



Année : 2012

Faculté des Sciences
Département de Mathématiques

MEMOIRE

Présenté en vue de l'obtention du diplôme de : **MAGISTER**

Thème

***SUR LE COMPORTEMENT ASYMPTOTIQUE
DU SYSTEME DE FILES D'ATTENTE M/G/1
AVEC RAPPELS***

OPTION

SYSTEMES DYNAMIQUES ET CALCUL STOCHASTIQUE

Par :

Doghmane Samia

DIRECTEUR DE MEMOIRE: Dr. N. Djellab M.C Université d'Annaba

DATE DE SOUTENANCE: 28 juin 2012

DEVANT LE JURY

PRESIDENT: Mr. H. Boutabia Prof U.B.M.ANNABA

EXAMINATEURS:

Mme. N. Saddik Ameer Prof U.B.M.ANNABA

Mme. A. Chadli M. C U.B.M.ANNABA

Dédicace

Je dédie ce modeste travail à mes parents qui m'ont aidé pour arriver à ce que je suis aujourd'hui.

A mon très cher époux ***AHMED*** pour tout son aide, son soutien et ses encouragements.

A mes enfants ***WASSIM*** et ***ANIS*** à qui je souhaite une longue vie pleine de bonheur et réussite.

A mes frères et sœurs à qui je souhaite une réussite dans la vie.

Remerciements

Je remercie vivement mon encadreur Madame N. Djellab, docteur de l'université d'Annaba pour m'avoir proposé le sujet de ce mémoire, pour sa patience et ses conseils qui m'ont été d'un grand apport, et encouragement rendant toute la durée de l'élaboration de ce travail.

Je tiens également à remercier:

* Le président de jury, d'avoir eu l'amabilité d'accepter la présidence du jury de soutenance.

* Les membres de jury pour leur participation et leur dévouement.

Résumé

Dans ce travail, nous nous sommes intéressés à l'étude du comportement asymptotique du système de files d'attente M/G/1 avec rappels.

Dans un premier temps, nous avons effectué une étude bibliographique sur les systèmes de files d'attente classiques, et avons présenté les modèles les plus connus. Nous nous sommes ensuite présentés les modèles d'attente avec rappels, en particulier le système M/G/1 avec rappels. Puis, nous avons passé à l'étude sa propriété de décomposition stochastique.

Dans un deuxième temps, nous avons étudié le comportement asymptotique du système de files d'attente M/G/1 avec rappels, où nous avons considéré des propriétés asymptotiques pour étudier le caractère de la distribution du nombre de clients en orbite.

Mots clés : Processus de Poisson, Système avec rappels, Trafic chargé, Intensité des rappels.

Summary

In this work, we add ourselves concerned persons to the survey of the asymptotic behavior of the $M/G/1$ retrial queue.

In a first time, we did a bibliographic survey on the systems classic, and presented the most known models. We add ourselves presented the retrial queues, in particular the $M/G/1$ retrial queue. Then, we studied its stochastic decomposition property.

In a second time, we studied the asymptotic behavior of the $M/G/1$ retrial queue, where we considered asymptotic properties to study the character of the distribution of the number of customers in orbit.

Key words: Poisson process, repeated calls, heavy traffic, low rate of retrials.

ملخص

لقد ركزنا في هذا العمل لدراسة السلوك المقارب للأجهزة بخطوط الانتظام $G/M/1$ مع الإعادة.

في البداية قمنا بدراسة مرجعية للنظم الكلاسيكية الأكثر شيوعا، بعدها قدمنا النماذج مع الإعادة و بصفة خاصة النظام $G/M/1$ مع التكرار. بعد ذلك قدمنا دراسة لخاصية التحليل للأنظمة.

في ثاني مرحلة قمنا بدراسة السلوك المقارب للنظام $G/M/1$ مع الإعادة، حيث استخدمنا الخصائص المقاربة لدراسة ميزة توزيع عدد الزبائن في المدار.

Table des matières

Introduction	4
1 Notions préliminaires	6
1.1 Introduction.....	6
1.2 Classification des systèmes de files d'attente.....	6
1.3 Processus stochastique.....	7
1.4 Processus de naissance et de mort	8
1.5 Analyse mathématique.....	9
1.6 Système de files d'attente M/M/c.....	10
1.7 Cas particuliers du système M/M/c.....	11
1.7.1 Système M/M/1.....	11
1.7.2 Système M/M/ ∞	12
1.8 Système de files d'attente M/G/1.....	12
1.9 Système de files d'attente G/G/1.....	16
1.10 Conclusion.....	18
2 Systèmes de files d'attente avec rappels	19
2.1 Introduction.....	19
2.2 Exemples.....	20
2.2.1 Réseaux de communication par paquet.....	20
2.2.2 CSMA non-persistant.....	20
2.2.3 Réseaux locaux avec évitement de collision.....	21
2.2.4 Système mémoire sur les disques magnétiques.....	21
2.2.5 Réseaux locaux en anneau.....	21
2.3 Modèle général.....	22
2.4 Modèles markoviens.....	23
2.4.1 Modèle M/M/1 avec rappels.....	23
2.4.2 Modèle M/M/2 avec rappels.....	25

2.5	Système de files d'attente M/G/1 avec rappels.....	26
2.5.1	Description du modèle.....	26
2.5.2	Chaîne de Markov induite.....	26
2.5.3	Distribution stationnaire de l'état du système.....	29
2.5.4	Période d'activité.....	32
2.5.5	Temps d'attente.....	32
2.5.6	Processus de départ.....	33
2.5.7	Mesures de performance.....	34
2.6	Conclusion.....	35
3	Propriété de décomposition stochastique des systèmes de files d'attente avec rappels	36
3.1	Introduction.....	36
3.2	Propriété de décomposition stochastique des modèles avec vacances.....	36
3.2.1	Cas d'un service exhaustif.....	38
3.2.1.1	Système de files d'attente M/G/1 avec vacation unique.....	38
3.2.1.2	Système de files d'attente M/G/1 avec vacances multiples.....	41
3.2.2	Cas d'un service non-exhaustif.....	43
3.3	Propriété de décomposition stochastique des modèles avec rappels.....	44
3.3.1	Système de files d'attente M/G/1 avec rappels.....	44
3.3.2	Autres modèles.....	46
3.4	Applications de la propriété de décomposition stochastique.....	50
3.4.1	Formules explicites des moments du nombre de clients en orbite.....	50
3.4.2	Taux de convergence vers le système M/G/1 classique.....	52
3.5	Conclusion.....	53
4	comportement asymptotique du système de files d'attente M/G/1 avec rappels	
4.1	Introduction.....	54
4.2	Trafic chargé.....	54

4.3 Taux faible des rappels.....	56
4.4 Taux fort des rappels.....	58
4.5 Conclusion.....	61
5 Application	62
5.1 Introduction.....	62
5.2 Transformées de Laplace-Stieltjes pour les systèmes de files d'attente classique.....	62
5.2.1 Système M/M/1.....	62
5.2.2 Système M/E ₂ /1.....	63
5.2.3 Système M/H ₂ /1.....	64
5.3 Transformées de Laplace-Stieltjes pour les systèmes de files d'attente avec rappels.....	64
5.3.1 Système M/M/1 avec rappels.....	64
5.3.2 Système M/E ₂ /1 avec rappels.....	65
5.3.3 Système M/H ₂ /1 avec rappels.....	65
5.4 Comportement asymptotique des systèmes avec rappels.....	66
5.4.1 Système M/M/1 avec rappels.....	66
5.4.2 Système M/E ₂ /1 avec rappels.....	66
5.4.3 Système M/H ₂ /1 avec rappels.....	67
Conclusion générale	71
Annexes	72
Annexe A.....	73
Annexe B.....	75
Bibliographie	77

Introduction

Les évolutions récentes de l'environnement technologique dues à l'accroissement des besoins liés à l'industrialisation font des modèles de files d'attente l'un des outils indispensables dans la modélisation mathématique. En effet, une grande classe de modèles mathématiques provient de la théorie de files d'attente dont les bases présentées dans les ouvrages de L.Kleinrock ([KLE75] et [KLE76]) s'ouvrent à de nombreuses applications. Un modèle typique de files d'attente nécessite la définition des processus d'inter arrivées et de durée de service des clients, la taille de la file qui peut être finie ou non, ainsi que la discipline de service. Tous ces paramètres sont indiqués dans la notation dite de Kendall. Depuis la première étude probabiliste réalisée en 1917 par l'ingénieur Danois Erlang pour modéliser le central téléphonique de Copenhague, la modélisation et l'évaluation des performances des systèmes complexes par les modèles de files d'attente n'ont cessé d'être un domaine de recherche d'un intérêt croissant. Quoique bon nombre de formules *magnifiques* furent élaborées et suggérées comme étant des solutions analytiques de certains types de problèmes, les travaux effectués dans le domaine des files d'attente restent jusque là limité en raison de l'inexploitabilité pratique de la plupart des résultats connus. La théorie est assez développée pour l'étude d'une file simple. En revanche, il n'existe que très peu de résultats dans la littérature, pour le cas où le nombre de files dépasse l'unité. Or c'est ce dernier cas qui permet de modéliser l'essentiel de beaucoup de systèmes réels.

Les modèles d'attente développés ces dernières décennies tentent de prendre en considération le phénomène de répétition de demande de service. Ce phénomène affecte les indices de performance des systèmes réels. Une nouvelle classe de modèles de files d'attente avec rappels, où un client arrivant dans le système et trouvant l'espace de service occupé tente de nouveau son service après une durée du temps aléatoire, est donc apparue. Pour identifier un système de files d'attente avec rappels, il faut ajouter une nouvelle spécification concernant le processus de répétition de demandes de service.

Habituellement, lors de l'étude des problèmes de la théorie des modèles des files d'attente avec rappels, on peut toujours trouver la distribution de la taille de l'orbite. Cependant dans les applications, on rencontre souvent des cas où le calcul de cette distribution devient difficile. C'est pourquoi nous nous intéressons aux propriétés asymptotiques de la variable aléatoire représentant le nombre de clients en orbite.

A cet effet, la distribution de la taille de l'orbite est obtenue dans les cas de taux faible et fort des rappels ainsi que dans le cas d'un régime chargé.

Ce mémoire est organisé en cinq chapitres. Dans le premier, nous présentons quelques rappels sur les systèmes de files d'attente classiques. Nous passons en revue quelques systèmes les plus connus.

Le deuxième chapitre comprend une synthèse sur les systèmes de files d'attente avec rappels, en particulier le modèle $M/G/1$ avec rappels.

Dans le troisième chapitre, nous présentons un aperçu sur la propriété de décomposition stochastique du système de files d'attente $M/G/1$ avec rappels.

Le quatrième chapitre concerne l'étude du comportement asymptotique du système $M/G/1$ avec rappels. Cette étude doit être appliquée dans le cas d'un système faiblement chargé, moyennement chargé ou fortement chargé.

Le dernier chapitre est consacré à l'application des résultats théoriques obtenus sur les systèmes $M/M/1$, $M/E_2/1$ et $M/H_2/1$ avec rappels.

Chapitre 1

Notions préliminaires

1.1 Introduction

Les files d'attente peuvent être considérées comme un phénomène caractéristique de la vie quotidienne. Ce phénomène se manifeste dans les domaines d'activités les plus divers : les guichets de la poste, les stations téléphoniques, la technique militaire,...etc.

Le modèle général d'un phénomène d'attente peut être résumé comme suit : des clients arrivent suivant un processus quelconque à un intervalle du temps aléatoire pour acquérir un service auprès d'un serveur. A l'arrivée d'un client, si un dispositif de service (serveur) est libre, il se dirige immédiatement vers ce dispositif où il est servi. Dans le cas contraire, le client prend place dans une file d'attente, sinon il quitte le système. La durée du service auprès de chaque serveur est aussi aléatoire.

1.2 Classification des systèmes de files d'attente

La classification des systèmes des files d'attente se base principalement sur trois éléments suivants : la nature stochastique du processus des arrivées, le mécanisme de service et la discipline d'attente.

Le processus des arrivées spécifie les instants auxquels les clients arrivent dans le système. Ce processus peut être régulier: c'est -à-dire à chaque unité du temps, on a une arrivée, ou aléatoire (processus de Poisson).

Le mécanisme de service comprend le nombre de serveurs et la distribution des durées de service. Dans la majorité des cas, on suppose que la population des clients est homogène, c'est-à-dire que les temps de service nécessaires au traitement des clients sont identiquement distribués selon une loi de probabilité commune. En particulier, on rencontre la distribution exponentielle qui est la plus simple à manipuler mathématiquement. Sa propriété la plus importante est l'absence de mémoire, qui pourrait être caractérisée par le fait que le temps résiduel d'un service est indépendant du temps déjà écoulé de ce dernier.

La capacité de l'espace d'attente peut être illimitée ou non. Dans le second cas, certains clients qui arrivent vers le système n'ont pas la possibilité d'y entrer [39].

Pour la classification des systèmes d'attente, on a recours à *la notation symbolique* induite par Kendall au début des années cinquante. Cette notation comprend quatre symboles rangés dans l'ordre $A/B/s/N$,

où

A : distribution des temps entre deux arrivées successives,

B : distribution des durées de service,

s : nombre de postes de service montés en parallèle,

N : capacité du système.

On peut toutefois faire abstraction du dernier symbole lorsque $N = \infty$. Pour spécifier les distributions A et B , les symboles suivants sont utilisés :

M : distribution exponentielle,

E_k : distribution d'Erlang d'ordre k ,

H_k : distribution hyperexponentielle d'ordre k ,

G : distribution générale.

Au sein de chaque file, le prochain client à servir est sélectionné sur la base d'une règle prédéterminée appelée discipline de service dont les plus courantes sont *FIFO*,

LIFO,...etc.

1.3 Processus stochastique

Un processus stochastique $\{X(t)\}_{t \in T}$ est une fonction du temps dont la valeur à chaque instant dépend de l'issue d'une expérience aléatoire. A chaque instant $t \in T$,

$X(t)$ est donc une variable aléatoire. Un processus stochastique peut être considéré comme une famille de variables généralement non indépendantes. L'ensemble des temps T peut être discret ou continu. $X(t)$ définit l'état du processus à un instant donné t . L'ensemble noté S des valeurs que peut prendre le processus à chaque instant est appelé *espace d'états* et peut, de même que T , être discret (fini ou infini) ou continu. En fonction des valeurs possibles de T et S , on classe les processus stochastiques de la façon suivante :

- processus à temps discret et à espace d'état discret.
- processus à temps continu et à espace d'état discret.
- processus à temps discret et à espace d'état continu.
- processus à temps continu et à espace d'état continu.

1.4 Processus de naissance et de mort

Ces processus permettent de façon générale de décrire l'évolution temporelle de la taille d'une population d'un type donné. Dans le cas d'un système d'attente, on considère par exemple des populations comprenant tous les clients qui sont dans le système à l'instant t .

Les processus de naissance et de mort sont des processus stochastiques à temps continu et à espace d'états discret $n = 0, 1, 2, \dots$. Ils sont caractérisés par deux conditions importantes : ils sont sans mémoire, et à partir d'un état donné n , des transitions ne sont possibles que vers l'un ou l'autre des états voisins $(n+1)$ et $(n-1)$ pour $n \geq 1$.

Alors, soit $\{N(t), t \geq 0\}$ un processus de naissance et de mort à états discrets et homogène dans le temps, c'est-à-dire :

$$p(N(t+s) = j / N(s) = i) = p_{ij}(t)$$

ne dépend pas de s . Ce processus est de naissance et de mort si :

$$p_{i,i+1}(\Delta t) = \lambda_i \Delta t + o(\Delta t), \quad i \geq 0;$$

$$p_{i,i-1}(\Delta t) = \mu_i \Delta t + o(\Delta t), \quad i \geq 1;$$

$$p_{i,i}(\Delta t) = 1 - (\lambda_i + \mu_i)\Delta t + o(\Delta t), \quad i \geq 0;$$

$$p_{i,j}(\Delta t) = o(\Delta t), \quad |i-j| \geq 2;$$

$$p_{i,j}(0) = \delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}.$$

λ_i et μ_i sont appelés taux de transition (taux de naissance et de mort).

1.5 Analyse mathématique

L'étude mathématique d'un système de files d'attente se fait par l'introduction d'un processus stochastique qui décrit l'évolution temporelle du système. En fonction des quantités qui déterminent la structure du système, on cherche à calculer les probabilités d'état définissant le régime transitoire du processus en question, puis le régime stationnaire. Les probabilités d'état définissant le régime transitoire dépendent de la distribution initiale du processus étudié. Ainsi, le calcul explicite du régime transitoire s'avère pénible, voir impossible, pour la plupart des modèles.

A partir de la distribution stationnaire, on peut calculer les caractéristiques du système, telles que : le temps d'attente d'un client, le temps de séjour d'un client dans le système, le taux d'occupation des dispositifs de service, la durée de la période d'activité, et également les mesures de performance suivantes:

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

Ces mesures permettent de juger le comportement opérationnel d'un système d'attente. Elles ne sont pas indépendantes les unes des autres, mais sont liées par les relations suivantes :

$$\bar{n} = \lambda_e \bar{W}_s ;$$

$$\bar{n}_f = \lambda_e \bar{W} ;$$

$$\bar{W}_s = \bar{W} + \frac{1}{\mu} ;$$

$$\bar{n} = \bar{n}_f + \frac{\lambda_e}{\mu} ;$$

où λ_e et $\frac{1}{\mu}$ sont respectivement le taux d'arrivées des clients et la durée moyenne de service ($\mu > 0$). Une autre mesure importante d'un système de files d'attente, celle qui mesure le degré de saturation du système, est l'intensité du trafic ρ . Elle est définie par :

$$\rho = \frac{\text{temps moyen de service}}{\text{temps moyen entre deux arrivées successives}} .$$

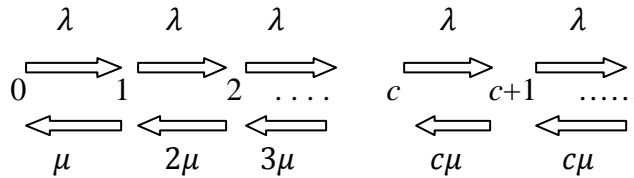
1.6 Système de files d'attente M/M/c

Les clients arrivent au système selon un processus de Poisson de taux $\lambda > 0$, la loi de service est une loi exponentielle de taux $\mu > 0$, le service est assuré par $c \geq 1$ serveurs en parallèle, la file d'attente est de capacité infini. Si l'un des serveurs est libre, le client qui arrive se dirige immédiatement vers ce serveur. Dans le cas contraire, le client prend sa place dans une file d'attente commune pour tous les serveurs. Lorsqu'un serveur se libère, le client en tête de la file occupe ce serveur. Par conséquent, la discipline d'attente est *FIFO*. L'état du système à la date t peut être décrit par le processus stochastique

$$\{N(t), t \geq 0\}. \quad (1.1)$$

où $N(t)$ représente le nombre de clients dans le système à l'instant t . Ce dernier est un processus de naissance et de mort dont les taux de transitions sont [51],[39] :

$$\begin{cases} \lambda_n = \lambda ; & n = 0, 1, 2, \dots \\ \mu_n = \min \{n, c\} \mu ; & n = 1, 2, \dots \end{cases}$$



Evaluation de l'état dans la file d'attente M/M/c.

Du diagramme, on déduit les résultats qui suivent. L'analyse du système en régime stationnaire, à l'aide de la procédure des équations de Chapman-Kolmogorov aboutit aux équations suivantes :

$$\lambda p_0 = \mu p_1 ;$$

$$(\lambda + n\mu)p_n = \lambda p_{n-1} + (n+1)\mu p_{n+1} \quad 1 \leq n < c ;$$

$$(\lambda + c\mu)p_n = \lambda p_{n-1} + c\mu p_{n+1} \quad n \geq c ;$$

avec

$$\sum_{n=0}^{\infty} p_n = 1.$$

La résolution du système ci-dessus présente la distribution stationnaire suivante :

$$p_n = \frac{\rho^n}{n!} p_0, \quad 0 \leq n \leq c; \quad (1.2)$$

$$p_n = \frac{\rho^c}{c!} (A)^{n-c} p_0, \quad n \geq c, \quad (1.3)$$

où

$$p_0 = \left[\sum_{n=0}^{c-1} \frac{\rho^n}{n!} + \frac{\rho^c}{c!} \sum_{n=c}^{\infty} \rho^{n-c} \right]^{-1}, \quad \rho = \frac{\lambda}{\mu} \text{ et } A = \frac{\lambda}{c\mu}.$$

Cette dernière existe si $\lambda < c\mu$.

