

République Algérienne Démocratique et Populaire
Ministre de l'Enseignement Supérieur et de la Recherche Scientifique
Université Abderrahmane Mira-Bejaia



Faculté des Sciences Exactes
Département de Recherche Opérationnelle

MÉMOIRE DE FIN DE CYCLE

présenté par :

Drifa Mohamedi

THÈME

*File d'Attente avec Serveurs Hétérogènes,
Dérobade et Abandon de Clients avec
Seuil : Analyse par la Théorie des Jeux*

devant le jury composé de :

S. Adjabi	Prof	président
K. Adel	Prof	Encadreur
N. Khimoum	MCA	Co-Encadreur
S. Hakmi	MCA	Examinatrice
N. Bernine	MCB	Examinatrice

Année universitaire : 2024-2025

Remerciement

Je tiens tout d'abord à remercier Dieu, le Tout-Puissant, qui m'a accordé la force, le courage et la patience d'accomplir ce Modeste travail.

Je souhaite adresser mes sincères remerciements à mes encadrants, Monsieur **Khimoum Nourdine** et Madame **Adel Karima**, pour leur accompagnement précieux tout au long de ce mémoire. Leur disponibilité, leurs conseils éclairés, ainsi que leur rigueur scientifique ont été d'un grand soutien et ont grandement contribué à l'aboutissement de ce travail.

Mes remerciements sont aussi adressés aux membres du jury qui m'ont fait l'honneur d'accepter de consacrer leur temps à évaluer mon travail.

Je remercie également tous les enseignants du département de Recherche Opérationnelle pour la qualité de leur enseignement tout au long de mon parcours universitaire, ainsi que tous ceux qui, de près ou de loin, m'ont encouragé et soutenu durant cette période.

Je remercie du fond du cœur mes chers parents ainsi que l'ensemble de ma famille que Dieu les protège pour leur soutien inestimable tout au long de mes années d'études.

Dédicaces

Je dédie ce travail

- *À ma merveilleuse mère ,source d'inspiration inépuisable, En ce jour important , je te dédie cette déclaration.Tu es ma force et mon soutien , je te remercie jamais assez pour tout ton amour et ton soutien.*
- *À mon père, que le destin a rappelé trop tôt. Que ce mémoire soit un hommage à ta mémoire et à tout ce que tu as représenté pour moi. Ce travail, je l'aurais tant aimé le partager avec toi.*
- *À mon frère Chafaa et ma sœur Widade, Cette dédicace est pour vous est adressée, source de joie infinie.*
- *À toutes mes amies,et tout particulièrement à Fahem, mon compagnon et mon soutien précieux, merci d'avoir été à mes côtés tout au long de ce chemin.*

Drifa Mohamedi

Table des matières

Liste des figures	6
Introduction Générale	6
Introduction Générale	7
1 File d'attente classiques	9
1.1 Introduction	9
1.2 Description d'une file d'attente simple	9
1.2.1 File d'attente simple	9
1.2.2 Notation de Kendall	10
1.2.3 Loi de Little	12
1.3 Quelques Modèles de Files d'Attente	12
1.3.1 Modèle d'attente $M/M/1$	12
1.3.2 Modèle d'attente $M/M/1/N$	14
1.3.3 Modèle d'attente $M/M/C$	16
1.4 Conclusion	19
2 Etat de l'art	20
2.1 Introduction	20
2.2 Système $M/M/2/N$ avec serveurs hétérogènes, dérobade et abandon	20
2.3 Systèmes de files d'attente avec clients impatientes	21
2.4 Analyse d'un Modèle $M/M/c/N$ avec Dérobade et abandon	22
2.5 Analyse d'un Système $M/M/2$ sans Attente	22
2.6 gestion optimale des files d'attentes	23
2.7 Equilibre de Nash dans les réseaux de files d'attente	23
2.8 Strategic Behavior and Social Optimization in Markovian Queues . .	24
2.9 Modélisation des Files d'attente $M/M/c$ avec Clients Stratégiques . .	24
2.10 Analyse d'un Système $M/M/c$ Feedback et Clients Impatients	25
2.11 Conclusion	25
3 Modèle $M/M/2/N$ avec seuil de dérobade et abandon	26
3.1 Introduction	26
3.1.1 Systèmes de files d'attente avec dérobade	27
3.1.2 Systèmes de files d'attente avec abandon	30
3.2 Modèle $M/M/2/N$ avec dérobade et abandon	33

3.2.1	Description du modèle	33
3.3	Equations d'équilibre et leur solutions	34
3.4	Résolution des équations	35
3.4.1	Résolution pour différents cas de seuils	36
3.5	Paramètres de performance	37
3.6	Résultats numériques	38
4	Modélisation par la théorie des jeux	42
4.1	Introduction	42
4.2	Définitions et généralités	42
4.2.1	Qu'est ce qu'un jeu?	42
4.2.2	Représentation des jeux	44
4.2.3	Classification des jeux	44
4.2.4	Jeux finis et jeux infinis	46
4.2.5	Les jeux symétriques	46
4.2.6	Equilibre de Nash	46
4.3	Modélisation du système $M/M/2/N$	47
4.3.1	Joueurs	47
4.3.2	Taux d'arrivée et dérobade	48
4.3.3	Stratégies	48
4.3.4	Fonction d'utilité du client	50
4.4	Équilibre de Nash	50
4.4.1	Stratégies Optimales	50
4.4.2	Seuils d'entrée et d'abandon à l'équilibre	51
4.5	Résultats numériques	51
4.6	Conclusion	55
	Conclusion Générale	56
	Bibliographie	58

Table des figures

1.1	Système de file d'attente simple	10
1.2	Représentation schématique d'une file $M/M/c$	17
3.1	modèle $M/M/2/N$ avec seuils de dérobade et abandon	34
3.2	Diagramme d'états-transition	35
3.3	Probabilités d'occupation des états du système $M/M/2/N$ avec dérobade et abandon	39
3.4	Indicateurs de performance du système en régime permanent	40
3.5	Évolution des taux d'abandon et de dérobade en fonction des seuils K_1 et K_2	41
4.1	Exemple d'un Taux effectif d'arrivée en fonction de l'état n	48
4.2	Arbre de décision du comportement des clients dans le système	49
4.3	Évolution du seuil stratégique de dérobade K_1^* en fonction du nombre de clients	52
4.4	Évolution de l'utilité en fonction du nombre de clients dans le système.	53
4.5	Évolution de l'utilité et du temps d'attente moyen selon le nombre de clients dans le système	54

Introduction Générale

L'attente constitue aujourd'hui une composante essentielle des systèmes de service dans de nombreux domaines tels que les services bancaires, les télécommunications, les centres d'appels, les systèmes informatiques, les services publics, ou encore les environnements hospitaliers. Ces systèmes, soumis à des contraintes de ressources limitées face à une demande souvent aléatoire, nécessitent une analyse rigoureuse afin d'optimiser leur performance et garantir un niveau de service satisfaisant. L'analyse de ces phénomènes a conduit au développement d'un domaine de recherche spécialisé : la théorie des files d'attente.

Cette théorie des files d'attente constitue un outil fondamental à une branche de la recherche opérationnelle développée dès le début du XXe siècle par les travaux de l'ingénieur danois Agner Krarup Erlang (1878, 1929) sur le trafic téléphonique de Copenhague pour déterminer le nombre de circuits nécessaires afin de fournir un service téléphonique acceptable. Ensuite, les files d'attente ont été intégrés dans la modélisation des systèmes informatiques et aux réseaux de communication. Cette intégration, étendue à d'autres domaines, a contribué à l'évolution de la théorie, notamment en ce qui concerne l'évaluation des paramètres de performance de ces systèmes. Aujourd'hui, ce sont les applications liées à l'analyse de performance des réseaux tels que (la téléphonie mobile, Internet ou les systèmes multimédias, ...) qui suscitent le plus de travaux.

Les modèles classiques de files d'attente partent généralement du principe que les clients attendent jusqu'à obtention du service. Cependant, dans les situations réelles, les clients peuvent faire preuve d'impatience, ce qui se traduit par deux types de comportements : la dérobade (balking), où un client refuse d'entrer dans le système en constatant une file trop longue, et l'abandon (reneging), lorsqu'il décide de quitter la file après un certain temps d'attente sans avoir été servi. Ces comportements affectent de manière significative les performances globales du système.

Dans ce mémoire on s'intéresse à un système de files d'attente intégrant des phénomènes de dérobade, d'abandon et serveurs hétérogènes soumis à un seuil. L'attente peut rendre les clients impatients, ce qui nuit à la qualité du service. Cette impatience est directement liée au nombre de clients en attente. Pour limiter ces pertes, certaines mettent en place des méthodes d'amélioration continue visant à réduire, ou éliminer l'attente.

Par ailleurs, nous enrichissons notre cadre d'analyse en y intégrant la théorie des jeux, afin de modéliser les comportements stratégiques des clients. Cette étude vise à modéliser des situations où chaque client cherche à optimiser son propre gain

(réduire son temps d'attente ou son coût), tout en anticipant les choix des autres. L'objectif est de caractériser les équilibres de Nash entre les stratégies des clients et d'explorer des leviers d'optimisation sociale du système.

Les modèles de files d'attente avec dérobage, abandon, ou les deux, étaient traités par plusieurs chercheurs. Haight, [10] est le premier qui a étudié le modèle d'attente $M/M/1$ avec dérobage. Ensuite, Haight, [11] a traité ce même modèle $M/M/1$ mais avec abandon. Abou-EI-Ata et Hariri [1] ont considéré le modèle $M/M/c/N$ avec dérobage, abandon et plusieurs serveurs.

Ce mémoire se compose d'une introduction générale, de trois chapitres, d'une conclusion générale.

Dans le premier chapitre, nous commencerons par un rappel des notions fondamentales sur les files d'attente. Introduit les modèles classiques tels que $M/M/1$, $M/M/c$ et $M/M/2/N$, ainsi que les outils nécessaires à l'analyse des systèmes d'attente, en mettant l'accent sur les paramètres de performance.

Dans le deuxième chapitre, Nous présenterons quelques travaux existants intégrant les notions d'impatience et de comportements stratégiques.

Le troisième chapitre développe un modèle $M/M/2/N$ intégrant des seuils de dérobage et d'abandon. Une analyse théorique est menée pour établir les équations d'équilibre du système et en déduire les indicateurs de performance à l'aide de méthodes analytiques.

Enfin, le quatrième chapitre introduit la théorie des jeux comme outil d'analyse des comportements stratégiques des clients. Chaque client y est considéré comme un joueur rationnel, cherchant à maximiser son utilité individuelle. Des équilibres de Nash sont identifiés et étudiés, en vue de comprendre les interactions individuelles et leurs effets sur la performance globale du système. Le travail se conclut par une conclusion.

CHAPITRE

1

File d'attente classiques

1.1 Introduction

La théorie des files d'attente a commencé en 1909 ,à l'époque où Agner Krarup Erlang (1878,1929) a posé les bases dans ses recherches sur le trafic téléphonique. Par la suite, ses travaux ont été intégrés dans la recherche opérationnelle [8].

La théorie des files d'attente est une technique de la recherche opérationnelle qui permet de modéliser un système admettant un phénomène d'attente, de calculer ses performances et déterminer ses caractéristiques pour aider les praticiens dans leurs prises de décisions [28].

L'objet de ce chapitre est d'analyser deux paramètres essentiels :le nombre de clients dans le système et et la durée totale de leur présence.Cette dernière se décompose en un temps d'attente et un temps de service.

1.2 Description d'une file d'attente simple

1.2.1 FILE D'ATTENTE SIMPLE

La théorie des files d'attente vise à modéliser et à analyser de nombreuses situations différentes en apparence. Une file d'attente, est un système stochastique composé d'un espace d'attente (fini ou infini), d'un ou plusieurs serveurs et de clients

qui arrivent, attendent, sont servis selon des règles de priorité spécifiques, puis quittent le système. [15]

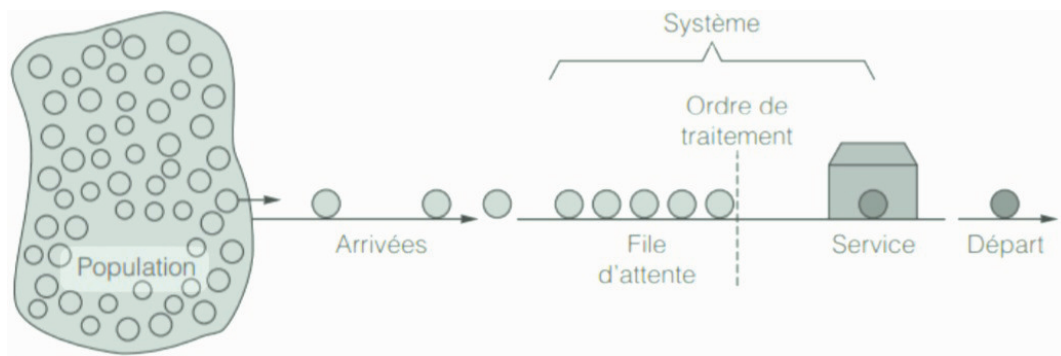


FIGURE 1.1 – Système de file d’attente simple .

Afin de spécifier un système de file d’attente, on se base sur trois éléments :

Processus d’arrivée

Pour le processus d’arrivée, on s’intéresse aux instants d’arrivés des clients dans la file. qui sont généralement aléatoires. Certaines hypothèses sont faites en se basant sur leurs lois. Tout d’abord, un seul client arrive à la fois. Ensuite, le processus est supposé homogène dans le temps, ce qui signifie que les temps interarrivées des clients suivent une même loi et sont indépendants. De plus, cette loi est supposée connue. Le cas le plus fréquent est celui où les temps interarrivées suivent une loi exponentielle, Dans ce cas, le modèle des temps d’arrivée des clients est un processus de Poisson. Evidemment d’autres distributions peuvent être envisagées, comme des temps interarrivées constantes, une loi uniforme, normale ou encore gamma.

Processus de service

Il comprend le nombre de serveurs et la loi probabiliste décrivant la durée des services.

Structure et discipline de la file

La discipline de service définit l’ordre selon lequel les clients sont placés dans la file et en sont extraits pour être servis.

1.2.2 NOTATION DE KENDALL

La notation suivante, appelée la notation de Kendall, est largement utilisée pour classer les différents systèmes de files d’attente [13] :

$$T/Y/C/K/m/Z$$

où

1. **T** : représente le processus d'arrivée des clients (distribution d'interarrivée). Les symboles utilisés sont :
 - (a) **M** (Markov) : inter-arrivées des clients suivent une loi exponentielle identique. Ce processus correspond à un processus de Poisson ponctuel, (propriété sans mémoire).
 - (b) **D** (Répartition déterministe) : Les temps inter-arrivées des clients ou les temps de service sont constants et toujours les mêmes.
 - (c) **GI** (général indépendant) : Inter-arrivées des clients ont une distribution générale (il n'y a aucune hypothèse sur la distribution mais les Inter-arrivées sont indépendantes et identiquement distribuées).
 - (d) **G** (général) : Inter-arrivées des clients ont une distribution générale (il n'y a aucune hypothèse sur la distribution mais les inter-arrivées sont indépendantes et identiquement distribuées).
 - (e) E_k : Ce symbole représente un processus dans lequel les intervalles de temps entre deux arrivées successives sont des variables aléatoires indépendantes et identiquement distribuées suivant une loi d'Erlang d'ordre k .
2. **Y** : Il représente la distribution des temps de service d'un client, en utilisant les mêmes codes que T .
3. **C** : nombre de serveurs.
4. **K** : capacité de la file correspond au nombre total de places disponibles dans le système, en d'autre terme le nombre maximal de clients pouvant être présents simultanément, y compris ceux en cours de service.
5. **m** : population des usagers.
6. **Z** : La discipline de service définit l'ordre selon lequel les clients sont priorisés pour être servis. Les codes utilisés sont les suivants :
7. **FIFO** (first in, first out) ou **FCFS** (first come first served) : c'est la file standard dans laquelle les clients sont servis dans leur ordre d'arrivée. Notons que les disciplines FIFO et FCFS ne sont pas équivalentes lorsque la file contient plusieurs serveurs.

