

N° d'ordre :

REPUBLIQUE ALGERIENNE DEMOCRATIQUE & POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR & DE LA RECHERCHE
SCIENTIFIQUE



UNIVERSITE DJILLALI LIABES
FACULTE DES SCIENCES EXACTES
SIDI BEL ABBES

THESE DE DOCTORAT

Présentée par :
ELHADDAD Meriem

Domaine : Mathématiques Informatique

Filière : Mathématiques

*Intitulé de la formation : Statistique, Mathématiques
appliquées à l'économie et à la finance*

Intitulée

Approximate analysis of M/M/c retrial queue

Soutenue le : 28/10/2020

Devant le jury composé de :

Président :

Mr GHERIBALLAH Abdelkader Professeur

à L'Université S.B.A

Examineurs :

Mr KANDOUZI Abdeldjebbar Professeur

à L'Université SAIDA

Mr BENCHIKH Tawfik Professeur

à L'Université S.B.A

Mr MECHAB Boubaker Maître de Conférence A

à L'Université S.B.A

Directeur de thèse :

Mme LIMAM-BELARBI Faiza Professeur

à L'Université S.B.A

Année universitaire : 2019/2020.

*Cette thèse est dédiée à toute ma famille...
et mes meilleurs amies...*

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 tiens d'exprimer toute ma reconnaissance à ma directrice de thèse **Prof. LIMAM-BELARBI Faiza** de m'avoir encadré, orienté, aidé et conseillé, et ainsi pour la confiance et la liberté qu'elle m'a accordé, sans oublier ses efforts fournis pendant ces longues d'années.

Je remercie sincèrement Monsieur le Professeur **GHERIBALLAH Abdelkader** pour l'honneur qu'il ma fait en présidant le jury. Je suis très honoré de sa présence.

Mes vifs remerciements à Messieurs **KANDOUCI Abdeldjebbar** , **BENCHIKH Tawfik** et **MECHAB Boubaker** qui m'ont fait l'honneur de participer à mon jury. Je les remercie vivement pour l'attention qu'ils ont accordé à ce travail.

J'adresse mes sincères remerciements à tous les professeurs qui m'ont enseigné.

Je remercie les membres du Laboratoire de Statistique et Processus Stochastiques de l'université Djillali Liabès de Sidi Bel Abbès. J'ai toujours trouvé le soutien et l'encouragement.

Je voudrais exprimer ma reconnaissance envers mes amies et collègues qui m'ont apporté leur support moral et intellectuel tout au long de ma démarche.

Ce travail n'aurait pu arriver à terme sans le soutien, la prière et l'encouragement familial surtout mon marie et ma chère mère que dieu les garde pour moi en bon santé.

Merci à tous ceux qui ont contribué, de près ou de loin, à l'aboutissement de ce travail.

Résumé

Les réseaux de files d'attente sont des outils importants pour modéliser et évaluer les performances de toute une variété de systèmes réels. Cependant, les résultats analytiques ne sont pas toujours disponibles sous forme explicite. Pour la résolution de ces problèmes, on fait souvent recours à des méthodes d'approximation qui consistent à assimiler le système complexe à un système plus simple, analytiquement exploitable, qui lui est proche dans un certain sens. Dans ce travail, nous nous sommes intéressés à évaluer les performances d'un système de files d'attente avec rappels de source infinie $M/M/s$ avec impatience où tous les serveurs sont sujets à des pannes. En appliquant l'algorithme de phase merging, un modèle simplifié du système est construit sur la base de l'espace de phase fusionné de l'état du système, Cela nous a permis d'approximer l'état stationnaire de la répartition conjointe du nombre de clients en orbite et du nombre de clients dans la zone de service.

Abstract

Queue networks are important tools for modeling and assessing performance of a variety of real systems. However, the analytical results are not always available in explicit form. To solve these problems, we often use approximation methods which consist in assimilating the Complex system to a simpler, analytically exploitable system, which is close to him in a certain sense. In this work, we are interested in evaluating the performance of an $M/M/s$ Markovian multiserver infinite source retrial queuing with impatience, in which all the servers are subject to breakdown and repairs. By applying the merging phase algorithm, a simplified model of the system is built on the basis of the merged phase space of the system state, This allowed us to approximate the stationary state of the joint distribution of the customers in orbit and of the number of customers in the service area.

Table des matières

Introduction	7
1 Systèmes de files d'attente	11
1.1 Files d'attente classiques	11
1.1.1 Introduction	11
1.1.2 Structure des systèmes de files d'attente	12
1.1.3 Gestion de la file d'attente	12
1.1.4 Caractéristiques d'un système de files d'attente	13
1.1.5 Analyse mathématique d'un système de files d'attente	14
1.1.6 Quelques systèmes de files d'attente classiques	14
1.1.7 Les systèmes de files d'attente complexes	20
1.2 Files d'attente avec rappels	20
1.2.1 Définition	20
1.2.2 Politiques d'accès au serveur à partir de l'orbite	21
1.3 Interruption de service dans les systèmes d'attente avec rappels	22
1.3.1 Quelques cas modélisés par des systèmes de files d'attente avec rappels	23
1.4 Files d'attente avec des clients impatientes	25
1.4.1 Revue de littérature	26
2 The Phase Merging Algorithm	29
2.1 Introduction	29
2.2 L'approche APM	30
2.2.1 Introduction	30
2.2.2 Description de l'algorithme de fusion de phase APM	32

3	Approximate analysis of an unreliable $M/M/c$ retrial queue with phase merging algorithm	35
3.1	Introduction	37
3.2	Model description	37
3.3	Main results	39
3.4	Proof	40
3.4.1	The Phase-Merging Algorithm	40
3.4.2	Approximation Using the Phase-Merging Algorithm	45
3.5	Illustrative example : $M/M/3$	49
3.6	Numerical Experiments	54
3.7	Conclusion	55
4	On the analysis of unreliable Markovian multiserver queue with re-trials and impatience	57
4.1	Introduction	59
4.2	Model description	60
4.3	Underlying Markov Chain	60
4.4	Application of the Phase Merging Algorithm	62
4.5	results	66
4.6	Model Generalization	66
4.7	Numerical Results	71
	Conclusion	77

Introduction

Depuis le dernier siècle, les planificateurs militaires ont cherché à obtenir un avantage pour la guerre en rassemblant rapidement des informations sur le lieu où se trouvaient les ennemis et leurs objectifs. Pour faciliter le partage d'informations entre les principaux acteurs, des réseaux de communication élaborés sont nécessaires et utilisent plusieurs types de supports de transmission pour inclure des réseaux informatiques, des émetteurs audio et vidéo (sur terre, en mer ou dans les airs). Ces réseaux doivent également avoir la capacité de prendre en charge plusieurs types de données, y compris du texte standard, audio et vidéo.

Un partage efficace des informations via ces types de configurations de réseaux est essentiel pour la mise en œuvre du concept de guerre réseau-centrée (network-centric warfare NCW). La capacité de recueillir les bonnes informations et de les partager en temps opportun est l'objectif de NCW. la bonne mise en œuvre de ce dernier mène directement à l'atteinte de la supériorité de l'information, qui à son tour fournit au combattant de guerre et au commandant une connaissance de la situation partagée aidant à la réussite de la mission. La compréhension du mouvement de l'information à travers les réseaux de traitement de l'information est utile au processus de modélisation du flux de données de l'expéditeur au récepteur. Plusieurs supports de transmission sont nécessaires pour traiter avec précision et rapidité la suite d'informations en ce qui concerne NCW. Chacun des supports de transmission peut être sensible aux perturbations dues à la congestion du trafic, aux effets des conditions météorologiques, aux dommages dus aux tirs ennemis ou aux pannes mécaniques. Considérons, par exemple, que les informations transmises prennent la forme de paquets de données sensibles au temps qui arrivent de manière aléatoire sur différents supports de transmission, si un support est occupé à transmettre d'autres paquets de données ou subit une interruption, les paquets de données arrivant sont retardés. En

outre, un support qui échoue pendant la transmission entraîne également des paquets de données retardés. Selon l'importance et la sensibilité temporelle des informations, elles peuvent être retransmises ultérieurement ou éventuellement abandonnées. Par exemple, il est possible que la localisation d'un groupe terroriste soit connue au cours des prochaines minutes, si un message d'attaque est retardé, l'emplacement du groupe peut changer au moment où la retransmission se produit, dans ce cas, le message pourrait être rendu inutile. Naturellement, la retransmission ne peut avoir lieu qu'après que le support perturbé est à nouveau opérationnel et disponible.

De nombreux systèmes de services stochastiques du monde réel, y compris les réseaux de partage d'informations susmentionnés, peuvent être modélisés comme des systèmes de files d'attente avec rappel à serveurs non fiables. Certains d'entre eux comprennent les réseaux de téléphonie cellulaire, les réseaux informatiques et les centres de contact clients. Ces centres emploient plusieurs opérateurs pour répondre aux demandes de service des clients avec une garantie de qualité de service. Cependant, si le service n'est pas initié en temps opportun, les clients peuvent choisir d'abandonner leurs demandes. En outre, le service peut être interrompu par des événements aléatoires de la baie tels que la congestion du réseau des appels mal dirigés, des pannes mécaniques ou d'autres circonstances imprévues. Cela peut tous conduire à l'insatisfaction des clients, ce qui peut entraîner l'abandon du client suite à une perturbation.

Parce que les organisations de services sont les plus intéressées à fournir un niveau élevé de satisfaction du client, il est impératif que ces systèmes soient analysés afin d'évaluer leurs performances et de fournir un moyen par lequel ils peuvent être conçus de manière optimale, doté en personnel et exploité. La littérature sur les files d'attente contient une quantité importante de travail consacrée à l'analyse des performances des systèmes de files d'attente avec rappel à un seul serveur peu fiables. Cependant, à notre connaissance, le cas de plusieurs serveurs est resté relativement inexploré, l'objectif principal de cette recherche est d'analyser formellement un système de files d'attente avec rappel $M/M/s$ à s serveurs non fiables et des clients impatientes. Comme on le verra aux chapitres 3 et 4, une analyse exacte d'un tel système est complexe et peut-être intraitable. C'est pourquoi nous concentrerons notre attention sur une analyse approximative.

Les systèmes de files d'attente avec rappel à multi-serveurs, en général, sont difficiles

à analyser d'un point de vue mathématique. Les résultats exacts des probabilités en régime permanent des systèmes fiables ne sont donnés que pour les cas à un et deux serveurs. Dans le modèle peu fiable, il n'y a pas de solutions exactes lorsque le nombre de serveurs dépasse un. Par conséquent, nous cherchons à approximer la distribution conjointe en régime permanent du nombre de clients en orbite et de l'état des serveurs pour le cas de processus d'arrivée des clients et de processus de service markovien. Nous fournissons également des expressions approximatives pour plusieurs mesures de performances de mise en files d'attente. Notre approche pour déterminer les probabilités approximatives en régime permanent utilise un algorithme de fusion de phases décrit par Korolyuk et korolyuk [28] et Courtois [17].

En plus de l'introduction, le document se compose de quatre chapitres, une conclusion générale et une bibliographie.

- Dans le premier chapitre, nous donnons une synthèse sur la théorie de files d'attente classiques et avec rappels et priorité.

- Dans le deuxième chapitre nous présentons une description de l'algorithme qui est l'outil de base de notre étude pour l'évaluation de performances de nos deux systèmes considérés dans les chapitres trois et quatre.

- Dans le troisième chapitre, nous fournissons la description formelle du notre premier modèle $M/M/c$ et énonçons les hypothèses nécessaires à la mise en œuvre de la procédure d'approximation en l'appliquant à notre modèle, nous dérivons une approximation des probabilités de régime permanent et plusieurs d'autres mesures de performance de notre file d'attente standard, ensuite nous évaluons la qualité de l'approximation en comparant l'efficacité numérique avec celle obtenu avec la méthode de troncature directe [39] en prenant le cas de $M/M/3$.

Dans le chapitre 4, notre deuxième modèle $M/M/s$, en lui introduisant la notion d'impatience des clients, nous fournissons la description du modèle et nous analysons l'effet simultané de balking et de l'impatience de clients sur les performances du système en employant l'algorithme de fusion de phases.

En conclusion, nous résumons les principales contributions de la thèse. Enfin, quelques idées de recherche sont suggérées et pourraient faire avancer le domaine des systèmes de file d'attente pour les procès non fiables.

Chapitre 1

Systèmes de files d'attente

1.1 Files d'attente classiques

1.1.1 Introduction

On parle de phénomène d'attente, chaque fois que certaines unités appelées clients se présentent d'une manière aléatoire à des stations, afin de recevoir un service dont la durée est généralement aléatoire. Les phénomènes d'attente surviennent naturellement dans la plupart des environnements rencontrés dans la vie courante et particulièrement dans les systèmes informatiques, les systèmes de production et les réseaux de télécommunication. Ces phénomènes peuvent être modéliser par une technique de la recherche opérationnelle qui est la théorie de files d'attente. Cette théorie vise à optimiser les ressources disponibles et gérer le temps d'attente des clients sollicitant un service précis pour une durée de temps déterminée. Elle a été élaboré pour la première fois par l'ingénieur danois Agner Krarup Erlang en 1917, alors qu'il s'intéressait à la modélisation du réseau téléphonique de Copenhague, afin d'en réduire les phénomènes d'attente et elle s'est développée par la suite grâce notamment aux travaux de Palm, de Kolmogorov et de Khintchine.

Un modèle de files d'attente est un outil puissant pour l'analyse et l'optimisation des performances des systèmes à flux discret tels que les systèmes informatiques et les réseaux de communication. Pour un modèle typique de files d'attente, il est né-

cessaire de définir les processus d'inter-arrivées, la durée de service, la taille de la file qui peut être finie ou non et la discipline de service. L'objectif est de caractériser le degré de performance du système en répondant à des questions du type suivant :

- En moyenne, combien de temps attend un client avant d'être servi ?
- Quel est le nombre moyen de clients dans le système ?
- Quel est le taux d'occupation du serveur ?

1.1.2 Structure des systèmes de files d'attente

Pour traiter l'étude des files d'attente, nous allons utiliser un modèle abstrait dénommé station de service. Cette station est composée d'un ou plusieurs serveurs représentant la ressource et d'une file d'attente contenant à tout instant les clients en attente de service (serveurs occupés). Si un poste de service est libre, le client qui arrive se dirige immédiatement vers ce poste où il est servi, sinon, il prend sa place dans une file d'attente dans laquelle les clients se rangent suivant leur ordre d'arrivée.

1.1.3 Gestion de la file d'attente

- **La capacité de la file d'attente** : peut être limitée, par exemple la salle d'attente d'un cabinet médical ou illimité, par exemple les péages de l'autoroute.
- **Discipline de service** : Les clients peuvent être servis suivant les règles de gestion
 - FIFO (Le premier arrivé est le premier servi).
 - LIFO (Le dernier arrivé sera le premier servi).
 - choix suivant les règles de priorité.
- **Comportement des clients** : les clients peuvent être *patients* s'ils attendent patiemment leur tour ou peuvent être *impatients* s'ils refusent à rejoindre la file longue.

La notation suivante, introduite par Kendall permet de désigner de manière synthétique un système de files d'attente : $A/B/C/D/E$, ses termes sont les suivants :

A : la loi des inter-arrivées ou du processus d'arrivée.

B : la loi de service.

C : le nombre de serveurs.

D : la capacité d'accueil de la file d'attente .

E : la discipline de service.

La forme abrégée de la file : $A/B/C$ signifie que D est infinie et que la discipline est FIFO.

Exemple :

- La file $M/M/1/\infty/FIFO$

M processus d'arrivée : Poisson de paramètre $\lambda > 0$.

M distribution du temps de service : exponentielle de paramètre $\mu > 0$.

- La file $M/G/1/\infty/FIFO$

M processus d'arrivée : Poisson de paramètre $\lambda > 0$.

G distribution du temps de service : général.

1.1.4 Caractéristiques d'un système de files d'attente

A partir de la distribution stationnaire, on pourra obtenir d'autres caractéristiques d'exploitation du système telles que :

- le nombre moyen L de clients dans le système ;
- le nombre moyen L_q de clients dans la file d'attente ;
- la durée d'attente moyenne W_q d'un client ;
- la durée de séjour moyenne W dans le système (attente + service).

Ces valeurs sont liées les unes aux autres par les relations suivantes[4] :

$$\begin{aligned} L &= \lambda_e W; \\ L_q &= \lambda_e W_q; \\ W &= W_q + \frac{1}{\mu}; \\ L &= L_q + \frac{\lambda_e}{\mu} \end{aligned}$$

avec

λ_e : taux d'arrivée dans le système ;

μ : taux de service ;

$\frac{1}{\lambda_e}$: intervalle de temps moyen séparant deux arrivées consécutives ;

$\rho = \frac{\lambda_e}{\mu}$: taux d'occupation du système.

Les deux premières relations sont appelées **formules de little** qui sont des relations très générales qui s'appliquent à une grande classe de systèmes. Aucune hypothèse sur les variables aléatoires qui caractérisent le système (temps d'interarrivées, temps de service, etc...). La seule condition d'application de la loi de Little est que le système soit stable.

1.1.5 Analyse mathématique d'un système de files d'attente

L'étude mathématique d'un système de files 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 :

- les probabilités d'état $\mathbb{P}_n(t) = P(X(t) = n)$ qui définissent le régime transitoire du processus stochastique $X(t), t \geq 0$. Il est évident que les fonctions $\mathbb{P}_n(t)$ dépendent de l'état initial ou de la distribution initiale du processus.
- Le régime stationnaire du processus stochastique est défini par :

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

où, $\pi_{n(n \geq 0)}$ est appelée distribution stationnaire du processus $X(t), (t \geq 0)$.

Le calcul explicite du régime transitoire s'avère généralement pénible, voir impossible, pour la plupart des modèles donnés. On se contente donc de déterminer le régime stationnaire.

