

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université A/MIRA de Béjaia
Faculté des Sciences exactes
Département de Mathématique



MEMOIRE DE MASTER

En
Mathématiques
Option
Statistique et Analyse Décisionnelle

Thème :
Sur les files d'attente avec impatience

Présenté par :

Meziani Tinhinane

Devant le jury composé de :

Président	Mr M. Boualem	M.C.A	U. A. Mira, Béjéïa
Examinatrice	Mme N. Hamadouche	M.C.B	U. A. Mira, Béjéïa
Rapporteur	Mme L. Bouraine	M.C.A	U. A. Mira, Béjéïa

Béjaïa, 30 Juin 2016.

Table des matières

Listes des figures	4
Introduction générale	6
1 Systèmes de files d'attente classiques	9
1.1 Systèmes de files d'attente markoviens	9
1.1.1 Notations	9
1.1.2 Classification des systèmes d'attente	10
1.1.3 Caractéristiques d'un système de files d'attente	12
1.2 Le système $M/M/1$	12
1.2.1 Régime transitoire	12
1.2.2 Régime stationnaire	13
1.2.3 Caractéristiques du système $M/M/1$	14
1.2.4 Distribution du temps de séjour	15
1.3 Le système $M/M/1/K$	15
1.3.1 Les caractéristiques du système $M/M/1/K$	16
1.4 Le système $M/M/s$	17
1.4.1 Distribution stationnaire	18
1.4.2 Les caractéristiques du système $M/M/s$	18
1.5 Le système $M/M/\infty$	19
1.5.1 Les caractéristiques du système $M/M/\infty$	19
1.6 Le système $M/M/s/s$ (avec perte)	20
1.6.1 Les caractéristiques du système $M/M/s/s$	21
1.7 Le système $M/M/s/K$	21
1.7.1 Les caractéristiques du système	22
2 Files d'attente avec impatience	24
2.1 Files d'attente avec impatience : Une classe de clients	24
2.1.1 Introduction	24
2.2 Notations	24
2.3 Disciplines de service	25
2.4 Les mesures de performances	26
2.5 Les processus de performances	26

2.6	Le système $M/M/1+M$	29
2.6.1	Calcul des mesures de performances	30
2.7	Le système $M/M/1+D$	31
2.8	Encadrement de la probabilité de perte d'un client :	32
2.8.1	Application du résultat :	34
2.9	Le système $M/M/s+M$	34
2.9.1	Calcul des mesures de performances	36
2.10	Le système $M/M/s+D$	37
2.11	Files d'attente avec impatience : Deux classes de clients	38
2.11.1	Cas particulier	39
2.11.2	Cas avec priorité stricte et avec préemption	40
2.11.3	Mesures de performances	43
3	Etude du système $M/M/s/(s + N)$ avec clients impatientes	45
3.1	Description du modèle	45
3.2	Les mesures de performances	48
3.2.1	La probabilité de blocage	48
3.2.2	Nombre moyen de clients dans la file	49
3.2.3	La probabilité d'abandon	51
3.2.4	Nombre moyen de clients en service	52
3.2.5	Probabilité d'atteinte du service	54
3.2.6	Le rapport de blocage, d'abandon et de clients servis	55
3.2.7	Nombre de clients dans le système vu par une arrivée acceptée	56
3.2.8	Probabilité d'attente	57
3.2.9	La durée moyenne d'attente	58
	Conclusion générale	59
	Annexe	60
	Bibliographie	61

Table des figures

1.1	Shéma général d'un système de file d'attente	11
1.2	Graphe de file d'attente M/M/1	14
1.3	Graphe de file d'attente M/M/1/K	16
1.4	Graphe de la file M/M/s	18
1.5	Graphe de la file M/M/s/s	20
1.6	Graphe de la file M/M/s/K	22
2.1	Graphe des premiers états de la file M/M/1+M	29
2.2	Graphe de transition du système M/M/s+M	35
2.3	Un serveur avec deux classes de clients qui peuvent abandonner	39
3.1	Shéma du système $M/M/s/(s+N)$ avec impatience	46
3.2	Graphe de transition du système M/M/s/(s+N) avec impatience	46
3.3	La distribution $\{P_n, 0 \leq n \leq s+N\}$, pour le nombre moyen de clients dans le système	48
3.4	La probabilité de blocage $P\{\mathcal{B}\}$	49
3.5	Le nombre moyen de clients dans la file	50
3.6	La probabilité d'abandon $\mathbb{P}\{Ab Nb\}$	52
3.7	Le nombre moyen de clients en service $\mathbb{E}[Z]$	53
3.8	La probabilité $\mathbb{P}\{Sr Nb\}$	55
3.9	La probabilité d'attente $\mathbb{P}\{W' > 0\}$	57
3.10	Le temps d'attente moyen W	58

Remerciements

A terme de mon modeste travail, je tiens à remercier avant tout, le Dieu le tout puissant de m'avoir accordé la volonté, force, santé, détermination et courage afin d'accomplir mon mémoire.

Un grand merci aux membres de ma famille pour leurs présences, leurs préoccupations et leurs soucis qu'ils se fassent pour nous, leur encouragement et leur suivi avec patience du déroulement de ce travail.

Je tient également à remercier Mme L. Bouraine ; ma promotrice pour l'honneur qu'elle m'a fait en assurant la direction du présent mémoire, je la remercie pour ses précieux conseils et orientations.

Je remercie les membres de jury qui nous ont fait l'honneur de juger mon travail.

Enfin, merci à tous ceux qui ont contribué de près ou de loin à la réalisation de ce mémoire.

Dédicaces

Je dédie ce mémoire,

À **mes chers parents** qui représentent pour moi le symbole de la bonté par excellence, qui n'ont pas cessé de m'encourager et de prier pour moi. Vos prières et vos bénédictions m'ont été d'un grand secours pour mener à bien mes études.

Je vous dédie ce travail en témoignage de mon profond amour. Puisse Dieu, le tout puissant, vous préservez et vous accordez santé, longue vie et bonheur.

À **mes chères sœurs** Bida et Yasmine en témoignage de l'attachement, de l'amour et de l'affection que je porte pour vous. Je vous dédie ce travail avec tous mes vœux de bonheur, de santé et de réussite.

À **mes grands-parents** paternelle et maternelle qui honorent notre famille. A tous les membres de ma famille, petits et grands surtout mes très chères cousines Aliouche Meriem et Assma veuillez trouver dans ce modeste travail l'expression de mon affection.

À **mes chers ami(e)s** Achoui Tinhinane qui est ma meilleure amie sans oublier les autres amis un par un, sans oublier tous mes camarades de classes, vous êtes pour moi des frères, sœurs et des amis sur qui je peux compter.

En témoignage de l'amitié qui nous uni et des souvenirs de tous les moments que nous avons passé ensemble, je vous dédie ce travail et je vous souhaite une vie pleine de santé et de bonheur.

Pour conclure, un remerciement particulier pour mon cher fiancé Zemmoura Hani ainsi que toute ma belle famille pour son aide et précieux conseils qui m'ont servis dans ce mémoire.

Introduction générale

La théorie des files d'attente est l'un des outils analytiques les plus puissants pour la modélisation de systèmes de logistiques et de communications. Cette théorie a pour objet l'étude de systèmes et réseaux où des entités, appelées clients, cherchent à accéder à des ressources, généralement limitées, afin d'en obtenir un service.

L'origine des études sur les phénomènes d'attente remonte aux années 1909 – 1920 avec les travaux d'A.K Erlang concernant les réseaux téléphoniques. La théorie mathématique s'est ensuite développée notamment grâce aux contributions de Palm, Kolmogorov, Khintchine, Pollaczek . . . et fait actuellement toujours l'objet de nombreuses publications scientifiques. Le modèle de file d'attente classique consiste en un système dans lequel les serveurs traitent un flux de requêtes en suivant des stratégies bien précises. De nombreuses applications illustrent ce modèle, en particulier dans les domaines des télécommunications, du transport routier et des réseaux informatiques.

Prenons par exemple le cas d'un opérateur téléphonique et qui veut mettre en œuvre un service d'assistance technique pour sa clientèle. Bien évidemment, cet opérateur cherche à acquérir/garder une bonne réputation en résolvant les problèmes rencontrés par ses clients. Par contre, ses clients sont dans la majorité des cas impatients car ils ne peuvent pas attendre longtemps pour être mis en relation avec un téléconseiller. Une solution naïve serait d'embaucher autant de télé conseillers que de clients mais on voit rapidement que cette solution n'est ni envisageable ni optimale.

Ainsi, la notion d'impatience demeure une contrainte prépondérante dans la vie courante et au quelle il faut trouver des solutions pratiques mais efficaces.

Pour rendre compte de la contrainte 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 impatients. 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.

Plusieurs chercheurs ont déjà étudié le cas d'une seule classe de clients mais avec différentes angles d'approche. Ainsi, Boxma, Perry et Stadje [4] ont étudié récemment la file $M/G/1 + G$ avec un temps de service et une probabilité d'abandon de distribution générale. Ils étudient deux cas :

1. Le nouvel arrivant dispose d'informations sur les clients présents dans le système, comme leur temps d'attente, leurs priorités ... etc, et peut décider d'entrer ou non dans la file d'attente.
2. Le nouvel arrivant ne possède aucune information et rejoint directement la file.

De la même façon, Economou, Gomez-Corral et Kanta [20] ont fait eux aussi cette distinction en suivant un peu près la même démarche. D'autre part, la littérature reste un peu réduite en ce qui concerne le cas de deux classes de clients car la majorité des ouvrages spécialisés traitent le problème mais sans possibilité d'abandon. Ainsi, Gross et M. Harris [21] trouvent des résultats analytiques pour certaines mesures de performance. Aussi, Iravani [22] a consacré toute une étude sur ce cas de figure en traitant la file $M/GI/1 + M$.

Dans notre travail, d'une part on a redémontré et détaillé quelques résultats pour les files d'attente avec clients impatientes dans le cas markovien à une seule classe de clients. Nous avons aussi étudié le cas d'une file d'attente avec une seule classe de clients avec un délai d'attente maximal déterministe pour chaque client, nous avons redémontré l'encadrement de la probabilité de perte qui a la particularité d'être générale à toute distribution des arrivées et du temps de service et nous donnons quelques applications de cette inégalité pour des files particulières. Cet encadrement est d'une grande utilité car il permet d'estimer à l'avance la probabilité de perte, ce qui contribuera au choix de la bonne configuration du système. Par la suite, on s'est intéressé au système markovien avec plusieurs serveurs, à capacité limitée et clients impatientes vu son importance dans la modélisation des centres téléphoniques. Une étude détaillée a été faite avec es illustrations numériques pour les mesures de performances. D'autre part, il est clair qu'en ajoutant cette possibilité d'abandon à un système de deux classes de clients, le calcul devient plus compliqué. Cependant, dans le cas particulier où on a le même taux d'abandon γ et avec priorité stricte, nous avons développé une démarche très simple qui permet de généraliser les résultats trouvés avec une seule classe à n classes de clients tout en offrant la possibilité d'une programmation récurrente qui est d'une grande utilité pour ce genre de problème.

Ce mémoire est composé d'une introduction générale, de trois chapitres et une conclusion générale.

Le chapitre 1 résume les résultats importants de la théorie de file d'attente classique.

Le deuxième chapitre traite les files d'attente avec impatience dans le cas markovien ainsi que le cas des délais déterministe une classe et deux classes de clients.

Dans le chapitre 3 on étudie le système markovien à plusieurs serveurs, capacité limitée et avec clients impatientes.

La conclusion synthétise la contribution de ce mémoire et donne quelques perspectives de recherche.

Chapitre 1

Systemes de files d'attente classiques

Introduction

L'objet de ce chapitre est la modélisation des files d'attente par des processus aléatoires. Les applications de la théorie des files d'attente sont, au-delà des files à un guichet ou des embouteillages sur la route, la gestion du trafic aérien (décollage, routage, atterrissage des avions), les réseaux informatiques (comment dimensionner un serveur pour qu'il puisse répondre à un certain nombre de requêtes), les call-centers (combien d'opérateurs pour assurer une certaine qualité de service?), etc. Les réponses que l'on cherche à apporter concernent par exemple la dimension de la file et du processus de service (combien de places dans la file d'attente, combien de serveurs inoccupés, . . . etc).

1.1 Systemes de files d'attente markoviens

1.1.1 Notations

Soient des clients qui entrent dans un système à des instants aléatoires, en demandant un service de durée aleatoire. Ces clients attendent dans une file d'attente que l'un des serveurs soit libre pour pouvoir être servis. On se place sur un espace probabilisé (Ω, \mathcal{F}, P) . On appelle C_1, C_2, \dots les clients dans l'ordre de leur arrivées, et on note pour tout n , T_n l'instant d'arrivées de C_n et Y_n , la durée du service demandé par C_n . On suppose que p.s les arrivées sont simples (un seul client 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

$$N_t = \sum_{n \in \mathbb{N}^*} \mathbf{1}_{\{T_n \leq t\}}. \quad (1.1)$$

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

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

Soit $X(t)$ "Le nombre de clients dans le système à l'instant t ".

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 $\{Y_n, n \in \mathbb{N}^*\}$, 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 Y une v.a ayant pour loi la loi des temps de services. On suppose de plus que ces v.a sont intégrables :

$$\mathbb{E}[U] < \infty \quad \text{et} \quad \mathbb{E}[Y] < \infty.$$

On note en outre :

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

et on définit la charge ρ du système par

$$\rho = \frac{\lambda}{\mu} = \frac{\mathbb{E}[Y]}{\mathbb{E}[U]}. \quad (1.2)$$

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 le nombre de serveurs, la capacité de la file d'attente, et sa discipline de service.

Les disciplines de service les plus usuelles sont [26] :

- **FCFS (First Come, First Served)** si on sert toujours le plus ancien client arrivé. En particulier, s'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 Come, First Served)** si on sert toujours le dernier client arrivé. En particulier, s'il n'ya qu'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.

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 client qui lui est prioritaire, non préemptive sinon.

1.1.2 Classification des systèmes d'attente

Un système d'attente est constitué de " clients ", qui entrent dans une file d'attente, avant d'accéder à des " serveurs " qui délivrent un service pendant un certain temps. " Client " et " serveur " sont ici des termes génériques.

Un système d'attente (dénommé "file" d'attente par abus de langage) est caractérisé par

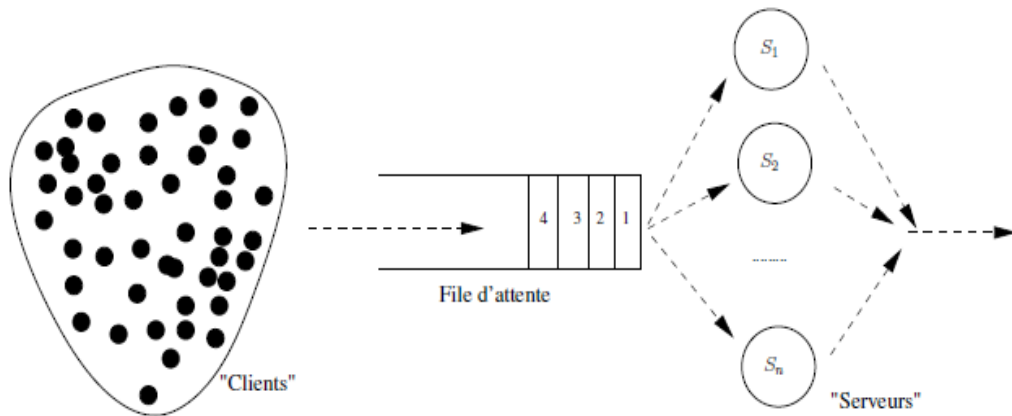


FIGURE 1.1 – Shéma général d'un système de file d'attente

- La loi d'arrivée des clients dans la file (déterministe ou stochastique).
- La loi de la durée des services (déterministe ou stochastique).
- Le nombre de serveurs travaillant en parallèle dans le centre de service.
- La taille de la file d'attente (finie ou infinie).
- Discipline de service.

Les notations de Kendall (1953) permettent de décrire le système d'attente de manière succincte. Avec ces notations, un système d'attente est décrit par : $A/B/s/K/DS$ où :

A est la distribution des arrivées : stochastique (on précise alors la loi) ou déterministe.