Dans la première, le premier client arrivé sera le premier à quitter la file alors que la deuxième, il sera le premier à commencer son service. Rien n'empêche alors qu'un client qui commence son service après lui, dans un autre serveur, termine avant lui..
8. **LIFO** (last in, first out) ou **LCFS** (last come, first served) : Ce principe correspond à une pile, où le dernier client arrivé (placé en haut de la pile) est le premier à être servi. Une fois encore, les disciplines LIFO et LCFS ne sont équivalentes que pour une file mono serveur.
9. **SIRO** (Served In Random Order) : les clients sont servis aléatoirement.
10. **PNPN** (Priority service) : Les clients sont servis en fonction de leur priorité : ceux ayant la priorité la plus élevée sont traités en premier, suivis des clients de priorité inférieure, et ainsi de suite.
11. **PS** (Processor Sharing) : Les clients sont servis de manière équitable, la capacité du système étant répartie entre eux.

Remarque 1. Dans sa forme abrégée, seuls les trois premiers symboles $T/Y/C$ sont utilisés. Dans ce cas, il est supposé que la file suit une discipline FIFO et que le nombre de places d'attente ainsi que le nombre total de clients pouvant au système sont illimités.

1.2.3 LOI DE LITTLE

La loi de Little est une relation très général qui s'applique à une grande classe de systèmes.

Elle ne concerne que le régime permanent du système. Aucune hypothèse sur les variables aléatoires qui caractérisent le système (temps d'inter arrivées, temps de service,...) n'est nécessaire. La seule condition d'application de la loi de Little est que le système soit stable. le débit du système est alors indifféremment, soit le débit de sortie : $d_s = d_e = d$ La loi de Little s'exprime telle qu'elle est décrite dans le théorème suivant, qui est l'un des résultats les plus généraux et utiles dans la théorie des files d'attente.

Théorème 1. *Le nombre moyen de clients L , le temps moyen passé dans le système W et le débit moyen d d'un système stable en régime permanent sont liés par la relation suivante :*

$$L = W \times d$$

Remarque 2. *La loi de Little s'applique à tous les modèles de file d'attente rencontrés en pratique (pas seulement à la file $M/M/1$).*

1.3 Quelques Modèles de Files d'Attente

1.3.1 MODÈLE D'ATTENTE $M/M/1$

Le système de file d'attente $M/M/1$ est un modèle caractérisé par des arrivées suivant un processus de Poisson de taux λ , des temps de service suivant une loi exponentielle de paramètre μ , et un serveur unique. Les clients arrivent à la station selon un processus de Poisson de taux λ . Si le serveur est disponible, le client est immédiatement pris en charge ; sinon, il rejoint la file d'attente, dont la capacité est illimitée et qui convient à une discipline FIFO. Les temps inter-arrivées sont indépendants, tout comme les temps de service, qui suivent une distribution exponentielle de paramètre μ [17]. De plus, toutes les variables aléatoires impliquées sont supposées indépendantes les unes des autres. Les clients sont servis dans

l'ordre de leur arrivée. Cette file est un cas particulier du processus de naissance et de mort :

$$\forall n \in \mathbb{N}, \lambda_n = \lambda, \quad \mu_n = \mu,$$

c'est-à-dire, indépendants du nombre de clients dans le système. Sous l'hypothèse $\lambda < \mu$ (le taux d'arrivée est plus petit que le taux de service), on aura :

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

c'est l'intensité du trafic, avec $\rho < 1$:

$$\begin{cases} \pi_n = \pi_0 \rho^n \\ \pi_0 = \frac{1}{\sum_{n=0}^{\infty} \rho^n} = 1 - \rho \end{cases}$$

Remarque 3. π_0 représente la probabilité d'être servi immédiatement à l'arrivée.

La condition $\lambda < \mu$ est équivalent à $\rho < 1$. Ainsi, un régime stationnaire ne peut exister que si l'intensité de trafic est inférieure à 1. On peut également observer l'importance de ce paramètre lorsqu'on calcule le nombre moyen de clients en régime stationnaire, avec :

$$\sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu} \right)^n = \frac{1}{1 - \frac{\lambda}{\mu}}$$

π_n : c'est la probabilité d'avoir n client dans le système.

$$\pi_n = \rho^n \pi_0$$

avec :

$$\pi_0 = 1 - \rho$$

Débit d : Le débit du système (noté d) représente la probabilité pour que la file ne soit pas vide donc :

Ici, $d = \lambda$ car $\lambda_n = \lambda$ pour tout $n \geq 0$.

$$d = \sum_{n=0}^{\infty} \pi_n \mu = (1 - \pi_0) \mu = \rho \mu = \lambda$$

donc :

$$d = d_e = d_s$$

Taux d'utilisation du serveur U

Le taux d'utilisation du serveur (noté U) est la probabilité pour que le serveur de la file soit occupé.

$$U = 1 - \pi_0 = \rho \tag{1.1}$$

Nombre moyen de clients dans le système L

$$\sum_{n=0}^{\infty} n\pi_n = \frac{\rho}{1-\rho} = \frac{\frac{\lambda}{\mu}}{1-\frac{\lambda}{\mu}} = \frac{\lambda}{\mu-\lambda} \quad (1.2)$$

Temps moyen de séjour W

Le paramètre W est obtenu en utilisant la loi de Little :

$$W = \frac{L}{d} = \frac{1}{\mu(1-\rho)}, \quad (1.3)$$

qui peut être décomposé en :

$$W = \frac{1}{\mu} + \frac{\rho}{\mu(1-\rho)}. \quad (1.4)$$

Durée moyenne d'attente en régime stationnaire

$$E(W) = \frac{\rho}{\lambda(1-\rho)}. \quad (1.5)$$

Proposition : Soit $0 < \lambda < \mu$ Dans le système $M/M/1$, la variable aléatoire L , représentant la durée de séjour des clients dans le système $\mu(1-\rho) = \mu - \lambda$.

Temps moyen passé dans la file d'attente W_q

$$W_q = \frac{\rho}{\mu(1-\rho)}. \quad (1.6)$$

Nombre moyen de clients dans la file d'attente L_q

$$L_q = \lambda W_q = \frac{\rho^2}{1-\rho}. \quad (1.7)$$

1.3.2 MODÈLE D'ATTENTE $M/M/1/N$

On considère un système à serveur simple, similaire à un modèle $M/M/1/N$, à l'exception près que la capacité du fichier d'attente est limitée. Ainsi, les hypothèses de base sont les suivantes :

- Les arrivées des clients suivent un processus de Poisson avec un taux d'arrivée λ .
- Les temps de service suivent une loi exponentielle avec un taux de service μ .

- La capacité totale du système, c'est-à-dire la somme des clients dans le fichier d'attente et en cours de service, est limitée à N .

Lorsqu'un client arrive alors que le système contient déjà N clients, il ne peut pas entrer et est immédiatement perdu. Ce type de système est désigné sous le nom de fichier $M/M/1/N$. Dans ce contexte, l'espace des états E est borné et peut être décrit par :

$$E = \{0, 1, 2, \dots, N\} \quad (1.8)$$

où chaque état n représente le nombre de clients présents dans le système, incluant celui en cours de service. Etant donné que la capacité de la file est limitée, même si le taux d'arrivée des clients est nettement supérieur au taux de traitement du serveur, une fois la file pleine, les nouveaux arrivants sont immédiatement rejetés. Par conséquent, le nombre de clients dans le système ne peut jamais tendre vers l'infini [22]. De plus, dès qu'un client est admis dans le système, il finira par en sortir après un temps fini, correspondant au temps de service des clients qui le précèdent, ce nom étant essentiellement né par N . Ainsi, le débit de sortie du système est toujours égal au débit d'entrée, garantissant ainsi la stabilité inconditionnelle du système. Donc ce processus est considéré comme un processus de naissance et de mort, avec

- Un taux de naissance $\lambda_n = \lambda$ pour tout $n < N$.
- Un taux de mortalité $\mu_n = \mu$ pour tout $n \neq 0$.

Soit π_n , $n = 0, 1, \dots, N$, la probabilité pour qu'il ait n clients dans le système à l'instant t .

$$\begin{aligned} \lambda\pi_0 &= \mu\pi_1 \\ (\lambda + \mu)\pi_n &= \lambda\pi(n-1) + \mu\pi(n+1), \quad n = 1, 2, \dots, N-1 \\ \lambda\pi(N-1) &= \mu\pi(N) \end{aligned} \quad (1.9)$$

Le calcul de π_n se fait de la manière suivante :

$$\pi_n = \begin{cases} \frac{(1-\rho)\rho^n}{1-\rho^{N+1}} & \text{pour } n \leq N, \\ 0 & \text{pour } n > N. \end{cases} \quad (1.10)$$

Dans le cas particulier où $\rho = 1$, on a :

$$\pi_n = \begin{cases} \frac{1}{N+1}, & \text{pour } n \leq N, \\ 0, & \text{pour } n > N. \end{cases} \quad (1.11)$$

Taux d'utilisation du serveur U

Le taux d'utilisation du serveur, (noté U), représente la probabilité que le serveur soit occupé, c'est-à-dire la probabilité qu'il y ait au moins un client dans la file.

$$U = \sum_{n=1}^N \pi_n = 1 - \pi_0 = \frac{\rho(1-\rho^N)}{1-\rho^{N+1}} \quad (1.12)$$

Remarque 4.

$$\lim_{N \rightarrow +\infty} U(N) = \begin{cases} \rho, & \text{si } \rho < 1, \\ 1, & \text{si } \rho > 1. \end{cases}$$

qui représente le taux d'utilisation du serveur dans le cas de la file $M/M/1$.

Nombre moyen de clients L

$$\begin{aligned} L &= \sum_{n=0}^N n\pi_n = \frac{(1-\rho)}{1-\rho^{N+1}} \sum_{n=0}^N n\rho^n \\ &= \frac{\rho(1-\rho)}{1-\rho^{N+1}} \sum_{n=1}^N n\rho^{n-1} \\ &= \frac{\rho}{1-\rho} \frac{1-(N+1)\rho^N + N\rho^{N+1}}{1-\rho^{N+1}} \end{aligned}$$

Pareil que le taux d'utilisation du serveur, lorsque $N \rightarrow +\infty$ on obtient :

$$L = \frac{\rho}{1-\rho} \tag{1.13}$$

qui représente le nombre moyen de clients pour la file $M/M/1$.

Temps moyen de séjour W

Le temps moyen de séjour W dans la file est déterminé en utilisant la formule de Little, qui établit une relation entre le nombre moyen de clients L dans le système et le débit d de la file :

$$W = \frac{L}{d} \tag{1.14}$$

1.3.3 MODÈLE D'ATTENTE $M/M/C$

Le modèle d'attente $M/M/c$ est un système formé d'une file à capacité infinie, avec c serveurs identiques et la discipline de service est FIFO. Supposons que le flot des arrivées est poissonien de paramètre λ , la durée de service est exponentielle de paramètre. On a le processus $\{X(t), t \geq 0\}$ du nombre de clients dans le système à

l'instant t est un processus de naissance et de mort de taux de transition :

$$\lambda_n = \lambda, \quad \forall n \geq 0 \quad (1.15)$$

$$\mu_n = \begin{cases} n\mu, & \text{pour } n = 1, 2, \dots, c-1 \\ c\mu, & \text{pour } n \geq c \end{cases} \quad (1.16)$$

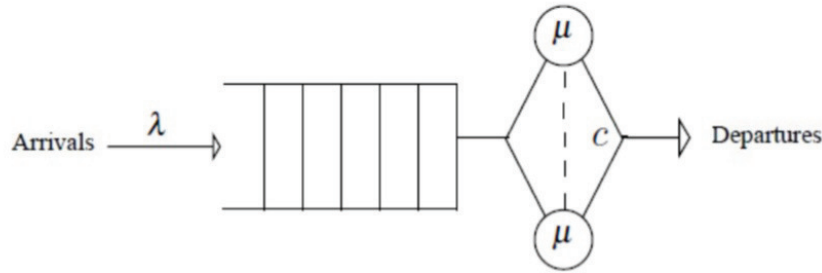


FIGURE 1.2 – Représentation schématique d'une file M/M/c.

Remarque 5. Pour calculer π_n , le système doit être stable $\lambda < c\mu$ donc $\rho < c$.

Cela traduit le fait que le nombre moyen de clients arrivant à la file par unité de temps doit être inférieur à la capacité moyenne de traitement des serveurs sur la même période, Alors :

$$\pi_n = \begin{cases} \pi_0 \frac{\rho^n}{n!} & \text{pour } n = 1, 2, \dots, c-1 \\ \pi_0 \frac{\rho^n c^{c-n}}{c!} & \text{pour } n \geq c \end{cases} \quad (1.17)$$

avec

$$\pi_0 = \left[\sum_{n=0}^{c-1} \frac{\rho^n}{n!} + \frac{\rho^c}{c!} \frac{1}{1 - \frac{\rho}{c}} \right]^{-1} \quad (1.18)$$

Remarque 6. pour la valeur de $c = 1$ on obtient :

$$\pi_n = (1 - \rho)\rho^n \quad (1.19)$$

qui représente le cas de la file M/M/1

Débit d

Lorsque le nombre de clients dans le système est inférieur à C , ils sont pris en charge par les C serveurs à un taux de service de $n\mu$. En revanche, lorsque le

système compte plus de C clients, le taux de service devient $c\mu$.

$$d = \sum_{n=1}^{c-1} \pi_n n \mu + \sum_{n=c}^{\infty} \pi_n c \mu \quad (1.20)$$

En remplaçant les expressions obtenues pour les probabilités π_n et π_0 , on trouve bien que la file est stable

$$d = \lambda$$

Remarque 7. Pour des raisons de calcul, dans le cas de la file $M / M / c$, il est plus pratique de déterminer d'abord le temps moyen de séjour, puis d'en déduire le nombre moyen de clients dans le système.

Temps moyen de séjour W

Le temps moyen de séjour d'un client se compose en deux parties : le temps moyen passé dans la file d'attente et le temps moyen de service. Il suffit alors d'appliquer la loi de Little pour le déterminer.

$$W = W_q + W_s = \frac{L_q}{d} + \frac{1}{\mu} = \frac{L_q}{\lambda} + \frac{1}{\mu}$$

Il reste alors à calculer le nombre moyen de clients en attente dans la file, L_q :

$$L_q = \sum_{n=C}^{+\infty} (n - c) \pi_n = \frac{\rho^{c+1}}{(C-1)!(C-\rho)^2} \pi_0$$

On en déduit l'expression du temps moyen de séjour

$$W = \frac{\rho^c}{\mu(c-1)!(c-\rho)^2} \pi_0 + \frac{1}{\mu}$$

Nombre moyen de clients L

La loi de Little permet de trouver le nombre moyen de clients dans la file :

$$L = W \times d = W \times \lambda = \frac{\rho^{c+1}}{(c-1)!(c-\rho)^2} \pi_0 + \rho$$

1.4 Conclusion

Dans ce chapitre, nous avons présenté les principaux concepts de la théorie des files d'attente, en explorant leurs structures, leurs notations et leurs modèles classiques. L'analyse des systèmes a permis de comprendre leur comportement et leurs performances sous différentes configurations. Ces bases théoriques permettent d'aborder des modèles plus complexes, comme celui étudié dans le chapitre suivant, intégrant des phénomènes tels que la dérobade et l'abandon des clients.

CHAPITRE

2

Etat de l'art

2.1 Introduction

La modélisation des files d'attente constitue un outil essentiel pour analyser les performances des systèmes de service.