1.1.6 Quelques systèmes de files d'attente classiques

Modèles markoviens

La file M/M/1 : Une file d'attente M/M/1 est définie par le processus stochastique suivant : Une arrivée poissonnienne d'intensité λ et des temps de service qui sont indépendants du processus d'arrivée et suivent la loi exponentielle de paramètre μ . C'est le modèle qui a une seule file pour offrir un service.

Ce système réunit certaines conditions qui sont :



FIGURE 1.1 – La file M/M/1

- Les clients sont servis selon la règle FIFO, comme il est supposé que chaque client rejoint la file au moment de son arrivée et qu'il ne le quitte qu'après avoir été servi.
- Les processus d'arrivées des clients sont indépendants les uns des autres, mais le taux moyen d'arrivée est constant.
- L'arrivée aléatoire des clients suit la loi de Poisson de paramètre λ par unité de temps, ainsi que les autres qui reçoivent le service sont d'une population infinie ou très grande.
- La durée de service des clients diffère d'un client à autre et le temps moyen de service est connu et défini.
- Le temps de service suit la loi exponentielle de taux μ par unité de temps.
- Condition de stabilité : Le taux d'arrivée est inférieur au taux de service, c'est la condition de stabilité $\lambda < \mu$.

La file peut être considérée comme un processus de naissance et de mort pour lequel $\lambda_n = \lambda \quad \forall n \geq 0$

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

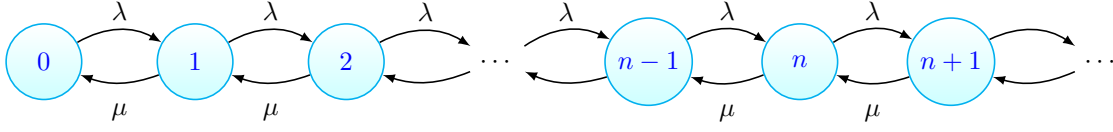


FIGURE 1.2 – CMTC associée à la file M/M/1.

• Régime stationnaire : Il est difficile d'étudier la variable aléatoire $N(t)$ représentant le nombre de clients au temps t dans le système. On s'intéresse alors à $N = \lim_{t \rightarrow \infty} N(t)$. On parle alors d'analyse en régime stationnaire (ou à l'équilibre). Le système d'équations linéaires et homogènes est le suivant :

$$\begin{aligned} \mu p_1 &= \lambda p_0 \\ (\lambda p_{n-1} + \mu) p_{n+1} &= (\lambda + \mu) p_n \end{aligned}$$

Avec

$$\sum_{n=0}^{\infty} p_n = 1$$

Sa résolution est élémentaire ; en additionnant les $(n + 1)$ premières équations, on obtient

$$\mu p_{n+1} = \lambda p_n$$

où

$$p_n = \left(\frac{\lambda}{\mu}\right)^n, \quad n = 1, 2, \dots$$

En se servant du fait que les p_n définissent une distribution de probabilité, on obtient finalement

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

à condition que $\lambda < \mu$. Le régime stationnaire du système d'attente M/M/1 est donc gouverné par la loi géométrique.

Lorsque $\lambda < \mu$, la CMTC associée est donc telle que tous ses états sont transitoires ou récurrents nuls. Dans ces conditions, les probabilités stationnaires $p(n)$ pour que la file contienne n clients sont toutes nulles ($p(n) = 0$ pour tout n), donc avec une

probabilité de 1 pour que la file contienne une infinité de clients. Cela correspond bien à la notion d'instabilité de la file.

La file M/M/c : Les modèles d'attente de type M/M/c sont des modèles où les clients arrivent selon un processus homogène de Poisson de taux λ . Le service des clients est assuré par c serveurs identiques et indépendants les uns des autres. Les durées de service suivent une loi exponentielle de taux μ (pour chacun des serveurs). Le taux de transition de l'état $n - 1$ de la CMTC associé est donc bien égal à $n\mu$. Lorsque $n > c$, seuls c clients sont en service et sont donc susceptibles de quitter la file, donc de faire passer le processus de l'état n à l'état $n - 1$. Le taux de transition correspondant est donc égal à $c\mu$, tandis qu'une transition d'un état n vers un état $n + 1$ correspondant à une arrivée de client, soit en un temps dt , à une probabilité $\lambda dt + o(dt)$. Le taux de transition est donc égal à λ . Par conséquent, la discipline d'attente est FIFO.

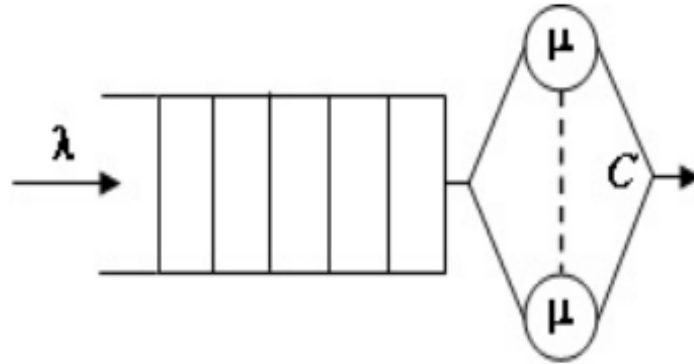


FIGURE 1.3 – La file M/M/c

L'état du système à la date t peut être décrit par le processus stochastique $N(t)$, $t \geq 0$ où $N(t)$ représente le nombre de clients dans le système à l'instant t . Ce dernier est un processus de naissance et de mort dont les taux de transitions sont

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

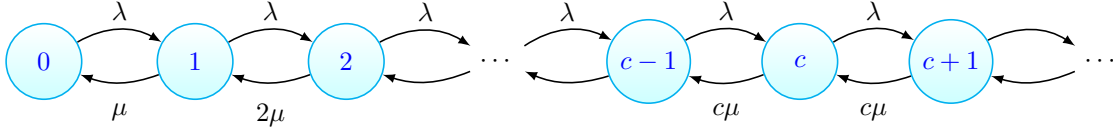


FIGURE 1.4 – CMTC associée à la file M/M/c

$$\mu_n = \begin{cases} 0, & \text{si } n = 0, \\ n\mu, & \forall n = 1, \dots, c, \\ c\mu, & \forall n \geq c. \end{cases}$$

Du diagramme on déduit les résultats qui suivent. L'analyse du système en régime stationnaire, à l'aide de la procédure des équations de Chapman-Kolmogorov aboutit aux équations suivantes :

$$\begin{aligned} \lambda p_0 &= \mu p_1 \\ (\lambda + n\mu)p_n &= \lambda p_{n-1} + (n+1)\mu p_{n+1} \quad 1 \leq n < c, \\ (\lambda + c\mu)p_n &= \lambda p_{n-1} + c\mu p_{n+1} \quad n \geq c, \end{aligned}$$

avec

$$\sum_{n=0}^{\infty} p_n = 1.$$

La résolution du système ci-dessus présente la distribution stationnaire suivante :

$$p_n = \frac{\rho^n}{n!} p_0; \quad 0 \leq n \leq c,$$

$$p_n = \frac{\rho^c}{c!} (A)^{n-c} p_0; \quad n \geq c,$$

où

$$p_0 = \left[\sum_{n=0}^{c-1} \frac{\rho^n}{n!} + \frac{\rho^c}{c!} \sum_{n=c}^{\infty} \rho^{n-c} \right]^{-1}, \quad \rho = \frac{\lambda}{\mu} \quad \text{et} \quad A = \frac{\lambda}{c\mu}$$

Cette dernière existe si $\lambda < c\mu$

File d'attente M/M/ ∞ : On considère un système composé d'un nombre illimité de serveurs identiques et indépendants les uns des autres. Dès qu'un client arrive, il rentre donc instantanément en service. Dans cette file particulière, il n'y a donc pas d'attente. On suppose toujours que le processus d'arrivée des clients est poissonien de taux μ et que les temps de service sont exponentiels de taux μ (pour tous les serveurs). Ce système est connu sous le nom de file M/M/ ∞ .

Condition de stabilité : De façon intuitive, la capacité de traitement de la file est infinie puisque tout nouveau client se présentant à l'entrée de la file est instantanément traité. La condition de stabilité exprimant que "le nombre moyen de clients arrivant à la file par unité du temps doit être inférieur à la capacité de traitement de la file" est donc toujours satisfaite.

Régime stationnaire : Soit p_n la probabilité stationnaire d'être dans l'état n . Les équations d'équilibre nous donnent

$$\lambda p_{n-1} = n\mu p_n$$

On peut écrire aussi

$$p_n = \frac{\rho}{n} p_{n-1}; \quad \forall n \geq 1;$$

où, $\rho = \frac{\lambda}{\mu}$.

On peut alors exprimer toutes les probabilités en fonction de p_0

$$p_n = \frac{\rho^n}{n!} p_0$$

Avec la condition de normalisation, on obtient p_0

$$\sum_{n=0}^{\infty} p_n = 1 \Rightarrow p_0 = e^{-\rho}.$$

On obtient finalement

$$p_n = \frac{\rho^n}{n!} e^{-\rho}.$$

1.1.7 Les systèmes de files d'attente complexes

Les systèmes d'attente complexes les plus élaborés sont :

- Les systèmes de files d'attente avec arrivées négatives
- Les systèmes de files d'attente avec vacances.
- Les systèmes de files d'attente avec rappel.
- Les systèmes de files d'attente avec serveurs non fiables.

Dans cette thèse, nous nous intéressons à l'étude des système avec rappel et serveurs non fiable afin d'avoir l'influence de la non-fiabilité des serveurs et l'impacience de client sur l'évaluation des caractéristiques du système.

1.2 Files d'attente avec rappels

1.2.1 Définition

La naissance des modèles de files d'attente avec rappels a été entraînée par des phénomènes d'appels répétés dans des réseaux informatiques et téléphoniques. Elles sont caractérisées par l'hypothèse de base suivante : le client qui arrive et trouve le serveur occupé doit rejoindre une file supplémentaire de clients appelée "orbite", et réessaye ultérieurement de faire des tentatives de service après une durée de temps aléatoire d'après une politique particulière de rappels. Si par contre à l'arrivée du client il y a un ou plusieurs serveurs libres et en bon état, le client sera servi immédiatement et quitte le système après son service. L'orbite peut être de capacité finie ou infinie, dans le premier cas, si l'orbite est pleine, le client doit quitter le système sans être servi dès qu'il trouve tous les serveurs et les positions d'attente occupés.

Ces modèles sont d'un usage courant dans la modélisation des systèmes téléphoniques, où un appel trouvant la ligne occupée rappelle ultérieurement après un temps aléatoire. On connaît des applications dans les systèmes informatiques et de communications, et aussi dans la modélisation de l'atterrissage des avions dans une piste où un avion ne trouvant pas la piste libre reste en "orbite" un temps aléatoire jusqu'à ce que la tour de contrôle lui donne le signal d'atterrissage. Il existe dans la littérature des modèles avec rappels essentiellement trois politiques de rappels : La politique linéaire classique, la politique constante et la politique versatile.

1.2.2 Politiques d'accès au serveur à partir de l'orbite

• **Politique linéaire classique :** Le protocole le plus décrit dans la théorie classique des files d'attente avec rappels est La politique de rappels linéaire classique, elle est caractérisée par le fait que chaque client en orbite engendre sa propre tentative de joindre le serveur indépendamment des autres clients en orbite. Un modèle illustratif est le modèle de type $M/G/1$ où les clients arrivent de l'extérieur selon un processus de Poisson de taux λ sachant qu'ils sont identifiés comme clients primaires. Si le serveur est libre à l'instant d'arrivée d'un client primaire, ce client obtient son service immédiatement et quitte le système après être servi. D'autre part, si le serveur est occupé à l'arrivée d'un client primaire, ce dernier doit rejoindre l'orbite. Chaque client en orbite engendre un processus de Poisson de taux μ de tentatives de joindre le serveur jusqu'à ce qu'il trouve le serveur libre pour prendre son service et quitter le système. Les clients primaires et ceux provenant de l'orbite ont la même distribution du temps de service. On note par σ_n le n -ème temps de service et on suppose que la suite σ_n est i.i.d. avec $0 < E\sigma_n < \infty$.

Sous des suppositions Markoviennes, la condition de stabilité de ce système est la même que le système $M/G/1$ classique (sans rappels)[27]

$$\lambda E\sigma_1 < 1$$

Politique de rappels constante : Cette politique a été introduite initialement par Fayolle [29] dans le cadre d'un modèle de type $M/M/1$, elle est caractérisée par le fait que l'orbite engendre un seul processus de poisson de taux θ et dès que ce processus trouve le serveur libre, un seul client en orbite (le premier en file où un autre choisi aléatoirement) reçoit son service et quitte le système. Ce qui veut dire que le taux de rappels ne dépend pas du nombre de clients en orbite.

La condition de stabilité de ce système de type $M/M/1$ dépend du taux de rappels θ et est donnée par

$$\lambda E\sigma_1 < \frac{\theta}{\lambda + \theta}$$

Il existe une variante de cette politique et qui est appelée politique de contrôle des rappels. Dans cette politique les clients en orbite se rangent en une file de capacité infinie selon un ordre "premier arrivé premier servis" FIFO et on permet seulement

au client à la tête de cette file de tenter de rejoindre le serveur. Elle est aussi assimilée à une politique de rappels où c'est le serveur qui cherche les clients en orbite à la fin d'un service d'un client. Cette politique de contrôle des rappels a été étudiée de manière intensive par Gomez-Corral [47] dans un cadre Markovien, i.e $M/GI/1$ avec supposition d'indépendance des temps de service.

Politique de rappels versatile : Cette politique est une combinaison des deux précédentes et a été introduite par Artalejo et Gomez-Corral [5]. L'orbite tente de trouver le serveur libre selon un processus de Poisson de taux θ indépendamment des clients en orbite et chaque client en orbite effectue sa propre tentative selon un processus de Poisson de taux μ . Le système est stable si l'une des conditions suivantes est satisfaite :

$$\begin{aligned} i) \quad & \theta > 0, \mu = 0 \quad \text{et} \quad \lambda E\sigma_1 < \frac{\theta}{\lambda + \theta} \\ ii) \quad & \theta \geq 0, \mu > 0 \quad \text{et} \quad \lambda E\sigma_1 < 1 \end{aligned}$$

Condition de stabilité d'un système de files d'attente avec rappels : Les files d'attente avec rappels se composent des clients primaires, des clients en orbite (clients secondaires) et d'un ou plusieurs serveurs. On considère un système de files d'attente avec rappels dont lequel le flot des arrivées des clients primaires suit un processus de Poisson du taux λ et la durée de service est de moyenne $\frac{1}{\mu}$.

$$\rho = \begin{cases} \frac{\lambda}{\mu} < 1, & \text{si la file est composé d'un seul serveur,} \\ \frac{\lambda}{c\mu} < 1, & \text{si la file est composé de } c \text{ serveurs.} \end{cases}$$

1.3 Interruption de service dans les systèmes d'attente avec rappels

Lors de l'étude d'un système de files d'attente, il est important de prendre en considération la non-fiabilité du serveur afin de voir l'impact des pannes sur les caractéristiques du système étudié (par exemple le nombre moyen de clients dans le système). L'influence de l'interruption de service ne peut pas être étudiée sans spécifier l'interaction entre le processus d'interruption et le processus de service. Cette interaction

peut être sans préemption, c'est-à-dire qu'elle n'intervient pas pendant le service d'un client, ou avec préemption dans le cas contraire. En outre, après chaque interruption avec préemption, le client reprend le service à partir du point d'interruption ou dès le début [19].

Les modèles de files d'attente à interruption de service ont plusieurs applications dans les réseaux de télécommunication et de nombreuses études ont été menées dans ce sens. En général, les interruptions peuvent être dues à diverses causes : la possibilité de vacance du serveur, les pannes du serveur ou l'arrivée de clients de haute priorité.

1.3.1 Quelques cas modélisés par des systèmes de files d'attente avec rappels

Il est bien connu que les files d'attente avec rappels apparaissent naturellement comme des modèles pratiques dans la vie quotidienne et dans les réseaux de communication, par exemple, en faisant des réservations, des réseaux de commutation de paquets, des accès CSMA non persistants (accès multiple du transporteur) et des systèmes en temps réel. Nous présentons quelques exemples de systèmes qui peuvent être modélisés comme une 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 la véritable volonté des abonnés au téléphone, et le flux de répétitions des appels, ce qui est la conséquence du manque de succès des précédentes tentatives. Les modèles standard 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 accessibles pour lui 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 rappel 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 de base basée sur la file d'attente $M/M/c$ (l'un des modèles de file d'attente les plus utilisés) Shekhar et al.(2016).

Exemple 2 : (Réseaux de communication par paquet)

Considérons un réseau de communications d'ordinateurs dans lequel on trouve un ensemble d'interfaces IMPs (Interface Message Processors) reliées entre elles par des câbles. Un ordinateur principal est connecté à l'une de ces interfaces. Si l'ordinateur veut envoyer un message à un autre ordinateur principal, il doit en premier lieu envoyer le message avec l'adresse de destination à l'interface à laquelle il est connecté. L'interface à son tour envoie le message à l'ordinateur destinataire directement si elle y est connectée, ou indirectement via d'autres interfaces. Considérons une interface à laquelle un ordinateur principal est connecté. Les messages arrivent de l'extérieur selon un processus aléatoire. Après la réception du message, l'ordinateur l'envoie immédiatement à son interface. S'il y a un tampon libre, le message est accepté. Dans le cas contraire, le message est rejeté et l'ordinateur doit réessayer une autre fois après une période de temps. S'il existe des tampons libres, le message rejeté sera stocké dans un tampon de l'ordinateur principal. Dans le cas contraire, le message est considéré comme perdu. On peut poser les questions suivantes :

- o Quelles sont les probabilités pour qu'un message soit rejeté par l'interface et par l'ordinateur principal ?
- o Quel est le nombre moyen de messages dans le tampon de IMP ?
- o Quel est le nombre moyen de messages dans le tampon de l'ordinateur principal ?
- o Quel est le temps d'attente d'un message dans le tampon de l'ordinateur principal ?

Le problème présenté peut être modélisé comme un système avec rappels à serveur unique (interface IMP) possédant des tampons (positions d'attente). Le nombre de tampons de l'ordinateur principal constitue la capacité de l'orbite.

