

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Abderrahmane Mira de Béjaïa
Faculté des Sciences Exactes
Département de Mathématiques



Mémoire de fin de cycle
En vue de l'obtention du diplôme de Master Académique en
Mathématiques

Spécialité : Probabilités Statistique et Applications

Thème

Etude comparative des systèmes
d'attente classiques et avec rappels

Présenté par :

M^{lle} OUALI Ilhem

Devant le Jury

Président : M^r OUAZINE Sofiane M.C.B Université de Béjaïa.

Promotrice : M^{me} BOURAINE Louiza Professeur Université de Béjaïa.

Examineur : M^r ZIOUI Arezki M.C.B Université de Béjaïa.

Septembre 2021

Remerciements

Tous les éloges et les remerciements sont à **ALLAH** de m'avoir donné la force et la puissance de pouvoir mener ce travail à terme.

Je remercie vivement mon encadreur Madame L. BOURAINE, pour m'avoir proposé le sujet de ce mémoire, pour sa patience et ses conseils qui m'ont été d'un grand apport, et encouragements pendant toute la durée de l'élaboration de ce travail.

Je tiens également à remercier :

Le président de jury Monsieur S. OUAZINE, d'avoir eu l'amabilité d'accepter la présidence du jury de soutenance.

Mes remerciements s'adressent également à Monsieur A. ZIOUI, de m'avoir honoré, en acceptant d'examiner ce travail.

Je remercie tous ceux qui ont participé de près ou de loin à l'élaboration de ce travail.

Je tiens également à remercier ma famille, tous mes enseignants, et tous mes amis(es).

Dédicaces

Je dédie ce modeste travail à tous ceux et celles qui m'ont aidé de proche ou loin, notamment :

À mes parents qui ne m'ont jamais laissé tomber dans toutes les circonstances. Que Dieu vous protège et que la réussite soit toujours à ma portée pour que je puisse vous combler ce bonheur.

À mon très cher fiancé Fouad pour tout son aide, son soutien et ses encouragements.

À Mon très cher frère Ali. Je lui souhaite un avenir radieux plein de réussite.

À mes très chères sœurs. Amani, Nawel et Souria, à qui je souhaite une longue vie pleine de bonheur et réussite.

À Toute ma famille et tous mes amis.

TABLE DES MATIÈRES

Liste des figures	6
Introduction générale	7
1 Systèmes d'attente classiques markovien et non markoviens	11
1.1 Classification des systèmes file d'attente	11
1.2 Notions élémentaires sur les files d'attente	12
1.2.1 Une file simple	12
1.2.2 Notation de Kendall	12
1.3 Distributions usuelles des inter-arrivées et de service	13
1.4 Analyse mathématique d'un système de file d'attente	15
1.5 Caractéristiques d'un système file d'attente	16
1.6 Modèles markoviens	16
1.6.1 Système $M/M/1$	17
1.6.2 Système $M/M/c$	19
1.7 Modèles non markoviens	22
1.7.1 Chaîne de Markov induite	22
1.7.2 Mesures de performance	24
2 Systèmes d'attente avec rappels de types $M/M/c$ et $M/G/1$	26
2.1 Modèle général d'un système de file d'attente avec rappels	28
2.2 Modèle d'attente $M/M/c$ avec rappels	29

2.2.1	Description mathématique du modèle	29
2.2.2	Distribution stationnaire	29
2.2.3	Mesures de performance	33
2.3	Le modèle d'attente $M/G/1$ avec rappels	34
2.3.1	Description mathématique du modèle	34
2.3.2	Chaîne de Markov induite du système	35
2.3.3	Mesures de performance	37
3	Analyse comparative des modèles d'attente classiques et des modèles d'attente avec rappels	39
3.1	Le modèle $M/M/c$	39
3.2	Le modèle $M/G/1$	45
3.3	Conclusion	54
	Conclusion générale	55
	Bibliographie	56

TABLE DES FIGURES

1.1	Représentation d'une file d'attente simple	12
1.2	Graphe de transitions du modèle $M/M/1$	17
1.3	Graphe de transitions du modèle $M/M/c$	19
1.4	Graphe de transitions du modèle $M/G/1$	24
2.1	Schéma générale d'un système avec rappels	29
2.2	Graphe de transition du modèle $M/M/c$ avec rappels	32
2.3	Graphe de transition du modèle $M/G/1$ avec rappels	35
3.1	Nombre moyen de clients dans le système (dans l'orbite)	45
3.2	Nombre moyen de clients dans le système (dans l'orbite)	51
3.3	Temps moyen d'attente	51
3.4	Nombre moyen de clients dans le système (dans l'orbite)	52
3.5	Temps moyen d'attente	53

INTRODUCTION GÉNÉRALE

La théorie des files d'attente, est l'un des outils analytiques les plus puissants pour la modélisation des systèmes dynamiques. Cette théorie tire son origine des recherches de l'ingénieur Danois Agner Krarup Erlang entre (1909, 1920), elle a été inspirée au Danemark avec le développement de la téléphonie. La compagnie de Copenhague souhaitait à l'époque mettre en place une plateforme permettant aux utilisateurs d'être mis en relation par l'intermédiaire d'opérateurs, mais ne savait pas quelle taille devait avoir une telle structure, ni combien d'appels elle aurait à gérer. Si le centre était trop gros, l'entreprise risquait la banqueroute. Si elle voyait trop petit les utilisateurs, faute d'être connectés auraient manifesté leurs mécontentement. La compagnie a donc demandé à l'un de ses meilleurs ingénieurs, A. K. Erlang, de travailler à une conceptualisation. Le sujet a inspiré et continue à inspirer de nombreux chercheurs comme en témoignent les nombreuses publications parues à ce jour dans le domaine [29, 21, 39]. C'est grâce aux apports des mathématiciens Kendall, Pollaczek que la théorie s'est vraiment développée.

Dés la fin des années 1940, des chercheurs ont mis en évidence les limites de la théorie classique des files d'attente qui ne permettait pas d'expliquer le comportement stochastique des systèmes réels de plus en plus complexes tels que les systèmes téléphoniques où les abonnés répétaient leurs appels en recomposant le numéro plusieurs fois jusqu'à l'obtention de la communication.

Ce phénomène de répétition de demandes du service a poussé certains chercheurs à étendre le modèle d'attente classique à celui dit avec rappels. Cette branche de la théorie

des files d'attente s'applique dans la modélisation stochastique de différents problèmes de télécommunication, et de réseaux informatique. Les systèmes de files d'attente avec rappels peuvent être utilisés pour résoudre des problèmes pratiques, tels que l'analyse du comportement des données dans les réseaux téléphoniques, l'évitement de collision dans les réseaux locaux, l'analyse du temps d'attente pour accéder à la mémoire sur les disques magnétiques.

L'étude des modèles de files d'attentes avec rappels remonte déjà aux travaux de Clos (1948)[23] et Wilkinson (1956)[37]. La première contribution sérieuse sur ce sujet a été publiée en 1957 dans "Philips Telecommunication Review" par Cohen [24]. Cependant, l'influence de ce phénomène a été longtemps négligée durant les décennies suivantes. Ce n'est que vers les années 1970-1980 qu'on a vu un net regain d'intérêt pour cette catégorie de modèles, avec l'avènement de nouvelles technologies, notamment dans les systèmes de télécommunication. Les progrès récents dans ce domaine sont résumés dans les articles de synthèse de Yang et Templeton (1987)[38], de Falin (1990)[26], de Aissani (1994)[1], de J. Kim et B. Kim (2016)[30] et de Phung-Duc (2019) qui a présenté une synthèse compréhensive sur la théorie et les applications des systèmes d'attente avec rappels [32], ainsi que dans les monographies de Falin et Templeton (1997) [28], Artalejo et Gómez (2008)[12]. Des classifications bibliographiques sont données dans les articles de Artalejo (1999) et (2010) [6, 7] ainsi que Shekhar et al.(2016)[34]. L'importance et l'actualité de ce domaine est également confirmée par l'organisation périodique d'une conférence internationale sur les systèmes d'attente avec rappels (International Workshop on Retrial Queues) : Madrid (Spain) (1998), Minsk (Belarus) (1999), Amsterdam (Netherlands) (2000), Cochin (India) (2002), Seoul (Korea) (2004), Miraflore de la Sierra (Spain) (2006), Athens (Greece) (2008), Beijing (China) (2010), Seville (Spain) (2012), Tokyo (Japan) (2014), Amsterdam (2016), Toursh (Russia) (2018).

L'analyse stochastique des systèmes markoviens avec rappels et plusieurs serveurs a fait l'objet de plusieurs publications (voir Artalejo 2008 et les références citées dans ce livre [12]). Artalejo (1996) [8], a donné l'analyse stationnaire des caractéristiques du système $M/M/2$ avec rappels constants.

La théorie analytiques des modèles $M/M/c$ avec rappels où $c > 2$, s'avère d'une portée limitée en raison de la complexité des résultats connus. En effet, dans la majorité des cas, on se

retrouve confronté à des équations complexes, sous forme de fonctions hypergéométrique. Par ailleurs, on peut citer le degré de difficulté pour l'obtention de certaines caractéristiques dans les modèles de type $M/M/c$ avec rappels. Pour pallier à toutes ces difficultés, les chercheurs ont recouru aux méthodes d'approximation en utilisant les modèles tranqués et les modèles tranqués généralisées (voir Anisimov et Artalejo (2002)[3]), Artalejo et Pozo (2002) [18] ont utilisé un algorithme numérique pour le calcul de la distribution stationnaire de $M/M/c$.

Le modèle $M/G/1$ avec rappels est le plus étudié par les spécialistes : Falin, Artalejo, Gomez, Kullcarni, Aissani, . . . et il existe une littérature abondante sur ses diverses propriétés. néanmoins, il existe aussi des difficultés pour l'obtention de certaines caractéristiques. Ces difficultés résident essentiellement dans l'utilisation des inverses des transformées de Laplace et des distributions marginales.

Plusieurs similitudes et différences existent entre les caractéristiques des modèles classiques et les modèles avec rappels [9]. Les modèles avec rappels peuvent être approximés par des modèles classiques.

L'objectif de notre travail concerne l'analyse comparative des systèmes de files d'attente classiques et des systèmes d'attente avec rappels à savoir les systèmes $M/M/c$ et $M/G/1$ avec rappels, et à partir de leurs caractéristiques principales : probabilité de blocage, moments de la période d'activité, le temps moyen d'attente et le nombre moyen de clients en orbite.

Ce mémoire est constitué de trois chapitres, d'une conclusion générale et une bibliographie.

- ✓ Le premier chapitre présente brièvement le formalisme des systèmes de file d'attente classiques. Nous nous focalisons surtout sur les systèmes $M/M/1$, $M/M/c$ et $M/G/1$ classiques.
- ✓ Le deuxième chapitre est consacré à l'analyse stochastique des systèmes $M/M/c$ et $M/G/1$ avec rappels.
- ✓ Le troisième chapitre concerne l'étude comparative entre modèles avec rappels et les modèles classiques.

✓ Le travail s'achève par une conclusion et quelques perspectives de recherche.

CHAPITRE 1

SYSTÈMES D'ATTENTE CLASSIQUES MARKOVIEN ET NON MARKOVIENS

On parle de phénomène d'attente chaque fois que certaines unités appelés "clients" se présentent d'une manière aléatoire à des stations à fin de recevoir un service dont la durée est généralement aléatoire.

Si le serveur est libre, le client est servi immédiatement sinon il prend sa place dans une file d'attente selon l'ordre d'arrivée.

Un système d'attente comprend donc : "un espace d'attente" dans le quel se former une éventuelle file d'attente. L'objectif de ce chapitre est l'analyse stochastique des modèles de files d'attente de base markoviens et non markoviens.

1.1 Classification des systèmes file d'attente

On identifie un système file d'attente par :

- La nature stochastique du processus des inter-arrivées définie par la distribution des intervalles séparant deux arrivées consécutives.
- La distribution du temps de service.
- Le nombre de serveurs " c ".
- La capacité N du système, si $N < +\infty$, la file d'attente ne peut dépasser une longueur de ' $N - c$ ' unités.

