



HAL
open science

Systemes de files d'attente strategiques avec information contrôlée

Tesnim Naceur

► **To cite this version:**

Tesnim Naceur. Systemes de files d'attente strategiques avec information contrôlée. Autre [cs.OH].
Université d'Avignon, 2020. Français. NNT : 2020AVIG0279 . tel-03211895

HAL Id: tel-03211895

<https://theses.hal.science/tel-03211895v1>

Submitted on 29 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Présentée à Avignon Université pour obtenir le diplôme de DOCTORAT

SPÉCIALITÉ : INFORMATIQUE

École Doctorale 536 « Agrosciences & Sciences »
Laboratoire Informatique d'Avignon (EA 4128)

Systèmes de files d'attente stratégiques avec information contrôlée

par
Tesnim Naceur

Soutenue publiquement le 29 octobre 2020 devant un jury composé de :

M ^{me}	Domique Quadri	Professeure, LRI, Université de Paris-Sud	Rapporteure
M.	Patrick Maillé	Professeur, IMT, Université de Rennes	Rapporteur
M ^{me}	Tania Jimenez	Ingénieure de recherche, LIA, Université d'Avignon	examinatrice
M.	Abderrahim Benslimane	Professeur, LIA, Université d'Avignon	examineur
M.	Oussama Habachi	Maître de Conférences, XLIM, Université de Limoges	examineur
M.	Yezekael Hayel	Maître de Conférences HDR, LIA, Université d'Avignon	Directeur de thèse

Sommaire

1	Introduction générale	3
1.1	Contexte	4
1.2	Organisation du manuscrit	6
1.3	Publications	7
2	Concepts de base et état de l'art	9
2.1	Introduction	10
2.2	Systèmes de files d'attente	10
2.2.1	Caractéristiques	10
2.2.2	Notion de classe de clients	12
2.2.3	Notation de Kendall	13
2.2.4	Systèmes de file d'attente Markovienne	13
2.3	Théorie de jeux sur les files d'attente	15
2.3.1	Notions de base	15
2.3.2	File d'attente observable	19
2.3.3	File d'attente non-observable	20
2.3.4	File d'attente avec une observation partielle	22
2.4	Conclusion	25
3	Tarification de l'information dans une file d'attente stratégique	27
3.1	Introduction	28
3.2	Description du modèle avec tarification de l'information	28
3.2.1	Modèle	28
3.2.2	Utilités espérées	30
3.3	Détermination de l'équilibre	32
3.4	Maximisation du revenu	35

3.4.1	Faire payer l'accès à la file d'attente	36
3.4.2	Faire payer l'information sur la taille de la file d'attente . . .	38
3.5	Politique de tarification optimale	39
3.6	Illustrations numériques	42
3.6.1	Charger l'accès à la file d'attente	42
3.6.2	Charger l'information	44
3.6.3	Politique de tarification optimale	44
3.7	Conclusion	45
4	Politique de partage d'information dans un système de file d'attente stratégique	47
4.1	Introduction	48
4.2	Description du modèle de file d'attente stratégique avec politique de partage de l'information	48
4.2.1	Modèle	48
4.2.2	Politique de partage d'informations basée sur un seuil fixe .	49
4.3	Bien-social en fonction de la PPI	50
4.3.1	Cas complètement observable	51
4.3.2	Cas partiellement observable	52
4.3.3	Comparaison entre la PPI partiellement et totalement observable	60
4.4	Système saturé	63
4.5	Illustrations numériques	67
4.5.1	Bien-être social espéré en fonction de la PPI	68
4.5.2	Bien-être social optimal	69
4.6	Conclusion	70
5	Politiques de partage d'information dans un système de file d'attente stratégique hétérogène	71
5.1	Introduction	72
5.2	Description du modèle de file d'attente hétérogène avec politique de partage d'information	72
5.2.1	Modèle	72
5.2.2	Politique de partage d'informations basée sur un seuil fixe .	73
5.3	Système de file d'attente avec deux classes de clients	74

5.3.1	Cas complètement observable	74
5.3.2	Cas partiellement observable	77
5.4	Illustrations numériques	84
5.4.1	Calcul de l'équilibre	84
5.4.2	Nombre moyen de clients	84
5.4.3	Temps moyen de séjour	86
5.5	Conclusion	87
6	Système de file d'attente avec déplacement et information partiellement observable	89
6.1	Introduction	90
6.2	Description du modèle de file d'attente avec déplacement	91
6.2.1	Modélisation du système et des décisions des clients	91
6.2.2	Modélisation des décisions dans la file d'attente de service	92
6.2.3	Modélisation des décisions dans la file d'attente de déplacement	93
6.3	Processus de naissance et de mort dans un système avec déplacement et partiellement observable	93
6.3.1	Transitions possibles	93
6.3.2	Calcul des probabilités stationnaires	97
6.4	Calcul des utilités attendues	99
6.4.1	Utilité des clients non-informés	99
6.4.2	Utilité des clients informés	99
6.5	Stratégie d'équilibre dans un système partiellement observable	102
6.5.1	Meilleure réponse pour les clients non-informés	102
6.5.2	Meilleure réponse pour les clients informés	103
6.5.3	Stratégie d'équilibre	103
6.5.4	Bien-être social	105
6.6	Illustrations numériques	105
6.6.1	Stratégies d'équilibre	105
6.6.2	Bien-être social	108
6.7	Conclusion	109

7 Conclusion générale et perspectives	111
7.1 Conclusion générale	112
7.2 Perspectives	113
Liste des figures	118
Liste des tableaux	119
Bibliographie	121
Ouvrages de référence	121

Tout vient à point à qui sait attendre.

Clément Marot

Chapitre 1

Introduction générale

Sommaire

1.1	Contexte	4
1.2	Organisation du manuscrit	6
1.3	Publications	7

1.1 Contexte

La théorie des files d'attente est l'étude analytique et scientifique la plus utilisée pour la modélisation de systèmes de files d'attente. Cette théorie a pour but d'étudier les systèmes où des entités veulent accéder à des ressources afin d'obtenir un service donné. En d'autres termes, une file d'attente représente toute situation dans laquelle des entités sont introduites dans un système où ils vont être traités et servis pendant une certaine période dans un certain ordre. De nombreux phénomènes réels, tels que les systèmes informatiques, les réseaux de communication, les serveurs Web, les systèmes de fabrication ainsi que les trafics routiers peuvent être représentés au moyen de modèles de files d'attente. L'analyse de tels systèmes doit prendre en considération les caractéristiques de la file telles que le flux d'arrivée, le serveur, la discipline de service, etc, pour caractériser et comprendre les décisions stratégiques des entités et leur impact sur les performances du système.

Ce domaine a été étudié depuis la révolution industrielle au Royaume-Uni par l'ingénieur électricien Erlang. Son travail a porté sur la gestion des réseaux téléphoniques de Copenhague (voir (ERLANG 1909; KINGMAN 2009; ASMUSSEN et al. 2009)). À cette époque, lors de la conception des systèmes téléphoniques, l'un des principaux problèmes concernant les critères de performance était de déterminer le nombre de lignes à fournir pour garantir une certaine qualité de service. L'utilisation de la théorie de file d'attente était très limitée au début, mais au fil des années, ce domaine de la recherche a beaucoup évolué et il a inspiré de nombreux chercheurs.

La période d'expansion économique après la guerre a offert aux chercheurs de nombreuses possibilités d'appliquer la théorie des files d'attente dans l'industrie, l'urbanisme, la construction des ponts, la construction des autoroutes et les loisirs. Le mathématicien, David George Kendall, a publié plusieurs articles sur le sujet qui ont considérablement façonné l'avenir de la théorie des files d'attente, y compris la notation abrégée $A/S/c$ pour décrire les modèles mathématiques des files d'attente (voir (KENDALL et al. 1951)). Kleinrock résume dans son livre (KLEINROCK 1975) plusieurs travaux de recherche sur les modèles d'attente (Voir aussi (KLEINROCK 1976)). Aux États-Unis, Thomas L. Saaty a publié son livre (SAATY 1961), qui est devenu un manuel standard pour enseigner la théorie des files d'attente jusqu'au début des années 1980.

Essentiellement, dans les files d'attente dite "stratégiques", les clients interagissent stratégiquement : ils décident de rejoindre la file ou de quitter le système

sans avoir le service, tout en anticipant les décisions des autres clients. Leurs décisions sont prises en fonction de leurs utilités espérées. La théorie des files d'attente stratégique se base sur deux modèles classiques : le modèle observable et le modèle non-observable. Dans les deux cas, le client arrive à le système avec serveur unique et il est face à deux possibilités : soit il rejoint la file d'attente, soit il quitte le système sans recevoir le service. Les prises de décisions de chaque client dépendent des paramètres du système et de son état (la longueur de la file) à un instant donné. La question à laquelle la théorie des files d'attente stratégique est confrontée dans la pratique est de savoir comment trouver un équilibre pour optimiser certaines fonctions objectifs telles que la maximisation du revenu du fournisseur, la maximisation du bien-être social des clients, etc. La contribution principale des travaux de recherche faits, était d'étudier et d'analyser les décisions stratégiques des clients dans les systèmes de file d'attente.

Dans les modèles de files d'attente réelles, le contrôle peut améliorer les performances du système en augmentant le débit d'arrivée des clients, en diminuant la surcharge, en réduisant la longueur des files d'attente, etc. Le contrôle est effectué par des actions adéquates qui peuvent être décrites dans des conditions de fonctionnement et être soumises à la détermination des stratégies de contrôle optimales. Un des contrôles effectués est le contrôle de l'accessibilité de l'information sur l'état du système pour un nouveau client qui se présente à une file d'attente. Parmi les nouveaux systèmes de files d'attente qui apparaissent en lien avec les nouvelles technologies de l'information, il y a les systèmes qui publient leur état en ligne et en temps réel. Les informations en ligne concernant le temps d'attente actuel et prévu pour le service sont de plus en plus accessibles aujourd'hui grâce aux nouvelles technologies et à l'émergence du big data. Les informations sont publiées par les fournisseurs de services sur les sites web ou leur application mobile pour faciliter l'accès à l'information et aider le client à prendre leur décision. Nous citons l'exemple du Google, qui offre avec une simple recherche aujourd'hui des informations sur les heures populaires et le temps d'attente prévu en temps réel de nombreux services (voir Figure 1.1) tels les restaurants, les aéroports, les grandes surfaces, les consulats, etc. De nombreuses recherches sont faits récemment afin de modéliser et d'étudier les systèmes de file d'attente avec une information contrôlée. Nous intéressons dans ce travail à ce domaine de recherche.

L'objectif de cette thèse est d'aider les usagers stratégiques de systèmes de files d'attente à prendre les meilleurs décisions. Nous proposons des nouveaux modèles de files d'attente et nous étudions l'impact de l'information sur le comportement

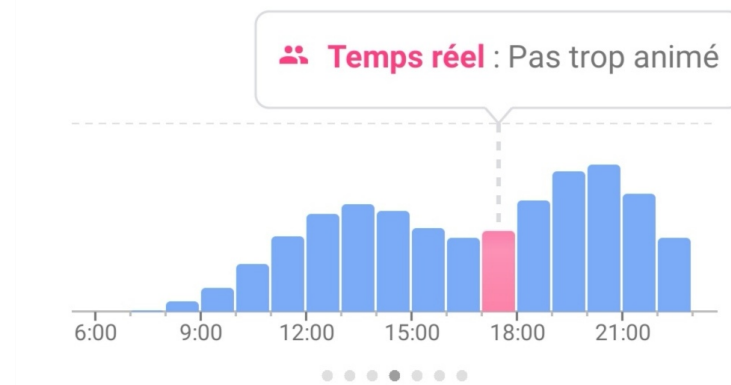


Figure 1.1 – Temps d’attente réel d’un restaurant affiché sur Google

des clients et sur les performance des systèmes étudiés.

Les modèles étudiés sont décrits dans la section suivante.

1.2 Organisation du manuscrit

Ce document est organisé en plusieurs chapitres.

Le chapitre 2 présente les concepts fondamentaux des systèmes stratégiques de files d’attente. Nous commençons par introduire les caractéristiques des systèmes de files d’attente et les notions de base. Ensuite, nous présentons les différents types de files d’attente stratégiques (observable, non-observable et partiellement observable).

Le chapitre 3 est divisé en deux parties. La première décrit une étude d’un modèle observable sur demande (les clients arrivent à un système non-observable et ils décident d’acquérir ou non l’information sur l’état du système avant de prendre leur décision). Nous étudions l’effet du coût d’inspection sur leurs décisions stratégiques. Dans la deuxième partie, nous comparons le premier modèle étudié avec un modèle qui facture l’accès au système et nous proposons une politique de tarification optimale. La contribution de ce chapitre a été publiée dans (NACEUR et al. 2019).

Dans le chapitre 4, nous présentons une étude qui combine les deux modèles fondamentaux de problèmes de files d’attente stratégiques en appliquant une politique de partage d’information. Nous étudions l’impact de cette politique sur la

performance globale du système lorsque les clients sont sensibles à cette information. Nous étudions aussi le même modèle dans le cas d'un système saturé. Ce travail de recherche a été publié dans le journal (NACEUR et al. 2020).

Le chapitre 5 est une généralisation du chapitre précédent. Nous considérons le même modèle du système avec une politique de partage de l'information, en supposant que des clients appartiennent aux différentes classes. Nous étudions l'impact de l'information sur les décisions stratégiques des clients et sur l'équilibre dans ce système hétérogène. Nous considérons un cas particulier où deux classes de clients se présentent au système. Ce travail est une collaboration avec le Professeur Manjesh K. Hanawal de l'institut indienne de technologie, Bombay dans le cadre du projet CEFIPRA qui a financé ma thèse de doctorat.

Enfin, dans le chapitre 6, nous étudions le comportement des clients dans un système avec déplacement. Les clients doivent se déplacer pour atteindre la file d'attente de service (exemple typique du service des urgences, restaurants, etc). Dans cette étude, la population est hétérogène et elle est divisée en deux types de clients : informés et non informés par l'état du système avant le voyage. Nous étudions l'impact de l'information dans ce système avec déplacement et partiellement observable. Cette contribution est en collaboration avec la Professeur Ricky Roet-Green de l'université de Rochester, New York. Ce travail a été effectué grâce à la bourse Perdiguier 2019 pour l'aide à la mobilité des doctorants de l'université d'Avignon qui m'a permis de faire un stage de recherche à l'université de Rochester.

Le manuscrit se termine par une synthèse du travail ainsi que des perspectives à explorer.

1.3 Publications

La liste de mes publications, y compris les conférences nationales et internationales ainsi qu'un journal, est la suivante :

– **Conférences nationales :**

- Tesnim Naceur, Yezekael Hayel, "Optimisation du bien-être social dans un système complexe avec attente et décision stratégique des usagers", ROADEF 2018, Lorient, France
- Tesnim Naceur, Yezekael Hayel, "Politiques de divulgation d'informations dans les systèmes de file d'attente stratégique afin de maximiser le gain social", ROADEF 2020, Montpellier, France

– **Conférence internationale :**

- Tesnim Naceur, Yezekael Hayel, "Revenue maximization with access and information pricing schemes in a partially-observable queueing game", VALUETOOLS 2019-12th EAI International Conference on Performance Evaluation Methodologies and Tools, Mar 2019, Palma, Spain.

– **Journal :**

- Tesnim Naceur, Yezekael Hayel, "Deterministic State based Information Disclosure Policies and Social Welfare Maximization in Strategic Queueing Systems", accepté dans Queueing Systems, special issue on Strategic Queueing, 2020.

Chapitre 2

Concepts de base et état de l'art

Sommaire

2.1	Introduction	10
2.2	Systèmes de files d'attente	10
2.2.1	Caractéristiques	10
2.2.2	Notion de classe de clients	12
2.2.3	Notation de Kendall	13
2.2.4	Systèmes de file d'attente Markovienne	13
2.3	Théorie de jeux sur les files d'attente	15
2.3.1	Notions de base	15
2.3.2	File d'attente observable	19
2.3.3	File d'attente non-observable	20
2.3.4	File d'attente avec une observation partielle	22
2.4	Conclusion	25

2.1 Introduction

La littérature sur la théorie des files d'attente a connu une croissance exponentielle au fil des années depuis le début du siècle dernier (Voir le livre (D. GROSS et al. 2008)). Ce domaine d'étude fournit un cadre mathématique efficace pour l'étude de plusieurs phénomènes de congestion dans divers domaines d'application. De nombreux phénomènes réels peuvent être représentés au moyen de modèles de files d'attente tels que les lignes de caisse dans les grandes surfaces, les lignes de production, les réseaux de communication, ainsi que les systèmes militaires. Le formalisme des files d'attente permet de modéliser convenablement les problèmes à étudier et de répondre à de nombreuses questions liées à notre problématique.

Dans ce chapitre, nous présentons, dans une première partie, les concepts fondamentaux des files d'attente tels que leurs caractéristiques, la notion de classe de clients, la notation de Kendall et les systèmes de file d'attente Markovienne. Dans une deuxième partie, nous abordons la théorie de jeux sur les files d'attente et nous présentons les principaux résultats dans la littérature des systèmes de files d'attente stratégiques.

2.2 Systèmes de files d'attente

2.2.1 Caractéristiques

On parle d'un phénomène d'attente chaque fois que certaines entités, appelées aussi clients, arrivent à un système afin de recevoir un service par un ou plusieurs serveurs et attendent d'être servis. Une discipline de service est utilisée afin de servir les clients.

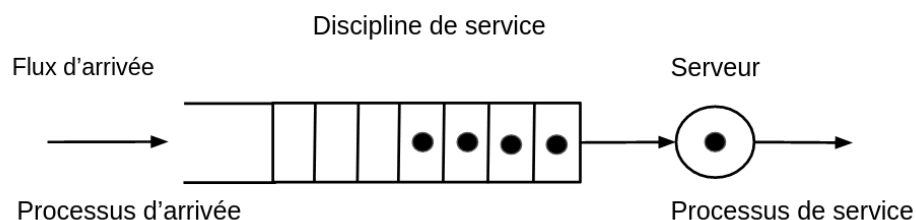


Figure 2.1 – Représentation d'une de file d'attente simple.

2.2.1.1 Flux d'arrivée

On est confronté à des systèmes de files d'attente, dès que nous parlons de flux d'arrivée de clients. Les arrivées aux files d'attente peuvent être déterministes ou aléatoires, dépendants ou indépendants, individuelles ou groupées, homogènes ou hétérogènes.

Notons que le processus de Poisson est le plus utilisé pour caractériser les processus d'arrivée dans les systèmes de file d'attente.

2.2.1.2 Serveur

Le flux d'arrivée se présente à un dispositif, appelé serveur qui fournit le service demandé. Le temps de service peut être décrit par une loi de probabilité ou être déterminé dès le début. Plusieurs facteurs agissent sur le temps de service tels que :

- le nombre de serveurs,
- le type de disposition (en parallèle, en série ou en réseaux),
- la classe de clients si la population est hétérogène,
- le moment de la journée, etc.

Notons que la distribution du temps de service la plus utilisée est la distribution exponentielle.

2.2.1.3 Discipline de service

La discipline de service est la méthode de sélection du prochain client lorsque le serveur termine le service du client courant. Un certain nombre de disciplines sont utilisées telles que :

- Premier arrivé, premier servi ou en anglais First come First served (FCFS) qui est équivalente à la discipline Premier entré, premier sorti ou en anglais First In First Out (FIFO) : cette discipline respecte l'ordre d'arrivée des clients. Elle est utilisée surtout lorsque le service est effectué sur un serveur unique.
- Dernier arrivé, premier servi ou en anglais Last In First Out (LIFO) : cette discipline consiste à servir en premier le dernier client arrivé. Elle est utilisée surtout pour la gestion du stock où le système de pile est appliqué : celui qui vient plus tard part plus tôt.

- **Priorité** : les clients sont servis, avec cette discipline, suivant leur importance dans la population. Une personne prioritaire sera servie avant une personne non-prioritaire, même si elle est arrivée avant. Cette discipline s'impose dans les services d'urgence, les aéroports, etc.
- **Aléatoire** : cette discipline consiste à sélectionner le client à servir aléatoirement.

2.2.2 Notion de classe de clients

Une population peut contenir des usagers hétérogènes qui appartiennent aux différentes classes de clients. Chaque classe est distinguée par :

- son processus d'arrivée,
- son processus de service,
- son ordre de service dans le système.

Avec la notion de classes, la discipline de service utilisée généralement est celle avec priorité. Dans ce cas, deux situations sont envisagées :

- **Un système non préemptif** : le service du client non-prioritaire continue jusqu'à son terme avant le traitement du client prioritaire. Citant l'exemple des hôpitaux et les urgences où il faut terminer une intervention chirurgicale avant de traiter un autre cas, même plus urgent.
- **Un système préemptif** : le service du client non-prioritaire est interrompu par l'arrivée d'un client prioritaire et il reprend dès que le traitement du client prioritaire se termine et s'il n'y a plus de personne prioritaire dans la file. C'est le cas notamment de certaines activités, qui donnent la priorité aux femmes enceintes et aux personnes handicapées.

Dans ce contexte, plusieurs travaux de recherche ont traité des systèmes de files d'attente stratégiques avec plusieurs classes de clients. Alan Cobham a écrit l'un des premiers articles (COBHAM 1954) bien connus sur les systèmes de files d'attente prioritaires. Adiri et Yechiali dans (ADIRI et al. 1974) se sont concentrés sur un système qui contient deux classes des clients. Ils ont traité les règles optimales d'achat de priorité pour un nouveau client qui arrive au système, et les politiques de tarification optimales pour les disciplines de priorité préemptive et non préemptive. L'article (ALPERSTEIN 1988) a considéré le même modèle, mais avec plusieurs niveaux de priorité gérés par un serveur maximisant le revenu. L'auteur a prouvé que le revenu augmente avec le nombre de priorités des clients. Hassin et Haviv

(HASSIN et HAVIV 1997) ont aussi poursuivi cette ligne de recherche et ils ont élargi l'ensemble des stratégies possibles pour inclure des stratégies mixtes au modèle étudié. Coffman et Mitrani (COFFMAN et al. 1980) caractérisent la région réalisable des files d'attente $M/M/1$ avec n classes des clients. Le service est préemptif dans leur modèle. Des résultats similaires ont été montrés pour des modèles plus généraux dans (FEDERGRUEN et al. 1988 ; SHANTHIKUMAR et al. 1992) qui proposent, avec une caractérisation du système, des solutions de maximisation des objectifs de performance dans le système. Dans (R. HASSIN et al. 2009), les auteurs ont étudié un système de file d'attente avec n classes de clients et ils ont étudié la minimisation des coûts d'attente associés pour chaque classe et en particulier lorsque le temps d'attente est limité.

2.2.3 Notation de Kendall

David George Kendall a créé une notation pour décrire un système de file d'attente. La nomenclature générale d'une file d'attente en utilisant cette notation est de la forme $A/S/c/K/D$ où :

- A décrit la distribution de probabilité des temps entre les arrivées des clients,
- S décrit la distribution de probabilité du temps de service d'un client,
- c représente le nombre de serveurs,
- K représente la capacité de la file d'attente : finie ou infinie,
- D décrit la discipline de la file : FIFO, LIFO, avec priorité, aléatoirement, etc.

La version la plus utilisée de la notation Kendall contient seulement trois paramètres $A/S/c$. Plusieurs types de distributions sont possibles pour les temps d'arrivée et les temps de service tel que la loi de Markov (M), la loi constante ou déterministe (D), et la loi générale (G).

La file d'attente $M/M/1$ est un exemple classique où les clients arrivent et sont servis suivant la loi de Markov, à un serveur unique.

2.2.4 Systèmes de file d'attente Markovienne

En 1906, A. A. Markov a commencé l'étude d'un nouveau type de processus stochastique où le résultat d'une expérience donnée peut affecter le résultat de l'expérience suivante (Voir (A.A.MARKOV 1906)). Ce type de processus est appelé

chaîne de Markov. La théorie des processus de Markov est un fondement théorique des systèmes de file d'attente qui a été abondamment étudié dans la littérature (Voir (G. BOLCH et al. 2005; BAUERLEET et al. 2011)).

Les modèles de files d'attente les plus simples à étudier et à analyser sont les modèles markoviens où la distribution des arrivées et de service suit la loi exponentielle.

2.2.4.1 Graphe d'une chaîne de Markov

Un graphe orienté est utilisé pour représenter l'évolution d'une chaîne de Markov. Il est défini à partir d'un ensemble d'états et d'une matrice de transition. Les nœuds du graphe représentent les états possibles dans une chaîne de Markov, et les flèches représentent les transitions d'un état à un autre. Un exemple élémentaire d'une chaîne de Markov est représenté dans la Figure 2.2.

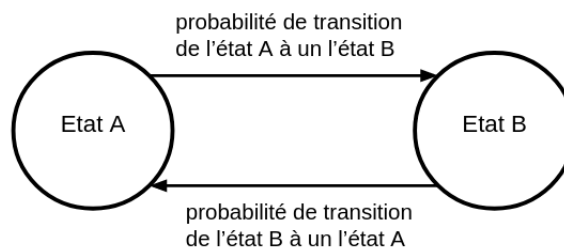


Figure 2.2 – Exemple élémentaire de chaîne de Markov

Une flèche allant de l'état A vers l'état B indique qu'avec une probabilité strictement positive $p(A,B)$, le prochain état sera l'état B et inversement, une flèche allant de l'état B vers l'état A indique qu'avec une probabilité strictement positive $p(A,B)$, le prochain état sera l'état A .

2.2.4.2 Stabilité

Nous parlons de la stabilité du système lorsque le nombre de clients qui arrivent au système est strictement plus petit que le nombre des clients qui partent (Voir (KERNANE 2007)).

Dans le cas d'une file d'attente Markovienne simple, qui dispose (k) serveur avec un taux d'arrivée (λ) et un taux de service (μ), le système est dit stable si et

seulement s'il satisfait la condition suivante :

$$\frac{\lambda}{\mu} < k$$

Pour des raisons de simplification du calcul, nous notons par $\rho = \frac{\lambda}{\mu}$ le taux d'utilisation du système.

Remarque. Lorsque, le nombre des clients qui arrivent au système est plus grand (ou égale) que le nombre des clients qui le quitte, c'est à dire $\rho = \frac{\lambda}{\mu} \geq k$, le système est dit saturé.

2.3 Théorie de jeux sur les files d'attente

2.3.1 Notions de base

La théorie des jeux est un cadre théorique pour traiter des situations sociales entre joueurs concurrents. Elle se base sur les interactions stratégiques au sein d'un groupe d'individus. Cette théorie a été traité pour la première fois dans (COURNOT 1838) puis dans (BERTRAND 1883). Les études de la théorie des jeux moderne ont commencé dans les année 1940 avec le mathématicien John von Neumann et l'économiste Oskar Morgenstern (Voir (NEUMANN et al. 1944)). Elles a été appliquées avec succès dans nombreux domaines tel les sciences l'économie, les sciences politiques, la biologie, etc.

2.3.1.1 Généralités

Dans la théorie des jeux, les joueurs sont supposés être des individus rationnels.

Définition 1. *Un chaque joueur a des objectifs bien définis par rapport à l'ensemble des résultats possibles et il met en œuvre la meilleure stratégie disponible pour les atteindre.*

La rationalité implique que les individus connaissent les stratégies disponibles et peuvent déterminer la meilleure stratégie pour eux-mêmes et la mettre en œuvre. Nous définissons, par ailleurs, la théorie de jeux comme suit :

Définition 2. *La théorie de jeux est une étude des interactions stratégiques d'un groupe d'individus rationnels.*

Nous donnons quelques termes couramment utilisés dans la théorie des jeux :

- Jeu : tout ensemble du système dont les joueurs, les stratégies, les informations disponibles.
- Ensemble des joueurs : des décideurs stratégiques dans le contexte du jeu.
- Ensemble de stratégies : un ensemble des actions possibles qui représentent la spécification complète du comportement d'un joueur dans n'importe quelle situation. Deux types de stratégies peuvent exister dans les jeux stratégiques : soit des choix de stratégies (stratégie pure), soit des distributions de probabilités (stratégie mixte).
- Gain : La récompense qu'un joueur reçoit en arrivant à un résultat particulier. Ce gain ne dépend pas seulement de sa stratégie choisie mais, aussi des stratégies choisies par le reste des joueurs.
- Ensemble d'informations : les informations disponibles à un moment donné du jeu.
- Équilibre : le point dans un jeu où les deux joueurs ont pris leurs décisions et un résultat est atteint.

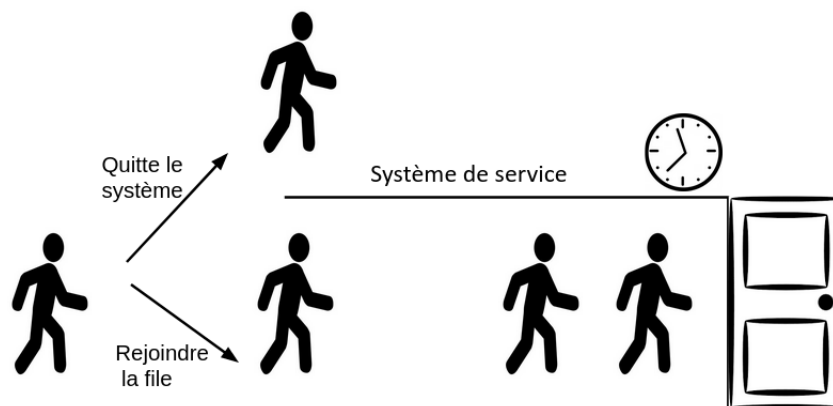


Figure 2.3 – Processus de prise de stratégie dans une file d'attente simple

Remarque. Dans un jeu stratégiques, toutes les informations du jeu sont fournies à tous les joueurs et chaque joueur sait que tous les autres joueurs savent tout sur le jeu.

2.3.1.2 Types de jeux

Il existe de nombreux types de jeux dans la théorie de jeux tels que les jeux simultanés et séquentiels, symétriques et asymétriques, mais les plus courantes et

les plus utilisées sont les jeux coopératifs et non-coopératifs (dits aussi stratégiques).

Définition 3. *Les jeux coopératifs modélisent les systèmes où les joueurs communiquent entre eux et prennent des actions de façon à atteindre un objectif commun.*

La théorie des jeux coopératifs se base sur la communication, l'échange et la coopération entre les joueurs. Par contre, les jeux non-coopératifs sont mieux adaptés pour modéliser les décisions décentralisées sur les performances des systèmes de files d'attente.

Définition 4. *Les jeux non-coopératifs modélisent les systèmes où chaque joueur prend ses décisions individuellement pour maximiser sa fonction d'utilité espérée.*

Dans ce genre de jeux, un joueur ne prend sa décision sans concertation préalable avec les autres joueurs. Généralement, la fonction d'utilité à optimiser correspond au profit individuel de chaque joueur.

2.3.1.3 L'équilibre de Nash

Le mathématicien John Nash est considéré comme le chercheur qui a fait la première extension significative des travaux de von Neumann et Morgenstern. Il a introduit l'un des concepts fondamentaux de la théorie des jeux (Voir (NASH 1950 ; NASH 1951)) : la conceptualisation du comportement des joueurs pour déterminer la solution d'équilibre dans un jeu. Sous l'équilibre de Nash, chaque joueur n'a aucun intérêt à changer sa stratégie initiale, en supposant que les autres joueurs gardent également leurs stratégies inchangées. Cet équilibre repose sur deux hypothèses :

- chaque individu choisit ses stratégies de manière rationnelle,
- ses croyances sur les stratégies des autres joueurs sont correctes.

Une fois que chaque individu agit conformément à l'équilibre de Nash, personne n'est incité à s'écarter et à prendre une autre décision. Chaque joueur joue une meilleure réponse contre les autres joueurs simultanément, dans le sens où une fois la décision prise, le joueur n'aura aucun regret concernant les décisions compte tenu des conséquences. L'équilibre de Nash est un concept principal dans les jeux non-coopératifs.

Un jeu stratégique peut avoir plusieurs équilibres de Nash, mais il peut aussi n'en avoir aucun (Voir (SMITH 1982)). L'article de (R.HASSIN et al. 2002) contient plus de détails sur l'équilibre de Nash dans les systèmes de files d'attente stratégiques.

2.3.1.4 Comportement des joueurs dans un jeu stratégique

Dans le cas où la meilleur réponse est strictement monotone en fonction de l'état du système, deux types de caractéristiques du système peuvent être constatés.

Définition 5. *Un comportement ATC (Avoid the crowd) : un comportement des joueurs dans un jeux donné qui a tendance à éviter la foule.*

Si la meilleur réponse est strictement décroissante le comportement des clients est un comportement ATC.

Définition 6. *Un comportement FTC (Follow the crowd) : un comportement des joueurs dans un jeux donné qui a tendance à suivre la foule.*

Si la meilleur réponse est strictement croissante le comportement des clients est un comportement FTC.

Des observations correspondent aux comportements de suivi de la foule (FTC) et d'évitement de la foule (ATC) sont illustrées dans la Figure 2.4.

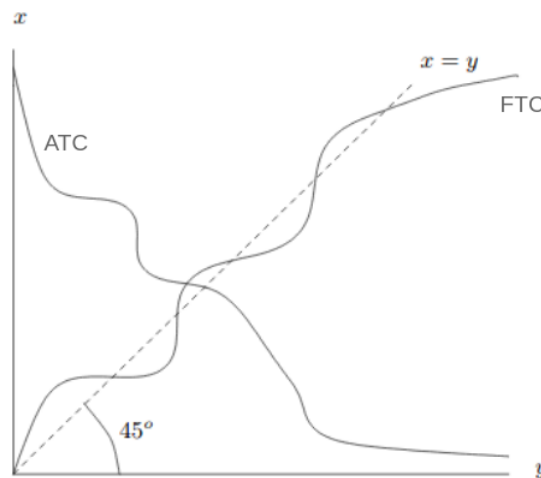


Figure 2.4 – Exemples d'un comportement ATC et FTC (Figure extraite de (HASSIN et HAVIV 2003))

Remarque. Dans le cas d'un comportent des clients de type ATC, il existe au plus un équilibre alors que des équilibres multiples sont possibles dans le cas du FTC.