Dans ce chapitre, nous présentons un panorama synthétique et structuré des travaux existants portant sur les systèmes $M/M/2/N$ intégrant des comportements réalistes comme la dérobade, l'abandon, et plus particulièrement les travaux ayant utilisé des approches de modélisation mathématique permettant d'analyser l'effet de l'impatience des clients sur la performance et la stabilité des systèmes de files d'attente.

2.2 Système $M/M/2/N$ avec serveurs hétérogènes, dérobade et abandon

Le mémoire de "Khadidja HAFSI"[25] s'intéresse à l'étude d'un système de files d'attente à capacité limitée, intégrant deux serveurs aux performances différentes et accueillant jusqu'à N clients simultanément. Dans ce système, les clients peuvent soit renoncer à entrer lorsqu'ils perçoivent une attente trop longue (dérobade), soit quitter la file après une certaine période sans avoir été servis (abandon). Ces comportements liés à l'impatience sont fréquents dans les contextes de service comme les centres d'appel, les administrations ou les services de santé.

Le modèle considéré est de type $M/M/2/N$, caractérisé par des arrivées suivant

un processus de Poisson et des durées de service exponentielles. L'originalité de l'étude réside dans la combinaison simultanée de la dérobade, de l'abandon et de l'hétérogénéité des serveurs dans un même cadre théorique. À l'aide des processus de naissance et de mort, les équations d'équilibre sont formulées puis résolues à travers une approche numérique.

Des expérimentations numériques permettent d'analyser l'effet de divers paramètres (taux d'arrivée, capacité du système, taux de service, etc.) sur les indicateurs de performance. Les résultats obtenus offrent un éclairage pertinent sur les impacts combinés de l'impatience des clients et des différences entre les serveurs, apportant ainsi des éléments utiles à la gestion et à l'optimisation des systèmes de service.

2.3 Systèmes de files d'attente avec clients impatients

Dans l'article de "KADI MOKHTAR" [19], l'auteur a présenté une étude approfondie de deux modèles avancés de files d'attente réalistes intégrant divers comportements d'impatience et de réitération propres aux clients : dérobade, abandon, rappels et feedback. Le premier modèle est une file $M/M/2/N$ à deux serveurs hétérogènes avec capacité finie, dans laquelle les clients peuvent refuser de joindre la file (dérobade), quitter sans être servis (abandon), et la rétroaction, où les clients peuvent revenir dans la file après un service incomplet. À cela s'ajoute la possibilité de rétention, c'est-à-dire la capacité à convaincre certains clients impatients de rester malgré leur intention initiale de partir. Les auteurs s'appuient sur la théorie des chaînes de Markov pour établir les équations du régime stationnaire, permettant de calculer les probabilités d'état stable du système. Ces résultats sont ensuite utilisés pour évaluer différentes mesures de performance telles que le nombre moyen de clients dans le système et dans la file, les temps d'attente moyens, ainsi que le taux moyen de clients servis. Les résultats obtenus confirment les comportements attendus : par exemple, une augmentation du taux d'arrivée ou une forte probabilité de rétroaction tend à augmenter la congestion du système, tandis qu'un taux élevé de renoncement ou de refus d'entrée réduit la taille de la file d'attente. Ce modèle enrichi, combinant plusieurs dynamiques complexes, offre un outil pertinent pour l'analyse et la gestion de systèmes de service dans lesquels le comportement des clients joue un rôle déterminant.

Un autre travail est celui de "BENDAHMA SAIDA" [24], où l'auteur s'est intéressé à des systèmes de files d'attente, en mettant l'accent sur les modèles probabilistes utilisés pour analyser leur stabilité et leur performance. L'objectif central de cette étude est d'analyser les effets de certains comportements clients, notamment la dérobade (balking) et l'abandon (reneging), sur la dynamique et la stabilité des systèmes de type $M/M/c/N$, avec un intérêt particulier pour le cas $M/M/2/N$. Ces comportements sont modélisés à l'aide de seuils, selon lesquels les clients prennent des décisions stratégiques : soit ils choisissent de ne pas entrer dans la file s'ils la file trop longue, soit ils la quittent après un certain temps d'attente. Ces comportements influencent directement la dynamique du système, particulièrement la

structure des équations d'équilibre de type Chapman-Kolmogorov.

La modélisation repose sur des outils tels que les chaînes de Markov, les processus de naissance et de mort, et les martingales, permettant d'analyser les conditions de stabilité modifiées par l'impatience des clients.

Ce travail apporte ainsi une contribution significative à la modélisation des files d'attente à serveurs multiples intégrant des comportements stratégiques dépendants de seuils, et ouvre des perspectives pour l'optimisation des performances dans des environnements contraints en capacité et en qualité de service.

2.4 Analyse d'un Modèle M/M/c/N avec Dérobade et abandon

Dans cet article de "Abou El-Ata et Hariri"[1] proposent une analyse approfondie d'un système de file d'attente M/M/c/N, caractérisé par un nombre limité de places (N), plusieurs serveurs (c), des arrivées de clients selon un processus de Poisson, et des temps de service suivant une distribution exponentielle. Le modèle se distingue par l'intégration de deux comportements réalistes des clients : le balking, les clients choisissent de ne pas rejoindre le système lorsqu'il perçoit la file comme trop longue, et le renoncement (reneging), qui correspond à l'abandon de la file après une attente excessive.

Les auteurs formalisent ce système à l'aide d'un modèle markovien et établissent les équations permettant d'atteindre l'état stationnaire. À partir de là, ils déduisent plusieurs mesures de performance, telles que les probabilités d'entrée et d'abandon, la distribution du nombre de clients dans le système, ainsi que des indicateurs de performance globale. Leur analyse permet de quantifier l'influence du comportement des clients sur le fonctionnement du système, en mettant en évidence l'impact de la capacité, du nombre de serveurs, du taux d'arrivée et de la patience des clients sur la qualité du service.

Ce travail fournit ainsi un cadre analytique pertinent pour modéliser et optimiser des systèmes de service à capacité limitée, dans lesquels les décisions des clients influencent directement les performances du système.

2.5 Analyse d'un Système M/M/2 sans Attente

L'article de "BOUCHENTOUF et MESSABIHI"[6] propose une analyse d'un système de file d'attente à deux serveurs hétérogènes, sans espace pour une file d'attente. Les clients arrivent selon un processus de Poisson et peuvent quitter le service s'ils deviennent impatients avant d'être servis. Un mécanisme permet toutefois de retenir certains clients avec une certaine probabilité. Une fois le service terminé, un client peut soit quitter définitivement le système, soit y revenir pour un nouveau service selon une rétroaction de type Bernoulli. En s'appuyant sur la théorie des processus de Markov, les auteurs établissent les équations du régime permanent et déterminent les probabilités d'état stationnaire. À partir de ces résultats, plusieurs mesures de performance sont calculées, telles que le nombre moyen

de clients dans le système, le taux de renoncement, de rétention, d'abandon, et la proportion de clients perdus. Des exemples numériques illustrent l'impact de différents paramètres sur le comportement du système, montrant notamment l'effet des taux de service, de renoncement et de rétroaction.

2.6 gestion optimale des files d'attentes

Dans cet article de "Naor"[21], s'intéresse à une méthode visant à améliorer la gestion des files d'attente : faire payer un droit d'entrée aux clients. À travers un modèle mathématique, il analyse le comportement des clients qui arrivent de façon aléatoire à une station-service. Chaque client doit décider s'il rejoint la file d'attente en supportant un coût lié au temps d'attente ou s'il renonce au service sans rien perdre ni gagner. Naor montre que lorsque les clients agissent selon leur intérêt personnel, cela peut provoquer une surcharge du système et nuire à son efficacité. Pour éviter cette congestion, il propose l'instauration d'un péage à l'entrée de la file, destiné à dissuader les clients dont la participation n'est pas socialement bénéfique. Ce péage joue un rôle de filtre économique, améliorant ainsi la fluidité et le bien-être collectif. Cependant, si le péage vise uniquement à maximiser les revenus d'un gestionnaire, il peut devenir excessif et nuire à l'efficacité globale. Naor conclut que les péages sont utiles à condition qu'ils servent un objectif d'optimisation sociale, et suggère même que, dans certains cas, des subventions peuvent être préférables aux péages.

2.7 Equilibre de Nash dans les réseaux de files d'attente

L'article de "d'Altman et Shimkin"[2] analyse un système où des clients arrivent séquentiellement et doivent choisir entre deux options : un serveur partagé (MF) dont la performance diminue avec le nombre d'utilisateurs, ou un serveur privé (PC) avec un coût fixe. Chaque client cherche à minimiser son propre temps de service, ce qui rend sa décision dépendante des choix futurs des autres utilisateurs. Ce problème est modélisé comme un jeu dynamique non coopératif.

Les auteurs démontrent que la stratégie optimale de chaque client suit une règle de seuil, et qu'un équilibre de Nash symétrique unique existe. Ce seuil détermine si le client rejoint ou non le serveur partagé en fonction de la charge observée. L'article étudie aussi comment ce seuil peut être calculé, notamment dans le cas d'arrivées de Poisson.

Ensuite, un modèle d'apprentissage est introduit, dans lequel les clients adaptent leurs décisions en fonction des statistiques de performance passées, sans connaissance a priori du système. Ce modèle converge vers l'équilibre de Nash, montrant que l'équilibre peut émerger de manière naturelle.

Enfin, l'étude compare l'équilibre individuel avec l'optimum social, montrant que les décisions individuelles peuvent entraîner une surcharge du serveur partagé par rapport à une politique socialement optimale.

2.8 Strategic Behavior and Social Optimization in Markovian Queues

Ce livre de "Refael Hassin"[12] explore l'application de la théorie des jeux à la gestion des files d'attente, en mettant l'accent sur les comportements stratégiques des clients dans des systèmes markoviens. L'auteur modélise les décisions des clients (rejoindre, attendre, abandonner) comme des jeux non coopératifs où chacun maximise son utilité individuelle. Il examine d'abord le cas classique $M/M/1$, puis étend l'analyse aux modèles multi-serveurs ($M/M/c$), aux files à capacité finie ($M/M/1/N$), et à divers comportements comme le balking (refus d'entrer) et le reneging (abandon en cours d'attente).

Hassin introduit les concepts d'équilibre de Nash, en les adaptant aux environnements stochastiques des files d'attente. Il distingue entre les systèmes à information complète (où les clients connaissent la taille de la file) et à information incomplète. Il analyse aussi les conséquences sociales de ces comportements stratégiques : les décisions optimales pour les individus mènent souvent à un équilibre sous-optimal pour la société (perte d'efficacité collective).

Le livre aborde également des mécanismes d'incitation (tarification, priorités, pénalités) pour aligner les intérêts individuels et sociaux. Enfin, des extensions vers les modèles avec impatience, priorité, ou incertitude sur le service sont proposées.

2.9 Modélisation des Files d'attente M/M/c avec Clients Stratégiques

Cet article de "Economou et Kanta"[7] étudie un système de files d'attente Markovien multi-serveurs dans lequel les clients choisissent de rejoindre la file la plus courte, avec la possibilité de dérobage (balking) et abandon (reneging). Chaque client prend une décision stratégique basée sur la longueur actuelle des files et son propre niveau d'impatience. Le système est modélisé comme un processus de décision stochastique, avec une dynamique de type $M/M/c$. modélisation le système comme une chaîne de Markov à états finis, intégrant la dynamique des arrivées, des départs, des balkings et des renegings. Ils calculent les probabilités d'équilibre où aucun client ne bénéficie d'un changement unilatéral de stratégie, c'est-à-dire un équilibre de Nash en stratégie mixte ou pure.

L'étude met en lumière l'influence de ces comportements sur la performance globale du système, En particulier sur le temps moyen d'attente et la probabilité de perte (clients qui abandonnent ou ne rentrent pas). Les résultats montrent que l'introduction du reneging et du balking modifie fortement la dynamique du système, rendant nécessaires des adaptations dans la gestion des files d'attente afin d'optimiser la qualité de service.

2.10 Analyse d'un Système $M/M/c$ Feedback et Clients Impatients

L'article [14] analyse un modèle de file d'attente $M/M/c$ dans lequel les serveurs prennent des vacances synchronisées, soit uniques, soit multiples. Les clients sont impatients : ils peuvent refuser d'entrer (balking) ou quitter la file avant d'être servis (reneging), mais certains peuvent être retenus pour attendre leur tour. Le système prévoit aussi un mécanisme de retour (feedback), permettant à certains clients de revenir après leur service. En utilisant des fonctions génératrices de probabilité, déterminer les distributions stationnaires et calculent plusieurs indicateurs de performance, comme le nombre moyen de clients ou le taux d'abandon. Un modèle économique permet d'évaluer les coûts, revenus et bénéfices. Des simulations numériques illustrent l'effet des paramètres sur la performance globale du système. L'étude met en évidence l'intérêt d'une bonne gestion des vacances et de la rétention des clients.

2.11 Conclusion

Ce chapitre a présenté un aperçu des études sur les files d'attente intégrant l'impatience des clients. Il constitue une base pour développer dans le prochain chapitre étude d'un modèle $M/M/2/N$ avec seuil de dérobade et abandon.

CHAPITRE

3

Modèle $M/M/2/N$ avec seuil de déro- bade et abandon

3.1 Introduction

Dans de nombreux domaines, les clients impatients, découragés soit par la qualité de service soit par la longueur de la file d'attente ou abandonnés carrément la file, ont fait l'objet de nombreuses études. Les systèmes comportant des clients impatients ont entraîné des pertes économiques significatives pour de nombreuses entreprises [3]. Dans ce chapitre on s'intéresse aux files d'attente :

- Avec dérobage : Un client impatient qui décide ne pas rejoindre le système.
- Avec abandon : Après un temps passé dans la file, le client impatient décide de quitter le système sans avoir le service.

Les files d'attente avec dérobage, abandon (lorsqu'un client impatient quitte la file après un certain temps sans être servi) ou les deux ont suscité de nombreuses études. Haight, 1957, est le premier qui a étudié le modèle d'attente $M/M/1$ avec dérobage [?], puis, en 1959, il a traité ce même modèle $M/M/1$ avec abandon [11]. Ancker et Gafarian [3, 4] ont ensuite étudié l'impact de la combinaison du dérobage et de l'abandon sur un fichier $M/M/1/N$. Les auteurs dans le papier [1], ont étendu cette analyse au modèle $M/M/c/N$, intégrant plusieurs serveurs ainsi que les phénomènes de dérobage et d'abandon. Enfin, Wang et Chang [27] ont approfondi ces travaux en examinant le modèle $M/M/c/N$ intégrant dérobage, abandon et pannes de serveurs. Actuellement, les recherches se concentrent principalement sur les applications liées à l'analyse de performance des réseaux, notamment dans les domaines de la téléphonie mobile, d'Internet et du multimédia.

3.1.1 SYSTÈMES DE FILES D'ATTENTE AVEC DÉROBADE

Supposons qu'un client arrive dans un système et constate la présence de n clients devant lui. Il a le choix :

- Soit il rejoint la file avec une probabilité P_n .
- Soit il part avec une probabilité $q_n = 1 - P_n$.

Si la longueur de la file d'attente décourage les clients, alors P_n est une fonction décroissante de n . Un cas particulier de ce phénomène se produit lorsque le système dispose d'une salle d'attente de capacité c . Dans ce cas, nous pouvons supposer que :

$$P_n = \begin{cases} 1, & \text{si } n < c, \\ 0, & \text{si } n \geq c. \end{cases}$$

Cela signifie que, dès que la salle d'attente atteint sa capacité maximale, aucun nouveau client ne peut entrer dans le système.