- La discipline de service.

1.2 Notions élémentaires sur les files d'attente

1.2.1 Une file simple

Une file d'attente simple est un système où des clients se présentent à un dispositif de service, appelé serveur. Puisqu'un client occupe le serveur pendant un certain temps, les autres clients doivent attendre avant d'être servis, formant ainsi une file d'attente. Afin de spécifier complètement une file d'attente simple, on doit caractériser le processus des arrivées des clients, le temps de service ainsi que la structure et la discipline de service.

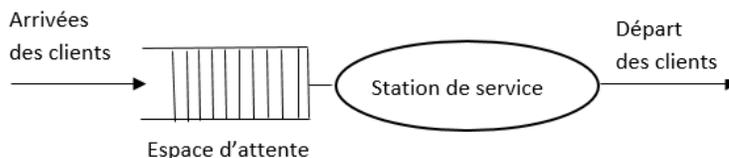


FIGURE 1.1 – Représentation d'une file d'attente simple

1.2.2 Notation de Kendall

Un système de file d'attente est généralement représenté par la notation de Kendall définie comme suit : $A/B/c(N/K/D_s)$ où

A : distribution des temps inter-arrivées : L'arrivée des clients est décrite par un processus stochastique de comptage $\{N(t), t \geq 0\}$.

Si A_n : est la variable aléatoire mesurant l'instant d'arrivée du $n^{\text{ème}}$ client dans le système, on aura ainsi : $A_0 = 0$ et $A_n = \inf\{t/N(t) = n\}$.

Si T_n : désigne la variable aléatoire mesurant le temps séparant deux arrivées consécutives, alors $T_n = A_n - A_{n-1}$.

B : distribution des durées de service.

c : nombre de serveurs en parallèle.

N : capacité du système (file+service).

K : taille de la population source.

D_s : discipline de service.

Les principales disciplines de service sont les suivantes :

- * FIFO (First-In-First-Out) Cette discipline est la plus utilisée. Le premier client arrivé est le premier servi.
- * LIFO (Last-In-First-Out) Le dernier client dans la file est le premier servi.
- * SIRO (Served In Random Order) Tout les clients ont la même probabilité d'être servis en premier.
- * RR (Round Robin) Les clients sont servis à tour de rôle pendant un intervalle de temps fixe, appelé **quantum (Q)** et sont replacés dans la file, jusqu'à ce que leur service soit totalement accompli.
- * PS (Processor Sharing) Les clients sont servis au même temps, mais avec une vitesse inversement proportionnelle au nombre de clients présents.

L'orsque les trois derniers éléments ne sont pas mentionnés, il est sous-entendu que la capacité du système et la population source sont infinis et que la discipline de service est FIFO.

1.3 Distributions usuelles des inter-arrivées et de service

1. Distribution Exponentielle (M)

La distribution exponentielle de paramètre λ sur $[0; +\infty[$ modélise la durée de vie d'un phénomène sans vieillissement a pour densité la fonction f définie sur $[0; +\infty[$ par :

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0, \quad \lambda > 0,$$

et sa fonction de répartition

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0 & \text{sinon.} \end{cases}$$

D'espérance :

$$\mathbb{E}(X) = \frac{1}{\lambda}.$$

La loi exponentielle est caractérisée par sa propriété d'absence de mémoire (Memoryless).

2. Distribution Hyper-Exponentielle (H)

La distribution hyper-exponentielle est une loi de probabilité mélangeant plusieurs lois exponentielle. Elle dépend de trois paramètres : n le nombre de lois exponentielles indépendantes, $(\lambda(i), 1 \leq i \leq n)$ les paramètres de ces lois exponentielles et $(p_i, 1 \leq i \leq n)$ une pondération de ces lois vérifiant $\sum_{i=1}^n p_i = 1$. Sa densité de probabilité est donnée par :

$$f(x) = \sum_{i=1}^n p_i \lambda_i e^{-\lambda_i x}, \quad x \geq 0.$$

D'espérance :

$$\mathbb{E}(X) = \sum_{i=1}^n \frac{p_i}{\lambda_i}.$$

Sa fonction de répartition est donnée par :

$$F(x) = 1 - \sum_{i=1}^n p_i e^{-\lambda_i x}, \quad x \geq 0.$$

Sa transformée de Laplace est :

$$\beta^*(s) = \sum_{i=1}^n \frac{p_i \lambda_i}{s + \lambda_i}.$$

3. Distribution Gamma ($\gamma(\alpha, \beta)$)

Une variable aléatoire X est de loi gamma, si sa densité de probabilité est donnée par :

$$f(x, \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta x} x^{\alpha-1}, \quad x \geq 0.$$

Γ est la fonction gamma définie sur R_+^* par :

$$\Gamma(x) = \int_0^{+\infty} e^{-t} t^{x-1} dt, \quad \text{et } \Gamma(\alpha) = (\alpha - 1)! \text{ pour } \alpha \in N^*.$$

D'espérance :

$$\mathbb{E}(X) = \frac{\alpha}{\beta}.$$

De transformé de Laplace :

$$\beta^*(s) = \mathbb{E}(e^{-sX}) = \left(\frac{s}{s + \beta}\right).$$

4. Distribution Déterministe (D)

Le temps inter-arrivées des clients ou le temps de service sont des constants et toujours les même avec une moyenne m . Sa transformé de Laplace est donnée par :

$$\beta^*(s) = \exp(-sm).$$

1.4 Analyse mathématique d'un système de file d'attente

L'étude mathématique d'un système de file d'attente se fait généralement par l'introduction d'un processus stochastique défini de façon appropriée. On s'intéresse principalement au nombre de clients $X(t)$ se trouvant dans le système à l'instant t ($t \geq 0$).

En fonction des quantités qui définissent le système, on cherche à déterminer :

Régime transitoire

Les probabilités d'états $P_n(t) = \mathbb{P}(X(t) = n)$, qui définissent le régime transitoire du processus $\{X(t), t \geq 0\}$. Il est évident que les fonctions $P_n(t)$ dépendent de l'état initial ou de la distribution initiale du processus.

Régime stationnaire

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

$(P_n)_{n \geq 0}$: Distribution stationnaire du processus $\{X(t), t \geq 0\}$.

1.5 Caractéristiques d'un système file d'attente

A partir de la distribution stationnaire du processus $\{X(t), t \geq 0\}$, on peut calculer les caractéristiques suivantes :

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

Ces caractéristiques ne sont pas indépendantes les unes des autres, mais sont liées par les relations suivantes qui sont vérifiées sous la condition de stabilité du système ($\frac{\lambda}{\mu} < 1$) :

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

Les deux premières formules sont appelées **formules de Little**, où $\lambda_e < \lambda$ est le taux d'entrée dans le système.

1.6 Modèles markoviens

Les modèles markovien caractérisent les systèmes dans lesquels les deux quantités stochastiques principales, qui sont le temps des inter-arrivées et la durée de service, sont des variables aléatoires indépendantes et exponentiellement distribuées. L'étude de tels systèmes se fait par l'introduction d'un processus stochastique approprié. Ce processus est souvent le processus $\{X(t), t \geq 0\}$ défini comme étant le nombre de clients dans le système à l'instant t . Dans cette section on s'intéresse spécialement au modèle markovien à c serveurs.

1.6.1 Système $M/M/1$

La file d'attente $M/M/1$ est un système à serveur unique et capacité infinie, la discipline de service est FIFO. Les instants d'arrivées des clients sont distribués selon un processus de poisson de taux λ , et les temps de service sont indépendants suivant la loi exponentielle de taux μ . Le processus $(X(t))_{t \geq 0}$ décrivant le nombre de clients à l'instant t est un processus de naissance et de mort de taux de transition :

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

Graphe de transition

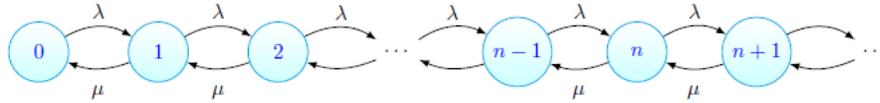


FIGURE 1.2 – Graphe de transitions du modèle $M/M/1$

Régime stationnaire du système

On dit que le système est stable si et seulement si $\rho = \frac{\lambda}{\mu} < 1$.

Les équations de Balance sont données par :

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

D'où

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

Comme

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

Mesures de performance

1. Le nombre moyen de clients dans le système :

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

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

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

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

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

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

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

1.6.2 Système $M/M/c$

Le modèle d'attente $M/M/c$ est un système formé d'une file à capacité infinie, avec c serveurs identiques et la discipline de service est FIFO. Supposons que le flot des arrivées est poissonien de paramètre λ , la durée de service est exponentielle de paramètre μ .

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

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

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

Graphe de transition

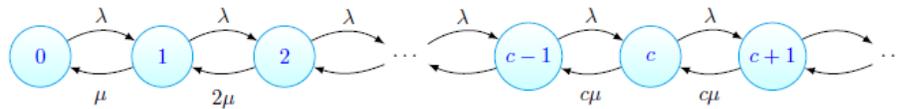


FIGURE 1.3 – Graphe de transitions du modèle $M/M/c$

Distribution stationnaire de $M/M/c$

Les équations de balance

1^{er} cas $n \leq c$

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

D'où

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

2^{ème} Cas $n > c$

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

D'où

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

Donc

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

Avec

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

Les caractéristiques du système $M/M/c$

1- Nombre moyen de clients dans la file :

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

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

$$L_q = p_c\rho \frac{1}{(1-\rho)^2}.$$

2- Nombre moyen de clients dans le système :

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

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

3- Temps moyen d'attente :

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

4- Temps moyen de séjour :

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

1.7 Modèles non markoviens

En l'absence de l'exponentialité ou lorsque l'on s'écarte de l'hypothèse d'exponentialité de l'une des deux quantités stochastiques : le temps des inter-arrivées et la durée de service, ou en prenant en compte certaines spécificités des problèmes par introduction de paramètres supplémentaires, on aboutit à un modèle non markovien. Le processus $\{X(t), t \geq 0\}$ représentant le nombre de clients dans le système à l'instant t , n'est plus markovien. Dans cette section, on s'intéresse spécialement au système $M/G/1$ où les clients arrivent selon un processus de poisson de taux $\lambda > 0$. De ce fait, le temps entre deux arrivées successives suit une loi exponentielle de moyenne $\frac{1}{\lambda}$. Le service est assuré par un seul serveur. A l'arrivée d'un client, si le serveur est libre, le client sera pris en charge immédiatement. Dans le cas contraire, il rejoint la file d'attente (de capacité illimitée et de discipline FIFO), les durées de service Y sont des variables aléatoires indépendantes et identiquement distribuées de loi générale dont la fonction de répartition F , d'espérance mathématique $\frac{1}{\mu}$.

Pour l'analyse de ce système on procède par la méthode de la chaîne de markov induite.

1.7.1 Chaîne de Markov induite

Nous introduisons le processus stochastique $\{X(t), t \geq 0\}$ qui n'est pas un processus de Markov. Pour le rendre markovien, nous utiliserons la méthode des chaîne de Markov induites.

Soit le processus à temps discret $\{X_n = X(t_n), n \in \mathbb{N}\}$ où t_n est l'instant où le $n^{\text{ème}}$ client a fini son service et quitte le système. Vérifions que cette suite de variables définit bien une chaîne de Markov.