A partir de la distribution stationnaire du processus (1.1), on peut calculer les caractéristiques du système. En effet,

$$\bar{n} = \rho + \frac{\rho^{c+1}}{c \cdot c! (1-A)^2} p_0;$$

$$\bar{n}_f = \frac{\rho^{c+1}}{c \cdot c! (1-A)^2} p_0;$$

$$\bar{W} = \frac{c\mu\rho^c}{c!(c\mu-\lambda)^2} p_0;$$

$$\bar{W}_s = \frac{1}{\mu} + \frac{\rho^c}{\mu \cdot c \cdot c! (1-A)^2} p_0.$$

1.7 Cas particuliers du système M/M/c

1.7.1 Système M/M/1

Le modèle M/M/1 permet d'illustrer les concepts fondamentaux liés à l'attente devant un serveur. Il est décrit par : les clients arrivent dans le système selon un processus de Poisson de taux $\lambda > 0$, les durées de service suivent une loi exponentielle de paramètre $\mu > 0$, la discipline d'attente est *FIFO*, la file d'attente est de capacité infinie. Dans ce cas le processus (1.1) est de naissance et de mort dont les taux de transition sont : $\lambda_n = \lambda, n \geq 0$ et $\mu_n = \mu, n \geq 1$.

Pour trouver la distribution stationnaire et les mesures de performance de ce système, on remplace c par 1 dans les formules du système M/M/c; on obtient :

$$\text{La distribution stationnaire : } p_n = (1 - \rho)\rho^n. \quad (1.4)$$

Les mesures de performance :

$$\bar{n} = \frac{\rho}{1-\rho} ;$$

$$\bar{n}_f = \frac{\lambda^2}{\mu(\mu-\lambda)} ;$$

$$\bar{W} = \frac{\bar{n}_f}{\lambda} ;$$

$$\bar{W}_s = \frac{\bar{n}}{\lambda} = \frac{1}{\mu-\lambda} .$$

1.7.2 Système M/M/ ∞

Dans le cas $c = \infty$, nous obtenons le système de files d'attente M/M/ ∞ . Il est évident qu'aucune file d'attente ne se forme : chaque client est servi dès son arrivée. La distribution stationnaire est donnée par :

$$p_n = \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} e^{-\frac{\lambda}{\mu}}, \quad n \geq 1.$$

Donc, on a les mesures de performance suivantes:

$$\bar{n} = \rho ;$$

$$\bar{W}_s = \frac{1}{\mu} ; \text{ tandis que } \bar{n}_f = 0 \text{ et } \bar{W} = 0.$$

1.8 Système de files d'attente M/G/1

Pour décrire l'état d'un système de type M/G/1 à la date t , il faut connaître non seulement le nombre de clients $N(t)$ qui se trouvent dans le système à la date t , mais également le temps de service déjà écoulé $R(t)$ du client qui est en train d'être servi. On peut alors montrer que le processus bidimensionnel $\{N(t), R(t), t \geq 0\}$ est à nouveau de type markovien. Cependant, le calcul de son régime transitoire ferait intervenir des équations aux dérivées partielles. Par conséquent, on choisit une autre méthode qui ramène l'étude du processus non markovien $\{N(t), t \geq 0\}$ [35] à celle d'une chaîne de Markov à temps discret associée dont elle permet de calculer le régime stationnaire.

Dans ce système, on a affaire à des arrivées poissonniennes de taux $\lambda > 0$. Le service est assuré par un seul serveur. Les durées de service sont des variables aléatoires Se positives, mutuellement indépendantes et distribuées selon une loi générale de fonction de répartition $B(t)$, de transformée de Laplace-Stieltjes $\tilde{B}(s)$, $R(s) > 0$, de moyenne finie $E[Se] = \frac{1}{\mu}$ et de $E[Se^2]$. Les durées entre deux arrivées consécutives et

les durées de service sont également mutuellement indépendantes. La discipline d'attente est *FIFO*.

Soit $\{N(t), t \geq 0\}$. Montrons que $\{N(t)\}$ ne définit pas une chaîne de Markov. Soient t_d et t_f les dates de début et de fin d'un service, t_a l'instant d'arrivée d'un nouveau client. Si $t_d < t_a < t_f$, la probabilité qu'un départ s'effectue dans l'intervalle $[t_a, t_a + \Delta t]$ ne dépend pas seulement de Δt , mais de la date t_d à laquelle le service en cours a commencé. Comme le temps résiduel du service ($t_f - t_a$) dépend du passé, alors le chaîne $\{N(t)\}$ n'est pas markovienne. Par conséquent, on utilise la méthode de la chaîne de Markov induite. A cet effet, on considère $N(t)$ aux instants $\xi_1, \xi_2, \dots, \xi_n, \dots$ où les clients terminent leur service et quittent le système. On définit ainsi un processus stochastique à temps discret

$$(1.5) \quad \{N_n = N(\xi_n), \quad n \geq 1\}.$$

Pour vérifier que cette suite de variables aléatoires est une chaîne de Markov à temps discret, on considère le nombre A_n de clients qui entrent dans le système pendant que le $n^{\text{ième}}$ client est servi. Les variables A_n sont indépendantes entre elles, leur distribution commune est donnée par :

$$P(A_n = k) = a_k = \int_0^\infty e^{-\lambda t} \frac{(\lambda t)^k}{k!} dB(t), \text{ où } a_k > 0 \text{ et } k > 0.$$

Alors,

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

L'équation fondamentale de la chaîne vaut donc :

$$(1.6) \quad N_{n+1} = N_n - \delta_n + A_{n+1},$$

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

N_{n+1} ne dépend que de N_n et de A_{n+1} et non pas des valeurs prises par N_{n-1}, N_{n-2}, \dots . La suite $\{N_n, n \geq 1\}$ est une chaîne de Markov induite du processus $\{N(t), t \geq 0\}$. Ses probabilités de transition $p_{ij} = P(N_{n+1} = j / N_n = i)$ se calcule par :

$$\begin{cases} p_{0j} = a_j & \text{pour } j \geq 0 ; \\ p_{ij} = a_{j-i+1} & \text{pour } 1 \leq i \leq j+1; \end{cases}$$

$$p_{ij} = 0 \quad \text{ailleurs.}$$

La matrice des transitions est

$$P = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & \dots & \dots \\ a_0 & a_1 & a_2 & a_3 & \dots & \dots \\ 0 & a_0 & a_1 & a_2 & \dots & \dots \\ 0 & 0 & a_0 & a_1 & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

Vu qu'on peut passer de chaque état vers n'importe quel autre état, il s'agit d'une chaîne de Markov irréductible. De plus, la matrice n'est pas décomposable (est apériodique). La chaîne de Markov est donc ergodique. La distribution stationnaire de la chaîne existe si $\rho = \frac{\lambda}{\mu} < 1$.

Pour les variables aléatoires A_n , nous disposons de quelques résultats importants :

$$E[A_n] = \lambda E[S_e] = \frac{\lambda}{\mu} = \rho ;$$

la fonction génératrice

$$\begin{aligned} A(z) &= \sum_{k=0}^{\infty} a_k z^k = \sum_{k=0}^{\infty} z^k \int_0^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} dB(t) = \int_0^{\infty} e^{-\lambda t} \left(\sum_{k=0}^{\infty} \frac{(\lambda t z)^k}{k!} \right) dB(t) \\ &= \int_0^{\infty} e^{-\lambda t} e^{\lambda t z} dB(t) = \int_0^{\infty} e^{-(\lambda - \lambda z)t} dB(t). \end{aligned}$$

Soit $\tilde{B}(s) = \int_0^{\infty} e^{-st} dB(t)$. Alors, $A(z) = \tilde{B}(\lambda - \lambda z)$.

Supposons que $\rho < 1$. Le système se trouve dans un régime stationnaire.

Soit $\Pi = [\pi_0, \pi_1, \dots]$ la distribution stationnaire de la chaîne de Markov induite ($\pi_j = \lim_{n \rightarrow \infty} P(N(\xi_n) = j)$).

Par conséquent,

$$\Pi = \Pi \cdot P, \text{ où } \pi_j = \sum_{i=0}^{\infty} \pi_i p_{ij}, \quad j \geq 0 ;$$

$$\pi_j = a_j \pi_0 + \sum_{i=1}^{j+1} a_{j-i+1} \pi_i = a_j \pi_0 + \sum_{i=0}^{j+1} a_{j-i+1} \pi_i - a_{j+1} \pi_0, \quad j \geq 0.$$

A présent, on applique la méthode des fonctions génératrices. En effet,

$$\sum_{j=0}^{\infty} \pi_j z^j = \pi_0 \sum_{j=0}^{\infty} a_j z^j + \frac{1}{z} \sum_{j=0}^{\infty} c_{j+1} z^{j+1} - \frac{\pi_0}{z} \sum_{j=0}^{\infty} a_{j+1} z^{j+1},$$

où

$$c_{j+1} = \sum_{i=0}^{j+1} a_{j-i+1} \pi_i.$$

En introduisant les fonctions génératrices suivantes :

$$\Pi(z) = \sum_{i=0}^{\infty} \pi_i z^i ; A(z) = \sum_{i=0}^{\infty} a_i z^i ; C(z) = \sum_{j=0}^{\infty} c_j z^j = \Pi(z)A(z), \text{ nous}$$

Obtenons:

$$\Pi(z) = \pi_0 A(z) + \frac{1}{z} [C(z) - c_0] - \frac{\pi_0}{z} [A(z) - a_0].$$

D'où

$$\Pi(z) = \pi_0 \frac{A(z)(z-1)}{z-A(z)}. \quad (1.7)$$

De la relation (1.6), on trouve :

$$E[N_{n+1}] = E[N_n], E[A(z)] = E[\delta_n] = P(\delta_n > 0) = P(N_n > 0) = 1 - P(N_n = 0). \text{ D'où } \pi_0 = 1 - \rho.$$

Si $\rho < 1$, la fonction génératrice de la distribution stationnaire est donnée par :

$$\Pi(z) = (1 - \rho) \frac{A(z)(z-1)}{z-A(z)}, \text{ pour } |z| < 1 \text{ et } z \neq 0. \quad (1.8)$$

La formule (1.6) rend possible à calculer le nombre moyen de clients dans le système en régime stationnaire. En l'élevant au carré et en observant que $\delta_n^2 = \delta_n$ et $\delta_n N_n = N_n$, on obtient :

$$N_{n+1}^2 = N_n^2 + \delta_n + A_{n+1}^2 - 2N_n - 2\delta_n A_{n+1} + 2N_n A_{n+1}.$$

On a que : A_{n+1} est indépendante de N_n et de δ_n ; $E[N_{n+1}^2] = E[N_n^2]$; $E[A_n] = \rho = \frac{\lambda}{\mu}$.

Alors, $E[N_{n+1}^2] = E[N_n^2] + E[\delta_n] + E[A_{n+1}^2] - 2E[N_n] + 2E[A_{n+1}]E[N_n - \delta_n]$, ou bien

$$0 = \rho + E[A_{n+1}^2] - 2E[N_n] + 2\rho(E[N_n] - \rho).$$

D'où

$$E[N_n] = \frac{\rho + E[A_{n+1}^2] - 2\rho^2}{2(1 - \rho)}. \quad (1.9)$$

Pour trouver $E[A_{n+1}^2]$, considérons la régime stationnaire.

$$\begin{aligned} \lim_{n \rightarrow \infty} E[A_{n+1}^2] &= E[A^2] = \int_0^{\infty} E[A^2 / T=t] dB(t) = \lambda \int_0^{\infty} t dB(t) + \lambda^2 \int_0^{\infty} t^2 dB(t) \\ &= \frac{\lambda}{\mu} + \lambda^2 \left(\text{Var}[Se] + \left(\frac{1}{\mu}\right)^2 \right). \end{aligned}$$

Enfin, la formule (1.9) devient :

$$\lim_{n \rightarrow \infty} E[N_n] = E[N] = \rho + \frac{\rho^2 + \lambda^2 \text{Var}[Se]}{2(1 - \rho)}.$$

D'après les formules de Little, on déduit :

$$\bar{W}_s = \frac{1}{\mu} + \frac{\lambda(\text{Var}[Se] + 1/\mu^2)}{2(1-\rho)},$$

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

1.9 Système de files d'attente G/G/1

Le système G/G/1 est beaucoup plus complexe que ceux que nous avons considéré ci-dessus, et peu de résultats exacts sont connus à son sujet. De plus, ces résultats sont peu exploitables. Par contre, de nombreuses formules approximatives sont obtenues, notamment celle dérivant directement de l'équation de Pollaczek-Khintchine :

$\bar{W}_{G/G/1} \approx \left(\frac{(ca)^2 + (cs)^2}{2} \right) \bar{W}_{M/M/1}$, où ca et cs sont des coefficients de variation des lois de probabilité du temps entre deux arrivées successives et du temps de service, respectivement.

Supposons que le système est régi par la discipline *FIFO*, l'étude (exacte) du temps d'attente peut être faite à l'aide d'une relation classique : l'équation de Lindley. Afin d'établir cette relation, nous allons considérer une séquence d'arrivées de clients indexée par n et utiliser les notations suivantes : $C_{\bar{n}}$ le $n^{\text{ième}}$ client arrivé dans le système, $a_{\bar{n}}$ la date d'arrivées de C_n , $d_{\bar{n}}$ la date de départ de C_n , $t_{\bar{n}}$ l'intervalle de temps entre l'arrivée de C_{n-1} et C_n , $s_{\bar{n}}$ le temps de service de C_n , $w_{\bar{n}}$ le temps d'attente de C_n .

Le processus qui nous intéresse est $\{w_n, n \geq 0\}$. Afin d'établir qu'il s'agit d'un processus markovien. Remarquons premièrement que la date de départ du client C_{n+1} est

$$(1.10) \quad d_{n+1} = \begin{cases} d_n + s_{n+1}, & d_n \geq a_{n+1}; \\ a_{n+1} + s_{n+1}, & d_n \leq a_{n+1}. \end{cases}$$

La première alternative correspond au cas où C_{n+1} doit attendre le départ de C_n pour accéder au serveur alors que la seconde s'applique lorsque C_{n+1} arrive dans un système vide.

Par définition, le temps d'attente de C_{n+1} est $w_{n+1} = d_{n+1} - a_{n+1} - s_{n+1}$; ainsi, en soustrayant $(a_{n+1} + s_{n+1})$ de (1.10), il vient

$$w_{n+1} = \begin{cases} d_n - a_{n+1}, & d_n - a_{n+1} \geq 0; \\ 0, & d_n - a_{n+1} \leq 0. \end{cases}$$

En utilisant l'identité $w_n = d_n - a_n - s_n$, cette dernière expression s'écrit aussi,

$$w_{n+1} = \begin{cases} w_n + s_n - t_{n+1}, & w_n + s_n - t_{n+1} \geq 0; \\ 0, & w_n + s_n - t_{n+1} \leq 0. \end{cases}$$

L'introduction de la variable $u_n = s_n - t_{n+1}$ aboutit finalement à la relation fondamentale suivante :

$$w_{n+1} = \begin{cases} w_n + u_n, & w_n + u_n \geq 0; \\ 0, & \text{sinon.} \end{cases}$$

Alors, le processus $\{w_n, n \geq 0\}$ est markovien.

Les suites $\{t_n, n \geq 0\}$ et $\{s_n, n \geq 0\}$ étant indépendantes entre elles et formées de variables aléatoires indépendantes et identiquement distribuées, la suite $\{u_n, n \geq 0\}$ est également composée de variables aléatoires indépendantes et identiquement distribuées.

Soit $U_n(x)$ la fonction de répartition de u_n , nous étudions cette fonction en fonction de celle des intervalles de temps entre les arrivées successives de clients, notée $A(t)$, et de celle des temps de service, notée $B(t)$. Par définition, pour $x \in \mathbb{R}$,

$U_n(x) = P(u_n = s_n - t_{n+1} \leq x)$. On obtient :

$$\begin{aligned} U_n(x) &= P(u_n = s_n - t_{n+1} \leq x) = \int_{t=0}^{\infty} P(s_n \leq x + 1/t_{n+1}=t) dA(t) \\ &= \int_{t=0}^{\infty} B(x + t) dA(t). \end{aligned} \quad (1.11)$$

Il faut noter que le système est en régime stationnaire si et seulement si la variable u_n a une espérance négative car $E[u_n] = E[s_n - t_{n+1}] = E[s_n] - E[t_{n+1}] = E[Se] - E[A] = E[A](\rho - 1)$, où $E[Se]$ et $E[A]$ représentent les espérances des temps de service et des intervalles entre les arrivées successives de clients. Ainsi, l'espérance de u_n est négative si est seulement si $\rho = E[Se]/E[A] < 1$. Notant $W_n(y)$ la fonction de répartition du temps d'attente w_n de C_n , nous obtenons, pour $y \geq 0$ et grâce à l'équation (1.11),

$$\begin{aligned} W_n(y) &= P(w_n + u_n \leq y) = \int_0^{\infty} P(u_n \leq y - w/w_n = w) dW_n(w) \\ &= \int_0^{\infty} U(y - w) dW_n(w). \end{aligned}$$

Le système est en régime stationnaire, le processus $\{w_n, n \geq 0\}$ est ergodique et admet une distribution invariante $W(y)$ vérifiant l'équation fonctionnelle

$W(y) = \int_0^\infty U(y-w)dW(w)$, pour $y \geq 0$. Par ailleurs, le temps d'attente étant une grandeur non négative, il est clair que $W(y) = 0$ pour $y < 0$.

Pour un système de files d'attente G/G/1, en régime stationnaire, la distribution stationnaire du temps d'attente satisfait donc l'équation intégrale de Lindley

$$W(y) = \begin{cases} \int_0^\infty U(y-w)dW(w), & \text{pour } y \geq 0 ; \\ 0, & \text{pour } y < 0. \end{cases}$$

Cette équation de type Wiener-Hopf ne peut être résolue explicitement que dans de cas très simples mais se laisse fort bien traiter par les méthodes numériques.

1.10 Conclusion

Dans ce chapitre, nous avons présenté une synthèse sur les systèmes de files d'attente ordinaires. Ils sont utilisés pour modéliser et évaluer les performances de différents systèmes réels. Les modèles d'attente développés ces dernières décennies tentent de prendre en considération des phénomènes de répétition de demande de service. Il s'agit donc des systèmes de files d'attente avec rappels.

Chapitre 2

Systemes de files d'attente avec rappels

2.1 Introduction

Dans la théorie des files d'attente classique, il est supposé qu'un client qui ne peut pas obtenir son service immédiatement dès son arrivée, rejoint la file d'attente ou quitte le système définitivement. Les systèmes de file d'attente développés tentent de prendre en considération des phénomènes de répétition de demandes de service, et ceci après une durée du temps aléatoire. Un tel système est connu comme « système de files d'attente avec rappels ».

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

2.2 Exemples

2.2.1 Réseaux de communication par paquet

Considérons un réseau de communications d'ordinateurs dans lequel on trouve un ensemble d'interfaces IMPs (Interface Message Processors) reliées entre elles par des câbles. Un ordinateur principal est connecté à l'une de ces interfaces. Si l'ordinateur veut envoyer un message à un autre ordinateur principal, il doit en premier lieu envoyer le message avec l'adresse de destination à l'interface à laquelle il est connecté. L'interface à son tour envoie le message à l'ordinateur destinataire directement si elle y est connectée, ou indirectement via d'autres interfaces.

Considérons une interface à laquelle un ordinateur principal est connecté. Les messages arrivent de l'extérieur selon un processus aléatoire. Après la réception du message, l'ordinateur l'envoie immédiatement à son interface. S'il y a un tampon libre, le message est accepté. Dans le cas contraire, le message est rejeté et l'ordinateur doit réessayer une autre fois après une période de temps. S'il existe des tampons libres, le message rejeté sera stocké dans un tampon de l'ordinateur principal. Dans le cas contraire, le message est considéré comme perdu. On peut poser les questions suivantes:

- Quelles sont les probabilités pour qu'un message soit rejeté par l'interface et par l'ordinateur principal?
- Quel est le nombre moyen de messages dans le tampon de IMP?
- Quel est le nombre moyen de messages dans le tampon de l'ordinateur principal?
- Quel est le temps d'attente d'un message dans le tampon de l'ordinateur principal?

Le problème présenté peut être modélisé comme un système avec rappels à serveur unique (interface IMP) possédant des tampons (positions d'attente). Le nombre de tampons de l'ordinateur principal constitue la capacité de l'orbite.

2.2.2 CSMA non-persistant

Dans les réseaux locaux (LAN), l'un des protocoles de communication les plus utilisés est CSMA (Carrier-Sence Multiple Accés) non-persistant. Supposons qu'un réseau local est composé de n stations connectées par un seul bus. Les messages de longueurs variables arrivent aux stations du monde extérieur. En recevant le message, la station le découpe en un nombre fini de paquets de longueur fixe et consulte le bus pour voir s'il est occupé. Si le bus est libre, l'un des paquets est transmis via ce bus à la station de destination, et les autres paquets sont stockés dans le tampon pour une transmission ultérieure. Autrement, tous les paquets sont stockés dans le tampon et la

station peut consulter le bus après une certaine durée aléatoire. Les questions concernant ce problème sont:

- Quel est le temps moyen d'attente d'un paquet?
- Quel est le nombre moyen de messages (paquets) dans le tampon d'une station?