Soit $X(t)$ le nombre de clients présents dans le système à l'instant t . Le processus d'arrivée des clients suit un processus de Poisson de taux λ . Lorsqu'un client arrive et constate la présence de n clients dans le système, il décide d'y entrer avec une probabilité P_n . Ainsi, le processus de naissance correspondant est donné par un taux effectif d'arrivée de :

$$\lambda_n = \lambda P_n, \quad \text{pour } n = 0, 1, \dots$$

Dans le cas d'un système à serveur unique, le taux de service est constant et noté $\mu_n = \mu$ pour $n = 1, 2, \dots$. Nous pouvons alors déterminer la distribution stationnaire de la longueur du fichier d'attente. Dans les systèmes avec abandon, tous les clients entrants accèdent initialement au système, et certains sont perdus. À l'état stationnaire, le taux effectif d'entrée des clients dans le système est donné par :

$$\lambda_n = \lambda \sum_{n=0}^{\infty} \pi_n P_n$$

Le taux de perte des clients est donné par :

$$\lambda \sum_{n=0}^{\infty} \pi_n q_n$$

Le taux de perte des clients à long terme est donné par :

$$\sum_{n=0}^{\infty} \pi_n q_n$$

Nous analysons à présent en détail le cas d'un système $M/M/c$. Un client ne peut entrer dans le système que s'il ya au moins un serveur est libre. Pour celà, nous

avons :

$$\lambda_k = \begin{cases} \lambda, & \text{pour } k = 0, 1, \dots, c-1, \\ 0, & \text{pour } k = c. \end{cases}$$

et

$$\mu_k = k\mu \quad \text{pour } k = 0, 1, \dots, c$$

Afin de trouver la distribution, nous procédons à l'analyse suivante :

$$\pi_k = \frac{1}{k!} \left(\frac{\lambda}{\mu} \right)^k \quad \text{pour } k = 0, 1, \dots, c$$

Par conséquent, nous obtenons ainsi le résultat suivant :

$$\pi_k = \frac{\frac{1}{k!} \left(\frac{\lambda}{\mu} \right)^k}{\sum_{j=0}^c \frac{1}{j!} \left(\frac{\lambda}{\mu} \right)^j}, \quad \text{pour } k = 0, 1, \dots, c.$$

La fraction des clients qui ne peuvent pas entrer dans le système est donnée par :
 $\pi_c q_c = \pi_c$ d'où $q_c = 1$.

Taux de service

De manière similaire, on peut considérer un système où le taux de service dépend du nombre de clients dans le système. Par exemple, dans un système initialement doté d'un seul serveur, un deuxième serveur peut être ajouté dès que la longueur de la file d'attente dépasse un seuil critique. Si les arrivées suivent un processus de Poisson et que les taux de service possèdent la propriété d'absence de mémoire, alors les paramètres de naissance et de mort appropriés sont les suivants :

$$\lambda_k = \lambda, \quad \text{pour tout } k \geq 0.$$

et

$$\mu_k = \begin{cases} \mu, & \text{si } n \leq \eta, \\ 2\mu, & \text{si } n > \eta. \end{cases}$$

De manière plus générale, nous considérons un processus d'arrivées de Poisson avec $\lambda_k = \lambda$ pour tout $k \geq 0$ et des taux de service arbitraires μ_k pour $k \geq 0$. Dans ce cas, la distribution stationnaire est donnée par :

$$\pi_k = \pi_0 \frac{\lambda^k}{\mu_1 \mu_2 \cdots \mu_k}, \quad \text{pour } k \geq 1.$$

où La condition de normalisation nous conduit à l'expression suivante :

$$\pi_0 = \left[1 + \sum_{k=1}^{\infty} \frac{\lambda^k}{\mu_1 \mu_2 \cdots \mu_k} \right]^{-1} .$$

Modèle de files d'attente $M/M/1$ avec dérobade

À son arrivée dans le système, le client découragé par la longueur de la file d'attente, il décide de ne pas rejoindre la file. Ce phénomène, appelé dérobade. Cette impatience a motivé de nombreux mathématiciens à étudier ce phénomène pour trouver des solutions adaptées [9].

Description du modèle Dans ce modèle, lorsqu'un client arrive et trouve k clients devant lui, il peut devenir impatient et choisir de ne pas rejoindre la file d'attente. Notons b_k la probabilité qu'un client rejoigne le système lorsqu'il ya k clients présents à son arrivée. Ce modèle a été étudié dans [?] Ainsi, le nombre de clients dans le système suit un processus de naissance et de mort, avec des taux de naissance définis par :

pour $k = 0, 1, \dots$

$$\lambda_k = \lambda b_k$$

tel que

$$b_k = \frac{1}{k+1}$$

et

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

c'est l'intensité du trafic. ainsi

$$\pi_k = \frac{\rho^k}{k!} \pi_0$$

(d'après Chapman-Kolmogorov)[26]

En appliquant la condition de normalisation, nous obtenons l'expression suivante :

$$\pi_k = \frac{\rho^k}{k!} e^{-\rho}$$

La condition de stabilité dans ce cas est $E(\rho) < \infty$. Contrairement au système $M / M / 1$, il n'est pas nécessaire d'imposer la condition $\rho < 1$.

Les paramètres de performance sont définis comme suit :

Taux d'utilisation du serveur U :

$$U_s = 1 - \pi_0 = 1 - e^{-\rho}$$

$$= \frac{E(\delta)}{\frac{1}{\lambda} + E(\delta)}$$

avec $E(\delta)$ est la période moyenne d'occupation du serveur.
Par conséquent :

$$E(\delta) = \frac{1}{\lambda} \cdot \frac{1 - e^{-\rho}}{e^{-\rho}}$$

Nombre moyen de clients dans le système L :

$$L = \rho$$

et

$$\text{Var}(L) = \rho$$

Nombre moyen de clients dans la file d'attente L_q :

$$L_q = L - U_s = \rho - (1 - e^{-\rho}) = \rho + e^{-\rho} - 1$$

Temps moyen de séjour W :

$$W = \frac{\rho}{\mu(1 - e^{-\rho})}$$

et on déduit le temps moyen passé dans la file d'attente

$$W_q = W - \frac{1}{\mu} = \frac{1}{\mu} \left(\frac{\rho + e^{-\rho} - 1}{1 - e^{-\rho}} \right)$$

Débit d : on a aussi

$$d = \sum_{k=0}^{\infty} \lambda_k \pi_k = \sum_{k=1}^{\infty} \mu_k \pi_k = \mu(1 - e^{-\rho})$$

d'où

$$d \cdot W = \mu(1 - e^{-\rho}) \cdot \frac{\rho}{\mu(1 - e^{-\rho})} = \rho = L$$

qui est la formule de Little pour ce système [18].

3.1.2 SYSTÈMES DE FILES D'ATTENTE AVEC ABANDON

La théorie des files d'attente avec abandon joue un rôle essentiel dans la modélisation de nombreux problèmes du monde réel. Ces applications s'étendent à divers domaines tels que (l'informatique, les télécommunications, l'industrie), ainsi que les secteurs de la santé et des sciences médicales, entre autres.

Dans cette section, nous étudions un modèle de file d'attente avec abandon [23]. Dans ce modèle, un client peut quitter le système s'il estime avoir attendu trop longtemps sans être servi.

Le modèle est basé sur les hypothèses suivantes :

- Le système comporte un serveur unique, Les clients arrivent selon un processus de Poisson avec un taux dépendant du nombre de clients présents, noté $\lambda(n+1)$.
- Les temps de service suivent une loi exponentielle de paramètre μ et sont indépendants les uns des autres.
- Les clients sont servis selon la discipline FIFO .
- La capacité du système est limitée de taille N .
- Lorsqu'un client arrive dans le système, il patiente un certain temps pour être servi. Si ce service ne commence pas, il peut devenir impatient et choisir de quitter le système avec une probabilité P , ou bien rejoint dans la file avec une probabilité $q = 1 - P$ Le temps avant l'abandon suit une distribution exponentielle de paramètre α .

À l'état stable, la dynamique du système est modélisée par un processus de naissance et de mort . Les équations différentielles associées concernent l'évolution des probabilités P_n , représentant la probabilité qu'il y ait n clients dans le système à un instant donné t .

$$\frac{d}{dt}P_0(t) = -\lambda P_0(t) + \mu P_1(t) \quad (3.1)$$

$$\frac{d}{dt}P_n(t) = -[\lambda(n+1) + \mu + (n-1)\alpha P] P_n(t) + (\mu + n\alpha P)P_{n+1}(t) + \left(\frac{\lambda}{n}\right)P_{n-1}(t), \quad n = 1, \dots, N-1 \quad (3.2)$$

$$\frac{d}{dt}P_N(t) = -[\mu + (N-1)\alpha P] P_N(t) + \left(\frac{\lambda}{N}\right)P_{N-1}(t) \quad (3.3)$$

En état d'équilibre, $\lim_{t \rightarrow +\infty} P_n(t) = P_n$ et, $\frac{d}{dt}P_n(t) = 0$ et donc, la solution des équations (1) à (3) donne :

$$-\lambda P_0 + \mu P_1 = 0 \quad (3.4)$$

$$-[\lambda(n+1) + \mu + (n-1)\alpha P] P_n + (\mu + n\alpha P)P_{n+1} + \left(\frac{\lambda}{n}\right)P_{n-1}, \quad n = 1, \dots, N-1 = 0 \quad (3.5)$$

$$-[\mu + (N-1)\alpha P] P_N + \left(\frac{\lambda}{N}\right)P_{N-1} = 0 \quad (3.6)$$

En résolvant les équations de (4) à (6), nous obtenons

$$P_n(t) = - \left[P_n(t) = \frac{1}{n!} \prod_{k=1}^n \frac{\lambda}{\mu + (k-1)\alpha P} \right] P_0 \quad n = 1, \dots, N \quad (3.7)$$

En utilisant la condition de normalisation, $\sum_{n=0}^N P_n = 1$ nous obtenons

$$P_0 = \frac{1}{1 + \sum_{n=1}^N \left(\frac{1}{n!} \prod_{k=1}^n \frac{\lambda}{\mu + (k-1)\alpha P} \right)} \quad (3.8)$$

Par conséquent, la distribution stationnaire du processus permet de déterminer d'autres paramètres de performance du système, tels que :

Nombre moyen de clients dans le système L_s :

$$L_s = \sum_{n=1}^N n \left[\frac{1}{n!} \prod_{k=1}^n \frac{\lambda}{\mu + (k-1)\alpha P} \right] P_0$$

Nombre moyen de clients dans la file L_q :

$$L_q = \left[\sum_{n=1}^N n \left[\frac{1}{n!} \prod_{k=1}^n \frac{\lambda}{\mu + (k-1)\alpha P} \right] P_0 \frac{\lambda}{\mu} \right]$$

Temps moyen de séjour dans le système W_s :

$$W_s = \left[\frac{1}{\lambda} \sum_{n=1}^N n \left[\frac{1}{n!} \prod_{k=1}^n \frac{\lambda}{\mu + (k-1)\alpha P} \right] \right] P_0$$

Temps moyen de séjour dans la file W_q :

$$W_q = \left[\frac{1}{\lambda} \sum_{n=1}^N n \left[\frac{1}{n!} \prod_{k=1}^n \frac{\lambda}{\mu + (k-1)\alpha P} \right] \right] P_0 - \frac{1}{\lambda}$$

Le nombre moyen de clients servis $E(\text{client servis})$:

$$E(\text{clients servis}) = \sum_{n=1}^N n \mu P_n$$

$$E(\text{clients servis}) = \sum_{n=1}^N n \mu \left[\frac{1}{n!} \prod_{k=1}^n \frac{\lambda}{\mu + (k-1)\alpha P} \right] P_0$$

Taux d'abandon R_{abond} :

$$R_{abond} = \lambda \sum_{n=0}^N P_n - E(\text{clients servis})$$

$$R_{\text{abond}} = \lambda - \sum_{n=1}^N n\mu \left[\frac{1}{n!} \prod_{k=1}^n \frac{\lambda}{\mu + (k-1)\alpha P} \right] P_0$$

Le nombre de clients en attente E(client en attente) :

$$E(\text{clients en attente}) = \frac{\sum_{n=2}^N (n-1)P_n}{\sum_{n=2}^N P_n}$$

$$E(\text{clients en attente}) = \frac{\sum_{n=2}^N (n-1) \left[\frac{1}{n!} \prod_{k=1}^n \frac{\lambda}{\mu + (k-1)\alpha P} \right] P_0}{\sum_{n=2}^N \left[\frac{1}{n!} \prod_{k=1}^n \frac{\lambda}{\mu + (k-1)\alpha P} \right] P_0}$$

La probabilité de la période d'occupation P(période d'occupation) :

$$P(\text{période d'occupation}) = \sum_{n=1}^N \left[\frac{1}{n!} \prod_{k=1}^n \frac{\lambda}{\mu + (k-1)\alpha P} \right] P_0$$

Où P_0 est la valeur calculée en (3.8).

3.2 Modèle $M/M/2/N$ avec dérobade et abandon

Dans cette section, on va étudier un modèle de files d'attente $M/M/2/N$, où la capacité du système est limitée à N clients. Ce modèle intègre une fonction de dérobade et d'abandon, ainsi que deux serveurs hétérogènes. Les clients sont servis selon la discipline FCFS (First Come, First Served). Ensuite, nous calculerons les paramètres de performance de ce système.

- **Avec dérobade :** le client qui refuse d'entrer dans la file d'attente s'il voit que la file d'attente est très grande ou le temps d'attente est trop long.

- **Avec abandon :** Après un temps passé dans la file, le client impatient décide de quitter le système sans avoir le service.

3.2.1 DESCRIPTION DU MODÈLE

Les clients arrivent dans le système selon un processus de Poisson avec un taux d'arrivée $\lambda(n)$ qui dépend du nombre total de clients n déjà présents dans le système. Les dispositions suivantes sont applicables :

Taux d'arrivée et dérobade

1. Si le nombre de clients n est inférieur à un seuil K_1 , le taux d'arrivée effectif de paramètre $\lambda(n) = \lambda$.
2. Si $n \geq K_1$, le taux d'arrivée est réduit à $\lambda(n) = \lambda(1 - \epsilon)$ où ϵ est la probabilité qu'un client renonce à joindre la file.

Abandon

1. Si le nombre de clients n est inférieur à un seuil $K_2, \gamma(n) = 0$ pas d'abandon.
2. Si $n \geq K_2, \gamma(n) = \gamma \cdot (n-1)$ taux d'abandon proportionnel au nombre de clients en attente, où γ est le taux d'impatience individuelle).

La probabilité qu'un client abandonne pendant un intervalle de temps court Δt est : $r_n = \gamma(n)\Delta t = \gamma \cdot (n-1)\Delta t$. Les temps de service des clients suivent une distribution exponentielle avec des taux spécifiques :

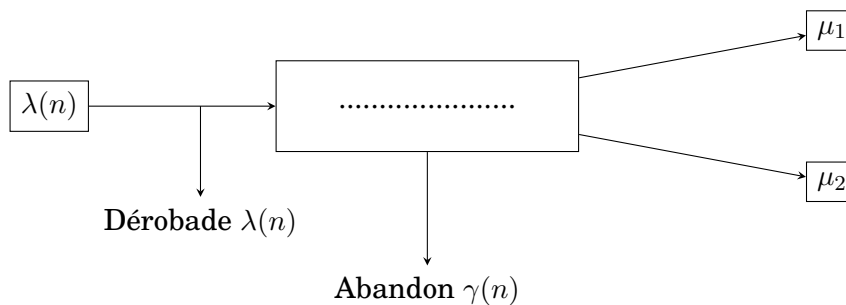
μ_1 : Pour le premier serveur.

μ_2 : Pour le deuxième serveur.

($\mu_3 = \mu_1 + \mu_2$) : est le taux de service combiné des deux serveurs.

Le service des clients suit les règles suivantes :

1. Si les deux serveurs sont occupés, les nouveaux clients rejoignent la file d'attente.
2. Si l'un des deux serveurs est libre, le client à la tête de la file l'occupe.
3. Si les deux serveurs sont libres, le client en tête de la file choisit :
 - (a) Le serveur 1 avec une probabilité π_1 .
 - (b) Le serveur 2 avec une probabilité π_2 .
 - (c) Avec la contrainte $\pi_1 + \pi_2 = 1$.