Soit (A_n) des variables aléatoires indépendantes représentant le nombre de clients arrivants pendant le $n^{\text{ème}}$ service avec la distribution commune

$$P(A_n = k) = a_k = \int_0^\infty \frac{\exp(-\lambda t)(\lambda t)^k}{k!} F(t) dt,$$

avec $a_k > 0$ ($k = 0, 1, 2, \dots$),

alors

$$X_{n+1} = \begin{cases} X_n - 1 + A_{n+1} & \text{si } X_n > 0, \\ A_{n+1} & \text{si } X_n = 0. \end{cases}$$

Soit la variable aléatoire

$$\delta_n = \begin{cases} 1 & \text{si } X_n > 0, \\ 0 & \text{si } X_n = 0. \end{cases}$$

Alors l'équation fondamentale de la chaîne devient :

$$X_{n+1} = X_n - \delta_n + A_{n+1}.$$

Il est évident que X_{n+1} dépend de X_n et A_{n+1} seulement et non pas de X_{n-1}, X_{n-2}, \dots . D'où la suite $\{X_n, n \geq 1\}$ est une chaîne de Markov induite du processus $\{X(t), t \geq 0\}$ avec les probabilités de transitions $\mathbb{P}(X_{n+1} = j | X_n = i) = p_{ij}$ qui s'expriment de la manière suivante :

$$\begin{cases} p_{0j} = a_j & \text{si } j \geq 0, \\ p_{ij} = a_{j-i+1} & \text{si } 1 \leq i \leq j+1, \\ p_{ij} = 0 & \text{ailleurs.} \end{cases}$$

Par conséquent, la matrice de transition P est donnée par

$$P = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & \dots & \dots \\ a_0 & a_1 & a_2 & a_3 & \dots & \dots \\ 0 & a_0 & a_1 & a_2 & \dots & \dots \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & 0 & a_0 & a_1 & \dots \end{pmatrix}.$$

Puisque nous pouvons passer d'un état à n'importe quel autre, alors la chaîne de Markov est irréductible.

Et comme $P_{00} = a_0 > 0$ donc la chaîne est apériodique donc elle est ergodique.

Graphe de transitions : Ce graphe donne les transitions de sortie de l'état 0 et les transitions de sortie d'un état $i > 0$.

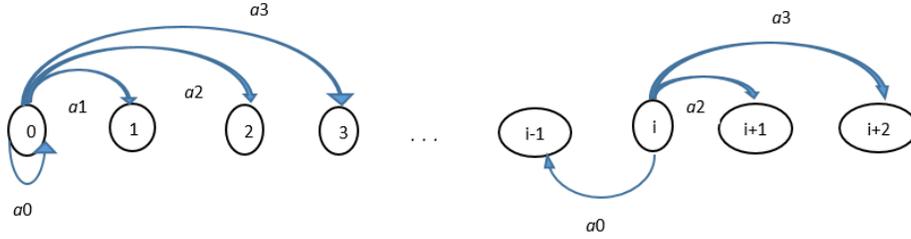


FIGURE 1.4 – Graphe de transitions du modèle $M/G/1$

La distribution stationnaire $\pi = (\pi_1, \dots, \pi_n, \dots)$ de la chaîne de Markov induite existe si $\rho = \frac{\lambda}{\mu} < 1$. Elle est donnée par sa fonction génératrice

$$\pi(z) = \sum_{n=0}^{\infty} z^n \pi_n = \frac{(1 - \rho)\beta^*(\lambda - \lambda z)(1 - z)}{\beta^*(\lambda - \lambda z) - z}.$$

Avec $\beta^*(z) = \mathbb{E}(e^{-zY})$ est la transformée de Laplace-Stieltjes de la durée de service.

1.7.2 Mesures de performance

- Nombre moyen de clients dans le système :

$$L = \rho + \frac{\rho^2 + \lambda^2 \text{Var}(Y)}{2(1 - \rho)}.$$

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

$$L_q = L - \rho \Rightarrow L_q = \frac{\rho^2 + \lambda^2 \text{Var}(Y)}{2(1 - \rho)}.$$

- Temps moyen de séjour d'un client dans le système :

$$W = \frac{L}{\lambda} = \frac{1}{\mu} + \frac{\lambda(\text{Var}(Y) + \frac{1}{\mu^2})}{2(1 - \rho)}.$$

-
- Temps moyen d'attente d'un client dans la file :

$$W_q = \frac{L_q}{\lambda} = \frac{\lambda(\text{Var}(Y) + \frac{1}{\mu^2})}{2(1 - \rho)}.$$

Conclusion

Dans ce chapitre, nous avons présenté la théorie des files d'attente classiques qui offre deux possibilités pour résoudre le conflit qui apparaît lorsqu'un client qui arrive et trouve l'espace de service est occupé : soit le client quitte le système définitivement (système d'Erlang), soit il rejoint la file d'attente (système classique de file d'attente), une possibilité alternative permet au client de renouveler sa demande après une durée aléatoire. Il s'agit des systèmes d'attente avec rappels (voir chapitre 2).

CHAPITRE 2

SYSTÈMES D'ATTENTE AVEC RAPPELS DE TYPES $M/M/C$ ET $M/G/1$

Les systèmes de files d'attente avec rappels (Retrial Queues) sont caractérisés par la propriété suivante : un client arrivant dans le système et qui trouve tous les serveurs occupés quitte le système définitivement ou entre en "orbite" pour rappeler ultérieurement à des instants aléatoires.

La description d'un système classique se fait avec ses éléments principaux : le processus des arrivées, processus de service et sa structure (nombre de serveurs, discipline d'attente). Pour un système avec rappels on doit ajouter un élément décrivant la loi des rappels.

Dans ce chapitre, on s'intéresse aux principaux modèles $M/M/c$ et $M/G/1$ avec rappels.

Exemple de systèmes de file d'attente avec rappels

Exemple 1 : (Systèmes téléphoniques)

Il est bien connu qu'un abonné au téléphone qui obtient un signal occupé répète habituellement l'appel jusqu'à ce que la connexion requise soit établie. En conséquence, le flux des appels circulant dans un réseau téléphonique se compose de deux parties : le flux des appels primaires, qui reflète les véritables souhaits des abonnés au téléphone, et le flux des répétitions des appels, ce qui est la conséquence du manque de succès des précédentes tentatives. Les modèles classiques des systèmes téléphoniques, des systèmes de file d'attente avec pertes, ne

tiennent pas compte de cette structure de flux réel des appels et ne peuvent donc pas être appliqués à la résolution d'un nombre de problèmes pratiquement importants.

Ces considérations mettent en évidence la nécessité des files d'attente avec rappels en tant que modélisation appropriée du comportement du client dans les systèmes téléphoniques classiques.

Cette classe de files d'attente est caractérisée par la caractéristique suivante : un client qui arrive et tous les serveurs sont occupés quitte la zone de service mais après un moment aléatoire répète sa demande. Notons que le but principal de tout centre d'appels est de fournir un service de qualité via un appareil, dans notre cas cet appareil étant un téléphone. De plus, lorsque la structure, la planification, la gestion et l'exécution d'un centre d'appels typique sont soigneusement étudiées, il est immédiatement clair comment la modélisation des systèmes téléphoniques en tant que files d'attente avec rappels clarifie le mieux le comportement des clients dans un tel scénario. Presque tous les principaux acteurs de l'industrie des télécommunications utilisent les centres d'appels comme leur principal moyen de communication et d'interaction avec leurs clients. Pour avoir une perspective, un centre d'appels peut être considéré comme une structure de file d'attente basée sur la file d'attente $M/M/c$ (l'un des modèles de file d'attente les plus utilisés) (voir Shekhar et al. (2016)[34]).

Exemple 2 : (Système mémoire sur les disques magnétiques)

Considérons un système mémoire où K unités de disques partagent un contrôleur des disques et transmettent l'information quand elles trouvent ce dernier libre. Les demandes insatisfaites sont répétées après une rotation du disque. Ce système peut être présenté comme un système de files d'attente avec rappels. Le serveur est le contrôleur des disques. L'arrivée dans le système est quasi-aléatoire : le nombre de sources est fini et égale à K . La rotation du disque dans le cas de la répétition de demande peut être considérée comme un intervalle entre deux rappels consécutifs de durée constante.

2.1 Modèle général d'un système de file d'attente avec rappels

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

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

Le schéma générale d'un système avec rappels est donné dans la Figure 2.1

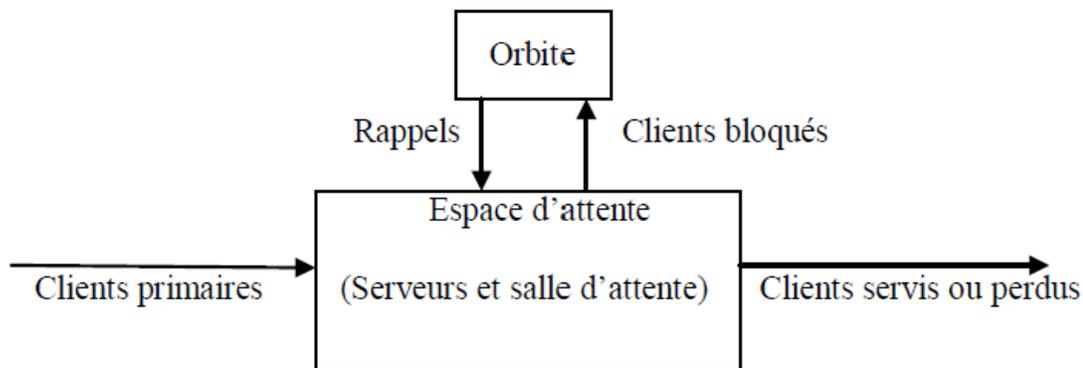


FIGURE 2.1 – Schéma générale d'un système avec rappels

2.2 Modèle d'attente $M/M/c$ avec rappels

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

2.2.1 Description mathématique du modèle

Nous considérons un système $M/M/c$ avec rappels à c serveurs identiques. Les clients primaires arrivent selon un processus de Poisson de taux λ , la durée de service est exponentielle de paramètre μ . Si un client primaire trouve au moins un serveur libre, il l'occupe immédiatement et quitte le système après le service. D'autre part, tout client qui trouve tous les serveurs occupés à son arrivée, il est obligé de quitter la zone de service, mais il rappelle ultérieurement. La durée entre deux rappels consécutifs est exponentielle de paramètre θ [20]. On utilise la politique de rappels classique et on suppose que les temps inter-arrivées, les durées de service et temps inter-rappels sont des variables aléatoire mutuellement indépendantes.

2.2.2 Distribution stationnaire

Le système peut-être décrit par le processus $X = \{C(t), N(t); t \geq 0\}$, où $C(t)$ est le nombre de serveurs occupés, $N(t)$ est le nombre de clients en orbite la date t ,

Supposons que le régime stationnaire existe ($\rho = \frac{\lambda}{c\mu} < 1$) le processus $X(t)$ est une chaîne de Markov régulière à temps continu d'espace d'états $S = \{0, 1, \dots, c\} \times \mathbb{N}$.

Soit

$$P_{ij}(t) = \mathbb{P}(C(t) = i, N(t) = j)$$

avec $i = 0, 1, \dots, c, \quad j \geq 0$.

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

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

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

pour $i = c$,

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

Dans le cas $c = 1$ et sous la condition $\rho = \frac{\lambda}{\mu} < 1$, la distribution stationnaire est donnée par [28]

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

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

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

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

Toutes les mesures de performance s'obtiennent en utilisant les fonctions génératrices (voir par exemple dans [26]).

Pour $c \geq 2$, il n'existe pas de formules analytiques explicites, mais il existe des méthodes d'approximation. Parmi ces méthodes on trouve la méthode d'approximation géométrique matricielle utilisant les processus quasi-naissance et mort (Q.B.D), cette dernière elle est utilisée dans [12].

En ordonnant les états comme $S = \{(0, 0), \dots, (c, 0), (0, 1), \dots, (c, 1), \dots\}$

Nous pouvons exprimer le générateur infinitésimal Q du processus X sous la forme d'une matrice- bloc suivante :

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

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

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

$$A_{jj} = \begin{pmatrix} a_{0j} & \lambda & & & \\ \mu & a_{1j} & \lambda & & \\ & 2\mu & a_{2j} & \lambda & \\ & & \ddots & \ddots & \ddots \\ & & & (c-1)\mu & a_{c-1,j} & \lambda \\ & & & & c\mu & a_{cj} \end{pmatrix},$$

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

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

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

Geométriquement, le comportement asymptotique du processus X , peut être représenté par

le diagramme de transition donné dans la figure 2.2.

