



# *Remerciement*

Nous tenons tout d'abord à remercier Dieu le tout puissant, qui nous a donné la force et la patience d'accomplir ce modeste travail.

Je tiens à remercier tout particulièrement mon encadreur, *Dr. Ouahiba benzatout*, pour ses conseils, sa grande disponibilité et sa générosité. La pertinence de ses questions et de ses remarques ont toujours su me motiver et me diriger.

Qu'elle trouve ici le témoignage de ma profonde gratitude.

Je voudrais également remercier tous les membres de jury, *Dr. S. Rahmani* de m'avoir honorée en acceptant de présider le jury, *Pr. A. Kandouci* et *Dr. M. Kadi* d'avoir accepté d'examiner ce modeste travail, merci pour toutes leurs remarques et critiques.

J'exprime mes remerciements à tous mes enseignants du département de Mathématiques qui m'ont initié aux valeurs authentiques, en signe d'un profond respect et d'un profond amour, ainsi que le personnel de l'administration.

Je remercie tous les membres de laboratoire des Modèles Stochastique, Statistique et Application pour leur accueil et leur sympathie.

Je remercie chaleureusement toute *ma famille*, qui m'a soutenu, encouragé et poussé durant toutes mes années d'étude.

Tous ceux qui m'ont aidé ou soutenu de toute manière que ce soit.

*Merci à tous*

# *Dédicaces*

**Je dédie ce travail :**

Mes chère parents que dieu me les garde et qui m'ont toujours encouragé et m'aidé à terminer mes études.

A mes chères soeurs et mes chers frères.

Une spéciale dédicace à mon adorable amie *Benkhira Fadila* qui m'a bien aidé le long de mon parcours universitaire et bien sur sans oublier mon adorable amie *Fatima*.

A tous les membres de ma famille, petits et grands, veuillez trouvez dans ce modeste travail l'expression de mon affection.

A tous mes collègues de la promotion de Master A.S.S.P.A 2019.

A toutes les personnes qui m'aiment.

# Table des matières

<b>Introduction</b>	<b>6</b>
<b>1 Définitions et notations</b>	<b>9</b>
1.1 La file d'attente avec clients impatients . . . . .	9
1.2 Disciplines de service temps-réel . . . . .	13
1.3 Variables remarquables et processus de performances . . . . .	14
<b>2 Problématique et résultats existants</b>	<b>18</b>
2.1 La file FCFS : résultats et méthodes existants . . . . .	19
2.2 Comparaison des disciplines de service . . . . .	22
2.3 La file <i>EDF</i> : problématique probabiliste . . . . .	23
2.3.1 Description du système . . . . .	23
2.3.2 Calcul numérique des performances d' <i>EDF</i> . . . . .	24
2.3.3 Le descripteur à valeurs mesures ponctuelles . . . . .	25
<b>3 Stabilité</b>	<b>28</b>
3.1 Régénérativité dans le cas $GI/GI/1/1 + GI$ . . . . .	29
3.1.1 Comparaison avec la file " délai pur " . . . . .	29
3.1.2 Résultat de régénérativité . . . . .	30
3.2 Stationnarité pour la file $G/G/1/1 + G$ . . . . .	31
3.2.1 Un résultat de stationnarité . . . . .	31
3.2.2 Application à la file $G/G/1/1 + G$ . . . . .	39
<b>4 Probabilité de perte</b>	<b>46</b>
4.1 Deux encadrements . . . . .	47
4.1.1 Encadrement à forte charge . . . . .	47
4.1.2 Cas $GI/M/1/1 + D/FIFO(EDF)$ . . . . .	49
4.2 Comparaison entre disciplines de service . . . . .	53

---

4.2.1	Cas général $G/M/1/1 + G$ . . . . .	53
4.2.2	Cas $M/M/1/1 + G$ . . . . .	63
	<b>Bibliographie</b>	<b>68</b>

# *Introduction*

Dans ce travail nous étudions le modèle probabiliste *de la file d'attente avec clients impatientes*. Le modèle classique de la file d'attente consiste en un système dans lequel des serveurs sont soumis à un flux de requêtes qu'ils doivent traiter. Il a un grand nombre d'applications dans les réseaux de télécommunications, dans les réseaux informatiques, les analyses de trafic routier ou même dans de " vraies " files d'attente, au magasin, au cinéma... Il permet de répondre à des questions de temps de traitement, de structuration de réseaux ou de dimensionnement.

Le développement du temps-réel est aujourd'hui une préoccupation majeure dans les réseaux de télécommunication. Dans le contexte multimédia notamment, on cherche à transmettre les données en un temps très bref. Des flux de données de types parfois très différents doivent cohabiter, l'intégrité de ces flux doit être respectée et les réseaux qui les transmettent doivent être à la fois flexibles et rapides. Toute donnée doit alors avoir une " durée de vie " très limitée dans le système, puisque son traitement doit être instantané.

Pour rendre compte de la contrainte du temps-réel dans les réseaux, on doit enrichir le modèle classique de la file d'attente par un nouveau paramètre, *le délai* des requêtes qui entrent dans la file. On considère qu'une requête est perdue dès qu'elle dépasse ce délai sans avoir commencé son traitement. En termes idéalistes, on parle *de files d'attente avec clients impatientes*. Les clients ont une patience pour entrer en service, au delà de laquelle ils choisissent de quitter la file. Ce modèle a initialement été construit pour décrire les réseaux téléphoniques où les clients mis en attente raccrochaient au bout d'un certain temps si leur appel n'était pas pris en compte. On comprend qu'il est pertinent, plus largement, pour décrire les réseaux dits " temps-réel ".

On représente usuellement le comportement des files d'attente par des processus stochastiques qui les décrivent de manière exhaustive. On cherche alors à savoir à quelle condition ces processus atteignent le régime stationnaire et leurs comportements asymptotiques permettent souvent de calculer les " performances " des systèmes (temps de réponse, probabilité de perte pour des files à capacité limitée...).

Dans le cas particulier de la file d'attente avec clients impatientes, le modèle probabiliste comporte un paramètre supplémentaire avec le délai des clients. La première question sur ce nouveau

modèle est de savoir si les processus qui décrivent habituellement les files d'attente classiques y sont adaptés et sinon, de les représenter différemment. On cherchera ensuite à répondre aux questions suivantes :

- **stabilité** : à quelle condition sur ses paramètres la file d'attente atteint-elle un régime stationnaire ?
- **calcul des performances** : comment calculer les différents indicateurs des performances du système, au premier rang desquels, la probabilité de perte d'un client en fonction des paramètres du système ?
- **dimensionnement** : comment mieux dimensionner la file (nombre de serveurs, vitesse d'exécution) pour diminuer cette perte ?
- **choix de la discipline de service** : suivant quelle discipline servir les clients pour améliorer ces performances et en particulier, pour en perdre le moins possible ?

Le chapitre 1 est dévolu à la description du modèle des files d'attente avec clients impatient. Nous y introduisons les principales définitions et notations que nous utilisons tout au long de cette travail. En particulier, nous définissons les processus stochastiques décrivant le système et les disciplines de services que nous étudions.

Le chapitre 2 est consacré à la problématique d'ensemble et aux résultats et méthodes existants sur la stabilité et les performances des files d'attente avec clients impatient. Nous commençons par rappeler les résultats de stabilité, de congestion à l'état stationnaire, de temps d'attente pour des files d'attente dirigées par la discipline FCFS et rappelons les formules analytiques existantes pour le calcul de la probabilité de perte et du temps d'attente des clients.

Ensuite, nous mettons l'accent sur la comparaison des performances entre les différentes disciplines de services et en particulier sur l'optimalité de la discipline EDF (Earliest Deadline First).

Après avoir pointé l'impossibilité d'appliquer la méthodologie adaptée aux files FCFS dans le cas où la discipline est EDF, nous décrivons une méthode originale existant dans la littérature pour décrire l'évolution d'un système soumis à EDF, mais sans perte de clients.

Tous les chapitres suivants sont dévolus à nos contributions.

Dans le chapitre 3, nous complétons les résultats de stabilité existants et décrits dans le chapitre 2 pour une discipline de service quelconque et dans un cadre plus général pour les lois des paramètres du système. Nous mettons en outre en évidence une méthode ergodique adaptée du

calcul de Palm, pour déterminer à quelle condition le système atteint le régime stationnaire.

Dans le chapitre 4, nous nous intéressons au calcul de la probabilité de perte et à la comparaison entre les disciplines de service de ce point de vue. Nous donnons deux encadrements pour la probabilité de perte, dans le cas  $M/G/1$  à forte charge avec délais quelconques et dans le cas  $G/M/1$  avec délais déterministes. Nous donnons ensuite une nouvelle preuve de l'optimalité de la discipline EDF pour la comparaison stochastique sur la perte des clients, montrons l'optimalité d'EDF en terme de probabilité de perte, et donnons un majorant du gain d'EDF pour la probabilité de perte.

Nous prouvons par ailleurs que la discipline de service LDF (Latest Deadline First) est la " moins bonne " discipline de service suivant ces critères.

# Chapitre 1

## Définitions et notations

Nous introduisons ici les définitions que nous allons utiliser tout au long de ce travail. Nous définissons tout d'abord en partie (1.1), à partir de la définition générale d'une file d'attente, la file d'attente avec clients impatientes. Nous définissons en partie (1.2) la discipline de service *EDF*, qui jouera un rôle central par la suite. Nous évoquons enfin en partie (1.3) les différentes variables et processus de performances, qui permettront de décrire le comportement des différents modèles étudiés.

### 1.1 La file d'attente avec clients impatientes

Nous rappelons ici la définition générale du modèle probabiliste de la file d'attente, en introduisant les processus de base aux quelles nous ferons référence tout au long de ce travail.

#### **File d'attente**

Soient des clients qui entrent dans un magasin à des instants aléatoires, en demandant un service de durée aléatoire. Ces clients attendent dans une file d'attente que l'un des serveurs du magasin soit libre pour pouvoir être servis.

On se place sur un espace probabilisé  $(\Omega, \mathcal{F}, \mathbb{P})$ . On appelle  $C_1, C_2, \dots$  les clients dans l'ordre de leurs arrivées, et on note pour tout  $n$ ,  $T_n$  l'instant d'arrivées de  $C_n$ , et  $\sigma_n$ , la durée du service demandé par  $C_n$ .

On suppose que **p.s.** les arrivées sont simples (un seul client entre à la fois), la suite des instants d'arrivées  $(T_n, n \in \mathbb{N}^*)$  est donc une suite de variables aléatoires **p.s.** strictement croissante. On note  $(N_t, t \geq 0)$ , le processus stochastique qui compte, à chaque instant  $t$ , le nombre de clients

entrés dans le système jusqu'à  $t$ . Autrement dit, pour tout  $t$ ,

$$N_t = \sum_{n \in \mathbb{N}^*} \mathbb{1}_{\{T_n \leq t\}}.$$

On notera par ailleurs  $U_1 = T_1$  et pour tout  $n \geq 1$ , la  $(n + 1)$ -ième interarrivée

$$U_{n+1} = T_{n+1} - T_n.$$

On dit alors que  $(N_t, t \geq 0)$  est le processus des arrivées,  $(U_n, n \in \mathbb{N}^*)$ , est la suite des interarrivées et  $(\sigma_n, n \in \mathbb{N}^*)$  est la suite des temps de services. On suppose par ailleurs que les suites des interarrivées et des temps de services sont stationnaires et on note alors  $U$ , une variable aléatoire (v.a.) générique ayant pour loi la loi des interarrivées et  $\sigma$ , une v.a. ayant pour loi la loi des temps de services. On suppose de plus que ces v.a. sont intégrables :

$$\mathbb{E}[U] < \infty \text{ et } \mathbb{E}[\sigma] < \infty,$$

on note en outre

$$\lambda = (\mathbb{E}[U])^{-1} \text{ et } \mu = (\mathbb{E}[\sigma])^{-1}$$

et on définit la charge  $\rho$  du système par :

$$\rho = \frac{\lambda}{\mu} = \frac{\mathbb{E}[\sigma]}{\mathbb{E}[U]}.$$

La file d'attente correspondante est donc l'objet probabiliste entièrement déterminé par ce processus des entrées, par cette suite des temps de service, et par les propriétés du magasin. Nommément, ces propriétés sont le nombre de serveurs, la capacité de la file d'attente (ou buffer), c'est à dire, le nombre de clients qui peuvent être mis en même temps en attente, et sa discipline de service, i.e. le critère suivant lequel les serveurs choisissent le client prioritaire à servir lorsqu'ils sont disponibles et qu'il y a des clients en attente. Les disciplines de service les plus usuelles sont :

- *FCFS* (First came, First served) si on sert toujours le plus ancien client arrivé. En particulier, si il n'y a qu'un serveur, la discipline revient à *FIFO* (First in, First out), puisqu'alors un client entré avant un autre sort forcément avant lui sous *FCFS*,
- *LCFS* (Last came, First served) si on sert toujours le dernier client arrivé. En particulier, si il y a un serveur, la discipline revient à *LIFO* (Last in, First out),
- *RANDOM* si on sert les clients au hasard suivant un tirage uniforme sur la population de la file d'attente,

- *PS* (Processor Sharing) si un " serveur " sert simultanément tous les clients présents dans le système, à une vitesse divisée par le nombre de clients servis à chaque instant,
- *Files avec priorités* : les clients sont des représentants de différentes classes  $C_1, C_2, \dots$  et le serveur, si il a le choix, sert un client de la classe  $C_i$  avant un client de la classe  $C_j$  si  $i < j$ .

Dans toutes les disciplines de services exceptée *FCFS* et *PS*, un client qui est déjà dans le système peut voir entrer après lui un client qui lui est prioritaire. On dira alors qu'un serveur sert de façon préemptive si il peut interrompre le service d'un client pour commencer à en servir un autre qui lui est prioritaire, non-préemptive sinon.

Les lois de probabilités et les valeurs des paramètres qui caractérisent entièrement une file d'attente sont résumés par la *nomenclature de Kendall*, notation universellement reconnue dans la théorie des files d'attente quelconques :

### Définition 1.1.1

Une file d'attente est entièrement caractérisée par le quintuple  $A/Y/s/s + B/X$  où :

- $A$  représente la loi de la suite des interarrivées ( $U_n, n \in \mathbb{N}^*$ ),
- $Y$  représente la loi de la suite ( $\sigma_n, n \in \mathbb{N}^*$ ) des temps de services demandés par les clients,
- $s$  représente le nombre (éventuellement infini) de serveurs,
- $B$  est la taille (éventuellement infinie) du buffer (nombre de clients maximal qu'il peut contenir),
- $X$  représente la discipline de service.

En particulier, les deux premiers paramètres  $A$  et  $Y$  caractérisant les lois des suites des interarrivées et des temps de services prennent essentiellement les valeurs suivantes :

- $G$  (general) si la suite est stationnaire ergodique
- $GI$  (general independent) si elle est **i.i.d.**,
- $M$  (memoryless) si la suite est **i.i.d.** de loi exponentielle (si  $A = M$ , le processus des entrées est donc un processus de Poisson),
- $D$  (deterministic) si la suite est déterministe et constante.

En général, si aucune ambiguïté n'est possible, on notera  $././s$ , une file avec  $s$  serveurs, de buffer de capacité infinie, régie par la discipline *FCFS*. Par exemple, la file  $M/M/1$  est la file d'attente où les clients entrent suivant un processus de Poisson, en demandant un service de loi exponentielle, avec un serveur, un buffer de taille infinie et traitée en *FIFO*.

## Modèle avec clients impatientes

Pour modéliser le transfert de données en temps-réel, on doit ajouter une contrainte au système en spécifiant que les tâches sont perdues si le temps qu'elles passent dans le système est plus grand qu'un délai qui leur est alloué.

Le modèle que nous considérons est donc une file d'attente où les clients ont un délai (ou patience) pour être servis, au delà duquel ils sortent du magasin sans avoir été satisfaits. Autrement dit, le  $n$ -ième client  $C_n$  entre dans la file d'attente affecté du délai  $D_n$ , qui est une variable aléatoire **p.s.** strictement positive. Si il n'a pas pu atteindre le serveur à la fin de sa patience (**i.e.** à l'instant  $T_n + D_n$ ), il est éliminé du système. Les délais seront donc considérés éliminatoires et jusqu'au début du service.

Par ailleurs, la suite  $(D_n, n \in \mathbb{N}^*)$  sera toujours supposée stationnaire de loi générique la loi de  $D$ . On supposera de plus que

$$\mathbb{E}[D] < \infty$$

et on notera

$$\Gamma = (\mathbb{E}[D])^{-1}.$$

Par ailleurs, si le client  $C_n$  est toujours dans le système à l'instant  $t$ , on appellera son délai résiduel à  $t$   $DR_n(t)$  le temps restant à  $t$  avant l'expiration de son délai. On a alors :

$$DR_n(t) = D_n - (t - T_n).$$

On peut décrire la file d'attente avec clients impatientes avec la nomenclature suivante, introduite par Barrer ([1]) :

### Définition 1.1.2

*Une file d'attente avec clients impatientes est caractérisée par la donnée de  $A/Y/s/s + D/X$ , où :*

- $A, Y$  et  $s$  représentent, comme dans la nomenclature de Kendall, respectivement la loi du processus des arrivées, des temps de services et le nombre de serveurs,
- $D$  représente la loi générique des délais des clients, éliminatoires et jusqu'au début du service,
- $X$  est la discipline de service.

### Remarque 1.1.1

*Si le contraire n'est pas précisé, on supposera toujours que la discipline de service est non-préemptive.*

La définition précédente indique que l'on suppose toujours que le buffer est de taille infinie. Le paramètre aléatoire  $D$  (le délai) remplace alors dans la nomenclature de Kendall classique le

paramètre  $B$  (la taille du buffer) comme facteur limitant la capacité du système. La file d'attente avec clients impatientes est donc un modèle à perte, au même titre par exemple qu'une file à buffer fini  $G/G/s/s + B$  (où un client est perdu si il entre dans le système alors que le buffer est déjà plein).

Il est par ailleurs à noter que dans la littérature, les délais peuvent aussi ne pas être éliminatoires (auquel cas ils ne servent qu'à donner des priorités aux clients). Un client reste alors dans le système après expiration de sa patience, même si il n'a pas atteint le serveur.

Enfin, les délais peuvent aussi être éliminatoires, mais jusqu'à la fin du service : dans ce cas un client est éliminé si il n'a pas terminé son service à l'instant  $T_n + D_n$ . Mais nous nous limiterons dans ce travail aux cas précisés dans la définition (1.1.2)

## 1.2 Disciplines de service temps-réel

Les disciplines de services mentionnées dans la partie précédente sont bien sûr applicable à une file d'attente avec clients impatientes. La dynamique de la file d'attente est alors celle d'une file d'attente classique, modifiée par les éliminations de clients, qui ne sont plus à servir. Ceci étant, aucune d'entre elles ne tient a priori compte des délais des clients. Nous présentons ici deux disciplines de services, définies dans le cas où les clients sont affectés de délais et qui joueront un rôle central dans cette étude : les disciplines  $EDF$  et  $LDF$ .

### Définition 1.2.1

1. La discipline de service d'une file d'attente avec clients impatientes est dite  $EDF$  (*Earliest Deadline First*) non préemptive (notée  $EDF$ ) si à chaque fin de service, le serveur sert le client dont le délai résiduel à cet instant est le plus court. Il sert alors ce client jusqu'à complétion de son service.
2. La discipline de service est dite  $EDF$  préemptive (notée  $EDF_p$ ) si le serveur sert à chaque instant le client dont le délai résiduel à cet instant est le plus court.
3. La discipline est dite  $LDF$  (*Largest Deadline First*) non-préemptive ( $LDF$ ) si le serveur sert à chaque fois qu'il complète un service le client dont le délai résiduel à cet instant est le plus long.
4. La discipline est dite  $LDF$  préemptive ( $LDF_p$ ) si le serveur sert à chaque instant le client dont le délai résiduel à cet instant est le plus long.

Comme nous l'avons précisé dans la remarque précédente, nous n'étudierons en détail que les disciplines de services non-préemptives, notées ici  $EDF$  et  $LDF$ . Nous reviendrons largement

par la suite sur les propriétés de ces disciplines de service et sur les problèmes que pose l'étude des files d'attente correspondantes.

### 1.3 Variables remarquables et processus de performances

Définissons ici les différentes variables aléatoires et processus stochastiques caractérisant une file d'attente avec clients impatientes.