Seuil de dérobade :

$$\lambda(n) = \lambda \quad \text{si } n < K_1$$

$$\lambda(n) = \lambda(1 - \varepsilon) \quad \text{si } n \geq K_1$$

Taux d'abandon :

$$\gamma(n) = 0 \quad \text{si } n < K_2$$

$$\gamma(n) = \gamma \cdot (n-1) \quad \text{si } n \geq K_2$$

FIGURE 3.1 – modèle $M/M/2/N$ avec seuils de dérobade et abandon

3.3 Equations d'équilibre et leur solutions

Nous définissons les probabilités d'équilibre comme suit :

- $P_0 = P(\text{Il n'y a aucun client dans le système})$.
- $P_{1'} = P(\text{Il y a un client entrain d'être servi par le serveur 1})$.
- $P_{1''} = P(\text{Il y a un client entrain d'être servi par le serveur 2})$.

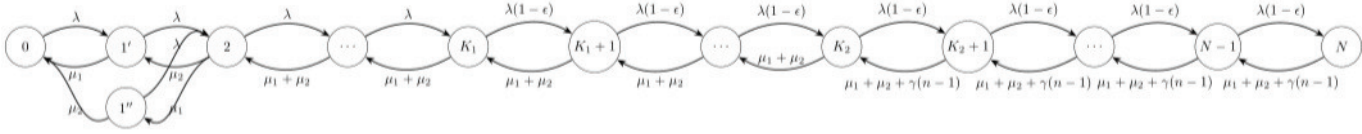


FIGURE 3.2 – Diagramme d'états-transition

- $P_n = \text{P(II y a } n \text{ clients dans le système)}$, $n = 2, 3, \dots, N$.

De plus, on pose :

- $P_1 = P_{1'} + P_{1''}$ P(Il y a un client soit en service, quel que soit le serveur).

Par conséquent, les équations de probabilité à l'état stable sont :

Pour $n = 0$:

$$\lambda(0)P_0 = \mu_1 P_{1'} + \mu_2 P_{1''} \quad (3.9)$$

Pour $n = 1$:

$$\begin{cases} [\lambda(1) + \mu_1]P_{1'} = \mu_2 P_2 + \lambda(0)\pi_1 P_0 \\ [\lambda(1) + \mu_2]P_{1''} = \mu_1 P_2 + \lambda(0)\pi_2 P_0 \end{cases} \quad (3.10)$$

où $\pi_1 + \pi_2 = 1$ représentent les probabilités de choisir chaque serveur lorsque les deux sont libres.

Pour $n = 2$:

$$[\lambda(2) + \mu_3]P_2 = [\mu_3 + \gamma(3)]P_3 + \lambda(1)P_1 \quad (3.11)$$

Pour $3 \leq n < N$:

$$[\lambda(n) + \mu_3 + \gamma(n)]P_n = [\mu_3 + \gamma(n+1)]P_{n+1} + \lambda(n-1)P_{n-1} \quad (3.12)$$

Pour $n = N$:

$$[\mu_3 + \gamma(N)]P_N = \lambda(N-1)P_{N-1} \quad (3.13)$$

3.4 Résolution des équations

En résolvant l'équation pour $n = 1$ et en utilisant celle pour $n = 0$, on déduit :

$$P_{1'} = \frac{\lambda(0)[\lambda(1) + \pi_1 \mu_3]}{\mu_1 [2\lambda(1) + \mu_3]} P_0 \quad (3.14)$$

$$P_{1''} = \frac{\lambda(0)[\lambda(1) + \pi_2 \mu_3]}{\mu_2 [2\lambda(1) + \mu_3]} P_0 \quad (3.15)$$

Donc :

$$P_1 = P_{1'} + P_{1''} = \theta P_0 \quad (3.16)$$

où :

$$\theta = \frac{\lambda(0)[\lambda(1)\mu_3 + \mu_3(\pi_1\mu_2 + \mu_1\pi_2)]}{\mu_1\mu_2[\lambda(1) + \mu_3]} \quad (3.17)$$

Pour les valeurs de $n \geq 2$, nous devons tenir compte des seuils K_1 et K_2 .

3.4.1 RÉSOLUTION POUR DIFFÉRENTS CAS DE SEUILS

Cas 1 : $n < \min(K_1, K_2)$

Dans ce cas, il n'y a ni dérobade ni abandon. Les équations d'équilibre se simplifient :

$$[\lambda + \mu_3]P_n = \mu_3 P_{n+1} + \lambda P_{n-1} \quad (3.18)$$

En posant $\rho = \frac{\lambda}{\mu_3}$, c'est l'intensité du trafic, avec $\rho < 1$ on obtient :

$$P_n = \rho^{n-1} \cdot \theta P_0, \quad 2 \leq n < \min(K_1, K_2) \quad (3.19)$$

Cas 2 : $K_1 \leq n < K_2$

Dans ce cas, il y a dérobade mais pas d'abandon :

$$[\lambda(1 - \epsilon) + \mu_3]P_n = \mu_3 P_{n+1} + \lambda(1 - \epsilon)P_{n-1} \quad (3.20)$$

En posant $\rho' = \frac{\lambda(1-\epsilon)}{\mu_3}$, c'est l'intensité du trafic, avec $\rho' < 1$ on obtient :

$$P_n = \rho^{K_1-1} \cdot (\rho')^{n-K_1} \cdot \theta P_0, \quad K_1 \leq n < K_2 \quad (3.21)$$

Cas 3 : $K_2 \leq n < K_1$

Dans ce cas, il y a abandon mais pas de dérobade :

$$[\lambda + \mu_3 + \gamma(n)]P_n = [\mu_3 + \gamma(n+1)]P_{n+1} + \lambda P_{n-1} \quad (3.22)$$

La résolution devient plus complexe et nécessite une approche récursive :

$$P_n = \rho^{n-1} \cdot \prod_{i=K_2}^{n-1} \frac{\lambda}{\mu_3 + \gamma(i+1)} \cdot \theta P_0, \quad K_2 \leq n < K_1 \quad (3.23)$$

Cas 4 : $n \geq \max(K_1, K_2)$

Dans ce cas, il y a à la fois dérobade et abandon :

$$[\lambda(1 - \epsilon) + \mu_3 + \gamma(n)]P_n = [\mu_3 + \gamma(n+1)]P_{n+1} + \lambda(1 - \epsilon)P_{n-1} \quad (3.24)$$

Qui donne :

$$P_n = \rho^{K_1-1} \cdot (\rho')^{n-K_1} \cdot \prod_{i=K_2}^{n-1} \frac{\lambda(i)}{\mu_3 + \gamma(i+1)} \cdot \theta P_0, \quad n \geq \max(K_1, K_2) \quad (3.25)$$

La probabilité P_0 est calculée en utilisant la condition de normalisation :

$$\sum_{n=0}^N P_n = 1 \quad (3.26)$$

Alors, la probabilité d'état vide P_0 s'exprime par :

$$P_0 = \left[1 + \Theta \cdot \left(1 + \sum_{n=2}^N f(n) \right) \right]^{-1}$$

où $f(n)$ est défini par cas :

$$f(n) = \begin{cases} \rho^{n-1} & \text{si } n < \min(K_1, K_2) \\ \rho^{K_1-1} \cdot (\rho')^{n-K_1} & \text{si } K_1 \leq n < K_2 \\ \rho^{n-1} \cdot \prod_{i=K_2}^{n-1} \frac{\lambda}{\mu_3 + \gamma(i+1)} & \text{si } K_2 \leq n < K_1 \\ \rho^{K_1-1} \cdot (\rho')^{n-K_1} \cdot \prod_{i=K_2}^{n-1} \frac{\lambda(1-\varepsilon)}{\mu_3 + \gamma(i+1)} & \text{si } n \geq \max(K_1, K_2) \end{cases}$$

Ce modèle M/M/2/N avec seuils de dérobade et d'abandon s'inspire des travaux de Abou El-Ata et Hariri [1], Bendahma Saida [24] et Kadi Mokhtar [19], qui ont étudié des systèmes à capacité limitée intégrant des comportements d'impatience.

3.5 Paramètres de performance

Nombre moyen de clients dans le système (L_s) :

$$\begin{aligned} L_s &= \sum_{n=1}^N n P_n \\ &= \theta \left[1 + \sum_{n=2}^N n f(n) \right] P_0 \end{aligned}$$

Nombre moyen de clients dans la file d'attente (L_q) :

$$L_q = \sum_{n=2}^N (n-2) P_n$$

$$= \theta \left[1 + \sum_{n=2}^N (n-2)f(n) \right] P_0$$

Taux d'arrivée effectif :

$$\lambda_{eff} = \sum_{n=0}^{N-1} \lambda(n)P_n$$

Temps moyen de séjour dans le système (W_s) :

$$W_s = \frac{L_s}{\lambda_{eff}}$$

Temps moyen de séjour dans la file (W_q) :

$$W_q = \frac{L_q}{\lambda_{eff}}$$

Taux de d'abandon :

$$R_{aband} = \sum_{n=K_2}^N \gamma(n-1)P_n$$

Taux de dérobage :

$$R_{dérob} = \lambda\epsilon \sum_{n=K_1}^N P_n$$

3.6 Résultats numériques

Cette section présente les résultats issus de la résolution numérique du modèle $M/M/2/N$ avec serveurs hétérogènes, seuil de dérobage K_1 , seuil d'abandon K_2 , ainsi que les comportements stochastiques associés (probabilité de dérobage ϵ et taux d'abandon γ).

La Figure (3.3), illustre la distribution stationnaire des probabilités d'état P_n , représentant la probabilité qu'il y ait exactement n clients dans le système, pour $n \in \{0, 1, \dots, 10\}$. Cette distribution a été obtenue en résolvant le système de files d'attente $M/M/2/N$ avec des seuils comportementaux (seuil de dérobade K_1 et seuil d'abandon K_2).

On observe que la probabilité est maximale pour les états $n = 1$ et $n = 2$, ce qui signifie que le système est le plus souvent faiblement chargé. La probabilité décroît ensuite rapidement à mesure que le nombre de clients augmente. Au-delà du seuil de dérobade ($n \geq 4$) et du seuil d'abandon ($n \geq 6$), cette décroissance accentuée résulte de l'effet combiné de la dérobade et de l'abandon. Cela indique une autorégulation du système, où les comportements des clients limitent la surcharge.

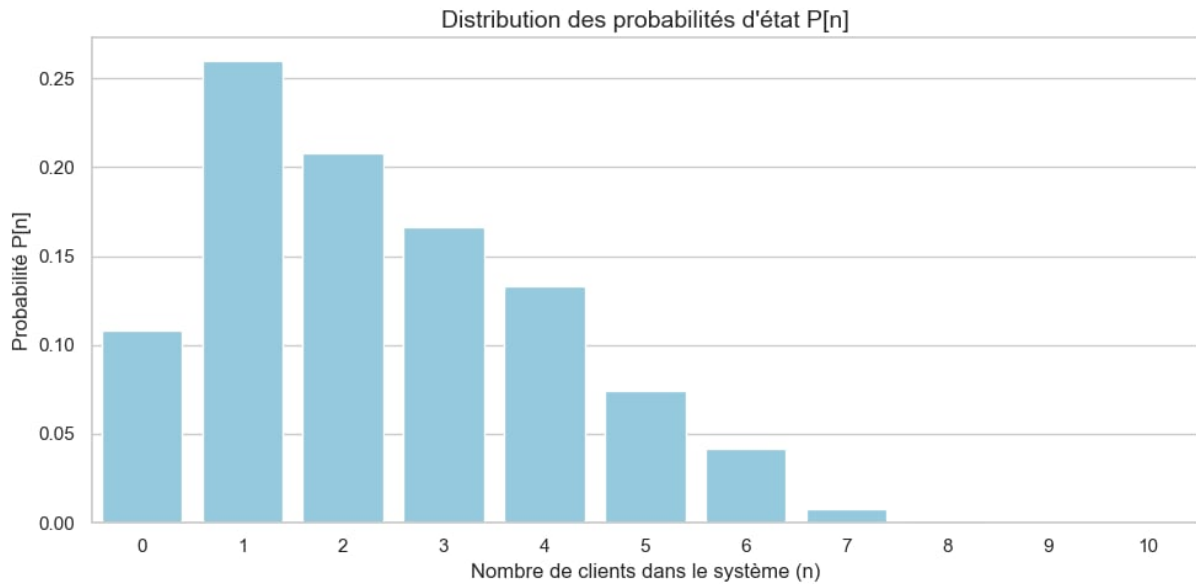


FIGURE 3.3 – Probabilités d'occupation des états du système $M/M/2/N$ avec dérobade et abandon .

La Figure (3.4), présente les principaux indicateurs de performance dérivés de la distribution stationnaire :

- La moyenne du nombre de clients dans le système L_s est estimée à environ 2,40 clients, ce qui indique un niveau de charge modéré.
- La moyenne du nombre de clients en file d'attente L_q est de 0,87, ce qui suggère que la majorité des clients sont en cours de service.
- Le taux d'arrivée effectif $\lambda_{\text{eff}} \approx 3,69$ est inférieur au taux d'arrivée brut $\lambda = 4$, en raison de la dérobade induite par le seuil K_1 . Cela signifie qu'une fraction non négligeable des clients renonce à entrer dans le système.
- Le temps moyen passé dans le système W_s est de 0,649 unités de temps, tandis que le temps d'attente moyen en file W_q est de 0,236, soit environ 36 % du temps total.

Cela indique que le service est relativement fluide, avec un temps d'attente raisonnable.

- Enfin, les taux de dérobade et d'abandon sont respectivement de 0,310 et 0,133.

Le taux de dérobade est plus élevé, ce qui souligne l'importance de ce phénomène dans la dynamique du système lorsque le seuil K_1 est atteint.

Ces résultats confirment que le système fonctionne de manière globalement efficace, en maintenant un équilibre satisfaisant entre la charge, la capacité de service et les comportements stratégiques des clients. Le modèle proposé permet ainsi de mesurer l'influence des seuils comportementaux sur la performance globale du système.

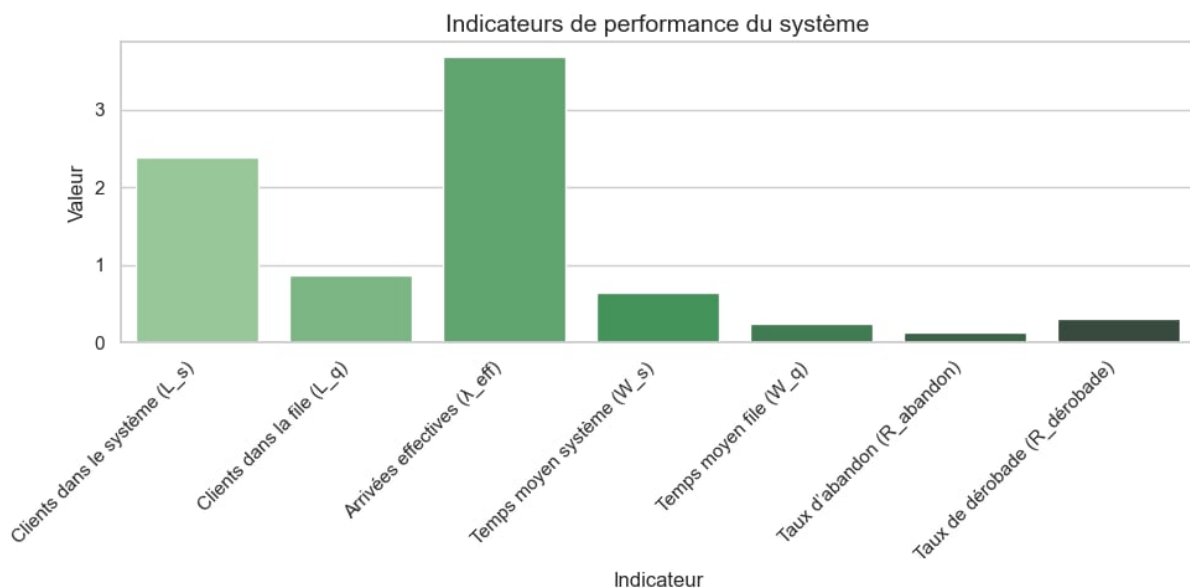


FIGURE 3.4 – Indicateurs de performance du système en régime permanent .