Le comportement stratégique du client dans les systèmes de service non-observables est un mélange d'ATC et de FTC, tandis que dans les systèmes de service observables, les clients évitent toujours la foule (un comportement ATC). Plusieurs travaux de recherches ont traité les comportements des clients dans des systèmes de files d'attente

stratégiques tels que (HASSIN et ROET-GREEN 2017; BOUNTALI et al. 2017; M. ZHOU et al. 2018).

2.3.2 File d'attente observable

Dans cette partie, nous considérons des files d'attente observables : à leurs arrivées, les clients observent le nombre de personnes en attente et en service et ils prennent leur décision de rejoindre ou non le système selon l'état du système.

2.3.2.1 Modèle du Naor

L'étude de Naor (NAOR 1969), se base sur le contrôle d'un système FCFS de type $M/M/1$ où les clients réagissent en définissant une stratégie pure selon l'état du système. Deux choix sont possible dans ce modèle : rejoindre la file ou quitter sans service. La décision de chaque client ne dépend pas explicitement de la stratégie adoptée par les autres clients mais de l'état du système. Le modèle de Naor se base sur plusieurs hypothèses :

- λ est le taux d'arrivée des clients,
- μ est le taux de service,
- C est le coût d'attente par unité de temps,
- R est l'évaluation de service.

À son arrivée, le client examine l'état de la file d'attente et décide s'il va rejoindre le système ou non. Celui qui décide de quitter le système, ne revient jamais et ne reçoit pas le service. Un client décide de rejoindre la file d'attente, lorsque i clients sont déjà dans le système, si son utilité espérée est strictement positive. Cette utilité est donnée par l'expression suivante :

$$U(i) = R - \frac{(i+1)C}{\mu}.$$

Le client rejoint le système si et seulement si son utilité espérée est positive. Le cas d'un client qui arrive à un système vide est le cas le plus favorable pour avoir le service. Par contre, si le nombre de clients qui attendent dans le système est élevé, un nouvel arrivant aura tendance à éviter la file d'attente. Une stratégie de seuil noté par n_e est défini par Naor telle que

$$n_e = \lfloor \frac{R\mu}{C} \rfloor.$$

En appliquant cette stratégie pure, un client rejoint la file d'attente s'il observe au plus $n_e - 1$ clients dans le système, sinon il la quitte. Le nombre maximal possible de clients est par la suite n_e clients. D'où, dans le cas observable, la file d'attente est de type $M/M/1/n_e$.

2.3.2.2 Généralisation du modèle de Naor

Le modèle de Naor a été développé et utilisé dans plusieurs travaux de recherche tels que les enquêtes réalisées par Johansen et Stidham (JOHANSEN et al. 1980), Stidham (STIDHAM 1985) et Mendelson et Whang (MENDELSON et al. 1990). Les conditions générales d'optimiser des politiques de type seuil pour contrôler les systèmes d'entrée-sortie de Poisson ont été données par Hassin et Henig (HASSIN et HENIG 1986).

2.3.3 File d'attente non-observable

Dans cette partie, nous considérons le modèle des files d'attente non-observables où les clients n'observent pas l'état du système avant de prendre leur décision de rejoindre la file ou de partir sans service.

2.3.3.1 Le modèle d'Edelson et Hildebrand

Les propriétés de la file d'attente à un serveur unique non-observable ont été étudié la première fois par Edelson et Hildebrand (EDELSON et al. 1975). Ils ont adopté toutes les hypothèses du modèle observable de Naor, en ajoutant quelques modifications. Au moment de faire leur choix, les clients ne sont pas informés de l'état du système. Ils prennent leurs décisions de rejoindre la file d'attente ou de la quitter sans avoir l'accès à cette information. Une fois que leur décision est prise, ils n'ont pas le droit de la changer. Le temps d'attente espéré d'un client est noté par W . Le choix d'avoir le service et de rejoindre la file dépend du choix des autres clients et il est fait en fonction de l'utilité espérée pour chaque client.

2.3.3.2 Équilibre de Nash

Deux stratégies pures sont disponibles pour un nouveau client qui arrive au système : rejoindre la file d'attente ou la quitter. Nous désignons la probabilité à l'équilibre de rejoindre le système par q_e . Le taux d'arrivée effectif à l'équilibre est

noté λ_e , avec $\lambda_e = q_e \lambda < \mu$. L'utilité espérée pour un client qui rejoint la file d'attente est donnée par l'expression suivante :

$$\bar{V} = \sum_{i=0}^{\infty} (R - C \frac{i+1}{\mu}) \Pi_i$$

avec Π_i la probabilité stationnaire qu'il y ait i clients dans le système à son arrivée. Cette probabilité dépend des probabilité q_e . La meilleure réponse de chaque client sera prise en fonction du gain et nous distinguons trois possibilités d'équilibre. Le premier cas est lorsque le gain est négatif, c'est-à-dire $\bar{V} < 0$. La stratégie de rejoindre le système à l'équilibre est alors la probabilité $q_e = 0$ qui est la seule stratégie d'équilibre. Ensuite, le deuxième cas est lorsque le gain est positif, c'est-à-dire $\bar{V} > 0$. Dans ce cas, même si tous les clients rejoignent le système, leur utilité n'est pas négatif. Par conséquent, la seule stratégie de rejoindre le système à l'équilibre est la probabilité $q_e = 1$. Enfin, le troisième cas, lorsque $\bar{V} = 0$. Dans ce cas, la stratégie prise par les clients à l'équilibre est une stratégie mixte avec $q_e \in]0,1[$.

La Figure 2.5 illustre la fonction de meilleure réponse et le point d'équilibre dans ce modèle non-observable. Sachant que la probabilité de rejoindre le système est q , la meilleure réponse unique est 1 si $q < q_e(p)$, elle est entre 0 et 1 si $q = q_e(p)$ et elle est égale à 0 si $q > q_e(p)$.

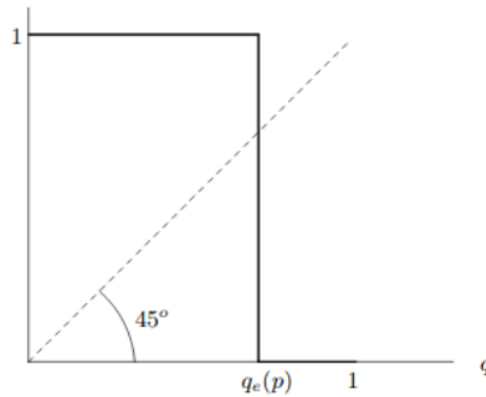


Figure 2.5 – Meilleure réponse (Figure extraite de (HASSIN et HAVIV 2003))

Remarque. Étant donné que, la meilleure réponse est une fonction monotone décroissante, le modèle a un comportement du type ATC. Il existe alors un équilibre unique.

2.3.3.3 Généralisation du modèle d'Edelson et Hildebrand

Plusieurs travaux de recherche ont discuté les résultats de (EDELSON et al. 1975). Chen et Frank (CHEN et al. 2004) ont démontré dans leur travail (CHEN et al. 2004), qui traite une file non-observable, qu'un coût socialement optimal peut donner un bien-être du client. Balachandran dans (BALACHANDRAN 1991) a considéré un modèle $M/G/1$ non-observable avec un coût fixe pour gérer le service. Il a étudié l'impact des coûts sur le taux d'arrivée à l'équilibre et il a prouvé que lorsque le taux d'arrivée augmente, le bien-être d'un client sera affecté parce que le coût d'attente attendu augmente et que le taux d'arrivée à l'équilibre est unique.

2.3.4 File d'attente avec une observation partielle

2.3.4.1 Contexte

Les modèles de file d'attente observable et non-observable sont deux modèles mathématiques qui ont été utilisés avec succès dans de nombreuses études de formalisation de processus décisionnels. L'impact de ces études amène les chercheurs à étudier des systèmes plus complexes. Des nombreuses recherches sont apparues, qui traitent des modèles de files d'attente partiellement observable où l'état des files est contrôlé et accessible seulement à une partie de la population.

Une file d'attente avec une observation partielle est un processus décisionnel de Markov dans lequel seulement une partie de la population connaît l'état réel du système. Le client qui a accès à l'état de la file, applique le modèle de (NAOR 1969) d'une file d'attente observable et le client qui n'a pas accès à l'information, prend sa décision en fonction des informations disponibles sur le système et les stratégies prises par les autres clients. La formalisation de ces systèmes de file d'attente permet de modéliser des situations plus réalistes inspirées de la vie de tous les jours où les clients n'ont pas toutes les mêmes opportunités pour accéder à l'état du système donné d'une file d'attente. Nous citons dans ce contexte, l'exemple des files d'attente pour environ un million de restaurants dans le monde où leurs temps d'attente moyens sont désormais visibles en temps réel sur Google Maps. Tout client peut obtenir des informations sur les performances du système par une simple recherche sur Google, comme l'occupation moyenne pour une période particulière dans la journée et le temps de séjour moyen d'un client. Tous ces nouveaux systèmes d'information ont un impact sur la décision stratégique des clients et sur les performances du système lui-même. Par contre, seulement les clients qui ont les moyens et qui font la recherche, peuvent accéder à toutes ces informations. Dans ce cas, la population

est divisée en deux parties : clients informés et clients non informés.

2.3.4.2 Généralisation du modèle avec une observation partielle

Beaucoup de recherches sont faites sur l'impact de l'information dans les systèmes de files d'attente. L'article (HASSIN et HENIG 1986) est l'un des premiers travaux liés à la comparaison de l'impact de l'information sur les performances du système. L'auteur a considéré les deux structures d'information extrêmes, les modèles complètement observables et non-observables, et il a comparé deux paramètres de performance : le bien-être social et le revenu du fournisseur. Cette étude a donné un premier aperçu sur l'importance de l'information dans les systèmes de files d'attente stratégiques. Haviv et Hassin dans leur article (HASSIN et HAVIV 1994) ont étudié un mélange des modèles observables et non-observables, en supposant que seulement une fraction de la population peut voir l'état du système à leur arrivée. Les clients arrivent à un système de deux serveurs et ils peuvent obtenir les informations sur la file d'attente la plus courte en payant un certain coût. Un client qui n'achète pas ces informations choisit une file d'attente au hasard. Les auteurs ont également abordé la question de savoir si les clients achètent la quantité d'informations socialement souhaitée en équilibre.

Récemment, il y a de nouvelles recherches qui envisagent les modèles avec une information contrôlée. Dans l'article (M. HU et al. 2018), les auteurs supposent que seulement une proportion de clients est informée. Ils donnent une caractérisation de l'équilibre en fonction de cette proportion et montrent que le taux d'arrivée effectif est unimodal en fonction de cette proportion. Un modèle similaire est étudié dans (ECONOMOU et al. 2015) où les auteurs ont prouvé qu'il existe une fraction optimale de clients informés pour le bien-être social du système. Récemment, R. Hassin et R. Roet-Green (HASSIN et ROET-GREEN 2017) ont présenté une analyse approfondie et complète du modèle de mise en file d'attente $M/M/1$ où les clients choisissent de : rejoindre la file d'attente, la quitter sans service ou inspecter la longueur de la file d'attente à un coup fixe. L'analyse de l'équilibre est assez complexe dans cette étude mais les auteurs ont donné certaines caractérisations et ils ont prouvé l'existence de l'équilibre. Ils ont analysé l'impact du coût d'inspection sur l'équilibre, les revenus et le bien-être social. Ils ont discuté à la fois de cas clients homogènes et hétérogènes. Un autre type de modèle d'information est proposé dans (ALTMAN et al. 2013) où un signal binaire est fourni aux clients si le nombre de clients qui existent déjà dans la file d'attente est inférieur à un seuil donné. Les auteurs étudient les politiques optimales de contrôle d'admission et ils prouvent que la politique optimale pour le problème d'informations partielles ne coïncide pas avec la politique optimale dans

un modèle complètement observable. L'article (G. ALLON et al. 2011) présente aussi des différents signaux fournis aux clients afin d'influencer leur comportement dans une telle file d'attente stratégique. Les auteurs prouvent que seul un signal binaire conduit à une stratégie optimale pour le prestataire de services. Plusieurs types d'équilibres à deux signaux de la décision des clients sont possibles et les auteurs les étudient en détail. Dans (IBRAHIM 2018), Ibrahim a décrit les différents articles sur le sujet de l'information dans les files d'attente stratégiques. Dans (E. SIMHON et al. 2015), les auteurs étudient le modèle dans lequel le fournisseur partage les informations de longueur de file d'attente si le nombre actuel de clients dans la file d'attente est inférieur à une valeur donnée. Ils supposent que les clients sont homogènes et ils prouvent qu'informer les clients lorsque la longueur de la file d'attente est courte et ne pas les informer lorsqu'elle est longue n'est jamais optimal pour maximiser les revenus du fournisseur. Dans (KIM et al. 2017), les auteurs prouvent qu'avec une politique de partage d'information dans un système de file d'attente de type $M/M/1$ le système peut maximiser le revenu en informant tous les clients de la longueur de la file d'attente lorsque la longueur de la file d'attente est au-dessus du seuil spécifié. Le travail récent (Y. DIMITRAKOPOULOS 2019) considère un modèle alternatif dans lequel il y a des périodes de temps pendant lesquelles le système est observable et des périodes de temps qui ne le sont pas. Ces périodes observables et non-observables suivent des distributions exponentielles, et les auteurs montrent que cette structure d'information alternée a des impacts importants sur les performances des systèmes comme le débit d'équilibre et le bien-être social. Dans (ROET-GREEN et al. 2020), les auteurs ont comparé les performances du système, le débit et le bien-être social, dans deux types de systèmes : une file d'attente entièrement observable et une file d'attente partiellement observable. Dans (HASSIN et GREEN 2020), un système de service qui affiche l'information sur la longueur de file d'attente avant que les clients se déplacent pour y atteindre, a été étudié. Les auteurs ont comparé le modèle avec des clients informés et le modèle avec des clients non-informés par l'état du système avant le voyage.

Des enquêtes ont été faites et décrites les différents articles sur le sujet de l'information dans les files d'attente stratégiques (Voir chapitre 3 dans (HASSIN 2016) et (IBRAHIM 2018)).

2.4 Conclusion

Ce chapitre a un rôle introductif au domaine des files d'attente et de la théorie des jeux sur les files d'attente. Nous avons présenté, dans une première partie, les concepts fondamentaux des systèmes de file d'attente. Un aperçu a été fait pour les mesures de performance, la notation de Kendall, la notion des clients et les processus de Markov. Ensuite, nous avons introduit les files d'attente stratégiques. Nous avons tout d'abord, présenté les jeux sous formes stratégiques, et par la suite, nous avons parlé des deux systèmes d'informations classiques de file d'attente : observable et non-observable et nous avons conclu par les systèmes partiellement observables sur lesquels se base une grande partie des études faites dans cette thèse de doctorat.

Chapitre 3

Tarification de l'information dans une file d'attente stratégique

Sommaire

3.1	Introduction	28
3.2	Description du modèle avec tarification de l'information	28
3.2.1	Modèle	28
3.2.2	Utilités espérées	30
3.3	Détermination de l'équilibre	32
3.4	Maximisation du revenu	35
3.4.1	Faire payer l'accès à la file d'attente	36
3.4.2	Faire payer l'information sur la taille de la file d'attente	38
3.5	Politique de tarification optimale	39
3.6	Illustrations numériques	42
3.6.1	Charger l'accès à la file d'attente	42
3.6.2	Charger l'information	44
3.6.3	Politique de tarification optimale	44
3.7	Conclusion	45

3.1 Introduction

Aujourd'hui, dans certaines files d'attente non-observables, les clients peuvent acquérir l'information sur l'état de la file avant de prendre la décision d'y rejoindre ou non tels certains hôpitaux et centres de santé aux États-Unis et au Canada qui affichent sur leurs sites web le temps d'attente réel (Voir Figure 3.1). Les décisions

Site d'urgence	Capacité	Occupation		Séjour		Attente Hosp.
		Nb.	Écart	+ 24 h	+ 48 h	
CHU de Québec						
Hôpital St-Sacrement	22	13	-9	7	0	3
Hôpital l'Enfant-Jésus	48	57	+9	27	7	27
CHUL	43	40	-3	7	2	13
Hôpital St-François-d'Assise	34	35	+1	6	2	9
L'Hôtel-Dieu de Québec	15	17	+2	3	1	5

Figure 3.1 – Situation en temps réel dans les salles d'urgence de la région de Québec

sont prises selon leurs contraintes et leurs besoins et en prenant en compte les paramètres et l'état du système. Le coût de l'information joue un rôle très important pour la prise des décisions des usagers dans ce genre du système, et varie selon le contexte.

Dans ce chapitre, nous étudions dans un premier lieu, un processus de décision dans lequel le client peut inspecter l'information sur la longueur d'une file d'attente. Nous nous intéressons à l'impact du coût d'inspection, sur la prise de décision des clients. L'existence et l'unicité de l'équilibre ont été prouvées dans ce chapitre. Dans un second temps, nous comparons cette méthode de tarification de l'information avec la méthode classique de tarification à l'accès au système. Nous allons voir que la facturation de l'accès n'est pas toujours la meilleure politique de tarification pour maximiser le revenu du fournisseur.

3.2 Description du modèle avec tarification de l'information

3.2.1 Modèle

Nous considérons un système de file d'attente avec un serveur unique du type $M/M/1$ non-observable. Les clients arrivent selon un processus de Poisson avec un taux λ et sont servis de manière exponentielle avec un taux μ . Le coût d'attente pour

chaque unité de temps passée dans le système est noté $C \geq 0$. Pour simplifier le modèle, nous supposons que la population est homogène, c'est-à-dire que tous les clients sont identiques et ont la même évaluation du service $R > 0$. Les usagers qui arrivent au système sont confrontés à deux possibilités :

- prendre la décision d'inspecter l'information sur la longueur de la file d'attente, avant d'y rejoindre (ou non),
- prendre la décision de ne pas inspecter l'information et de rejoindre la file directement.

Nous notons par p la probabilité qu'un nouveau client décide d'inspecter l'état de la file d'attente.

Par la suite, en fonction de leur décision d'acquérir ou non l'information de la longueur de la file d'attente, chaque client décide de rejoindre ou non le système. La Figure 3.2 représente ce processus de décision pour dans le système.

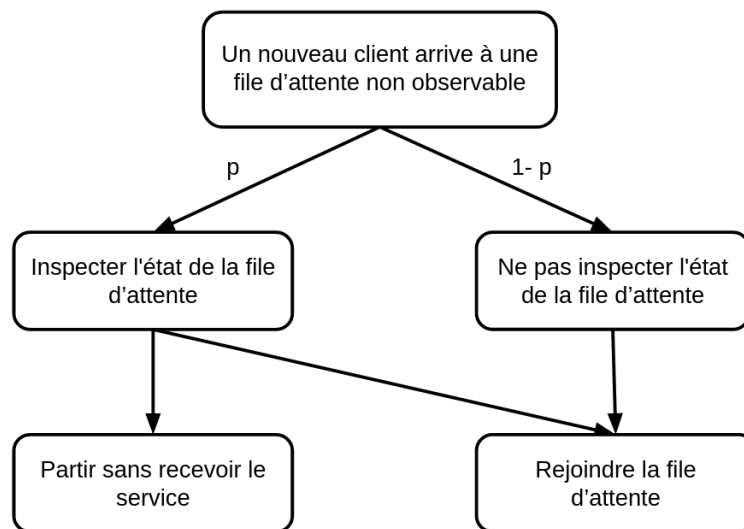


Figure 3.2 – Processus de prise de décision

Le coût d'inspection est noté par $C_I \geq 0$. L'accès à l'information sur l'état de la file d'attente en temps réel nécessite des outils spécifiques tels que un téléphone portable avec l'accès à Internet, le téléchargement d'une application mobile, le pré-enregistrement en ligne, etc. Ainsi, l'accès à l'information se fait à un certain coût.

3.2.2 Utilités espérées

Dans le cas où le client a inspecté la file d'attente, il décide de la rejoindre ou non en fonction de la longueur de la file d'attente, le taux de services, le coût d'attente et l'évaluation du service. Lorsqu'un nouveau client arrive au système, sachant que i clients sont déjà dans la file d'attente, il la rejoint si et seulement si $i \leq n_e - 1$, où :

$$n_e = \lfloor \frac{R\mu}{C} \rfloor. \quad (3.1)$$

Le taux d'utilisation du système est défini par $\rho = \frac{\lambda}{\mu}$ où $\lambda < \mu$.

Le taux d'arrivée effectif λ_e de clients dans le système dépend du nombre actuel de clients i dans la file d'attente, et il est défini comme suit :

$$\lambda_e(i) = \begin{cases} \lambda & \text{si } i \leq n_e - 1, \\ (1-p)\lambda & \text{sinon.} \end{cases}$$

Pour les états $i \geq n_e$, seuls les clients qui n'inspectent pas entrent directement dans la file. Afin de déterminer la distribution stationnaire du processus de Markov (Voir Figure 3.3), nous commençons par donné l'équation d'équilibre pour les états $i \in \{0, \dots, n_e - 1\}$:

$$\lambda\pi_i = \mu\pi_{i+1},$$

où π_i est la probabilité stationnaire dans l'état i . Pour les états $i \geq n_e$, l'équation d'équilibre est donnée par :

$$(1-p)\lambda\pi_i = \mu\pi_{i+1}.$$

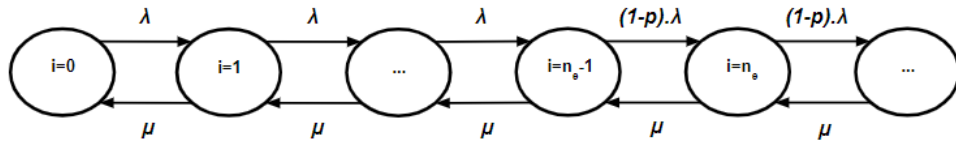


Figure 3.3 – Représentation en chaîne de Markov de l'évolution du système

Pour simplifier le calcul, nous définissons $\eta = 1 - (1-p)\rho$. La distribution stationnaire pour tous les états $i > 0$, selon la probabilité stationnaire π_0 , est la suivante :

$$\pi_i = \begin{cases} \rho^i \pi_0 & \text{si } i \leq n_e - 1, \\ \rho^{n_e} (1 - \eta)^{i-n_e} \pi_0 & \text{sinon.} \end{cases}$$

En sommant à un, nous obtenons la probabilité stationnaire π_0 lorsque le système de file d'attente est vide :

$$\pi_0 = \left[\sum_{i=0}^{n_e-1} \rho^i + \sum_{i=n_e}^{\infty} (1-\eta)^{i-n_e} \right]^{-1} = \left[\frac{1-\rho^{n_e}}{1-\rho} + \frac{\rho^{n_e}}{\eta} \right]^{-1}.$$

Remarque. Si $p = 0$, nous retrouvons un résultat connu : $\pi_0 = 1 - \rho$

Compte tenu du processus de décision illustré sur la Figure 3.2, l'utilité moyenne espérée U_I pour un client qui a inspecté la file d'attente est donnée par :

$$U_I = \sum_{i=0}^{n_e-1} \pi_i \left(R - C \frac{i+1}{\mu} \right) - C_I.$$

L'utilité moyenne espérée U_I peut être simplifiée par l'expression :

$$U_I = \pi_0 \left[R \frac{1-\rho^{n_e}}{1-\rho} - \frac{C}{\mu} \frac{1 - (n_e+1)\rho^{n_e} + n_e\rho^{n_e+1}}{(1-\rho)^2} \right] - C_I.$$

L'utilité moyenne espérée U_{NI} , pour un client qui n'a pas inspecté l'état de la file et rejoint directement la file d'attente, est donnée par :

$$U_{NI} = \sum_{i=0}^{\infty} \pi_i \left(R - C \frac{i+1}{\mu} \right).$$

L'utilité moyenne espérée U_{NI} peut être simplifiée par :

$$U_{NI} = \left(R - \frac{C}{\mu} \pi_0 \left[\frac{1 - (n_e+1)\rho^{n_e} + n_e\rho^{n_e+1}}{(1-\rho)^2} + \frac{\rho^{n_e}(n_e\eta+1)}{\eta^2} \right] \right).$$

Observation 1. Lorsque le coût d'inspection de l'information est nul, c'est-à-dire que $C_I = 0$, alors on a :

$$\forall p \in [0,1], \quad U_{NI}(p) < U_I(p).$$

Preuve

L'utilité espérée pour un client qui n'a pas inspecté l'état de la file, peut être écrit comme suit :

$$\begin{aligned} U_{NI} &= \sum_{i=0}^{\infty} \pi_i \left(R - C \frac{i+1}{\mu} \right), \\ &= U_I + C_I + \sum_{i=n_e}^{\infty} \pi_i \left(R - C \frac{i+1}{\mu} \right). \end{aligned}$$

Notez que $\forall i \geq n_e$ nous avons $R < C \frac{i+1}{\mu}$ par la définition du seuil de Naor n_e donné dans l'équation (3.1). Ensuite, lorsque $C_I = 0$ et pour tout $p \in [0,1]$, nous avons :

$$U_{NI}(p) < U_I(p).$$

■

Cette observation est un résultat assez intuitif qui montre que tout client décidera d'inspecter l'état de la file d'attente si cette opération ne coûte rien. Nous en déduisons le corollaire suivant.

Corollaire 1. Lorsque $C_I = 0$, l'équilibre dans le système est $p^* = 1$.

3.3 Détermination de l'équilibre

Auparavant, nous avons pu donner un premier résultat intuitif sur l'équilibre dans un contexte très spécifique où le coût d'inspection est nul. Nous recherchons maintenant à calculer l'équilibre dans le cas général.

Nous pouvons étudier l'équilibre en considérant l'utilité d'un client marqué par rapport à l'ensemble de la population. En fait, comme la population est infinie, un seul client n'a pas impact sur les performances globales du système (HASSIN 2016). Par conséquent, nous désignons par $U(p, p')$ l'utilité d'un client qui décide d'inspecter l'état de la file avec la probabilité p lorsque tous les autres clients décident d'inspecter l'état de la file avec la probabilité p' . Elle s'écrit comme suit :

$$U(p, p') = pU_I(p') + (1 - p)U_{NI}(p').$$

Définition 7. Un équilibre est une stratégie p^* tel que :

$$p^* = \arg \max_p U(p, p^*).$$

Un équilibre est par définition la meilleure réponse contre lui-même et cette notion d'équilibre correspond dans ce cas à l'équilibre de Nash (Voir (NASH 1951)). Sur la base de la définition formelle d'équilibre, nous donnons dans le théorème

ci-dessous une expression explicite de la proportion de clients qui décident d'obtenir l'information de la longueur de la file d'attente à l'équilibre. Pour simplifier la présentation du résultat, nous définissons les grandeurs mathématiques suivantes :

$$\begin{aligned}
 K_1 &= \rho^{n_e}(C(n_e + 1) - R.\mu), \\
 K_2 &= C_I(1 - \rho^{n_e+1}), \\
 K_3 &= \mu C_I(1 - \rho^{n_e})\rho^2, \\
 K_4 &= -(C_I(1 - \rho^{n_e})\rho\mu + K_2\rho\mu + (1 - \rho)(\rho^{n_e+1}C - K_1\rho)), \\
 K_5 &= K_2\mu - K_1(1 - \rho), \\
 \Delta &= K_4^2 - 4K_3K_5.
 \end{aligned}$$

Théorème 1. *Pour l'ensemble des paramètres du système, un équilibre p^* existe et est défini comme suit :*

$$p^* = \frac{2\rho - 1}{2\rho} - \frac{(1 - \rho^{n_e+1})}{2\rho(1 - \rho^{n_e})} - \frac{(1 - \rho)(\rho^{n_e+1}C - K_1\rho) - \sqrt{\Delta}}{2\mu C_I(1 - \rho^{n_e})\rho^2}.$$

Preuve

À partir de la définition d'un équilibre p^* , nous devons avoir la propriété à l'équilibre de principe d'indifférence ((HASSIN 2016)), qui s'exprime par l'équation suivante :

$$U_{NI}(p^*) = U_I(p^*),$$

ce qui est équivalent à

$$\begin{aligned}
 \sum_{i=n_e}^{\infty} \pi_i(p^*) \left(R - C \frac{i+1}{\mu} \right) &= -C_I, \\
 \sum_{i=n_e}^{\infty} \pi_i(p^*) \left(C \frac{i+1}{\mu} - R \right) &= C_I,
 \end{aligned}$$

avec

$$\pi_i(p) = \frac{\rho^i(1-p)^{i-n_e}}{\frac{1-\rho^{n_e}}{1-\rho} + \frac{\rho^{n_e}}{1-(1-p)\rho}}.$$

Nous devons résoudre l'équation suivante :

$$\sum_{i=n_e}^{\infty} \rho^i(1-p)^{i-n_e} \left(C \frac{i+1}{\mu} - R \right) = C_I \left(\frac{1 - \rho^{n_e}}{1 - \rho} + \frac{\rho^{n_e}}{1 - (1-p)\rho} \right).$$

De la partie gauche, nous obtenons :

$$\begin{aligned}
 & \rho^{n_e} \sum_{i=0}^{\infty} (\rho(1-p))^i \left(C \frac{i + n_e + 1}{\mu} - R \right) \\
 &= \rho^{n_e} \sum_{i=0}^{\infty} (\rho(1-p))^i C \frac{i}{\mu} + \left(\frac{C(n_e + 1)}{\mu} - R \right) \rho^{n_e} \sum_{i=0}^{\infty} (\rho(1-p))^i \\
 &= \frac{\rho^{n_e} C(1-p)\rho}{\mu(1-(1-p)\rho)^2} + \frac{\rho^{n_e} C(n_e + 1) - R\rho^{n_e} \mu}{\mu(1-(1-p)\rho)}.
 \end{aligned}$$

Et, de la partie droite, nous obtenons :

$$\begin{aligned}
 & C_I \left(\frac{1 - \rho^{n_e}}{1 - \rho} + \frac{\rho^{n_e}}{1 - (1-p)\rho} \right) \\
 &= \frac{C_I(1 - \rho^{n_e})(1 - (1-p)\rho) + C_I \rho^{n_e}(1 - \rho)}{(1 - (1-p)\rho)(1 - \rho)}.
 \end{aligned}$$

Ensuite, nous obtenons :

$$\begin{aligned}
 & \frac{\rho^{n_e} C(1-p)\rho + (\rho^{n_e} C(n_e + 1) - R\rho^{n_e} \mu)(1 - (1-p)\rho)}{\mu(1 - (1-p)\rho)^2} \\
 &= \frac{C_I(1 - \rho^{n_e})(1 - (1-p)\rho) + C_I \rho^{n_e}(1 - \rho)}{(1 - (1-p)\rho)(1 - \rho)}.
 \end{aligned}$$

Ce qui équivaut à :

$$\begin{aligned}
 & (1-p)^2 (\mu C_I(1 - \rho^{n_e})\rho^2) - (1-p)(C_I(1 - \rho^{n_e})\rho\mu + K_2\rho\mu) \\
 & + (1-\rho)(\rho^{n_e+1}C - K_1\rho) + K_2\mu - K_1(1-\rho) = 0.
 \end{aligned}$$

avec $K_1 = \rho^{n_e}(C(n_e + 1) - R.\mu)$ et $K_2 = C_I(1 - \rho^{n_e+1})$.

Pour simplifier le calcul suivant, nous définissons :

$$P = (1 - p).$$

Donc, nous arrivons à résoudre une équation du deuxième degré :

$$K_3 P^2 + K_4 P + K_5 = 0.$$

Si $\Delta > 0$, les solutions de l'équation du deuxième degré sont :

$$P_1 = \left(\frac{-K_4 - \sqrt{\Delta}}{2K_3} \right) \text{ and } P_2 = \left(\frac{-K_4 + \sqrt{\Delta}}{2K_3} \right).$$

Ce qui équivaut à :

$$P_1 = \frac{1}{2\rho} + \frac{(1 - \rho^{n_e+1})}{(1 - \rho^{n_e})2\rho} + \frac{(1 - \rho)(\rho^{n_e+1}C - K_1\rho)}{2\mu C_I(1 - \rho^{n_e})\rho^2} - \frac{\sqrt{\Delta}}{2\mu C_I(1 - \rho^{n_e})\rho^2},$$

$$P_2 = \frac{1}{2\rho} + \frac{(1 - \rho^{n_e+1})2\rho}{(1 - \rho^{n_e})} + \frac{(1 - \rho)(\rho^{n_e+1}C - K_1\rho)}{2\mu C_I(1 - \rho^{n_e})\rho^2} + \frac{\sqrt{\Delta}}{2\mu C_I(1 - \rho^{n_e})\rho^2}.$$

Nous voulons par la suite, déterminer le signe de p_2 . On a :

$$\begin{aligned} \frac{1}{2\rho} &> \frac{1}{2} \text{ parce que } \rho < 1 \Rightarrow \frac{1}{\rho} > 1, \\ \frac{(1 - \rho^{n_e+1})}{(1 - \rho^{n_e})} \frac{1}{2\rho} &> \frac{1}{2} \text{ parce que } (1 - \rho^{n_e+1}) > (1 - \rho^{n_e}) \\ \frac{\sqrt{\Delta}}{2\mu C_I(1 - \rho^{n_e})\rho^2} &> 0 \text{ parce que } \sqrt{\Delta} > 0 \text{ et } (1 - \rho^{n_e}) > 0 \\ \frac{(1 - \rho)(\rho^{n_e+1}C - K_1\rho)}{2\mu C_I(1 - \rho^{n_e})\rho^2} &> 0 \text{ parce que} \\ \frac{(1 - \rho)(\rho^{n_e+1}C - K_1\rho)}{2\mu C_I(1 - \rho^{n_e})\rho^2} &= \frac{(1 - \rho)\rho^{n_e+1}(R\mu - C[\frac{R\mu}{C}])}{2\mu C_I(1 - \rho^{n_e})\rho^2} > 0. \end{aligned}$$

On peut en déduire que $P_2 > 1$, et comme $p = 1 - P$, on obtient $p_2 = 1 - P_2 < 0$. Par conséquent, pour ensemble de paramètres du système, un équilibre unique p peut exister défini par :

$$p^* = p_1 = 1 - P_1$$

ce qui est égale à,

$$p^* = \frac{2\rho - 1}{2\rho} - \frac{(1 - \rho^{n_e+1})}{2\rho(1 - \rho^{n_e})} - \frac{(1 - \rho)(\rho^{n_e+1}C - K_1\rho) - \sqrt{\Delta}}{2\mu C_I(1 - \rho^{n_e})\rho^2}.$$

■

3.4 Maximisation du revenu

Dans cette section, nous étudions la maximisation du revenu du fournisseur de services. Nous nous intéressons principalement au problème du coût d'accès au système. Le problème le plus connu est le mécanisme optimal de tarification du péage, qui s'applique principalement dans les jeux de routage (VERHOEF 2002). Dans notre problème, les informations données aux clients affectent leur comportement et leur décision et, par conséquent, un fournisseur peut optimiser ses revenus en facturant l'information sur la longueur de la file au lieu de facturer l'accès à la file.