2.2.3 Réseaux locaux avec évitement de collision

Les stations au sol transmettent vers un satellite unique des paquets d'information, tous de même longueur. Le temps est découpé en segments. L'émission d'un paquet par une station ne peut commencer qu'au début de chaque segment (ces instants sont donnés par des tops de l'horloge du satellite). La durée de transmission d'un paquet est celle d'un segment. Il se peut que lors d'un top donné, plusieurs stations émettent simultanément des paquets. Les collisions entre les messages sont donc inévitables. Les collisions provoquent la destruction (perte) de l'information, et par conséquent la qualité de performance diminue. Les paquets détruits devront être réémis ultérieurement jusqu'à ce qu'ils soient transmis sans erreurs.

2.2.4 Système mémoire sur les disques magnétiques

Considérons un système mémoire où K unités disques partagent un contrôleur des disques, et transmettent l'information quand elles trouvent ce dernier libre. Les demandes insatisfaites sont répétées après une rotation du disque.

Ce système peut être présenté comme un système de files d'attente avec rappels. Le serveur est le contrôleur des disques. L'arrivée dans le système est quasi-aléatoire: le nombre de sources est fini et égal à K . La rotation du disque dans le cas de la répétition de demande peut être considérée comme un intervalle entre deux rappels consécutifs de durée constante.

2.2.5 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 sont: anneau à jeton (Token ring network) et anneau à "slots" (Cambridge ring network, Orwell) [Hopper et al., 1986]. Le premier a été modélisé comme un réseau fermé avec un serveur circulant [Watson, 1984]. Une procédure a été proposée par Mitrani (1987) pour calculer le temps de réponse. Le second a été décrit comme un serveur "processeur partagé" (au niveau de hardware), ainsi que comme un réseau de files d'attente fermé multiclassés (au niveau du protocole Bloc de Base). La particularité de ces modèles est que le taux de service d'un récepteur dépend du nombre de transmissions en cours (le nombre de "slots" peut être > 1) [Mitrani, 1987].

Cependant, ces systèmes peuvent être modélisés comme les systèmes de files d'attente avec rappels, vacances du serveur. L'orbite est constituée des tampons des stations attachées au réseau, le serveur est la permission (jeton, "slots"). Les vacances du serveur sont les passages de la permission libérée par une station vers une autre.

2.3 Modèle général

Les systèmes de files d'attente avec rappels ou avec répétition d'appels se caractérisent par le fait suivant: un client qui arrive dans le système et trouve tous les serveurs occupés, quitte le système définitivement, ou rappelle ultérieurement à des instants aléatoires. Un client qui attend pour rappeler est dit *en orbite* et devient *source d'appels secondaires*.

Un système de files d'attente avec rappels contient un espace de service composé de $c \geq 1$ dispositifs de service et d'un espace d'attente ayant $m-c$ ($m \geq c$) positions d'attente. A l'arrivée d'un client primaire, s'il y a un ou plusieurs serveurs libres, le client sera immédiatement pris en charge. Sinon, s'il y a une position d'attente libre, le client rejoint la file d'attente. Lorsque tous les serveurs et positions d'attente sont occupés, le client quitte le système, soit définitivement avec une probabilité $1-H_0$ soit temporairement avec une probabilité H_0 et rappelle ultérieurement, après un temps aléatoire. La capacité O de l'orbite peut-être finie ou infinie. Dans le cas où O est finie, et si l'orbite est pleine, le client quitte le système définitivement. Chaque client de l'orbite forme un processus d'arrivées secondaires de taux θ et il est traité de la même manière qu'un client primaire avec une probabilité H_k (s'il s'agit de la $k^{\text{ième}}$ tentative échouée). Le schéma général d'un système avec rappels est donné dans la figure suivante:

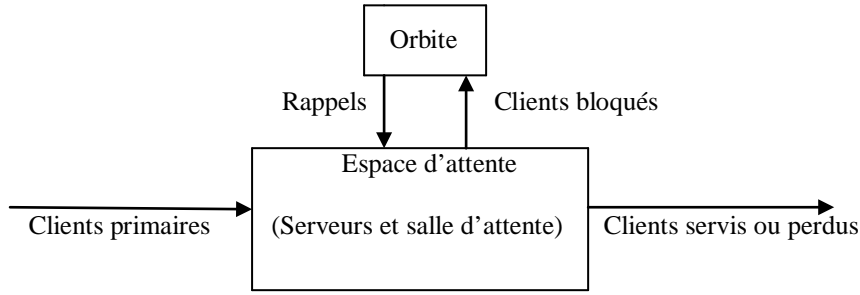


FIG.2.1- Schéma général d'un système avec rappels

D'après cette figure on peut déduire que le système de files d'attente avec rappels peut être vu comme un type spécial des réseaux de files d'attente. Dans sa forme la plus générale, ces réseaux contiennent des nœuds : un nœud spécial pour le service et l'attente, et un nœud de délai pour les rappels.

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

2.4 Modèles markoviens

2.4.1 Modèle M/M/1

On considère un système de files d'attente sans positions d'attente. Le service est assuré par un seul serveur. Les clients primaires arrivent selon un processus de Poisson de taux $\lambda > 0$. Les durées de service suivent une loi exponentielle de fonction de répartition $B(x) = 1 - e^{-\mu x}$, $x \geq 0$ et de moyenne finie $\frac{1}{\mu}$. Les temps entre deux rappels consécutifs sont également exponentiels de paramètre $\theta > 0$ (la fonction de répartition $T(x) = 1 - e^{-\theta x}$, $x \geq 0$). Nous admettons que les durées de service, les durées entre deux rappels consécutifs ainsi que entre deux arrivées primaires successives sont mutuellement indépendantes. L'état du système peut être décrit par le processus

$$\{C(t), N_o(t), t \geq 0\}, \quad (2.1)$$

Où $C(t)$ est égale à 0 ou 1 selon le fait que le serveur est libre ou non, $N_o(t)$ est le nombre de clients en orbite l'instant t . Supposons que le régime stationnaire existe ($\rho = \frac{\lambda}{\mu} < 1$). Le processus (2.1) est de Markov d'espace d'états $S = \{0,1\} \times \mathbb{N}$.

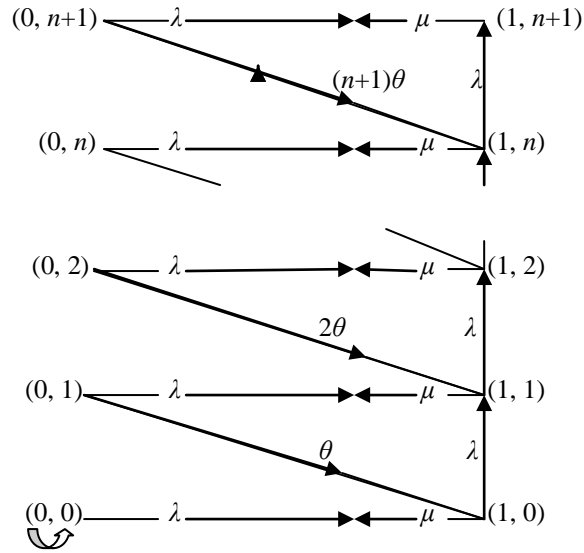


FIG. 2.2– La structure générale du modèle M/M/1 avec rappels.

Les équations d'équilibre statistique sont :

$$\begin{cases} (\lambda + n\theta) p_{0n} = \mu p_{1n}; & (2.2) \\ (\lambda + \mu)p_{1n} = \lambda p_{0n} + (n+1)\theta p_{0,n+1} + \lambda p_{1,n-1} . & (2.3) \end{cases}$$

Ici, $p_{in} = \lim_{t \rightarrow \infty} P(C(t) = i, N_o(t) = n)$, $i = 0, 1$ et $n \geq 0$, représentent la distribution stationnaire conjointe de l'état du serveur et du nombre de clients en orbite. Introduisons les fonctions génératrices suivantes :

$$P_0(z) = \sum_{n=0}^{\infty} z^n p_{0n};$$

$$P_1(z) = \sum_{n=0}^{\infty} z^n p_{1n}.$$

A l'aide de ses fonctions et à partir des équations (2.2) et (2.3), on obtient :

$$P_0(z) = (1 - \rho) \left(\frac{1 - \rho}{1 - z\rho} \right)^{\frac{\lambda}{\theta}}; \quad (2.4)$$

$$P_1(z) = \rho \left(\frac{1 - \rho}{1 - z\rho} \right)^{\frac{\lambda}{\theta} + 1}. \quad (2.5)$$

Les transformées inverses des (2.4) et (2.5) nous donnent les formules analytiques explicites [25] :

$$p_{0n} = \frac{\rho}{n! \theta^n} \prod_{k=0}^{n-1} (1 + k\theta) (1 - \rho)^{\frac{\lambda}{\theta} + 1};$$

$$p_{1n} = \frac{\rho^{n+1}}{n! \theta^n} \prod_{k=1}^n (\lambda + k\theta) (1 - \rho)^{\frac{\lambda}{\theta} + 1};$$

2.4.2 Modèle M/M/2 avec rappels

Nous considérons un système de files d'attente avec rappels où l'espace de service comprend $c = 2$ serveurs. Les clients primaires arrivent dans le système selon un processus de Poisson de taux $\lambda > 0$. Si un client primaire trouve au moins un serveur libre, il commence son service. Sinon, il entre en orbite. Les durées de service et les durées entre deux rappels consécutives sont exponentiellement distribuées de moyennes finies, respectivement $\frac{1}{\mu}$ et $\frac{1}{\theta}$. Nous supposons que toutes les variables aléatoires introduites sont mutuellement indépendantes.

L'état du système à la date t peut être décrit par le processus (2.1), dont l'espace d'états est $S = \{0, 1, 2\} \times \mathbb{N}$. Les probabilités d'état sont

$$p_{in} = P(C(t) = i, N_o(t) = n), \quad (i, n) \in S.$$

Les transitions possibles sont données dans la figure FIG.2.3.

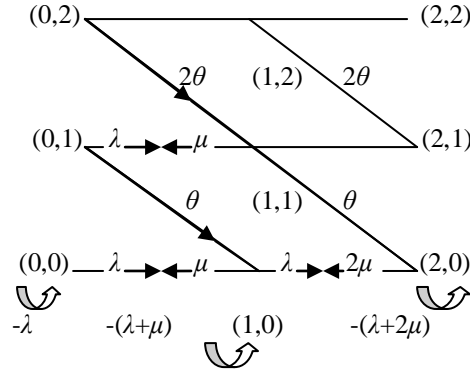


FIG 2.3 – La structure générale du modèle M/M/2 avec rappels

La condition d'existence d'un régime stationnaire est $\lambda < 2\mu$ [19].

Supposons que $\rho = \frac{\lambda}{2\mu} < 1$ et $\mu = 1$. A partir du graphe des transitions, il est possible d'obtenir les équations d'équilibre statistique, telles que:

$$\begin{cases} (\lambda + n\theta) p_{0n} = p_{1n}; & (2.6) \end{cases}$$

$$\begin{cases} (\lambda + 1 + n\theta) p_{1n} = \lambda p_{0n} + (n+1)\theta p_{0,n+1} + 2p_{2n}; & (2.7) \end{cases}$$

$$\begin{cases} (\lambda + 2) p_{2n} = \lambda p_{1n} + (n+1)\theta p_{1,n+1} + \lambda p_{2,n-1}; & (2.8) \end{cases}$$

Ainsi que l'équation de normalisation:

$$\sum_{n=0}^{\infty} p_{0n} + \sum_{n=0}^{\infty} p_{1n} + \sum_{n=0}^{\infty} p_{2n} = 1.$$

pour $p_{in} = \lim_{t \rightarrow \infty} p_{in}(t)$.

La résolution des équations (2.6)-(2.7)-(2.8) nous donne

$$p_{0n} = \frac{\lambda^n}{n!\theta^n} \prod_{k=0}^{n-1} \frac{(\lambda+k\theta)^2+k\theta}{2+3\lambda+2\theta+2k\theta} p_{00};$$

$$p_{1n} = (\lambda+n\theta) \frac{\lambda^n}{n!\theta^n} \prod_{k=0}^{n-1} \frac{(\lambda+k\theta)^2+k\theta}{2+3\lambda+2\theta+2k\theta} p_{00};$$

$$p_{2n} = [1+\lambda+(n+1)\theta] \frac{\lambda^n}{n!\theta^n} \prod_{k=0}^n \frac{(\lambda+k\theta)^2+k\theta}{2+3\lambda+2\theta+2k\theta} p_{00}.$$

2.5 Système de file d'attente M/G/1 avec rappels

2.5.1 Description de modèle

Les clients primaires arrivent dans ce type de système selon un processus homogène de Poisson de taux $\lambda > 0$. Le service est assuré par un seul serveur. Les durées de service suivent une loi générale de fonction de répartition $B(x)$, de transformée de Laplace-Stieltjes $\tilde{B}(s) = \int_0^\infty e^{-sx} dB(x)$, $Re(s) > 0$. Soient les moments d'ordre k , $\beta_k = (-1)^k \tilde{B}^{(k)}(0)$, l'intensité du trafic $\rho = \lambda\beta_1$ et $\mu = \frac{1}{\beta_1}$. La durée entre deux rappels successifs d'une même source secondaire est exponentiellement distribuée de paramètre $\theta > 0$: $T(x) = 1 - e^{-\theta x}$.

Le système évolue de la manière suivante: On suppose que le $(n-1)^{i\text{ème}}$ client termine son service à l'instant ξ_{n-1} (les clients sont numérotés dans l'ordre de service) et le serveur devient libre. Même s'il y a des clients dans le système, ils ne peuvent pas occuper le dispositif de service immédiatement à cause de leur ignorance de ce dernier. C'est pourquoi le $n^{i\text{ème}}$ client suivant n'entre en service qu'après un intervalle de temps R_n durant lequel le serveur est libre. A l'instant $\eta_n = \xi_{n-1} + R_n$ le $n^{i\text{ème}}$ client débute le service durant un temps t_s . Les clients primaires arrivent dans le système pendant ce temps deviennent sources de clients secondaires. Tous les clients qui arrivent durant ce temps de service n'influent pas sur le processus. A l'instant $\xi_n = \eta_n + t_s$ le $n^{i\text{ème}}$ client achève son service, le serveur devient libre et ainsi de suite.

2.5.2 Chaîne de Markov induite

Considérons le processus $\{C(t); N_o(t); t \geq 0\}$. En général ce processus n'est pas processus de Markov, mais il possède une chaîne de Markov induite. Cette chaîne a été décrite pour la première fois par Choo et Conolly (1979) [12]. On dénote par

$q_n = N_o(\xi_n)$ le nombre de clients en orbite après le $n^{i\text{ème}}$ départ (l'instant de départ du $n^{i\text{ème}}$ client). La suite des variables aléatoires q_n forme une chaîne de Markov induite, dont l'équation fondamentale est:

$$q_{n+1} = q_n - \delta_{q_n} + \nu_{n+1}.$$

La variable aléatoire v_{n+1} représente le nombre de clients primaires arrivant dans le système durant le service du $(n+1)^{ième}$ client. Elle ne dépend pas des événements qui se sont produits avant l'instant η_{n+1} du début du service du $(n+1)^{ième}$ client. Sa distribution est donnée par [39],[7]:

$P(v_n = i) = k_i = \int_0^\infty \exp(-\lambda x) \frac{(\lambda x)^i}{i!} dB(x)$; où $k_i > 0, i \geq 0$; avec les résultats suivants:

$$\text{si } v = \lim_{n \rightarrow \infty} v_n, E[v] = \rho;$$

$$K(z) = \sum_{i=0}^{\infty} k_i z^i = \tilde{B}(\lambda - \lambda z).$$

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

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

Elle dépend de q_n , et sa distribution est donnée par:

$$P(\delta_{q_n} = 1/q_n = k) = \frac{k\theta}{\lambda + k\theta}; \quad P(\delta_{q_n} = 0/q_n = k) = \frac{\lambda}{\lambda + k\theta}.$$

Les probabilités de transition à une étape de la chaîne sont :

$$r_{ij} = \frac{i\theta}{\lambda + i\theta} k_{j-i+1} + \frac{\lambda}{\lambda + i\theta} k_{j-i}.$$

Notons que $r_{ij} \neq 0$ uniquement pour $i = 0, 1, \dots, j + 1$. Pour démontrer la dernière formule, on utilise l'équation fondamentale de la chaîne:

$$\begin{aligned} r_{ij} &= P(q_{n+1} = j/q_n = i) = P(q_n - \delta_{q_n} + v_{n+1} = j/q_n = i) \\ &= P(v_{n+1} = j - i + \delta_{q_n} / q_n = i) \\ &= P(v_{n+1} = j - i/q_n = i, \delta_{q_n} = 0)P(\delta_{q_n} = 0/q_n = i) \\ &+ P(v_{n+1} = j - i + 1/q_n = i, \delta_{q_n} = 1)P(\delta_{q_n} = 1/q_n = i) \\ &= P(v_{n+1} = j - i) P(\delta_{q_n} = 0/q_n = i) \\ &+ P(v_{n+1} = j - i + 1) P(\delta_{q_n} = 1/q_n = i) \\ &= k_{j-i} \frac{\lambda}{\lambda + i\theta} + k_{j-i+1} \frac{i\theta}{\lambda + i\theta}. \end{aligned}$$

La condition d'existence du régime stationnaire peut être obtenue de façon suivante :

L'accroissement moyen de la chaîne vaut :

$$E[q_{n+1} - q_n / q_n = k] = E[v_{n+1}] - E[\delta_{q_n} / q_n = k] = \rho - \frac{k\theta}{\lambda + k\theta}.$$

Si $\rho < 1$, alors $E[q_{n+1} - q_n / q_n = k] \rightarrow \rho - 1 < 0$ lorsque $k \rightarrow \infty$. La chaîne est donc ergodique. Si par contre $\rho \geq 1$, alors $E[q_{n+1} - q_n / q_n = k] \geq \frac{\lambda}{\lambda + k\theta}$, et la chaîne induite n'est pas ergodique. Si $\rho > 1$, la chaîne $\{q_n\}$ est statistiquement supérieure à la chaîne induite du système de files d'attente M/G/1 classique, par conséquent, elle est transitoire.

