

الجمهورية الجزائرية الديمقراطية الشعبية
République algérienne démocratique et populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'enseignement supérieur et de la recherche scientifique
جامعة عين تموشنت بلحاج بوشعيب
Université –Ain Temouchent- Belhadj Bouchaib
Faculté des Sciences et de Technologie
Département - Mathématiques et Informatique



Projet de Fin d'Etudes
Pour l'obtention du diplôme de Master en : Mathématiques
Domaine : Mathématiques et Informatique
Filière : Mathématiques
Spécialité : Probabilités et Statistique appliquées
Thème

Analyse d'un système de file d'attente M/M/s avec rappels.

Présenté Par :

1) Melle Ourci Nesrine

Devant le jury composé de :

Dr. HAMMOUDI Ahmed	Pr	UAT.B.B (Ain Temouchent)	Président
Dr. MESSABIHI Aicha	MCB	UAT.B.B (Ain Temouchent)	Examineur
Dr. SAKHI Hanane	MCB	UAT.B.B (Ain Temouchent)	Encadrant

Année Universitaire 2023/2024

Remerciement

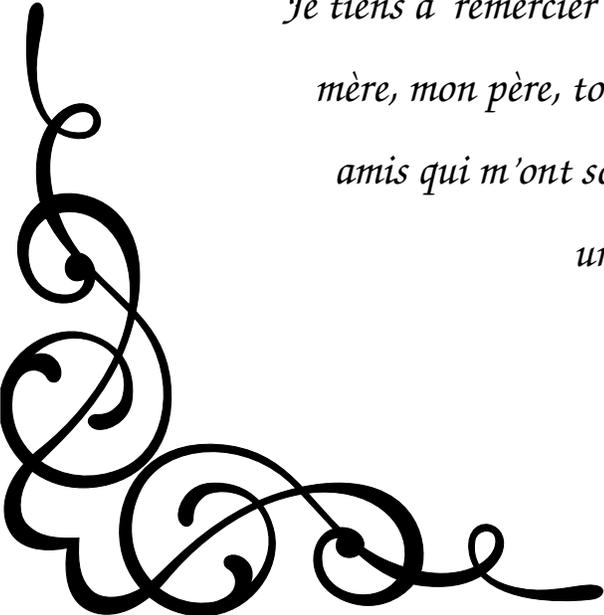


*Je remercie, avant tout Allah le tout puissant de
m'avoir donné la force et la volonté nécessaire
pour achever ce travail.*

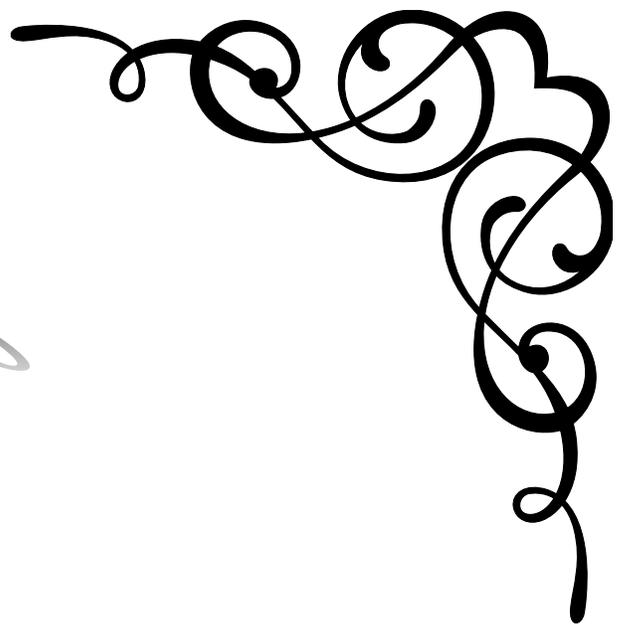
*Je remercie aussi mon encadrante Madame Sakfi
Hanane, pour son aide, ses conseils, son
savoir-faire et ses compétences.*

*Je remercie aussi les membres du jury Dr
Hammoudi Ahmed, et Dr Messabih Aicha de
m'avoir fait l'honneur d'en faire partie et d'avoir
accepté l'évaluation de ce mémoire.*

*Je tiens à remercier mon unique source de joie, ma
mère, mon père, toute ma famille, et toute mes
amis qui m'ont soutenu, durant mon cursus
universitaire.*



Dédicace



*Je dédie ce modeste travail, À mes deux
parents.*

*À mes soeurs, qui m'ont soutenu durant
tout mon cursus universitaire.*

À tous mes amis.

À tous ceux que j'aime.



Table des matières

Introduction	5
I Préliminaires	8
I.1 Description d'un modèle de file d'attente	8
I.2 Classification des systèmes de files d'attente	8
I.3 Analyse mathématique d'un système de files d'attente	9
I.4 Processus de naissance et de mort	9
I.5 Notation de kendall	11
I.6 Discipline de la file d'attente	12
I.7 Mesures de Performances d'une files d'attente	12
I.8 Systèmes de file d'attente Markoviens	13
I.8.1 Modèle d'attente M/M/1	13
I.8.2 Modèle d'attente M/M/s	15
II Les files d'attente avec rappels	19
II.1 Modèle général d'un système de file d'attente avec rappels	19
II.2 Modèles markoviens	20
II.2.1 Modèle M/M/1 avec rappels	20
II.2.2 Modèle M/M/2 avec rappels	22
II.2.3 Modèle M/M/s avec rappels	23
II.3 Les limites fluides d'une classe de processus de Markov	28
II.4 Relation entre renormalisation et ergodicité	31
III Système de file d'attente M/M/s avec deux orbites de reconnects et deux orbites de redials	33
III.1 Le modèle	33

III.2 Les limites fluides	35
III.3 La stationnarité du modèle	42
III.4 Discussion des résultats	43
Conclusion	44
Bibliographie	45

Introduction

Les files d'attente ont une application importante dans un grand nombre de domaines, elles sont également très répandues dans l'évaluation et la conception des réseaux informatique, dans les télécommunications, les réseaux informatique et les réseaux sans fil.

Les files d'attente sont des situations courantes dans de nombreux contextes où les ressources ou les services sont limités et doivent être distribués parmi un grand nombre de demandeurs. La gestion des files d'attente est un domaine d'étude important, souvent abordé dans la théorie des files d'attente, qui analyse comment les files se forment, se comportent et comment elles peuvent être optimisées.

Dans la théorie des files d'attente, un mécanisme dans lequel les clients rejetés reviennent à intervalles aléatoires jusqu'à ce qu'ils soient servis, s'appelle une file d'attente avec rappels. Une file d'attente avec rappels est similaire à tout système de file d'attente ordinaire dans la mesure où on a un processus d'arrivée, et un ou plusieurs serveurs. Les différences élémentaires sont :

- Premièrement, les clients qui rentrent pendant une période d'inactivité ou d'occupation ou des serveurs peuvent tenter à nouveau d'être servis à un moment donné dans le futur.
- Deuxièmement une salle d'attente, qui est connue sous le nom de "salle primaire", n'est pas obligatoire dans le contexte des files d'attente avec rappels. La salle d'attente ordinaire est remplacée par une file d'attente que nous l'appelons dans ce présent travail, "orbite de re-dial", vers laquelle les clients se dirigent après l'échec de la tentative de service, et à partir de laquelle ils tentent à nouveau d'être servis selon une politique probabiliste ou déterministe donnée.

En raison de l'utilité et des propriétés mathématiques intéressantes des modèles de file d'attente, une abondante littérature sur le sujet a vu le jour au cours des dernières décennies. Pour une étude des files d'attente et un résumé de nombreux résultats, le lecteur peut se référer aux travaux de Falin,1990 ; Gharbi Kim [12] & Loualalen,2006 ; Falin & Artalejo,1998 [16] ; Ebrahimi,2006 ; Libman & Orda,2002[20] ; Walrand,1991 ; Choi & Kim,1998[4] et les références qui s'y rapportent.

Le système avec rappels avec deux classes de clients, un seul serveur, pas de salle d'attente, des arrivées par lots et le schéma avec rappels classique a été présenté et analysé dans (Kulkarni,1986 [18]). Ensuite, dans (Falin, 1988[8]), l'auteur a étendu l'analyse du modèle de (Kulkarni,1986[18]) au cadre multi-classes avec un nombre arbitraire de classes. Dans (Grishechkin,1992[14]) l'auteur a établi l'équivalence entre les files d'attente d'arrivée par lots multi-classes avec une politique classique avec rappels et les processus de branchement avec immigration. Dans (Mouzoukis & Langaris,1996[22]) un mécanisme de priorité non préemptive a été ajouté au modèle de (Falin, 1988[8], Kulkarni,1986[18]). Dans (Langaris & Dimitriou, 2010[21]) les auteurs ont envisagé un système avec rappels multi-classes dans lequel les classes avec rappels sont associées à différentes phases de service.

A l'heure actuelle, les centres d'appels deviennent un moyen de communication important avec les clients. Par conséquent, le temps de réponse des centres d'appel est crucial pour la satisfaction des clients. Ensuite, prendre les bonnes décisions en matière de serveur est fondamental pour les coûts et les performances des centres d'appels. De nombreux modèles ont été élaborés pour déterminer le nombre adéquat de serveurs, voir (Gans et al.,2002[25]), (Halfin & Whitt,1981[31]), et les références qui s'y rapportent.

Ainsi, l'étude des comportements de redial des client, dans les centres d'appels est assez significative (Gans et al.,2002[25]), (Aguir et al.,2008[24]), (Sze,1984[6]) et autres références. En plus de redial (file d'attente), le reconnect est un autre aspect important, ces deux derniers devraient être pris en compte et modélisés, voir (Gans et al.,2002[25]), (Ding, Van der & Koole,2013[33]). De manière appropriée, nous pouvons dire que négliger l'impact des redials et des reconnects conduira soit à une sous-estimation, soit à une sur-estimation. En cas de surcharge, les performances du centre d'appel seront bonnes, mais à des coûts inutilement élevés. En de sous-estimation, la performance du centre d'appel sera dégradée, ce qui peut conduire à l'insatisfaction des clients. Lorsque le système est fortement sollicité, il en résulte de mauvais niveaux de service. Cependant, pour les grands centres d'appels, surtout pendant les heures d'affluence où le volume d'appels entrants est très important, il est possible que les niveaux de service visés puissent être atteints même en cas de forte charge du système, pour plus de détails voir (Garnett et al.,2002) et (Borst et al.,2004).

Les modèles fluides pour les centres d'appels ont été largement étudiés, (Whitt,2006[34]), (Mandelbaum et al.,2002[3]). Dans (Mandelbaum et al.,1999[2]), l'approximation fluide et l'approximation diffusion pour les file d'attente multiserveurs variables dans le temps avec abandon et reprises,

il a été montré que l'approximation fluide et la diffusion peuvent être obtenus en résolvant des ensembles d'équations différentielles non linéaires. Dans (Mandelbaum et al.,1998[1]) des résultats théoriques plus résultats théoriques plus généraux pour l'approximation fluide et diffusion pour les réseaux de services markoviens. Dans (Aguir et al.,2004[23]), les auteurs ont étendu le modèle en autorisant le comportement d'hésitation du client. Les modèles fluides ont également été appliqués à l'annonce des retards des clients dans les centres d'appel (Ibrahim & Whit,2009[29];Ibrahim & Whit,2011[30]).

Récemment, dans (Ding et al.,2013), les auteurs étudient les centres d'appels avec un redial et une orbite, en utilisant la limite fluide ils calculent le taux d'arrivée total attendu, qui est ensuite donné en entrée à l'Erlang. Un modèle pour calculer les niveaux de service et les taux d'abandon. La performance d'une telle procédure est validée dans le cas d'intervalles uniques et d'intervalles multiples avec des paramètres changeants.