B est la distribution des temps de service.

s est le nombre de serveurs.

K est le nombre maximum de clients dans le système (clients attendant dans la file + clients en cours de service).

DS est la discipline de service (la manière de sortir de la file d'attente pour les clients) : First In First Out (FIFO), Last In First Out (LIFO), aléatoire (RAND), Round Robin 1 ...

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

- **M (memoryless)** si la suite des variables aléatoires sont indépendantes et identiquement distribuées de loi exponentielle. (si $A=M$, le processus des entrées est donc un processus de Poisson.)
- **G (general)** si la suite des variables aléatoires est de loi générale.
- **GI (general independent)** si les variables aléatoires sont de loi générale et indépendantes.
- **D (deterministe)** si la suite est déterministe et constante.

1.1.3 Caractéristiques d'un système de files d'attente

À partir de la distribution stationnaire du processus $\{X(t), t > 0\}$ représentant le nombre de clients dans le système à l'instant t , on peut calculer les caractéristiques d'un système d'attente telles que :

L nombre moyen de clients dans le système.

L_q nombre moyen de clients dans la file d'attente.

W temps moyen de séjour d'un client dans le système.

W_q temps d'attente moyen d'un client dans la file.

Ces valeurs sont liées par des relations suivantes appelées *formules de Little* :

$$L = \lambda_e W \quad (1.3)$$

$$L_q = \lambda_e W_q \quad (1.4)$$

où λ_e taux d'entrée dans le système.

On a aussi les relations suivantes :

$$L = L_q + \frac{\lambda_e}{\mu}. \quad (1.5)$$

$$W = W_q + \frac{1}{\mu}. \quad (1.6)$$

Remarque 1.1. Si la capacité du système $N = \infty$, alors, $\lambda_e = \lambda$.

Dans le cas contraire, certains clients doivent s'en aller sans être servis, d'où $\lambda_e < \lambda$. Dans ce chapitre on étudie les modèles de file d'attente markoviens .

1.2 Le système $M/M/1$

1.2.1 Régime transitoire

Pour le système de file d'attente $M/M/1$ (plus simple), le flot des arrivées est poissonien de paramètre λ , la durée de service est exponentielle de paramètre μ , la capacité de la file est illimitée, il y a un seul serveur. Grâce aux propriétés du processus de Poisson et de la loi exponentielle, nous avons pour un petit intervalle de temps Δ_t les probabilités suivantes :

$$\mathbb{P}(\text{exactement une arrivée pendant } \Delta_t) = \lambda \Delta_t + o(\Delta_t).$$

$$\mathbb{P}(\text{aucune arrivée pendant } \Delta_t) = 1 - \lambda \Delta_t + o(\Delta_t).$$

$$\mathbb{P}(2 \text{ arrivées ou plus pendant } \Delta_t) = o(\Delta_t).$$

$$\mathbb{P}(\text{un départ pendant } \Delta_t | X(t) \geq 1) = \mu \Delta_t + o(\Delta_t).$$

$$\mathbb{P}(\text{aucun départ pendant } \Delta_t | X(t) \geq 1) = 1 - \mu \Delta_t + o(\Delta_t).$$

$\mathbb{P}(2 \text{ départs ou plus pendant } \Delta_t) = o(\Delta_t)$.

Remarque 1.2. Les probabilités ne dépendent ni du temps t ni de l'état $X(t)$ dans lequel le système se trouve.

Soit $P_{ij}(\Delta_t) = \mathbb{P}(X(t+\Delta_t) = j | X(t) = i)$ la probabilité (conditionnelle) que le processus $X(t)$ fasse une transition de i vers j pendant la durée Δ_t .

On a :

$$P_{m+1}(\Delta_t) = \lambda\Delta_t + o(\Delta_t).$$

$$P_{nn}(\Delta_t) = 1 - (\lambda + \mu)\Delta_t + o(\Delta_t).$$

$$P_{00}(\Delta_t) = (1 - \lambda\Delta_t) + o(\Delta_t).$$

$$P_{n+1n}(\Delta_t) = (\mu\Delta_t) + o(\Delta_t).$$

$$P_{nm}(\Delta_t) = o(\Delta_t), \text{ pour } |n - m| \geq 2.$$

À partir d'un état donné, le processus markovien $\{X(t), t > 0\}$ ne peut donc passer que vers l'un des états voisins. C'est donc un processus de naissance et de mort.

Les équations du régime transitoire du système sont déterminées par le calcul de

$$P_n(t + \Delta_t) = \mathbb{P}(X(t + \Delta_t) = n) \text{ et } P_0(t + \Delta_t) = \mathbb{P}(X(t + \Delta_t) = 0).$$

Les équations de Kolmogorov seront données alors par le système suivant :

$$\begin{cases} P'_n(t) = \lambda P_{n-1}(t) + \mu P_{n+1}(t) - (\lambda + \mu)P_n(t), & \forall n \geq 1; \\ P'_0(t) = -\lambda P_0(t) + \mu P_1(t). \end{cases}$$

1.2.2 Régime stationnaire

$$P_n = \lim_{t \rightarrow \infty} P_n(t), \quad \lim_{t \rightarrow 0} P'_n(t) = 0.$$

Le régime stationnaire du système $M/M/1$ est obtenu par la résolution du système d'équations linéaires de balance suivant :

$$\lambda P_{n-1} + \mu P_{n+1} = (\lambda + \mu)P_n.$$

$$\mu P_1 = \lambda P_0$$

$$P_n = \rho^n (1 - \rho), \quad \forall n \geq 0, \text{ avec : } \rho = \frac{\lambda}{\mu} < 1. \quad (1.7)$$

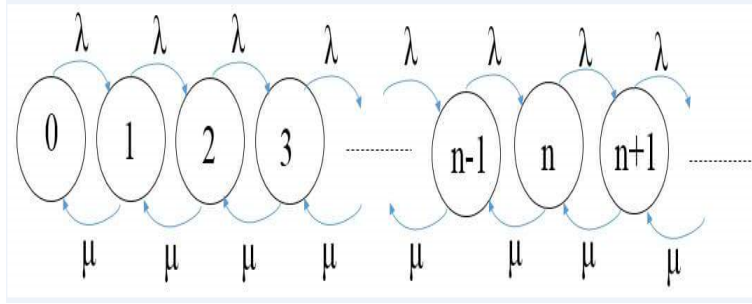


FIGURE 1.2 – Graphe de file d’attente M/M/1

Remarque 1.3. Le régime stationnaire du système M/M/1 est gouverné par la loi géométrique de paramètre $1 - \rho$.

ρ est appelé le coefficient d’utilisation du système où intensité du trafic correspondant. Il constitue une mesure de saturation du système.

$\rho = \lambda \times \frac{1}{\mu}$ (nombre moyen d’arrivées par la durée moyenne de service).

Si $\lambda \geq \mu$, on aura : $\lim_{t \rightarrow \infty} P_n(t) = P_n = 0, \forall n \geq 0$ (la longueur de la file dépasse toute limite, donc le système est instable).

1.2.3 Caractéristiques du système M/M/1

$$L = \mathbb{E}(X) = \sum_{n=0}^{\infty} nP_n = \frac{\lambda}{\mu - \lambda} = \frac{\rho}{(1 - \rho)}. \quad (1.8)$$

Soit X_q ”le nombre de clients se trouvant dans la file”.

$$X_q = \begin{cases} 0, & X = 0; \\ X - 1, & X \geq 1. \end{cases}$$

$$L_q = \mathbb{E}|X_q| = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{\rho}{1 - \rho} - \rho. \quad (1.9)$$

W et W_q peuvent être calculés, soit par les formules de Little où par la distribution stationnaire du système.

$$W = \frac{L}{\lambda} = \frac{1}{\mu - \lambda}. \quad (1.10)$$

$$W_q = \frac{L_q}{\lambda} = \frac{\lambda}{\mu(\mu - \lambda)}. \quad (1.11)$$

1.2.4 Distribution du temps de séjour

Soient T la durée de séjour d'un client dans le système.

T_q le temps d'attente d'un client dans la file (Somme des temps de service des n clients se trouvant devant lui et dont l'un est entrain d'être servi ($n = 1, 2, \dots$))

On a

$T_q = 0$, si le système est vide.

$$\mathbb{P}(T_q = 0) = \mathbb{P}(X = 0) = P_0 = 1 - \rho. \quad (1.12)$$

$$F_{T_q}(t) = \mathbb{P}(T_q \leq t) = \mathbb{P}(T_q = 0) + \mathbb{P}(0 < T_q \leq t). \quad (1.13)$$

Pour $n \geq 1$, la distribution conditionnelle de T_q . ie : $T_q|A_n \rightsquigarrow \gamma(n, \mu)$ avec A_n "Il ya n clients dans le système à l'instant d'entrée".

On obtient,

$$\mathbb{P}(0 < T_q \leq t|A_n) = \int_0^t \frac{\mu(\mu u)^{n-1}}{(n-1)!} e^{-\mu u} du. \quad (1.14)$$

$$\mathbb{P}(0 < T_q \leq t) = \rho[1 - e^{-\mu t(1-\rho)}]. \quad (1.15)$$

D'où,

$$F_{T_q}(t) = \mathbb{P}(T_q < t) = 1 - \rho e^{-\mu t(1-\rho)}, \quad t \geq 0. \quad (1.16)$$

Alors, la durée d'attente d'un client obéit dans le cas du régime stationnaire à une loi exponentielle à laquelle est superposée la probabilité $(1 - \rho)$ concentrée à l'origine.

1.3 Le système $M/M/1/K$

Ce système est identique au système $M/M/1$ exceptée que la capacité est finie. $\{X(t), t \geq 0\}$ est un processus de naissance et de mort dont le taux de transitions sont

$$\lambda_n = \begin{cases} \lambda, & n < K; \\ 0, & n = K. \end{cases}$$

$$\mu_n = \begin{cases} \mu, & n \neq 0; \\ 0, & n = 0. \end{cases}$$

D'après la distribution stationnaire d'un processus de Naissance et de Mort :

$$P_n = P_0 \prod_{i=1}^n \frac{\lambda_{i-1}}{\mu_i}$$

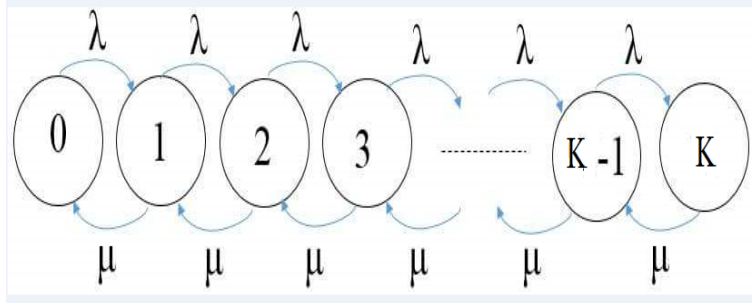


FIGURE 1.3 – Graphe de file d'attente M/M/1/K

$$P_n = \begin{cases} \left(\frac{\lambda}{\mu}\right)^n P_0, & n \leq K; \\ 0, & \text{sinon.} \end{cases}$$

On aura :

$$P_0 = \begin{cases} \frac{1-\rho}{1-\rho^{K+1}}, & \lambda \neq \mu; \\ \frac{1}{K+1}, & \lambda = \mu. \end{cases}$$

$$P_n = \begin{cases} \rho^n \frac{1-\rho}{1-\rho^{K+1}}, & \lambda \neq \mu; \\ \frac{1}{K+1}, & \lambda = \mu. \end{cases}$$

Remarque 1.4. Si $K \rightarrow +\infty$. On retrouve les résultats du système M/M/1.

1.3.1 Les caractéristiques du système M/M/1/K

Si $\lambda \neq \mu$

$$L = \mathbb{E}(X) = \frac{\rho}{1-\rho} \frac{1 - (K-1)\rho^n + K\rho^{K+1}}{1 - \rho^{K+1}}.$$

quand $n \rightarrow +\infty$

$$L = \frac{\rho}{1-\rho}. \tag{1.17}$$

$$L_q = \mathbb{E}(X-1) = \sum_{n=1}^K (n-1)P_n. \tag{1.18}$$

D'après la formule de Little suivante

$$L = \lambda_e W \Rightarrow W = \frac{L}{\lambda_e}.$$

Calcul de λ_e :

$$\begin{aligned}\lambda_e &= \lambda \mathbb{P}(\text{file est non pleine aux instants d'arrivées}) \\ &= \lambda(1 - P_K).\end{aligned}\tag{1.19}$$

avec,

$$P_K = \frac{(1 - \rho)\rho^K}{1 - \rho^{K+1}}.\tag{1.20}$$

$$\lambda_e = \lambda \left(\frac{1 - \rho^K}{1 - \rho^{K+1}} \right).$$

Remarque 1.5. λ_e est aussi appelée débit d'entrée dans le système. Noté d_e , ou aussi débit de sortie du système noté d_s donné par :

$$d_s = \frac{\rho - \rho^{K+1}}{1 - \rho^{K+1}} \mu. \text{ avec : } \rho = \frac{\lambda}{\mu}.$$

Si $\lambda = \mu$,

$$P_n = P_0 = \frac{1}{K+1}, \quad n \leq K.\tag{1.21}$$

$$L = \frac{K}{2}.\tag{1.22}$$

$$\lambda_e = \frac{K\lambda}{K+1}.$$

$$W = \frac{L}{\lambda_e} = \frac{K+1}{2\lambda}.\tag{1.23}$$

1.4 Le système $M/M/s$

Ce système comprend "s" stations de service en parallèle, les durées de services sont exponentielles (μ).

Les inter-arrivées sont de variables aléatoires indépendantes et identiquement distribuées selon $\exp(\lambda)$.

La discipline FIFO, capacité du système est illimitée.

$\{X(t), t \geq 0\}$ est un processus de naissance et de mort dont les taux de transitions sont

$$\lambda_n = \lambda, \forall n \geq 0.\tag{1.24}$$

$$\mu_n = \begin{cases} n\mu, & 1 \leq n \leq s; \\ s\mu, & n \geq s. \end{cases}$$

Remarque 1.6. $s\mu$ est le taux de service globale du système.

$\rho = \frac{\lambda}{s\mu}$ est l'intensité du trafic global.

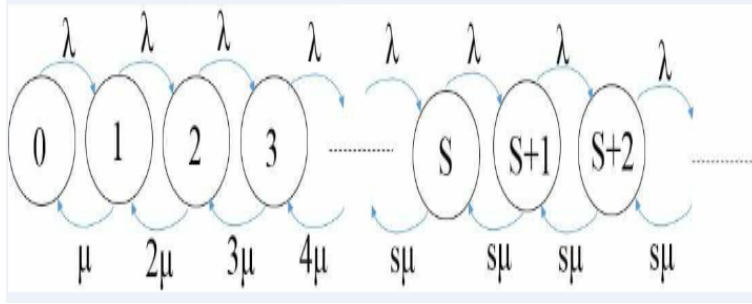


FIGURE 1.4 – Graphe de la file M/M/s

1.4.1 Distribution stationnaire

$$P_n = \begin{cases} \frac{(\frac{\lambda}{\mu})^n}{n!} P_0 & \text{si, } n \leq s; \\ \frac{(\frac{\lambda}{\mu})^n}{s^{n-s} s!} P_0 & \text{si, } n \geq s. \end{cases}$$

Pour chercher P_0 , On a : $\sum_{n=0}^{\infty} P_n = 1$.

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

$$P_0 = \left[\sum_{n=0}^s \frac{(\frac{\lambda}{\mu})^n}{n!} + \frac{(\frac{\lambda}{\mu})^{s+1}}{(s - \frac{\lambda}{\mu})s!} \right]^{-1}$$

Remarque 1.7. Pour $s = 1$, On retrouve les résultats de $M/M/1$, $P_n = \rho^n P_0$, $n \geq 0$.