Théorème 2.1 [17] *Soit $\rho < 1$. La distribution stationnaire de la chaîne de Markov induite possède la fonction génératrice suivante:*

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

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

$$\pi_n = \sum_{m=0}^n \pi_m k_{n-m} + \sum_{m=1}^{n+1} \pi_m k_{n-m+1}; n = 0, 1, \dots$$

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

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

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,

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

$$\psi'(z) \theta [K(z) - z] = \psi(z) \lambda [1 - K(z)].$$

Sous la condition que $\rho < 1$,

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

On obtient

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

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

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

2.5.3 Distribution stationnaire de l'état du système

L'état du système peut être décrit par le processus :

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

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

Notons,

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

Théorème 2.2 *Les fonctions génératrices de la distribution conjointe de l'état du serveur et de la taille de l'orbite en régime stationnaire ($\rho = \lambda\beta_1 < 1$) sont données par:*

$$P_0(z) = \sum_{n=0}^{\infty} z^n p_{0n} = (1-\rho) \exp \left\{ \frac{\lambda}{\theta} \int_1^z \frac{1-\tilde{B}(\lambda-\lambda u)}{\tilde{B}(\lambda-\lambda u)-u} d(u) \right\} ; \quad (2.9)$$

$$P_1(z) = \sum_{n=0}^{\infty} z^n p_{1n} = \frac{1-\tilde{B}(\lambda-\lambda z)}{\tilde{B}(\lambda-\lambda z)-z} P_0(z). \quad (2.10)$$

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

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

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

Le système d'équations de balance devient:

$$\lambda \sum_{n=0}^{\infty} z^n p_{0n} + \theta \sum_{n=0}^{\infty} z^n n p_{0n} = \int_0^{\infty} \sum_{n=0}^{\infty} z^n p_{1n}(x) b(x) dx;$$

$$\sum_{n=0}^{\infty} z^n p'_{1n}(x) = -(\lambda + b(x)) \sum_{n=0}^{\infty} z^n p_{1n}(x) + \lambda \sum_{n=0}^{\infty} z^n p_{1,n-1}(x);$$

$$\sum_{n=0}^{\infty} z^n p_{1n}(0) = \lambda \sum_{n=0}^{\infty} z^n p_{0n} + \theta \sum_{n=0}^{\infty} z^n (n+1) p_{0,n+1}.$$

Ou bien

$$(2.11) \quad \left\{ \begin{array}{l} \lambda P_0(z) + \theta z P'_0(z) = \int_0^{\infty} P_1(z, x) b(x) dx; \\ P'_1(z, x) = (\lambda z - \lambda - b(x)) P_1(z, x); \end{array} \right.$$

(2.12)

$$(2.13) \quad P_1(z, 0) = \lambda P_0(z) + \theta P'_0(z).$$

De l'équation (2.12), on a

$$(2.14) \quad P_1(z, x) = P_1(z, 0) [1 - B(x)] \exp(-(\lambda - \lambda z)x).$$

A l'aide de l'équation (2.14), l'équation (2.11) devient

$$(2.15) \quad \begin{aligned} \lambda P_0(z) + \theta z P'_0(z) &= \int_0^{\infty} P_1(z, 0) [1 - B(x)] \exp(-(\lambda - \lambda z)x) b(x) dx \\ &= P_1(z, 0) \int_0^{\infty} B'(x) \exp(-(\lambda - \lambda z)x) dx \\ &= P_1(z, 0) \int_0^{\infty} \exp(-(\lambda - \lambda z)x) dB(x) \\ &= P_1(z, 0) \tilde{B}(\lambda - \lambda z) = P_1(z, 0) K(z). \end{aligned}$$

A partir des équations (2.13) et (2.15), on a

$$P_1(z, 0) K(z) = \lambda P_0(z) + \theta z \left(\frac{P_1(z, 0)}{\theta} - \frac{\lambda}{\theta} P_0(z) \right);$$

$$P_1(z, 0) (K(z) - z) = (\lambda - \lambda z) P_0(z);$$

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

A l'aide de l'équation (2.14), on trouve

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

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

$$\begin{aligned} P_1(z) &= \int_0^\infty P_1(z, x) dx = \frac{\lambda - \lambda z}{K(z) - z} P_0(z) \int_0^\infty [1 - B(x)] \exp(-(\lambda - \lambda z)x) dx \\ &= \frac{\lambda - \lambda z}{K(z) - z} P_0(z) \frac{1 - \tilde{B}(\lambda - \lambda z)}{\lambda - \lambda z} = P_0(z) \frac{1 - K(z)}{K(z) - z}. \end{aligned}$$

Pour trouver $P_0(z)$, on utilise les équations (2.13) et (2.15):

$$\begin{aligned} \lambda P_0(z) + \theta z P_0'(z) &= K(z) [\lambda P_0(z) + \theta z P_0'(z)]; \\ \theta [K(z) - z] P_0'(z) &= \lambda [1 - K(z)] P_0(z). \end{aligned} \quad (2.16)$$

Considérons $f(z) = K(z) - z$. On a:

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

Par conséquent, $f'(1) = -\lambda \tilde{B}'(0) - 1 = \rho - 1 < 0$ et $f''(z) = \lambda^2 \tilde{B}''(\lambda - \lambda z) \geq 0$. La fonction $f(z)$ est décroissante sur $[0, 1]$ et positive:

pour $\rho < 1$ et $z \in [0, 1]$, $z < K(z) \leq 1$. Encore,

$$\lim_{z \rightarrow 1-0} \frac{1 - K(z)}{K(z) - z} = \frac{K'(1)}{1 - K'(1)} = \frac{\rho}{1 - \rho} < \infty. \text{ De ce fait, la fonction } \frac{1 - K(z)}{K(z) - z} \text{ peut être}$$

définie au point $z = 1$ comme $\frac{\rho}{1 - \rho}$. Donc

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

La résolution de cette équation nous donne l'équation (2.9). Encore, à partir de

l'équation (2.10) nous obtenons $P_1(1) = P_0(1) \frac{\rho}{1 - \rho}$. De plus, vu que

$$P_0(1) + P_1(1) = 1, \text{ on trouve } P_1(1) = \rho \text{ et } P_0(1) = 1 - \rho.$$

Conséquences

1. La distribution marginale du nombre de serveurs occupés s'exprime de la manière suivante:

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

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

$$P(z) = P_0(z) + P_1(z) = \frac{1-z}{\tilde{B}(\lambda - \lambda z) - z} P_0(z). \quad (2.17)$$

3. La fonction génératrice du nombre de clients dans le système se présente comme:

$$Q(z) = P_0(z) + z P_1(z) = \frac{(1-\rho)(1-z)\tilde{B}(\lambda - \lambda z)}{\tilde{B}(\lambda - \lambda z) - z} \exp \left\{ \frac{1}{\theta} \int_1^z \frac{1 - \tilde{B}(\lambda - \lambda u)}{\tilde{B}(\lambda - \lambda u) - u} du \right\}.$$

2.5.4 Période d'activité

La période d'activité d'un système de files d'attente avec rappels est définie comme un intervalle de temps qui commence à l'arrivée d'un client dans le système vide et se termine à l'instant de départ quand le système redevient vide. Falin [19] a obtenu la transformée de Laplace-Stieljes de la durée L de la période et du nombre de clients I servis pendant L :

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

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

En outre,

- a) Si $\rho > 1$, alors $p[L = \infty] = p[I = \infty] > 0$;
- b) Si $\rho = 1$, alors $E[L] = E[I] = \infty$;
- c) Si $\rho < 1$, alors

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

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

2.5.5 Temps d'attente

Le temps d'attente d'un client peut être mesuré par le temps physique et 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. Un effort important a été fait par Falin et Fricker (1991) [23], pour obtenir l'expression de la transformée de Laplace- Stieljes 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 $f(t)$ est définie par :

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

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

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

Dans les systèmes de files d'attente avec rappels, on s'intéresse aussi à une autre mesure qui est 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 :

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

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

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

2.5.6 Processus de départ

Ce processus est défini par une suite $\{\xi_1, \xi_2, \dots\}$ d'instants de départ des clients après le service, ou une suite $\{T_i = \xi_i - \xi_{i-1}\}$ d'intervalles inter-départs. Une particularité importante du processus de départ d'un système avec rappels est que ce processus n'est jamais un processus de renouvellement, à l'exception d'un service instantané [16]. Pourtant, les conditions d'approximation par un processus de renouvellement sont définies [17] : Soit $\rho < 1$. Etant donné un entier k , tel que : $i_1 < i_2 < \dots < i_k$;

$i_2 - i_1, \dots, i_k - i_{k-1} \rightarrow \infty$, alors indépendamment de l'état initial on a :

$p(T_{i_1} < x_1, \dots, T_{i_k} < x_k) \rightarrow F(x_1) \dots F(x_k)$, où $F(x_k) = B(x) (1 - e^{-\lambda x} \varphi(e^{-\theta x}))$ est la fonction de répartition des intervalles inter-départs.

Pour le système M/G/1 avec rappels, la transformée de Laplace- Stieltjes et la moyenne des intervalles inter-départs peuvent être calculées à l'aide des formules suivantes [20] :

$$E[e^{-sT_i}] = \tilde{B}(s) \left[1 - \frac{s}{\theta} (1-\rho) \int_0^1 u^{\frac{s+1}{\theta}} \tilde{B}(\lambda - \lambda u) \frac{1-u}{\tilde{B}(\lambda - \lambda u) - u} \right. \\ \left. \times \exp \left\{ \frac{\lambda}{\theta} \int_1^u \frac{1 - \tilde{B}(\lambda - \lambda v)}{\tilde{B}(\lambda - \lambda v) - v} dv \right\} du \right]; \\ E[T_i] = \frac{1}{\lambda}.$$

2.5.7 Mesures de performance

Le système M/G/1 peut être caractérisés par les mesures de performance, telles que :

- nombre moyen de clients dans le système

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

- nombre moyen de clients en orbite

$$\bar{n}_0 = P'(1) = \frac{\lambda^2 \beta_2}{2(1-\rho)} + \frac{\lambda \rho}{\theta(1-\rho)};$$

- temps moyen d'attente d'un client

$$\bar{W} = \frac{\bar{n}_0}{\lambda} = \frac{\lambda\beta_2}{2(1-\rho)} + \frac{\lambda\beta_1}{\theta(1-\rho)}$$

- nombre moyen de rappels par client

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

2.6 Conclusion

En raison des difficultés d'analyse des systèmes de files d'attente avec rappels et la complexité des résultats analytiques obtenus, plusieurs chercheurs ont tenté de développer des méthodes approximatives d'analyse des phénomènes ou d'appliquer la propriété de décomposition stochastique (si le modèle possède la propriété en question).

Chapitre 3

Propriété de décomposition stochastique des systèmes de files d'attente avec rappels

3.1 Introduction

Parmi les approches développées ces dernières années permettant d'analyser les systèmes de files d'attente avec rappels, on trouve celle basée sur la propriété de décomposition stochastique que peut posséder un modèle. Elle offre les avantages de simplification de résolution de modèles complexes. Le concept général de la propriété de décomposition stochastique d'un système d'attente M/G/1 est défini de la manière suivante : le nombre de clients se trouvant dans le système à une date aléatoire est distribué comme la somme de deux variables aléatoires indépendantes ou plus ; l'une de ces variables représente le nombre de clients se trouvant dans le système M/G/1 ordinaire à une date aléatoire (le serveur est toujours disponible). Les systèmes évoqués sont en régime stationnaire. Ce type de décomposition stochastique a été observé auparavant pour les systèmes d'attente avec vacances. Ces derniers sont caractérisés par le fait que le temps inoccupé du serveur peut être utilisé pour des tâches extérieures (tâches prioritaires ou de maintenance).

Fuhrmann et Copper (1985)[28] définissent une série d'hypothèses caractérisant les systèmes de files d'attente vérifiant la propriété de décomposition stochastique, particulièrement pour les systèmes d'attente avec vacances généralisées.

Ces hypothèses sont :

- Les clients arrivent dans le système selon un processus de Poisson.
- La durée de service suit une loi générale. Les temps de service des différents clients sont indépendants les uns des autres, et sont indépendants du processus des arrivées. Chaque durée de service est aussi indépendante de la suite des périodes des vacances qui la précède.
- Tous les clients arrivant dans le système sont éventuellement servis, et $\rho < 1$.
- Les clients sont servis dans un ordre qui est indépendant de leur durée de service.
- Les règles qui déterminent quand le serveur débute et termine les vacances n'ont aucune influence sur le processus des arrivées.
- Le nombre de clients arrivant dans le système durant les vacances est indépendant du nombre de clients présents dans le système au début des vacances.

La décomposition stochastique, établie dans [28], stipule que la distribution stationnaire de la taille du système peut être décomposée en deux variables; l'une de ces variables correspond à la taille du système d'attente ordinaire sans vacances, la seconde variable aléatoire est liée à la taille du système étant donné que le serveur est en vacances. Dans un certain sens, les modèles d'attente avec rappels peuvent être considérés comme un type particulier des modèles avec vacances, où les vacances commencent après chaque durée de service et le serveur se met à nouveau selon la compétition entre deux flux indépendants. L'un de ces flux est poissonnien, et correspond aux appels primaires; l'autre (correspondant aux appels répétés) possède une structure complexe et son intensité dépend du nombre de clients en orbite. Par conséquent, le modèle sans vacances est le système d'attente classique sans rappels, et les vacances du serveur sont occasionnées par les tentatives répétées. Pour les modèles avec vacances, la décomposition stochastique est valable aussi bien pour la distribution stationnaire de la taille du système que pour le temps virtuel d'attente. Dans [29], sous l'hypothèse que le temps inter-arrivées est de loi générale, une formule importante a été obtenue. Cette dernière lie le temps d'attente d'un système avec vacances au temps d'attente d'un système GI/G/1 ordinaire. Pour les modèles avec rappels, la validité de décomposition pour le temps d'attente est une conjoncture.

Dans ce chapitre, nous passons en revue la propriété de décomposition stochastique des systèmes de files d'attente avec vacances. Puis, nous présentons les résultats existant dans la littérature sur la propriété de décomposition stochastique des systèmes de files d'attente de type M/G/1 avec rappels.

3.2 Propriété de décomposition stochastique des modèles avec vacances

Pour les modèles de type M/G/1 avec vacances, la propriété en question a été établie pour la première fois par Fuhrmann et Cooper (1985) [28] à l'aide des outils de la théorie de renouvellement, puis confirmée par Doshi (1986) [15].

On distingue les vacances du serveur dans le cas d'un service exhaustif et celles dans le cas d'un service non-exhaustif. Dans la première situation, le serveur prend ses vacances lorsque le système est vide. En outre, le serveur est en vacances jusqu'à ce qu'il trouve (en revenant des vacances) au moins un client dans le système (vacances multiples), ou bien il prend une seule vacation après chaque période d'activité (vacation unique). Dans la seconde situation, le serveur est autorisé à prendre des vacances en présence des clients dans le système. Encore, la distribution des vacances peut être indépendante du processus de service, ou dépendante. Il est également possible que les durées des vacances sont mutuellement indépendantes, ou elles dépendent des vacances précédentes.

Ajoutons une suite de vacances $\{V_k\}$ à la description du modèle M/G/1 (voir chapitre 1), et supposons qu'elle est indépendante du processus des arrivées et du processus de service. Il s'agit d'une suite composée de variables aléatoires positives, indépendantes et identiquement distribuées.

3.2.1 Cas d'un service exhaustif

3.2.1.1 Système de files d'attente M/G/1 avec vacation unique

Considérons le système de files d'attente M/G/1 décrit dans le chapitre 1 et supposons que à la fin d'une période d'activité interrompue, le serveur prend une vacation [45],[27]. Ajoutons une suite de vacances $\{V_k\}$ à la description du modèle M/G/1, et supposons qu'elle est indépendante du processus des arrivées et du processus de service. Il s'agit d'une suite composée de variables aléatoires positives, indépendantes et identiquement distribuées.

Notons par $V(x)$ la fonction de répartition des durées de vacances. Et par $\tilde{V}(s)$, $Re(s) > 0$ sa transformée de Laplace-Stieltjes.

Supposons que le serveur revient au système après la vacation. A cet instant, s'il trouve des clients dans le système, il commence immédiatement une période d'activité jusqu'à ce que le système devient vide. Sinon, il attend le premier client qui arrive et puis commence une période d'activité. On s'intéresse au processus stochastique

$$\{N(t); t \geq 0\}. \quad (3.1)$$

où $N(t)$ est le nombre de clients dans le système à l'instant t . Il n'est pas markovien. Par conséquent, on considère le système aux époques de la fin de vacances ou de l'achèvement du service.

Soient :

ζ_n : l'instant où le service du $n^{\text{ième}}$ client s'achève ;

$q_n = N(\zeta_n)$: le nombre de clients dans le système après le $n^{\text{ième}}$ départ ;

v_n : le nombre de clients arrivant dans le système durant le temps de service du $n^{\text{ième}}$ client ;

b : le nombre de clients qui arrivent pendant la vacation.

La variable $I_n = I(\zeta_n)$ est

$$I_n = \begin{cases} 0 & \text{fin de vacance;} \\ 1 & \text{fin de service.} \end{cases} \quad (3.2)$$

Le vecteur $\{(q_n, I_n) : q_n = 0,1,2,\dots ; I_n = 0,1\}$ décrit l'état du $n^{\text{ième}}$ époque, c'est-à-dire:

si $I_n = 0$, alors q_n est le nombre de clients dans le système à l'instant de la fin de vacance ;

si $I_n = 1$, alors q_n est le nombre de clients dans le système immédiatement après l'achèvement du service.

Par conséquent à l'instant ζ_n , la séquence $\{(q_n, I_n)\}$ détermine une chaîne de Markov induite gouvernée par l'équation fondamentale suivante :

$$(3.3) \quad (q_{n+1}, I_{n+1}) = \begin{cases} (q_n + v_{n+1} - 1, 1) & q_n \geq 1; \\ (v_{n+1}, 1) & (q_n, I_n) = (0,0); \\ (b, 0) & (q_n, I_n) = (0,1). \end{cases}$$

Les variables aléatoires v_n sont indépendantes entre elles, leur distribution commune est:

$$(3.4) \quad P(v_n = k) = a_k = \int_0^\infty \exp(-\lambda x) \frac{(\lambda x)^k}{k!} dB(x); \quad k = 0,1,2,\dots$$

De même :

$$(3.5) \quad P(b = j) = b_j = \int_0^\infty \exp(-\lambda t) \frac{(\lambda t)^j}{j!} dV(t).$$

La distribution de l'état stationnaire est

$$p_{ij} = \lim_{n \rightarrow \infty} P(I_n = i, q_n = j) , i = 0,1 \text{ et } j \geq 0, \quad (3.6)$$

Elle satisfait les équations suivantes :

$$p_{0j} = p_{1,0} b_j \quad j = 0,1,2,\dots; \quad (3.7)$$

$$p_{1j} = p_{00} a_j + \sum_{n=1}^{j+1} p_{.,j} a_{j-n+1} \quad j = 0,1,2,\dots; \quad (3.8)$$

$$\sum_{(i,j)} p_{ij} = 1 ; \quad (3.9)$$

où

$$p_{.,j} = p_{0j} + p_{1j}.$$

Définissons les fonctions génératrices suivantes :

$$A(z) = \sum_{j=0}^{\infty} a_j z^j = \tilde{B}(\lambda - \lambda z) ; V(z) = \sum_{j=0}^{\infty} b_j z^j = \tilde{V}(\lambda - \lambda z) ; \quad (3.10)$$

$$Q_i(z) = \sum_{j=0}^{\infty} p_{ij} z^j , \quad i = 0,1;$$

$$Q(z) = Q_0(z) + Q_1(z) = \sum_{j=0}^{\infty} p_{.,j} z^j . \quad (3.11)$$

En utilisons les équations (3.7) et (3.8), on obtient :

$$Q_1(z) = \frac{b_0 \tilde{B}(\lambda - \lambda z)(z+1) + \tilde{B}(\lambda - \lambda z)[\tilde{V}(\lambda - \lambda z) - 1]}{z - \tilde{B}(\lambda - \lambda z)} p_{1,0} ;$$

$$Q(z) = \frac{z \tilde{V}(\lambda - \lambda z) \tilde{B}(\lambda - \lambda z) + b_0 \tilde{B}(\lambda - \lambda z)(z-1)}{z - \tilde{B}(\lambda - \lambda z)} p_{1,0},$$

où $b_0 = \tilde{V}(\lambda)$.

Il est clair que pour $A(z) = z$, il n'y a pas de solution dans l'intervalle $[0,1]$. D'où, le

système est en régime stationnaire si $A'(1) = \sum_{n=0}^{\infty} n a_n = \frac{\lambda}{\mu} = \rho < 1$.

La fonction génératrice du nombre de clients dans le système est :

$$Q(z) = \frac{(1-\rho)(1-z)\tilde{B}(\lambda - \lambda z)}{\tilde{B}(\lambda - \lambda z) - z} \frac{1 - \tilde{V}(\lambda - \lambda z) + (1-z)\tilde{V}(\lambda)}{[\tilde{V}(\lambda) + \lambda E[V]](1-z)} \quad (3.12)$$

$$= \prod(z)R(z),$$

où $\Pi(z) = \frac{(1-\rho)(1-z)\tilde{B}(\lambda-\lambda z)}{\tilde{B}\lambda-\lambda z-z}$ est la fonction génératrice du nombre de clients dans le système M/G/1 classique ; et $R(z) = \frac{1-\tilde{V}(\lambda-\lambda z)+(1-z)\tilde{V}(\lambda)}{[\tilde{V}(\lambda)+\lambda E[V](1-z)}$ est la fonction génératrice du nombre de clients qui arrivent dans le système durant les

vacances. Donc, le résultat de la décomposition stochastique est vérifié pour le nombre de clients dans le système.

Pour le temps d'attente d'un client arbitraire dans le système, on a :

$$\Pi(z) = \tilde{W}^{M/G/1}(\lambda-\lambda z)\tilde{B}(\lambda-\lambda z).$$

Par conséquent,

$$\tilde{W}^{M/G/1}(s) = \frac{\Pi(1-\frac{s}{\lambda})}{\tilde{B}(s)}.$$

De même, on a :

$$\tilde{W}(s) = \frac{\Pi(1-\frac{s}{\lambda})}{\tilde{B}(s)}. \quad (3.13)$$

A partir de cette équation, on obtient la transformée de Laplace-Stieltjes de la distribution du temps d'attente d'un client arbitraire dans le système :

$$\begin{aligned} \tilde{W}(s) &= \tilde{W}^{M/G/1}(s) \frac{1-\tilde{V}(s)+s\frac{\tilde{V}(\lambda)}{\lambda}}{\frac{s}{\lambda}} \\ &= \tilde{W}^{M/G/1}(s)\tilde{W}_v(s), \end{aligned} \quad (3.14)$$

où $\tilde{W}^{M/G/1}$ est la transformée de Laplace-Stieltjes de la distribution du temps d'attente dans le système M/G/1 classique, $\tilde{W}_v(s)$ est la TL-S de la distribution du temps d'attente liée aux vacances. L'équation (3.14) exprime la propriété de décomposition stochastique pour le temps d'attente.