Dans la suite du chapitre, nous voulons répondre à cette question : est-il plus rentable pour le fournisseur de facturer l'information plutôt que l'accès au système? Nous verrons que la réponse dépend en grande partie de la sensibilité des clients à leur temps d'attente.

Nous notons par R_I (respectivement R_A) le revenu du fournisseur en fonction du prix de l'information C_I (respectivement du prix de l'accès C_{Acc}). Notez que la variable C_I qui a été désignée comme un coût dans la section précédente est maintenant contrôlée et devient un prix imposé par le fournisseur de services.

En considérant ce mécanisme de tarification, les expressions des revenus peuvent s'écrire de la manière suivante :

$$R_I(C_I) = \lambda \cdot p^*(C_I) \cdot C_I, \quad (3.2)$$

où $p^*(C_I)$ est la proportion de clients qui acceptent de payer pour avoir l'information, et

$$R_A(C_{Acc}) = \lambda \cdot q^*(C_{Acc}) \cdot C_{Acc}, \quad (3.3)$$

où $q^*(C_{acc})$ est la proportion de clients qui paient l'accès à la file d'attente.

Notez que dans ce dernier contexte, les clients qui ne paient pas n'ont pas accès au service, tandis que dans l'autre cas, lorsque le fournisseur de services facture les informations relatives à la longueur de la file d'attente, certains clients ne rejoignent pas le service même s'ils ont payé pour obtenir l'information.

3.4.1 Faire payer l'accès à la file d'attente

Lorsque le fournisseur décide de facturer l'accès à la file d'attente, la fonction objectif de maximisation du revenu est écrite comme suit :

$$\max_{C_{Acc}} R_A(C_{Acc}) = \max_{C_{Acc}} \lambda \cdot q^*(C_{Acc}) \cdot C_{Acc}.$$

Selon le prix d'accès C_{acc} , une première proportion de la population rejoint la file d'attente et une deuxième proportion quitte le système sans obtenir le service, étant donné que les clients ne connaissent pas la longueur de la file d'attente. Ce problème correspond au problème de mise en file d'attente stratégique non-observable étudié pour la première fois dans (EDELSON et al. 1975).

L'équilibre, pour la proportion de clients qui rejoignent la file d'attente, en fonc-

tion des paramètres du système et en particulier le coût d'accès est donné par :

$$q^*(C_{Acc}) = \begin{cases} 1 & \text{si } C_{Acc} \leq R - \frac{C}{\mu - \lambda}, \\ \frac{\mu - \frac{C}{R - C_{Acc}}}{\lambda} & \text{si } R - \frac{C}{\mu - \lambda} \leq C_{Acc} \leq R - \frac{C}{\mu}, \\ 0 & \text{si } C_{Acc} > R - \frac{C}{\mu}. \end{cases}$$

Observation 2. L'équilibre $q^*(C_{Acc})$ décroît en fonction du coût d'accès C_{Acc}

Nous observons aussi, que le revenu est nul si le coût d'accès est nul et, au contraire, lorsque le coût d'accès va à l'infini, le revenu tend également à être nul. Ensuite, par continuité de la fonction, il existe presque un prix d'accès optimal C_{Acc}^* qui maximise les revenus du fournisseur $R_A(C_{Acc})$ lorsque le fournisseur décide de facturer l'accès aux clients de la file d'attente. En fait, on peut facilement prouver que ce maximum est unique et on le note $R_A^* := \max_{C_{Acc}} R_A(C_{Acc})$.

Proposition 1. Lorsque le fournisseur décide de facturer l'accès à la file d'attente aux clients, le revenu maximal R_A^* est donné par :

$$R_A^* = \max\left\{R_A\left(R - \frac{C}{\mu - \lambda}\right), R_A\left(R - \sqrt{\frac{RC}{\mu}}\right)\right\}.$$

Preuve

On distingue trois cas pour obtenir le revenu :

$$R_A(C_{Acc}) =$$

$$\begin{cases} \lambda C_{Acc} & \text{si } C_{Acc} \leq R - \frac{C}{\mu - \lambda}, \\ C_{Acc}\left(\mu - \frac{C}{R - C_{Acc}}\right) & \text{si } R - \frac{C}{\mu - \lambda} \leq C_{Acc} \leq R - \frac{C}{\mu}, \\ 0 & \text{si } C_{Acc} > R - \frac{C}{\mu}. \end{cases}$$

Par conséquent, nous obtenons les différents cas suivants :

-Lorsque $C_{Acc} \leq R - \frac{C}{\mu - \lambda}$, nous avons :

$$C_{Acc}^* = R - \frac{C}{\mu - \lambda}.$$

-Lorsque $R - \frac{C}{\mu - \lambda} \leq C_{Acc} \leq R - \frac{C}{\mu}$, on a :

$$\begin{aligned} R_A(C_{Acc}) &= \lambda q^*(C_{Acc}) C_{Acc}, \\ &= C_{Acc} \left(\mu - \frac{C}{R - C_{Acc}} \right), \\ &= C_{Acc} \mu - \frac{C_{Acc} C}{R - C_{Acc}}. \end{aligned}$$

Cette fonction est C^∞ et nous avons :

$$\begin{aligned} R'_A(C_{Acc}) = 0 &\Leftrightarrow \mu - \frac{C(R - C_{Acc}) + C_{Acc}C}{(R - C_{Acc})^2} = 0 \\ &\Leftrightarrow \mu - \frac{RC}{(R - C_{Acc})^2} = 0 \\ &\Leftrightarrow (R - C_{Acc})^2 = \frac{RC}{\mu} \\ &\Leftrightarrow C_{Acc}^* = R - \sqrt{\frac{RC}{\mu}}. \end{aligned}$$

-Lorsque $C_{Acc} > R - \frac{C}{\mu}$, nous avons : $C_{Acc}^* = 0$.

Ensuite, les coûts d'accès optimaux C_{Acc}^* sont :

$$R - \frac{C}{\mu - \lambda} \text{ ou } R - \sqrt{R \frac{C}{\mu}}.$$

Enfin, le coût optimal C_{Acc}^* est égal à l'argument maximum des deux expressions suivantes :

$$C_{Acc}^* = \operatorname{argmax} \left\{ R_A \left(R - \frac{C}{\mu - \lambda} \right), R_A \left(R - \sqrt{R \frac{C}{\mu}} \right) \right\}.$$

Dans la section suivante, nous étudions l'optimisation des revenus du fournisseur lorsqu'il décide de facturer aux clients, l'information plutôt que l'accès à la file d'attente.

3.4.2 Faire payer l'information sur la taille de la file d'attente

Dans un système où le fournisseur de services facture l'information de la longueur de la file, la variable de contrôle est le coût d'inspection C_I . Le fournisseur

détermine le prix d'information optimal C_I^* qui optimise son revenu. Cette relation est décrite par l'expression suivante :

$$R_I^*(C_I) = \lambda \cdot p^*(C_I) \cdot C_I,$$

où $p^*(C_I)$ est l'équilibre (proportion de clients qui paient pour avoir l'information sur la longueur de la file d'attente) qui dépend des coût d'information.

Pour résoudre ce problème, nous proposons l'algorithme itératif 1 qui se base sur la redéfinition du coût d'inspection à chaque itération, ce qui influence l'équilibre puis la valeur du revenu.

Algorithme 1 Maximisation du revenu R_I^*

Initialisation : $C_I = 0, R_I = 0, R_I^* = 0, \epsilon = 0.001$

Calcul $p^*(C_I)$

$R_I = \lambda p^*(C_I) C_I$

while ($R_I^* < R_I$) **do**

$R_I^* = R_I$

$C_I = C_I + \epsilon$

 Calcul $p^*(C_I)$

$R_I = \lambda p^*(C_I) C_I$

end while

Retourner R_I^*

En raison de la propriété de concavité stricte du revenu, l'algorithme initialise le coût d'inspection à zéro et, étape par étape, il augmente cette valeur jusqu'à ce que le revenu commence à diminuer. À cette étape, l'algorithme retourne la valeur maximale des revenus indiquée par R_I^* .

Sur la base des résultats d'optimisation des deux mécanismes de tarification précédents, le fournisseur peut déterminer le meilleur mécanisme en fonction d'un paramètre important, le coût d'attente C par chaque unité du temps. Ce paramètre décrit la sensibilité des clients à l'attente dans le système.

3.5 Politique de tarification optimale

Nous supposons que le fournisseur souhaite déterminer quel mécanisme de tarification entre la facturation de l'accès et de l'information conduit au meilleur revenu. En particulier, nous nous intéressons à la sensibilité de ce choix en fonction d'un paramètre important qui est le coût d'attente C de clients dans la file d'attente.

Nous montrons d'abord que le revenu optimal décroît en fonction du coût d'attente lorsque le fournisseur facture l'accès à la file d'attente. Ensuite, lorsqu'il facture l'information relative à la longueur de la file, le revenu optimal est égal à zéro lorsque C tend vers zéro et va à l'infini lorsque C tend vers l'infini. En corollaire de ces résultats, nous montrons que lorsque le coût d'attente est faible (c'est-à-dire que les clients ne sont pas très sensibles au temps d'attente), le fournisseur doit facturer l'accès. Par contre, lorsque le coût d'attente est élevé, le fournisseur doit facturer l'information.

Notez que l'estimation du coût de l'attente peut être obtenue par les analyses statistiques des données de l'expérience des clients extraites pour les enquêtes. Ce problème est lié à de nombreux travaux qui ont été réalisés en tenant en compte l'impatience dans les décisions des clients dans des systèmes de file d'attente aussi complexes (voir (D. GROSS et al. 2008)).

Proposition 2. *Lorsque le fournisseur de services facture l'accès à la file d'attente, son revenu optimal diminue strictement avec le coût d'attente.*

Preuve

Lorsque $C_{Acc} \leq R - \frac{C}{\mu - \lambda}$, l'équation de revenu en fonction du coût d'attente s'écrit comme suit :

$$\begin{aligned} R_A(C_{Acc}(C), C) &= \lambda q^*(C_{Acc}^*(C)) C_{Acc}^*(C), \\ &= \left(R - \frac{C}{\mu - \lambda} \right) \lambda, \\ &= R\lambda - \frac{\lambda C}{\mu - \lambda} := \check{R}(C). \end{aligned}$$

La dérivée de cette équation donne :

$$\check{R}'(C) = \frac{-\lambda}{\mu - \lambda} < 0.$$

Lorsque $R - \frac{C}{\mu - \lambda} \leq C_{Acc} \leq R - \frac{C}{\mu}$, nous avons l'équation de revenu suivante en fonction du coût d'attente :

$$\begin{aligned}
 R_A(C_{Acc}(C), C) &= \lambda q^*(C_{Acc}^*(C)) C_{Acc}^*(C), \\
 &= \left(\mu - \frac{C}{\sqrt{\frac{RC}{\mu}}} \right) \left(R - \sqrt{\frac{RC}{\mu}} \right), \\
 &= C - \sqrt{C} \left(R \sqrt{\frac{\mu}{R}} - \mu \sqrt{\frac{R}{\mu}} \right) + \mu R, \\
 &= C - 2\sqrt{\mu R} \sqrt{C} + \mu R := \check{R}(C).
 \end{aligned}$$

La dérivée de cette équation donne :

$$\begin{aligned}
 \check{R}'(C) &= 1 - \frac{\sqrt{\mu R}}{\sqrt{C}}. \\
 \check{R}'(C) < 0 &\Leftrightarrow 1 < \frac{\sqrt{\mu R}}{\sqrt{C}} \Leftrightarrow \sqrt{C} < \sqrt{\mu R}
 \end{aligned}$$

Lorsque,

$$C_{Acc} > 0 \Leftrightarrow R > \sqrt{\frac{RC}{\mu}} \Leftrightarrow \sqrt{R\mu} > \sqrt{C}$$

■

Nous examinons ensuite le comportement du revenu optimal lorsque le fournisseur de services décide de facturer l'information donnée aux clients entrants.

Proposition 3. *Lorsque le fournisseur de services facture l'information, son revenu optimal tend vers zéro lorsque C tend vers zéro et tend vers l'infini lorsque C tend vers l'infini.*

Preuve

Nous avons l'équation de revenu suivante en fonction du coût d'attente :

$$R_I(C_I(C), C) = \lambda p^*(C_I^*(C)) C_I^*(C).$$

Lorsque C tend vers zéro et pour tout C_I , nous avons :

$$\begin{aligned}
 U_I &= \sum_{i=0}^{n_e-1} \pi_i \left(R - C \frac{i+1}{\mu} \right) - C_I \xrightarrow{C \rightarrow 0} R \sum_{i=0}^{n_e-1} (\pi_i) - C_I, \\
 U_{NI} &= \sum_{i=0}^{\infty} \pi_i \left(R - C \frac{i+1}{\mu} \right) \xrightarrow{C \rightarrow 0} R.
 \end{aligned}$$

En conséquence, nous obtenons :

$$R_I(C_I(C), C) \xrightarrow{C \rightarrow 0} 0.$$

parce que,

$$U_{NI} > U_I \implies p^* = 0.$$

Lorsque C tend vers l'infini et pour tout C_I , on a :

$$U_{NI} = U_I + \sum_{i=n_e}^{+\infty} \pi_i (R - C \frac{i+1}{\mu}) + C_I \xrightarrow{C \rightarrow 0} -\infty,$$

$$U_{NI} - U_I < 0 \implies p^* = 1.$$

En conséquence, nous obtenons :

$$R_I(C_I(C), C) \xrightarrow{C \rightarrow +\infty} +\infty.$$

■

Sur la base des deux propositions précédentes, en corollaire, en fonction du coût de l'attente (le niveau de sensibilité des clients au temps d'attente dans la file d'attente), le fournisseur doit proposer un mécanisme de tarification différent afin de maximiser ses revenus. Plus précisément, lorsque cette sensibilité est faible, le fournisseur a plutôt intérêt à facturer l'accès alors que, lorsque la sensibilité est élevée, il préférera facturer les informations.

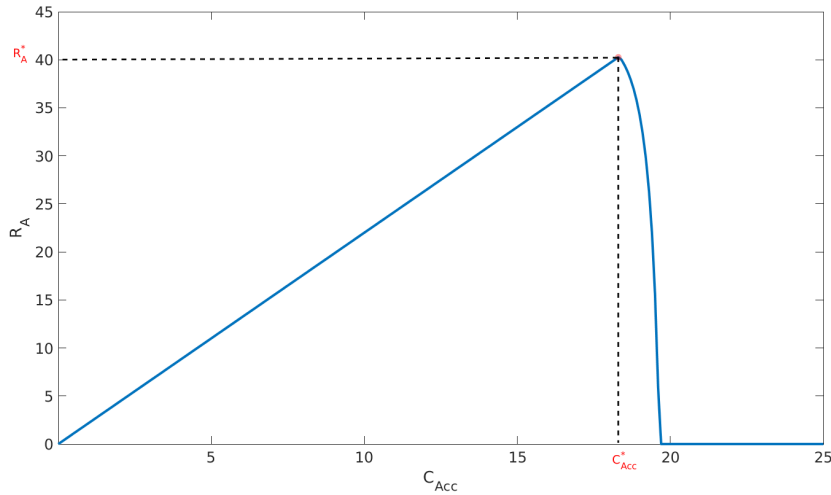
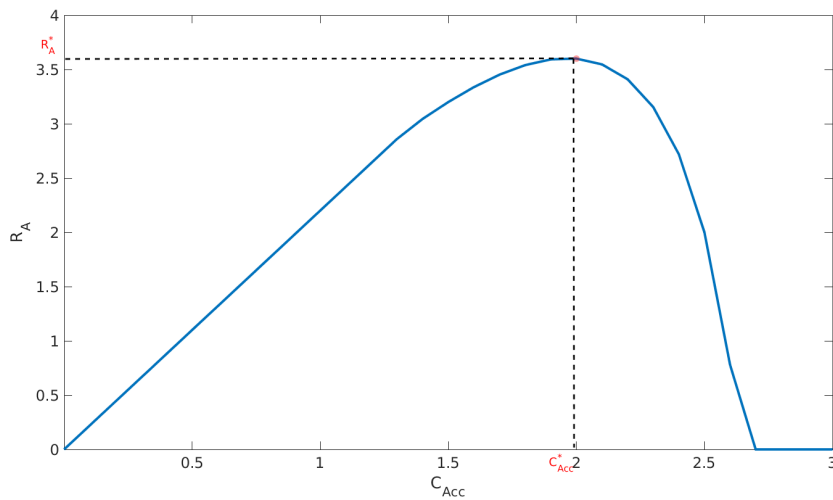
3.6 Illustrations numériques

Dans cette section, nous analysons les deux méthodes de facturation en fonction des paramètres du système lorsque le fournisseur facture l'accès à la file d'attente et l'inspection de l'inspection. Nous illustrons la politique de tarification en fonction du coût d'attente C .

3.6.1 Charger l'accès à la file d'attente

La Figure 3.4 illustre les revenus du fournisseur de services en fonction du coût d'accès.

Dans la Figure 3.4a, nous obtenons que $C_{Acc}^* = (R - \frac{C}{\mu - \lambda}) = 18.33$. Ce revenu maximum est obtenu lorsque tous les clients rejoignent la file d'attente, c'est-à-dire

(a) $\mu = 2.8, \lambda = 2.2, R = 20, C = 1$ (b) $\mu = 2.8, \lambda = 2.2, R = 3, C = 1$ Figure 3.4 – Revenu en fonction du coût d'accès C_{Acc}

$q^* = 1$.

Observation 3. Pour tout ensemble de paramètres, le revenu R_A possède un maximum unique.

Le revenu R_{IA} est une fonction strictement concave en fonction du coût d'inspection. À une valeur du coût d'accès au système C_{Acc}^* , il atteint sa valeur maximale R_A^* .

Dans la Figure 3.4b, nous considérons une valeur de récompense égale $R = 3$. Nous obtenons $C_{Acc}^* = (R - \sqrt{\frac{R \cdot C}{\mu}}) = 1,96$. Dans ce cas, le revenu maximum est obtenu lorsqu'une partie des clients rejoint la file d'attente, c'est-à-dire $0 < q^* < 1$.

3.6.2 Charger l'information

La Figure 3.5 représente l'évolution du revenu du fournisseur de services en fonction du coût d'inspection.

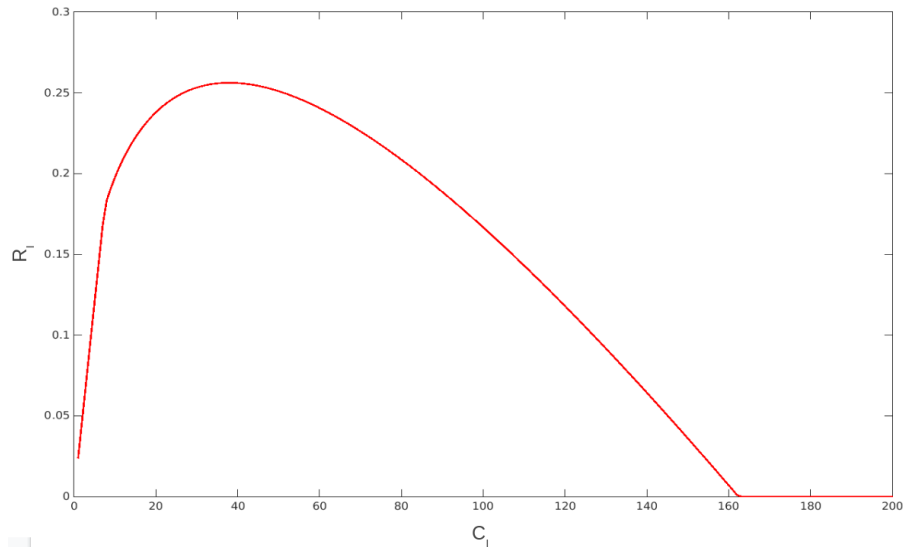


Figure 3.5 – Revenu en fonction du coût d'accès avec $\mu = 2.8$, $\lambda = 2.2$, $R = 3$ et $C = 1$

Observation 4. *Pour tout ensemble de paramètres, le revenu R_I possède un maximum unique.*

Le revenu R_I est une fonction strictement concave en fonction du coût d'inspection. À une valeur du coût d'inspection, il atteint sa valeur maximale R_I^* .

3.6.3 Politique de tarification optimale

La Figure 3.6 donne un exemple numérique du revenu maximum obtenu par le fournisseur de services en fonction du mécanisme de tarification utilisé et en fonction du coût d'attente C . Nous observons que le revenu maximum basé sur les coût d'accès est en diminution monotone : ceci est dû au coût croissant du temps d'attente, ce qui rend les clients peu encouragés à entrer ; ils ne sont peut-être pas disposés à attendre le service en évitant de payer trop cher. Le résultat a été prouvé théoriquement dans la section précédente. Nous observons par ailleurs que le revenu optimal basé sur les coûts d'inspection augmentent en façon monotone, ce qui signifie que les clients sont plus motivés à obtenir l'information de l'état du système à mesure que le coût de l'attente augmente. Cela montre qu'il existe un seuil unique qui détermine quelle politique de tarification est la meilleure.

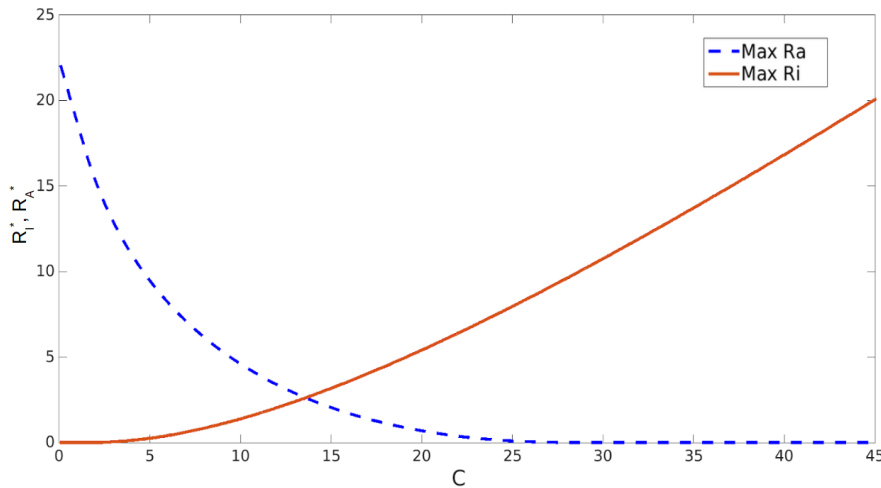


Figure 3.6 – Revenus optimaux en fonction du coût d'attente avec $\mu = 2.8$, $\lambda = 2.2$ et $R = 10$

Ce résultat montre qu'il existe un seuil unique qui détermine quelle politique de tarification est la meilleure pour maximiser le revenu du fournisseur de services.

3.7 Conclusion

Ce chapitre a présenté dans un premier temps, une file d'attente non-observable dans laquelle les clients peuvent obtenir l'information avant de prendre leur décision de rejoindre ou non le système de service. Nous avons prouvé que le système possède un équilibre unique et nous avons donné son expression exacte. Dans un deuxième temps, nous avons comparé deux politiques de tarification : la facturation de l'information sur l'état du système et la facturation de l'accès au système. Nous avons proposé par la suite une politique optimale pour maximiser le revenu du fournisseur de services. Nous avons vu que lorsque les clients sont très impatientes et stratégiques, il est préférable de facturer l'information plutôt que l'accès au service.

Chapitre 4

Politique de partage d'information dans un système de file d'attente stratégique

Sommaire

4.1	Introduction	48
4.2	Description du modèle de file d'attente stratégique avec politique de partage de l'information	48
4.2.1	Modèle	48
4.2.2	Politique de partage d'informations basée sur un seuil fixe	49
4.3	Bien-social en fonction de la PPI	50
4.3.1	Cas complètement observable	51
4.3.2	Cas partiellement observable	52
4.3.3	Comparaison entre la PPI partiellement et totalement observable	60
4.4	Système saturé	63
4.5	Illustrations numériques	67
4.5.1	Bien-être social espéré en fonction de la PPI	68
4.5.2	Bien-être social optimal	69
4.6	Conclusion	70

4.1 Introduction

Les décisions des clients stratégiques suscitent beaucoup d'intérêt dans la conception de systèmes complexes de files d'attente. Des tels systèmes sont généralement étudiés d'un point de vue économique (prix d'accès, valeur de service, capacités, coût de l'information, etc.) afin d'optimiser les résultats globaux obtenus dans le système tel que le temps d'attente, le bien-être social et le revenu. Ces informations aident les clients à décider de rejoindre ou de quitter le système, et d'améliorer leur satisfaction.

Dans ce chapitre, nous combinons les deux modèles fondamentaux de problèmes de files d'attente stratégiques : file d'attente observable et file d'attente non-observable. Nous étudions l'impact d'une politique de partage de l'information sur l'état du système, qui permet à partager l'information sur la taille de la file avec les clients qui arrivent au système jusqu'à un seuil bien défini, sur les performances du système lorsque les clients sont stratégiques et réagissent à cette information.

4.2 Description du modèle de file d'attente stratégique avec politique de partage de l'information

4.2.1 Modèle

Nous considérons une population homogène d'individus (leur récompense et leur coût sont les mêmes) qui décident rationnellement de rejoindre ou non un système de file d'attente. Le renoncement n'est pas possible dans notre système, c'est-à-dire que tout client qui a décidé de rejoindre le système ne peut pas le quitter. La décision des clients est stratégique. Les arrivées suivent un processus de Poisson avec un taux λ . Les clients qui décident de rejoindre la file, attendent le service en suivant une politique de discipline de service du premier arrivé, premier servi (FCFS). Le temps de service suit une distribution exponentielle avec un taux μ . Nous notons par ρ le facteur d'utilisation du système donné par $\rho = \frac{\lambda}{\mu}$ avec l'hypothèse de stabilité $\lambda < \mu$. Tout client obtiendra une récompense R pour le service s'il rejoint la file d'attente. De plus, le coût d'attente par unité de temps passé dans la file d'attente est noté par C . Un nouveau client rejoindra la file d'attente en suivant une règle simple : rejoindre la file d'attente si et seulement si la récompense du service R est supérieure au coût du temps de séjour prévu $C \times W$, où W est la durée de séjour prévue du client dans la file d'attente. En d'autres termes, un client entre si

$R - C \times W$ est non négatif. Cette différence est appelée le bénéfice social net d'un client qui rejoint la file d'attente.

4.2.2 Politique de partage d'informations basée sur un seuil fixe

Le fournisseur de services impose une politique du partage de l'information (PPI) de l'état du système en temps réel, c'est-à-dire le nombre de clients qui sont déjà dans le système (en attente et en service). Sous cette PPI, la longueur de la file d'attente est observable si son état est inférieur ou égal à une valeur du seuil fixée. Le fournisseur contrôle la disponibilité de cette information pour tout nouveau client qui arrive, et en se basant sur l'existence de l'information ou non, chaque client réagit différemment. Lorsque l'information de la longueur de file d'attente est disponible, les clients sont dit des clients informés et la file d'attente est dite observable, sinon ils sont dits des clients non-informés et la file est dite non-observable.

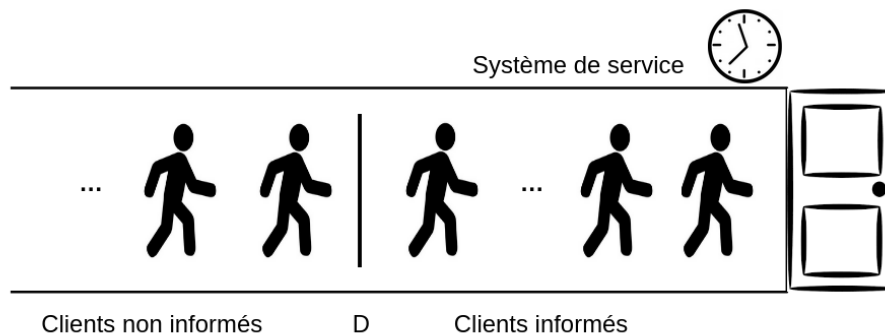


Figure 4.1 – Système de file d'attente avec politique de partage d'information.

Dans le système observable, un nouveau client décidera de rejoindre ou non la file d'attente suivant la règle de seuil du Naor (NAOR 1969) : un client observant i clients dans le système, rejoindra la file d'attente si la différence entre la récompense et le coût associé au temps de séjour prévu est positive. Le seuil de Naor est noté par $n_e := \lfloor \frac{R\mu}{C} \rfloor$ tel qu'un nouveau client rejoindra la file d'attente si et seulement si la taille de la file i est strictement inférieure à n_e . Un client non informé prendra sa décision en fonction du temps de séjour prévu et en supposant qu'il connaît la PPI utilisée par le fournisseur de services.

Le fournisseur de services connaît l'état de la file d'attente à chaque instant et il applique une PPI en se basant sur l'état considéré comme un mécanisme de contrôle dans notre modèle de système. Cette PPI a été proposée dans (E. SIMHON et al. 2015; KIM et al. 2017) afin de maximiser le profit du fournisseur de services. D'autres politiques sont proposées dans d'autres travaux tel que une PPI basé sur un intervalle de

temps dans (Y. DIMITRAKOPOULOS 2019) où l'information de longueur de file d'attente alterne entre observable et non-observable. Dans cette dernière référence, les périodes ne sont pas nécessairement contrôlées. Une PPI peut également dépendre du type de clients dans le cas d'une population hétérogène.

Une telle PPI basée sur l'état du système est définie telle que pour chaque état du système est associée une valeur binaire qui détermine si la longueur de la file d'attente va être partagée avec un nouveau client qui arrive ou non. Notre PPI est déterministe dans le sens où elle correspond à une règle binaire, c'est-à-dire que l'information est donnée ou non. Elle est notée par $u_D(\cdot)$ et paramétrée avec un entier non négatif $D \in \mathbb{N}$ tel que :

$$u_D(i) = \begin{cases} 1 & \text{si } i \leq D, \\ 0 & \text{si } i > D, \end{cases}$$

où i désigne l'état de la file d'attente. Cette politique est stationnaire car elle ne dépend pas du temps et elle est construite de telle sorte que l'information de la longueur de file d'attente n'est révélée que lorsque l'occupation de la file d'attente est inférieure à D . Cela semble intuitif afin d'augmenter le taux d'entrée des clients et par conséquent le revenu du fournisseur de services. Dans ce chapitre, nous étudions le problème de maximisation du bien-être social dans un système de file d'attente qui applique ce type de politique de partage de l'information.

4.3 Bien-social en fonction de la PPI

Le bien-être social du système noté par $SW(D)$, dépend de la PPI u_D utilisée par le fournisseur de services, pour contrôler la disponibilité de l'information selon l'état du système. Nous notons par q_D le vecteur de probabilités qu'un client rejoint le système lorsque i clients sont déjà dans la file d'attente pour tout $i \in \{0, \dots, \infty\}$ si le fournisseur applique la PPI. La distribution stationnaire π_D peut être explicitement calculée en fonction du taux d'arrivée λ , le taux de service μ et la probabilité q_D . Cette distribution stationnaire π_D est donnée par : La distribution stationnaire π_D est donnée par :

$$\pi_D(i) = \begin{cases} \pi_D(0) \left(\frac{\lambda}{\mu}\right)^i & \text{si } i \leq D + 1, \\ \pi_D(0) \frac{\lambda^i q_D^{i-D-1}}{\mu^i} & \text{si } i > D + 1, \end{cases} \quad (4.1)$$

où $\pi_D(i)$ est la probabilité stationnaire dans l'état du système i . En additionnant tous les termes à 1, la probabilité stationnaire inactive (c'est-à-dire la probabilité que la file d'attente soit vide) peut être déterminé en fonction de D et de la probabilité avec laquelle un client non informé décide de rejoindre le système q_D comme suit :

$$\pi_D(0) = \frac{(\lambda - \mu)(\lambda q_D - \mu)}{\lambda^2(-1 + q_D)(\lambda/\mu)^D - \lambda q_D \mu + \mu^2}.$$

Par la suite, leur bien-être social attendu par unité de temps est défini par l'expression suivante :

$$\begin{aligned} SW(D) &= \sum_{i=0}^D \lambda q_D(i) \left(R - \frac{i+1}{\mu} C \right) \pi_D(i) \\ &+ \sum_{i=D+1}^{\infty} \lambda q_D(i) \left(R - \frac{i+1}{\mu} C \right) \pi_D(i), \end{aligned} \quad (4.2)$$

avec $\pi_D(i)$ est la probabilité stationnaire que la longueur de la file d'attente soit égale à i lorsque la PPI u_D est utilisée. Lorsque le seuil D est assez grand, le système équivaut à une file d'attente observable et donc le bien-être social attendu peut-être calculé explicitement et nous pouvons montrer qu'il ne dépend plus du paramètre D .