La figure (3.5), représente l'impact des seuils K_1 (seuil de dérobade) et K_2 (seuil d'abandon) sur les taux de dérobade et d'abandon dans un système de files d'attente à capacité finie $M/M/2/N$, avec deux serveurs hétérogènes.

Les paramètres utilisés pour cette simulation sont : un taux d'arrivée $\lambda = 4,0$, des taux de service $\mu_1 = 3,0$ et $\mu_2 = 2,0$, une probabilité de dérobade $\epsilon = 0,3$, et un taux d'abandon $\gamma = 0,5$. La capacité maximale du système est fixée à $N = 10$.

Deux courbes sont tracées :

- En rouge, l'évolution du taux d'abandon R_{abandon} en fonction du seuil K_2 : on observe que ce taux est nul tant que le seuil K_2 est supérieur à la population maximale atteinte dans le système. À partir d'un certain seuil, le taux d'abandon commence à croître, traduisant une augmentation du nombre de clients atteignant un niveau d'impatience suffisant pour quitter la file.
- En bleu, l'évolution du taux de dérobade $R_{\text{dérobade}}$ en fonction du seuil K_1 : de manière similaire, ce taux reste nul pour des valeurs faibles de K_1 , puis augmente lorsque davantage de clients décident de ne pas entrer dans le système en raison d'une congestion perçue.

Cette analyse met en évidence l'intérêt d'une modulation stratégique des seuils comportementaux pour optimiser les performances du système tout en intégrant la tolérance des clients face à l'attente.

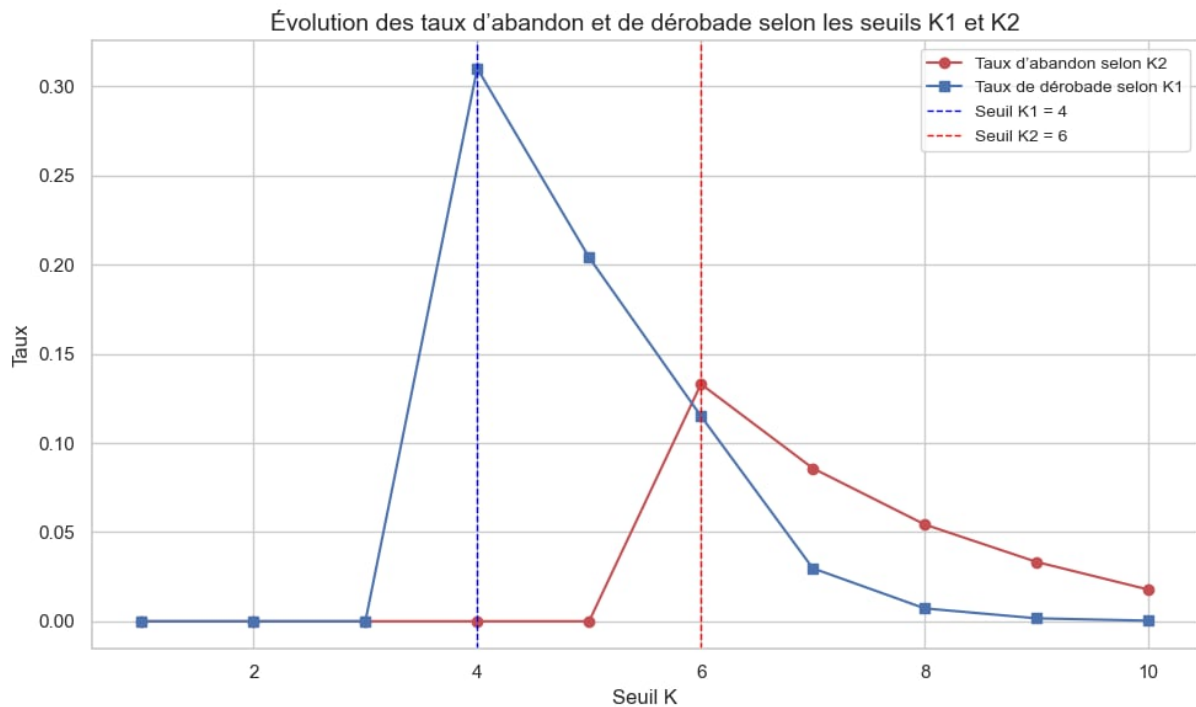


FIGURE 3.5 – Évolution des taux d'abandon et de dérobade en fonction des seuils K_1 et K_2 .

CHAPITRE

4

Modélisation par la théorie des jeux

4.1 Introduction

Dans ce qui suit, nous présenterons les notions de base de la théorie des jeux, notamment les concepts clés qui nous permettront de modéliser par la théorie des jeux, une file d'attente $M/M/2/N$, où les clients adoptent des comportements stratégiques en choisissant d'entrer ou non dans le système en fonction de la congestion observée, avec possibilité de renoncement (dérobade) ou d'abandon une fois dans la file. Pour cela, on introduit une fonction d'utilité et on caractérise les conditions d'équilibre de Nash dans un tel système.

Dans ce chapitre, nous appliquons la théorie des jeux pour analyser les comportements stratégiques des clients dans un système de file d'attente avec dérobade et abandon. Les clients, considérés comme des joueurs rationnels, doivent choisir entre rejoindre la file d'attente ou la quitter, en évaluant les coûts et les bénéfices attendus de chaque option. Elle permet de modéliser les interactions entre les clients et d'identifier les équilibres stratégiques.

4.2 Définitions et généralités

4.2.1 QU'EST CE QU'UN JEU ?

Un jeu est une situation où des individus (les joueurs), doivent effectuer des choix parmi un ensemble d'actions possibles, dans un cadre défini à l'avance par

(les règles de jeu) précises, L'issue du jeu, résultant de l'ensemble des décisions prises, détermine un gain (ou une perte) pour chaque participant. Bien que le hasard puisse parfois intervenir, les résultats dépendent avant tout des interactions stratégiques entre les joueurs.

Joueurs

Un joueur est l'entité de base de la théorie des jeux pouvant être interprété comme des individus ou un groupe d'individus prenant une décision. On notera N l'ensemble des joueurs participant à un même jeu : $N = \{1, \dots, N\}$ Où N désigne leur nombre ($N \geq 2$).

Stratégies

Une stratégie est un plan d'actions complet défini pour chaque joueur, précisant les décisions qu'il prendra à chaque étape du jeu et dans chaque situation possible. Elle décrit de manière exhaustive le comportement du joueur tout au long du déroulement du jeu. Il existe différents types de stratégies :

Stratégies pures :

Une stratégie pure du joueur i est une action, ou un ensemble d'actions, choisies avec certitude.

Stratégies mixtes :

Une stratégie mixte du joueur i est une distribution de probabilité α^i sur l'ensemble des stratégies pures de i^{eme} joueur. Une stratégie mixte pour le joueur i est représentée par un vecteur α défini comme suit :

$$\Delta_i = \left\{ \alpha_i = (\alpha_i^1, \alpha_i^2, \dots, \alpha_i^{n_i}) \in \mathbb{R}^{n_i} : 0 \leq \alpha_j^i \leq 1 \forall j \in \{1, \dots, n_i\}, \sum_{j=1}^{n_i} \alpha_j^i = 1 \right\}$$

où n_i est le nombre de stratégies pures du joueur i , et α_j^i représente la probabilité que le i^{eme} joueur choisisse la j^{eme} stratégie.

Fonction d'utilité

L'utilité est la mesure de chaque issue du jeu aux yeux de chaque joueur, elle est représentée par une fonction de gain u_i , telle que :

$$u_i : \mathcal{S} \rightarrow \mathbb{R}, \\ s = (s_1, s_2, \dots, s_N) \mapsto u_i(s)$$

Ainsi, il est nécessaire de connaître les utilités que chaque joueur obtiendra à l'issue du jeu pour chaque combinaison de stratégies. Une situation de jeu se caractérise par le fait que l'utilité de chaque joueur dépend non seulement de sa propre stratégie, mais également de celles adoptées par les autres joueurs.

On considère qu'un joueur adopte un comportement rationnel s'il est conscient des différentes alternatives et choisit délibérément la stratégie qui lui est la plus avantageuse parmi l'ensemble des options disponibles.

4.2.2 REPRÉSENTATION DES JEUX

Les jeux peuvent être représentés sous deux formes[5].

Jeux sous forme stratégique

Un jeu sous forme normale (ou bien stratégique) s'écrit de la manière suivante :

$$\langle I, \{X_i\}_{i \in I}, \{f_i\}_{i \in I} \rangle$$

- $I = \{1, 2, \dots, n\}$ est l'ensemble des joueurs.
- X_i : l'ensemble des stratégies du joueur i . L'issue x du jeu correspond au vecteur $x = (x_1, x_2, \dots, x_n)$, où $x_i \in X_i$ est la stratégie choisie par le joueur i .
- Les gains (ou utilités) des joueurs sur l'ensemble $X = \prod_{i=1}^n X_i$ des issues $x \in X$ sont représentés par une fonction

$$f_i : X_1 \times X_2 \times \dots \times X_n \rightarrow \mathbb{R}.$$

Jeux sous forme extensive

Un jeu sous forme extensive est défini par les éléments suivants :

- Un ensemble de joueurs $I = \{1, 2, \dots, N\}$.
- Un arbre fini composé de noeuds représentant les états du jeu et de branches illustrant les actions possibles à chaque étape.
- Une fonction de nommage qui indique à chaque noeud quel est le joueur qui doit jouer.
- Une fonction de gain qui associe à chaque noeud terminal un vecteur de réels représentant les gains (ou utilités) obtenus par chacun des joueurs.
- Une partition des noeuds en un ensemble d'information représentant les croyances des joueurs [16].

4.2.3 CLASSIFICATION DES JEUX

Il existe plusieurs classifications possibles des jeux. Nous allons définir quelques catégories :

I. Selon l'ordre

I.1 Jeux statiques

Un jeu est qualifié de statique lorsque tous les joueurs prennent leurs décisions simultanément, et reçoivent ensuite leurs gains respectifs. Ce type de jeu est représenté de manière appropriée sous la forme normale [5].

I.2 Jeux dynamiques

On dit qu'un jeu est dynamique lorsque les joueurs choisissent leurs actions alternativement. La représentation la plus appropriée pour ce type de jeu est la forme extensive, qui décrit l'ordre des actions et les informations disponibles à chaque étape.

II. Selon les relations entre les joueurs

II.1 Jeux coopératifs un jeu est dit coopératif lorsque les joueurs ont la possibilité de conclure des accords contraignants entre eux, leur permettant de coordonner leurs stratégies de manière collective dans le but d'optimiser les gains de l'ensemble des membres de la coalition [20].

II.2 Jeux non coopératifs Dans un jeu non coopératif, chaque joueur cherche à maximiser sa propre fonction d'utilité tout en prenant en considération les stratégies des autres joueurs, il n'est donc pas possible de former des coalitions.

III. Selon les gains des joueurs Les jeux non coopératifs se divisent en deux grandes catégories : les **jeux à somme nulle** et les **jeux à somme non nulle**.

Un jeu à somme nulle est un jeu dans lequel, pour chaque issue possible, la somme des gains de tous les joueurs est strictement égale à zéro. Ce qui est gagné par un joueur est perdu par les autres. La somme des fonctions d'utilité est donc nulle, c.à.d :

$$\sum_{i \in I} u_i(s) = 0, \quad \forall s \in S$$

En revanche, si la somme des gains n'est pas toujours nulle. S'il existe, au moins, une issue $s' \in S$, telle que :

$$\sum_{i \in I} u_i(\bar{s}) \neq 0$$

alors, le jeu sera dit à somme non nulle.

IV. Selon l'information

IV.1 Jeux à information parfaite / imparfaite

Un jeu est dit à information parfaite si, au moment de prendre une décision, les joueurs sont au courant de tous les choix passés de leurs rivaux. lorsqu'un joueur atteint un point de décision où il ne connaît pas toutes les actions des autres joueurs qui l'ont précédés alors on est dans un jeu à information imparfaite.

IV.2 Jeu à information complète/ incomplète

Un jeu est dit à information complète si tous les joueurs connaissent parfaitement la structure du jeu, l'ensemble des joueurs, l'ensemble de ses

stratégies ainsi que l'ensemble des stratégies des autres joueurs et les motivations ou les fonctions objectifs de tous les autres joueurs . Le jeu est dit à information incomplète si, au moins, un des joueurs ne connaît pas entièrement la structure du jeu.

4.2.4 JEUX FINIS ET JEUX INFINIS

Un jeu est dit fini lorsque chaque joueur dispose d'un nombre fini de stratégies, c'est-à-dire que pour tout $i \in I$ l'ensemble S_i est fini. Ainsi, le jeu est infini s'il existe au moins un joueur $i \in I$ dont l'ensemble des stratégies $i \in I$ est infini.

4.2.5 LES JEUX SYMÉTRIQUES

Un jeu est dit symétrique, si tous les joueurs possèdent les mêmes stratégies, et la même fonction d'utilité.

$$S_1 = S_2 = S$$

et

$$u_1(s_1, s_2) = u_2(s_2, s_1), \quad \forall (s_1, s_2) \in S$$

4.2.6 EQUILIBRE DE NASH

En 1950, le mathématicien John Nash a introduit la notion d'équilibre. Cet équilibre décrit une situation où aucun des joueurs n'a d'intrêt à changer unilatéralement de stratégie [20].

a- Equilibre de Nash en stratégies pures

Une situation $s^* = \{s_1^*, s_2^*, \dots, s_N^*\} \in S$ est un équilibre de Nash en stratégies pures du jeu , si pour chaque joueur $i \in \mathbb{N}$, on a :

$$U_i(s_i^*; s_{-i}^*) \geq U_i(s_i; s_{-i}^*), \quad \forall s_i \in S_i, \forall i \in I$$

où $s_{-i} = (s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_N)$.

En d'autres termes, un équilibre de Nash est une issue du jeu dont aucun joueur ne gagne à changer seul de stratégie.[5]

b- Equilibre de Nash en stratégies mixtes

Un équilibre de Nash en stratégies mixtes pour le jeu est un ensemble de stratégies mixtes α tel que :

$$U_i(\alpha_i^*; \alpha_{-i}^*) \geq U_i(\alpha_i; \alpha_{-i}^*), \quad \forall \alpha_i \in \Delta^{m_i}, \forall i \in I$$

Où :

$$\Delta^{m_i} = \left\{ \alpha = (\alpha_1, \dots, \alpha_{m_i}) \in \mathbb{R}^{m_i}, \alpha_j \geq 0 \forall j = 1, \dots, m_i, \sum_{j=1}^{m_i} \alpha_j = 1 \right\}$$

Ainsi, il a été démontré que tout jeu fini sous forme stratégique admet un équilibre de Nash en stratégies mixtes [5].

4.3 Modélisation du système $M/M/2/N$

Le système de file d'attente $M/M/2/N$ peut être interprété comme un jeu non coopératif entre clients arrivant individuellement dans un environnement stochastique. Plus précisément, il s'agit d'un jeu statique à décisions séquentielles avec un joueur représentatif myope, c'est-à-dire qu'un client prend sa décision d'entrer ou non dans le système uniquement en fonction de l'état courant n (nombre de clients présents) sans anticiper les choix futurs ni passés des autres. Chaque joueur (client) arrive selon un processus de Poisson et observe l'état du système au moment de son arrivée. Étant donné que les arrivées sont indépendantes, on peut se ramener à un jeu à un seul joueur représentatif, qui décide rationnellement s'il souhaite entrer dans le système en fonction de son utilité attendue.