- le temps d'attente proposé  $W_n$  au client  $C_n$  est le temps qu'aura à attendre  $C_n$  avant d'être servi. Ce temps peut être supérieur à  $D_n$ , auquel cas  $C_n$  est perdu (il n'a pas pu atteindre le serveur),
- le temps d'attente  $\mathcal{T}_n^a$  d'un client  $C_n$  est le temps effectivement passé par  $C_n$  dans le buffer :

$$\mathcal{T}_n^a = W_n \mathbb{1}_{\{W_n < D_n\}} + D_n \mathbb{1}_{\{W_n \geq D_n\}},$$

- le temps de séjour  $\mathcal{T}_n^s$  du client  $C_n$  est le temps que passe effectivement  $C_n$  dans le système (buffer + service) :

$$\mathcal{T}_n^s = (W_n + \sigma_n) \mathbb{1}_{\{W_n < D_n\}} + D_n \mathbb{1}_{\{W_n \geq D_n\}}.$$

Pour tout système de files d'attente à perte, la probabilité de perte  $\pi_k$  du client  $C_k$  est la probabilité que ce client soit perdu. Par exemple, pour une file d'attente classique à buffer fini  $M/M/1/1+B$ , cette probabilité serait donnée par la probabilité qu'il y ait  $1+B$  clients dans le système à l'entrée de  $C_k$ . Dans le cas d'une file d'attente avec clients impatientes, la condition de perte est plus complexe :  $\pi_k$  est donnée par la probabilité que  $C_k$  se voit proposer un temps d'attente supérieur à son délai :

$$\pi_k = \mathbb{P}[C_k \text{ est perdu}] = \mathbb{P}[W_k > D_k] \quad (1.1)$$

Définissons maintenant les processus de performances :

- le processus de congestion  $(X_t, t \geq 0)$  est le processus comptant le nombre de clients dans le système à chaque instant  $t$ ,
- le processus d'affluence du buffer  $(Q_t, t \geq 0)$  est le processus qui compte le nombre de clients dans le buffer à l'instant  $t$ . Ainsi, pour tout  $t$  :

$$Q_t = (X_t - s)^+,$$

- pour tout  $n \geq 1$ , on note par  $\tilde{T}_n$ , le  $n$ -ième instant de sortie du serveur, c'est-à-dire, l'instant de la  $n$ -ième fin de service. Le processus des fins de services  $(S_t, t \geq 0)$  est alors le processus ponctuel (à trajectoires p.s. càdlàg) comptant le nombre de services terminés jusqu'à  $t$ . Autrement dit, pour tout  $t$  :

$$S_t = \sum_{n \in \mathbb{N}^*} \mathbb{1}_{\tilde{T}_n \leq t},$$

- pour tout  $n \geq 1$ , on note par  $\tilde{\sigma}_n$ , la durée du  $n$ -ième service rendu par le serveur, **i.e.** la durée du service qui se termine à  $\tilde{T}_n$ ,
- pour tout  $n \geq 1$ , on note par  $\check{T}_n$ , le  $n$ -ième instant où un client est perdu. Le processus des pertes  $(P_t, t \geq 0)$  est alors le processus càdlàg comptant le nombre de pertes jusqu'à  $t$ . Pour tout  $t$  :

$$P_t = \sum_{n \in \mathbb{N}^*} \mathbb{1}_{\check{T}_n \leq t},$$

- le processus des départs du système  $(D_t, t \geq 0)$  est le processus comptant les instants de sorties du système (services ou pertes) jusqu'à  $t$ . Autrement dit, pour tout  $t$  :

$$D_t = S_t + P_t.$$

Si on réordonne les délais résiduels des  $Q_t$  clients présents dans le buffer à  $t$  par ordre croissant (pour les servir en *EDF*, donc), on note ces délais résiduels :

$$R_1(t) < R_2(t) < \dots < R_{Q_t}(t).$$

$R_1(t)$  est donc le plus petit délai résiduel à l'instant  $t$ . A un instant de fin de service  $\tilde{T}_n$ , le serveur sert donc en *EDF* le client de délai résiduel  $R_1(\tilde{T}_n)$ . On notera également le délai initial du client prioritaire à l'instant  $t$  par  $D(t)$ . Supposons que la file est traitée en *EDF*. Soit  $n_0$ , le numéro du client prioritaire à l'instant  $t$  dans l'ordre des arrivées (ce client est donc  $C_{n_0}$ ). On a alors

$$R_1(t) = DR_{n_0}(t) \quad \text{et} \quad D(t) = D_{n_0} = R_1(t) + t - T_{n_0}.$$

On notera également

$$Z_t = R_{Q_t}(t),$$

le plus grand délai résiduel à  $t$  d'un client dans le buffer à  $t$ . A chaque fin de service  $\tilde{T}_n$ , on sert donc en *LDF* le client dont le délai résiduel à  $\tilde{T}_n$  est  $Z_{\tilde{T}_n}$ . Le processus  $(Z_t, t \in \mathbb{R}^+)$  est donc le

processus des délais résiduels maximaux.

On a vu que, tout au long de ce travail, tous les processus et les suites de bases du modèle (entrées, services, délais) seront au moins considérés stationnaires. Ceci étant, rien ne dit pour l'instant que le système lui-même, et par-là même les variables de performances comme les temps d'attente et de séjour et les processus de performances  $X$  et  $Q$ , soient des processus stationnaires.

La zone de stabilité de la file d'attente est l'ensemble des valeurs des paramètres de la file d'attente (lois des arrivées, des demandes de services et des délais) telles que le système puisse atteindre le régime stationnaire (et dans ce cas, il l'atteindra forcément un jour). En particulier, la loi du processus  $X$  devient stationnaire, et un client qui entre dans le système trouve toujours le système à l'état stationnaire, sans perturber cet état. La version stationnaire de  $X$  est alors la loi de probabilité sur  $\mathbb{N}$  telle que pour tout  $t$  à partir de l'atteinte de l'état stationnaire,  $X_t$  soit distribué suivant cette loi.

On notera, à l'état stationnaire :

- $X_\infty$ , la version stationnaire de  $X_t$
- $Q_\infty$ , la version stationnaire de  $Q_t$ ,
- $\mathcal{T}^a$ , la version stationnaire du temps d'attente,
- $\mathcal{T}^s$ , la version stationnaire du temps de séjour.
- $W$ , la version stationnaire du temps d'attente proposé.

On appelle point de construction d'une file d'attente stable, un point de renouvellement pour son processus de congestion, c'est à dire, un instant où un client entre dans un système vide. On appelle par ailleurs busy cycle, le temps qui s'écoule entre deux points de construction successifs. Un busy cycle se décompose en

$$\text{busy cycle} = \text{busy period} + \text{idle period}$$

où la busy period est le temps passé entre le point de construction et le premier instant de vidage et  $Y$  idle period est le temps passé entre cet instant de vidage et le point de construction suivant. Par définition de la stabilité d'une file d'attente, une file d'attente stable a **p.s.** une infinité de points de construction, et les durées des busy periods successives forment une suite **i.i.d.**, d'espérance finie.

A supposer que le système atteigne un jour l'état stationnaire, on définit un important indicateur de performances pour une file d'attente à perte, la probabilité de perte à l'état stationnaire, ou probabilité de perte.

### Définition 1.3.1

Pour toute file d'attente à perte, la probabilité de perte est donnée par :

$$\pi = \mathbb{P}[\text{un client entrant dans le système à l'état stationnaire est perdu}]. \quad (1.2)$$

Pour une file d'attente avec clients impatientes,  $\pi$  est donnée par la probabilité à l'état stationnaire qu'un client doit attendre plus longtemps que son délai pour être servi :

$$\pi = \lim_{n \rightarrow \infty} \pi_n = \lim_{n \rightarrow \infty} \mathbb{P}[W_n > D_n] = \lim_{n \rightarrow \infty} \mathbb{P}[W_n > D] = \mathbb{P}[W > D]. \quad (1.3)$$

## Chapitre 2

# Problématique et résultats existants

Nous exposons ici les principaux résultats existants sur le sujet des files avec clients impatients. Nous présentons en outre les difficultés méthodologiques soulevées par ce modèle, comme une introduction à nos contributions, qui suivront ce chapitre.

La littérature, abondante en ce qui concerne le domaine général des files d'attente, est en réalité assez réduite dès lors que l'on pose des limites sur le temps d'attente des clients. On distinguera ici trois types d'approches, auxquelles correspondront les trois parties de ce chapitre d'état de l'art du problème.

La première consiste à adapter la théorie classique des files d'attente *FCFS* au cas où les clients sont impatients. On cherche dans ce cadre à savoir à quelles conditions les processus qui sont les métriques habituelles d'étude du système (congestion, charge de travail...) restent ergodiques et à en déduire les caractéristiques du système à l'état stationnaire. La théorie est assez robuste : en supposant au moins que les interarrivées, les temps de service et les délais forment des suites **i.i.d.**, on peut adapter certaines descriptions de la file d'attente classique *FCFS* et donner des formules explicites pour la stabilité, la probabilité de perte ou le temps d'attente.

La deuxième approche pose le problème des files d'attente avec clients impatients en terme d'ordonnancement : quelle discipline de service doit-on appliquer pour perdre le moins de clients possible ? Apparaît ici l'idée qu'une autre discipline de service que *FCFS*, qui prenne en compte les délais des clients, peut traiter la demande de manière plus performante.

Il émerge des travaux réalisés sur le sujet que la discipline de service *EDF* est optimale à plusieurs titres. Ceci étant, ces résultats sont pour l'heure, soit posés en termes totalement déterministes (on connaît le scénario des requêtes, on ordonnance le traitement des tâches pour donner

la meilleure réponse), soit purement expérimentaux : il apparaît que les arguments valables pour l'étude probabiliste d'une file avec clients impatients gérée par *FCFS* ne sont pas adaptables à un système *EDF* par exemple.

Nous présentons donc une troisième approche, celle de la Real Time Queueing Theory (RTQT) ([14]) : elle propose un processus stochastique original pour décrire une file d'attente avec délais non éliminatoires régie par *EDF* .

Ce chapitre s'organise comme suit : en partie (2.1), nous passons en revue les résultats existants sur le problème de la file *FCFS* avec clients impatients. En partie (2.2), nous présentons l'état de l'art sur la question de la comparaison des performances entre les différentes disciplines de service et en particulier l'optimalité de l'algorithme *EDF*. En partie(2.3), nous présentons la problématique de l'étude probabiliste d'une file d'attente régie par *EDF* et la complexité de ce modèle. Nous voyons en section (2.3.2) les résultats expérimentaux existant sur le calcul des performances d'*EDF* et en section (2.3.3), nous évoquons brièvement la méthode originale de la RTQT et donnons l'intérêt théorique que cette méthode peut présenter dans le cadre de notre problème.

## 2.1 La file FCFS : résultats et méthodes existants

Depuis longtemps (Erlang, [16]), on sait étudier avec des processus markoviens simples, l'état stationnaire d'une file d'attente " classique "  $M/M/s/s + B$  gérée par *FCFS* où  $B$  est la taille, limitée et déterministe, du buffer. Ces modèles sont les modèles à pertes les plus simples. On décrit parfaitement le comportement du processus de congestion  $(X_t, t \geq 0)$  et on en déduit que le système atteint la stationnarité sans condition sur ses paramètres. Par ailleurs, la propriété *PASTA* avec l'ergodicité de  $X$  permet de donner la probabilité de perte à l'état stationnaire par la probabilité à l'état stationnaire qu'il y ait  $s + B$  clients dans le système.

Le délai (ou patience, suivant les auteurs) affecté aux clients apparaît comme une limitation (mais de taille aléatoire!) de la capacité du système. Barrer ([1]) a, le premier, soulevé le problème pour une file d'attente  $M/M/s/s + D/FCFS$  : peut-on adapter le résultat précédent pour trouver la probabilité de perte d'un client ? La réponse est non, puisqu'alors le processus de congestion  $X$  n'est pas markovien. Barrer construit alors un modèle ad hoc pour lever cette indétermination. Ce faisant, il exhibe la loi stationnaire du nombre de clients dans le système et donne  $\pi$  la probabilité de perte :

$$\pi = \frac{1}{s!} \rho^s \exp^{(\lambda - s\mu)d} \mathbb{P}[X_\infty = 0], \quad (2.1)$$

où  $d$  est la valeur des délais initiaux. Gnedenko et Kovalenko ([20]) ont pointé et corrigé des erreurs dans la preuve de (2.1), tout en validant la formule. Jurkevic ([23]) a ensuite élargi ce résultat et donné la loi du nombre de clients à l'état stationnaire lorsque les délais sont le minimum d'une constante et d'une loi exponentielle. Par ailleurs, Finch a donné la loi du temps d'attente dans une file  $GI/M/1/1 + D/FIFO$  ([17]). Des considérations du même type ont mené au calcul de la probabilité de perte d'une file  $M/G/1/1 + D/FIFO$  (Kok et Tijms, [12]) et d'une file  $M/G/1/1 + D/FIFO$  avec délais jusqu'à la fin du service ([13]). Citons encore [10], [19] et [37] pour d'autres résultats sur des modèles de ce type.

Pour initier la généralisation de ces résultats à une file  $G/G/1/1 + G/FIFO$ , [30] donne une borne supérieure heuristique pour la probabilité de perte :

$$\pi \leq \mathbb{P}[\sigma > U] \quad (2.2)$$

Pour une file d'attente classique  $G/G/1$  régie par  $FIFO$ , le descripteur d'état est la suite des temps d'attente proposés aux clients (ou workload)  $(W_n, n \in \mathbb{N})$ . L'équation de Lindley donne pour tout  $n$  :

$$W_{n+1} = [W_n - U_{n+1} + \sigma_n]^+ \quad (2.3)$$

L'étude de la stabilité de la file  $G/G/1$  due à Loynes ([31]) utilise ce processus remarquable. Loynes prouve qu'à la condition

$$\mathbb{E}[U] > \mathbb{E}[\sigma] \quad (2.4)$$

le système atteint un régime stationnaire, et décrit le régime stationnaire du système avec  $W$ , la loi stationnaire du workload.

Revenons au temps-réel. Baccelli et Hébuterne ([7]) et Baccelli, Boyer et Hébuterne ([5]) montrent que le workload d'une file d'attente  $GI/GI/1/1 + GI/FIFO$  est une chaîne de Markov à partir d'une observation simple : le workload est le même que celui d'un système où les clients, qui connaissent à leur arrivée la valeur du workload (aware customers), ne rentrent effectivement dans le système que si le workload est plus petit que leur délai (auquel cas ils seront forcément servis). Une file avec aware customers n'endure donc pas de perte de clients rentrés dans le système : un client peut savoir dès le début s'il pourra atteindre le service ou non avant l'expiration de son

délai, et ne rentre pas si ça n'est pas le cas. Le workload de la file avec aware customers (et donc, celui de la file avec unaware customers) vérifie alors une équation markovienne du même type que (2.3) :

$$\begin{cases} W_{n+1} = [W_n + \sigma_n - U_{n+1}]^+ & \text{si } D_n > W_n \\ W_{n+1} = [W_n - U_{n+1}]^+ & \text{sinon.} \end{cases}$$

Il est alors aisé de donner la condition pour l'atteinte du régime stationnaire :

**Théorème 2.1.1** (cf [5])

A la condition  $\mathbb{P}[\sigma < U] > 0$  la chaîne de Markov  $(W_n, n \in \mathbb{N})$  est ergodique si et seulement si :

$$1 - \rho(1 - \lim_{n \rightarrow \infty} (1 - \mathbb{P}[D > x])) > 0.$$

De plus, les périodes d'activité sont d'espérance finie.

En particulier, ce théorème implique qu'à la condition  $\mathbb{P}[\sigma < U] > 0$ , si la loi des délais ne charge pas  $+\infty$ , la file est stable dans le sens où les périodes d'activité sont **p.s.** finies. Remarquons d'ailleurs que cette condition de stabilité est plus faible que la condition de Loynes (2.4), ce à quoi on pouvait s'attendre, puisque l'on perd des clients.

L'étude du comportement stationnaire de la file donne aussi une relation simple entre la probabilité de perte de la file et la probabilité que le système soit vide à l'état stationnaire. [5] montre que :

$$\pi = 1 - \frac{1}{\rho} + \frac{\mathbb{P}[X_\infty = 0]}{\rho}. \quad (2.5)$$

Considérons maintenant le cas  $M/M/s/s + G/FCFS$ . [5] donne la loi du temps d'attente de la file  $M/M/s/s + GI/FCFS$  grâce à l'argument précédent. Récemment, des méthodes markoviennes nouvelles ont permis de donner des formules exactes pour des métriques de performance du système, telles que la probabilité de perte ou le nombre de clients à l'état stationnaire. Brandt et Brandt ([4]) et Movaghar ([32]) donnent la loi du nombre de clients à l'état stationnaire  $X_\infty$  dans une file  $M/M/s/s + GI/FCFS$  (et même, où les paramètres des lois exponentielles des interarrivées et des services dépendent du nombre de clients dans le système, s'iaie dépendant imput, [32]). Ils utilisent un argument de "conservation de flux" (entrées en sorties) dû à Kleinrock ([24]) : à l'état stationnaire, le nombre de clients dans le système a la même loi suivant qu'il est vu d'un client qui sort du système (éliminé ou servi) ou d'un client qui entre dans le système. Les calculs sont alors permis par des propriétés spécifiques du processus de Poisson (*PASTA...*).

Ils donnent en particulier pour la file  $M/M/1/1 + GI/FIFO$  le terme  $\mathbb{P}[X_\infty = 0]$  par :

$$\mathbb{P}[X_\infty = 0] = 1 + \lambda \int_0^\infty e^{\lambda \int_0^y \mathbb{P}[D > 0] dx - \mu y} dy \quad (2.6)$$

ce qui avec (2.5) donne explicitement la probabilité de perte pour la file  $M/M/1/1 + G/FIFO$  (notons que le résultat est généralisé à la file  $M/M/s/s + GI/FCFS$  pour  $s \geq 1$ ).

Dans le cas d'une file  $M/M/s/s + D/FCFS$  où les délais initiaux déterministes sont tous égaux à  $d$ , [9] propose une approche sensiblement différente et notamment le calcul explicite de performances (probabilité de perte, temps d'attente...) par l'étude de la loi stationnaire du processus markovien à 1 ancienneté  $(K_t, t \geq 0)$ , à valeurs dans l'espace  $(\{0\} \times \{0, 1, \dots, s\}) \cup (\{1\} \times [0, d])$  :

$$\begin{cases} K_t = (0, c) & \text{si il y a } c \text{ clients dans le système, } c \leq s, \\ K_t = (1, Y_t) & \text{s'il y a au moins } s + 1 \text{ clients dans le système.} \end{cases}$$

où  $Y_t$  est le temps passé dans le système à  $t$  par le client prioritaire à cet instant (le plus ancien). On peut alors calculer la loi stationnaire de  $(K_t, t \geq 0)$  et donc la densité  $f$  de la version stationnaire de  $Y$ . En vertu de la propriété PASTA, la probabilité de perte est alors donnée par :

$$\pi = \frac{f(d)}{\lambda} \quad (2.7)$$

où  $\lambda$  est l'intensité du processus des entrées, et on retrouve le résultat de Barrer, (2.1). Remarquons que le résultat précédent est généralisé à des délais courant jusqu'à la fin du service, où le modèle est adaptable.

## 2.2 Comparaison des disciplines de service

Voyons maintenant dans les grandes lignes où en est le spectre de la connaissance concernant la question suivante : étant fixé un scénario d'arrivées de requêtes, affectées de délais d'exécution donnés, comment ordonnancer le traitement des tâches de manière à obtenir de meilleures performances ?

Un scénario temps-réel (scénario de trafic avec délais) est dit faisable par un ordonnanceur s'il peut traiter toutes les requêtes sans en perdre une seule. Dertouzos [11] prouve alors le résultat suivant :

**Théorème 2.2.1** (cf [11])

L'algorithme EDF est optimal en terme de faisabilité parmi tous les algorithmes préemptifs.

L'optimalité d'EDF en termes de faisabilité signifie qu'il n'existe pas de scénario de trafic temps-réel faisable par un autre algorithme préemptif, mais pas par EDF. D'autre part, Leboucher [25] montre qu'EDF est également optimal en terme de faisabilité si on considère les délais jusqu'à la fin du service, et dans d'autres cas connexes.

En des termes probabilistes, les résultats sont très réduits. Comme on va le voir tout au long de ce travail, la dynamique du système sous EDF est très complexe. Panwar, Towsley et Wolf ([35]) montrent cependant le résultat suivant :

**Théorème 2.2.2** (cf [35]) ;

Soit  $\gamma$ , une discipline de service quelconque. En notant  $P_t^\gamma$ , le nombre de clients perdus à  $t$  sous  $\gamma$ ,  $B_n^\gamma$ , la durée de la  $n$ -ième busy period sous  $\gamma$  et  $\mathcal{T}^\gamma$  le débit de service sous  $\gamma$  :

1. pour toute file  $G/M/s/s + GI$ , pour tout  $t$ ,  $P_t^{EDF} \leq_{st} P_t^\gamma$ ,
2. pour toute file  $M/G/s/s + GI$ ,  $\mathcal{T}^\gamma \leq \mathcal{T}^{EDF}$  et pour tout  $n$ ,  $B_n^{EDF} \geq_{st} B_n^\gamma$ ,

où on note :

$$X \geq_{st} Y \Leftrightarrow \mathbb{P}[X \leq x] \leq \mathbb{P}[Y \leq x], \text{ pour tout } x \in \mathbb{R}$$

Il est à noter que ces résultats sont obtenus par stricte comparaison des files d'attente trajectoire par trajectoire (i.e., pour un scénario de trafic donné) et ces arguments ne permettent pas a priori de calculer ou même d'estimer la loi du processus des pertes sous EDF ( $P_t^{EDF}$ ,  $t \geq 0$ ) et la probabilité de perte sous EDF,  $\pi^{EDF}$ . Néanmoins, ils confirment le résultats précédent de faisabilité en tant qu'ils désignent aussi EDF comme la discipline optimale.

## 2.3 La file EDF : problématique probabiliste

### 2.3.1 Description du système

Comme on l'a remarqué dans la section précédente et mis à part le dernier résultat, les comparaisons de performances entre les disciplines de service ont souvent été posées en termes d'ordonnancement pur, et rarement en termes probabilistes. Elles désignent EDF comme la discipline optimale. Cependant, alors que l'étude d'un système temps-réel sous FCFS est très robuste comme on l'a vu en partie (3.1), il n'en va pas de même lorsque la discipline de service est EDF.

En effet, du point de vue du client qui rentre dans le système, la situation est beaucoup plus complexe qu'en *FCFS*. Sachant son délai initial et le workload du système à l'instant de son arrivée, un client ne peut pas savoir en *EDF* s'il sera servi ou non. Sa situation dans la file d'attente va évoluer de manière instable et incertaine, puisque des clients peuvent entrer dans le système après lui et passer devant lui si leur délai initial est plus court que son délai résiduel à l'instant de leur arrivée. Ensuite, la situation du même client peut brutalement s'améliorer s'il s'avère que ces clients qui l'ont préempté sont éliminés du système à cause d'un délai trop court. En allant plus loin, il est très clair en *FCFS* qu'un client entrant dans un système à l'état stationnaire a d'autant moins de chances d'être perdu que son délai initial est long. Par contre, sous *EDF*, il semble a priori beaucoup plus difficile de répondre à cette question : un client qui a un délai initial très long et qui, de fait, a beaucoup de chances de passer longtemps dans le système n'a pas forcément plus de chances d'être servi qu'un client qui entre dans le système avec un délai initial très court, et qui a donc une probabilité grande d'être prioritaire (dans le cas non-préemptif), ou même d'être directement servi (dans le cas préemptif).

Ces remarques montrent donc bien la difficulté de décrire l'évolution de la file d'attente sous *EDF* par un descripteur d'état classique : entre autres, les remarques que nous venons de faire montrent que le processus de congestion  $(X_t, t \in \mathbb{R}^+)$ , la suite des workloads aux instants d'entrées  $(W_n, n \in \mathbb{N})$  ou encore le processus d'ancienneté  $(K_t, t \geq 0)$  introduit par [9] pour la file *FCFS* ne sont pas markoviens dans le cas d'*EDF*.