Dans ce présent mémoire, une analyse du modèle de file d'attente M/M/s avec rappel; un centre d'appel avec deux orbites de redials, deux orbites de reconnects et deux flux exogènes de différents types est effectuée. L'objectif de ce travail, est de présenter un type de centre d'appel où l'importance des orbites de reconnect et de redial est cruciale dans tout système de télécommunication.

Décrivons à présent la structure de ce mémoire. Après l'introduction, nous introduisons dans le premier chapitre, les outils essentiels pour atteindre notre objectif dont la théorie des files d'attente, les modèles markoviens et leurs caractéristiques. Dans le second chapitre, nous présentons le modèle général d'un système des files d'attente avec rappels, on s'intéresse aux propriétés génériques des limites fluides d'une classe de processus de Markov et à la fin du chapitre on expose la relation entre la renormalisation et l'ergodicité. Dans le troisième chapitre, qui fait objet de notre travail, nous proposons un modèle fluide, qui est un analogue déterministe du modèle stochastique. Nous prouvons que le modèle stochastique original converge vers le modèle fluide sous une échelle appropriée. Ainsi, en gros, nous utilisons un modèle fluide pour dériver des approximations de premier ordre pour le nombre de clients dans les redials orbites et les reconnects orbites dans le trafic lourd, la limite fluide d'un tel modèle est la solution unique d'un système de trois équations différentielles.

Le mémoire se termine par une conclusion générale et quelques références bibliographiques.

Préliminaires

La théorie des files d'attente a de nombreuses applications, notamment dans les télécommunications et les réseaux informatiques. Dans ce chapitre, nous allons présenter les concepts fondamentaux de cette théorie à savoir les différents modèles de files d'attente, les modèles markoviens et leurs caractéristiques.

I.1 Description d'un modèle de file d'attente

Une file d'attente est un système stochastique composé d'un certain nombre (fini ou non) de places d'attente pour un ou plusieurs serveurs et bien sûr de clients qui arrivent, attendent, sont servis selon des règles de priorités données et quittent ensuite le système. La description ci-dessus des files d'attente, dont un diagramme schématique est présenté à la figure ne reprend pas toutes les caractéristiques des différents modèles de la littérature, mais identifie les éléments clés qui permettent de classer la majorité des files d'attente simples.



FIGURE I.1 – Le système de files d'attente.

Pour identifier le système de file d'attente, il est nécessaire de se baser sur des spécifications.

I.2 Classification des systèmes de files d'attente

- Propriété probabiliste du processus d'arrivée définie par la distribution de l'intervalle séparant deux arrivées consécutives ;

- Distribution des temps de service aléatoires ;
- le nombre m de serveurs (stations de service M/M/s) qui sont montées en parallèle. En général, on suppose que les temps de service correspondants suivent la même distribution et que les clients qui arrivent forment une seule file d'attente (si elle est homogène) ;
- La capacité N du système. Si $N < \infty$, la file d'attente ne doit pas dépasser une longueur de $N-m$ unités. Dans ce cas, certains clients arrivant dans le système peuvent ne pas être admis ;
- Sources de clients potentiels

L'étude mathématique d'un système se fait le plus souvent par l'introduction d'un processus stochastique défini de façon appropriée.

I.3 Analyse mathématique d'un système de files d'attente

En premier lieu, on s'intéresse ici au nombre $N(t)$ de clients se trouvant dans le système à l'instant t ($t \geq 0$).

En suite, en fonction des quantités qui définissent la structure du système, on cherche alors à calculer :

1/ les probabilités d'état $P_n(t) = \mathbb{P}(N(t) = n)$ qui définissent le régime transitoire du processus stochastique $\{N(t); t \geq 0\}$; il va de soi que les fonctions $P_n(t)$ dépendent de l'état initial où de la distribution initiale du processus,

2/ le régime stationnaire du processus stochastique, défini par

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

Pour modéliser l'évolution temporelle des populations, on utilise Les processus de saut de Markov les plus simples sont les processus de naissance et de mort.

I.4 Processus de naissance et de mort

On appelle processus de naissance et de mort toute chaîne de Markov en temps continu a valeurs dans \mathbb{N} et dont le générateur infinitésimal Q est de la forme :

$$a_{ij} \neq 0 \Leftrightarrow j \in \{i - 1, i, i + 1\}$$

Dans un tel processus, les transitions possibles à partir de i sont soit vers $i+1$ à la naissance, soit vers $i-1$ à la mort. Le taux de transition de droite est noté $\lambda_i, i \in \mathbb{N}$, et le taux de transition de gauche est noté $\mu_i, i \in \mathbb{N}^*$ (comme le processus est à valeurs dans \mathbb{N} , on a forcément $\mu_0 = 0$). On appelle les λ_i taux de naissance et les μ_i taux de mort sont donnés par une matrice triangulaire :

$$Q = \begin{pmatrix} -\lambda_0 & \lambda_0 & & & 0 \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & & \\ & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & \\ & & \mu_3 & \ddots & \ddots \\ 0 & & \ddots & \ddots & \ddots \end{pmatrix}$$

avec le graphe :

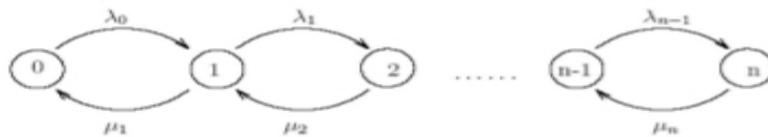


FIGURE I.2 – Graphe de transition d'un processus de naissance et de mort.

Ce graphique montre le passage d'un état à un autre. La transition vers la droite représente la naissance, la transition vers la gauche représente la mort.

- Si tous les λ_i sont nuls, parlez du processus de mort.
- Si tous les μ_i sont nuls, parlez du processus de naissance.

Comme λ_i et μ_i sont tous strictement positifs, tous les états communiquent entre eux et la chaîne est irréductible, de sorte que le processus à une distribution stationnaire P . Peut être trouvé en résolvant le système d'équations $PQ=0$:

$$\begin{cases} \lambda_0 P_0 - \mu_1 P_1 = 0 \\ \lambda_{i-1} P_{i-1} - (\lambda_i + \mu_i) P_i + \mu_{i+1} P_{i+1} = 0 \quad \forall i \geq 1. \end{cases}$$

D'où :

$$P_n = \frac{\lambda_1 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} P_0,$$

or

$$\sum_{i=0}^{+\infty} P_i = 1.$$

l'existence d'une loi stationnaire dépend de :

$$\begin{aligned} P_0 &= \left[1 + \sum_{i=0}^{+\infty} \prod_{j=1}^i \frac{\lambda_{j-1}}{\mu_j} \right]^{-1} \\ &= \left[\sum_{i=0}^{+\infty} \prod_{j=1}^i \frac{\lambda_{j-1}}{\mu_j} \right]^{-1}, \end{aligned}$$

avec la convention usuelle qu'un produit vide

$$\prod_{j=1}^0 = 1.$$

Les files d'attente de Markov sont un cas particulier très important des processus de naissance et de mort.

Pour désigner un système de file d'attente, la notation de kendall est généralement adoptée.

I.5 Notation de kendall

La notation de Kendall pour un système de file d'attente est décrit par six symboles : A/B/s/K/N/D. avec :

A : est la distribution des temps des inter-arrivées.

B : est la distribution des temps de service.

s : est le nombre de serveurs.

K : la capacité de la file.

N : la taille de la population.

D : est la discipline de service (FIFO , LIFO,...).

Pour désigné des distribution A et B, les symboles suivants sont utilisés :

M : Loi markovienne.

G : Loi générale.

D : Loi déterministe.

E_k : Loi de Erlang.

H_k : Loi hyperexponentielle.

I.6 Discipline de la file d'attente

Il y a plusieurs discipline de service :

- FIFO : First in First out.
- LIFO : Last in First out.
- RANDOM : aléatoire.
- HL (Hold On Line) : si un client important arrive, il prend la première place dans la file d'attente.
- PR (Preemption) : si un client important arrive, il est servi directement et le client moins important (en train d'être servi) est remis dans la file.
- ...

L'étude des files d'attente pour but de calculer ou d'estimer la performance du système dans des conditions d'exploitation données.

I.7 Mesures de Performances d'une files d'attente

Définition I.7.1 *On définit les mesures de performance d'une file d'attente par :*

- $L = \mathbb{E}(X)$: nombre moyen de clients dans le système.
- L_q : nombre moyen de clients dans la file d'attente.
- W : le temps moyen de séjour d'un clients dans le système.
- W_q : le temps moyen d'attente d'un client dans la file.

Ces mesure sont liées par la loi de Little

- $L = \lambda_e W$
- $L_q = \lambda_e W_q$
- $W = W_q + \frac{1}{\mu}$
- $L = L_q + \rho$

avec λ_e : taux d'entrée réel dans le système.

$\frac{1}{\mu}$: est la durée moyen 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

$$\rho = \frac{\lambda}{\mu}.$$

avec λ : le taux moyen d'arrivée.

μ : le taux moyen de service.

Dans la section suivante, on définit les systèmes de files d'attente Markoviens dont M/M/1 et M/M/s.

I.8 Systèmes de file d'attente Markoviens

I.8.1 Modèle d'attente M/M/1

La file d'attente M/M/1 est un système à serveur unique avec une capacité infinie et une discipline de service FIFO. Les temps d'arrivée des clients sont distribués selon une distribution de Poisson de proportion λ , et les temps de service sont indépendants selon une distribution exponentielle de proportion μ . Le processus $(N(t))_{t \geq 0}$ représentant le nombre de clients à l'instant t est le processus de naissance et de mort du taux de transition :

$$\lambda_n = \lambda, \forall n \geq 0, \text{ et } \mu_n = \mu, \forall n \geq 0.$$

On peut présenter la figure I.3.

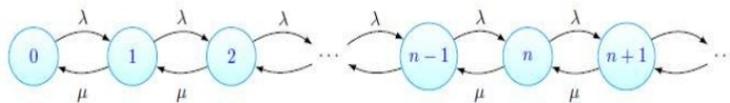


FIGURE I.3 – Graphe de transitions du modèle M/M/1.

Il est difficile d'étudier la variable aléatoire $N(t)$ représentant le nombre de clients au temps t dans le système. On s'intéresse plutôt à $N = \lim_{t \rightarrow \infty} N(t)$. On parle alors d'analyse en régime stationnaire.

- **Régime stationnaire du système**

Le système est dit stable si et seulement si $\rho = \frac{\lambda}{\mu} < 1$., il est aisé d'obtenir

Les équations de la balance sont données par :

$$\left\{ \begin{array}{l} \lambda p_0 = \mu p_1, \\ \lambda p_1 + \mu p_1 = \lambda p_0 + \mu p_2, \\ \cdot \\ \cdot \\ \cdot \\ \lambda p_n + \mu p_n = \lambda p_{n-1} + \mu p_{n+1} \quad \forall n \geq 1. \end{array} \right.$$

D'où

$$\left\{ \begin{array}{l} p_1 = \frac{\lambda}{\mu} p_0, \\ p_2 = \frac{\lambda}{\mu} p_1 = \left(\frac{\lambda}{\mu}\right)^2 p_0, \\ p_3 = \frac{\lambda}{\mu} p_2 = \left(\frac{\lambda}{\mu}\right)^3 p_0, \\ \cdot \\ \cdot \\ p_n = \frac{\lambda}{\mu} p_{n-1} = \left(\frac{\lambda}{\mu}\right)^n p_0. \end{array} \right.$$

Comme

$$\begin{aligned} \sum_{n=0}^{\infty} p_n = 1 &\Rightarrow \sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^n p_0 = 1 \\ &\Rightarrow p_0 \sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^n = 1 \\ &\Rightarrow p_0 = \frac{1}{\sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^n} \\ &\Rightarrow p_0 = 1 - \frac{\lambda}{\mu}. \\ p_n &= \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^n = (1 - \rho) \rho^n. \end{aligned}$$

Une importante caractéristique des systèmes de files d'attente est d'estimer la performance.