4.3.1 Cas complètement observable

Le système est entièrement observable lorsque $D \geq n_e - 1$. Dans ce cas, tout client non informé ne rejoindra jamais la file d'attente car son temps de séjour attendu est forcément supérieur au rapport entre sa récompense et son coût ($W > \frac{R}{C}$) (voir Figure 4.2). En fait, les clients non informés connaissent la PPI utilisée par le fournisseur et lorsque $D \geq n_e - 1$ un client non informé sait que la longueur réelle de la file d'attente i est strictement supérieure à $n_e - 1 = \lfloor \frac{\mu R}{C} \rfloor - 1$, et donc son temps de séjour prévu $\frac{i+1}{\mu}$ sera forcément supérieur à $\frac{R}{C}$. Par conséquent, dans ce cas, le système équivaut à une file d'attente de type $M/M/1/n_e$ et nous notons π_e sa distribution stationnaire qui ne dépend pas de D et qui s'exprime par :

$$\forall i = 0, \dots, n_e, \quad \pi_D(i) := \pi_e(i) = \frac{\rho^i}{\sum_{j=0}^{n_e} \rho^j}.$$

Dans ce cas, tous les clients qui rejoignent le système, sont informés et le bien-

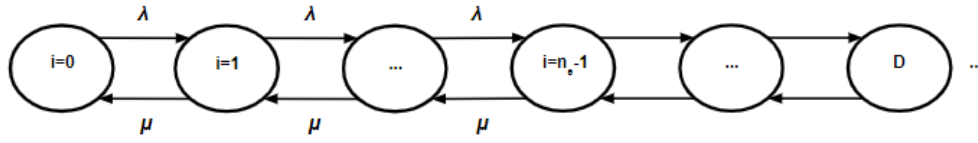


Figure 4.2 – Chaîne de Markov associée au modèle complètement observable

être social attendu $SW(D)$ ne dépend plus de D . Nous le désignons par SW_O et son expression est donnée par :

$$\forall D \geq n_e - 1, \quad SW(D) := SW_O = \lambda R \frac{1 - \rho^{n_e}}{1 - \rho^{n_e+1}} - C \left(\frac{\rho}{1 - \rho} - \frac{(n_e + 1)\rho^{n_e+1}}{1 - \rho^{n_e+1}} \right), \quad (4.3)$$

avec $\rho = \frac{\lambda}{\mu}$ et $n_e = \lfloor \frac{R\mu}{C} \rfloor$.

L'une des principales hypothèses concernant les paramètres du modèle de Naor provient du comportement stratégique tel qu'un nouveau client qui trouve le système vide devrait avoir un intérêt à le rejoindre. En d'autres termes, nous devrions avoir la relation suivante $\frac{R\mu}{C} \geq 1$. Dans la section suivante, les PPI u_D avec $D \in \{0, \dots, n_e - 2\}$ sont considérées. Afin d'avoir des équilibres non triviaux car le système sera complètement observable, nous faisons l'hypothèse ci-dessous qui vient de la condition $n_e - 2 \geq 0$.

Hypothèse 1. Les paramètres système satisfont la relation suivante :

$$\frac{R\mu}{C} \geq 2.$$

4.3.2 Cas partiellement observable

Le système est dit partiellement observable lorsque seulement une partie des nouveaux clients observent la longueur de la file d'attente avant de prendre leur décision de rejoindre ou non la file d'attente. En fait, la politique $u_D(\cdot)$ utilisée est telle que l'information de l'état de la file d'attente est disponible pour un nouveau client si et seulement si la longueur de la file d'attente à l'instant d'arrivée est inférieure ou égale à D . Lorsque $D \in \{0, \dots, n_e - 2\}$, un nouveau client qui arrive peut-être confronté à deux situations différentes :

- la longueur de la file d'attente est observable. Dans ce cas, la longueur réelle

de la file d'attente i est inférieure ou égale à D . Comme $i \leq D < n_e - 1$, ce nouveau client rejoint la file d'attente avec une probabilité $q_D = 1$ et son utilité espérée est donnée par l'expression suivante : $R - \frac{i+1}{\mu}C$,

- la longueur de la file d'attente n'est pas observable. Dans ce cas, la longueur réelle de la file d'attente i est supérieure à D . Un nouveau client rejoint la file d'attente avec une probabilité q_D .

Dans ce dernier cas, la probabilité q_D avec laquelle un nouveau client rejoindra la file d'attente dépend de son temps de séjour attendu noté W_D .

En se basant sur les observations précédentes, le processus aléatoire $X(t)$ qui correspond au nombre de clients dans le système à un instant t est un processus de naissance et de mort avec le taux de natalité λ pour tous les états $i = 0, \dots, D$ et λq_D pour tous les états $i \geq D + 1$. Le taux de mortalité est de μ pour tous les états (voir Figure 4.3).

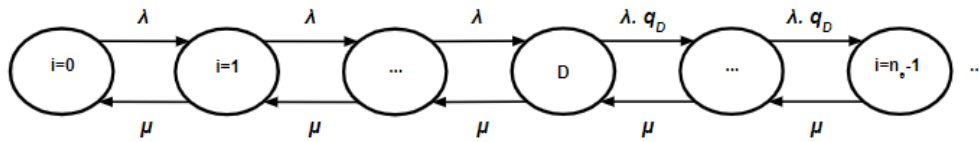


Figure 4.3 – Chaîne de Markov associée au modèle partiellement observable

Proposition 4. Si le fournisseur de services utilise la PPI $u_D(\cdot)$ avec $D \in \{0, \dots, n_e - 2\}$ alors tout client non informé rejoindra la file d'attente avec une probabilité q_D^* définie par :

$$q_D^* = \frac{\mu(R\mu - C(D+2))}{\lambda(R\mu - C(D+1))}. \quad (4.4)$$

Preuve

Lorsque la longueur de la file d'attente est $i \leq D$, le système est dit observable et comme $D < n_e$, nous avons $q_D = 1$ pour tout $i \in \{0, \dots, D\}$. Lorsque la longueur de la file d'attente est $i > D$, le système est dit non-observable et les nouveaux clients ne sont pas informés par l'état du système à l'instant de leur arrivée mais ils savent que la longueur de la file d'attente est supérieure à D et potentiellement inférieure à n_e . Par la suite, ils peuvent avoir intérêt à rejoindre la file d'attente. Nous notons par $\pi_D(i|i > D)$ la probabilité que la longueur de la file d'attente est i étant donné que la longueur de la file d'attente est supérieure à D . En utilisant des règles

de probabilité conditionnelle, nous obtenons la relation suivante :

$$\pi_D(i|i > D) = \frac{\pi_D(i)}{\sum_{j=D+1}^{\infty} \pi_D(j)} \quad (4.5)$$

À l'équilibre, les clients rejoignent la file d'attente avec la probabilité q_D^* tel que le temps de séjour attendu $W^U(D, q_D^*)$ est exactement égal au rapport $\frac{R}{C}$. Il est donné par l'expression suivante :

$$W^U(D, q_D^*) = \sum_{i=D+1}^{\infty} \frac{i+1}{\mu} \pi_D(i|i > D). \quad (4.6)$$

En utilisant l'équation 4.1 et l'équation 4.5, nous pouvons exprimer le temps de séjour attendu :

$$\begin{aligned} W^U(D, q_D^*) &= \sum_{i=D+1}^{\infty} \frac{i+1}{\mu} \frac{\pi_D(i)}{\sum_{j=D+1}^{\infty} \pi_D(j)}, \\ &= \frac{D+1}{\mu} + \frac{1}{\mu - \lambda q_D^*} = \frac{R}{C}. \end{aligned} \quad (4.7)$$

En isolant q_D^* , nous obtenons :

$$q_D^* = \frac{\mu(R\mu - C(D+2))}{\lambda(R\mu - C(D+1))}.$$

■

Nous obtenons donc la valeur exacte de la probabilité, correspondant à une stratégie d'équilibre mixte, avec laquelle un client non informé rejoindra la file d'attente, tout en sachant que l'information n'est révélée que jusqu'à l'état de la file atteint le seuil D .

Lemme 1. La probabilité de décision stratégique à l'équilibre q_D^* diminue en fonction de D lorsque $D \in [0, \infty[$.

Preuve

Nous considérons la version continue de la probabilité de décision stratégique à l'équilibre q_D^* . Sa dérivée est donnée par :

$$q^{*'}(D) = \frac{-\mu C^2}{\lambda(R\mu - C(D+1))^2} < 0$$

La dérivé de q_D^* est strictement négative, ce qui signifie que l'augmentation du seuil D induit une réduction de la proportion de clients non informés qui rejoignent la file d'attente. ■

Ce résultat est assez intuitif parce que d'autant plus que D est proche de la valeur $n_e - 1$, moins les clients non informés ont de la chance de trouver l'état de la file d'attente inférieur à $n_e - 1$ pour la rejoindre.

Sur la base de ce résultat, nous sommes en mesure de déterminer explicitement le bien-être social attendu $SW(D)$ lorsque le fournisseur de services utilise la PPI $u_D(\cdot)$ avec $D \in \{0, \dots, n_e - 2\}$.

Proposition 5. Lorsque le fournisseur de services utilise une PPI $u_D(\cdot)$ avec $D \in \{0, \dots, n_e - 2\}$, le bien-être social attendu $SW(D)$ est donné par l'expression suivante :

$$SW(D) = \lambda \pi_D(0) \left(R \frac{1 - \rho^{D+1}}{1 - \rho} - \frac{C}{\mu} \frac{1 - \rho^{D+2} - (D+2)\rho^{D+1}(1 - \rho)}{(1 - \rho)^2} \right) \quad (4.8)$$

où $\pi_D(0) = \frac{C}{C(\frac{1-\rho^{D+2}}{1-\rho}) + \rho^{D+1}(R\mu - C(D+2))}$ représente la probabilité stationnaire que la file d'attente soit vide.

Preuve

En remplaçant l'équilibre par l'expression obtenue dans la proposition 4, nous obtenons donc :

$$\begin{aligned} \pi_D(0) &= \frac{(\mu - \lambda)C}{\mu C - \lambda \rho^D [\rho(R\mu - C(1+D)) - (R\mu - C(2+D))]} \\ &= \frac{C}{C(\frac{1-\rho^{D+2}}{1-\rho}) + \rho^{D+1}(R\mu - C(D+2))}. \end{aligned} \quad (4.9)$$

Le bien-être social attendu $SW(D)$ est défini dans l'équation (4.2) par :

$$\begin{aligned} SW(D) &= \lambda \sum_{i=0}^D q_D(i) \left(R - \frac{i+1}{\mu} C \right) \pi_D(i) \\ &\quad + \lambda \sum_{i=D+1}^{\infty} q_D(i) \left(R - \frac{i+1}{\mu} C \right) \pi_D(i). \end{aligned}$$

Ensuite, le bien-être social espéré est égal à :

$$\begin{aligned} SW(D) &= \lambda \sum_{i=0}^D \left(R - \frac{i+1}{\mu} C \right) \pi_D(i) + \lambda \sum_{i=D+1}^{\infty} q_D^* \left(R - \frac{i+1}{\mu} C \right) \pi_D(i) \\ &= \lambda R \sum_{i=0}^D \pi_D(i) - \lambda C \sum_{i=0}^D \frac{i+1}{\mu} \pi_D(i) + \lambda q_D^* R \sum_{i=D+1}^{\infty} \pi_D(i) - \lambda q_D^* C \sum_{i=D+1}^{\infty} \frac{i+1}{\mu} \pi_D(i). \end{aligned}$$

Basé sur l'expression du temps de séjour espéré $W^U(D, q_D^*)$ calculée, nous obtenons la relation suivante :

$$\sum_{i=D+1}^{\infty} \frac{i+1}{\mu} \pi_D(i) = \frac{R}{C} \sum_{j=D+1}^{\infty} \pi_D(j).$$

La différence des deux derniers termes dans l'expression du bien-être social espéré est égale à 0. En fait, le bien-être social est égal à zéro à l'équilibre dans une file d'attente non-observable en raison de la propriété d'indifférence de l'équilibre (voir (HASSIN et HAVIV 2003)). Ensuite, le bien-être social espéré $SW(D)$ pour $i \in \{0, \dots, D\}$, est calculé comme suit :

$$\begin{aligned} SW(D) &= \lambda R \pi_D(0) \sum_{i=0}^D \rho^i - \lambda \frac{C}{\mu} \pi_D(0) \sum_{i=0}^D (i+1) \rho^i, \\ &= \lambda \pi_D(0) \left(R \frac{1 - \rho^{D+1}}{1 - \rho} - \frac{C}{\mu} \frac{1 - \rho^{D+2} - (D+2)\rho^{D+1}(1 - \rho)}{(1 - \rho)^2} \right), \end{aligned}$$

avec la probabilité stationnaire dans un système vide donnée par :

$$\pi_D(0) = \frac{C}{C \left(\frac{1 - \rho^{D+2}}{1 - \rho} \right) + \rho^{D+1} (R\mu - C(D+2))}.$$

■

Suite à la proposition précédente, nous avons une expression explicite du bien-être social espéré en fonction de la valeur seuil D de la PPI $u_D(\cdot)$ proposée par le fournisseur de services. Sur la base de cette expression, nous montrons que le bien-être social espéré augmente en fonction du seuil D pour tout $D \in \{0, \dots, n_e - 2\}$. Par la suite, nous obtenons le théorème suivant qui indique que le bien-être social espéré est maximisé pour $D = n_e - 2$.

Théorème 2. *Le bien-être social espéré $SW(D)$ atteint son maximum sur l'intervalle des seuils $D \in \{0, \dots, n_e - 2\}$ lorsque $D = n_e - 2$ pour un taux d'arrivée λ et taux de service μ tel que $\rho = \frac{\lambda}{\mu} \in [0, 1[$.*

Preuve

En se basant sur la proposition 5, pour tout seuil $D \in \{0, \dots, n_e - 2\}$, le bien-être social espéré est donné par :

$$SW(D) = \lambda \pi_D(0) \left(R \frac{1 - \rho^{D+1}}{1 - \rho} - \frac{C}{\mu} \frac{1 - \rho^{D+2} - (D+2)\rho^{D+1}(1 - \rho)}{(1 - \rho)^2} \right),$$

$$\text{avec } \pi_D(0) = \frac{C}{C \left(\frac{1 - \rho^{D+2}}{1 - \rho} \right) + \rho^{D+1}(R\mu - C(D+2))}.$$

Nous notons que cette expression provient du calcul suivant :

$$SW(D) = \lambda R \pi_D(0) \sum_{i=0}^D \rho^i - \lambda \frac{C}{\mu} \pi_D(0) \sum_{i=0}^D (i+1) \rho^i.$$

Définissons la fonction de différence $\Delta(D) = SW(D+1) - SW(D)$ pour tout $D \in \{0, \dots, n_e - 3\}$. Nous notons que cette fonction de différence est donnée par :

$$\begin{aligned} \Delta(D) &= \lambda \sum_{i=0}^{D+1} \left(R - \frac{i+1}{\mu} C \right) \pi_{D+1}(i) - \lambda \sum_{i=0}^D \left(R - \frac{i+1}{\mu} C \right) \pi_D(i) \\ &= \lambda (\pi_{D+1}(0) - \pi_D(0)) \sum_{i=0}^D \left(R - \frac{i+1}{\mu} C \right) \rho^i + \lambda \left(R - \frac{D+2}{\mu} C \right) \pi_{D+1}(D+2). \end{aligned}$$

Premièrement, nous avons $D \leq n_e - 2$ qui implique que $\left(R - \frac{D+2}{\mu} C \right) \geq 0$.

Deuxièmement, $\sum_{i=0}^D \left(R - \frac{i+1}{\mu} C \right) \rho^i \geq 0$ car c'est une somme de termes positifs.

Nous pouvons étudier la différence $(\pi_{D+1}(0) - \pi_D(0))$ pour $D \in \{0, \dots, n_e - 3\}$. Nous notons que la probabilité stationnaire lorsque le système est vide $\pi_D(0)$ est donnée par l'équation (4.9) :

$$\pi_D(0) = \frac{(\mu - \lambda)C}{f(D)},$$

avec

$$\begin{aligned} f(D) &:= \mu C - \lambda \rho^D [\rho(R\mu - C(1+D)) - (R\mu - C(2+D))], \\ &= \mu C + \lambda(1 - \rho)R\mu\rho^D - \lambda(1 - \rho)C(D+1)\rho^D - \lambda C\rho^D. \end{aligned}$$

En considérant la fonction de relaxation continue f définie sur l'intervalle $[0, n_e - 2]$,

nous obtenons la dérivée suivante :

$$\begin{aligned}
 f'(D) &= \lambda(1-\rho)R\mu \log(\rho)\rho^D - \lambda(1-\rho)C\rho^D \\
 &\quad - \lambda(1-\rho)C(D+1) \log(\rho)\rho^D - \lambda C \log(\rho)\rho^D, \\
 &= \lambda(1-\rho) \log(\rho)\rho^D \left(R\mu - \frac{C}{\log(\rho)} - C(D+1) - \frac{C}{1-\rho} \right) \\
 &= \lambda(1-\rho) \log(\rho)\rho^D \left(R\mu - C(D+1) - C \left(\frac{1}{\log(\rho)} + \frac{1}{1-\rho} \right) \right).
 \end{aligned}$$

Nous notons la fonction h suivante définie sur l'intervalle $]0,1[$:

$$h(\rho) = \frac{1}{\log(\rho)} + \frac{1}{1-\rho}.$$

Sa dérivée est donnée par :

$$\begin{aligned}
 h'(\rho) &= \left(\frac{1}{\log(\rho)} + \frac{1}{1-\rho} \right)' = \left(\frac{-1}{(\log \rho)^2 \rho} + \frac{1}{(1-\rho)^2} \right), \\
 &= \frac{\rho(\log(\rho))^2 - (1-\rho)^2}{(\log \rho)^2 \rho (1-\rho)^2}, \\
 &= \frac{[\sqrt{\rho}(\log(\rho)) - (1-\rho)] \times [(\sqrt{\rho}(\log(\rho)) + (1-\rho))]}{(\log \rho)^2 \rho (1-\rho)^2}
 \end{aligned}$$

Nous définissons :

$$\forall \rho \in]0,1[, \quad G(\rho) = \sqrt{\rho}(\log(\rho)) - (1-\rho).$$

Nous avons :

$$\lim_{\rho \rightarrow 0} G(\rho) = -1 \quad \text{and} \quad \lim_{\rho \rightarrow 1} G(\rho) = 0.$$

De plus, pour tout $\rho \in]0,1[$, la dérivée de G est donnée par :

$$G'(\rho) = 1 + \frac{\log(\rho)}{2\sqrt{\rho}} + \frac{\sqrt{\rho}}{\rho} = 1 + \frac{\log(\rho) + 2}{2\sqrt{\rho}} = \frac{2\sqrt{\rho} + \log(\rho) + 2}{2\sqrt{\rho}}.$$

Nous notons les fonctions suivantes sur l'intervalle $]0,1[$:

$$\mathcal{F}(\rho) = 1 + \sqrt{\rho} \quad \text{and} \quad \mathcal{G}(\rho) = -\frac{\log(\rho)}{2}.$$

Alors $G'(\rho) = 0$ si et seulement si $\mathcal{F}(\rho) = \mathcal{G}(\rho)$. De plus, nous avons cela :

$$\forall \rho \in]0,1[, \quad \mathcal{F}'(\rho) = \frac{1}{2\sqrt{\rho}} > 0 \quad \text{and} \quad \mathcal{G}'(\rho) = -\frac{1}{2\rho} < 0.$$

Nous avons également les valeurs extrêmes :

$$\lim_{\rho \rightarrow 0} \mathcal{F}(\rho) = 1 < \lim_{\rho \rightarrow 0} \mathcal{G}(\rho) = +\infty,$$

et

$$\lim_{\rho \rightarrow 1} \mathcal{F}(\rho) = 2 > \lim_{\rho \rightarrow 1} \mathcal{G}(\rho) = 0.$$

Alors, comme les deux fonctions \mathcal{F} et \mathcal{G} sont C^∞ sur l'intervalle $]0,1[$, il existe une valeur $\rho^* \in]0,1[$, tel que $\mathcal{F}(\rho^*) = \mathcal{G}(\rho^*)$ signifiant que $G'(\rho^*) = 0$. En particulier, nous avons :

$$\forall \rho \in]0, \rho^*], \quad \mathcal{G}(\rho) \geq \mathcal{F}(\rho) \iff G'(\rho) \leq 0,$$

et

$$\forall \rho \in [\rho^*, 1[, \quad \mathcal{G}(\rho) \leq \mathcal{F}(\rho) \iff G'(\rho) \geq 0.$$

Ainsi la fonction $G(\rho) := \sqrt{\rho}(\log(\rho)) - (1 - \rho) \leq \lim_{\rho \rightarrow 1} G(\rho) = 0$ pour tout $\rho \in]0,1[$. De plus, pour tout $\rho \in]0,1[$, nous avons :

$$\sqrt{\rho}(\log(\rho)) + (1 - \rho) \geq 0,$$

et

$$(\log \rho)^2 \rho (1 - \rho)^2 \geq 0,$$

alors pour tout $\rho \in]0,1[$:

$$h'(\rho) = \frac{[\sqrt{\rho}(\log(\rho)) - (1 - \rho)] \times [(\sqrt{\rho}(\log(\rho)) + (1 - \rho))]}{(\log \rho)^2 \rho (1 - \rho)^2} \leq 0.$$

Donc, la fonction $h(\rho) := \frac{1}{\log(\rho)} + \frac{1}{1 - \rho}$ est décroissante sur l'intervalle $]0,1[$. Enfin, nous avons la limite suivante :

$$\lim_{\rho \rightarrow 0} h(\rho) = 1$$

et puis comme la fonction h est C^∞ sur l'intervalle $]0,1[$, nous avons :

$$\forall \rho \in]0,1[, \quad C \left(\frac{1}{\log(\rho)} + \frac{1}{1 - \rho} \right) \leq C. \quad (4.10)$$

Enfin, rappelons que $D \leq n_e - 2$ avec $n_e = \lfloor \frac{R\mu}{C} \rfloor$. Donc nous avons :

$$D \leq \lfloor \frac{R\mu}{C} \rfloor - 2 \leq \frac{R\mu}{C} - 2.$$

Enfin,

$$\frac{R\mu}{C} - D \geq 2 \iff R\mu - C(D + 1) \geq C. \quad (4.11)$$

Puis, en combinant les inégalités (4.10) et (4.11) nous obtenons pour tout $D \in [0, n_e - 2]$:

$$R\mu - C(D + 1) \geq C \geq C \left(\frac{1}{\log(\rho)} + \frac{1}{1 - \rho} \right),$$

Donc, nous avons prouvé que pour tout $D \in [0, n_e - 2]$:

$$f'(D) := \lambda(1 - \rho) \log(\rho) \rho^D \left(R\mu - C(D + 1) - \frac{C}{\log(\rho)} - \frac{C}{1 - \rho} \right) \leq 0.$$

Alors, la probabilité stationnaire $\pi_D(0) = \frac{(\mu - \lambda)C}{f(D)}$ augmente avec le seuil D , et donc $(\pi_{D+1}(0) - \pi_D(0)) \geq 0$ pour tout $D \in \{0, \dots, n_e - 3\}$.

Nous obtenons pour tout $D \in \{0, \dots, n_e - 3\}$, $SW(D + 1) \geq SW(D)$. Ainsi le bien-être social espéré atteint sa valeur maximale à $SW(n_e - 2)$ pour une valeur seuil $D = n_e - 2$. ■

Le théorème précédent montre que la meilleure PPI sur l'ensemble $\{u_0, \dots, u_{n_e-2}\}$ est u_{n_e-2} afin de maximiser le bien-être social espéré. Néanmoins, lorsque $D \geq n_e - 1$, nous avons montré que le système est complètement observable et que l'expression du bien-être social espéré ne dépend plus de D et est donnée par l'équation (4.3). Même si nous avons obtenu des expressions explicites du bien-être social espéré $SW(n_e - 2)$ et SW_O , la comparaison analytique n'est pas simple et nous observons numériquement que le maximum entre les deux n'est pas toujours le même. Nous avons un résultat important qui est que la PPI optimale pour un taux d'arrivée λ et un taux de service μ tel que $\lambda < \mu$, est soit u_{n_e-2} soit u_{n_e-1} . Notez que des conditions explicites telles que la PPI u_{n_e-2} donne le bien-être social maximum espéré peuvent être obtenues en considérant un système surchargé, c'est-à-dire lorsque $\lambda = \mu$.

4.3.3 Comparaison entre la PPI partiellement et totalement observable

Dans cette section, nous voulons répondre à la question posée précédemment. Nous allons prouver que dans certaines conditions spécifiques, le bien-être social espéré $SW(n_e - 2)$ (dans le cas partiellement observable) est supérieur au bien-être social espéré $SW_O = SW(n_e - 1)$ (dans le cas complètement observable). Cette condition est prouvée dans le théorème ci-dessous.

Théorème 3. *S'il existe un entier non négatif δ strictement supérieur à 1 tel que $\mu = \frac{\delta C}{R}$, alors :*

$$SW(n_e - 2) \geq SW(n_e - 1).$$

Preuve

Nous considérons qu'il existe un entier non négatif δ strictement supérieur à 1 tel que $\mu = \frac{\delta C}{R}$. Nous avons alors $n_e := \lfloor \frac{R\mu}{C} \rfloor = \delta$. Ainsi, à partir de la preuve de la proposition 1, le bien-être social espéré est donné par :

$$SW(n_e - 2) = SW(\delta - 2) = \lambda R \pi_{\delta-2}(0) \sum_{i=0}^{\delta-2} \rho^i - \frac{\lambda C}{\mu} \pi_{\delta-2}(0) \sum_{i=0}^{\delta-2} (i+1) \rho^i,$$

avec la probabilité stationnaire inactive de l'équation (4.9) donnée par :

$$\pi_{\delta-2}(0) = \frac{C}{C(\frac{1-\rho^\delta}{1-\rho}) + \rho^{\delta-1}(R\mu - C\delta)} = \frac{1-\rho}{1-\rho^\delta}.$$

D'après les résultats de la section 4.3.1 pour un cas complètement observable, lorsque $D \geq n_e - 1 = \delta - 1$ le bien-être social espéré ne dépend pas de D et il est donné par :

$$\begin{aligned} SW(\delta - 1) &= \lambda R \frac{1-\rho^\delta}{1-\rho^{\delta+1}} - C \left(\frac{\rho}{1-\rho} - \frac{(\delta+1)\rho^{\delta+1}}{1-\rho^{\delta+1}} \right), \\ &= \lambda R \frac{1-\rho}{1-\rho^{\delta+1}} \sum_{i=0}^{\delta-1} \rho^i - \frac{\lambda C}{\mu} \frac{1-\rho}{1-\rho^{\delta+1}} \sum_{i=0}^{\delta-1} (i+1) \rho^i. \end{aligned}$$

Ensuite, la différence entre le bien-être social espéré en considérant la PPI u_{n_e-2} et le cas complètement observable ($D \geq n_e - 1$) lorsque $\mu = \frac{\delta C}{R}$ est donnée par :

$$\begin{aligned} SW(\delta - 2) - SW(\delta - 1) &= \lambda R \left(\frac{1-\rho}{1-\rho^\delta} - \frac{1-\rho}{1-\rho^{\delta+1}} \right) \sum_{i=0}^{\delta-2} \rho^i \\ &\quad - \frac{\lambda R}{\delta} \left(\frac{1-\rho}{1-\rho^\delta} - \frac{1-\rho}{1-\rho^{\delta+1}} \right) \sum_{i=0}^{\delta-2} (i+1) \rho^i \\ &\quad - \lambda \rho^{\delta-1} \frac{1-\rho}{1-\rho^{\delta+1}} \left(R - \frac{C\delta}{\mu} \right). \end{aligned}$$

Et comme $\mu = \frac{\delta C}{R}$, le dernier terme vaut zéro et la différence devient :

$$\begin{aligned} SW(\delta - 2) - SW(\delta - 1) &= \lambda R \left(\frac{1 - \rho}{1 - \rho^\delta} - \frac{1 - \rho}{1 - \rho^{\delta+1}} \right) \sum_{i=0}^{\delta-2} \left(\rho^i - \frac{(i+1)}{\delta} \rho^i \right), \\ &= \frac{\lambda R}{\delta} \left(\frac{1 - \rho}{1 - \rho^\delta} - \frac{1 - \rho}{1 - \rho^{\delta+1}} \right) \sum_{i=0}^{\delta-2} \rho^i (\delta - (i+1)) \geq 0. \end{aligned}$$

■

Informations sur l'inactivité du serveur Nous notons par u_0 une PPI particulière intéressante (lorsque $D = 0$). Dans ce cas, les clients ne sont informés que lorsque le serveur est inactif (vide). Sachant que le fournisseur ne donne pas accès à l'information lorsque le serveur est occupé, les clients non informés savent alors que le serveur est occupé mais ils ne connaissent pas le nombre de clients qui attendent dans la file d'attente. Une telle PPI permet de mettre en œuvre une expression explicite du bien-être social espéré $SW(0)$ donnée par :

$$SW(0) = \lambda \frac{(\mu - \lambda)C}{\mu C - \lambda [\rho(R\mu - C) - (R\mu - 2C)]} \left(R - \frac{C}{\mu} \right).$$

Notre approche est différente à celle de (NAOR 1969) où l'auteur impose un frais d'entrée ou celle de (HASSIN et HENIG 1986) où un système avec un service de renonciation et préemptif est supposé. Notre politique de contrôle repose sur une technique de révélation d'information. Cette technique peut entraîner une légère perte de bien-être social optimal lorsque $n^* \neq n_e - 2$ en la comparant aux autres techniques. Nous savons que le bien-être social augmente en fonction de D jusqu'à $n_e - 2$ et le bien-être social avec une PPI qui mène à un système entièrement observable coïncident lorsque $D \geq n_e - 1$. Mais une question intéressante se pose dans ce cas : pourquoi est-il socialement préférable de laisser les clients rejoindre la file d'attente jusqu'à une valeur supérieure à n^* alors que Naor a prouvé que le bien-être social diminue après cette valeur ? La réponse vient de l'ambiguïté de la décision stratégique de clients non informés. En raison de la structure d'information particulière proposée, en diminuant D de $n_e - 2$ à n^* (en supposant que $n^* < n_e - 2$) plus d'ambiguïté est créée pour les clients non informés et cela conduit à une probabilité d'équilibre accrue q_D^* . Ceci est directement observable avec l'expression explicite de la probabilité q_D^* à l'équilibre donnée dans la proposition 4. Ensuite, cela aboutit à davantage de clients indésirables et donc à une diminution du bien-être social.

Le théorème précédent montre que dans des circonstances particulières, la PPI

u_{n_e-2} donne un meilleur bien-être social espéré par rapport au système entièrement observable. Ce résultat est prouvé pour un taux d'arrivée arbitraire λ et un taux de service μ tel que $\lambda < \mu$. Dans le cas particulier où $\lambda = \mu$, qui est bien connu dans la théorie des files d'attente comme un système surchargé (KLEINROCK 1975), le même type de comparaison entre un système partiellement observable et système entièrement observable peut être effectué et bien plus de détails peuvent être donnés. Une analyse approfondie est décrite dans la section suivante.

4.4 Système saturé

Dans la théorie des files d'attente, une file d'attente est dite surchargée ou saturé si le taux d'arrivée λ des clients est supérieur ou égal au taux de service μ .

Hypothèse 2. *Le taux d'arrivée λ et le taux de service μ sont égaux tels que $\rho = \frac{\lambda}{\mu} = 1$.*

1. *Compte tenu de cette contrainte, le système est dit surchargé.*

En supposant que ces deux débits sont égaux, et donc $\rho = 1$, alors on peut parler d'une file d'attente surchargée. Dans ce cas, le bien-être social espéré \overline{SW}_O lorsque le fournisseur de services utilise la PPI u_D avec $D \geq n_e - 1$ est relativement simple à calculer car les probabilités stationnaires sont simplement $\bar{\pi}_e(i) = \frac{1}{n_e + 1}$ pour tous les états $i \in \{0, \dots, n_e\}$. Notant que pour le reste du chapitre, la notation sur-lignée signifie que c'est dans le contexte d'un système surchargé. Le résultat de Naor (Equation (4.3)) pour un cas entièrement observable peut également être proposé pour un système surchargé. Dans ce contexte, lorsque $\rho = 1$, on obtient :

$$\overline{SW}_O = \frac{n_e}{2(n_e + 1)} [2R\mu - C(n_e + 1)]. \quad (4.12)$$

Cela correspond au bien-être social espéré pour tous les PPI $u_D(\cdot)$ telles que $D \geq n_e - 1$ qui correspondent aux systèmes entièrement observables.

Lorsque $D \leq n_e - 2$, la proposition suivante montre que le bien-être social espéré $\overline{SW}(D)$ peut également être exprimé par une expression plus simple que dans le cas général.