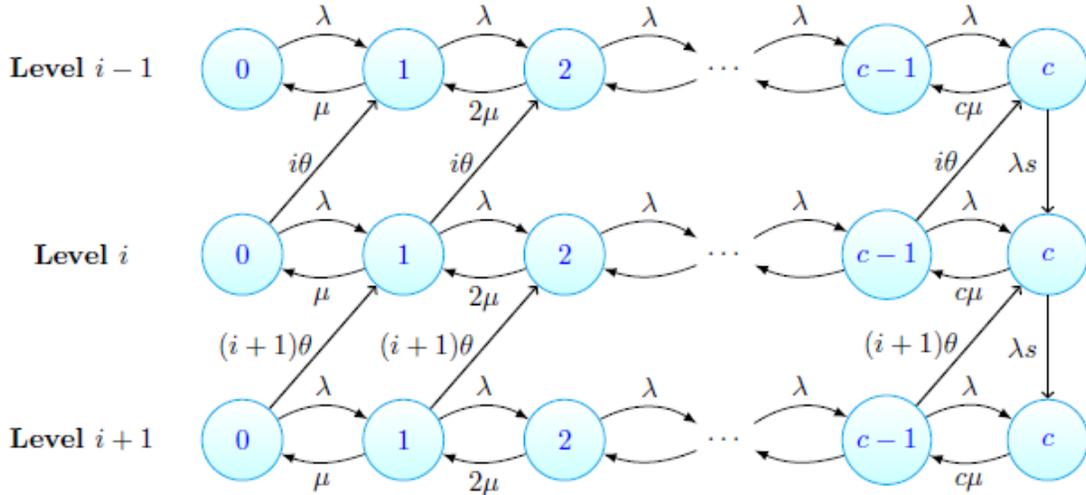


FIGURE 2.2 – Graphe de transition du modèle $M/M/c$ avec rappels

Notons que la distribution stationnaire du processus X est : $p = \{P_{ij}; (i, j) \in S\}$, le calcul de cette distribution se fait généralement à l'aide des équations de Kolmogorov $pQ = 0'$, où $0'$ indique la transposée du vecteur colonne nul. En partitionnant le vecteur de probabilité stationnaire p comme $p = (p(0), p(1), \dots)$, où $p(j) = (p_{0j}, \dots, p_{cj})$, nous pouvons écrire les équations de Kolmogorov sous forme d'une matrice :

$$p(j-1)A_{j-1,j} + p(j)A_{jj} + p(j+1)A_{j+1,j} = 0'_{c+1}, \quad j \geq 0, \quad (2.1)$$

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

On peut introduire les fonctions génératrices suivantes :

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

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

$$\mu \frac{dp(z)}{dz} A(z) = p(z) B(z), \quad (2.2)$$

— Le nombre moyen de serveurs occupés

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

En utilisant les formules de Little, on obtient

— le temps moyen de séjour dans le système, donné par

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

— le temps moyen d'attente, donné par

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

2.3 Le modèle d'attente $M/G/1$ avec rappels

Le modèle $M|G|1$ avec rappels est le modèle le plus étudié par les spécialistes. Falin [26] et Artalejo [5] ..., Il existe une littérature abondante sur ses diverses propriétés.

2.3.1 Description mathématique du modèle

Pour ce modèle, qui est composé d'un seul serveur, le processus d'arrivée des clients dans le système est de poisson de taux $\lambda > 0$: $P(\tau_n^e \leq x) = 1 - e^{-\lambda x}$. La durée de service τ est de loi générale, de moyenne $\frac{1}{\mu}$, $P(\tau_n^e \leq x) = F(x)$ et de transformée de Laplace-Stieltjes $\beta^*(s)$. Soient les moments $\beta_k = (-1)^k \beta^{*(k)}(0)$, l'intensité du trafic $\rho = \lambda \beta_1$ et $\mu = \frac{1}{\beta_1}$. La durée entre deux rappels successifs d'une même source secondaire est exponentiellement distribuée de paramètre $\theta > 0$.

$$T(x) = P(\tau_n^r \leq x) = 1 - e^{-\theta x}.$$

La description du système est la suivante : On suppose que le $(i - 1)^{\text{ème}}$ client termine son service à l'instant ξ_{n-1} (les clients sont numérotés dans l'ordre de service) et le serveur devient libre, même s'il y a des clients dans le système, ils ne peuvent pas occuper le serveur immédiatement à cause de leur ignorance de l'état de ce dernier. Donc il existe un intervalle de temps R_n durant lequel le serveur reste libre avant que le $n^{\text{ème}}$ client entre en service. A

l'instant $\xi_n = \eta_n + R_n$ le $n^{\text{ème}}$ client débute son service durant un temps τ_n^s . Les rappels qui arrivent durant ce temps de service n'influent pas sur ce processus. A l'instant $\xi_n = \eta_n + \tau_n^s$ le $n^{\text{ème}}$ client achève son service, le serveur devient libre et ainsi de suite.

Graphe de transition

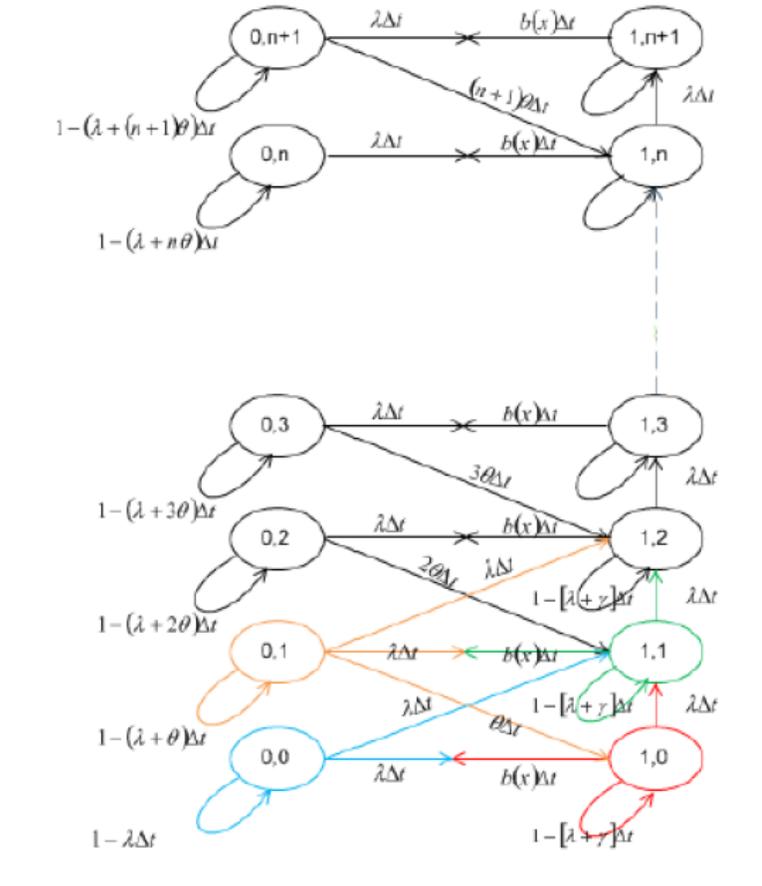


FIGURE 2.3 – Graphe de transition du modèle $M/G/1$ avec rappels

2.3.2 Chaîne de Markov induite du système

Considérons le processus $\{C(t), N(t), t \geq 0\}$. En générale ce processus n'est pas un processus de Markov, mais il possède une chaîne de Markov induite. Cette chaîne a été décrite pour la première fois par Choo et Conolly (1979)[22].

$C(t)$ est la variable aléatoire indiquant l'état du serveur à l'instant t

$$C(t) = \begin{cases} 0, & \text{si le serveur est libre;} \\ 1, & \text{si le serveur est occupé.} \end{cases}$$

Soit $q_n = N(\xi_n)$ le nombre de clients en orbite après le $n^{\text{ème}}$ départ (l'instant de départ du $n^{\text{ème}}$ client). La suite des variables aléatoires q_n forme une chaîne de Markov induite, dont l'équation fondamentale est :

$$q_{n+1} = q_n - \delta_{q_n} + V_{n+1}.$$

La variable aléatoire V_n représente le nombre de clients primaires arrivant dans le système durant le service de $(n)^{\text{ème}}$ client. Sa distribution est donnée par :

$$P(V_n = i) = a_i = \int_0^\infty \exp(-\lambda x) \frac{(\lambda x)^i}{i!} dB(x); \quad a_i > 0, \quad i \geq 0,$$

avec les résultats suivants :

$$\text{si } V = \lim_{n \rightarrow \infty} V_n, \quad E[V] = \rho.$$

$$A(z) = \sum_{i=0}^{\infty} a_i z^i = \beta^*(\lambda - \lambda z) = \int_0^\infty e^{-x(\lambda - \lambda z)} dB(x).$$

La variable aléatoire δ_{q_n} est une variable de Bernoulli définie par :

$$\delta_{q_n} = \begin{cases} 1, & \text{si le } (n+1)^{\text{ème}} \text{ client servi provient de l'orbite;} \\ 0, & \text{si le } (n+1)^{\text{ème}} \text{ client est primaire.} \end{cases}$$

Elle dépend de q_n , et sa distribution est donnée par :

$$P(\delta_{q_n} = 1/q_n = i) = \frac{i\theta}{\lambda + i\theta},$$

$$P(\delta_{q_n} = 0/q_n = i) = \frac{\lambda}{\lambda + i\theta},$$

Les probabilités de transition en une étape s'écrivent alors :

$$P_{ij} = \frac{i\theta}{\lambda + i\theta} a_{j-i+1} + \frac{\lambda}{\lambda + i\theta} a_{j-i}.$$

Pour démontrer la dernière formule, on utilise l'équation fondamentale de la chaîne :

$$\begin{aligned}
P_{ij} &= \mathbb{P}(q_{n+1} = j/q_n = i) = \mathbb{P}(q_n - \delta_{q_n} + V_{n+1} = j/q_n = i) \\
&= \mathbb{P}(V_{n+1} = j - i + \delta_{q_n}/q_n = i) \\
&= \mathbb{P}(V_{n+1} = j - i/q_n = i, \delta_{q_n} = 0)P(\delta_{q_n} = 0/q_n = i) \\
&+ \mathbb{P}(V_{n+1} = j - i + 1/q_n = i, \delta_{q_n} = 1)P(\delta_{q_n} = 1/q_n = i) \\
&= \mathbb{P}(V_{n+1} = j - i)P(\delta_{q_n} = 0/q_n = i) + \mathbb{P}(V_{n+1} = j - i + 1)P(\delta_{q_n} = 1/q_n = i) \\
&= \frac{i\theta}{\lambda + i\theta}a_{j-i+1} + \frac{\lambda}{\lambda + i\theta}a_{j-i}.
\end{aligned}$$

Distribution stationnaire de la chaîne de Markov induite

Si $\rho < 1$, la distribution stationnaire de la chaîne de Markov induite existe, et possède la fonction génératrice suivante :

$$Q(z) = \sum_{n=0}^{\infty} \pi_n z^n = \frac{(1-\rho)(1-z)\beta^*(\lambda-\lambda z)}{\beta^*(\lambda-\lambda z) - z} \exp \left\{ -\frac{\lambda}{\theta} \int_1^z \frac{1-\beta^*(\lambda-\lambda x)}{\beta^*(\lambda-\lambda x)} dx \right\},$$

$$Q(z) = \frac{(1-\rho)(1-z)\beta^*(\lambda-\lambda z)}{\beta^*(\lambda-\lambda z) - z} \frac{\phi(z)}{\phi(1)},$$

où

$$\phi(z) = \exp \left\{ \frac{\lambda}{\theta} \int_1^z \frac{1-\beta^*(\lambda-\lambda x)}{x-\beta^*(\lambda-\lambda x)} dx \right\},$$

en posant

$$\pi(z) = \frac{(1-\rho)(1-z)\hat{\beta}(\lambda-\lambda z)}{\hat{\beta}(\lambda-\lambda z) - z},$$

d'où

$$Q(z) = \pi(z) \frac{\phi(z)}{\phi(1)}.$$

Cette formule est appelée "décomposition stochastique" du système $M/G/1$ avec rappels [10].