- Mesures de performance

1. Le nombre moyen de clients dans le système :

$$L = \frac{\rho}{1 - \rho}.$$

2. Le nombre moyen de clients dans la file d'attente :

$$L_q = L - \frac{\lambda}{\mu} = \frac{\rho^2}{1 - \rho}.$$

3. Le temps moyen de séjour d'un client dans le système :

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

4. Le temps moyen d'attente d'un client dans la file :

$$W_q = \frac{\rho}{\mu(1 - \rho)}.$$

I.8.2 Modèle d'attente M/M/s

Le modèle M/M/s est un système avec "s" serveurs identiques et une file d'attente à capacité infinie, avec une discipline de service FIFO. En supposant que le flux d'arrivée est une fonction de Poisson avec le paramètre λ , la durée de service est exponentielle avec le paramètre μ .

On a le processus $\{N(t), t \geq 0\}$ du nombre de clients dans le système à l'instant t est un processus de naissance et de mort de taux de transition :

$$\lambda_n = \lambda \quad \forall n \geq 0$$

$$\mu_n = \begin{cases} n\mu & \text{si } n \leq s \\ s\mu & \text{si } n > s \end{cases}$$

Le graphe des transitions est donnée dans la figure I.4.

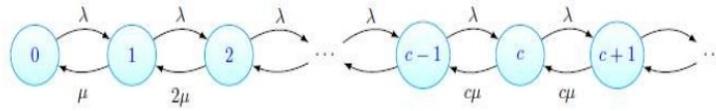


FIGURE I.4 – Graphe de transitions du modèle M/M/s où M/M/c.

- Le régime stationnaire de la file d'attente M/M/s

Soit $P_n = \lim_{t \rightarrow \infty} P_n(t)$, $n \geq 0$. Cette distribution stationnaire satisfait les équations de la balance

- 1^{er} Cas : $n \leq s$

$$\left\{ \begin{array}{l} \mu p_1 = \lambda p_0 \\ \lambda p_0 + 2\mu p_2 = \lambda p_1 + \mu p_1 \\ \cdot \\ \cdot \\ \cdot \\ \lambda p_{s-1} + s\mu p_{s+1} = \lambda p_s + n\mu p_s \end{array} \right.$$

D'où

$$\left\{ \begin{array}{l} p_1 = \frac{\lambda}{\mu} p_0, \\ p_2 = \frac{\lambda}{2\mu} p_1 = \frac{1}{2} \left(\frac{\lambda}{\mu}\right)^2 p_0, \\ p_3 = \frac{\lambda}{3\mu} p_2 = \frac{1}{2 \times 3} \left(\frac{\lambda}{\mu}\right)^3 p_0, \\ \cdot \\ \cdot \\ p_s = \frac{\lambda}{s\mu} p_{s-1} = \frac{1}{s!} \left(\frac{\lambda}{\mu}\right)^s p_0. \end{array} \right.$$

- 2^{me} Cas : $n > s$

$$\left\{ \begin{array}{l} \lambda p_s = s\mu p_{s+1}, \\ \lambda p_{s+1} + s\mu p_{s+1} = \lambda p_s + s\mu p_{s+2}, \\ \cdot \\ \cdot \\ \cdot \end{array} \right.$$

D'où

$$\left\{ \begin{array}{l} p_s = \frac{1}{s!} \left(\frac{\lambda}{\mu}\right)^s p_0, \\ p_{s+1} = \frac{1}{s!s} \left(\frac{\lambda}{\mu}\right)^{s+1} p_0, \\ p_{s+2} = \frac{1}{s!s^2} \left(\frac{\lambda}{\mu}\right)^{s+2} p_0, \\ \cdot \\ \cdot \\ p_n = \frac{\lambda}{s!s^{n-s}\mu} \left(\frac{\lambda}{\mu}\right)^n p_0, \end{array} \right.$$

Donc

$$p_n = \begin{cases} p_0 \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!} & \text{si } 0 \leq n \leq s, \\ p_0 \rho^n \frac{s^s}{s!} & \text{si } n > s. \end{cases}$$

On remarque que pour $n = s$, les deux formules donnent la même valeurs. Pour calculer la probabilité pour que le système est vide P_0 , on applique l'équation de normalisation $\sum_{n=0}^{\infty} P_n = 1$.

En effet

$$p_0 = \left[\frac{s^s \rho^{s+1}}{s!(1-\rho)} + \sum_{n=0}^s \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} \right]^{-1}$$

- Les mesures de performances

1. *Nombre moyen de clients dans la file :*

$$\begin{aligned} L_q &= \mathbb{E}(X_q) \\ &= \sum_{n=s+1}^{\infty} (n-s)p_n. \\ &= \sum_{n=s+1}^{\infty} (n-s)\rho^{n-s}p_s. \\ &= p_s \sum_{n=s+1}^{\infty} (n-s)\rho^{n-s}. \end{aligned}$$

On pose $n-s = j \Rightarrow L_q = p_s \sum_{j=1}^{\infty} j\rho^j = p_s \rho \sum_{j=1}^{\infty} j\rho^{j-1}$.

$$L_q = p_s \rho \frac{1}{1-\rho^2}.$$

2. Nombre moyen de clients dans le système :

$$\begin{aligned} L &= L_q + \frac{\lambda}{\mu} \\ &= s\rho + \frac{\rho p_s}{(1-\rho)^2}. \end{aligned}$$

A l'aide des formules de Little, on trouve :

3. Temps moyen d'attente :

$$W_q = \frac{p_s}{s\mu(1-\rho)^2}.$$

4. Temps moyen de séjour :

$$W = \frac{p_s}{s\mu(1-\rho)^2} + \frac{1}{\mu}.$$

Remarque I.8.1 En générale, une file est stable si et seulement si le nombre moyen d'arrivées de clients par unité de temps, noté λ , est inférieur au nombre moyen de client pouvant être servis par unité de temps. Si chaque serveur peut traiter μ clients par unité de temps et si le nombre de serveurs est s , une file est stable si et seulement si :

$$\lambda < s\mu \Leftrightarrow \rho = \frac{\lambda}{s\mu} < 1.$$

Les files d'attente avec rappels

On introduit dans ce présent chapitre, la notion de file d'attente avec rappel est souvent utilisée dans la théorie des files d'attente pour modéliser des situations où les clients peuvent être rappelés s'ils ne sont pas servis immédiatement.

II.1 Modèle général d'un système de file d'attente avec rappels

Le modèle général d'un système de files d'attente avec rappels (Retrial Queue), peut s'écrire comme suit : le système est contient de " $s \geq 1$ " dispositifs de service et " $m - s$ " ($m \geq s$) positions d'attente. Les clients arrivent dans le système selon un processus aléatoire avec une loi de probabilité donnée, et forment un flux d'appels primaires. A l'arrivée d'un client, s'il y a une position d'attente libre, le client rejoint la file d'attente. Dans le cas contraire, il quitte l'espace de service temporairement avec une probabilité H_0 pour tenter sa chance après une durée de temps aléatoire, ou il quitte le système définitivement avec une probabilité $1 - H_0$. Entre les tentatives, le client est dit "en orbite" et devient source d'appels répétés ou d'appels secondaires. La capacité O de l'orbite peut être finie ou infinie. Dans le cas où O est finie et si l'orbite est pleine, le client quitte le système pour toujours. Lorsqu'un client rappelle de l'orbite, il est traité de la même manière qu'un client primaire, si tous les serveurs sont occupés il quitte le système avec une probabilité H_k (s'il s'agit de le k ème tentative échouée), ou quitte le système définitivement avec une probabilité $1 - H_k$.

Pour désigner les systèmes avec rappels, on utilise la notation de kendall $A/B/s/m/O/H$, où A et B décrivant respectivement la distribution du temps inter-arrivées et la distribution du temps de service, s représente le nombre de serveurs identiques et en parallèles, m est le nombre de positions d'attente plus le nombre de serveurs, et O est la capacité de l'orbite, la séquence $H = \{H_i, i \geq 0\}$ est la fonction persévérance, où H_i est la probabilité qu'un client fasse une $(i+1)$ ème tentative de

rappels, après une i ème tentative échouée. Si m , O et H sont absents dans la notation de Kendall, alors $m=s$, $O = \infty$, pour tout $i \geq 0$.

Le schéma générale d'un système avec rappels est donné dans la figure II.1.

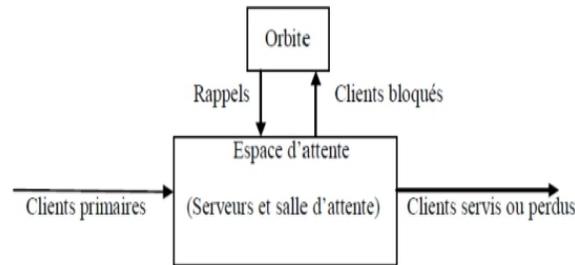


FIGURE II.1 – Schéma générale d'un système avec rappels

On définit ainsi, dans ce qui suit modèles markoviens $M/M/1$ avec rappels et $M/M/s$ avec rappels.

II.2 Modèles markoviens

II.2.1 Modèle $M/M/1$ avec rappels

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

$$\{C(t), N_o(t), t \geq 0\}, \quad (\text{II.1})$$

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 (II.1) est de Markov d'espace d'états $S = \{0, 1\} \times \mathbb{N}$.

Le graphe de transition du modèle M/M/1 avec rappels peut être présenté dans le schéma ci-dessous.

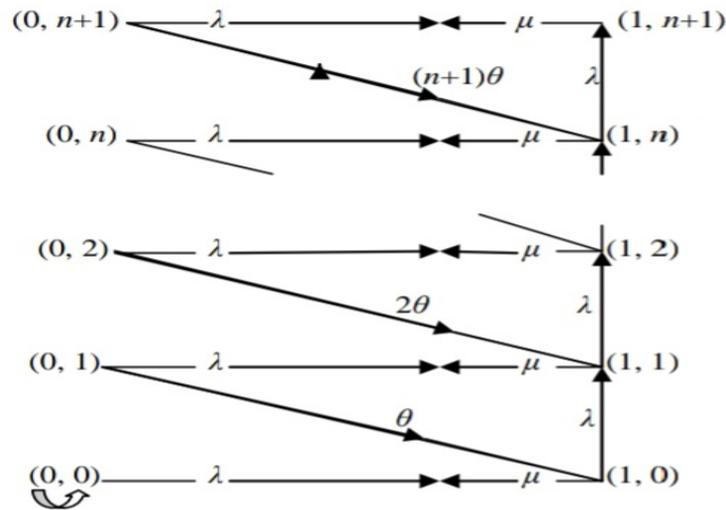


FIGURE II.2 – Graphe de transition du modèle M/M/1 avec rappels.

• Le régime stationnaire

Les équations d'équilibre statistique sont :

$$(\lambda + j\theta)p_{0j} = \mu p_{1j}; \quad (\text{II.2})$$

$$(\lambda + \mu)p_{1j} = \lambda p_{0j} + (j + 1)\theta p_{0,j+1} + \lambda p_{1,j-1}. \quad (\text{II.3})$$

Ici, $p_{ij} = \lim_{t \rightarrow \infty} P(C(t) = i, N_o(t) = j)$, $i = 0, 1$ et $j \geq 0$, représentent la distribution stationnaire conjointe de l'état du serveur et du nombre de clients en orbite.

Introduisons les fonctions génératrices suivantes :

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

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

A l'aide de ses fonctions et à partir des équations (II.2) et (II.3), on obtient :

$$P_0(z) = (1 - \rho) \left(\frac{1 - \rho}{1 - z\rho} \right)^{\frac{\lambda}{\theta}}; \quad (\text{II.4})$$

$$P_1(z) = \rho \left(\frac{1 - \rho}{1 - z\rho} \right)^{\frac{\lambda}{\theta} + 1}. \tag{II.5}$$

Les transformées inverses des (II.4) et (II.5) nous donnent les formules analytiques explicites :

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

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

II.2.2 Modèle M/M/2 avec rappels

Nous considérons un système de files d'attente avec rappels où l'espace de service contient $s = 2$ 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 introduits sont mutuellement indépendantes.