3.2.1.2 Système de files d'attente M/G/1 avec vacances multiples

Considérons le système de files d'attente M/G/1 avec vacances. A présent, lorsque le serveur trouve le système vide à la fin des vacances, il prend immédiatement une autre vacance, et continue de cette manière jusqu'à ce qu'il trouve un client en attente à son retour des vacances.

Nous définissons une chaîne de Markov induite de la même manière que dans le cas précédent sauf que dans ce modèle l'état (0,0) n'existe pas. Alors, à l'instant ξ_n l'équation fondamentale de la chaîne de Markov est:

$$(3.15) \quad (q_{n+1}, I_{n+1}) = \begin{cases} (q_n + v_{n+1} - 1, 1) & q_n \geq 1; \\ (b^*, 0) & (q_n, I_n) = (0,1). \end{cases}$$

où b^* est le nombre de clients présents à la fin de la période des vacances. Il est clair que

$$P[b^* = j] = \frac{b_j}{1 - b_0} = \frac{b_j}{1 - \tilde{V}(\lambda)}, \quad j = 1, 2, \dots$$

Les probabilités $p_{ij} = \lim_{n \rightarrow \infty} P(I_n = i, q_n = j)$, $i = 0, 1$ et $j = 0, 1, 2, \dots$, satisfont les équations suivantes[15] :

$$p_{0j} = \frac{b_j}{1 - \tilde{V}(\lambda)} p_{1,0}, \quad j = 1, 2, \dots; \quad (3.16)$$

$$p_{1j} = \sum_{n=1}^{j+1} p_{,j} a_{j-n+1}, \quad j = 0, 1, 2, \dots \quad (3.17)$$

A l'aide des mêmes procédures que celles-ci-dessus, on trouve

$$Q_0(z) = \frac{\tilde{V}(\lambda - \lambda z) - \tilde{V}(\lambda)}{1 - \tilde{V}(\lambda)} p_{1,0}; \quad (3.18)$$

$$Q_1(z) = \frac{\tilde{B}(\lambda - \lambda z) \frac{[\tilde{V}(\lambda - \lambda z) - 1]}{[1 - \tilde{V}(\lambda)]}}{z - \tilde{B}(\lambda - \lambda z)} p_{1,0}. \quad (3.19)$$

Encore, il est évident que la condition d'existence des probabilités p_{ij} est $\tilde{B}(\lambda - \lambda z) \neq z$ sur l'intervalle $[0, 1]$. On a :

$$Q_1(1) = \frac{\lambda E[V]}{(1 - \tilde{V}(\lambda))(1 - \rho)} p_{1,0}.$$

Puisque $Q_1(1) + p_{1,0} = 1$, on obtient

$$p_{1,0} = \frac{(1 - \rho)(1 - \tilde{V}(\lambda))}{1 - \rho + \lambda E[V]}. \quad (3.20)$$

En substituant (3.20) dans (3.18) et (3.19), on trouve :

$$(3.21) \quad Q_0(z) = \frac{[\tilde{V}(\lambda - \lambda z) - \tilde{V}(\lambda)](1 - \rho)}{1 - \rho + \lambda E[V]},$$

$$Q_1(z) = \frac{\tilde{B}(\lambda - \lambda z)[\tilde{V}(\lambda - \lambda z) - 1](1 - \rho)}{[z - \tilde{B}(\lambda - \lambda z)](1 - \rho + \lambda E[V])}. \quad (3.22)$$

D'après l'équation (3.11), la fonction génératrice du nombre de clients dans le système avec vacances multiples se présente sous la forme de décomposition stochastique

$$\begin{aligned} Q(z) &= \frac{(1 - \rho)(1 - z)\tilde{B}(\lambda - \lambda z)}{\tilde{B}(\lambda - \lambda z) - z} \frac{1 - \tilde{V}(\lambda - \lambda z)}{\lambda E[V](1 - z)} \\ &= \prod(z)R(z), \end{aligned} \quad (3.23)$$

où $\prod(z)$ est la fonction génératrice du nombre de clients dans le système M/G/1 ordinaire; $R(z)$ est la fonction génératrice pour le nombre de clients qui vont arriver dans le système durant les vacances.

En ce qui concerne le temps d'attente d'un client arbitraire dans le système, en se basant sur l'équation (3.11), on trouve:

$$\tilde{W}(s) = \tilde{W}^{M/G/1}(s) \frac{1 - \tilde{V}(s)}{sE[V]}, \quad (3.24)$$

qui est la transformée de Laplace-Stieltjes du temps d'attente d'un client arbitraire. Cette dernière est le produit de la TL-S du temps d'attente dans le système ordinaire M/G/1 et de la TL-S du temps d'attente des clients arrivant pendant les vacances.

3.2.2 Cas d'un service non-exhaustif

Dans ce type des modèles, les vacances peuvent commencer quand les clients sont présents dans le système. Deux types de problèmes apparaissent : le serveur laisse la priorité aux vacances (vacances avec préemption), et les vacances ont lieu lorsque le service ou la vacation sont achevés (vacances sans préemption).

Les vacances avec préemption sont observées dans les systèmes où le serveur est sujet à des pannes. Leur première étude a été réalisée par Gaver (1962)[29]. Ici, les vacances correspondent à des durées d'interruption (ou réparation), et on parle de la durée d'accomplissement du service. On définit la durée d'accomplissement du service comme l'intervalle aléatoire de temps depuis le début du service d'un client jusqu'à l'instant où le service du client suivant peut commencer. Ainsi, on est amené à considérer cette période comme une durée de service (si aucune interruption n'a lieu), ou comme une durée de service effective plus la durée de réparation.

Si les durées de fonctionnement du serveur sont exponentiellement distribuées, les vacances ont essentiellement deux effets sur les clients arrivant dans le système :

- Elles étendent réellement la durée de service d'un client à la durée d'accomplissement du service. Ceci est analogue à la durée d'accomplissement du service utilisée lors de l'étude des modèles avec priorité et arrivées poissoniennes. La différence réside dans le fait qu'au cours d'une vacation, il n'est pas possible d'avoir d'autres interruptions.

- Après une période d'inactivité, le premier client arrivé peut trouver le serveur en vacance, et la période d'activité peut commencer avec plus d'un client.

Donc, si nous remplaçons la durée de service par le temps d'accomplissement du service, le modèle se réduit à un modèle avec service exhaustif et la décomposition stochastique pour le nombre de client dans le système se présente de la manière suivante :

$$Q(z) = \Pi(z)R(z),$$

où $R(z)$ est la fonction génératrice de la distribution stationnaire du nombre de clients dans le système étant donné que le serveur est en vacances. Ce résultat est connu comme celui pour tout système d'attente avec vacances. Il rend possible de se concentrer uniquement sur l'étude des effets des vacances sur le nombre de clients dans le système, étant donné que le serveur est en vacances.

Quand les vacances correspondent à des clients secondaires, les vacances peuvent arriver même au cours d'une vacance. Un tel cas a été considéré par Ou(1984)[49], où le serveur sert deux types de clients : le premier possède des arrivées poissoniennes et temps de service général, et le second a des arrivées et temps de service général. Le client de type 2 peut être considéré comme une vacance. Le résultat de la décomposition stochastique est dérivé pour le nombre des clients de type 1 quand le temps inter-arrivées des clients de type 2 est constant. Un cas spécial des modèles avec vacances sans préemption et modèles à service limité a été étudié par Ali et Neuts (1984)[48]. L'analyse réalisée par les auteurs révèle une double décomposition aboutissant à trois composantes pour la fonction génératrice de la distribution stationnaire du nombre de clients dans le système à un instant arbitraire d'accomplissement du service :

$$Q(z) = \Pi(z)D(z)R(z),$$

où $D(z)$ est une fonction génératrice pour le nombre de clients se trouvant dans le système à une date arbitraire où les vacances commencent.

3.3 Propriété de décomposition stochastique des modèles avec rappels

3.3.1 Système de files d'attente M/G/1 avec rappels

La propriété de décomposition stochastique du système M/G/1 avec rappels dans le cas de la distribution exponentielle du temps inter-rappels a été établie par Yang et Templeton (1987)[51]. Les auteurs ont appliqué la méthode de la chaîne de Markov induite aux instants de départ. On s'intéresse au processus stochastique :

$$\{N(t) = C(t) + N_0(t), t \geq 0\},$$

où $N(t)$ est le nombre de clients dans le système, $C(t)$ est 0 ou 1 selon le fait que le serveur est libre ou occupé, $N_0(t)$ est le nombre de clients en orbite à la date t . Ce n'est pas un processus de Markov. Cependant son étude peut être ramenée à celle

d'une chaîne de Markov induite aux instants ξ_n où le serveur devient libre pour la $n^{\text{ième}}$ fois, $\{N(\xi_n), n \geq 1\} = \{q_n, n \geq 1\}$. Supposons que le système est en régime stationnaire ($\rho < 1$).

Considérons la relation fondamentale de la chaîne de Markov induite :

$$q_{n+1} = q_n - \delta_{q_n} + a_{n+1}. \quad (3.25)$$

Etant donné que les variables aléatoires $q_n - \delta_{q_n}$ et a_{n+1} sont indépendantes, on obtient :

$$z^{q_{n+1}} = z^{q_n - \delta_{q_n}} z^{a_{n+1}} ;$$

$$\text{Lim}_{n \rightarrow \infty} E[z^{q_{n+1}}] = \lim_{n \rightarrow \infty} (E[z^{q_n - \delta_{q_n}}] E[z^{a_{n+1}}]) .$$

D'où

$$Q(z) = \frac{(1-z)[D(z) - d_0]}{[z - z\tilde{B}(\lambda - \lambda z)]} \tilde{B}(\lambda - \lambda z), \quad (3.26)$$

où

$$Q(z) = \sum_{k=0}^{\infty} p_k z^k, \text{ avec } p_k = \lim_{t \rightarrow \infty} p(N(t) = k) = \lim_{n \rightarrow \infty} p(N(\xi_n) = k);$$

$$D(z) = \sum_{k=0}^{\infty} d_k z^k \text{ et } d_0 = 1 - \sum_{k=1}^{\infty} d_k.$$

Ainsi

$$E[q_{n+1}] = E[q_n] - E[\delta_{q_n}] + E[a_{n+1}];$$

$$\text{Lim}_{n \rightarrow \infty} E[q_{n+1}] = \lim_{n \rightarrow \infty} (E[q_n] - E[\delta_{q_n}] + E[a_{n+1}]). \quad (3.27)$$

Notons que $\lim_{n \rightarrow \infty} a_n = a$, $\lim_{n \rightarrow \infty} \delta_{q_n} = \delta_q$ et $\lim_{n \rightarrow \infty} q_n = q$. L'équation (3.27) devient:

$$E[a] = E[\delta_q].$$

Alors,

$$E[\delta_q] = \sum_{k=0}^{\infty} [0 \cdot P(\delta_q = 0 / q = k) + 1 \cdot P(\delta_q = 1 / q = k)] \cdot P(q = k)$$

$$= \sum_{k=1}^{\infty} \frac{k\theta}{\lambda + k\theta} p_k = \sum_{k=1}^{\infty} d_k = \rho.$$

Donc, $d_0 = 1 - \rho$. En prenant en considération la relation entre les variables aléatoires q et d , on trouve :

$$Q'(z) - D'(z) = \frac{\lambda}{\theta z} [D(z) - d_0],$$

$$D'(z) + \frac{\lambda}{\theta z} D(z) = Q'(z) + \frac{1}{z} \frac{\lambda}{\theta} (1 - \rho),$$

Dont la solution pour $D(z)$ est :

$$D(z) = \alpha z^{-\frac{\lambda}{\theta}} + z^{-\frac{\lambda}{\theta}} \int z^{\frac{\lambda}{\theta}} Q'(z) dz + d_0, \quad (3.28)$$

où α est une constante indéterminée. En combinant (3.6) et (3.28), on obtient l'équation différentielle de premier ordre pour $Q(z)$:

$$\left[S(z) - z^{\frac{\lambda}{\theta}} \right] Q'(z) + S'(z) Q(z) = 0, \quad (3.29)$$

où

$$S(z) = \frac{z^{\frac{\lambda}{\theta}+1} [1 - \tilde{B}(\lambda - \lambda z)]}{\tilde{B}(\lambda - \lambda z)(1 - z)}.$$

En résolvant (3.29) pour $Q(z)$ et en appliquant la condition de normalisation $Q(1) = 1$, on trouve l'expression de la propriété de décomposition stochastique du système de files d'attente M/G/1 avec rappels et distribution exponentielle du temps inter-rappels :

$$Q(z) = \frac{(1 - \rho)(1 - z)\tilde{B}(\lambda - \lambda z)}{\tilde{B}(\lambda - \lambda z) - z} \frac{\Phi(z)}{\Phi(1)}, \quad (3.30)$$

où

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

Ici $\frac{(1 - \rho)(1 - z)\tilde{B}(\lambda - \lambda z)}{\tilde{B}(\lambda - \lambda z) - z}$ est la fonction génératrice qui définit la distribution stationnaire du nombre de clients dans le système M/G/1 classique à espace d'attente illimité (équation de Pollaczek-Khintchine). Elle est indépendante du temps inter-rappels, et $\frac{\Phi(z)}{\Phi(1)}$ est la fonction génératrice de la distribution stationnaire du nombre de clients dans le système M/G/1 avec rappels étant donné que le serveur est libre.

Yang et al.(1994)[50] ont exploré la propriété de décomposition du système M/G/1 avec rappels, et ont prouvé que cette propriété est toujours vraie pour la distribution générale du temps inter-rappels :

$$Q(z) = \frac{(1-\rho)(1-z)\tilde{B}(\lambda-\lambda z)}{\tilde{B}(\lambda-\lambda z)-z} \frac{F_0(z)}{(1-\rho)},$$

où $\frac{F_0(z)}{(1-\rho)}$ est la fonction génératrice du nombre de clients dans le système avec rappels étant donné que le serveur est libre.

3.3.2 Autres modèles

La validité de la décomposition stochastique a été étendue aux modèles avec rappels et arrivées par groupes [37],[24],[32],[44]. Dans ce type de systèmes, il est supposé que les clients arrivent par groupes. A l'arrivée d'un groupe, si le serveur est libre, un client commence son service et les autres deviennent sources d'appels répétés; dans le cas contraire tout le groupe rejoint l'orbite. Chaque client dans l'orbite produit un flux poissonien de rappels de taux $\theta > 0$. Les flux d'arrivées et de rappels ainsi que les temps de service sont mutuellement indépendants.

On s'intéresse au processus stochastique $\{N(t), t \geq 0\}$. La suite $\{N(\xi_n); n \geq 1\}$ forme une chaîne de Markov induite, soit $q_n = N(\xi_n)$. Considérons l'équation fondamentale suivante:

$$q_{n+1} = q_n - \Delta \delta_{q_n} + \Delta - 1 + a_{n+1}, \quad (3.31)$$

où a_{n+1} est le nombre de clients primaires arrivant dans le système durant le service du $(n+1)^{ième}$ client; δ_{q_n} est 0 ou 1 selon le fait que le $(n+1)^{ième}$ client en service est primaire ou non; Δ est la taille du premier groupe qui arrive après ξ_n .

Soient $\lim_{t \rightarrow \infty} P[\Delta = k] = g_k$; $\rho = \bar{g} \frac{\lambda}{\mu} < 1$ où \bar{g} est la taille moyenne d'un groupe.

Les équations de Kolmogorov pour la distribution stationnaire

$p_k = \lim_{n \rightarrow \infty} P(N(\xi_n) = k)$ sont :

$$p_k = \sum_{m=0}^k p_m \frac{\lambda}{\lambda+m\theta} \sum_{i=1}^{k-m+1} g_i a_{k-m-i} + \sum_{m=1}^{k+1} p_m \frac{m\theta}{\lambda+m\theta} a_{k-m+1}, \quad k = 0, 1, \dots \quad (3.32)$$

où

$$a_k = P[a_n = k] = \int_0^\infty \sum_j \frac{(\lambda x)^j}{j!} e^{-\lambda x} g_k^{(j)} dB(x)$$

avec "j-fold convolution" pour la séquence g_i notée $g_k^{(j)}$.

Pour $n \rightarrow \infty$, nous avons :

$$E[a] = \rho = 1 - E[\Delta] + E[\Delta] \sum_{k=0}^{\infty} \frac{k\theta}{\lambda+k\theta} p_k.$$

On définit les fonctions génératrices suivantes :

$$\begin{aligned} Q(z) &= \sum_{k=0}^{\infty} p_k z^k; & G(z) &= \sum_{k=0}^{\infty} g_k z^k; \\ D(z) &= \sum_{k=0}^{\infty} d_k z^k; & A(z) &= \sum_{k=0}^{\infty} a_k z^k = \tilde{B}(\lambda - \lambda G(z)). \end{aligned}$$

$$\text{où } \sum_{k=1}^{\infty} d_k = 1 - \frac{1-\rho}{\bar{g}} \text{ et } d_0 = \frac{1-\rho}{\bar{g}}.$$

Alors pour l'équation (3.31), on a :

$$Q(z) = \frac{[1-G(z)][D(z)-d_0]\tilde{B}(\lambda-\lambda G(z))}{z-\tilde{B}(\lambda-\lambda G(z))} \quad (3.33)$$

$$\text{ainsi que } Q'(z) - D'(z) = \frac{\lambda}{\theta z} D(z) - \frac{\lambda}{\theta z} d_0$$

Remplaçons (3.28) dans (3.33), on obtient :

$$Q'(z) \left[Z^{\frac{\lambda}{\theta}} - S(z) \right] - S'(z)Q(z) = 0 \quad (3.34)$$

avec

$$S(z) = z^{\frac{\lambda}{\theta}} \frac{z-\tilde{B}(\lambda-\lambda G(z))G(z)}{[1-G(z)]\tilde{B}(\lambda-\lambda G(z))}.$$

En résolvant (3.34) pour $Q(z)$ et en appliquant la condition de normalisation $Q(1) = 1$, on obtient l'expression de la propriété de décomposition stochastique du système M/G/1 avec rappels, arrivées par groupe et distribution d'inter-rappels exponentielle :

$$Q(z) = \frac{(1-\rho)[1-G(z)]\tilde{B}(\lambda-\lambda G(z))}{\bar{g}(\tilde{B}(\lambda-\lambda G(z))-z)} \frac{\Psi(z)}{\Psi(1)} \quad (3.35)$$

où

$$\Psi(z) = \exp \left\{ -\frac{\lambda}{\theta} \int_0^z \frac{t-G(t)\tilde{B}(\lambda-\lambda G(t))}{t-[t-\tilde{B}(\lambda-\lambda G(t))]} dt \right\}.$$

Le facteur $\frac{(1-\rho)[1-G(t)]\tilde{B}(\lambda-\lambda G(t))}{\bar{g}(\tilde{B}(\lambda-\lambda G(t))-t)}$ est la fonction génératrice qui définit la distribution stationnaire du nombre de clients dans le système M/G/1 avec arrivées par groupe.

Soient η_n l'instant du début du $n^{\text{ième}}$ service, $q'_n = N(\eta_n)$ le nombre de clients présents dans le système à l'instant η_n . Par conséquent, nous avons :

$$q'_{n+1} = q'_n + \Delta'(1 - \delta_{q'_n}).$$

Notons par $P_{q'}(z)$ la fonction génératrice pour $q' = \lim_{n \rightarrow \infty} q'_n$. En utilisant la méthode ci-dessus, on trouve :

$$P_{q'}(z) = G(z) Q(z) + [1 - G(z)][D(z) - d_0].$$

De (3.33) et (3.35), on obtient :

$$P_{q'}(z) = \frac{(1 - \rho) z [1 - G(z)] \Psi(z)}{\bar{g} [\tilde{B}(\lambda - \lambda G(z)) - z] \Psi(1)}.$$

Soit $P_B(z)$ la fonction génératrice pour le nombre de clients dans le système à l'arrivée d'un client (le serveur est occupé), et $P_I(z)$ la fonction génératrice pour le nombre de clients dans le système à l'arrivée d'un client (le serveur est libre).