Exemple 3 : (Système mémoire sur les disques magnétiques)

Considérons un système mémoire où K unités 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 égal 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.

Exemple 4 : (La radio cognitive)

La radio cognitive (CR) est actuellement l'une des technologies de transmission de l'information les plus prometteuses pour faire face au problème de la rareté du spectre et de la sous-utilisation du spectre dans les communications sans fil. Les réseaux CR visent à améliorer l'efficacité du spectre pour répondre aux demandes toujours croissantes des utilisateurs finaux. Le principe est de donner la possibilité aux utilisateurs sans licence (utilisateurs secondaires, SU) d'accéder temporairement et dynamiquement à la bande passante inutilisée ou peu utilisée tout en veillant à ce qu'elle n'interfère ni ne dégrade les performances des titulaires de licence titulaires, communément appelés utilisateurs principaux (PU). Cela soulève plusieurs défis à relever dans les réseaux CR et la performance des utilisateurs secondaires . Nous proposons de concevoir des réseaux CR comme un système de files d'attente avec rappels où les PU ont priorité sur les SU.

1.4 Files d'attente avec des clients impatientes

Un client serait impatient s'il a tendance à rejoindre la file d'attente seulement quand une courte attente est attendue et a tendance à rester dans la ligne si son attente a été suffisamment courte. Les clients impatientes, découragés par la longueur de la file, sont devenus le but de plusieurs études. Si nous parlons d'un point de vue commercial,

les clients impatients ont fait des pertes importantes à l'économie de plusieurs firmes (Ancker et Gafarian (1963)).

Des exemples pratiques des systèmes de files d'attente avec des clients impatients inclure des systèmes de télécommunication en temps réel dans lesquels les données reçues après une date butoir est inutile, les réseaux de télécommunications dans lesquels les abonnés abandonnent en raison de l'impatience avant que la connexion demandée ne soit établie, et les systèmes d'inventaire des produits périssables dans lesquels la durée de vie des produits expire est retiré de l'inventaire; voir (Baccelli (1984) et Boots (1999)). D'autres exemples peuvent être trouvés dans les centres d'appels, les systèmes d'inventaire de production et de nombreux autres domaines.

L'impatience prend généralement trois formes : la première '*balking*', décidant de ne pas du tout rejoindre la file d'attente à l'arrivée, la seconde '*reneging*', la réticence à rester dans la file d'attente après avoir rejoint et attendu et la troisième '*jockeying*' entre les lignes lorsque chacun d'un certain nombre de canaux de service parallèles a sa propre file d'attente (Gupta et Garg (2012)).

1.4.1 Revue de littérature

(Haight, (1957)) est le premier qui a étudié le modèle d'attente $M/M/1$ avec dérobade. (Haight (1959)) envisage une file d'attente $M/M/1$ avec des clients revenant à se produire lorsque le serveur n'est pas disponible pendant la panne du système ou les périodes de vacances. L'auteur se concentre sur la dérivation des solutions d'état stationnaire en utilisant une technique de variable supplémentaire et calcule la longueur moyenne de la file d'attente et le temps d'attente moyen. Les effets combinés du rechignage et du retour en arrière dans une file d'attente $M/M/1/N$ ont été étudiés par (Ancker et Gafarian (1963a), (1963b)). Pour la première fois, Cohen (1957) a envisagé un modèle prenant en compte la non-persistance des abonnés appelants pour la file d'attente avec rappel de type $M/M/c$. C'est une généralisation du modèle ordinaire avec des demandes impatientes et suppose que des sources d'appels répétés quitter le système avec une certaine intensité δ .

Boyer, Dupuis et Khelladi(1988), Fayolle et Brune (1988) ont envisagé un système à serveur unique avec une mémoire tampon illimitée. Tous les clients qui arrivent rejoignent la mémoire tampon et sont ensuite servis conformément à une discipline (telle que FIFO). Cependant, les clients sont impatientes, c'est-à-dire que leurs délais d'attente sont limités. Si le délai d'attente expire, le client en attente quitte la mémoire tampon. Après cela, avec une certaine probabilité, le client quitte le système pour toujours et avec une probabilité complémentaire tente à nouveau sa chance après un délai aléatoire. En supposant que toutes les variables aléatoires impliquées dans la description du modèle soient indépendantes et distribuées de manière exponentielle, le fonctionnement du système peut être décrit par un processus de Markov à deux variables $(C(t), N(t))$, où $C(t)$ est le nombre de clients dans la mémoire tampon (à la fois en attente et servi)

et $N(t)$ est le nombre des sources de appels répétés.

(Abou-El-Ata et Hariri (1992)) ont considéré un système de files d'attente $M/M/c/N$ à multi-serveurs avec balking et reneging.

N. K. Boots et H. Tijms (1999), ont considéré un système de files d'attente à multi-serveurs, qu'un client quitte lorsque son service n'a pas commencé dans un intervalle fixe après son arrivée. La probabilité de perte peut être exprimée dans une formule simple impliquant les probabilités de temps d'attente dans la file d'attente standard $M/M/c$.

Altman et Yechiali (E. Altman (2006)) analysent des modèles où les clients ne s'impatientent que lorsque le serveur est en vacances et indisponible pour le service. Les files d'attente $M/M/1$, $M/G/1$ et $M/M/c$ sont étudié et diverses mesures du rendement sous forme fermée sont obtenu.

Yechiali (U.Yechiali (2007)) étudie le modèle $M/M/c$ avec des clients impatientes, où les serveurs subissent des pannes désastreuses, entraînant la perte de tous les clients présents. Pendant qu'un processus de réparation est en cours, l'afflux de nouveaux clients se poursuit mais ils s'impatientent car aucun serveur n'est disponible.

(Chen et al. (2008)) ont étudié la file d'attente $M/M/m/k$ avec reprise préemptive et impatience des clients prioritaires et ils ont dérivé la distraction de la longueur de la file d'attente à l'état stationnaire et les mesures de performance en utilisant la méthode d'analyse matricielle.

(Patrick Wüchner and Hermann de Meer(2009)) ont étudié les files d'attente avec rappels de source finie avec plusieurs serveurs homogènes et recherche de clients sur l'orbite. Le principal objectif de l'analyse est de montrer l'effet du rechute, de l'impatience et de la taille du tampon sur les mesures de performance en régime permanent. Seedy, El-Sherbinya et Ammar (2009) ont étudié une analyse pour la file d'attente $M/M/c$ avec rechute et reniement. Les auteurs ont supposés que les clients arrivant rechignent avec une probabilité fixe et reviennent selon une distribution exponentielle négative.

(Perel et Yechiali (2010)) ont examiné le cas où l'impatience des clients est due à une phase de service lente, le serveur continue de fonctionner mais à un rythme plus lent qu'auparavant car il pourrait être occupé par d'autres tâches de priorité plus élevée, mais n'est pas totalement indisponible. Afin d'analyser le modèle, les auteurs envisagent une file d'attente $M/M/c$ fonctionnant dans un environnement aléatoire à 2 phases.

Kapodistria (2011) étudie une file d'attente markovienne à serveur unique avec clients impatients et considère les situations où les clients abandonnent le système simultanément. Il considère deux scénarios d'abandon. Dans le premier, tous les clients actuels deviennent impatients et effectuent des abandons synchronisés ; tandis que dans le deuxième scénario, le client en service est exclu de la procédure d'abandon. Kapodistria étend son analyse dans (Kapodistria (2011)) au File d'attente $M/M/c$ dans le deuxième scénario d'abandon également.

(Kumar et Sharma (2014)) ont présenté une étude d'un multi-serveurs Markovien modèle à capacité finie avec rechute, renégation et rétention des clients renégés.

Chapitre 2

The Phase Merging Algorithm

2.1 Introduction

Dans cette thèse, nous souhaitons obtenir la distribution limite de l'état du système de files d'attente avec rappels $M/M/s$, car d'importants descripteurs du système, tels que le nombre moyen de clients dans le système et le nombre moyen de clients faisant des tentatives répétées, peuvent être exprimés en termes de cette loi limite. Dans quelques modèles simples, il est possible de dériver des expressions de forme fermée mais très souvent il est difficile, voire impossible, de déterminer la distribution limite de l'état du système. Les méthodes algorithmiques et les techniques approximatives sont de puissants outils pour résoudre les problèmes analytiquement insolubles. Nous complétons l'analyse des performances des files d'attente avec rappels en étudiant une variété de mesures de qualité, y compris le temps d'attente, la période occupée et d'autres descripteurs spécifiques résultant de la fonction de rappels.

Habituellement, lors de l'étude des problèmes de la théorie des files d'attente classique ou celle avec rappels, on supposait que le service s'effectue d'une manière continue et bien déterminée (sans interruptions). Cependant en pratique, on rencontre souvent des cas où les serveurs sont sujets à des pannes aléatoires, ou le service d'un client est interrompu à cause de l'arrivée d'un client prioritaire .

L'objectif de notre travail consistera à l'évaluation des mesures de performances principales d'un modèle de file d'attente avec rappel $M/M/s$ à s serveurs non fiables et

des clients impatients en utilisant la méthode algorithmique de fusion de phase "The Phase merging Algorithm" .

2.2 L'approche APM

2.2.1 Introduction

Une solution analytique aux probabilités d'état des modèles de files d'attente à multi-serveurs non fiables est extrêmement difficile à obtenir, comme mentionné précédemment, aucun résultat n'existe pour les modèles avec plus de deux serveurs dans le cas fiable. les modèles peu fiables contribuent encore plus à la complexité analytique, principalement en raison de la préemption des clients rejoignant l'orbite.

Il existe plusieurs méthodes d'approximation qui ont été développées pour la résolution de modèles stochastiques lorsque la résolution mathématique d'un problème donné n'est pas possible. Cela permet de modéliser des situations très complexes dont la solution analytique s'avère difficile voir impossible. Parmi ces méthodes d'approximation on peut citer :

- La méthode de troncature directe,
- La méthode de troncature généralisée,
- L'approximation géométrique matricielle,
- L'approximation avec l'algorithme de fusion de phase (The Phase Merging Algorithm.)

La méthode de troncature directe : Dans cette méthode, on peut tronquer le système d'équations pour une valeur suffisamment grande du nombre de clients dans l'orbite, disons M . c'est-à-dire que la taille de l'orbite est limitée à M de sorte que tout client arrivant trouvant l'orbite pleine est considéré comme perdu. la valeur de M peut être choisie pour que la probabilité de perte soit petite. En raison de la nature intrinsèque du système, le seul choix disponible pour étudier M est à travers des méthodes algorithmiques. Alors qu'un certain nombre d'approches sont disponibles pour déterminer le point de coupure, M , celle qui semble bien fonctionner est d'augmenter M jusqu'à ce que le plus grand changement individuel dans les éléments de X pour les valeurs successives soit inférieur à une valeur infinitésimale prédéterminée.

L'utilisation d'un modèle tronqué pour la résolution numérique des équations de Kolmogorov pour le modèle d'origine avec un nombre illimité de sources a été suggérée par Wilkinson (1956). Il a également noté que les probabilités stationnaires du modèle tronqué peuvent être calculées récursivement à partir des équations de Kolmogorov, mais n'a pas donné d'algorithme détaillé. Le premier algorithme basé sur cette idée a été publié par Jonin et Sedol (1970a, 1970b, 1970c). Dans le livre Jonin et Sedol (1970c), ils ont publié des tableaux de performance caractéristiques de la file d'attente de reprise multiserveur avec des abonnés non persistants (modèle de Cohen). Falin (1983c) a suggéré une approximation à l'aide du modèle dans laquelle le taux de répétition est égal à l'infini lorsque le nombre de clients en orbite dépasse un certain niveau. Greenberg (1986) a suggéré une approximation à l'aide du modèle dans laquelle le taux de répétition reste constant lorsque le nombre de clients en orbite dépasse un certain niveau. Une analyse détaillée de ce modèle d'approximation est réalisée par Neuts et Rao (1990).

Les tentatives de calcul de l'erreur d'approximation de la file d'attente avec rappels par le modèle tronqué correspondant ou, de manière équivalente, par évaluer la limite de troncature M qui garantit l'erreur prédéfinie ont été effectuées pendant une longue période. Wilkinson (1956) a suggéré de choisir la limite de troncature M de telle sorte que la probabilité de perte dans le modèle tronqué soit égale à la probabilité $1 - H$ qu'un abonné appelant ne fasse plus tentative après échec de la tentative principale. Cette recommandation n'est pas naturelle car, dans le modèle original, les abonnés appelants sont supposés être absolument persistants. Cohen (1957) a étudié l'erreur numériquement et en a déduit certaines propriétés de la dépendance de l'erreur aux paramètres du modèle. Cependant, son enquête est limitée et ne donne aucune recommandation concernant le choix de la limite de troncature. Falin (1983d) a montré que, dans certains domaines de paramètres (trafic intense ou faible taux de nouvelles tentatives), cette méthode n'est pas correcte. Quelques estimations de l'erreur ont été obtenues par Stepanov (1983a).

L'Approximation géométrique matricielle :

Les méthodes d'analyse matricielle sont populaires comme outils de modélisation car elles donnent la possibilité de construire et d'analyser une large classe de modèles

de files d'attente de manière unifiée et exploitable par des algorithmes. La première procédure numérique, connue sous le nom de la méthode de la matrice géométrique, a été proposée par Neuts dans sa monographie (Voir Vladimir(1999)). Les méthodes proposées par Latouche et Ramaswami (1999), Naumov (2000) et Naumov et al.(1996) sont des versions améliorées de la méthode classique de la matrice géométrique. Dans ces études, une analyse profonde et une interprétation probabiliste concernant les processus QBD (Quasi Birth and Death) ont été effectuées. Latouche et Ramaswami ont proposé un algorithme pour le calcul de la matrice des taux qui a été amélioré dans la suite par Naumov et al.(1996). Ces méthodes sont devenues populaires et ont été largement appliquées dans plusieurs travaux.

La méthode de la matrice géométrique s'applique aux types de systèmes qui ont une certaine structure régulière. Particulièrement l'espace d'état du système est subdivisé en une infinité de niveaux dont chaque niveau est subdivisé en m phases. L'idée principale de cette méthode de résolution est que le vecteur de solution p possède une forme de matrice géométrique.

2.2.2 Description de l'algorithme de fusion de phase APM

"The Phase Merging Algorithm" est un outil mathématique qui élimine les complications ou la complexité dans la construction d'un modèle mathématique requis pour une description adéquate de systèmes réels (Neuts(1990)) en appliquant cet algorithme, un modèle simplifié du système est construit sur la base de l'espace de phase fusionné de l'état du système.

La classification et les propriétés de base des systèmes stochastiques sont considérés en utilisant une approche plutôt élémentaire. La simplification de tels systèmes est discutée en tant que principe de fusion de phases. Le schéma de fusion de phase consiste à scinder l'espace de phase des états et à fusionner les classes d'états distinctes dans les différents états fusionnés du système fusionné. Voir Figure 2.1

Dans notre modèle considéré, commençant par un CTMC sur un espace d'état qui décrit complètement un système de files d'attente avec rappels.

L'objectif de l'algorithme de fusion de phases est d'approximer la probabilité du régime permanent de distribution de $(N(t), B(t), F(t)) : t \geq 0$ where $N(t)$ is the random variable which represents the number of customers in the orbit at time t , $C(t)$

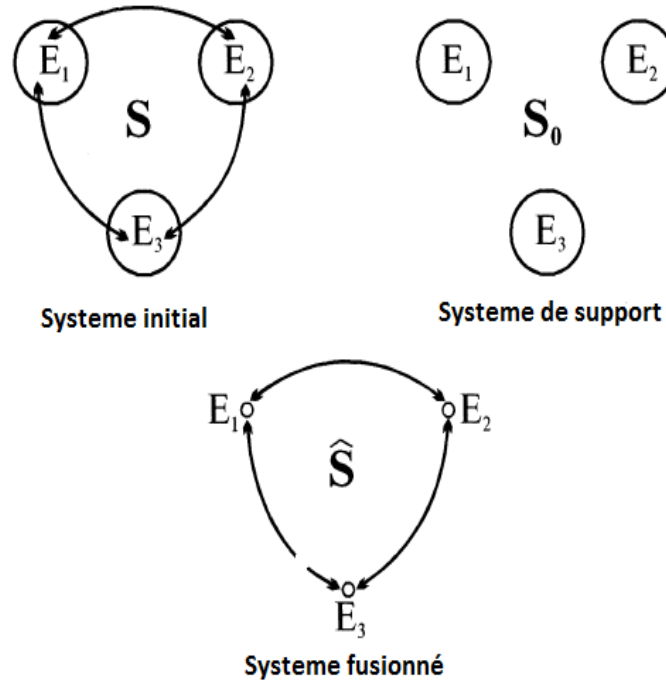


FIGURE 2.1 –

is the random variable which represents the number of customers in the service area at time t and $F(t)$ represents the number of failed servers at time t en approximant la distribution de probabilité conditionnelle du statut des serveurs et en approximant la distribution de probabilité marginale du nombre de clients en orbite. L'algorithme procède en partitionnant l'espace d'état en ensembles disjoints et mutuellement exhaustifs qui correspondent aux niveaux de l'orbite. Chaque niveau est analysé comme un CTMC à partir duquel les probabilités conditionnelles approximatives sont obtenues. Chaque niveau est ensuite considéré comme un état d'un CTMC agrégé où les taux de transition entre les niveaux correspondent aux clients entrant ou quittant l'orbite.

L'analyse de ce système de «macrostats» donne la distribution de probabilité marginale approximative du nombre de clients sur l'orbite. Le produit des probabilités conditionnelles et marginales est donc la probabilité conjointe approximative du niveau de l'orbite et de l'état des serveurs. En utilisant cette distribution conjointe,

nous approximations ensuite les mesures de performances de la file d'attente standard.

Chapitre 3

Approximate analysis of an unreliable $M/M/c$ retrial queue with phase merging algorithm

Ce chapitre fait l'objet d'une publication accepté dans *International Journal of New Trends in Mathematical Sciences*.