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

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

Les transitions possibles sont données dans la figure II.3.

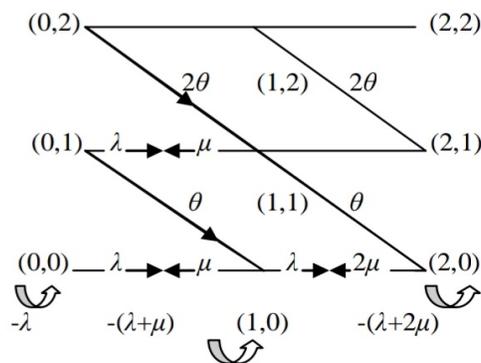


FIGURE II.3 – Graphe de transition du modèle M/M/2 avec rappels.

- Le régime stationnaire

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

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, telles que :

$$(\lambda + j\theta)p_{0j} = p_{1j}; \quad (\text{II.6})$$

$$(\lambda + 1 + j\theta)p_{1j} = \lambda p_{0j} + (j + 1)\theta p_{0,j+1} + 2p_{2j}; \quad (\text{II.7})$$

$$(\lambda + 2)p_{2j} = \lambda p_{1j} + (j + 1)\theta p_{1,j+1} + \lambda p_{2,j-1}; \quad (\text{II.8})$$

Ainsi que l'équation de normalisation :

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

pour $p_{ij} = \lim_{t \rightarrow \infty} p_{ij}(t)$, la résolution des équations (II.6)-(II.7)-(II.8) nous donne

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

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

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

II.2.3 Modèle M/M/s avec rappels

Le modèle d'attente M/M/s avec rappels est un système où les temps d'inter-arrivées primaires, les durées de service et les temps d'inter-rappels sont des variables aléatoires indépendantes et exponentiellement distribuées.

- Description mathématique du modèle

Nous considérons un système M/M/s avec rappels à "s" serveurs identiques. Les clients primaires arrivent selon un processus de Poisson de taux λ , la durée de service est exponentielle de

paramètre μ . Si un client primaire trouve au moins un serveur libre, il l'occupe immédiatement et quitte le système après le service. D'autre part, tout client qui trouve tous les serveurs occupés à son arrivée, il est obligé de quitter la zone de service, mais il rappelle ultérieurement. La durée entre deux rappels consécutifs est exponentielle de paramètre θ [19]. On utilise la politique de rappels classique et on suppose que les temps d'inter-arrivées, les durées de service et temps d'inter-rappels sont des variables aléatoire mutuellement indépendantes.

- **La distribution stationnaire**

Le système peut-être décrit par le processus (II.1), où $C(t)$ est le nombre de serveurs occupés, $N_o(t)$ est le nombre de clients en orbite au temps t ,

Supposons que le régime stationnaire existe ($\rho = \frac{\lambda}{s\mu} < 1$), le processus (II.1) est une chaîne de Markov régulière à temps continu d'espace $S = \{0, 1, \dots, s\} \times \mathbb{N}$.

Soit $P_{ij} = \mathbb{P}(C(t) = i, N(t) = j)$, avec $i = 0, 1, \dots, s, j \geq 0$.

Les probabilités de transitions à l'état stationnaire sont données par :

- Pour $0 \leq i \leq s - 1$,

$$P_{ij}(n, m) = \begin{cases} \lambda & \text{si } (n, m) = (i + 1, j), \\ i\mu & \text{si } (n, m) = (i - 1, j), \\ j\theta & \text{si } (n, m) = (i + j, j - 1), \\ -(\lambda + i\mu + j\theta) & \text{si } (n, m) = (i, j), \\ 0 & \text{sinon.} \end{cases}$$

- Pour $i=s$,

$$P_{sj}(n, m) = \begin{cases} \lambda & \text{si } (n, m) = (s, j + 1), \\ s\mu & \text{si } (n, m) = (s - 1, j), \\ -(\lambda + s\mu) & \text{si } (n, m) = (s, j), \\ 0 & \text{sinon.} \end{cases}$$

- **Cas particulier**

Dans le cas $s=1$ et sous la condition $\rho = \frac{\lambda}{\mu} < 1$, la distribution stationnaire est donnée par [15]

$$P_{0j} = \frac{\rho^j}{j!\theta^j} (1 - \rho)^{1 + \frac{\lambda}{\theta}} \prod_{k=0}^{j-1} (\lambda + k\theta), \quad j \geq 0.$$

$$P_{1j} = \frac{\rho^{j+1}}{j!\theta^j} (1 - \rho)^{1 + \frac{\lambda}{\theta}} \prod_{k=1}^j (\lambda + k\theta), \quad j \geq 0.$$

Ainsi nous avons l'expression suivante de la distribution stationnaire du nombre total de clients dans le système :

$$P_{0j} + (1 - \delta_{0j})P_{1,j-1} = \frac{\rho^j}{j!\theta^j} (1 - \rho)^{1+\frac{\lambda}{\theta}} \prod_{k=1}^j (\lambda + k\theta), \quad j \geq 0.$$

Toutes les mesures de performance s'obtiennent en utilisant les fonctions génératrices (voir par exemple dans [13]).

Pour $s \geq 2$, il n'existe pas de formules analytiques explicites, mais il existe des méthodes d'approximation. Parmi ces méthodes on trouve la méthode d'approximation géométrique matricielle utilisant les processus quasi-naissance et mort (Q.N.M), cette dernière elle est utilisé dans [10]. En ordonnant les états comme $S = (0,0), \dots, (s,0), (0,1), \dots, (s,1), \dots$

Nous pouvons exprimer le générateur infinitésimal M du processus (II.1) sous la forme d'une matrice- bloc suivante :

$$M = \begin{pmatrix} A_{00} & A_{01} & & & \\ A_{10} & A_{11} & A_{12} & & \\ & A_{21} & A_{22} & A_{23} & \\ & & \ddots & \ddots & \ddots \end{pmatrix},$$

avec $A_{j,j-1}$, A_{jj} et $A_{j,j+1}$ sont des matrices carrées d'ordre $s + 1$, définies comme suit :

$$A_{j,j-1} = \begin{pmatrix} 0 & j\theta & & & \\ & 0 & j\theta & & \\ & & \ddots & \ddots & \\ & & & 0 & j\theta \\ & & & & 0 \end{pmatrix},$$

$$A_{j,j} = \begin{pmatrix} a_{0j} & \lambda & & & \\ \mu & a_{1j} & \lambda & & \\ & 2\mu & a_{2j} & \lambda & \\ & & \vdots & \ddots & \ddots \\ & & & (s-1)\mu & a_{s-1,j} & \lambda \\ & & & & s\mu & a_{sj} \end{pmatrix},$$

$$A_{j,j+1} = \text{diag}(0, \dots, 0, \lambda),$$

avec $a_{ij} = -(\lambda + i\mu + (1 - \delta_{is})j\theta)$, pour $0 \leq i \leq s$ et $j \geq 0$, où δ_{ij} est défini par :

$$\delta_{ij} = \begin{cases} 1 & \text{si } i = j, \\ 0 & \text{sinon.} \end{cases}$$

Géométriquement, le comportement asymptotique du processus (II.1), peut être représenté par le diagramme de transition donné dans la figure II.4.

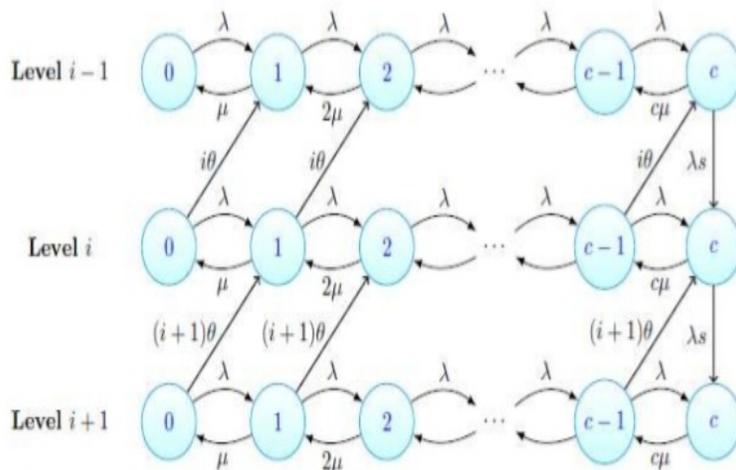


FIGURE II.4 – Graphe de transition du modèle M/M/s où M/M/c avec rappels.

Notons que la distribution stationnaire du processus (II.1) est : $p = \{P_{ij}; (i, j) \in S\}$, le calcul de cette distribution se fait généralement à l'aide des équations de kolmogorov $pQ = 0$, où 0 indique la transposée du vecteur colonne nul.

En partitionnant le vecteur de probabilité stationnaire p comme $p = (p(0), p(1), \dots)$, où $p(j) = (p_{0j}, \dots, p_{sj})$, nous pouvons écrire les équations de kolmogorov sous forme d'une matrice :

$$p(j-1)A_{j-1,j} + p(j)A_{jj} + p(j+1)A_{j+1,j} = 0_{s+1}, \tag{II.9}$$

où $p(-1)$ et $A_{-1,0}$ sont nuls par définition, et 0_{s+1} est le vecteur colonne nul d'ordre $s+1$.

On peut introduire les fonctions génératrices suivantes :

$$p_i(z) = \sum_{j=0}^{\infty} z^j p_{ij}, \quad 0 \leq i \leq s,$$

et transformer les équations de kolmogorov en un ensemble d'équations différentielles

$$\mu \frac{dp(z)}{dz} A(z) = p(z) B(z), \quad (\text{II.10})$$

où $p(z) = (P_0(z), \dots, P_s(z))$, $A(z)$ et $B(z)$ sont des matrices carrées d'ordre $s+1$:

$$A(z) = \begin{pmatrix} z & -1 & & & \\ & z & -1 & & \\ & & \ddots & \ddots & \\ & & & z & -1 \\ & & & & 0 \end{pmatrix},$$

$$B(z) = \begin{pmatrix} a_{00} & \lambda & & & \\ \mu & a_{10} & \lambda & & \\ & 2\mu & a_{20} & \lambda & \\ & & \vdots & \ddots & \ddots \\ & & & (s-1)\mu & a_{s-1,0} & \lambda \\ & & & & s\mu & a_{sj} \end{pmatrix}.$$

- Les mesures de performances

- Pour le cas $s=1$, la probabilité de blocage et le nombre moyen de clients dans l'orbite sont données dans la monographie de Falin [15] par

$$B = \frac{\lambda}{\mu} \text{ et } E(N) = \frac{1 + \theta}{\theta} \frac{\lambda^2}{1 - \lambda}$$

- Pour le cas $s=2$, les caractéristiques sont données dans Falin, en termes des fonctions hyper-géométrique [15].

