phénomène d'interruptions de service dans les systèmes d'attente avec rappel fait l'objet du second chapitre. Nous considérons les modèles avec serveur non fiable, les modèles prioritaires et ceux avec vacances.

Le troisième chapitre contient une présentation de quelque système réelles modélisables par des systèmes d'attente avec rappels et interruptions du service.

Les chapitre quatre et cinq sont concernés par nos investigations sur le modèle d'attente M/G/1 avec rappels et priorité absolue.



Systemes de files d'attente avec rappels

1.1. Introduction

Dans ce chapitre nous allons faire une synthèse de certains résultats concernant une nouvelle branche de la théorie de files d'attente, dite *théorie des systèmes de files d'attente avec rappels*. Un système de files d'attente avec rappels est caractérisé par l'hypothèse de base suivante : « *Un client qui arrive et trouve l'espace de service et celui d'attente (s'il existe) occupés peut faire des tentatives de service après une durée de temps aléatoire* ». Son étude est motivée par diverses applications pratiques dans le domaine de télécommunication [23].

1.2. Classification

Les systèmes de files d'attente avec rappels sont identifiés par 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.

La notation de Kendall correspondante à ce type de système (avec rappels) 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 de la file d'attente, O est la capacité de l'orbite et H représente une fonction de persistance $H = \langle H_k, k \geq 0 \rangle, 0 < H_k < 1$. Si m, O, 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.

1.3. Modèle général

En général, un système de files d'attente avec rappels contient un espace de service composé de $c \geq 1$ dispositifs de service et un espace d'attente 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 un ou plusieurs serveurs libres, le client sera immédiatement pris en charge. Sinon, s'il y a des positions d'attente libres, le client va rejoindre la file d'attente selon la discipline de service adaptée dans le système. D'autre part, si un client arrive et trouve tous les serveurs et positions d'attentes occupés, il quitte le système définitivement avec une probabilité $1 - H_0$ ou fait des tentatives de service après une durée de temps aléatoire avec une probabilité H_0 . Entre les tentatives, le client est en « orbite » et devient « source d'appels secondaires » (rappels). La capacité de l'orbite O peut être finie ou infinie. Dans le cas où O est finie et si l'orbite est pleine, le client quitte le système pour toujours. Lorsqu'un client est rappelé de l'orbite, il est traité de la même manière qu'un client primaire avec une probabilité H_k (s'il s'agit de la $k^{\text{ème}}$ tentative échouée).

1.4. Modèle d'attente avec rappels de type M/G/1

1.4.1. Description du modèle

Les modèles d'attente de type $M/G/1$ sont des modèles où les appels arrivent selon un processus homogène de Poisson de taux $\lambda > 0$. Le service des clients est assuré par un serveur. Les durées de service suivent une loi générale de fonction de

répartition $B(x)$ de transformée de Laplace-Stieltjes $\tilde{B}(s)$, $Re(s) > 0$. Soient les moments d'ordre k , $\beta_k = (-1)^k \tilde{B}^{(k)}(0)$, l'intensité du trafic $\rho = \lambda\beta_1$ et $\mu = 1/\beta_1$. La durée entre deux rappels successifs d'une même source secondaire est exponentiellement distribuée de paramètre $\theta > 0$.

L'état du système à la date t peut être décrit par le processus stochastique suivant

$$\{C(t); N_o(t); \xi(t); t \geq 0\}, \quad (1.1)$$

où $C(t)$ est le nombre de serveurs occupés (0 ou 1), $N_o(t)$ est le nombre de clients en orbite, $\xi(t)$ est une variable supplémentaire à valeurs dans \mathbb{R}^+ et désignant la durée de service écoulée à la date t .

1.4.2. Chaîne de Markov induite

Le modèles d'attente avec rappels de type M/G/1 possède une chaîne de Markov induite qui a été décrite pour la première fois par Choo et Conolly [15]. Dénotons par $N_i = N_o(\xi_i)$ le nombre de clients en orbite à la date ξ_i du $i^{\text{ème}}$ départ. Il est facile de voir que

$$N_i = N_{i-1} - B_i + v_i,$$

où B_i est une variable de Bernoulli, prenant la valeur 1 si le $i^{\text{ème}}$ client servi provient de l'orbite et 0 si le $i^{\text{ème}}$ client servi est primaire, v_i est une variable aléatoire représentant le nombre de clients primaires arrivant dans le système durant le temps de service du $i^{\text{ème}}$ client. La variable aléatoire B_i dépend de N_{i-1} , sa distribution conditionnelle est donnée par

$$P(B_i = 1 \setminus N_{i-1} = n) = \frac{n\theta}{\lambda + n\theta};$$

$$P(B_i = 0 \setminus N_{i-1} = n) = \frac{\lambda}{\lambda + n\theta}.$$

La variable aléatoire v_i ne dépend pas d'événements qui se sont produits avant l'instant ξ_i et a comme distribution

$$k_n = P(v_i = n) = \int_0^{\infty} \exp(-\lambda x) \frac{(\lambda x)^n}{n!} dB(x),$$

avec une fonction génératrice

$$K(z) = \sum_{n=0}^{\infty} k_n z^n = \tilde{B}(\lambda - \lambda z),$$

et une moyenne

$$E[v_i] = K'(1) = \sum_{n=0}^{\infty} nk_n = \rho.$$

Les remarques en haut impliquent que la suite des variables aléatoires $\{N_i\}$ forme une chaîne de Markov induite pour le système en question. Les probabilités de transition à une étape de la chaîne $r_{nm} = P(N_i = m \setminus N_{i-1} = n)$ sont données par la formule

$$r_{nm} = \frac{\lambda}{\lambda + n\theta} k_{m-n} + \frac{n\theta}{\lambda + n\theta} k_{m-n+1}.$$

Notons que $r_{nm} \neq 0$ uniquement pour $n = 0, 1, \dots, (m+1)$.

1.4.2.1. Condition d'ergodicité

Théorème 1: La chaîne de Markov induite $\{N_i\}$ est ergodique si et seulement si $\rho = \lambda\beta_1 < 1$.

Preuve :

Comme la structure de l'équation fondamentale est réursive, il est plus conviviale d'utiliser le critère de Foster [35]. D'après le critère de Foster, une condition suffisante pour qu'une chaîne de Markov $\{N_i\}$ apériodique et irréductible d'espace des états S soit ergodique, est l'existence d'une fonction non négative dite la fonction de Lyapunov $f(S), s \in S$ et $\varepsilon > 0$ tel que l'accroissement moyen de la chaîne $x_s \equiv E(f(N_i) - f(N_{i-1}) \mid N_{i-1} = s)$ est finie pour tout $s \in S$ et $x_s \leq -\varepsilon$ pour tout $s \in S$ sauf peut être pour un nombre fini. Quand l'espace d'états S est l'ensemble \mathbb{Z}_+ d'entiers non négatifs, il est suffisant de considérer la fonction $f(n) = n$. Cela veut dire, qu'une chaîne de Markov avec espace d'états $S = \mathbb{Z}_+$ est ergodique si son accroissement moyen $x_n \equiv E(N_i - N_{i-1} \mid N_{i-1} = n) \leq -\varepsilon$ pour tout $n \geq N$ où N est un nombre entier suffisamment grand. Alors, si existe là la limite $x = \lim_{n \rightarrow +\infty} x_n$, cette condition est vérifiée si est seulement si $x < 0$. Pour la chaîne de Markov décrite dans ce chapitre, l'accroissement moyen de la chaîne vaut :

$$\begin{aligned}
x_n &= E(-B_i + v_i \mid N_{i-1} = n) \\
&= -E(B_i \mid N_{i-1} = n) + E(v_i \mid N_{i-1} = n) \\
&= -E(B_i = 1 \mid N_{i-1} = n) + E(v_i) \\
&= -\frac{n\theta}{\lambda + n\theta} + \rho,
\end{aligned}$$

$\lim_{n \rightarrow +\infty} x_n = -1 + \rho$. Cette limite est négative si $\rho < 1$. Puisque $r_{nm} = 0$ pour $n > m + 1$, $\rho < 1$ est aussi une condition nécessaire pour l'ergodicité. Pour prouver que $\rho \geq 1$ donne non ergodicité à notre chaîne de Markov à étudier, on applique le critère de Kaplan [38] qui dit : une chaîne de Markov apériodique et irréductible d'espace des états $S = \mathbb{Z}_+$ est non-ergodique si l'accroissement moyen $x_n \equiv E(N_i - N_{i-1} \mid N_{i-1} = n)$ est fini pour tout $n \in \mathbb{Z}_+$, et l'accroissement moyen $\delta_n \equiv \sum_{m < n} (m - n) \cdot r_{nm} \geq -const$, i.e. borné en bas, et qu'il existe

N tel que $x_n \geq 0$ pour $n \geq N$. Pour le système $M/G/1$ considéré dans ce chapitre, δ_n est borné en bas quand $N_i - N_{i-1} \geq -1$, et si $\rho \geq 1$ alors

$$x_n = -\frac{n\theta}{\lambda + n\theta} + \rho \geq -\frac{n\theta}{\lambda + n\theta} + 1 = \frac{\lambda}{\lambda + n\theta} > 0.$$

1.4.2.2. Distribution stationnaire

Théorème 2 [19] : Soit $\rho < 1$. La distribution stationnaire de la chaîne de Markov induite possède la fonction génératrice suivante :

$$\varphi(z) = \sum_{n=0}^{\infty} z^n \pi_n = \frac{(1-\rho)(1-z)\tilde{B}(\lambda-\lambda z)}{\tilde{B}(\lambda-\lambda z)-z} \exp\left\{\frac{\lambda}{\theta} \int_0^z \frac{1-\tilde{B}(\lambda-\lambda u)}{\tilde{B}(\lambda-\lambda u)-u} du\right\}.$$

Preuve :

Soit $\pi_n = \lim_{i \rightarrow \infty} P(N_o(\xi_i) = n)$. Les équations de Kolmogorov se présentent de la manière suivante :

$$\pi_n = \sum_{m=0}^n \pi_m \frac{\lambda}{\lambda + m\theta} k_{n-m} + \sum_{m=1}^{n+1} \pi_m \frac{m\theta}{\lambda + m\theta} k_{n-m+1}$$

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

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

Encore

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

Par conséquent,

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

Sous la condition que $\rho < 1$,

$$\psi(z) = \psi(1) \exp \left\{ \frac{\lambda}{\theta} \int_0^z \frac{1 - \tilde{B}(\lambda - \lambda u)}{\tilde{B}(\lambda - \lambda u) - u} du \right\}.$$

On obtient

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

Vu que $\varphi(1) = 1$, on a $\psi(1) = \frac{1 - \rho}{\lambda}$. Enfin, on obtient la fonction génératrice

$$\varphi(z) = \sum_{n=0}^{\infty} z^n \pi_n.$$

1.4.3. Distribution conjointe de l'état du serveur et du nombre de clients en orbite

Le premier résultat sur les modèles de type M/G/1 a été trouvé par Keilson, Cozzolino et Young [30]. Les auteurs ont utilisé la méthode des variables supplémentaires. Leur résultat peut être résumé comme suit : L'état du système à la date t est décrit par le processus (1.1). Lorsque $\rho < 1$, le système est en régime stationnaire. Dénotons par

$$\begin{aligned}p_{0n} &= \lim_{t \rightarrow \infty} P(C(t) = 0; N_0(t) = n), \\ p_{1n}(x) &= \lim_{t \rightarrow \infty} \frac{d}{dx} P(C(t) = 1; \xi(t) \leq x; N_0(t) = n)\end{aligned}$$

les probabilités qui vérifient le système d'équations d'équilibre statistique

$$\begin{cases} (\lambda + n\theta)p_{0n} = \int_0^{\infty} p_{1n}(x) b(x) dx; \\ p'_{1n}(x) = -(\lambda + b(x))p_{1n}(x) + \lambda p_{1,n-1}(x); \\ p_{1n}(0) = \lambda p_{0n} + (n+1)\theta p_{0,n+1}; \end{cases} \quad (1.9)$$

où $b(x) = (B'(x)) \setminus ((1 - B(x)))$ est l'intensité instantanée du service (en supposant que x est la durée de service écoulée).

A l'aide des fonctions génératrices telles que

$$P_0(z) = \sum_{n=0}^{\infty} z^n p_{0n} \text{ et } P_1(z, x) = \sum_{n=0}^{\infty} z^n p_{1n}(x),$$

le système (1.9) devient :

$$\begin{cases} \lambda P_0(z) + \theta z \dot{P}_0(z) = \int_0^{\infty} P_1(z, x) b(x) dx; \\ \dot{P}_1(z, x) = (\lambda z - \lambda - b(x))P_1(z, x); \\ P_1(z, 0) = \lambda P_0(z) + \theta \dot{P}_0(z). \end{cases} \quad (1.10)$$

La résolution de ce dernier fournit les résultats suivants :

$$P_0(z) = (1 - \rho) \exp \left\{ \frac{\lambda}{\theta} \int_1^z \frac{1 - \tilde{B}(\lambda - \lambda u)}{\tilde{B}(\lambda - \lambda u) - u} du \right\}; \quad (1.11)$$

$$P_1(z) = \int_0^{\infty} P_1(z, x) dx = \sum_{n=0}^{\infty} z^n P_{1n} = \frac{1 - \tilde{B}(\lambda - \lambda z)}{\tilde{B}(\lambda - \lambda z) - z} P_0(z). \quad (1.12)$$

La forme explicite de p_{0n} et p_{1n} existe pour le cas d'une distribution exponentielle du temps de service.

A partir des équations (1.11)-(1.12), on peut définir les fonctions génératrices suivantes :

- Fonction génératrice de la distribution marginale de la taille de l'orbite

$$P(z) = P_0(z) + P_1(z) = \frac{1-z}{\tilde{B}(\lambda - \lambda z) - z} P_0(z).$$

- Fonction génératrice du nombre de clients dans le système

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

- Distribution marginale du nombre de serveurs occupés

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

1.4.4. Période d'activité et temps d'attente

La période d'activité d'un système de files d'attente avec rappels est définie comme un intervalle de temps qui commence à l'instant d'arrivée d'un client dans le système vide et se termine à l'instant de départ quand le système redevient vide. Une analyse détaillée de cette période a été faite par Falin [24]. L'auteur a obtenu la transformée de Laplace-Stieljes conjointe de la durée L de la période en question et du nombre de clients I servis pendant L :

$$E[e^{-sL}y^I] = \frac{s+\lambda}{\lambda} - \frac{\theta}{\lambda} \left[\int_0^{\pi_\infty(s,y)} \exp \left\{ -\frac{1}{\theta} \int_0^u \frac{s+\lambda - \lambda y \tilde{B}(s+\lambda - \lambda v)}{y \tilde{B}(s+\lambda - \lambda v) - v} dv \right\} \frac{du}{y \tilde{B}(s+\lambda - \lambda u) - u} \right]^{-1},$$

où $\pi_\infty(s, y)$ et la solution continue de l'équation $y\tilde{B}(s + \lambda - \lambda\pi_\infty) = \pi_\infty$ dans l'intervalle $[0,1]$.

En outre,

- Si $\rho > 1$, alors $p[L = \infty] = p[I = \infty] > 0$;
- Si $\rho = 1$, alors $E[L] = E[I] = \infty$;

c) Si $\rho < 1$, alors

$$E[I] = \frac{1}{1-\rho} \exp \left\{ \frac{\lambda}{\theta} \int_0^1 \frac{1 - \tilde{B}(\lambda - \lambda u)}{\tilde{B}(\lambda - \lambda u) - u} du \right\},$$

$$E[L] = \frac{E[I] - 1}{\lambda} < \infty.$$

Le temps d'attente d'un client peut être mesuré par le temps physique et par le nombre de rappels effectués. Le temps virtuel d'attente $W(t)$ est la durée qui s'écoule entre l'instant d'arrivée d'un client primaire et l'instant de son départ (achèvement du service). Les chercheurs n'ont pas réussi à obtenir une expression explicite de la distribution $W(t)$. Par contre, un effort important a été fait par Falin et Fricher [22]. Les auteurs ont obtenu l'expression de la transformée de Laplace-Stieltjes du temps virtuel d'attente en régime stationnaire:

$$E[e^{-sW(t)}] = 1 - \rho + \frac{1 - \rho}{s} \int_1^{\pi_\infty(s,1)} \left[\frac{\lambda(1-u)[\tilde{B}(\lambda - \lambda u) - \tilde{B}(s + \lambda - \lambda u)]}{[\tilde{B}(\lambda - \lambda u) - u][\tilde{B}(s + \lambda - \lambda u) - u]} \right]$$

$$\times \exp \left\{ \int_u^1 \frac{(\theta + s)[\tilde{B}(\lambda - \lambda v) - v] + (\lambda - \lambda v)[\tilde{B}(\lambda - \lambda v) - \tilde{B}(s + \lambda - \lambda v)]}{\theta[\tilde{B}(\lambda - \lambda v) - v][\tilde{B}(s + \lambda - \lambda v) - v]} dv \right\} du ,$$

Où $\pi_\infty(s, 1)$ est la solution continue de l'équation $\tilde{B}(s + \lambda - \lambda\pi_\infty) = \pi_\infty$ dans l'intervalle $[0,1]$.

Ce résultat est trouvé à l'aide d'une variable aléatoire T_n qui représente le temps d'attente d'un client étiqueté parmi $n \geq 1$ clients se trouvant en orbite à l'instant de départ d'un client, et sa distribution (densité de probabilité) $f(t)$ est donnée par :

$$f(t) = \begin{cases} [1 - (1 - \rho)t]^{\frac{\rho}{1-\rho}}, & 0 \leq t \leq \frac{1}{1-\rho} & \text{si } \rho < 1, \\ e^{-1} & & \text{si } \rho = 1, \\ [1 + (\rho - 1)t]^{-\frac{\rho}{1-\rho}} & & \text{si } \rho > 1. \end{cases}$$

Si la distribution du temps de service est exponentielle, le premier moment est :

$$E[W(t)] = \frac{\rho}{1-\rho} \left(\frac{1}{\mu} + \frac{1}{\theta} \right).$$

Dans les systèmes de files d'attente avec rappels, on s'intéresse aussi à une autre mesure dite le nombre de rappels $R(t)$ effectués par un client primaire, arrivant dans le système à la date t , avant le début de son service. La fonction génératrice de la distribution stationnaire de $R(t)$ est [22]:

$$E[x^{R(t)}] = 1 - \rho + \frac{1 - \rho}{\theta - \theta x} x \int_1^{\pi_\infty(\theta - \theta x, 1)} \frac{[\lambda(1-u)[\tilde{B}(\lambda - \lambda u) - \tilde{B}(\theta - \theta x + \lambda - \lambda u)]}{[\tilde{B}(\lambda - \lambda u) - u][\tilde{B}(\theta - \theta x + \lambda - \lambda u) - u]} \\ \times \exp \left\{ \int_u^1 \frac{\theta[\tilde{B}(\lambda - \lambda v) - v] + (\lambda - \lambda v)[\tilde{B}(\lambda - \lambda v) - \tilde{B}(\theta - \theta x + \lambda - \lambda v)]}{\theta[\tilde{B}(\lambda - \lambda v) - v][\tilde{B}(\theta - \theta x + \lambda - \lambda v) - v]} dv \right\} du.$$

Dans le cas d'une distribution du temps de service exponentielle, le premier moment est donnée par :

$$E[R(t)] = \frac{\rho}{1-\rho} \left(1 + \frac{\theta}{\mu} \right).$$

1.5. Les propriétés de monotonie de la chaine de Markov induite

Souvent, il y a une grande difficulté d'obtenir des résultats explicites pour la plupart des systèmes d'attente. Cette difficulté repose sur la complexité intrinsèque de l'espace d'états et de la dynamique d'évolution sur cet espace. Il est alors crucial de pouvoir comparer ces systèmes à des systèmes de performances bien connues. Pour cela, il est nécessaire de disposer d'outils de comparaison et, comme la comparaison se fait sur des systèmes aléatoires, la théorie adoptée est celle de la comparaison stochastique.

L'idée clé de la comparaison stochastique, est de concevoir une nouvelle chaîne de Markov telle que les indices de celle-ci soient une borne supérieure ou inférieure des indices de performance de la chaîne de Markov originale. La nouvelle chaîne de Markov est celle d'un modèle simplifié du modèle original. En fait, ces bornes et les critères de simplification sont basés sur la notion d'ordres stochastiques.

Un ordre stochastique est simplement, une relation d'ordre définie sur un espace des fonctions de répartition. Nous nous intéressons particulièrement aux ordres dits intégraux, définis par une famille de fonction. Ces ordres sont importants pour l'évaluation de performances car ils permettent de comparer différentes fonctions de récompense (par exemple le nombre de clients dans un système, le taux de perte, la probabilité de panne,...).

L'ouvrage de référence, ouvrage historique écrit par Stoyan [40], développe les bases de cette théorie. Il présente les principaux outils qui seront utilisés dans ce chapitre. Plus récemment, les ouvrages de Szekli [41] et Shaked et Shantikumar [39] exposent les derniers résultats de recherche concernant cette théorie.

Dans la littérature, on trouve plusieurs applications de la théorie générale des ordres stochastiques, parmi lesquelles on trouve l'étude des propriétés de monotonie du système $M/G/1$ avec rappels relativement aux ordres stochastiques : fort (\leq_{st}), laplacien (\leq_L) et convexe (\leq_v). Ces propriétés sont établies par Khalil et Falin [31]. Leurs résultats seront annancer dans ce chapitre. On trouve également le travail de Boualem, Djellab et Aissani [9] où des bornes stochastiques pour le système $M/G/1$ avec rappels constants et vacances du serveur ont été établies.

Dans tous ce qui suit, on dénote par $M^{(1)}$ et $M^{(2)}$ deux systèmes d'attente avec rappels de type $M/G/1$ définis par $\lambda^{(1)}, \theta^{(1)}, B^{(1)}(x), k_n^{(1)}$ and $\lambda^{(2)}, \theta^{(2)}, B^{(2)}(x), k_n^{(2)}$ respectivement. Pour plus de détails, nous rappelons la définition des susdits ordres stochastiques[40]. Supposons que nous ayons $F_1(x)$ et $F_2(x)$ deux fonctions de distribution de variables aléatoires non négatives. Alors

$$- F_1 \leq_{st} F_2 \text{ si et seulement si } F_1(x) \geq F_2(x) \text{ pour tout } x \geq 0.$$

- $F_1 \leq_v F_2$ si et seulement si $\int_x^\infty (1 - F_1(u)) du \leq \int_x^\infty (1 - F_2(u)) du$ pour tout $x \geq 0$.
- $F_1 \leq_L F_2$ si et seulement si $\int_0^\infty e^{-sx} dF_1(x) \geq \int_0^\infty e^{-sx} dF_2(x)$ pour tout $s \geq 0$.

Si ces variables sont discrètes et $p^{(1)} = (p_n^{(1)})$, $p^{(2)} = (p_n^{(2)})$ sont leurs distributions correspondantes, les susdites définitions prennent la forme suivante :

- $p^{(1)} \leq_{st} p^{(2)}$ si et seulement si $\sum_{n=m}^\infty p_n^{(1)} \leq \sum_{n=m}^\infty p_n^{(2)}$ pour tout m .
- $p^{(1)} \leq_v p^{(2)}$ si et seulement si $\sum_{n=m}^\infty \sum_{k=n}^\infty p_k^{(1)} \leq \sum_{n=m}^\infty \sum_{k=n}^\infty p_k^{(2)}$ pour tout m .
- $p^{(1)} \leq_L p^{(2)}$ si et seulement si $\sum_{n=0}^\infty z^n p_n^{(1)} \geq \sum_{n=0}^\infty z^n p_n^{(2)}$ pour tout $z \in [0, 1]$.

1.5.1. Monotonie interne et externe de l'opérateur de transition

Soit T l'opérateur de transition de la chaîne de Markov induite $\{N_i, i \geq 1\}$ du système $M/G/1$ avec rappels classiques¹, qu'à chaque distribution $p = (p_n)$ associe une distribution $T_p = q = (q_m)$ telle que $q_m = \sum_n p_n p_{nm}$. On a le théorème suivant :

Théorème 1. *L'opérateur T est monotone par rapport à l'ordre stochastique fort \leq_{st} (à l'ordre stochastique convexe \leq_v), c'est-à-dire pour n'importe quelles deux distributions $p^{(1)}$ et $p^{(2)}$ l'inégalité $p^{(1)} \leq_{st} p^{(2)}$ (l'inégalité $p^{(1)} \leq_v p^{(2)}$) implique $Tp^{(1)} \leq_{st} Tp^{(2)}$ ($Tp^{(1)} \leq_v Tp^{(2)}$).*

1. Les rappels sont dits classiques lorsque la probabilité de rappel dans un intervalle de temps $]t, t + \Delta t[$, sachant que n clients sont en orbite à l'instant t , est $n\theta\Delta t + o(\Delta t)$ où θ est le paramètre de la distribution de la durée entre deux rappels successifs d'une même source de l'orbite

Ce théorème implique que le nombre de clients dans le système aux instants de départ forme une suite croissante par rapport à l'ordre stochastique fort \leq_{st} et convexe \leq_v si le système est vide à l'instant $t = 0$.

Le lemme suivant donne les conditions sous lesquelles la comparaison des distributions de probabilité du nombre de client arrivant pendant la durée de service d'un client des deux systèmes $M^{(1)}$ et $M^{(2)}$ aura lieu. En plus, il est utile dans la démonstration de la comparabilité de deux opérateurs de transition (autrement dit : la monotonie externe).

Lemme 1. *Si $\lambda^{(1)} \leq \lambda^{(2)}$ et $B^{(1)}(x) \leq_s B^{(2)}(x)$ alors $\{k_n^{(1)}\} \leq_s \{k_n^{(2)}\}$, où S est st , v ou bien L , et*

$$k_n^{(i)} = \int_0^\infty \frac{(\lambda^{(i)} x)^n}{n!} e^{-\lambda^{(i)} x} dB^{(i)}(x), \quad i \in \{1, 2\}.$$

Dénotons par $T^{(1)}$, $T^{(2)}$ les opérateurs de transition de la chaîne de Markov induite du système $M^{(1)}$ et $M^{(2)}$, respectivement. Le théorème suivant donne les conditions de comparabilité de $T^{(1)}$ et $T^{(2)}$ par rapport aux ordres stochastique et convexe.

Théorème 2. *Si $\lambda^{(1)} \leq \lambda^{(2)}$, $\theta^{(1)} \geq \theta^{(2)}$ et $B^{(1)}(x) \leq_s B^{(2)}(x)$, où S est st ou bien v , alors $T^{(1)} \leq_s T^{(2)}$, i.e. pour n'importe quelle distribution p nous avons $T^{(1)}p \leq_s T^{(2)}p$.*

1.5.2. Inégalités stochastiques pour la distribution stationnaire du nombre de clients dans le système. Les propriétés de monotonie de la chaîne de Markov induite

Dénotons par $\{N_i^{(1)}, i \geq 1\}$, $\{N_i^{(2)}, i \geq 1\}$ les chaînes de Markov induites correspondantes aux systèmes $M^{(1)}$ et $M^{(2)}$, et par $\pi_n^{(1)}$, $\pi_n^{(2)}$ leurs distributions stationnaires respectivement. Le théorème suivant montre les conditions sous lesquelles les distributions $\pi_n^{(1)}$ et $\pi_n^{(2)}$ sont comparables.

Théorème 3. Si $\lambda^{(1)} \leq \lambda^{(2)}$, $\theta^{(1)} \geq \theta^{(2)}$ et $B^{(1)}(x) \leq_S B^{(2)}(x)$, où S est st ou bien v , alors les inégalités $\{\pi_n^{(1)}\} \leq_S \{\pi_n^{(2)}\}$ ont lieu.

A l'aide du théorème 3, nous pouvons établir des bornes stochastiques insensibles pour la distribution stationnaire du nombre de clients dans le système.

Théorème 4. Si dans le modèle $M/G/1$ avec rappels classiques, la distribution du temps de service $B(x)$ est $NBUE^1$ (ou bien $NWUE^2$), alors la distribution stationnaire du nombre de clients dans le système est inférieure (supérieure) par rapport à l'ordre convexe \leq_v à la distribution stationnaire du nombre de clients dans le système $M/M/1$ avec rappels classiques.

1.5.3. Inégalités stochastiques pour le nombre moyen de clients servis durant la période d'activité

Dénotons par $I^{(1)}$ et $I^{(2)}$ le nombre de clients servis durant une période d'activité dans les systèmes $M^{(1)}$ et $M^{(2)}$ respectivement. Le théorème suivant donne les conditions sur les paramètres de deux systèmes $M^{(1)}$ et $M^{(2)}$, sous lesquelles la moyenne de $I^{(1)}$ et celle de $I^{(2)}$ sont comparables.

Théorème 5. Si $\lambda^{(1)} \leq \lambda^{(2)}$, $\theta^{(1)} \geq \theta^{(2)}$ et $B^{(1)}(x) \leq_L B^{(2)}(x)$, alors l'inégalité $E(I^{(1)}) \leq E(I^{(2)})$ a lieu.

A l'aide du théorème 5, nous pouvons établir une borne stochastique insensible pour le nombre moyen de clients servis durant la période d'activité dans le système $M/G/1$ avec rappels classiques.