Ce jeu peut être formellement défini par le triplet :

$$G = \langle N, X(n), U(n) \rangle, \quad (4.1)$$

où :

- N : Ensemble des joueurs (population anonyme de clients potentiels),
- $X(n)$: Ensemble des stratégies disponibles pour le joueur lorsqu'il observe n clients dans le système,
- $U(n)$: Utilité attendue du joueur à l'état n .

Ce jeu appartient à la famille des jeux de file d'attente stratégiques, où chaque joueur ne subit l'influence des autres qu'à travers la congestion induite dans le système.

4.3.1 JOUEURS

Dans la modélisation de la file d'attente $M/M/2/N$ par la théorie des jeux, les joueurs sont les clients supposés identiques, qui arrivent dans le système en prenant une décision stratégique basée sur l'état observé de la file (nous supposons dans le cadre de cette étude que l'état de la file est observable).

Remarque 8. *Dans le cas non observable, les décisions des clients sont basées sur des croyances .*

4.3.2 TAUX D'ARRIVÉE ET DÉROBADE

Le taux effectif d'arrivée des clients dans le système dépend du nombre n de clients présents et de la probabilité de dérobade :

$$\lambda(n) = \begin{cases} \lambda, & \text{si } n \leq K_1 \\ \lambda \cdot (1 - \varepsilon), & \text{si } n > K_1 \end{cases}$$

où ε est la probabilité qu'un client refuse d'entrer lorsque n clients sont déjà présents. Autrement dit, en dessous du seuil K_1 , tous les clients entrent, mais au-delà, certains clients se dérobent, ce qui réduit le flux effectif.

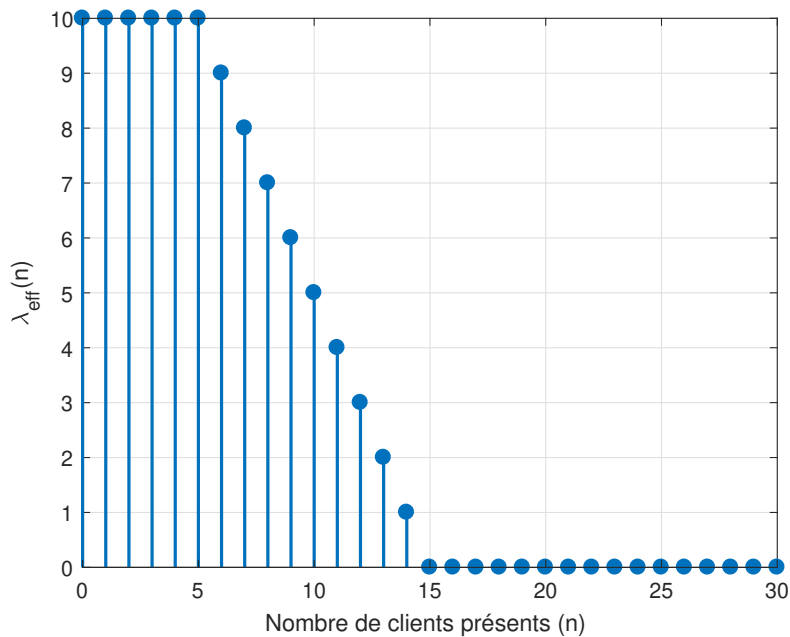


FIGURE 4.1 – Exemple d'un Taux effectif d'arrivée en fonction de l'état n

4.3.3 STRATÉGIES

Dans ce modèle, chaque client arrivant observe l'état courant du système (nombre n total de clients, dont au plus 2 en service) et décide d'entrer ou de dérober, et en espérant être servi avant d'abandonner. Il prend ainsi ses décisions à deux niveaux :

Niveau 1 : décision d'entrer

Le client observe le nombre n de clients présents. Il décide d'entrer avec une probabilité $1 - \varepsilon(n)$ ou de se dérober avec une probabilité $\varepsilon(n)$. Ces probabilités

peuvent être modélisées de façon déterministe par un seuil K_1 :

$$\varepsilon(n) = \begin{cases} 0, & n \leq K_1; \\ \varepsilon, & n > K_1. \end{cases}$$

Ainsi, un client entre toujours lorsque la congestion est faible (inférieure ou égale à K_1), mais au-delà de ce seuil, il peut décider de ne pas entrer avec une probabilité strictement positive ε .

Niveau 2 : décision d'abandon après entrée

Une fois dans la file, le client peut soit patienter jusqu'à être servi, soit abandonner le système avant d'être pris en charge. Cette décision dépend du niveau de congestion n observé une fois entré, et elle est modélisée à l'aide d'un seuil K_2 :

$$\gamma(n) = \begin{cases} 0, & n \leq K_2; \\ \gamma \cdot (n - 1), & n > K_2. \end{cases}$$

où γ représente le taux d'impatience d'un client. Autrement dit, tant que le nombre total de clients dans le système reste inférieur ou égal à K_2 , les clients sont prêts à attendre. Au-delà de ce seuil, le taux global d'abandon devient croissant, proportionnel au nombre de clients présents, traduisant ainsi une augmentation de l'impatience collective.

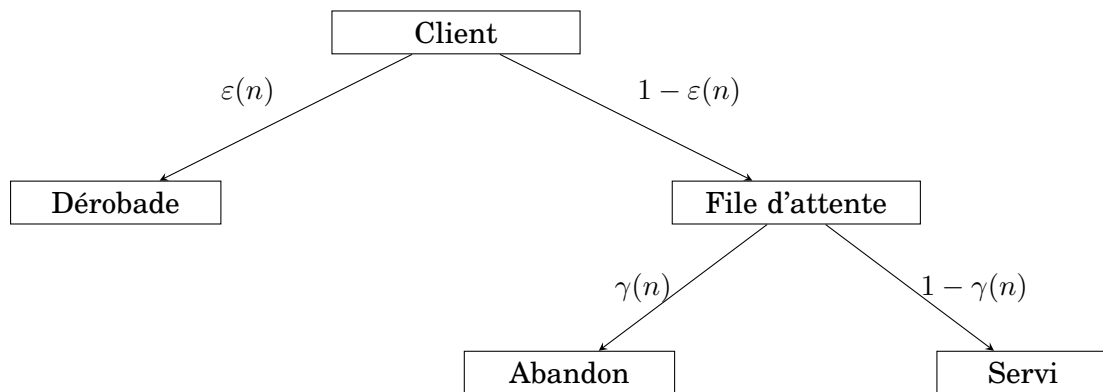


FIGURE 4.2 – Arbre de décision du comportement des clients dans le système

Remarque 9. Le seuil K_2 peut être interprété comme une tolérance maximale de la file d'attente, au-delà, les clients estiment que l'attente probable est trop longue par rapport au gain espéré du service. Plus n est grand, plus l'impatience globale croît, rendant l'abandon plus probable.

4.3.4 FONCTION D'UTILITÉ DU CLIENT

Un client qui arrive dans le système alors que n clients y sont déjà présents, évalue l'utilité qu'il pourrait tirer d'une entrée en tenant compte du gain espéré s'il est servi et du coût potentiel lié à l'attente, ou alors de refuser d'entrer auquel cas il ne subit ni attente ni perte. Cette évaluation conditionne sa décision stratégique d'entrer ou non dans la file. Son utilité espérée à l'entrée est alors :

$$U(n) = (1 - \varepsilon(n)) \cdot [R \cdot P(n) - C \cdot W(n)] \quad (4.2)$$

où :

- $\varepsilon(n)$ est la probabilité de dérobade,
- $P(n)$ est la probabilité d'être servi,
- $W(n)$ est le temps moyen d'attente,
- R et C sont le gain espéré et le coût unitaire.

Si un client choisit d'entrer, il obtient un gain R avec probabilité $P(n)$ d'être servi, mais subit un coût d'attente C proportionnel au temps moyen passé dans le système, même s'il abandonne par la suite. Si le client choisit de se dérober, il ne subit aucun coût d'attente, ni ne reçoit de gain, sa fonction d'utilité est $U_e(n) = 0$.

Remarque 10. Bien que le taux d'abandon $\gamma(n)$ n'apparaisse pas explicitement dans l'expression de l'utilité $U(n)$, il influence indirectement la décision du client. En effet, $\gamma(n)$ affecte à la fois la probabilité d'être servi $P(n)$ et la durée moyenne passée dans le système $W(n)$. Une impatience plus élevée (grande valeur de γ) tend à réduire ces deux quantités, modifiant ainsi l'équilibre entre le gain espéré et le coût anticipé.

4.4 Équilibre de Nash

Un équilibre stratégique (au sens de l'équilibre de Nash) dans une file d'attente observable est une stratégie telle qu'aucun client n'a intérêt à modifier unilatéralement sa décision (entrer ou se dérober), étant donné le comportement des autres clients. Dans ce qui suit, nous allons caractériser l'équilibre stratégique résultant de ces comportements rationnels dans le cadre du modèle $M/M/2/N$.

4.4.1 STRATÉGIES OPTIMALES

La stratégie d'un client rationnel consiste à entrer dans le système uniquement si son utilité espérée est positive. Cela signifie que, pour chaque état n du système, le client compare son utilité s'il se dérobe et son utilité s'il entre dans le système. Ainsi, la décision d'entrer est rationnelle si et seulement si :

$$U(n) > 0 \iff R \cdot P(n) > C \cdot W(n) \quad (4.3)$$

4.4.2 SEUILS D'ENTRÉE ET D'ABANDON À L'ÉQUILIBRE

Dans ce contexte, l'équilibre de Nash correspond à l'existence de seuils optimaux K_1^* tel que les clients acceptent d'entrer si $n \leq K_1^*$ et K_2^* tel que les clients abandonnent la file si $n > K_2^*$. :

$$\varepsilon(n) = \begin{cases} 0, & \text{si } n \leq K_1^* \quad (\text{entrer}) \\ 1, & \text{si } n > K_1^* \quad (\text{dérober}) \end{cases} \quad (4.4)$$

Le seuil K_1^* est le plus grand n tel que l'utilité attendue d'entrée est positive. Il est déterminé par la condition d'indifférence :

$$U(K_1^*) = 0 \iff R \cdot P(K_1^*) = C \cdot W(K_1^*) \quad (4.5)$$

Le seuil K_2^* est défini comme le plus grand entier tel que les clients acceptent encore de patienter une fois entrés dans la file :

$$\gamma(n) = \begin{cases} 0, & \text{si } n \leq K_2^* \quad (\text{pas d'abandon}) \\ > 0, & \text{si } n > K_2^* \quad (\text{abandon possible}) \end{cases} \quad (4.6)$$

Ainsi, chaque client suit une stratégie rationnelle de seuils, ce qui structure l'équilibre global.

Remarque 11. *Le couple (K_1^*, K_2^*) constitue un équilibre lorsque, pour tout n , les décisions d'entrée et de patience maximisent l'utilité espérée des clients, compte tenu du comportement stratégique collectif. Ces seuils peuvent être déterminés analytiquement ou numériquement en résolvant l'équation $U(n) = 0$ pour K_1^* et l'évolution de $\gamma(n)$ pour K_2^* .*

4.5 Résultats numériques

Analyse le comportement stratégique des clients dans un système de files d'attente M/M/2/N, nous présentons une série d'illustration numérique mettant en évidence les interactions entre l'utilité individuelle, le temps d'attente, et les seuils de décision.

La figure (4.3) illustre l'évolution du seuil stratégique de dérobade K_1^* en fonction du nombre de clients présents dans le système n , pour un seuil d'abandon fixé à K_2 .

On observe que :

- Pour les faibles valeurs de n , le seuil K_1^* croît de manière presque linéaire : les clients acceptent d'entrer plus facilement lorsque le système est peu chargé.
- À partir d'un certain seuil, K_1^* se stabilise en dessous de K_1 , traduisant une anticipation stratégique du comportement d'abandon : les clients refusent d'entrer si la file est trop longue, même avant le seuil K_1 , pour éviter une utilité négative.
- La ligne horizontale rouge correspond à K_1 , le seuil au-delà duquel les clients dans la file abandonnent s'ils ne sont pas servis à temps.

Ce comportement découle du compromis entre le gain espéré du service et le coût anticipé du temps d'attente, pondéré par le risque d'abandon.

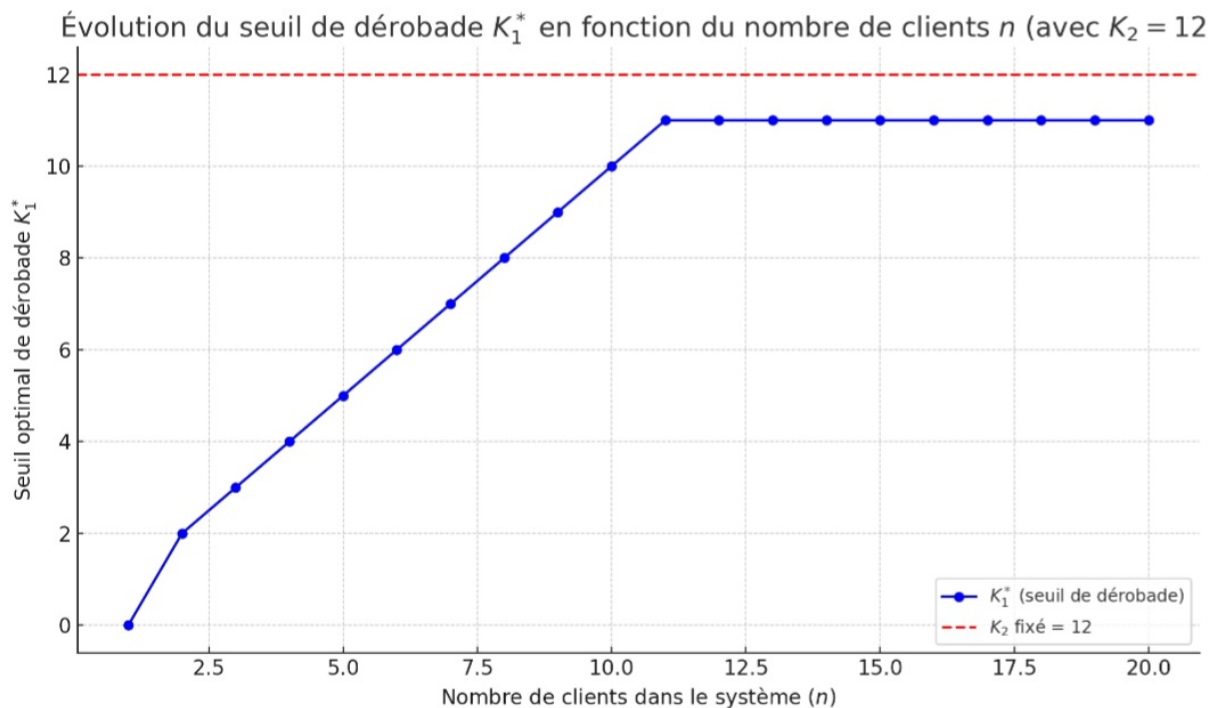


FIGURE 4.3 – Évolution du seuil stratégique de dérobade K_1^* en fonction du nombre de clients .

Dans la figure (4.4), nous analysons le comportement stratégique d'un client rationnel confronté à un système de files d'attente de type $M/M/2/N$ avec serveurs hétérogènes, abandon, et dérobade. Plus précisément, nous cherchons à déterminer un seuil optimal de dérobade K_1^* , au-delà duquel un client préfère ne pas entrer dans le système, car l'utilité espérée devient négative.