- Pour le cas général ($s > 2$), en utilisant les équations (II.9) et (II.10), Artalejo [9], a obtenu les valeurs moyennes stationnare suivantes :

— *La probabilité de blocage stationnaire*

$$B = \lim_{t \rightarrow \infty} P(C(t)).$$

— *Le nombre moyen des clients en orbite*

$$\mathbb{E}(N) = \lim_{t \rightarrow \infty} \mathbb{E}(N(t)) = \frac{(\mu + \theta)(\lambda + \mu \text{Var}(C))}{\theta(s\mu - \lambda)}.$$

— *Le nombre moyen de serveurs occupés*

$$\mathbb{E}(C) = \lim_{t \rightarrow \infty} \mathbb{E}(C(t)) = \frac{\lambda}{\mu}.$$

En utilisant les formules de Little, on obtient

— *le temps moyen de séjour dans le système, donnée par*

$$W = \frac{\mathbb{E}(N)}{\lambda}.$$

— *le temps moyen d'attente, donnée par*

$$W_B = \frac{W}{B}.$$

Dans le contexte des files d'attente, les "limites fluides" se réfèrent à un état stationnaire où les caractéristiques du système, telles que le nombre moyen de clients en attente ou le temps moyen d'attente, se stabilisent. Les limites fluides peuvent nous aider à optimiser notre temps, à gérer les interruptions et à rendre le flux de mouvements plus fluide et efficace. Pour décrire l'évolution d'un système au fil du temps, où les changements d'état se produisent à des moments aléatoires plutôt qu'à des intervalles réguliers, on utilise les "processus de sauts" comme modèles. En incorporant des processus de sauts dans la modélisation d'un service, nous pouvons mieux tenir compte de la nature aléatoire et imprévisible de certaines activités d'attente, ce qui pourrait vous aider à planifier notre temps de manière plus réaliste et à mieux gérer les décisions.

II.3 Les limites fluides d'une classe de processus de Markov

Dans cette section, on s'intéresse aux propriétés génériques des limites fluides d'une classe assez générale de processus de sauts à valeurs dans \mathbb{N}^d . Si d'autres types de processus rencontrés en pratique ne rentrent pas dans ce formalisme, les techniques utilisées ici s'appliquent très souvent.

À partir de maintenant $d \in \mathbb{N}$ est fixé, la norme L_1 sur \mathbb{Z}^d est notée $\|\cdot\|$,

i.e. $\|m\| = |m_1| + \dots + |m_d|$ si $m = (m_k) \in \mathbb{Z}^d$.

Définition II.3.1 *Un processus Markovien de sauts $(X(t))$ appartient à la classe (C) si sa loi est celle de la solution d'une équation différentielle stochastique du type*

$$dX(t) = \sum_{i \in I} 1_{\{X(t-) \in C_i\}} \int_{\mathbb{Z}^d} m V_i(dt, dm), \quad (\text{II.11})$$

- I est un ensemble au plus dénombrable et, pour $i \in I$,
- C_i est un sous ensemble de \mathbb{R}^d ;
- V_i est un processus de poisson sur $\mathbb{R}_+ \times \mathbb{Z}^d$ d'intensité $\lambda_i dt \otimes v_i(dm)$ avec λ_i positif. Les processus ponctuels V_i , $i \in I$ sont indépendant;
- $\sum_{i \in I} \lambda_i \int_{\mathbb{Z}^d} \|m\|^2 v_i(dm) < +\infty$.

L'intégrale sur \mathbb{Z}^d dans l'équation (II.11) porte bien entendu sur la variable m . Remarquer que pour t fixé, le processus ponctuel marqué $V_i(dt, dm)$ est soit la mesure nulle soit une masse de Dirac.

L'équation différentielle (II.11) est vectorielle, si $j=1, \dots, d$, pour la j -ième coordonnée, elle vaut :

$$dX_j(t) = \sum_{i \in I} 1_{X(t-) \in C} \int_{\mathbb{Z}^d} m V_i(dt, dm), \quad (\text{II.12})$$

La combinaison de "limites fluides" et de "renormalisation" peut être une approche puissante pour modéliser et optimiser des systèmes complexes.

Proposition II.3.1 [28] *Si $(X(t))$ est un processus de Markov de classe (C), pour $T > 0$ les variables aléatoires*

$$\left\{ \frac{\|X(x, \|x\|T)\|}{\|x\|}; x \in \mathbb{N}^d \right\} = \left\{ \|X\|_f(x, T); x \in \mathbb{N}^d \right\}$$

sont uniformément intégrables.

Proposition II.3.2 *Si $(X(x, t))$ est un processus de Markov de classe (C), l'ensemble des processus renormalisés*

$$\left\{ (\bar{X}(x, t)); x \in \mathbb{N}^d \right\} = \left\{ \frac{1}{\|x\|} (X(x, \|x\|t)); x \in \mathbb{N}^d \right\}$$

est relativement compact et toutes ses valeurs d'adhérence, i.e. les limites fluides associées, sont des processus continus.

Le processus

$$(\bar{M}(x, t)) = \left(\bar{X}(x, t) - \frac{x}{\|x\|} - \sum_{i \in I} \lambda_i \int_{\mathbb{Z}^d} m v_i(dm) \int_0^t 1_{\{X(x, \|x\|s) \in C_i\}} ds \right) \quad (\text{II.13})$$

est une martingale qui converge en probabilité vers 0 uniformément sur les compacts quand $\|x\|$ tend vers l'infini, i.e. pour tout $T \geq 0$ et $\varepsilon > 0$,

$$\lim_{\|x\| \rightarrow +\infty} P\left(\sup_{0 \leq s \leq T} \|\bar{M}(x, s)\| \geq \varepsilon\right) = 0.$$

• Caractérisation des limites fluides :

Les processus renormalisés étant solutions d'une équation différentielle stochastique, il est naturel d'essayer de caractériser les limites fluides comme les solutions d'une équation différentielle.

Si les ensembles (C_i) sont des cônes de \mathbb{R}^d , i.e. pour $i \in I$, si $x \in C_i$ alors $\alpha x \in C_i$ pour tout $\alpha > 0$, l'égalité¹ peut se réécrire de la façon suivante

$$\bar{X}(x, t) = \frac{x}{\|x\|} + \bar{M}(x, t) + \sum_{i \in I} \lambda_i \int_{\mathbb{Z}^d} m v_i(dm) \int_0^t 1_{\{\bar{X}(x, s) \in C_i\}} ds. \quad (\text{II.14})$$

En supposant que le passage à la limite soit valide et que $x/\|x\| \rightarrow \alpha$ quand $\|x\|$ tend vers l'infini, une limite fluide $(\bar{z}(t))$ devrait satisfaire l'équation

$$\bar{z}(t) = \alpha + \sum_{i \in I} \lambda_i \int_{\mathbb{Z}^d} m v_i(dm) \int_0^t 1_{\{\bar{z}(s) \in C_i\}} ds,$$

ou encore, sous forme différentielle,

$$d\bar{z}(t) = \sum_{i \in I} \lambda_i \int_{\mathbb{Z}^d} m v_i(dm) 1_{\{\bar{z}(s) \in C_i\}} dt.$$

La proposition ci-dessous montre que cette équation est satisfaite si la limite fluide est en dehors des points de discontinuité de l'équation différentielle, i.e. tant que $\bar{z}(t)$ n'est pas sur le bord d'un des $C_i, i \in I$.

1. $\bar{X}_j(x, t) = \frac{x_j}{\|x\|} + \bar{M}_j(x, t) + \sum_{i \in I} \lambda_i \int_{\mathbb{Z}^d} m_j v_i(dm) \int_0^t 1_{\{X(x, \|x\|s) \in C_i\}} ds.$

Proposition II.3.3 [28] *Si $(\bar{z}(t))$ est une limite fluide d'un processus de la classe (C) , si les ensembles $(C_i; i \in I)$ sont des cônes de \mathbb{R}^d et s'il existe un intervalle $[0, t_0]$ tel qu'avec probabilité 1, pour tout $t \in [0, t_0]$, le processus $(\bar{z}(t))$ n'est sur la frontière d'aucun des ensembles (C_i) , alors presque sûrement*

$$\bar{z}(t) = \bar{z}(0) + \sum_{i \in I} \lambda_i \int_{\mathbb{R}^d} m v_i(dm) \int_0^t 1_{\{\bar{z}(u) \in C_i\}} du, \quad (\text{II.15})$$

pour tout $t \in [0, t_0]$.

Dans le contexte de la modélisation des systèmes complexes, la renormalisation et l'ergodicité sont deux concepts distincts mais qui peuvent parfois être liés.

II.4 Relation entre renormalisation et ergodicité

On revient dans cette section au cas d'un processus de Markov général à valeurs dans un espace d'états dénombrable. Le résultat suivant est dû à Rybko et Stolyar (quand le temps d'arrêt τ est déterministe).

Théorème II.4.1 [28] *S'il existe une fonction f positive sur S , des constantes A, ε strictement positives et un temps d'arrêt τ tels que*

$$\limsup_{f(x) \rightarrow +\infty} \frac{\mathbb{E}_x(\tau)}{f(x)} \leq A, \quad (\text{II.16})$$

$$\limsup_{f(x) \rightarrow +\infty} \mathbb{E}_x \left(\frac{f(X(\tau))}{f(x)} \right) \leq 1 - \varepsilon, \quad (\text{II.17})$$

si la variable $f(X(1))$ est intégrable et si l'ensemble

$$F_K = \{x \in S / f(x) \leq K\}$$

est fini pour tout $K \geq 0$, le processus de Markov $(X(t))$ est ergodique .

Pour K suffisamment grand, le temps d'atteinte T_{F_K} de F_K par $(X(t))$ satisfait l'inégalité

$$\frac{\mathbb{E}_x(T_{F_K})}{f(x)} \leq \frac{4A}{\varepsilon}, \quad (\text{II.18})$$

Pour $x \in S$ tout tel que $f(x) > K$.

Proposition II.4.1 [28]

Si une fonction positive f sur S est telle que

a) pour $K \geq 0$, l'ensemble $\{x \in S / f(x) \leq K\}$ est fini,

b) la variable $f(X(1))$ est intégrable, et s'il existe $T > 0$ et $\varepsilon > 0$ tels que

$$\limsup_{f(x) \rightarrow +\infty} \mathbb{E}_x(\|X\|_f(x, t)) = \limsup_{f(x) \rightarrow +\infty} \frac{\mathbb{E}_x(f(X(f(x)T)))}{f(x)} \leq 1 - \varepsilon$$

le processus de Markov $(X(t))$ est ergodique.

Le remarque suivant montre si toute les limites fluides tendent vers 0 convenablement alors le processus de Markov est ergodique.

Remarque II.4.1 [28]

Si les conditions a) et b) de la proposition précédente sont satisfaites et s'il existe T tel que

c) la variable $\|X\|_f(x, T)$ converge en loi vers 0 quand $f(x)$ tend vers l'infini,

d) la suite $(\|X\|_f(x, T); x \in S)$ est uniformément intégrable

le processus de Markov est ergodique. La condition c) peut être remplacée par

c') le processus $(\|X\|_f(x, t))$ converge pour la topologie de Skorokhod vers $(x(t))$ quand $f(x)$ tend vers l'infini et $x(t) = 0$ pour tout $t \geq T/2$.

Système de file d'attente M/M/s avec deux orbites de reconnects et deux orbites de redials

Dans ce chapitre nous exposons une approche souvent utilisée dans les situations où une tâche nécessite plusieurs étapes ou processus successifs pour être accomplie. Les systèmes de files d'attente avec orbites sont un concept utilisé dans la théorie des files d'attente pour modéliser des systèmes où les clients peuvent passer par plusieurs étapes de service avant d'être complètement servis.

III.1 Le modèle

Considérons un modèle de file d'attente avec deux orbites de reconnects et deux orbites de redials et s serveurs (figure III.1). Deux flux indépendants de poisson de clients se retrouvent dans une file d'attente interne commune. Un client arrivant de type i , $i=1,2$ est pris en charge par un serveur disponible, s'il y en a ; sinon, il attend dans une file d'attente infinie. Les clients sont traités dans l'ordre d'arrivée (FIFO), le temps de service requis pour chaque client est indépendant de son type.