Cela pose problème, car nous avons vu qu'*EDF* est précisément la discipline de service la plus intéressante en tant qu'elle maximise de nombreux critères de performances. C'est bien là toute la problématique de ce travail : la théorie probabiliste est robuste lorsqu'il s'agit d'étudier les propriétés à l'état stationnaire d'un système régi par *FCFS*, mais les mêmes arguments ne peuvent plus s'appliquer s'agissant d'un système *EDF*, qui par ailleurs présente toutes les propriétés d'optimalité.

### 2.3.2 Calcul numérique des performances d'*EDF*

La simulation des performances d'*EDF* ne peut déjà être réalisée qu'en mettant en évidence un bon descripteur du système. Ainsi, citons quelques exemples montrant comment simuler numériquement la perte du système sous *EDF* en mettant en évidence une chaîne de Markov [[22], [34], [33], [35]]. Cela peut sembler surprenant, puisqu'on a vu qu'aucune chaîne de Markov simple ne peut a priori décrire un système sous *EDF*. Précisément, des approximations du système par un autre sont nécessaires pour rendre le système markovien et ces études ont des limites :

- Certaines d’entre elles utilisent un modèle simplifié, comme par exemple la file  $F/ML(n)$  ([22]), qui n’applique EDF qu’aux  $n$  premiers (ou derniers, [34]) clients, les autres étant servis en *FIFO*.
- Le calcul de  $\pi$  impose des limites pratiques sur le nombre d’états du système, et requiert une spécification restrictive des données, comme les  $n$  clients choisis dans [22], ou une borne sur les délais ([35]).

Par ailleurs, Chen et Decreusefond donnent dans [8] une méthode originale adaptée de la loi de conservation de Kleinrock pour calculer numériquement le temps d’attente moyen pour une file d’attente  $M/G/1/1 + G/EDF$ , où  $n$  clients appartiennent à  $N$  classes différentes qui déterminent leur délai. Ce calcul s’effectue par la résolution d’un système non linéaire de dimension  $N$ , faisable numériquement.

De manière générale, la mise en place de l’algorithme EDF pose comme on vient de le voir de vrais problèmes de tractabilité pour les calculs numériques et les simulations, et présente de plus un coup algorithmique en terme de complexité. Une autre question sous-jacente devient donc : le gain de performance d’EDF est-il suffisamment important par rapport au système FCFS par exemple, pour compenser cette plus grande complexité ? A notre connaissance, il n’existe pas dans la littérature de réels éléments permettant de répondre à cette question. L’un des objectifs de ce travail est précisément de donner des pistes pour le calcul des performances sous EDF.

### 2.3.3 Le descripteur à valeurs mesures ponctuelles

Nous introduisons ici l’idée et la méthode de Doytchinov, Lehoczky et Shreve (*RTQT*, [14]) pour décrire une file d’attente à un serveur régie par EDF. L’idée centrale de ce travail consiste à considérer comme processus descripteur du système le processus à valeurs mesures ponctuelles qui garde la trace à l’instant  $t$  de tous les délais résiduels de tous les clients présents dans le système à l’instant  $t$ . Les trajectoires du processus prennent donc leur valeur dans un espace d’état beaucoup plus complexe que la droite réelle (ici, un espace de dimension infinie), mais semblent être les seules qui rendent compte de l’évolution du système de manière exhaustive.

Soit le système suivant :

- une file d’attente  $GI/GI/1$ ,
- les clients entrent affectés de délais **i.i.d.** et **p.s.** bornés et sont servis suivant la discipline de service EDF préemptive,
- aucun client n’est perdu : les clients restent dans la file même si leur délai est expiré, jusqu’à

l'achèvement de leur service.

Attardons-nous sur les deux différences majeures avec les systèmes que nous étudions dans cette mémoire :

- d'une part, les délais ne sont pas éliminatoires, ce qui bouleverse fondamentalement la dynamique du système : celui-ci ne présente pas la même condition de stabilité, ne se vide pas aux mêmes instants, sert des clients que le système EDF avec élimination aurait perdus. Dans ce cas, les délais des clients ne servent qu'à les classer dans la file,
- d'autre part, la discipline de service est préemptive, ce qui signifie que le service d'un client peut être interrompu à tout moment si un client plus pressé que lui entre dans le système.

Remarquons enfin que l'hypothèse que les délais initiaux sont **p.s.** bornés n'est qu'une hypothèse technique. D'une part, cette hypothèse est très plausible en pratique et d'autre part, les simulations réalisées dans ces travaux indiquent que les résultats sont toujours vérifiés si cette condition n'est pas remplie. Cela dit, nous verrons que nous pourrons nous passer de cette hypothèse dans notre cas.

On définit alors comme descripteur du système, le processus des profils à l'instant  $t$  par, pour tout  $t$  et pour tout borélien  $\mathfrak{B}$  :

$$\mathcal{Q}_t(\mathfrak{B}) = \{\text{nombre de clients dans le file à } t \text{ ayant à } t \text{ un délai résiduel dans } \mathfrak{B}\}$$

On note par ailleurs

$$\mathcal{W}_t(\mathfrak{B}) = \{\text{travail associé aux clients dans le file à } t \text{ ayant à } t \text{ un délai résiduel dans } \mathfrak{B}\}$$

le processus du workload associé.

L'objet de [14] est l'étude à forte charge ( $\rho = 1$ ) d'un tel système par changement d'échelle en espace et en temps.;

On considère ainsi une suite de systèmes tels que  $\lambda^{(n)}\mathbb{E}[\sigma^{(n)}] \rightarrow 1$ , où les clients entrent dans le système affectés de délais initiaux  $D^{(n)}$  de fonction de répartition  $G^{(n)}$ . On suppose qu'il existe une fonction de répartition  $G$  telle que  $G^{(n)}(\sqrt{nx}) = G(x)$  pour tout  $n$ .

On note pour le  $n$ -ième système,  $\mathcal{Q}_t^{(n)}$  et  $\mathcal{W}_t^{(n)}$ , les processus des profils et du travail correspondant, et

- le Netput process  $N_t^{(n)} = \text{travail entré jusqu'à } t$ ,
- l'Idleness process  $I_t^{(n)} = - \min_{0 \leq s \leq t} N_s^{(n)} = \text{temps idle jusqu'à } t$

– le Workload process  $W_t^{(n)} = N_t^{(n)} + I_t^{(n)}$ .

On normalise ces processus de la façon suivante, par exemple :

$$\bar{Q}_t^{(n)}(\mathfrak{B}) = \frac{1}{\sqrt{n}} Q_{nt}^{(n)}(\sqrt{n}\mathfrak{B}).$$

Rappelons maintenant le théorème central limite fonctionnel pour les files d'attente sans perte, réalisé pour une charge limite 1. Au sens de la convergence en loi des processus à valeurs réelles :

**Théorème 2.3.1**

$$(\hat{N}^{(n)}, \hat{I}^{(n)}, \hat{W}^{(n)}) \Rightarrow (N^*, I^*, W^*),$$

où  $N^*$  est un mouvement Brownien avec dérive, et

$$I_t^* = - \min_{0 \leq s \leq t} N_s^*,$$

$$W_t^* = N_t^* + I_t^*.$$

En utilisant ce dernier résultat, Doytchinov et *al.* démontrent alors le théorème suivant :

**Théorème 2.3.2** (cf [14])

Les processus à valeurs mesures  $\bar{W}^{(n)}$  et  $\bar{Q}^{(n)}$  convergent en distribution vers les processus  $\bar{W}^*$  et  $\bar{Q}^*$  définis pour tout  $t$  et tout  $\mathfrak{B}$  par :

$$\hat{W}_t^*(\mathfrak{B}) = \int_{\mathfrak{B} \cap [F^*(t), \infty[} (1 - G(x)) dx, \quad \hat{Q}_t^*(\mathfrak{B}) = \lambda \hat{W}_t^*(\mathfrak{B}),$$

où  $F^*$  est le processus défini pour tout  $t$  par :

$$F^*(t) := H^{-1}(W_t^*), \quad \text{où } H(y) := \int_y^\infty (1 - G(x)) dx.$$

En projetant pour tout  $t$  et pour tout  $n$  continûment la mesure  $\hat{Q}_t^{(n)}$  sur la droite réelle, on a en particulier :

**Corollaire 2.3.1** ([14])

La longueur de la file renormalisée  $Q^{(n)}$  converge en distribution au sens des processus vers  $\lambda W^*$ .

**Remarque 2.3.1** L'idée de prendre un processus à valeurs mesures ponctuelles comme indicateur de l'état d'un système à la dynamique instable est également développée par Gromoll, Puha et Williams ([21]) dans le cas de la file d'attente Processor Sharing (PS) et par Limic ([29]) pour une file LIFO, dans les deux cas pour des systèmes sans pertes à forte charge ( $\rho = 1$ ).

## Chapitre 3

# Stabilité

Ce chapitre, et le chapitre suivant, sont dévolus à nos contributions. Nous posons ici la question de la stabilité du système. A quelle condition sur les paramètres du système, celui-ci atteint-il un régime stationnaire? Comme on l'a vu dans le chapitre précédent, théorème (2.1.1), une condition nécessaire et suffisante de stabilité pour une file d'attente  $GI/GI/1/1 + GI/FIFO$  où les délais sont **p.s.** finis est donnée par

$$\mathbb{P}[\sigma < U] > 0.$$

Nous nous proposons ici d'élargir ce résultat à un cas plus général :

- en supposant seulement que l'on est dans le cas  $G/G/1$ , où les délais forment une suite ergodique,
- en considérant une discipline de service quelconque.

Il est à noter en effet sur ce dernier point que les résultats de stabilité existant à notre connaissance ne sont établis que pour la discipline de service FIFO. Une méthode liée au calcul de Palm permet comme on va le voir de donner des résultats pour le cas  $G/G/1/1 + G$  et pour une discipline de service quelconque.

Ce chapitre s'organise comme suit. En partie (3.1) nous montrons la stabilité dans le cas  $GI/GI/1/1 + GI$  avec discipline de service quelconque, en comparant le processus de congestion de la file avec clients impatientes avec celui de la file " délai pur "  $G/G/\infty$ .

Nous donnons en partie (3.2) un résultat de stationnarité adapté du calcul de Palm et d'un schéma de récurrence arrière pour une file avec clients impatientes  $G/G/1/1 + G$  avec discipline de service quelconque.

### 3.1 Régénérativité dans le cas $GI/GI/1/1 + GI$

#### 3.1.1 Comparaison avec la file " délai pur "

Nous considérons ici une file d'attente délai pur  $G/G/\infty$ , c'est-à-dire :

- les suites des interarrivées et des temps de service  $(U_n, n \in \mathbb{N})$  et  $(\sigma_n, n \in \mathbb{N})$  sont ergodiques et intégrables,
- il y a une infinité de serveurs.

Ainsi, chaque client qui entre dans le système est directement pris en charge par un serveur (il y en a toujours un disponible) et reste dans le système pendant la durée de son service.

Nous noterons alors pour un tel système :

- $\sigma_n^{DP}$ , le temps de service demandé par le  $n$ -ième client entré dans la queue,
- $X_n^{DP}$ , le nombre de clients à l'instant  $t$ .

#### **Théorème 3.1.1**

*Soit une file  $G/G/1/1 + G$  gérée par une discipline de service non-préemptive quelconque, où  $(\sigma_n, n \in \mathbb{N})$  représente la suite des temps de service et  $(D_n, n \in \mathbb{N})$  est la suite des délais initiaux des clients. Alors :*

1. *Le nombre de clients dans le système est à tout instant  $t$  inférieur à celui  $X_t^{DP}$  de la file  $G/G/\infty$  qui traite les demandes des mêmes clients, où les temps de services demandés par les clients sont donnés par*

$$\sigma_n^{DP} = \sigma_n + D_n.$$

2. *Le nombre de clients dans le buffer est à tout instant  $t$  inférieur au nombre de clients  $\hat{X}_t^{DP}$  dans la file  $G/G/\infty$  qui traite les demandes des mêmes clients et où les temps de service demandés par les clients sont donnés par :*

$$\hat{\sigma}_n^{DP} = D_n.$$

**Preuve.** Supposons que la file avec impatiences et la file délai pur partent du même état initial.

Ensuite, elles traitent les demandes des mêmes clients (arrivés aux mêmes instants).

1. Dans la file  $G/G/1/1 + G$ , le client  $C_n$  reste dans le système pendant un temps au plus égal à  $\sigma_n + D_n$ . Par ailleurs, le même client  $C_n$  entre dans la file délai pur en demandât un service

$\sigma_n^{DP} = \sigma_n + D_n$ . Les suites  $(D_n, n \in \mathbb{N})$  et  $(\sigma_n, n \in \mathbb{N})$  étant ergodiques,  $(\sigma_n^{DP}, n \in \mathbb{N})$  l'est aussi et donc le système est bien de type  $G/G/\infty$ . On a de plus :

$$X_t \leq \sum_{n \in \mathbb{N}} \mathbb{1}_{\{T_n \leq t, T_n + \sigma_n + D_n > t\}} = X_t^{DP}.$$

2. Le client  $C_n$  passe dans le buffer de la file  $G/G/1/1/G$  un temps inférieur ou égal à  $D_n$ . Si on considère maintenant que les mêmes clients entrent dans le système délai pur en demandant un service de durée

$$\hat{\sigma}_n^{DP} = D_n,$$

ce système est bien de type  $G/G/\infty$  et on a :

$$Q_t \leq \sum_{n \in \mathbb{N}} \mathbb{1}_{\{T_n \leq t, T_n + D_n > t\}} = \hat{X}_t^{DP}$$

### 3.1.2 Résultat de régénérativité

La question de la stabilité est souvent posée en terme de finitude presque sûre des busy periods, ou autrement dit, de régénérativité du processus de congestion  $(X_t, t \geq 0)$ . On dit qu'un processus est régénératif lorsque 0 est récurrent positif pour ce processus. Dans ce cas,  $(X_t, t \geq 0)$  atteint **p.s.** un régime stationnaire et le temps passé entre les instants successifs de visite en 0 forment une suite **i.i.d.**. Dans le cas particulier où  $(X_t, t \geq 0)$  représente le processus de congestion d'une file d'attente, la régénérativité de  $X$  implique en particulier que la file d'attente atteint un régime stationnaire et que les durées des busy periods successives forment une suite **i.i.d.**. En particulier, le processus  $X$  atteint le régime stationnaire et la file est stable.

#### Théorème 3.1.2

*Soit une file d'attente avec clients impatientes  $GI/GI/1/1 + GI$  gérée par une discipline de service non-préemptive quelconque, où les interarrivées sont notées  $(U_n, n \in \mathbb{N})$ , les temps de services  $(\sigma_n, n \in \mathbb{N})$  et les délais initiaux des clients sont notés  $(D_n, n \in \mathbb{N})$ . On suppose que :*

$$\mathbb{P}[D + \sigma < U] > 0. \tag{3.1}$$

*Alors, le processus de congestion  $(X_t, t \geq 0)$  est régénératif.*

**Preuve.** Considérons la file  $GI/GI/\infty$  où les interarrivées sont de loi celle de  $U$ , telle que les demandes de services des clients sont la somme des délais et des temps de services demandés par

les clients de la file avec clients impatientes (**i.e.**,  $(\sigma_n^{DP} = \sigma_n + D_n, n \in \mathbb{N})$ ). Soit  $\mathbb{P}_N^0$ , la probabilité de Palm associée au processus des entrées  $N$ . (3.1) implique que

$$\mathbb{P}_N^0[\sigma^{DP} < U] = \mathbb{P}_N^0[\sigma + D < U] > 0.$$

On peut donc appliquer le résultat de Thorisson ([38], p.370) : le processus  $(\hat{X}_t^{DP}, t \in \mathbb{R}^+)$  du nombre de clients dans la file  $GI/GI/\infty$  est régénératif. On applique alors la première partie du théorème (3.1.1) : le processus  $(X_t, t \in \mathbb{R}^+)$  du nombre de clients dans la file avec impatiences  $GI/GI/1/1 + GI$  atteint 0 **p.s.** une infinité de fois.

### Remarque 3.1.1

*Ce résultat généralise le théorème (2.1.1) à une discipline de service quelconque.*

## 3.2 Stationnarité pour la file $G/G/1/1 + G$

Nous considérons ici le cas général d'une file d'attente avec clients impatientes  $G/G/1/1 + G$  gérée par une discipline de service non-préemptive quelconque.

### 3.2.1 Un résultat de stationnarité

Soient :

- $(\hat{S}_t, t \geq 0)$ , un processus ponctuel stationnaire de points  $(\hat{T}_n, n \in \mathbb{N})$ . Pour tout  $t$  :

$$\hat{S}_t = \sum_{n \in \mathbb{N}} \mathbb{1}_{\{\hat{T}_n \leq t\}},$$

- pour tout  $n \in \mathbb{N}$ ,

$$\hat{\sigma}_n = \hat{T}_{n+1} - \hat{T}_n.$$

La suite  $(\hat{\sigma}_n, n \in \mathbb{N})$  est donc une suite stationnaire ergodique de loi générale celle de  $\hat{\sigma}$ ,

- $(\hat{\alpha}_n, n \in \mathbb{N})$  une suite de marques du processus  $(\hat{S}_t, t \geq 0)$ . La suite  $(\hat{\alpha}_n, n \in \mathbb{N})$  est donc une suite stationnaire ergodique de loi générale celle de  $\hat{\alpha}$ .

Nous nous plaçons sur l'espace de Palm correspondant au processus  $\hat{S}$ . Plus précisément, nous munissons l'espace probabilisé  $(\Omega, \mathcal{F}, \mathbb{P})$  du flot  $\theta = \theta_{\hat{T}_1}$  et de la probabilité de Palm  $\mathbb{P}_{\hat{S}}^0$  associés à  $\hat{S}$ .

L'espace de Palm  $(\Omega, \mathcal{F}, \mathbb{P}_{\hat{S}}^0, \theta)$  est tel que :

–  $\theta$  est stationnaire : Pour tout  $\mathcal{A}$  dans  $\mathcal{F}$ ,

$$\mathbb{P}_{\hat{S}}^0[\theta^{-1}(\mathcal{A})] = \mathbb{P}_{\hat{S}}^0[\mathcal{A}],$$

–  $\theta$  est ergodique : Pour tout  $\mathcal{A}$  dans  $\mathcal{F}$ ,

$$A = \theta^{-1}(\mathcal{A}) \Rightarrow \mathbb{P}_{\hat{S}}^0[\mathcal{A}] = 0 \text{ ou } 1,$$

–  $\hat{T}_0 = 0$ ,  $\mathbb{P}_{\hat{S}}^0$  – **p.s.**

Suivant cette définition, la suite  $(\hat{\sigma}_n, n \in \mathbb{N})$  est compatible avec le flot ergodique  $\theta$  et autrement dit, en notant pour tout  $n$  dans  $\mathbb{N}$ ,

$$\theta^n = \theta \circ \theta \circ \dots \circ \theta,$$

alors,  $\mathbb{P}_{\hat{S}}^0$  – **p.s.**, pour tout  $n$  dans  $\mathbb{N}$  :

$$\hat{\sigma}_n = \hat{\sigma} \circ \theta^n.$$

De même,  $(\hat{\alpha}_n, n \in \mathbb{N})$  étant une suite de marques du processus  $(\hat{S}_t, t \geq 0)$ , elle est également compatible avec le flot. C'est donc une suite stationnaire ergodique et pour tout  $n$  dans  $\mathbb{N}$  :

$$\hat{\alpha}_n = \hat{\alpha} \circ \theta^n.$$

Nous supposons de plus que :

$$\mathbb{E}_{\hat{S}}^0[\hat{\sigma}] < \infty \text{ et } \mathbb{E}_{\hat{S}}^0[\hat{\alpha}] < \infty.$$

Par ailleurs, nous supposons que le flot  $\theta$  est bijectif, et notons :

$$\theta^{-1}(\mathcal{A}) = \mathfrak{B} \in \mathcal{F} \text{ tel que } \theta(\mathfrak{B}) = \mathcal{A}.$$

Ainsi, pour tout  $n$  dans  $\mathbb{N}$ ,

$$\theta^{-n} = \theta^{-1} \circ \theta^{-1} \circ \dots \circ \theta^{-1}.$$

Nous décrivons une suite définie par une équation de récurrence de type équation de Lindley ([31]) :

### Définition 3.2.1

Soit  $Y$ , une variable aléatoire **p.s.** finie. La suite  $(\hat{Y}_n^Y, n \in \mathbb{N})$  à valeurs  $\mathbb{R}^+$  est définie par :

$$\begin{cases} \hat{Y}_0^Y = Y, \\ \hat{Y}_{n+1}^Y = [\max\{\hat{Y}_n^Y, \hat{\alpha}_n\} - \hat{\sigma}_n]^+, \quad \text{pour tout } n \in \mathbb{N}. \end{cases}$$

Nous cherchons ici à établir à quelle condition  $(\hat{Y}_n^Y, n \in \mathbb{N})$  atteint un régime stationnaire. Autrement dit, nous cherchons une variable aléatoire  $Y$ , **p.s.** finie, telle que  $Y$  soit la version stationnaire de la suite  $(Y_n, n \in \mathbb{N})$ , ou encore :

$$\hat{Y}_n^Y = Y \circ \theta^n \text{ pour tout } n \in \mathbb{N}.$$

En particulier, avec la définition (3.2.1), on aura alors :

$$Y \circ \theta = [\max\{Y, \hat{\alpha}\} - \hat{\sigma}]^+. \quad (3.2)$$

On a le résultat suivant :

**Théorème 3.2.1**

*Il existe une unique solution  $Y$  de (3.2), finie **p.s.**. De plus :*

$$\mathbb{P}_{\hat{\sigma}}^0[Y \leq \hat{\alpha}] > 0. \quad (3.3)$$

**Preuve.** Soit  $Y$  une solution de (3.2). Vérifions (3.3) :

- Si  $\mathbb{P}_{\hat{\sigma}}^0[Y = 0] > 0$ , il n'y a rien à montrer.
- Supposons que  $\mathbb{P}_{\hat{\sigma}}^0[Y = 0] = 0$ . Alors, sur un événement de probabilité 1,

$$Y \circ \theta > 0 \Leftrightarrow [\max\{Y, \hat{\alpha}\} - \hat{\sigma}]^+ > 0 \Leftrightarrow Y \circ \theta = \max\{Y, \hat{\alpha}\} - \hat{\sigma}. \quad (3.4)$$

Si on suppose que (3.3) est fautive,  $\mathbb{P}_{\hat{\sigma}}^0 - p.s.$ ,  $Y > \hat{\alpha}$  et donc  $Y \circ \theta = Y - \hat{\sigma}$ . On a alors :

$$\mathbb{E}_{\hat{\sigma}}^0[Y \circ \theta - Y] = -\mathbb{E}_{\hat{\sigma}}^0[\hat{\sigma}] < 0,$$

ce qui est absurde par le lemme ergodique. (3.3) est vérifiée.