2.3.3 Mesures de performance

Les caractéristiques du modèle $M/G/1$ avec rappels sont données dans [20].

Nombre moyen de clients dans le système :

$$L = Q'(1) = \rho + \frac{\lambda^2 \beta_2}{2(1-\rho)} + \frac{\lambda \rho}{\theta(1-\rho)}, \quad \text{avec } \beta_2 \text{ est le second moment de service}$$

Nombre moyen de client en orbite :

$$L_0 = L - \rho = \frac{\lambda^2 \beta_2}{2(1-\rho)} + \frac{\lambda \rho}{\theta(1-\rho)}.$$

Temps moyen d'attente d'un client :

$$W = \frac{\lambda \beta_2}{2(1-\rho)} + \frac{\rho}{\theta(1-\rho)}.$$

Le nombre moyen de rappels par client :

$$R = \theta W = \frac{\lambda \theta \beta_2}{2(1-\rho)} + \frac{\rho}{1-\rho}.$$

Conclusion

Dans ce chapitre nous avons présenté les mesures de performance des systèmes avec rappels dont les formules sont complexes. Dans les situations réelles et pratiques, on peut approximer les systèmes avec rappels par les systèmes classiques (voir chapitre 3).

CHAPITRE 3

ANALYSE COMPARATIVE DES MODÈLES D'ATTENTE CLASSIQUES ET DES MODÈLES D'ATTENTE AVEC RAPPELS

Dans ce chapitre, on s'intéresse aux similitudes et aux différences entre les modèles avec rappels et leurs homologues en modèles classiques.

3.1 Le modèle $M/M/c$

Soit le processus $X = \{C(t), N(t); t \geq 0\}$ décrivant le système $M/M/c$ avec rappels et on considère le processus $Y = \{Q(t); t \geq 0\}$ indiquant le nombre de clients dans le système classique $M/M/c$.

le processus Y est un processus de naissance et de mort de taux de transitions

$$\lambda_n = \lambda, \quad n \geq 0 \quad ; \quad \mu_n = \min(n, c)\mu, \quad n \geq 1.$$

Les deux processus sont récurrents positifs (ergodiques) si et seulement si $\rho = \frac{\lambda}{c\mu} < 1$.

Pour le processus Y , la démonstration utilise les résultats des processus de naissance et de mort [31], pour le processus X on utilise le critère de Foster basé sur les accroissements moyens [28].

Pour $\rho = 1$, le processus Y est récurrent nul.

Le comportement de X pour $\rho = 1$, dépend du taux des rappels.

Pour $c = 1$, $\rho = 1$, X est récurrent si et seulement si $\theta > \mu$ [28].

On remarque que la condition de la récurrence positive est indépendante du taux des rappels.

Si $\rho < 1$, la distribution stationnaire de $M/M/c$ est donnée par (Voir chapitre 1) :

$$p_n = \begin{cases} p_0 \frac{(c\rho)^n}{n!} & \text{si } 0 \leq n \leq c, \\ p_0 \frac{c^c \rho^n}{c!} & \text{si } n > c. \end{cases}$$

Où

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

En particulier pour $M/M/1$, on a la distribution géométrique $P_n = (1-\rho)\rho^n$, $n \geq 0$.

La distribution stationnaire du système $M/M/1$ avec rappels est donnée par[28] :

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

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

Donc l'expression de la distribution stationnaire du nombre total de clients dans le système est :

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

Qui peut être identifié comme la distribution binomiale négative.

Pour $c = 2$, on aura la distribution hypergéométrique. Toute fois, la considération de plus de deux serveurs compliquera les transitions d'états. Plusieurs méthodes numériques, pour le calcul de P_{ij} ont été considérées dans plusieurs articles ([27],[33],[35],[37]). Parmi ces méthodes, celle des modèles tronqués généralisés propose d'approximer un système infini, qui ne peut être résolu directement à l'aide d'un autre système infini, dont l'analyse quantitative est possible.

Période d'activité

Parmi les caractéristiques importantes, on trouve la période d'activité qui est définie comme la période commençant à l'arrivée d'un client qui trouve le système vide et se terminant au premier instant de fin de service pour lequel le système redevient vide.

Notons par L_∞ la période d'activité du système $M/M/c$ classique.

Les études existantes (voir[15]) portent principalement sur l'existence de solutions approchées pour le cas $c \leq 2$. La solution est donnée, en terme de la fonction Bessel du premier type d'ordre r , définie comme suit

$$I_r(x) = \sum_{n=0}^{\infty} \frac{\left(\frac{x}{2}\right)^{r+2n}}{(n+r)!n!}.$$

Pour $c = 1$, la fonction de densité de probabilité de L_∞ est :

$$f_{L_\infty}(x) = \frac{e^{-(\lambda+\mu)x}}{x\rho^{\frac{1}{2}}} I_1(2x\sqrt{\lambda\mu}), \quad x > 0.$$

Pour $c = 2$, nous avons

$$f_{L_\infty}(x) = \frac{e^{-(\lambda+2\mu)x}}{x} \sum_{r=0}^{\infty} (r+1) \left(\frac{\mu}{2\lambda}\right)^{\frac{r+1}{2}} I_{r+1}(2x\sqrt{2\lambda\mu}), \quad x > 0.$$

Les premiers moments de L_∞ sont donnés par

$$\mathbb{E}[L_\infty] = \frac{1}{\mu(1-\rho)},$$

$$\mathbb{E}[L_\infty^2] = \frac{2-\rho}{\mu^2(1-\rho)^3}.$$

Pour $c > 2$, il n'existe pas de formules explicites. Toutefois, d'après la théorie des processus régénératifs, la probabilité stationnaire du système vide est donnée par

$$P_0 = (1 + \lambda E(L_\infty))^{-1}.$$

Dans l'article de Artalejo et al. [11], l'investigation de la transformée de Laplace de L_∞ et le calcul de ces moments sont réduits à la solution récursive de quelques systèmes linéaires.

En comparaisant avec les résultats ci-dessus, l'étude de la période d'activité L_θ du système $M/M/c$ avec rappels, peut-être considéré comme un problème ouvert. Pour $c = 1$, les mo-

ments de L_θ sont donnés récursivement dans [22].

Temps virtuel d'attente

On considère le temps virtuel d'attente, $W(t)$, d'un client arrivant dans le système à l'instant t . $W(t)$ est par définition, le temps d'attente d'un client dans la file (modèles classique) ou dans l'orbite (modèle avec rappels). Nous traitons le système à l'état stable, donc nous indiquons $W(t)$ par W .

Dans la suite, les temps d'attente pour les files d'attente classiques et avec rappels sont respectivement notés par W_∞ et W_θ .

La fonction de distribution du temps d'attente dans la file d'attente $M/M/c$ classique peut être exprimée sous la forme d'une intégrale qui peut être développée en utilisant les polynôme orthogonaux de lagrange ou les méthodes de la série McLaurin, le résultat est le suivant (voir la référence [31])

$$P(W_\infty > x | W_\infty > 0) = 1 - c\mu x \frac{1-\rho}{\rho} \ln \frac{1}{1-\rho} + \frac{(c\mu x)^2}{2} (1-\rho) \left(2 - \frac{1-\rho}{\rho} \ln \frac{1}{1-\rho}\right) - \dots$$

D'autre part, l'analyse approximative de W_θ dans le système $M/M/c$ avec rappels est donnée dans l'article de Artalejo et Gomez [13].

Comportement asymptotique

On s'intéresse au comportement asymptotique de la file d'attente $M/M/c$ avec rappels dans le cas du taux de rappels élevé, $\theta \rightarrow \infty$, (i.e que les intervalles entre les rappels successifs tendent vers 0). La file d'attente $M/M/c$ avec rappels peut être alors considérée comme la file d'attente classique. Cette remarque peut être transformée en un résultat mathématique régoureux. Pour ce faire, notons par $P_{ij}(\infty)$ la probabilité stationnaire que le nombre de serveurs occupés soit égal à i et la longueur de la file d'attente égale j , dans la file d'attente $M/M/c$ classique, et par $P_{ij}(\theta)$ la probabilité correspondante, dans la file d'attente $M/M/c$ avec rappels. Alors

$$\lim_{\theta \rightarrow \infty} P_{ij}(\theta) = P_{ij}(\infty), \quad 0 \leq i \leq c, \quad j \geq 0,$$

où

$$P_{00}(\infty) = \left(\frac{c^c \rho^{c+1}}{c!(1-\rho)} + \sum_{k=0}^c \frac{(c\rho)^k}{k!} \right)^{-1},$$

$$P_{ij}(\infty) = \begin{cases} P_{00}(\infty) \frac{(c\rho)^i}{i!} & \text{si } 0 \leq i \leq c-1, \quad j=0, \\ P_{00}(\infty) \frac{\rho^{c+j} c^c}{c!} & \text{si } i=c, \quad j \geq 0, \\ 0 & \text{sinon.} \end{cases}$$

Considérons la probabilité de blocage B_θ dans $M/M/c$ avec rappels qui est définie comme la probabilité que tous les serveurs sont occupés, on a :

$$B_\theta = B_\infty + \frac{(c-1)\mu - \lambda + \mu B_\infty \ln(1-\rho)}{\theta} B_\infty + o\left(\frac{1}{\theta}\right), \quad (3.1)$$

où B_∞ est la probabilité de blocage dans le système $M/M/c$ classique est donnée par :

$$B_\infty = \frac{\frac{(c\rho)^c}{(c-1)!}}{\frac{(c\rho)^c}{(c-1)!} + c(1-\rho) \sum_{i=0}^{c-1} \frac{(c\rho)^i}{i!}}.$$

et le second terme de la formule décrit l'influence des rappels.

Nombre de clients limite en orbite

Soit N_θ , le nombre de clients en orbite dans $M/M/c$ avec rappels. Une propriété asymptotique de N_θ est sous la forme :

$$\mathbb{E}(N_\theta) = \mathbb{E}(N_\infty) + Z\mathbb{E}(N_\infty) + o\left(\frac{1}{\theta}\right).$$

où $\mathbb{E}(N_\infty) = \frac{\rho}{1-\rho} B_\infty$ correspond au nombre moyen de clients dans le système $M/M/c$ classique.

et

$$Z = \frac{\lambda - (c(c-1)\mu + \lambda(\lambda\mu^{-1} - B_\infty - 2(c-1)))\ln(1-\rho)}{\lambda\theta\mu^{-1}}$$

Tableau récapitulatif de l'étude comparative

Le tableau suivant résume les principales caractéristiques du système $M/M/c$ classique et $M/M/c$ avec rappels.

condition de stabilité $\rho = \frac{\lambda}{c\mu} < 1$	$M/M/c$ classique	$M/M/c$ avec rappels
Distribution stationnaire $p_n =$	$\begin{cases} p_0 \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!} & \text{si } 0 \leq n \leq c, \\ p_0 \rho^n \frac{c^c}{c!} & \text{si } n > c. \end{cases}$	pour $c = 1$ $P_j = \frac{\rho^j}{j!\theta^j} (1-\rho)^{1+\frac{\lambda}{\theta}} \prod_{k=1}^j (\lambda + k\theta)$, $j \geq 0$
Probabilité de blocage	$B_\infty = \frac{(c\rho)^c}{(c-1)! \left(\frac{(c\rho)^c}{(c-1)!} + c(1-\rho) \sum_{i=0}^{c-1} \frac{(c\rho)^i}{i!} \right)}$	$B_\theta = B_\infty + \frac{(c-1)\mu - \lambda + \mu B_\infty}{\theta} \ln(1-\rho) B_\infty + o\left(\frac{1}{\theta}\right)$
Nombre moyen de clients dans le système (dans l'orbite)	$\mathbb{E}(N_\infty) = \frac{\rho}{1-\rho} B_\infty$	$\mathbb{E}(N_\theta) = \mathbb{E}(N_\infty) + Z \mathbb{E}(N_\infty) + o\left(\frac{1}{\theta}\right)$ avec $Z = \frac{\lambda - (c(c-1)\mu + \lambda(\lambda\mu^{-1} - B_\infty - 2(c-1))) \ln(1-\rho)}{\lambda\theta\mu^{-1}}$

TABLE 3.1 – Tableau comparative : cas markovien

Illustrations numériques

Une illustration numérique peut être envisagée pour montrer l'approximation du système $M/M/c$ avec rappels par $M/M/c$ classique et montrer ainsi l'influence du taux des rappels θ et l'intensité de trafic ρ sur les valeurs moyennes de W_θ , L_θ et la probabilité de blocage.