-
1. Soient X et X_t des variables aléatoires représentant respectivement la durée de vie d'un élément et la durée de vie résiduelle, et soient F et F_t ($F_t(x) = P(X \leq x+t \setminus X > t)$) leur distributions respective. On dit que F est $NBUE$ (New Better Used in Expectation), si $E[X_t] \leq E[X]$ pour tout t .
 2. Soient X et X_t des variables aléatoires représentant respectivement la durée de vie d'un élément et la durée de vie résiduelle, et soient F et F_t ($F_t(x) = P(X \leq x+t \setminus X > t)$) leur distributions respective. On dit que F est $NWUE$ (New Worse than Used in Expectation), si $E[X] \leq E[X_t]$ pour tout t .

Théorème 6. *Pour un système $M / G / 1$ avec rappels,*

$$E(I) \leq \frac{1}{1-\rho} \left\{ \frac{\lambda}{\mu} \int_0^1 \frac{1-e^{\rho(u-1)}}{e^{\rho(u-1)}-u} du \right\}.$$

Si $B(x)$ est NBUE, alors $E(I) \geq (1-\rho)^{-\lambda/\mu-1}$.

1.6. Autres modèles

L'analyse des modèles avec arrivées par groupe se trouve dans [25], et celle des systèmes avec rappels et clients non-persistants est réalisée dans [19]. Les systèmes à multiserveurs sont décrits par Hanschke [28] ainsi que par Boots et Tijms [8]. Les modèles avec population des clients potentiels finie ont fait l'objet des investigations de Artalejo [2] et également de Falin et Artalejo [21]. Ces dernières années, on observe un grand intérêt pour les systèmes de files d'attente avec rappels et arrivées négatives, où l'arrivée d'un client négatif a pour effet la suppression d'un ou plusieurs clients dans le système, Artalejo [3], Artalejo et Gomez-Corral [4]. Un autre axe de recherche concerne les systèmes de files d'attente avec rappels où le flux d'appels primaires est modélisé comme BMAP (Batch Markovian Arrival Process), Dudin et Klimenok [18].

1.7. Conclusion

Nous avons présenté quelques modèles de files d'attente avec rappels. On constate que leur analyse présente de grandes difficultés analytiques. Cette analyse devient encore plus difficile lorsqu'on ajoute les interruptions du service au phénomène de rappels

2

Interruption de service dans les systèmes d'attente avec rappels

2.1. Introduction

L'interruption de service est un événement qui empêche la continuité du service des clients. Son influence ne peut pas être étudiée sans spécifier l'interaction entre le processus d'interruption et le processus de service. Cette interaction peut être *sans préemption*, c'est-à-dire qu'elle n'intervient pas pendant le service d'un client, ou *avec préemption* dans le cas contraire. En outre, après chaque interruption avec préemption, le client reprend le service à partir du point d'interruption ou dès le début Gaver [26].

Les modèles de files d'attente à interruption de service ont plusieurs applications dans les réseaux de télécommunication, et de nombreuses études ont été menées dans ce sens. En général, les interruptions peuvent être dues à diverses causes : la possibilité de vacance du serveur, les pannes du serveur ou l'arrivée de clients de haute priorité.

Dans ce chapitre, nous passons en revue les systèmes de files d'attente avec rappels et interruptions de service. Nous commençons par les modèles avec vacances (ils représentent le cas le plus général). Ici, le serveur prend occasionnellement une vacance d'une durée aléatoire, qui peut être utilisée pour accomplir une ou plusieurs tâches secondaires (prioritaires ou de maintenance). Puis, nous considérons les modèles avec priorité ainsi que les systèmes de files d'attente avec serveurs non fiables.

2.2. Modèles d'attente avec rappels et vacances du serveur

Les systèmes de files d'attente avec vacances peuvent être classifiés selon la discipline de service. En effet, on a

- **Discipline de service exhaustif (complet) :**
Chaque fois que le serveur retourne d'une vacance, il sert tous les clients se trouvant dans la file d'attente, et ceci avant de commencer une autre vacance.
- **Discipline de service avec barrière :**
Lorsque le serveur retourne d'une vacance, il sert seulement les clients qui étaient en attente à son arrivée.
- **Discipline de service limité (N-Policy) :**
A son retour de la vacance, le serveur servira au plus N clients et puis commencera une autre vacance.
- **Discipline de service limité :**
Après la période d'activité d'une durée fixe, le serveur prend une vacance même s'il y a des clients dans la file.

En outre, dans le cas d'un service complet, si le serveur, qui retourne d'une vacance, trouve la file d'attente vide, il exécute l'une des deux actions suivantes :

- Sous le schéma de *vacances multiples*, le serveur commence immédiatement une autre vacance et continue à prendre des vacances successives jusqu'à ce qu'il trouve au moins un client dans la file ;
- Sous le schéma de *vacance unique*, le serveur attend jusqu'à la fin de la prochaine période d'activité pendant laquelle un client au moins sera servi, avant de commencer une autre vacance (autrement dit, il y a exactement une seule vacance à la fin de chaque période d'activité).

Une comparaison des différentes politiques de service dans les modèles avec vacances est donnée dans [42].

2.2.1. Description du modèle

Considérons un système avec rappels de type M/G/1. Le flux d'arrivée est engendré par un nombre fini m de sources ($2 \leq m < \infty$). Une source peut être active ou inactive à la date t . Lorsqu'une source est active, elle peut engendrer un appel primaire durant un intervalle $]t, t + \Delta t[$ avec une probabilité $\lambda \Delta t + o(\Delta t)$, et elle devient immédiatement inactive après cela. Après la fin du service du client correspondant, elle devient encore active. À tout moment, le serveur peut être seulement dans l'un des trois états : libre, occupé ou en vacance. Si le serveur est occupé ou en vacance au moment d'arrivée d'un appel primaire, la source engendre un flux poissonnien d'appels répétés (appels secondaires) de taux $\theta > 0$, jusqu'à ce qu'elle trouve le serveur libre. Si le serveur est libre au moment d'arrivée d'un appel primaire ou secondaire, il commence le service avec une probabilité α_k ou prend une vacation avec une probabilité $\bar{\alpha}_k = 1 - \alpha_k$ (k est le nombre de clients en orbite à l'instant de la décision). La durée de service suit une loi générale de fonction de répartition $B(x)$ de transformée de Laplace-Stieltjes $\tilde{B}(s)$, $Re(s) > 0$. On a les moments d'ordre k , $\beta_k = (-1)^k \tilde{B}^{(k)}(0)$. Les durées des vacances sont des variables aléatoires indépendantes de fonction de répartition $V(x)$, de moyenne v . Toutes les variables précédemment introduites sont supposées mutuellement indépendantes.

2.2.2. Analyse du modèle

L'état du système à la date t peut être décrit par le processus stochastique $\{C(t); N_o(t); \xi(t); t \geq 0\}$, où $N_o(t)$ est le nombre de clients en orbite, $C(t)$ est l'état du serveur à l'instant t

$$C(t) = \begin{cases} 0 & \text{si le serveur est libre} \\ 1 & \text{si le serveur est occupé} \\ 2 & \text{si le serveur est en vacance} \end{cases},$$

et $\xi(t)$ est la durée écoulée de service en cours (si $C(t) = 1$) ou celle de vacation (si $C(t) = 2$). Ce modèle peut être résolu à l'aide de la méthode des caractéristiques pour trouver la solution des équations différentielles aux dérivées partielles [34].

Soient

$$p_{0n} = \lim_{t \rightarrow \infty} p(C(t) = 0, N_o(t) = n) \quad 0 \leq n \leq m;$$

$$p_{1n}(x) = \lim_{t \rightarrow \infty} \frac{d}{dx} p(C(t) = 1, \xi(t) \leq x, N_o(t) = n) \quad 0 \leq n \leq m - 1;$$

$$p_{2n}(x) = \lim_{t \rightarrow \infty} \frac{d}{dx} p(C(t) = 2, \xi(t) \leq x, N_o(t) = n) \quad 0 \leq n \leq m.$$

Ces distributions s'obtiennent à l'aide du schéma suivant :

$$p_{00} = \left(\sum_{n=0}^m (1 + \lambda(m-n)(\alpha_n \beta_1 + \bar{\alpha}_n v) + n\theta(\alpha_{n-1} \beta_1 + \bar{\alpha}_{n-1} v)) \varphi_n \right)^{-1};$$

$$p_{0n} = \varphi_n p_{00} \quad 1 \leq n \leq m;$$

$$p_{1n}(x) = (1 - B(x)) \sum_{k=0}^n \left\{ (C_{m-n-1}^{m-k-1}) e^{-(m-n-1)\lambda x} (1 - e^{-\lambda x})^{n-k} \right. \\ \left. \times (\lambda(m-k)\varphi_k + (k+1)\theta\varphi_{k+1})\alpha_k p_{00} \right\}, \quad 0 \leq n \leq m - 1$$

$$p_{2n}(x) = (1 - V(x)) \sum_{k=0}^n \left\{ (C_{m-n}^{m-k}) e^{-(m-n)\lambda x} (1 - e^{-\lambda x})^{n-k} \right. \\ \left. \times (\lambda(m-k+1)\varphi_{k-1} + k\theta\varphi_k)\bar{\alpha}_{k-1} p_{00} \right\}, \quad 0 \leq n \leq m;$$

Où $\{\varphi_i\}_{i=0}^k$ vérifie

$$\varphi_{i+1} = \frac{1}{q_{i+1}} \sum_{k=0}^i \sum_{j=i+1}^m q_{kj} \varphi_k, \quad 0 \leq i \leq m - 1, \varphi_0 = 1;$$

$$q_{kj} = a_{ki} \alpha_k (m-k)\lambda + a_{k-1,i} \alpha_{k-1} k\theta + b_{k+1,i} \bar{\alpha}_k (m-k)\lambda$$

$$\begin{aligned}
& +b_{k,i}\bar{\alpha}_{k-1} k\theta, & 0 \leq k \leq i + 1; \\
a_{ki} = & \begin{cases} C_{m-i-1}^{m-k-1} \int_0^{\infty} e^{-(m-i-1)\lambda x} (1 - e^{-\lambda x})^{i-k} dB(x) & \text{si } 0 \leq k \leq i < m ; \\ 0 & \text{sinon} \end{cases} \\
b_{ki} = & \begin{cases} C_{m-i}^{m-k} \int_0^{\infty} e^{-(m-i)\lambda x} (1 - e^{-\lambda x})^{i-k} dV(x) & \text{si } 0 \leq k \leq i < m . \\ 0 & \text{sinon} \end{cases}
\end{aligned}$$

2.2.3. Mesures de performance [2]

- Probabilité p_1 d'utilisation du serveur

$$p_1 = \beta_1 p_{00} \sum_{n=0}^m (\alpha_n \lambda (m - n) + \alpha_{n-1} n \theta) \varphi_n.$$

- Nombre moyen de sources en orbite \bar{n}_o

$$\bar{n}_o = m - \frac{vp_1}{\lambda}.$$

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

$$\bar{w} + \beta_1 = \frac{m}{vp_1} - \frac{1}{\lambda}.$$

2.3. Modèles d'attente avec rappels et priorité

Ces modèles ont les mêmes caractéristiques que le modèle précédemment étudié dans le chapitre 1. La seule différence c'est qu'il ne s'agit pas d'un seul type de clients. On a deux types de clients qui arrivent vers le système : prioritaires et non prioritaires. En général, la priorité peut être relative ou absolue.

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.

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 entre immédiatement dans le serveur. Le client, dont le service est interrompu par l'arrivée du client prioritaire, quitte le système ou entre en orbite.

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

Par la suite nous étudions un modèle de type $M_1M_2/G_1G_2/1$ avec rappels et priorité.

2.3.1. Modèle d'attente avec rappels de type $M_1M_2/G_1G_2/1$ et priorité relative

2.3.1.1. Description du modèle

Il s'agit d'un modèle où deux types de clients (type1-prioritaire et type2-non prioritaire) arrivent dans le système selon un processus de Poisson des taux λ_1 et λ_2 respectivement. Le taux total est $\lambda = \lambda_1 + \lambda_2$. Le comportement d'un client, qui trouve à son arrivée le serveur occupé par le service d'un client, dépend de son type de priorité. S'il est prioritaire, il prend place dans la file d'attente. S'il est non prioritaire, il entre en orbite et devient source d'appels secondaires (rappels). La durée entre deux rappels successifs d'une même source secondaire est exponentiellement distribuée de paramètre $\theta > 0$. Les clients en orbite peuvent être servis lorsque la file d'attente est vide. Les durées de service suivent une loi générale de fonction de répartition $B_i(x)$ de transformée de Laplace-Stieltjes $\tilde{B}_i(s)$, $Re(s) > 0$ 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 (intensité instantanée du service) est

$$b_i(x) = \frac{B_i'(x)}{1 - B_i(x)}, \quad i = 1, 2.$$

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 de type i . De Choi et Park [14], la condition pour que le système se trouve en régime stationnaire est $\rho < 1$. 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 de deux types arrivant dans le système durant le temps de service d'un client de type $i = 1, 2$. Sa fonction génératrice est donnée par :

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

2.3.1.2. Distributions stationnaires

L'état du système à la date t peut être décrit à l'aide d'un processus de Markov à temps continu, tel que

$$\{C(t); N_f(t); N_o(t); \xi(t) \quad t \geq 0\},$$

où

$$C(t) = \begin{cases} 0 & \text{si le serveur est libre} \\ 1 & \text{si le serveur est occupé (client de type 1),} \\ 2 & \text{si le serveur est occupé (client de type 2)} \end{cases}$$

$N_f(t)$ est le nombre de clients (type1) dans la file d'attente, $N_o(t)$ est le nombre de clients (type2) en orbite, $\xi(t)$ est une variable aléatoire supplémentaire à valeurs dans \mathbb{R}^+ et désignant la durée de service écoulé à la date t .

Supposons $\rho < 1$. Les probabilités d'état sont définies par

$$p_{0,i,j} = \lim_{t \rightarrow \infty} p(C(t) = 0, N_f(t) = i, N_o(t) = j);$$

$$p_{1,i,j}(x) = \lim_{t \rightarrow \infty} \frac{d}{dx} p(C(t) = 1, N_f(t) = i, N_o(t) = j, \xi(t) \leq x);$$

$$p_{2,i,j}(x) = \lim_{t \rightarrow \infty} \frac{d}{dx} p(C(t) = 2, N_f(t) = i, N_o(t) = j, \xi(t) \leq x);$$

$$p_{1,i,j} = \int_0^{\infty} p_{1,i,j}(x) dx;$$

$$p_{2,i,j} = \int_0^{\infty} p_{2,i,j}(x) dx.$$

Les probabilités $p_{0,i,j}$, $p_{1,i,j}(x)$ et $p_{2,i,j}(x)$ vérifient le système d'équations de balance

$$(\lambda_1 + \lambda_2 + j\theta)p_{0,0,j} = \int_0^\infty (p_{1,0,j}(x)b_1(x) + p_{2,0,j}(x)b_2(x)) dx;$$

$$p_{0,i,j} = 0 \quad \text{si } i \geq 1;$$

$$p'_{1,i,j}(x) = -(\lambda_1 + \lambda_2 + b_1(x))p_{1,i,j}(x) + \lambda_1 p_{1,i-1,j}(x) + \lambda_2 p_{1,i,j-1}(x);$$

$$p'_{2,i,j}(x) = -(\lambda_1 + \lambda_2 + b_2(x))p_{2,i,j}(x) + \lambda_1 p_{2,i-1,j}(x) + \lambda_2 p_{2,i,j-1}(x);$$

$$p_{1,i,j}(0) = \lambda_1 p_{0,0,j} \delta_{i,0} + \int_0^\infty p_{1,i+1,j}(x)b_1(x) dx + \int_0^\infty p_{2,i+1,j}(x)b_2(x) dx;$$

$$p_{2,i,j}(0) = \begin{cases} 0 & \text{si } i \geq 1 \\ \lambda_2 p_{0,0,j} + (j+1)\theta p_{0,0,j+1} & \text{si } i = 0 \end{cases}$$

où $\delta_{i,0}$ est la fonction de Kronecker.

Comme la résolution de ce système est très compliquée, nous utilisons la méthode des fonctions génératrices. A cet effet, on introduit les fonctions génératrices partielles suivantes

$$P_0(z_1, z_2) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p_{0,i,j} z_1^i z_2^j;$$

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

$$P_2(z_1, z_2) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p_{2,i,j} z_1^i z_2^j;$$

$$P_2(z_1, z_2, x) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p_{2,i,j}(x) z_1^i z_2^j;$$

$$P_2(z_1, z_2) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p_{2,i,j} z_1^i z_2^j.$$

Vu que $N_f(t) = 0$ si $C(t) = 0$, alors $p_{0,i,j} = 0$ pour $i \geq 1$. Par conséquent, $P_0(z_1, z_2)$ ne dépend pas de z_1 : $P_0(z_1, z_2) = P_0(0, z_2) = P_0(z_2)$. Le système d'équations de balance devient

$$\begin{aligned}
& \theta z_2 \frac{dP_0(z_2)}{dz_2} \\
&= -(\lambda_1 + \lambda_2)P_0(z_2) + \int_0^\infty P_1(0, z_2, x) b_1(x) dx \\
&+ \int_0^\infty P_2(0, z_2, x) b_2(x) dx; \tag{2.1}
\end{aligned}$$

$$\frac{\partial P_1(z_1, z_2, x)}{\partial x} = -(\lambda_1 - \lambda_1 z_1 + \lambda_2 - \lambda_2 z_2 + b_1(x))P_1(z_1, z_2, x); \tag{2.2}$$

$$\frac{\partial P_2(z_1, z_2, x)}{\partial x} = -(\lambda_1 - \lambda_1 z_1 + \lambda_2 - \lambda_2 z_2 + b_2(x))P_2(z_1, z_2, x); \tag{2.3}$$

$$\begin{aligned}
& z_1 P_1(z_1, z_2, 0) \\
&= \int_0^\infty (P_1(z_1, z_2, x) - P_1(0, z_2, x)) b_1(x) dx + \int_0^\infty (P_2(z_1, z_2, x) - P_2(0, z_2, x)) b_2(x) dx \\
&+ \lambda_1 z_1 P_0(z_2); \tag{2.4}
\end{aligned}$$

$$P_2(z_1, z_2, 0) = \lambda_2 P_0(z_2) + \theta \frac{dP_0(z_2)}{dz_2}. \tag{2.5}$$

La résolution du système (2.1) – (2.5) donne

$$\begin{aligned}
P_1(z_1, z_2, x) &= P_0(z_2) \left[[(\lambda_1 - \lambda_1 h(z_2) + \lambda_2 - \lambda_2 K_2(h(z_2), z_2))(K_2(h(z_2), z_2) - z_2) \right. \\
&\quad \left. + (\lambda_1 - \lambda_1 z_1 + \lambda_2 - \lambda_2 K_2(z_1, z_2))(z_2 - K_2(h(z_2), z_2))] \right. \\
&\quad \left. / (K_2(h(z_2), z_2) - z_2)(z_1 - K_1(z_1, z_2)) \right] \\
&\quad \times (1 - B_1(x)) e^{-(\lambda_1(1-z_1) + \lambda_2(1-z_2))x};
\end{aligned}$$

$$P_2(z_1, z_2, x) = P_0(z_2) \frac{\lambda_1 - \lambda_1 h(z_2) + \lambda_2 - \lambda_2 z_2}{K_2(h(z_2), z_2) - z_2} \times (1 - B_2(x)) e^{-(\lambda_1(1-z_1) + \lambda_2(1-z_2))x};$$

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

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

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

avec

$$P_0(1) = 1 - \rho_1 - \rho_2.$$

2.3.1.3. Mesures de performance

A l'aide des fonctions génératrices $P_0(z_2)$, $P_1(z_1, z_2)$ et $P_2(z_1, z_2)$, on obtient les mesures de performance suivantes :

- Probabilité $P_1(P_2)$ que le serveur est occupé par le service d'un client type1 (d'un client type2)

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

- Nombre moyen de clients dans la file d'attente \bar{n}_f

$$\bar{n}_f = E[N_f(t)] = \lim_{t \rightarrow \infty} E[N_f(t)] = \frac{\lambda_1(\lambda_1\beta_{1,2} + \lambda_2\beta_{2,2})}{2(1 - \rho_1)}.$$

- Nombre moyen de clients en orbite \bar{n}_o

$$\bar{n}_o = E[N_o(t)] = \lim_{t \rightarrow \infty} E[N_o(t)] = \frac{\lambda_2(\lambda_1\beta_{1,2} + \lambda_2\beta_{2,2})}{2(1 - \rho_1)(1 - \rho_1 - \rho_2)} + \frac{\lambda_2(\rho_1 + \rho_2)}{\theta(1 - \rho_1 - \rho_2)}.$$

- Temps moyen d'attente \bar{W}_i d'un client type $i = 1, 2$ (d'après la formule de Little)

$$\bar{W}_1 = \frac{(\lambda_1\beta_{1,2} + \lambda_2\beta_{2,2})}{2(1 - \rho_1)}; \bar{W}_2 = \frac{(\lambda_1\beta_{1,2} + \lambda_2\beta_{2,2})}{2(1 - \rho_1)(1 - \rho_1 - \rho_2)} + \frac{(\rho_1 + \rho_2)}{\theta(1 - \rho_1 - \rho_2)}.$$

2.4. Modèles avec rappels et serveur sujet à des pannes aléatoires

Ces modèles ont les mêmes caractéristiques que les modèles précédemment étudiés dans le chapitre 1. Ce qui les distingue est le risque des pannes. Ces pannes peuvent se produire aussi bien lorsque le serveur est libre (panne passive) que lorsqu'il est occupé par le service d'un client (panne active). Dans le cas des pannes actives, le client dont le service est interrompu soit il abandonne le serveur, soit il continue à le tenir.

Dans le cas où le client décide d'abandonner le serveur, il prend l'un des deux chemins : quitter le système (panne avec perte définitive) ou entrer en orbite (panne avec perte momentanée). Nous incorporons ces deux possibilités en introduisant un facteur de

récupération c . Ce dernier désigne la fraction du nombre de clients qui persistent en orbite après l'interruption afin de récupérer leur service, sachant que ces clients perdent toute la mémoire de leurs services antérieurs et que leur comportement ne diffère en rien de celui des clients qui se trouvent déjà en orbite.

Dans le cas où le client reste auprès du serveur jusqu'à ce que la réparation soit faite, le client en question reprend son service à partir du point d'interruption (panne conservatrice) ou il reprend à zéro (panne non conservatrice).

Par la suite, nous étudions le modèle M/G/1 avec rappels et serveur sujet à des pannes actives et passives. Il existe deux approches pour décrire ce type de modèle : la première approche est basée sur la théorie des processus de Markov par morceaux [1]; la seconde utilise les outils de la théorie des processus régénérateurs [32]. Nous allons utiliser la deuxième approche.

2.4.1. Modèle d'attente avec rappels de type M/G/1 et serveur non fiable

2.4.1.1. Description du modèle

Un modèle d'attente de type M/G/1 avec rappels et serveur non fiable est celui où les appels primaires arrivent selon un processus homogène de Poisson de taux $\lambda > 0$. Le service des clients est assuré par un serveur. Les durées de service suivent une loi générale de fonction de répartition $B(x)$ de transformée de Laplace-Stieltjes $\tilde{B}(s)$, $Re(s) > 0$. Le taux d'accomplissement conditionnel (intensité du service instantanée) est

$$b(x) = \frac{B'(x)}{1 - B(x)}.$$

Le serveur est sujet à des pannes aléatoires qui peuvent être actives ou passives et qui surviennent à des taux exponentiels $\alpha_1 > 0$ et $\alpha_0 > 0$, respectivement.

Lorsque le serveur tombe en panne, il sera immédiatement réparé. Dans le cas d'une panne active, la durée exigée à la réparation est une variable aléatoire d_1 de fonction de répartition $R_1(x)$, de transformée de Laplace-Stieltjes $\tilde{R}_1(s)$. Dans le cas d'une panne passive, la même durée est une variable aléatoire d_0 de fonction de répartition $R_0(x)$, de transformée de Laplace-Stieltjes $\tilde{R}_0(s)$.

Si un appel primaire arrive et trouve le serveur libre et opérationnel, il occupe immédiatement le serveur. Sinon, s'il le trouve occupé ou en panne, il rejoint l'orbite et devient source d'appels secondaires (rappels). La durée entre deux rappels successifs d'une même source secondaire est exponentiellement distribuée de paramètre $\theta > 0$. Un client, dont le service a été interrompu par une panne active, quitte le serveur soit pour entrer en orbite avec une probabilité c , soit pour abandonner définitivement le système avec une probabilité $(1 - c)$. Toutes les variables précédemment décrites sont supposées mutuellement indépendantes.

Nous nous intéressons au comportement stationnaire du processus stochastique $\{C(t); N_o(t); t \geq 0\}$, où $N_o(t)$ est le nombre de clients en orbite à la date t et $C(t)$ (l'état du serveur à la date t) est défini comme suit :

$$C(t) = \begin{cases} 0 & \text{si le serveur est libre et opérationnel à la date } t ; \\ 1 & \text{si le serveur est occupé et opérationnel à la date } t ; \\ 2 & \text{si le serveur est en panne à la date } t . \end{cases}$$

Ce processus n'est pas un processus de Markov. Pour le rendre markovien, on introduit une variable supplémentaire $\xi(t)$ à valeurs dans \mathbb{R}^+ , désignant la durée de service écoulée (si $C(t) = 1$) ou la durée de réparation écoulée (si $C(t) = 2$) à la date t . D'une autre manière, on étudie sa chaîne de Markov induite $\{N_n; n \geq 0\}$, où $N_n = N_o(\xi_n)$ est le nombre de clients en orbite à la date ξ_n lorsque le serveur devient libre et opérationnel pour la $n^{\text{ème}}$ fois ($C(\xi_n) = 0; \forall n \geq 0$). La fonction génératrice de la distribution stationnaire $\{\pi_k = \lim_{n \rightarrow \infty} p(N_n = N(\xi_n) = k); k \geq 0\}$ de la chaîne de Markov induite est :

$$\Phi(z) = K \frac{\alpha(z)}{\beta(z)} \times \exp \left\{ -\frac{1}{\theta} \int_1^z \frac{1}{u} \frac{(\alpha_0 + \lambda)\beta(u) - \alpha(u)}{\beta(u)} du \right\}, \quad (2.6)$$

où

$$\begin{aligned} \beta(z) &= (1 - \tilde{\beta}(z)), \\ \alpha(z) &= \tilde{\alpha}(z) - (\alpha_0 + \lambda)\tilde{\beta}(z), \end{aligned}$$

avec

$$\tilde{\alpha}(z) = \alpha_0 \tilde{R}_0(\lambda - \lambda z) + \lambda \alpha_1 \frac{(1 - \tilde{B}(\lambda - \lambda z + \alpha_1)) \tilde{R}_1(\lambda - \lambda z)(1 - c(1 - z))}{\lambda - \lambda z + \alpha_1} + \lambda \tilde{B}(\lambda - \lambda z + \alpha_1)$$

et

$$\tilde{\beta}(z) = \alpha_1 \frac{(1 - \tilde{B}(\lambda - \lambda z + \alpha_1)) \tilde{R}_1(\lambda - \lambda z)}{\lambda - \lambda z + \alpha_1} \left(c + \frac{1 - c}{z} \right) + \frac{1}{z} \tilde{B}(\lambda - \lambda z + \alpha_1).$$

Ici, K est une constante arbitraire qui s'obtient à l'aide de la condition de normalisation $\Phi(1) = 1$ et l'application de la règle de l'Hopital. En effet,

$$K = \frac{1 - \rho}{\lambda(1 + \alpha_0 E[d_0]) + \alpha_0(1 - \rho)},$$

où

$$\rho = \lambda \frac{1 - \tilde{\beta}(\alpha_1)}{\alpha_1} \left[1 + \alpha_1 \left(E[d_1] + \frac{c}{\lambda} \right) \right].$$

Il est clair que cette chaîne est en régime stationnaire (récurrente et positive) si $\rho < 1$. De plus, cette condition est indépendante du taux de rappels θ .

2.4.1.2. Distributions stationnaires

Le processus $\{N_o(t); t \geq 0\}$ est un processus de Markov régénérateur avec le processus de Markov de renouvellement induit $\{(N_{on}, \xi_n); n \geq 0\}$.