1.4.2 Les caractéristiques du système M/M/s

La probabilité q'un client qui entre doit attendre :

$$\mathbb{P}(\text{attente}) = \mathbb{P}(X \geq s) = \sum_{n=s}^{\infty} \frac{P_n}{1 - \rho}, \text{ avec } \rho = \frac{\lambda}{s\mu} < 1 \text{ et } P_s = \frac{(\frac{\lambda}{\mu})^s}{s!} P_0. \quad (1.25)$$

$$L_q = \mathbb{E}(X_q) = \sum_{n=s+1}^{\infty} (n - s) P_n = \frac{\rho}{(1 - \rho)^2} P_s. \quad (1.26)$$

$$L = L_q + \frac{\lambda}{\mu}. \quad (1.27)$$

$$W_q = \frac{P_s}{s\mu(1 - \rho)^2}. \quad (1.28)$$

$$W = \frac{P_s}{s\mu(1 - \rho)^2} + \frac{1}{\mu}. \quad (1.29)$$

1.5 Le système $M/M/\infty$

C'est un système comprenant une infinité de stations de service identiques, aucune file d'attente ne se forme, chaque client est servi dès son entrée dans le système.

Ce système possède non seulement un intérêt théorique mais il permet des études approximatives de phénomène d'attente de type $M/M/s$ et $M/M/s/s$ comprenant un grand nombre de stations en parallèle.

Pour le processus de naissance et de mort de ce système

$$\begin{cases} \lambda_n = \lambda, & \forall n \geq 0; \\ \mu_n = n\mu, & \forall n \geq 1. \end{cases}$$

avec λ et μ sont les taux d'entrée et de service du système.

$$P_n = \prod_{i=1}^n \frac{\lambda_{i-1}}{\mu i} P_0 = \frac{(\frac{\lambda}{\mu})^n}{n!} P_0. \quad (1.30)$$

$$\sum_{n=0}^{\infty} P_n = 1 \Rightarrow P_0 = \frac{1}{\sum_{n=0}^{\infty} \frac{(\frac{\lambda}{\mu})^n}{n!}} = e^{-\frac{\lambda}{\mu}}, \quad \forall \lambda, \mu.$$

D'où,

$$P_n = \frac{(\frac{\lambda}{\mu})^n}{n!} e^{-\frac{\lambda}{\mu}}, \quad \forall n \geq 0.$$

Donc, le régime stationnaire du système $M/M/\infty$ est régi par la loi $\mathcal{P}(\frac{\lambda}{\mu})$.

1.5.1 Les caractéristiques du système $M/M/\infty$

$$L = \mathbb{E}(X) = \frac{\lambda}{\mu}. \quad (1.31)$$

$$L_q = W_q = 0. \quad (1.32)$$

$$W = \frac{1}{\mu}. \quad (1.33)$$

Remarque 1.8. La distribution du temps de séjour dans $M/M/\infty$ est identique à la distribution de service, donc elle suit une loi exponentielle de paramètre μ .

1.6 Le système $M/M/s/s$ (avec perte)

C'est un système qui exclu toute possibilité d'attente i.e un client qui arrive ne peut être entré dans le système si les "s" stations de services sont toutes occupées.

$\{X(t)\}_{t \geq 0}$ est un processus de naissance et de mort dont les taux de transitions sont

$$\lambda_n = \begin{cases} \lambda, & 0 \leq n \leq s; \\ 0, & \text{sinon.} \end{cases}$$

$$\mu_n = \begin{cases} n\mu, & 1 \leq n \leq s; \\ 0, & \text{sinon.} \end{cases}$$

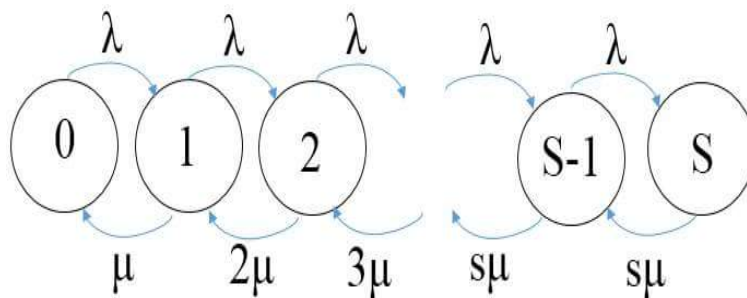


FIGURE 1.5 – Graphe de la file $M/M/s/s$

La distribution stationnaire est donnée par :

$$P_n = \prod_{i=1}^n \frac{\lambda_{i-1}}{\mu i} P_0.$$

$$P_n = \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} P_0, \quad 0 \leq n \leq s, \quad \text{avec } P_0 = \frac{1}{\sum_{n=0}^s \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!}}.$$

$$P_n = \frac{\rho^n}{n!} P_0, \quad \text{avec } \rho = \frac{\lambda}{\mu} \text{ existe sans aucune condition sur } \lambda \text{ et } \mu.$$

Remarque 1.9. P_n est appelée distribution de poisson tronquée.

Cas particulier

Le système $M/M/1/1$

$$P_0 = \frac{1}{\sum_{n=0}^1 \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!}} = \frac{1}{1 + \rho}, \quad \rho = \frac{\lambda}{\mu}. \quad (1.34)$$

$$P_1 = \frac{\rho}{1 + \rho}. \quad (1.35)$$

1.6.1 Les caractéristiques du système $M/M/s/s$

$$W_q = L_q = 0 \text{ (pas d'attente)}. \quad (1.36)$$

$$W = \frac{1}{\mu}. \quad (1.37)$$

$$L = \rho(1 - P_s) = \lambda_e W. \quad (1.38)$$

$$\mathbb{P}(\text{perte}) = \mathbb{P}(\text{ tous les serveurs sont occupés }) = P_s. \quad (1.39)$$

$$\lambda_e = \lambda(1 - P_s) = \mu(1 - P_0).$$

1.7 Le système $M/M/s/K$

Soit $(X(t))_{t \geq 0}$ est un processus de naissance et de mort dont les taux de transitions sont

$$\lambda_n = \begin{cases} \lambda, & 0 \leq n < K; \\ 0, & \text{sinon.} \end{cases}$$

$$\mu_n = \begin{cases} n\mu, & 1 \leq n \leq s; \\ s\mu, & s \leq n \leq K. \end{cases}$$

Avec $K > s$:

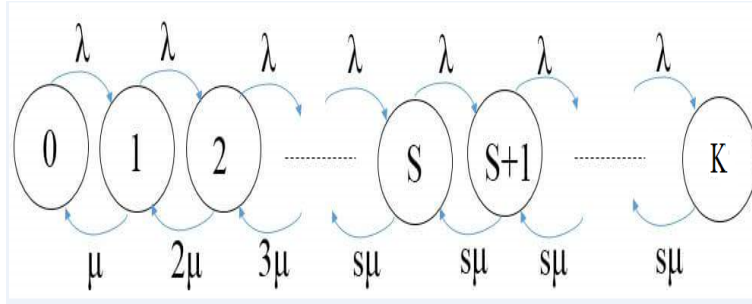


FIGURE 1.6 – Graphe de la file M/M/s/K

$$P_n = \begin{cases} \frac{(\frac{\lambda}{\mu})^n}{n!} P_0, & 1 \leq n \leq s; \\ \frac{(\frac{\lambda}{\mu})^n}{s^{n-s} s!} P_0, & s \leq n \leq K. \end{cases}$$

avec

$$P_0 = \frac{1}{\sum_{n=0}^s \frac{(\frac{\lambda}{\mu})^n}{n!} + \sum_{n=s+1}^K \frac{(\lambda\mu)^n}{s^{n-s} s!}}$$

et on pose : $\rho = \frac{\lambda}{\mu}$

1.7.1 Les caractéristiques du système

$$\mathbb{P}(\text{Perte}) = P_K = \frac{\rho^K}{s! s^{K-s}} P_0. \quad (1.40)$$

$$\begin{aligned} \mathbb{P}(\text{qu'un client doit attendre}) &= \mathbb{P}(s \leq X \leq K) \\ &= P_0 \sum_{n=s}^{K-1} \frac{(\lambda\mu)^n}{s^{n-s} s!} \\ &= \frac{P_0 \rho^s}{s!} \frac{1 - (\frac{\rho}{s})^{K-s}}{1 - \frac{\rho}{s}}. \end{aligned}$$

$$\begin{aligned}
L_q &= \mathbb{E}(X - s) = \sum_{n=s}^K (n - s)P_K \\
&= L - s - \sum_{n=0}^s (n - s)P_K.
\end{aligned}$$

$$\begin{aligned}
L &= \sum_{n=0}^K nP_n \\
&= L_q + s + \sum_{n=0}^{s-1} (n - s) \frac{\rho^n}{n!}
\end{aligned}$$

$$W = \frac{L}{\lambda_e}$$

$$W_q = \frac{L_q}{\lambda_e}$$

Avec, $\lambda_e = \lambda(1 - P_K)$.

Conclusion

Les files d'attentes classiques sont des outils mathématiques utilisés dans la vie réelle. De nombreuses applications illustrent ce modèle en particulier dans les domaines des télécommunications, du transport routier et des réseaux informatiques. Les clients de ce domaine sont dans la majorité des cas impatientes car ils ne peuvent pas attendre longtemps pour être mis en service. Pour mieux développer les systèmes de file d'attente, on doit rajouter une contrainte bien spécifique qui va représenter un délai d'attente dans le chapitre suivant qui est dit un système de files d'attente avec des clients impatientes.

Chapitre 2

Files d'attente avec impatience

2.1 Files d'attente avec impatience : Une classe de clients

2.1.1 Introduction

Pour modéliser un système de file d'attente avec impatience, on doit ajouter une contrainte au système en spécifiant que les clients sont perdus si le temps qu'ils 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 système sans avoir été satisfait. Ce chapitre, décrit les modèles d'attente avec impatience et une classe de clients où plusieurs mesures de performances, à savoir la probabilité de perte, le temps d'attente moyen et le nombre moyen de clients dans le système seront étudiés.

2.2 Notations

Considérons le $n^{\text{ième}}$ client C_n entrant dans la file d'attente est affecté du délai D_n , qui est une variable aléatoire p.s strictement positive. S'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 ce jusqu'au début de 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 que,

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

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 [2] $A/B/s + D - DS$, où :

A, B 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 services.

DS est la discipline de service.

Remarque 2.1. • La notation de Barrer indique que l'on suppose toujours que le buffer (la file) est de taille infinie.

Le paramètre aléatoire D (le délai) remplace alors dans la nomenclature de Kendall classique le paramètre N 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 $G/G/s/N$. (où un client est perdu s'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 s'il n'a pas atteint le serveur. Enfin, les délais peuvent aussi être éliminatoire, mais jusqu'à la fin du service, dans ce cas un client est éliminé s'il n'a pas terminé son service à l'instant $T_n + D_n$

2.3 Disciplines de service

Les disciplines de services mentionnées dans le chapitre 1 sont biensûr 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 à 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 2.1. [28]

- La discipline de service d'une file d'attente avec clients impatientes est dite EDF (Earliest Dead line 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.
- 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.

- Définition 2.2.**
- La discipline est dite LDF (Largest Dead line 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. Il sert alors ce client jusqu'à complétion de son service.
 - 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.

2.4 Les mesures de performances

Définissons à présent les différentes variables aléatoire caractérisant notre problème et permettant d'évaluer ses performances

- Le nombre de client moyen L dans le système (file + serveur), avec $L = \mathbb{E}[N]$.
- Le temps d'attente proposé V_n au client C_n est le temps qu'aura à attendre C_n avant d'être servis. Ce temps peut être supérieur à D_n , auquel cas C_n est perdu (il n'a pas pu atteindre le serveur).
- Le temps de séjour W_n est le temps que passe effectivement le client C_n dans le système (file + serveur),

$$W_n = (V_n + s_n)\mathbf{1}_{[V_n < D_n]} + D_n\mathbf{1}_{[V_n \geq D_n]}.$$

- Le temps d'attente W_n^q d'un client C_n est le temps effectivement passé par C_n dans la file,

$$W_n^q = V_n\mathbf{1}_{[V_n < D_n]} + D_n\mathbf{1}_{[V_n \geq D_n]}.$$

Pour tout système de file d'attente à perte, la probabilité de perte π_k de client C_k est la probabilité que ce client soit perdu. Par exemple, pour une file d'attente classique M/M/1/N fini, cette probabilité serait donnée par la probabilité qu'il y ait N 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 π_k est donnée par la probabilité que C_k se voit proposer un temps d'attente supérieur à son délai.

$$\pi_k = \mathbb{P}[C_k \text{ est perdu}] = P[V_n > D_n]. \quad (2.1)$$

2.5 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 de la file $\{Q_t, t \geq 0\}$ est le processus qui compte le nombre de clients dans la file à l'instant t . Ainsi pour tout t ,

$$Q_t = (X_t - S)^+ = \sup_{t \geq 0} (X_t - s, 0).$$

- Pour tout $n \geq 1$, on note par \hat{T}_n , le $n^{\text{ième}}$ instant de sortie du service, i.e l'instant de la $n^{\text{ième}}$ fin de service.

Le processus des fins de services $\{S_t, t \geq 0\}$ est alors le processus ponctuel comptant le nombre de services terminés jusqu'à l'instant t .

Autrement dit, pour tout t ,

$$S_t = \sum_{n \in \mathbb{N}^*} \mathbf{1}_{\hat{T}_n \leq t}. \quad (2.2)$$

- Pour tout $n \geq 1$, on note par \tilde{S}_n , la durée du $n^{\text{ième}}$ service rendu par le serveur, i.e la durée du service qui se termine à \hat{T}_n .
- Pour tout $n \geq 1$, on note par \tilde{T}_n , le $n^{\text{ième}}$ instant où un client est perdu. Le processus des pertes $\{P_t, t \geq 0\}$ est alors le processus comptant le nombre de pertes jusqu'à l'instant t .

pour tout t ,

$$P_t = \sum_{n \in \mathbb{N}^*} \mathbf{1}_{\tilde{T}_n \leq t}. \quad (2.3)$$

- Le processus des départs du système $\{d_t, t \geq 0\}$ est le processus comptant les instants de sorties du système (service ou pertes) jusqu'à l'instant t .

Autrement dit, pour tout t ,

$$d_t = S_t + P_t. \quad (2.4)$$

Si on réordonne les délais résiduels des Q_t clients présents dans la file à t par ordre croissant (pour les servir en EDF), 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 .

À 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 l'instant t d'un client dans la file d'attente à t .

À 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-la même les mesures 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 (loi 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 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_∞ La version stationnaire de X_t .

Q_∞ La version stationnaire de Q_t .

W^a La version stationnaire du temps d'attente.

W La version stationnaire du temps de séjour.

V 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, i.e 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 constructions 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 l'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.

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

Définition 2.3. Pour toute la file d'attente à perte, la probabilité de perte est donnée par,

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

Pour que la file d'attente avec clients impatientes, Π est donnée par la probabilité à l'état stationnaire qu'un client doive attendre plus longtemps que son délai pour être servi.

$$\begin{aligned} \Pi &= \lim_{n \rightarrow +\infty} \Pi_n = \lim_{n \rightarrow +\infty} \mathbb{P}[V_n > D_n]. \\ &= \lim_{n \rightarrow +\infty} \mathbb{P}[V_n > D] = \mathbb{P}[V > D]. \end{aligned} \tag{2.5}$$

Soit $X(t)$ le nombre de clients dans le système à l'instant t .

2.6 Le système M/M/1+M

Considérons la file M/M/1+M où les arrivées sont markoviens de taux λ , le temps de service est exponentiel de taux μ et le délai de patience est exponentiel de taux γ . Le processus $X(t)$ est un processus de naissance et de mort, dont les taux de transitions sont voir le graphe ci-dessous,

$$\lambda_n = \lambda, \quad \forall n > 0. \quad \text{et} \quad \mu_n = \mu + n\gamma, \quad \forall n > 1.$$

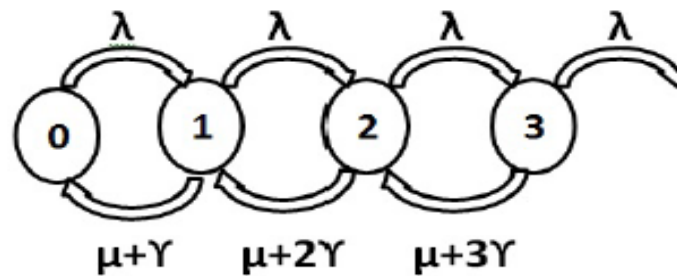


FIGURE 2.1 – Graphe des premiers états de la file M/M/1+M

Posons P_n la probabilité associée à l'état n . Ainsi, et par conservation de flux nous obtenons la formule de récurrence suivante :

$$\lambda P_n = [\mu + (n + 1)\gamma]P_{n+1}.$$

Et on montrera aisément par récurrence que,

$$P_n = \left(\prod_{k=1}^n \frac{\lambda}{\mu + k\gamma} \right) P_0. \quad (2.6)$$

Par souci de simplification posons,

$$\rho_k = \frac{\lambda}{\mu + k\gamma}, \quad \forall k \geq 1.$$

Ainsi,

$$P_n = \left(\prod_{k=1}^n \rho_k \right) P_0. \quad (2.7)$$

D'autre part, et comme les probabilités somment à 1, alors, on déduit que,

$$P_0 = \frac{1}{1 + \sum_{n=1}^{\infty} \prod_{k=1}^n \rho_k}.$$

P_0 existe car $\sum_{n=1}^{\infty} \prod_{k=1}^n \frac{\lambda}{\mu + k\gamma}$ converge.