Nous allons exhiber une solution  $Y$  de l'équation (3.2) par un schéma de type schéma arrière de Loynes ([31]).

Nous notons pour tout  $n$  dans  $\mathbb{N}$ ,

$$\hat{\alpha}_{-n} = \hat{\alpha} \circ \theta^{-n} \text{ et } \hat{\sigma}_{-n} = \hat{\sigma} \circ \theta^{-n}.$$

Définissons la suite  $(M_n, n \in \mathbb{N})$  (suite arrière correspondant à  $(\hat{Y}_n^0, n \in \mathbb{N})$ ) :

$$\begin{cases} M_0 = 0, \text{ p.s.}, \\ M_n = \left[ \max_{j=1}^n \left( \hat{\alpha}_{-j} - \sum_{i=1}^j \hat{\sigma}_{-i} \right) \right]^+ \end{cases} \text{ pour tout } n \in \mathbb{N}.$$

Remarquons que  $(M_n, n \in \mathbb{N})$  est croissante, puisque pour tout  $n$  dans  $\mathbb{N}$  :

$$\begin{aligned} M_{n+1} &= \left[ \max_{J=1}^{n+1} \left( \hat{\alpha}_{-j} - \sum_{i=1}^j \hat{\sigma}_{-i} \right) \right]^+ \\ &= \left[ \max \left\{ \max_{J=1}^n \left( \hat{\alpha}_{-j} - \sum_{i=1}^j \hat{\sigma}_{-i} \right), \hat{\alpha}_{-(n+1)} - \sum_{i=1}^{n+1} \hat{\sigma}_{-i} \right\} \right]^+ \\ &= \max \left\{ M_n, \hat{\alpha}_{-(n+1)} - \sum_{i=1}^{n+1} \hat{\sigma}_{-i} \right\}, \end{aligned}$$

en remarquant que  $\{\max(A, B)\}^+ = \max\{A^+, B\}$ , pour tous  $A$  et  $B$ .

Donc, il existe une variable aléatoire (éventuellement infinie)  $M_\infty$ , telle que  $M_\infty$  soit la limite presque sûre de la suite  $(M_n, n \in \mathbb{N})$ . Plus précisément,  $M_\infty$  est donnée par :

$$M_\infty = \left[ \sup_{j \in \mathbb{N}^*} \left( \hat{\alpha}_{-j} - \sum_{i=1}^j \hat{\sigma}_{-i} \right) \right]^+. \quad (3.5)$$

Par ailleurs, on peut remarquer que :

**Lemme 3.2.1**

*Pour tout  $n$  dans  $\mathbb{N}$ ,*

$$M_{n+1} \circ \theta = [\max\{M_n, \hat{\alpha}\} - \hat{\sigma}]^+.$$

**Preuve du lemme (3.2.1).** On fait une récurrence sur  $n$  :

$$\begin{aligned} M_{n+1} &= \left[ \max_{J=1}^{n+1} \left( \hat{\alpha}_{-j} - \sum_{i=1}^j \hat{\sigma}_{-i} \right) \right]^+ \\ &= \left[ \max \left\{ \max_{J=2}^{n+1} \left( \hat{\alpha}_{-j} - \sum_{i=1}^j \hat{\sigma}_{-i} \right), \hat{\alpha}_{-1} - \hat{\sigma}_{-1} \right\} \right]^+ \\ &= \left[ \max \left\{ \max_{J=1}^n \left( \hat{\alpha}_{-j-1} - \sum_{i=0}^j \hat{\sigma}_{-i-1} \right), \hat{\alpha}_{-1} - \hat{\sigma}_{-1} \right\} \right]^+ \\ &= \left[ \max \left\{ \max_{J=1}^n \left( \hat{\alpha}_{-j} - \sum_{i=1}^j \hat{\sigma}_{-i} \right) \circ \theta^{-1} - \hat{\sigma}_{-1}, \hat{\alpha}_{-1} - \hat{\sigma}_{-1} \right\} \right]^+ \\ &= \left[ \max \left\{ \left[ \max_{J=1}^n \left( \hat{\alpha}_{-j} - \sum_{i=1}^j \hat{\sigma}_{-i} \right) \right]^+ \circ \theta^{-1}, \hat{\alpha}_{-1} \right\} - \hat{\sigma}_{-1} \right]^+ \\ &= \left[ \max \left\{ M_n, \hat{\alpha} \right\} - \hat{\sigma} \right]^+ \circ \theta^{-1} \end{aligned}$$

Par continuité **p.s.** de la fonction  $[\max\{., \hat{\alpha}\} - \hat{\sigma}]^+ \circ \theta^{-1}$ , la limite  $M_\infty$  vérifie donc :

$$M_\infty \circ \theta = [\max\{M_\infty, \hat{\alpha}\} - \hat{\sigma}]^+,$$

ce qui correspond à (3.2).  $M_\infty$  est donc un bon " candidat " pour la v.a. stationnaire  $Y$ . Il reste à vérifier à quelle condition  $M_\infty$  est **p.s.** finie.

**Lemme 3.2.2**

$M_\infty$  est **p.s.** finie.

**Preuve du lemme (3.2.2).** L'évènement  $\{M_\infty = \infty\}$  est  $\theta$ -invariant. En effet,

$$\begin{aligned} \theta^{-1}\{M_\infty = \infty\} &= \{M_\infty \circ \theta = \infty\} \\ &= \{[\max\{M_\infty, \hat{\alpha}\} - \hat{\sigma}]^+ = \infty\} = \{M_\infty = \infty\} \end{aligned}$$

puisque  $\hat{\alpha}$  et  $\hat{\sigma}$  sont **p.s.** finies.

Maintenant, remarquons que :

$$M_\infty = \left[ \sup_{j \in \mathbb{N}} \left( \hat{\alpha}_{-j} - \sum_{i=1}^j \hat{\sigma}_{-i} \right) \right]^+ = \left[ \sup_{j \in \mathbb{N}} \left( \sum_{i=1}^j (\hat{\alpha}_{-i} - \hat{\alpha}_{-i+1} - \hat{\sigma}_{-i}) + \hat{\alpha}_0 \right) \right]^+. \quad (3.6)$$

**Théorème 3.2.2** (cf [2])

Soit l'espace probabilisé  $(\Omega, \mathcal{F}, \mathbb{P})$ . Une suite de variables aléatoires  $(Y_n, n \in \mathbb{N}^*)$  de  $L^1(\Omega, X)$  est ergodique si et seulement si

$$\frac{1}{n} \sum_{i=1}^n F(Y_i) \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[F(Y_0)]$$

pour toute fonction  $F$  mesurable et intégrable de  $X$  dans  $\mathbb{R}$ .

Or, d'après le théorème de Birkhoff (appliqué aux suites ergodiques et intégrables  $(\hat{\alpha}_i, i \in \mathbb{N})$  et  $(\hat{\sigma}_i, i \in \mathbb{N})$ ) :

$$\frac{1}{j} \sum_{i=1}^j (\hat{\alpha}_{-i} - \hat{\alpha}_{-i+1} - \hat{\sigma}_{-i}) \xrightarrow[j \rightarrow \infty]{} \mathbb{E}_S^0[\hat{\alpha} - \hat{\alpha} - \hat{\sigma}] = \mathbb{E}_S^0[\hat{\sigma}] \mathbf{p.s.}$$

Le dernier terme est strictement négatif et donc en définissant pour tous  $N \in \mathbb{N}$  et  $\eta > 0$ , l'évènement

$$\mathcal{E}_{N,\eta} = \left\{ \forall j \geq N, \sum_{i=1}^j (\hat{\alpha}_{-i} - \hat{\alpha}_{-i+1} - \hat{\sigma}_{-i}) \leq j \left( -\mathbb{E}_S^0[\hat{\sigma}] + \eta \right) \right\},$$

pour tout  $\varepsilon > 0$ , il existe  $N \in \mathbb{N}$  et  $\eta > 0$ , tels que :

$$\mathbb{P}_{\hat{S}}^0[\mathcal{E}_{N,\eta}] \geq 1 - \varepsilon$$

Pour  $\varepsilon$  tel que  $-\mathbb{E}_{\hat{S}}^0[\hat{\sigma}] + \eta < 0$ , sur l'évènement  $\varepsilon_{N,\eta}$ , pour tout  $j \geq N$ ,

$$\sum_{i=1}^j (\hat{\alpha}_{-i} - \hat{\alpha}_{-i+1} - \hat{\sigma}_{-i}) + \hat{\alpha}_0 \leq j(-\mathbb{E}_{\hat{S}}^0[\hat{\sigma}] + \eta) < 0$$

et donc

$$M_\infty = \sup_{j \in \mathbb{N}} \sum_{i=1}^j (\hat{\alpha}_{-i} - \hat{\alpha}_{-i+1} - \hat{\sigma}_{-i}) + \hat{\alpha}_0 < \infty.$$

Donc, pour tout  $\varepsilon > 0$ ,

$$\mathbb{P}_{\hat{S}}^0[M_\infty < \infty] \geq 1 - \varepsilon,$$

ou encore :

$$\mathbb{P}_{\hat{S}}^0[M_\infty = \infty] \leq \varepsilon.$$

Comme l'évènement  $\{M_\infty = \infty\}$  est  $\theta$ -invariant, par l'ergodicité du flot  $\theta$ , on conclut :

$$\mathbb{P}_{\hat{S}}^0[M_\infty = \infty] = 0.$$

Donc, on a trouvé une version stationnaire de la suite  $(Y_n, n \in \mathbb{N})$ , **i.e.** une solution de (3.2) : il s'agit de la variable aléatoire **p.s.** finie

$$Y = M_\infty.$$

Nous prouvons finalement que  $M_\infty$  est l'unique solution de (3.2). Soit  $Z$  une autre solution finie.

D'une part,  $Z \geq 0 = M_0$  et si on suppose que  $Z \geq M_n \mathbb{P}_{\hat{S}}^0 - \mathbf{p.s.}$ , alors,

$$Z \circ \theta = [\max\{Z, \hat{\alpha}\} - \hat{\sigma}]^+ \geq [\max\{M_n, \hat{\alpha}\} - \hat{\sigma}]^+ = M_{n+1} \circ \theta, \mathbb{P}_{\hat{S}}^0 - \mathbf{p.s.}$$

et donc  $Z \geq M_{n+1} \mathbb{P}_{\hat{S}}^0 - \mathbf{p.s.}$

Finalement, pour tout  $n$ ,  $\mathbb{P}_{\hat{S}}^0[Z \geq M_n] = 1$  et on passe à la limite :

$$\mathbb{P}_{\hat{S}}^0[Z \geq M_\infty] = 1 \tag{3.7}$$

Par ailleurs, sur l'évènement  $\{Z \leq M_\infty\}$ ,

$$\begin{aligned} Z \circ \theta &= [Z - \hat{\sigma}]^+ \mathbb{1}_{\{Z > \hat{\alpha}\}} + [\hat{\alpha} - \hat{\sigma}]^+ \mathbb{1}_{\{Z \leq \hat{\alpha}\}} \\ &\leq [M_\infty - \hat{\sigma}]^+ \mathbb{1}_{\{Z > \hat{\alpha}\}} + [\hat{\alpha} - \hat{\sigma}]^+ \mathbb{1}_{\{Z \leq \hat{\alpha}\}} \\ &\leq [M_\infty \circ \theta]^+ \mathbb{1}_{\{Z > \hat{\alpha}\}} + [M_\infty \circ \theta]^+ \mathbb{1}_{\{Z \leq \hat{\alpha}\}} = M_\infty \circ \theta. \end{aligned}$$

L'évènement  $\{Z \leq M_\infty\}$  est donc  $\theta$ -contractant, et donc de probabilité 0 ou 1 en vertu de l'ergodicité du flot. Comme

$$\{Z \leq \alpha\} \subseteq \{Z \circ \theta \leq M_\infty \circ \theta\},$$

(3.3) implique que

$$\mathbb{P}_{\hat{S}}^0[Z \circ \theta \leq M_\infty \circ \theta] = \mathbb{P}_{\hat{S}}^0[Z \leq M_\infty] > 0$$

ce qui donne finalement :

$$\mathbb{P}_{\hat{S}}^0[Z \leq M_\infty] = 1$$

soit avec (3.7) :

$$\mathbb{P}_{\hat{S}}^0[Z = M_\infty] = 1$$

Nous déduisons de ce résultat deux corollaires importants sur le comportement asymptotique de la suite  $(\hat{Y}_n, n \in \mathbb{N})$  :

**Corollaire 3.2.1** *Partant de 0,*

$$\hat{Y}_n^0 \xrightarrow[n \rightarrow \infty]{\mathcal{L}} Y$$

**Preuve.** On peut en particulier donner la forme développée de  $Y_n^0$  pour tout  $n \in \mathbb{N}$  :

$$\hat{Y}_n^0 = \left[ \max_{j=0}^{n-1} \left( \hat{\alpha}_j - \sum_{i=j}^{n-1} \hat{\sigma}_i \right) \right]^+. \quad (3.8)$$

En effet, pour le premier terme, la définition (3.2.1) implique :

$$\hat{Y}_1^0 = [\max\{0, \hat{\alpha}_0\} - \hat{\sigma}_0]^+ = [\hat{\alpha}_0 - \hat{\sigma}_0]^+$$

et (3.8) est vérifiée.

Ensuite, par récurrence, si  $Y_n^0$  satisfait (3.8), alors :

$$\begin{aligned}
\hat{Y}_{n+1}^0 &= \left[ \max \left\{ \left[ \max_{j=0}^{n-1} \left( \hat{\alpha}_j - \sum_{i=j}^{n-1} \hat{\sigma}_i \right) \right]^+, \hat{\alpha}_n \right\} - \hat{\sigma}_n \right]^+ \\
&= \left[ \max \left\{ 0, \max_{j=0}^{n-1} \left( \hat{\alpha}_j - \sum_{i=j}^{n-1} \hat{\sigma}_i \right), \hat{\alpha}_n \right\} - \hat{\sigma}_n \right]^+ \\
&= \left[ \max \left\{ -\hat{\sigma}_n, \max_{j=0}^{n-1} \left( \hat{\alpha}_j - \sum_{i=j}^{n-1} \hat{\sigma}_i \right) - \hat{\sigma}_n, \hat{\alpha}_n - \hat{\sigma}_n \right\} \right]^+ \\
&= \left[ \max \left\{ \max_{j=0}^{n-1} \left( \hat{\alpha}_j - \sum_{i=j}^n \hat{\sigma}_i \right), \hat{\alpha}_n - \hat{\sigma}_n \right\} \right]^+ \\
&= \left[ \max_{j=0}^n \left( \hat{\alpha}_j - \sum_{i=j}^n \hat{\sigma}_i \right) \right]^+.
\end{aligned}$$

Montrons maintenant par récurrence que pour tout  $n$  :

$$M_n^0 = \hat{Y}_n^0 \circ \theta^{-n}.$$

D'une part, pour  $n = 0$ , on a clairement :

$$M_0 = 0 = \hat{Y}_0^0.$$

Ensuite, en supposant que pour  $n$ , la propriété soit vraie, on a :

$$\begin{aligned}
M_{n+1} &= \left[ \max \left\{ M_n \circ \theta^{-1}, \hat{\alpha}_{-1} \right\} - \hat{\sigma}_{-1} \right]^+ \\
&= \left[ \max \left\{ \hat{Y}_n^0 \circ \theta^{-n-1}, \hat{\alpha}_{-1} \right\} - \hat{\sigma}_{-1} \right]^+ \\
&= \left[ \max \left\{ \hat{Y}_n^0, \hat{\alpha}_n \right\} - \hat{\sigma}_n \right]^+ \circ \theta^{-n-1} \\
&= \left[ \max \left\{ \max_{j=0}^{n-1} \left( \hat{\alpha}_j - \sum_{i=1}^{n-1} \hat{\sigma}_i \right), \hat{\alpha}_n \right\} - \hat{\sigma}_n \right]^+ \circ \theta^{-n-1} \\
&= \left[ \max \left\{ \max_{j=0}^{n-1} \left( \hat{\alpha}_j - \sum_{i=1}^n \hat{\sigma}_i \right), \hat{\alpha}_n - \hat{\sigma}_n \right\} \right]^+ \circ \theta^{-n-1} \\
&= \left[ \max_{j=0}^n \left( \hat{\alpha}_j - \sum_{i=1}^n \hat{\sigma}_i \right) \right]^+ \circ \theta^{-n-1} \\
&= Y_{n+1}^0 \circ \theta^{-(n+1)}.
\end{aligned}$$

La propriété est donc vraie au rang  $n + 1$ . Elle est donc vérifiée pour tout  $n$ . Maintenant, soit  $F$ , une fonction réelle continue et bornée. Alors :

$$\begin{aligned} \mathbb{E}_{\hat{S}}^0[F(\hat{Y}_n^0)] &= \mathbb{E}_{\hat{S}}^0[F(\hat{Y}_n^0 \circ \theta^{-n})] \\ &= \mathbb{E}_{\hat{S}}^0[F(M_n)] \xrightarrow{n \rightarrow \infty} \mathbb{E}_{\hat{S}}^0[F(M_\infty)] \\ &= \mathbb{E}_{\hat{S}}^0[F(Y)]. \end{aligned}$$

□

### Corollaire 3.2.2

*A la condition*

$$\mathbb{P}_{\hat{S}}^0 \left[ \sup_{j \in \mathbb{N}^*} \left( \hat{\alpha}_{-j} - \sum_{i=1}^j \hat{\sigma}_{-i} \right) \leq 0 \right] > 0 \quad (3.9)$$

il existe  $\mathbb{P}_{\hat{S}}^0 - \mathbf{p.s.}$  une infinité d'indices tels que  $\hat{Y}_n^Y = Y \circ \theta^n = 0$ .

**Preuve.** La condition (3.2.2) équivaut à

$$\mathbb{P}_{\hat{S}}^0[Y = 0] > 0$$

où  $Y = M_\infty$  est donnée par (3.5). La suite  $(\hat{Y}_n^Y, n \in \mathbb{N})$  " retombe " donc  $\mathbb{P}_{\hat{S}}^0 - \mathbf{p.s.}$  une infinité de fois en 0 puisque 0 est chargé par sa loi stationnaire.

□

### 3.2.2 Application à la file $G/G/1/1 + G$

Nous considérons une file d'attente avec clients impatients  $G/G/1/1 + G$  régie par une discipline de service non-préemptive quelconque. Comme défini en chapitre d'introduction, nous notons par  $Z_t$ , le délai résiduel à l'instant  $t$  du client le " moins pressé " présent dans le buffer à cet instant. Nous majorons ici aux instants de fins de service, le processus  $(Z_t, t \geq 0)$  par une méthode de type " borne stochastique ". Nous montrons alors la stationnarité du processus majorant grâce au théorème précédent et donc, celle du processus  $(Z_t, t \geq 0)$ .

Rappelons que l'on note par  $\tilde{T}_n$ , le  $n$ -ième instant de sortie d'un client du service et  $(S_t, t \geq 0)$ , le processus des sorties du serveur, c'est-à-dire :

$$S_t = \sum_{\mathbb{N}} \mathbb{1}_{\{\tilde{T}_n \leq t\}} \text{ pour tout } t.$$

Soit pour tout  $n$ ,  $\tilde{\sigma}_n$  la durée du  $n$ -ième service rendu par le serveur. Il est clair, comme la suite des temps de service demandés dans l'ordre des arrivées  $(\sigma_n, n \in \mathbb{N})$  est indépendante des autres paramètres du système et en particulier de la discipline de service, que

$$(\sigma_n, n \in \mathbb{N}) \stackrel{\mathcal{L}}{=} (\tilde{\sigma}_n, n \in \mathbb{N}).$$

En d'autres termes, la "statistique" du système (loi des variables aléatoires et processus qui le caractérisent) est la même suivant que c'est le serveur ou le client qui décide de la durée du service qui commence. Nous considérons alors que c'est le serveur qui décide de la durée de chaque service qui commence, sans perte de généralités :

$$\text{pour tout } n, \text{ durée du } n\text{-ième service} = \tilde{\sigma}_n.$$

Alors,  $\tilde{\sigma}_n$  est de même loi que la durée de service d'un client, **i.e.** celle de la **v.a.**  $\sigma$ .

### Lemme 3.2.3

Soit pour tout  $n$  dans  $\mathbb{N}$  ;

$$\mathcal{A}_n = \max_{k/T_k \in (\tilde{T}_n - \tilde{\sigma}_n, \tilde{T}_n]} \{(D_k - (\tilde{T}_n - T_k))^+\} \mathbb{1}_{\{N(\tilde{T}_n - \tilde{\sigma}_n, \tilde{T}_n) > 0\}}.$$

Alors,  $\mathcal{A}_n$  désigne le plus grand délai résiduel à  $\tilde{T}_n$  des clients entrés dans le système durant le  $n$ -ième service et conservés à l'instant  $\tilde{T}_n$  si il y en a, ou 0 sinon.

#### Preuve.

Si au moins un client est entré pendant le  $n$ -ième service (à la condition  $N(\tilde{T}_n - \tilde{\sigma}_n, \tilde{T}_n) > 0$ ), le plus grand délai résiduel à  $\tilde{T}_n$  parmi ces clients nouveaux entrés sera donné par le plus grand des  $(D_k - (\tilde{T}_n - T_k))^+$ .

Si la précédente quantité est nulle, cela signifie que des clients sont entrés dans le système durant le  $n$ -ième service, mais qu'ils ont tous atteint leur délai à  $\tilde{T}_n$ .  $\square$

Définissons pour tout  $n$  dans  $\mathbb{N}$  la quantité  $\alpha_n$  par :

$$\mathcal{A}_n = [\alpha_n - \tilde{\sigma}_n]^+,$$

ou autrement dit :

$$\alpha_n = \max_{k/T_k \in (\tilde{T}_n - \tilde{\sigma}_n, \tilde{T}_n]} \{D_k + T_k - (\tilde{T}_n - \tilde{\sigma}_n)\} \mathbb{1}_{\{N(\tilde{T}_n - \tilde{\sigma}_n, \tilde{T}_n) > 0\}}.$$

On peut remarquer que

$$\alpha_n \geq 0 \text{ p.s. pour tout } n \in \mathbb{N}, \tag{3.10}$$

puisque pour tout  $k$  tel que  $T_k \in (\tilde{T}_n - \tilde{\sigma}_n, \tilde{T}_n]$ ,  $T_k + \tilde{\sigma}_n > \tilde{T}_n$ .