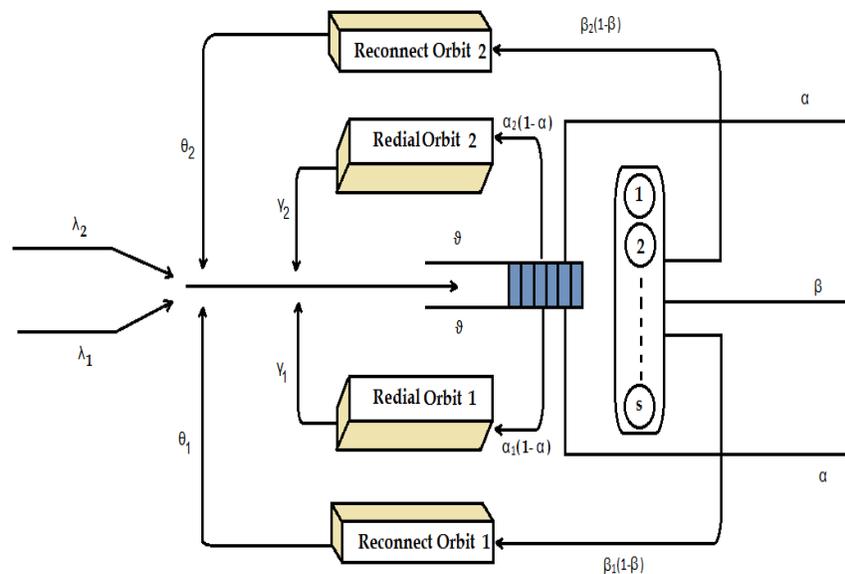


FIGURE III.1 – Un modèle de file d’attente avec deux orbites de redials et deux orbites de reconnects.

Un client en attente de type i qui n’a pas été connecté à un serveur perdra patience (client impatient) et abandonnera au bout d’un temps exponentiellement distribué ψ . Nous supposons que $E(\psi) = \frac{1}{\vartheta} < \infty$, où ϑ est le taux d’abandon. Le client abandonné quitte le système de manière négative avec une probabilité α et il est considéré comme un client perdu, soit il décide de continuer à demander un service avec une probabilité $(1 - \alpha)$. Ainsi, si le client décide de ne pas quitter le système, il se dirigera vers une file d’attente de redial de type i séparée qui tente de redistribuer ses travaux avec une probabilité $(1 - \alpha)\alpha_i, (i = 1, 2)(\alpha_1 + \alpha_2 = 1)$. Ce qui signifie que le client doit choisir entre deux orbites de redial (par exemple, un client de type 2 peut être acheminé vers l’orbite de redial 1 avec la probabilité $(1 - \alpha)\alpha_1$, où vers l’orbite de redial 2 avec la probabilité $(1 - \alpha)\alpha_2$, le choix est aléatoire et ne dépend pas du seuil des orbites de redial ou du type de client. Notons qu’une orbite de redial est une salle d’attente ordinaire où un client venant de la file d’attente primaire attendra et recomposera son numéro après un temps exponentiellement distribué $\omega_i, (i = 1, 2)$ avec $E(\omega_i) = \gamma_i < \infty$.

Nous supposons que le temps de service d’un client de type i a une distribution exponentielle avec une moyenne de $\frac{1}{\mu}$. Nous supposons que les temps de service sont indépendants, et que le temps de service requis pour chaque client est indépendant de son type. Après avoir été servi, le client peut quitter le système avec une probabilité β ou décider de revenir dans le système pour un autre service avec une probabilité $(1 - \beta)$. Le client de type i entrera dans l’une des files d’attente de reconnect de type i avec la probabilité $(1 - \beta)\beta_i$, avec $(\beta_1 + \beta_2 = 1)$. Comme cela a été

mentionné plus haut (dans le cas des orbites de redial), notons que le client après avoir obtenu le service peut revenir vers le système, en choisissant entre deux orbites de reconnect (par exemple, le client de type 2 peut être acheminé vers l'orbite de reconnect 1 avec la probabilité $(1 - \beta)\beta_1$, où vers l'orbite de reconnect 2 avec la probabilité $(1 - \beta)\beta_2$, le choix est aléatoire et ne dépend pas du seuil des orbites de reconnects ni du type de client. L'orbite de reconnect est une salle d'attente ordinaire où un client revenant du dernier service et demandant un autre service sera attendu et se reconnectera après un temps φ_i exponentiellement distribué, avec $E(\varphi_i) = \theta_i < \infty$. Ainsi, un client abandonné de type i doit choisir entre l'une des deux orbites de redials, et celui qui décide de revenir pour un autre service a le choix entre l'une des deux orbites de reconnect, cela signifie qu'un client de type 1 (resp. de type 2) n'est pas dirigé automatiquement vers les orbites de redial ou de reconnect de type 1 (resp. de type 2).

Nous supposons que α et β ne dépendent pas de l'expérience des clients dans le système. Ces expériences comprennent le temps d'attente et le nombre de fois où les clients ont déjà essayé d'obtenir un service. Cela crée un système avec les files d'attente dépendantes les unes des autres. Un tel système de file d'attente sert de modèle à deux flux de travail concurrents dans un système d'accès multiple à détection de porteuse.

On a proposé un modèle fluide, qui est un analogue déterministe du modèle stochastique. Dans ce qui suit nous prouvons que le modèle stochastique original converge vers le modèle fluide sous une échelle appropriée.

III.2 Les limites fluides

Dans cette section, nous calculons $\mathbb{E}(Z_Q(t))$; $\mathbb{E}(Z_R(t))$ et $\mathbb{E}(Z_O(t))$, où $Z_Q(t)$ est le nombre de clients dans la file d'attente plus le nombre de client en service à l'instant t .

- $\mathbb{E}(Z_R(t)) = \mathbb{E}(Z_{R_1}(t)) + \mathbb{E}(Z_{R_2}(t))$ est le nombre de clients dans les files d'attente de redial 1 et 2 à l'instant t .
- $\mathbb{E}(Z_O(t)) = \mathbb{E}(Z_{O_1}(t)) + \mathbb{E}(Z_{O_2}(t))$ est le nombre de clients dans les orbites de reconnect 1 et 2 à l'instant t .

Ainsi, on pose

$$\mathbb{E}\Delta(t) = \lambda_1(t) + \lambda_2(t) + (\gamma_1 + \gamma_2)E(Z_R(t)) + (\theta_1 + \theta_2)E(Z_O(t)); \quad (\text{III.1})$$

où

- $\Delta(t)$ représente le nombre attendu d'arrivées jusqu'à l'instant t , qui est un processus stochastique.
- $\lambda_1(t) + \lambda_2(t)$ représente le taux d'arrivée externe au temps t .
- $(\gamma_1 + \gamma_2)\mathbb{E}(Z_R(t))$ représentent le taux d'arrivées moyen de radials orbites au temps t .
- $(\theta_1 + \theta_2)\mathbb{E}(Z_O(t))$ représentent le taux d'arrivée moyen de reconnects orbites au temps t .

Par conséquent, une fois que $\mathbb{E}(Z_Q(t))$; $\mathbb{E}(Z_R(t))$ et $\mathbb{E}(Z_O(t))$, $\mathbb{E}\Delta(t)$ peut être obtenu par l'équation (III.1).

Notons que $Z_Q(t)$ n'apparaît pas dans l'équation (III.1), mais nous verrons plus tard que $Z_R(t)$ et $Z_O(t)$ dépendent de $Z_Q(t)$.

En fait, il est facile de prouver que notre processus stochastique $\{Z(t); t \geq 0\}$, qui est désigné par

$$Z(t) := (Z_Q(t), Z_{R_1}(t), Z_{R_2}(t), Z_{O_1}(t), Z_{O_2}(t))^T; \quad (\text{III.2})$$

est un processus de Markov à 5 dimensions, L'espace d'état de ce processus de Markov est \mathbb{Z}_+^5 . Ce processus stochastique peut être à un processus de Markov tridimensionnel

$$(Z_Q(t); Z_R(t); Z_O(t))^T,$$

avec $Z_R(t) = Z_{R_1}(t) + Z_{R_2}(t)$ et $Z_O(t) = Z_{O_1}(t) + Z_{O_2}(t)$.

Ensuite, tout au long de ce chapitre nous concentrerons sur l'étude du processus de Markov tridimensionnel, car nous pensons qu'il n'est pas nécessaire d'étudier séparément chacune des orbites de reconnect et de radial.

Comme il s'agit d'un processus de Markov, nous pouvons tronquer le système à certains grands états, et obtenir numériquement la distribution de l'état stable de $Z(t)$ en résolvant les équations d'équilibre globales. Mais, pour le modèle que nous considérons, il est difficile de formuler et de résoudre les équations de l'équilibre global, et leur solution n'apporte pas d'informations sur le système. Par conséquent, pour des raisons pratiques, nous n'envisagerons pas de résoudre ce processus de Markov, mais d'autres méthodes d'approximation.

Nous présentons le modèle fluide, dont nous montrons qu'il apparaît comme la limite sous une mise à l'échelle appropriée du modèle stochastique présenté dans la figure III.1. Considérons un seul intervalle, les taux d'arrivée externe restant constants pendant cet intervalle ($\lambda_1(t) + \lambda_2(t) = \lambda_1 + \lambda_2, t \geq 0$).

Les équations de flux suivantes sont valables pour ce modèle stochastique

$$Z_Q(t) = Z_Q(0) + \Delta_{\lambda_1}(t) + \Delta_{\lambda_2}(t) + D_R(t) + D_O(t) - D_s(t) - D_a(t). \quad (\text{III.3})$$

Avec

$$D_{s_1}(t) + D_{s_2}(t) = D_s(t) \text{ et } D_{a_1}(t) + D_{a_2}(t) = D_a(t).$$

$$Z_R(t) = Z_R(0) + \sum_{j=1}^{D_a(t)} B_j(1 - \alpha) - D_R(t). \quad (\text{III.4})$$

$$Z_O(t) = Z_O(0) + \sum_{j=1}^{D_s(t)} B_j(1 - \beta) - D_O(t). \quad (\text{III.5})$$

Où

- $\Delta_{\lambda_1}(t) + \Delta_{\lambda_2}(t)$ est le nombre d'arrivées externes de type 1 et 2 dans l'intervalle de temps $[0; t)$. Les clients de type 1 et 2 arrivant au système suivant un processus de poisson de paramètre λ_1 et λ_2 respectivement.
- $D_R(t) = D_{R_1}(t) + D_{R_2}(t)$ est le nombre de clients dans redials 1 et 2 dans l'intervalle de temps $[0; t)$.
- $D_O(t) = D_{O_1}(t) + D_{O_2}(t)$; est le nombre de clients dans reconnects 1 et 2 dans l'intervalle $[0; t)$.
- $D_s(t) = D_{s_1}(t) + D_{s_2}(t)$; est le nombre de clients de type 1 et 2 qui sont servis dans l'intervalle $[0; t)$.
- $D_a(t) = D_{a_1}(t) + D_{a_2}(t)$ est le nombre de clients de type 1 et 2 abandonnés dans l'intervalle $[0; t)$.

Finalement $B_j(1 - \alpha)$ est une variable aléatoire de Bernoulli avec une probabilité $1 - \alpha$, $j = 1, \dots, D_{a_i}(t)$, tel que

$$B_j(1 - \alpha) = \begin{cases} 1, & \text{si le } j^{\text{ème}} \text{ client abandonné décide de rester dans} \\ & \text{le système et il rentre dans l'un des redials orbites;} \\ 0, & \text{sinon.} \end{cases}$$

Par conséquent, pour un $D_a(t)$ donné, nous avons $\sum_{j=1}^{D_a(t)} B_j(1 - \alpha) \rightsquigarrow Bin(D_a(t), 1 - \alpha)$.

Pour un $D_s(t)$ donné on a $\sum_{j=1}^{D_s(t)} B_j(1 - \beta) \rightsquigarrow Bin(D_s(t), 1 - \beta)$, ($Bin(.,.)$ est une distribution Binomiale).

Soit Δ_i des processus de Poisson indépendants de taux 1.