En effet,

$$\begin{aligned} \mu + k\gamma > k\gamma &\implies \frac{1}{\mu + k\gamma} \leq \frac{1}{k\gamma} \\ &\implies \prod_{k=1}^n \frac{1}{\mu + k\gamma} \leq \left(\frac{\lambda}{\gamma}\right)^n \prod_{k=1}^n \frac{1}{k} = \left(\frac{\lambda}{\gamma}\right)^n \frac{1}{n!} \\ &\implies \sum_{n=1}^{\infty} \prod_{k=1}^n \frac{1}{\mu + k\gamma} \leq \sum_{n=1}^{\infty} \left(\frac{\lambda}{\gamma}\right)^n \frac{1}{n!} = e^{\frac{\lambda}{\gamma}} < \infty. \end{aligned}$$

Enfin et par souci de simplification nous garderons toujours la constante P_0 dans l'expression de P_n .

2.6.1 Calcul des mesures de performances

En gardant les simplifications déjà faites nous trouverons les résultats suivants :

$$L = \sum_{n=1}^{\infty} nP_n = n \sum_{n=1}^{\infty} P_0 \prod_{k=1}^n \rho_k. \quad (2.8)$$

En appliquant la loi de Little $L = \lambda W$, on trouve

$$\begin{aligned} W &= \frac{L}{\lambda} \\ &= \frac{P_0}{\lambda} \sum_{n=1}^{\infty} n \prod_{k=1}^n \rho_k. \end{aligned} \quad (2.9)$$

Probabilité de perte d'un client à l'état stationnaire,

$$\Pi = \mathbb{P}[V > D] = \frac{\gamma}{\lambda} \sum_{n=1}^{\infty} P_0 \prod_{k=1}^n \rho_k = \frac{\gamma}{\lambda} (1 - P_0). \quad (2.10)$$

Les résultats obtenus, lors du calcul des performances sont purement théoriques. Ainsi, dans la pratique nous utilisons des approches algorithmiques beaucoup plus intéressantes, comme la troncature des sommes infinies, afin de réduire le temps d'exécution des programmes les implémentant [32].

2.7 Le système M/M/1+D

Considérons à présent la file $M/M/1 + D$ où les arrivées sont markoviens de taux λ , le temps de service est aussi markovien de taux μ et le délai de patience est lui déterministe. Dans le cas où les délais initiaux sont égaux à d , Choi, Kim et Chung [7] proposent le calcul explicite des performances par l'étude de la loi stationnaire du processus markovien multidimensionnel $\{X(t), t \geq 0\}$ à valeurs dans l'espace $(\{0\} \times \{0, 1\}) \cup (\{1\} \times \{0, d\})$:

$$\begin{cases} X(t) = (0, c) \text{ s'il y a } c \text{ clients dans le système,} & \text{avec } c \leq 1; \\ X(t) = (1, Z_t) \text{ s'il y a au moins 2 clients dans le système.} \end{cases}$$

où Z_t est le temps passé dans le système à l'instant t par le client en service à cet instant. Ainsi, on peut calculer la loi stationnaire de $X(t)$ et par la suite, la densité f de Z , la version stationnaire de Z_t .

Par ailleurs, la propriété PASTA¹ nous permet de trouver la probabilité de perte,

$$\Pi = \frac{f(d)}{\lambda}.$$

où λ est l'intensité du processus des entrées et on retrouve bien le résultat de Barrer [2],

$$\Pi = \rho e^{(\lambda - \mu)d} \mathbb{P}[X_{\infty} = 0]. \quad (2.11)$$

1. Poisson Arrivals See Time Averages (Wolff 1982) : Avec le théorème ergodique, cela signifie qu'en régime permanent un nouveau client verra n clients devant lui avec une probabilité égale à la proportion du temps pendant laquelle le système est occupé par n clients.

2.8 Encadrement de la probabilité de perte d'un client :

L'absence de mémoire de la loi exponentielle permet d'exhiber des modèles markoviens simples pour le calcul des performances et donne en particulier une formule close pour la probabilité de perte Π .

Soit

$$\Pi_k = \mathbb{P}(\text{ le client } C_k \text{ est perdu} \mid \text{ sort impatient}). \quad (2.12)$$

Notons T'_k l'instant où 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 $T'_k \in [T_k, T_k + d]$ sinon, on a bien $T'_k = T_k + d$.

Soient aussi les événements suivants :

A_j^t : " C_j est en service à l'instant t "

B_j : " C_j est servi"

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

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

Ainsi,

$$\begin{aligned} \Pi_k &= \sum_{j=0}^{k-1} \mathbb{P}(A_j^{T_k+d}) \\ &= \sum_{j=0}^{k-1} \mathbb{P}(A_j^{T_k+d} \cap A_j^{T_{k-1}+d}), \quad (\text{car : } A_j^{T_k+d} \subset A_j^{T_{k-1}+d}) \\ &= \sum_{j=0}^{k-1} \mathbb{P}(A_j^{T_k+d} \mid A_j^{T_{k-1}+d}) \mathbb{P}(A_j^{T_{k-1}+d}). \end{aligned} \quad (2.13)$$

Comme les temps de service sont exponentiels, la probabilité qu'un client soit encore en service à l'instant $T_k + d$ sachant qu'il était déjà en service à l'instant $T_{k-1} + d$ est donnée par,

$$\mathbb{P}(Y > T_k - T_{k-1}) = \mathbb{P}(Y > U).$$

Ainsi,

$$\Pi_k = \sum_{j=0}^{k-1} \mathbb{P}(Y > U) \mathbb{P}(A_j^{T_{k-1}+d}). \quad (2.14)$$

$$= \mathbb{P}(Y > U) \sum_{j=0}^{k-1} \mathbb{P}(A_j^{T_{k-1}+d}). \quad (2.15)$$

$$= \mathbb{P}(Y > U) \left[\sum_{j=0}^{k-2} \mathbb{P}(A_j^{T_{k-1}+d}) + \mathbb{P}(A_{k-1}^{T_{k-1}+d}) \right]. \quad (2.16)$$

$$= \mathbb{P}(Y > U) \Pi_{k-1} + \mathbb{P}(Y > U) \mathbb{P}(A_{k-1}^{T_{k-1}+d}). \quad (2.17)$$

Or,

$$A_{k-1}^{T_{k-1}+d} \subset B_{k-1}.$$

Donc,

$$\begin{aligned} \Pi_k &\leq \mathbb{P}(Y > U) \Pi_{k-1} + \mathbb{P}(Y > U) \mathbb{P}(B_{k-1}). \\ \Pi_k &\leq \mathbb{P}(Y > U) \Pi_{k-1} + \mathbb{P}(Y > U) (1 - \Pi_{k-1}). \end{aligned} \quad (2.18)$$

Comme le système est stable alors, on déduit que,

$$\Pi = \lim_{k \rightarrow +\infty} \Pi_k \leq \mathbb{P}(Y > U). \quad (2.19)$$

D'autre part, on a :

$$\begin{aligned} \Pi_k &= P(Y > U) \Pi_{k-1} + P(Y > U) P(A_{k-1}^{T_{k-1}+d}). \\ &= P(Y > U) [\Pi_{k-1} + P(B_{k-1}) - P(\text{le service de } C^k \text{ se termine avant } T_{k-1} + d)]. \\ &= P(Y > U) [T_{k-1} + 1 - P(B_{k-1} \cap \{T'_{k-1} + Y_{k-1} < T_{k-1} + d\})]. \\ &\geq P(Y > U) [1 - (1 - \Pi_{k-1}) P(Y_{k-1} < d)]. \\ &\geq P(Y > U) - P(Y > U) P(Y_{k-1} < d) (1 - \Pi_{k-1}). \end{aligned}$$

Ainsi et par passage à la limite, on déduit que,

$$\Pi \geq \mathbb{P}(Y > U) - \mathbb{P}(Y > U) \mathbb{P}(Y < d) (1 - \Pi).$$

Où encore,

$$\frac{\mathbb{P}(Y > U) \mathbb{P}(Y > d)}{1 - \mathbb{P}(Y > U) \mathbb{P}(Y < d)} \leq \Pi \quad (2.20)$$

Finalement de (2.19) et (2.20), on déduit que :

$$\frac{\mathbb{P}(Y > U) \mathbb{P}(Y > d)}{1 - \mathbb{P}(Y > U) \mathbb{P}(Y < d)} \leq \Pi \leq \mathbb{P}(Y > U).$$

2.8.1 Application du résultat :

Cet encadrement est très utile, de fait de sa généralité, pour d'autres types de système :

- Cas de la file $D/M/1 + D$ où les arrivés sont périodiques de période p ($U=p$ p.s).

$$\frac{\mathbb{P}(Y > p) \mathbb{P}(Y > d)}{1 - \mathbb{P}(Y > p) \mathbb{P}(Y < d)} \leq \Pi \leq \mathbb{P}(Y > p).$$

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

- Cas de la file $M/M/1 + D$,

$$\frac{\mathbb{P}(Y > U) \mathbb{P}(Y > d)}{1 - \mathbb{P}(Y > U) \mathbb{P}(Y < d)} \leq \Pi \leq \mathbb{P}(Y > U).$$

On calcule d'abord $\mathbb{P}(Y > U)$,

D'après la formule des probabilités totale, on aura :

$$\mathbb{P}(Y > U) = \int_0^{+\infty} \mathbb{P}(Y > U | Y = x) f_Y(x) dx \quad (2.21)$$

$$= \int_0^{+\infty} \mathbb{P}(x > U | Y = x) \mu e^{-\mu x} dx \quad (2.22)$$

$$= \int_0^{+\infty} \mathbb{P}(x > U) \mu e^{-\mu x} dx \quad (\text{Car } Y \text{ et } U \text{ sont indépendantes}) \quad (2.23)$$

$$= \int_0^{+\infty} (1 - e^{-\lambda x}) \mu e^{-\mu x} dx \quad (2.24)$$

$$= \int_0^{+\infty} [\mu e^{-\mu x} - \mu e^{-(\lambda+\mu)x}] dx = 1 - \frac{\mu}{\mu + \lambda}. \quad (2.25)$$

D'où

$$\mathbb{P}(Y > U) = \frac{\lambda}{\lambda + \mu}.$$

$$\frac{\frac{\lambda}{\lambda + \mu} e^{-\mu d}}{1 - \frac{\lambda}{\lambda + \mu} (1 - e^{-\mu d})} \leq \Pi \leq \frac{\lambda}{\lambda + \mu}. \quad (2.26)$$

2.9 Le système $M/M/s + M$

Considérons la file $M/M/s + M$ où les arrivées sont markoviens de taux λ , le temps de service est exponentiel de taux μ , le nombre de serveurs est s et le délai de patience est exponentiel de taux γ .

Le processus $X(t)$ est un processus de naissance et de mort sur \mathbb{N} , dont les taux de transitions sont,

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

$$\mu_n = \begin{cases} n\mu, & 1 \leq n \leq s; \\ s\mu + (n-s)\gamma, & n \geq s \end{cases}$$

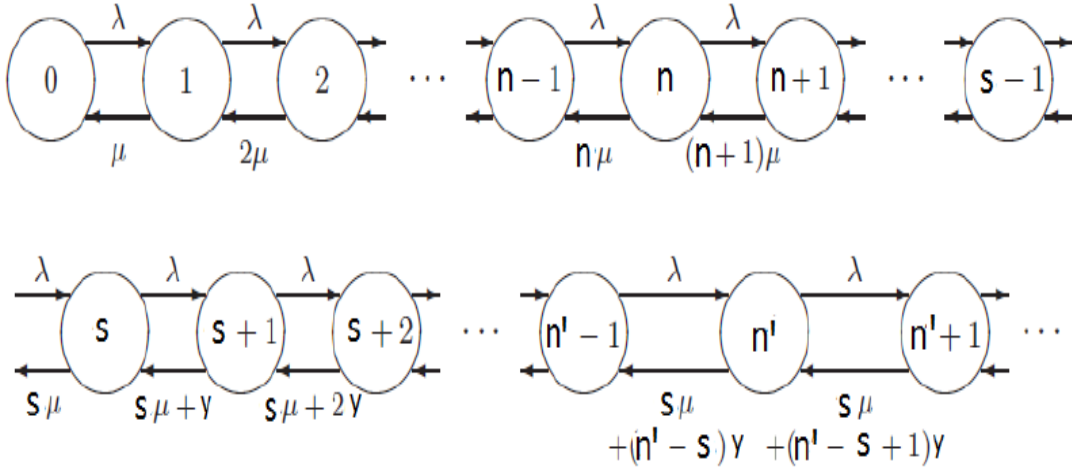


FIGURE 2.2 – Graphe de transition du système M/M/s+M

Si $0 \leq n < s$ Posons P_n la probabilité associée à l'état n . Ainsi, on a

$$P_n = \left(\prod_{i=0}^{n-1} \frac{\lambda_{i+1}}{\mu_{i+1}} \right) P_0.$$

On trouve,

$$P_n = \frac{\left(\frac{\lambda}{\mu + \gamma} \right)^n}{n!} P_0. \quad (2.27)$$

Si $n \geq s$

$$\lambda P_s = (s\mu + \gamma) P_{s+1}$$

$$P_{s+1} = \frac{\lambda}{(s\mu + \gamma)} P_s.$$

Et on montrera aisément par récurrence que,

$$P_n = \left(\frac{\lambda}{(s\mu + \gamma)} \right)^{n-s} P_s \quad (2.28)$$

On pose : $\rho = \frac{\lambda}{(s\mu + \gamma)}$

Alors,

$$P_n = \rho^{n-s} P_s, \quad \text{avec, } P_s = \frac{\left(\frac{\lambda}{\mu}\right)^s}{s!} P_0.$$

D'où,

$$P_n = \begin{cases} \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} P_0, & \text{si, } 0 \leq n \leq s; \\ \left(\frac{\lambda}{s\mu + \gamma}\right)^{n-s} \frac{\left(\frac{\lambda}{\mu}\right)^s}{s!} P_0, & \text{si, } n \geq s. \end{cases}$$

D'autre part, et comme les probabilités somment à 1, alors, on déduit que,

$$P_0 = \left[\sum_{n=0}^s \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} + \sum_{n=s+1}^{\infty} \left(\frac{\lambda}{s\mu + \gamma}\right)^{n-s} \frac{\left(\frac{\lambda}{\mu}\right)^s}{s!} \right]^{-1}. \quad (2.29)$$

P_0 existe si $\rho < 1$.