**Définition 3.2.2**

Pour tout  $n \in \mathbb{N}$ , nous notons

$$\tilde{Z}_n = Z_{\tilde{T}_n},$$

le plus grand délai résiduel à  $\tilde{T}_n$  d'un client présent dans le système à cet instant.

En particulier, en rappelant (cf chapitre 1) que

$$R_1(t) < R_2(t) < \dots < R_{Q(t)}(t)$$

désigne la suite des délais résiduels à  $t$  des clients présents dans le buffer à cet instant, on a alors :

$$\tilde{Z}_n = R_{Q_{\tilde{T}_n}}(\tilde{T}_n) \mathbb{1}_{\{Q_{\tilde{T}_n} > 0\}} = R_{X_{\tilde{T}_n}}(\tilde{T}_n) \mathbb{1}_{\{X_{\tilde{T}_n} > 0\}}, \quad (3.11)$$

puisque  $\tilde{T}_n$  étant l'instant d'une fin de service, à cet instant la population du buffer est la population de système tout entier .

En d'autres termes, la suite  $(\tilde{Z}_n)$  caractérise la quantité suivante :

- si plus d'un client est dans le système à l'instant  $\tilde{T}_n$  (en particulier,  $\tilde{T}_n + 1$  sera égal à  $\tilde{T}_n + \tilde{\sigma}_{n+1}$ , puisque le système ne se vide pas à  $\tilde{T}_n$ ),  $\tilde{Z}_n$  indique le plus grand délai résiduel des clients dans le système à  $\tilde{T}_{n-1}$ , le maximum étant pris y compris sur le client qui commence à être servi,
- si  $X_{\tilde{T}_n} = 1$ , il n'y a qu'un seul client dans le système à  $\tilde{T}_n$  et  $\tilde{Z}_n$  est le délai résiduel à cet instant de ce client qui précisément va être servi et à cet instant son délai ne compte plus. Il faudra l'effacer du descripteur par la suite,
- si  $X_{\tilde{T}_n} = 0$  il n'y a aucun client dans le système à la fin du  $n$ -ième service, le système est vide et  $\tilde{Z}_n = 0$ .

Par conséquent, on a :

$$X_{\tilde{T}_n} = 0 \Leftrightarrow \tilde{Z}_n = 0 \text{ pour tout } n \in \mathbb{N}. \quad (3.12)$$

**Lemme 3.2.4**

Pour toute discipline de service,

$$\tilde{Z}_{n+1} \leq [\max\{\tilde{Z}_n \mathbb{1}_{\{X_{\tilde{T}_n} \neq 1\}}, \alpha_{n+1}\} - \tilde{\sigma}_n]^+. \quad (3.13)$$

De plus, les deux termes sont égaux sous EDF.

**Preuve.** Nous nous plaçons dans les trois cas possibles pour la récurrence, suivant que  $X_{\tilde{T}_n}$  est égal à 0, 1 ou  $\geq 2$  :

- Si  $X_{\tilde{T}_n} \geq 2$ , en particulier,  $\tilde{Z}_n \neq 0$ , le système n'est pas vide à  $\tilde{T}_n$  et  $\tilde{T}_{n+1} = \tilde{T}_n + \tilde{\sigma}_n$ . Alors, le plus grand délai résiduel des clients dans le système à  $\tilde{T}_{n+1}$  est inférieur ou égal au  $\max\{\tilde{Z}_n - \tilde{\sigma}_n, \mathcal{A}_{n+1}\}$ . Il y a égalité dans le cas d'EDF, puisqu'alors le client relatif à ce maximum, s'il était déjà dans le système à  $\tilde{T}_n$  (et donc, il correspond à  $\tilde{Z}_n - \tilde{\sigma}_n$ ), n'a pas été servi à cet instant et est donc toujours dans le système à  $\tilde{T}_{n+1}$ . Dans tous les cas,  $\tilde{T}_{n+1} = \tilde{T}_n + \tilde{\sigma}_n$  et on a :

$$\tilde{Z}_{n+1} \leq \max\{\tilde{Z}_n - \tilde{\sigma}_n, \mathcal{A}_{n+1}\} = [\max\{\tilde{Z}_n, \alpha_{n+1}\} - \tilde{\sigma}_n]^+$$

- Si  $X_{\tilde{T}_n} = 1$ , le seul client dans le système à l'instant  $\tilde{T}_n$  commence son service à cet instant et donc son délai résiduel  $\tilde{Z}_n$  à  $\tilde{T}_n$  doit disparaître du descripteur puisqu'il ne sera plus en concurrence avec les autres clients à  $\tilde{T}_{n+1}$ , date de la fin de son service. Dans ce cas le plus grand délai résiduel à  $\tilde{T}_{n+1}$ , s'il y a au moins un client dans le système, est apporté par un client entré pendant le  $n + 1$ -ième service. Ici aussi,  $\tilde{T}_{n+1} = \tilde{T}_n + \tilde{\sigma}_n$  et on a :

$$\tilde{Z}_{n+1} = \mathcal{A}_{n+1} = [\alpha_{n+1} - \tilde{\sigma}_n]^+.$$

- Si  $X_{\tilde{T}_n} = 0$ , le système est vide à  $\tilde{T}_n$ . Dans ce cas, il faut attendre l'arrivée d'un client pour que le  $n + 1$ -ième service commence. Donc,  $\tilde{T}_{n+1} \neq \tilde{T}_n + \tilde{\sigma}_n$ , mais

$$\begin{aligned} & \max_{k/T_k \in (\tilde{T}_{n+1} - \tilde{\sigma}_{n+1}, \tilde{T}_{n+1}]} \left( \left\{ D_k + T_k - (\tilde{T}_{n+1} - \tilde{\sigma}_{n+1}) \right\} \cdot \mathbb{1}_{\{N(\tilde{T}_{n+1} - \tilde{\sigma}_{n+1}, \tilde{T}_{n+1}) > 0\}} \right) \\ &= \max_{k/T_k \in (\tilde{T}_n, \tilde{T}_{n+1}]} \left( \left\{ D_k + T_k - (\tilde{T}_{n+1} - \tilde{\sigma}_{n+1}) \right\} \cdot \mathbb{1}_{\{N(\tilde{T}_n, \tilde{T}_{n+1}) > 0\}} \right) \end{aligned}$$

puisque aucun client n'est entré entre  $\tilde{T}_n$  et  $\tilde{T}_{n+1} - \tilde{\sigma}_{n+1}$ . On a là-encore :

$$\mathcal{A}_{n+1} = [\alpha_{n+1} - \tilde{\sigma}_n]^+$$

donc ,

$$\tilde{Z}_{n+1} = \mathcal{A}_{n+1} = [\alpha_{n+1} - \tilde{\sigma}_n]^+$$

Nous donnons une condition pour la récurrence de la suite  $(\tilde{Z}_n, n \in \mathbb{N})$  en 0. Or, (3.13) induit deux difficultés. La première est que la récurrence est définie par une inégalité dans le cas général (discipline de service quelconque), la seconde est que le terme  $\mathbb{1}_{\{X_{\tilde{T}_n} \neq 0\}}$ , qui dépend de tout le

passé de la file et pas seulement de  $\tilde{Z}_n$ , montre que l'inégalité ne met pas directement en évidence une suite ergodique.

Nous exhibons donc une suite de variable aléatoire qui soit définie par une égalité de récurrence du type de l'équation de Lindley et qui permette par comparaison de donner le comportement asymptotique de  $(\tilde{Z}_n, n \in \mathbb{N})$ .

**Définition 3.2.3**

On définit la suite  $(Y_n, n \in \mathbb{N})$  par :

$$\begin{cases} Y_0 = \tilde{Z}_0, \\ Y_{n+1} = [\max\{Y_n, \alpha_n\} - \tilde{\sigma}_n]^+, \quad \text{pour tout } n \in \mathbb{N}. \end{cases}$$

En particulier, il est clair que :

$$Y_n \geq \tilde{Z}_n \quad \text{pour tout } n \in \mathbb{N} \quad (3.14)$$

avec une récurrence immédiate et l'inégalité (3.13).

Soit maintenant la suite  $(\hat{T}_n, n \in \mathbb{N})$  suivante :

$$\begin{cases} \hat{T}_0 = \tilde{T}_0, \\ \hat{T}_n = \hat{T}_{n-1} + \tilde{\sigma}_n, \quad \text{pour tout } n \in \mathbb{N}^*. \end{cases}$$

Nous notons par ailleurs le processus suivant :

$$\hat{S}_t = \sum_{\mathbb{N}} \mathbb{1}_{\{\hat{T}_n \leq t\}} \quad \text{pour tout } t.$$

La suite  $(\tilde{\sigma}_n, n \in \mathbb{N})$  étant stationnaire et ergodique,  $(\hat{S}_t, t \geq 0)$  l'est aussi et on peut considérer l'espace de Palm  $(\Omega, \mathcal{F}, \mathbb{P}_{\hat{S}}^0, \theta)$  relatif au processus  $(\hat{S}_t, t \geq 0)$ , en notant  $\theta = \theta_{\hat{T}_n}$  le flot stationnaire ergodique correspondant. Sous cette hypothèse,  $(\tilde{\sigma}_n, n \in \mathbb{N})$  est donc une suite de marques du processus  $\hat{S}$ . En particulier elle forme une suite ergodique, puisqu'elle est compatible avec le flot.  $\mathbb{P}_{\hat{S}}^0 - \mathbf{p.s.}$  :

$$\tilde{\sigma}_n(\omega) = \tilde{\sigma}(\theta_{\hat{T}_n}(\omega)) = \tilde{\sigma} \circ \theta^n(\omega)$$

où  $\tilde{\sigma}$  est la durée du service qui se termine  $\mathbb{P}_{\hat{S}}^0 - \mathbf{p.s.}$  à  $t = 0$ , de loi celle de  $\sigma$ .

Nous définissons par ailleurs la variable aléatoire

$$\hat{\alpha} = \max_{k/T_k \in (\hat{T}_0, \hat{T}_1]} \left\{ D_k + T_k - (\hat{T}_0) \right\} \cdot \mathbb{1}_{\{N(\hat{T}_0, \hat{T}_1) > 0\}}$$

Le processus des entrées marqué par les délais initiaux des clients  $(N, D)$  étant indépendant de  $\hat{S}$ , on peut construire la suite de marques de  $\hat{S}(\hat{\alpha}_n, n \in \mathbb{N})$  de la manière suivante, pour tout  $n \in \mathbb{N}$ ,  $\mathbb{P}_{\hat{S}}^0 - \mathbf{p.s.}$  :

$$\hat{\alpha}_n(\omega) = \hat{\alpha}(\theta_{T_n}(\omega)) = \hat{\alpha} \circ \theta^n(\omega)$$

On a en particulier :

$$\hat{\alpha}_n \geq 0 \quad \mathbb{P}_{\hat{S}}^0 - \mathbf{p.s.} \quad \text{pour tout } n \in \mathbb{N}.$$

Finalement, nous définissons pour toute  $\mathbf{v.a.} X$  :

$$\begin{cases} \hat{Y}_0^X = X, \\ \hat{Y}_{n+1}^X = [\max\{\hat{Y}_n^X, \hat{\alpha}_n\} - \tilde{\sigma}_n]^+, \quad \text{pour tout } n \in \mathbb{N}. \end{cases}$$

### Théorème 3.2.3

1. Partant de  $Y_0 = 0$ , Il existe une unique version stationnaire  $Y$   $\mathbf{p.s.}$  finie telle que :

$$Y_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} Y.$$

2. Si de plus

$$\mathbb{P}_{\hat{S}}^0 \left[ \sup_{j \in \mathbb{N}^*} \left( \hat{\alpha}_{-j} - \sum_{i=1}^j \tilde{\sigma}_{-i} \right) \leq 0 \right] > 0$$

la file d'attente se vide  $\mathbf{p.s.}$  une infinité de fois.

### Preuve.

1. Les hypothèses du théorème (3.2.1), 1 sont vérifiées pour le processus  $(\hat{S}_t, t \geq 0)$  et les suites  $(\hat{\alpha}_n, n \in \mathbb{N})$  et  $(\tilde{\sigma}_n, n \in \mathbb{N})$ . En particulier on peut appliquer le corollaire (3.2.1) : il existe une unique variable aléatoire  $Y$   $\mathbf{p.s.}$  finie telle que :

$$\hat{Y}_n^0 \xrightarrow[n \rightarrow \infty]{\mathcal{L}} Y \tag{3.15}$$

Remarquons maintenant que  $(\hat{\alpha}_n, n \in \mathbb{N})$  a la même loi que  $(\alpha_n, n \in \mathbb{N})$ . Donc, partant de  $Y_0 = \tilde{Z}_0 = 0$ , la suite  $(\hat{Y}_n^0, n \in \mathbb{N})$  a la même loi que  $(Y_n, n \in \mathbb{N})$  et on conclut avec (3.15).

2. La condition (3.9) est satisfaite et on peut appliquer le corollaire (3.2.2) : il existe  $\mathbb{P}_{\hat{S}}^0 - \mathbf{p.s.}$  une infinité d'indices tels que  $\hat{Y}_n^Y = 0$ . Partons de l'état  $Y_0 = \tilde{Z}_0 = Y$ . D'une part, comme

la suite  $(\hat{Y}_n^Y, n \in \mathbb{N})$  la même loi que  $(Y_n, n \in \mathbb{N})$ , il y a **p.s.** une infinité d'indices tels que  $Y_n = 0$ . D'autre part, à tout instant  $\tilde{T}_n$  et pour toute discipline de service :

$$\{Y_n = 0\} \subseteq \{\tilde{Z}_n = 0\} = \{X_{\tilde{T}_n} = 0\}$$

avec (3.12) et (3.14). 0 est donc récurrent pour la suite  $(X_{\tilde{T}_n})$  et il l'est aussi pour le processus  $(X_t, t \geq 0)$  (par définition,  $X_t$  ne peut valoir 0 qu'à un instant de fin de service).

### Remarque 3.2.1

Il est intéressant de rapprocher ce résultat à la condition de stabilité de la file  $G/G/1/1$ , donnée par [18] ([3], propriété (2.6.1). p.123), de la même forme. Cependant, dans la file  $G/G/1/1$ , lorsque cette condition n'est pas vérifiée, l'unicité, et même l'existence de la solution stationnaire ne sont pas garanties ([3], exemples (2.6.1) et (2.6.2), p. 122). Elle le sont dans le cas  $G/G/1/1 + G..$

### Corollaire 3.1

La file d'attente avec clients impatientes  $M/GI/1/1 + GI$  se vide **p.s.** une infinité de fois pour toute discipline de service.

### Preuve

D'une part, il est immédiat de voir avec l'expression de  $\hat{\alpha}_n$  pour tout  $n$  que si le processus des entrées est un processus de Poisson et les temps de service et les délais forment deux suites **i.i.d.**, la suite  $(\hat{\alpha}_n, n \in \mathbb{N})$  est **i.i.d.**. D'autre part, pour tout  $n$ ,

$$\begin{aligned} \mathbb{P}_{\hat{S}}^0[\hat{\alpha} \leq \tilde{\sigma}] &\geq \mathbb{P}_{\hat{S}}^0[\hat{\alpha} = 0] \\ &\geq \mathbb{P}[N(\hat{T}_{n-1}, \hat{T}_n) = 0] = \mathbb{P}[N(\tilde{T}_n - \tilde{\sigma}_n, \tilde{T}_n) = 0] > 0 \end{aligned}$$

Donc,

$$\begin{aligned} \mathbb{P}_{\hat{S}}^0 \left[ \sup_{j \in \mathbb{N}^*} \left( \hat{\alpha}_{-j} - \sum_{i=1}^j \tilde{\sigma}_{-i} \right) \leq 0 \right] &= \mathbb{P}_{\hat{S}}^0 \left[ \sup_{j \in \mathbb{N}^*} \left( \hat{\alpha} - \sum_{i=1}^j \tilde{\sigma}_{-i} \right) \leq 0 \right] \\ &= \mathbb{P}_{\hat{S}}^0[\hat{\alpha} - \tilde{\sigma} \leq 0] > 0 \end{aligned}$$

On peut donc appliquer le théorème (3.2.3), partie (2).

### Remarque 3.2.2

Dans ce dernier cas, on a affaibli les hypothèses du théorème (3.1.2). Remarquons de plus que le théorème (3.2.3), 2 s'applique dans un cadre plus général, ses hypothèses étant plus larges.

## Chapitre 4

# Probabilité de perte

La probabilité de perte d'un client à l'état stationnaire (probabilité pour un client entrant dans le système à l'état stationnaire de ne pas atteindre le serveur avant l'expiration de son délai) est la principale métrique de performance d'un système à perte. L'estimation de cette probabilité par une formule close dépendant des autres paramètres du système peut permettre de répondre à des questions de dimensionnement : étant données des contraintes de trafic (flux des entrées et loi des délais par exemple), on peut en particulier déduire de cette formule les valeurs des paramètres (nombre de serveurs, débit de service), qui garantissent une exigence fixée en terme de perte pour une discipline de service donnée. Nous avons vu dans le chapitre (2) une formule explicite de cette probabilité pour un système  $M/M/s/s + G/FCFS$  ([32], par exemple).

Dans la première partie de ce chapitre, nous exposons dans quelques cas plus généraux, d'autres résultats et méthodes de calcul pour estimer la probabilité de perte.

L'autre aspect du problème est précisément le choix de la discipline de service : laquelle choisir afin de minimiser la probabilité de perte ? C'est l'objet de la deuxième partie de ce chapitre, qui présente quelques résultats de comparaison entre disciplines de service pour un jeu de paramètre donné. Ce faisant nous précisons et complétons certains résultats de Panwar, Towsley et Wolf dans [35] sur la comparaison de la perte entre les différentes disciplines de service pour toute file d'attente avec clients impatientes. Comme on l'a rappelé dans le chapitre (2), théorème (2.2.2), Panwar et *al.* ont montré qu'EDF est optimale en terme de perte pour toutes les disciplines de services non-préemptives, en le sens qu'à chaque instant  $t$ , la perte dans une file d'attente  $G/M/s/s + G/EDF$  est stochastiquement inférieure à celle endurée par le même système sous toute autre discipline de service non-préemptive. Nous donnons ici une autre preuve de ce résultat, en allant plus loin : nous montrons que la probabilité de perte à l'état stationnaire  $\pi$  est inférieure en EDF par rapport à toute autre discipline de service non-préemptive. Par le même

type de raisonnement, nous montrons également que la discipline LDF maximise la perte pour l'ordre stochastique et la probabilité de perte.

L'algorithme *EDF* est coûteux en terme d'ordonnancement : il requiert une réactualisation complète du buffer tout entier à chaque arrivée d'un client. Il est donc naturel de se demander si le gain réalisé en appliquant EDF est suffisant pour justifier ce coût. Ne disposant pas d'une formule close pour la probabilité de perte sous *EDF*, on cherche donc à estimer le gain réalisé en terme de probabilité de perte entre *EDF* et toute autre discipline de service non-préemptive. Nous proposons ici une première estimation.

Ce chapitre s'organise comme suit. En partie (4.1) nous présentons deux encadrements de la probabilité de perte pour une file avec clients impatients : un encadrement à forte charge ( $\rho > 1$ ) pour un système  $M/G/1/1 + G$  avec discipline de service quelconque en section (4.1.1) et un encadrement pour un système  $GI/M/1/1 + D/FIFO$  en section (4.1.2) . En partie (4.2), nous comparons les différentes disciplines de service du point de vue de la perte. Nous donnons en section (4.2.1) une nouvelle preuve du résultat de Panwar et *al.* ([35]) qui donne *EDF* comme discipline optimale en terme de perte dans le cas  $G/M/1/1 + G$  pour l'ordre stochastique. Nous montrons également que *LDF* maximise la perte pour l'ordre stochastique. Nous donnons en section (4.2.2) un encadrement de la probabilité de perte sous toute discipline de service non-préemptive (et donc pour *FIFO*, par exemple) en fonction de la probabilité de perte sous *EDF* pour un système  $M/M/1/1 + G$ . Nous montrons au passage qu'*EDF* est dans ce cas optimal pour la probabilité de perte, et donnons un majorant du gain d'*EDF* en terme de probabilité de perte.

## 4.1 Deux encadrements

### 4.1.1 Encadrement à forte charge

Nous proposons un premier encadrement, pour une file d'attente avec clients impatients régie par une discipline de service quelconque, de la probabilité de perte dans le cas où la charge du système est supérieure à 1.

#### Lemme 4.1.1

Soit  $\gamma$  une discipline de service quelconque. Pour toute file d'attente  $G/G/1/1 + G/\gamma$  soumise à la charge  $\rho > 1$ ,

$$\pi > \frac{\rho - 1}{\rho}.$$

**Preuve.** Soit  $N^s$ , le processus des arrivées des clients qui seront servis.  $N^s$  est toujours jointement stationnaire avec  $N$  et son intensité est donnée par :

$$\lambda^s = \lambda(1 - \pi).$$

Considérons le " sous-système " restreint aux clients qui seront servis. Le sous-système est donc une file  $G/G/1$ , avec un processus stationnaire d'intensité  $\lambda(1 - \pi)$  en entrée, stable puisque le système d'ensemble est stable lui-même. D'après le théorème de Loynes, on a donc :

$$\lambda(1 - \pi)\mathbb{E}[\sigma] < 1$$

et autrement dit,

$$1 - \frac{1}{\rho} < \pi.$$

□

On a par ailleurs :

**Lemme 4.1.2**

*Pour toute discipline de service  $\gamma$ , pour toute file d'attente  $M/G/1/1 + G/\gamma$  soumise à la charge  $\rho$ ,*

$$\pi < \frac{\rho}{\rho + 1}.$$

**Preuve.** La probabilité de perte d'un client est inférieure à celle d'une file d'attente classique sans buffer  $M/G/1/1$  (les clients ne sont servis que si ils entrent dans un système vide). Pour un tel système, la formule d'Erlang-B donne la probabilité de perte  $\pi (M/G/1/1)$  :

$$\pi (M/G/1/1) = \frac{\rho}{\rho + 1}$$

et on a le résultat. □

Nous pouvons donc déduire des deux lemmes précédents un encadrement de la probabilité de perte pour toute file d'attente  $M/G/1/1 + G$ , avec discipline de service quelconque :

**Théorème 4.1.1**

*Pour toute discipline de service  $\gamma$ , la probabilité de perte de la file  $M/G/1/1 + G/\gamma$  soumise à une charge  $\rho > 1$  vérifie*

$$\frac{\rho - 1}{\rho} < \pi < \frac{\rho}{\rho + 1}. \quad (4.1)$$

Comme on peut s'y attendre, la probabilité de perte tend alors vers 1 quand la charge tend vers l'infini, puisque le système tend à s'engorger, et un client à avoir très peu de chances d'être choisi.