Ainsi,

$$D_s(t) = \Delta_1 \left(\int_0^t \mu \min \{s, Z_Q(u)\} du \right), \quad (III.6)$$

$$D_a(t) = \Delta_2 \left(\int_0^t \vartheta(Z_Q(u) - s)^+ du \right), \quad (III.7)$$

$$D_R(t) = \Delta_3 \left(\int_0^t (\gamma_1 + \gamma_2) Z_R(u) du \right), \quad (III.8)$$

$$D_O(t) = \Delta_4 \left(\int_0^t (\theta_1 + \theta_2) Z_O(u) du \right). \quad (III.9)$$

Nous n'avons pas besoin de donner la preuve de ces affirmations, il suffit de se référer à (Pang et al., 2007).

Pour introduire la limite fluide, nous considérons une séquence de modèles comme dans la figure III.1 telle que, dans la n ème modèle, les taux d'arrivées externe sont $\lambda_1 n$, $\lambda_2 n$ et le nombre de serveurs est ns . Nous ajoutons l'exposant " n " à toutes les notations du n -ième modèle. De manière similaire à (III.3)-(III.4)-(III.5), nous avons alors dans le n -ième modèle :

$$Z_Q^{(n)}(t) = Z_Q^{(n)}(0) + \Delta_{\lambda_1 n}^{(n)}(t) + \Delta_{\lambda_2 n}^{(n)}(t) + D_R^{(n)}(t) + D_O^{(n)}(t) - D_s^{(n)}(t) - D_a^{(n)}(t), \quad (III.10)$$

$$Z_R^{(n)}(t) = Z_R^{(n)}(0) + \sum_{j=1}^{D_a^{(n)}(t)} B_j(1 - \alpha) - D_R^{(n)}(t). \quad (\text{III.11})$$

$$Z_O^{(n)}(t) = Z_O^{(n)}(0) + \sum_{j=1}^{D_s^{(n)}(t)} B_j(1 - \beta) - D_O^{(n)}(t). \quad (\text{III.12})$$

Maintenant, en divisant par n les deux côtés des équations (III.10)-(III.11)-(III.12), nous avons

$$\bar{Z}_Q^{(n)}(t) = \bar{Z}_Q^{(n)}(0) + G_Q^{(n)}(\bar{Z}^{(n)})(t) + \int_0^t H_Q(\bar{Z}^{(n)})(u) du, \quad (\text{III.13})$$

$$\bar{Z}_R^{(n)}(t) = \bar{Z}_R^{(n)}(0) + G_R^{(n)}(\bar{Z}^{(n)})(t) + \int_0^t H_R(\bar{Z}^{(n)})(u) du, \quad (\text{III.14})$$

$$\bar{Z}_O^{(n)}(t) = \bar{Z}_O^{(n)}(0) + G_O^{(n)}(\bar{Z}^{(n)})(t) + \int_0^t H_O(\bar{Z}^{(n)})(u) du, \quad (\text{III.15})$$

Où

$$\begin{aligned} G_Q^{(n)}(\bar{Z}^{(n)})(t) &= \left(\frac{\Delta_{\lambda_1 n}^{(n)}(t) + \Delta_{\lambda_2 n}^{(n)}(t)}{n} - (\lambda_1 + \lambda_2)t \right) \\ &- \left(\bar{D}_s^{(n)}(t) - \int_0^t \mu \min \{ s, \bar{Z}_Q^{(n)}(u) \} du \right) \\ &- \left(\bar{D}_a^{(n)}(t) - \int_0^t \vartheta (\bar{Z}_Q^{(n)}(u) - s)^+ du \right) \\ &+ \left(\bar{D}_R^{(n)}(t) - \int_0^t (\gamma_1 + \gamma_2) \bar{Z}_O^{(n)}(u) du \right) \\ &+ \left(\bar{D}_O^{(n)}(t) - \int_0^t (\theta_1 + \theta_2) \bar{Z}_O^{(n)}(u) du \right) \end{aligned} \quad (\text{III.16})$$

$$\begin{aligned} G_R^{(n)}(\bar{Z}^{(n)})(t) &= \left(\sum_{j=1}^{n\bar{D}_a^{(n)}(t)} B_j(1 - \alpha)/n - \int_0^t (1 - \alpha) \vartheta (\bar{Z}_Q^{(n)}(u) - s)^+ du \right) \\ &- \left(\bar{D}_R^{(n)}(t) - \int_0^t (\gamma_1 + \gamma_2) \bar{Z}_R^{(n)}(u) du \right) \end{aligned} \quad (\text{III.17})$$

$$G_O^{(n)}(\bar{Z}^{(n)})(t) = \left(\sum_{j=1}^{n\bar{D}_s^{(n)}(t)} B_j(1-\beta)/n - \int_0^t (1-\beta)\mu \min\{s, \bar{Z}_Q^{(n)}(u)\} du \right) - \left(\bar{D}_O^{(n)}(t) - \int_0^t (\theta_1 + \theta_2)\bar{Z}_O^{(n)}(u) du \right). \quad (\text{III.18})$$

Et

$$\bar{D}_s^{(n)}(t) = \Delta_1 \left(n \int_0^t \mu \min\{s, \bar{Z}_Q^{(n)}(u)\} du \right) / n \quad (\text{III.19})$$

$$\bar{D}_a^{(n)}(t) = \Delta_2 \left(n \int_0^t \vartheta(\bar{Z}_Q^{(n)}(u) - s)^+ du \right) / n \quad (\text{III.20})$$

$$\bar{D}_R^{(n)}(t) = \Delta_3 \left(n \int_0^t (\gamma_1 + \gamma_2)\bar{Z}_R^{(n)}(u) du \right) / n \quad (\text{III.21})$$

$$\bar{D}_O^{(n)}(t) = \Delta_4 \left(n \int_0^t (\theta_1 + \theta_2)\bar{Z}_O^{(n)}(u) du \right) / n \quad (\text{III.22})$$

Et

$$\int_0^t H_Q(\bar{Z}^{(n)})(u) du = \int_0^t \lambda_1 + \lambda_2 + (\gamma_1 + \gamma_2)\bar{Z}_R^{(n)}(u) + (\theta_1 + \theta_2)\bar{Z}_O^{(n)}(u) - \mu \min\{s, \bar{Z}_Q^{(n)}(u)\} - \vartheta(\bar{Z}_Q^{(n)}(u) - s)^+ du \quad (\text{III.23})$$

$$\int_0^t H_R(\bar{Z}^{(n)})(u) du = \int_0^t (1-\alpha)\vartheta(\bar{Z}_Q^{(n)}(u) - s)^+ - (\gamma_1 + \gamma_2)\bar{Z}_R^{(n)}(u) du \quad (\text{III.24})$$

$$\int_0^t H_O(\bar{Z}^{(n)})(u) du = \int_0^t (1-\beta)\mu \min\{s, \bar{Z}_Q^{(n)}(u)\} - (\theta_1 + \theta_2)\bar{Z}_O^{(n)}(u) du \quad (\text{III.25})$$

Pour des raisons de commodité de notation, nous réécrivons les équations (III.13)-(III.14)-(III.15) sous la forme vectorielle

$$\bar{Z}^{(n)}(t) = \bar{Z}^{(n)}(0) + G^{(n)}(\bar{Z}^{(n)})(t) + \int_0^t H(\bar{Z}^{(n)})(u) du \quad (\text{III.26})$$

Où

$$G^{(n)}(\bar{Z}^{(n)})(t) = (G_Q^{(n)}(\bar{Z}^{(n)})(t), G_R^{(n)}(\bar{Z}^{(n)})(t), G_O^{(n)}(\bar{Z}^{(n)})(t))^T$$

$$H^{(n)}(\bar{Z}^{(n)})(u) = (H_Q^{(n)}(\bar{Z}^{(n)})(u), H_R^{(n)}(\bar{Z}^{(n)})(u), H_O^{(n)}(\bar{Z}^{(n)})(u))^T$$

Maintenant, nous allons définir le processus

$$\bar{Z}^n(t) = (\bar{Z}_Q^{(n)}(t), \bar{Z}_R^{(n)}(t), \bar{Z}_O^{(n)}(t))^T$$

Où

$$\bar{Z}_Q^{(n)}(t) = \frac{Z_Q^{(n)}(t)}{n}, \quad \bar{Z}_R^{(n)}(t) = \frac{Z_R^{(n)}(t)}{n}, \quad \bar{Z}_O^{(n)}(t) = \frac{Z_O^{(n)}(t)}{n}.$$

Soit $D([0, \infty), \mathbb{R}^3)$ l'espace des fonctions continues à droite avec limites à gauche dans \mathbb{R}^3 ayant pour domaine $[0, \infty)$. Nous dotons $D([0, \infty), \mathbb{R}^3)$ de la topologie habituelle de Skorokhod J_1 . Supposons que $\{X^{(n)}\}_{n=1}^\infty$ est une séquence de processus stochastique, alors $X^{(n)} \rightarrow^d x$ signifie que $X^{(n)}$ converge faiblement vers le processus stochastique X .

Définition III.2.1 *S'il existe une limite de distribution pour le processus échelonné $\{\bar{Z}^{(n)}(\cdot)\}_{n=1}^\infty$, i.e. $\bar{Z}^{(n)}(\cdot) \rightarrow^d z(\cdot)$, alors $z(\cdot)$ est appelé la limite fluide du modèle stochastique original.*

Déterminons maintenant la limite supérieure du modèle stochastique d'origine ; ainsi si

$$H_0 : (\bar{Z}_Q^{(n)}(0), \bar{Z}_R^{(n)}(0), \bar{Z}_O^{(n)}(0)) \rightarrow^d (z_Q(0), z_R(0), z_O(0)) \text{ lorsque } n \rightarrow \infty,$$

est vérifié, alors la limite fluide du modèle stochastique original est l'unique solution du système d'équations suivant

$$\begin{aligned} z_Q(t) = z_Q(0) + (\lambda_1 + \lambda_2)t + (\gamma_1 + \gamma_2) \int_0^t z_R(u)du + (\theta_1 + \theta_2) \int_0^t z_O(u)du - \\ \mu \int_0^t \min\{s, z_Q(u)\} du - \vartheta \int_0^t (z_Q(u) - s)^+ du \end{aligned} \tag{III.27}$$

$$z_R(t) = z_R(0) + (1 - \alpha)\vartheta \int_0^t (z_Q(u) - s)^+ du - (\gamma_1 + \gamma_2) \int_0^t z_R(u) du. \quad (\text{III.28})$$

$$z_O(t) = z_O(0) + (1 - \beta)\mu \int_0^t \min \{s, z_Q(u)\} du - (\theta_1 + \theta_2) \int_0^t z_O(u) du. \quad (\text{III.29})$$

Pour démontrer cette dernière, nous devons nous référer à (Ding et al.,2013), la preuve sera donnée de manière très similaire.

Maintenant, dans la section suivante, en utilisant les équations (III.27)-(III.28)-(III.29) données ci-dessus nous dérivons la limite d'utilité dans la stationnarité.

III.3 La stationnarité du modèle

Dans cette section, nous développons les conditions sous lesquelles $z(t)$ est constant. En dérivant les équations (III.27)-(III.28)-(III.29), nous obtenons

$$\lambda_1 + \lambda_2 = \mu \min \{s, z_Q(\infty)\} + \vartheta(z_Q(\infty) - s)^+ - (\gamma_1 + \gamma_2)z_R(\infty) - (\theta_1 + \theta_2)z_O(\infty). \quad (\text{III.30})$$

$$(\gamma_1 + \gamma_2)z_R(\infty) = (1 - \alpha)\vartheta(z_Q(\infty) - s)^+. \quad (\text{III.31})$$

$$(\theta_1 + \theta_2)z_O(\infty) = (1 - \beta)\mu \min \{s, z_Q(\infty)\}. \quad (\text{III.32})$$

Où $z_Q(\infty) = \lim_{t \rightarrow \infty} z_Q(t)$, $z_R(\infty) = \lim_{t \rightarrow \infty} z_R(t)$, $z_O(\infty) = \lim_{t \rightarrow \infty} z_O(t)$.