2.9.1 Calcul des mesures de performances

La probabilité q'un client qui entre doit attendre :

$$\begin{aligned} \mathbb{P}(\text{attente}) &= \mathbb{P}(X \geq s) = \sum_{n=s}^{\infty} \mathbb{P}_n. \\ &= \sum_{n=s}^{\infty} \left(\frac{\lambda}{s\mu + \gamma}\right)^{n-s} P_s \\ &= \sum_{n=s}^{\infty} \rho^{n-s} P_s \\ &= P_s \sum_{j=0}^{\infty} \rho^j \\ &= \frac{P_s}{1 - \rho}. \end{aligned}$$

$$\mathbb{P}(\text{attente}) = \mathbb{P}(X \geq s) = \sum_{n=s}^{\infty} \mathbb{P}_n = \frac{P_s}{1 - \rho}, \quad \text{avec } \rho = \frac{\lambda}{(s\mu + \gamma)} < 1 \quad \text{et } P_s = \frac{\left(\frac{\lambda}{\mu}\right)^s}{s!} P_0. \quad (2.30)$$

$$\begin{aligned}
L_q &= \mathbb{E}(X_q) = \sum_{n=s+1}^{\infty} (n-s)P_n. \\
&= \sum_{n=s+1}^{\infty} (n-s)\rho^{n-s}P_s. \\
&= P_s \sum_{n=s+1}^{\infty} (n-s)\rho^{n-s}. \\
&= P_s \rho \sum_{j=1}^{\infty} j\rho^{j-1}. \\
&= P_s \rho \frac{1}{(1-\rho)^2}. \\
L_q &= \frac{\rho}{(1-\rho)^2} P_s = \frac{\rho}{(1-\rho)^2} \frac{(\frac{\lambda}{\mu})^s}{s!} P_0. \\
L_q &= \mathbb{E}(X_q) = \sum_{n=s+1}^{\infty} (n-s)P_n = \frac{\rho}{(1-\rho)^2} P_s. \tag{2.31}
\end{aligned}$$

$$W_q = \frac{P_s}{s\mu(1-\rho)^2}. \tag{2.32}$$

$$W = \frac{P_s}{s\mu(1-\rho)^2} + \frac{1}{\mu}. \tag{2.33}$$

La probabilité de Perte

La probabilité de perte Π , qui est le rapport entre le taux de perte et le taux d'arrivée à l'état d'équilibre, est donné par,

$$\Pi = \frac{\gamma}{\lambda} \sum_s^{\infty} P_n = \frac{\gamma}{\lambda} \frac{P_s}{1-\rho}. \tag{2.34}$$

Pour plus de détails sur ce système voir les livres [11], [19], [30] et [35]. Ce système a été largement utilisé comme modèle de base pour centres d'appels [16],[27], [36] et [37].

2.10 Le système M/M/s+D

Depuis longtemps (Erlang [10]), on sait étudier avec des processus markoviens simples, l'état stationnaire d'une file d'attente classique $M/M/s + N$ gérée par FCFS où N est la capacité de la file. 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 qu'il y ait $s + N$ clients dans le système. Le délai affecté aux clients apparait comme une limitation (mais de taille aléatoire) de la capacité du système. Barrer[2] est le premier qui a soulevé le problème pour une file d'attente $M/M/s + D - FCFS$: 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. Il exhibe la loi stationnaire du nombre de clients dans le système et donne Π la probabilité de perte :

$$\Pi = \frac{1}{s!} \rho^s e^{(\lambda - s\mu)d} \mathbb{P}[X_\infty = 0] \quad (2.35)$$

où d est la valeur des délais initiaux. Gnedenko et Kovalenko [19] ont pointé et corrigé des erreurs dans la preuve de (2.35), 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 [14] a donné la loi du temps d'attente dans une file $GI/M/1 + D - FIFO$. Des considérations du même type ont mené au calcul de la probabilité de perte d'une file $M/G/1 + D - FIFO$ (Kok et Tijims,[12]) et d'une file $M/G/1 + D - FIFO$ avec délais jusqu'à la fin de service [13]. Citons encore [8], [18] et [33] pour d'autres résultats sur des modèles de ce type.

2.11 Files d'attente avec impatience : Deux classes de clients

Afin de modéliser notre problème nous allons étudier le cas de deux classes de clients avec les contraintes suivantes

- Les clients de la classe i arrivent suivant un processus de Poisson de taux λ_i .
- Le temps de service est le même pour tous les clients et suit une loi exponentielle de paramètre μ .
- Les clients prioritaires ont une priorité stricte par rapport à ceux non prioritaires. i.e qu'à l'arrivée d'un client prioritaire, le client non prioritaire en service quitte le serveur et rejoint la file d'attente et son temps de service déjà effectué est gardé en mémoire.
- Un client peut abandonner le système après un temps d'attente distribué suivant une loi exponentielle de taux γ_i et ce, jusqu'au début de son service. Autrement dit, il ne pourra plus quitter le système après être pris en charge par le serveur.

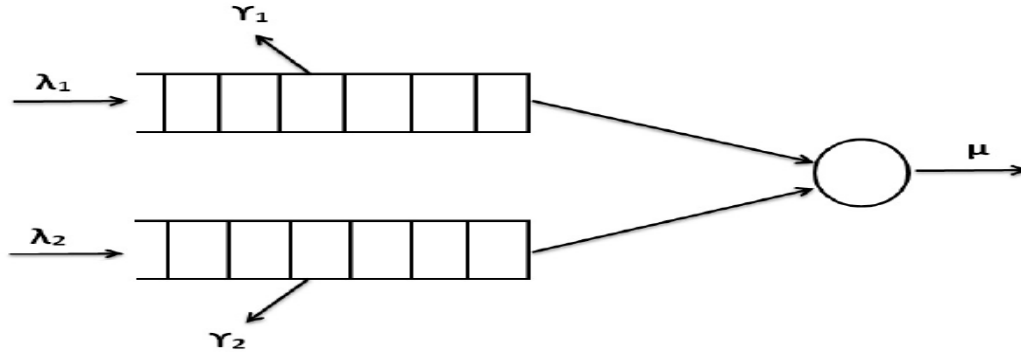


FIGURE 2.3 – Un serveur avec deux classes de clients qui peuvent abandonner

Nous étudions ici plusieurs mesures de performance, à savoir la probabilité d’abandon, le temps d’attente moyen et le nombre de clients moyen dans le système.

Ce modèle, bien qu’il paraisse simple, représente vraiment le pilier de toute approche de résolution de ce problème car il permet de modéliser le besoin tout en permettant une généralisation dans certains cas de figures.

2.11.1 Cas particulier

Dans cette section nous allons étudier le cas où on aura deux classes de clients de même taux d’abandon γ et avec une priorité stricte [1].

Notons $X(t) = X_1(t) + X_2(t)$, qui est, le nombre de clients dans le système caractérisé par,

$$\lambda = \lambda_1 + \lambda_2$$

$$\mu = \mu_1 + \mu_2$$

$$\gamma = \gamma_1 + \gamma_2$$

Ainsi, cette considération nous permet de constater que X et X_1 évoluent comme un système à une seule classe de clients. De plus, en exploitant les résultats déjà trouvés précédemment on pourra déduire les mesures de performances de la deuxième classe comme suit

$$L_2 = L - L_1$$

$$W_2 = W - W_1$$

Ainsi, nous pourrions généraliser ce procédé à n classes de clients avec un calcul de proche en proche.

2.11.2 Cas avec priorité stricte et avec préemption

Dans le cas d'une priorité stricte, nous considérons que la première classe est prioritaire par rapport à la deuxième en se plaçant toujours dans le cas des délais éliminatoires jusqu'à début du service. D'autre part, nous supposons que les clients de la classe 1 ont une priorité préemptive sur ceux de la classe 2 et que ces derniers sont patients indéfiniment après le début de leur service. Autrement dit, un client de la classe 2 n'est servi que s'il n'y a aucun client de la classe 1 dans le système et il quitte immédiatement le serveur, pour rejoindre la file, dès l'arrivée d'un client de la classe 1 même s'il n'a pas encore terminé son service.

Il est clair que le comportement de la classe prioritaire peut être approché par le cas d'une seule classe de clients. On pourra ainsi, reprendre tous les résultats déjà trouvés précédemment. Par contre, on ne peut rien dire a priori sur le comportement de l'autre classe.

D'autre part, nous aurons besoin de la distribution de la période où le système est occupé par les clients de la classe la plus prioritaire (classe 1).

Ainsi, cette période est définie, lorsqu'elle est générée par k clients de la classe 1, comme étant la durée écoulée entre le moment où le serveur commence à servir le premier client parmi ces k clients et le moment où aucun clients de la classe 1 n'est plus présent dans le système. Du moment où les clients de la classe 1 se comportent comme étant les seuls présents dans le système, nous pouvons reprendre les résultats établis par Rao [29] qui a étudié la file $M/G/1 + M$ avec une seule classe de clients impatientes. Posons $B_k(t)$ la fonction de répartition de cette période initiée par k clients de la classe 1 dans la file et $b_k(t)$ sa densité ayant $\mathcal{L}b_k(p)$ comme transformée de Laplace. D'autre part, notons $S(t)$ la fonction de répartition du temps de service, $\hat{S}(t) = 1 - S(t)$ celle du temps de service complémentaire et $\mathcal{L}S(p)$ sa transformée de Laplace.

Comme le temps de service est Markovien de taux μ alors on a,

$$S(t) = 1 - e^{-\mu t} \Rightarrow \mathcal{L}S(p) = \frac{1}{p} - \frac{1}{\mu + p} \quad (2.36)$$

Ainsi les résultats de Rao [29] se simplifient comme suit

$$\mathcal{L}b_0(p) = \frac{\frac{\mu}{\mu+p} + \sum_{i=1}^{\infty} \frac{v^i}{i!} \psi(i-1, p) \frac{\mu}{\mu+p+i\gamma_1}}{1 + \sum_{i=1}^{\infty} \frac{v^i}{i!} \psi(i-1, p)}. \quad (2.37)$$

$$\mathcal{L}b_k(p) = \frac{\frac{\mu}{\mu+p} + \sum_{i=1}^{\infty} (-1)^i \psi(i-1, p) \frac{\mu}{\mu+p+i\gamma_1} l(k, i)}{1 + \sum_{i=1}^{\infty} \frac{v^i}{i!} \psi(i-1, p)}, \quad \forall k \in \mathbb{N}^*. \quad (2.38)$$

Avec,

$$v = \frac{\lambda_1}{\gamma_1}$$

$$l(k, i) = \begin{cases} \sum_{j=0}^i \frac{(-1)^j v^j}{j!} \binom{k}{i-j}, & \text{si, } i \leq k; \\ \sum_{j=i-k}^i \frac{(-1)^j v^j}{j!} \binom{k}{i-j}, & \text{si, } i > k. \end{cases}$$

$$\psi(i, p) = \prod_{j=0}^i \frac{p + j\gamma_1}{\mu + p + j\gamma_1}$$

D'autre part, posons $g_1(x)$ (resp. $g_2(x)$) la densité du temps d'attente d'un client de la classe 1 (resp. la classe 2) jusqu'à ce que son service sera initié pour la première fois. Comme la préemption est autorisée, g_2 n'affecte pas g_1 et par conséquent, cette dernière peut être retrouvée à partir de plusieurs ouvrages (voir [1, 9, 32]).

Notons au passage que ne nous aurons besoin ici que de $b_0(t)$ qui est la densité de la période active initiée par un client de la classe 1 entrant dans un système vide.

Notre système alterne entre une phase inactive (PI) où il n'y a aucun client dans le système et une phase active (PA) qui est initiée par l'arrivée d'un client.

De plus, nous appelons cycle de phase, l'union de ces deux phases. Ainsi, une PA s'arrête au moment où tout les clients ont quitté le système, dû à des services ou à des abandons. Selon que le client de la classe 2 arrive durant une PI ou une PA, la fonction g_2 est donnée par,

$$g_2(x) = g_2(x/PI)\mathbb{P}(PI) + g_2(x/PA)\mathbb{P}(PA), \quad \forall x \in \mathbb{R}^+ \quad (2.39)$$

D'autre part, comme chaque client qui arrive durant une PI sera automatiquement pris en charge par le serveur, la transformée de Laplace de $g_2(x/PI) = 1$ sera égale à 1. Par contre, pour un client qui arrive durant une PA, il y a deux cas de figures à traiter.

1. Il arrive juste après l'initiation de la PA et doit attendre dans la file mais son temps d'attente dépend du type du client qui a initié la PA. Si la PA a été initié par un client de la classe 1, il doit attendre jusqu'à ce que ce client soit servi ainsi que tous les autres clients de la classe 1 rejoignant la file. Ce temps d'attente est égal à la période initiée par un client arrivant quand le serveur est vide dans une file $M/G/1 + M$ avec une seule classe de clients impatientes qui est largement étudiée dans la littérature.

Sinon, il doit attendre jusqu'à ce que le client de la classe 2, qui a initié la PA, soit servi ainsi que tous les éventuels clients de la classe 1 qui sont entrés dans le système. Nous allons définir ce temps d'attente comme le temps de complétion de service d'un client de la classe 2, qui est le temps écoulé entre l'instant où il a été pris en charge par le serveur pour la première fois et l'instant où il termine son service. Posons $C_2(t)$ la fonction de répartition de ce temps d'attente et $c_2(t)$ sa densité ayant $\mathcal{L}c_2(p)$ comme transformée de Laplace et qui peut être trouvé par [17],

$$\mathcal{L}c_2(p) = \frac{\mu}{\mu + p + \lambda_1 - \lambda_1 \mathcal{L}b_0(p)} \quad (2.40)$$

Notons au passage, que la PA sera initiée par un client de la classe 1 avec une probabilité de $\frac{\lambda_1}{\lambda_1 + \lambda_2}$ sinon par un client de la classe 2 avec une probabilité $\frac{\lambda_2}{\lambda_1 + \lambda_2}$. Posons $I_j(t)$ la fonction de répartition du premier saut du de la phase active. Ainsi,

$$I_J(t) = \frac{\lambda_1}{\lambda_1 + \lambda_2} B_0(t) + \frac{\lambda_2}{\lambda_1 + \lambda_2} C_2(t), \quad \forall t \in \mathbb{R} \quad (2.41)$$

Nous constatons donc que la durée de PA décroît avec un taux de 1 jusqu'à ce qu'on atteigne la PI ou jusqu'à ce qu'un autre client de la classe 2 entre dans le système. De plus, chaque fois qu'un client de la classe 2 arrive durant la PA, la durée de PA saute avec le temps de complétion de service de ce client et s'il n'y a aucun client, alors la PA arrive à son bout et on entre en PI.

Afin de calculer le nombre moyen avec lequel décroît la PA par rapport au nombre moyen avec lequel elle croît au niveau x , nous pouvons nous baser sur les résultats donnés par Brill [5, 6],

$$\mathbb{E}[PA]g_2(x/PA) = 1 - I_J(x) + \lambda_2 \mathbb{E}[PA] \int_0^x \hat{C}_2(x-t)e^{-\gamma_2 t} g_2(t/PA) dt. \quad (2.42)$$

Notons au passage que la partie droite de cette équation indique que le niveau x peut être dépassé si le saut initial I_J dépasse x avec une probabilité de $1 - I_J(x)$ ou si un client de la classe 2 arrivant durant la PA est suffisamment patient pour attendre les clients qui sont devant lui et ce avec un nombre moyen de $\lambda_2 \mathbb{E}[PA]$. Ainsi,

$$g_2(x/PA) = \frac{1 - I_J(x)}{\mathbb{E}[PA]} + \lambda_2 \int_0^x \hat{C}_2(x-t)e^{-\gamma_2 t} g_2(t/PA) dt. \quad (2.43)$$

Cette dernière équation est une équation intégrale de Volterra au second degré et peut être résolu par récurrence (voir [3]).

Après plusieurs étapes de récurrence et comme $\mathbb{E}[b_0]$ est le temps moyen que dure la période, initiée par un client de la classe 1, où le système est occupé, et comme $\mathbb{E}[PI] = \frac{1}{(\lambda_1 + \lambda_2)}$.

Alors,

$$P[PI] = \frac{\mathbb{E}[PI]}{\mathbb{E}[PI] + \mathbb{E}[PA]} = \frac{1}{1 + (\lambda_1 + \lambda_2)\mathbb{E}[PA]} \quad (2.44)$$

Ainsi,

$$P[PA] = 1 - \mathbb{E}[PI] = \frac{(\lambda_1 + \lambda_2)\mathbb{E}[PA]}{1 + (\lambda_1 + \lambda_2)\mathbb{E}[PA]} \quad (2.45)$$

En combinant les différents résultats trouvés et en travaillant dans le domaine de Laplace, on trouve finalement

$$\begin{aligned}\mathcal{L}g_2(p) &= \frac{\lambda_1 + \lambda_2}{1 + (\lambda_1 + \lambda_2)\mathbb{E}[PI]} \\ &+ \left[\frac{1}{\lambda_1 + \lambda_2} + \frac{1 - \mathcal{L}I_J(p)}{p} + \lambda_2 \mathcal{L}C_2(p) \mathcal{L}\psi_{n-1}(p + \gamma_2) \right]\end{aligned}$$

2. Il arrive à un instant arbitraire après l'initiation de la PA mais comme le système est Markovien sans mémoire, on se ramène au premier cas.