Nos simulations indiquent qu'à  $\rho > 1$  fixé, la probabilité de perte tend vers la borne supérieure de l'intervalle précédent lorsque  $1/\mathbb{E}[D]$  tend vers l'infini, donc en cas de délais d'exécution très petits et par-là même, de forte perte.

#### 4.1.2 Cas $GI/M/1/1 + D/FIFO(EDF)$

Nous nous plaçons ici dans un cas sensiblement plus complexe, où on ne suppose plus que le processus des entrées est un processus de Poisson, mais simplement un processus de renouvellement. Nous supposons par ailleurs dans cette section que les délais initiaux sont déterministes.

Nous considérons donc une file d'attente  $GI/M/1/1 + D/FIFO$ , où :

- les clients entrent dans le système suivant un processus de renouvellement. Autrement dit, les interarrivées  $U_{n+1} = T_{n+1} - T_n$  sont indépendantes et identiquement distribuées,
- Les clients demandent un service de durée exponentielle  $\sigma \sim \varepsilon(\mu)$ ,
- les délais éliminatoires des clients sont déterministes, initialement tous égaux à  $d$ .

Remarquons tout d'abord que dans ce cas, les disciplines de service  $EDF$  et  $FIFO$  sont équivalentes. En effet, le délai résiduel du client  $C_k$  à un instant  $t$  auquel il est encore en attente dans le système est donné par :

$$\begin{aligned} DR_k(t) &= D_k - (t - T_k) \\ &= d + T_k - t. \end{aligned}$$

Ainsi :

$$\left\{ \text{le client } C_i \text{ est prioritaire à } t \text{ sur le client } C_j \text{ sous EDF} \right\} \iff (*)$$

$$(*) \iff 0 < DR_i(t) < DR_j(t)$$

$$\iff 0 < d + T_i - t < d + T_j - t$$

$$\iff T_i < T_j, \text{ et } C_i \text{ et } C_j \text{ sont en attente à } t$$

$$\iff \left\{ \text{le client } C_i \text{ est prioritaire à } t \text{ sur le client } C_j \text{ sous FIFO} \right\}.$$

Le résultat donné ici, exprimé pour une file gérée en *FIFO*, est par conséquent toujours valable si la file est gérée en *EDF* : il s'agit exactement de la même file d'attente.

**Remarque 4.1.1**

*On montre de la même façon que dans ce cas où les délais sont déterministes, la discipline LDF revient à la discipline LIFO.*

Nous donnons un encadrement de la probabilité de perte d'un client à l'état stationnaire. Alors que l'absence de mémoire de la loi exponentielle permettait dans le cas  $M/M/s/s + G/FCFS$  d'exhiber des modèles markoviens simples pour le calcul de performances et donnaient en particulier une formule close pour la probabilité de perte, un tel cadre d'étude ne fonctionne plus ici et il faut trouver un cadre *ad hoc* pour pouvoir estimer cette probabilité.

Rappelons que l'on note pour tout  $k$ ,

$$\pi_k = \mathbb{P}[\text{le client } C_k \text{ est perdu }].$$

Nous donnons une équation de récurrence simple pour  $(\pi_k, k \in \mathbb{N})$  et en déduisons la probabilité de perte en passant à la limite sur l'expression générale de  $\pi_k$ .

**Théorème 4.1.2**

*La probabilité de perte à l'état stationnaire d'une file d'attente  $GI/M/1/1 + D/FIFO$  (ou *EDF*) vérifie :*

$$\frac{\mathbb{P}[\sigma > U]e^{-\mu d}}{1 - \mathbb{P}[\sigma > U](1 - e^{-\mu d})} \leq \pi \leq \mathbb{P}[\sigma > U].$$

**Preuve.** Nous notons  $T'_k$ , l'instant où le client  $C_k$  quitte la file d'attente. S'il n'est pas perdu,  $T'_k$  est donc l'instant où il entre en service et  $T'_k$  se trouve dans l'intervalle de temps  $[T_k, T_k + d[$ . Sinon, on a  $T'_k = T_k + d$ .

Notons par ailleurs les évènements suivants :

$$\left\{ \begin{array}{l} I_j^t = \left\{ C_j \text{ est en service à l'instant } t \right\}, \\ G_j = \left\{ C_j \text{ est servi} \right\}. \end{array} \right.$$

Le client  $C_k$  est éliminé lorsque le serveur est occupé à l'instant  $T_k + d$ . Cet évènement est réalisé si l'un des clients  $C_0, C_1, \dots, C_{k-1}$  entrés avant  $C_k$  est en service à l'instant  $T_k + d$ . Dans ce cas,

ce client en service à  $T_k + d$  l'était déjà à  $T_{k-1} + d$ , sinon il aurait été éliminé à cet instant :

$$\begin{aligned}\pi_k &= \sum_{j=0}^{k-1} \mathbb{P} \left[ I_j^{T_k+d} \right] \\ &= \sum_{j=0}^{k-1} \mathbb{P} \left[ I_j^{T_k+d} \cap I_j^{T_{k-1}+d} \right] \\ &= \sum_{j=0}^{k-1} \mathbb{P} \left[ I_j^{T_k+d} \mid I_j^{T_{k-1}+d} \right] \mathbb{P} \left[ I_j^{T_{k-1}+d} \right].\end{aligned}$$

Les temps de service étant exponentiels, la probabilité qu'un client soit encore en service à  $T_k + d$  sachant qu'il était déjà en service à  $T_{k-1} + d$  est donnée par  $\mathbb{P}[\sigma > T_k - T_{k-1}] = \mathbb{P}[\sigma > U]$ . L'égalité précédente donne donc :

$$\pi_k = \sum_{j=0}^{k-1} \mathbb{P}[\sigma > U] \mathbb{P} \left[ I_j^{T_{k-1}+d} \right] = \mathbb{P}[\sigma > U] \pi_{k-1} + \mathbb{P}[\sigma > U] \mathbb{P} \left[ I_{k-1}^{T_{k-1}+d} \right]. \quad (4.2)$$

Comme  $I_{k-1}^{T_{k-1}+d} \subset G_{k-1}$ , on a alors :

$$\pi_k \leq \mathbb{P}[\sigma > U] \pi_{k-1} + \mathbb{P}[\sigma > U] (1 - \pi_{k-1}) = \mathbb{P}[\sigma > U].$$

On passe à la limite sur  $k$  dans l'inégalité précédente. Comme on atteint un jour l'état stationnaire,

$$\lim_{k \rightarrow \infty} \pi_k = \pi.$$

et on a la borne supérieure pour  $\pi$ .

Considérons maintenant la borne inférieure. (4.2) donne pour tout client  $C_k$  :

$$\begin{aligned}\pi_k &= \mathbb{P}[\sigma > U] \pi_{k-1} + \mathbb{P}[\sigma > U] \mathbb{P} \left[ I_{k-1}^{T_{k-1}+d} \right] \\ &= \mathbb{P}[\sigma > U] \left( \pi_{k-1} + \mathbb{P} [G_{k-1}] - \mathbb{P} \left[ \text{le service de } C_{k-1} \text{ se termine avant } T_{k-1} + d \right] \right) \\ &= \mathbb{P}[\sigma > U] \left( \pi_{k-1} + 1 - \pi_{k-1} - \mathbb{P} \left[ G_{k-1} \cap \{T_{k-1}^\wedge + \sigma_{k-1} < T_{k-1} + d\} \right] \right) \\ &\geq \mathbb{P}[\sigma > U] \left( 1 - (1 - \pi_{k-1}) \mathbb{P}[\sigma_{k-1} < d] \right) \\ &= \mathbb{P}[\sigma > U] - \mathbb{P}[\sigma > U] \mathbb{P}[\sigma < d] (1 - \pi_{k-1}).\end{aligned}$$

En d'autres termes, pour tout  $k$  :

$$\pi_k \geq \mathbb{P}[\sigma > U] \mathbb{P}[\sigma > d] + \mathbb{P}[\sigma > U] \mathbb{P}[\sigma < d] \pi_{k-1}.$$

et en passant à la limite sur  $k$  :

$$\pi \geq \mathbb{P}[\sigma > U] \mathbb{P}[\sigma > d] + \mathbb{P}[\sigma > U] \mathbb{P}[\sigma < d] \pi.$$

La minoration est démontrée. □

Ce résultat complète celui de [30], (2.2), qui ne donnait qu'une borne supérieure pour la probabilité de perte. Par ailleurs, il est clairement beaucoup plus fin que l'encadrement (4.1) et présente l'avantage d'être vrai pour des interarrivées **i.i.d.** de loi quelconque et pour n'importe quelle charge.

Précisons à quoi revient cet encadrement dans trois cas de lois pour les interarrivées, déterministe, uniforme et exponentielle :

#### Corollaire 4.1

1. Dans une file  $D/M/1/1 + D/FIFO$  (EDF), où les arrivées sont périodiques de période  $p$  (autrement dit  $U = p$  **p.s.**),

$$\frac{e^{-\mu(p+d)}}{1 - e^{-\mu}(1 - e^{-\mu d})} \leq \pi \leq e^{-\mu p}. \quad (4.3)$$

2. Dans une file  $GI/M/1/1 + D/FIFO$  (EDF), où les interarrivées forme une suite de variables uniformément distribuées sur  $[0, M]$ ,

$$\frac{\frac{1-e^{-\mu M}}{\mu M} e^{-\mu d}}{1 - \frac{1-e^{-\mu M}}{\mu M} (1 - e^{-\mu d})} \leq \pi \leq \frac{1 - e^{-\mu M}}{\mu M}. \quad (4.4)$$

3. Dans une file  $M/M/1/1 + D/FIFO$  (EDF), où le processus des arrivées est un processus de Poisson de paramètres  $\lambda$ ,

$$\frac{\frac{\lambda}{\lambda+\mu} e^{-\mu d}}{1 - \frac{\lambda}{\lambda+\mu} (1 - e^{-\mu d})} \leq \pi \leq \frac{\lambda}{\lambda + \mu}. \quad (4.5)$$

On peut remarquer en particulier que le dernier résultat (4.5) est cohérent avec celui de [9], (2.7), puisque les bornes précédentes encadrent la valeur exacte de  $\pi$ .

## 4.2 Comparaison entre disciplines de service

Plaçons-nous maintenant dans le cas général stationnaire pour la loi des délais initiaux des clients. Ici, les disciplines *EDF* et *FIFO* par exemple, ne sont a priori plus équivalentes, et nous cherchons à les comparer du point de vue de la perte. Dans cette partie, nous redémontrons le résultat de Panwar et al. ([35]), théorème (2.2.2), qui donne l'optimalité d'*EDF* en terme de perte pour l'ordre stochastique. Nous montrons également que *LDF* est la moins bonne discipline de service avec ce critère. En outre, nous allons plus loin : nous montrons pour un système équivalent qu'*EDF* a la plus faible probabilité de perte à l'état stationnaire, que *LDF* a la plus forte probabilité de perte à l'état stationnaire entre toutes les disciplines de service non-préemptives et donnons un majorant du gain d'*EDF* en terme de probabilité de perte par rapport à toute autre discipline de service.

### 4.2.1 Cas général $G/M/1/1 + G$

Soit une file d'attente avec impatiences  $G/M/1/1 + G$ . Soit  $\tilde{\sigma}_n$ , la durée du  $n$ -ième service rendu par le serveur. Nous supposons comme dans le chapitre précédent que c'est le serveur et non l'identité du client qui entre en service à  $\tilde{T}_{n-1}$ , qui décide de la valeur de  $\tilde{\sigma}_n$ . Cette hypothèse ne change pas la statistique du système.

On note pour toute discipline de service  $(\cdot)$  :

- pour tout  $t$ ,  $P_t^{(\cdot)}$  le nombre de clients perdus et  $S_t^{(\cdot)}$  servis par  $(\cdot)$  à l'instant  $t$ ,
- à l'état stationnaire :
  - $P^{(\cdot)}$ , le nombre de clients perdus pendant une *busy period* de  $(\cdot)$ ,
  - $E^{(\cdot)}$ , le nombre de clients entrés pendant une *busy period* de  $(\cdot)$ ,
  - $S^{(\cdot)}$ , le nombre de clients servis pendant une *busy period* de  $(\cdot)$ ,
  - $B^{(\cdot)}$ , la durée d'une *busy priod* pour  $(\cdot)$ ,
  - $\pi^{(\cdot)}$ , la probabilité de perte pour  $(\cdot)$ .

Nous comparons, pour un scénario de trafic (une trajectoire de la file d'attente), les performances du système pour trois disciplines de service différentes : *EDF*, *LDF* et une discipline de service quelconque  $\gamma$ .

On se place donc sur une trajectoire  $(t_n, n \in \mathbb{N})$  de la suite des temps d'arrivées  $(T_n, n \in \mathbb{N})$ ,  $(s_n, n \in \mathbb{N})$  de la suite des temps de service  $(\tilde{\sigma}_n, n \in \mathbb{N})$  et  $(d_n, n \in \mathbb{N})$  une trajectoire de la

suite des délais initiaux  $(D_n, n \in \mathbb{N})$ . On notera, pour cette trajectoire et pour toute discipline de service  $(.)$  :

- pour tout  $t$ ,  $p_t^{(.)}$  le nombre de clients perdus à l'instant  $t$  par  $(.)$ ,
- pour tout  $n$ ,  $s_n^{(.)}$  le nombre de service pendant la  $n$ -ième *busy period* de  $(.)$ .

Soit  $t_0$ , l'instant d'arrivée du premier client dans la file (point de construction). On note également :

- pour tout  $n = 1 \dots \max\{s_1^\gamma, s_1^{EDF}, s_1^{LDF}\}$ ,  $\tilde{t}_n$ , la date de fin du  $n$ -ième service,  $V_n^{(.)}$ , la population présente dans la file  $(.)$  à  $\tilde{t}_n$ ,  $x_n^{(.)} = (V_n^{(.)})$ , le nombre de clients présents dans le système  $(.)$  à  $\tilde{t}_n$  et  $l_n^{(.)} = p_{\tilde{t}_n}^{(.)}$ , le nombre de clients perdus par  $(.)$  à  $\tilde{t}_n$ ,
- pour tout  $n = 1 \dots \min\{s_1^\gamma, s_1^{EDF}\}$ , pour  $(.) = \gamma$  et  $EDF$ , On sépare à  $\tilde{t}_n$  les populations  $V_n^{(.)}$  de la façon suivante :

$$V_n^{EDF} = M_n^{\gamma, EDF} + R_n^{EDF/\gamma}$$

et

$$V_n^\gamma = M_n^{\gamma, EDF} + R_n^{\gamma/EDF},$$

où  $M_n^{\gamma, EDF}$  est la population présente à la fois dans le système  $EDF$  et dans le système  $\gamma$  à  $\tilde{t}_n$ ,  $R_n^{EDF/\gamma}$  est la population présente dans le système  $EDF$  à  $\tilde{t}_n$ , qui est absente à cet instant dans le système  $\gamma$  et  $R_n^{\gamma/EDF}$  est la population présente dans le système  $\gamma$  à  $\tilde{t}_n$ , qui est absente à cet instant dans le système  $EDF$ ,

- pour tout  $n = 1 \dots \min\{s_1^{LDF}, s_1^\gamma\}$ , pour  $(.) = \gamma$  et  $LDF$ ,

$$V_n^{\gamma/LDF} = M_n^{\gamma, LDF} + R_n^{\gamma/LDF}$$

et

$$V_n^{LDF/\gamma} = M_n^{\gamma, LDF} + R_n^{LDF/\gamma},$$

où  $M_n^{\gamma, LDF}$  est la population présente à la fois dans le système  $LDF$  et dans le système  $\gamma$  à  $\tilde{t}_n$ ,  $R_n^{\gamma/LDF}$  est la population présente à  $\tilde{t}_n$  dans le système  $LDF$  qui n'est pas présente au même instant dans le système  $\gamma$  et vice-versa pour  $R_n^{LDF/\gamma}$ .

**Lemme 4.2.1**

Pour tout  $n$  jusqu'à  $\min\{s_1^\gamma, s_1^{EDF}\}$ , si  $R_n^{\gamma/EDF} \neq \emptyset$ , il existe une application injective :

$$\Phi_n : \begin{cases} R_n^{\gamma/EDF} \longrightarrow R_n^{EDF/\gamma} \\ C_j \longmapsto C_k \text{ tel que } T_j + D_j < T_k + D_k. \end{cases}$$

**Preuve.** Nous raisonnons par récurrence. A  $\tilde{t}_1$ ,

$$R_1^{\gamma/EDF} = R_1^{EDF/\gamma} = \emptyset.$$

et la condition est vérifiée. On suppose que la propriété est vraie à l'instant  $\tilde{t}_n$  et on appelle  $C_l$ , le client servi par  $\gamma$  à cet instant.

1. Si  $R_n^{\gamma/EDF} = \emptyset$ , tous les clients de  $V_n^\gamma$  sont dans  $V_n^{EDF}$ .
  - si  $EDF$  sert aussi  $C_l$ ,  $V_{n+1}^\gamma \subset V_{n+1}^{EDF}$  et  $R_{n+1}^{\gamma/EDF} = \emptyset$ ,
  - si  $EDF$  sert  $C_j \in R_n^{EDF}$ , il sert un client qui de toutes façons n'est pas dans le système  $\gamma$  à  $\tilde{t}_n$  et on a bien  $R_{n+1}^{\gamma/EDF} = \emptyset$ ,
  - si  $EDF$  sert un client  $C_j \in M_n^{\gamma,EDF}$ ,
    - si  $C_j$  est le client  $C_l$ , le même client est servi à  $\tilde{t}_n$  dans les deux systèmes et donc  $R_{n+1}^{\gamma/EDF} = \emptyset$ ,
    - si  $C_j \neq C_l$  et  $T_l + D_l < \tilde{t}_{n+1}$ ,  $T_j + D_j < T_l + D_l < \tilde{t}_{n+1}$ ,  $C_j$  est donc éliminé du système régi par  $\gamma$  à  $\tilde{t}_{n+1}$ ,  $C_j$  n'est pas dans  $R_{n+1}^{\gamma/EDF}$  et on a bien  $R_{n+1}^{\gamma/EDF} = \emptyset$ ,
    - si  $C_j \neq C_l$  et  $T_l + D_l \geq \tilde{t}_{n+1}$ ,
      - \* si  $T_j + D_j < \tilde{t}_{n+1}$ ,  $C_j$  est donc éliminé du système  $\gamma$  à  $\tilde{t}_{n+1}$  et  $R_{n+1}^{\gamma/EDF} = \emptyset$ ,
      - \* si  $T_j + D_j > \tilde{t}_{n+1}$ ,  $C_j$  est donc toujours dans le système  $\gamma$  à  $\tilde{t}_{n+1}$  et  $R_{n+1}^{\gamma/EDF} = \{C_j\}$ . Or,  $C_l \in V_n^\gamma \subset V_n^{EDF}$  et  $T_l + D_l > T_j + D_j > \tilde{t}_{n+1}$ , donc  $C_l \in R_{n+1}^{EDF/\gamma}$  et on peut noter  $\Phi_n(C_j) = C_l$ .

Donc, dans tous les cas l'hypothèse de récurrence est vérifiée si  $R_n^{\gamma/EDF} = \emptyset$ .

2. On suppose maintenant que  $R_n^{\gamma/EDF} \neq \emptyset$  et que l'application injective  $\Phi_n$  existe. On suppose de plus que  $R_{n+1}^{\gamma/EDF} \neq \emptyset$ . Soit  $C_j \in R_{n+1}^{\gamma/EDF}$ . Alors,  $C_j \in V_n^\gamma$  à  $\tilde{t}_n$  et  $C_j \neq C_l$  car  $\gamma$  sert  $C_l$  à  $\tilde{t}_n$ .

- si  $C_j \in R_n^{\gamma/EDF}$ , soit  $C_k = \Phi_n(C_j) \in R_n^{EDF/\gamma}$ . Alors,  $T_k + D_k > T_j + D_j > \tilde{t}_{n+1}$ ,
  - si  $C_k$  n'est pas le client servi par  $EDF$  à  $\tilde{t}_n$  (et donc  $T_k + D_k \neq \min\{T_i + D_i, C_i \in V_n^{EDF}\}$ ), alors,  $C_k \in R_{n+1}^{EDF/\gamma}$  car il n'est pas éliminé à  $\tilde{t}_{n+1}$  et  $\Phi_{n+1}(C_j) = C_k$ ,
  - si  $EDF$  sert  $C_k$  à  $\tilde{t}_n$ ,
    - \* si  $C_l \in M_n^{\gamma,EDF}$ ,  $T_l + D_l > T_k + D_k > \tilde{t}_{n+1}$ ,  $C_l$  n'est pas éliminé par  $EDF$  à  $\tilde{t}_{n+1}$  et donc  $C_l \in R_{n+1}^{EDF/\gamma}$ .  
Alors on note  $\Phi_{n+1}(C_j) = C_l$ .
    - \* si  $C_l$  n'est pas dans  $M_n^{\gamma,EDF}$ ,  $C_l \in R_n^{\gamma/EDF}$ ,  $C_m = \Phi_n(C_l) \in R_n^{EDF/\gamma}$  et  $C_m \neq C_k$  car  $C_l \neq C_j$ . Comme  $C_k$  est le client servi par  $EDF$  à  $\tilde{t}_n$ ,  $T_m + D_m > T_k + D_k > \tilde{t}_{n+1}$  et donc  $C_m \in R_{n+1}^{EDF/\gamma}$ .  
Alors on note  $\Phi_{n+1}(C_j) = C_m$ ,
- si  $C_j \in M_n^{\gamma,EDF}$ ,  $C_j$  est le client servi par  $EDF$  à  $\tilde{t}_n$ , et donc  $T_j + D_j = \min\{T_i + D_i, C_i \in V_n^{EDF}\}$ ,
  - si  $C_l \in M_n^{\gamma,EDF}$ ,  $T_l + D_l > T_j + D_j > \tilde{t}_{n+1}$ ,  $C_l$  est toujours dans le système  $EDF$  à  $\tilde{t}_{n+1}$ , plus précisément  $C_l \in R_{n+1}^{EDF/\gamma}$  et on note  $\Phi_{n+1}(C_j) = C_l$ ,
  - si  $C_l \in R_n^{\gamma/EDF}$ , on note à nouveau  $C_m = \Phi_n(C_l) \in R_n^{EDF/\gamma} \subseteq V_n^{EDF}$  et  $T_m + D_m > T_j + D_j > \tilde{t}_{n+1}$ . Donc,  $C_m \in R_{n+1}^{EDF/\gamma}$  car à  $\tilde{t}_{n+1}$  il n'est ni servi ni éliminé par  $EDF$ .  
On note alors  $\Phi_{n+1}(C_j) = C_m$ .