Proposition 6. Dans un système surchargé, c'est-à-dire $\rho = 1$, pour tout $D \in \{0, \dots, n_e - 2\}$, le bien-être social espéré $\overline{SW}(D)$ est donné par :

$$\overline{SW}(D) = \frac{C(D+1)}{2R\mu} [2R\mu - C(D+2)]. \quad (4.13)$$

Preuve

Quand $\rho = 1$ et considérant la PPI u_D avec $D \in \{0, \dots, n_e - 2\}$, le vecteur de distribution stationnaire $\bar{\pi}_D$ est donné par :

$$\bar{\pi}_D(i) = \begin{cases} \bar{\pi}_D(0) & \text{si } i \leq D+1, \\ \bar{\pi}_D(0)q^{i-D-1} & \text{si } i > D+1, \end{cases}$$

avec q est la probabilité qu'un client non informé rejoigne la file d'attente. Et comme $\sum_{i=0}^{\infty} \bar{\pi}_D(i) = 1$, la probabilité stationnaire $\bar{\pi}_D(0)$ pour l'état vide est :

$$\begin{aligned} \bar{\pi}_D(0) &= \left(\sum_{i=0}^{D+1} 1 + \sum_{i=D+2}^{\infty} q^{i-D-1} \right)^{-1}, \\ &= \left(\sum_{i=0}^D 1 + \sum_{i=0}^{\infty} q^i \right)^{-1} = \left((D+1) + \frac{1}{1-q} \right)^{-1}, \\ &= \frac{1-q}{(D+1)(1-q) + 1}. \end{aligned}$$

Par la suite, le bien-être social espéré devient :

$$\begin{aligned} \overline{SW}(D) &= \lambda R \sum_{i=0}^D \bar{\pi}_D(i) - \lambda C \sum_{i=0}^D \frac{i+1}{\mu} \bar{\pi}_D(i), \\ &= \lambda R \sum_{i=0}^D \bar{\pi}_D(0) - \lambda C \sum_{i=0}^D \frac{i+1}{\mu} \bar{\pi}_D(0), \\ &= \lambda R \bar{\pi}_D(0)(D+1) - \lambda C \bar{\pi}_D(0) \sum_{i=0}^D \frac{i+1}{\mu}, \\ &= \frac{\lambda R(1-q)(D+1)}{(D+1)(1-q) + 1} - \frac{\lambda C(1-q)}{(D+1)(1-q) + 1} \frac{(D+1)(D+2)}{2\mu}, \\ &= \frac{\lambda(D+1)(1-q)}{2((D+1)(1-q) + 1)} \left[2R - \frac{C(D+2)}{\mu} \right], \\ &= \frac{(D+1)}{2((D+1) + \frac{1}{1-q})} [2R\mu - C(D+2)] \end{aligned}$$

À l'équilibre, basé sur la proposition 4, chaque nouveau client non informé rejoint la file d'attente avec probabilité $q_D^* = \frac{R\mu - (D+2)C}{R\mu - (D+1)C}$, ce qui donne

$$\frac{1}{1 - q_D^*} = \frac{R\mu - (D+1)C}{C}.$$

En remplaçant cette expression dans $\overline{SW}(D)$, nous obtenons :

$$\forall D \in \{0, \dots, n_e - 2\}, \quad \overline{SW}(D) = \frac{C(D+1)}{2R\mu} [2R\mu - C(D+2)].$$

■

Ensuite, nous avons le corollaire suivant qui provient directement du résultat précédent.

Corollaire 2. Dans un système surchargé, le bien-être social espéré $\overline{SW}(D)$ est strictement croissant en fonction de $D \in \{0, \dots, n_e - 3\}$.

Preuve

Nous considérons une valeur $D \in \{0, \dots, n_e - 3\}$, nous calculons la différence suivante : $\Delta(D) := \overline{SW}(D+1) - \overline{SW}(D)$:

$$\begin{aligned} \Delta(D) &= \frac{C(D+2)}{2R\mu} [2R\mu - C(D+3)] - \frac{C(D+1)}{2R\mu} [2R\mu - C(D+2)], \\ &= \frac{C}{2R\mu} [(D+2)2R\mu - C(D+2)(D+3) - (D+1)2R\mu + C(D+1)(D+2)], \\ &= \frac{C}{2R\mu} [2R\mu(D+2 - D - 1) + C(D+2)(D+1 - D - 3)], \\ &= \frac{C}{R\mu} [R\mu - C(D+2)]. \end{aligned}$$

Notant que $D < n_e - 2$, ensuite nous avons :

$$R\mu - C(D+2) > R\mu - n_e C$$

et par définition :

$$n_e = \lfloor \frac{R\mu}{C} \rfloor \leq \frac{R\mu}{C}.$$

Ensuite, $R\mu - n_e C \geq 0$ et $\Delta > 0$.

Pour conclure, lorsque le système est surchargé, c'est-à-dire $\rho = 1$, pour chaque

$D \in \{0, \dots, n_e - 2\}$, $\overline{SW}(D)$ est strictement croissant en fonction du paramètre D . ■

Ce résultat peut être montré d'une manière différente en considérant une relaxation continue du problème. En considérant que la fonction $\overline{SW}(\cdot)$ est continue et D est un nombre réel non négatif, nous pouvons montrer que cette fonction atteint sa valeur maximale à D^* qui est supérieur à $n_e - 2$.

Cette dernière proposition est en accord avec la remarque intuitive qui dit que fournir plus d'informations est bénéfique pour les clients, et donc pour le bien-être social du système (voir l'introduction du chapitre 3 dans (HASSIN 2016) lié à informations dans la file d'attente stratégique).

Dans ce qui suit, nous prouvons que pour certaines conditions spécifiques, le bien-être social espéré peut-être plus élevé en révélant l'information de longueur de file d'attente jusqu'à une longueur de file d'attente particulière au lieu de toujours fournir cette information.

Sur la base de la proposition précédente, on peut déduire que, lorsque le système est surchargé, le bien-être social espéré est maximisé en appliquant la politique u_{n_e-2} ou en donnant des informations complètes. La proposition suivante donne une preuve formelle que, dans certaines conditions spécifiques, il n'est pas toujours préférable de donner une information complète aux nouveaux clients afin de maximiser le bien-être social.

Théorème 4. *Dans un système surchargé, c'est-à-dire $\rho = 1$, si les paramètres système R , μ et C sont tels que :*

$$\lfloor \frac{R\mu}{C} \rfloor - \frac{1}{\lfloor \frac{R\mu}{C} \rfloor} > \frac{R\mu}{C} \left(1 - \frac{1}{2\frac{R\mu}{C} - \lfloor \frac{R\mu}{C} \rfloor} \right), \quad (4.14)$$

alors il est socialement préférable pour le système que le fournisseur de services révèle les informations jusqu'à l'état $n_e - 2$ au lieu de toujours donner l'information de la longueur de la file d'attente aux nouveaux clients entrants.

Preuve

Nous considérons la PPI particulière u_{n_e-2} qui signifie que l'information de la longueur de file d'attente est donnée aux nouveaux clients jusqu'à la longueur de file d'attente $n_e - 2$. Dans ce cas et dans un contexte du système surchargé, basé

sur l'équation (4.14), le bien-être social espéré est donné par :

$$\overline{SW}(n_e - 2) = \frac{C(n_e - 1)}{2R\mu} [2R\mu - Cn_e].$$

Ensuite, $\overline{SW}(n_e - 2)$ est supérieur à \overline{SW}_O qui correspond au bien-être social espéré lorsque le fournisseur de services révèle toujours l'information, si et seulement si :

$$\begin{aligned} & \overline{SW}(n_e - 2) > \overline{SW}_O \\ \Leftrightarrow & \frac{C(n_e - 1)}{R\mu} [2R\mu - Cn_e] > \frac{n_e}{(n_e + 1)} [2R\mu - C(n_e + 1)], \\ \Leftrightarrow & \frac{C(n_e - 1)(n_e + 1)}{R\mu n_e} > \frac{2R\mu - C(n_e + 1)}{2R\mu - Cn_e}, \\ \Leftrightarrow & \frac{C}{R\mu} \left(n_e - \frac{1}{n_e} \right) > 1 - \frac{C}{2R\mu - Cn_e}, \\ \Leftrightarrow & n_e - \frac{1}{n_e} > \frac{R\mu}{C} \left(1 - \frac{1}{2\frac{R\mu}{C} - n_e} \right). \end{aligned}$$

■

Ce théorème donne une condition pour laquelle la PPI u_{n_e-2} est meilleure qu'un système surchargé complètement observable en termes de bien-être social espéré. Des paramètres spécifiques vérifient cette condition. Par exemple, en considérant l'exemple de base de $\frac{R\mu}{C} = \delta$ où δ est un entier supérieur à 2. Ensuite, nous obtenons le seuil $n_e = \delta$. Par conséquent, la partie gauche de l'inégalité (4.14) est égale à $\delta - \frac{1}{\delta}$ et la partie droite est égale à $\delta - 1$. Mais comme $\delta \geq 2 > 1$, le côté gauche est strictement plus grand que le côté droit et la condition (4.14) est satisfaite. L'analyse précédente du bien-être social espéré a montré qu'un choix judicieux de la personne déplacée conduit à un bien-être social maximum espéré pour le système, et donc très bénéfique pour les clients.

4.5 Illustrations numériques

Cette section illustre dans un premier lieu les résultats du bien-être social espéré en fonction de la PPI proposée par le fournisseur de services. Dans un deuxième temps, l'accent est mis sur l'optimisation du bien-être social espéré. Enfin, des investigations numériques sont effectuées pour le contexte particulier d'un système surchargé. Dans toutes les expériences numériques faites, nous avons considéré un coût d'attente par unité de temps donné par $C = 3$.

4.5.1 Bien-être social espéré en fonction de la PPI

Tout d'abord, nous illustrons le comportement du bien-être social espéré SW du système en considérant une PPI avec des valeurs de seuils différentes. Nous fixons la récompense $R = 20$ et le taux du service est $\mu = 2$. Compte tenu de ce choix de paramètres du système, le seuil de Naor est $n_e = 13$ et la condition (4.14) est satisfaite. Par conséquent, chaque fois que $D \geq 12$ le bien-être social espéré ne dépend plus de D comme illustré sur la figure 4.4. Sur cette figure, nous illustrons également l'impact du taux d'arrivée λ sur le bien-être social espéré. Nous avons montré dans la section 4.4 du système surchargé (lorsque $\lambda = \mu$), la PPI u_{n_e-2} a un meilleur bien-être social espéré par rapport à une information complète si la condition définie dans le théorème 4 est satisfaite. Les illustrations numériques corroborent ce résultat. Notant que les PPI u_{10} et u_9 rapportent également un bien-être social espéré plus élevé lorsque le système est surchargé. Au contraire, lorsque λ est inférieur à 1.6, le pic n'est pas aussi clair et encore plus pour $\lambda = 1$, il n'y a pas de pic. Dans ce cas, informer les clients est la meilleure politique pour le bien-être social espéré. Enfin, une dernière observation est que le bien-être social espéré est strictement croissant en fonction de $D \in \{0, \dots, n_e - 2\}$ a été prouvé en corollaire 2.

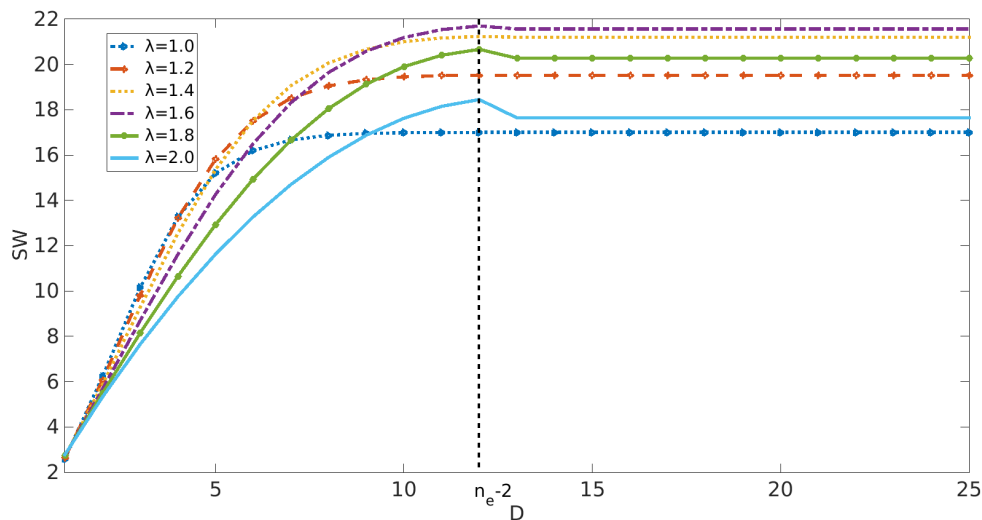


Figure 4.4 – Bien-être social espéré SW en fonction de la PPI u_D pour des taux d'arrivée différents $\lambda \in \{1, 1.2, 1.4, 1.6, 1.8, 2\}$.

Dans ces cas, en fournissant plus d'informations, dans le sens qu'en augmentant le seuil d'information sur la longueur maximale de la file d'attente D , le bien-être social espéré est amélioré. Le fournisseur a intérêt à fixer D à $n_e - 2$ pour obtenir la valeur optimale du bien-être social dans le système.

4.5.2 Bien-être social optimal

L'optimisation du bien-être social espéré en ajustant de manière appropriée les paramètres du système est une question importante. Le taux d'arrivée λ est un paramètre du système intéressant qui peut être contrôlé. Sur la base des observations précédentes, la PPI u_{n_e-2} peut améliorer le bien-être social espéré, en particulier lorsque la file d'attente proche du régime surchargé. Nous illustrons dans la figure 4.5 le bien-être social en fonction du taux d'arrivée λ dans l'intervalle $[1,2]$, considérant le taux de service $\mu = 2$. Sur la base de la figure précédente 4.4, le bien-être social espéré est unimodal avec un taux d'arrivée λ . Une autre observation de la figure 4.5 qui est assez intuitive est que le maximum du bien-être social espéré augmente en fonction de la récompense R . Nous notons également que la valeur du taux d'arrivée optimal λ^* change pour chaque valeur de récompense R .

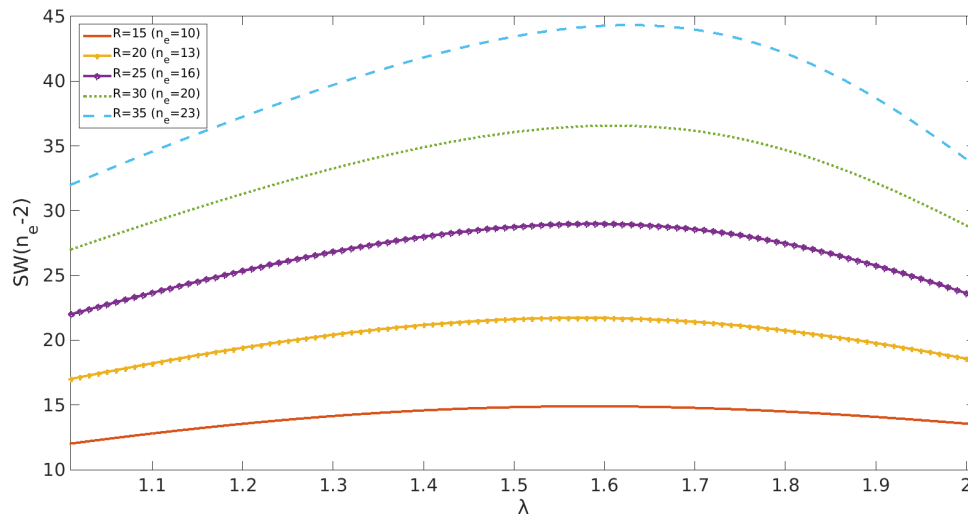


Figure 4.5 – Bien-être social espéré $SW(n_e - 2)$ en fonction du taux d'arrivée λ pour différentes valeurs de récompense R .

Le résultat le plus intéressant est que pour chaque valeur de R , comme le bien-être social espéré est strictement concave, il existe un taux d'arrivée optimal unique. Nous le désignons par λ^* . Sur la base d'une remarque précédente, un mécanisme de contrôle d'admission peut être proposé par le fournisseur de services : lorsque $\lambda > \lambda^*$, le fournisseur de services peut réduire le flux entrant de clients jusqu'à atteindre λ^* en refusant les clients ou en imposant un prix d'accès. Au contraire, lorsque $\lambda < \lambda^*$ le fournisseur de services améliorera le bien-être social espéré du système en augmentant la congestion (comme le paradoxe de Braess (B. CALVERT et al. 1997)) en ajoutant une fausse demande λ_0 dans le système. Un résultat similaire peut être obtenu en réduisant le tarif du service μ mais pas de manière aussi simple car μ a un impact sur la valeur seuil n_e .

Enfin la condition nécessaire donnée par le théorème 4 pour un système surchargé est illustrée sur la dernière figure 4.6 avec deux valeurs différentes du taux service (une valeur est juste sous un entier et l'autre juste au-dessus). Comme cela a été montré dans le théorème 4, lorsque la condition (4.14) n'est pas satisfaite, le bien-être social espéré est meilleur en fournissant toujours des informations sur la longueur de la file d'attente. Alors que lorsque $\mu = 8.1$, la condition (4.14) est satisfaite et par la suite fournir des informations sur la longueur de la file d'attente jusqu'à l'état 6 implique un meilleur bien-être social espéré.

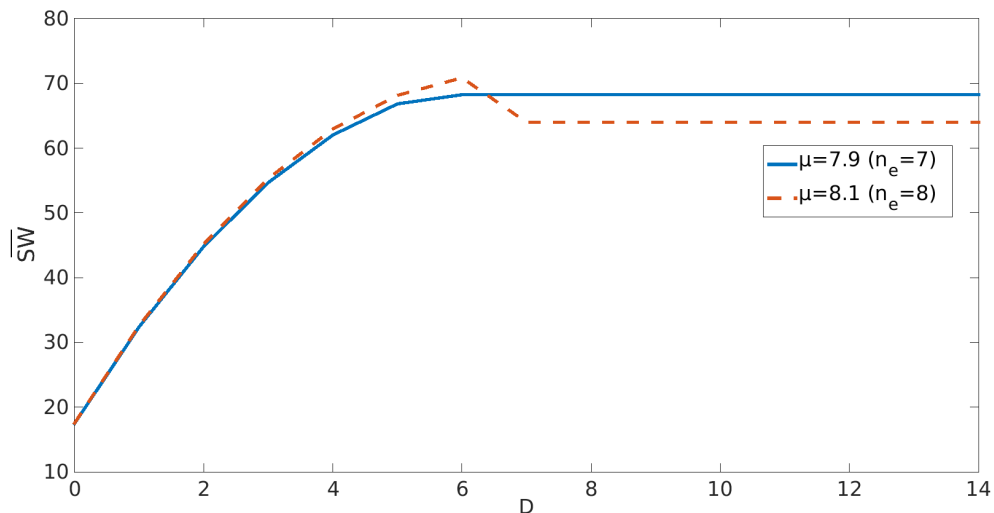


Figure 4.6 – Bien-être social espéré en régime surchargé \overline{SW} en fonction de D avec $R = C = 20$.

4.6 Conclusion

Dans ce chapitre, nous avons vu que les informations sur la longueur de la file d'attente, fournies aux clients ont un impact important sur leur décision de rejoindre les systèmes de file d'attente stratégiques. Des politiques déterministes de partage d'informations (PPI) basées sur l'état du système ont été prises en compte dans ce travail pour contrôler le comportement des clients et leur impact sur le bien-être social espéré est étudié. Le principal résultat de ce chapitre est la preuve que l'application d'une PPI qui révèle les informations de longueur de file d'attente jusqu'à ce qu'un seuil particulier maximise le bien-être social espéré pour des taux d'arrivée et de service arbitraires. De plus, un paradoxe est montré, affirmant que le bien-être social espéré est plus élevé compte tenu de ce type de structure d'information par rapport au cas d'une information complète. Enfin, des scénarios numériques ont également montré que des mécanismes de contrôle des admissions peuvent être proposés afin de maximiser le bien-être social espéré.

Chapitre 5

Politiques de partage d'information dans un système de file d'attente stratégique hétérogène

Sommaire

5.1	Introduction	72
5.2	Description du modèle de file d'attente hétérogène avec politique de partage d'information	72
5.2.1	Modèle	72
5.2.2	Politique de partage d'informations basée sur un seuil fixe	73
5.3	Système de file d'attente avec deux classes de clients	74
5.3.1	Cas complètement observable	74
5.3.2	Cas partiellement observable	77
5.4	Illustrations numériques	84
5.4.1	Calcul de l'équilibre	84
5.4.2	Nombre moyen de clients	84
5.4.3	Temps moyen de séjour	86
5.5	Conclusion	87

5.1 Introduction

Dans notre vie quotidienne, les clients qui se présentent aux systèmes de file d'attente sont généralement hétérogènes. Nous citons l'exemple des files d'attente dans les supermarchés qui contiennent des clients classiques et des clients fidèles et les files d'attente à l'aéroport où il existe des voyageurs de la classe économique et d'autres de la classe business. Aussi, dans le péage d'autoroute, plusieurs types de véhicules peuvent passer, tels que les véhicules légers, les véhicules intermédiaires, les poids lourds, les deux-roues, etc. Cette hétérogénéité est prise en compte à travers des coûts de service différents pour chaque type de client.

Dans ce chapitre, nous modélisons le même problème étudié dans le chapitre précédent du système de file d'attente stratégique avec une politique de partage d'information, mais en supposant que les clients sont hétérogènes dans la file d'attente. Nous considérons plusieurs classes de clients dans le système et nous étudions l'impact de la PPI dans une telle file d'attente sur l'équilibre du système.

5.2 Description du modèle de file d'attente hétérogène avec politique de partage d'information

5.2.1 Modèle

Nous considérons une population hétérogène d'individus, qui contient J classes des clients, qui se présente à une file d'attente avec un serveur unique de type $M/M/1$. Dans ce système, tout type de client qui a décidé de rejoindre le système ne peut pas le quitter et sa décision, de rejoindre la file ou de la quitter sans être servi, est stratégique. Les arrivées suivent un processus de Poisson avec un taux λ_j pour chaque classe j de clients, avec $j \in \{1, \dots, J\}$. Les clients qui décident de rejoindre la file, sont servis selon la politique de discipline du premier arrivé, premier servi (FCFS). Le temps de service suit une distribution exponentielle avec le même taux μ pour tous les types de clients. Nous notons, pour tout $j \in \{1, \dots, J\}$, par ρ_j le facteur d'utilisation du système donné par : $\rho_j = \frac{\lambda_j}{\mu}$, associé pour chaque classe de client j et sachant que nous faisons l'hypothèse que $\lambda_j < \mu$ pour tout j . Tout type de client obtiendra une récompense R strictement positif pour le service s'il rejoint la file d'attente. De plus, le coût d'attente par unité de temps passé dans un tel système de file d'attente pour un client de type j est noté par C_j . Pour simplifier le modèle étudié, nous supposons que les coûts d'attente associés pour chaque j classe sont

croissantes, ce qui signifie que les clients de classe 1 sont les moins sensibles au temps d'attente et les clients de classe J sont les plus sensibles à l'attente. Cette relation est décrite comme suit :

$$C_1 < C_2 < C_3 < \dots < C_J. \quad (5.1)$$

Un nouveau client rejoindra la file d'attente si et seulement si la récompense du service est supérieure au coût du temps de séjour prévu.

5.2.2 Politique de partage d'informations basée sur un seuil fixe

Nous appliquons la même politique de partage d'information (PPI) utilisée dans le modèle étudié dans le chapitre précédent. Précédemment, nous avons supposé que sous une PPI, l'état du système est partagé avec les clients qui arrivent au système, si son état est inférieur ou égal à une valeur du seuil fixée par le fournisseur de services. Par la suite, dans notre modèle de multi-classes, un nouveau client qui appartient à une classe quelconque et qui arrive au système sera face à deux types de systèmes possibles : observable et non-observable.

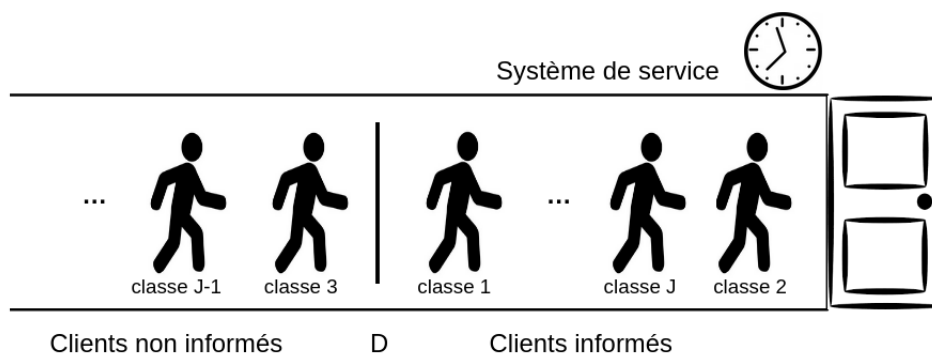


Figure 5.1 – Système de file d'attente hétérogène en appliquant une PPI

Un client de type j informé par l'état du système, décidera de rejoindre ou non la file d'attente en appliquant la règle de seuil du Naor (NAOR 1969) avec le coût d'attente correspondant à sa classe. Par la suite, un seuil de Naor est associé pour chaque type j de client dans ce système et il est donné par l'expression suivante :

$$n_j = \lfloor \frac{\mu R}{C_j} \rfloor,$$

tel qu'un nouveau client de classe j rejoindra la file d'attente si et seulement si la taille de la file i à son arrivée est strictement inférieure à n_j . En se basant sur la relation 5.1, les seuils de Naor associé à chaque classe j sont décroissants en fonction

du type j de classe et ils sont donnés par la relation suivante :

$$n_1 > n_2 > n_3 > \dots > n_J. \quad (5.2)$$

Un client de classe j non informé prendra sa décision en fonction du son temps de séjour prévu W_j et en supposant qu'il connaît la PPI utilisée par le fournisseur de services.

La PPI est définie telle que pour chaque état du système est associée une valeur binaire qui détermine si la longueur de la file d'attente va être partagée avec un nouveau j client qui arrive ou non. Elle est notée par $u_D(\cdot)$ et paramétrée avec un entier non négatif $D \in \mathbb{Z}$ tel que :

$$u_D(i) = \begin{cases} 1 & \text{si } i \leq D, \\ 0 & \text{si } i > D, \end{cases}$$

où i désigne le nombre des clients qui existent dans la file d'attente.

Pour la simplification du calcul et la lisibilité des expressions mathématiques, nous traitons le modèle avec deux classes de clients. Ce cas peut être par la suite généralisé en se basant sur les résultats trouvés.

5.3 Système de file d'attente avec deux classes de clients

Dans le reste du chapitre, nous étudions le cas où deux classes de clients, i.e. $J = 2$, se présentent à un système de file d'attente avec une PPI telle que définie dans la section précédente.

5.3.1 Cas complètement observable

Pour le modèle où deux classes de clients peuvent exister, le système est entièrement observable lorsque $D \geq n_1 - 1$. Dans ce cas, tout client de type j non informé ne rejoindra jamais la file d'attente car son temps de séjour attendu est forcément supérieur au rapport entre sa récompense et son coût ($W_D > \frac{R}{C_j}$). Par conséquent, dans le cas où le seuil D de la PPI est supérieur à $n_1 - 1$, le système équivaut à une file d'attente de type $M/M/1/n_1$. Nous considérerons l'hypothèse ci-dessous.

Hypothèse 3. Les paramètres du système satisfont la relation suivante :

$$\frac{R\mu}{C_j} \geq 2, \text{ pour tout } j \in \{1, \dots, J\}.$$

La figure 5.2 représente l'évolution de système, dans le cas complètement observable, par une chaîne de Markov. Les clients de classe 1 et 2, rejoignent le système avec un taux $\lambda_1 + \lambda_2$ jusqu'à atteindre le seuil de Naor n_2 pour la classe 2 (sachant que $n_2 < n_1$). À partir de cette valeur, les clients de classe 2, ne vont plus rejoindre le système (parce que leur utilité espérée devient négative). Par la suite, seulement les clients de classe 1 vont rejoindre le système avec un taux λ_1 .

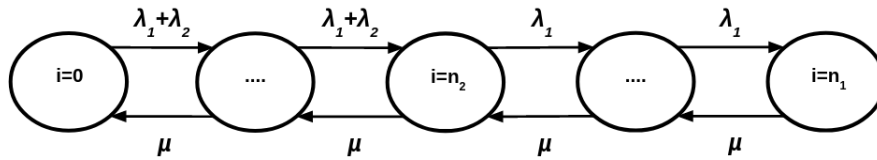


Figure 5.2 – Chaîne de Markov associée au modèle complètement observable

Nous notons par π_i sa distribution stationnaire de la chaîne de Markov décrite dans la figure 5.2. Elle satisfait la propriété suivante :

$$\forall i \in \{0, \dots, n_1 - 1\}, \quad \pi_i = \begin{cases} (\rho_1 + \rho_2)^i \pi_0 & \text{si } i \leq n_2 \\ (\rho_1 + \rho_2)^{n_2} \rho_1^{i-n_2} \pi_0 & \text{si } n_2 < i \leq n_1 \end{cases} \quad (5.3)$$

Proposition 7. La probabilité stationnaire d'avoir aucun client dans le système est donnée par :

$$\pi_0 = \frac{(1 - (\rho_1 + \rho_2))(1 - \rho_1)}{1 - \rho_1 - (\rho_1 + \rho_2)^{n_2+1}(1 - \rho_1^{n_1-n_2+1}) + (\rho_1 + \rho_2)^{n_2} \rho_1 (1 - \rho_1^{n_1-n_2})}$$

Preuve

En se basant sur l'équation 5.3, nous additionnons tous les termes à 1, la probabilité stationnaire lorsque la file d'attente est vide peut être calculée comme suit :

$$\pi_0 \left[\sum_{i=0}^{n_2} (\rho_1 + \rho_2)^i + \sum_{i=n_2+1}^{n_1} (\rho_1 + \rho_2)^{n_2} \rho_1^{i-n_2} \right] = 1$$

Les sommes sont calculées comme suit :

$$\sum_{i=0}^{n_2} (\rho_1 + \rho_2)^i = \frac{1 - (\rho_1 + \rho_2)^{n_2+1}}{1 - (\rho_1 + \rho_2)}$$

$$\sum_{i=n_2+1}^{n_1} (\rho_1 + \rho_2)^{n_2} \rho_1^{i-n_2} = (\rho_1 + \rho_2)^{n_2} \rho_1 \sum_{i=0}^{n_1-n_2-1} \rho_1^i$$

$$= (\rho_1 + \rho_2)^{n_2} \rho_1 \frac{1 - \rho_1^{n_1-n_2}}{1 - \rho_1}$$

Nous obtenons alors :

$$\pi_0 \left[\frac{1 - (\rho_1 + \rho_2)^{n_2+1}}{1 - (\rho_1 + \rho_2)} + (\rho_1 + \rho_2)^{n_2} \rho_1 \frac{1 - \rho_1^{n_1-n_2}}{1 - \rho_1} \right] = 1$$

Après quelques calculs algébriques, nous obtenons :

$$\pi_0 \left[\frac{1 - \rho_1 - (\rho_1 + \rho_2)^{n_2+1} (1 - \rho_1^{n_1-n_2+1}) + (\rho_1 + \rho_2)^{n_2} \rho_1 (1 - \rho_1^{n_1-n_2})}{(1 - (\rho_1 + \rho_2))(1 - \rho_1)} \right] = 1$$

Donc, l'expression de π_0 est donnée par :

$$\pi_0 = \frac{(1 - (\rho_1 + \rho_2))(1 - \rho_1)}{1 - \rho_1 - (\rho_1 + \rho_2)^{n_2+1} (1 - \rho_1^{n_1-n_2+1}) + (\rho_1 + \rho_2)^{n_2} \rho_1 (1 - \rho_1^{n_1-n_2})}$$

■

Nombre moyen de clients Lorsque $D \geq n_1 - 1$, le nombre espéré de clients dans le système est donné par :

$$N = \sum_{i=1}^{n_1} i \pi_i$$

$$= \pi_0 \left[\sum_{i=1}^{n_2} i (\rho_1 + \rho_2)^i + (\rho_1 + \rho_2)^{n_2} \sum_{i=n_2+1}^{n_1} i \rho_1^i \right] \quad (5.4)$$

Telle que la somme $\sum_{i=1}^{n_2} i(\rho_1 + \rho_2)^i$ est calculée comme suit :

$$\begin{aligned}
 \sum_{i=1}^{n_2} i(\rho_1 + \rho_2)^i &= (\rho_1 + \rho_2) \sum_{i=1}^{n_2} i(\rho_1 + \rho_2)^{i-1} \\
 &= (\rho_1 + \rho_2) (1 + 2(\rho_1 + \rho_2) + 3(\rho_1 + \rho_2)^2 + \cdots + n_2(\rho_1 + \rho_2)^{n_2-1}) \\
 &= (\rho_1 + \rho_2) (1 + (\rho_1 + \rho_2) + (\rho_1 + \rho_2)^2 + (\rho_1 + \rho_2)^3 + \cdots + (\rho_1 + \rho_2)^{n_2})' \\
 &= (\rho_1 + \rho_2) \left(\frac{1 - (\rho_1 + \rho_2)^{n_2+1}}{1 - (\rho_1 + \rho_2)} \right)' \\
 &= (\rho_1 + \rho_2) \left(\frac{(1 - (\rho_1 + \rho_2))(n_2 + 1)(\rho_1 + \rho_2)^{n_2} + (1 - (\rho_1 + \rho_2)^{n_2+1})}{(1 - (\rho_1 + \rho_2))^2} \right)
 \end{aligned}$$

et avec le même principe, la somme $\sum_{i=n_2+1}^{n_1} i\rho_1^i$ est calculée comme suit :

$$\sum_{i=n_2+1}^{n_1} i\rho_1^i = \rho_1^{n_2+2} \left(\frac{(1 - \rho_1)(n_1 + 1)\rho_1^{n_1} + (1 - \rho_1^{n_1+1})}{(1 - \rho_1)^2} \right)$$

Ensuite, nous obtenons l'expression suivante :

$$\begin{aligned}
 N = \pi_0 &\left[(\rho_1 + \rho_2) \left(\frac{(1 - (\rho_1 + \rho_2))(n_2 + 1)(\rho_1 + \rho_2)^{n_2} + (1 - (\rho_1 + \rho_2)^{n_2+1})}{(1 - (\rho_1 + \rho_2))^2} \right) \right] \\
 &+ \pi_0 \left[(\rho_1 + \rho_2)^{n_2} \rho_1^{n_2+2} \left(\frac{(1 - \rho_1)(n_1 + 1)\rho_1^{n_1} + (1 - \rho_1^{n_1+1})}{(1 - \rho_1)^2} \right) \right]
 \end{aligned}$$

Dans la section suivante, seulement les PPI u_D avec $D \in \{0, \dots, n_1 - 2\}$ sont considérées.

5.3.2 Cas partiellement observable

Le système est dit partiellement observable lorsque seulement une partie des nouveaux clients, quelles que soient leurs classes, observent la longueur de la file d'attente avant de prendre leur décision de rejoindre ou non la file d'attente. Lorsque $D \in \{0, \dots, n_1 - 2\}$, un nouveau client de type j qui arrive peut-être confronté à deux situations différentes décrites en détails ci-dessous.