2.11.3 Mesures de performances

Il est clair que le système est stable (voir [22]). Posons pour $i \in \{1, 2\}$:

$W_{i,S}$ Le temps d'attente moyen des clients qui rejoignent le serveur.

$W_{i,A}$ Le temps d'attente moyen des clients qui abandonnent la file.

$W_{i,T}$ Le temps d'attente moyen de tous les clients.

Π_i La probabilité de perte d'un client de la classe i .

$F_i(t)$ La fonction de répartition de la charge du travail apportée par la classe i .

Il est clair que le serveur est soit vide avec la probabilité $P[PI]$ soit il sert un client de la classe 1 ou de la classe 2. Mais comme la présence des clients de la classe 2 n'a aucun impact sur les clients de la classe 1, alors en posant P_0 la probabilité qu'aucun client de la classe 1 n'est présent dans le système, le serveur servira un client de la classe 1 avec la probabilité $1 - P_0$. Ainsi, la probabilité qu'un client de la classe 2 soit admis en service, $P_{2,S}$, est égale à $P_0 - P[PI]$. Ainsi, Stanford [32] nous permet d'écrire :

$$\Pi_1 = 1 - \frac{1 - P_0}{\rho_1} \quad (2.46)$$

$$\Pi_2 = 1 - \frac{P_0 - P[PI]}{\rho_2} \quad (2.47)$$

Avec :

$$\rho_1 = \frac{\lambda_1}{\mu} \quad \text{et} \quad \rho_2 = \frac{\lambda_2}{\mu}.$$

De plus, Stanford montre que,

$$P[W_{i,S} = 0] = \frac{F_i(0)}{1 - \Pi_i} \quad \text{avec} \quad F_1(0) = P_0. \quad (2.48)$$

Comme un client de la classe 2 ne peut commencer immédiatement son service que si le système est vide alors $F_2(0) = P[PI]$. Ainsi, Stanford permet de trouver :

$$W_{i,S}(x) = \int_0^x \frac{dF_i(t)e^{-\gamma t}}{1 - \Pi_i} dt. \quad (2.49)$$

$$W_{i,T}(x) = 1 - e^{-\gamma_2 x} + e^{-\gamma_2 x} F_i(x). \quad (2.50)$$

Notons au passage que $W_{i,S}$ peut être numériquement calculé s'il est exprimé en domaine de Laplace.

$$\mathcal{L}W_{i,S}(p) = \frac{\mathcal{L}F_i(\gamma_i + p)}{p(1 - \Pi_i)} \quad (2.51)$$

Enfin on a,

$$W_{i,A}(x) = \frac{(1 - \Pi_i)W_{i,S}(x) - W_{i,T}(x)}{\Pi_i}. \quad (2.52)$$

Chapitre 3

Etude du système $M/M/s/(s + N)$ avec clients impatientes

Les modèles de file d'attente ont été largement utilisés pour modéliser les performances des centres d'appel avec des clients impatientes, ce qui signifie que les clients en attente peuvent quitter la file avant d'obtenir un service.

La file d'attente $M/M/s$ avec des clients impatientes est appelé " un modèle d'Erlang-A, "A" pour abandon " par Maudelbaum et Zeltyn en contraste avec le modèle bien connu Erlang-N, $M/M/s$ avec seulement des clients patients. Pour une synthèse sur les modèles de files d'attente des modèles pour les centres d'appels, voir par exemple ,Gans et al. [15] et Koole et Maudelbaum [25].

3.1 Description du modèle

On considère le système de file d'attente $M/M/s/(s + N)$ avec clients impatientes où les arrivées forment un processus de Poisson de taux λ , les temps de services sont exponentielles de moyenne $\frac{1}{\mu}$. La capacité de la file N est finie, le délai de patience de chaque client est exponentielle de moyenne $\frac{1}{\gamma}$. Une fois le service est commencé, il ne quitte pas le système.

Ce système est représenté dans la figure 3.1.

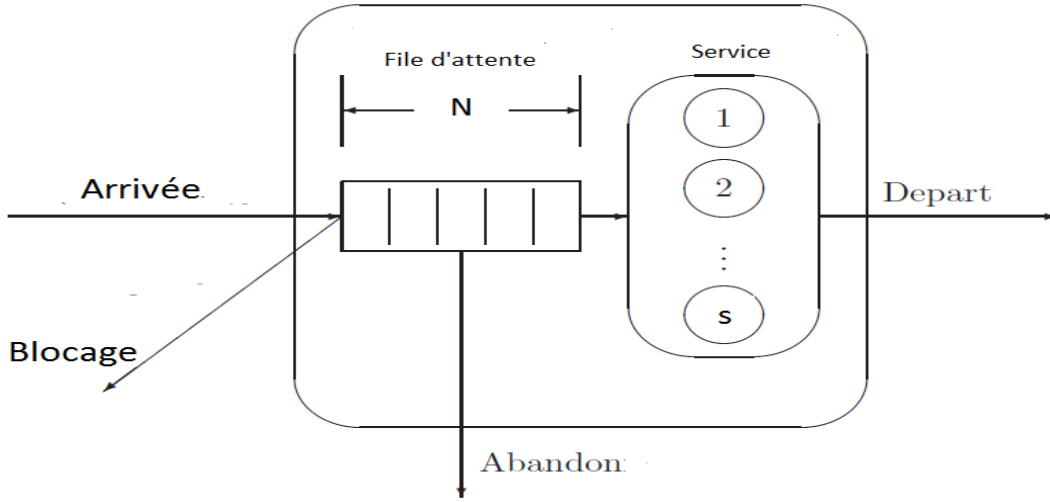


FIGURE 3.1 – Shéma du système $M/M/s/(s+N)$ avec impatience

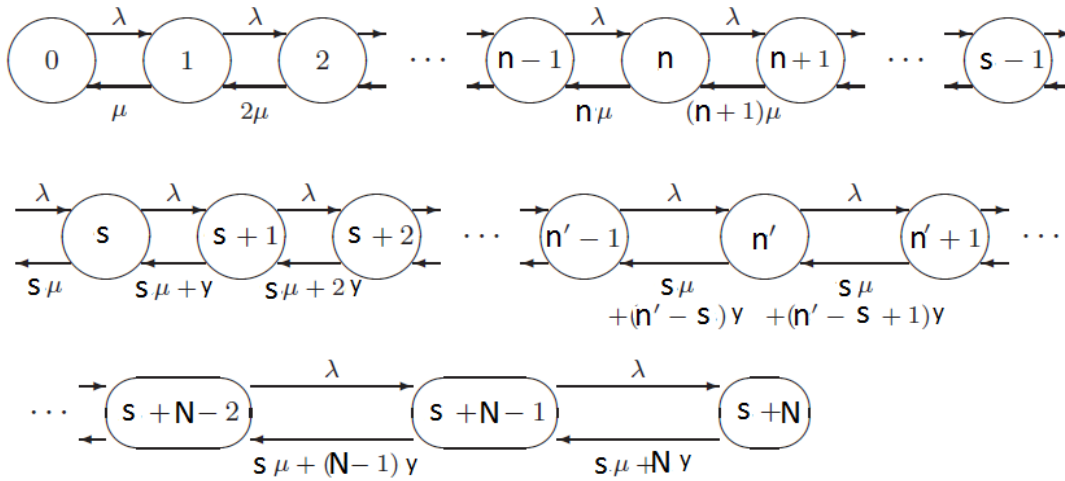


FIGURE 3.2 – Graphe de transition du système $M/M/s/(s+N)$ avec impatience

On a la distribution stationnaire du système est définie par :

$$P_n = \lim_{t \rightarrow +\infty} P\{X(t) = n\}, \quad 0 \leq n \leq s + N.$$

L'ensemble des équations de balances est donnée par :

$$\lambda P_0 = \mu P_1.$$

$$(\lambda + n\mu)P_n = \lambda P_{n-1} + (n+1)\mu P_{n+1}, \quad 1 \leq n \leq s-1.$$

$$(\lambda + s\mu + (n - s)\gamma)P_n = \lambda P_{n-1} + [s\mu + (n - s + 1)\gamma]P_{n+1}, \quad s \leq n \leq s + N - 1.$$

$$(s\mu + N\gamma)P_{s+N} = \lambda P_{s+N-1}$$

Avec La condition de Normalisation est donnée par :

$$\sum_{n=0}^{s+N} P_n = 1$$

L'ensemble des équations de balances est équivalent à l'ensemble des équations suivantes :

$$\lambda P_{n-1} = \begin{cases} n\mu P_n, & 1 \leq n \leq s - 1; \\ [s\mu + (n - s)\gamma]P_n, & s \leq n \leq s + N. \end{cases}$$

Donc, pour $1 \leq n \leq s$:

$$P_n = P_0 \frac{\rho^n}{n!} \quad \text{avec,} \quad \rho = \frac{\lambda}{\mu} \quad (3.1)$$

Pour $1 \leq n \leq s + N$, nous avons,

$$\begin{aligned} P_{n+s} &= \frac{\lambda}{s\mu + n\gamma} P_{s+n-1} = \frac{\lambda}{s\mu[1 + \frac{n\gamma}{s\mu}]} P_{s+n-1} \\ &= \frac{\frac{\rho}{s}}{1 + n\frac{\xi}{s}} P_{s+n-1} = \frac{\frac{\rho}{s}}{1 + n\frac{\xi}{s}} P_{s+n-2} \\ &= \dots \\ &= P_0 \frac{\rho^s}{s!} \frac{(\frac{\rho}{s})^n}{\prod_{j=1}^n (1 + j\frac{\xi}{s})}. \end{aligned}$$

$$P_{n+s} = \frac{\lambda}{s\mu + n\gamma} P_{s+n-1} = P_0 \frac{\rho^s}{s!} \frac{(\frac{\rho}{s})^n}{\prod_{j=1}^n (1 + j\frac{\xi}{s})}. \quad (3.2)$$

Où $\xi = \frac{\gamma}{\mu}$: est le taux moyen de patience par le temps moyen de service. En utilisant condition de normalisation, on détermine :

$$\frac{1}{P_0} = \sum_{n=0}^{s-1} \frac{\rho^n}{n!} + \frac{\rho^s}{s!} \frac{(\frac{\rho}{s})^n}{\prod_{j=1}^n (1 + j\frac{\xi}{s})}. \quad (3.3)$$

Ainsi, on trouve trouvé $\{P_n, 0 \leq n \leq s + N\}$ explicitement.

Pour les exemples d'illustration, nous supposons que, $s = 5, N = 10, \mu = 1$.

Dans la figure 3.3 représente $\{P_n, 0 \leq n \leq (s + N)\}$. pour $\rho = 5$ et pour différentes valeurs de $\xi = 0, 0.5, 1$ et 2.

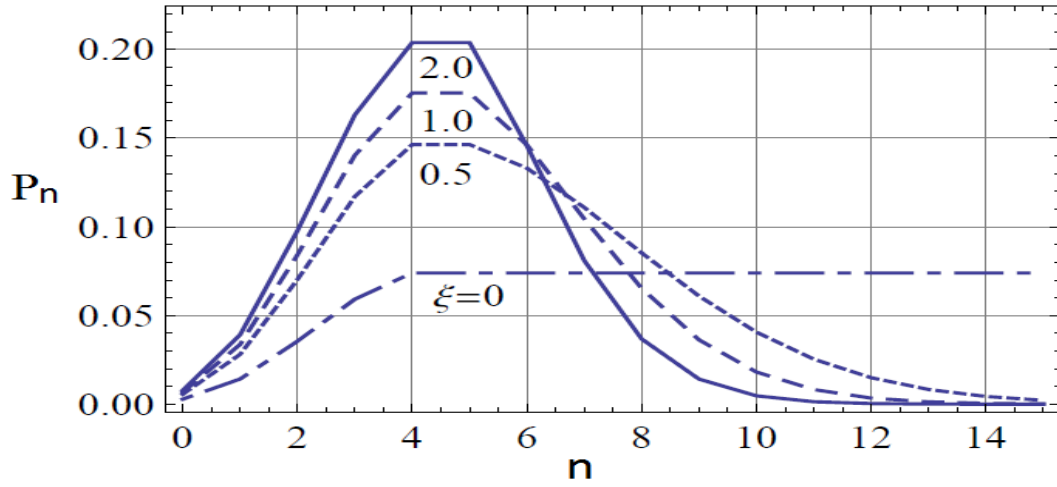


FIGURE 3.3 – La distribution $\{P_n, 0 \leq n \leq s + N\}$, pour le nombre moyen de clients dans le système

Remarque 3.1. le cas $\xi = 0$ signifie que les clients n'abandonne jamais, (on retrouve le résultat pour le système $M/M/s/K$).

3.2 Les mesures de performances

3.2.1 La probabilité de blocage

Les clients qui arrivent au système quand toute la file est pleine, à savoir il y a $(s + N)$ clients dans le système, sont bloqués. Ils sont perdus à jamais.

D'après la propriété de PASTA la distribution de probabilité à l'heure d'arrivée est la même à un moment quelconque. Par conséquent, la probabilité de blocage \mathcal{B} des clients est :

$$\mathbb{P}\{\mathcal{B}\} = P_{s+N} = P_0 \frac{\rho^s}{s!} \frac{\left(\frac{\rho}{s}\right)^n}{\prod_{j=1}^n (1 + j \frac{\xi}{s})}. \quad (3.4)$$

Dans la figure 3.4, On représente $\mathbb{P}\{\mathcal{B}\}$ en fonction de ρ pour différentes valeurs de $\xi = 0, 0.5, 1$ et 2 .

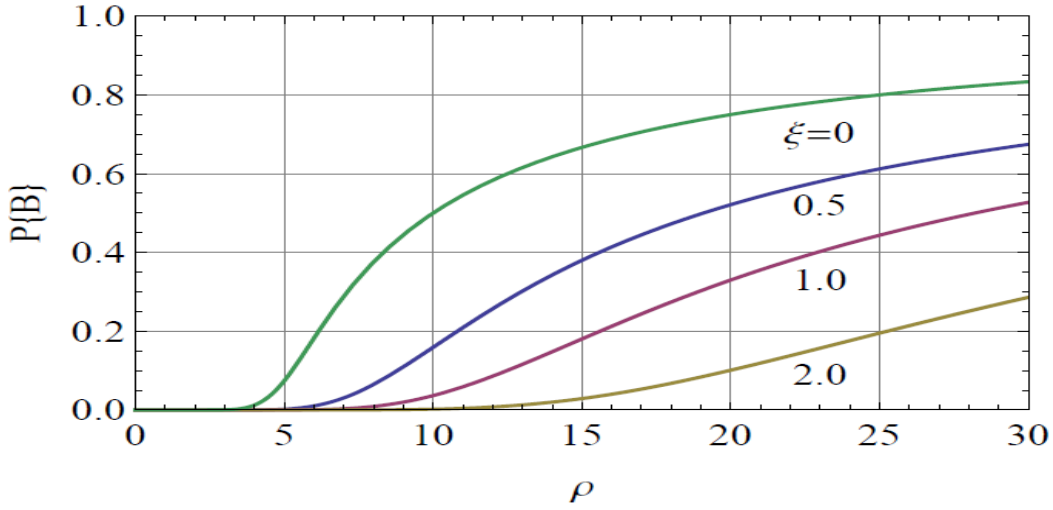


FIGURE 3.4 – La probabilité de blocage $P\{\mathcal{B}\}$

Remarque 3.2. on remarque que la probabilité de blocage $P\{\mathcal{B}\}$ augmente à mesure que ξ diminue.