L'hypothèse de récurrence est donc vérifiée si  $R_n^\gamma \neq \emptyset$ .

On peut de même donner le résultat suivant en comparant maintenant, pour le même scénario de trafic, le système géré par la discipline de service quelconque  $\gamma$  avec celui géré par  $LDF$  :

#### Lemme 4.2.2

Pour tout  $n$  jusqu'à  $\min\{s_1^{LDF}, s_1^\gamma\}$ , si  $R_n^{LDF/\gamma} \neq \emptyset$ , il existe une application injective :

$$\Psi_n : \begin{cases} R_n^{LDF/\gamma} \longrightarrow R_n^{\gamma/LDF} \\ C_j \longmapsto C_k \text{ tel que } T_j + D_j < T_k + D_k. \end{cases}$$

**Preuve.** Nous raisonnons de la même façon que dans le lemme précédent, par récurrence. A  $\tilde{t}_1$ ,

$$R_1^{LDF/\gamma} = R_1^{\gamma/LDF} = \emptyset.$$

On suppose que la propriété est vraie à l'instant  $\tilde{t}_n$  et on appelle  $C_l$ , le client servi par  $LDF$  à cet instant.

1. Si  $R_n^{LDF/\gamma} = \emptyset$ ,  $V_n^{LDF} \subseteq V_n^\gamma$ .
  - si  $\gamma$  sert aussi  $C_l$ ,  $V_{n+1}^{LDF} \subseteq V_{n+1}^\gamma$  et  $R_{n+1}^{LDF/\gamma} = \emptyset$ ,
  - si  $\gamma$  sert  $C_j \in R_n^{\gamma/LDF}$ , il sert un client qui n'est pas dans le système  $LDF$  à  $\tilde{t}_n$  et on a bien  $R_{n+1}^{LDF/\gamma} = \emptyset$ ,
  - si  $\gamma$  sert un client  $C_j \in M_n^{\gamma,LDF}$ ,
    - si  $C_j = C_l$ ,  $R_{n+1}^{LDF/\gamma} = \emptyset$ ,
    - si  $C_j \neq C_l$  et  $T_l + D_l < \tilde{t}_{n+1}$ , alors  $T_j + D_j < T_l + D_l < \tilde{t}_{n+1}$ ,  $C_j$  est donc éliminé du système  $LDF$  à  $\tilde{t}_{n+1}$  et on a bien  $R_{n+1}^{LDF/\gamma} = \emptyset$ ,
    - si  $C_j \neq C_l$  et  $T_l + D_l \geq \tilde{t}_{n+1}$ ,
      - \* si  $T_j + D_j < \tilde{t}_{n+1}$ , et donc  $C_j$  est éliminé par  $LDF$  à  $\tilde{t}_{n+1}$  et  $R_{n+1}^{LDF/\gamma} = \emptyset$ ,
      - \* si  $T_j + D_j > \tilde{t}_{n+1}$ ,  $C_j$  n'est pas éliminé du système  $LDF$  à  $\tilde{t}_{n+1}$  et  $R_{n+1}^{LDF/\gamma} = \{C_j\}$ .  
Or,  $C_l \in V_n^{LDF} \subset V_n^\gamma$  et  $T_l + D_l > T_j + D_j > \tilde{t}_{n+1}$ , donc  $C_l \in R_{n+1}^{EDF/\gamma}$  et on peut noter  $\Psi_n(C_j) = C_l$ .
2. On suppose que  $R_n^{LDF/\gamma} \neq \emptyset$ , que l'application injective  $\Psi_n$  existe et que  $R_{n+1}^{LDF/\gamma} \neq \emptyset$ . Soit  $C_j \in R_{n+1}^{LDF/\gamma}$ .  $C_j \in V_n^{LDF}$ .
  - si  $C_j \in R_n^{LDF/\gamma}$ ,  $C_k = \Psi_n(C_j) \in R_n^{EDF}$  et  $T_k + D_k > T_j + D_j > \tilde{t}_{n+1}$ ,
    - si  $C_k$  n'est pas le client servi par  $\gamma$  à  $\tilde{t}_n$ ,  $C_k \in R_{n+1}^{\gamma/LDF}$  et  $\Psi_{n+1}(C_j) = C_k$ ,
    - si  $\gamma$  sert  $C_k$  à  $\tilde{t}_n$ ,
      - \* si  $C_l \in M_n^{\gamma,LDF}$ , comme  $LDF$  sert  $C_l$  à  $\tilde{t}_n$ ,  $T_l + D_l > T_k + D_k > \tilde{t}_{n+1}$  et donc  $C_l \in R_{n+1}^{\gamma/LDF}$ . On note  $\Psi_{n+1}(C_j) = C_l$ ,
      - \* si  $C_l \in R_n^{LDF/\gamma}$ , on note  $C_m = \Psi_n(C_l) \in R_n^{\gamma/LDF}$  et  $T_m + D_m > T_l + D_l$ . Or,  $T_l + D_l > T_j + D_j > \tilde{t}_{n+1}$  car  $C_j$  et  $C_l$  sont dans  $V_n^{LDF}$  et  $LDF$  sert  $C_l$ . Donc,  $C_m$  n'est pas éliminé à  $\tilde{t}_{n+1}$  et n'a pas été servi par  $\gamma$  :  $C_m \in R_{n+1}^{\gamma/LDF}$  et on note  $\Psi_{n+1}(C_j) = C_m$ .
  - Si  $C_j \in M_n^{\gamma,LDF}$ ,  $\gamma$  sert  $C_j$  à  $\tilde{t}_n$ ,

- si  $C_l \in M_n^{\gamma, LDF}$ ,  $T_l + D_l > T_j + D_j > \tilde{t}_{n+1}$ ,  $C_l \in R_{n+1}^{\gamma/LDF}$  et on note  $\Psi_{n+1}(C_j) = C_l$ ,
- si  $C_l \in R_n^{LDF/\gamma}$ , on note à nouveau  $C_m = \Psi_n(C_l)$ .  $C_j$  et  $C_l$  sont dans  $V_n^{EDF}$  et  $LDFsertC_l$ , donc  $T_l + D_l > T_j + D_j > \tilde{t}_{n+1}$  et comme  $T_m + D_m > T_l + D_l$ ,  $C_m$  n'est pas servi par  $\gamma$  à  $\tilde{t}_n$  et n'est pas non-plus éliminé du système  $\gamma$  à  $\tilde{t}_{n+1}$ . Donc  $C_m \in R_{n+1}^{\gamma/LDF}$  et on note  $\Psi_{n+1}(C_j) = C_m$ .

□

On en déduit :

### Corollaire 4.2.1

Partant d'un point de construction commun aux trois systèmes et répondant au même trafic,

1. le système régi par  $\gamma$  se vide avant le système régi par EDF et après celui géré par LDF :

$$s_1^{LDF} \leq s_1^\gamma \leq s_1^{EDF},$$

2. pour tout  $n$  jusqu'à  $s_1^\gamma$ ,

$$(R_n^{EDF/\gamma}) \leq n - 1, \quad (4.6)$$

$$p_{\tilde{t}_{s_1^\gamma}}^{EDF} \leq p_{\tilde{t}_{s_1^\gamma}}^\gamma \leq p_{\tilde{t}_{s_1^\gamma}}^{EDF} + s_1^\gamma - 1, \quad (4.7)$$

3. pour tout  $n$  jusqu'à  $s_1^{LDF}$ ,

$$(R_n^{\gamma/LDF}) \leq n - 1, \quad (4.8)$$

$$p_{\tilde{t}_{s_1^{LDF}}}^\gamma \leq p_{\tilde{t}_{s_1^{LDF}}}^{LDF} \leq p_{\tilde{t}_{s_1^{LDF}}}^\gamma + s_1^{LDF} - 1. \quad (4.9)$$

### Preuve

1. Le lemme (4.2.1) implique en particulier que

$$(R_n^{\gamma/EDF}) \leq (R_n^{EDF/\gamma}) \text{ pour tout } n \leq \min\{S_1^\gamma, S_1^{EDF}\} \quad (4.10)$$

et donc :

$$(V_n^\gamma) \leq (V_n^{EDF}) \text{ pour tout } n \leq \min\{S_1^\gamma, S_1^{EDF}\}.$$

Donc, le système EDF ne peut se vider avant que le système  $\gamma$  ne se vide et autrement dit  $s_1^\gamma < s_1^{EDF}$ .

Par ailleurs, le lemme (4.2.2) implique de même que

$$(R_n^{LDF/\gamma}) \leq (R_n^{\gamma/LDF}) \text{ pour tout } n \leq \min\{S_1^{LDF}, S_1^\gamma\}$$

et donc :

$$(V_n^{LDF}) \leq (V_n^\gamma) \text{ pour tout } n \leq \min\{S_1^{LDF}, S_1^\gamma\}.$$

Le système LDF se vide donc avant le système  $\gamma$  :  $s_1^{LDF} < s_1^\gamma$ .

2. Par récurrence : à  $\tilde{t}_1$ ,  $(R_1^{EDF}) = 0$ . On suppose la propriété vraie à  $\tilde{t}_n$ . On note encore  $C_l$ , le client servi par  $\gamma$  à cet instant. Alors,

$$R_{n+1}^{EDF/\gamma} \subseteq \begin{cases} \left\{ \begin{array}{l} \{C_l\} \cup \\ \left\{ R_n^{EDF/\gamma} \cap \left\{ T_k + D_k \geq \left( \tilde{t}_{n+1} \vee \min_{C_i \in V_n^{EDF}} \{T_i + D_i\} \right) \right\} \right\} \\ \text{si } C_l \in M_n^{\gamma, EDF}, \end{array} \right\} \\ \left\{ \begin{array}{l} \left\{ R_n^{EDF/\gamma} \cap \left\{ T_k + D_k \geq \left( \tilde{t}_{n+1} \vee \min_{C_i \in V_n^{EDF}} \{T_i + D_i\} \right) \right\} \right\} \\ \text{si } C_l \in R_n^{\gamma/EDF}. \end{array} \right\} \end{cases}$$

Donc, pour tout  $n$  jusqu'à  $S_1^\gamma$ ,

$$(R_{n+1}^{EDF/\gamma}) \leq (R_n^{EDF/\gamma}) + 1$$

et on a (4.6). Par ailleurs :

$$p_{\tilde{t}_n}^{EDF} = N(0, \tilde{t}_n] - n - x_n^{EDF} = p_{\tilde{t}_n}^\gamma + (R_n^{\gamma/EDF}) - (R_n^{EDF/\gamma})$$

et avec (4.10) et (4.6)

$$p_{\tilde{t}_n}^{EDF} \leq p_{\tilde{t}_n}^\gamma \leq p_{\tilde{t}_n}^{EDF} + n - 1 - (R_n^{\gamma/EDF}). \quad (4.11)$$

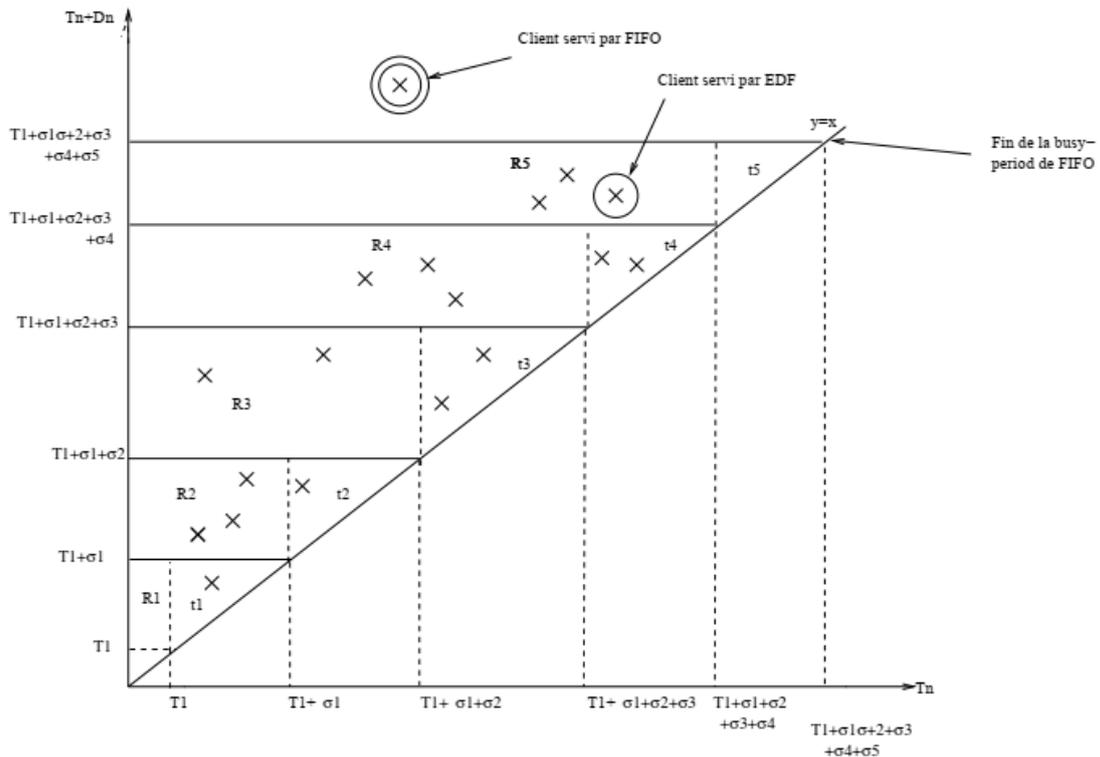
On a donc (4.7) pour  $n = s_1^\gamma$ .

3. On applique le même argument à la comparaison  $\gamma/LDF$  : pour tout  $n \leq s_1^{LDF}$ , en notant  $C_j$  le client servi par LDF et  $C_l$  à  $\tilde{t}_n$  :

$$R_{n+1}^{\gamma/LDF} \subseteq \begin{cases} \left\{ \begin{array}{l} \{C_j\} \cup \left\{ R_n^{\gamma/LDF} \cap \{T_k + D_k \geq \tilde{t}_{n+1}\} \cap \{C_l\}^c \right\} \\ \text{si } C_j \in M_n^{\gamma, LDF}, \end{array} \right\} \\ \left\{ \begin{array}{l} \left\{ R_n^{\gamma/LDF} \cap \{T_k + D_k \geq \tilde{t}_{n+1}\} \cap \{C_l\}^c \right\} \\ \text{si } C_j \in R_n^{LDF/\gamma}. \end{array} \right\} \end{cases}$$

et on conclut de la même façon.

La figure (4.2.1) illustre ces résultats, en représentant une busy period pour EDF et  $\gamma = FIFO$ . Les busy periods pour les deux systèmes, soumis aux mêmes demandes, aux mêmes instants d'arrivées, commencent au même instant. En abscisses, on note les instants d'arrivées ( $T_n, n \in \mathbb{N}^*$ ) des clients



**Figure 4.1** - Busy period pour EDF et FIFO

et en ordonnées leurs dates d'élimination ( $T_n + D_n, n \in \mathbb{N}^*$ ). Chaque croix représente donc un client entré dans le système. Les barres horizontales représentent les instants de fins de service pour les deux systèmes, pour ce scénario de trafic. Un client est encore dans le système, et susceptible d'être servi à l'instant  $\tilde{t}_n = \tilde{t}_1 + \sum_{i=1}^n \tilde{\sigma}_i$  si la croix correspondante est au dessus de la  $n$ -ième barre horizontale, et si il n'a pas encore été servi sous la discipline de service considérée. Par exemple, le serveur peut servir à  $\tilde{t}_n$  toutes les croix se trouvant dans le rectangle  $R_{n+1}$  (clients qui si ils ne sont pas servis à  $\tilde{t}_n$  seront éliminés à  $\tilde{t}_{n+1}$ ), ou au dessus.

Les croix se trouvant dans les triangles  $t_i$  sont des clients éliminés avant même la fin du service pendant lequel ils sont entrés (clients " mort-nés "). EDF sert toujours le client correspondant à la croix la plus basse parmi celles se trouvant au dessus de la barre horizontale correspondant à l'instant de fin de service. FIFO sert au même instant, le client dont la croix se trouve la plus à gauche (LDF servirait, lui, toujours la croix la plus haute). A  $\tilde{t}_4$ , EDF sert le client entouré une fois et FIFO, le client entouré deux fois. La busy period sous FIFO se termine donc à  $\tilde{t}_5$ , alors que celle d'EDF se poursuit : le client entouré deux fois n'a pas encore été servi à  $\tilde{t}_5$  par EDF et n'a pas atteint son délai à cet instant : il est servi.

On peut alors retrouver et compléter le résultat de Panwar et *al.* :

### Corollaire 4.2.2

Pour toute file  $G/M/1/1 + G$ , pour tout  $t$  et toute discipline de service  $\gamma$ ,

$$P_t^{EDF} \leq_{st} P_t^\gamma \leq_{st} P_t^{LDF}.$$

**preuve** On part de  $\tilde{t}_0^0 = \tilde{t}_0$ . A cet instant un client entre dans les trois systèmes vides et pour les trois disciplines de service :

$$s_{\tilde{t}_0^0}^{(\cdot)} = p_{\tilde{t}_0^0}^{(\cdot)} = 0$$

(aucun service n'a été complété et aucun client n'a été éliminé).

Commençons par considérer les disciplines  $\gamma$  et EDF. D'après le corollaire (4.2.1),  $\gamma$  termine sa première busy period avant EDF, et à tout instant  $t \leq s_1^\gamma$ , d'après (4.11) :

$$p_t^{EDF} \leq p_t^\gamma.$$

A la fin de la busy period de  $\gamma$  (à  $s_1^\gamma$ ), si le système EDF n'est pas vide, EDF commence son  $(s_1^\gamma + 1)$ -ième service et

$$p_{\tilde{t}_{s_1^\gamma}}^{EDF} \leq p_{\tilde{t}_{s_1^\gamma}}^\gamma, \quad (4.12)$$

puisqu'à cet instant autant de client sont entrés et ont été servis dans les deux systèmes, mais il reste des clients dans le système EDF. Si EDF termine sa busy period avant l'arrivée d'un nouveau client dans le système  $\gamma$  vide, les deux systèmes sont vides en même temps et on peut ré-initier le même raisonnement en repartant d'un système vide pour les deux disciplines (point de construction commun) dès que ce nouveau client  $C_i$  entre dans le système.

Sinon, on définit la discipline de service modifiée  $\tilde{\gamma}$  :  $\tilde{\gamma}$  bloque " le serveur (il le rend inactif) pendant toute la durée du  $(s_1^\gamma + 1)$ -ième service d'EDF (et ainsi de suite à chaque début d'un

nouveau service d'EDF tant qu'aucun nouveau client n'est entré dans le système  $\gamma$ ). Le premier client  $C_i$  à entrer dans le système à partir de là voit sa patience " gelée " jusqu'à ce qu'il rentre en service, à la fin du service en cours pour EDF. Ainsi, le système  $\tilde{\gamma}$  ré-initie une période d'activité avec le service de  $C_i$  nécessairement à un instant de début de service du système EDF. Soit  $\tilde{t}_0^1$  cet instant. On a alors :

$$p_{\tilde{t}_0^1}^{EDF} \leq p_{\tilde{t}_0^1}^{\tilde{\gamma}} \quad (4.13)$$

puisque les clients éventuellement perdus par EDF pendant l'*idle period* de  $\tilde{\gamma}$  sont soit :

- des clients qui ont figuré dans le système  $\tilde{\gamma}$  (perdus ou servis),
- des clients entrés après  $C_i$  et éliminés avant  $\tilde{t}_0^1$  auquel cas ils sont aussi éliminés par  $\tilde{\gamma}$ ,
- $C_i$  lui-même (qui dans ce cas ne doit sa présence dans le système  $\tilde{\gamma}$  à  $\tilde{t}_0^1$  qu'au fait que son délai a été gelé) et dans ce cas  $\tilde{\gamma}$  perd un client de moins qu'EDF.

Dans tous les cas,  $\tilde{\gamma}$  perd au plus un nouveau client de moins qu'EDF pendant son *idle period* et donc (4.13) découle de (4.12).

Ensuite, à l'instant  $\tilde{t}_0^1$ , une *busy period* reprend pour  $\tilde{\gamma}$ , couplée avec les services d'EDF : on réitère le même raisonnement pour cette nouvelle *busy period* de  $\tilde{\gamma}$  et pour les clients entrés à partir de cet instant en appliquant le corollaire (4.2.1) partie 2 et l'équation (4.11).

On peut donc coupler les deux systèmes à l'infini en s'assurant que la perte du système  $\tilde{\gamma}$  soit plus forte que celle d'EDF : pour ce scénario de trafic, pour tout  $t$ ,

$$p_t^{EDF} \leq p_t^{\tilde{\gamma}}.$$

On généralise ce raisonnement à tous les scénarios de trafic : p.s., pour tout  $t$ ,

$$P_t^{EDF} \leq P_t^{\tilde{\gamma}}. \quad (4.14)$$

Maintenant, en vertu de la propriété d'oubli de la loi exponentielle, la durée du service du client  $C_i$  qui entre dans le système  $\gamma$  vide est de même loi que la durée résiduelle du service en cours pour EDF,  $\tilde{\sigma}_{n_0}$ . Donc, le système  $\gamma$  a la même statistique que le système  $\tilde{\gamma}$ . En particulier, pour tout  $t$ ,

$$P_t^{\tilde{\gamma}} \stackrel{\mathcal{L}}{=} P_t^{\gamma},$$

et on a la première majoration stochastique avec (4.14).