Tous les calculs ont été réalisés avec le logiciel MATLAB.

En générant une suite de variables aléatoires $\theta = 5, 10, 15, \dots, 40$ pour $c = 2$ et $\rho = 0.25, 0.5, 0.75$. Les résultats sont données dans la figure suivante :

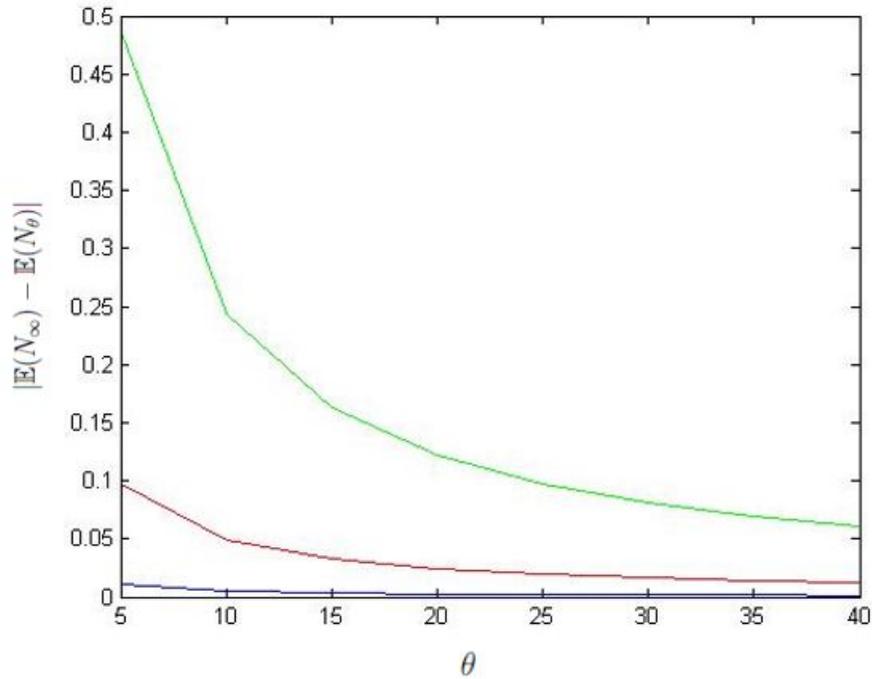


FIGURE 3.1 – Nombre moyen de clients dans le système (dans l’orbite)

Interprétations :

D’après la figure 3.1, on remarque que lorsque le taux des rappels croît ($\theta \rightarrow \infty$), la valeur $|\mathbb{E}(N_\infty) - \mathbb{E}(N_\theta)|$ converge vers zéro i.e que le nombre moyen de clients dans le système (dans l’orbite) de $M/M/c$ avec rappels peut-être approximée par celle correspondant au $M/M/c$ classique.

3.2 Le modèle $M/G/1$

Dans ce système à serveur unique on considère que la distribution du temps de service suit une loi générale. On suppose que $\rho = \lambda\beta_1 < 1$, donc le système est stable et les probabilités stationnaire des deux systèmes ($M/G/1$ classique et avec rappels) existent, et sont données respectivement par :

$$P_j = \lim_{t \rightarrow \infty} P(Q(t) = j), \quad j \geq 0,$$

$$P_{ij} = \lim_{t \rightarrow \infty} P(C(t) = i, N(t) = j), \quad (i, j) \in \{0, 1\} \times \mathbb{Z}_+,$$

Les deux distribution $\{P_j, j \geq 0\}$ et $\{P_{ij}, i \in \{0, 1\}, j \geq 0\}$ peuvent être calculés de manière récursive à l'aide des équations suivantes (voir Artalejo [12])

$$P_0 = 1 - \rho, \quad P_1 = \frac{1 - a_0}{a_0} P_0, \quad P_2 = \frac{1 - a_0 - a_1}{a_0} (P_0 + P_1),$$

$$P_{j+1} = \frac{1 - \sum_{i=0}^j a_i}{a_0} \sum_{i=0}^j P_i + \sum_{i=2}^j P_i \sum_{k=j-i+2}^j \frac{a_k}{a_0}, \quad j \geq 2,$$

$$P_{0j} = \frac{\lambda}{\lambda + j\theta} \pi_j, \quad P_{1j} = \frac{(j+1)\theta}{\lambda} P_{0,j+1}, \quad j \geq 0,$$

$$\pi_j = \sum_{i=0}^j \pi_i \frac{\lambda}{\lambda + i\theta} a_{j-i} + \sum_{i=1}^{j+1} \pi_i \frac{i\theta}{\lambda + i\theta} a_{j-i+1}, \quad j \geq 0,$$

$$\pi_0 = (1 - \rho) \exp\left\{-\frac{\lambda}{\theta} \int_0^1 \frac{1 - \beta^*(\lambda - \lambda x)}{\beta^*(\lambda - \lambda x) - x} dx\right\},$$

$$\pi_j = \sum_{i=0}^j \pi_i \frac{\lambda}{\lambda + i\theta} a_{j-i} + \sum_{i=1}^{j+1} \pi_i \frac{i\theta}{\lambda + i\theta} a_{j-i+1}, \quad j \geq 0,$$

où $\beta(\cdot)$ est la transformée de Laplace de la durée de service.

et

$$a_j = \int_0^\infty e^{-\lambda x} \frac{(\lambda x)^j}{j!} dB(x), \quad j \geq 0.$$

est la probabilité que exactement j clients arrivent pendant un temps de service. En effet, $\{\pi_j, j \geq 0\}$ correspond à la distribution de la chaîne de Markov induite. Une solution alternative en termes de fonction génératrices sont données par :

$$P(z) = \sum_{j=0}^{\infty} z^j P_j,$$

$$P_i(z) = \sum_{j=0}^{\infty} z^j P_{ij}, \quad i \in \{0, 1\}.$$

$$P(z) = \frac{(1 - \rho)(1 - z)\beta(\lambda - \lambda z)}{\beta(\lambda - \lambda z) - z}. \quad (3.2)$$

$$P_0(z) = (1 - \rho) \exp\left\{-\frac{\lambda}{\theta} \int_z^1 \frac{1 - \beta^*(\lambda - \lambda x)}{\beta^*(\lambda - \lambda x) - x} dx\right\}, \quad (3.3)$$

$$P_1(z) = \frac{1 - \beta^*(\lambda - \lambda z)}{\beta^*(\lambda - \lambda z) - z} P_0(z). \quad (3.4)$$

Notons que la solution pour les files d'attente classiques et avec rappels est donnée en terme de la transformation Laplace-Stieltjes du temps de service, mais le modèle avec rappels présente une expression plus complexe, principalement en raison de l'intégrale dans la formule (3.3).

En particulier, les valeurs moyennes sont données par :

$$\mathbb{E}(Q) = \rho + \frac{\lambda^2 \beta_2}{2(1 - \rho)}.$$

$$\mathbb{E}(C) = \rho.$$

$$\mathbb{E}(N) = \frac{\lambda^2}{1 - \rho} \left(\frac{\beta_1}{\theta} + \frac{\beta_2}{2} \right).$$

où $\mathbb{E}(Q) = \lim_{t \rightarrow \infty} \mathbb{E}(Q(t))$ et β_1, β_2 sont le 1^{er} et le 2^{ème} moment de service.

Soit (C_θ, N_θ) et (C_∞, N_∞) l'état stationnaire de $M/G/1$ avec rappels et $M/G/1$ classique.

D'après les formules (3.2), (3.3), (3.4), (C_θ, N_θ) se décompose somme de deux variables aléatoire indépendantes tel que $(C_\theta, N_\theta) = (C_\infty, N_\infty) + (0, N_\theta^0)$, où N_θ^0 : représente le nombre de clients en orbite étant donné que le serveur est libre, de fonction génératrice

$$\frac{P_0(z)}{1 - \rho}.$$

De plus, on peut estimer la mesure de proximité entre les distributions limites de $M/G/1$ avec et sans rappels [12] suivante

$$D = \sum_{i=0}^1 \sum_{j=0}^{\infty} |P_{ij}(\theta) - P_{ij}(\infty)|,$$

par

$$2(1 - \rho - \pi_0) < D < 2\left(1 - \frac{\pi_0}{1 - \rho}\right).$$

Période d'activité

La période d'activité L_∞ du système $M/G/1$ classique a été analysée par plusieurs méthodologies. Kleinrock dans son livre [31], a donné la transformé de Laplace-Stieltjes de L_∞ :

$$L_\infty^*(s) = \beta(s + \lambda - \lambda L_\infty^*(s)). \quad (3.5)$$

Le 1^{er} et le second moment de L_∞ sont donnés par :

$$\mathbb{E}(L_\infty) = \frac{\beta_1}{1 - \rho}.$$

$$\mathbb{E}(L_\infty^2) = \frac{\beta_2}{(1 - \rho)^3}.$$

D'autre part, la structure de la période d'activité L_θ du système $M/G/1$ avec rappels et sa transformée de Laplace, ont été étudiés par plusieurs méthodes [28]. Donc, on a

$$L_\theta^*(s) = \frac{\int_0^{L_\infty^*(s)} \frac{\beta^*(s+\lambda-\lambda x)}{e(s,x)(\beta^*(s+\lambda-\lambda x)-x)} dx}{\int_0^{L_\infty^*(s)} \frac{dx}{e(s,x)(\beta^*(s+\lambda-\lambda x)-x)}} \quad s > 0,$$

où $L_\infty^*(s)$ est donnée dans la formule (3.5) et $e(s, x)$ est définie par

$$e(s, x) = \exp\left\{\frac{1}{\theta} \int_0^x \frac{s + \lambda - \lambda\beta^*(s + \lambda - \lambda y)}{\beta^*(s + \lambda - \lambda y) - y} dy\right\}, \quad 0 \leq x < L_\infty^*(s).$$

La formule ci-dessus fournit une solution théorique limitée en pratique.