3.2.2 Nombre moyen de clients dans la file

Soit X_q : " le nombre de clients dans la file d'attente à un temps arbitraire ".
S'il y a $n + s$ clients dans le système, il y a n clients dans la file d'attente. Par conséquent, nous avons :

$$\mathbb{P}\{X_q = 0\} = \sum_{n=0}^s P_n = \sum_{n=0}^s P_0 \frac{\rho^n}{n!} \quad (3.5)$$

$$\mathbb{P}\{X_q = n\} = P_{s+n} = P_0 \frac{\rho^s}{s!} \frac{\binom{\rho}{s}^n}{\prod_{j=1}^n (1 + j \frac{\rho}{s})}, \quad 1 \leq n \leq s. \quad (3.6)$$

Le nombre moyen de clients dans la file d'attente est donnée par :

$$L_q = \sum_{n=1}^N n P_{s+n} = P_0 \frac{\rho^s}{s!} \sum_{n=1}^N \frac{n \binom{\rho}{s}^n}{\prod_{j=1}^n (1 + j \frac{\rho}{s})} \quad (3.7)$$

Quand on trace L_q en fonction de ρ , on a les limites suivantes :

$$\lim_{\rho \rightarrow 0} L_q = 0 ; \quad \lim_{\rho \rightarrow \infty} L_q = N$$

Dans la figure 3.5, on représente L_q en fonction de ρ pour différentes valeurs de $\xi = 0, 0.5, 1$ et 2 .

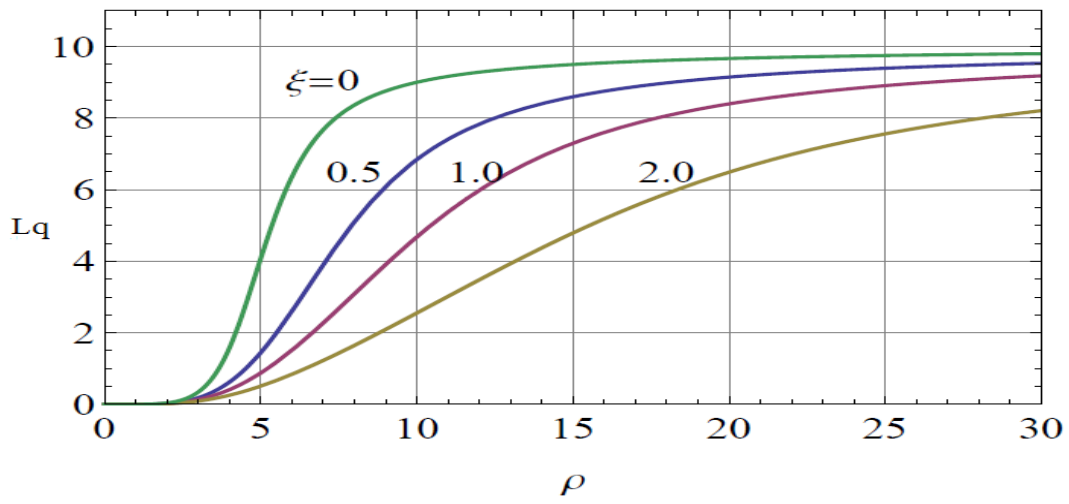


FIGURE 3.5 – Le nombre moyen de clients dans la file

Remarque 3.3. Le nombre moyen L_q diminue à mesure que ξ augmente car : beaucoup de clients quittent le système.

3.2.3 La probabilité d'abandon

Etant donné que chaque client quitte la file d'attente avec un taux γ , la fraction des clients qui abandonnent est donné par :

$$\mathbb{P}\{Ab\} = \frac{1}{\lambda} \sum_{n=1}^N (n\gamma) P_{s+n} = \frac{\gamma}{\lambda} P_0 \sum_{n=1}^N \frac{n \left(\frac{\rho}{s}\right)^n}{\prod_{j=1}^n (1 + j \frac{\xi}{s})} \quad (3.8)$$

Ainsi, nous avons la relation suivante :

$$\gamma L_q = \lambda \mathbb{P}\{Ab\}$$

Après, nous dérivons la relation :

$$\lambda(1 - \mathbb{P}\{B\}) = \gamma L_q + \mu \mathbb{E}[Z].$$

avec $s + N$ est le nombre de clients dans le système.

La probabilité qu'un client arbitraire abandonne est donnée par :

$$\mathbb{P}\{Ab|Nb\} = \frac{\mathbb{P}\{Ab\}}{1 - \mathbb{P}\{B\}} = \frac{\gamma L_q}{\gamma L_q + \mu \mathbb{E}[N]} \quad (3.9)$$

où l'événement Nb est qu'un client qui arrive soit acceptée.

Depuis,

$$\begin{aligned} \lim_{\rho \rightarrow +\infty} \mathbb{P}\{Ab|Nb\} &= \frac{\gamma N}{\gamma N + \mu s} \\ &= \frac{1}{1 + \frac{s}{(N\xi)}}. \end{aligned}$$

Dans la figure 3.6, On représente $P\{Ab|Nb\}$ en fonction de ρ pour différentes valeurs de $\xi = 0.5, 1$ et 2 .

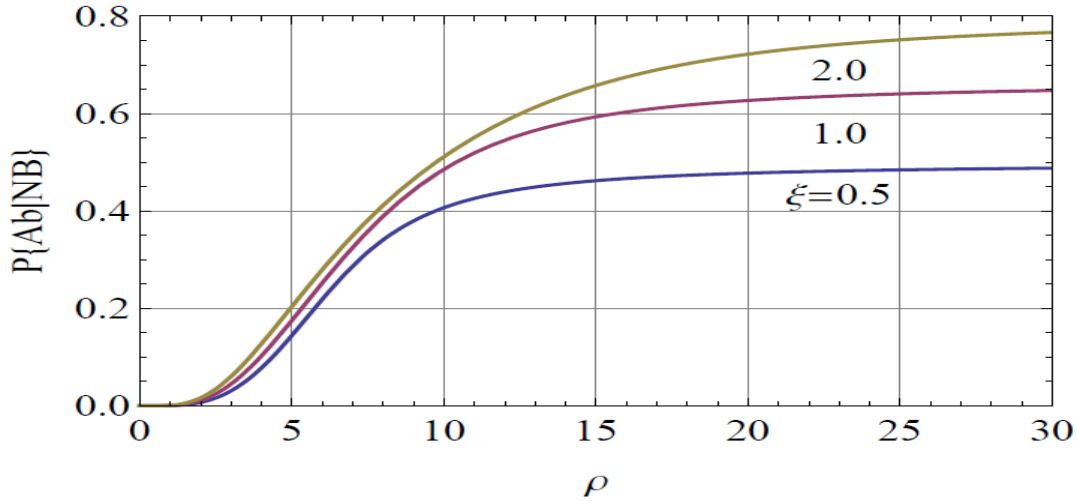


FIGURE 3.6 – La probabilité d’abandon $\mathbb{P}\{Ab|Nb\}$

Remarque 3.4. On note que $\mathbb{P}\{Ab|Nb\}$ augmente à mesure que ξ augmente.

3.2.4 Nombre moyen de clients en service

Soit Z : ”le nombre de clients en service à l’instant t ”.

Après on a

$$\mathbb{P}\{Z = n\} = P_n = P_0 \frac{\rho^n}{n!}, \quad 1 \leq n \leq s - 1. \quad (3.10)$$

$$\mathbb{P}\{Z = s\} = \sum_{n=0}^N P_{s+n} = P_0 \frac{\rho^s}{s!} \sum_{n=0}^N \frac{\binom{\rho}{s}^n}{\prod_{j=1}^n (1 + j \frac{\xi}{s})}. \quad (3.11)$$

Le nombre moyen de clients en service est égale au nombre moyen de serveurs occupés, qui est donné par :

$$\mathbb{E}[Z] = \sum_{n=1}^s n P_n + s \sum_{n=1}^N P_{s+n}. \quad (3.12)$$

Quand : $\lambda P_{n-1} = n \mu P_n$ pour $1 \leq n \leq s$, on a :

$$\mu \sum_{n=1}^s n P_n = \lambda \sum_{n=1}^s P_{n-1} = \lambda \sum_{n=0}^{s-1} P_n.$$

Par conséquent, on aura :

$$\mathbb{E}[Z] = \rho \sum_{n=1}^s P_{n-1} = \lambda \sum_{n=0}^{s-1} P_n. \quad (3.13)$$

Quand on trace $\mathbb{E}[Z]$ en fonction de ρ , on a les limites suivantes :

$$\lim_{\rho \rightarrow 0} \mathbb{E}[Z] = 0 ; \quad \lim_{\rho \rightarrow \infty} \mathbb{E}[Z] = s.$$

Dans la figure 3.7, on représente $\mathbb{E}[Z]$ pour différentes valeurs de $\xi = 0, 0.5, 1$ et 2 .

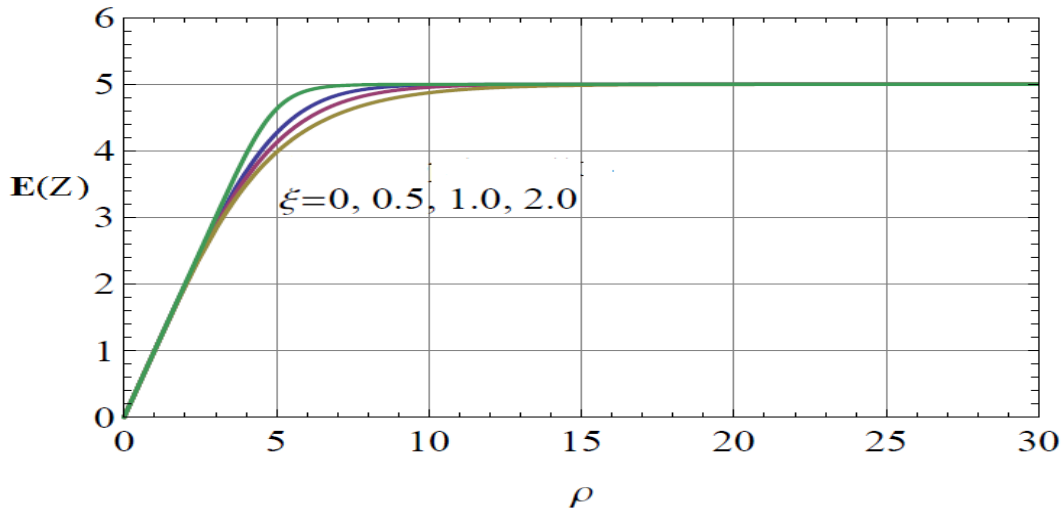


FIGURE 3.7 – Le nombre moyen de clients en service $\mathbb{E}[Z]$

Remarque 3.5. Le nombre moyen de clients en service ne semble pas beaucoup sensible à ξ .

3.2.5 Probabilité d'atteinte du service

Le débit du système noté λ' est défini comme étant le nombre moyen de clients servis par unité de temps.

D'autre part, la probabilité qu'un service soit réussi, noté $\mathbb{P}\{Sr\}$ est donnée par la fraction d'arriver des clients que le client atteint son service. Par conséquent, on aura la relation suivante :

$$\lambda' = \lambda \mathbb{P}\{Sr\} = \mu \mathbb{E}[Z].$$

Par conséquent, on obtient :

$$\mathbb{P}\{Sr\} = \frac{\mathbb{E}[Z]}{\rho} = \sum_{n=0}^{s-1} P_n + \frac{s}{\rho} \sum_{n=1}^N P_{s+n}. \quad (3.14)$$

La probabilité qu'un client arbitraire soit servi est donnée par :

$$\mathbb{P}\{Sr/Nb\} = \frac{\mathbb{P}\{Sr\}}{1 - \mathbb{P}\{\mathcal{B}\}} = \frac{\mu \mathbb{E}[Z]}{\gamma L_q + \mu \mathbb{E}[Z]} \quad (3.15)$$

Qui a pour limite :

$$\lim_{\rho \rightarrow \infty} \mathbb{P}\{Sr|Nb\} = \frac{\mu s}{\gamma N + \mu s} = \frac{1}{1 + N \frac{\xi}{s}}.$$

Dans la figure 3.8, on représente $\mathbb{P}\{Sr|Nb\}$ en fonction de ρ pour différentes valeurs de $\xi = 0.5, 1$ et 2 .

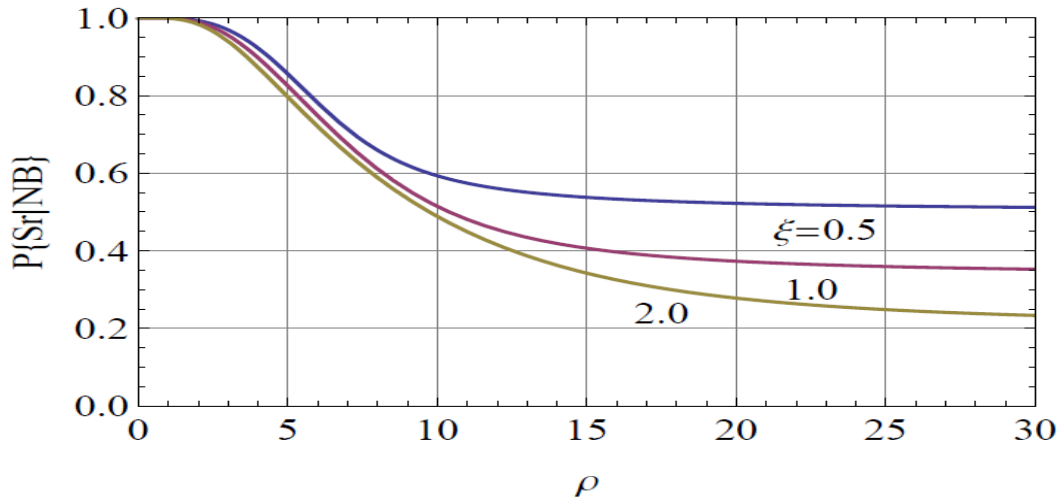


FIGURE 3.8 – La probabilité $\mathbb{P}\{Sr|Nb\}$

Remarque 3.6. notons que $\mathbb{P}\{Sr|Nb\}$ diminue à mesure que ξ augmente.

3.2.6 Le rapport de blocage, d'abandon et de clients servis

Un client qui arrive trouve trois cas devant lui, soit il sera bloqué à l'arrivée avec une probabilité $\mathbb{P}\{\mathcal{B}\}$, il entre mais abandonne avec probabilité $\mathbb{P}\{Ab\}$, où il sera servi avec succès avec la probabilité $\mathbb{P}\{Sr\}$:

$$\mathbb{P}\{\mathcal{B}\} + \mathbb{P}\{Ab\} + \mathbb{P}\{Sr\} = 1 \quad (3.16)$$

On multiplie les deux côtés par λ , on obtient :

$$\lambda\mathbb{P}\{\mathcal{B}\} + \gamma L_q + \mu\mathbb{E}[Z] = \lambda$$

Sur le côté gauche, $\lambda\mathbb{P}\{\mathcal{B}\}$ est le nombre moyen de clients qui sont bloqué à l'arrivée par unité de temps, γL_q le nombre moyen de clients qui quittent la file par unité de temps, et $\lambda' = \mu\mathbb{E}[Z]$ est le nombre moyen de clients qui sont servis avec succès par unité de temps.

En utilisant la condition de normalisation $\sum_{n=0}^{s+N} P_n = 1$, on obtient,

$$\begin{aligned}
\lambda \mathbb{P}\{\mathcal{B}\} + \mathbb{P}\{Ab\} &= 1 - \mathbb{P}\{Sr\} = 1 - \sum_{n=0}^{s-1} P_n - \frac{s}{\rho} \sum_{n=0}^N P_{s+n} \\
&= 1 - \sum_{n=0}^{s-1} P_n - \frac{s}{\rho} (1 - \sum_{n=0}^s P_n) \\
&= (1 - \frac{s}{\rho}) (1 - \sum_{n=0}^s P_n) + \frac{s}{\rho} P_s \\
&= (1 - \frac{s}{\rho}) (1 - \sum_{n=0}^s P_n) + P_{s-1}.
\end{aligned}$$

$$\lambda \mathbb{P}\{\mathcal{B}\} + \mathbb{P}\{Ab\} = 1 - \mathbb{P}\{Sr\} = (1 - \frac{s}{\rho}) (1 - \sum_{n=0}^s P_n) + P_{s-1}. \quad (3.17)$$

qui est la probabilité qu'un client arbitraire est perdu, soit en étant bloqué à l'arrivée ou par l'abandon dans la file d'attente.