APPROXIMATE ANALYSIS OF AN UNRELIABLE $M/M/c$ RETRIAL QUEUE WITH PHASE MERGING ALGORITHM

M.ELHADDAD and F. LIMAM-BELARBI

Laboratory of Statistic and Processus Stochastic

University of Djillali Liabes

L.P 89, Sidi Bel Abbes 22000, Algeria

e-mail : *meriem.elhaddad61@gmail.com* and *faiza_belarbi@yahoo.fr*

Abstract : In this paper, we investigate an approximate analysis of unreliable $M/M/c$ retrial queue with $c \geq 3$ in which all servers are subject to breakdowns and repairs. Arriving customers that are unable to access a server due to congestion or failure can choose to enter a retrial orbit for an exponentially distributed amount of time and persistently attempt to gain access to a server, or abandon their request and depart the system. Once a customer is admitted to a service station, he remains there for a random duration until service is complete and then depart the system. However, if the server fails during service, i.e., an active breakdown, the customer may choose to abandon the system or proceed directly to the retrial orbit while the server begins repair immediately. In the unreliable model, there are no exact solutions when the number of servers exceeds one. Therefore, we seek to approximate the steady-state joint distribution of the number of customers in orbit and the status of the c servers for the case of Markovian arrival and service times. Our approach to deriving the approximate steady-state probabilities employs a phase-merging algorithm.

keywords : Retrial queue, Multi server, Breakdown and repair of service, Phase merging algorithm.

3.1 Introduction

Queueing systems in which arriving customers who find all servers and waiting positions (if any) occupied may retry for service after a period of time are called retrial queues. Retrial queues have been widely used to model many problems in telephone switching systems, telecommunication networks, computer networks and computer systems. For detailed survey of retrial queues and bibliographical information see Falin [21], Artalejo [8], [9], [11], monograph by Falin and Templeton [24] and Gomez-Corral [10]. Retrial queues with unreliable servers have been studied by Kulkarni and Choi [33], Aissani and Artalejo [2] and Brian [14]. There are a great number of numerical and approximations methods available. In this paper we will place more emphasis on the solutions by phase merging algorithm outlined by Korolyuk [28].

Our paper is organized as follows. In section 2, we provide the formal model description and state the assumptions that are needed to implement the approximation procedure. In sections 3 and 4, the algorithm is formally reviewed. In section 5, we give an illustrate example. Applying it to our model, we derive approximations for the steady-state probabilities and several standard queueing performance measures. In section 6, we assess the quality of the approximations by comparing results with those obtained using direct truncation method by M.G. Subramanian, Ayyappan and G. Sekar [39].

3.2 Model description

We consider an unreliable $M/M/c$ retrial queueing system in which customers arrive according to a Poisson process with rate λ ($\lambda > 0$). If upon arrival, the customer finds one of the servers idle and not failed, he occupies him immediately. However, if customer does not find any available servers (busy or failed) may join the retrial orbit with probability q_a or abandons the system with probability $1 - q_a$ ($0 \leq q_a \leq 1$). Customers who enter the orbit wait for an exponentially distributed time with rate θ ($\theta > 0$) before attempting to access a server again. The service times are assumed to be exponentially distributed with mean $1/\mu$. Failures for the c servers occur independently via a Poisson process with rate ξ ($\xi > 0$) and the repair times for each server are

exponentially distributed with rate α ($\alpha > 0$). Furthermore, interarrival times, service times, retrial times, interfailure times and repair times are mutually independent.

This model accounts for both active and idle breakdowns. For active breakdowns, the customer that is preempted by a server failure enters the retrial orbit with probability q_f or abandons the service with probability $1 - q_f$.

The state of the system can be described by a trivariate stochastic process in continuous time, $\{(N(t), B(t), F(t)) : t \geq 0\}$, where $N(t)$ is the number of customers in the orbit at time t , $B(t)$ is the number of busy servers at time t and $F(t)$ is the number of failed servers at time t .

Since all the random times are exponentially distributed, the stochastic process is a continuous-time Markov chain (CTMC) on the state space $S = \{(i, j, k) :$

$i \geq 0, j + k \leq c, j, k \in \{0, 1, 2, \dots, c\}\}$. We assume that as $t \rightarrow \infty$ the steady-state distribution of $\{(N(t), B(t), F(t)) : t \geq 0\}$ exists.

Define $p(i, j, k)$ as the limiting probability that the system is in the state (i, j, k) where $(i, j, k) \in S$. Defined mathematically,

$$p(i, j, k) = \lim_{t \rightarrow \infty} P(N(t) = i, B(t) = j, F(t) = k).$$

Note that a set of only $\sum_{n=1}^{c-1} n + (2c + 1)$ pairs of (j, k) are needed to completely characterize the status of the servers at any time. Let us introduce the following algorithm which gives the ordered pairs represents the number of busy servers and the number of failed servers, respectively.

Var c : Integer ;

begin

read (c) ;

for $j = 0$ to c Do

for $k = 0$ to $c - j$ Do

write (j, k) ;

end.

3.3 Main results

• Mean Orbit Length

For the retrial queueing model with c servers, it was shown that the steady-state distribution is Poisson with parameter $\hat{\lambda}/\hat{\theta}$. Therefore, the long-run mean orbit length is approximately the expected value of this Poisson random variable. Denoting N as the steady-state number of customers in orbit, the mean orbit size is approximated by

$$E[N] \approx \frac{\hat{\lambda}}{\hat{\theta}} \frac{\lambda q_a p_{0,c} + \sum_{\substack{j+k < c \\ j \neq 0}} j \xi q_f p_{j,k} + \sum_{\substack{j+k = c \\ j \neq 0}} (\lambda q_a + j \xi q_f) p_{j,k}}{\theta \sum_{j+k \neq c} p_{j,k}}.$$

• Mean Number of customer in Service

Let N_s be defined as the random number of customers at the servers, the approximate expression for the expected number of customers at the servers can be computed using the approximate steady-state joint probabilities derived in the last step of the algorithm and it is given by

$$E[N_s] = \sum_{i=0}^{\infty} \left(\sum_{j \neq 0} p(i, j, k) \right)$$

• Steady-State System Size and Sojourn Time

Let us define L as the steady-state number of customers in the system, to calculate it, we simply sum the expressions for $E[N]$ and $E[N_s]$.

The steady-state mean sojourn time, W , follows directly from Little's law.

$$\begin{aligned} L &\approx E[N] + E[N_s] \\ W &\approx \frac{L}{\lambda}. \end{aligned}$$

• Total Expected Time in Orbit

Due to server failure and blocking when making a retrial attempt, customers may enter the orbit more than once. Therefore, the expected time a customer spends in orbit is $1/\theta$ times the expected number of retrial attempts before gaining access to the server. Define Y as the random number of retrials a customer performs until it gains access to a server. Then Y is a geometric random variable with parameter p_u , the steady-state probability that at least one server is available. The approximation for p_u is given by

$$p_u = \sum_{i=0}^{\infty} \left(\sum_{j+k \neq c} p(i, j, k) \right).$$

The expected number of retrials performed, $E[Y]$, is therefore, $1/p_u$ and letting W_r be the random time spent in orbit once they are there we have,

$$E[W_r] \approx (\theta p_u)^{-1}.$$

3.4 Proof

To prove these results, a phase-merging algorithm developed by Korolyuk [28] and Courtois [17] will be employed and is summarized in the following sections :

3.4.1 The Phase-Merging Algorithm

Beginning with a CTMC on a state space that completely describes a retrial queueing system. The objective of the phase-merging algorithm is to approximate the steady-state probability distribution of $\{(N(t), B(t), F(t)) : t \geq 0\}$ by approximating the conditional probability distribution of the status of the servers and by approximating the marginal probability distribution of the number of customers in orbit. The algorithm proceeds by partitioning the state space into disjoint and mutually exhaustive sets that correspond to levels of the orbit. Each level is analysed as a CTMC from which the approximate conditional probabilities are obtained. Each level itself is subsequently considered as a state of an aggregated CTMC where the transition rates

between levels correspond to customers entering or leaving the orbit. Analyzing this system of 'macrostates' yields the approximate marginal probability distribution of the number of customers in the orbit. The product of the conditional and marginal probabilities is, therefore, the approximate joint probability distribution of the level of the orbit and status of the servers. Using this joint distribution, we then approximate standard queueing performance measures.

To begin, we reduce the dimensionality of the state space by defining $X(t)$ as the status of the servers at time t (outlined in Table 1), such that $X(t) \in \{1, 2, 3, 4, \dots, \sum_{n=1}^{c-1} n + (2c + 1)\}$.

State (j, k)	Index
(0,0)	1
(1,0)	2
(2,0)	3
.	.
.	.
.	.
(c,0)	$c + 1$
.	.
.	.
.	.
(0,1)	
(0,2)	
.	.
.	.
.	.
(0,c)	$\sum_{n=1}^{c-1} n + (2c + 1)$

TABLE 3.1 – Substitution for server status

The algorithm proceeds in the following manner. First, partition the state space, S , into disjoint sets that are conditional upon i such that,

$$S \equiv \bigcup_{i=0}^{\infty} S_i, \quad S_i \cap S_j = \emptyset, \quad i \neq j$$

where $S_i = (i, l) : l = 1, 2, \dots, \sum_{n=1}^{c-1} n + (2c + 1), i \geq 0$. This step results in an infinite number of classes (or levels) which can be analyzed individually.

Next we obtain the steady-state distribution of each class or level, S_i , by determining the infinitesimal generator matrix, Q_i defined by

$$q_{l,m} = \begin{cases} q_{(i,l),(i,m)} & l \neq m \\ -\sum_{l \neq m} q_{(i,l),(i,m)} & l = m \\ 0 & \text{otherwise} \end{cases}$$

Denote by $q_{l/i}$ the steady-state conditional probability that the status of the servers is state l , given there are i customers in orbit, $i \geq 0, l = 1, 2, \dots, \sum_{n=1}^{c-1} n + (2c + 1)$. Letting $p_i = [p_{l/i}]$, we solve the system of equations $p_i Q_i = 0$ and $p_i e = 1$ (where e is a column vector of ones) to obtain the approximate probability distribution.

Following this, we merge, or aggregate, all states within the class S_i , into one state corresponding to the level of orbit, i . These 'macrostates' form the overall state space of the merged model which are defined as $\hat{S} \equiv \{i : i \geq 0\}$. The infinitesimal generator, Q_M , of the merged model is

$$q_{i,j} \equiv \sum_{(i,l) \in S_i} p_{l/i} \left(\sum_{(j,m) \in S_j} q_{(i,l),(i,m)} \right)$$

Denote π_i as the marginal probability that there are i customers in the orbit. Letting the infinite-dimensional vector $\pi \equiv [\pi_0, \pi_1, \pi_2, \dots]$, we solve the system of equations

$\pi Q_M = 0$ and $\pi e = 1$ to obtain the approximate steady-state marginal probabilities.

$$p(i, l) \approx \hat{p}(i, l) = p_{l|i} \times \pi_i \quad i \geq 0, \quad l = 0, 1, 2, \dots, \sum_{n=1}^{c-1} n + (2c + 1)$$

Making use of these joint probabilities, we can obtain approximations for the performance measures for the unreliable c -servers retrial queue.

To illustrate the algorithm, we consider a reliable $M/M/c$ retrial queue whose customers arrive according to a Poisson process with rate λ . Each customer brings an exponential service requirement with mean time $1/\mu$.

Customers who find the c servers busy enter the orbit with probability s or leave the system with probability $1 - s$. The time between retrials is exponentially distributed with mean $1/\theta$. All times are assumed to be mutually independent. Define the continuous-time stochastic process as $\{(N(t), B(t)) : t \geq 0\}$ where $N(t)$ is the number of customers in the orbit at time t and $B(t)$ is the number of busy servers at time t . The process is a CTMC on the state space $S = \{(i, j) : i \geq 0, j = 0, 1, 2, \dots, c\}$. For the purpose of illustrating the phase-merging algorithm, we will assume the system is stable and denote $p(i, j) = \lim_{t \rightarrow \infty} P(N(t) = i, B(t) = j)$ as the limiting probability that the system is in state $(i, j), i \geq 0, j = 0, 1, 2, \dots, c$. Figure 1.

To make use of the algorithm we proceed as follows : First, partition the state space into individual levels where the index of each level corresponds to the number of customers in the orbit. Denote this as class S_i for level $i, i \geq 0$. Note that each class has an identical structure and, therefore, the generator matrices, $\mathbf{Q}_i$ are identical for all $i \geq 0$. This fact will be extremely useful for analyzing the case of unreliable servers. Next we compute the steady-state conditional distribution of the status of the servers given there are i customers in orbit. Denote these probabilities by $p_{j|i}, j = 0, 1, 2, \dots, c$, and it is easy to obtain the following conditional probabilities for all $i \geq 0$ by solving the balance equation system using the Formal Maple Calculation Software.

The next step is to aggregate the states of each class to form a series of merged states, i where $i \geq 0$ and investigate the transitions between them. The elements of

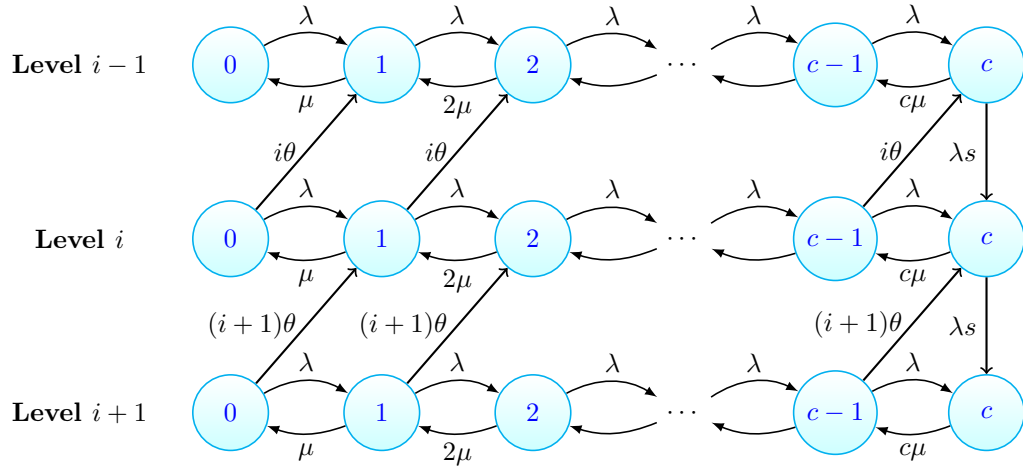


FIGURE 3.1 – Transition rate diagram for a reliable $M/M/c$ retrial queue.

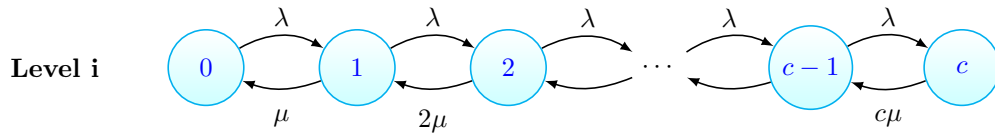


FIGURE 3.2 – Transition rate diagram for level i : reliable $M/M/c$ retrial queue .

the infinitesimal generator matrix for the merged states are :

$$q_{i,j} = \begin{cases} \lambda s p_{c|i} & i \geq 0, \quad j = i + 1 \\ i\theta \left(\sum_{j=0}^{c-1} p_{j|i} \right) & i \geq 0, \quad j = i - 1 \\ -[\lambda s p_{c|i} + i\theta \left(\sum_{j=0}^{c-1} p_{j|i} \right)] & i = j \\ 0 & \text{otherwise} \end{cases}$$

Using the substitutions, $\hat{\lambda} = \lambda s p_{c|i}$ and $\hat{\theta} = \theta \left(\sum_{j=0}^{c-1} p_{j,k} \right)$, we see that the analysis of this system is analogous to the $M/M/\infty$ queueing system. Thus, defining the steady-state marginal probability vector as $\pi = [\pi_0, \pi_1, \pi_2, \dots]$ we have,

$$\pi_i = \frac{1}{i!} \left(\frac{\lambda}{\theta} \right)^i e^{-\lambda/\theta}, \quad i \geq 0$$

Finally, the approximate steady-state distribution of $\{(N(t), B(t)) : t \geq 0\}$ is given by

$$\begin{aligned} p(i, j) &\approx \hat{p}(i, j) = p_{j|i} \times \pi_i \\ &= \frac{p_{j|i}}{i!} \left(\frac{\lambda}{\theta} \right)^i e^{-\lambda/\theta} \quad i \geq 0, \quad j = 0, 1, 2, \dots, c. \end{aligned}$$

3.4.2 Approximation Using the Phase-Merging Algorithm

We apply the phase-merging algorithm described in [28] and [17] to the unreliable $M/M/c$ retrial queue. Recall that the interarrival times, service and repair times, time between failures and time between retrials are all exponentially distributed with the parameters defined previously. Since the number of customers in the orbit can theoretically reach infinity, the state space of the system can be partitioned into a countable number of classes. As noted previously, the state space S is partitioned as the countable union

$$S = \bigcup_{i=0}^{\infty} S_i, \quad S_i \cap S_j = \emptyset, \quad i \neq j$$

where $S_i = (i, l) : l = 1, 2, \dots, \sum_{n=1}^{c-1} n + (2c + 1), i \geq 0$. Just as in the reliable $M/M/c$ retrial queue, each class is identical in structure so that only one class needs to be analyzed.

To determine the steady-state probabilities for the class S_i define the stochastic process $\{(B(t), F(t)) : t \geq 0\}$ where $B(t)$ represents the number of busy servers and $F(t)$ represents the number of failed servers at time t . Clearly, the process is a **CTMC** on the state space E defined previously. Using the notation defined in Table 1 we denote $p_{l|i}$ as the limiting conditional probability of the servers being in state l given that there are i customers in orbit,

$$p_{l|i} = \lim_{t \rightarrow \infty} P(X(t) = l / N(t) = i), l = 1, 2, \dots, \sum_{n=1}^{c-1} n + (2c + 1)$$

For each $i \geq 0$, the transition rates for this process are described in the generator matrix Q_i .