Alors, la fonction génératrice du nombre de clients dans le système à l'arrivée d'un client est

$$P_{q'}(z) = \rho P_B(z) + (1 - \rho) P_I(z). \quad (3.36)$$

Maintenant, supposons que le groupe arrive durant l'intervalle $[t, t + dt]$, où t désigne le temps écoulé après le début de la période de service dont la longueur est x . La probabilité que le serveur est occupé est [50] :

$$P_B(z) = \int_{x=0}^{\infty} \int_{t=0}^x P_{q_{b(t)}}(z) \frac{b(x)}{\bar{x}} dt dx, \quad (3.37)$$

avec $q_{b(t)} = q' + B(t)$ le nombre de clients dans le système à la date t après le début du service et $b(t) = \frac{B'(x)}{1 - B(x)}$. On a :

$$P_{q_{b(t)}}(z) = P_{q'}(z) \exp[-\lambda t(1 - G(z))].$$

Les équations (3.37) et (3.36), donnent :

$$P_B(z) = \frac{1 - \tilde{B}(\lambda - \lambda G(z))}{\lambda \bar{x} [1 - G(z)]} P_{q'}(z). \quad (3.38)$$

De (3.33) et (3.35), nous obtenons :

$$P_I(z) = \frac{\Psi(z)}{\Psi(1)}. \quad (3.39)$$

Donc,

$$P_{\dot{q}}(z) = \frac{(1-\rho)(1-z)\tilde{B}(\lambda-\lambda G(z))}{\tilde{B}(\lambda-\lambda G(z))-z} \frac{\Psi(z)}{\Psi(1)}, \quad (3.40)$$

Puisque le processus des arrivées est poissonien, $P(z) = P_{\dot{q}}(z)$.

Pour résumer les résultats de la décomposition stochastique pour le modèle M/G/1 avec rappels et arrivées par groupe, soient :

M : modèle M/G/1 classique avec arrivées par groupe ;

M_{θ} : modèle M/G/1 avec rappels et arrivées par groupe ;

N : nombre de clients dans M (régime stationnaire) ;

$N^{-} = N(\eta)$: nombre de clients dans M à l'instant du début du service (régime stationnaire) ;

$N^{+} = N(\zeta)$: nombre de clients dans M à l'instant du départ (régime stationnaire).

De même, soient N_{θ} , N_{θ}^{-} et N_{θ}^{+} correspondants et liés au modèle M_{θ} , alors :

$$N_{\theta} = N + \psi ; \quad N_{\theta}^{-} = N^{-} + \psi ; \quad N_{\theta}^{+} = N^{+} + \psi,$$

avec ψ un nombre entier qui admet la fonction génératrice $\frac{\Psi(z)}{\Psi(1)}$.

3.4 Applications de la propriété de décomposition stochastique

La propriété de décomposition est extrêmement utile pour analyser les systèmes de files d'attente M/G/1 avec rappels. Elle permet de résoudre les problèmes d'obtention des moments de $N_{\theta}(t)$ et de convergence d'un système M/G/1 avec rappels vers un système limite lorsque le taux de rappels $\theta \rightarrow \infty$. Ces applications, dans le cas de la distribution exponentielle du temps inter-rappels, sont réalisées par Artalejo et Falin (1994) [4].

La propriété de décomposition du système M/G/1 avec rappels s'écrit :

$$\{C_{\theta}(t); N_{\theta\theta}(t)\} = \{C_{\infty}(t); N_{f\infty}(t)\} + \{0; R_{\theta}(t)\}. \quad (3.41)$$

Les processus $\{C_{\theta}(t); N_{\theta\theta}(t)\}$ et $\{0; R_{\theta}(t)\}$ sont associés au modèle M/G/1 avec rappels de taux de rappels $\theta > 0$, où $R_{\theta}(t)$ représente le nombre de clients en orbite à la date t étant donné que le serveur est libre. Le processus $\{C_{\infty}(t); N_{f\infty}(t)\}$ est associé

au modèle $M/G/1$ ordinaire de capacité d'attente illimitée, où $N_{f\infty}(t)$ est le nombre de clients en attente à la date t .

3.4.1 Formules explicites des moments du nombre de clients en orbite

Soit un moment factoriel d'ordre n : $M_n = E[(N_{a\infty})_n]$, où $(a)_n = a(a-1)\dots(a-n+1)$.

Soient n et j des entiers non-négatifs. On définit un ensemble $X_j(n)$ de tous les vecteurs $\vec{x} = (x_1, \dots, x_j) \in \mathbb{Z}^j$, tel que $x_i > 0$ et $x_1 + \dots + x_j = n$. On désigne $\vec{x}! = (x_1! \dots x_j!)$.

Théorème 3.4.1 : *Le moment factoriel d'ordre n du nombre de clients en orbite $M_n(\theta) = E[(N_{o\theta})_n]$, où $N_{o\theta} = \lim_{t \rightarrow \infty} N_{o\theta}(t)$, est définie par :*

$$M_n(\theta) = M_n + \sum_{k=1}^n \binom{n}{k} M_{n-k} k! \sum_{j=1}^k \frac{1}{j!} \left(\frac{\lambda}{\theta(1-\rho)} \right)^j \sum_{\vec{x} \in X_j(k)} \frac{1}{x!} \prod_{i=1}^j (M_{x_i-1} - (1-\rho)\delta_{x_i;1}). \quad (3.42)$$

Preuve : De l'équation (3.41), on a :

$$N_{o\theta} = N_{f\infty}(t) + R_\theta(t).$$

En utilisant le théorème binomial de Vandermonde et l'indépendance des variables aléatoires $N_{f\infty}(t)$ et $R_\theta(t)$, on obtient

$$M_n(\theta) = \sum_{k=0}^n \binom{n}{k} M_{n-k} E[(R_\theta)_k]. \quad (3.43)$$

Pour obtenir $E[(R_\theta)_k]$, nous considérons la fonction suivante :

$$Q_\theta(z) = \ln R_\theta(z) = \frac{\lambda}{\theta} \int_1^z \frac{1 - \tilde{B}(\lambda - \lambda u)}{\tilde{B}(\lambda - \lambda u) - u} du \quad (3.44)$$

avec

$$R_\theta(z) = E[z^{R_\theta}] = \exp \left\{ \frac{\lambda}{\theta} \int_1^z \frac{1 - \tilde{B}(\lambda - \lambda u)}{\tilde{B}(\lambda - \lambda u) - u} du \right\}. \quad (3.45)$$

Soit $P_\infty(z) = \frac{(1-\rho)(1-z)}{\tilde{B}(\lambda - \lambda z) - z}$ la fonction génératrice pour la variable aléatoire $N_{f\infty}(t)$.

Alors l'expression (3.44) peut être réécrite comme suit :

$$Q_\theta(z) = \frac{\lambda}{\theta(1-\rho)} \int_1^z P_\infty(u) du + \frac{\lambda}{\theta} (1-z). \quad (3.46)$$

En dérivant (3.46) par-rapport à z au point $z = 1$, on obtient :

$$Q_\theta^{(n)}(1) = \frac{\lambda}{\theta(1-\rho)} P_\infty^{(n-1)}(1) - \frac{\lambda}{\theta} \delta_{n1}, \quad n = 1, 2, \dots \quad (3.47)$$

$Q_\theta^{(n)}(1)$ est le $n^{\text{ième}}$ cumulatif factoriel de $R_\theta(t)$, et $P_\infty^{(n-1)}(1)$ est le $(n-1)^{\text{ième}}$ moment factoriel M_{n-1} de la variable aléatoire $N_{f_\infty}(t)$. Le moment factoriel $E[(R_\theta)_k]$ de la variable aléatoire $R_\theta(t)$ peut être exprimé en termes du cumulatif factoriel comme suit :

$$E[(R_\theta)_k] = k! \sum_{j=1}^k \frac{1}{j!} \sum_{\vec{x} \in X_j(k)} \frac{1}{x!} \prod_{i=1}^j Q_\theta^{(x_i)}(1), \quad k \geq 1. \quad (3.48)$$

De (3.47) et (3.48), on trouve:

$$E[(R_\theta)_k] - k! \sum_{j=1}^k \frac{1}{j!} \left(\frac{\lambda}{\theta(1-\rho)} \right)^j \sum_{\vec{x} \in X_j(k)} \frac{1}{x!} \prod_{i=1}^j (M_{x_i-1} - (1-\rho)\delta_{x_i,1}), \quad k \geq 1. \quad (3.49)$$

Finalement, en mettant l'équation (3.49) dans (3.43), nous obtenons (3.42).

Remarque 3.4.1.1 : A l'aide de ce théorème, il est possible de déduire les formules pour le nombre moyen et la variance du nombre de clients en orbite :

$$E[N_{o\theta}] = M_1(\theta) = \frac{\lambda^2 \beta_2}{2(1-\rho)} + \frac{\lambda \rho}{\theta(1-\rho)};$$

$$\begin{aligned} \text{Var}[N_{o\theta}] &= M_2(\theta) + M_1(\theta) + (M_1(\theta))^2 \\ &= \frac{\lambda^4 \beta_2^2}{4(1-\rho)^2} + \frac{\lambda^3 \beta_3}{3(1-\rho)} + \frac{\lambda^2 \beta_2}{2(1-\rho)} + \frac{\lambda^3 \beta_2}{2\theta(1-\rho)^2} + \frac{\lambda \rho}{\theta(1-\rho)}. \end{aligned}$$

Ainsi, les mesures, telles que : temps moyen d'attente, nombre moyen de rappels par client, peuvent être facilement calculées (d'après les formules de Little).

Remarque 3.4.1.2 : Dans le cas particulier où le temps de service est exponentiellement distribué avec taux γ , nous obtenons :

$$M_n(\theta) = \frac{\rho^n (\lambda + n\rho\theta)}{(1-\rho)^n \theta^n} \prod_{k=1}^{n-1} (\lambda + k\theta), \quad n \geq 1,$$

avec $\rho = \frac{\lambda}{\gamma}$.

3.4.2 Taux de convergence vers le système M/G/1 classique

En général, quand $\theta \rightarrow \infty$ la distribution stationnaire de l'état d'un système de files d'attente avec rappels converge vers une limite. Dans notre cas, cette limite est le système M/G/1 classique (*FIFO*) [4].

Théorème 3.4.2 : *Quand $\theta \rightarrow \infty$, la distance*

$$D = \sum_{i=0}^1 \sum_{n=0}^{\infty} |p_{in}(\theta) - p_{in}(\infty)| = 0\left(\frac{1}{\theta}\right).$$

L'inégalité suivante a lieu :

$$2(1 - \rho)(1 - R_0(\theta)) < D < 2(1 - R_0(\theta)), \quad (3.50)$$

où

$$R_0(\theta) = \exp \left\{ -\frac{\lambda}{\theta} \int_0^1 \frac{1 - \tilde{B}(\lambda - \lambda u)}{\tilde{B}(\lambda - \lambda u) - u} du \right\}. \quad (3.51)$$

Preuve 3.4.2 : La preuve est basée sur la propriété de décomposition stochastique (3.41).

Si $p_{in}(\theta)$ est la distribution stationnaire de $\{C_\theta(t); N_{o\theta}(t); t \geq 0\}$ et $p_{in}(\infty)$ est celle de $\{C_\infty(t); N_{f\infty}(t); t \geq 0\}$, alors $p_{in}(\theta)$ est une convolution des distributions $p_{in}(\infty)$ et $R_0(\theta) = \lim_{t \rightarrow \infty} P(R_\theta(t) = n)$:

$$p_{in}(\theta) = \sum_{k=0}^n p_{ik}(\infty) R_{n-k}(\theta). \quad (3.52)$$

Alors,

$$p_{in}(\theta) - p_{in}(\infty) = p_{in}(\infty)R_0(\theta) - p_{in}(\infty) + (1 - \delta_{n0}) \sum_{k=0}^{n-1} p_{ik}(\infty) R_{n-k}(\theta).$$

En utilisant l'équation (3.52), pour obtenir

$$\begin{aligned} |p_{in}(\theta) - p_{in}(\infty)| &< p_{in}(\infty)(1 - R_0(\theta)) + (1 - \delta_{n0}) \sum_{k=0}^{n-1} p_{ik}(\infty) R_{n-k}(\theta) \\ &= p_{in}(\infty)(1 - R_0(\theta)) + p_{in}(\theta) - p_{in}(\infty)R_0(\theta) \\ &= p_{in}(\infty)(1 - 2R_0(\theta)) + p_{in}(\theta). \end{aligned}$$

Cependant, en additionnant tous les états, nous trouvons que

$$D < (1 - 2R_0(\theta)) \sum_{i=0}^1 \sum_{n=0}^{\infty} p_{in}(\infty) + \sum_{i=0}^1 \sum_{n=0}^{\infty} p_{in}(\theta). \quad (3.53)$$

Puisque $p_{in}(\infty)$ et $p_{in}(\theta)$ sont des probabilités, alors la somme des sommes à l'aide de l'équation (3.53) égal a 1; d'où on obtient la borne supérieure pour l'équation (3.50). Pour trouver la borne inférieure de (3.50), on utilise l'inégalité $|a - b| \geq a - b$.

Par conséquent, nous avons :

$$\begin{aligned}
D &\geq \sum_{i=0}^1 |p_{i0}(\theta) - p_{i0}(\infty)| + \sum_{i=0}^1 \sum_{n=1}^{\infty} (p_{in}(\theta) - p_{in}(\infty)) \\
&= (1 - R_0(\theta)) \sum_{i=0}^1 p_{i0}(\infty) + 1 - \sum_{i=0}^1 p_{i0}(\theta) - 1 + \sum_{i=0}^1 p_{i0}(\infty) \\
&= 2(1 - R_0(\theta)) \sum_{i=0}^1 p_{i0}(\infty) \\
&= 2(1 - R_0(\theta)) \frac{(1-\rho)}{\bar{B}(\lambda)} \\
&> 2(1 - R_0(\theta))(1 - \rho).
\end{aligned}$$

Enfin, on obtient la formule (3.50).

3.5 Conclusion

Dans ce chapitre, nous avons présenté le concept général de la propriété de la décomposition stochastique des systèmes de files d'attente avec vacances ainsi que de ceux avec rappels. Dans un certain sens, les modèles d'attente avec rappels peuvent être considérés comme un type particulier des modèles avec vacances où les vacances commencent après chaque service, et leurs durées dépendent du processus des arrivées et de l'état du système.

L'application de cette propriété simplifie la résolution d'un certain nombre de problèmes (calcul des moments de la taille de l'orbite, vitesse de convergence vers un système limite).

Dans le prochain chapitre, nous étudions le comportement asymptotique du système M/G/1 avec rappels.

Chapitre 4

Comportement asymptotique du système de files d'attente M/G/1 avec rappels

4.1 Introduction

Bien que les caractéristiques de performance du système M/G/1 avec rappels sont disponibles sous forme explicite, elles sont encombrantes et d'une utilisation restreinte du point de vue pratique. Prenons l'expression de fonction génératrice du nombre de clients en orbite obtenue dans le chapitre 2. Elle ne permet pas d'étudier le caractère de la distribution de $N_o(t)$ dans certains cas particuliers importants dans les applications: régime chargé, intensité faible des rappels et intensité forte des rappels.

Dans ce chapitre, nous analysons le comportement asymptotique du nombre de clients en orbite associé au système de files d'attente M/G/1 avec rappels, et ceci dans le cas des valeurs extrêmes de certaines paramètres.

4.2 Trafic chargé

Nous considérons le cas où le taux d'arrivée λ augmente de telle manière que $\rho \rightarrow 1 - 0$.

Théorème 4.1 *Si le système de files d'attente M/G/1 avec rappels est dans un régime stationnaire ($\rho = \frac{\lambda}{\mu} < 1$) et $\beta_2 < \infty$, alors :*

$$\lim_{\lambda \rightarrow \frac{1}{\beta_1} - 0} E[e^{-s(1-\rho)N_o(t)}] = \left(1 + \frac{\beta_2}{2\beta_1^2} s\right)^{-1 - \frac{2\beta_1}{\theta\beta_2}}, \quad (4.1)$$

C'est-à-dire en régime chargé la longueur de l'orbite $N_o(t)$ asymptotiquement a une distribution Gamma.

Preuve:

Soit $\varepsilon = 1 - \rho$. La transformée de Laplace de la variable aléatoire

$\varepsilon N_o(t) \equiv (1 - \rho)N_o(t)$ peut être obtenue de l'équation (2.17) en mettant $z = e^{-\varepsilon s}$:

$$E[e^{-s\varepsilon N_o(t)}] = \varepsilon \frac{1 - e^{-\varepsilon s}}{K(e^{-\varepsilon s}) - e^{-\varepsilon s}} \times \exp \left\{ \frac{\lambda}{\theta} \int_1^{e^{-\varepsilon s}} \frac{1 - K(u)}{K(u) - u} du \right\}. \quad (4.2)$$

Comme $\lambda \rightarrow \frac{1}{\beta_1} - 0$, la variable $t = \lambda(1 - e^{-\varepsilon s}) \rightarrow 0$. Autrement dit, si $\beta_2 < \infty$ alors, comme $t \rightarrow 0$, l'expression asymptotique suivante a lieu:

$$\tilde{B}(t) = 1 - \beta_1 t + \frac{\beta_2}{2} t^2 + o(t^2).$$

Ainsi, $K(e^{-\varepsilon s}) \equiv \tilde{B}(\lambda(1 - e^{-\varepsilon s}))$

$$= 1 - \varepsilon s + \varepsilon^2 \left[s^2 \left(\frac{1}{2} + \frac{\beta_2}{2\beta_1^2} \right) + s \right] + o(\varepsilon^2).$$

Cela implique que

$$\lim_{\lambda \rightarrow \frac{1}{\beta_1} - 0} \varepsilon \frac{1 - e^{-\varepsilon s}}{K(e^{-\varepsilon s}) - e^{-\varepsilon s}} = \frac{1}{1 + \frac{\beta_2}{2\beta_1^2} s}. \quad (4.3)$$

Pour trouver la limite du terme exponentiel dans le coté droit de l'équation (4.2), on doit trouver

$$\lim_{\lambda \rightarrow \frac{1}{\beta_1} - 0} \int_1^{e^{-\varepsilon s}} \frac{1 - K(u)}{K(u) - u} du. \quad (4.4)$$

En introduisant une nouvelle variable $v = \frac{1-u}{1-e^{-\varepsilon s}}$, on peut transformer l'intégral dans (4.4), ce qui donne:

$$\int_0^1 \frac{1 - \tilde{B}(\lambda(1 - e^{-\varepsilon s})v)}{\tilde{B}(\lambda(1 - e^{-\varepsilon s})v) - 1 + (1 - e^{-\varepsilon s})v} \cdot (e^{-\varepsilon s} - 1) dv. \quad (4.5)$$

Si $\beta_2 < \infty$, alors dans un intervalle fini $[0; S]$ on a :

$$\tilde{B}(\varepsilon s) = 1 - \beta_1 \varepsilon s + \frac{\beta_2}{2} \varepsilon^2 s^2 + \varepsilon^2 \cdot 0(1).$$

Considérons la fonction $f(t) = \tilde{B}(t) - 1 + \beta_1 t - \frac{\beta_2}{2} t^2$. Nous avons que :

- a) $f(0) = 0, f'(0) = 0, f''(0) = 0,$
- b) $f'''(t) = \tilde{B}'''(t)$ et alors $f'''(t)$ est finie et négative pour $t > 0$.

Par conséquent, $f''(t)$ est décroissante lorsque $t \geq 0$. Ceci implique que $f''(t) < 0$ pour $t > 0$, c'est-à-dire : $f'(t)$ est décroissante pour $t > 0$. Donc, $f(t)$ est décroissante et $f(t) < 0$. Alors pour $t \leq T$, on a $f(T) \leq f(t) \leq 0$. En outre, l'existence de β_2 implique que $\lim_{t \rightarrow 0} \frac{f(t)}{t^2} = 0$.

Soit $t = \varepsilon s, T = \varepsilon S, 0 \leq s \leq S$. Par conséquent:

$$S^2 \frac{f(\varepsilon S)}{\varepsilon^2 S^2} \leq \frac{f(\varepsilon s)}{\varepsilon^2} \leq 0.$$

Comme $\varepsilon \rightarrow 0$, la variable $\frac{f(\varepsilon S)}{\varepsilon^2 S^2} \rightarrow 0$ alors: $\frac{f(\varepsilon s)}{\varepsilon^2} \rightarrow 0$ uniformément pour

$s \in [0; S]$.