3.2.7 Nombre de clients dans le système vu par une arrivée acceptée

Soit N Nombre de clients dans le système vu par une arrivée acceptée. La distribution de \hat{N} est donnée par :

$$\hat{P}_n = P_{\hat{N} = n} = \frac{P_n}{1 - P\{\mathbb{B}\}}, \quad 0 \leq n \leq s + N - 1. \quad (3.18)$$

Ce résultat peut s'écrire comme suit :

$$\hat{P}_n = \hat{P}_0 \frac{\rho^n}{n!}, \quad 1 \leq n \leq s. \quad (3.19)$$

$$\hat{P}_{s+n} = \hat{P}_0 \frac{\rho^s}{s!} \frac{\binom{\rho}{s}^n}{\prod_{j=1}^n (1 + j \frac{\rho}{s})}, \quad 1 \leq n \leq N - 1. \quad (3.20)$$

Avec,

$$\frac{1}{\hat{P}_0} = \frac{1 - P\{\mathbb{B}\}}{P_0} = \sum_{n=0}^{s-1} \frac{\rho^n}{n!} + \frac{\rho^s}{s!} \sum_{n=0}^{N-1} \frac{\binom{\rho}{s}^n}{\prod_{j=1}^n (1 + j \frac{\rho}{s})}. \quad (3.21)$$

3.2.8 Probabilité d'attente

On note W' le temps d'attente des clients qui sont acceptés soit servi ou abandonné dans la file d'attente.

Plus tard, nous considérons séparément les temps d'attente pour les clients servis, les clients abandonnés et les clients restés dans la file d'attente.

Un client arrivant qui est accepté doit attendre si tous les serveurs sont occupés à savoir s'il y a s clients ou plus dans le système immédiatement avant l'arrivée.

Par conséquent, la probabilité d'attente d'un client arbitraire accepté est donné par :

$$\mathbb{P}\{W' > 0\} = \sum_{n=0}^{s-1} \hat{P}_{s+n} = \hat{P}_0 \frac{\rho^s}{s!} \sum_{n=0}^{N-1} \frac{\binom{\rho}{s}^n}{\prod_{j=0}^n (1 + j \frac{\xi}{s})}. \quad (3.22)$$

Dans la figure 3.9, on représente $\mathbb{P}\{W' > 0\}$ en fonction de ρ pour différentes valeurs de $\xi = 0, 0.5, 1$ et 2 .

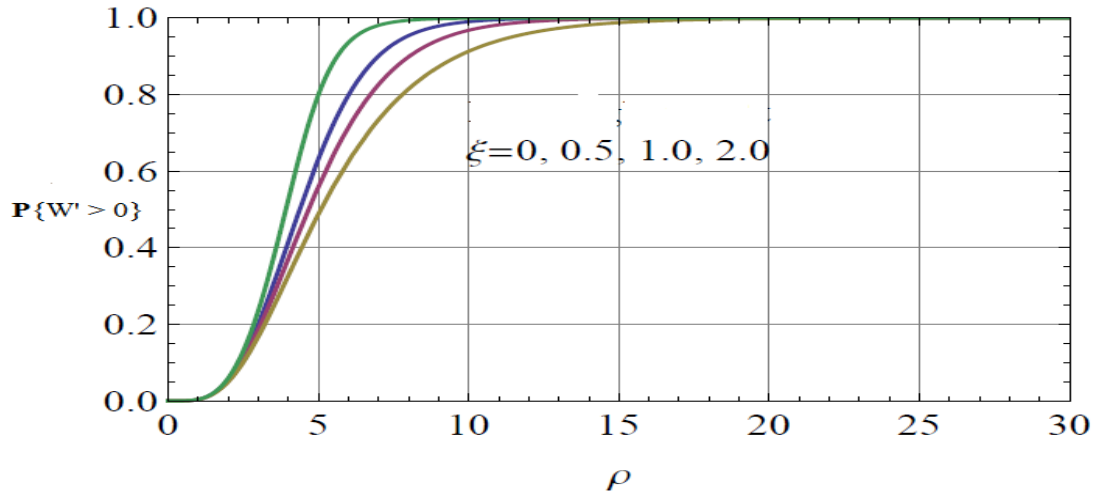


FIGURE 3.9 – La probabilité d'attente $\mathbb{P}\{W' > 0\}$

Remarque 3.7. Nous constatons que $\mathbb{P}\{W' > 0\}$ diminue à mesure que ξ augmente.

La probabilité d'aucune attente est donnée par :

$$\mathbb{P}\{W' = 0\} = 1 - \mathbb{P}\{W' > 0\} \quad (3.23)$$

$$= \sum_{n=0}^{s-1} \hat{P}_n = \hat{P}_0 \sum_{n=0}^{s-1} \frac{\rho^n}{n!}. \quad (3.24)$$

3.2.9 La durée moyenne d'attente

Il résulte des formules de Little appliquées aux clients admis que,

$$W = \frac{L_q}{\lambda(1 - \mathbb{P}\{\mathcal{B}\})} = \frac{\hat{P}_0 \rho^s}{\lambda s!} \sum_{n=1}^N \frac{n \binom{\rho}{s}^n}{\prod_{j=0}^n (1 + j \frac{\xi}{s})}. \quad (3.25)$$

Par conséquent, nous avons la relation suivante :

$$\mathbb{P}\{Ab|Nb\} = \gamma W.$$

$$\lim_{\rho \rightarrow \infty} W = \frac{N}{\gamma N + s\mu}.$$

Dans la figure 3.10, on représente W en fonction de ρ avec $\mu = 1$ pour différentes valeurs de $\xi = 0, 0.5, 1$ et 2 .

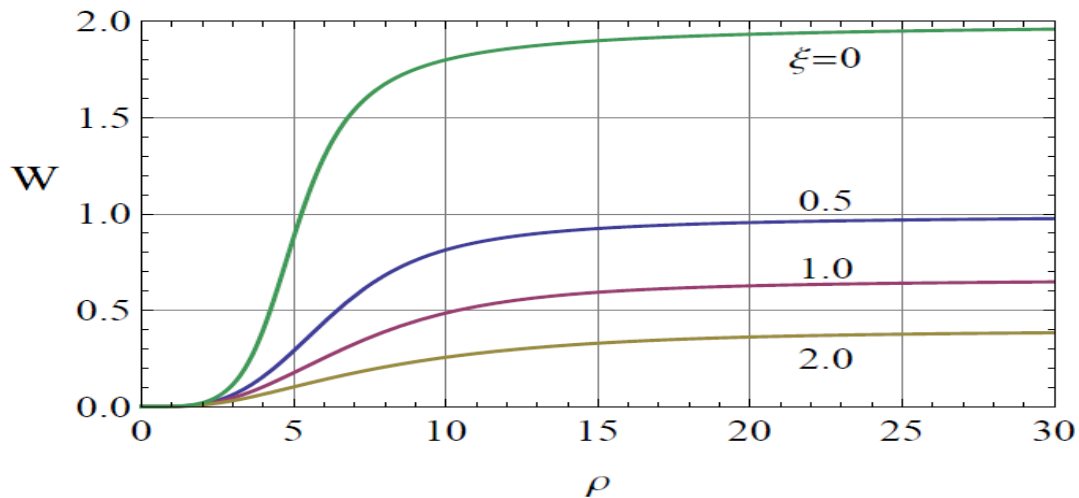


FIGURE 3.10 – Le temps d'attente moyen W

Remarque 3.8. On constate que, W diminue lorsque ξ augmente parce qu'il y a moins de clients dans ce système.

Conclusion

Nous avons modélisé un système de file d'attente avec impatience et avec une seule classe de clients, où on a calculé plusieurs mesures de performances, à savoir la probabilité de perte, le temps d'attente moyen et le nombre moyen de clients dans le système.

Conclusion générale

Dans ce travail, nous avons étudié deux modèles qui sont largement présent dans la vie courante. Dans le cas d'une seule classe de clients, nous avons redémontré un encadrement de la probabilité de perte, qui a la particularité d'être général, c'est à dire indépendant de la distribution qui caractérise le temps des arrivés, chose qui peut aider les administrateurs des systèmes d'information à choisir le bon type qui correspond au besoin du client. D'autre part, dans le cas de deux classes de clients avec priorité stricte, c'est à dire la première classe est plus prioritaire sur la deuxième, nous nous sommes intéressés au cas particulier où les temps d'abandon et de service sont les mêmes pour les clients des deux classes, et qui nous permet de profiter des résultats trouvés dans le cas d'une seule classe pour trouver ceux de la deuxième classe, chose qui se réitère par récurrence afin de généraliser les résultats à n classes de clients. Enfin, nous signalons que le cas sans préemption reste difficile à traiter car il demande une étude très approfondie afin d'explorer toutes ses possibilités. Nous avons les perspectives suivantes :

- Etude des S.F.A avec impatience et deux classes de clients dans le cas sans préemption.
- Modéliser des problèmes réels, (cas des centres d'appels), en utilisant les modèles de F.A avec impatience.

Rappels

loi de Poisson et loi exponentielle

Ces résultats ont été établis dans le cours de statistique en deuxième année.

Définition 3.1. Une variable aléatoire X à valeurs entières suit une loi de Poisson de paramètre.

$\lambda > 0$, si $\forall k \in \mathbb{N}$,

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda k} \quad (3.26)$$

Proposition 3.1. Une variable aléatoire qui suit une loi de Poisson de paramètre λ a une espérance et une variance égales à λ .

Définition 3.2. Une variable aléatoire Y à valeurs réelles strictement positives suit une loi exponentielle de paramètre $\mu > 0$, si $\forall t > 0$, la densité f est donnée par :

$$f_Y(t) = \mu e^{-\mu t} \quad (3.27)$$

Proposition 3.2. Une variable aléatoire Y qui suit une loi exponentielle de paramètre μ a une espérance égale à $\frac{1}{\mu}$ et une variance égale à $\frac{1}{\mu^2}$. Sa fonction de répartition est donnée par :

$$\forall t > 0, \quad \mathbb{P}(Y < t) = 1 - e^{-\mu t}. \quad (3.28)$$

Proposition 3.3. Une variable aléatoire Y suit une loi exponentielle si et seulement si :

$$\forall t, s > 0, \mathbb{P}(Y > s + t | Y > t) = \mathbb{P}(Y > s) \quad (3.29)$$

On dit qu'une variable exponentielle est "sans mémoire" si un événement ne s'est pas produit entre 0 et t , la probabilité qu'il se produise entre t et $t + s$ est la même que la probabilité pour qu'il se produise entre 0 et s l'instant de départ n'a pas d'influence et la variable ne "garde pas de mémoire" du passé.

Bibliographie

- [1] F. Baccelli, P. Boyer, and G. Hebuterne. *single-server queues with impatient customers*. Advances In Applied Probability, 16(4) :887905, 1984.
- [2] D.Y. Barrer. *Queuing with impatient customers and ordered service*, Oper. Res. 5, 650-656.. 1957.
- [3] S.A. Belbas and Y. Bulka. *Numerical solution of multiple nonlinear volterra integral equations*. Applied Mathematics and Computation, 217(9) :4791 4804, 2011.
- [4] O. Boxma, D. Perry, and W. Stadjc. *The M/G/1+G queue revisited*. QUEUEING SYSTEMS, 67(3) :207220, MAR 2011.
- [5] H. Brill. *System point theory in exponential queues*. Dissertation Abstracts International Part B : Science and Engineering [DISS. ABST. INT.PT. B - SCI. ENG.], vol. 49, no. 11 :242, 1989.
- [6] P. H. Brill. *An embedded level crossing technique for dams and queues*. Journal of Applied Probability, Vol. 16, No. 1 :174186, Mar., 1979.
- [7] BD. Choi, B. Kim, and J. Chung. *M/M/1 queue with impatient customers of higher priority*. Queuing Systems, 38(1) :4966, 2001.
- [8] J. W. Cohen. *Single server queue with uniformly bounded virtual waiting time*. J. Appl. Probability, 5 :93-122, 1968.
- [9] D. J. Daley. *General customer impatience in the queue gi/g/1*. Journal of Applied Probability, Vol. 2, No. 1 :186205, Jun, 1965.
- [10] A. K. Erlang. *Solutions of some problems in the theory of probabilities of significance in automatic telephone exchange*. Traduction Anglaise : P. O. Elec. Eng. J., 10 :189-197, 1917.
- [11] M. Fujiki and E. Gambe, *Teletraffic Theory*, Maruzen, Tokyo, 1980 (in Japanese).
- [12] A. G. De Kok and H. C. Tijms. *A queueing system with imaptient 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] P.D. Finch. *Deterministic customer impatience in the queueing system GI/M/1*. Biometrika, 47 (1,2) :45-52, 1960.

- [15] N. Gans, G. Koole and A. Mandelbaum, *Telephone call centers : tutorial, review, and research prospects*, Manufacturing Service Operations Management, 5, No.2, 79-141. DOI : <http://dx.doi.org/10.1287/msom.5.2.79.16071>, 2003.
- [16] O. Garnett, A. Mandelbaum, and M. Reiman, *Designing a call center with impatient customers*, Manufacturing Service Operations Management, 4, No.3, 208-227. DOI : <http://dx.doi.org/10.1287/msom.4.3.208.7753>, 2002.
- [17] Jr D. P. Gaver. *A waiting line with interrupted service, including priorities*. Journal of the Royal Statistical Society. Series B (Methodological), Vol. 24, No. 1 :7390, 1962.
- [18] A. Ghosal. *Queues with finite waiting time*. *Operations Res.*, 11 :919-921, 1963.
- [19] B. V. Gnedenko and I. N. Kovalenko. *Introduction to queueing theory*. Translated from Russian by R. Kondor. Translation edited by D. Louvish. Israel Program for Scientific Translations, Jerusalem, 1968.
- [20] A. Economou, A. Gómez-Corral, and S. Kanta. *Optimal balking strategies in single-server queues with general service and vacation times*. PERFORMANCE EVALUATION, 68(10) :967982, OCT 2011.
- [21] D. Gross and C.M. Harris. *Fundamentals of Queueing Theory*. Library of Congress Cataloging-in-Publication Data, 1998.
- [22] F. Iravani and B. Balcioglu. *On priority queues with impatient customers*. Queuing systems, 58(4) :239260, APR 2008.
- [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] P. Khudyakov, P. D. Feign, and A. Mandelbaum, *Designing a call center with an IVR (interactive voice response)*, Queueing Systems, 66, No.3, 225-237. 2010 DOI : <http://dx.doi.org/10.1007/s1134-010-9193-y>
- [25] G. Koole and A. Mandelbaum, *Queueing models of call centers : an introduction*, *Annals of Operations Research*, 113, Nos.1-4, 41-59. DOI : <http://dx.doi.org/10.1023/A:1020949626017>, 2002.
- [26] L. Kleinrock *Queueing systems. Volume I*. John Ziley & Sonns, 1975.
- [27] A. Mandelbaum and S. Zeltyn, *Service engineering in action : The Palm/Erlang-A queue, with applications to call centers*. In D. Spath and K.-P. Fährnich (eds.) : Advances in Services Innovations, Springer, Berlin, 17-45. 2007
- [28] P. Moyal. *Contributions à l'étude des files d'attente avec clients impatientes*. *Mathematics [math]. Télécom ParisTech*, 2005.
- [29] S. Rao. *Queueing with balking and reneging in M—G—1 systems*. *Metrika*, 12 :173188,10.1007/BF02613493. 1967.
- [30] J. Riordan, *Stochastic Service Systems*, John Wiley and Sons, 1962.
- [31] A. Ruegg. *Processus stochastiques*. Press Polytechnique Romande. 1989.
- [32] R. E. Stanford. *Reneging phenomena in single channel queues*. *Mathematics of Operations Research*, Vol. 4, No. 2 :162178, May, 1979.

- [33] L. Takàcs. *A single-server queue with limited virtual waiting time*. J. Appl. Probability, 11 :612-617, 1974.
- [34] H. Takagi *International Journal of Pure and Applied Mathematics*. Vol : 90 No. 4, 519-559.2014
- [35] Takahashi, Y., H. Yamamoto, H. Yoshino, and A. Toda, *Queueing System Primer : Theory and Practice*, The Institute of Electronics, Information and Communication Engineers, Tokyo, 2003 (in Japanese).
- [36] W. Whitt, *Engineering solution of a basic call-center model*, Management Science, 51, No.2, 221-235. DOI : <http://dx.doi.org/10.1287/mnsc.1040.0302>, 2005
- [37] S. Zeltyn, *Call centers with impatient customers : Exact analysis and manyserver asymptotics of the M/M/n+G queue*. Research thesis, Technion - Israel Institute of Technology, Haifa, Israel, October 2004.