Soit

$\tau_{ij} = E(\text{duree de sejour du processus } \{N_o(t); t \geq 0\} \text{ dans l'état } j \text{ pendant } [0, \xi_1) \setminus (0) = i)$, et $\tau_i = E(\xi_1 \setminus N_o(0) = i)$. Comme $\{N_o(t); t \geq 0\}$ est un processus de Markov régénérateur, la distribution limite du nombre de clients en orbite existe lorsque $t \rightarrow \infty$, et est donnée par [32]

$$p_j = \lim_{t \rightarrow \infty} p(N_o(t) = j \mid N_o(0) = i) = \frac{\sum_{i=0}^{\infty} \pi_i \tau_{ij}}{A},$$

où

$$A = \sum_{i=0}^{\infty} \pi_i \tau_i.$$

Pour calculer sa fonction génératrice, on doit trouver les fonctions génératrices des probabilités limites $p_{(j,k)} = \lim_{t \rightarrow \infty} p((N_o(t), C(t)) = (j, k)), k = 0, 1, 2$ et $j \geq 0$, car $p_j = p_{(j,0)} + p_{(j,1)} + p_{(j,2)}$. Ces probabilités limites sont données par

$$p_{(j,k)} = \frac{1}{A} \sum_{i=0}^{\infty} \pi_i \tau_{i(j,k)},$$

où

$$\tau_{i(j,k)} = E(\text{durée de séjour du processus } \{N_o(t); C(t); t \geq 0\} \text{ dans l'état } (j, k) \text{ pendant } [0, \xi_1] \mid N_o(0) = i).$$

Leurs fonctions génératrices sont données par les expressions suivantes [32]

$$P_0(z) = \sum_{j=0}^{\infty} p_{(j,0)} z^j = \frac{1}{A} M(z);$$

$$P_1(z) = \sum_{j=0}^{\infty} p_{(j,1)} z^j = \frac{1}{A} \left(\frac{1 - \tilde{B}(\alpha_1 + \lambda - \lambda z)}{\alpha_1 + \lambda - \lambda z} \right) (\lambda M(z) + \theta M'(z));$$

$$\begin{aligned} P_2(z) &= \sum_{j=0}^{\infty} p_{(j,2)} z^j \\ &= \frac{1}{A} \left[\frac{1 - \tilde{R}_0(\lambda - \lambda z)}{\lambda - \lambda z} \alpha_0 M(z) + M(z) N(z) (\lambda(1 - c) + \lambda cz) \right. \\ &\quad \left. + M'(z) N(z) (\theta(1 - c) + \theta cz) \right]. \end{aligned}$$

où

$$N(z) = \alpha_1 \frac{1 - \tilde{B}(\alpha_1 + \lambda - \lambda z)}{\alpha_1 + \lambda - \lambda z} \frac{1 - \tilde{R}_1(\lambda - \lambda z)}{\lambda - \lambda z},$$

$$M(z) = \sum_{j=0}^{\infty} \frac{\pi_j z^j}{\lambda + \alpha_0 + j\theta}.$$

Enfin, la fonction génératrice de la distribution limite du nombre de clients en orbite p_j est

$$\begin{aligned} Q(z) &= P_0(z) + P_1(z) + P_2(z) \\ &= \frac{1}{A} \left[M(z) + \frac{1 - \tilde{B}(\alpha_1 + \lambda - \lambda z)}{\alpha_1 + \lambda - \lambda z} (\lambda M(z) + \theta M'(z)) + \frac{1 - \tilde{R}_0(\lambda - \lambda z)}{\lambda - \lambda z} \alpha_0 M(z) \right. \\ &\quad + M(z)N(z)(\lambda(1 - c) + \lambda cz) \\ &\quad \left. + M'(z)N(z)(\theta(1 - c) + \theta cz) \right]. \end{aligned} \quad (2.7)$$

Du fait que $Q(1) = 1$, nous obtenons

$$A = K(1 + \alpha_0 E[d_0]) + (1 - \tilde{B}(\alpha_1)) (1 - \alpha_0 K) \left(E[d_1] + \frac{1}{\alpha_1} \right).$$

La distribution stationnaire de l'état du serveur s'obtient à partir des fonctions génératrices partielles $P_0(z), P_1(z), P_2(z)$ et de la constante A :

$$p_0 = P_0(1) = \lim_{t \rightarrow \infty} p(C(t) = 0) = \frac{K}{A};$$

$$p_1 = P_1(1) = \lim_{t \rightarrow \infty} p(C(t) = 1) = \frac{(1 - \alpha_0 K) (1 - \tilde{B}(\alpha_1))}{A \alpha_1};$$

$$p_2 = P_2(1) = \lim_{t \rightarrow \infty} p(C(t) = 2) = \frac{1}{A} \left[(1 - \tilde{B}(\alpha_1)) (1 - \alpha_0 K) E[d_1] + \alpha_0 K E[d_0] \right].$$

2.4.1.3. Mesures de performance

A l'aide de la fonction génératrice de la distribution stationnaire de la chaîne de Markov induite (2.6) et celle de la taille de l'orbite (2.7), on obtient les mesures de performance suivantes :

- Nombre moyen de clients dans le système (aux instants où le serveur devient libre et opérationnel) \bar{n}

$$\bar{n} = \Phi(1) = K \frac{(1-\rho) \left(\tilde{\alpha}^{(2)}(1) - \lambda \tilde{\beta}^{(2)}(1) \right) + \lambda \tilde{\beta}^{(2)}(1) (1 + \alpha_0 E[d_0])}{2(1-\rho)^2} - \frac{1}{\theta} \left((\lambda + \alpha_0) - \frac{1}{K} \right),$$

où

$$\tilde{\alpha}^{(2)}(1) - \lambda \tilde{\beta}^{(2)}(1) = \lambda^2 \alpha_0 (E[d_0])^2 + 2\lambda^2 (1 - \tilde{B}(\alpha_1)) E[d_1]$$

$$+ \frac{2\lambda^2 (1 - \tilde{B}(\alpha_1))}{\alpha_1} + \frac{2\lambda^3 (1 - \tilde{B}(\alpha_1)) E[d_1]}{\alpha_1}$$

$$- 2\lambda(1-c) (1 - \tilde{B}(\alpha_1)) - 2\lambda \tilde{B}(\alpha_1),$$

$$\tilde{\beta}^{(2)}(1) = \tilde{B}'(\alpha_1) 2\lambda^2 E[d_1] + 2\lambda c \tilde{B}'(\alpha_1) + \lambda^2 (E[d_0])^2 (1 - \tilde{B}(\alpha_1))$$

$$+ 2(1-c) (1 - \tilde{B}(\alpha_1)) (\lambda E[d_1] - 1) + 1 + \frac{2\lambda^2 (1 - \tilde{B}(\alpha_1)) E[d_1]}{\alpha_1}$$

$$+ \frac{2\lambda(1-c) (1 - \tilde{B}(\alpha_1))}{\alpha_1} + \frac{2\lambda^2}{\alpha_1} \left(\tilde{R}_1(\alpha_1) + \frac{1 - \tilde{B}(\alpha_1)}{\alpha_1} \right) + 2\tilde{B}(\alpha_1).$$

– Nombre moyen de clients en orbite \bar{n}_o

$$\begin{aligned} \bar{n}_o = Q'(1) &= \frac{1}{A} \left[\frac{\alpha_0 + \lambda + \theta N_n}{\theta} + \frac{\lambda \tilde{B}'(\alpha_1) \alpha_1 + \lambda (1 - \tilde{B}(\alpha_1))}{\alpha_1^2} (1 - \alpha_0 K) \right. \\ &+ \frac{(1 - \tilde{B}(\alpha_1))}{\alpha_1} \left(\frac{\lambda}{\theta} (\alpha_0 + \lambda + \theta N_n) + \theta M^{(2)}(1) \right) + \frac{\lambda \alpha_0 K}{2} (E[d_0])^2 \\ &+ \frac{\lambda}{\theta} (\alpha_0 + \lambda + \theta N_n) E[d_0] + \frac{\lambda}{\theta} (\alpha_0 + \lambda + \theta N_n) (1 - \tilde{B}(\alpha_1)) E[d_1] \\ &+ \lambda K N'(1) + \lambda c K (1 - \tilde{B}(\alpha_1)) E[d_1] + \theta M^{(2)}(1) (\alpha_0 + \lambda + \theta N_n) E[d_1] \\ &\left. + (\alpha_0 + \lambda + \theta N_n) N'(1) + c E[d_1] (1 - \tilde{B}(\alpha_1)) (\alpha_0 + \lambda + \theta N_n) \right], \end{aligned}$$

où

$$M^{(2)}(1) = \frac{1}{\theta} \left[\bar{n} - (\lambda + \alpha_0) \frac{(\alpha_0 + \lambda + \theta N_n)}{\theta} - (\alpha_0 + \lambda + \theta N_n) \right];$$

$$N'(1) = \alpha_1 \left[\frac{E[d_1] (\lambda \alpha_1 \tilde{B}(\alpha_1) + \lambda(\alpha_0 + \lambda + \theta N_n))}{\alpha_1^2} + \frac{\lambda(\alpha_0 + \lambda + \theta N_n)(E[d_0])^2}{2\alpha_1} \right].$$

- Temps moyen d'attente d'un client (à l'aide de la formule de Little) \bar{W}

$$\bar{W} = \frac{\bar{n}_O}{\lambda}.$$

- Nombre moyen de rappels d'une même source secondaire \bar{R}

$$\bar{R} = \bar{W}\theta.$$

2.4.2. Modèle M/G/1 avec rappels et deux types de pannes actives

2.4.2.1. Description du modèle

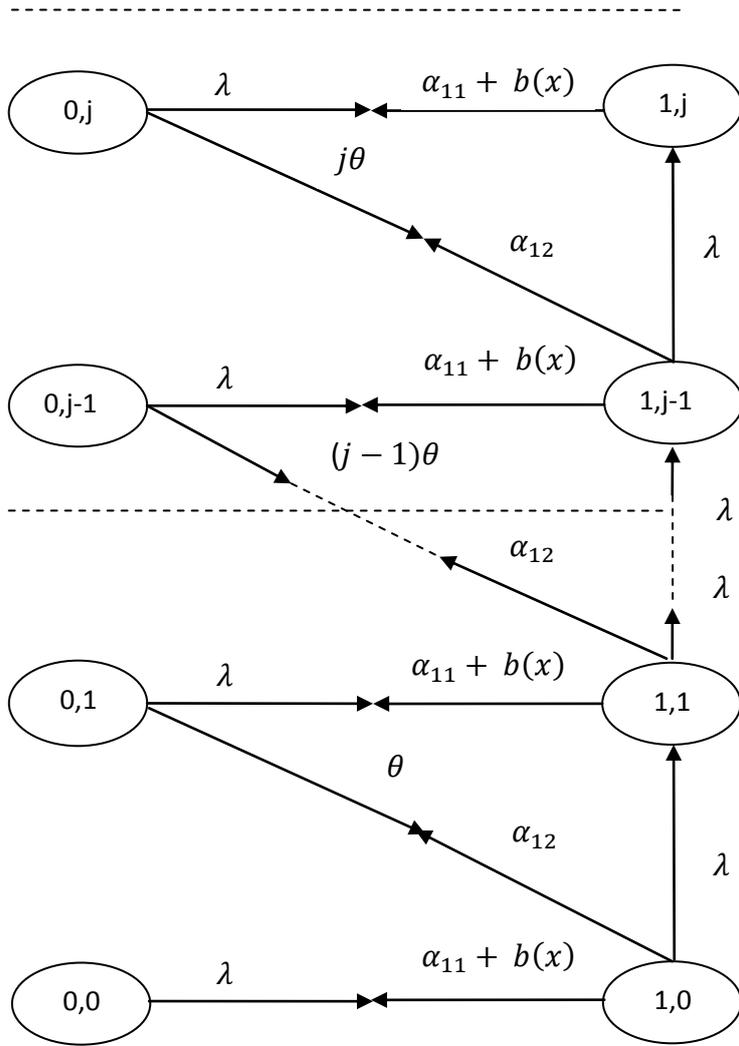
Ce modèle est le cas particulier du modèle de Kulkarni et Choi [32]: on a uniquement les pannes actives (le taux des pannes passives $\alpha_0 = 0$), la durée de réparation (panne active) $d_1 = 0$.

On a donc deux types de pannes actives [12]:

- 1) l'intervalle du temps entre deux pannes suit une loi exponentielle de paramètre $\alpha_{11} > 0$ (client quitte le système) ;
- 2) l'intervalle du temps entre deux pannes suit une loi exponentielle de paramètre $\alpha_{12} > 0$ (client entre en orbite).

L'état du système à la date t peut être décrit par le processus $\{C(t); N_o(t); \xi(t); t \geq 0\}$, où $C(t)$ est 0 ou 1 selon le fait que le serveur est libre ou occupé, $N_o(t)$ est le nombre de clients en orbite. Si $C(t) = 1$, $\xi(t)$ est une variable supplémentaire à valeurs dans \mathbb{R}^+ et désignant la durée de service écoulée à la date t .

Les transitions entre les états sont



De Artalejo et Gomez-Corral [5], on a que ce système est en régime stationnaire si

$$\rho = \frac{\lambda (1 - \tilde{B}(\alpha_{11} + \alpha_{12}))}{\alpha_{11} + \alpha_{12} \tilde{B}(\alpha_{11} + \alpha_{12})} < 1.$$

2.4.2.2. Analyse de la distribution stationnaire

Soit $\rho < 1$ et soient

$$p_{ij} = \lim_{t \rightarrow \infty} p(C(t) = i, N_o(t) = j), \text{ où } (i, j) \in S = \{0,1\} \times \mathbb{N} \quad (2.8)$$

Nous utilisons l'approche régénératrice. Les instants de régénération sont ceux où le serveur devient libre et aucun client n'est en orbite. On définit la durée de service effective comme l'intervalle aléatoire de temps depuis le début du service d'un client jusqu'à l'instant où le service en question s'achève avec succès ou avec interruption. Soient encore

T - la longueur d'un cycle de régénération ;

T_{ij} - la durée de séjour du processus $\{C(t), N_o(t), t \geq 0\}$ dans l'état (i, j) durant T ;

N_j^s - le nombre de services effectifs durant T aboutissant à un succès et laissant j clients en orbite ;

N_j^k - le nombre de services effectifs durant T aboutissant à une interruption et laissant j clients en orbite ($k = 1, 2$ -type de l'interruption).

Selon la théorie des processus régénérateurs Ross [36],

$$p_{ij} = \frac{E[T_{ij}]}{E[T]}, (i, j) \in S \quad (2.9)$$

A partir du graphe des transitions, en égalisant le taux du flux d'entrée et le taux du flux de sortie des états $(0, j)$ et $\{(i, m): i \in \{0, 1\}, j \geq m \geq 0\}$ respectivement, on obtient les équations d'équilibre statistique suivantes :

$$(\lambda + j\theta)E[T_{0j}] = E[N_j^s] + \alpha_{11}E[T_{1j}] + (1 - \delta_{0j})\alpha_{12}E[T_{1,j-1}], j \geq 0; \quad (2.10)$$

$$(j + 1)\theta E[T_{0,j+1}] = (\lambda + \alpha_{12})E[T_{1j}], j \geq 0. \quad (2.11)$$

Prenons l'équation (2.11) et divisons par $E[T]$.

D'après la définition (2.9), on a

$$(j + 1)\theta p_{0,j+1} = (\lambda + \alpha_{12})p_{1j}, j \geq 0. \quad (2.12)$$

Par conséquent, le problème est réduit à celui du calcul de la suite $\{p_{1j}, j \geq 0\}$. Dans le but de trouver la relation entre $E[T_{1j}]$, $E[N_j^s]$ et $E[N_j^k]$, nous introduisons la quantité

auxiliaire A_{kj} : le temps moyen, tel que durant un service effectif j clients sont en orbite étant donné que le service effectif précédent a laissé k clients en orbite.

D'après le théorème de Wald,

$$E[T_{1j}] = \sum_{k=0}^{j+1} (E[N_k^S] + E[N_k^1] + E[N_k^2]) A_{kj}, \quad j \geq 0 \quad (2.13)$$

En outre, vu que les interruptions sont exponentiellement distribuées, on trouve que :

$$E[N_k^1] = \alpha_{11} E[T_{1k}] \quad k \geq 0; \quad (2.14)$$

$$E[N_k^2] = \alpha_{12} E[T_{1,k-1}] \quad k \geq 0. \quad (2.15)$$

En combinant (2.10) – (2.11) et (2.13) – (2.15), on obtient

$$E[T_{1j}] = A_{0j} + (\lambda + \alpha_{12}) \sum_{k=1}^{j+1} \left(1 + \frac{\lambda}{k\theta}\right) A_{kj} E[T_{1,k-1}] \quad j \geq 0. \quad (2.16)$$

Il est évident que $E[T_{00}] = \frac{1}{\lambda}$, alors $E[T] = \frac{1}{\lambda p_{00}}$.

En divisant (2.16) par $E[T]$, nous obtenons la relation de récurrence

$$p_{1j} = \lambda p_{00} A_{0j} + (\lambda + \alpha_{12}) \sum_{k=1}^{j+1} \left(1 + \frac{\lambda}{k\theta}\right) A_{kj} p_{1,k-1} \quad j \geq 0. \quad (2.17)$$

Les probabilités $\{p_{1j}, j \geq 0\}$ peuvent donc être calculées en terme de p_{00} . A cet effet, il est nécessaire d'évaluer les quantités A_{kj} . Enfin, on obtient p_{00} à l'aide de l'équation de normalisation

$$\sum_{(i,j) \in S} p_{ij} = 1.$$

Pour calculer A_{kj} , définissons B_{kj} : le temps moyen, tel que durant un service effectif j clients sont en orbite étant donné que ce service a débuté avec k clients.

A l'aide de la démarche définie par Artalejo et Gomez Corral[5], on a que

$$A_{j+1,j} = \frac{(j+1)\theta}{\lambda + (j+1)\theta} B_{jj}, j \geq 0; \quad (2.18)$$

$$A_{kj} = \frac{k\theta}{\lambda + k\theta} B_{k-1,j} + \frac{\lambda}{\lambda + k\theta} B_{kj}, j \geq k \geq 0; \quad (2.19)$$

$$B_{kj} = \int_0^{\infty} e^{-\lambda t} \frac{(\lambda t)^{j-k}}{(j-k)!} (1 - B(t)) e^{-(\alpha_{11} + \alpha_{12})t} dt, j \geq k \geq 0. \quad (2.20)$$

La contribution de l'intervalle $(t, t + \Delta t)$ pour B_{kj} est :

- i) le temps de service n'est pas écoulé avant le temps t avec une probabilité $(1 - B(t))$;
- ii) une interruption de service n'a pas eu lieu avant t avec une probabilité $\{e^{-(\alpha_{11} + \alpha_{12})t}\}$;
- iii) $(j - k)$ clients primaires arrivent dans le système durant $(0, t)$.

Par conséquent, nous avons un schéma stable permettant de calculer la distribution stationnaire p_{ij} :

Algorithme

1. Calculer B_{kj} d'après (2.20) ;
2. Calculer A_{kj} d'après (2.18) – (2.19) ;
3. Exprimer p_{1j} en termes de p_{00} à l'aide de (2.17) ;
4. Trouver p_{00} à l'aide de l'équation de normalisation.

A présent, supposons que $B(t) = 1 - e^{-\mu t}$, $\mu > 0, t \geq 0$. Donc, on a un processus de Markov $\{C(t), N_o(t), t \geq 0\}$

Théorème[12]

La distribution stationnaire du processus $\{C(t), N_o(t), t \geq 0\}$ est donnée par

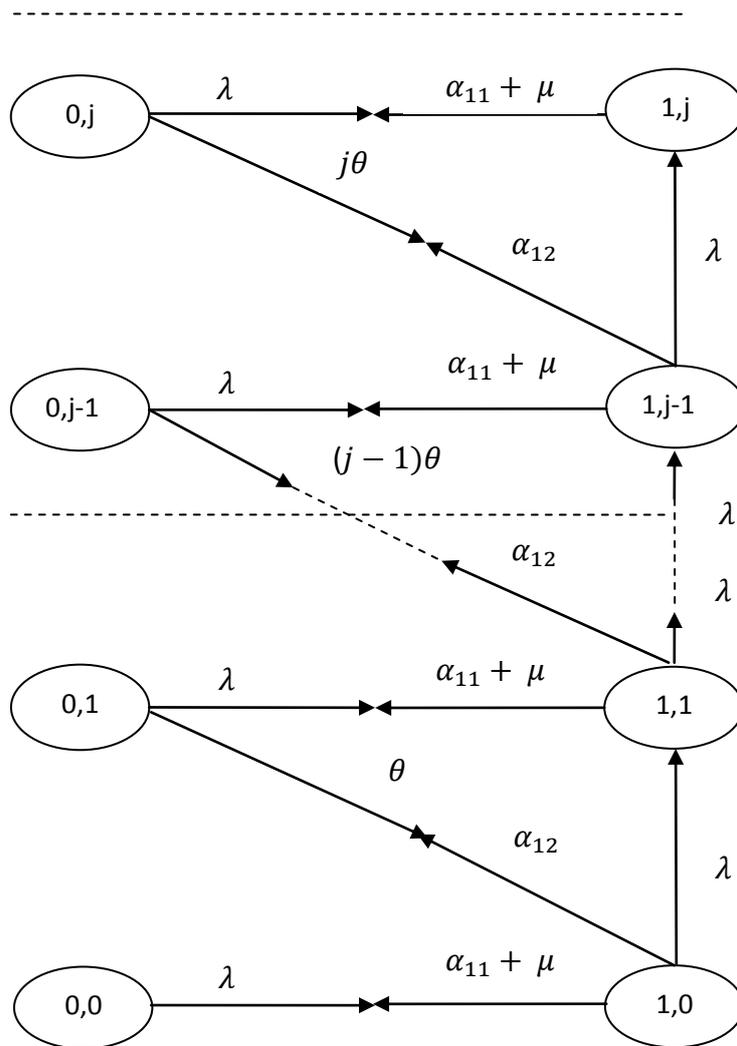
$$p_{0j} = \frac{(\lambda + \alpha_{12})}{\theta} \frac{\left(1 + \frac{(\lambda + \alpha_{12})}{\theta}\right)_{j-1}}{(1)_j} \rho^j p_{00}, j \geq 1; \quad (2.21)$$

- **Système d'équations d'équilibre statistique**

Lorsque $t \rightarrow \infty$, le système d'équations de Kolmogorov ci-dessus devient

$$\begin{cases} (\lambda + j\theta)p_{0j} = (\alpha_{11} + \mu)p_{1j} + \alpha_{12}(1 - \delta_{0j})p_{1,j-1} & j \geq 0; \quad (2.25) \\ (\lambda + \alpha_{11} + \alpha_{12} + \mu)p_{1j} = \lambda p_{0j} + \lambda(1 - \delta_{0j})p_{1,j-1} + (j+1)\theta p_{0,j+1} & j \geq 0. \quad (2.26) \end{cases}$$

- **Graphe des transitions**



De l'équation (2.12), on a que

$$p_{1j} = \frac{(j+1)\theta}{(\lambda + \alpha_{12})} p_{0,j+1}; \quad (2.27)$$

$$p_{0,j+1} = \frac{\lambda + \alpha_{12}}{(j+1)\theta} p_{1j}. \quad (2.28)$$

En substituant (2.27) dans (2.26), on obtient

$$\begin{aligned} & (\lambda + \alpha_{11} + \alpha_{12} + \mu) \frac{(j+1)\theta}{(\lambda + \alpha_{12})} p_{0,j+1} \\ & = \lambda p_{0j} + (j+1)\theta p_{0,j+1} + \lambda \frac{j\theta}{(\lambda + \alpha_{12})} p_{0,j}. \end{aligned} \quad (2.29)$$

Posons $j = j - 1$ dans (2.29). Ce qui donne comme résultat

$$(\lambda + \alpha_{11} + \alpha_{12} + \mu) \frac{j\theta}{(\lambda + \alpha_{12})} p_{0,j} = \lambda p_{0,j-1} + j\theta p_{0,j} + \lambda \frac{(j-1)\theta}{(\lambda + \alpha_{12})} p_{0,j-1}. \quad (2.30)$$

L'équation (2.30) peut être réécrite de la manière suivante

$$p_{0,j} = \frac{\lambda}{j\theta(\alpha_{11} + \mu)} [\lambda + \alpha_{12} + (j-1)\theta] p_{0,j-1}. \quad (2.31)$$

Vu que la distribution du temps de service est exponentielle, l'intensité du trafic ρ devient :

$$\rho = \frac{\lambda}{(\alpha_{11} + \mu)}.$$

L'équation (2.31) devient

$$p_{0,j} = \frac{\rho}{j\theta} [\lambda + \alpha_{12} + (j-1)\theta] p_{0,j-1},$$

ou bien

$$p_{0,j} = \frac{\rho^j}{j! \theta^j} \prod_{k=0}^{j-1} [\lambda + \alpha_{12} + k\theta] p_{00}. \quad (2.32)$$

Effectuons les calculs suivants

$$\begin{aligned}
p_{0,j} &= \frac{\rho^j \theta^j}{j! \theta^j} \prod_{k=0}^{j-1} \left[\frac{\lambda + \alpha_{12}}{\theta} + k \right] p_{00} = \\
&= \frac{\rho^j}{j!} \frac{\lambda + \alpha_{12}}{\theta} \left(\frac{\lambda + \alpha_{12}}{\theta} + 1 \right) \left(\frac{\lambda + \alpha_{12}}{\theta} + 2 \right) \dots \left(\left(\frac{\lambda + \alpha_{12}}{\theta} + 1 \right) + (j-1) - 1 \right) p_{00}.
\end{aligned}$$

A l'aide de la définition du symbole de Pochhammer et sachant que $j! = (1)_j$, on obtient l'équation (2.21) :

$$p_{0j} = \rho^j \frac{\lambda + \alpha_{12}}{\theta} \frac{\left(\frac{\lambda + \alpha_{12}}{\theta} + 1 \right)_{(j-1)}}{(1)_j} p_{00}.$$

A présent, considérons l'équation (2.25). A l'aide de (2.28) cette dernière devient

$$(\lambda + j\theta) \frac{\lambda + \alpha_{12}}{j\theta} p_{1,j-1} = \alpha_{12}(1 - \delta_{0j})p_{1,j-1} + (\alpha_{11} + \mu)p_{1j} \quad j \geq 0.$$

D'où

$$\begin{aligned}
p_{1j} &= \frac{\lambda^2 + \lambda\alpha_{12} + \lambda j\theta + \alpha_{12}j\theta - \alpha_{12}(1 - \delta_{0j})j\theta}{j\theta(\alpha_{11} + \mu)} p_{1,j-1} = \\
&= \frac{\lambda^j}{j! \theta^j (\alpha_{11} + \mu)^j} \prod_{k=1}^j [\lambda + \alpha_{12} + k\theta] p_{10} \\
&= \frac{\rho^j}{j! \theta^j} \prod_{k=1}^j [\lambda + \alpha_{12} + k\theta] p_{10}.
\end{aligned}$$

Encore, en appliquant la définition du symbole de Pochhammer, on trouve l'équation (2.22)

$$p_{1j} = \rho^j \frac{\left(1 + \frac{\lambda + \alpha_{12}}{\theta} \right)_j}{(1)_j} p_{10}.$$

Pour trouver la relation entre p_{00} et p_{10} , nous faisons les calculs suivants :

De l'équation (2.12), on a

$$\theta p_{01} = (\lambda + \alpha_{12})p_{10}.$$

De plus, l'équation (2.32) fournit

$$p_{01} = \frac{\rho}{\theta}(\lambda + \alpha_{12})p_{00}.$$

Par conséquent,

$$\theta \frac{\rho}{\theta}(\lambda + \alpha_{12})p_{00} = (\lambda + \alpha_{12})p_{10}$$

ou bien

$$\rho p_{00} = p_{10} \text{ et } p_{00} = \rho^{-1}p_{10}.$$

La probabilité p_{00} (ou celle p_{10}) peut être obtenue à l'aide de l'équation de normalisation

$$\sum_{j=0}^{\infty} p_{0j} + \sum_{j=0}^{\infty} p_{1j} = 1.$$

En effet,

$$\sum_{j=0}^{\infty} \rho^j \frac{\lambda + \alpha_{12}}{\theta} \frac{\left(\frac{\lambda + \alpha_{12}}{\theta} + 1\right)_{j-1}}{(1)_j} p_{00} + \sum_{j=0}^{\infty} \rho^j \frac{\left(1 + \frac{\lambda + \alpha_{12}}{\theta}\right)_j}{(1)_j} \rho p_{00} = 1.$$

D'où

$$p_{00} = \left(\sum_{j=0}^{\infty} \frac{\left(1 + \frac{\lambda + \alpha_{12}}{\theta}\right)_j}{(1)_j} \rho^j \right)^{-1}.$$

Considérons

$$\sum_{j=0}^{\infty} \frac{\left(1 + \frac{\lambda + \alpha_{12}}{\theta}\right)_j}{(1)_j} \rho^j = \sum_{j=0}^{\infty} \frac{(1)_j \left(\frac{\lambda + \alpha_{12}}{\theta}\right)_j \rho^j}{(1)_j j!}.$$

A l'aide de la définition de la fonction hypergéométrique, on obtient

$$p_{00} = \left[F \left(1, 1 + \frac{\lambda + \alpha_{12}}{\theta}; 1; \rho \right) \right]^{-1}.$$

La distribution stationnaire p_{ij} , $i = 0, 1$ et $j \geq 0$, permet de calculer les mesures de performance :

- *Nombre moyen de clients en orbite \bar{n}_o*

$$\bar{n}_o = \lim_{t \rightarrow \infty} E[N_o(t)] = \sum_{i=0}^1 \sum_{j=0}^{\infty} j p_{ij}.$$

- *Nombre moyen de serveurs occupés \bar{c}*

$$\bar{c} = \lim_{t \rightarrow \infty} E[C(t)] = \sum_{i=0}^1 \sum_{j=0}^{\infty} i p_{ij}.$$

2.5. Conclusion

Les modèles étudiés dans ce chapitre ont un nombre important d'applications potentielles. Dans le chapitre suivant, nous présenterons quelques exemples de problèmes qui peuvent être résolus à l'aide des systèmes de files d'attente avec rappels. Ces derniers comprennent soit le phénomène de priorité, soit les vacances du serveur, soit les pannes du serveur.

3

Systemes de files d'attente avec rappels et interruptions de service dans la modélisation des phénomènes réels

3.1. Introduction

Dans ce chapitre, nous donnons plusieurs exemples des systèmes réels qui peuvent être modélisés comme des systèmes d'attente avec rappels et serveur sujet à des interruptions de service.