Nous considérons que les clients arrivent selon un processus de Poisson de taux $\lambda = 4$. Le système comporte deux serveurs hétérogènes avec des taux de service respectifs $\mu_1 = 3$ et $\mu_2 = 2$, ce qui donne un taux global de service $\mu_3 = 5$. Les clients peuvent décider de ne pas entrer dans le système s'ils estiment que l'attente est trop longue (comportement de dérobade), avec une probabilité d'entrée égale à $1 - \varepsilon = 0.7$. De plus, une fois dans le système, un client peut abandonner la file s'il n'est pas servi assez rapidement, selon un taux d'abandon $\gamma = 0.9$, applicable à partir d'un seuil d'impatience fixé à $K_2 = 30$.

Le graphe montre que l'utilité diminue à mesure que n augmente, ce qui traduit un accroissement du coût d'attente et une baisse de la probabilité de service. Le seuil optimal K_1^* est identifié comme la plus petite valeur de n pour laquelle $U(n) < 0$. Ce seuil correspond à une stratégie rationnelle de dérobade : tout client observant $n \geq K_1^*$ choisira de ne pas entrer dans le système, car l'utilité attendue devient négative.

Cette représentation met en lumière la dimension décisionnelle du comportement des clients, modélisée ici dans un cadre de théorie des jeux où chaque individu agit en maximisant son propre gain espéré face à un environnement incertain.

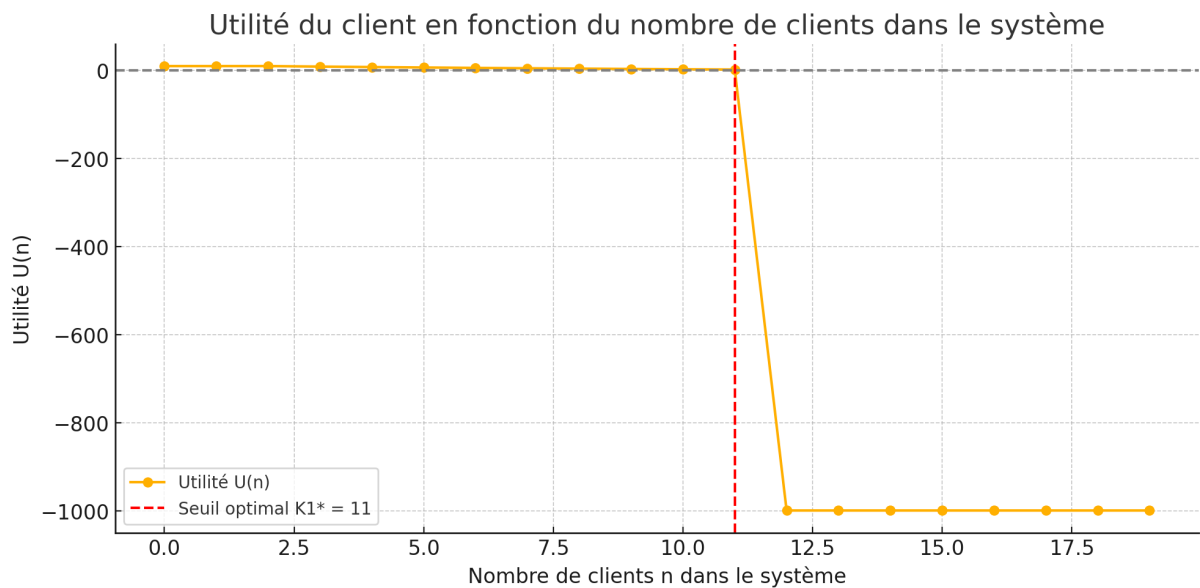


FIGURE 4.4 – Évolution de l'utilité en fonction du nombre de clients dans le système.

La figure (4.5) illustre le comportement stratégique d'un client face à l'état du système de file d'attente M/M/2/N, en considérant à la fois son utilité attendue $U(n)$ et le temps d'attente moyen $W(n)$ en fonction du nombre de clients présents n . Deux seuils clés sont mis en évidence :

Le seuil de dérobade optimal K_1^* (trait bleu pointillé), au-delà duquel le client préfère ne pas entrer dans le système ;

Le seuil d'abandon K_2^* (trait noir pointillé), à partir duquel les clients présents dans la file choisissent de quitter s'ils n'ont pas encore été servis.

L'axe vertical gauche représente l'utilité $U(n)$, définie comme la différence entre le gain potentiel R du service et le coût d'attente estimé, fonction croissante du nombre de clients et descendante. On observe que :

- Tant que $n \leq K_1^*$, l'utilité reste positive : il est rationnel d'entrer dans le système.
- Pour $n > K_1^*$, l'utilité chute brutalement, traduisant un désavantage net à l'entrée, principalement dû au temps d'attente excessif.

L'axe vertical droit montre l'évolution du temps moyen d'attente $W(n)$. Celui-ci reste raisonnable jusqu'à atteindre K_1^* , mais augmente fortement à l'approche du seuil K_2^* , au-delà duquel le comportement d'abandon est activé.

Ainsi, cette figure met en évidence l'équilibre stratégique des décisions client dans un système avec impatience, en montrant que la décision de rejoindre ou non la file dépend directement de l'évaluation individuelle du bénéfice net à chaque niveau d'encombrement du système.

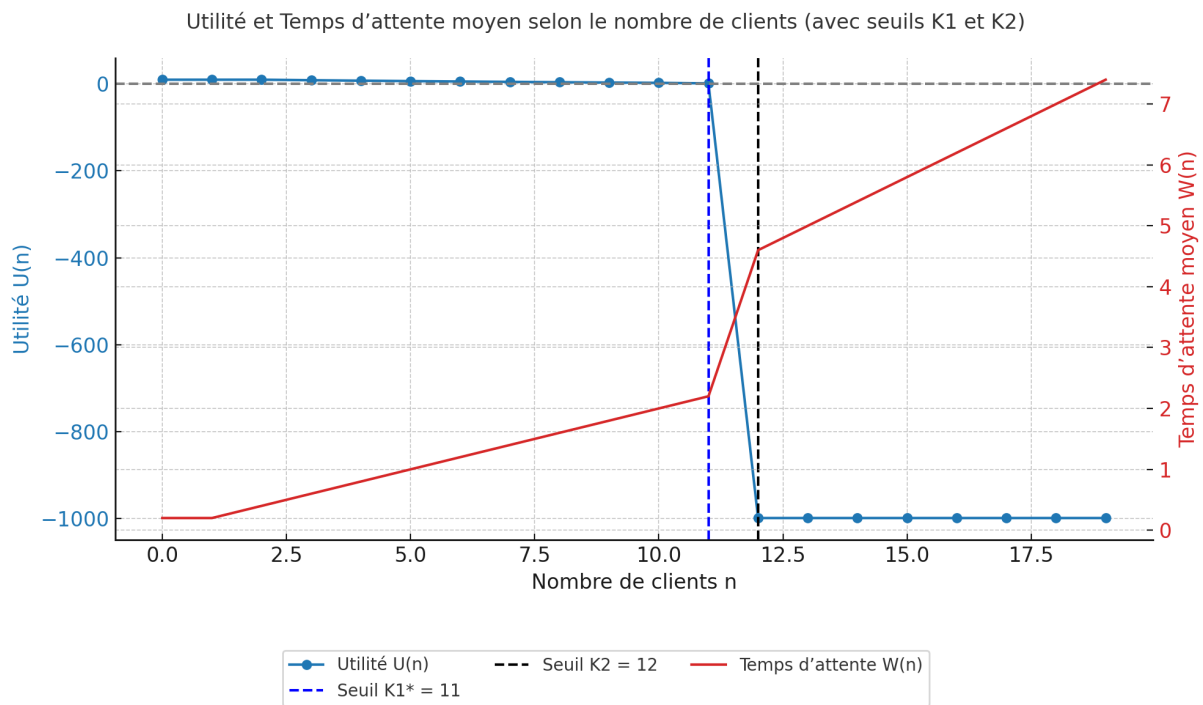


FIGURE 4.5 – Évolution de l'utilité et du temps d'attente moyen selon le nombre de clients dans le système .

4.6 Conclusion

Dans ce chapitre, nous avons utilisé la théorie des jeux pour analyser le comportement stratégique des clients dans un système $M/M/2/N$ avec dérobade et abandon. L'équilibre de Nash obtenu reflète les choix optimaux des clients en fonction de leur impatience et de la longueur de la file. Cette approche offre une meilleure compréhension des interactions individuelles et de leur impact sur la performance globale du système.

Conclusion Générale

Dans ce mémoire, nous analysons le modèle : $M/M/2/N$ intégrant des serveurs hétérogènes, des seuils de dérobade et d'abandon. Des équations d'équilibre ont été établies afin de déterminer la distribution stationnaire du système. Nous avons également dérivé des expressions analytiques pour diverses mesures de performance telles que les probabilités d'état, le taux d'occupation, les taux d'abandon et de dérobade, ainsi que le nombre moyen de clients dans le système.

L'originalité de notre démarche réside dans l'intégration de la théorie des jeux pour modéliser les comportements stratégiques des clients. À travers cette approche, nous avons mis en évidence l'existence d'un équilibre de Nash entre les stratégies individuelles des clients, en confrontant leur intérêt personnel à la performance globale du système.

Les résultats obtenus confirment l'utilité d'une approche conjointe entre la théorie des files d'attente et celle des jeux pour mieux modéliser et gérer les systèmes de service confrontés à l'impatience des clients. Ils ouvrent également la voie à de futures recherches sur l'optimisation sociale, la tarification incitative ou encore l'étude de systèmes à information partielle.

Ce travail contribue à une meilleure compréhension des interactions entre l'impatience, la stratégie individuelle et la performance collective, dans un cadre réaliste et opérationnel.

Bibliographie

- [1] M. O. Abou El-Ata and A. M. A. Hariri. The m/m/c/n queue with balking and reneging. *Computers and Operations Research*. Vol. 19(13), pp. 713-716, 1992.
- [2] Eitan Altman and Nahum Shimkin. Équilibre individuel et apprentissage dans les systèmes de partage de processeurs. *Operations Research*, Vol. 46(6), pp. 776-784, 1998.
- [3] C. J. Ancker Jr. and A. V. Gafarian. Some queuing problems with balking and reneging : I. *Operations Research*, Vol. 11(1), pp. 88-100, 1963.
- [4] C. J. Ancker. Jr and A. V. Gafarian. Some queuing problems with balking and reneging : II. *Operations Research*, Vol 11(6) , pp. 928-937, 1963.
- [5] Emmanuel Bonzon. *Modélisation des interactions entre agents rationnels : les jeux booléens*. PhD thesis, Université Paul Sabatier de Toulouse 3, 2007.
- [6] Amina Angelika Bouchentouf and Aicha Messabihi. A heterogeneous two-server queueing system with reneging and no waiting line. *ProbStat Forum*, Vol.(11), pp 67-76, 2018.
- [7] Antonis Economou and Aikaterini Kanta. Joining the shortest queue with reneging and balking in markovian multi-server systems. *Applied Mathematics and Computation*, Vol. 204 (1), pp. 28-49, 2008.
- [8] B. Filipowicz and J. Kwiecien. Queueing systems and networks : Models and applications. *Bulletin of the Polish Academy of Sciences. Technical Sciences*, Vol. 56(4), pp. 100-200, 2008.
- [9] O. Garnett, A. Mandelbaum, and M. Reiman. Designing a call center with impatient customers. *Manufacturing & Service Operations Management*. Vol. 4(3), pp. 208-227, 2002.
- [10] E. Gelenbe and G. Pujolle. *Introduction to Queueing Networks*. John Wiley and Sons, Chichester, England, 1987.
- [11] D. Gross and C. M. Harris. *Fundamentals of Queueing Theory*. Wiley, New York, 1975.
- [12] Pengfei Guo and R. Hassin. Strategic behavior and social optimization in markovian vacation queues. *Operations Research*, Vol. 59 (4), pp. 986-997, 2011.
- [13] Claudie H. CHABRIAC. *Éléments de Théorie des files d'attente*. 2008.

- [14] M. Kadi, A. Angelika Bouchentouf, and Lahcene Yahiaoui. On a multiserver queueing system with customers' impatience until the end of service under single and multiple vacation policies. *Operations Research Letters*, Vol. 15(2), pp. 740-763, 2020.
- [15] K.B. and G.K. *Techniques de modélisation : Méthodes analytiques*. 2020.
- [16] S. Konieczny. Introduction à la théorie des jeux. Technical report, Université d'Artois.
- [17] R. Kumar and S.K. Sharma. An $m/m/1/n$ queueing model with retention of renege customers and balking. *American Journal of Operational Research*, Vol. 2(1), pp. 1-5, 2012.
- [18] Roberts J Lubacz. J. A new approach to the single-server repeated attempt system with balking. In *Proceedings of the Third International Seminar on Teletraffic Theory, Moscow*, PP. 290-293, 1984.
- [19] M.KADI. Estimation et analyse d'un système de files d'attente avec dérobades. Master's thesis, Université de Sidi Bel Abbès - Djillali Liabes, 2017.
- [20] Hervé Moulin. *Théorie des jeux pour l'économie et la politique*. Herman, Paris, 1981.
- [21] P. Naor. The regulation of queue size by levying tolls. *Econometrica*, Vol. 37(1), pp. 15-24, 1969.
- [22] G. F. Newell. *Applications of Queueing Theory*. Chapman and Hall, Second edition, London, 1982.
- [23] Bocquet. S. Queueing theory with reneging. Technical Report DSTO-TR-1772, Defence Science and Technology Organisation, 2005.
- [24] B. Saida Safaa. Etude de la stabilité des files d'attente. Master's thesis, Université de Ain Temouchent-Belhadj Bouchaib, 2024.
- [25] Hammache Saouaguia. Systèmes de file d'attente avec découragement. Master's thesis, Université de Saida - Dr Moulay Tahar, 2018.
- [26] Kangzhou Wang, Na Li, and Zhibin Jiang. Queueing system with impatient customers : A review. In *IEEE International Conference on Service Operations and Logistics and Informatics (SOLI)*, pp 82-87, 2010.
- [27] K.H. Wang and Y.C. Chang. Cost analysis of a finite $m/m/r$ queueing system with balking, reneging and server breakdowns. *Mathematical Methods of Operations Research*. Vol. 56(2), pp. 169-180, 2002.
- [28] A. Willig. A short introduction to queueing theory. *Technical University Berlin, Telecommunication Networks Group*, Vol. 21, 1999.

Résumé

Ce mémoire porte sur l'étude d'un système de files d'attente du type M/M/2/N intégrant des serveurs hétérogènes, des comportements impatientes des clients dérobadé (refus d'entrée) ou d'abandon (quitter la file), ainsi qu'un seuil de tolérance.

L'objectif est de modéliser et analyser ce système afin d'en extraire les performances et d'en déduire des stratégies optimales du point de vue individuel et collectif. Une approche mathématique est développée à partir des chaînes de Markov pour établir les équations d'équilibre, ensuite résolues numériquement. Afin de mieux comprendre les comportements stratégiques des clients, la théorie des jeux est introduite. Le modèle permet de déterminer les équilibres de Nash entre les clients et d'évaluer les effets des seuils de dérobadé et d'abandon sur les indicateurs de performance du système. Ce travail offre un cadre analytique pertinent pour la conception et l'optimisation des systèmes de service à capacité limitée, principalement dans les domaines des télécommunications, des services hospitaliers ou des centres d'appels.

Abstract

This thesis focuses on the study of a M/M/2/N queueing system incorporating heterogeneous servers, impatient customer behavior (balking and renegeing), and a threshold policy. The goal is to model and analyze this system in order to derive its performance measures and determine optimal strategies from both individual and social perspectives. A mathematical framework based on Markov chains is developed to formulate and solve the equilibrium equations numerically. To better capture the strategic behavior of customers, game theory is employed. The model allows for the characterization of Nash equilibria among customers and evaluates the impact of balking and renegeing thresholds on system performance metrics. This work provides a relevant analytical framework for designing and optimizing limited-capacity service systems, particularly in fields such as telecommunications, healthcare services, and call centers.