Let p_i be the steady-state conditional probability vector where $p_i = [p_{l|i}]$,

$$l = 1, 2, \dots, \sum_{n=1}^{c-1} n + (2c + 1).$$

Solving the equations $p_i Q_i = 0$ and $p_i e = 1$ yields the following system,

$$\begin{aligned}
(\lambda + c\xi)p_{(0,0)|i} &= \mu p_{(1,0)|i} + \alpha p_{(0,1)|i} \\
(\lambda + (c-1)\xi + \mu)p_{(1,0)|i} &= \lambda p_{(0,0)|i} + 2\mu p_{(2,0)|i} + \alpha p_{(1,1)|i} \\
(\lambda + (c-2)\xi + 2\mu)p_{(2,0)|i} &= \lambda p_{1,0|i} + 3\mu p_{(3,0)|i} + \alpha p_{(2,1)|i} \\
&\vdots \\
&\vdots \\
(\lambda + \xi + (c-1)\mu)p_{(c-1,0)|i} &= \lambda p_{(c-2,0)|i} + c\mu p_{(c,0)|i} + \alpha p_{(c-1,1)|i} \\
c\mu p_{(c,0)|i} &= \lambda p_{(c-1,0)|i} \\
(\lambda + (c-(j+k))\xi + j\mu + k\alpha)p_{j,k|i} &= (j+1)\mu p_{j+1,k|i} + (k+1)\alpha p_{j,k+1|i} + \lambda p_{j-1,k|i} \\
&\quad + (c-(j+k-1))\xi p_{j,k-1|i} \quad \text{with } j+k < c, j \neq 0, k \neq 0 \\
(j\mu + k\alpha)p_{j,k|i} &= \lambda p_{j-1,k|i} + (c-(j+k-1))\xi p_{j,k-1|i} \quad \text{with } j+k = c, j \neq 0, k \neq 0 \\
(\lambda + (c-1)\xi + \alpha)p_{(0,1)|i} &= c\xi p_{(0,0)|i} + \mu p_{(1,1)|i} + 2\alpha p_{(0,2)|i} \\
(\lambda + (c-2)\xi + 2\alpha)p_{(0,2)|i} &= (c-1)\xi p_{0,1|i} + \mu p_{(1,2)|i} + 3\alpha p_{(0,3)|i} \\
(\lambda + (c-3)\xi + 3\alpha)p_{(0,3)|i} &= (c-2)\xi p_{0,2|i} + \mu p_{(1,3)|i} + 4\alpha p_{(0,4)|i} \\
&\vdots \\
&\vdots \\
&\vdots \\
(\lambda + \xi + (c-1)\alpha)p_{(0,c-1)|i} &= 2\xi p_{0,c-2|i} + \mu p_{(1,c-1)|i} + c\alpha p_{(0,c)|i} \\
c\alpha p_{(0,c)|i} &= \xi p_{(0,c-1)|i} \\
\sum_{l=1}^{10} p_{l|i} &= 1
\end{aligned}$$

The solutions to the conditional probabilities are obtained for all $i \geq 0$ by solving the balance equations system using the Formal Maple Calculation Software.

Aggregating the states of each class S_i yields a system of macro-states which denote as i , ($i \geq 0$). The rates of transitions between the "macro-states" are expressed in the infinitesimal generator matrix with elements,

$$q_{i,j} = \begin{cases} \lambda q_a p_{0,c} + \sum_{\substack{j+k < c \\ j \neq 0}} j \xi q_f p_{j,k} + \sum_{\substack{j+k=c \\ j \neq 0}} (\lambda q_a + j \xi q_f) p_{j,k} & i \geq 0, \quad j = i + 1 \\ i \theta \left(\sum_{j+k \neq c} p_{j,k} \right) & i \geq 0, \quad j = i - 1 \\ \lambda q_a p_{0,c} + \sum_{\substack{j+k < c \\ j \neq 0}} j \xi q_f p_{j,k} + \sum_{\substack{j+k=c \\ j \neq 0}} (\lambda q_a + j \xi q_f) p_{j,k} + i \theta \left(\sum_{j+k \neq c} p_{j,k} \right) & i = j \\ 0 & \text{otherwise .} \end{cases}$$

To simplify the analysis of the merged states we use the following substitutions for $i, i \geq 0$

$$\begin{aligned} \hat{\lambda} &= \lambda q_a p_{0,c} + \sum_{\substack{j+k < c \\ j \neq 0}} j \xi q_f p_{j,k} + \sum_{\substack{j+k=c \\ j \neq 0}} (\lambda q_a + j \xi q_f) p_{j,k} \\ \hat{\theta} &= \theta \sum_{j+k \neq c} p_{j,k} \end{aligned}$$

Making the substitutions, the elements of the generator matrix are,

$$q_{i,j} = \begin{cases} \hat{\lambda} & i \geq 0, \quad j = i + 1 \\ i \hat{\theta} & i \geq 0, \quad j = i - 1 \\ -[\hat{\lambda} + i \hat{\theta}] & i = j \\ 0 & \text{otherwise} \end{cases}$$

This new model is a state dependent birth-and-death process, the analysis of which is analogous to the $M/M/\infty$ queueing system. Using the method of arc cuts, we recursively solve for the steady-state probability vector, $\pi = [\pi_0, \pi_1, \pi_2, \dots]$

$$\begin{aligned} \hat{\lambda} \pi_0 &= \hat{\theta} \pi_1 \Rightarrow \pi_1 = \frac{\hat{\lambda}}{\hat{\theta}} \pi_0 \\ \hat{\lambda} \pi_1 &= 2 \hat{\theta} \pi_2 \Rightarrow \pi_2 = \frac{1}{2} \left(\frac{\hat{\lambda}}{\hat{\theta}} \right)^2 \pi_0 \\ \hat{\lambda} \pi_2 &= 3 \hat{\theta} \pi_3 \Rightarrow \pi_3 = \frac{1}{6} \left(\frac{\hat{\lambda}}{\hat{\theta}} \right)^3 \pi_0 \\ \hat{\lambda} \pi_3 &= 4 \hat{\theta} \pi_4 \Rightarrow \pi_4 = \frac{1}{24} \left(\frac{\hat{\lambda}}{\hat{\theta}} \right)^4 \pi_0 \end{aligned}$$

Continuing inductively, it can easily be shown that,

$$\pi_i = \frac{1}{i!} \left(\frac{\hat{\lambda}}{\hat{\theta}} \right)^i \pi_0, \quad (i \geq 0). \quad (3.4.1)$$

Using the normalization equation , $\sum_{j=0}^{\infty} \pi_j = 1$, the solution for π_0 is obtained by

$$\begin{aligned} \pi_0 + \frac{1}{2} \left(\frac{\hat{\lambda}}{\hat{\theta}}\right)^2 \pi_0 + \frac{1}{6} \left(\frac{\hat{\lambda}}{\hat{\theta}}\right)^3 \pi_0 + \dots &= 1 \\ \pi_0 \left(1 + \frac{\hat{\lambda}}{\hat{\theta}} + \frac{1}{2} \left(\frac{\hat{\lambda}}{\hat{\theta}}\right)^2 + \frac{1}{6} \left(\frac{\hat{\lambda}}{\hat{\theta}}\right)^3 \pi_0 + \dots\right) &= 1 \end{aligned}$$

$$\pi_0 \left(\sum_{j=0}^{\infty} \frac{1}{j!} \left(\frac{\hat{\lambda}}{\hat{\theta}}\right)^j \right) = 1. \quad (3.4.2)$$

The infinite series of (4.4.2) is the Maclaurin power series expansion for $e^{-\hat{\lambda}/\hat{\theta}}$. Thus, we see that

$$\pi_0 = e^{\hat{\lambda}/\hat{\theta}}$$

Substituting π_0 into equation (4.4.1) we have the following expression,

$$\pi_i = \frac{1}{i!} \left(\frac{\hat{\lambda}}{\hat{\theta}}\right)^i e^{-\hat{\lambda}/\hat{\theta}}$$

which is the probability mass function for a poisson distributed random variable with rate parameter $\frac{\hat{\lambda}}{\hat{\theta}}$.

Finally, we approximate the steady-state distribution of $\{(N(t), X(t)) : t \geq 0\}$ by

$$\begin{aligned} p(i, l) \approx \hat{p}(i, l) &= p_{l|i} \times \pi_i \\ &= \frac{p_{l|i}}{i!} \left(\frac{\lambda}{\theta}\right)^i e^{-\lambda/\theta} \quad i \geq 0, \quad l = 0, 1, 2, \dots, \sum_{n=1}^{c-1} n + (2c + 1). \end{aligned}$$

3.5 Illustrative example : $M/M/3$.

We reduce the dimensionality of the state space by defining $X(t)$ as the status of the servers at time t , such that $X(t) \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. For each $i \geq 0$, the transition rates for this process are described in the following generator matrix Q_i .

State (j,k)	Index
(0,0)	1
(1,0)	2
(2,0)	3
(3,0)	4
(1,1)	5
(2,1)	6
(1,2)	7
(0,1)	8
(0,2)	9
(0,3)	10

TABLE 3.2 – Substitution for server status

$$Q_i = \begin{pmatrix} -(\lambda + 3\xi) & \lambda & 0 & 0 & 0 & 0 & 0 & 3\xi & 0 & 0 \\ \mu & -(\lambda + \mu + 2\xi) & \lambda & 0 & 2\xi & 0 & 0 & 0 & 0 & 0 \\ 0 & 2\mu & -(\lambda + 2\mu + \xi) & \lambda & 0 & \xi & 0 & 0 & 0 & 0 \\ 0 & 0 & 3\mu & -3\mu & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \alpha & 0 & 0 & -(\lambda + \mu + \alpha + \xi) & \lambda & \xi & \mu & 0 & 0 \\ 0 & 0 & \alpha & 0 & 2\mu & -(\alpha + 2\mu) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2\alpha & 0 & -(2\alpha + \mu) & 0 & \mu & 0 \\ \alpha & 0 & 0 & 0 & \lambda & 0 & 0 & -(\lambda + 2\xi) & 2\xi & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda & 2\alpha & -(\lambda + \xi + 2\alpha) & \xi \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3\alpha & -3\alpha \end{pmatrix}$$

Let p_i be the steady-state conditional probability vector where $p_i = [p_{l|i}]$,

$l = 1, 2, \dots, 10$

Solving the equations $p_i Q_i = 0$ and $p_i e = 1$ yields the following system,

$$(\lambda + 3\xi)p_{1|i} = \mu p_{2|i} + \alpha p_{8|i} \quad (3.5.1)$$

$$(\lambda + 2\xi + \mu)p_{2|i} = \lambda p_{1|i} + 2\mu p_{3|i} + \alpha p_{5|i} \quad (3.5.2)$$

$$(\lambda + \xi + 2\mu)p_{3|i} = \lambda p_{2|i} + 3\mu p_{4|i} + \alpha p_{6|i} \quad (3.5.3)$$

$$3\mu p_{4|i} = \lambda p_{3|i} \quad (3.5.4)$$

$$(\lambda + \xi + \mu + \alpha)p_{5|i} = 2\mu p_{6|i} + 2\alpha p_{7|i} + \lambda p_{8|i} + 2\xi p_{2|i} \quad (3.5.5)$$

$$(2\mu + \alpha)p_{6|i} = \lambda p_{5|i} + \xi p_{3|i} \quad (3.5.6)$$

$$(\mu + 2\alpha)p_{7|i} = \lambda p_{9|i} + 2\xi p_{5|i} \quad (3.5.7)$$

$$(\lambda + 2\xi + \alpha)p_{8|i} = 3\xi p_{1|i} + \mu p_{5|i} + 2\alpha p_{9|i} \quad (3.5.8)$$

$$(\lambda + \xi + 2\alpha)p_{9|i} = 2\xi p_{8|i} + \mu p_{7|i} + 3\alpha p_{10|i} \quad (3.5.9)$$

$$3\alpha p_{10|i} = \xi p_{9|i} \quad (3.5.10)$$

$$\sum_{l=1}^{10} p_{l|i} = 1. \quad (3.5.11)$$

Replacing Equation (3.5.9) with the normalization equation (3.5.11), the solution to the conditional probabilities are obtained for all $i \geq 0$, using the Maple Calculation Software

$$\begin{aligned} p_{1/i} = & D^{-1}(6(-4\alpha\xi^3 + 2\mu\xi^3 - 2\xi^2\alpha\lambda + 5\mu\xi^2\lambda - 4\alpha^2\xi^2 - 4\alpha\mu\xi^2 + 7\mu^2\xi^2 + 7\mu^3\xi + 4\lambda^2\mu\xi + 9\lambda\alpha\mu\xi \\ & + 10\alpha\mu^2\xi + 4\alpha^2\mu\xi + 11\lambda\mu^2\xi + 4\lambda\mu^3 + 3\lambda^2\mu^2 + 7\alpha^2\mu^2 + 9\alpha\mu^2\lambda + 2\alpha^3\mu + 7\alpha\mu^3 + \lambda^3\mu \\ & + 4\lambda^2\alpha\mu + 5\alpha^2\mu\lambda + 2\mu^4)\alpha^3\mu^2) \end{aligned}$$

$$\begin{aligned} p_{2/i} = & D^{-1}(6(2\mu^3 + 9\mu^2\xi + 7\alpha\mu^2 + 4\lambda\mu^2 + 16\mu\xi^2 + 15\alpha\mu\xi + 15\lambda\mu\xi + 9\lambda\alpha\mu + 7\mu\alpha^2 + 3\mu\lambda^2 + 6\xi^3 \\ & + \lambda^3 + 11\lambda\xi^2 + 10\alpha\xi^2 + 14\alpha\lambda\xi + 6\lambda^2\xi + 6\alpha^2\xi + 5\alpha^2\lambda + 4\lambda^2\alpha + 2\alpha^3)\lambda\alpha^3\mu^2) \end{aligned}$$

$$\begin{aligned} p_{3/i} = & D^{-1}(3(2\mu^3 + 4\lambda\mu^2 + 11\mu^2\xi + 7\alpha\mu^2 + 9\lambda\alpha\mu + 18\mu\lambda\xi + 20\alpha\mu\xi + 3\mu\lambda^2 + 7\mu\alpha^2 + 24\mu\xi^2 \\ & + 5\alpha^2\lambda + 8\alpha^2\xi + 17\alpha\lambda\xi + 7\lambda^2\xi + 4\lambda^2\alpha + 16\lambda\xi^2 + 18\alpha\xi^2 + 2\alpha^3 + \lambda^3 + 12\xi^3)\lambda^2\alpha^3\mu) \end{aligned}$$

$$\begin{aligned} p_{4/i} = & D^{-1}((2\mu^3 + 4\lambda\mu^2 + 11\mu^2\xi + 7\alpha\mu^2 + 9\lambda\alpha\mu + 18\mu\lambda\xi + 20\alpha\mu\xi + 3\mu\lambda^2 + 7\mu\alpha^2 + 24\mu\xi^2 + 5\alpha^2\lambda \\ & + 8\alpha^2\xi + 17\alpha\lambda\xi + 7\lambda^2\xi + 4\lambda^2\alpha + 16\lambda\xi^2 + 18\alpha\xi^2 + 2\alpha^3 + \lambda^3 + 12\xi^3)\lambda^3\alpha^3) \end{aligned}$$

$$\begin{aligned} p_{5/i} = & D^{-1}(6(6\mu^3 + 27\mu^2\xi + 19\alpha\mu^2 + 10\lambda\mu^2 + 44\alpha\mu\xi + 18\lambda\alpha\mu + 28\mu\lambda\xi + 5\mu\lambda^2 + 16\mu\alpha^2 + 36\mu\xi^2 \\ & + 16\alpha^2\xi + \lambda^3 + 24\alpha\xi^2 + 16\lambda\xi^2 + 12\xi^3 + 22\alpha\lambda\xi + 7\lambda^2\xi + 8\alpha^2\lambda + 5\lambda^2\alpha + 4\alpha^3)\alpha^2\mu^2\xi\lambda) \end{aligned}$$

$$\begin{aligned} p_{6/i} = & D^{-1}(3(6\mu^3 + 10\lambda\mu^2 + 27\mu^2\xi + 17\alpha\mu^2 + 28\mu\lambda\xi + 36\alpha\mu\xi + 36\mu\xi^2 + 5\mu\lambda^2 + 15\alpha\lambda\mu + 11\mu\alpha^2 \\ & + 5\alpha^2\lambda + \lambda^3 + 8\alpha^2\xi + 17\alpha\lambda\xi + 7\lambda^2\xi + 4\lambda^2\alpha + 16\lambda\xi^2 + 18\alpha\xi^2 + 2\alpha^3 + 12\xi^3)\xi\mu\lambda^2\alpha^2) \end{aligned}$$

$$p_{7/i} = D^{-1}(3(12\mu^3 + 42\mu^2\xi + 14\lambda\mu^2 + 42\alpha\mu^2 + 34\mu\alpha^2 + 33\mu\lambda\xi + 42\mu\xi^2 + 84\alpha\mu\xi + 33\lambda\alpha\mu + 6\mu\lambda^2 + 8\alpha^3 + 32\alpha\lambda\xi + 7\lambda^2\xi + 36\alpha\xi^2 + 14\alpha^2\lambda + 32\alpha^2\xi + 7\lambda^2\alpha + 12\xi^3 + 16\lambda\xi^2 + \lambda^3)\alpha\lambda\xi^2\mu^2$$

$$p_{8/i} = D^{-1}(6\alpha^2(13\alpha\mu\lambda\xi + 24\mu^2\lambda\xi + 6\xi^3\mu + 6\lambda^2\mu\xi + 11\lambda\xi^2\mu - 2\lambda^2\xi\alpha - 10\alpha\lambda\xi^2 + 21\mu^3\xi + 5\lambda^2\mu^2 + 21\mu^2\xi^2 + 7\mu\lambda^2\alpha - 12\alpha\xi^3 + \mu\lambda^3 + 6\mu^4 + 10\lambda\mu^3 + 30\xi\mu^2\alpha - 12\alpha\mu\xi^2 + 22\alpha\mu^2\lambda + 21\alpha\mu^3 - 4\lambda\xi\alpha^2 - 12\alpha^2\xi^2 + 12\alpha^2\mu\xi + 13\mu\alpha^2\lambda + 21\mu^2\alpha^2 + 6\alpha^3\mu)\mu^2\xi)$$