Ensuite, on applique le même raisonnement en comparant maintenant les systèmes  $\gamma$  et LDF. On définit la discipline  $L\tilde{D}F$  qui couple avec les *busy periods* de  $\gamma$  lorsque le système LDF se vide mais pas le système  $\gamma$ , et on applique le corollaire (4.2.1) partie 3 : on a la deuxième majoration.

**Remarque 4.2.1** *Panwar et al. montrent dans [35], d'une manière différente, la première majoration stochastique*

#### 4.2.2 Cas $M/M/1/1 + G$

Nous considérons dans cette partie que le processus des entrées est un processus de Poisson. Nous montrons dans ce cas l'optimalité d'EDF en terme de **probabilité de perte à l'état stationnaire** par rapport à toute autre discipline de service non-préemptive. Ce faisant, on peut aller plus loin que dans la section précédente et donc que Panwar et al, qui donnaient cette optimalité sur les pertes instantanées en terme de majoration stochastique pour un système  $G/M/1/1 + G$ .

##### **Théorème 4.2.1**

*Pour toute file d'attente avec clients impatients  $M/M/1/1+G$  où  $\lambda$  est l'intensité du processus des entrées et les demandes de services sont de durée moyenne  $1/\mu$ , pour toute discipline de service non préemptive  $\gamma$ , la probabilité de perte  $\pi^\gamma$  vérifie :*

$$\pi^{EDF} \leq \pi^\gamma \leq \pi^{EDF} + \frac{(\lambda \mathbb{E}[B^{EDF}] + 1)(\mu \mathbb{E}[B^{LDF}] - 1)}{(\lambda \mathbb{E}[B^{LDF}] + 1)\lambda \mathbb{E}[B^{EDF}]}.$$

**Preuve.** Comme dans la preuve du corollaire (4.2.2), on part d'un point de construction commun aux trois systèmes  $\tilde{t}_0^0 = \tilde{t}_0$ . D'après le corollaire (4.2.1),  $\gamma$  termine sa première busy period avant EDF, et à cet instant  $s_1^\gamma$ , on a (4.7). A  $s_1^\gamma$ , si le système EDF n'est pas vide, on considère comme plus haut la discipline de service modifiée  $\tilde{\gamma}$ . Si EDF termine sa busy period avant l'arrivée du client  $C_i$  qui ré-initie la busy period de  $\tilde{\gamma}$ , les deux systèmes sont vides en même temps et on note  $\nu = 1$ .

Sinon, on note à nouveau  $\tilde{t}_0^1$ , le premier instant où le serveur est libre sous EDF et les deux systèmes sont non-vides. A cet instant, on a :

$$p_{\tilde{t}_0^1}^{EDF} \leq p_{\tilde{t}_0^1}^{\tilde{\gamma}} \leq p_{\tilde{t}_0^1}^{EDF} + s_1^{\tilde{\gamma}} - 1. \quad (4.15)$$

En effet, (4.13) donne la minoration et d'autre part, tout client perdu par  $\tilde{\gamma}$  pendant sa période d'inactivité est un client " mort-né ", entré après  $C_i$  pendant le service d'EDF et éliminé avant la fin de celui-ci. Un tel client est donc aussi perdu par EDF et (4.7) donne la majoration.

A  $\tilde{t}_0^1$ , une nouvelle busy period commune à EDF et  $\tilde{\gamma}$  commence et à cet instant, il n'y a aucun client présent dans le système  $\tilde{\gamma}$  qui ne soit pas présent dans le système EDF. On réitère le même raisonnement que pour la première busy period jusqu'à l'instant où le système  $\tilde{\gamma}$  se vide

à nouveau. Si le système EDF n'est pas vide à cet instant, on appelle  $t_0^{\tilde{\gamma}}$  l'instant de reprise de la période d'activité pour  $\tilde{\gamma}$ , etc... jusqu'à l'instant (dont on sait qu'il existe puisque le système est stable) où le système EDF se vide. Cet instant est nécessairement un instant où le système  $\tilde{\gamma}$  se vide aussi : c'est la fin de la  $\nu$ -ième busy period de  $\tilde{\gamma}$ . Par récurrence, pour tout  $1 < n \leq \nu$ , en appliquant le corollaire (4.2.1), (4.7) et (4.15) à la  $n$ -ième busy period de  $\tilde{\gamma}$ , on généralise (4.15) :

$$\sum_{i=1}^n \left( p_{t_0^{\tilde{\gamma}}}^{EDF} - p_{t_0^{\tilde{\gamma}-1}}^{EDF} \right) \leq \sum_{i=1}^n \left( p_{t_0^{\tilde{\gamma}}}^{\tilde{\gamma}} - p_{t_0^{\tilde{\gamma}-1}}^{\tilde{\gamma}} \right) \leq \sum_{i=1}^n \left( \left( p_{t_0^{\tilde{\gamma}}}^{EDF} - p_{t_0^{\tilde{\gamma}-1}}^{EDF} \right) + s_i^{\tilde{\gamma}} - 1 \right).$$

D'autre part, en notant  $b^{EDF}$ , la durée totale de la busy period d'EDF, pour  $i = 1, \nu$ ,  $b_i^{\tilde{\gamma}}$ , la durée de la  $i$ -ième busy period de  $\tilde{\gamma}$  pendant celle d'EDF et pour  $i = 1, \dots, \nu - 1$  ( $\nu \geq 2$ ),  $i_i^{\tilde{\gamma}}$  la durée de la  $i$ -ième idle period de  $\tilde{\gamma}$  pendant que le serveur est actif sous EDF :

$$b^{EDF} = \sum_1^{\nu} b_i^{\tilde{\gamma}} + \sum_1^{\nu-1} i_i^{\tilde{\gamma}},$$

si  $\nu \geq 2$ .

D'après les deux équations précédentes, on a donc p.s. pour deux systèmes EDF et  $\tilde{\gamma}$  partant d'un point de construction commun et répondant à la même demande de service :

$$\sum_{i=1}^{\nu^{\tilde{\gamma}/EDF}} \left( P_{T_0^i}^{EDF} - P_{T_0^{i-1}}^{EDF} \right) \leq \sum_{i=1}^{\nu^{\tilde{\gamma}/EDF}} \left( P_{T_0^i}^{\tilde{\gamma}} - P_{T_0^{i-1}}^{\tilde{\gamma}} \right) \leq \sum_{i=1}^{\nu^{\tilde{\gamma}/EDF}} \left( P_{T_0^i}^{EDF} - P_{T_0^{i-1}}^{EDF} + S_i^{\tilde{\gamma}} \right) - \nu^{\tilde{\gamma}/EDF}, \quad (4.16)$$

$$B^{EDF} = \sum_1^{\nu^{\tilde{\gamma}/EDF}} B_i^{\tilde{\gamma}} + \sum_1^{\nu^{\tilde{\gamma}/EDF}} I_i^{\tilde{\gamma}}, \quad (4.17)$$

où :

- $\nu^{\tilde{\gamma}/EDF}$  est le nombre de busy period que fait  $\tilde{\gamma}$  pendant une busy period d'EDF,
- pour  $i = 1, \dots, \nu^{\tilde{\gamma}/EDF}$ ,  $B_i^{\tilde{\gamma}}$  est la durée de la  $i$ -ième busy period de  $\tilde{\gamma}$  pendant celle d'EDF et  $S_i^{\tilde{\gamma}}$  est le nombre de service complétés sous  $\tilde{\gamma}$  pendant cette  $i$ -ième busy period,
- pour  $i = 1, \dots, \nu^{\tilde{\gamma}/EDF} - 1$ ,  $I_i^{\tilde{\gamma}}$  est la durée de la  $i$ -ième idle period de  $\tilde{\gamma}$  pendant la busy period d'EDF.

D'une part, soit  $\lambda$ , l'intensité du processus Poissonien des entrées. Le système étant supposé à l'état stationnaire, l'ergodicité et la propriété PASTA appliquées à ce processus impliquent que  $\pi^{(\cdot)}$  est la limite du quotient du taux de perte par  $\lambda$  :

$$\pi^{(\cdot)} = \frac{1}{\lambda} \limsup_{t \rightarrow \infty} \mathbb{E} \left[ \frac{P_t^{(\cdot)}}{t} \right]. \quad (4.18)$$

Par ailleurs, les points de construction étant des instants de renouvellement du système stable, d'après ([36], théorème (3.6.1).), (4.18) revient à :

$$\pi^\gamma = \frac{\mathbb{E}[P^\gamma]}{\mathbb{E}[\mathbb{E}^\gamma]} = \frac{\mathbb{E}[P^\gamma]}{\lambda \mathbb{E}[B^\gamma]}. \quad (4.19)$$

(4.19) implique :

$$\begin{aligned} \pi^{EDF} &= \frac{\mathbb{E}[P^{EDF}]}{\mathbb{E}[\mathbb{E}^{EDF}]} \\ &= \frac{\mathbb{E} \left[ P_{\tilde{T}_0^{\nu^{\tilde{\gamma}}/EDF}}^{EDF} - P_{\tilde{T}_0^0}^{EDF} \right]}{\lambda \mathbb{E} \left[ \tilde{T}_0^{\nu^{\tilde{\gamma}}/EDF} - \tilde{T}_0^0 \right]} \\ &= \frac{\mathbb{E} \left[ \sum_{i=1}^{\nu^{\tilde{\gamma}}/EDF} \left( P_{\tilde{T}_0^i}^{EDF} - P_{\tilde{T}_0^{i-1}}^{EDF} \right) \right]}{\lambda \mathbb{E}[B^{EDF}]}, \end{aligned}$$

$$\begin{aligned} \pi^{\tilde{\gamma}} &= \frac{\mathbb{E}[P^{\tilde{\gamma}}]}{\mathbb{E}[\mathbb{E}^{\tilde{\gamma}}]} \\ &= \frac{\mathbb{E} \left[ P_{\tilde{T}_0^1}^{\tilde{\gamma}} - P_{\tilde{T}_0^0}^{\tilde{\gamma}} \right]}{\lambda \mathbb{E} \left[ \tilde{T}_0^1 - \tilde{T}_0^0 \right]} \\ &= \frac{\mathbb{E} \left[ \nu^{\tilde{\gamma}}/EDF \right] \mathbb{E} \left[ P_{\tilde{T}_0^1}^{\tilde{\gamma}} - P_{\tilde{T}_0^0}^{\tilde{\gamma}} \right]}{\lambda \mathbb{E} \left[ \nu^{\tilde{\gamma}}/EDF \right] \mathbb{E} \left[ \tilde{T}_0^1 - \tilde{T}_0^0 \right]} \\ &= \frac{\mathbb{E} \left[ \sum_{i=1}^{\nu^{\tilde{\gamma}}/EDF} \left( P_{\tilde{T}_0^i}^{\tilde{\gamma}} - P_{\tilde{T}_0^{i-1}}^{\tilde{\gamma}} \right) \right]}{\lambda \mathbb{E} \left[ \tilde{T}_0^{\nu^{\tilde{\gamma}}/EDF} - \tilde{T}_0^0 \right]} \\ &= \frac{\mathbb{E} \left[ \sum_{i=1}^{\nu^{\tilde{\gamma}}/EDF} \left( P_{\tilde{T}_0^i}^{\tilde{\gamma}} - P_{\tilde{T}_0^{i-1}}^{\tilde{\gamma}} \right) \right]}{\lambda \mathbb{E} \left[ B^{EDF} \right]}, \end{aligned}$$

avec l'identité de Wald. Ces deux équations donnent avec (4.16) :

$$\pi^{EDF} \leq \pi^{\tilde{\gamma}} \leq \pi^{EDF} + \frac{\mathbb{E}\left[\nu^{\tilde{\gamma}/EDF}\right] \mathbb{E}\left[S^{\tilde{\gamma}}\right] - \mathbb{E}\left[\nu^{\tilde{\gamma}/EDF}\right]}{\lambda \mathbb{E}\left[B^{EDF}\right]}. \quad (4.20)$$

D'autre part, avec (4.17) et l'identité de Wald :

$$\mathbb{E}\left[B^{EDF}\right] = \mathbb{E}\left[\nu^{\tilde{\gamma}/EDF}\right] \mathbb{E}\left[B^{\tilde{\gamma}}\right] + \mathbb{E}\left[\nu^{\tilde{\gamma}/EDF}\right] \mathbb{E}\left[I\right] - \mathbb{E}\left[I\right],$$

où  $I$  est la variable aléatoire du temps à attendre avant l'arrivée d'un nouveau client dans un système vide. L'équation précédente se réécrit :

$$\mathbb{E}\left[\nu^{\tilde{\gamma}/EDF}\right] = \frac{\mathbb{E}\left[B^{EDF}\right] + \mathbb{E}\left[I\right]}{\mathbb{E}\left[B^{\tilde{\gamma}}\right] + \mathbb{E}\left[I\right]} = \frac{\mathbb{E}\left[B^{EDF}\right] + 1/\lambda}{\mathbb{E}\left[B^{\tilde{\gamma}}\right] + 1/\lambda}. \quad (4.21)$$

(4.20) et (4.21) donnent donc :

$$\begin{aligned} \pi^{EDF} &\leq \pi^{\tilde{\gamma}} \\ &\leq \pi^{EDF} + \frac{\mathbb{E}\left[B^{EDF}\right] + 1/\lambda}{\mathbb{E}\left[B^{\tilde{\gamma}}\right] + 1/\lambda} (\mathbb{E}\left[S^{\tilde{\gamma}}\right] - 1). \end{aligned}$$

La statistique du système sous  $\tilde{\gamma}$  étant la même que celle du système sous  $\gamma$ , la suite d'inégalités précédente équivaut à :

$$\pi^{EDF} \leq \pi^{\gamma} \leq \pi^{EDF} + \frac{\mathbb{E}\left[B^{EDF}\right] + 1/\lambda}{\mathbb{E}\left[B^{\gamma}\right] + 1/\lambda} (\mathbb{E}\left[S^{\gamma}\right] - 1). \quad (4.22)$$

(4.22) est vraie pour toute discipline de service non-préemptive  $\gamma$ . Elle est vraie pour  $\gamma = LDF$  et donc en particulier :

$$\pi^{EDF} \leq \pi^{LDF} \leq \pi^{EDF} + \frac{\mathbb{E}\left[B^{EDF}\right] + 1/\lambda}{\mathbb{E}\left[B^{LDF}\right] + 1/\lambda} (\mathbb{E}\left[S^{LDF}\right] - 1). \quad (4.23)$$

Maintenant, en refaisant le même raisonnement pour la comparaison des probabilités de perte entre LDF et la discipline quelconque  $\gamma$  à partir de l'équation (4.9), on aboutit à l'équation analogue à (4.22) en remplaçant  $\gamma$  par LDF et EDF par  $\gamma$  :

$$\pi^{\gamma} \leq \pi^{LDF} \leq \pi^{\gamma} + \frac{\mathbb{E}\left[B^{\gamma}\right] + 1/\lambda}{\mathbb{E}\left[B^{LDF}\right] + 1/\lambda} (\mathbb{E}\left[S^{LDF}\right] - 1). \quad (4.24)$$

En combinant la première inégalité de (4.24) et celle de (4.22) avec (4.23), on obtient :

$$\begin{aligned}\pi^{EDF} &\leq \pi^\gamma \\ &\leq \pi^{EDF} + \frac{\frac{\mathbb{E}[B^{EDF}] + 1/\lambda}{\mathbb{E}[B^{LDF}] + 1/\lambda} (\mathbb{E}[S^{LDF}] - 1)}{\lambda \mathbb{E}[B^{EDF}]}\end{aligned}$$

Finalement, l'espérance de la durée d'une busy period vérifie

$$\begin{aligned}\mathbb{E}[B^{LDF}] &= \mathbb{E}[\sigma] \mathbb{E}[S^{LDF}] \\ &= \frac{1}{\mu} \mathbb{E}[S^{LDF}]\end{aligned}$$

et l'équation précédente donne le résultat.

# Bibliographie

- [1] D. Y. Barrer. Queuing with impatient customers and ordered service. *Operations Res.*, **5** : 650-656, 1957.
- [2] F. Baccelli and P. Brémaud. *Eléments of Queueing Theory : Palm- Martingale Calculus and Stochastic Recurrences*. Springer-Verlag, August, 1993.
- [3] F. Baccelli and P. Brémaud. *Eléments of queueing theory*. Springer, 1994.
- [4] A. Brandt and M. Brandt. On the  $m(n)/m(n)/s$  queues with impatient calls. *Performance évaluation*, **35** : 1-18, 1999.
- [5] F. Baccelli, P. Boyer, and G. Hébuterne. Single-server queues with impatient customers. *Adv. in Appl. Probab.*, **16(4)** : 887-905, 1984.
- [6] R.O. Baldwin, N.J. Davis, J. E. Kobza, and S.F. Midkiff. Real- time queueing theory : A tutorial présentation with an admission control application. *Queueing Systems and their appl.*, **35** : 1-21, 2000.
- [7] F. Baccelli and G. Hébuterne. On queues with impatient customers. In *Performance'81*, pages 159-179, 1981.
- [8] K. Chen and L. Decreusefond. An approximate analysis of waiting time in multi-class  $m/g/1/edf$  queues. *Performance Evaluation*, **24** : 190-199, 1996.
- [9] B. D. Choi, B. Kim, and J. Chung.  $M/M/1$  queue with impatient customers of higher priority. *Queueing Syst. Theory Appl.*, **38(1)** : 49- 66, 2001.
- [10] J. W. Cohen. Single server queue with uniformly bounded Virtual waiting time. *J. Appl. Probability*, **5** : 93-122, 1968.
- [11] M. Dertouzos. Control robotics : The procédural control of physical processus. In *Proc. IFIP Congress*, 1974.
- [12] A.G. De Kok and H.C. Tijms. A queueing System with impatient customers. *J.Appl.Pro.*, **22** : 688-696, 1985.

- [13] A.G. De Kok and H.C. Tijms. A two-moments approximation for a buffer design problem requiring a small rejection probability. *Perf. Eval.*, **5** 77-84, 1985.
- [14] B. Doytchinov, J.P. Lehoczky, and S. Shreve. Real-time queues in heavy-traffic with earliest deadline first queue discipline. *Ann. of Appl. Probab.*, **11(2)** : 332-378, 2001.
- [15] L. Decreusefond and P. Moyal. Eléments for the study of stability and performances in real time queues. In *IEEE proceedings IGT'03*, pages 1126-1132, 2003.
- [16] A.K. Erlang. Solutions of some problems in the theory of probabilities of significance in automatic téléphoné exchange. *Traduction Anglaise : P.O. Elec. Eng. J.*, **10** : 189-197, 1917.
- [17] P.D. Finch. Deterministic customer impatience in the queueing System  $gi/m/1$ . *Biometrika*, **47(1,2)** : 45-52. 1960.
- [18] D. Flipo. Steady state of loss Systems. *Comptes rendus de l'Académie des Sciences*, (297,6), 1983.
- [19] A. Ghosal. Queues with finite waiting time. *Operations Res.*, **11** : 919-921, 1963.
- [20] B. V. Gnedenko and I. N. Kovalenko. *Introduction to queueing theory*. Translated from Russian by R. Kondor. Translation edited by D. Louvish. Israël Program for Scientific Translations, Jérusalem, 1968.
- [21] C. Gromoll, A. Puha, and R. Williams. The fluid limit of a processor sharing queue. *Sto. Proc. and Appl.*, 2001.
- [22] JWT. Hong, Tan X.N., and Towsley D. A performance analysis of minimum laxity and earliest deadline scheduling in a real-time System. *IEEE Trans. Computer*, **38(12)** : 1736-1744, december 1988.
- [23] O.M. Jurkevic. On the investigation of many-server queueing Systems with bounded waiting time. *Izv.Akad.Nauk. SSSR Techniceskaja Kibernetika*, **4** : 39-46, 1971.
- [24] L. Kleinrock. *Queueing Systems*. J. Wiley and sons, 1976.
- [25] L. Leboucher. Algorithmique et modélisation pour la qualité de service des systèmes répartis temps réel. *These de doctorat*, 1998.
- [26] J.P. Lehoczky. Real time queueing theory. In *Proceedings of the IEEE Real-time symposium*, pages 186-195, 1996.
- [27] J.P. Lehoczky. Real time queueing network theory. In *Proceedings of the IEEE Real-time symposium*, pages 58-67, 1997
- [28] J.P Lehoczky. Using real-time queueing theory to control lateness in real-time Systems. *Performances Evaluation Reviewis*, **25** : 158-168, 1997.

- 
- [29] V. Limic. A lifo queue in heavy traffic. *Ann. of Appl. Prob.*, **11 (2)** : 301-331, 2001.
- [30] C.L. Liu and J.W. Layland. Scheduling algorithms for multiprogramming in hard real-time environment. *Journal of the ACM*, **20(1)** : 846- 61, 1973.
- [31] R.M. Loynes. The stability of queues with non-independent interarrivals and service times. *Proceedings of the Cambridge Philosophical Society*, **58** : 497-520, 1962.
- [32] A. Movaghar. On queueing with customer impatience until the beginning of service. *Queueing Systems Theory Appl.*, 29(2-4) : 337- 350, 1998.
- [33] P. Nain and Towsley D. Comparison of hybrid minimum laxity/first- in-first-out scheduling policies for real-time multiprocessors. *IEEE Trans. Computer*, **41 (10)** : 1271-1278, october 1992.
- [34] S. Pingali and J. Kurose. On scheduling two classes of real time traffic with identical deadlines. In *Globecom '91*,1991.
- [35] S. Panwar and D. Towsley. Optimal scheduling policies for a class of queues with customer deadlines to the beginning of service. *Journal ofthe ACM.*, **35(4)** : 832-844, 1988.
- [36] K. Ross. *Multiservice loss models for broadband télécommunications models*.Springer-Verlag, 1995.
- [37] L. Takàcs. A single-server queue with limited Virtual waiting time. *J. Appl. Probability*, **11** : 612-617, 1974.
- [38] H. Thorisson. *Coupling, stationarity, and régénération*. Springer- Verlag, 2000.