3.2. Systèmes réels modélisables comme systèmes d'attente avec rappels et priorité relative

3.2.1. Modèle 1

Prenons le modèle avec rappels et priorité relative précédemment présenté dans le chapitre 2 et faisons les modifications suivantes : A l'arrivée d'un client primaire type 2, si le serveur est libre, le client en question est pris en charge. Dans le cas contraire, le client quitte le système définitivement avec une probabilité $1 - H_1$, ou entre en orbite avec une probabilité $H_1 < 1$ et devient source d'appels secondaires. A l'arrivée d'un client type 2 de l'orbite, si le serveur est encore occupé, le client quitte le système sans recevoir le service avec une probabilité $1 - H_2$, ou revient en orbite avec une probabilité H_2 . Supposons que $H_2 = H_3 = \dots$. Encore, les cas $H_2 = 1$ et $H_2 < 1$ suggèrent des solutions différentes.

A l'aide de la méthode des variables supplémentaires, Choi et Chang[13] ont obtenu les mesures de performance pour le modèle $M_1, M_2 / G / 1$ avec rappels, deux types de clients et $H_2 = 1$. Ces mesures sont :

- Nombre moyen de clients dans la file d'attente \bar{n}_f

$$\bar{n}_f = E[N_f] = \frac{\lambda_1 \lambda \beta_2}{2(1 - \rho_1)(1 + \rho_2(1 - H_1))}.$$

- Nombre moyen de clients en orbite \bar{n}_o

$$\bar{n}_o = E[N_o] = \frac{\lambda_1 \lambda H_1}{(1 - \rho_1 - H_1 \rho_2)} \left(\frac{\beta_1}{\theta} + \frac{\beta_2}{2(1 - \rho_1)(1 + \rho_2(1 - H_1))} \right).$$

- La distribution du nombre de serveurs occupés P_i avec ($i = 0, 1$) obtenue est

$$P_0 = \frac{1 - \rho_1 - H_1 \rho_2}{(1 + \rho_2(1 - H_1))}; \quad P_1 = \frac{\rho}{(1 + \rho_2(1 - H_1))}; \quad \text{où } \rho = \rho_1 + \rho_2.$$

Ces auteurs ont également trouvé la formule qui permet de calculer le nombre moyen de clients type 1 \bar{n}_f dans le système $M_1, M_2 / G_1, G_2 / 1$ avec rappels, deux types de clients et $H_2 < 1$:

$$\bar{n}_f = E[N_f] = \frac{\lambda_1^2 \beta_{1,2} \beta_{2,1} + (\rho - \lambda_1 \beta_{1,1}) \lambda_1 \beta_{2,2}}{2\beta_{2,1}(1 - \lambda_1 \beta_{1,1})}; \text{ où } \rho = \lambda_1 \beta_{1,1} + \lambda_2 \beta_{2,1}.$$

3.2.2. Modèle 2

Prenons encore une fois, le modèle avec rappels et priorité relative précédemment présenté dans le chapitre 2 et faisons les modifications suivantes : le taux d'arrivée des clients prioritaire est $q\lambda > 0$ et celui des clients non prioritaire est $p\lambda > 0$ où q représente la probabilité avec laquelle le client prend une place dans la file d'attente et $p = 1 - q$ représente la probabilité avec laquelle le client entre en orbite. La durée entre deux rappels successifs d'un même client en orbite suit une loi arbitraire de fonction de distribution de probabilité commune $A(x)$ et transformée de Laplace-Stieltjes $\alpha(s)$.

Les mesures de performance de ce modèle ont été trouvées par Atencia et Moreno[6] à l'aide de la méthode des variables supplémentaires. En voici quelques unes :

- Nombre moyen de clients dans la file d'attente \bar{n}_f

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

- Nombre moyen de clients en orbite \bar{n}_o

$$\bar{n}_o = E[N_o] = \frac{\lambda p}{2(1 - \rho q)} \left(\frac{2\beta_1(1 - \rho q)[1 - \alpha(\lambda)] + \lambda\beta_2[p + q\alpha(\lambda)]}{\alpha(\lambda) - [p + q\alpha(\lambda)]\rho} \right).$$

- Temps moyen d'attente \bar{W}_i d'un client type $i = 1, 2$ (d'après la formule de Little)

$$\bar{W}_1 = \frac{E[N_f]}{\lambda q} = \frac{\lambda \beta_2}{2(1 - \rho q)};$$

$$\bar{W}_2 = \frac{E[N_o]}{\lambda p} = \left(\frac{2\beta_1(1 - \rho q)[1 - \alpha(\lambda)] + \lambda\beta_2[p + q\alpha(\lambda)]}{2(1 - \rho q)(\alpha(\lambda) - [p + q\alpha(\lambda)]\rho)} \right).$$

3.2.3. Exemples d'application

Considérons quelques exemples d'application.

3.2.3.1. Systèmes téléphoniques modernes

Dans les échanges téléphoniques modernes, les lignes des abonnés sont connectées aux modules qui traitent les appels arrivants et ceux sortants. Dans le cas du blocage (les canaux sont occupés), les appels sortants sont placés en attente dans un tampon (de capacité infinie) tandis que les appels arrivants sont refusés et plus tard réinitialisés pour établir la connexion. Lorsque l'un des canaux se libère, un appel sortant (s'il se trouve dans le tampon) l'occupe immédiatement. De ce fait, les appels arrivants ne réussissent pas à établir la connexion aussi longtemps qu'il y a des appels dans le tampon. Ce comportement implique que les appels sortants possèdent une priorité relative sur les appels arrivants.

3.2.3.2. Réseaux LAN avec le protocole CSMA non persistant et CSMA persistant

Supposons qu'un LAN possède des utilisateurs non persistants et des utilisateurs persistants connectés par un bus. Les utilisateurs persistants sont contrôlés par l'unité centrale de façon que chaque fois lorsque le canal est libre, un utilisateur persistant l'occupe pour envoyer ces paquets. Les utilisateurs non persistants essaient la retransmission indépendamment après une durée de temps. Ce système se présente comme un système avec rappels et priorité, où les clients type 1 sont ceux persistants, les clients type 2 sont ceux non persistants et le serveur est le bus.

3.2.3.3. Réseau cellulaire mobile

On considère un réseau cellulaire mobile, où chaque cellule est servie par une station de base différente. Une cellule particulière peut s'occuper de plusieurs communications actives simultanément. La station de base dans une cellule manipule deux types d'appels. Le premier est un appel d'origine (initié dans la cellule en question). Habituellement, un abonné avec un appel d'origine bloqué refait sa tentative après une durée de temps. Un autre type d'appels, "handoff", apparaît lorsqu'un abonné détenteur de la ligne de communication entre en cellule en question des cellules adjacentes. Si la station de base ne parvient pas à attribuer un canal libre jusqu'à ce que l'abonné sorte de la région de recouvrement des cellules, il souffre de la panne durant la conversation. La

dégradation de la qualité de service téléphonique dans ce cas est plus sérieuse que dans le cas d'un appel d'origine. Par conséquent, la station de base peut donner la priorité aux appels "handoff" en attribuant une file. Dans un réseau cellulaire mobile, la perte d'un appel "handoff" est un facteur important pour la qualité de service. Ce système peut être modélisé comme un système avec rappels, deux types de clients et perte géométrique.

3.2.3.4. Téléphone dans une banque

Prenant l'exemple d'un banquier qui doit s'occuper d'une ligne d'attente et d'un téléphone. Un tel employé prête attention au téléphone seulement quand il n'y a aucun client dans la banque. Un client arrivant et trouvant ce banquier inoccupé sera servi immédiatement ; autrement, il attend avec une probabilité q dans une file d'attente ou avec la probabilité complémentaire $p=1-q$ décide de téléphoner plus tard jusqu'à ce qu'il obtienne son service. Ce système peut être modelé comme un système $M_1, M_2 / G / 1$ avec rappels et priorité relative où les clients attendant dans la banque sont considérés comme des clients prioritaires et les clients faisant des appels téléphoniques comme des clients non prioritaires.

3.3. Systèmes réels modélisables comme systèmes d'attente avec rappels et priorité absolue

Dans cette partie, nous présentons deux exemples de systèmes réels: le premier est modélisé par le système étudié dans le chapitre 4 et 5 et l'autre est modélisé par le système $M / G / 1$ avec rappels, service à deux phases et politique résumé préemptive [33].

3.3.1. Bureaux d'information

Le modèle étudié dans le chapitre 4 et 5 peut être utilisé pour modéliser une situation (fréquemment observée dans quelques bureaux d'information), où un agent seul répond aux appels téléphoniques et sert le client présent. Dans ce contexte, les appels téléphoniques ont une priorité absolue par rapport au client présent.

3.3.2. Réseaux de communication

Dans des réseaux de communication, des messages de longueur variables arrivent à une station de service qui les traite dans deux étapes par un seul serveur. A l'étape de prétraitement (le service préliminaire), le serveur peut interrompre le message actuel du processus en cours pour satisfaire le nouveau message prioritaire en arrivée. Après l'achèvement de prétraitement, si le message est qualifié, le traitement principal commencera et si autrement, il doit quitter le système. Le serveur est bloqué pendant le traitement principal pour qu'un message arrivant soit acheminé à l'orbite pour faire des rappels. Ce système a été modélisé par un système M/G/1, service à deux phases et priorité absolue dans [33].

3.4. Systèmes réels modélisables comme systèmes de files d'attente avec rappels et vacances du serveur

Nous considérons un système d'attente avec rappels de type M/G/1, où le serveur applique la politique de vacances multiples quand aucun client n'est enregistré dans l'orbite. On donne la description détaillée du modèle comme suit :

1. De nouveaux clients arrivent de l'extérieur du système selon un processus de Poisson de taux $\lambda > 0$. Nous supposons qu'il n'y a aucun espace d'attente, et donc si un client arrivant trouve le serveur libre il commencera son service. Si un client arrivant trouve le serveur occupé ou en vacance, il rejoint l'orbite avec une probabilité b ou bien quitte le système avec une probabilité $(1 - b)$. Les clients de l'orbite essaient d'exiger leurs services plus tard et les temps inter-rappels suivent une distribution arbitraire $A(x)$ de transformée de Laplace Stieltjes $\alpha(s)$.
2. Le service est assuré par un seul serveur. Dès qu'un client est servi complètement, il décidera de rejoindre l'orbite pour avoir un autre service avec une probabilité p ou quitter le système avec la probabilité $(1 - p)$. Les temps de service suivent une loi générale de fonction de distribution $B(x)$ de transformée de Laplace Stieltjes $\tilde{B}(s)$.
3. À chaque fois que l'orbite est vide, le serveur prend des vacances d'une longueur aléatoire V . Si aucun client n'apparaît dans l'orbite lorsque le serveur retourne des

vacances, il repart pour de nouvelles vacances de la même longueur. Ce schéma se répète jusqu'à ce qu'il trouve au moins un client enregistré dans l'orbite à son retour ou qu'il a déjà pris J vacances. Si l'orbite est vide à la fin de la $J^{\text{ème}}$ vacance, le serveur reste en attente des clients dans le système. Si à une époque d'achèvement de vacances l'orbite est non vide, le serveur attend les clients éventuels de l'orbite, ou l'arrivée de nouveaux. Le temps des vacances V a une fonction de distribution $V(x)$ de transformée de Laplace Stieltjes $\tilde{V}(s)$.

4. Les processus stochastiques impliqués dans ce système sont indépendants les uns des autres.

L'analyse détaillée de ce système a été faite par Jau-Chuan Ke et Fu-Min Chang [29] à l'aide de la méthode des variables supplémentaires. Les auteurs ont réussi de trouver les expressions explicites des mesures de performance. Ces mesures sont :

- La probabilité p_1 que le serveur est inoccupé pendant le temps de rappel

$$p_1 = \frac{\{1 - \alpha(\lambda)\} \{N^{(1)}(1) + p + \lambda b E[S]\}}{N^{(1)}(1) \{1 - p + (1 - b)\lambda E[S]\} + p[1 - \alpha(\lambda)] + \{b[1 - \alpha(\lambda)] + \alpha(\lambda)\} \lambda E[S]}$$

- La probabilité p_2 que le serveur est occupé

$$p_2 = \frac{[N^{(1)}(1) + \alpha(\lambda)] \lambda E[S]}{N^{(1)}(1) \{1 - p + (1 - b)\lambda E[S]\} + p[1 - \alpha(\lambda)] + \{b[1 - \alpha(\lambda)] + \alpha(\lambda)\} \lambda E[S]}$$

- La probabilité p_3 que le serveur est en vacance

$$p_3 = \frac{N^{(1)}(1) \alpha(\lambda) - p - \lambda b E[S]}{N^{(1)}(1) \{1 - p + (1 - b)\lambda E[S]\} + p[1 - \alpha(\lambda)] + \{b[1 - \alpha(\lambda)] + \alpha(\lambda)\} \lambda E[S]}$$

- Nombre moyen de clients dans le système $E(N)$

$$E(N) = P_0 \left\{ 2 \alpha(\lambda) (1 - \alpha(\lambda)) (N^{(1)}(1) + p + \lambda b E[S]) (1 - p + (1 - b)\lambda E[S]) + N^{(2)}(1) (\alpha(\lambda) - p - \lambda b E[S]) (1 - p + (1 - b)\lambda E[S]) + \lambda^2 b (N^{(1)}(1) + \alpha(\lambda)) \{E[S^2] (\alpha(\lambda) - p) + 2pE[S]^2\} + \lambda b (1 - \alpha(\lambda)) (N^{(1)}(1) + \alpha(\lambda)) (2pE[S] + \lambda b E[S]^2) \right\} / \{2(\alpha(\lambda) - p - \lambda b E[S])^2 + \{N^{(1)}(1) + \alpha(\lambda)\} \lambda E[S]\} / \{N^{(1)}(1) \{1 - p + (1 - b)\lambda E[S]\} + p[1 - \alpha(\lambda)] + \{b[1 - \alpha(\lambda)] + \alpha(\lambda)\} \lambda E[S]\}.$$

avec

$$N^{(1)}(1) = \frac{\{1 - [\tilde{V}(\lambda b)]^j\} \lambda b E[V]}{[\tilde{V}(\lambda b)]^j (1 - \tilde{V}(\lambda b))};$$

$$N^{(2)}(1) = \frac{\{1 - [\tilde{V}(\lambda b)]^j\} \lambda b^2 E[V^2]}{[\tilde{V}(\lambda b)]^j (1 - \tilde{V}(\lambda b))};$$

$$P_0 = \frac{\alpha(\lambda) - p - \lambda b E[S]}{N^{(1)}(1) \{1 - p + (1 - b) \lambda E[S]\} + p [1 - \alpha(\lambda)] + \{b [1 - \alpha(\lambda)] + \alpha(\lambda)\} \lambda E[S]}.$$

3.4.1. Exemples d'application

Considérons quelques exemples d'application.

3.4.1.1. *Système de transfert de courrier électronique*

Dans le modèle de système de transfert de courrier électronique, le système de courrier utilise le protocole de transfert de courrier simple (le SMTP) pour livrer les messages entre les serveurs de courrier. Quand un programme de transfert de courrier contacte un serveur sur une machine à distance, il forme une connexion sur laquelle il communique. Une fois que la connexion est en place, les deux programmes suivent le SMTP qui permet à l'expéditeur de s'identifier, spécifier un destinataire et transférer un message électronique. Lorsque l'expéditeur dépose le courrier électronique dans son propre serveur de courrier, le serveur de courrier peut à plusieurs reprises essayer d'envoyer un message de contact pour cibler le serveur jusqu'à ce que le serveur cible devienne opérationnel. Typiquement, les messages de contact arrivent au serveur selon un processus de Poisson. Quand les messages arrivent au serveur de courrier, un message est choisi pour être servi et les autres rejoindront le tampon. Dans le tampon, chaque message reste une certaine durée de temps avant de faire une nouvelle tentative de service. Il y a un programme de démon mis en œuvre au serveur de courrier pour gérer les demandes de service provenant du tampon. Le serveur cible est comme celui de l'expéditeur et le message d'envoi sera probablement retourné au serveur pour demander le service de réception. Pour garder le serveur de courrier bien fonctionnant, le scan antivirus est une activité de maintenance importante. Il peut être exécuté quand le

serveur de courrier est inoccupé. Ce type de maintenance peut être programmé pour fonctionner régulièrement. Cependant, ces activités de maintenance ne se répètent pas continuellement. Quand ces activités sont finies, le serveur de courrier entrera dans l'état inoccupé de nouveau et attendra l'arrivée des messages de contact. Dans ce scénario, le tampon dans le serveur de courrier expéditeur, le serveur de courrier récepteur, la politique de transmission et les activités de maintenance correspondent à l'orbite, le serveur, la discipline des rappels et la politique de vacances, respectivement, dans la terminologie de la théorie des files d'attente [29].

3.4.1.2. Le serveur www (world wide web)

Dans le modèle opérationnel du serveur www, des requêtes http (HyperText Transfert Protocol) arrivent au serveur www selon un processus de Poisson, tandis que ces requêtes http peuvent être terminées par un utilisateur auparavant arrivant au serveur www. Lorsque les requêtes arrivent au serveur www, une requête est choisie pour être servie et le reste des requêtes entrera dans le tampon qui est placé à l'intérieure du serveur www. Dans le tampon, chaque requête observe une certaine durée de temps avant de faire une nouvelle tentative de service. Il y a un programme de démon mis en œuvre au serveur www pour gérer les requêtes provenant du tampon. Si la page web cible est placée au même serveur www, la requête sera probablement retournée au serveur. Pour garder le serveur www bien fonctionnant, des activités de maintenance tel que le scan antivirus peut être exécuté lorsque le serveur www est inoccupé. Ce type de maintenance peut être programmé pour fonctionner régulièrement. Ce pendant, ces activités de maintenance ne se répètent pas continuellement. Lorsque ces activités de maintenance sont finies, le serveur www entrera dans l'état inoccupé de nouveau et attendra l'arrivée des nouvelles requêtes. Dans ce scénario, le tampon dans le serveur www, le serveur www, la politique de retransmission et les activités de maintenance correspondent à l'orbite, le serveur, la discipline des rappels et la politique des vacances, respectivement, dans la terminologie de la théorie des files d'attente[29].

Les deux exemples réels suivants, sont des exemples pratiques du model $M / G / 1$ avec rappels et vacances présenté dans le chapitre 2.

3.4.1.3. Réseau local avec CSMA/CD non persistant

Dans un tel système, un nombre fini m de clients (terminaux actifs) sont reliés par un seul bus. Sous le protocole CSMA/CD, si un terminal a un message à transmettre, il consulte le bus pour voir s'il est occupé ou au repos. Si le canal est occupé, le terminal consulte le bus de nouveau après une durée exponentiellement distribuée. Par contre, si le canal est libre, il commence à transmettre le message. Vu que le délai de propagation n'est pas égal à zéro, pendant un certain temps, les autres terminaux avec messages peuvent également trouver le bus libre, et par conséquent transmettre leurs messages. Ce phénomène est référé comme une collision. Chaque terminal impliqué dans la collision abandonne sa transmission et consulte le bus après une durée de temps (comme dans le cas du bus occupé). Habituellement, une collision dure un certain temps pendant lequel aucun terminal n'est autorisé à transmettre. Ce système se conduit comme un système de files d'attente avec rappels, nombre de source fini et vacances du serveur, où les sources sont les terminaux, les vacances sont les durées de collision et le serveur est le bus.

3.4.1.4. Réseaux locaux en anneau

Ces systèmes sont caractérisés par le fait qu'une permission de transmettre circule dans l'anneau. Une station, ayant des messages à transmettre, attend l'arrivée de la permission libre. Si c'est le cas, elle change l'indicateur de la permission de libre à occupée, et le message part pour la station destination. Après la réception du message, la permission est marquée libre et passe vers une autre station. Dans le cas contraire (la permission n'est pas libre ou la destination est occupée par la réception d'un autre message), le message est stocké dans le tampon pour une transmission ultérieure. Ces systèmes peuvent être modélisés comme les systèmes de files d'attente avec rappels, vacances du serveur et nombre de source fini. L'orbite est constituée des tampons des stations attachées au réseau, le serveur est la permission. Les vacances du serveur sont les passages de la permission libérée par une station vers une autre.

3.5. Systèmes réels modélisables comme un système de files d'attente avec rappels et serveur sujet à des pannes

3.5.1. Transmission des messages à travers des réseaux fac-similes

Le modèle $M / G / 1$ avec rappels et deux types de pannes présenté dans le chapitre 2 peut être utilisé pour étudier le problème provenant des possibilités "auto-repeat" pour la transmission des messages à travers des réseaux fac-similes. Considérons une entreprise de communication. Cette dernière réserve une machine fac-simile, spécifique pour envoyer des messages aux destinataires se trouvant à l'extérieur de l'entreprise en question. A l'arrivée d'un utilisateur désirant envoyer un message, si le médium de transmission est libre, la transmission commence. Dans le cas contraire, le message est soit stocké dans un tampon et la demande sera réinitialisée après une certaine durée aléatoire ou quitte le médium pour toujours. Si la transmission d'un message ne se termine pas avec succès pour des raisons, telles que : un blocage dans les lignes extérieurs ou une erreur de transmission, le message en question quitte le médium et, également, rejoint le tampon. Il est utile d'indiquer que l'équipement fac-simile reste opérationnel et les interruptions du service sont dues aux facteurs extérieurs.

La correspondance avec le modèle est la suivante : les clients en orbite, la distribution du temps inter-rappels et la distribution du temps de service sont, respectivement, les messages bloqués/interrompus et stockés dans le tampon, le protocole de retransmission et le temps de transmission.

3.5.2. Réseaux informatiques avec infection virale

Quand il n'y a aucun virus, les réseaux informatiques peuvent être modélisés et analysés en utilisant des réseaux de files d'attente conventionnels avec ou sans rappels. Les clients représentent l'exécution des opérations algorithmiques et logiques, et les nœuds représentent les dispositifs CPUS, I/O,...etc. Quand un virus entre dans un nœud, un ou plusieurs fichiers peuvent être infectés et l'administrateur système devrait passer un certain nombre de sauvegardes à fin de récupérer les fichiers infectés. Dans certains cas, ils ne peuvent pas être récupérables. Un virus peut émaner de l'extérieure du réseau, par exemple, disquette, ou peut venir d'un autre nœud dans le réseau, par exemple, un

courrier électronique. Le temps de récupération des fichiers infectés peut être considéré comme le temps de réparation du serveur (le nœud) par suite d'une arrivée de client négatif (le virus). Ce cas réel a été modélisé par un système Géo/Géo/1 avec rappels, clients négatifs et serveur non fiable dans [43] où les auteurs ont étudié l'impact des clients négatifs et le facteur de manque de fiabilité sur les performances du système. La description mathématique de ce modèle peut être donnée comme suit :

Nous considérons un système d'attente avec rappels à un seul serveur où l'axe des temps est divisé en des intervalles égaux dites slots et toutes les activités du système (arrivées, départs, rappels, pannes et réparations) se produisent aux bornes du slot. L'axe est supposé marqué par $0, 1, \dots, m, \dots$. Considérons l'époque m . Supposons que les départs et des achèvements des réparations se produisent dans $]m^-, m[$ où m^- est l'époque avant un départ tandis que les arrivées positifs, les arrivées négatifs, les rappels et les débuts des réparations se réalisent dans $]m, m^+[$ où m^+ est l'époque après une arrivée. Deux types de clients, positifs et négatifs, arrivent selon des processus d'arrivée géométriques avec des probabilités p et q , respectivement. Il n'y a aucun espace d'attente devant le serveur et donc, si un client positif arrive et trouve le serveur inoccupé, il commence son service immédiatement et quitte le système après l'achèvement de son service. Autrement, si le serveur est occupé ou en réparation à l'époque d'arrivée, le client positif arrivant quitte l'espace de service et rejoint l'orbite afin de tenter sa chance à nouveau plus tard. Chaque client dans l'orbite génère un flux de Bernoulli de taux $\bar{r} = 1 - r$, où r est la probabilité qu'un client en orbite ne fait pas un nouveau rappel dans un slot, et ceci indépendamment des autres clients en orbite. Le serveur tombera en panne à l'arrivée des clients négatifs, et sera envoyé à la réparation immédiatement. Les temps de réparations suivent une distribution géométrique de paramètre $\bar{\alpha} = 1 - \alpha$, *i.e.*, α est la probabilité que la réparation ne se termine pas dans un slot. Les clients négatifs feront le client étant dans le service perdu, mais n'ont aucun effet sur l'orbite. Les temps de service sont indépendants et géométriquement distribués avec une probabilité $\bar{s} = 1 - s$, où s est la probabilité qu'un client ne conclut pas son service dans un slot. Finalement, les processus stochastiques impliqués dans ce système sont supposés être indépendants les uns des autres. Pour éviter les cas insignifiants, il est supposé que $0 < p < 1, 0 < q < 1,$

$0 \leq r < 1, 0 < \alpha \leq 1$ et $0 < s < 1$. Les mesures de performances de ce système sont données par les expressions suivantes [43] :

- La probabilité p_1 que le serveur est libre

$$p_1 = 1 - \frac{q}{\bar{\alpha} \bar{q} + q} - \frac{p \bar{\alpha} \bar{q} + \alpha pq}{(\bar{\alpha} \bar{q} + q)(\bar{s} + s\bar{p}q)}.$$

- La probabilité p_2 que le serveur est occupé

$$p_2 = \frac{p \bar{\alpha} \bar{q} + \alpha pq}{(\bar{\alpha} \bar{q} + q)(\bar{s} + s\bar{p}q)}.$$

- La probabilité p_3 que le serveur est en panne

$$p_3 = \frac{q}{\bar{\alpha} \bar{q} + q}.$$

- Le nombre moyen de clients en orbite $E(N_o)$

$$\begin{aligned} E(N_o) &= \sum_{n=1}^{\infty} \frac{\hat{G}(r^n)}{G(r^n)} r^n \\ &+ \left\{ 1 - \frac{q}{\bar{\alpha} \bar{q} + q} - \frac{p \bar{\alpha} \bar{q} + \alpha pq}{(\bar{\alpha} \bar{q} + q)(\bar{s} + s\bar{p}q)} \right\} \left\{ \frac{\alpha p \bar{q} p + q s \bar{q} \bar{p}}{\bar{\alpha} \bar{q} (\bar{p} - s p q - s \bar{q}) - \alpha p q} \right. \\ &+ \left. \frac{[(1 - s \bar{q} - p \bar{q}) \alpha p + s p \bar{q} (\bar{\alpha} \bar{q} + \bar{\alpha} \bar{q} p)] [(\alpha q + \bar{\alpha} \bar{q}) p + q (1 - s \bar{q} - s p q)]}{[\alpha \bar{q} (\bar{p} - s p q - s \bar{q}) - \alpha p q]^2} \right\}, \end{aligned}$$

avec

$$G(z) = \bar{p} \bar{q} \frac{sp \bar{q} \alpha p z^2 + [\alpha p (s \bar{p} \bar{q} - 1) - sp \bar{\alpha} \bar{q}] z + [\bar{s} (\alpha p + \bar{\alpha} \bar{q}) + \bar{\alpha} (\bar{s} q + s \bar{p} q)]}{sp \bar{q} \alpha p z^2 - [sp \bar{q} (\bar{\alpha} \bar{q} + \alpha pq) + (1 - s \bar{p} \bar{q} - s \bar{p} \bar{q}) \alpha p] + W},$$

où $W = \bar{s} \bar{p} \bar{q} (\alpha p + \bar{\alpha} \bar{q}) + \bar{\alpha} \bar{p} \bar{q} (\bar{s} q + s \bar{p} q)$.

- Le nombre moyen de clients dans le système $E(N)$

$$E(N) = E(N_o) + \frac{p \bar{\alpha} \bar{q} + \alpha pq}{(\bar{\alpha} \bar{q} + q)(\bar{s} + s\bar{p}q)}.$$

3.5.3. Interaction hôte-parasite (application biologique)

Considérons le cas d'un seul hôte (un organisme qui héberge un parasite) qui est né, sans parasite, au temps $t=0$ et dans sa durée de vie est exposé aux parasites de temps en temps dans une zone infectée. A un instant d'exposition, l'hôte acquiert un seul parasite indépendamment d'une exposition à une autre. Le nombre de parasites dans l'hôte peut augmenter en raison de la reproduction de parasites et peut provoquer la mortalité de l'hôte (quand il atteint un certain nombre m) à l'âge t_0 . L'intérêt du biologiste est au nombre de parasites acquis par l'hôte jusqu'à l'instant de temps t quand l'hôte a été déplacé vers une zone non infectée à un certain âge $t_0 < t$. L'hôte entre dans la zone non infectée seulement s'il est vivant et infecté à l'âge t_0 . Dans la zone non infectée, l'hôte vit dans des conditions non infectieuses. Cela signifie que l'hôte n'acquiert pas de nouveaux parasites, subit un traitement concret pour diminuer le nombre de parasites acquis jusqu'à l'instant de déplacement.

L'interaction précédemment décrite peut être modélisé par un système de files d'attente avec rappels (les tentatives faites par un parasite à fin de s'héberger dans l'hôte qui lui résiste), une orbite (la zone infectée) de capacité infinie, un seul serveur (l'hôte) sujet à des pannes (infection par m ou plus parasites) réparables (le traitement de diminution de parasites représente la réparation de l'hôte) [27].

3.6. Conclusion

Dans ce chapitre nous avons présenté quelques phénomènes réels qui peuvent être modélisés par les systèmes de files d'attente avec rappels et service sujet à des interruptions.