$$p_{9/i} = D^{-1}(3\alpha\mu^2\xi^2(-14\alpha\mu\lambda\xi + 33\mu^2\lambda\xi + 12\mu\xi^3 - 2\lambda^3\alpha + 7\lambda^2\xi\mu + 16\lambda\xi^2\mu - 14\lambda^2\xi\alpha - 32\alpha\lambda\xi^2 + 42\mu^3\xi + 6\mu^2\lambda^2 + 42\xi^2\mu^2 - \mu\lambda^2\alpha - 24\alpha\xi^3 + \mu\lambda^3 + 12\mu^4 + 14\lambda\mu^3 + 60\mu^2\alpha\xi - 24\mu\alpha\xi^2 + 21\mu^2\alpha\lambda + 42\mu^3\alpha - 24\xi\alpha^2\lambda - 6\alpha^2\lambda^2 - 24\xi^2\alpha^2 + 24\alpha^2\mu\xi + 8\lambda\mu\alpha^2 + 42\alpha^2\mu^2 - 4\lambda\alpha^3 + 12\mu\alpha^3))$$

$$p_{10/i} = D^{-1}(\mu^2\xi^3(-14\xi\mu\alpha\lambda + 33\xi\lambda\mu^2 + 12\xi^3\mu - 2\lambda^3\alpha + 7\lambda^2\mu\xi + 16\lambda\xi^2\mu - 14\alpha\xi\lambda^2 - 32\xi^2\alpha\lambda + 42\mu^3\xi + 6\lambda^2\mu^2 + 42\xi^2\mu^2 - \lambda^2\alpha\mu - 24\alpha\xi^3 + \mu\lambda^3 + 12\mu^4 + 14\lambda\mu^3 + 60\xi\alpha\mu^2 - 24\xi^2\mu\alpha + 21\alpha\mu^2\lambda + 42\alpha\mu^3 - 24\lambda\alpha^2\xi - 6\lambda^2\alpha^2 - 24\xi^2\alpha^2 + 24\xi\alpha^2\mu + 8\lambda\alpha^2\mu + 42\alpha^2\mu^2 - 4\alpha^3\lambda + 12\alpha^3\mu))$$

where the constant D is given by

$$\begin{aligned}
D = & 252\mu^5\xi^2\alpha^2 + 14\lambda\mu^5\xi^3 + 168\mu^5\xi^3\alpha + 500\alpha^2\mu^3\xi^3\lambda + 78\lambda\alpha\mu^5\xi^2 + 96\lambda\alpha^2\mu^5\xi + 42\mu^5\xi^4 + 36\alpha\mu^6\xi^2 \\
& + 60\alpha^5\mu^3\xi + 36\alpha^2\mu^6\xi + 12\mu^6\xi^3 + 4\lambda\alpha\mu^2\xi^5 + 19\lambda^3\alpha\mu^2\xi^3 + 108\lambda^2\alpha^2\mu^4\xi + 72\lambda\alpha^5\mu^3 + 51\lambda^2\alpha^5\mu^2 \\
& + 2\lambda^3\alpha^6 + 22\lambda^3\alpha^5\mu + 66\alpha^2\mu^3\xi\lambda^3 + 3\lambda^4\mu^2\xi^2\alpha + 34\lambda^2\mu^2\xi^4\alpha + 160\lambda\mu^3\xi^4\alpha + 48\lambda^2\alpha^3\mu^4 + 42\alpha^5\mu^4 \\
& + 225\xi\alpha^4\lambda^2\mu^2 + 60\lambda^2\mu^4\xi^2\alpha + 246\lambda\mu^4\xi^3\alpha + 21\lambda^3\mu^3\xi^2\alpha + 12\mu^6\alpha^3 + 42\alpha^4\mu^5 + 42\xi^5\mu^4 + 6\lambda^2\alpha^6\mu \\
& + 4\lambda^5\alpha^4 + 33\lambda\mu^4\xi^4 + 6\lambda^2\mu^4\xi^3 + 186\alpha\mu^4\xi^4 + \lambda^3\mu^3\xi^3 + 7\lambda^2\mu^3\xi^4 + 16\lambda\mu^3\xi^5 + 12\alpha^6\lambda\mu^2 + 12\mu^3\xi^6 \\
& + 12\alpha\mu^3\xi^5 - 12\alpha^2\mu^3\xi^4 - 24\alpha\mu^2\xi^6 + 348\alpha^2\mu^4\xi^3 + 24\alpha^3\mu^3\xi^3 - 96\alpha^2\mu^2\xi^5 + 348\alpha^3\mu^4\xi^2 + 8\lambda^3\alpha^5\xi \\
& + 38\lambda^3\alpha^3\mu^3 + 99\lambda^2\alpha^4\mu^3 + 84\alpha^4\mu^3\xi^2 - 144\alpha^3\mu^2\xi^4 + 36\lambda\alpha^3\mu^5 + 96\lambda\alpha^4\mu^4 + 19\lambda^4\alpha^3\mu^2 + 12\alpha^6\mu^3 \\
& + 6\lambda^5\alpha^3\mu + 7\lambda^5\alpha^3\xi + 16\lambda^4\alpha^3\xi^2 + 12\lambda^3\alpha^3\xi^3 + 21\lambda^4\alpha^4\mu + 18\lambda^3\alpha^4\xi^2 + 119\lambda^2\alpha\mu^3\xi^3 + 30\xi\alpha^5\lambda^2\mu \\
& + 17\lambda^4\alpha^4\xi + 495\lambda\alpha^2\mu^4\xi^2 + 594\alpha^3\mu^3\xi^2\lambda + 252\xi^3\lambda^2\alpha^2\mu^2 + 60\alpha^2\mu^2\lambda\xi^4 + 348\lambda^2\mu^3\xi\alpha^3 + 3\lambda^5\alpha^2\mu\xi \\
& + 141\lambda^3\mu^2\xi^2\alpha^2 + 366\lambda\alpha^3\mu^4\xi + 318\lambda\alpha^4\mu^3\xi + 390\lambda^2\alpha^3\mu^2\xi^2 + 140\lambda\alpha^3\mu^2\xi^3 + 176\alpha^3\mu^2\lambda^3\xi + \lambda^6\alpha^3 \\
& + 21\lambda^4\alpha^2\mu^2\xi + 36\mu\lambda^2\alpha^2\xi^4 + 48\mu\lambda^3\alpha^2\xi^3 + 123\mu\lambda^3\alpha^3\xi^2 + 90\mu\lambda^2\alpha^3\xi^3 + 21\lambda^4\alpha^2\mu\xi^2 + 51\lambda^4\alpha^3\mu\xi \\
& + 86\lambda^3\alpha^4\xi\mu + 78\lambda^2\alpha^4\xi^2\mu - 24\alpha^5\mu^2\xi^2 + 5\lambda^4\alpha^5 + 186\alpha^4\mu^4\xi + 60\alpha^5\mu^2\lambda\xi - 96\alpha^4\mu^2\xi^3 + 58\alpha^4\mu^2\lambda^3 \\
& + 168\alpha^3\mu^5\xi + 381\lambda^2\alpha^2\mu^3\xi^2 + 132\lambda\alpha^4\mu^2\xi^2.
\end{aligned}$$

Transition

rate diagram for a reliable $M/M/3$ retrial queue

Aggregating the states of each class S_i , The rates of transition between the 'macro-states' are expressed in the infinitesimal generator matrix

$$q_{i,j} = \begin{cases} \lambda q_a p_{10|i} + \xi q_f p_{2|i} + 2\xi q_f p_{3|i} + \xi q_f p_{5|i} + (\lambda q_a + 3\xi q_f) p_{4|i} & \\ + (\lambda q_a + 2\xi q_f) p_{6|i} + (\lambda q_a + \xi q_f) p_{7|i} & i \geq 0, \quad j = i + 1 \\ i\theta(p_{1|i} + p_{2|i} + p_{3|i} + p_{5|i} + p_{8|i} + p_{9|i}) & i \geq 0, \quad j = i - 1 \\ -[\lambda q_a p_{10|i} + \xi q_f p_{2|i} + 2\xi q_f p_{3|i} + \xi q_f p_{5|i} + (\lambda q_a + 3\xi q_f) p_{4|i} + (\lambda q_a + 2\xi q_f) p_{6|i} + & \\ (\lambda q_a + \xi q_f) p_{7|i} + i\theta(p_{1|i} + p_{2|i} + p_{3|i} + p_{5|i} + p_{8|i} + p_{9|i})] & j = i \\ 0 & \text{otherwise} \end{cases}$$

3.6 Numerical Experiments

In this section, we assess the quality of the phase-merging approximate for the unreliable $M/M/3$ retrial queue. Using a direct truncation method [39], we will compare results for congestion.

ξ	Phase-Merging Algorithm	Direct Truncation Method
10.0000	0.0934	0.1326
5.0000	0.0448	0.0768
2.500	0.0236	0.0497
1.2500	0.0139	0.0364
0.6250	0.0083	0.0297
0.3125	0.0057	0.0264
0.1563	0.0045	0.0248
0.0781	0.0038	0.0240
0.0391	0.0035	0.0235
0.0195	0.0033	0.0233
0.0098	0.0033	0.0232

TABLE 3.3 – Breakdown rate (ξ) and mean number of customers in the orbit for $c = 3, \lambda = 8, \mu = 10, \alpha = 100, \theta = 100$ and $q_a = q_f = 1$

θ	phase-merging	Direct Truncation Method
10	0.935	0.9677
30	0.3117	0.3475
50	0.1870	0.2234
70	0.1336	0.1701
90	0.1042	0.1404
100	0.0935	0.1300
300	0.0311	0.0670
500	0.0187	0.0541
600	0.0155	0.0509
700	0.0133	0.0486
800	0.0116	0.0468
900	0.0103	0.0455
1000	0.0093	0.0444
2000	0.0046	0.0394
3000	0.0031	0.0378
4000	0.0023	0.0369
5000	0.0018	0.0364
6000	0.0015	0.0361

TABLE 3.4 – Retrial rate (θ) and mean number of customers in the orbit for $c = 3$, $\lambda = 30$, $\mu = 40$, $\xi = 10$, $\alpha = 100$ and $q_a = q_f = 1$

3.7 Conclusion

The primary aim of this work was to provide a formal analysis of the unreliable $M/M/c$ retrial queueing system with $c \geq 3$. Applying a phase merging algorithm due to Korolyuk [28] and Courtois [17], we showed that the steady state orbit length is approximately Poisson distributed. Using this result, we approximated the joint probability distribution of the number of customers in the orbit and the status of the servers. This enabled us to derive approximate expressions for the steady state mean orbit length, mean number of customers in service, mean number of customers in the system, the mean system sojourn time and the mean orbit sojourn time. The numerical examples show that that means number of customers in the orbit using the phase-merging algorithm decrease more quickly than direct truncation method's. Better yet, these results remain valid if $q_a < 1$ and $q_f < 1$.

Chapitre 4

On the analysis of unreliable Markovian multiserver queue with retrials and impatience

Ce chapitre fait l'objet d'une publication accepté dans l' *International Journal of Mathematical Sciences and Applications E-Notes*.

Mathematical Sciences and Applications E-Notes, mois 2019, Volume :..., Issue Number :..., Pages -
ISSN(.....)

ON THE ANALYSIS OF UNRELIABLE MARKOVIAN MULTISERVER QUEUE WITH RETRIALS AND IMPATIENCE

M.ELHADDAD and F. LIMAM-BELARBI

Laboratory of Statistic and Processus Stochastic

University of Djillali Liabes

L.P 89, Sidi Bel Abbes 22000, Algeria

e-mail : *meriem.elhaddad61@gmail.com* and *faiza_belarbi@yahoo.fr*

Abstract : This paper concerns an approximate analysis of a Markovian multiserver infinite source retrial queuing with impatience, in which all the servers are subject to breakdown and repairs. Customer who find the total number of busy and failed servers equal to s may balk, i.e, he is given to choice to enter a retrial orbit for an random amount of time before attempting to recess an available server or enter the queue of size q . Customer waiting in the queue start being served as an idle or repaired server assigned to them, they can also leave the queue and enter orbit due to impatience. Customers whose service is interrupted by a failure may have the option of leaving the system entirely or returning to the orbit to repeat or resume service. We assume that each server has its own dedicated repair person, and repairs begin immediately following a failure and all process are assumed to be mutually independent.

The simultaneous effect of customer balking, impatience and retrials is analyzed. We try to approximate the steady-state joint distribution of the number of customers in orbit and the number of customers in the service area using a phase-merging Algorithm.

Keywords : Retrial queue; Multi server; Impatient; Breakdown and repair of service; Phase merging algorithm.

Mathematics Subject Classification : 60K25,90B22,68M20.

4.1 Introduction

Queuing systems in which arriving customers who find all servers and waiting positions occupied may retry for service after a period of time is called Retrial Queues. Retrial queue are devoted by large literature, his study is motivated by the advent of new technologies including applications of switching systems phones and also to computer and telecommunication net-works. One of the earliest papers in retrial queue was "On the influence of repeated calls in the theory of probabilities of blocking" by Kosten [32]. In 1957, Cohen [16] published "Basic problem of telephone traffic theory and the influence of repeated calls in which he considered the more general $M/M/c$ retrial queue with impatient customers. Aguir and al. [1] and Altalejo and al. [10] modeled a call center as the $M/M/m$ retrial queue with exponential impatience time. Shin and Choo [38] considered an $M/M/s$ retrial queue with balking and reneging in which authors assumed that probabilities that the balking customers and reneging customers join orbit may depend on the number of customer in service facility. For detailed description of $M/M/s$ queue with impatient retrial see Shin [38].

Due to the complexity of the analytical results obtained during the analysis of call-back queue systems, various authors have tried to develop numerical and approximate methods in order to be able to analyze these systems with repeated calls. Aguir and al. [1], estimate the total arrival rate, the sum of external arrival rate and retrial rate for an $M/M/s$ retrial queue with balking and reneging by using fluid flow approximation. Subramanian and al. [39] , published " $M/M/s$ retrial queuing system with breakdown and repair of services" by using Direct Truncation Method. Falin [23] consider two approximations in order to study the system behavior for low retrial intensity. For detailed survey of retrial queues and bibliographical information see Falin [21], Artalejo [8], [9], [11], monograph by Falin and Templeton [24], Gomez-Corral [10], Melikov [29] and Elhaddad [20].

In this paper, we applied an approximate method in order to investigate an unreliable $M/M/s$ retrial queues with failed servers and discuss the influence of customer balking and impatience as well as the effect of the service area's on the mean orbit length using a phase-merging algorithm due to Korolyuk [28] and Courtois [17]. These kinds of models are readily available in real life situations, for example, consider the transmission of messages in facsimile "net-works" having the autorepeat facilities. If transmission is not successfully completed for some reasons such as a power failure or a transmission error, then the message leaves the server and joins the buffer with some probability and leaves the system with the complimentary probability. Our paper is organized as follows : In section 2, we provide

the formal model description and state the assumptions that are needed to implement the approximation procedure. In sections 3 and 4, the algorithm is formally reviewed for $q = 1$ and $s = 1$. In section 5, we derive approximations for the steady-state probabilities and several standard queueing performance measures. In section 6, we applied the Algorithm for a higher values of s and q and in section 7, based on the results obtained in section 5 for the case $s = 1$ and $q = 1$ we shall present some numerical example by varying the values of parameters model, which are chosen to satisfy the stationary condition given in section 6, to illustrate how the various parameters of the model influence on the mean orbit length.

4.2 Model description

The unreliable homogeneous multiservers infinite source retrial queueing system considered is described as follows : There are an infinite of sources each of which generates customers according to a Poisson process with rate λ ($\lambda > 0$). The customer who finds one of the servers idle and not failed occupies it immediately. However, if he finds the total number of busy and failed servers equal to s may balk, i.e., he is given the choice to enter the retrial orbit with probability f_e^- ($0 \leq f_e^- \leq 1$) or enter the queue of size q with probability $f_e^+ = 1 - f_e^-$. Let c be the finite capacity of the customer in service area ($c = s + q$). Customers located in the orbit wait for an exponentially distributed time and retry to enter the service area with retrial rate ν ($\nu > 0$). Busy Server complete requests with service rate μ and customers leave the system. Failures for the s servers occur via a Poisson process with rate ξ ($\xi > 0$) and the repair times for each server are exponentially distributed with rate α ($\alpha > 0$).

Customers waiting in the queue start being served as soon as an idle or repaired server assigned to them, they can also leave the queue and enter orbit with rate η due to impatience. The customer whose service is interrupted by a failure may enter the retrial orbit with probability q_f ($0 \leq q_f \leq 1$) or abandons the service with probability $1 - q_f$. Customers are lost if they decide not to join the orbit.

All times involved in the model are assumed to be mutually independent of each other.

4.3 Underlying Markov Chain

The system state of our model as described in Section 2 can be represented by a trivariate stochastic process, $\{(N(t), C(t), F(t)) : t \geq 0\}$, where $N(t)$ is the random variable which

represents the number of customers in the orbit at time t , $C(t)$ is the random variable which represents the number of customers in the service area at time t and $F(t)$ represents the number of failed servers at time t .

Since all the random times are exponentially distributed, the stochastic process is a continuous-time Markov chain (CTMC) on the state space $A = \{(i, j, k) : i \geq 0; j \in \{0, 1, 2, \dots, q+s\}; k \in \{0, 1, 2, \dots, s\}\}$. We assume that as $t \rightarrow \infty$ the steady-state distribution of $\{(N(t), C(t), F(t)) : t \geq 0\}$ exists.