La moyenne de L_θ ne peut-être donnée directement par différentiation. La moyenne est donnée via les processus régénératifs [12], par :

$$\mathbb{E}(L_\theta) = \frac{1}{\lambda} \left(\frac{1}{P_{00}} - 1 \right). \quad (3.6)$$

Artalejo et Lopez- Herrero [16], ont calculé le second moment :

$$\begin{aligned} \mathbb{E}(L_\theta^2) &= \frac{1}{P_{00}} \left[\frac{1}{(1 - \rho)^2} \left(\frac{2\rho\beta_1}{\theta} + \beta_2 \right) - \int_0^1 \frac{2}{\lambda\theta(\beta^*(\lambda - \lambda x) - x)} \right. \\ &\times \left. \left(1 - \frac{\lambda(1 - x)\beta'(\lambda - \lambda x)}{\beta^*(\lambda - \lambda x) - x} - \frac{1}{1 - \rho} \exp\left\{ \frac{\lambda}{\theta} \int_x^1 \frac{1 - \beta^*(\lambda - \lambda y)}{\beta^*(\lambda - \lambda y) - y} dy \right\} \right) \right] dx. \end{aligned}$$

Temps d'attente

Le temps d'attente de la file d'attente $M/G/1$ classique est donné par la transformée de Laplace-Stieltjes suivante :

$$W_{\infty}^*(s) = 1 - \rho + \frac{\lambda(1 - \rho)}{s} \int_{L_{\infty}^*(s)}^1 \frac{(1 - x)(\beta^*(\lambda - \lambda x) - \beta^*(s + \lambda - \lambda x))}{(\beta^*(\lambda - \lambda x) - x)(x - \beta^*(s + \lambda - \lambda x))} \\ \times \exp\left\{-\int_x^1 \frac{dy}{y - \beta^*(s + \lambda - \lambda y)}\right\} dx$$

Les premiers moments de W_{∞} sont donnés par :

$$\mathbb{E}(W_{\infty}) = \frac{\lambda\beta_2}{2(1 - \rho)},$$

$$\mathbb{E}(W_{\infty}^2) = \frac{2\lambda\beta_3}{3(1 - \rho)(2 - \rho)} + \frac{\lambda^2\beta_2^2}{(1 - \rho)^2(2 - \rho)}.$$

Il existe des algorithmes pour le calcul numérique $P(W_{\infty} > t | W_{\infty} > 0)$. Néanmoins, à l'exception de quelques répartitions spéciales de temps de service, il s'agit de procédures complexes (voir [25]). La formule de transformée de Laplace-Stieltjes de W_{θ} dans la file d'attente $M/G/1$, est donnée par :

$$W_{\theta}^*(s) = 1 - \rho + \frac{\lambda(1 - \rho)}{s} \int_{L_{\infty}^*(s)}^1 \frac{(1 - x)(\beta^*(\lambda - \lambda x) - \beta^*(s + \lambda - x))}{(\beta^*(\lambda - \lambda x) - x)(x - \beta^*(s + \lambda - \lambda x))} \\ \times \left\{ \frac{1}{\theta} \int_x^1 \left(\frac{s + \theta + \lambda - \lambda y}{\beta^*(s + \lambda - \lambda y) - y} - \frac{\lambda - \lambda y}{\beta^*(\lambda - \lambda y) - y} \right) dy \right\} dx.$$

Le temps d'attente moyen peut être obtenu facilement à l'aide de la formule de Little :

$$\mathbb{E}(W_{\theta}) = \mathbb{E}(W_{\infty}) + \frac{\rho}{\theta(1 - \rho)}.$$

Récemment, Artalejo et al. [11], on obtenu la formule suivante pour le second moment de W_{θ}

$$\mathbb{E}(W_{\theta}^2) = \mathbb{E}(W_{\infty}^2) + \frac{\lambda\beta_2}{\theta} \left(\frac{2}{(1 - \rho)^2(2 - \rho)} + \frac{\rho}{(1 - \rho)^2} \right) + \frac{2\rho}{\theta^2(1 - \rho)^2}.$$

Tableau récapitulatif de l'étude comparative

Le tableau 3.2 résume les principales caractéristiques du système $M/G/1$ classique et $M/G/1$ avec rappels.

condition de stabilité $\rho = \frac{\lambda}{\mu} < 1$	$M/G/1$ classique	$M/G/1$ avec rappels
Fonction génératrice de la distribution stationnaire	$P(z) = \frac{(1-\rho)\beta^*(\lambda-\lambda z)(1-z)}{\beta^*(\lambda-\lambda z)-z}$	$P_0(z) = (1-\rho)\exp\left\{-\frac{\lambda}{\theta} \int_z^1 \frac{1-\beta^*(\lambda-\lambda x)}{\beta^*(\lambda-\lambda x)-x} dx\right\}$ et $P_1(z) = \frac{1-\beta^*(\lambda-\lambda z)}{\beta^*(\lambda-\lambda z)-z} P_0(z)$
Nombre moyen de clients dans le système (dans l'orbite)	$\mathbb{E}(Q) = \rho + \frac{\lambda^2 \beta_2}{2(1-\rho)}$	$\mathbb{E}(N) = \frac{\lambda^2}{1-\rho} \left(\frac{\beta_1}{\theta} + \frac{\beta_2}{2} \right)$
Période d'activité	$\mathbb{E}(L_\infty) = \frac{\beta_1}{1-\rho}$	$\mathbb{E}(L_\theta) = \frac{1}{\lambda} \left(\frac{1}{P_{00}} - 1 \right)$ avec $P_{00} = \exp\left\{-\frac{\lambda}{\theta} \int_0^1 \frac{1-\beta^*(\lambda-\lambda x)}{\beta^*(\lambda-\lambda x)-x} dx\right\}$
Temps moyen d'attente	$\mathbb{E}(W_\infty) = \frac{\lambda \beta_2}{2(1-\rho)}$	$\mathbb{E}(W_\theta) = \mathbb{E}(W_\infty) + \frac{\rho}{\theta(1-\rho)}$

TABLE 3.2 – Tableau comparative : cas non markovien

Illustrations numériques

Dans cette section, quelques résultats numériques sont présentés pour illustrer l'approximation du système $M/G/1$ classique par $M/G/1$ avec rappels. Tout les calculs ont été réalisés avec le logiciel MATLAB.

Premier cas

On considère que la distribution du temps de service suit une loi exponentielle de paramètre μ . En générant une suite de variables aléatoires $\theta = 5, 10, 15, \dots, 40$ et $\rho = 0.25, 0.5, 0.75$. Les résultats obtenus sont exhibés dans les figures suivantes :

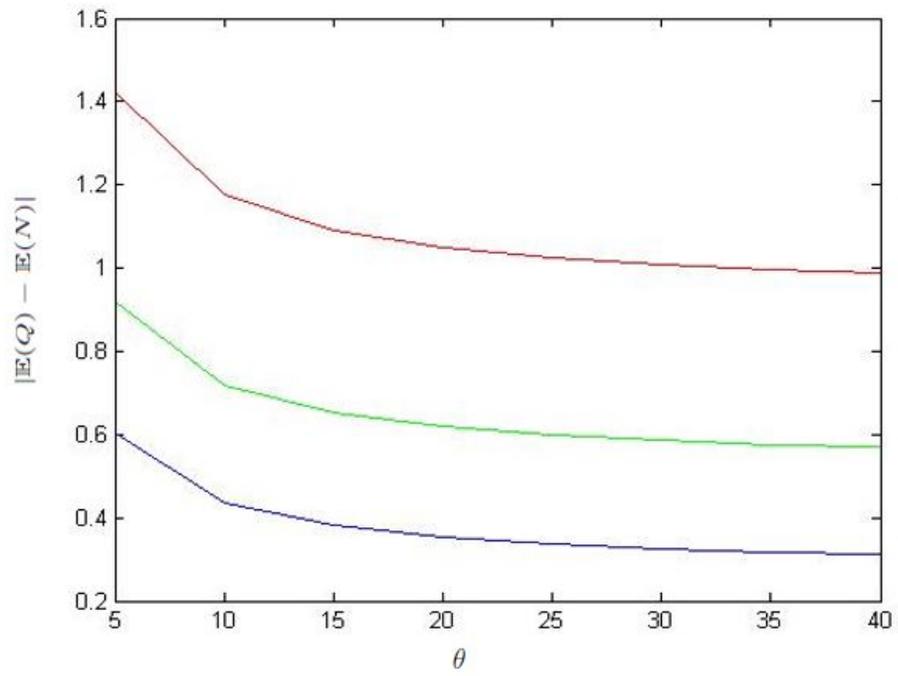


FIGURE 3.2 – Nombre moyen de clients dans le système (dans l'orbite)

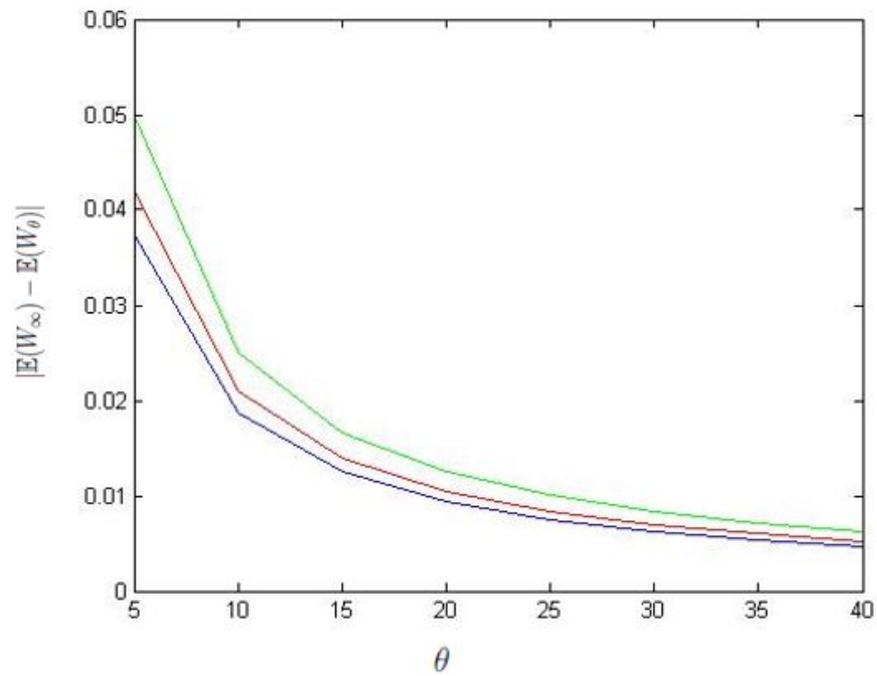


FIGURE 3.3 – Temps moyen d'attente

Interprétations

- Lorsque le taux des rappels croît ($\theta \rightarrow \infty$), la valeur $|\mathbb{E}(Q) - \mathbb{E}(N)|$ converge vers zéro i.e que le nombre moyen de clients dans le système (dans l'orbite) de $M/G/1$ avec rappels peut-être approximé par celle correspondant au $M/G/1$ classique. Ce qui est logique le montre la figure 3.2 .

- D'après la figure 3.3. On remarque que lorsque le taux des rappels croît ($\theta \rightarrow \infty$), la valeur $|\mathbb{E}(W_\infty) - \mathbb{E}(W_\theta)|$ converge vers zéro i.e que le temps moyen d'attente du système $M/G/1$ avec rappels peut être approximé par celle correspondant au système $M/G/1$ classique.

Deuxième cas

On considère que la distribution du temps de service suit une loi de Erlang. En générant une suite de variables aléatoires $\theta = 5, 10, 15, \dots, 40$ et $\rho = 0.25, 0.5, 0.75$. Les résultats obtenus sont exhibés dans les figures suivantes :

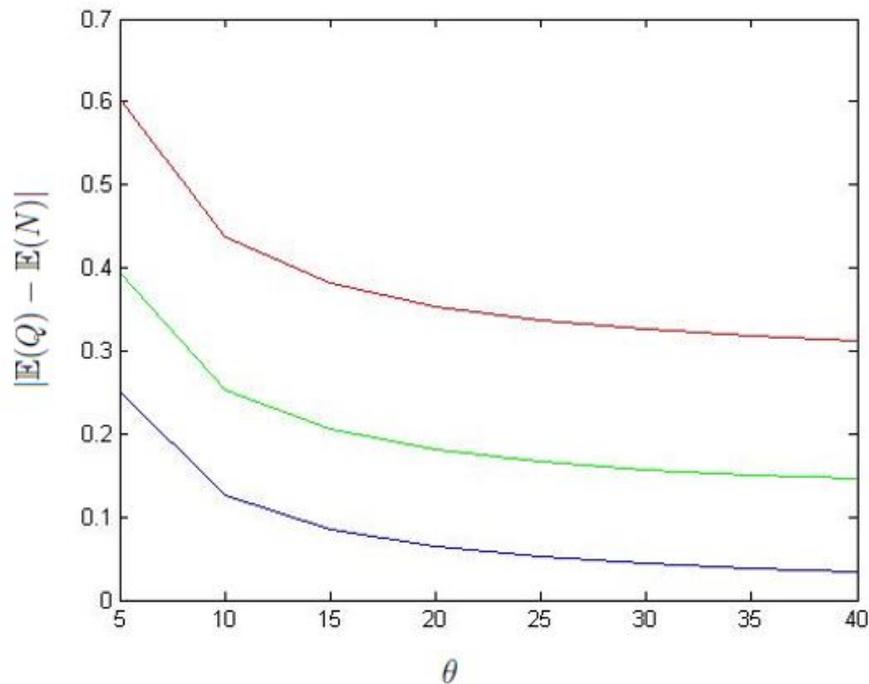


FIGURE 3.4 – Nombre moyen de clients dans le système (dans l'orbite)