4

Modèle d'attente avec rappels et priorité absolue

4.1. Introduction

Il devrait être indiqué que presque tous les modèles avec rappels et priorité y compris les modèles avec priorité qu'on a présenté dans le chapitre 2 opèrent sous la règle qui met toujours les clients non prioritaires en orbite. Dans ce chapitre, nous considérons un système de files d'attente de type $M_1, M_2/G_1, G_2/1$ avec rappels et priorité absolue de politique „preemptive resume” qui opère sous la règle contraire. Notre modèle peut être utilisé pour modéliser une situation (fréquemment observée dans quelques bureaux d'information), où un agent seul répond aux appels téléphoniques et sert le client présent. Dans ce contexte, les appels téléphoniques ont une priorité absolue par rapport au client présent. Pour le modèle en question, nous discutons le problème d'ergodicité. Puis en utilisant la méthode des variables supplémentaires, nous obtenons les fonctions génératrices partielles de la distribution stationnaire de l'état du système et quelques

mesures de performance pertinentes. Ces résultats font l'objet de l'étude réalisée dans [11].

4.2. Description mathématique du modèle

Nous considérons un système de files d'attente à un seul serveur, auquel deux types différents de clients primaires arrivent selon des processus de Poisson indépendants de taux λ_1 et λ_2 , respectivement. Les clients du premier flux de taux $\lambda_1 > 0$ ont une priorité absolue sur les clients du deuxième flux de taux $\lambda_2 > 0$. Ainsi, les règles suivantes gouvernent la dynamique des clients :

- Tout client primaire arrivant et trouvant le serveur inoccupé, occupe immédiatement le serveur et quitte le système après l'achèvement du service.
- Tout client prioritaire arrivant et trouvant le serveur occupé par le service d'un autre client prioritaire, rejoint l'orbite. La durée entre deux rappels successifs d'un même client en orbite suit une loi exponentielle de taux $\theta > 0$.
- Tout client prioritaire arrivant (primaire ou d'orbite) et trouvant le serveur occupé par le service d'un client non prioritaire, entre directement dans le serveur. Le client non prioritaire, dont le service a été interrompu, persiste dans la zone du serveur jusqu'à l'achèvement du service du client prioritaire pour reprendre son service d'où il a été interrompu (la politique „preemptive resume”).
- Tout client primaire non prioritaire arrivant et trouvant le serveur occupé quitte le système sans service.

Les durées de service des deux types de clients suivent une distribution générale de fonction de répartition $B_l(x), l \in \{1, 2\}$, de transformée de Laplace-Stieltjes

$\tilde{B}_l(s) = \int_0^{\infty} \exp(-sx) dB_l(x), l \in \{1, 2\}$ et $\text{Re}(s) > 0$. Soient les moments d'ordre k ,

$\beta_{l,k} = (-1)^k \tilde{B}_l^{(k)}(0)$ du durée de service à partir de l'origine; $b_l(x) = \frac{B_l'(x)}{1 - B_l(x)}, l \in \{1, 2\}$,

est l'intensité instantanée du service du client de type l , étant donné que la durée de

service écoulee est égale à x . $K_l(z_1, z_2) = \tilde{B}_l(\lambda_1 - \lambda_1 z_1 + \lambda_2 - \lambda_2 z_2)$ est la fonction génératrice du nombre de clients primaires de deux types arrivant dans le système durant le temps de service d'un client de type l ($l \in \{1, 2\}$). Finalement, nous admettons l'hypothèse de l'indépendance mutuelle entre toutes les variables aléatoires définies précédemment.

L'état du système à la date t peut être décrit à l'aide d'un processus de Markov à temps continu, tel que

$$\{C(t), N_o(t), \xi_1(t), \xi_2(t), t \geq 0\}, \quad (4.1)$$

Où $N_o(t)$ est le nombre de clients prioritaires en orbite et $C(t)$ représente l'état de station-service à la date t . $C(t)$ est égal à 0, 1, 2 ou 3 selon le fait que le serveur est inoccupé, un client prioritaire est servi et il n'y a d'aucun client non prioritaire dont le service a été interrompu dans la station-service, un client non prioritaire est servi ou un client prioritaire est servi et il y a un client non prioritaire dont le service a été interrompu dans la station-service. Si $C(t) \in \{1, 3\}$ (si $C(t) = 2$) $\xi_1(t)$ ($\xi_2(t)$) représente le temps de service écoulee du client prioritaire (du client non prioritaire) dans le serveur à la date t .

Les transitions entre les états sont définies comme suit :

- $C(t) = 0, i \geq 0$

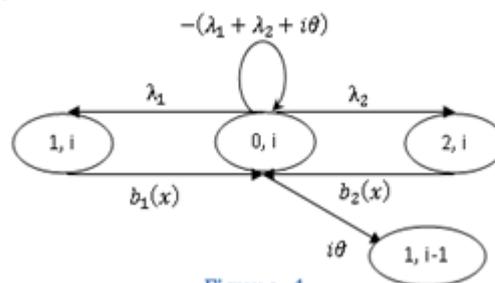


Figure: 1

- $C(t) = 1, i \geq 0$

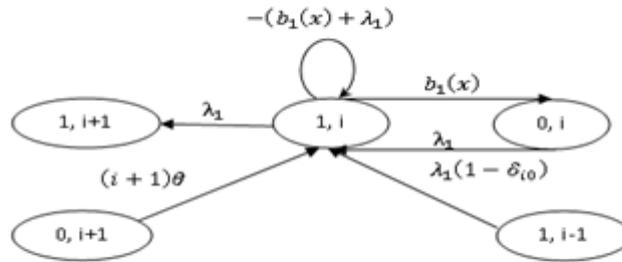


Figure: 2

- $C(t) = 2, i \geq 0$

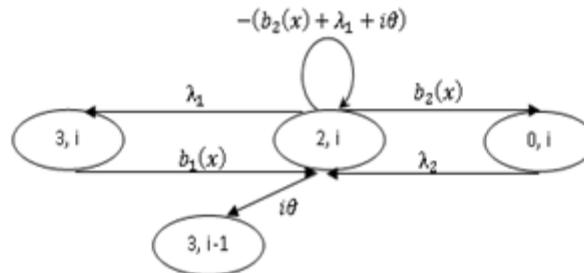


Figure: 3

- $C(t) = 3, i \geq 0$

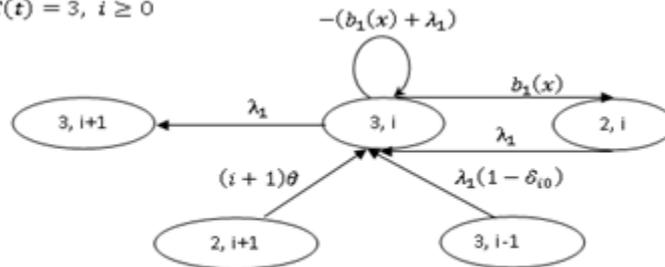


Figure: 4

4.3. Condition d'ergodicité

Dénotons par t_d la date du d 'ème départ, $N_{1,d}$ ($N_{2,d}$) le nombre de clients prioritaires (de clients non prioritaires) dans l'orbite (dans la station-service) juste avant la date t_d . Nous avons les équations fondamentales suivantes :

$$N_{1,d} = N_{1,d-1} - V_d + Y_{1,d}; \quad (4.2)$$

$$N_{2,d} = N_{2,d-1} \times B_d, \quad Y_{2,d} = 0. \quad (4.3)$$

où V_d est égale à 0 ou 1, selon si le d 'ème client servi est un client primaire ou secondaire; $Y_{1,d}$ ($Y_{2,d}$) est le nombre de clients prioritaires (non prioritaires) arrivant au système pendant le temps de service du d 'ème client; B_d est égale à 0 ou 1 selon si le d 'ème client servi est non prioritaire ou prioritaire. Dénotons par U_d le type du d 'ème client servi (qui peut être 1 ou 2 selon si le client en question est un client prioritaire ou non prioritaire).

Le vecteur aléatoire (U_d, V_d) dépend de l'histoire du système avant le temps t_d seulement par le vecteur $(N_{1,d-1}, N_{2,d-1})$. Sa distribution conditionnelle est définie par

$$P\left(\left(U_d = 1, V_d = 1\right) \setminus \left(\left(N_{1,d-1}, N_{2,d-1}\right) = (m, n)\right)\right) = \frac{m\theta}{\lambda_1 + \lambda_2 + m\theta};$$

$$P\left(\left(U_d = 1, V_d = 0\right) \setminus \left(\left(N_{1,d-1}, N_{2,d-1}\right) = (m, n)\right)\right) = \frac{\lambda_1}{\lambda_1 + \lambda_2 + m\theta};$$

$$P\left(\left(U_d = 2, V_d = 1\right) \setminus \left(\left(N_{1,d-1}, N_{2,d-1}\right) = (m, n)\right)\right) = 0;$$

$$P\left(\left(U_d = 2, V_d = 0\right) \setminus \left(\left(N_{1,d-1}, N_{2,d-1}\right) = (m, n)\right)\right) = \frac{\lambda_2}{\lambda_1 + \lambda_2 + m\theta}.$$

Le vecteur aléatoire $(Y_{1,d}, Y_{2,d})$ dépend des événements qui sont arrivés pendant le temps de service du d 'ème client seulement par U_d et a la distribution conditionnelle suivante :

$$P\left(\left(\left(Y_{1,d}, Y_{2,d}\right) = (m, n)\right) \setminus \left(U_d = l\right)\right) = \int_0^{\infty} \exp(-\lambda_1 x) \frac{(\lambda_1 x)^m}{m!} \exp(-\lambda_2 x) \frac{(\lambda_2 x)^n}{n!} dB_l(x) = k_{l,m,n} \quad \text{avec}$$

$$l \in \{1, 2\}, \quad m \geq 0 \quad \text{et} \quad n \in \{0, 1\}.$$

La suite $\{X_d = (U_d, N_{1,d}, N_{2,d}), d \geq 1\}$ forme une chaîne de Markov à espace d'états $S = \{1, 2\} \times Z_+ \times \{0, 1\}$, qui est une chaîne de Markov induite de notre système d'attente. Ses probabilités de transition à un pas $r_{(l,m,n),(c,i,j)} = P(X_d = (c, i, j) \mid X_{d-1} = (l, m, n))$, où $c \in \{1, 2\}$, sont données par :

$$r_{(l,m,n),(2,i,j)} = \frac{\lambda_2}{\lambda_1 + \lambda_2 + m\theta} k_{2,i-m,0}, \quad n \in \{0, 1\} \text{ et } m = i; \quad (4.4)$$

$$r_{(l,m,n),(1,i,j)} = \frac{\lambda_1}{\lambda_1 + \lambda_2 + m\theta} k_{1,i-m,0} + \frac{m\theta}{\lambda_1 + \lambda_2 + m\theta} k_{1,i-m+1,0}. \quad (4.5)$$

Dans le but de prouver la déclaration précédente, nous avons besoin d'examiner les deux cas de la valeur de n comme suit :

Si $n = 1$, $r_{(l,m,n),(2,i,j)} = \frac{\lambda_2}{\lambda_1 + \lambda_2 + m\theta} k_{2,i-m,j-n+1}$; tandis que si $n = 0$,
 $r_{(l,m,n),(2,i,j)} = \frac{\lambda_2}{\lambda_1 + \lambda_2 + m\theta} k_{2,i-m,j-n}$, pour $c = 0$ et $j = 0$. Ainsi
 $r_{(l,m,n),(2,i,j)} = \frac{\lambda_2}{\lambda_1 + \lambda_2 + m\theta} k_{2,i-m,0}$ pour $i = m$ et $r_{(l,m,n),(2,i,j)} = 0$ pour $m - i \neq 0$. La
probabilité de transition $r_{(l,m,n),(1,i,j)}$ devient $\frac{\lambda_1}{\lambda_1 + \lambda_2 + m\theta} k_{1,i-m,j-n} + \frac{m\theta}{\lambda_1 + \lambda_2 + m\theta} k_{1,i-m+1,j-n}$,
pour $c = 1$ et $j = n$. Par conséquent, nous obtenons (4.5).

Le problème à étudier par la suite est celui de l'ergodicité de notre chaîne. Comme les équations fondamentales (4.2)-(4.3) ont une structure récursive, nous utilisons le critère de Foster [16]. Considérons la fonction de Lyapunov suivante sur l'espace d'état S défini ci-dessus : $f(l, m, n) = m + (1 - \rho_1)n$, où $\rho_1 = \lambda_1 \beta_{1,1}$. L'accroissement moyen de la chaîne $x_{(l,m,n)} = E[(f(X_d) - f(X_{d-1})) \mid X_{d-1} = (l, m, n)]$ peut être obtenu de la façon suivante :

i) Si $n = 0$, les équations (4.2)-(4.3) deviennent

$$N_{1,d} = N_{1,d-1} - V_d + Y_{1,d} = m - V_d + Y_{1,d}; \quad N_{2,d} = 0.$$

Dans ces circonstances, $f(X_{d-1}) = m$ et $f(X_d) = m - V_d + Y_{1,d}$, ce qui implique que

$$\begin{aligned} x_{(l,m,n)}^{(1)} &= E\left[(-V_d + Y_{1,d}) \setminus \left((N_{1,d-1}, N_{2,d-1}) = (m, 0)\right)\right] \\ &= E\left[Y_{1,d} \setminus \left((N_{1,d-1}, N_{2,d-1}) = (m, 0)\right)\right] - E\left[V_d \setminus \left((N_{1,d-1}, N_{2,d-1}) = (m, 0)\right)\right] \\ &= \frac{\lambda_1 + m\theta}{\lambda_1 + \lambda_2 + m\theta} \lambda_1 \beta_{1,1} - \frac{m\theta}{\lambda_1 + \lambda_2 + m\theta}. \end{aligned}$$

ii) Si $n = 1$, de (4.2)-(4.3) nous avons

$$N_{1,d} = N_{1,d-1} - V_d + Y_{1,d} = m - V_d + Y_{1,d}, \quad N_{2,d} = B_d$$

$$\text{et } f(X_d) - f(X_{d-1}) = -V_d + Y_{1,d} + (1 - \rho_1)(B_d - 1).$$

Ainsi

$$\begin{aligned} x_{(l,m,n)}^{(2)} &= E\left[(-V_d + Y_{1,d} + (1 - \rho_1)(B_d - 1)) \setminus \left((N_{1,d-1}, N_{2,d-1}) = (m, 1)\right)\right] \\ &= E\left[(-V_d + Y_{1,d}) \setminus \left((N_{1,d-1}, N_{2,d-1}) = (m, 1)\right)\right] \\ &\quad + (1 - \rho_1) E\left[U_d \setminus \left((N_{1,d-1}, N_{2,d-1}) = (m, 1)\right)\right] - (1 - \rho_1) \\ &= -\frac{m\theta}{\lambda_1 + \lambda_2 + m\theta} + \frac{\lambda_1 + m\theta}{\lambda_1 + \lambda_2 + m\theta} \lambda_1 \beta_{1,1} + (1 - \rho_1) \left(\frac{\lambda_1 + m\theta}{\lambda_1 + \lambda_2 + m\theta} - 1\right). \end{aligned}$$

A présent, considérons $x^{(1)} = \lim_{m \rightarrow \infty} x_{(l,m,n)}^{(1)} = \rho_1 - 1$ et $x^{(2)} = \lim_{m \rightarrow \infty} x_{(l,m,n)}^{(2)} = \rho_1 - 1$. Alors

$x^{(1)} = x^{(2)} < 0$ si $\rho_1 < 1$. Donc, la condition suffisante est $\rho_1 < 1$. Puisque $r_{(l,m,n)(1,i,j)} = r_{(l,m,n)(2,i,j)} = 0$ pour $i < m - 1$, $\rho_1 < 1$ est aussi une condition nécessaire pour l'ergodicité (selon la condition de Kaplan[38], $\rho_1 \geq 1$ donne non ergodicité de notre chaîne de Markov induite). Finalement, $\{X_d, d \geq 1\}$ est ergodique si et seulement si $\rho_1 < 1$.

4.4. Distribution stationnaire de l'état du système

A présent, nous étudions la distribution stationnaire du processus (4.1) en utilisant la méthode des variables supplémentaires (voir par exemple [23]). Pour ce faire, nous supposons que $\rho_1 < 1$ et nous introduisons les probabilités suivantes :

$$P_{0,i} = \lim_{t \rightarrow \infty} P(C(t) = 0, N_o(t) = i), \quad i \geq 0;$$

$$P_{1,i}(x) = \lim_{t \rightarrow \infty} \frac{d}{dx} P(C(t) = 1, \xi_1(t) \leq x, N_o(t) = i), \quad i \geq 0 \text{ et } x \geq 0;$$

$$P_{2,i}(y) = \lim_{t \rightarrow \infty} \frac{d}{dy} P(C(t) = 2, \xi_2(t) \leq y, N_o(t) = i), \quad i \geq 0 \text{ et } y \geq 0;$$

$$P_{3,i}(x, y) = \lim_{t \rightarrow \infty} \frac{d}{dx} P(C(t) = 3, \xi_1(t) \leq x, \xi_2(t) \leq y, N_o(t) = i), \quad i \geq 0, \quad x \geq 0 \text{ et } y \geq 0.$$

Alors, en suivant la méthode des variables supplémentaires, nous constatons que les probabilités considérées satisfont les équations d'équilibre statistique suivantes :

$$(\lambda_1 + \lambda_2 + i\theta)P_{0,i} = \int_0^{\infty} P_{2,i}(y)b_2(y)dy + \int_0^{\infty} P_{1,i}(x)b_1(x)dx; \quad (4.6)$$

$$P'_{1,i}(x) = -(b_1(x) + \lambda_1)P_{1,i}(x) + \lambda_1(1 - \delta_{i0})P_{1,i-1}(x); \quad (4.7)$$

$$P_{1,i}(0) = \lambda_1 P_{0,i} + (i+1)\theta P_{0,i+1}; \quad (4.8)$$

$$P'_{2,i}(y) = -(b_2(y) + \lambda_1 + i\theta)P_{2,i}(y) + \int_0^{\infty} P_{3,i}(x, y)b_1(x)dx; \quad (4.9)$$

$$P_{2,i}(0) = \lambda_2 P_{0,i}; \quad (4.10)$$

$$P'_{3,i}(x, y) = -(b_1(x) + \lambda_1)P_{3,i}(x, y) + \lambda_1(1 - \delta_{i0})P_{3,i-1}(x, y); \quad (4.11)$$

$$P_{3,i}(0, y) = \lambda_1 P_{2,i}(y) + (i+1)\theta P_{2,i+1}(y). \quad (4.12)$$

La résolution du système (4.6)-(4.12) est très compliquée, voire impossible. A l'aide des fonctions génératrices $P_0(z_1) = \sum_{i=0}^{\infty} z_1^i P_{0,i}$, $P_1(z_1, x) = \sum_{i=0}^{\infty} z_1^i P_{1,i}(x)$, $P_2(z_1, y) = \sum_{i=0}^{\infty} z_1^i P_{2,i}(y)$ et

$P_3(z_1, x, y) = \sum_{i=0}^{\infty} z_1^i P_{3,i}(x, y)$, le système d'équations(4.6)-(4.12) devient

$$\begin{aligned}
& \bullet \sum_{i=0}^{\infty} z_1^i (\lambda_1 + \lambda_2 + i\theta) P_{0,i} = \sum_{i=0}^{\infty} z_1^i \int_0^{\infty} P_{2,i}(y) b_2(y) dy + \sum_{i=0}^{\infty} z_1^i \int_0^{\infty} P_{1,i}(x) b_1(x) dx; \\
& (\lambda_1 + \lambda_2) \sum_{i=0}^{\infty} z_1^i P_{0,i} + \sum_{i=0}^{\infty} z_1^i i \theta P_{0,i} = \int_0^{\infty} \sum_{i=0}^{\infty} z_1^i P_{2,i}(y) b_2(y) dy + \int_0^{\infty} \sum_{i=0}^{\infty} z_1^i P_{1,i}(x) b_1(x) dx; \\
& (\lambda_1 + \lambda_2) P_0(z_1) + z_1 \theta \frac{dP_0(z_1)}{dz_1} = \int_0^{\infty} P_2(z_1, y) b_2(y) dy + \int_0^{\infty} P_1(z_1, x) b_1(x) dx; \\
& z_1 \theta \frac{dP_0(z_1)}{dz_1} = -(\lambda_1 + \lambda_2) P_0(z_1) + \int_0^{\infty} P_2(z_1, y) b_2(y) dy + \int_0^{\infty} P_1(z_1, x) b_1(x) dx; \tag{4.13}
\end{aligned}$$

$$\begin{aligned}
& \bullet \sum_{i=0}^{\infty} z_1^i P'_{1,i}(x) = \sum_{i=0}^{\infty} (-(b_1(x) + \lambda_1)) P_{1,i}(x) z_1^i + \sum_{i=0}^{\infty} \lambda_1 (1 - \delta_{i0}) P_{1,i-1}(x) z_1^i; \\
& \sum_{i=0}^{\infty} z_1^i P'_{1,i}(x) = -(b_1(x) + \lambda_1) \sum_{i=0}^{\infty} P_{1,i}(x) z_1^i + \lambda_1 \sum_{i=0}^{\infty} (1 - \delta_{i0}) P_{1,i-1}(x) z_1^i; \\
& \frac{\partial P_1(z_1, x)}{\partial x} = -(b_1(x) + \lambda_1) P_1(z_1, x) + \lambda_1 z_1 P_1(z_1, x); \\
& \frac{\partial P_1(z_1, x)}{\partial x} = (-b_1(x) - \lambda_1 + \lambda_1 z_1) P_1(z_1, x); \tag{4.14}
\end{aligned}$$

$$\begin{aligned}
& \bullet \sum_{i=0}^{\infty} z_1^i P_{1,i}(0) = \sum_{i=0}^{\infty} z_1^i \lambda_1 P_{0,i} + \sum_{i=0}^{\infty} z_1^i (i+1) \theta P_{0,i+1}; \\
& \sum_{i=0}^{\infty} z_1^i P_{1,i}(0) = \lambda_1 \sum_{i=0}^{\infty} z_1^i P_{0,i} + \theta \sum_{i=0}^{\infty} z_1^i (i+1) P_{0,i+1}; \\
& P_1(z_1, 0) = \lambda_1 P_0(z_1) + \theta \frac{dP_0(z_1)}{dz_1}; \tag{4.15}
\end{aligned}$$

$$\begin{aligned}
& \bullet \sum_{i=0}^{\infty} z_1^i P'_{2,i}(y) = \sum_{i=0}^{\infty} z_1^i (-(b_2(y) + \lambda_1 + i\theta)) P_{2,i}(y) + \sum_{i=0}^{\infty} z_1^i \int_0^{\infty} P_{3,i}(x, y) b_1(x) dx; \\
& \sum_{i=0}^{\infty} z_1^i P'_{2,i}(y) = -(b_2(y) + \lambda_1) \sum_{i=0}^{\infty} z_1^i P_{2,i}(y) - z_1 \theta \sum_{i=0}^{\infty} i z_1^{i-1} P_{2,i}(y) + \int_0^{\infty} \sum_{i=0}^{\infty} z_1^i P_{3,i}(x, y) b_1(x) dx; \\
& \frac{\partial P_2(z_1, y)}{\partial y} = -(b_2(y) + \lambda_1) P_2(z_1, y) - \theta z_1 \frac{\partial P_2(z_1, y)}{\partial z_1} + \int_0^{\infty} P_3(z_1, x, y) b_1(x) dx; \tag{4.16}
\end{aligned}$$

$$\begin{aligned}
& \bullet \sum_{i=0}^{\infty} z_1^i P_{2,i}(0) = \sum_{i=0}^{\infty} z_1^i \lambda_2 P_{0,i}; \\
& P_2(z_1, 0) = \lambda_2 P_0(z_1); \tag{4.17}
\end{aligned}$$

$$\begin{aligned}
\bullet \sum_{i=0}^{\infty} z_1^i P'_{3,i}(x, y) &= \sum_{i=0}^{\infty} z_1^i (-(b_1(x) + \lambda_1)) P_{3,i}(x, y) + \sum_{i=0}^{\infty} z_1^i \lambda_1 (1 - \delta_{i0}) P_{3,i-1}(x, y); \\
\sum_{i=0}^{\infty} z_1^i P'_{3,i}(x, y) &= -(b_1(x) + \lambda_1) \sum_{i=0}^{\infty} z_1^i P_{3,i}(x, y) + \lambda_1 z_1 \sum_{i=0}^{\infty} z_1^{i-1} (1 - \delta_{i0}) P_{3,i-1}(x, y); \\
\frac{\partial P_3(z_1, x, y)}{\partial x} &= -(b_1(x) + \lambda_1) P_3(z_1, x, y) + \lambda_1 z_1 P_3(z_1, x, y); \\
\frac{\partial P_3(z_1, x, y)}{\partial x} &= (-b_1(x) - \lambda_1 + \lambda_1 z_1) P_3(z_1, x, y); \tag{4.18}
\end{aligned}$$

$$\begin{aligned}
\bullet \sum_{i=0}^{\infty} z_1^i P_{3,i}(0, y) &= \sum_{i=0}^{\infty} z_1^i \lambda_1 P_{2,i}(y) + \sum_{i=0}^{\infty} z_1^i (i+1) \theta P_{2,i+1}(y); \\
\sum_{i=0}^{\infty} z_1^i P_{3,i}(0, y) &= \theta \sum_{i=0}^{\infty} z_1^i (i+1) P_{2,i+1}(y) + \lambda_1 \sum_{i=0}^{\infty} z_1^i P_{2,i}(y); \\
P_3(z_1, 0, y) &= \theta \frac{\partial P_2(z_1, y)}{\partial z_1} + \lambda_1 P_2(z_1, y). \tag{4.19}
\end{aligned}$$

La résolution du système (4.13)-(4.19) est moins compliquée que celle du système (3.6)-(4.12) et peut être faite en procédant de la manière suivante : de l'équation (3.14) et (4.18) on trouve que $P_1(z_1, x)$ et $P_3(z_1, x, y)$ dépendent de x comme suit : l'intégration des deux parties de l'équation (4.14) donne

$$\begin{aligned}
\int_0^x \frac{dP_1(z_1, x)}{P_1(z_1, x)} &= \int_0^x (-b_1(x) - \lambda_1 + \lambda_1 z_1) dx; \\
\int_0^x \frac{dP_1(z_1, x)}{P_1(z_1, x)} &= \int_0^x \left(-\frac{B_1'(x)}{1 - B_1(x)} \right) dx + \int_0^x (-\lambda_1 + \lambda_1 z_1) dx; \\
\left[\ln(P_1(z_1, x)) - \ln(P_1(z_1, 0)) \right] &= \left[\ln(1 - B_1(x)) - \ln(1 - B_1(0)) \right] + (-\lambda_1 + \lambda_1 z_1)x; \\
\ln\left(\frac{P_1(z_1, x)}{P_1(z_1, 0)} \right) &= \ln(1 - B_1(x)) + (-\lambda_1 + \lambda_1 z_1)x, \left((B_1(0) = 0, \text{ par convention}) \Rightarrow (\ln(1 - B_1(0)) = 0) \right).
\end{aligned}$$

D'où

$$P_1(z_1, x) = P_1(z_1, 0) \cdot (1 - B_1(x)) \cdot \exp\{(\lambda_1 z_1 - \lambda_1)x\}. \tag{4.20}$$

De la même façon: l'intégration des deux parties de l'équation (4.18) donne

$$\begin{aligned} \int_0^x \frac{dP_3(z_1, x, y)}{P_3(z_1, x, y)} &= \int_0^x (-b_1(x) - \lambda_1 + \lambda_1 z_1) dx; \\ \int_0^x \frac{dP_3(z_1, x, y)}{P_3(z_1, x, y)} &= \int_0^x \left(-\frac{B_1'(x)}{1-B_1(x)} \right) dx + \int_0^x (-\lambda_1 + \lambda_1 z_1) dx; \\ [\ln(P_3(z_1, x, y)) - \ln(P_3(z_1, 0, y))] &= [\ln(1-B_1(x)) - \ln(1-B_1(0))] + (-\lambda_1 + \lambda_1 z_1)x; \\ \ln\left(\frac{P_3(z_1, x, y)}{P_3(z_1, 0, y)}\right) &= \ln(1-B_1(x)) + (-\lambda_1 + \lambda_1 z_1)x, \left((B_1(0) = 0, \text{ par convention}) \Rightarrow (\ln(1-B_1(0)) = 0) \right). \end{aligned}$$

D'où

$$P_3(z_1, x, y) = P_3(z_1, 0, y) \cdot (1-B_1(x)) \cdot \exp\{(\lambda_1 z_1 - \lambda_1)x\}. \quad (4.21)$$

En remplaçant $P_3(z_1, x, y)$ par son expression donnée par (4.21) dans l'équation (4.16), on obtient

$$\begin{aligned} \frac{\partial P_2(z_1, y)}{\partial y} &= -(b_2(y) + \lambda_1)P_2(z_1, y) - \theta z_1 \frac{\partial P_2(z_1, y)}{\partial z_1} \\ &\quad + \int_0^\infty P_3(z_1, 0, y)(1-B_1(x)) \exp\{(\lambda_1 z_1 - \lambda_1)x\} b_1(x) dx; \\ \frac{\partial P_2(z_1, y)}{\partial y} &= -(b_2(y) + \lambda_1)P_2(z_1, y) - \theta z_1 \frac{\partial P_2(z_1, y)}{\partial z_1} + P_3(z_1, 0, y) \int_0^\infty \exp\{(\lambda_1 z_1 - \lambda_1)x\} dB_1(x); \\ \frac{\partial P_2(z_1, y)}{\partial y} &= -(b_2(y) + \lambda_1)P_2(z_1, y) - \theta z_1 \frac{\partial P_2(z_1, y)}{\partial z_1} + P_3(z_1, 0, y) K_1(z_1, 1). \end{aligned} \quad (4.22)$$