Define $p(i, j, k)$ as the limiting probability that the system is in the state (i, j, k) where $(i, j, k) \in A$. Defined mathematically

$$p(i, j, k) = \lim_{t \rightarrow \infty} P(N(t) = i, C(t) = j, F(t) = k).$$

In the following, we restrict our study to the case $s = 1$ and $q = 1$, the state space $A = \{(i, j, k) : i \geq 0; j \in \{0, 1, 2\}; k \in \{0, 1\}\}$. The state transition diagram of the corresponding CTMC is shown in Figure 4.1. The levels correspond directly to the size of the orbit and the pairs represent the number of customers in service area and the number of failed servers, respectively.

The statistical equilibrium equations of the system can be describe as follows :

for $i = 0$

$$(\lambda + \xi)p(0, 0, 0) = \mu p(0, 1, 0) + \alpha p(0, 1, 1)$$

and for $i \geq 1$

$$(\lambda + \xi + i\nu)p(i, 0, 0) = \mu p(i, 1, 0) + \alpha p(i, 1, 1)$$

$$(\lambda f_e^+ + \mu + i\nu f_e^+ + \lambda f_e^-)p(i, 1, 0) = \lambda p(i, 0, 0) + \mu p(i, 2, 0) + \alpha p(i, 1, 1) + (i+1)\nu p(i+1, 0, 0)$$

$$(\mu + \xi q_f + \eta + \lambda)p(i, 2, 0) = \lambda f_e^+ p(i, 1, 0) + i\nu f_e^+ p(i+1, 1, 0) + \lambda p(i-1, 2, 0)$$

$$(\lambda f_e^+ + \alpha + i\nu f_e^+ + \lambda f_e^-)p(i, 0, 1) = \xi p(i, 0, 0) + \eta p(i-1, 1, 1) + \lambda f_e^- p(i-1, 0, 1) + \xi q_f p(i-1, 1, 0)$$

$$(\alpha + \eta + \lambda)p(i, 1, 1) = \lambda f_e^+ p(i, 0, 1) + (i+1)\nu f_e^+ p(i+1, 0, 1) + \xi q_f p(i-1, 2, 0) + \lambda p(i-1, 1, 1)$$

$$\sum_{i=0}^{\infty} [p(i, 0, 0) + p(i, 1, 0) + p(i, 2, 0) + p(i, 0, 1) + p(i, 1, 1)] = 1$$

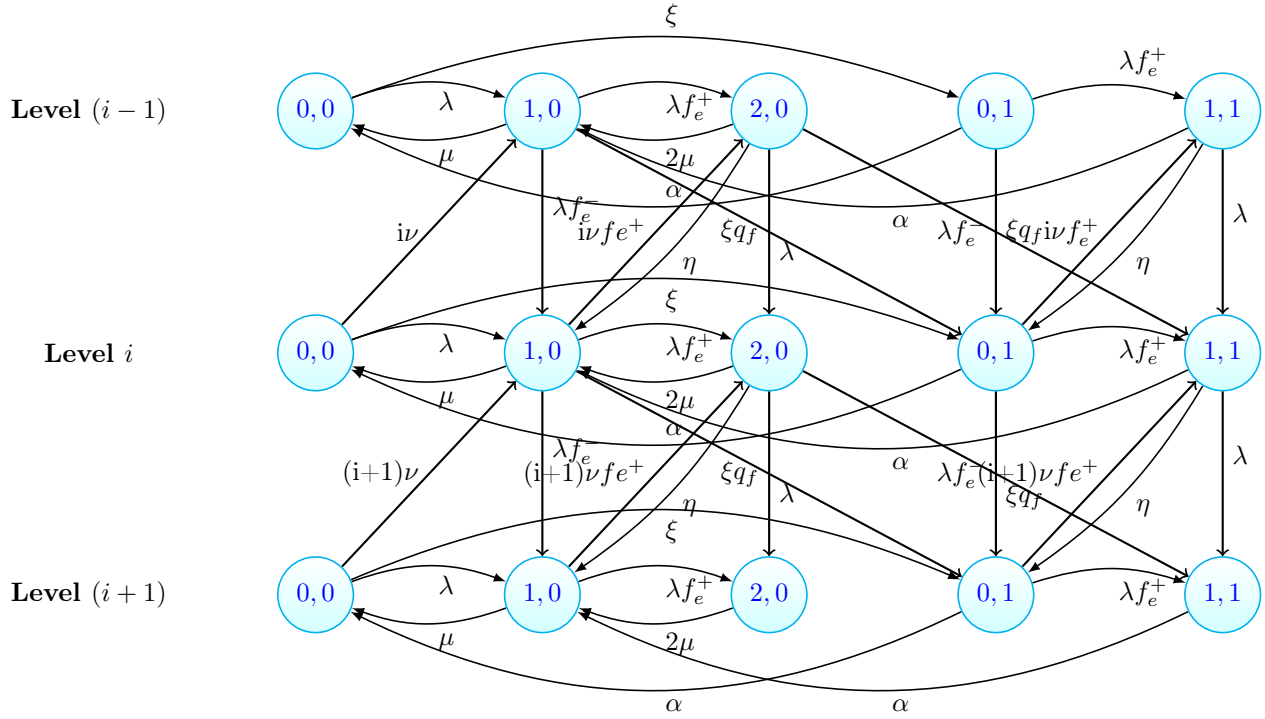


FIGURE 4.1 – State transition diagram of infinite source retrial queue for $s=1$, $q=1$

Note that is not easy to solve for the state probabilities recursively or by the method of generating functions, for this reason an approximate analysis is instead using a Phase Merging Algorithm developed by Korolyuk [28] and Courtois [17].

4.4 Application of the Phase Merging Algorithm

The objective of the phase-merging algorithm is to approximate the steady-state probability distribution of $\{(N(t), C(t), F(t)) : t \geq 0\}$ by approximating the conditional probability distribution of the status of the service area, given the level of the orbit and by approximating the marginal probability distribution of the number of customers in orbit, for more detailed, see Brian [14].

To make use of the algorithm, we proceed as follows : First, partition the state space into individual levels where the index of each level corresponds to the number of customers in the orbit. This step results in an infinite number of classes (or levels) which can be analyzed individually. Denote this as class A_i , for level $i, i \geq 0$.

Note that each class has an identical structure and is analyzed as a CTMC from which the approximate conditional probabilities are obtained, see Figure 4.2

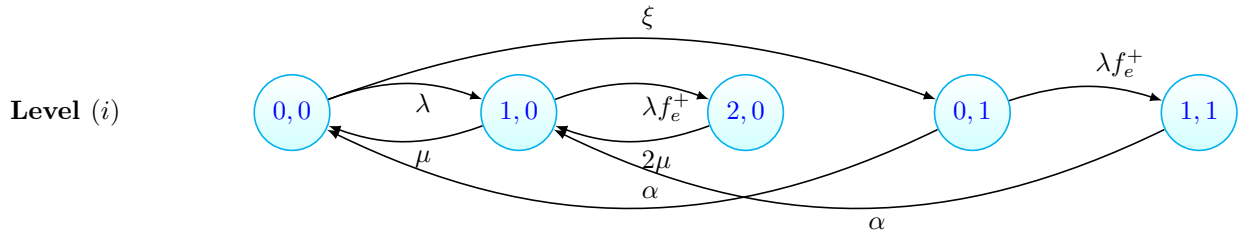


FIGURE 4.2 – The class A_i .

For each $i \geq 0$, the transition rates for this process are described in the following generator matrix Q_i

$$Q_i = \begin{pmatrix} -(\lambda + \xi) & \lambda & 0 & \xi & 0 \\ \mu & -(\lambda f_e^+ + \mu) & \lambda f_e^+ & 0 & 0 \\ 0 & \mu & -\mu & 0 & 0 \\ \alpha & 0 & 0 & -(\alpha + \lambda f_e^+) & \lambda f_e^+ \\ 0 & \alpha & 0 & 0 & -\alpha \end{pmatrix}$$

The following system yields,

$$\begin{aligned} (\lambda + \xi)p_{(0,0)|i} &= \mu p_{(1,0)|i} + \alpha p_{(1,1)|i} \\ (\lambda f_e^+ + \mu)p_{(1,0)|i} &= \lambda p_{(0,0)|i} + \mu p_{(2,0)|i} + \alpha p_{(1,1)|i} \\ \mu p_{(2,0)|i} &= \lambda f_e^+ p_{(1,0)|i} \\ (\lambda f_e^+ + \alpha)p_{(0,1)|i} &= \xi p_{(0,0)|i} \\ \alpha p_{(1,1)|i} &= \lambda f_e^+ p_{(0,1)|i} \end{aligned}$$

$$\sum_{i=0}^{\infty} [p(i, 0, 0) + p(i, 1, 0) + p(i, 2, 0) + p(i, 0, 1) + p(i, 1, 1)] = 1.$$

Solving the equations $p_{(j,k)|i} Q_i = 0$, ($j \in 0, 1, 2 ; k \in 0, 1$) and $p_i e = 1$ (where e is a column vector of ones),

the conditional probabilities are obtained for all $i \geq 0$, using the Maple Calculation Software

$$\begin{aligned}
 p_{(0,0)/i} &= D^{-1}(2\mu^2\alpha^2 + 2\mu^2\alpha\lambda f_e^+) \\
 p_{(1,0)/i} &= D^{-1}(\mu\alpha(2\lambda + \xi)(\alpha + \lambda f_e^+)) \\
 p_{(2,0)/i} &= D^{-1}(\lambda f_e^+\alpha(2\lambda + \xi)(\alpha + \lambda f_e^+)) \\
 p_{(0,1)/i} &= D^{-1}(2\alpha\mu^2\xi) \\
 p_{(1,1)/i} &= D^{-1}(\mu^2\xi(\alpha + \lambda f_e^+)).
 \end{aligned}$$

The constant D is given by the term

$$\begin{aligned}
 D = & 3\mu^2\alpha\xi + \mu^2\xi\lambda f_e^+ + 2\mu^2\alpha^2 + 2\mu^2\alpha\lambda f_e^+ + \mu\alpha^2\xi + \mu\alpha\xi\lambda f_e^+ + 2\alpha^2\lambda\mu + 2\mu\alpha\lambda^2 f_e^+ + \alpha^2\xi\lambda f_e^+ \\
 & + \alpha\xi\lambda^2(f_e^+)^2 + 2\lambda^2\alpha^2 f_e^+ + 2\alpha\lambda^3(f_e^+)^2.
 \end{aligned}$$

Next, we merge all states within the class A_i into one state corresponding to the level of orbit, i .

The infinitesimal generator, of the merged model is

$$q_{i,j} = \begin{cases} (\lambda f_e^- + \xi q_f)p_{(1,0)|i} + (\lambda + \eta + \xi q_f)p_{(2,0)|i} + \lambda f_e^- p_{(0,1)|i} \\ \quad + (\lambda + \eta)p_{(1,1)|i} & i \geq 0, \quad j = i + 1 \\ i\nu p_{(0,0)|i} + i\nu f_e^-(p_{(1,0)|i} + p_{(0,1)|i}) & i \geq 1, \quad j = i - 1 \\ -[(\lambda f_e^- + \xi q_f)p_{(1,0)+|i} + (\lambda + \eta + \xi q_f)p_{(2,0)|i} \\ \quad + \lambda f_e^- p_{(0,1)|i} + (\lambda + \eta)p_{(1,1)|i} + i\nu p_{(0,0)|i} + i\nu f_e^-(p_{(1,0)|i} + p_{(0,1)|i})] & i = j \\ 0 & \text{otherwise} \end{cases}$$

Using the substitutions, $\hat{\lambda} = (\lambda f_e^- + \xi q_f) p_{(1,0)|i} + (\lambda + \eta + \xi q_f) p_{(2,0)|i} + \lambda f_e^- p_{(0,1)|i} + (\lambda + \eta) p_{(1,1)|i}$ and $\hat{\nu} = \nu p_{(0,0)|i} + \nu f_e^- (p_{(1,0)|i} + p_{(0,1)|i})$, our infinitesimal generator, $q_{i,j}$, of the merged model become

$$q_{i,j} = \begin{pmatrix} -\hat{\lambda} & \hat{\nu} & 0 & 0 & 0 & . & . & . \\ \hat{\lambda} & -(\hat{\lambda} + \hat{\nu}) & 2\hat{\nu} & 0 & 0 & . & . & . \\ 0 & \hat{\lambda} & -(\hat{\lambda} + 2\hat{\nu}) & 3\hat{\nu} & 0 & . & . & . \\ 0 & 0 & \hat{\lambda} & -(\hat{\lambda} + 3\hat{\nu}) & 4\hat{\nu} & 0 & . & . \\ 0 & 0 & 0 & \hat{\lambda} & -(\hat{\lambda} + 4\hat{\nu}) & 5\hat{\nu} & 0 & . \\ \vdots & & & & & & & \end{pmatrix}$$

We see that the analysis of this system is analogous to the $M/M/\infty$ queueing system. Using the method of arc cuts, we recursively solve for the steady-state probability vector, $\pi = [\pi_0, \pi_1, \pi_2, \dots]$ solving recursively the system equations, we obtain

$$\begin{aligned} \hat{\lambda}\pi_0 &= \hat{\nu}\pi_1 \Rightarrow \pi_1 = \frac{\hat{\lambda}}{\hat{\nu}}\pi_0, \\ \hat{\lambda}\pi_1 &= 2\hat{\nu}\pi_2 \Rightarrow \pi_2 = \frac{1}{2}\left(\frac{\hat{\lambda}}{\hat{\nu}}\right)^2\pi_0. \end{aligned}$$

Continuing, it can be shown that

$$\pi_i = \frac{1}{i!} \left(\frac{\hat{\lambda}}{\hat{\nu}}\right)^i \pi_0, \quad i \geq 0. \quad (4.4.1)$$

Using the normalization equation, $\sum_{j=0}^{\infty} \pi_j = 1$, the solution for π_0 is obtained by

$$\begin{aligned} \pi_0 + \frac{\hat{\lambda}}{\hat{\nu}}\pi_0 + \frac{1}{2}\left(\frac{\hat{\lambda}}{\hat{\nu}}\right)^2\pi_0 + \frac{1}{6}\left(\frac{\hat{\lambda}}{\hat{\nu}}\right)^3\pi_0 + \dots &= 1 \\ \pi_0(1 + \frac{\hat{\lambda}}{\hat{\nu}} + \frac{1}{2}\left(\frac{\hat{\lambda}}{\hat{\nu}}\right)^2 + \frac{1}{6}\left(\frac{\hat{\lambda}}{\hat{\nu}}\right)^3 + \dots) &= 1 \end{aligned}$$

$$\pi_0 \left(\sum_{j=0}^{\infty} \frac{1}{j!} \left(\frac{\hat{\lambda}}{\hat{\nu}}\right)^j \right) = 1. \quad (4.4.2)$$

The infinite series of (4.4.2) is the Maclaurin power series expansion for $e^{\hat{\lambda}/\hat{\nu}}$. Thus, we see that

$$\pi_0 = e^{-\hat{\lambda}/\hat{\nu}}$$

Substituting π_0 into Equation (4.4.1) we have the following expression

$$\pi_i = \frac{1}{i!} \left(\frac{\hat{\lambda}}{\hat{\nu}} \right)^i e^{-\hat{\lambda}/\hat{\nu}}, \quad i \geq 0$$

which is the probability mass function for a poisson distributed random variable with rate parameter $\frac{\hat{\lambda}}{\hat{\nu}}$.

Finally, we approximate the steady-state distribution of $\{(N(t), C(t), F(t)) : t \geq 0\}$ by

$$\begin{aligned} p(i, j, k) &= p_{(j,k)|i} \times \pi_i \\ &= \frac{p_{(j,k)|i}}{i!} \left(\frac{\hat{\lambda}}{\hat{\nu}} \right)^i e^{-\hat{\lambda}/\hat{\nu}} \quad i \geq 0; j \in \{0, 1, 2\}; k \in \{0, 1\}. \end{aligned}$$

4.5 results

Using this joint distribution, we can approximate the standard queueing performance measures.

Mean Orbit Length

$$\begin{aligned} E[N] &\approx \frac{\hat{\lambda}}{\hat{\nu}} \\ &= \frac{(\lambda f_e^- + \xi q_f) p_{(1,0)+|i} + (\lambda + \eta + \xi q_f) p_{(2,0)|i} + \lambda f_e^- p_{(0,1)|i} + (\lambda + \eta) p_{(1,1)|i}}{\nu p_{(0,0)} + \nu f_e^- (p_{(1,0)|i} + p_{(0,1)|i})}. \end{aligned}$$

Mean Number of Customer in Service Area

The approximate expression for the expected number of customer at the Service Area is given by :

$$E[A] = \sum_{i=0}^{\infty} [p(i, 1, 0) + p(i, 2, 0) + p(i, 1, 1)].$$

Mean Queue Length

The mean queue length MQL is given by :

$$MQL = \sum_{i=0}^{\infty} [p(i, 1, 1) + p(i, 2, 0)].$$

4.6 Model Generalization

In section 3, the investigation was restricted for $q = 1$ and $s = 1$ to show the principles, in this section we apply the Algorithm for a higher values of s and q and to characterize the status of

the service area at any time, we need a set of $\sum_{n=1}^s n + (s+1)(q+1)$ pairs of (j, k) . Recall that all the random times are exponentially distributed and the stochastic process is a continuous-time Markov chain (CTMC) on the state space $A = \{(i, j, k) : i \geq 0, j \in \{0, 1, 2, \dots, q+s\}; k \in \{0, 1, 2, \dots, s\}\}$. We assume that as $t \rightarrow \infty$, the steady-state distribution of $\{(N(t), C(t), F(t)) : t \geq 0\}$ exists and we need the sufficient condition for ergodicity, see Falin [24]

$$\frac{\lambda}{s\mu} \left(1 + \frac{\xi}{\alpha} + \eta \right) < 1.$$

Define $p(i, j, k)$ as the limiting probability that the system is in the state (i, j, k) where $(i, j, k) \in A$. Defined mathematically

$$p(i, j, k) = \lim_{t \rightarrow \infty} P(N(t) = i, C(t) = j, F(t) = k).$$

First, we reduce the dimensionality of the state space by defining $X(t)$ as the status of the service area at time t (outlined in Table 1), such that $X(t) \in \{1, 2, 3, \dots, \sum_{n=1}^s n + (s+1)(q+1)\}$.