Supposons $s = v, \varepsilon = \lambda(1 - e^{-\varepsilon s})$. Vu que $\lambda \rightarrow 1/\beta_1 - 0$ uniformément par rapport à $v \in [0; 1]$, on obtient :

$$\tilde{B}(\lambda(1 - e^{-\varepsilon s})v) = 1 - \varepsilon s v + \varepsilon^2 \left(s v + \frac{s^2 v}{2} + \frac{\beta_2}{2\beta_1^2} s^2 v^2 \right) + \varepsilon^2 \cdot 0(1). \quad (4.6)$$

Substituons l'équation (4.6) dans (4.5). On obtient que la fonction sous le signe de l'intégral (4.5) converge uniformément vers la fonction

$$- \frac{s}{1 + \frac{\beta_2}{2\beta_1^2} s v}.$$

Ainsi,
$$\lim_{\lambda \rightarrow \frac{1}{\beta_1} - 0} \int_0^1 \frac{1 - \tilde{B}(\lambda(1 - e^{-\varepsilon s})v)}{\tilde{B}(\lambda(1 - e^{-\varepsilon s})v) - 1 + (1 - e^{-\varepsilon s})v} \cdot (e^{-\varepsilon s} - 1) dv$$

$$= - \int_0^1 \frac{s}{1 + \frac{\beta_2}{2\beta_1^2} s v} dv = - \frac{2\beta_1^2}{\beta_2} \ln \left(1 + \frac{\beta_2}{2\beta_1^2} s \right). \quad (4.7)$$

4.3 Taux faible des rappels

Le théorème suivant décrit la distribution du nombre de clients on orbite lorsque le taux de rappels est faible: $\theta \rightarrow 0$.

Théorème 4.2 Si $\beta_2 < \infty$, alors pour $\theta \rightarrow 0$ le nombre de clients on orbite suit asymptotiquement une loi normale de l'espérance $\frac{\lambda\rho}{(1-\rho)\theta}$ et de la variance

$$\frac{\lambda^3\beta_2 + 2\lambda\rho - 2\lambda\rho^2}{2(1-\rho)^2\theta}.$$

Preuve:

Soit

$$N_0^*(u) = \frac{N_0(u) - \frac{\lambda\rho}{(1-\rho)\theta}}{\frac{1}{\sqrt{\theta}}} = \sqrt{\theta} N_0(u) - \frac{\lambda\rho}{(1-\rho)\sqrt{\theta}}.$$

La fonction caractéristique $E[e^{itN_0^*(u)}]$ peut être exprimée sous forme de la fonction génératrice $P(z)$ comme suit :

$$\begin{aligned} E[e^{itN_0^*(u)}] &= P(e^{it\sqrt{\theta}}) \cdot e^{-it\frac{\lambda\rho}{(1-\rho)\sqrt{\theta}}} \\ &= \exp\left\{\frac{\lambda}{\theta} \int_1^{e^{it\sqrt{\theta}}} \frac{1-K(u)}{K(u)-u} du - it\frac{\lambda\rho}{(1-\rho)\sqrt{\theta}}\right\} \\ &\quad \times (1-\rho) \frac{1-e^{it\sqrt{\theta}}}{K(e^{it\sqrt{\theta}})-e^{it\sqrt{\theta}}}. \end{aligned} \quad (4.8)$$

Si $\theta \rightarrow 0$, alors $z = e^{it\sqrt{\theta}} \rightarrow 1$. Ainsi $(1-\rho) \frac{1-e^{it\sqrt{\theta}}}{K(e^{it\sqrt{\theta}})-e^{it\sqrt{\theta}}} \rightarrow 1$.

Pour calculer

$$\lim_{\theta \rightarrow 0} \exp\left\{\frac{\lambda}{\theta} \int_1^{e^{it\sqrt{\theta}}} \frac{1-K(u)}{K(u)-u} du - it\frac{\lambda\rho}{(1-\rho)\sqrt{\theta}}\right\} \times (1-\rho) \frac{1-e^{it\sqrt{\theta}}}{K(e^{it\sqrt{\theta}})-e^{it\sqrt{\theta}}}, \text{ il}$$

faut transformer l'argument de la fonction \exp comme suit:

$$\begin{aligned} &\frac{\lambda}{\theta} \int_1^{e^{it\sqrt{\theta}}} \frac{1-k(u)}{k(u)-u} du - it\frac{\lambda\rho}{(1-\rho)\sqrt{\theta}} \\ &= \frac{\lambda}{\theta} \int_1^{e^{it\sqrt{\theta}}} \left(\frac{1-K(u)}{K(u)-u} - \frac{\rho}{1-\rho}\right) du + \frac{\lambda\rho}{1-\rho} \cdot \frac{e^{it\sqrt{\theta}} - 1 - it\sqrt{\theta}}{\theta}. \end{aligned}$$

$$\text{On a: } \lim_{\theta \rightarrow 0} \frac{\lambda\rho}{1-\rho} \frac{e^{it\sqrt{\theta}} - 1 - it\sqrt{\theta}}{\theta} = -\frac{\lambda\rho t^2}{2(1-\rho)}. \quad (4.9)$$

Pour calculer $\lim_{\theta \rightarrow 0} \frac{\lambda}{\theta} \int_1^{e^{it\sqrt{\theta}}} \left(\frac{1-K(u)}{K(u)-u} - \frac{\rho}{1-\rho} \right) du$, nous introduisons la fonction suivante:

$$f(\theta) = \int_1^{e^{it\sqrt{\theta}}} \left(\frac{1-K(u)}{K(u)-u} - \frac{\rho}{1-\rho} \right) du.$$

Il est clair que :

- a) $f(0) = 0$;
- b) $f'(0) = - \frac{\lambda^2 \beta_2}{4(1-\rho)^2} t^2$.

De cette manière, lorsque $\theta \rightarrow 0$, on obtient :

$$f(\theta) = f(0) + \theta f'(0) + o(\theta) = - \theta \frac{\lambda^2 \beta_2}{4(1-\rho)^2} t^2 + o(\theta).$$

$$\text{Donc, } \frac{\lambda}{\theta} \int_1^{e^{it\sqrt{\theta}}} \left(\frac{1-K(u)}{K(u)-u} - \frac{\rho}{1-\rho} \right) du = \frac{\lambda}{\theta} f(\theta) = - \frac{\lambda^3 \beta_2}{4(1-\rho)^2} t^2 + o(1).$$

Par conséquent,

$$\lim_{\theta \rightarrow 0} E[e^{itN_0^*(u)}] = \exp \left\{ - \frac{t^2}{2} \cdot \frac{\lambda^3 \beta_2 + 2\lambda\rho - 2\lambda\rho^2}{2(1-\rho)^2} \right\}, \text{ où}$$

$$\exp \left\{ - \frac{t^2}{2} \cdot \frac{\lambda^3 \beta_2 + 2\lambda\rho - 2\lambda\rho^2}{2(1-\rho)^2} \right\}, \text{ est la fonction caractéristique d'une variable}$$

aléatoire normale ayant l'espérance mathématique égale à 0 et la variance égale à $\frac{\lambda^3 \beta_2 + 2\lambda\rho - 2\lambda\rho^2}{2(1-\rho)^2}$.

4.4 Taux fort des rappels

Dans les situations réelles, les abonnés des réseaux de téléphone répètent leurs appels immédiatement. Alors, une investigation sur le comportement asymptotique d'un système de files d'attente avec rappels sous l'intensité forte de répétition est d'un intérêt pratique spécial. Habituellement, comme $\theta \rightarrow \infty$, une caractéristique stationnaire $a(\theta)$ du système de files d'attente avec rappels converge à une limite $a(\infty)$, qui est la caractéristique stationnaire correspondante d'un certain système «limite». Ce dernier peut être facilement identifié. Par exemple, quand $\theta \rightarrow \infty$, le modèle générale M/G/1 avec rappels peut être considéré comme le système ordinaire M/G/1/ ∞ . Par conséquent, si on dénote par $q_n(\infty)$ la distribution du nombre de clients dans la file d'attente associée au système M/G/1/ ∞ classique, alors:

$$q_n^{(\infty)} = \lim_{\theta \rightarrow \infty} q_n, \text{ où } q_n = \lim_{t \rightarrow \infty} p(N_0(t) = n) \quad (4.10)$$

Cette formule est équivalente à la limite correspondante pour les fonctions génératrices :

$$\lim_{\theta \rightarrow \infty} P(z) = P^{(\infty)}(z).$$

Vu que $P^{(\infty)}(z) = (1 - \rho) \frac{(1-z)\tilde{B}(\lambda-\lambda z)}{\tilde{B}(\lambda-\lambda z)-z}$ (la formule de Pollaczek-Khinchine), cette relation s'ensuit de la formule (2.17) pour $P(z)$.

La distribution $q_n^{(\infty)}$ est étudié en détail est peut être considérée comme ordinaire (comme Gamma, Gaussien, etc.).

Le résultat suivant du taux de convergence de la distribution q_n à la distribution $q_n^{(\infty)}$ est plus intéressante que la relation précédente (4.10).

Théorème 4.3. *Quand $\theta \rightarrow \infty$ la distance*

$$\sum_{n=0}^{\infty} |q_n - q_n^{(\infty)}|$$

entre les distributions q_n et $q_n^{(\infty)}$ est $o(\frac{1}{\theta})$. Pour être plus exacte, on a les inégalités suivantes :

$$\begin{aligned} & 2 \frac{1-\rho}{\tilde{B}(\lambda)} \left[1 - \exp \left\{ -\frac{\lambda}{\theta} \int_0^1 \frac{1-K(u)}{K(u)-u} du \right\} \right] \\ & \leq \sum_{n=0}^{\infty} |q_n - q_n^{(\infty)}| \\ & \leq 2 \left[1 - \exp \left\{ -\frac{\lambda}{\theta} \int_0^1 \frac{1-K(u)}{K(u)-u} du \right\} \right]. \end{aligned} \quad (4.11)$$

Preuve:

La preuve est basée sur la propriété de décomposition stochastique du nombre de clients en orbite:

$$N_{0\theta} = N_{f\infty} + R_{\theta},$$

qui peut être présentée sous forme des distributions suivantes :

$$q_n = \lim_{t \rightarrow \infty} p(N_0(t) = n); \quad q_n^{(\infty)} = \lim_{t \rightarrow \infty} p(N_{f\infty}(t) = n);$$

$$\hat{q}_n = \lim_{t \rightarrow \infty} p(R_{\theta}(t) = n)$$

Où $R_{\theta}(t)$ est le nombre de clients en orbite étant donné que le serveur est libre:

$$\hat{q}_n = \lim_{t \rightarrow \infty} p(N_0(t) = n / C(t) = 0) = \frac{p_{0n}}{1-\rho}.$$

Cette propriété veut dire que q_n est une convolution des distributions $q_n(\infty)$ et \hat{q}_n :

$$q_n = \sum_{k=0}^n q_k(\infty) \hat{q}_{n-k} . \quad (4.12)$$

La fonction génératrice de \hat{q}_n est définie par:

$$\sum_{n=0}^{\infty} z^n \hat{q}_n = \exp \left\{ \frac{\lambda}{\theta} \int_1^z \frac{1-K(u)}{K(u)-u} \right\} du . \quad (4.13)$$

A présent, la formule (4.12) devienne :

$$q_n - q_n(\infty) = q_n(\infty)\hat{q}_0 - q_n(\infty) + \sum_{k=0}^{n-1} q_k(\infty) \hat{q}_{n-k} .$$

On a donc,

$$\begin{aligned} |q_n - q_n(\infty)| &\leq |q_n(\infty)\hat{q}_0 - q_n(\infty)| + \sum_{k=0}^{n-1} q_k(\infty)\hat{q}_{n-k} \\ &= q_n(\infty)(1 - \hat{q}_0) + q_n - q_n(\infty)\hat{q}_0 \\ &= q_n(\infty)(1 - 2\hat{q}_0) + q_n . \end{aligned}$$

Par conséquent,

$$\sum_{n=0}^{\infty} |q^n - q_n(\infty)| \leq (1 - 2\hat{q}_0) \sum_{n=0}^{\infty} q_n(\infty) + \sum_{n=0}^{\infty} q_n .$$

Vu que $\sum_{n=0}^{\infty} q_n(\infty) = 1$ et $\sum_{n=0}^{\infty} q_n = 1$,

$$\sum_{n=0}^{\infty} |q_n - q_n(\infty)| \leq 2(1 - \hat{q}_0) \quad (4.14)$$

En posant $z = 0$ dans l'équation (4.13), on obtient la formule suivante pour \hat{q}_0 :

$$\hat{q}_0 = \exp \left\{ - \frac{\lambda}{\theta} \int_0^1 \frac{1-K(u)}{K(u)-u} \right\} du . \quad (4.15)$$

A présent, de (4.14) et (4.15) on obtient la première partie de l'inégalité (4.11).

A l'aide de $|a - b| \geq a - b$, nous calculons

$$\begin{aligned} \sum_{n=0}^{\infty} |q_n - q_n(\infty)| &= |q_0 - q_0(\infty)| + \sum_{n=1}^{\infty} |q_n - q_n(\infty)| \\ &\geq |q_0 - q_0(\infty)| + \sum_{n=1}^{\infty} (q_n - q_n(\infty)) \\ &= |q_0 - q_0(\infty)| + q_0(\infty) - q_0 . \end{aligned}$$

Cependant,

$$q_0(\infty) = \frac{1 - \rho}{\tilde{B}(\lambda)} ,$$

$$q_0 = q_0(\infty)\hat{q}_0 < q_0(\infty).$$

Par conséquent,

$$|q_0 - q_0(\infty)| = q_0(\infty) - q_0 = q_0(\infty) (1 - \hat{q}_0).$$

$$\sum_{n=0}^{\infty} |q_n - q_n(\infty)| \geq 2(q_0(\infty) - q_0) = 2 \frac{1-\rho}{\bar{B}(\lambda)} (1 - \hat{q}_0).$$

La probabilité \hat{q}_0 peut être estimée à l'aide d'une approche basée sur les ordres stochastiques sous forme simple. En particulier, on a toujours [15]:

$$1 - \hat{q}_0 \leq 1 - \exp \left\{ - \frac{\lambda}{\theta} \int_0^1 \frac{1 - e^{-\rho(1-u)}}{e^{-\rho(1-u)} - u} du \right\}$$

4.5 Conclusion:

Dans ce chapitre, les investigations ont été portées sur le comportement asymptotique de la variable aléatoire représentant le nombre de clients en orbite, et ceci pour le modèle M/G/1 avec rappels. Il a été établi que la variable en question se comporte comme une variable de loi Gamma lorsque l'intensité du trafic $\rho \rightarrow 1$ et comme une variable normale lorsque le taux de rappels $\theta \rightarrow 0$.

Nous avons également estimé la proximité entre le modèle M/G/1 avec rappels et celui ordinaire correspondant.

Chapitre 5

Application

5.1 Introduction:

La majorité des études sur les modèles de files d'attente avec rappels considère le système M/G/1 avec rappels, où le calcul de la fonction génératrice de la taille de l'orbite soit disponible. L'expression de cette fonction ne permet pas d'étudier le caractère de la distribution du nombre de clients en orbite dans les cas qui nous avons analysé dans le chapitre précédent.

Dans ce chapitre, nous appliquons les résultats obtenus sur les systèmes M/M/1, M/E₂/1 et M/H₂/1 avec rappels.

5.2 Transformées de Laplace-Stieltjes pour les systèmes de files d'attente classiques

5.2.1 Système M/M/1

Ce système est décrit par le processus des arrivées poissonniennes de taux λ , la loi exponentielle de service de taux γ , la file d'attente de capacité infinie. Les fonctions de densité de probabilité et de répartition sont données par:

$$b(t) = \gamma e^{-\gamma t} \text{ et } B(t) = 1 - e^{-\gamma t}, t \geq 0.$$

Trouvons TL-S :

$$\begin{aligned}\tilde{B}(t) &= \tilde{B}(\lambda - \lambda Z) = \int_0^\infty e^{-(\lambda - \lambda Z)t} \gamma e^{-\gamma t} dt = \gamma \int_0^\infty e^{-(\lambda - \lambda Z + \gamma)t} dt \\ &= \frac{\gamma}{\lambda(1-Z) + \gamma} = \frac{1}{\frac{\lambda}{\gamma}(1-Z) + 1}.\end{aligned}$$

D'où

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

La 1^{ère} équation de Pollaczek-Khintchine pour le nombre de clients dans le système est définie par:

$$\pi(Z) = \frac{\tilde{B}(\lambda - \lambda Z)(1-\rho)(Z-1)}{Z - \tilde{B}(\lambda - \lambda Z)} = \frac{(1-\rho)(Z-1)}{Z[1 + \frac{\rho(1-Z)}{2}]^2 - 1}.$$

5.2.2 Système M/E₂/1

Dans ce système, la durée de service suit une loi d'Erlang d'ordre 2 et de moyenne finie $\frac{1}{\gamma}$. $B(t) = 1 - \gamma t e^{-\gamma t} - e^{-\gamma t}$ et $b(t) = \gamma^2 t e^{-\gamma t}$ sont respectivement la fonction de répartition et la fonction de densité de probabilité. $\tilde{B}(t)$ est TL-S telle que:

$$\begin{aligned}\tilde{B}(t) &= \tilde{B}(\lambda - \lambda Z) = \int_0^\infty e^{-(\lambda - \lambda Z)t} \gamma^2 t e^{-\gamma t} dt \\ &= \gamma^2 \int_0^\infty e^{-(\lambda - \lambda Z)t} t e^{-\gamma t} dt \\ &= \gamma^2 \int_0^\infty t e^{-(\lambda - \lambda Z + \gamma)t} dt.\end{aligned}$$

On utilise l'intégration par partie.

Posons:

$$U(t) = t \text{ et } V'(t) = e^{-(\lambda - \lambda Z + \gamma)t}.$$

$$\text{Alors, } U'(t) = dt \text{ et } V(t) = \frac{-1}{\lambda - \lambda Z + \gamma} e^{-(\lambda - \lambda Z + \gamma)t}.$$

Donc,

$$\tilde{B}(\lambda - \lambda Z) = \frac{\gamma^2}{\lambda - \lambda Z + \gamma} \int_0^\infty e^{-(\lambda - \lambda Z + \gamma)t} dt = \frac{\gamma^2}{[\lambda(1-Z) + \gamma]^2}$$

D'où

$$\tilde{B}(\lambda - \lambda Z) = \left[\frac{\gamma}{\lambda(1-Z) + \gamma} \right]^2 = \left[\frac{1}{\rho(1-Z) + 1} \right]^2.$$

La 1^{ère} équation de Pollaczek-Khintchine pour le nombre de clients dans le système est donnée par:

$$\pi(Z) = \frac{(1-\rho)(Z-1)}{Z\left[1+\frac{\rho(1-Z)}{2}\right]^2-1}.$$

5.2.3 Système M/H₂/1

La durée de service suit une loi hyperexponentielle d'ordre 2 dont la fonction de répartition est définie par: $B(t) = 1 - p_1 e^{-\gamma_1 t} - p_2 e^{-\gamma_2 t}$ où p_1, p_2, γ_1 et γ_2 vérifient:

$$p_1 + p_2 = 1 \text{ et } \frac{1}{\gamma_1} + \frac{1}{\gamma_2} = \frac{1}{\gamma} \text{ et } \frac{1}{\gamma} \text{ est la durée moyenne de service.}$$

Calculons TL-S:

$$\begin{aligned} \tilde{B}(t) &= \tilde{B}(\lambda - \lambda Z) = \int_0^\infty e^{-(\lambda - \lambda Z)t} dB(t) \\ &= p_1 \gamma_1 \int_0^\infty e^{-(\lambda - \lambda Z + \gamma_1)t} dt + p_2 \gamma_2 \int_0^\infty e^{-(\lambda - \lambda Z + \gamma_2)t} dt \\ &= \frac{p_1 \gamma_1}{\lambda(1-Z) + \gamma_1} + \frac{p_2 \gamma_2}{\lambda(1-Z) + \gamma_2}. \end{aligned}$$

Alors,

$$\tilde{B}(t) = \frac{p_1}{\rho_1(1-Z) + 1} + \frac{p_2}{\rho_2(1-Z) + 1}.$$

Par conséquent,

$$\pi(Z) = \frac{(1-\rho)[1+(\rho_1+\rho_2-\rho)(1-Z)]}{\rho_1 \rho_2 Z^2 - (\rho_1 + \rho_2 + \rho_1 \rho_2 Z) + 1 + \rho_1 + \rho_2 - \rho}$$

$$\text{où } \rho_1 = \frac{\lambda}{\gamma_1} \text{ et } \rho_2 = \frac{\lambda}{\gamma_2} \text{ et } \rho = \frac{\lambda}{\gamma}.$$

5.3 Transformées de Laplace- Stieltjes pour les systèmes de files d'attente avec rappels

5.3.1 Système M/M/1 avec rappels

La fonction de répartition est définie par: $B(t) = 1 - e^{-\gamma t}$, $t \geq 0$.

Alors,

$$\begin{aligned} \tilde{B}(t) &= \tilde{B}(\lambda - \lambda Z) = \int_0^\infty e^{-(\lambda - \lambda Z)t} dB(t) = \gamma \int_0^\infty e^{-(\lambda - \lambda Z + \gamma)t} dt \\ &= \frac{\gamma}{\lambda(1-Z) + \gamma} = \frac{1}{\frac{\lambda}{\gamma}(1-Z) + 1} \end{aligned}$$

Où $\rho = \lambda \beta_1$ et $\gamma = \frac{1}{\beta_1}$.