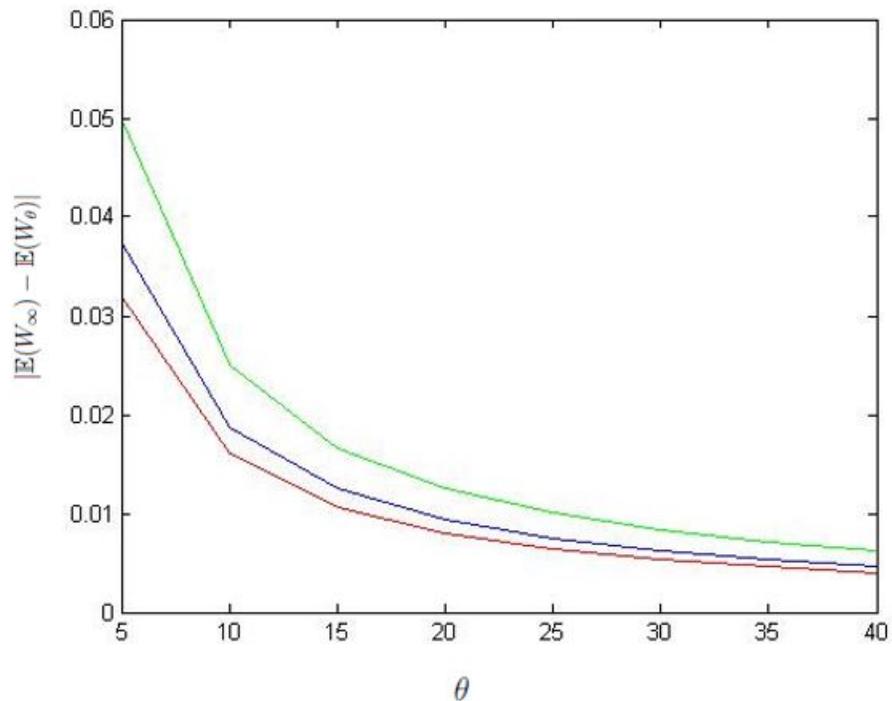


FIGURE 3.5 – Temps moyen d’attente

Interprétations

- D’après la figure 3.4. On remarque que lorsque le taux des rappels croit ($\theta \rightarrow \infty$), la valeur $|\mathbb{E}(Q) - \mathbb{E}(N)|$ converge vers zéro i.e que le nombre moyen de clients dans le système (dans l’orbite) de $M/G/1$ avec rappels peut-être approximé par celle correspondant au $M/G/1$ classique.

- Lorsque le taux des rappels croit ($\theta \rightarrow \infty$), la valeur $|\mathbb{E}(W_\infty) - \mathbb{E}(W_\theta)|$ converge vers zéro i.e que le temps moyen d’attente du système $M/G/1$ avec rappels peut être approximé par celle correspondant au système $M/G/1$ classique. Ce qui est logique le montre la figure 3.5 .

3.3 Conclusion

Dans ce chapitre on a comparé les caractéristiques principales des modèles markoviens $M/M/c$ et $M/M/c$ avec rappels ainsi que les caractéristiques des modèles non markoviens $M/G/1$ et $M/G/1$ avec rappels lorsque le taux des rappels $\theta \rightarrow \infty$, une similitude des résultats est constatée.

CONCLUSION GÉNÉRALE

Dans ce travail, nous nous sommes intéressées à l'étude comparative des systèmes de files d'attente avec rappels, et classiques.

Nous avons comparé leurs principales caractéristiques, à savoir la probabilité de blocage, les moments de la période d'activité, le temps d'attente moyen et le nombre moyen de clients en orbite.

Nous avons étudié aussi le comportement asymptotique des deux types de modèles et nous avons constaté que lorsque le taux des rappels $\theta \rightarrow \infty$, les modèles avec rappels sont approximés par les modèles classiques.

Comme perspectives de ce travail :

- ✓ Elargir l'étude de la période d'activité à d'autres systèmes avec rappels à savoir : systèmes avec rappels constants et linéaires.
- ✓ Application aux systèmes réels dans le domaine informatique ou télécommunication.

BIBLIOGRAPHIE

- [1] A. Aissani. A Survey on Retrial Queueing Models. Actes Journées Statistiques Appliquées, U.S.T.H.B. Alger. p.1-11. 1994.
- [2] J. Amador and J. R. Artalejo. The $M/G/1$ retrial queue : New descriptors of the customer's behavior. Journal of computational and Applied Mathematics 223. p.15-26. 2009.
- [3] V. V. Anisimov and J. R. Artalejo. Approximation of multiserver retrial queues by means of generalized truncated models. Top 10. p.51-66. 2002.
- [4] N.K. Arrar. Problèmes de convergence, optimisation d'algorithmes et analyse stochastique de systèmes de files d'attente avec rappels. Thèse de Doctorat en Mathématiques Appliquées. Université de Annaba. 2012.
- [5] J. R. Artalejo. Analysis of an $M/G/1$ queue with constant repeated attempts and server vacations. Computers. 24(N6) : 493-504. 1997.
- [6] J. R. Artalejo. Accessible bibliography on retrial queues. Mathematical and computer Modelling 30. p.1-6. 1999.
- [7] J. R. Artalejo. Accessible bibliography on retrial queues. Mathematical and computer Modelling. p.187-211. 2010.
- [8] J. R. Artalejo, Stationnary analysis of the characteristics of the $M/M/2$ queue with constant repeated attempts. Opsearch 33. 83-95. 1996.
- [9] J. R. Artalejo and G. I. Falin. Standard and retrial queueing systems : A comparative analysis. Revista Mathematica Complutense. p.101-129. 2002.

-
- [10] J. R. Artalejo and G. I. Falin. Stochastic decomposition for retrial queues. Top 2(2) : p.01-14, 1994.
- [11] J. R. Artalejo, G. I. Falin and M. J. Lopez-Herrero. A second order analysis of the waiting time in the $M/G/1$ retrial queue, Asia-Pacific Journal of Operational Research 19. p.131-148. 2002.
- [12] J. R. Artalejo and A. Gomez-Corral. Retrial Queueing systems. A Computational approach. Springer. 2008.
- [13] J. R. Artalejo and A. Gomez-Corral. Waiting time in the $M/M/c$ queue with finite retrial group. Bulletin of Karala Mathematics Association 2, p.1-17, 2005.
- [14] J. R. Artalejo and M. J. Lopez-Herrero. On the busy period of the $M/G/1$ retrial queue, Naval Research Logistics 47. p.115-127. 2000.
- [15] J. R. Artalejo and M. J. Lopez-Herrero. Analysis of the busy period for the $M/M/c$ queue : an algorithmic approach. Journal of Applied Probability 38. p.209-222. 2001.
- [16] J. R. Artalejo and M. J. Lopez-Herrero. On the $M/G/1$ queue with quadratic repeated attempts. Statistical method, 3, p.60-78. 2001.
- [17] J. R. Artalejo and V. Pla. On the impact of customer balking, impatience and retrials in telecommunication systems. Computers Mathematics with Applications, 57, p. 217-229. 2009.
- [18] J. R. Artalejo and M. Pozo. Numerical calculation of the stationnary distribution of the main multiserver retrial queue. Analyse of Operations Research 116. p. 41- 56. 2002.
- [19] K. Avrachenkov and U. Yechiali. Retrial networks with finite buffers and their application to internet data traffic. Probability in the Engineering and Informational Sciences, 22, p.519-536. 2008.
- [20] L. Berdjoudj. Stabilité forte dans les systèmes de files d'attente avec rappels. Thèse de magister en Mathmatiques Appliquées. Université de Bejaia. 2000.
- [21] U. N. Bhat. An Introduction to Queueing Theory : Modeling and Analysis in Applications. Boster. 2008.
- [22] Q. H. Choo and B.Conolly. New results in the theory of repeated orders queueing systems. Journal of applied probability 16. p.335-356. 1979.

-
- [23] C. Clos. An aspect of the dialing behaviour of subscribers and its effect on the trunk plant. *Bell Syst. Tech. J.* 27 : p. 424-445. 1948.
- [24] J. W. Cohen. Basic problem of telephone and influence of repeated calls. *Telecommunication review*, (2) : p. 49-100. 1957.
- [25] R. B. Cooper. *Introduction to Queueing Theory*. Edward Arnold. London. 1981.
- [26] G. I Falin. A survey of retrial queues. *Queueing systems* 7. p. 127-168. 1990.
- [27] G. I Falin. Calculation of probability characteristics of a multiline system with repeat calls. *Moscow University Computational Mathematics and Cybernetics* 1. p. 43-49. 1983.
- [28] G. I. Falin and J. G. C. Templeton, *Retrial Queues*. Monographs on statistics and Applied Probability, vol. 75. Chapman and Hall. London. 1997.
- [29] D. Gross, J. F. Shortle, J. M. Thompson and C. M. Harris. *Fundamentals of Queueing Theory*. Fourth Edition John Wiley Sons. 2008.
- [30] J. Kim and B. Kim. A survey of retrial queueing systems. *Annals of operations Research*. p. 1-34. 2016.
- [31] L. Kleinrock. *Queueing systems L. Kleinrock, Volume I : Theory*. John Wiley and Sons. 1975.
- [32] T. Phung-Duc. *Retrial Queueing Models : A survey on Theory and Application*. arXiv : 1902.09560v1[cs.PF] 23 Jun 2019.
- [33] M. F. Neuts and B.M.Rao. Numerical investigation of a multiserver retrial model. *Queueing System* 7. p. 169-190. 1990.
- [34] A. Shekhar, A. Raina and A. kumar. A brief review on retrial queue. *International Journal of Applied Sciences and Engineering Research*, 5(4). 2016.
- [35] S.N. Stepanov. Markov models with retrials : the calculation of stationary performance measures based on the concept of truncation. *Mathematical and Computer Modeling* 30. p. 207-228. 1999.
- [36] V. D. Tien, R. Chakha. An efficient method to compute the rate matrix for retrial queues with large number of servers. *Applied Mathematics Letters* 23. p. 638-643. 2010.
- [37] R.I. Wilkinson. Theories for toll traffic engineering in the USA. *The bell System Technical Journal* 35. p. 421-514. 1956.

-
- [38] T. Yang and J. G. C. Templeton, A survey on retrial queues. *Queueing System* 2. p. 201-233. 1987.
- [39] W. Yue, Y. Takahashi and H. Takagi. *Advanced in Queueing Theory and Network Application*. Springer. 2009.

Résumé

Dans ce travail, on s'est intéressé à l'étude comparative des modèles d'attente avec rappels et les modèles d'attente classiques de type $M/M/c$ et $M/G/1$.

Dans un premier temps, une synthèse des résultats existant dans la littérature sur les modèles considérés a été faite.

Dans un second temps une comparaison de leurs caractéristiques principales a été étudiée, à savoir : la distribution stationnaire, la probabilité de blocage, les moments de la période d'activité, le nombre moyen de clients dans le système (en orbite pour les systèmes avec rappels) et le temps moyen d'attente.

Le comportement asymptotique, lorsque le taux des rappels tend vers l'infini a été également discuté.

Mots clés : Files d'attente, rappels, orbite, période d'activité, temps d'attente, comportement asymptotique.

Abstract

In this work, we looked at the comparative study of retrial queueing systems and queueing systems of type $M/M/c$ and $M/G/1$.

Firstly, a synthesis of the results existing in the literature on the models considered was mod. Secondly, A comparison of their main characteristics was studied, to know : stationary distribution , the blocking probability, the moments of a busy period, the mean number of customers in the system (in orbit for the systems in retrial queues) and the waiting mean time.

The asymptotic behavior, when the recall rate tends to infinity was also discussed.

Key words : Queues, retrial, orbit, busy period, waiting time, asymptotic behavior.