En remplaçant $P_3(z_1, 0, y)$ par son expression donnée par (4.19), on obtient

$$\frac{\partial P_2(z_1, y)}{\partial y} = (-b_2(y) - \lambda_1 + \lambda_1 K_1(z_1, 1))P_2(z_1, y) + \theta(-z_1 + K_1(z_1, 1)) \frac{\partial P_2(z_1, y)}{\partial z_1}. \quad (4.23)$$

L'équation (4.23) admet comme solution

$$\begin{aligned}
P_2(z_1, y) &= \text{const} \cdot \exp \left\{ -\int_0^y b_2(y) \right\} \cdot \exp \left\{ \int_1^{z_1} \frac{\lambda_1 - \lambda_1 K_1(z_1, 1)}{\theta(K_1(z_1, 1) - z_1)} dz_1 \right\}; \\
&= \text{const} \cdot (1 - B_2(y)) \cdot \exp \left\{ \int_1^{z_1} \frac{\lambda_1 - \lambda_1 K_1(z_1, 1)}{\theta(K_1(z_1, 1) - z_1)} dz_1 \right\}.
\end{aligned} \tag{4.24}$$

En prenant $y = 0$ dans l'équation (4.24), on obtient une deuxième expression de $P_2(z_1, 0)$ (la première expression est donnée par l'équation (4.17))

$$P_2(z_1, 0) = \text{const} \cdot \exp \left\{ \int_1^{z_1} \frac{\lambda_1 - \lambda_1 K_1(z_1, 1)}{\theta(K_1(z_1, 1) - z_1)} dz_1 \right\}. \tag{4.25}$$

Pour avoir l'expression de $P_2(z_1)$, il suffit d'intégrer l'équation (4.25) par rapport à y

$$\begin{aligned}
P_2(z_1) &= \int_0^\infty P_2(z_1, y) dy \\
&= \text{const} \left(\int_0^\infty (1 - B_2(y)) dy \right) \exp \left\{ \int_1^{z_1} \frac{\lambda_1 - \lambda_1 K_1(z_1, 1)}{\theta(K_1(z_1, 1) - z_1)} dz_1 \right\}.
\end{aligned} \tag{4.26}$$

A l'aide des équations (4.25) et (4.17), on obtient

$$P_0(z_1) = \frac{\text{const}}{\lambda_2} \cdot \exp \left\{ \int_1^{z_1} \frac{\lambda_1 - \lambda_1 K_1(z_1, 1)}{\theta(K_1(z_1, 1) - z_1)} dz_1 \right\}. \tag{4.27}$$

A présent, en remplaçant $P_0(z_1)$ par son expression donnée par (4.27) dans l'équation (4.15), on obtient

$$P_1(z_1, 0) = \frac{\lambda_1}{\lambda_2} \cdot \text{const} \cdot \left(\frac{1 - z_1}{K_1(z_1, 1) - z_1} \right) \cdot \exp \left\{ \int_1^{z_1} \frac{\lambda_1 - \lambda_1 K_1(z_1, 1)}{\theta(K_1(z_1, 1) - z_1)} dz_1 \right\}; \tag{4.28}$$

en remplaçant $P_1(z_1, 0)$ par son expression donnée par (4.28) dans (4.20), on obtient

$$P_1(z_1, x) = \frac{\lambda_1}{\lambda_2} \text{const} \left(\frac{1-z_1}{K_1(z_1, 1) - z_1} \right) (1 - B_1(x)) \exp\{(\lambda_1 z_1 - \lambda_1)x\} \\ \times \exp \left\{ \int_1^{z_1} \frac{\lambda_1 - \lambda_1 K_1(z_1, 1)}{\theta(K_1(z_1, 1) - z_1)} dz_1 \right\}. \quad (4.29)$$

Pour avoir l'expression de $P_1(z_1)$, il suffit d'intégrer l'équation (4.29) par rapport à x :

$$P_1(z_1) = \int_0^{\infty} P_1(z_1, x) dx \\ = \frac{\lambda_1}{\lambda_2} \text{const} \left(\frac{1-z_1}{K_1(z_1, 1) - z_1} \right) \int_0^{\infty} (1 - B_1(x)) \exp\{(\lambda_1 z_1 - \lambda_1)x\} dx \\ \times \exp \left\{ \int_1^{z_1} \frac{\lambda_1 - \lambda_1 K_1(z_1, 1)}{\theta(K_1(z_1, 1) - z_1)} dz_1 \right\}.$$

Nous savons que

$$\int_0^{\infty} (1 - B_1(x)) \exp\{(\lambda_1 z_1 - \lambda_1)x\} dx = \frac{1 - K_1(z_1, 1)}{(\lambda_1 - \lambda_1 z_1)};$$

alors

$$P_1(z_1) = \frac{\text{const}}{\lambda_2} \left(\frac{1 - K_1(z_1, 1)}{K_1(z_1, 1) - z_1} \right) \exp \left\{ \int_1^{z_1} \frac{\lambda_1 - \lambda_1 K_1(z_1, 1)}{\theta(K_1(z_1, 1) - z_1)} dz_1 \right\}. \quad (4.30)$$

En remplaçant $P_2(z_1, y)$ par son expression donnée par l'équation (4.24) dans l'équation (4.19), on obtient

$$P_3(z_1, 0, y) = \lambda_1 \text{const} \left(\frac{1-z_1}{K_1(z_1, 1) - z_1} \right) (1 - B_2(y)) \exp \left\{ \int_1^{z_1} \frac{\lambda_1 - \lambda_1 K_1(z_1, 1)}{\theta(K_1(z_1, 1) - z_1)} dz_1 \right\}. \quad (4.31)$$

En remplaçant $P_3(z_1, 0, y)$ par son expression donnée par l'équation (4.31) dans l'équation (4.21), on obtient

$$\begin{aligned}
P_3(z_1, x, y) &= \lambda_1 \text{const} \left(\frac{1-z_1}{K_1(z_1, 1) - z_1} \right) (1-B_2(y))(1-B_1(x)) \\
&\quad \times \exp\{(\lambda_1 z_1 - \lambda_1)x\} \exp \left\{ \int_1^{z_1} \frac{\lambda_1 - \lambda_1 K_1(z_1, 1)}{\theta(K_1(z_1, 1) - z_1)} dz_1 \right\}.
\end{aligned} \tag{4.32}$$

Une intégration double de $P_3(z_1, x, y)$ par rapport à x et y donne

$$\begin{aligned}
P_3(z_1) &= \int_0^\infty \int_0^\infty P_3(z_1, x, y) dx dy \\
&= \text{const} \left(\frac{1-K_1(z_1, 1)}{K_1(z_1, 1) - z_1} \right) \left(\int_0^\infty (1-B_2(y)) dy \right) \exp \left\{ \int_1^{z_1} \frac{\lambda_1 - \lambda_1 K_1(z_1, 1)}{\theta(K_1(z_1, 1) - z_1)} dz_1 \right\}.
\end{aligned} \tag{4.33}$$

Pour rendre les expressions (4.26), (4.27), (4.30) et (4.33) définies continue au point $z_1 = 1$, il suffit de prolonger la fonction $\frac{1-K_1(z_1, 1)}{K_1(z_1, 1) - z_1}$ par continuité au point $z_1 = 1$.

Alors, on doit calculer la limite suivante

$$\begin{aligned}
\lim_{z_1 \rightarrow 1} \frac{1-K_1(z_1, 1)}{K_1(z_1, 1) - z_1} &= \lim_{z_1 \rightarrow 1} \frac{1-\tilde{B}_1(\lambda_1 - \lambda_1 z_1)}{\tilde{B}_1(\lambda_1 - \lambda_1 z_1) - z_1} \\
&= \lim_{z_1 \rightarrow 1} \frac{\lambda_1 \tilde{B}_1^{(1)}(\lambda_1 - \lambda_1 z_1)}{-\lambda_1 \tilde{B}_1^{(1)}(\lambda_1 - \lambda_1 z_1) - 1} \quad (\text{r\`egle de l'Hopital}) \\
&= \frac{-\lambda_1 \beta_{1,1}}{\lambda_1 \beta_{1,1} - 1} \\
&= \frac{\rho_1}{1 - \rho_1}.
\end{aligned}$$

Par suite, les expressions (4.26), (4.27), (4.30) et (4.33) deviennent au pont $z_1 = 1$

$$\begin{aligned}
P_0(1) &= \frac{\text{const}}{\lambda_2}; \quad P_1(1) = \frac{\text{const}}{\lambda_2} \frac{\rho_1}{1 - \rho_1}; \quad P_2(1) = \text{const} \left(\int_0^\infty (1-B_2(y)) dy \right); \\
P_3(1) &= \text{const} \frac{\rho_1}{1 - \rho_1} \left(\int_0^\infty (1-B_2(y)) dy \right).
\end{aligned}$$

À l'aide de l'équation de normalisation $P_0(1) + P_1(1) + P_2(1) + P_3(1) = 1$ et le logiciel **Maple**, on trouve que

$$\begin{aligned} const &= \frac{\lambda_2(1-\rho_1)}{1 + \lambda_2 \int_0^{\infty} (1-B_2(y)) dy} \\ &= \frac{1}{2(\lambda_1 + \lambda_2 + 1)}. \end{aligned}$$

À présent, nous pouvons trouver la fonction génératrice du nombre de clients en l'orbite

$$\begin{aligned} Q(z_1) &= P_0(z_1) + P_1(z_1) + P_2(z_1) + P_3(z_1) \\ &= \frac{(1-z_1) \left(1 + \lambda_2 \int_0^{\infty} (1-B_2(y)) dy \right)}{2(1+\lambda_1+\lambda_2)\lambda_2(K_1(z_1,1)-z_1)} \exp \left\{ \int_1^{z_1} \frac{\lambda_1 - \lambda_1 K_1(z_1,1)}{\theta(K_1(z_1,1)-z_1)} dz_1 \right\}. \end{aligned} \quad (4.34)$$

aussi bien que la fonction génératrice du nombre de clients prioritaires dans le système

$$\begin{aligned} H(z_1) &= P_0(z_1) + z_1 P_1(z_1) + P_2(z_1) + z_1 P_3(z_1) \\ &= \frac{K_1(z_1,1)(1-z_1) \left(1 + \lambda_2 \int_0^{\infty} (1-B_2(y)) dy \right)}{2(1+\lambda_1+\lambda_2)\lambda_2(K_1(z_1,1)-z_1)} \exp \left\{ \int_1^{z_1} \frac{\lambda_1 - \lambda_1 K_1(z_1,1)}{\theta(K_1(z_1,1)-z_1)} dz_1 \right\}. \end{aligned} \quad (4.35)$$

4.5. Mesures de performances

À l'aide des fonctions génératrices obtenues, nous pouvons calculer plusieurs mesures de performance de notre système, telles que :

- La probabilité p_1 que le serveur est occupé par le service d'un client prioritaire et il n'y a aucun client non prioritaire dont le service a été interrompu dans la station-service

$$p_1 = P_1(1) = \frac{\rho_1}{2\lambda_2(1-\rho_1)(1+\lambda_1+\lambda_2)}. \quad (4.36)$$

- La probabilité p_2 que le serveur est occupé par le service d'un client non prioritaire

$$p_2 = P_2(1) = \frac{1}{2(1+\lambda_1+\lambda_2)} \int_0^{\infty} (1-B_2(y)) dy. \quad (4.37)$$

- La probabilité p_3 que le serveur est occupé par le service d'un client prioritaire et il y a un client non prioritaire dont le service a été interrompu dans la station-service

$$p_3 = P_3(1) = \frac{\rho_1}{2(1+\lambda_1+\lambda_2)(1-\rho_1)} \int_0^{\infty} (1-B_2(y)) dy. \quad (4.38)$$

- Nombre moyen de clients prioritaires en l'orbite \bar{n}_o

$$\begin{aligned} \bar{n}_o &= Q'(1) \\ &= \frac{(2\rho_1\lambda_1 + \theta(1-\rho_1)\beta_{1,2} + 4\rho_1\lambda_1\beta_{1,1} + 2\lambda_1\beta_{1,1}^2)}{4\theta(1-\rho_1)(1+\lambda_1+\lambda_2)\lambda_2(1+\beta_{1,1})^2} \cdot \left(1 + \lambda_2 \int_0^{\infty} (1-B_2(y)) dy\right). \end{aligned} \quad (4.39)$$

- Nombre moyen de clients prioritaires dans le système \bar{n}

$$\begin{aligned} \bar{n} &= H'(1) \\ &= \frac{(2\rho_1\lambda_1 + \theta(1-\rho_1)\beta_{1,2} + 4\rho_1\lambda_1\beta_{1,1} + 2\lambda_1\beta_{1,1}^2)}{4\theta(1-\rho_1)(1+\lambda_1+\lambda_2)\lambda_2(1+\beta_{1,1})^2} \\ &\quad \times \left(1 + \lambda_2 \int_0^{\infty} (1-B_2(y)) dy\right) + \frac{\rho_1 \left(1 + \lambda_2 \int_0^{\infty} (1-B_2(y)) dy\right)}{2(1-\rho_1)\lambda_2(1+\lambda_1+\lambda_2)}. \end{aligned} \quad (4.40)$$

4.6. Illustration numérique

Dans ce paragraphe, nous présentons quelques résultats numériques pour illustrer l'effet du taux d'arrivée des clients prioritaires (il s'agit du taux qui définit la condition d'érgodicité) et celui de l'intensité des rappels sur les mesures de performance obtenues. A cette fin, nous considérons un système de files d'attente de type $M_1, M_2/G_1, G_2/1$ avec rappels et priorité de politique preemptive resume où le temps de service suit une distribution d'Erlang à deux étapes (E_2). Alors,

$$B_1(x) = 1 - e^{-\left(\frac{2}{\beta_{1,1}}\right)x} - \left(\frac{2}{\beta_{1,1}}\right)xe^{-\left(\frac{2}{\beta_{1,1}}\right)x}, x \geq 0;$$

$$B_2(y) = 1 - e^{-\left(\frac{2}{\beta_{2,1}}\right)y} - \left(\frac{2}{\beta_{2,1}}\right)ye^{-\left(\frac{2}{\beta_{2,1}}\right)y}, y \geq 0.$$

Tout au long de ce paragraphe, nous supposons que le temps moyen de service des deux types de client est $\beta_{1,1} = \beta_{2,1} = 1$ et que le taux d'arrivée des clients non prioritaire est $\lambda_2 = 0.3$. Alors,

$$p_1 = \frac{\lambda_1}{2 \cdot 0,3 \cdot (1 - \lambda_1)(1 + \lambda_1 + 0,3)}.$$

$$p_2 = \frac{1}{2(1 + \lambda_1 + 0,3)}.$$

$$p_3 = \frac{\lambda_1}{2(1 + \lambda_1 + 0,3)(1 - \lambda_1)}.$$

Dans la figure ci-dessous, nous présentons le comportement des probabilités p_1, p_2 et p_3 (donnée par (4.36)-(4.38)) par rapport aux différentes valeurs de λ_1 . Le choix des valeurs numériques des paramètres du système est effectué de façon assurant l'état d'équilibre du système.

| λ_1 | p_1 | p_2 | p_3 |
|-------------|--------------|--------------|---------------|
| 0.10 | 0.1190476191 | 0.3571428571 | 0.03571428572 |
| 0.25 | 0.2688172044 | 0.3225806452 | 0.08064516130 |
| 0.40 | 0.3921568627 | 0.2941176471 | 0.1176470588 |
| 0.60 | 0.5263157896 | 0.2631578947 | 0.1578947368 |

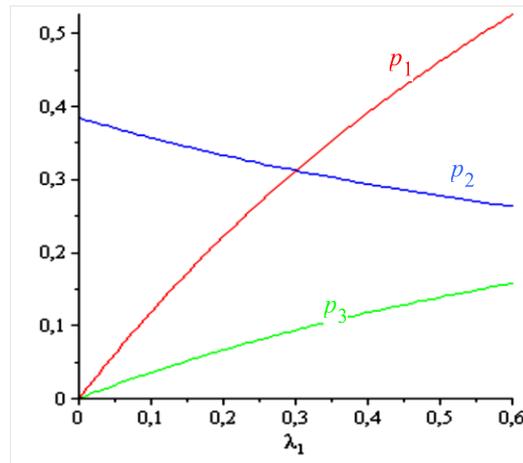


Figure1 : Le comportement des probabilités p_1 , p_2 et p_3 par rapport à λ_1 .

Comme prévu, l'augmentation du taux d'arrivée de clients prioritaire λ_1 abouti à une augmentation significative des probabilités p_1 et p_3 (liées aux clients prioritaires), et une diminution de la probabilité p_2 que le serveur est occupé par un client non prioritaire.

A présent, nous montrons l'influence du taux de rappels θ sur le nombre moyen de clients prioritaires en l'orbite \bar{n}_o et aussi dans le système \bar{n} (donnés par (4.39) et (4.40)).

L'expression de \bar{n}_o et de \bar{n} pour $\beta_{1,1} = 1, \rho_1 = \lambda_1 \beta_{1,1} = 0.6, \lambda_2 = 0.3$ deviennent

$$\bar{n}_o = \frac{0.3563596492(3.36 + \theta 0.6000000000)}{\theta},$$

$$\bar{n} = \frac{0.3563596492(3.36 + \theta 0.6000000000)}{\theta} + 0.6842105258.$$

| θ | \bar{n}_o | \bar{n} |
|----------|--------------|--------------|
| 1 | 1.411184211 | 2.095394737 |
| 5 | 0.4532894738 | 1.137500000 |
| 10 | 0.3335526317 | 1.017763158 |
| 100 | 0.2257894737 | 0.9099999995 |

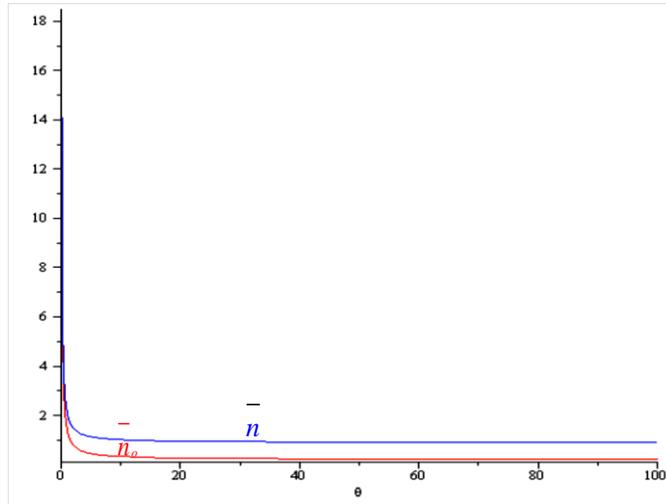


Figure2 : L'influence du taux de rappels θ sur les mesures de performance n_o et \bar{n}

Dans la dernière figure (où $\lambda_1 = \rho_1 = 0.6$), on remarque que l'augmentation du taux de rappels θ entraîne la diminution des mesures en question.

4.7. Conclusion

Dans ce chapitre, nous avons obtenu les fonctions génératrices partielles de la distribution stationnaire de l'état du système et l'expression de quelques mesures de performance. On constate que leurs expressions ne sont donc pas faciles à interpréter en pratique. Pour surmonter cette difficulté, nous ferons recours à une des méthodes d'approximation qui est la méthode de comparaison stochastique qui sera appliquée dans le chapitre suivant.

5

Monotonie de la chaîne de Markov induite d'un modèle M/G/1 avec rappels et priorité absolue

5.1 Introduction

L'objectif de ce chapitre est l'étude de la monotonie de l'opérateur de transition associé à la chaîne de Markov induite du modèle étudié dans le chapitre précédent par rapport à l'ordre stochastique fort et l'ordre convexe aussi bien que la comparabilité de deux opérateurs de transitions. Nous obtenons également les inégalités stochastiques pour

la distribution stationnaire du nombre de clients dans le système considéré. Ces résultats sont présentés dans [10].

5.2. Les propriétés de monotonie de la chaîne de Markov induite

Dans ce paragraphe, nous étudions la monotonie de l'opérateur de transition de la chaîne de Markov induite $\{X_d, d \geq 1\}$ relatif à l'ordre stochastique fort, \leq_{st} et l'ordre convexe, \leq_v . Nous donnons aussi les conditions de comparabilité de deux opérateurs de transition.

D'abord, nous rappelons la définition des susdits ordres stochastiques [11]. Supposons que nous ayons $F_1(x)$ et $F_2(x)$ deux fonctions de distribution de variables aléatoires non négatives. Alors

$$i) \quad F_1 \leq_{st} F_2 \text{ si et seulement si } F_1(x) \geq F_2(x) \text{ pour tout } x \geq 0.$$

$$ii) \quad F_1 \leq_v F_2 \text{ si et seulement si } \int_x^\infty (1 - F_1(u)) du \leq \int_x^\infty (1 - F_2(u)) du \text{ pour tout } x \geq 0.$$

Si ces variables sont discrets et $p^{(1)} = (p_n^{(1)})$, $p^{(2)} = (p_n^{(2)})$ sont leurs distributions correspondantes, les susdites définitions prennent la forme suivante :

$$i) \quad p^{(1)} \leq_{st} p^{(2)} \text{ si et seulement si } \sum_{n=m}^\infty p_n^{(1)} \leq \sum_{n=m}^\infty p_n^{(2)} \text{ pour tout } m.$$

$$ii) \quad p^{(1)} \leq_v p^{(2)} \text{ si et seulement si } \sum_{n=m}^\infty \sum_{k=n}^\infty p_k^{(1)} \leq \sum_{n=m}^\infty \sum_{k=n}^\infty p_k^{(2)} \text{ pour tout } m.$$

Soit T l'opérateur de transition de la chaîne de Markov induite $\{X_d, d \geq 1\}$, qu'à chaque distribution $p = (p_{l,m,n})$ associe une distribution $T_p = q = (q_{c,i,j})$ telle que

$$q_{c,i,j} = \sum_m \sum_n p_{l,m,n} r_{(l,m,n),(c,i,j)}.$$

Théorème 1. *L'opérateur T est monotone par rapport à l'ordre stochastique fort \leq_{st} (à l'ordre stochastique convexe \leq_v), c'est-à-dire pour n'importe quelles deux distributions $p^{(1)}$ et $p^{(2)}$ l'inégalité $p^{(1)} \leq_{st} p^{(2)}$ (l'inégalité $p^{(1)} \leq_v p^{(2)}$) implique $Tp^{(1)} \leq_{st} Tp^{(2)}$*

$$(Tp^{(1)} \leq_v Tp^{(2)}).$$

Preuve. D'après le théorème général 4.2.3 de Stoyan [11], l'opérateur T est monotone par rapport à \leq_{st} si et seulement si $\bar{r}_{(l,m-1,n-1),(c,i,j)} \leq \bar{r}_{(l,m,n),(c,i,j)}$, pour toutes les valeurs de m et n ;

et monotone par rapport à \leq_v si et seulement si $2\bar{r}_{(l,m,n),(c,i,j)} \leq \bar{r}_{(l,m-1,n-1),(c,i,j)} + \bar{r}_{(l,m+1,n+1),(c,i,j)}$,

pour toutes les valeurs de m et n , où $\bar{r}_{(l,m,n),(c,i,j)} = \sum_{h=i}^{\infty} \sum_{f=j}^{\infty} r_{(l,m,n),(c,h,f)}$ et

$$\bar{r}_{(l,m,n),(c,i,j)} = \sum_{h=i}^{\infty} \sum_{f=j}^{\infty} \bar{r}_{(l,m,n),(c,h,f)}.$$

En utilisant (4.4) - (4.5) (voir le chapitre 4) et $\bar{k}_{l,m,n} = \sum_{i=m}^{\infty} \sum_{j=n}^{\infty} k_{l,i,j}$, nous obtenons les

expressions suivantes pour $\bar{r}_{(l,m,n),(c,i,j)}$:

i) si $c=1$

$$\begin{aligned} \bar{r}_{(l,m,n),(1,i,j)} &= \frac{\lambda_1}{\lambda_1 + \lambda_2 + m\theta} \bar{k}_{1,i-m,0} + \frac{m\theta}{\lambda_1 + \lambda_2 + m\theta} \bar{k}_{1,i-m+1,0} \\ &= \frac{\lambda_1}{\lambda_1 + \lambda_2 + m\theta} \sum_{w=i-m}^{\infty} k_{1,w,0} + \frac{m\theta}{\lambda_1 + \lambda_2 + m\theta} \sum_{w=i-m+1}^{\infty} k_{1,w,0} \\ &= \frac{\lambda_1}{\lambda_1 + \lambda_2 + m\theta} \left(\sum_{w=i-m+1}^{\infty} k_{1,w,0} + k_{1,i-m,0} \right) + \frac{m\theta}{\lambda_1 + \lambda_2 + m\theta} \sum_{w=i-m+1}^{\infty} k_{1,w,0} \\ &= \frac{\lambda_1 + m\theta}{\lambda_1 + \lambda_2 + m\theta} \sum_{w=i-m+1}^{\infty} k_{1,w,0} + \frac{\lambda_1}{\lambda_1 + \lambda_2 + m\theta} k_{1,i-m,0} \\ &= \frac{\lambda_1 + m\theta}{\lambda_1 + \lambda_2 + m\theta} \bar{k}_{1,i-m+1,0} + \frac{\lambda_1}{\lambda_1 + \lambda_2 + m\theta} k_{1,i-m,0}. \end{aligned}$$

Aussi

$$\begin{aligned}
\bar{r}_{(l,m,n),(1,i,j)} &= \frac{\lambda_1}{\lambda_1 + \lambda_2 + m\theta} \bar{k}_{1,i-m,0} + \frac{m\theta}{\lambda_1 + \lambda_2 + m\theta} \bar{k}_{1,i-m+1,0} \\
&= \frac{\lambda_1}{\lambda_1 + \lambda_2 + m\theta} \sum_{w=i-m}^{\infty} k_{1,w,0} + \frac{m\theta}{\lambda_1 + \lambda_2 + m\theta} \sum_{w=i-m+1}^{\infty} k_{1,w,0} \\
&= \frac{\lambda_1}{\lambda_1 + \lambda_2 + m\theta} \sum_{w=i-m}^{\infty} k_{1,w,0} + \frac{m\theta}{\lambda_1 + \lambda_2 + m\theta} \left(\sum_{w=i-m}^{\infty} k_{1,w,0} - k_{1,i-m,0} \right) \\
&= \frac{\lambda_1 + m\theta}{\lambda_1 + \lambda_2 + m\theta} \sum_{w=i-m}^{\infty} k_{1,w,0} - \frac{m\theta}{\lambda_1 + \lambda_2 + m\theta} k_{1,i-m,0} \\
&= \frac{\lambda_1 + m\theta}{\lambda_1 + \lambda_2 + m\theta} \bar{k}_{1,i-m,0} - \frac{m\theta}{\lambda_1 + \lambda_2 + m\theta} k_{1,i-m,0}.
\end{aligned}$$

Alors, dans ce cas là

$$\begin{aligned}
\bar{r}_{(l,m,n),(1,i,j)} &= \frac{\lambda_1}{\lambda_1 + \lambda_2 + m\theta} \bar{k}_{1,i-m,0} + \frac{m\theta}{\lambda_1 + \lambda_2 + m\theta} \bar{k}_{1,i-m+1,0} \\
&= \frac{\lambda_1 + m\theta}{\lambda_1 + \lambda_2 + m\theta} \bar{k}_{1,i-m+1,0} + \frac{\lambda_1}{\lambda_1 + \lambda_2 + m\theta} k_{1,i-m,0} \\
&= \frac{\lambda_1 + m\theta}{\lambda_1 + \lambda_2 + m\theta} \bar{k}_{1,i-m,0} - \frac{m\theta}{\lambda_1 + \lambda_2 + m\theta} k_{1,i-m,0}.
\end{aligned} \tag{5.1}$$

ii) si $c = 2$

$$\bar{r}_{(l,m,n),(2,i,j)} = \frac{\lambda_2}{\lambda_1 + \lambda_2 + m\theta} \bar{k}_{2,i-m,0} \text{ pour } n \in \{0,1\} \text{ et } m = i. \tag{5.2}$$

Ainsi,

$$\begin{aligned}
\bar{r}_{(l,m,n),(1,i,j)} - \bar{r}_{(l,m-1,n-1),(1,i,j)} &= \frac{\lambda_1 + m\theta}{\lambda_1 + \lambda_2 + m\theta} \bar{k}_{1,i-m+1,0} + \frac{\lambda_1}{\lambda_1 + \lambda_2 + m\theta} k_{1,i-m,0} \\
&\quad - \frac{\lambda_1 + (m-1)\theta}{\lambda_1 + \lambda_2 + (m-1)\theta} \bar{k}_{1,i-m+1,0} + \frac{(m-1)\theta}{\lambda_1 + \lambda_2 + (m-1)\theta} k_{1,i-m+1,0} \\
&= \frac{\lambda_1}{\lambda_1 + \lambda_2 + m\theta} k_{1,i-m,0} + \frac{(m-1)\theta}{\lambda_1 + \lambda_2 + (m-1)\theta} k_{1,i-m+1,0} \\
&\quad \bar{k}_{1,i-m+1,0} \left(\frac{\lambda_1 + m\theta}{\lambda_1 + \lambda_2 + m\theta} - \frac{\lambda_1 + (m-1)\theta}{\lambda_1 + \lambda_2 + (m-1)\theta} \right) \\
&= \frac{\lambda_1}{\lambda_1 + \lambda_2 + m\theta} k_{1,i-m,0} + \frac{(m-1)\theta}{\lambda_1 + \lambda_2 + (m-1)\theta} k_{1,i-m+1,0} \\
&\quad + \frac{\theta\lambda_2}{(\lambda_1 + \lambda_2 + m\theta)(\lambda_1 + \lambda_2 + (m-1)\theta)} \bar{k}_{1,i-m+1,0} \geq 0.
\end{aligned}$$

et également

$$\bar{r}_{(l,m,n),(2,i,j)} - \bar{r}_{(l,m-1,n-1),(2,i,j)} = \frac{\lambda_2}{\lambda_1 + \lambda_2 + i\theta} \bar{k}_{2,0,0} - \frac{\lambda_2}{\lambda_1 + \lambda_2 + i\theta} \bar{k}_{2,0,0} = 0.$$