Donc,

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

5.3.2 Système M/E₂/1 avec rappels

La fonction de répartition est définie par: $B(t) = 1 - e^{-\gamma t} - \gamma t e^{-\gamma t}$, $t \geq 0$.

Trouvons TL-S:

$$\begin{aligned} \tilde{B}(t) &= \tilde{B}(\lambda - \lambda Z) = \int_0^\infty e^{-(\lambda - \lambda Z)t} dB(t) \\ &= \int_0^\infty e^{-(\lambda - \lambda Z)t} \gamma^2 t e^{-\gamma t} dt \end{aligned}$$

On utilise l'intégration par partie, on obtient:

$$\tilde{B}(\lambda - \lambda Z) = \left[\frac{1}{\rho(1-Z)+1} \right]^2.$$

5.3.3 Système M/H₂/1 avec rappels

L'expression de la fonction de répartition est donnée par:

$$B(t) = 1 - \zeta_1 e^{-\gamma_1 t} - \zeta_2 e^{-\gamma_2 t}, \quad t \geq 0. \quad \text{Où } \zeta_1 + \zeta_2 = 1 \text{ et } \frac{\zeta_1}{\gamma_1} + \frac{\zeta_2}{\gamma_2} = \frac{1}{\gamma}.$$

Calculons TL-S:

$$\begin{aligned} \tilde{B}(t) &= \tilde{B}(\lambda - \lambda Z) = \int_0^\infty e^{-(\lambda - \lambda Z)t} dB(t) \\ &= \zeta_1 \gamma_1 \int_0^\infty e^{-(\lambda - \lambda Z + \gamma_1)t} dt + \zeta_2 \gamma_2 \int_0^\infty e^{-(\lambda - \lambda Z + \gamma_2)t} dt \\ \tilde{B}(\lambda - \lambda Z) &= \frac{\zeta_1 \gamma_1}{\lambda(1-Z) + \gamma_1} + \frac{\zeta_2 \gamma_2}{\lambda(1-Z) + \gamma_2} \end{aligned}$$

On trouve:

$$\tilde{B}(\lambda - \lambda Z) = \frac{\zeta_1}{\frac{\lambda}{\gamma_1}(1-Z)+1} + \frac{\zeta_2}{\frac{\lambda}{\gamma_2}(1-Z)+1}.$$

5.4 Comportement asymptotique des systèmes avec rappels:

5.4.1 Système M/M/1 avec rappels

$$\text{On a: } \tilde{B}(\lambda - \lambda Z) = \frac{1}{\rho(1-Z)+1} \text{ avec } cv = 1$$

$$\text{Alors, } \tilde{B}'(\lambda - \lambda Z) = \frac{\rho}{[\rho(1-Z)+1]^2}$$

$$\text{Donc, } \tilde{B}''(\lambda - \lambda Z) = \frac{2\rho^2}{[\rho(1-Z)+1]^3} .$$

Cela implique que:

$$\beta_2 = 2\rho^2$$

Quand $\theta \longrightarrow 0$, la distribution du nombre de clients en orbite est de moyenne $m = \frac{\lambda\rho}{(1-\rho)\theta}$ et de variance $Var = \frac{\lambda^3\rho^2 + \lambda\rho - \lambda\rho^2}{(1-\rho)^2\theta}$.

5.4.2 Système M/E₂/1 avec rappels

Le coefficient de variation est donné par $cv \approx 0.7$.

A l'aide de TL-S, on obtient:

$$\tilde{B}'(\lambda - \lambda Z) = \frac{2\rho}{[\rho(1-Z)+1]^3}$$

Alors,

$$\tilde{B}''(\lambda - \lambda Z) = \frac{6\rho^2}{[\rho(1-Z)+1]^4}$$

Donc,

$$\beta_2 = 6\rho^2$$

Si: $\theta \longrightarrow 0$, la distribution du nombre de clients en orbite est de moyenne

$$m = \frac{\lambda\rho}{(1-\rho)\theta} \text{ et de variance } Var = \frac{3\lambda^3\rho^2 + \lambda\rho - \lambda\rho^2}{(1-\rho)^2\theta} .$$

5.4.3 Système M/H₂/1 avec rappels

La TL-S de se système donne:
$$\tilde{B}'(\lambda - \lambda Z) = \frac{\zeta_1 \frac{\lambda}{\gamma_1}}{[\frac{\lambda}{\gamma_1}(1-Z)+1]^2} + \frac{\zeta_2 \frac{\lambda}{\gamma_2}}{[\frac{\lambda}{\gamma_2}(1-Z)+1]^2} .$$

Alors,

$$\tilde{B}''(\lambda - \lambda Z) = \frac{2\zeta_1 \frac{\lambda^2}{\gamma_1^2}}{[\frac{\lambda}{\gamma_1}(1-Z)+1]^3} + \frac{2\zeta_2 \frac{\lambda^2}{\gamma_2^2}}{[\frac{\lambda}{\gamma_2}(1-Z)+1]^3} .$$

Donc,

Dans le cas d'une distribution hyperexponentielle de degré 2, on a:

$$\zeta_1 = \frac{1}{K+1} \quad \text{et} \quad \zeta_2 = \frac{K}{K+1} ,$$

où $K = (cv)^2 + [(cv)^4 - 1]^{\frac{1}{2}}$. Ainsi, si $cv = 1.5$, on obtient: $\zeta_1 = 0.19$ et $\zeta_2 = 0.81$.

Le tableau 5.1 présente les valeurs de la moyenne et la variance pour les modèles M/M/1 et M/E₂/1 avec rappels ainsi que pour le système M/H₂/1 avec rappels et $cv = 1.5$, lorsque $\theta \rightarrow 0$. A cet effet, nous utilisons les résultats fournis par le théorème 4.2. On choisit $\gamma = 1$ et $\lambda \in \{0.3; 0.5; 0.8\}$.

Tableau 5.1 Modèles avec rappels

ρ	θ	E	E ₂	H ₂
		$cv = 1$	$cv \approx 0.7$	$cv = 1.5$
		m	m	m
		Var	Var	Var
0.3	0.1	1.2857	1.2857	1.2857
		1.3353	1.4344	1.451
	0.01	12.8571	12.8571	12.8571
		13.353	14.3448	14.5102
	0.001	128.5714	128.5714	128.5714
		133.5306	143.4489	145.102
0.5	0.1	5	5	5
		6.25	8.75	7.5
	0.01	50	50	50
		62.5	87.5	75
	0.001	500	500	500
		625	875	750
0.8	0.1	32	32	32
		113.92	277.76	134.4
	0.01	320	320	320
		1139.2	277.76	134.4
	0.001	3200	3200	3200
		11392	27776	1344

Quand $\theta \rightarrow 0$, la moyenne et la variance dépendent de la distribution du temps de service (de son coefficient de variation). Lorsque le coefficient de variation est strictement supérieur ou strictement inférieur à 1, on observe une augmentation des valeurs numériques des caractéristiques en question. Avec accroissement du taux des rappels ainsi que de la charge du système (ρ), les valeurs de la moyenne et de la variance subissent une augmentation.

1) L'augmentation du taux des rappels entraîne la convergence plus rapide vers le système M/M/1 classique.

2) Le taux de convergence vers le modèle M/M/1 classique diminue avec l'augmentation de la charge du système (ρ).

Conclusion générale

Dans ce travail, nous avons analysé le comportement asymptotique du nombre de clients en orbite associé au système M/G/1 avec rappels.

En premier lieu, nous avons passé en revue certains résultats consacrés aux systèmes de files d'attente classiques. Ces derniers ne prennent pas en considération le phénomène de répétition de demandes de service : le phénomène en question est étudié par la théorie des files d'attente avec rappels.

En deuxième lieu, nous avons présenté l'une des approches permettant de faciliter les investigations sur les systèmes avec rappels, celle basée sur leur propriété de décomposition stochastique.

En suite, nous avons étudié le système de files d'attente M/G/1 avec rappels dans des cas particuliers : régime chargé, intensité faible des rappels et intensité forte des rappels.

Annexes

Annexe A :

- 1- Processus de Poisson.
- 2- Loi exponentielle.
- 3- Loi d'Erlang.
- 4- Loi hyperexponentielle.

Annexe B :

- 1- Fonction génératrice.
- 2- Transformée de Laplace-Stieltjes.

Annexe A

1 Processus de Poisson

Le processus en question est utilisé pour décrire la réalisation dans le temps d'évènements aléatoires d'un type donné. La description mathématique d'un flux d'évènements aléatoires peut se faire de deux manières différentes:

1. On considère le nombre d'évènements $X(t)$ se produisant dans $[0,t]$ et on cherche à déterminer la loi de probabilité de cette variable aléatoire discrète. Le processus $\{X(t), t \geq 0\}$ est appelé *processus de comptage*.
2. On considère les intervalles de temps qui séparent les instants d'apparition de deux évènements consécutifs. Ce sont des variables aléatoires continues, positives et en général indépendantes et identiquement distribuées.

On dit qu'un processus de comptage $\{X(t), t \geq 0\}$ est un processus de Poisson s'il satisfait aux 3 conditions suivantes:

- Le processus est homogène dans le temps: la probabilité d'avoir k évènements dans un intervalle de longueur t ne dépend que de t et non pas de la position de l'intervalle par rapport à l'axe temporel: $p_k(t) = P(X(t) = k)$.
- Pour tout système d'intervalles disjoints, les nombres d'évènements s'y produisant sont des variables aléatoires indépendantes.

$$\text{- La probabilité } p_k(\Delta t) = \begin{cases} 0(\Delta t) & \text{si } k \geq 2 \\ \lambda \Delta t + 0(\Delta t) & \text{si } k = 1, \text{ où } \lambda \text{ est la densité ou intensité} \\ 1 - \lambda \Delta t + 0(\Delta t) & \text{si } k = 0 \end{cases}$$

du processus (le nombre moyen d'évènements qui apparaissent par unité de temps).

Théorème 1: *Pour un processus de Poisson, on a:*

$$P(X(t) = k) = p_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \lambda > 0, k \geq 0 ;$$

$$E[X(t)] = \lambda t \text{ et } \text{Var}[X(t)] = \lambda t.$$

Ces relations définissent le régime transitoire du processus de Poisson. Aucun régime stationnaire n'existe vu que $p_k = \lim_{t \rightarrow \infty} p_k(t) = 0, \forall k \geq 0$.

Théorème 2: *Le temps V qui sépare un instant quelconque du prochain évènement est une variable aléatoire répartie selon une loi $\text{esp}(\lambda)$.*

2 Loi exponentielle

La densité de probabilité d'une loi exponentielle de paramètre λ est donnée par:

$f(t) = \lambda e^{-\lambda t}$, $t \geq 0$. La fonction de répartition est de la forme: $F(t) = 1 - e^{-\lambda t}$, $t \geq 0$.

3 Loi d'Erlang

Par définition, la loi d'Erlang $E_k(\lambda)$ est la loi suivie par la somme de k variables aléatoires indépendantes chacune distribuées selon la loi exponentielle de paramètre λ . Ainsi, si des évènements aléatoires indépendants arrivent selon un processus de Poisson, alors la date t_k d'arrivée du $k^{\text{ième}}$ évènement est gouvernée par la loi d'Erlang.

Les fonctions de densité de probabilités et de répartition sont données par :

$$f(t) = \frac{\lambda^k t^{k-1} e^{-\lambda t}}{(k-1)!} \quad \text{et} \quad F(t) = 1 - e^{-k\lambda t} \sum_{j=0}^{k-1} \frac{(k\lambda t)^j}{j!}, \quad t \geq 0.$$

4 Loi hyperexponentielle

Soit $\lambda_1, \dots, \lambda_n$, n réels positifs différents. Une variable aléatoire X continue à valeurs dans \mathbb{R}^+ suit une loi hyperexponentielle d'ordre n et de paramètres $\lambda_1, \dots, \lambda_n$ si sa fonction de densité de probabilités est donnée par:

$$f(t) = \sum_{i=1}^n a_i \lambda_i e^{-\lambda_i t} \mathbf{1}_{\mathbb{R}^+}(t) \quad \text{où} \quad a_i = \prod_{j=1, j \neq i}^n \frac{\lambda_j}{\lambda_j - \lambda_i} \quad \text{et} \quad i = 1, \dots, n.$$

Annexe B

1 Fonction génératrice

Soit X une variable aléatoire discrète non négative telle que $P(X = n) = p(n)$, $n \in \mathbb{N}$.

La fonction génératrice $P_X(z)$ de X est définie par :

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

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

$$P_X(0) = p(0), P_X(1) = 1, P_X'(1) = E(X).$$

Et d'une manière générale,

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

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

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

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

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

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

2 Transformée de Laplace-Stieltjes

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

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

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

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

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

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

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

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

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

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

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

Bibliographie

- [1] A. Aissani et J.R.Artalejo. On the single server retrial queue subject to breakdowns. *Queueing Systems*, 30 (3-4) : 309-321, 1998.
- [2] A. Aissani. A servery on retrial queueing models. *Actes des Journées Statistiques Appliquées*. U.S.T.H.B, Alger : 1-11, 1994.
- [3] J.R. Artalejo. Analysis of an M/G/1 queue with constant repeated attempts and server vacations. *Computers and Operations Research*, 24(6) : 493-504, 1997.
- [4] J.R. Artalejo et G.I. Falin. Stochastic decomposition for retrial queues. *TOP*, 2:329-342, 1994.
- [5] U.Narayan Bhat. An introduction to queueing theory. *Modeling and analysis in applications*. Mathematics Subject Classification, (2000).
- [6] Buzacott et Shanthikunrar. *Stochastic Models of Manufacturing Systems*. Preutice Hall, New Jersey, 1984.
- [7] J.W. Cohen. *The Single Server Queue*. North Holland, 2th ed , 1982.
- [8] B.D. Choi, K.B Choi et Y.W. Lee. M/G/1 retrial queueing system with tow types of calls and finite capacity. *Queueing Systems*, 19 : 215-229 (1995).
- [9] B.D. Choi et D.B. Zhu. The $M_1, M_2/G/1/k$ retrial queueing systems with priority. *Journal of the Korean Mathematical Society*. 35 : 691-712, 1998.
- [10] B.D. Choi. V.G. Kulkarni. Feedback retrial queueing systems. *Queueing and Related Models* : 93-105, 1992.
- [11] B.D. Choi. Y.C. Kim et Y.W. Lee. The M/M/c retrial queue with geometric loss and feedback . *Computer Appl*, 36(6) : 41-52, 1998.
- [12] Q.H. Choo et B. Conolly. New results in the theory of repeated orders queueing systems. *Journal of Applied Probability*, 16 : 631-640, 1979.
- [13] N.V. Djellab. M/G/1 retrial queue with random breakdowns and general retrial tims. *Procedings of the Int. Conference " Modern mathematical methods of investigating of the information networks"*. Minsk, pp 73-78, 2001.

- [14] N.V. Djellab. On the decomposition property of M/G/1 retrial queues. Proceedings of the Int. Conference " Modern mathematical methods of analysis and optimization of telecommunication networks". *Gomel Belarusse*, pp 75-79, 2003.
- [15] B.T. Doshi. Queueing systems with vacations. A survey. *Queueing Systems*, 1 : 29-66, 1986.
- [16] G.I. Falin. The exit flow of a single-line queueing system when there are secondary order. *Queueing Theory and probability*, 46-67, 1979.
- [17] G.I. Falin. Single-line repeated orders queueing systems. *Mathematische Operationsfor-
chung und Statistik, Optimization*, 5: 649-667, 1986.
- [18] G.I. Falin. A survey of retrial queues. *Queueing Systems*, 7 : 127-168, 1990.
- [19] G.I. Falin. A single line system with secondary orders. *Engineering Cybernetics
Review*, 17(2) : 76-93, 1979.
- [20] G.I. Falin. The output flow of a single line queueing system when there are secondary orders. *Engineering Cybernetics Review*, 16(5) : 64-67, 1978.
- [21] G.I. Falin et J.R. Artalejo. A finite source retrial queue. *European Journal of
Operational Research*, 108 : 409-424, 1998.
- [22] G.I. Falin et J.R. Artalejo et M. Martin. On the single server retrial queue with priority customers. *Queueing Systems*, 14 : 439-455, 1993.
- [23] G.I. Falin et C. Fricker. On the virtual waiting time in an M/G/1 retrial queue. *Journal of Applied Probability*, 28 : 446-460, 1991.
- [24] G.I. Falin et J.G.C. Templeton. *Retrial Queues*. Chapman and Hall, 1997.
- [25] A. Federgruen et L.Green. Queueing systems with service interruptions. *Research
Working. Columbia University*, 30 : 5-84, 1984.
- [26] A.A. Fredericks et G.A. Reisner. Approximations to stochastic service systems with an application to a retrial model. *The Bell System Technical Journal*, 58(3) : 557- 577, 1979.
- [27] S. Fuhrmann. A note on the M/G/1 queue with server vacations. *Operations
Research*, 32 : 1368-1373, 1984.
- [28] S. Fuhrmann et R.B. Cooper. Stochastic decomposition in the M/G/1 queue with generalized vacations. *Operations Research*, 33(5) : 1117-1129, 1985.
- [29] D.P. Gaver. A waiting line with interrupted service including priorities. *J. Roy.
Stat. Soc*, B25 : 73-90, 1962.
- [30] E. Gelenbe and R. Iasnogorodski. A queue with server of walking type. *Annales
de l'Institut Henri Poincaré (B)*, 16(1) :63-73, 1980.

- [31] E. Gelenbe et G. Pujole. *Introduction aux Réseaux de files d'attente*. Eyrolles, Paris, 1985.
- [32] S.A. Grishechkin. Multiclass batch arrival retrial queueing analyzed as branching process with immigration. *Queueing Systems*, 11 : 395-418, 1992.
- [33] Q. Gomez-Corral et M.F. Ramalhoto. The statistic analysis of a single server retrial queue with general retrial times. *Naval Research Logistics*, 46 : 561-581, 1999.
- [34] Q. Gomez-Corral et M.F. Ramalhoto. The stationary distribution of markovien process arising in the theory of multiserver retrial queueing systems. *Mathematical and Computer Modeling*, 30 :141-158, 1999.
- [35] D.P. Heymen. Optimal operating policies for M/G/1 queueing systems. *Oper. Res*, 16: 362-382, 1968.
- [36] N.K. Jaiswal. *Priority Queue*. Academic Press; New York, 1968.
- [37] G.I. Jonin et J.J. Sedol. Full-availability groups with repeated calls and time of advanced service. *Proc 7th Int. Teletraffic Congress, Stockholm*, 4 :1-137, 1973.
- [38] J. Keilson, J. Cozzolino et H. Young. A service system with unfilled system requested. *Operations Research*, 16 : 1126-1137, 1968.
- [39] L. Kleinrock. *Queueing Systems (Theory)*. Vol 1, John Wily and sons. Edition, 1975.
- [40] L. Kosten. On the influence of repeated calls in the theory of probabilities of blockings. *De Ingenieur*, 59 :1-25, 1947.
- [41] B. Krishma Kumar, S. Pavoï Madheswari et Vijayakumar. The M/G/1 retrial queue with feedback and starting failures. *Applied Mathematical Modeling*, 26(11) : 1057-1075, 2002.
- [42] A. Krishnamoorthy et P.V. Ushakumari. Reliability of a k-out-of-n system with repair and retrial failed units. *TOP*, 7 : 191-208, 1990.
- [43] V.G. Kulkarni et B.D. Choi. Retrial queue with server subject to breakdown and repairs. *Queueing systems*, 20 : 341-368, 1995.
- [44] C. Langaris et E. Moutzoukis. A retrial queue with structured batch arrivals priorities and server vacations. *Queueing systems*, 20 : 341-368, 1995.
- [45] Y. Levey et U. Yechiali. Utilization for the idle time in an M/G/1 queueing system. *Management Science*, 22 :202-211, 1975.
- [46] L.W. Miller. Alternating priorities in multi-class queues. *Cornell University, Ithaca, N.Y.* 4 : 354-373, 1964.
- [47] E. Moutzoukis et C. Langaris. Non-preemptive priorities and vacations in a multiclass retrial queueing system. *Stochastic Models*, 12 : 455-472, 1996.

- [48] M.F. Neuts. Markov chains with applications in queueing theory, with have a matrix geometric invariant probability vector. *Advances in Applied Probability*, 10 : 185-212, 1978.
- [49] T.J. Ou. On the M/G/1 queue with additional inputs. *J. Appl. Prob*, 21 : 129-142, 1984.
- [50] T. Yang et al. An approximation method for the M/G/1 retrial queue with general retrial times. *European Journal of Operational Research*, 76 : 552-562, 1994.
- [51] T. Yang et J.G.C. Templeton. A survey on retrial queues. *Queueing systems*, 2: 201-233, 1987.