- La longueur de la file d'attente est observable. Dans ce cas, la longueur réelle de la file d'attente i est inférieure ou égale à D . Ceci est le cas pour un client de classe j lorsque $i \leq D < n_j - 1$ tel que $j \in \{1, 2\}$. Dans ce cas, ce nouveau client de classe j avec $j \in \{1, 2\}$, rejoint la file d'attente avec une probabilité $q_j = 1$

et son utilité nette attendue, donnée par l'expression suivante : $R - \frac{i+1}{\mu}C_j$, est strictement positive.

- La longueur de la file d'attente n'est pas observable. Dans ce cas, la longueur réelle de la file d'attente i est supérieure à D . Un nouveau client de type j rejoint la file d'attente avec une probabilité notée q_j avec $j \in \{1,2\}$. Cette probabilité dépend de son temps de séjour attendu noté par W_D .

Dans la suite, nous allons étudier en détails les cas particuliers où $D \leq n_2 - 1$ et $n_2 - 1 < D \leq n_1 - 2$. Dans chacun des cas nous décrivons la structure de la chaîne de Markov associée et l'expression explicite de la probabilité stationnaire π_0 d'avoir le système vide.

5.3.2.1 Premier cas : $D \in \{0, n_2 - 1\}$

Dans ce cas, la longueur i de la file d'attente est observable par un nouveau client de type j lorsqu'elle est inférieure ou égale à $D \in \{0, n_2 - 1\}$.

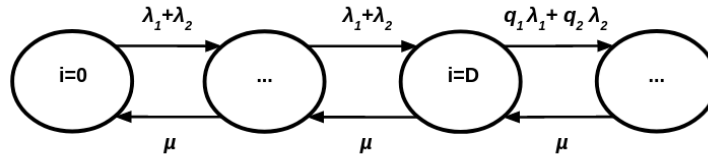


Figure 5.3 – Chaîne de Markov associée au modèle partiellement observable (cas $D \in \{0, n_2 - 1\}$)

Proposition 8. La probabilité stationnaire lorsque le système de mise en file d'attente est vide est donnée par :

$$\pi_0 = \frac{(1 - (\rho_1 + \rho_2))(\mu - (\lambda_1 q_1 + \lambda_2 q_2))}{\mu(1 - (\rho_1 + \rho_2)^{D+1}) - (\lambda_1 q_1 + \lambda_2 q_2) + (\rho_1 + \rho_2)^D (\lambda_1 q_1 + \lambda_2 q_2)}$$

Preuve

La probabilité stationnaire π_i que la longueur de la file d'attente soit égale à $i > 0$, est donnée par :

$$\forall i, \pi_i = \begin{cases} \pi_0 (\rho_1 + \rho_2)^i & \text{si } i \leq D \\ \pi_0 (\rho_1 + \rho_2)^D \left(\frac{\lambda_1 q_1 + \lambda_2 q_2}{\mu} \right)^{i-D} & \text{si } i > D \end{cases} \quad (5.5)$$

En additionnant tous les termes à 1, la probabilité stationnaire lorsque la file d'attente est vide peut être calculée comme suit :

$$\pi_0 \left[\sum_{i=0}^D (\rho_1 + \rho_2)^i + \sum_{i=D+1}^{\infty} (\rho_1 + \rho_2)^D \left(\frac{\lambda_1 q_1 + \lambda_2 q_2}{\mu} \right)^{i-D} \right] = 1$$

Les sommes sont calculées comme suit :

$$\sum_{i=0}^D (\rho_1 + \rho_2)^i = \frac{1 - (\rho_1 + \rho_2)^{D+1}}{1 - (\rho_1 + \rho_2)}$$

et,

$$\sum_{i=D+1}^{\infty} (\rho_1 + \rho_2)^D \left(\frac{\lambda_1 q_1 + \lambda_2 q_2}{\mu} \right)^{i-D} = (\rho_1 + \rho_2)^D \frac{\lambda_1 q_1 + \lambda_2 q_2}{\mu - (\lambda_1 q_1 + \lambda_2 q_2)}$$

Ensuite, nous obtenons :

$$\begin{aligned} \pi_0 \left[\frac{1 - (\rho_1 + \rho_2)^{D+1}}{1 - (\rho_1 + \rho_2)} + (\rho_1 + \rho_2)^D \frac{\lambda_1 q_1 + \lambda_2 q_2}{\mu - (\lambda_1 q_1 + \lambda_2 q_2)} \right] &= 1 \\ \pi_0 \left[\frac{(1 - (\rho_1 + \rho_2)^{D+1})(\mu - (\lambda_1 q_1 + \lambda_2 q_2)) + (\rho_1 + \rho_2)^D (\lambda_1 q_1 + \lambda_2 q_2) (1 - (\rho_1 + \rho_2))}{(1 - (\rho_1 + \rho_2))(\mu - (\lambda_1 q_1 + \lambda_2 q_2))} \right] &= 1 \\ \pi_0 &= \frac{(1 - (\rho_1 + \rho_2))(\mu - (\lambda_1 q_1 + \lambda_2 q_2))}{\mu(1 - (\rho_1 + \rho_2)^{D+1}) - (\lambda_1 q_1 + \lambda_2 q_2) + (\rho_1 + \rho_2)^D (\lambda_1 q_1 + \lambda_2 q_2)} \end{aligned}$$

■

Nombre moyen de clients Le nombre espéré de clients dans le système lorsque $D \in \{0, n_2 - 1\}$ est donné par :

$$\begin{aligned} N &= \sum_{i=1}^{\infty} i \pi_i \\ &= \pi_0 \left[\sum_{i=1}^D i (\rho_1 + \rho_2)^i + \sum_{i=D+1}^{\infty} i (\rho_1 + \rho_2)^D \left(\frac{\lambda_1 p_1 + \lambda_2 p_2}{\mu} \right)^{i-D} \right] \end{aligned} \quad (5.6)$$

5.3.2.2 Deuxième cas : $D \in \{n_2, n_1 - 2\}$

Dans ce cas, la longueur i de la file d'attente est observable par un nouveau client de type j lorsqu'elle est inférieure ou égale à $D \in \{n_2, n_1 - 2\}$.

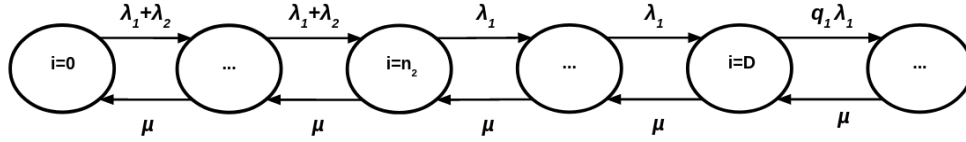


Figure 5.4 – Chaîne de Markov associée au modèle partiellement observable (cas $D \in \{n_2, n_1 - 2\}$)

Proposition 9. La probabilité stationnaire lorsque le système de mise en file d'attente est vide est donnée par :

$$\pi_0 = \frac{(1 - (\rho_1 + \rho_2))(1 - \rho_1)(1 - \rho_1 q_1)}{(1 - \rho_1 q_1)(1 - \rho_1 - \rho_2(\rho_1 + \rho_2)^{n_2}) + (\rho_1 + \rho_2)^{n_2} (q_1 - 1) \rho_1^{D-n_2+1} (1 - (\rho_1 + \rho_2))}$$

Preuve

La distribution stationnaire au sein d'une chaîne de Markov, lorsque la longueur de la file d'attente est égale à $i > 0$, est donnée par :

$$\forall i, \pi_i = \begin{cases} \pi_0 (\rho_1 + \rho_2)^i & \text{si } i \leq n_2 \\ \pi_0 (\rho_1 + \rho_2)^{n_2} \rho_1^{i-n_2} & \text{si } n_2 + 1 \leq i \leq D \\ \pi_0 (\rho_1 + \rho_2)^{n_2} \rho_1^{D-n_2} (q_1 \rho_1)^{i-D} & \text{si } i > D \end{cases}$$

En additionnant tous les termes à 1, la probabilité stationnaire lorsque la file d'attente est vide peut être calculée comme suit :

$$\pi_0 \left[\sum_{i=0}^{n_2} (\rho_1 + \rho_2)^i + \sum_{i=n_2+1}^D (\rho_1 + \rho_2)^{n_2} \rho_1^{i-n_2} + \sum_{i=D+1}^{\infty} (\rho_1 + \rho_2)^{n_2} \rho_1^{D-n_2} (q_1 \rho_1)^{i-D} \right] = 1$$

Les sommes sont calculées comme suit :

$$\sum_{i=0}^{n_2} (\rho_1 + \rho_2)^i = \frac{1 - (\rho_1 + \rho_2)^{n_2+1}}{1 - (\rho_1 + \rho_2)},$$

$$\sum_{i=n_2+1}^D (\rho_1 + \rho_2)^{n_2} \rho_1^{i-n_2} = (\rho_1 + \rho_2)^{n_2} \rho_1 \frac{1 - \rho_1^{D-n_2}}{1 - \rho_1},$$

et

$$\sum_{i=D+1}^{\infty} (\rho_1 + \rho_2)^{n_2} \rho_1^{D-n_2} (q_1 \rho_1)^{i-D} = (\rho_1 + \rho_2)^{n_2} \rho_1^{D-n_2} \frac{\rho_1 q_1}{1 - \rho_1 q_1}.$$

Ensuite, nous obtenons :

$$\pi_0 \left[\frac{1 - (\rho_1 + \rho_2)^{n_2+1}}{1 - (\rho_1 + \rho_2)} + (\rho_1 + \rho_2)^{n_2} \rho_1 \frac{1 - \rho_1^{D-n_2}}{1 - \rho_1} + (\rho_1 + \rho_2)^{n_2} \rho_1^{D-n_2} \frac{\rho_1 q_1}{1 - \rho_1 q_1} \right] = 1$$

Donc la probabilité stationnaire lorsque la file d'attente est vide est donnée avec l'expression suivante :

$$\pi_0 = \frac{(1 - (\rho_1 + \rho_2))(1 - \rho_1)(1 - \rho_1 q_1)}{(1 - \rho_1 q_1)(1 - \rho_1 - \rho_2(\rho_1 + \rho_2)^{n_2}) + (\rho_1 + \rho_2)^{n_2}(q_1 - 1)\rho^{D-n_2+1}(1 - (\rho_1 + \rho_2))}.$$

■

Nombre moyen de clients Le nombre espéré de clients dans le système lorsque $D \in \{n_2, n_1 - 2\}$ est donné par :

$$\begin{aligned} N &= \sum_{i=1}^{\infty} i\pi_i \\ &= \pi_0 \left[\sum_{i=1}^{n_2} i(\rho_1 + \rho_2)^i + \sum_{i=n_2+1}^D i(\rho_1 + \rho_2)^{n_2} \rho_1^{i-n_2} + \sum_{i=D+1}^{\infty} i(\rho_1 + \rho_2)^{n_2} \rho_1^{D-n_2} (\rho_1 \rho_2)^{i-D} \right] \end{aligned} \quad (5.7)$$

5.3.2.3 Détermination de l'équilibre

Le temps de séjour attendu W_D pour chaque client de type j est donné par la formule suivante :

$$W_D = \sum_{i=D+1}^{\infty} \pi_i (i | i > D) \frac{i+1}{\mu}.$$

La probabilité $\pi(i | i > D)$ est la probabilité que la longueur de la file d'attente soit égale à i sachant que cette longueur est supérieur à D . Cette probabilité est donnée par :

$$\pi(i | i > D) = \frac{\pi_i}{\sum_{k=D+1}^{\infty} \pi_k}, \quad (5.8)$$

Ensuite, nous obtenons l'expression suivante :

$$W_D(q_1, q_2) = \frac{D+1}{\mu} + \frac{1}{\mu - \sum_{j=1}^2 \lambda_j q_j}.$$

L'utilité attendue d'un client de type j non informé est donnée par l'expression suivante :

$$U_j(q) = R - C_j W_D$$

et dépend du vecteur $q = (q_1, q_2)$ des stratégies de chacune des classes de clients. Ensuite, une stratégie d'équilibre q_j^* associé pour chaque client de type j satisfait les

conditions suivantes :

$$q_j^* = \begin{cases} 1 & \text{si } U_j(q^*) > 0 \\ \in [0,1] & \text{si } U_j(q^*) = 0 \\ 0 & \text{si } U_j(q^*) < 0 \end{cases} \quad (5.9)$$

Nous avons la proposition suivante qui décrit l'équilibre en fonction des paramètres du système.

Proposition 10. *La stratégie d'équilibre $q^* = (q_1^*, q_2^*)$ peut être déterminée par l'un des cas suivants :*

- si $W_D(0,0) > \frac{R}{C_1}$ donc, $(q_1^*, q_2^*) = (0,0)$,
- si $W_D(1,1) < \frac{R}{C_2}$ donc, $(q_1^*, q_2^*) = (1,1)$,
- si $\frac{R}{C_2} < W_D(1,0) < \frac{R}{C_1}$ donc, $(q_1^*, q_2^*) = (1,0)$,
- si $W_D(1,0) > \frac{R}{C_1}$ donc, $(q_1^*, q_2^*) = (q_1^*, 0)$, avec $q_1^* \in]0,1[$ [tel que $W_D = \frac{R}{C_1}$,
- si $W_D(1,0) < \frac{R}{C_2}$ donc, $(q_1^*, q_2^*) = (1, q_2^*)$, avec $q_2^* \in]0,1[$ [tel que $W_D = \frac{R}{C_2}$.

avec les expressions des probabilités q_1^* et q_2^* sont données par :

$$q_1^* = \frac{\mu(R\mu - (D+2)C_1)}{\lambda_1(R\mu - (D+1)C_1)}$$

et

$$q_2^* = \frac{\mu(R\mu - (D+2)C_2)}{\lambda_2(R\mu - (D+1)C_2)} - \frac{\lambda_1}{\lambda_2}$$

Preuve

La preuve de cette proposition est organisée en plusieurs cas :

- Si $W_D(0,0) > \frac{R}{C_1}$, aucun client de classe 1 va rejoindre le système à l'équilibre ($q_1^* = 0$). Sachant que $C_1 < C_2$, alors $\frac{R}{C_1} > \frac{R}{C_2}$. Par ailleurs, $W_D(0,0) > \frac{R}{C_2}$, donc aucun client de classe 2 va rejoindre le système à l'équilibre ($q_2^* = 0$). Dans ce cas, la stratégie d'équilibre est $q^* = (0, 0)$.
- Si $W_D(1,1) < \frac{R}{C_2}$, les clients de classe 2 vont rejoindre le système à l'équilibre avec une probabilité $q_2^* = 1$. Sachant que $\frac{R}{C_2} < \frac{R}{C_1}$, les clients de classe 1 vont rejoindre le système à l'équilibre avec une probabilité $q_1^* = 1$. Dans ce cas, la stratégie d'équilibre est $q^* = (1, 1)$.

- Si $W_D(1,0) > \frac{R}{C_2}$, les clients de classe 2 ne vont pas rejoindre le système à l'équilibre ($q_2^* = 0$). De plus, si $W_D(1,0) < \frac{R}{C_1}$, les clients de classe 1 vont rejoindre le système à l'équilibre avec une probabilité $q_1^* = 1$. Dans ce cas, la stratégie d'équilibre est $q^* = (1, 0)$.
- Si $W_D(1,0) > \frac{R}{C_1}$ donc, $q^* = (q_1^*, 0)$, avec $q_1^* \in]0,1[$ tel que $W_D = \frac{R}{C_1}$.

$$W_D = \frac{D+1}{\mu} + \frac{1}{\mu - \lambda_1 q_1^*} = \frac{R}{C_1}.$$

En isolant q_1^* , nous obtenons :

$$q_1^* = \frac{\mu(R\mu - (D+2)C_1)}{\lambda_1(R\mu - (D+1)C_1)},$$

- Si $W_D(1,0) < \frac{R}{C_2}$ donc, $(q_1^*, q_2^*) = (1, q_2^*)$, avec $q_2^* \in]0,1[$ tel que $W_D = \frac{R}{C_2}$.

$$W_D = \frac{D+1}{\mu} + \frac{1}{\mu - \lambda_1 - \lambda_2 q_2^*} = \frac{R}{C_2}.$$

En isolant q_2^* , nous obtenons :

$$q_2^* = \frac{\mu(R\mu - (D+2)C_2)}{\lambda_2(R\mu - (D+1)C_2)} - \frac{\lambda_1}{\lambda_2}.$$

■

Lemme 2. Pour tout $j \in \{1,2\}$, la probabilité de décision stratégique à l'équilibre q_j^* diminue en fonction de D .

Preuve

La preuve se base sur l'étude de la fonction continue qui correspond à la relaxation de la contrainte discrète sur D . La dérivée de la probabilité de décision stratégique à l'équilibre, pour un client j est donnée par :

$$q_j^{*'}(D) = \frac{-\mu C_j^2}{\lambda_j (R\mu - C_j(D+1))^2} < 0$$

La dérivée de $q_j^{*'}$ est strictement négative, ce qui signifie que l'augmentation du seuil D induit une réduction de la proportion de clients de type j non-informés qui rejoignent la file d'attente. ■

Ce résultat est assez intuitif parce que d'autant plus que D est proche de la valeur $n_j - 1$, les clients de type j non informés ont moins de chance de trouver l'état de la file d'attente inférieur à $n_j - 1$ et donc la probabilité qu'ils rejoignent la file diminue.

5.4 Illustrations numériques

Nous illustrons, dans cette section, l'évolution des stratégies de clients à l'équilibre en fonction de la PPI proposée par le fournisseur de services. Nous illustrons ensuite le nombre moyen des clients dans le système en fonction de D . Dans toutes les expériences numériques, nous avons considéré l'ensemble de paramètres suivant : $R = 20$, $\mu = 2$, $\eta = 3$, $C_1 = 3$, $C_2 = 5$, $\lambda_1 = 0.8$ et $\lambda_2 = 1.1$. En fonction de ces valeurs, les seuils du Naor pour les deux classes de clients sont donc égales à $n_1 = 13$ et $n_2 = 8$.

5.4.1 Calcul de l'équilibre

La Figure 5.5 illustre les stratégies d'équilibre des clients de chaque classe dans le système en fonction du seuil D de la PPI appliquée par le fournisseur de services. Les clients rejoignent la file lorsque l'information est partagée avec eux et que la longueur réelle de la file d'attente est courte. Au contraire, les clients ne rejoignent pas le système lorsqu'ils observent que la longueur de file d'attente est assez grande. Au niveau intermédiaire, comme attendu, ce sont d'abord les clients les plus sensibles à l'attente (la seconde classe ici) qui ne rejoignent pas le système.

D'où l'observation suivante :

Observation 5. *Les stratégies d'équilibre des clients, q_1^* et q_2^* , sont strictement décroissantes en fonction du seuil D fixé par le fournisseur de services.*

Cette observation a été prouvé théoriquement dans la Lemme 2.

5.4.2 Nombre moyen de clients

La Figure 5.6 représente le nombre moyen de clients dans le système en fonction du seuil D de la PPI.

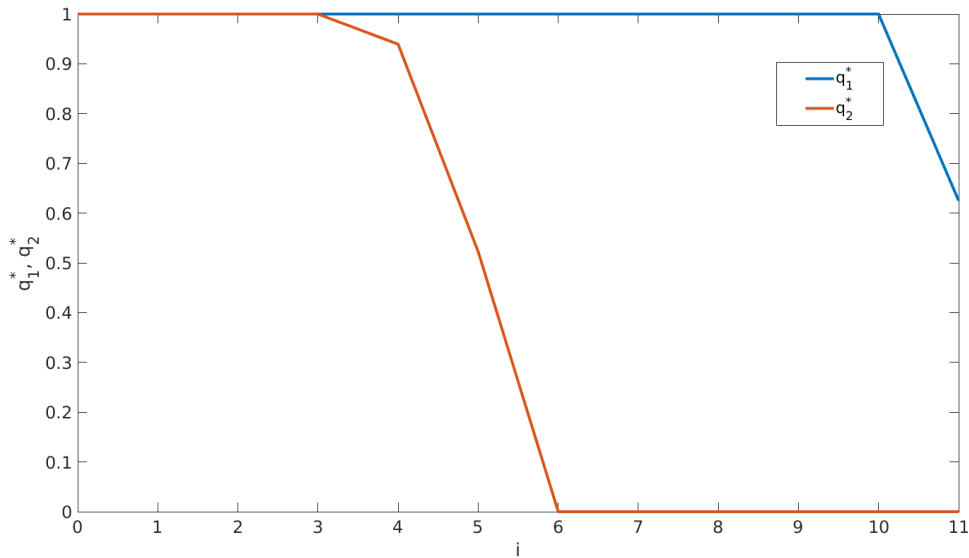


Figure 5.5 – Équilibres q_1^* et q_2^* dans un système de file d'attente à deux classes en fonction de D

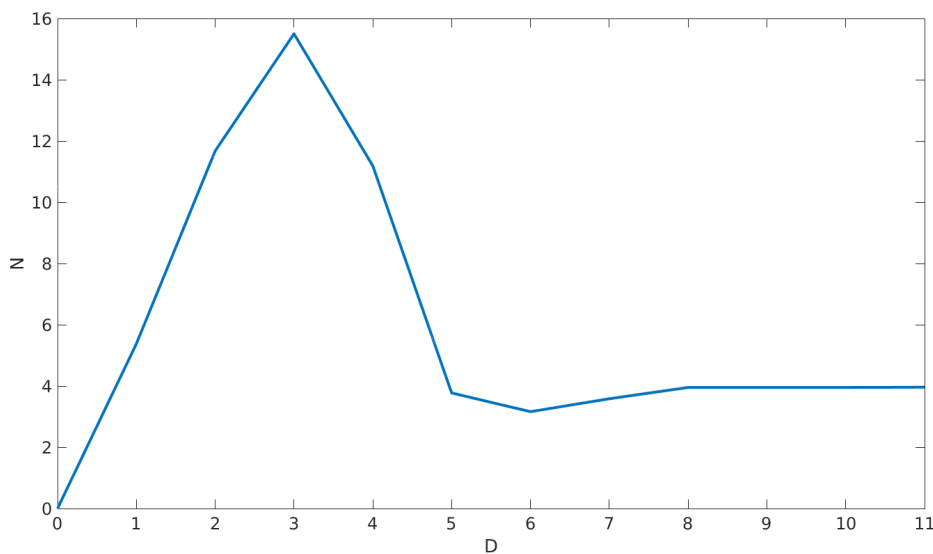


Figure 5.6 – Nombre moyen des clients dans le système en fonction de D

Observation 6. Le nombre moyen de clients atteint son maximum lorsque les clients, des deux classes, rejoignent le système ($q_1^* = q_2^* = 1$).

Observation 7. Le nombre moyen de clients atteint son minimum lorsque les clients de classe 2 ne rejoignent plus le système ($q_2^* = 0$).

Dans cet exemple, le nombre moyen de clients augmente dans le système jusqu'à la valeur du seuil $D = 3$ parce qu'ils rejoignent le système avec des probabilités q_1^* et q_2^* égales à 1. A partir de cette valeur, les clients de classe 2 commencent à diminuer et

décident de rejoindre le système avec une stratégie mixte $q_2^* \in]0,1[$. D'où, le nombre moyen diminue jusqu'à la valeur $D = 6$ à partir de laquelle les clients de classe 2 décident de ne plus rejoindre le système. À partir de cette valeur, le nombre moyen de clients reprend sa croissance car les clients les moins sensibles continuent toujours de rejoindre avec une grande proportion.

5.4.3 Temps moyen de séjour

La Figure 5.7 représente, avec le même ensemble de paramètres, le temps moyen de séjour $W_D(q_1^*, q_2^*)$ attendu dans le système en fonction du seuil D de la PPI.

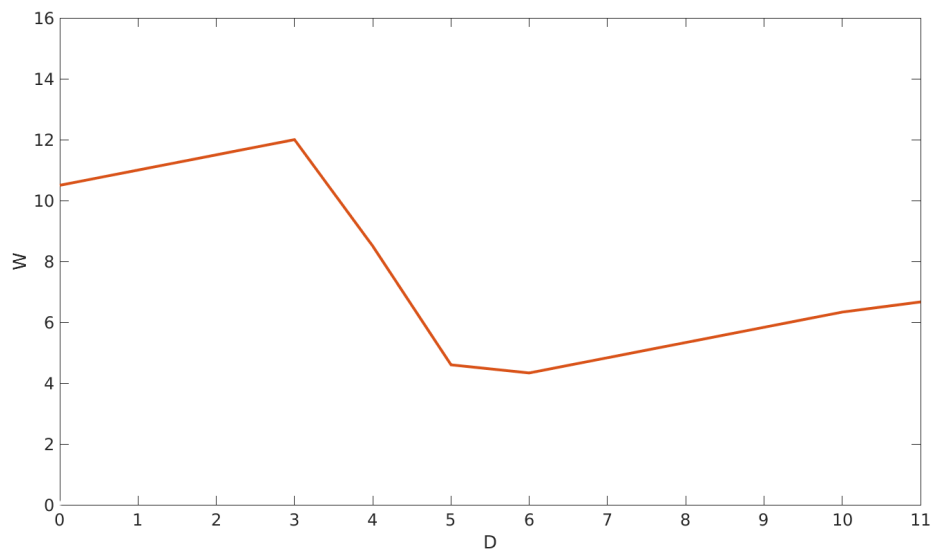


Figure 5.7 – Temps de séjour dans le système en fonction de D

Observation 8. *Le temps de séjour atteint son maximum lorsque les clients, des deux classes, rejoignent le système ($q_1^* = q_2^* = 1$).*

Observation 9. *Le temps de séjour atteint son minimum lorsque les clients de classe 2 ne rejoignent plus le système ($q_2^* = 0$). Ceci correspond également à la valeur de seuil D qui minimise le nombre moyen de clients.*

Globalement, nous pouvons donc remarquer que le temps moyen de séjour dans le système a la même allure que le nombre de clients dans le système.

5.5 Conclusion

Dans ce chapitre, nous avons traité un système similaire à celui étudié dans le chapitre précédent, mais avec une population hétérogène de clients. En particulier, les clients évaluent différemment le temps de séjour dans le système. En effet, dans les systèmes réels, certains sont impatients et d'autres moins. Nous avons traité une file d'attente qui impose une politique déterministe de partage de l'information basé sur l'état du système en temps réel et qui reçoit des classes différentes de clients. Nous nous sommes intéressés à l'impact de l'information sur les décisions stratégiques des clients et sur l'équilibre. Nous avons étudié tous les cas possibles, lorsque le système contient deux classes de clients, et nous avons donné les expressions explicites de l'équilibre pour chaque type de clients.

Chapitre 6

Systeme de file d'attente avec deplacement et information partiellement observable

Sommaire

6.1	Introduction	90
6.2	Description du modele de file d'attente avec deplacement	91
6.2.1	Modélisation du systeme et des decisions des clients	91
6.2.2	Modélisation des decisions dans la file d'attente de service	92
6.2.3	Modélisation des decisions dans la file d'attente de deplacement	93
6.3	Processus de naissance et de mort dans un systeme avec deplacement et partiellement observable	93
6.3.1	Transitions possibles	93
6.3.2	Calcul des probabilites stationnaires	97
6.4	Calcul des utilites attendues	99
6.4.1	Utilite des clients non-informes	99
6.4.2	Utilite des clients informes	99
6.5	Strategie d'equilibre dans un systeme partiellement observable	102
6.5.1	Meilleure reponse pour les clients non-informes	102
6.5.2	Meilleure reponse pour les clients informes	103
6.5.3	Strategie d'equilibre	103
6.5.4	Bien-etre social	105
6.6	Illustrations numeriques	105
6.6.1	Strategies d'equilibre	105
6.6.2	Bien-etre social	108
6.7	Conclusion	109

6.1 Introduction

Pour recevoir un service, les clients doivent dans la plupart du temps se déplacer au système de file d'attente qui fournit le service. Des nouveaux systèmes sont apparus récemment qui donnent aux clients la possibilité d'accéder à l'information sur l'état de la file en temps réel avant le déplacement tel que le site Web des douanes et de la protection des frontières des États-Unis¹ qui est dédié pour les clients qui voyagent des États-Unis au Canada et vice versa, et subissent souvent des retards à cause de l'attente à la frontière.

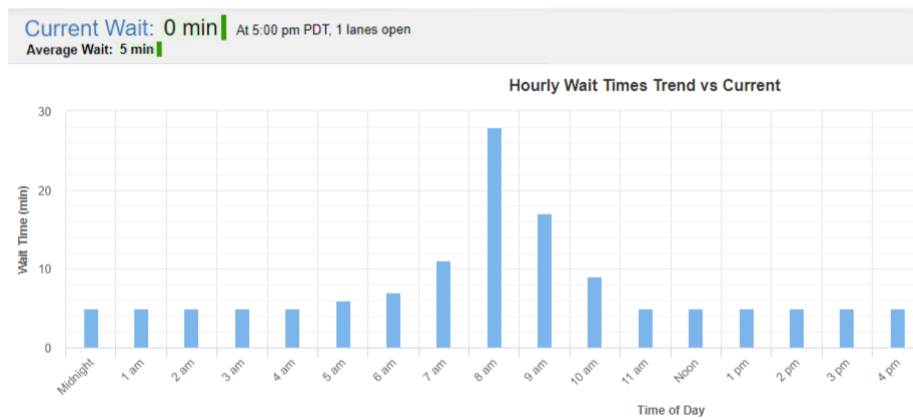


Figure 6.1 – Temps d'attente publié au site de la protection des douanes et des frontières américaines pour le passage de l'autoroute Blaine

Ce site donne la possibilité aux clients de vérifier le temps d'attente actuel sur le pont avant leur voyage. La Figure 6.1 affiche le temps d'attente, en temps réel, dans l'un des passages frontaliers les plus fréquentés qui relie les villes de Blaine à Washington, aux États-Unis, et de Surrey en Colombie-Britannique, au Canada. Dans ce genre de système, les clients prennent leur décision de voyager ou non en se basant sur l'information fournie. Dans ce chapitre, nous modélisons un problème du voyage où les clients se trouvent loin d'un service et ont besoin de voyager pour être servi. Nous supposons que seulement une partie de la population est informée par l'état de la file avant le voyage et nous étudions l'impact de l'information partagée avec cette fraction de la population sur l'équilibre et les performances globales du système.

1. voir : <https://explore.dot.gov/t/BTS/views/BTSBorderCrossingAnnualData/BorderCrossingTableDashboard>

6.2 Description du modèle de file d'attente avec déplacement

6.2.1 Modélisation du système et des décisions des clients

Nous considérons un système constitué par deux files d'attente successives. Les clients vont rejoindre une première file d'attente qui possède un nombre infini de serveurs ($M/M/\infty$). Le temps de déplacement des clients est donné par une variable aléatoire exponentielle de paramètre η . Ceci peut modéliser par exemple les différents temps de trajet pour atteindre un service particulier pour diverses clients. Le taux d'arrivée à cette file d'attente est un processus de Poisson de taux λ . Les clients dans ce système sont stratégiques : ils décident de voyager pour obtenir le service si et seulement si leur utilité attendue est positive. Après le voyage, les clients arrivent à une deuxième file d'attente (la file de service) qui possède un seul serveur avec un taux de service μ . Ils décident de rejoindre le système de service ou de le quitter sans recevoir le service, en se basant sur l'information de la longueur de cette file d'attente. Celle-ci est supposée observable ou l'information à propos de l'occupation de cette file est donnée à l'arrivée d'un client. Les clients qui décident de rejoindre la deuxième file d'attente (celle qui fournit le service) reçoivent une récompense noté par $R > 0$ et paient un coût d'attente $C > 0$ par unité de temps de séjour. La décision des clients de rejoindre ou non la file d'attente de service après un voyage ne prend pas en considération le coût du voyage, qui est considéré comme un coût irrécupérable.

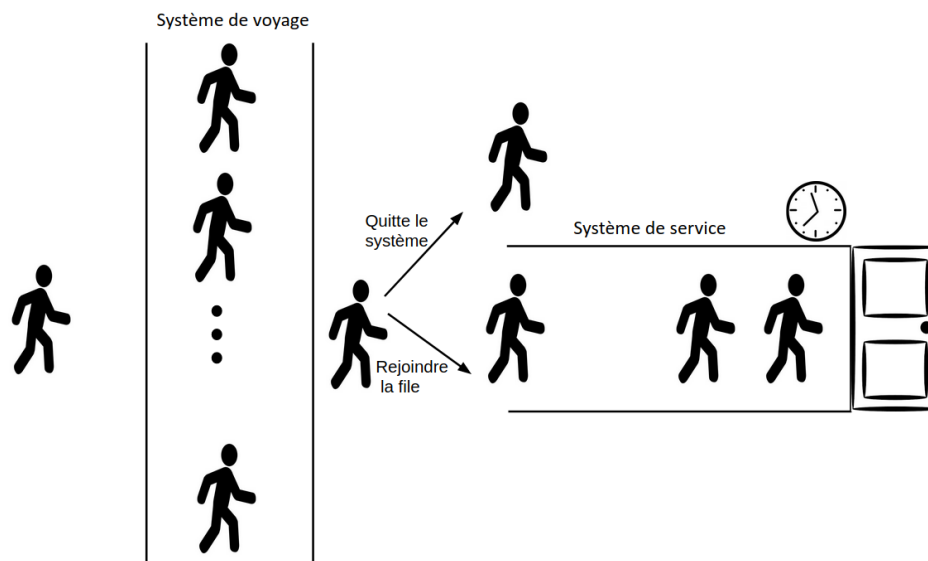


Figure 6.2 – Processus de prise de décisions dans un système de file d'attente avec déplacement

Nous modélisons ce problème en supposant que seulement une partie de la population est informée par la longueur de la file avant le voyage. Deux types de clients sont donc considérés dans le système : des clients informés et des clients non-informés par la longueur de la file d'attente de service avant le voyage. Nous étudions le comportement stratégique des clients dans un tel système de déplacement. Les deux fractions de la population de clients sont définies comme suit :

- La première fraction, notée par θ , représente la partie de population des clients non-informés par l'état de la file d'attente de service avant le voyage. Pour cette fraction, l'état du système de service est révélé aux clients que lorsqu'ils vont y arriver.
- La deuxième fraction, notée par $(1 - \theta)$, représente la partie de population des clients informés où l'information sur l'état de la file d'attente de service est fournie aux clients avant le voyage. Ils utilisent cette information pour prendre la décision de se déplacer ou non. Pour cette fraction, les clients n'obtiennent aucune information supplémentaire sur l'état de la file d'attente pendant le voyage.