D'où, T est monotone par rapport à l'ordre \leq_{st} . De plus

$$\begin{aligned}
\bar{r}_{(l,m,n),(1,i,j)} &= \frac{\lambda_1 + m\theta}{\lambda_1 + \lambda_2 + m\theta} \bar{k}_{1,i-m+1,0} + \frac{\lambda_1}{\lambda_1 + \lambda_2 + m\theta} \bar{k}_{1,i-m,0} \\
&= \frac{\lambda_1 + m\theta}{\lambda_1 + \lambda_2 + m\theta} \bar{k}_{1,i-m,0} - \frac{m\theta}{\lambda_1 + \lambda_2 + m\theta} \bar{k}_{1,i-m,0}.
\end{aligned}$$

Par conséquent,

$$\begin{aligned}
& \bar{\bar{r}}_{(l,m-1,n-1),(1,i,j)} + \bar{\bar{r}}_{(l,m+1,n+1),(1,i,j)} - 2\bar{\bar{r}}_{(l,m,n),(1,i,j)} = \frac{\lambda_1 + (m-1)\theta}{\lambda_1 + \lambda_2 + (m-1)\theta} \bar{\bar{k}}_{1,i-m+1,0} - \frac{(m-1)\theta}{\lambda_1 + \lambda_2 + (m-1)\theta} \bar{\bar{k}}_{1,i-m+1,0} \\
& \quad + \frac{\lambda_1 + (m+1)\theta}{\lambda_1 + \lambda_2 + (m+1)\theta} \bar{\bar{k}}_{1,i-m,0} + \frac{\lambda_1}{\lambda_1 + \lambda_2 + (m+1)\theta} \bar{\bar{k}}_{1,i-m-1,0} \\
& \quad - \left(\frac{\lambda_1 + m\theta}{\lambda_1 + \lambda_2 + m\theta} \bar{\bar{k}}_{1,i-m+1,0} + \frac{\lambda_1}{\lambda_1 + \lambda_2 + m\theta} \bar{\bar{k}}_{1,i-m,0} \right) \\
& \quad - \left(\frac{\lambda_1 + m\theta}{\lambda_1 + \lambda_2 + m\theta} \bar{\bar{k}}_{1,i-m,0} - \frac{m\theta}{\lambda_1 + \lambda_2 + m\theta} \bar{\bar{k}}_{1,i-m-1,0} \right) \\
& = \frac{\lambda_1 + (m-1)\theta}{\lambda_1 + \lambda_2 + (m-1)\theta} (\bar{\bar{k}}_{1,i-m,0} - \bar{\bar{k}}_{1,i-m-1,0}) - \frac{(m-1)\theta}{\lambda_1 + \lambda_2 + (m-1)\theta} (\bar{\bar{k}}_{1,i-m,0} - k_{1,i-m,0}) \\
& \quad + \frac{\lambda_1 + (m+1)\theta}{\lambda_1 + \lambda_2 + (m+1)\theta} \bar{\bar{k}}_{1,i-m,0} + \frac{\lambda_1}{\lambda_1 + \lambda_2 + (m+1)\theta} (\bar{\bar{k}}_{1,i-m,0} + k_{1,i-m-1,0}) \\
& \quad - \left(\frac{\lambda_1 + m\theta}{\lambda_1 + \lambda_2 + m\theta} (\bar{\bar{k}}_{1,i-m,0} - \bar{\bar{k}}_{1,i-m-1,0}) + \frac{\lambda_1}{\lambda_1 + \lambda_2 + m\theta} \bar{\bar{k}}_{1,i-m,0} \right) \\
& \quad - \left(\frac{\lambda_1 + m\theta}{\lambda_1 + \lambda_2 + m\theta} \bar{\bar{k}}_{1,i-m,0} - \frac{m\theta}{\lambda_1 + \lambda_2 + m\theta} \bar{\bar{k}}_{1,i-m-1,0} \right) \\
& = \left(\frac{\lambda_1 + (m-1)\theta}{\lambda_1 + \lambda_2 + (m-1)\theta} + \frac{\lambda_1 + (m+1)\theta}{\lambda_1 + \lambda_2 + (m+1)\theta} - 2 \frac{\lambda_1 + m\theta}{\lambda_1 + \lambda_2 + m\theta} \right) \bar{\bar{k}}_{1,i-m,0} \\
& \quad + \left(\frac{\lambda_1}{\lambda_1 + \lambda_2 + (m-1)\theta} + \frac{\lambda_1}{\lambda_1 + \lambda_2 + (m+1)\theta} + 2 \frac{\lambda_1 + m\theta}{\lambda_1 + \lambda_2 + m\theta} \right) \bar{\bar{k}}_{1,i-m-1,0} \\
& \quad + \frac{(m-1)\theta}{\lambda_1 + \lambda_2 + (m-1)\theta} k_{1,i-m,0} + \frac{\lambda_1}{\lambda_1 + \lambda_2 + (m+1)\theta} k_{1,i-m-1,0} \\
& = \bar{\bar{k}}_{1,i-m,0} \left(- \frac{2\theta^2 \lambda_2}{(\lambda_1 + \lambda_2 + (m+1)\theta)(\lambda_1 + \lambda_2 + (m-1)\theta)(\lambda_1 + \lambda_2 + m\theta)} \right) \\
& \quad + \bar{\bar{k}}_{1,i-m-1,0} \left(\frac{2\theta(m\theta\lambda_2 + \lambda_2\lambda_1 + \theta\lambda_1 + \lambda_2^2 + \theta\lambda_2)}{(\lambda_1 + \lambda_2 + m\theta)(\lambda_1 + \lambda_2 + (m-1)\theta)(\lambda_1 + \lambda_2 + (m+1)\theta)} \right) \\
& \quad + k_{1,i-m,0} \frac{(m-1)\theta}{\lambda_1 + \lambda_2 + (m-1)\theta} + k_{1,i-m-1,0} \frac{\lambda_1}{\lambda_1 + \lambda_2 + (m+1)\theta} \geq 0.
\end{aligned}$$

Nous trouvons également que

$$\bar{\bar{r}}_{(l,m-1,n-1),(2,i,j)} + \bar{\bar{r}}_{(l,m+1,n+1),(2,i,j)} - 2\bar{\bar{r}}_{(l,m,n),(2,i,j)} = 0.$$

Alors, T est monotone par rapport à l'ordre \leq_v .

Fin de preuve

A présent, nous présentons deux lemmes qui seront utilisés dans la démonstration de la comparabilité de deux opérateurs de transition.

Soient $M^{(1)}$ et $M^{(2)}$ deux systèmes d'attente avec rappels et priorité *absolue de politique preemptive resume* de type $M_1, M_2/G_1, G_2/1$ définis par $\lambda_l^{(1)}, \theta^{(1)}, B_l^{(1)}(x), k_{l,m,n}^{(1)}$ and $\lambda_l^{(2)}, \theta^{(2)}, B_l^{(2)}(x), k_{l,m,n}^{(2)}$ respectivement.

Lemme 1. Si $\lambda_l^{(1)} \leq \lambda_l^{(2)}$ et $B_l^{(1)}(x) \leq_{st} B_l^{(2)}(x)$ alors $\{k_{l,m,n}^{(1)}\} \leq_{st} \{k_{l,m,n}^{(2)}\}$, $l \in \{1, 2\}$.

Preuve. pour prouver que $\{k_{l,m,n}^{(1)}\} \leq_{st} \{k_{l,m,n}^{(2)}\}$, $l \in \{1, 2\}$, il est nécessaire d'établir l'inégalité

suivante: $\bar{k}_{l,i,j}^{(1)} = \sum_{m=i}^{\infty} \sum_{n=j}^{\infty} k_{l,m,n}^{(1)} \leq \bar{k}_{l,i,j}^{(2)} = \sum_{m=i}^{\infty} \sum_{n=j}^{\infty} k_{l,m,n}^{(2)}$, i.e. .

$$\int_0^{\infty} f_{l,i,j}(\lambda_1^{(1)}, \lambda_2^{(1)}, x) dB_l^{(1)}(x) \leq \int_0^{\infty} f_{l,i,j}(\lambda_1^{(2)}, \lambda_2^{(2)}, x) dB_l^{(2)}(x). \quad (5.3)$$

Ici, $f_{l,i,j}(\lambda_1, \lambda_2, x) = \sum_{m=i}^{\infty} \sum_{n=j}^{\infty} \frac{(\lambda_1 x)^m}{m!} e^{(-\lambda_1 x)} \frac{(\lambda_2 x)^n}{n!} e^{(-\lambda_2 x)}$. Ses dérivées partielles sont

$$\frac{\partial f_{l,i,j}(\lambda_1, \lambda_2, x)}{\partial \lambda_1} = \frac{x(\lambda_1 x)^{i-1} e^{(-\lambda_1 x)} (\Gamma(j) - \Gamma(j, \lambda_2 x))}{(i-1)!(j-1)!};$$

$$\frac{\partial f_{l,i,j}(\lambda_1, \lambda_2, x)}{\partial \lambda_2} = \frac{x(\lambda_2 x)^{j-1} e^{(-\lambda_2 x)} (\Gamma(i) - \Gamma(i, \lambda_1 x))}{(i-1)!(j-1)!};$$

$$\begin{aligned} \frac{\partial f_{l,i,j}(\lambda_1, \lambda_2, x)}{\partial x} &= \frac{\lambda_2 (\lambda_2 x)^{j-1} e^{(-\lambda_2 x)} (\Gamma(i) - \Gamma(i, \lambda_1 x))}{(i-1)!(j-1)!} \\ &+ \frac{\lambda_1 (\lambda_1 x)^{i-1} e^{(-\lambda_1 x)} (\Gamma(j) - \Gamma(j, \lambda_2 x))}{(i-1)!(j-1)!}. \end{aligned}$$

Pour déterminer leur signe, nous devons seulement déterminer le signe de $(\Gamma(j) - \Gamma(j, \lambda_2 x))$ et $(\Gamma(i) - \Gamma(i, \lambda_1 x))$. Il est connu que $\Gamma(i, \lambda_1 x) = \int_{\lambda_1 x}^{\infty} t^{i-1} e^{-t} dt$ est une fonction gamma incomplète supérieure et $\gamma(i, \lambda_1 x) = \int_0^{\lambda_1 x} t^{i-1} e^{-t} dt$ est une fonction gamma incomplète inférieure de paramètres réels i et $\lambda_1 x$. Une fonction gamma ordinaire de paramètre i est définie comme $\Gamma(i) = \int_0^{\infty} t^{i-1} e^{-t} dt$, $\gamma(i, \lambda_1 x) = \Gamma(i) - \Gamma(i, \lambda_1 x)$. De plus, $\gamma(i, \lambda_1 x)$ pour des paramètres positifs réels i et $\lambda_1 x$ peut être développé en une fonction holomorphe: $\gamma(i, \lambda_1 x) = \sum_{k=0}^{\infty} \frac{(\lambda_1 x)^i e^{(-\lambda_1 x)} (\lambda_1 x)^k}{i(i+1)\dots(i+k)}$. Elle est évidemment positive. Donc, $(\Gamma(i) - \Gamma(i, \lambda_1 x))$ a un signe positif. De la même façon nous pouvons prouver que $(\Gamma(j) - \Gamma(j, \lambda_2 x))$ a aussi un signe positif. Par conséquent, les susdites dérivées sont positives. Ceci implique que $f_{l,i,j}(\lambda_1, \lambda_2, x)$ est une fonction croissante par rapport à λ_1 , λ_2 et x . D'où, avec l'aide de théorème 1.2.2 donné dans Stoyan [11], nous trouvons que

$$\int_0^{\infty} f_{l,i,j}(\lambda_1^{(1)}, \lambda_2^{(1)}, x) dB_l^{(1)}(x) \leq \int_0^{\infty} f_{l,i,j}(\lambda_1^{(2)}, \lambda_2^{(2)}, x) dB_l^{(2)}(x), \quad l \in \{1, 2\}. \quad (5.4)$$

D'autre part, par la monotonie de la fonction $f_{l,i,j}(\lambda_1, \lambda_2, x)$ par rapport à λ_1 et λ_2 ,

$$\int_0^{\infty} f_{l,i,j}(\lambda_1^{(1)}, \lambda_2^{(1)}, x) dB_l^{(2)}(x) \leq \int_0^{\infty} f_{l,i,j}(\lambda_1^{(2)}, \lambda_2^{(2)}, x) dB_l^{(2)}(x), \quad l \in \{1, 2\}. \quad (5.5)$$

De (5.4) et (5.5), on trouve (5.3).

Fin de preuve.

Lemme 2. Si $\lambda_q^{(1)} \leq \lambda_q^{(2)}$ et $B_l^{(1)}(x) \leq_v B_l^{(2)}(x)$ alors $\{k_{l,m,n}^{(1)}\} \leq_v \{k_{l,m,n}^{(2)}\}$, $l \in \{1, 2\}$.

Preuve. Pour prouver que $\{k_{l,m,n}^{(1)}\} \leq_v \{k_{l,m,n}^{(2)}\}$, $l \in \{1, 2\}$, nous devons obtenir l'inégalité

suivante: $\bar{k}_{l,i,j}^{(1)} = \sum_{m=i}^{\infty} \sum_{n=j}^{\infty} \bar{k}_{l,m,n}^{(1)} \leq \bar{k}_{l,i,j}^{(2)} = \sum_{m=i}^{\infty} \sum_{n=j}^{\infty} \bar{k}_{l,m,n}^{(2)}$ avec $\bar{k}_{l,m,n}^{(1)} = \sum_{k=m}^{\infty} \sum_{h=n}^{\infty} k_{l,k,h}^{(1)}$ et

$$\bar{k}_{l,m,n}^{(2)} = \sum_{k=m}^{\infty} \sum_{h=n}^{\infty} k_{l,k,h}^{(2)} ; \text{ i.e.}$$

$$\int_0^{\infty} \sum_{m=i}^{\infty} \sum_{n=j}^{\infty} f_{l,m,n}(\lambda_1^{(1)}, \lambda_2^{(1)}, x) dB_l^{(1)}(x) \leq \int_0^{\infty} \sum_{m=i}^{\infty} \sum_{n=j}^{\infty} f_{l,m,n}(\lambda_1^{(2)}, \lambda_2^{(2)}, x) dB_l^{(2)}(x). \quad (5.6)$$

Considérons $\bar{f}_{l,i,j}(\lambda_1, \lambda_2, x) = \sum_{m=i}^{\infty} \sum_{n=j}^{\infty} f_{l,m,n}(\lambda_1, \lambda_2, x) = \sum_{m=i}^{\infty} \sum_{n=j}^{\infty} \sum_{k=m}^{\infty} \sum_{h=n}^{\infty} \frac{(\lambda_1 x)^k}{k!} e^{(-\lambda_1 x)} \frac{(\lambda_2 x)^h}{h!} e^{(-\lambda_2 x)}$.

Cette fonction a les dérivées partielles suivantes :

$$\frac{\partial \bar{f}_{l,i,j}}{\partial \lambda_1}(\lambda_1, \lambda_2, x) = \sum_{n=j}^{\infty} \frac{(\Gamma(n) - \Gamma(n, \lambda_2 x)) x (\Gamma(i-1) - \Gamma(i-1, \lambda_1 x))}{(n-1)!(i-2)!};$$

$$\frac{\partial \bar{f}_{l,i,j}}{\partial \lambda_2}(\lambda_1, \lambda_2, x) = \sum_{m=i}^{\infty} \frac{(\Gamma(m) - \Gamma(m, \lambda_1 x)) x (\Gamma(j-1) - \Gamma(j-1, \lambda_2 x))}{(m-1)!(j-2)!};$$

$$\frac{\partial \bar{f}_{l,i,j}}{\partial x}(\lambda_1, \lambda_2, x) = \sum_{m=i}^{\infty} \left(\sum_{n=j}^{\infty} \frac{\lambda_1 (\lambda_1 x)^{m-1} e^{(-\lambda_1 x)} (\Gamma(n) - \Gamma(n, \lambda_2 x))}{(m-1)!(n-1)!} \right.$$

$$\left. + \frac{\lambda_2 (\lambda_2 x)^{n-1} e^{(-\lambda_2 x)} (\Gamma(m) - \Gamma(m, \lambda_1 x))}{(m-1)!(n-1)!} \right);$$

$$\begin{aligned}
\frac{\partial^2 \bar{f}_{l,i,j}(\lambda_1, \lambda_2, x)}{\partial x^2} &= \sum_{m=i}^{\infty} \left(\sum_{n=j}^{\infty} \frac{\lambda_2 (\lambda_2 x)^{n-1} e^{(-\lambda_2 x)} (\Gamma(m+1) - m\Gamma(m, \lambda_1 x))}{m!(n-2)!x} \right. \\
&+ \frac{\lambda_2^2 (\lambda_2 x)^{n-1} e^{(-\lambda_2 x)} (m\Gamma(m, \lambda_1 x) - \Gamma(m+1))}{(n-1)!m!} + \frac{2\lambda_2 (\lambda_2 x)^{n-1} e^{(-(\lambda_2 + \lambda_1))} \lambda_1 (\lambda_1 x)^{m-1}}{(m-1)!(n-1)!} \\
&\left. + \frac{\lambda_1 (\lambda_1 x)^{m-1} e^{(-\lambda_1 x)} (\Gamma(n+1) - n\Gamma(n, \lambda_2 x))}{n!(m-2)!x} + \frac{\lambda_1^2 (\lambda_1 x)^{m-1} e^{(-\lambda_1 x)} (n\Gamma(n, \lambda_2 x) - \Gamma(n+1))}{n!(m-1)!} \right).
\end{aligned}$$

Puisque la fonction gamma incomplète inférieure (pour un paramètre positif) est positive, les dérivées partielles $\frac{\partial \bar{f}_{l,i,j}(\lambda_1, \lambda_2, x)}{\partial \lambda_1}$, $\frac{\partial \bar{f}_{l,i,j}(\lambda_1, \lambda_2, x)}{\partial \lambda_2}$ et $\frac{\partial \bar{f}_{l,i,j}(\lambda_1, \lambda_2, x)}{\partial x}$ sont aussi positives.

Par conséquent, la fonction $\bar{f}_{l,i,j}(\lambda_1, \lambda_2, x)$ est croissante par rapport à λ_1 , λ_2 et x . A l'aide du programme Maple nous avons prouvé que $\frac{\partial^2 \bar{f}_{l,i,j}(\lambda_1, \lambda_2, x)}{\partial x^2}$ a un signe positif. Donc, la fonction

$$\bar{f}_{l,i,j}(\lambda_1, \lambda_2, x) \text{ est convexe par rapport à } x. \text{ D'après le théorème 1.3.1 de Stoyan [40],}$$

$$\int_0^{\infty} \bar{f}_{l,i,j}(\lambda_1^{(1)}, \lambda_2^{(1)}, x) dB_l^{(1)}(x) \leq \int_0^{\infty} \bar{f}_{l,i,j}(\lambda_1^{(1)}, \lambda_2^{(1)}, x) dB_l^{(2)}(x), \quad l \in \{1, 2\}. \quad (5.7)$$

D'autre part, par la monotonie de $\bar{f}_{l,i,j}(\lambda_1, \lambda_2, x)$ par rapport à λ_1 et λ_2 ,

$$\int_0^{\infty} \bar{f}_{l,i,j}(\lambda_1^{(1)}, \lambda_2^{(1)}, x) dB_l^{(2)}(x) \leq \int_0^{\infty} \bar{f}_{l,i,j}(\lambda_1^{(2)}, \lambda_2^{(2)}, x) dB_l^{(2)}(x), \quad l \in \{1, 2\}. \quad (5.8)$$

À présent, nous pouvons conclure que (5.6) suit de (5.7) et (5.8).

Fin de preuve.

Dénotons par $T^{(1)}$, $T^{(2)}$ les operateurs de transitions de la chaîne de Markov induite des système $M^{(1)}$ et $M^{(2)}$, respectivement.

Théorème 2. Si $\lambda_l^{(1)} \leq \lambda_l^{(2)}$, $\theta^{(1)} \geq \theta^{(2)}$ et $B_l^{(1)}(x) \leq_s B_l^{(2)}(x)$, où s est st ou bien v , alors $T^{(1)} \leq_s T^{(2)}$, i.e. pour n'importe quelle distribution p nous avons $T^{(1)} p \leq_s T^{(2)} p$.

Preuve. D'après le théorème général 4.2.3 de Stoyan [11], nous devons établir que pour les probabilités de transition à une étape $r_{(l,m,n),(1,i,j)}^{(1)}$, $r_{(l,m,n),(1,i,j)}^{(2)}$, $r_{(l,m,n),(2,i,j)}^{(1)}$ et $r_{(l,m,n),(2,i,j)}^{(2)}$ correspondantes les inégalités numériques suivantes ont lieu :

$$\bar{r}_{(l,m,n),(1,i,j)}^{(1)} \leq_{st} \bar{r}_{(l,m,n),(1,i,j)}^{(2)}, \bar{r}_{(l,m,n),(2,i,j)}^{(1)} \leq_{st} \bar{r}_{(l,m,n),(2,i,j)}^{(2)}; \quad (5.9)$$

$$\bar{\bar{r}}_{(l,m,n),(1,i,j)}^{(1)} \leq_v \bar{\bar{r}}_{(l,m,n),(1,i,j)}^{(2)}, \bar{\bar{r}}_{(l,m,n),(2,i,j)}^{(1)} \leq_v \bar{\bar{r}}_{(l,m,n),(2,i,j)}^{(2)}. \quad (5.10)$$

Du lemme 1 et lemme 2, $\{k_{l,m,n}^{(1)}\} \leq_{st} \{k_{l,m,n}^{(2)}\}$ et $\{k_{l,m,n}^{(1)}\} \leq_v \{k_{l,m,n}^{(2)}\}$, $l \in \{1,2\}$. En plus,

$\lambda_l^{(1)} \leq \lambda_l^{(2)}$ et $\theta^{(1)} \geq \theta^{(2)}$ impliquent que $\frac{\lambda_1^{(1)} + \lambda_2^{(1)}}{\theta^{(1)}} \leq \frac{\lambda_1^{(2)} + \lambda_2^{(2)}}{\theta^{(2)}}$. Puisque la fonction $\frac{m}{x+m}$

est décroissante, il est évident que $\frac{m\theta^{(1)}}{\lambda_1^{(1)} + \lambda_2^{(1)} + m\theta^{(1)}} \geq \frac{m\theta^{(2)}}{\lambda_1^{(2)} + \lambda_2^{(2)} + m\theta^{(2)}}$. Aussi, $\lambda_1^{(1)} \leq \lambda_1^{(2)}$

et $\theta^{(1)} \geq \theta^{(2)}$ impliquent que $\frac{\lambda_1^{(1)}}{\theta^{(1)}} \leq \frac{\lambda_1^{(2)}}{\theta^{(2)}}$. Comme la fonction $\frac{x}{x+m}$ est croissante, il est

vrai que $\frac{\lambda_1^{(1)}}{\lambda_1^{(1)} + m\theta^{(1)}} \leq \frac{\lambda_1^{(2)}}{\lambda_1^{(2)} + m\theta^{(2)}}$. De la dernière inégalité nous pouvons trouver

$\lambda_1^{(1)} + m\theta_1^{(1)} \geq \lambda_1^{(2)} + m\theta_1^{(2)}$. Les inégalités $\lambda_2^{(1)} \leq \lambda_2^{(2)}$ et $\lambda_1^{(1)} + m\theta_1^{(1)} \geq \lambda_1^{(2)} + m\theta_1^{(2)}$ donnent

$\frac{\lambda_1^{(1)} + m\theta^{(1)}}{\lambda_2^{(1)}} \geq \frac{\lambda_1^{(2)} + m\theta^{(2)}}{\lambda_2^{(2)}}$. Comme la fonction $\frac{x}{x+1}$ est croissante, l'inégalité

$\frac{\lambda_1^{(1)} + m\theta^{(1)}}{\lambda_1^{(1)} + \lambda_2^{(1)} + m\theta^{(1)}} \geq \frac{\lambda_1^{(2)} + m\theta^{(2)}}{\lambda_1^{(2)} + \lambda_2^{(2)} + m\theta^{(2)}}$ a lieu. En utilisant (5.1), il est possible de trouver

$$\begin{aligned} \bar{r}_{(l,m,n),(1,i,j)}^{(1)} - \bar{r}_{(l,m,n),(1,i,j)}^{(2)} &= \frac{\lambda_1^{(1)} + m\theta^{(1)}}{\lambda_1^{(1)} + \lambda_2^{(1)} + m\theta} \bar{k}_{1,i-m,0}^{(1)} - \frac{m\theta^{(1)}}{\lambda_1^{(1)} + \lambda_2^{(1)} + m\theta^{(1)}} k_{1,i-m,0}^{(1)} \\ &\quad - \frac{\lambda_1^{(2)} + m\theta^{(2)}}{\lambda_1^{(2)} + \lambda_2^{(2)} + m\theta} \bar{k}_{1,i-m,0}^{(2)} + \frac{m\theta^{(2)}}{\lambda_1^{(2)} + \lambda_2^{(2)} + m\theta^{(2)}} k_{1,i-m,0}^{(2)} \end{aligned}$$

$$\begin{aligned} &\leq \frac{\lambda_1^{(1)} + m\theta^{(1)}}{\lambda_1^{(1)} + \lambda_2^{(1)} + m\theta} \bar{k}_{1,i-m,0}^{(1)} - \frac{m\theta^{(2)}}{\lambda_1^{(2)} + \lambda_2^{(2)} + m\theta^{(2)}} k_{1,i-m,0}^{(2)} \\ &\quad - \frac{\lambda_1^{(1)} + m\theta^{(1)}}{\lambda_1^{(1)} + \lambda_2^{(1)} + m\theta^{(1)}} \bar{k}_{1,i-m,0}^{(1)} + \frac{m\theta^{(2)}}{\lambda_1^{(2)} + \lambda_2^{(2)} + m\theta^{(2)}} k_{1,i-m,0}^{(2)} = 0. \end{aligned}$$

et

$$\begin{aligned} \bar{r}_{(l,m,n),(1,i,j)}^{(1)} - \bar{r}_{(l,m,n),(1,i,j)}^{(2)} &= \frac{\lambda_1^{(1)} + m\theta^{(1)}}{\lambda_1^{(1)} + \lambda_2^{(1)} + m\theta} \bar{k}_{1,i-m,0}^{(1)} - \frac{m\theta^{(1)}}{\lambda_1^{(1)} + \lambda_2^{(1)} + m\theta^{(1)}} \bar{k}_{1,i-m,0}^{(1)} \\ &\quad - \frac{\lambda_1^{(2)} + m\theta^{(2)}}{\lambda_1^{(2)} + \lambda_2^{(2)} + m\theta} \bar{k}_{1,i-m,0}^{(2)} + \frac{m\theta^{(2)}}{\lambda_1^{(2)} + \lambda_2^{(2)} + m\theta^{(2)}} \bar{k}_{1,i-m,0}^{(2)} \\ &\leq \frac{\lambda_1^{(1)} + m\theta^{(1)}}{\lambda_1^{(1)} + \lambda_2^{(1)} + m\theta} \bar{k}_{1,i-m,0}^{(1)} - \frac{m\theta^{(2)}}{\lambda_1^{(2)} + \lambda_2^{(2)} + m\theta^{(2)}} \bar{k}_{1,i-m,0}^{(2)} \\ &\quad - \frac{\lambda_1^{(1)} + m\theta^{(1)}}{\lambda_1^{(1)} + \lambda_2^{(1)} + m\theta^{(1)}} \bar{k}_{1,i-m,0}^{(1)} + \frac{m\theta^{(2)}}{\lambda_1^{(2)} + \lambda_2^{(2)} + m\theta^{(2)}} \bar{k}_{1,i-m,0}^{(2)} = 0. \end{aligned}$$

Donc,

$$\bar{r}_{(l,m,n),(1,i,j)}^{(1)} \leq_{st} \bar{r}_{(l,m,n),(1,i,j)}^{(2)},$$

$$\bar{r}_{(l,m,n),(1,i,j)}^{(1)} \leq_v \bar{r}_{(l,m,n),(1,i,j)}^{(2)}.$$

De la relation $\frac{\lambda_1 + \lambda_2 + m\theta}{\lambda_1 + \lambda_2 + m\theta} = 1$, on trouve que $\frac{\lambda_1 + m\theta}{\lambda_1 + \lambda_2 + m\theta} = 1 - \frac{\lambda_2}{\lambda_1 + \lambda_2 + m\theta}$. Alors

l'inégalité $\frac{\lambda_1^{(1)} + m\theta^{(1)}}{\lambda_1^{(1)} + \lambda_2^{(1)} + m\theta^{(1)}} \geq \frac{\lambda_1^{(2)} + m\theta^{(2)}}{\lambda_1^{(2)} + \lambda_2^{(2)} + m\theta^{(2)}}$ peut être réécrite comme suit :

$$1 - \frac{\lambda_2^{(1)}}{\lambda_1^{(1)} + \lambda_2^{(1)} + m\theta^{(1)}} \geq 1 - \frac{\lambda_2^{(2)}}{\lambda_1^{(2)} + \lambda_2^{(2)} + m\theta^{(2)}}. \text{ D'où, on constate que}$$

$$\frac{\lambda_2^{(1)}}{\lambda_1^{(1)} + \lambda_2^{(1)} + m\theta^{(1)}} \leq \frac{\lambda_2^{(2)}}{\lambda_1^{(2)} + \lambda_2^{(2)} + m\theta^{(2)}} \text{ et}$$

$$\bar{r}_{(l,m,n)(2,i,j)}^{-(1)} = \frac{\lambda_2^{(1)}}{\lambda_1^{(1)} + \lambda_2^{(1)} + m\theta^{(1)}} \bar{k}_{2,i-m,0}^{-(1)} \leq_{st} \frac{\lambda_2^{(2)}}{\lambda_1^{(2)} + \lambda_2^{(2)} + m\theta^{(2)}} \bar{k}_{2,i-m,0}^{-(2)} = \bar{r}_{(l,m,n)(2,i,j)}^{-(2)},$$

$$\bar{\bar{r}}_{(l,m,n)(2,i,j)}^{-(1)} = \frac{\lambda_2^{(1)}}{\lambda_1^{(1)} + \lambda_2^{(1)} + m\theta^{(1)}} \bar{\bar{k}}_{2,i-m,0}^{-(1)} \leq_v \frac{\lambda_2^{(2)}}{\lambda_1^{(2)} + \lambda_2^{(2)} + m\theta^{(2)}} \bar{\bar{k}}_{2,i-m,0}^{-(2)} = \bar{\bar{r}}_{(l,m,n)(2,i,j)}^{-(2)}.$$

L'inégalité (5.9) et (5.10) sont obtenues.