Les équations (III.30)-(III.31)-(III.32) peuvent être facilement résolues par rapport à $z_Q(\infty)$, $z_R(\infty)$ et $z_O(\infty)$, ce qui donne

$$z_Q(\infty) = \begin{cases} \frac{\lambda_1 + \lambda_2}{\mu\beta} & \text{si } \frac{\lambda_1 + \lambda_2}{s\mu} < \beta \\ \frac{\lambda_1 + \lambda_2 - \beta\mu s}{\vartheta\alpha} + s & \text{autrement} \end{cases} \quad (\text{III.33})$$

$$z_R(\infty) = \begin{cases} 0 & \text{si } \frac{\lambda_1 + \lambda_2}{s\mu} < \beta \\ \frac{(1-\alpha)\vartheta(z_Q(\infty) - s)}{\gamma_1 + \gamma_2} & \text{autrement} \end{cases} \quad (\text{III.34})$$

$$z_O(\infty) = \begin{cases} \frac{(1-\beta)\mu z_Q(\infty)}{\theta_1 + \theta_2} & \text{si } \frac{\lambda_1 + \lambda_2}{s\mu} < \beta \\ \frac{(1-\beta)\mu s}{\theta_1 + \theta_2} & \text{autrement} \end{cases} \quad (\text{III.35})$$

III.4 Discussion des résultats

D'abord, remarquons que $\frac{\lambda_1 + \lambda_2}{s\mu}$ est la charge du système due aux arrivées extérieures. $\frac{\lambda_1 + \lambda_2}{\mu\beta}$ est la charge totale lorsqu'il n'y a pas de redials. Remarquons ensuite que dans les expressions (III.33)-(III.34) et (III.35), la valeur de $\frac{\lambda_1 + \lambda_2}{\mu\beta}$ détermine si le modèle fluide est dans le trafic lourd ou non.

Ainsi, maintenant si $\frac{\lambda_1 + \lambda_2}{s\mu} < 1$, puisque la limite fluide est déterministe, nous obtenons $z_Q(\infty) < s$, et il n'y a pas de clients dans les deux orbites de redial ce qui signifie que $z_R(\infty) = 0$ serait valable, ce qui signifie également qu'il n'y a pas d'abandon du tout dans la limite d'utilisation.

En réalité, en raison de la variabilité de la durée du service et de la patience, les abandons ne seraient pas 0, mais très faible.

Maintenant, si $\frac{\lambda_1 + \lambda_2}{\mu\beta} > 1$, par l'équation (III.33), nous avons $z_Q(\infty) > s$.

Par conséquent, le modèle fluide indique qu'il y aura $z_Q(\infty) - s$ le nombre de clients en attente, chacun avec un taux ϑ , et que les clients seront acheminés vers les orbites de redial avec un taux $(1 - \alpha)\vartheta(z_Q(\infty) - s)$.

Conclusion

Les files d'attente sont un aspect inévitable de nombreux services, mais une gestion efficace peut considérablement améliorer l'expérience client. En utilisant des techniques d'optimisation basées sur la théorie des files d'attente, les entreprises peuvent réduire les temps d'attente, augmenter la satisfaction des clients, et améliorer leur efficacité opérationnelle.

Le concept de "redial" (rappel automatique) et "orbite" dans les files d'attente sont des aspects importants de la gestion des services clients, notamment dans les centres d'appels et les systèmes de support technique. La gestion efficace du redial et de l'orbite dans les files d'attente est cruciale pour améliorer l'expérience client et optimiser les opérations des centres d'appels. En utilisant des technologies avancées, des algorithmes optimisés et en augmentant les ressources disponibles, les entreprises peuvent réduire les temps d'attente, minimiser la frustration des clients et améliorer leur satisfaction globale.

Dans notre travail, nous avons étudié un centre d'appel avec deux redials orbites, deux reconnects orbites et deux types différents de flux exogènes. On a analysé un système de files d'attente $M/M/s$ avec rappels, deux orbites réservées aux clients impatientes dans le système dit "redial orbits" et deux autres réservées aux client revenant de l'extérieur pour d'autres services dits "reconnect orbits". Un modèle fluide est utilisé pour calculer l'approximation d'ordre premier pour le nombre de clients dans les orbites du système.

Les résultats obtenus, constituent une base de départ importante dans le domaine de communication, d'informatique et de fabrication.

Bibliographie

- [1] A. Mandelbaum, W. Massey.(1998). and M. Reiman. Strong approximations for markovian service networks. *Queueing Systems*, 30(1-2) :149-201 .
- [2] A. Mandelbaum, W. Massey, M. Reiman, and B. Rider.(1999). Time varying multiserver queues with abandonment and retrials. *In Proceedings of the 16th International Teletraffic Conference*, 4 :4-7.
- [3] A. Mandelbaum, W. Massey, M. Reiman, A. Stolyar, and B. Rider.(2002). Queue lengths and waiting times for multiserver queues with abandonment and retrials. *Telecommunication Systems*, 21(2-4) :149-171.
- [4] B. D. Choi and Y. C. Kim.(1998). The M/M/c Retrial Queue with Geometric Loss and Feedback, *Computers Math. Applic*, 36 (6) :41-52.
- [5] D.G.Kendall.(1951).Some problems in the theory of queues.J.Roy.Sta.Soc,13 :151 :-185.
- [6] D. Sze.(1984). Or practice - a queueing model for telephone operator staffing. *Operations Re – search*, 32(2) :229-249.
- [7] D. Samia.(2012). Sur le comportement asymptotique du système de files d’attente M/G/1 avec rappels. Mémoire de magister en Mathématiques. Université de Annaba.
- [8] Falin, G.I.(1988). On a multiclass batch arrival retrial queue. *Adv. Appl. Probab.* 20 :483-487.
- [9] J. R. Artalejo and G. I. Falin.(2002). Standard and retrial queueing systems : A comparative analysis. *Revista Mathematica Complutence*. p.101-129.
- [10] J. R. Artalejo and A. Gomez-Corral.(2008). Retrial Queueing systems. A Computational approach. Springer.
- [11] H. SAKHI et al.(2017). Analyse d’un réseau fluide multi-classe. PhD thesis.
- [12] G. Falin.(1990). A survey of retrial queues, *Queueing Systems : Theory and Applications*, 7(2) :127-167.
- [13] G. I Falin.(1990). A servey of retrial queues. *Queueing systems* 7. p. 127-168.

- [14] Grishechkin, S.A.(1992). Multiclass batch arrival retrial queues analyzed as branching processes with immigration. *Queueing Syst*, 11 :395-418.
- [15] G. I. Falin and J. G. C.(1997). Templeton, Retrial Queues. Monographs on statistics and Applied Probability, vol. 75. Chapman an Hall. London.
- [16] G. I. Falin and J. R. Artalejo.(1998). An infinite source retrial queue. *European Journal of Operational Research*, 108(2) :409-424.
- [17] G. Pang, R. Talreja, and W. Whitt.(2007) Martingale proofs of many-server heavy-traffic limits for markovian queues. *Probability Surveys*, 4 :193-267.
- [18] Kulkarni, V.G.(1986). Expected waiting times in a multiclass batch arrival retrial queue. *J. Appl. Probab*, 23 :144-154.
- [19] L. Berdjoudj.(2000). Stabilité forte dans les systèmes de files d’attente avec rappels. Thèse de magister en Mathématiques Appliquees. Universite de Bejaia.
- [20] L. Libman and A. Orda.(2002). Optimal retrial and timeout strategies for accessing network resources. *IEEE/ACM Transactions on Networking*, 10(4) :551-564.
- [21] Langaris, C., Dimitriou, I.(2010). A queueing system with n-phases of service and (n - 1)-types of retrial customers. *Eur. J. Oper. Res*, 205 :638-649.
- [22] Moutzoukis, E., Langaris, C.(1996). Non-preemptive priorities and vacations in a multiclass retrial queueing system. *Stoch Models*, 12(3) :455-472.
- [23] M.S. Aguir, F. Karaesmen, O.Z. Aksin, and F. Chauvet.(2004). The impact of retrials on call center performance. *OR Spectrum*, 26(3) :353-376
- [24] M.S. Aguir, O.Z. Akşin, F. Karaesmen, and Y. Dallery.(2008). On the interaction between retrials and sizing of call centers. *European Journal of Operational Research*, 191(2) :398-408.
- [25] N Gans, G Koole, and A Mandelbaum. (2002). *Telephone calls centers : a tutorial and literature review*.European Jour. Oper. Res.
- [26] N. Gharbi and M. Ioualalen.(2006). GSPN analysis of retrial systems with server breakdowns and repairs. *Applied Mathematics and Computation*, 174(2) :1151-1168.
- [27] O. Garnett, A. Mandelbaum, and M. Reiman.(2002). Designing a call center with impatient customers. *Manufacturing et Service Operations Management*, 4(3) :208-227.
- [28] P. Robert.(2000). Réseaux et files d’attente : méthodes probabilistes, Mathématiques & Applications, SMAI, Springer-Verlag Berlin Heidelberg.

-
- [29] R. Ibrahim and W. Whitt.(2009). Real-time delay estimation based on delay history. *Manufacturing & Service Operations Management*, 11(3) :397-415.
- [30] R. Ibrahim and W. Whitt.(2011). Wait-time predictors for customer service systems with time-varying demand and capacity. *Operations Research*, 59(5) :1106-1118.
- [31] S. Halfin and W. Whitt.(1981). Heavy-traffic limits for queues with many exponential servers. *Operations Research*, 29(3) :567-588.
- [32] S. Ding, R.D. van der Mei, and G. Koole.(2013). A method for estimation of redial and reconnect probabilities in call centers. In Proceedings of the 2013 Winter Simulation Conference, Winter Simulation Conference.
- [33] S. Ding , M. Remerova, Rob van der Mei, B. Zwart.(2013). *Fluid approximation of a Call Center model with redials and reconnects*, Cornell University Library,arXiv :1311.6248v1 [Math.PR].
- [34] W. Whitt.(2006). Fluid models for multiserver queues with abandonments. *Operations Research*, 54(1) :37-54.

ملخص

هذه الأطروحة مخصصة لتحليل نظام مركز اتصال تم نمذجته بنظام طابور $M/M/s$ مع عمليات إعادة الاستدعاء؛ مكونة من قائمة انتظار ذات حجم لا نهائي، نوعان من التدفقات تدفقات العملاء المستقلة التي تتبع عملية بواسون، الخادم s ، وهما مداران محجوزان للعملاء غير الصبورين، والمعروفين باسم مدار إعادة الاتصال المدار ومداران محجوزان للعملاء الذين لم يتسنى لهم الصبر ومداران محجوزان للعملاء العائدين من الخارج، ويطلق عليهما مدار إعادة الاتصال المدار. في هذه الدراسة، يتم استخدام نموذج مائع يُستخدم لحساب تقريب من الدرجة الأولى التقريبي من الدرجة الأولى لعدد العملاء في مدارات النظام.

Résumé

Ce mémoire est dédié à l'analyse d'un système de centre d'appel modélisé par un système de files d'attente $M/M/s$ avec rappels ; composé d'une file d'attente de taille infinie, deux type de flux de clients indépendants suivant le processus de Poisson, s serveur, deux orbites réservées aux flux de clients impatientes dites "redial orbit" et deux orbites réservées aux clients revenant de l'extérieur appelée "reconnect orbit". Dans cette étude, un modèle fluide est utilisé pour calculer une approximation de premier ordre pour le nombre de clients dans les orbites du système.

Abstract

This thesis is dedicated to the analysis of a call center system modeled by an $M/M/s$ queuing system with callbacks; composed of a queue of infinite size, two types of flow independent customer flows following the Poisson process, s server, two orbits reserved for impatient customers, known as "redial orbit" and two orbits reserved for customers for customers returning from abroad, called "reconnect orbit". In this study, a fluid model is used to calculate a first-order approximation for the number of customers in the system orbits.