La fraction de chaque type de clients n'est pas nécessairement définie par le fournisseur de services, elle peut dépendre de clients eux-mêmes. Certains clients peuvent accéder aux sites web ou aux applications mobile qui fournit l'information. D'autres clients n'ont pas les mêmes moyens et donc ils décident de voyager sans être informé par l'état du système. Dans ce système partiellement observable, l'utilité attendu d'un client, informé ou non-informé, qui décide de voyager est forcément non négative.

6.2.2 Modélisation des décisions dans la file d'attente de service

Le système de file d'attente de service est un système observable. En effet, tous les clients observent la longueur de la file d'attente à l'instant où ils arrivent au système de service. Un nouveau client qui arrive et observe i clients dans le système y compris celui qui est en service, peut regretter et quitte le système si la longueur de la file est égale au seuil du Naor (Voir (NAOR 1969)). Autrement dit, un client rejoint la file d'attente et attend son tour pour être servi si et seulement si la longueur de la file strictement inférieure à :

$$n_e := \lfloor \frac{R\mu}{C} \rfloor. \quad (6.1)$$

En d'autres termes, n_e est la longueur maximale de la file d'attente de service. Au-delà de cette valeur, l'utilité attendue d'un client qui va rejoindre le système sera négative. La file d'attente de service est donc une file d'attente observable du type $M/M/1/n_e$.

6.2.3 Modélisation des décisions dans la file d'attente de déplacement

Le système de file d'attente de voyage est modélisé par une file d'attente à un nombre infini de serveurs du type $(M/M/\infty)$. La sortie de cette file d'attente suit un processus du Poisson avec le taux λ multiplié par la probabilité par laquelle un client, informé ou non-informé, décide de voyager. Cette sortie constitue le taux d'arrivée à la deuxième file d'attente qui fournit le service.

La stratégie des clients est représentée par un vecteur de probabilités défini par :

$$q = (q_U, q_I(0), \dots, q_I(n_e)),$$

avec q_U représente la probabilité qu'un client non-informé se déplace à la file d'attente de service. Le reste du vecteur représente les probabilités de déplacement attribuées lors de l'observation de i clients dans la file d'attente de service, avec $i \in \{0, \dots, n_e\}$ pour les clients informés. La probabilité de voyage vers le système de service pour un client pris au hasard dans la population est donnée par l'expression suivante : $Q(i) = \theta q_U + (1 - \theta)q_I(i)$. Donc, le débit d'entrée dans la file de déplacement lorsque i clients sont dans la file de service est $\lambda Q(i)$.

6.3 Processus de naissance et de mort dans un système avec déplacement et partiellement observable

Nous modélisons dans ce système toutes les transitions possibles par un couple d'états (i, j) qui représentent respectivement le nombre de clients dans la file d'attente de service et le nombre de clients dans la file d'attente de voyage. Nous supposons que le nombre de clients qui voyagent peut-être infini ce qui signifie que $j \in \{0, \dots, \infty\}$. De plus, la file d'attente de service est une file observable, elle a donc une capacité finie notée par n_e ce qui signifie que $i \in \{0, \dots, n_e\}$.

6.3.1 Transitions possibles

6.3.1.1 Cas $(i < n_e, j < \infty)$:

La Figure 6.3 décrit l'ensemble des transitions possibles, sortantes et entrantes, de/vers le couple (i, j) lorsque $i < n_e$ et $j < N$. Nous expliquons la progression de l'état dans le système en fonction de ce couple comme suit :

- l'état du système change de (i, j) à $(i, j + 1)$ lorsqu'un client commence à se déplacer(rejoint la file d'attente de voyage) vers la file d'attente de service, donc avec un taux $\lambda Q(i) = \lambda \theta q_U + \lambda(1 - \theta)q_I(i)$,
- l'état du système change de (i, j) à $(i - 1, j)$ lorsque le client en service a obtenu son service souhaité et quitte le système, cela se produit avec un taux μ ,
- l'état du système passe de (i, j) à $(i + 1, j - 1)$ lorsqu'un client arrive dans la file d'attente de service après le voyage avec un taux η .
- l'état du système change de $(i + 1, j)$ à (i, j) lorsque le service est terminé pour le client qui est en cours de service, ce qui se produit avec le taux μ ,
- l'état du système change de $(i, j - 1)$ à (i, j) lorsqu'un client commence à voyager pour rejoindre la file d'attente de service, avec un taux $\lambda Q(i) = \lambda \theta q_U + \lambda(1 - \theta)q_I(i)$,
- l'état du système change de $(i - 1, j + 1)$ à (i, j) lorsqu'un client en voyage arrive dans la file d'attente de service avec taux η et il la trouve non complètement chargée,

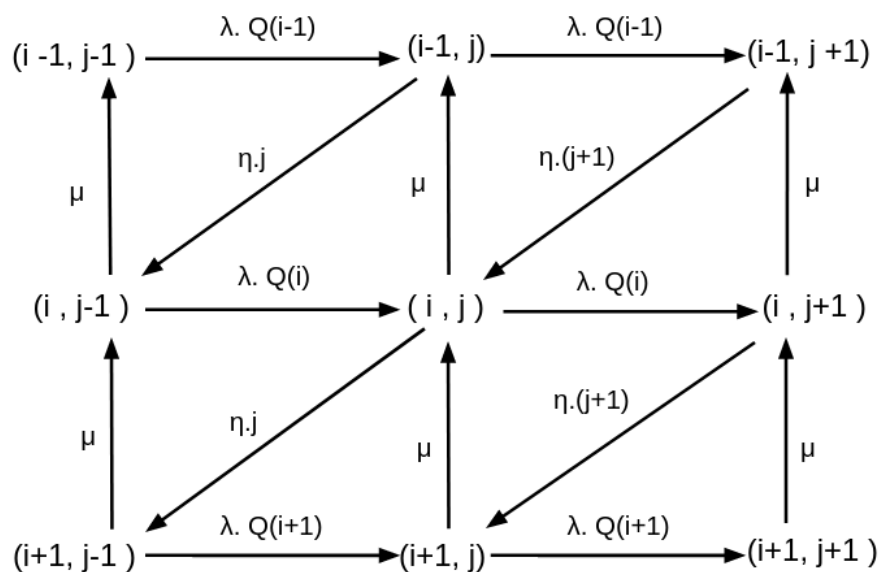


Figure 6.3 – États de transitions autour le couple d'états (i, j) .

Il y a quatre cas particuliers où les transitions autour le couple d'états (i, j) sont un peu différentes. Nous décrivons ci-dessous les cas extrêmes du système.

6.3.1.2 Cas ($i = 0, j = 0$) :

Lorsque la file de déplacement et la file de service sont vides, c'est-à-dire $i = 0$ et $j = 0$, les transitions autour cet état sont décrites comme suit :

- l'état change de $(0, 0)$ à $(0, 1)$ lorsque le premier client commence à se déplacer vers la file d'attente de service, avec un taux $\lambda Q(0)$,
- l'état change de $(1, 0)$ à $(0, 0)$ lorsque le seul client dans le système a obtenu le service et il a quitté le système de service, cela se produit avec un taux μ .

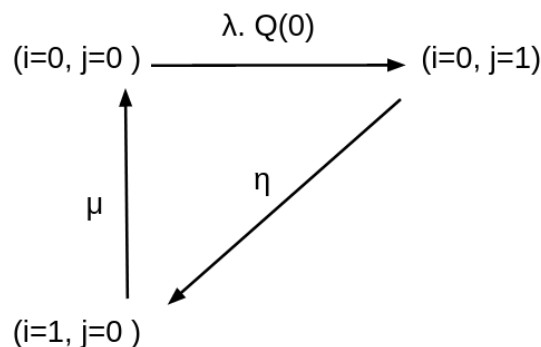


Figure 6.4 – États de transitions autour le couple d'états $(0,0)$.

La Figure 6.4 décrit les deux transitions sortante et entrante de/vers le couple d'états du système $(0,0)$.

Pour modéliser le problème, nous bornons artificiellement le nombre de clients dans la file d'attente en déplacement. Soit N le nombre maximal de clients pouvant être simultanément en déplacement vers la file d'attente de service.

6.3.1.3 Cas ($i = 0, j = N$) :

Lorsque la file de déplacement est pleine et la file de service est vide, c'est-à-dire $i = 0$ et $j = N$, les transitions autour cet état sont décrites comme suit :

- l'état passe de $(0, N)$ à $(1, N - 1)$ lorsqu'un client arrive dans la file d'attente de service après le voyage avec un taux η ,
- l'état change de $(0, N - 1)$ à $(0, N)$ lorsque le client numéro N décide de rejoindre le système de voyage pour atteindre la file d'attente de service, avec un taux $\lambda Q(0)$,
- l'état change de $(1, N)$ à $(0, N)$ lorsque le seul client dans le système a obtenu le service et il a quitté le système de service, cela se produit avec un taux μ .

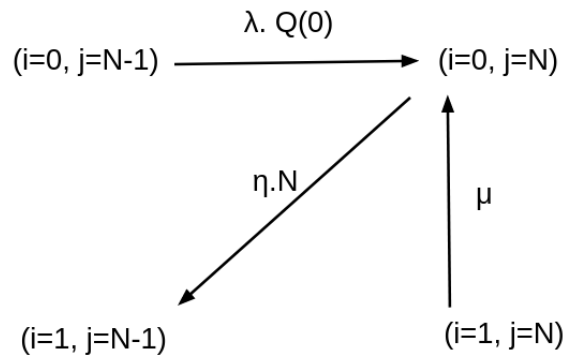


Figure 6.5 – États de transitions autour le couple d'états $(0, N)$.

La Figure 6.5 décrit les transitions sortantes et entrantes de/vers le couple d'états du système $(0, N)$.

6.3.1.4 Cas $(i = n_e, j = 0)$:

Lorsque la file de déplacement est vide et la file de service est pleine, c'est-à-dire $i = n_e$ et $j = 0$, les transitions autour cet état sont décrites comme suit :

- l'état change de $(n_e, 0)$ à $(n_e, 1)$ lorsqu'un premier client décide de rejoindre le système de voyage pour atteindre la file d'attente de service, avec un taux $\lambda Q(n_e)$,
- l'état change de $(n_e, 0)$ à $(n_e - 1, 0)$ lorsque le seul client dans le système a obtenu le service et il a quitté le système de service, cela se produit avec un taux μ ,
- l'état passe de $(n_e - 1, 1)$ à $(n_e, 0)$ lorsqu'un client arrive dans la file d'attente de service après le voyage avec un taux η ,
- l'état passe de $(n_e, 1)$ à $(n_e, 0)$ lorsqu'un client arrive dans la file d'attente de service après le voyage avec un taux η .

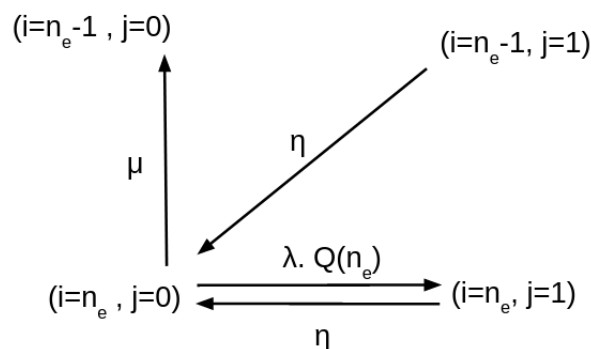


Figure 6.6 – États de transitions autour le couple d'états $(n_e, 0)$.

La Figure 6.6 décrit les transitions sortantes et entrantes de/vers le couple d'états du système $(n_e, 0)$.

6.3.1.5 Cas $(i = n_e, j = N)$:

Lorsque la file de déplacement et la file de service sont pleines, c'est-à-dire $i = n_e$ et $j = N$, les transitions autour cet état sont décrites comme suit :

- l'état change de (n_e, N) à $(n_e - 1, N)$ lorsque le service est terminé pour le client numéro n_e et qui a par la suite, quitté le système. Cela se produit avec un taux μ ,
- l'état passe de (n_e, N) à $(n_e, N - 1)$ lorsqu'un client arrive dans la file d'attente de service après le voyage avec un taux η ,
- l'état change de $(n_e, N - 1)$ à (n_e, N) lorsqu'un client décide de rejoindre le système de voyage, avec le d'indice N , pour atteindre la file d'attente de service, avec un taux $\lambda Q(n_e)$.

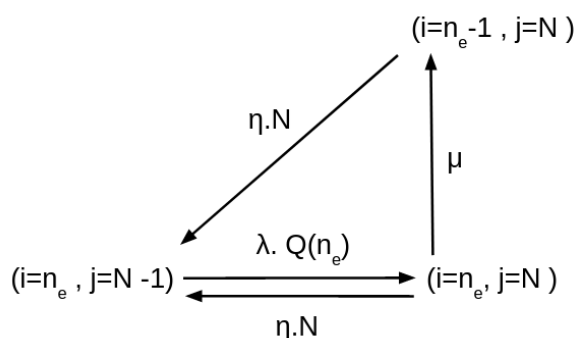


Figure 6.7 – États de transitions autour le couple d'états (n_e, N) .

La Figure 6.7 décrit les transitions sortantes et entrantes de/vers le couple d'états du système (n_e, N) .

6.3.2 Calcul des probabilités stationnaires

En se basant sur les expressions de transition vu précédemment, et sachant que la $\pi_{i,j}$ est la probabilité d'état stationnaire du système, d'après le théorème de coupes des chaînes de Markov, nous écrivons l'équation des probabilités stationnaires pour tout $i = \{0, 1, \dots, n_e\}$ et $j = \{0, 1, \dots, \infty\}$ comme :

$$\begin{aligned}
 & [\lambda \cdot ((1 - \theta)q_I(i) + \theta q_U) + \mu \cdot \mathbb{1}_{\{i>0\}} + \eta \cdot j] \pi_{i,j} \\
 & = \lambda \cdot ((1 - \theta)q_I(i) + \theta q_U) \cdot \mathbb{1}_{\{j>0\}} \cdot \pi_{i,j-1} + \mu \cdot \mathbb{1}_{\{i<n_e\}} \cdot \pi_{i+1,j} \\
 & + \eta \cdot (j + 1) [\mathbb{1}_{\{i>0\}} \cdot \pi_{i-1,j+1} + \mathbb{1}_{\{i=n_e\}} \cdot \pi_{i,j+1}],
 \end{aligned} \tag{6.2}$$

où

$$\mathbb{1} = \begin{cases} 1 & \text{si la condition est satisfaite,} \\ 0 & \text{sinon.} \end{cases}$$

Pour calculer la probabilité d'état stationnaire au système, nous avons proposé l'algorithme 2.

Algorithme 2 Calcul de la probabilité de stationnaire $\pi_{i,j}$

```

Données :  $\mu, \lambda, \eta, R, C, N$ ;
 $n_e := \lfloor \frac{R\mu}{C} \rfloor$ ;
//Création d'une matrice A :
 $A := [(n_e + 1) \times (N + 1)][(n_e + 1) \times (N + 1)]$ ;
//Création et initialisation d'un vecteur B :
 $B := \text{ones}[(n_e + 1) \times (N + 1)]$ ;
//Cas ( $i = 0, j = 0$ )
 $r = 1$ ;
 $A(r,r) = \lambda Q(0)$ ;  $A(r, N + 2) = -\mu$ ;
//Cas ( $i = 0, j = N$ )
 $r = N + 1$ ;
 $A(r,r) = \eta N$ ;  $A(r, r - 1) = -\lambda Q(0)$ ,  $A(r, 2r) = -\mu$ ;
//Cas ( $i = n_e, j = 0$ )
 $A((N + 1) \times n_e + 1, (N + 1) \times n_e + 1) = \lambda \times Q(n_e + 1) + \mu$ ,  $A((N + 1) \times n_e + 1, (N + 1) \times n_e + 1 + 1) = -\eta \times 1$ ;
//Cas ( $i = n_e, j = N$ )
 $A((N + 1) \times n_e + N + 1, (N + 1) \times n_e + N + 1) = \mu + \eta \times N$ ,  $A((N + 1) \times n_e + N + 1, (N + 1) \times n_e + N + 1 - 1) = -\lambda \times Q(n_e + 1)$ ;
//Cas ( $0 < i < n_e, 0 < j < N$ )
for ( $i = 1$  à  $n_e + 1$ ) do
  for ( $j = 1$  à  $N + 1$ ) do
     $r := (N + 1)i + j + 1$ 
     $A(r,r) := \lambda Q(i) + \mu + \eta j$ ,  $A(r, r - 1) := -\lambda Q(i)$ 
     $A(r, (N + 1)(i + 1) + j + 1) := -\mu$ ,  $A(r, (N + 1)(i - 1) + j + 2) := -\eta(j + 1)$ 
  end for
end for
//Calcul de la probabilité stationnaire :
 $\pi_{i,j} = \text{linsolve}(A, B)$ 
Retourner  $\pi_{i,j}$ .

```

Dans cet algorithme, nous avons créé une matrice, nous ont affecté à chaque case l'expression qui correspond à son état en se basant sur les transitions, décrites précédemment. Pour minimiser l'effet de la borne numérique sur les calculs d'équilibre, nous avons choisit N assez grand. Nous nous retrouvons donc avec $(n_e + 1) \times (N + 1)$ valeurs différentes. L'algorithme retourne la valeur $\pi_{i,j}$ en utilisant la fonction pré-définie *linsolve* qui permet de résoudre l'équation $\pi_{i,j} \times A = 1$.

6.4 Calcul des utilités attendues

L'utilité attendue d'un client dépend des stratégies des autres clients, informés et non-informés, dans le système. Il décide de voyager avec une certaine probabilité, lorsque tous les autres clients se déplacent vers le système du service avec une probabilité q (q représente le vecteur des probabilités dans le système). Les deux types des clients sont stratégiques : ils calculent leur utilité attendue dans le système, et ils voyagent si et seulement si cette utilité est non négatif. L'utilité d'un client qui décide de ne pas voyager est égale à 0.

6.4.1 Utilité des clients non-informés

Nous notons par $U(q)$ l'utilité attendue de voyage d'un client non-informé. Cette utilité est calculée comme la différence entre l'évaluation du service R et les coûts d'attente $C \frac{i+1}{\mu}$ et le coût du temps moyen de déplacement $\frac{C}{\eta}$. L'expression de cette utilité attendue est :

$$U(q) = U(q_U, q_I(0), \dots, q_I(n_e)) = \sum_{i=0}^{n_e-1} \pi_i \left(R - C \frac{i+1}{\mu} \right) - \frac{C}{\eta}, \quad (6.3)$$

avec $\pi_i = \sum_{j=0}^{\infty} \pi_{i,j}(q_U, q_I(i))$ est la probabilité stationnaire d'avoir i clients dans la file de service.

6.4.2 Utilité des clients informés

L'utilité attendue d'un client, qui se trouve à une certaine distance du service souhaité et qui connaît l'état du système de service, dépend forcément de la longueur de file d'attente observée au moment de son départ. Les clients utilisent cette

information pour estimer la longueur de file d'attente à leur arrivée et donc, pour vérifier si leur utilité attendue est positive. Ils prennent en considération aussi le trajet qu'ils vont faire pour arriver à la file d'attente de service. Si la congestion du système est faible et le temps de trajet est court, la longueur de la file d'attente à l'arrivée ne changera pas beaucoup par rapport à la longueur de la file d'attente observée. Par contre, pour une congestion plus élevée ou un temps de trajet plus long, la différence entre la longueur de la file d'attente de service avant son déplacement et celle qu'il va voir à son arrivée peut être grand et donc l'information sur la longueur de la file d'attente fournie avant le voyage est décisive afin que le client prenne la meilleure décision. Il faut noter que dans ce travail, la congestion sur le trajet pour atteindre la file de service est simplement pris en compte à travers une variable aléatoire exponentielle. Un modèle plus complexe pourrait être de prendre en considération une file d'attente avec interaction entre les clients (comme une file M/M/1) où le temps de séjour dans cette file dépend de son taux d'occupation et donc de la congestion sur le trajet.

Pour calculer l'utilité attendue dans ce cas, nous définissons la probabilité qu'un client arrive à la file d'attente de service après son déplacement et qu'il la trouve pleine, c'est-à-dire que n_e clients sont présents dans la file de service, sachant qu'il observe i clients dans la file de service à son départ et qu'il y a j clients en déplacement. Cette dernière information n'est pas connue par le client mais cette probabilité est utile pour le calcul de l'utilité attendue. Cette probabilité est notée par $r(i, j)$ et est déterminée comme suit :

$$\begin{aligned}
 r(i, j) = & \frac{\eta \cdot \mathbb{1}_{\{i=n_e\}}}{\lambda \cdot ((1 - \theta)q_I(i) + \theta q_U) + \mu \cdot \mathbb{1}_{\{i>0\}} + j \cdot \eta} \\
 & + \frac{\eta \cdot (j - 1)}{\lambda \cdot ((1 - \theta)q_I(i) + \theta q_U) + \mu \cdot \mathbb{1}_{\{i>0\}} + j \cdot \eta} \cdot [r(i + 1, j - 1) \cdot \mathbb{1}_{\{i < n_e\}} + r(i, j - 1) \cdot \mathbb{1}_{\{i=n_e\}}] \\
 & + \frac{\lambda \cdot ((1 - \theta)q_I(i) + \theta q_U)}{\lambda \cdot ((1 - \theta)q_I(i) + \theta q_U) + \mu \cdot \mathbb{1}_{\{i>0\}} + j \cdot \eta} \cdot r(i, j + 1) \\
 & + \frac{\mu \cdot \mathbb{1}_{\{i>0\}}}{\lambda \cdot ((1 - \theta)q_I(i) + \theta q_U) + \mu \cdot \mathbb{1}_{\{i>0\}} + j \cdot \eta} \cdot r(i - 1, j)
 \end{aligned} \tag{6.4}$$

Cette probabilité est la somme de quatre termes. Le premier terme représente la probabilité qu'un client va être le prochain à arriver au système de service et il va trouver la file d'attente complètement chargée (à l'instant de son arrivé, l'état i du système va être égal à n_e clients). Le deuxième terme représente la probabilité qu'un autre client en déplacement arrive dans la file d'attente. Si à son arrivée la file d'attente n'est pas pleine, le client va la rejoindre et la probabilité s'écrit $r(i + 1, j - 1)$. Sinon, si le client trouve la file d'attente complètement chargée à son arrivée, il va quitter le système et donc la probabilité s'écrit $r(i, j - 1)$. Le troisième terme représente la probabilité qu'un nouveau client décide de voyager pour atteindre le

système de service et la probabilité qu'il trouve la file d'attente pleine s'écrit $r(i, j + 1)$. Le quatrième terme représente la probabilité qu'une prestation de service se termine et la probabilité qu'un client qui arrive trouve la file d'attente complètement chargée est $r(i - 1, j)$.

Ensuite, nous calculons la probabilité $r(i)$ que la file d'attente de service soit pleine à son arrivée lorsqu'un client observe i clients dans la file d'attente de service à son départ. Cette probabilité est donnée par :

$$r(i) = \sum_{j=0}^{\infty} \frac{r(i, j + 1) \pi_{i,j}(q_U, q_I(i))}{\pi_i} \quad (6.5)$$

pour tout $i \in \{0, \dots, n_e\}$ et $\pi_i = \sum_{j=0}^{\infty} \pi_{i,j}(q_U, q_I(i))$.

Enfin, pour l'utilité attendue, nous calculons le temps de séjour prévu d'un client informé par l'état i de la file d'attente de service et qui se déplace pour y arriver, sachant que j clients sont déjà en voyage. Ce temps de séjour noté $E(i, j)$ peut-être calculé par la formule suivante :

$$\begin{aligned} E(i, j) = & \frac{1}{\lambda \cdot ((1 - \theta)q_I(i) + \theta q_U) + \mu \cdot \mathbb{1}_{\{i > 0\}} + j \cdot \eta} \\ & + \frac{\eta \cdot \mathbb{1}_{\{i < n_e\}}}{\lambda \cdot ((1 - \theta)q_I(i) + \theta q_U) + \mu \cdot \mathbb{1}_{\{i > 0\}} + j \cdot \eta} \cdot \frac{i + 1}{\mu} \\ & + \frac{\eta \cdot (j - 1)}{\lambda \cdot ((1 - \theta)q_I(i) + \theta q_U) + \mu \cdot \mathbb{1}_{\{i > 0\}} + j \cdot \eta} \cdot [E(i + 1, j - 1) \cdot \mathbb{1}_{\{i < n_e\}} + E(i, j - 1) \cdot \mathbb{1}_{\{i = n_e\}}] \\ & + \frac{\lambda \cdot ((1 - \theta)q_I(i) + \theta q_U)}{\lambda \cdot ((1 - \theta)q_I(i) + \theta q_U) + \mu \cdot \mathbb{1}_{\{i > 0\}} + j \cdot \eta} \cdot E(i, j + 1) \\ & + \frac{\mu \cdot \mathbb{1}_{\{i > 0\}}}{\lambda \cdot ((1 - \theta)q_I(i) + \theta q_U) + \mu \cdot \mathbb{1}_{\{i > 0\}} + j \cdot \eta} \cdot E(i - 1, j) \end{aligned} \quad (6.6)$$

Cette expression est la somme de cinq termes. Le premier représente le temps prévu jusqu'à ce que l'un des trois événements possibles se produise :

- un nouveau client rejoint la file d'attente de déplacement,
- un client arrive dans la file d'attente de service,
- le service est achevé.

Le deuxième terme représente le temps de séjour attendu si le client est le prochain à rejoindre la file d'attente. Le troisième terme représente le temps de séjour prévu si un des autres clients qui sont en route, arrive dans la file d'attente de service. Dans ce cas, il y a deux possibilité : soit le client qui arrive trouve la file d'attente de service complètement saturée et donc il va quitter le système, soit ce client trouve que la file d'attente de service n'est pas saturée et donc il la rejoint. Le quatrième terme représente le temps de séjour prévu si un nouveau client rejoint la file d'attente de

déplacement. Le dernier terme représente le temps de séjour prévu en cas où le service est achevé.

Enfin, l'utilité attendue d'un client informé, notée par $V(q, i)$ et sachant que q est le vecteur de probabilités de déplacement dans le système et i est l'état observé de la file d'attente de service avant le voyage, est donnée par l'expression suivante :

$$V(q, i) = R(1 - r(i)) - C \sum_{j=0}^{\infty} E(i, j + 1) \frac{\pi_{i, j}}{\pi_i}, \quad (6.7)$$

où $i \in \{0, \dots, n_e\}$ et $\pi_i = \sum_{j=0}^{\infty} \pi_{i, j}$. Cette utilité est associée à l'obtention du service qui a une valeur R , qui est conditionnée à une file d'attente qui n'est pas complètement chargée c'est à dire avec une probabilité $(1 - r(i))$, moins le coût de séjour prévu y compris le déplacement et l'attente dans le système de service $C \sum_{j=0}^{\infty} E(i, j + 1) \frac{\pi_{i, j}}{\pi_i}$.

6.5 Stratégie d'équilibre dans un système partiellement observable

Dans cette section, nous nous intéressons à la stratégie d'équilibre dans ce système de file d'attente où les clients sont stratégiques et hétérogènes (informés et non-informés) et doivent se déplacer pour atteindre le système de service. La stratégie d'un client informé qui se trouve loin du système de service est un vecteur de probabilités $q_I(i)$ avec $i \in \{0, \dots, n_e\}$. Cette stratégie dépend de l'occupation i de la file d'attente de service au moment du départ de l'utilisateur. D'autre part, la stratégie d'un client non-informé est donnée par une probabilité q_U de voyager pour rejoindre la file d'attente de service, sans avoir connaissance de l'état de la file de service à son départ. Le calcul des meilleures réponses dépendent des stratégies de tous les usagers, informés et non-informés. Nous supposons que chaque type de client choisit sa meilleure stratégie lorsque tous les autres clients dans le système ont déjà choisit leurs meilleurs stratégies. Nous utilisons par la suite, les deux meilleurs réponses des clients pour calculer l'équilibre dans le système.

6.5.1 Meilleure réponse pour les clients non-informés

Considérant que tous les usagers informés utilisent leur stratégie de meilleure réponse q_I^* , la stratégie de la meilleure réponse pour un usager non-informé notée

par q_U^* , satisfait la condition suivante :

$$q_U^* = \begin{cases} 1 & \text{if } U(1, q_I^*(0), \dots, q_I^*(n_e)) > 0, \\ q_U \in]0,1[& \text{if } U(q_U, q_I^*(0), \dots, q_I^*(n_e)) = 0, \\ 0 & \text{if } U(0, q_I^*(0), \dots, q_I^*(n_e)) < 0. \end{cases}$$

Un client non-informé aura intérêt à se déplacer si son utilité attendue U est strictement positive. Il ne se déplacera pas si son utilité attendue est strictement négative. Enfin, si son utilité attendue est nulle, il se déplacera avec une certaine probabilité strictement comprise entre 0 et 1.

6.5.2 Meilleure réponse pour les clients informés

De même que précédemment, lorsque tous les usagers non-informés utilisent leur stratégie de meilleure réponse q_U^* , la stratégie de la meilleure réponse pour un usager informé notée par q_I^* et pour tout $i \in \{0, \dots, n_e\}$, satisfait la condition suivante :

$$\forall i \in \{0, \dots, n_e\}, q_I^* = \begin{cases} 1 & \text{if } V(q_U^*, 1, i) > 0, \\ q_I \in]0,1[& \text{if } V(q_U^*, q_I, i) = 0, \\ 0 & \text{if } V(q_U^*, 0, i) < 0. \end{cases}$$

Un client informé aura intérêt à se déplacer si son utilité attendue V est strictement positive. Il décide de ne déplacera pas si son utilité attendue est strictement négative. Enfin, si son utilité attendue est nulle, il se déplacera avec une certaine probabilité strictement comprise entre 0 et 1.

6.5.3 Stratégie d'équilibre

Nous définissons la stratégie d'équilibre dans le système q^* pour tout $i \in \{0, \dots, n_e\}$ comme suit :

$$q^* = (q_U^*, q_I^*(0), \dots, q_I^*(n_e)) \quad (6.8)$$

Sachant que q_U^* est l'équilibre pour un client non-informé et le vecteur $(q_I^*(0), \dots, q_I^*(n_e))$ est l'équilibre d'un client informé par l'état i du système, tel que $i \in \{0, \dots, n_e\}$.

La stratégie d'équilibre est compliquée à calculer explicitement mais nous pouvons utiliser les stratégies de meilleures réponses et leur condition décrite précédemment pour proposer un algorithme qui permet, à l'aide d'une boucle itérative,

de déterminer la stratégie d'équilibre. Pour savoir si un équilibre existe ou non dans le modèle étudié, nous avons proposé l'algorithme 3.

Pour nos analyses numériques, nous avons utilisé une variable très petite proche de zéro, $\epsilon = 0.0005$. Nous proposons dans l'algorithme 3 le calcul de la stratégie d'équilibre dans ce modèle de file d'attente partiellement observable à partir de la fonction *Meilleur Reponse* qui retourne la meilleure réponse pour chaque client. Si l'utilité espérée de voyager lors de l'observation des clients est plus grande que ϵ , la meilleure réponse consiste à voyager ; s'il est inférieur à $-\epsilon$, la meilleure réponse est d'éviter de voyager, sinon, le client est indifférent et sa meilleure réponse est considérée comme la probabilité obtenue (q_U s'il est un client non-informé ou $q_I(i)$ s'il est un client informé par l'état i du système sachant que $i \in \{, \dots, n_e\}$).

Algorithme 3 Stratégie d'équilibre $q^* = \{q_U^*, q_I^*(0), \dots, q_I^*(n_e)\}$

Initialisation : $q_u := rand(1)$, $q_i = rand(1, n_e + 1)$, $\epsilon = 0.00005$

$q = [q_U, q_I]$

$MR = MeilleurReponse(N, \lambda, \mu, \eta, R, C, \epsilon, q_U, q_I, \theta)$

$Distance = pdist(MR, q)$

$t = 0$

while ($Distance > \epsilon$) **do**

$t = t + 1$

$\gamma = rand(1)$

if ($\gamma > 0.5$) **then**

$q^* = (1 - \gamma) \times MR + \gamma \times q$

else

$q^* = \gamma \times MR + (1 - \gamma) \times q$

end if

$q_U = q^*(1)$

$q_I = q^*(2 : n_e + 2)$

$MR = MR(N, \lambda, \mu, \eta, R, C, \epsilon, q_U, q_I, \theta)$

$q = q^*$

$Distance = pdist(MR, q)$

if ($t > 1000000$) **then**

 Affiche ("Erreur")

end if

end while

Retourner q^*

À partir d'une stratégie initiale arbitraire $q = [q_U, q_I(0), \dots, q_I(n_e)]$, nous calculons la meilleure réponse pour les deux types des clients, informés et non-informés, lorsque tous les autres jouent clients choisissent la stratégie q . Ensuite, nous comparons la stratégie de meilleure réponse avec la stratégie initiale, en calculant la

distance entre elles. Nous répétons ces itérations, jusqu'à ce que la distance entre les deux stratégies est inférieure à ϵ . Si ce critère d'arrêt est atteint, nous retournons la stratégie de l'équilibre dans le système. Le critère d'arrêt garantit que la stratégie finale est extrêmement proche de sa meilleure réponse. En d'autres termes, si l'algorithme converge vers une solution, le critère d'arrêt garantit que cette solution est un équilibre.