Fin de preuve.

L'utilité des résultats obtenus dans cette section sera montrée dans le paragraphe suivant.

5.3. Inégalités stochastiques pour la distribution stationnaire

Dans ce paragraphe, nous supposons encore une fois que nous ayons $M^{(1)}$ et $M^{(2)}$ comme deux systèmes d'attente avec rappels et priorité *absolue de politique preemptive resume* de type $M_1, M_2/G_1, G_2/1$ définis par $\lambda_i^{(1)}, \theta^{(1)}, B_i^{(1)}(x)$ and $\lambda_i^{(2)}, \theta^{(2)}, B_i^{(2)}(x)$ respectivement. Dénoteons par $\{X_d^{(1)}, d \geq 1\}, \{X_d^{(2)}, d \geq 1\}$ les chaînes de Markov induites correspondantes et par $\pi_{mn}^{(1)}, \pi_{mn}^{(2)}$ leurs distributions stationnaires respectivement.

Théorème 3. Si $\lambda_i^{(1)} \leq \lambda_i^{(2)}, \theta^{(1)} \geq \theta^{(2)}$ et $B_i^{(1)}(x) \leq_s B_i^{(2)}(x)$, où s est *st* ou bien *v*, alors les inégalités $\{\pi_{mn}^{(1)}\} \leq_s \{\pi_{mn}^{(2)}\}$ ont lieu.

Preuve. D'après théorème 2, les inégalités $\lambda_i^{(1)} \leq \lambda_i^{(2)}, \theta^{(1)} \geq \theta^{(2)}$ et $B_i^{(1)}(x) \leq_s B_i^{(2)}(x)$, où s est *st* ou bien *v*, impliquent que $T^{(1)} \leq_s T^{(2)}$, i.e. pour n'importe quelle distribution p nous avons l'inégalité suivante

$$T^{(1)} p \leq_s T^{(2)} p. \quad (5.11)$$

Du théorème 1, l'opérateur $T^{(2)}$ est monotone, i.e. pour n'importe quelles deux distributions $p_1^{(2)}, p_2^{(2)}$ telles que $p_1^{(2)} \leq_s p_2^{(2)}$ nous avons

$$T^{(2)} p_1^{(2)} \leq_s T^{(2)} p_2^{(2)}. \quad (5.12)$$

En mettant $p = p^{(1)}$ dans l'inégalité (5.11). On trouve

$$T^{(1)} p^{(1)} \leq_s T^{(2)} p^{(1)}. \quad (5.13)$$

Là, existe une probabilité $p_1^{(2)}$ telle que l'inégalité suivante a lieu

$$T^{(1)} p^{(1)} \leq_s T^{(2)} p_1^{(2)}. \quad (5.14)$$

De l'équation (5.12) et (5.14), on peut tirer le résultat suivant

$$T^{(1)} p^{(1)} \leq_s T^{(2)} p^{(2)} \quad (5.15)$$

pour deux distributions quelconques $p^{(1)}$ et $p^{(2)}$.

Par conséquent, les inégalités (5.15) peuvent être réécrites de la manière suivante

$$\begin{aligned} T^{(1)} p^{(1)} &= P(X_d^{(1)} = (0, m, n)) = P(X_d^{(1)} = (1, m, n)) \\ &\leq_s P(X_d^{(2)} = (0, m, n)) = P(X_d^{(2)} = (1, m, n)) = T^{(2)} p^{(2)}. \end{aligned}$$

Finalement, quand $d \rightarrow +\infty$, nous avons $\{\pi_{mn}^{(1)}\} \leq_s \{\pi_{mn}^{(2)}\}$.

Fin de preuve

A l'aide du théorème 3, il est possible de trouver des bornes stochastiques insensibles pour la distribution stationnaire du nombre de clients dans le système considéré.

Théorème 4 *Si dans le modèle $M_1, M_2 / G_1, G_2 / 1$ avec rappels et priorité absolue de politique preemptive resume, la distribution du temps de service $B_l(x)$, $l = 0, 1$ est NBUE (New Better than Used in Expectation), alors la distribution stationnaire du nombre de clients dans le système est inférieure par rapport à l'ordre convexe \leq_v à la distribution stationnaire du nombre de clients dans le système $M_1, M_2 / M_1, M_2 / 1$ avec rappels et priorité absolue de politique preemptive resume.*

Preuve. Soient $M^{(1)}$ et $M^{(2)}$ deux systèmes d'attente avec rappels et priorité absolue de politique preemptive resume, où les clients prioritaires rejoignent l'orbite après le blocage et les clients non prioritaires (dans la même situation) quittent le système sans service. De plus, le client non prioritaire dont le service a été interrompu persiste dans la zone du serveur jusqu'à l'achèvement du service du client prioritaire pour reprendre son service de nouveau d'où il a été interrompu. Les systèmes $M^{(1)}$ et $M^{(2)}$ sont définis par les paramètres $\lambda_l^{(1)} = \lambda_l$, $\theta^{(1)} = \theta$ et $\lambda_l^{(2)} = \lambda_l$, $\theta^{(2)} = \theta$ respectivement.

De plus, $B_l^{(1)}(x) = B_l(x)$ avec $\beta_{l,1}^{(1)} = \beta_{l,1}$ et $B_l^{(2)}(x) = B_l^*(x)$

$$\left(\text{où } B_l^*(x) = \begin{cases} 1 - e^{-\left(\frac{1}{\beta_{l,1}}\right)x} & , x \geq 0, \\ 0 & , x < 0. \end{cases} \right) \text{ avec } \beta_{l,1}^{(2)} = \beta_{l,1} .$$

Si $B_l(x)$ est *NBUE*, alors $B_l(x) \leq_v B_l^*(x)$ [40]. Ainsi, les conditions de théorème 3 sont satisfaites : $\lambda_l^{(1)} \leq \lambda_l^{(2)}$, $\theta^{(1)} \geq \theta^{(2)}$ et $B_l^{(1)}(x) = B_l(x) \leq_v B_l^{(2)}(x) = B_l^*(x)$. Donc la distribution stationnaire du nombre de clients dans le système $M^{(1)}$ est inférieure par rapport à l'ordre convexe \leq_v à la distribution correspondante dans le système $M^{(2)}$.

Fin de preuve

5.4. Applications numériques

Il s'ensuit (du théorème 4) que le nombre moyen de clients prioritaires dans $M^{(1)}$ est inférieure par rapport à l'ordre convexe \leq_v à celui dans le système $M^{(2)}$ (ce dernier présente une borne supérieure pour le nombre moyen de clients prioritaires dans $M^{(1)}$).

Supposons que le temps de service des deux types de clients suit une distribution exponentielle de paramètre $\mu_l^{(2)} = \frac{1}{\beta_{l,1}^{(2)}} = \mu_l = \frac{1}{\beta_{l,1}} > 0$, ($l = 1, 2$):

De [11], nous avons la formule suivante pour le nombre moyen de clients prioritaires dans $M^{(2)}$ avec les paramètres $\lambda_1^{(2)} = \lambda_1$, $\theta^{(2)} = \theta$:

$$\frac{(\mu_2^{(2)} + \lambda_2^{(2)})(\lambda_1^{(2)} + \theta^{(2)})(\mu_1^{(2)}\lambda_1^{(2)} + 1 + 2\lambda_1^{(2)})}{2\mu_2^{(2)}\theta^{(2)}\left(1 - \frac{\lambda_1^{(2)}}{\mu_1^{(2)}}\right)(1 + \lambda_1^{(2)} + \lambda_2^{(2)})\lambda_2^{(2)}(1 + \mu_1^{(2)})^2}$$

Dans ce qui suit, nous supposons toujours que le temps moyen de service des deux types de clients égal à une unité de temps et que le taux d'arrivée des clients non prioritaires est 0,3. En premier lieu, nous examinons le comportement de la borne par rapport au taux d'arrivée des clients primaires λ_1 . La figure ci-dessous montre que l'augmentation de λ_1 entraîne l'augmentation de la valeur de la borne (la valeur du taux de rappels θ a été fixée égale à 10).

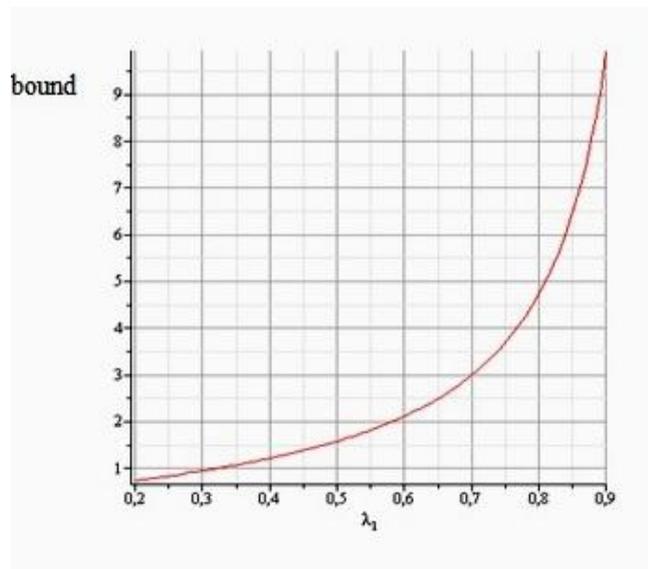


Figure1 : Le comportement de la borne par rapport à λ_1

Maintenant, nous voulons étudier les effets du taux de rappels des clients prioritaires θ sur la performance de la borne. Supposons que le taux d'arrivée de clients prioritaires $\lambda_1 = 0.8$. D'après les résultats numériques montrés dans la figure ci-dessous, on peut voir que l'augmentation du taux de rappels θ aboutit à une diminution de la valeur numérique de la borne.

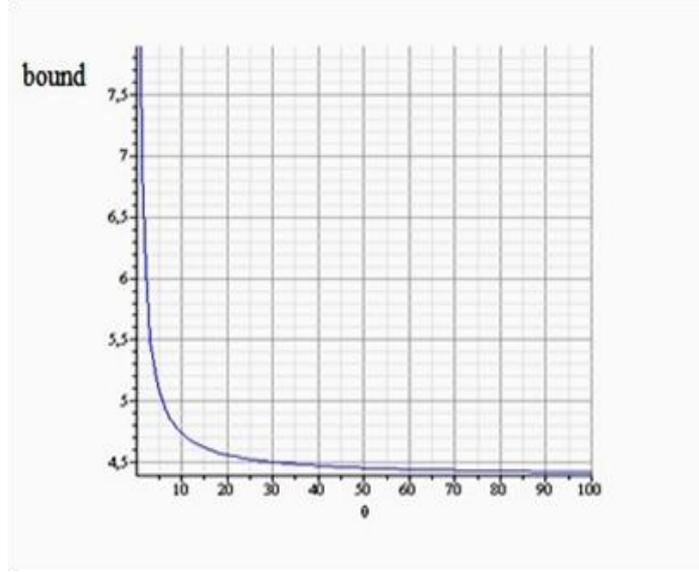


Figure2 : Le comportement de la borne par rapport à θ .

Encore une fois, à l'aide des résultats obtenus dans [11], nous pouvons déduire l'expression du nombre moyen de clients prioritaires dans $M^{(1)}$ avec les paramètres $\lambda_l^{(1)} = \lambda_l$, $\theta^{(1)} = \theta$, et ceci dans le cas où le temps de service des deux types de clients suit une distribution d' Erlang à deux étapes (E_2) avec $\mu_l^{(1)} = \frac{2}{\beta_{l,1}^{(1)}} = \mu_l = \frac{2}{\beta_{l,1}} > 0$, ($l = 1, 2$). En effet, la mesure en question se présente de la manière suivante :

$$\frac{\left((\lambda_1^{(1)} + \theta^{(1)}) \rho_1^{(1)} (\mu_1^{(1)})^2 + (3 + \rho_1^{(1)}) \theta^{(1)} + \rho_1^{(1)} \mu_1^{(1)} (4\lambda_1^{(1)} + 4\theta^{(1)}) + 4\lambda_1^{(1)} \right) (\mu_2^{(1)} + 2\lambda_2^{(1)})}{2\lambda_2^{(1)} \mu_2^{(1)} \theta^{(1)} (1 - \rho_1^{(1)}) (1 + \lambda_1^{(1)} + \lambda_2^{(1)}) (\mu_1^{(1)} + 2)^2}.$$

De la troisième figure ci-dessous, pour $\lambda_1 = 0.8$ et les valeurs différentes de θ , nous pouvons voir que le nombre moyen de clients prioritaires dans $M^{(1)}$ est inférieure par rapport à l'ordre convexe \leq_v au nombre moyen de clients prioritaires dans le système $M^{(2)}$.

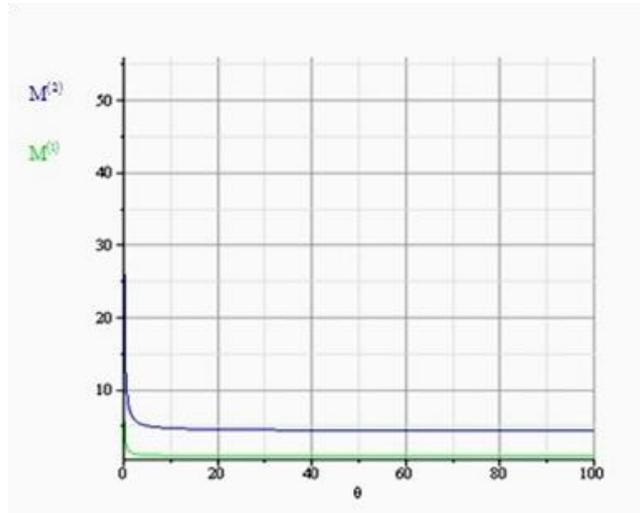


Figure3 : La comparaison entre le nombre moyen de clients prioritaires dans $M^{(1)}$ et dans $M^{(2)}$

5.5. Conclusion

Dans ce chapitre, nous avons obtenu les conditions sous lesquelles la monotonie interne et externe de l'opérateur de transition de la chaîne de Markov induite du modèle (considéré dans le chapitre 4) a lieu. En outre, nous avons prouvé que dans le cas où la distribution des temps de service est *NBUE*, la distribution stationnaire du nombre de clients dans le système est majorée par la distribution stationnaire du nombre de clients dans le système $M_1, M_2 / M_1, M_2 / 1$ avec rappels et priorité absolue de politique preemptive resume.

Conclusion Générale

Dans ce travail, nous nous sommes intéressés aux modèles d'attente avec rappels et plusieurs types de clients où certains jouissent d'une priorité de service.

Dans un premier temps, nous avons passé en revue certains résultats consacrés aux systèmes de files d'attente de type $M/G/1$ avec rappels ainsi que sur les approches utilisées par les chercheurs dans leurs investigations, notamment sur une approche basée sur la comparaison stochastique. Une attention particulière a été accordée aux modèles d'attente à interruptions de service : les modèles avec pannes aléatoires, les modèles avec vacances et les modèles avec priorité. Nous avons également présenté quelques situations réelles qui peuvent être étudiées à l'aide des systèmes de files d'attente avec rappels où la possibilité d'interruption du service est incluse.

Dans un deuxième temps, nous avons réalisé une analyse mathématique du système de files d'attente $M_1, M_2/G_1, G_2/1$ avec rappels et priorité absolue de politique « preemptive resume ». A cet effet, la chaîne de Markov induite est décrite, la condition d'ergodicité est établie à l'aide du critère de Foster et condition de Kaplan, la distribution stationnaire de l'état du système est trouvée en appliquant la méthode des variables supplémentaires et les mesures de performance sont déduites. En outre, nous avons étudié les propriétés de monotonie de la chaîne de Markov induite et, par la suite, obtenu les inégalités stochastiques pour la distribution stationnaire du nombre de clients dans le système. Nos investigations théoriques sont appuyées par des illustrations numériques.

Les travaux réalisés durant cette thèse et les résultats obtenus ouvrent un ensemble de perspectives. Pour nos futurs travaux, nous pensons orienter nos recherches vers les directions suivantes :

- Obtention de bornes stochastiques pour d'autres mesures de performance (par exemple, pour la durée moyenne de la période d'activité ou pour le temps moyen d'attente) ainsi que l'étude de l'influence de différents paramètres définissant le système donné sur la qualité des bornes.
- Application des résultats obtenus pour évaluer les performances des systèmes réels (par exemple, bureau d'un consultant financier dans une banque, système e-mail dans l'Internet).

- Extension du modèle étudié au cas où les clients prioritaires peuvent avoir la possibilité de quitter le système sans recevoir leur service, suivie par son analyse stochastique.

Bibliographie

- [1] A. AISSANI, *On the M/G/1 queueing system with repeated orders and unreliable server*, Journal of Technology, 6 (1988), pp. 98-123.
- [2] J. R. ARTALEJO, *Retrial queues with a finite number of sources*, Journal of the Korean Mathematical Society, 35(3) (1998), pp. 503-525.
- [3] J. R. ARTALEJO, *Retrial queues with negative arrivals*, Proceedings of the International Conference on Stochastic Processes, Cochin, pp (1996), pp. 159-167.
- [4] J. R. ARTALEJO and A. GOMEZ-CORRAL, *Performance analysis of a single-server queue with repeated attempts*, Mathematical and Computer Modelling, 30 (1999), pp. 79-88.
- [5] J. R. ARTALEJO and A. GOMEZ-CORRAL, *Unreliable retrial queues due to service interruptions arising from facsimile network*, Belgian Journal of Operations Research Statics and Computer Sciences, V.1, 38 (1) (1998), pp. 31-41.
- [6] I. ATENCIA and P. MORENO, *A single-server retrial queue with general retrial times and Bernoulli schedule*, Applied Mathematics and Computation 162 (2005), pp. 855–880.
- [7] 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 (1999), pp. 89-98.
- [8] N. K. BOOTS and H. C. TIJMS, *An M/M/c queue with impatient customers*, TOP, 7 (1999), pp. 213-220.
- [9] M. BOUALEM, N. DJELLAB and D. AISSANI, *Stochastic inequalities for M/G/1 retrial queues with vacations and constant retrial policy*, Mathematical and Computer Modelling, 50 (2009), pp. 207-212.
- [10] L. BOUTARFA and N. DJELLAB, *On the monotonicity properties of the embedded Markov chain associated to M1,M2/G1,G2/1 retrial queue with pre-emptive resume policy*, RAIRO (2014), pp. under reviews.
- [11] L. BOUTARFA and N. DJELLAB, *On the performance of the M1,M2/G1,G2/1 retrial queue with pre-emptive resume policy*, Yugoslav Journal of Operations Research, 25(1) (2015), pp. 153-164.

- [12] L. BOUTARFA and N. DJELLAB, *On the single server retrial queue with two types of breakdowns*, Campus journal: special issue, 2 (2008), pp. 221-225.
- [13] B. D. CHOI and Y. CHANG, *Single server retrial queues with priority calls*, Mathematical and Computer Modelling, 30 (1999), pp. 7-32.
- [14] B. D. CHOI and Y. CHANG, *Single server retrial queues with priority calls*, Mathematical and Computer Modelling, 30 (1999), pp. 7-32.
- [15] Q. H. CHOO and B. CONOLLY, *New results in the theory of repeated orders queueing systems*, Journal of Applied Probability, 16 (1979), pp. 631-640.
- [16] E. CINLAR, *Introduction to Stochastic Processes*, New York, 1975.
- [17] B. T. DOSHI, *Queueing systems with vacations- A survey -*, Queueing Systems, 1 (1986), pp. 29-66.
- [18] A. DUDIN and V. KLIMENOK, *Queueing system BMAP/G/1 with repeated calls*, Mathematical and Computer Modelling, 30 (1999), pp. 115-128.
- [19] G. I. FALIN, *On heavily loaded systems with repeated calls*, Sov. J. Comput. Syst. Sci, 24 (4) (1986).
- [20] G. I. FALIN, *Single-line repeated orders queueing systems*, Mathematische Operations Forschung und Statistik, Optimization, 5 (1986), pp. 649-667.
- [21] G. I. FALIN and J. R. ARTALEJO, *A finite source retrial queue*, European Journal of Operational Research, 108 (1998), pp. 409-424.
- [22] G. I. FALIN and C. FRICKER, *On the virtual waiting time in an M/G/1 retrial queue*, Journal of Applied Probability, 28 (1991), pp. 446-460.
- [23] G. I. FALIN and J. G. TEMPLETON, *Retrial Queues*, 1997.
- [24] G. I. FALIN., *A single-line system with secondary orders*, Eng. Cybernet. Rev, 17 (2) (1979), pp. 76-83.
- [25] S. W. FURHMANN and R. B.COOPER, *Stochastic decompositions in the M/G/1 queue with generalized vacations*, Oper Res, 33 (1985), pp. 1117-1129.
- [26] D. P. GAVER, *A waiting line with interrupted service including priorities*, J. Roy. Stat. Soc, B24 (1962), pp. 73-90.

- [27] A. GÓMEZ-CORRAL and M. L. GARCÍA, *Control strategies for a stochastic model of host–parasite interaction in a seasonal environment*, *Journal of Theoretical Biology*, 354 (2014), pp. 1-11.
- [28] J. L. HOLLEY, *Waiting line subject to priorities*, *Operations Research*, 2 (1954), pp. 341-343.
- [29] J.-C. KE and F.-M. CHANG, *Modified vacation policy for M/G/1 retrial queue with balking and feedback*, *Computers & Industrial Engineering*, 57 (2009), pp. 433–443.
- [30] J. KEILSON, J. COZZOLINO and H. YOUNG, *A service system with unfilled requests repeated*, *Oper. Res*, 16 (1968), pp. 1126-1137.
- [31] Z. KHALIL and G. FALIN, *Stochastic inequalities for M/G/1 retrial queues*, *Operations Research Letters*, 16 (1994), pp. 285-290.
- [32] V. G. KULKARNI and B. D. CHOI, *Retrial queue with server subject to breakdowns and re-pairs*, *Queueing Systems*, 7 (1990), pp. 191-208.
- [33] B. K. KUMAR, A. VIJAYAKUMAR and D. ARIVUDAINAMBI, *An M/G/1 Retrial Queueing System with Two-Phase Service and Preemptive Resume*, *Annals of Operations Research* 113 (2002), pp. 61–79.
- [34] H. LI and T. YANG, *A single-server retrial queue with server vacations and a finite number of input sources* *European Journal of Operational Research*, 85 (1995), pp. 149-160.
- [35] A. G. PAKES, *Some conditions for ergodicity and recurrence of Markov chains*, *Operations Research*, 17 (1969), pp. 1058-1061.
- [36] S. M. ROSS, *Stochastic Processes*, Willey, New York, 1983.
- [37] A. RUEGG, *Processus Stochastiques. Presses polytechniques romandes*, Lausannes, 1989.
- [38] L. I. SENOTT, P. A. HUMBLET and R. I. TWEEDIE, *Mean drifts and non ergodicity of Markov chains*, *Operations Research*, 31 (1983), pp. 783-789.
- [39] M. SHAKED and J. G. SHANTHIKUMAR, *Stochastic Orders and Their Applications*, AcademicPress, 1994.
- [40] D. STOYAN, *Comparison Methods for Queues and Other Stochastic Models*, New York, 1983.

- [41] R. SZEKL, *Stochastic Ordering and Dependence in Applied Probability*, Springer-Verlag, 1995.
- [42] H. TEDIJANITO, *Stochastic comparisons in vacation models* Commun Statistic-Stochastic Models, 7 (1991), pp. 125-135.
- [43] J. WANG and P. ZHANG, *A discrete-time retrial queue with negative customers and unreliable server*, Computer and Industrial Engineering 57 (2009), pp. 433–443.

Résumé

Dans ce travail, nous nous sommes intéressés à la modélisation de l'interruption de service due à une arrivée jouissante d'une priorité absolue dans les systèmes de files d'attente avec rappels.

Dans un premier temps, nous avons passé en revue certains résultats consacrés aux systèmes de files d'attente avec rappels ainsi que les approches utilisées par les chercheurs dans leurs investigations. Nous avons présenté les modèles d'attente à interruptions de service: les modèles avec pannes aléatoires, les modèles avec vacances et les modèles prioritaires, et ceci pour les cas où le phénomène de rappels est présent.

Dans un deuxième temps, nous avons présenté quelques systèmes réels qui peuvent être modélisés par les systèmes de files d'attente avec rappels et serveur sujet à des interruptions.

Finalement, nous avons réalisé l'analyse stochastique du système de files d'attente M/G/1 avec rappels et priorité absolue. Notre analyse inclut : description de la chaîne de Markov induite, condition d'ergodicité, distribution stationnaire de l'état du système, obtention de mesures de performance, étude des propriétés de monotonie de la chaîne de Markov induite. Des inégalités stochastiques pour la distribution stationnaire du nombre de clients dans le système considéré sont également introduites.

Mots-clés: file d'attente avec rappels, interruptions de service, client prioritaire, la politique de preemptive resume, chaîne de Markov induite, condition d'ergodicité, distribution stationnaire, mesures de performance, propriétés de monotonie, inégalités stochastiques.

Abstract

In this work, we are interested in the modeling of the service interruption due to the absolute priority arrivals in queuing systems with retrials.

First, we have reviewed some results devoted to the queuing systems with retrials and approaches used by researchers in their investigations. We have presented the waiting models with service interruptions: models with random failures, models with vacations and priority models, and this in cases where the retrial phenomenon is present.

Secondly, we presented some real systems which can be modeled by queuing system with retrials and server subject to interruptions.

Finally, we have performed a stochastic analysis of the $M / G / 1$ retrial queuing system with priority. Our analysis includes: description of the embedded Markov chain, ergodicity condition, stationary distribution of the system state, obtaining some performance measures, study of monotonicity properties of the embedded Markov chain. Stochastic inequalities for the stationary distribution of the number of customers in the system are also considered.

Keywords: retrial queue, service interruptions ,priority customer ,pre-emptive resume policy, embedded Markov chain , ergodicity condition, steady-state distribution, performance measure, monotonicity property, Stochastic inequalities .

تلخيص

في هذا العمل إهتمامنا كان منصبا على نمذجة إنقطاع الخدمة الذي يكون سببه قدوم زبون يتمتع بأولوية مطلقة في نظم طوابير الإنتظار التي لها خاصية إعادة التنبيه بطلب الخدمة.

أولا، قمنا بمراجعة بعض النتائج المكرسة لأنظمة الطابور مع إعادة التنبيه بطلب الخدمة والمناهج المستخدمة من قبل الباحثين في دراساتهم. قدمنا نماذج لنظم طوابير الإنتظار مع إنقطاع الخدمة: نماذج مع أعطال عشوائية، ونماذج مع العطلات ونماذج ذات الأولوية، وذلك في الحالات التي تكون فيها ظاهرة إعادة التنبيه بطلب الخدمة موجودة.

ثانيا عرضنا بغض النظم الموجودة على أرض الواقع و التي يمكن تمثيلها بنظم طوابير بخاصية إعادة التنبيه بطلب الخدمة وتملك وحدة خدمة تتعرض للإنقطاعات.

في الأخير أجرينا تحليل العشوائية لنظام $M / G / 1$ بخاصية إعادة التنبيه بطلب الخدمة والأولوية المطلقة. تحليل العشوائية الذي أجريناه يتضمن مايلي: وصف سلسلة ماركوف المضمنة، إعطاء شرط إحتمال حدوث الخدمة مرة أخرى، التوزيع الثابت لحالة النظام، الحصول على بعض مقاييس الأداء و دراسة خصائص الرتبة لسلسلة ماركوف المضمنة. المتراجحات العشوائية تم أيضا تحديدها.

كلمات البحث: نظم طوابير الإنتظار التي لها خاصية إعادة التنبيه بطلب الخدمة، إنقطاع الخدمة، زبون يتمتع بأولوية، سياسة مواصلة الخدمة، سلسلة ماركوف المضمنة، شرط إحتمال حدوث الخدمة مرة أخرى، التوزيع الثابت لحالة النظام، مقاييس الأداء، خصائص الرتبة، المتراجحات العشوائية.