6.5.4 Bien-être social

Le bien-être social représente l'utilité globale espérée d'un client à l'équilibre. Dans ce modèle, le bien-être social d'un client non-informé est donné par :

$$SW_U = \lambda q_U^* U,$$

et le bien-être social d'un client informé par l'état i du système de service est donné par :

$$SW_I = \lambda \sum_{i=0}^{n_e} q_I^*(i) V(i) \pi_i.$$

Par la suite, le bien-être social dans le système est sous forme d'une somme des deux expressions qui représente respectivement, la proportion des clients non-informés multiplié par leur bien-être social à l'équilibre et la proportion des clients informés multiplié par leur bien-être social à l'équilibre. Son expression est donc donnée par :

$$SW = \theta SW_U + (1 - \theta) SW_I \quad (6.9)$$

6.6 Illustrations numériques

6.6.1 Stratégies d'équilibre

Dans notre analyse numérique, nous normalisons le système en supposant que $\mu = 1$ et $C = 1$. L'observation 10 est une première interprétation sur la stratégie d'équilibre dans le système obtenu par l'algorithme précédent. Cette première observation porte sur l'existence d'un équilibre pour toute combinaison de paramètres du système.

Observation 10. *Pour tout ensemble de paramètres, une stratégie d'équilibre existe dans un système de déplacement partiellement observable.*

[title=]

L'algorithme 3 converge vers un équilibre, pour tout ensemble de paramètres. Des exemples de stratégies d'équilibre de clients non-informés sont présentés dans la Figure 6.8 pour différentes valeurs de θ . L'axe des abscisses représente le taux d'arrivée λ et l'axe des y représente la probabilité de voyager pour un client non-informé.

La Figure 6.8 représente des stratégies d'équilibre pour les clients non-informés en fonction du taux d'arrivée λ et pour différentes valeurs du paramètre θ .

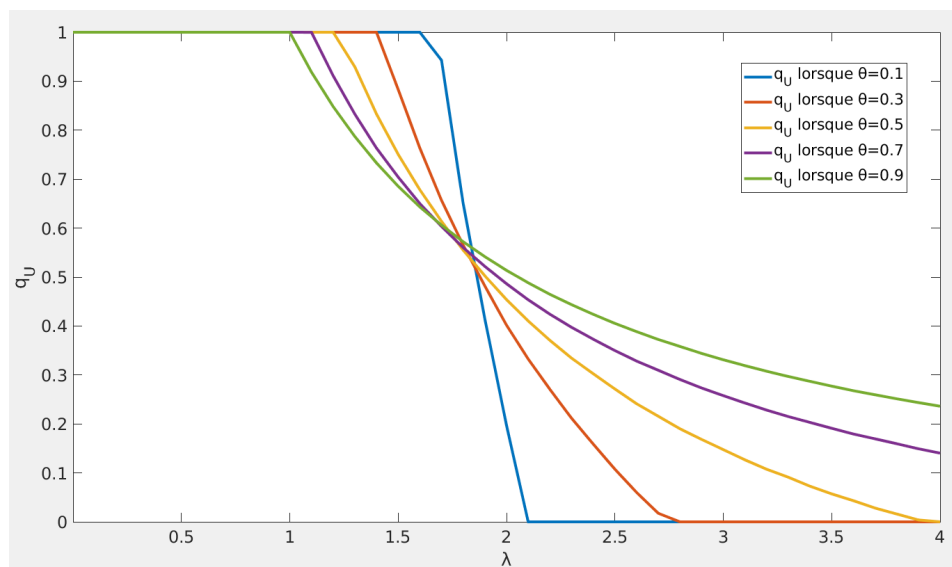


Figure 6.8 – Stratégie à l'équilibre q_U^* des clients non-informés en fonction de λ , pour $\eta = 3$ et $R = 6$.

Observation 11. *Les stratégies d'équilibre pour les clients non-informés diminuent en fonction de la charge dans le système.*

Nous remarquons que les clients non-informés décident de se déplacer ($q_U^* = 1$) lorsque la charge est faible, ils décident de ne se pas déplacer ($q_U^* = 0$) lorsque la charge est élevée. Enfin, ils se divisent entre se déplacer ou non ($q_U^* \in]0,1[$) lorsque la charge est ni faible ni élevée.

Dans la Figure 6.9, nous représentons des exemples de stratégies d'équilibre pour des clients informés en fonction de l'état i du système et pour des différentes

valeurs de θ .

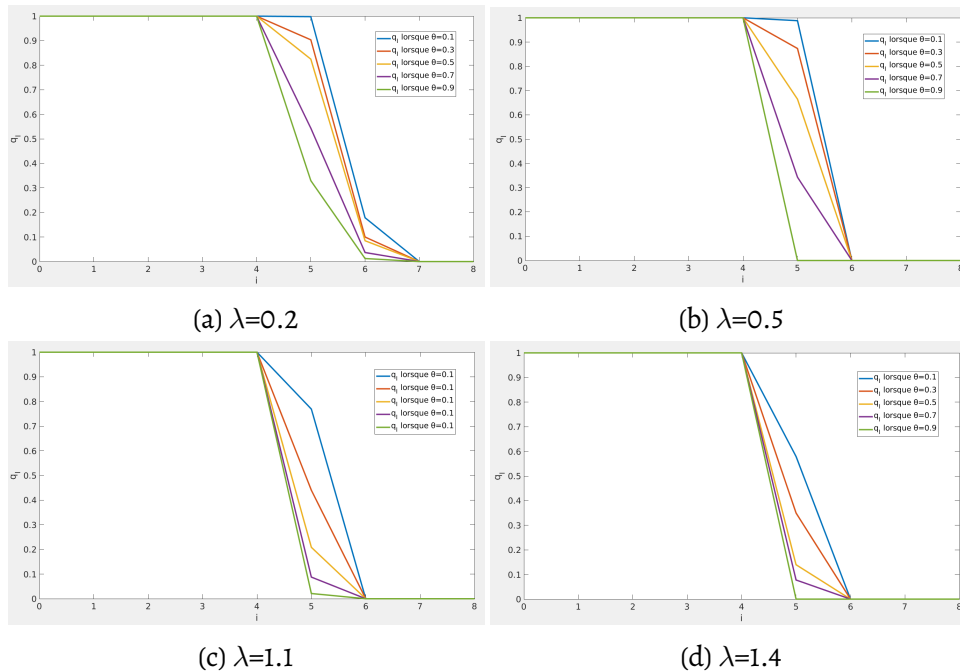


Figure 6.9 – Stratégie à seuil mixte en fonction de la longueur de la file d’attente observée avec $\eta = 0.7$, $\mu = 1$, $R = 8$ et $C = 1$

Observation 12. *Les stratégies d’équilibre pour les clients informés est monotone et décroissant en fonction de l’état de système observé avant le déplacement, quelque soit leur proportion dans la population.*

Observation 12 indique que les clients informés appliquent une stratégie d’équilibre mixte : ils voyagent lorsqu’ils observent une file d’attente courte, ne voyagent pas vers le système lorsqu’ils observent une longue file d’attente. Enfin, une partie des clients se déplacent lorsqu’ils observent une longueur intermédiaire de la file d’attente de service. Ce résultat est similaire à celui d’un système complètement observable étudié dans (HASSIN et GREEN 2020).

Observation 13. *En augmentant la proportion des clients informés dans la population ($1 - \theta$), la stratégie d’équilibre mixte augmente.*

La Figure 6.10 illustre les stratégies d’équilibre des clients informés en fonction de l’état du système i et pour différentes valeurs de λ .

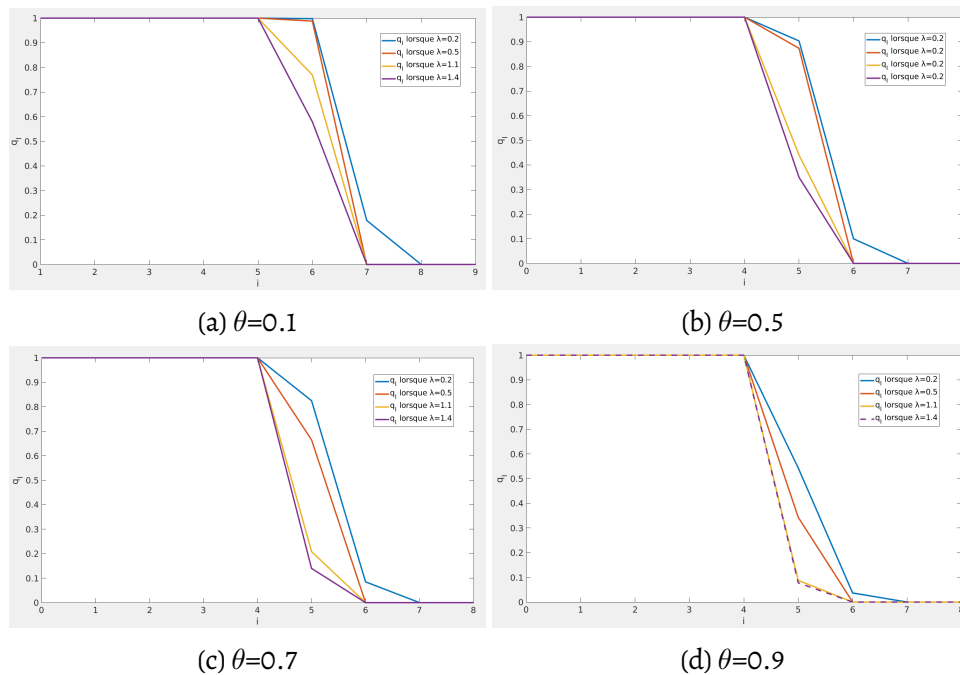


Figure 6.10 – Stratégie à seuil mixte en fonction de la longueur de la file d'attente observée avec $\eta = 0.7$, $\mu = 1$, $R = 8$ et $C = 1$

Observation 14. *Les stratégies d'équilibre pour les clients informés diminuent en fonction la charge dans le système.*

Avec le même raisonnement des clients non-informés, nous remarquons qu'en augmentant le taux d'arrivée λ , la stratégie d'équilibre mixte des clients informés diminue.

6.6.2 Bien-être social

Dans cette section, nous illustrons l'impact de la proportion θ sur le bien-être social du système à l'équilibre. Le résultat principal est donné dans l'observation ci-dessous.

Observation 15. *Pour tout ensemble de paramètres, et lorsque le taux λ augmente, le bien-être social augmente lorsque la proportion des clients non-informés θ dans le système augmente.*

Nous remarquons sur la figure 6.11 que lorsque le taux d'arrivée est faible, tous les clients se déplacent vers la file d'attente et donc, le bien-être social augmente en fonction du taux d'arrivée λ . À une certaine valeur de la congestion, les clients dé-

cident de ne plus voyager vers le système de service, ce qui implique une diminution du bien-être social. Nous remarquons que le bien-être social augmente de nouveau en fonction de λ . Le remarque importante est que le maximum du bien-être social est obtenu pour la même valeur de λ même en modifiant la proportion θ de clients non-informés.

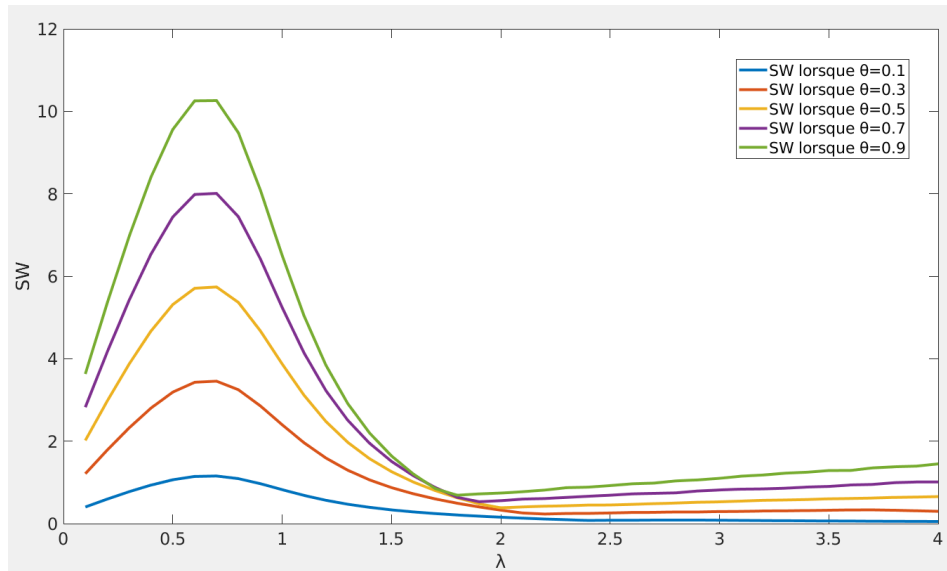


Figure 6.11 – Bien-être social SW en fonction de λ , pour $R = 10$, $\eta = 3$, $\mu = 1$ et $C = 1$.

6.7 Conclusion

Dans ce chapitre, nous avons analysé un système de déplacement avec une observation mixte : les clients doivent se déplacer pour atteindre le système de service et le modèle partiellement observable dans le sens où certains clients sont informés et d'autres sont non-informés par l'information de la taille de file d'attente de service avant le voyage. Nous avons étudié l'impact de cette information sur le comportement des clients et sur les performances du système. Nous avons observé, que pour tout ensemble de paramètre, un équilibre existe dans le système et que le bien-être social est plus important lorsque la proportion des clients informés est assez grande par rapport à la proportion des clients non-informés.

Chapitre 7

Conclusion générale et perspectives

Sommaire

7.1	Conclusion générale	112
7.2	Perspectives	113

7.1 Conclusion générale

Sous le renforcement et la pression de la concurrence, plusieurs systèmes de file d'attente proposent en exclusivité des avantages et des options pour les usagers. Parmi les avantages proposés, certains systèmes donnent la possibilité d'inspecter l'information de l'occupation du système par ses clients. Ils assurent la communication et la coordination de cette information via des interfaces et des applications en temps réel. Cette information permet d'aider les clients à prendre leurs décisions d'une façon optimale. Ceci favorise fortement les systèmes qui proposent ce genre d'option pour le bien-être des clients et la qualité de services proposés.

Dans cette thèse, nous avons utilisé la théorie des jeux pour modéliser et étudier des problèmes de file d'attente stratégiques. Nous nous sommes intéressés, plus précisément, à étudier l'impact de l'information sur le comportement des clients dans différents modèles de file d'attente. En étudiant la prise de décision dans des différents contextes, plusieurs modes de contrôle et de partage d'information ont été proposés. Nos résultats théoriques et analytiques, au cours de cette thèse, permettent de comprendre le fonctionnement de ces systèmes en fonction de l'accessibilité complète, partielle ou même l'absence de cette information. Plusieurs facteurs sont aussi pris en considération dans les études faites tel que le type du système, la discipline utilisée, l'homogénéité (ou non) de la population, etc. Nous avons prouvé l'existence de l'équilibre dans les modèles étudiés et nous avons évalué les performances des systèmes.

Les étapes de modélisation, d'étude et d'analyse sont appliquées à tous les problèmes étudiés. Nous résumons les contributions principales de cette thèse de doctorat :

- Le chapitre 3 a proposé, dans une première partie, une étude d'un système de file d'attente non-observable. Les clients, qui arrivent au système, peuvent choisir entre : obtenir l'information de la longueur de la file avant d'y accéder ou ne pas obtenir cette information et rejoindre directement le système. Nous avons prouvé que ce modèle de file d'attente possède un équilibre unique et nous avons donné son expression explicite. Dans un deuxième temps, une politique de tarification optimale pour maximiser les revenus du fournisseur a été proposée. Nous avons comparé deux systèmes de tarification : facturer l'information de l'état de système et facturer l'accès au système. Nous avons prouvé, dans cette partie, que la facturation de l'accès au système de service

n'est pas toujours la meilleure politique de tarification et que lorsque les clients sont très impatients, la meilleure méthode de tarification est de facturer l'information.

- Dans le chapitre 4, un système de file d'attente avec une politique de partage d'information (PPI) a été proposé. Cette PPI prend en compte l'état du système et révèle l'information de longueur de file d'attente jusqu'à un seuil particulier. Nous avons étudié dans ce chapitre l'impact de la PPI sur les décisions stratégiques des clients et sur le bien-être social attendu. Le principal résultat est que l'application d'une telle PPI maximise le bien-être social dans le système. De plus, nous avons montré que le bien-être social peut être plus élevé que dans un système où l'information est complètement observable.
- Le chapitre 5 a considéré le même modèle étudié dans le chapitre 4 avec une population qui contient différentes classes de clients. Une PPI est donc imposée par le fournisseur de service qui se base sur l'état du système en temps réel. Nous nous sommes intéressés à l'impact de l'information sur les décisions stratégiques des différents types de clients qui arrivent à la file et sur l'équilibre dans le système. Dans le cas de deux classes, nous avons donné les expressions explicites de l'équilibre pour chaque type de clients.
- Dans le chapitre 6, un système de file d'attente avec une population mixte et avec déplacement des clients a été étudié. Dans le modèle étudié, seulement une partie des clients sont informés par l'information sur la taille de la file de service avant le voyage. Nous avons étudié l'impact de cette information sur le comportement et les décisions stratégiques des clients, informés et non informés, et sur les performances du système. Nous avons constaté, que pour tout ensemble de paramètres, un équilibre existe dans le système.

7.2 Perspectives

Les travaux présentés peuvent ouvrir la voie à plusieurs axes de recherches. Nous présentons quelques suggestions pour de futurs travaux.

- L'étude d'autres politiques de partage et de contrôle d'information peut fournir un outil de décision très intéressant à traiter. Cet outil permet d'optimiser les fonctions objectifs dans les systèmes de file d'attente stratégiques.
- Un seul système dans le cas de saturation a été étudié. Les autres modèles vus dans cette thèse restent à traiter dans le cas de saturation ou à forte charge. L'étude de la stabilisation de ces systèmes saturés peut donner des résultats complètement différents et spécifiquement à leurs contextes.

- Une extension des travaux réalisés peut concerner le contexte spécifique multi-classes des clients dans les autres systèmes de file d'attente. Ce contexte est très important à étudier parce que la population est naturellement pas homogène et différentes classes de clients peuvent en exister. L'étude de l'hétérogénéité de la population peut prendre en considération plusieurs paramètres de système, et la performance globale est à étudier. La discipline de service "**priorité**" peut être appliqué dans ce genre de système et par la suite, un changement de la méthodologie et de traitement des ressources peut être induit. Des méta-heuristiques peuvent être proposées pour résoudre ses problèmes complexes et une comparaison entre les systèmes préemptifs et non-préemptifs peut se faire.
- L'impact d'autres paramètres du système sera aussi intéressant à étudier, tel que le taux de service, sur les décisions stratégiques des clients qui se présentent à un système de file d'attente.
- Une implémentation des méthodes et outils d'aide à la décision pour guider les usagers des systèmes de files d'attente à prendre les meilleures décisions et pour aider les fournisseurs de services à maximiser ses revenus. Ces outils doivent assurer la qualité de service et satisfaire les besoins des usagers par la mise à disposition aux meilleurs endroits et avec un meilleur coût.

Liste des figures

1.1	Temps d'attente réel d'un restaurant affiché sur Google	6
2.1	Représentation d'une de file d'attente simple.	10
2.2	Exemple élémentaire de chaîne de Markov	14
2.3	Processus de prise de stratégie dans une file d'attente simple . . .	16
2.4	Exemples d'un comportement ATC et FTC (Figure extraite de (HASSIN et HAVIV 2003))	18
2.5	Meilleure réponse (Figure extraite de (HASSIN et HAVIV 2003)) . .	21
3.1	Situation en temps réel dans les salles d'urgence de la région de Québec	28
3.2	Processus de prise de décision	29
3.3	Représentation en chaîne de Markov de l'évolution du système . .	30
3.4	Revenu en fonction du coût d'accès C_{Acc}	43
3.5	Revenu en fonction du coût d'accès avec $\mu = 2.8, \lambda = 2.2, R = 3$ et $C = 1$	44
3.6	Revenus optimaux en fonction du coût d'attente avec $\mu = 2.8, \lambda = 2.2$ et $R = 10$	45
4.1	Système de file d'attente avec politique de partage d'information. .	49
4.2	Chaîne de Markov associée au modèle complètement observable .	52
4.3	Chaîne de Markov associée au modèle partiellement observable . .	53
4.4	Bien-être social espéré SW en fonction de la PPI u_D pour des taux d'arrivée différents $\lambda \in \{1, 1.2, 1.4, 1.6, 1.8, 2\}$	68
4.5	Bien-être social espéré $SW(n_e - 2)$ en fonction du taux d'arrivée λ pour différentes valeurs de récompense R	69
4.6	Bien-être social espéré en régime surchargé \overline{SW} en fonction de D avec $R = C = 20$	70
5.1	Système de file d'attente hétérogène en appliquant une PPI	73

5.2	Chaîne de Markov associée au modèle complètement observable	75
5.3	Chaîne de Markov associée au modèle partiellement observable (cas $D \in \{0, n_2 - 1\}$)	78
5.4	Chaîne de Markov associée au modèle partiellement observable (cas $D \in \{n_2, n_1 - 2\}$)	80
5.5	Équilibres q_1^* et q_2^* dans un système de file d'attente à deux classes en fonction de D	85
5.6	Nombre moyen des clients dans le système en fonction de D	85
5.7	Temps de séjour dans le système en fonction de D	86
6.1	Temps d'attente publié au site de la protection des douanes et des frontières américaines pour le passage de l'autoroute Blaine	90
6.2	Processus de prise de décisions dans un système de file d'attente avec déplacement	91
6.3	États de transitions autour le couple d'états (i, j)	94
6.4	États de transitions autour le couple d'états $(0, 0)$	95
6.5	États de transitions autour le couple d'états $(0, N)$	96
6.6	États de transitions autour le couple d'états $(n_e, 0)$	96
6.7	États de transitions autour le couple d'états (n_e, N)	97
6.8	Stratégie à l'équilibre q_U^* des clients non-informés en fonction de λ , pour $\eta = 3$ et $R = 6$	106
6.9	Stratégie à seuil mixte en fonction de la longueur de la file d'attente observée avec $\eta = 0.7, \mu = 1, R = 8$ et $C = 1$	107
6.10	Stratégie à seuil mixte en fonction de la longueur de la file d'attente observée avec $\eta = 0.7, \mu = 1, R = 8$ et $C = 1$	108
6.11	Bien-être social SW en fonction de λ , pour $R = 10, \eta = 3, \mu = 1$ et $C = 1$	109

Liste des tableaux

Bibliographie

Ouvrages de référence

- [AAM06] A.A.MARKOV. « The law of large numbers to dependent quantities (in Russian) ». In : *Izvestiia Fiz.-Matem. Obsch. Kazan Univ.* 15 (1906), p. 135-156.
- [AB09] S. R. ASMUSSEN et O. J. BOXMA. « Editorial introduction ». In : *Queueing Systems* (2009).
- [AJ13] E. ALTMAN et T. JIMENEZ. « Admission Control to an M/M/1 Queue with Partial Information ». In : *proceedings of ASMTA* (2013).
- [Alp88] H. ALPERSTEIN. « Optimal pricing for the service facility offering a set of priority prices ». In : *Management Science* 34 (1988), p. 666-671.
- [AY74] I. ADIRI et U. YECHIALI. « Optimal priority purchasing and pricing decisions in nonmonopoly and monopoly queues ». In : *Operations Research* 22 (1974), p. 1051-1066.
- [Bal91] K.R. BALACHANDRAN. « Incentive and regulation in queues ». In : *Lecture Notes in Economics and Mathematical Systems* 370 (1991), p. 162-176.
- [BE17] O. BOUNTALI et A. ECONOMOU. « Equilibrium joining strategies in batch service queueing systems ». In : *European Journal of Operational Research* (2017).
- [Ber83] J. BERTRAND. « Théorie mathématique de la richesse sociale ». In : *Journal des Savants* (1883), p. 499-508.
- [BR11] Nicole BAUERLEET et Ulrich RIEDER. « Markov decision processes with applications to finance ». In : *Universitext, Springer, Heidelberg* (2011).
- [BZ97] W. Solomon B. CALVERT et I. ZIEDINS. « Braess's paradox in a queueing network with state dependent routing ». In : 34 (1997), p. 134-154.
- [CF04] H. CHEN et M. FRANK. « Monopoly pricing when customers queue ». In : *Working Paper, Faculty of Commerce and Business Administration, University of British Columbia.* (2004).
- [CM80] Edward G. Jr. COFFMAN et Isi MITRANI. « A characterization of waiting time performance realizable by single-server queues ». In : *Operations Research* 28.3 (1980), p. 810-821.
- [Cob54] Alan COBHAM. « Priority Assignment in Waiting Line Problems ». In : *Journal of the Operations Research Society of America* (1954), p. 70-76.

- [Cou38] A. COURNOT. « Recherches sur les Principes mathématiques de la théorie des richesses ». In : (1838).
- [DH08] J. Thompson D. GROSS J. Shortle et C. HARRIS. *Fundamentals of Queueing Theory*. 2008.
- [EG15] A. ECONOMOU et M. GRIGORIOU. « Strategic balking behavior in a queueing system with a mixed observation structure ». In : *proceedings of Stochastic Models of Manufacturing and Service Operations* (2015).
- [EH75] N. EDELSON et D. HILDEBRAND. « Congestion Tolls for Poisson queueing processes ». In : *Econometrica* 43 (1975), p. 81-92.
- [Erl09] A.K. ERLANG. « The Theory of Probabilities and Telephone Conversations ». In : *Nyt Tidsskrift for Matematik* 20 (1909).
- [ES15] Y. Hayel E. SIMHON et D. STAROBINSKI. « Optimal information disclosure policies in strategic queueing games ». In : *Operations Research Letters* 44 (2015), p. 109-113.
- [FG88] Awi FEDERGRUEN et Harry GROENEVELT. « Characterization and optimization of achievable performance in general queueing systems ». In : *Operations Research* 36.5 (1988), p. 733-741.
- [GG11] A. Bassamboo G. ALLON et I. GURVICH. « We will be right with you : managing customer expectations with vague promises and cheap talk ». In : *Operations Research* 59.6 (2011), p. 1382-1394.
- [GT05] H. de Meer G. BOLCH S. Greiner et K.S. TRIVEDI. « Queueing networks and Markov chains ». In : *Wiley-Interscience* (2005).
- [Has16] R. HASSIN. « Rational Queueing ». In : *CRC Press* (2016).
- [HG20] R. HASSIN et R. Roet GREEN. « On queue-length information when customers travel to a queue ». In : *Operations Management* (2020).
- [HH03] R. HASSIN et M. HAVIV. « To Queue or not to Queue : Equilibrium Behavior in Queueing ». In : *Kluwer Academic Publishers, Norwell, Massachusetts* (2003).
- [HH86] R. HASSIN et M. HENIG. « Control of arrivals and departures in a state-dependent input-output system ». In : *Operations Research Letters* 5.1 (1986), p. 33-36.
- [HH94] R. HASSIN et M. HAVIV. « Equilibrium strategies and the value of information in a two line queueing system with threshold jockeying ». In : *Communications in Statistics - Stochastic Models* 10 (1994), p. 415-436.
- [HH97] R. HASSIN et M. HAVIV. « Equilibrium threshold strategies : the case of queues with priorities ». In : *Operations Research* 45.6 (1997), p. 966-973.
- [HR17] R. HASSIN et R. ROET-GREEN. « The Impact of Inspection Cost on Equilibrium, Revenue, and Social Welfare in a Single-Server Queue ». In : *Operations Research* 65.3 (2017).
- [Ibr18] R. IBRAHIM. « Sharing delay information in service systems : a literature survey ». In : *Queueing Systems* 89 (2018), p. 49-79.

- [JS80] S.G. JOHANSEN et Jr. S. STIDHAM. « Control of arrivals to a stochastic input-output system ». In : *newblock Operations Research* (1980), p. 972-999.
- [Ker07] T. KERNANE. « Stabilité de modèles de files d'attente ». In : *Université des Sciences et de la Technologie USTHB*, (2007).
- [KG51] KENDALL et David G. « Some Problems in the Theory of Queues ». In : *Journal of the Royal Statistical Society* 13 (1951), p. 151-185.
- [Kin09] J. F. C. KINGMAN. « The first Erlang century—and the next ». In : *Queueing Systems* (2009).
- [KK17] B. KIM et J. KIM. « Optimal disclosure policies in a strategic queueing model ». In : *Operations Research Letters* 45 (2017), p. 181-186.
- [Kle75] L. KLEINROCK. *Queueing systems. Volume 1 : THEORY*. 1975.
- [Kle76] L. KLEINROCK. *Queueing systems. Volume 2 : Computer applications*. 1976.
- [MW18a] Y. Li M. HU et J. WANG. « Efficient Ignorance : Information Heterogeneity in a Queue ». In : *Management Science* 64 (2018), p. 2650-2671.
- [MW18b] X. Chai M. ZHOU L. Liu et Z. WANG. « Equilibrium strategies in a constant retrial queue with setup time and the N-policy ». In : *Communications in Statistics - Theory and Methods* (2018).
- [MW90] H. MENDELSON et S. WHANG. « Optimal incentive-compatible priority pricing for the M/M/1 queue ». In : *Operations Research* 38 (1990), p. 870-883.
- [Nao69] P. NAOR. « The regulation of queue size by levying tolls ». In : *Econometrica* 37.2 (1969), p. 15-24.
- [Nas50] J. NASH. « Equilibrium points in n-person games ». In : *Proceedings of the National Academy of Sciences* 36.1 (1950), p. 48-49.
- [Nas51] J. NASH. « Non-Cooperative Games ». In : *The Annals of Mathematics* 54.2 (1951), p. 286-295.
- [NH19] T. NACEUR et Y. HAYEL. « Revenue maximization with access and information pricing schemes in a partially-observable queueing game ». In : *Value Tools* (2019).
- [NH20] T. NACEUR et Y. HAYEL. « Deterministic State based Information Disclosure Policies and Social Welfare Maximization in Strategic Queueing Systems ». In : *Queueing Systems, special issue on Strategic Queueing* (2020).
- [NM44] J. von NEUMANN et O. MORGENSTERN. « Theory of games and economic behavior ». In : *Princeton University Press* (1944).
- [RF09] P. Justo R. HASSIN et R. Fern ande FRANCISCO. « The use of relative priorities in optimizing the performance of a queueing system ». In : *Operations Research* 193 (2009), p. 476-483.
- [RH02] R. HASSIN et M. HAVIV. « Nash equilibrium and subgame perfection : the case of observable queues ». In : *Annals of Operations Research* 113 (2002), p. 15-26.

- [RY20] R. ROET-GREEN et Y. YUAN. « Information visibility in omnichannel queues ». In : *Available at SSRN: <https://ssrn.com/abstract=3485810> or <http://dx.doi.org/10.2139/ssrn.3485810>* (2020).
- [Saa61] Thomas L. SAATY. « Elements of Queueing Theory, with Applications ». In : *Ann. Math. Statist.* 34.4 (1961), p. 1610-1612.
- [Smi82] M. SMITH. « Evolution and Game Theory ». In : *Cambridge University Press* (1982).
- [Sti85] S. STIDHAM. « Optimal control of admissions to a queueing system ». In : *IEEE Transactions on Automated Control* 30 (1985), p. 705-713.
- [SY92] J. George SHANTHIKUMAR et David D. YAO. « Multiclass queueing systems : polymatroidal structure and optimal scheduling control ». In : *Operations Research* 40.2 (1992), p. 293-299.
- [Ver02] E. VERHOEF. « Second-best congestion pricing in general networks. Heuristic algorithms for finding second-best optimal toll levels and toll points ». In : *Transportation research part B : Methodological* 36.8 (2002).
- [Y D19] S. Leonardos Y. DIMITRAKOPOULOS A. Economou. « Strategic Customers behavior in a queueing system with alternating information structure ». In : *Arxiv* (2019).

Systèmes de files d'attente stratégiques avec information contrôlée

par
Tesnim Naceur

Résumé :

FACE à des systèmes de file d'attente, les clients prennent des décisions stratégiques afin de rejoindre ou non ces systèmes d'une manière optimale. Un nouvel aspect intéressant est apparu et étudié ces dernières années : l'impact de l'information de la longueur de la file sur les décisions stratégiques des clients, sur l'équilibre et les performances du système. Les clients ne sont pas tous semblables face aux informations fournies par le système et cette hétérogénéité impacte également les équilibres et performances. Dans certains cas, le fournisseur de service peut avoir un intérêt à divulguer l'information sur l'état du système à certains clients

Mots-clés : Système de file d'attente, théorie de jeux, système Markovien, usagers stratégiques, information contrôlée.

Abstract :

FACED to queuing systems, customers can make their strategic decisions in order to join or not these systems. An interesting new aspect has emerged and studied in recent years, which is about the impact of current queue-length information on strategic decisions of customers, on the equilibrium and the performance of the system. Customers are not necessarily homogeneous in their behavior and their access to the information, which implies different equilibrium and performances solutions. In some cases, service provider may have an interest to give to customers the system state information and withholding it to others in order to optimize certain objectives. In other cases, ob-

et même de la cacher à d'autres afin d'optimiser certains objectifs. Dans d'autres cas, l'obtention de l'information est déterminée stratégiquement par les clients qui décident alors de l'inspecter ou non selon leurs contraintes et leurs souhaits. La motivation principale de cette thèse est d'étudier l'impact de l'information dans des systèmes d'attente avec usagers stratégiques et information contrôlée. Nos contributions permettent de déterminer l'équilibre et d'optimiser les performances des systèmes en fonction de l'accessibilité de l'information. Des solutions théoriques et analytiques ont été proposées pour résoudre les problèmes étudiés.

taining the information is mainly the choice of customers and therefore they have to decide to inspect or to collect the information or not, according to their constraints and their wishes. The main motivation for this thesis is to study the impact of the queue length information on the strategic decisions of customers and to analyze the performance of such strategic queuing systems with controlled information. Our contributions allow to determine the equilibrium and optimize the performance of the systems according to the queue length information. Theoretical and analytical solutions have been proposed to solve the studied problems.

Keywords : Queuing system, Game theory, Markovian systems, strategic customers, controlled information.