State (j, k)	Index(l)
(0,0)	1
(1,0)	2
(2,0)	3
$\vdots$	$\vdots$
(s,0)	s + 1
(s + 1,0)	s + 2
$\vdots$	$\vdots$
(c,0)	c+1
(0,1)	c+2
(1,1)	c+3
$\vdots$	$\vdots$
(q,s)	$\sum_{n=1}^s n + (s+1)(q+1)$

Table 1 : Substitution for service area status.

Since the number of customers in the orbit can theoretically reach infinity, the state space of the system can be partitioned into a countable number of classes. As noted previously, the

state space A is partitioned as the countable union,

$$A = \bigcup_{i=0}^{\infty} A_i, \quad A_i \cap A_j = \emptyset, \quad i \neq j$$

where $A_i = \{(i, l) : l = 1, 2, \dots, \sum_{n=1}^s n + (s+1)(q+1)\}, i \geq 0$. each class is identical in structure so that only one class needs to be analyzed.

We denote $p_{l|i}$ as the limiting conditional probability of the service area being in state l given that there are i customers in orbit

$$p_{l|i} = \lim_{t \rightarrow \infty} P(X(t) = l / N(t) = i), l = 1, 2, \dots, \sum_{n=1}^s n + (s+1)(q+1)$$

For each $i \geq 0$, the transition rates for this process are described in the generator matrix Q_i .

$$\begin{aligned} (\lambda + s\xi)p_{(0,0)|i} &= \mu p_{(1,0)|i} + \alpha p_{(0,1)|i} \\ (\lambda + (s-1)\xi + \mu)p_{(1,0)|i} &= \lambda p_{(0,0)|i} + 2\mu p_{(2,0)|i} + \alpha p_{(1,1)|i} \\ (\lambda + (s-2)\xi + 2\mu)p_{(2,0)|i} &= \lambda p_{(1,0)|i} + 3\mu p_{(3,0)|i} + \alpha p_{(2,1)|i} \\ &\vdots \\ (\lambda + \xi + (s-1)\mu)p_{(s-1,0)|i} &= \lambda p_{(s-2,0)|i} + s\mu p_{(s,0)|i} + \alpha p_{(s-1,1)|i} \\ (\lambda f_e^+ + s\mu)p_{(s,0)|i} &= \lambda p_{(s-1,0)|i} + s\mu p_{(s+1,0)|i} + \alpha p_{(s,1)|i} \\ (\lambda f_e^+ + s\mu)p_{(j,0)|i} &= \lambda f_e^+ p_{(j-1,0)|i} + s\mu p_{(j+1,0)|i} + \alpha p_{(j,1)|i} \text{ with } j \in \{s+1, s+2, \dots, s+q-1\} \\ s\mu p_{(s+q,0)|i} &= \lambda f_e^+ p_{(s+q-1,0)|i} \\ (\lambda + (s-(j+k))\xi + j\mu + k\alpha)p_{(j,k)|i} &= (j+1)\mu p_{(j+1,k)|i} + (k+1)\alpha p_{(j,k+1)|i} + \lambda p_{(j-1,k)|i} \text{ with } j+k < S; j \neq 0; k \neq 0 \\ (\lambda f_e^+ + (s-k)\mu + k\alpha)p_{(j,k)|i} &= \lambda f_e^+ p_{(j-1,k)|i} + (s-k)\mu p_{(j+1,k)|i} + (k+1)\alpha p_{(j,k+1)|i} \text{ with } s < j+k < C; k \neq 0, s \\ (\lambda f_e^+ + (s-k)\mu + k\alpha)p_{(j,k)|i} &= \lambda p_{(j-1,k)|i} + (s-k)\mu p_{(j+1,k)|i} + (k+1)\alpha p_{(j,k+1)|i} + \xi p_{(j,k-1)|i} \text{ with } j+k = s \\ ((s-k)\mu + k\alpha)p_{j,k|i} &= \lambda f_e^+ p_{(j-1,k)|i} \text{ with } j+k = s; k \neq 0; k \neq s \\ (\lambda + (s-1)\xi + \alpha)p_{(0,1)|i} &= s\xi p_{(0,0)|i} + \mu p_{(1,1)|i} + 2\alpha p_{(0,2)|i} \\ (\lambda + (s-2)\xi + 2\alpha)p_{(0,2)|i} &= (s-1)\xi p_{0,1|i} + \mu p_{(1,2)|i} + 3\alpha p_{(0,3)|i} \\ (\lambda + (s-3)\xi + 3\alpha)p_{(0,3)|i} &= (s-2)\xi p_{0,2|i} + \mu p_{(1,3)|i} + 4\alpha p_{(0,4)|i} \\ &\vdots \\ (\lambda + (s-1)\alpha + \xi)p_{(0,s-1)|i} &= 2\xi p_{0,s-2|i} + \mu p_{(1,s-1)|i} + s\alpha p_{(0,s)|i} \\ (\lambda f_e^+ + s\alpha)p_{(0,s)|i} &= \xi p_{0,s-1|i} \\ (\lambda f_e^+ + s\alpha)p_{(j,s)|i} &= \lambda f_e^+ p_{(j-1,s)|i} \text{ with } j \in 1, 2, \dots, q-1 \\ s\alpha p_{(q,s)|i} &= \lambda f_e^+ p_{(q-1,s)|i} \\ \sum_{n=1}^s n + (s+1)(q+1) &= 1 \\ \sum_{l=1} p_{l|i} &= 1 \end{aligned}$$

Aggregating the states of each class A_i . The rates of transitions between the "macro-states" are expressed in the infinitesimal generator matrix with elements,

$$q_{i,j} = \begin{cases} \sum_{j+k=c} \lambda p_{(j,k)|i} + \sum_{\substack{j+k \leq c \\ j > c}} q \eta p_{(j,k)|i} + \sum_{\substack{j+k < c \\ j > s}} \lambda f_e^- p_{(j,k)|i} + \sum_{\substack{j+k \leq c \\ j \neq 0, k < s}} j \xi q_f p_{(j,k)|i} & i \geq 0, \quad j = i + 1 \\ i \left[\nu \sum_{j+k \leq s} p_{(j,k)|i} + \nu f_e^+ \sum_{s < j+k \neq c} p_{(j,k)|i} \right] & i \geq 1, \quad j = i - 1 \\ - \left[\sum_{j+k=c} \lambda p_{(j,k)|i} + \sum_{\substack{j+k \leq c \\ j > c}} q \eta p_{(j,k)|i} + \sum_{\substack{j+k < c \\ j > s}} \lambda f_e^- p_{(j,k)|i} + \sum_{\substack{j+k \leq c \\ j \neq 0, k < s}} j \xi q_f p_{(j,k)|i} \right. \\ \left. + i \nu \sum_{j+k \leq s} p_{(j,k)|i} + i \nu f_e^+ \sum_{s < j+k \neq c} p_{(j,k)|i} \right] & i = j \\ 0 & \text{otherwise} . \end{cases}$$

We use the following substitutions for $i, i \geq 0$

$$\begin{aligned} \hat{\lambda} &= \sum_{j+k=c} \lambda p_{(j,k)|i} + \sum_{\substack{j+k \leq c \\ j > c}} q \eta p_{(j,k)|i} + \sum_{\substack{j+k < c \\ j > s}} \lambda f_e^- p_{(j,k)|i} + \sum_{\substack{j+k \leq c \\ j \neq 0, k < s}} j \xi q_f p_{(j,k)|i} \\ \hat{\nu} &= \nu \sum_{j+k \leq s} p_{(j,k)|i} + \nu f_e^+ \sum_{s < j+k \neq c} p_{(j,k)|i} . \end{aligned}$$

• Mean Orbit Length

For the infinite source multi service retrial queue, the steady-state distribution will be also Poisson with parameter $\hat{\lambda}/\hat{\nu}$. Denoting N as the steady-state number of customers in orbit when the mean orbit size is approximated by

$$\begin{aligned} E[N] &\approx \frac{\hat{\lambda}}{\hat{\nu}} \\ &= \frac{\sum_{j+k=c} \lambda p_{(j,k)|i} + \sum_{\substack{j+k \leq c \\ j > c}} q \eta p_{(j,k)|i} + \sum_{\substack{j+k < c \\ j > s}} \lambda f_e^- p_{(j,k)|i} + \sum_{\substack{j+k \leq c \\ j \neq 0, k < s}} j \xi q_f p_{(j,k)|i}}{\nu \sum_{j+k \leq s} p_{(j,k)|i} + \nu f_e^+ \sum_{s < j+k \neq c} p_{(j,k)|i}} . \end{aligned}$$

• Mean Number of customers in Service

Denoting N_e as the random number of customers at the servers, the approximate expression for the expected number of customers at the servers can be computed using the approximate

steady-state joint probabilities derived in the last step of the algorithm and it is given by

$$E[N_e] = \sum_{i=0}^{\infty} \sum_{\substack{j \neq 0 \\ k \neq s}} p(i, j, k)$$

• Mean Number of customers in Service Area

Denoting N_r as the random number of customers at the service area

$$E[N_r] = E[N_e] + \sum_{i=0}^{\infty} \sum_{\substack{j \neq 0 \\ k=s}} p(i, j, k).$$

• Mean Queue Length

The mean queue length MQL is given by :

$$MQL = \sum_{i=0}^{\infty} \left(\sum_{j+k=s+1}^c p(i, j, k) \right).$$

• Steady-State System Size and Sojourn Time

Denotig L as the steady-state number of customers in the system, to calculate it, we simply sum the expressions for $E[N]$ and $E[N_r]$.

The steady-state mean sojourn time, W , follows directly from Little's law.

$$\begin{aligned} L &\approx E[N] + E[N_r] \\ W &\approx \frac{L}{\lambda}. \end{aligned}$$

• The probability mass function of number of busy servers

$$P(\text{all } S \text{ servers are busy}) = \sum_{i=0}^{\infty} \sum_{j=s}^c p(i, j, 0).$$

- The probability mass function of number of servers in breakdown

$$P(\text{all } S \text{ servers are in breakdown}) = \sum_{i=0}^{\infty} \sum_{q=0}^{c-S} p(i, q, s).$$

4.7 Numerical Results

In this section, we will give some numerical results to illustrate how the various parameters of the model influence on the system performances measures for $s = 1$ and $q = 1$.

Table 1 : Retrial rate (ν) and impatience rate (η) against mean number of customers in the orbit for $\lambda = 0.5$, $\mu = 2$, $\alpha = 0.5$, $\xi = 0.05$, $f_e^- = f_e^+ = 0.5$ and $q_f = 0.5$.

ν	$\eta=0.02$	$\eta=0.04$	$\eta=0.06$	$\eta=0.08$	$\eta=0.1$	$\eta=0.2$	$\eta=0.4$	$\eta=0.6$	$\eta=0.8$	$\eta=1$
0.02	5.698	5.770	5.841	5.912	5.983	6.338	7.050	7.761	8.472	9.183
0.04	2.849	2.885	2.920	2.956	2.991	3.169	3.524	3.880	4.236	4.591
0.06	1.899	1.923	1.947	1.970	1.994	2.113	2.350	2.587	2.824	3.061
0.08	1.424	1.442	1.460	1.478	1.495	1.584	1.762	1.940	2.118	2.295
0.1	1.139	1.154	1.168	1.182	1.196	1.267	1.410	1.552	1.694	1.836
0.2	0.569	0.577	0.584	0.591	0.598	0.633	0.705	0.776	0.847	0.918
0.4	0.284	0.288	0.292	0.295	0.299	0.317	0.352	0.388	0.423	0.459
0.6	0.189	0.192	0.194	0.197	0.199	0.211	0.235	0.258	0.282	0.306
0.8	0.142	0.144	0.146	0.147	0.149	0.158	0.176	0.194	0.211	0.229
1	0.114	0.115	0.116	0.118	0.119	0.126	0.140	0.155	0.169	0.183

$P_{(0,0) i}$	$P_{(1,0) i}$	$P_{(2,0) i}$	$P_{(0,1) i}$	$P_{(1,1) i}$
0.708	0.185	0.023	0.047	0.035

From Figure 4.3, it can be seen that the mean number of customers in Orbit decreases rapidly as ν and η become higher and exceeds 0.1 .

Table 2 : Retrial rate ν and impatience rate η and probability f_e^+ against mean number of

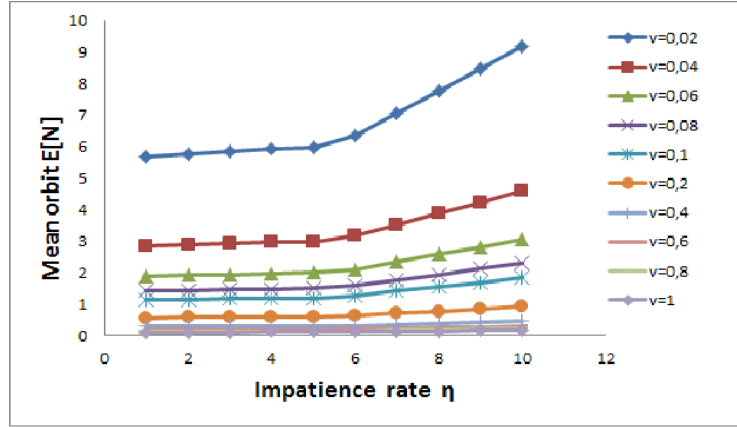


FIGURE 4.3 –

customers in the orbit for $\lambda = 0.5$, $\mu = 2$, $\alpha = 0.5$, $\xi = 0.05$ and $q_f = 0.5$.

$\nu = \eta$	$f_e^+ = 0$	$f_e^+ = 0.2$	$f_e^+ = 0.4$	$f_e^+ = 0.6$	$f_e^+ = 0.8$	$f_e^+ = 1$
0.02	17.581	13.866	11.112	8.963	7.225	5.783
0.04	8.815	6.978	5.618	4.558	3.701	2.991
0.06	5.893	4.683	3.787	3.090	2.527	2.060
0.08	4.432	3.535	2.872	2.355	1.939	1.594
0.1	3.556	2.846	2.322	1.915	1.587	1.315
0.3	1.218	1.009	0.857	0.740	0.647	0.570
0.5	0.751	0.642	0.564	0.505	0.459	0.421
0.7	0.550	0.484	0.439	0.405	0.378	0.357

From table 2, it is also interested in the sensitivity of the approximation procedure to vary f_e^+ from 0 to 1 in increments of 0.2 and provide the results for the same values of $\lambda, \mu, \alpha, \xi$ of table 1 and it can be shown that along the increases of ν and η for the same values, the mean number of customers in orbit decreases also when the values of f_e^+ become larger, this is because there is a high probability ($p_{(0,0)|i} = 0.708$) that the customer find the server idle when he retry to enter the server again for small value of λ .

Table 3 : Retrial rate ν , impatience rate η and λ against mean number of customers in the orbit for $\mu = 15$, $\alpha = 0.5$, $\xi = 0.05$, $f_e^- = f_e^+ = 0.5$ and $q_f = 0.5$.

λ	$\nu = \eta = 0.5$	$\nu = \eta = 0.6$	$\nu = \eta = 0.7$	$\nu = \eta = 0.8$	$\nu = \eta = 0.9$	$\nu = \eta = 1$
7	4.662	3.907	3.367	2.962	2.647	2.395
6	3.488	2.925	2.522	2.220	1.986	1.798
5	2.502	2.100	1.812	1.597	1.430	1.296
4	1.696	1.425	1.232	1.088	0.975	0.885
3	1.061	0.895	0.776	0.687	0.617	0.562
2	0.589	0.500	0.436	0.388	0.351	0.321
1	0.263	0.227	0.202	0.182	0.168	0.156

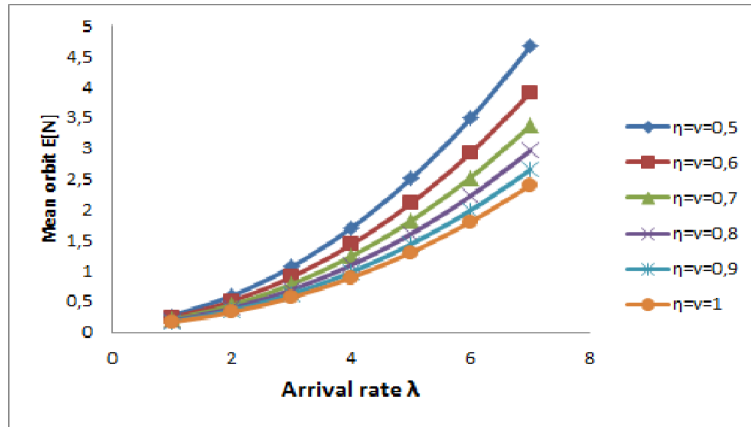


FIGURE 4.4 –

From Figure 4.4, it is shown that the mean number of customers in orbit increases and the server keep busy for more time as λ increase. High generation rate lead to high server utilization.

Table4 : Retrial rate ξ and repair rate α against mean number of customers in the orbit for $\lambda = 0.5$, $\mu = 2$, $\nu = 0.5$, $\eta = 0.5$, $f_e^- = f_e^+ = 0.5$ and $q_f = 0.5$.

ξ	$\alpha = 2$	$\alpha = 2.5$	$\alpha = 3$	$\alpha = 3.5$	$\alpha = 4$	$\alpha = 4.5$	$\alpha = 5$
0.02	0.186	0.184	0.182	0.181	0.180	0.180	0.179
0.04	0.206	0.202	0.199	0.196	0.195	0.193	0.192
0.06	0.226	0.220	0.215	0.212	0.209	0.207	0.205
0.08	0.246	0.237	0.231	0.227	0.223	0.221	0.219
0.1	0.266	0.255	0.247	0.242	0.238	0.235	0.232
0.3	0.463	0.433	0.413	0.398	0.387	0.378	0.371
0.5	0.657	0.612	0.581	0.559	0.542	0.529	0.518
0.7	0.847	0.791	0.753	0.725	0.703	0.686	0.672
0.9	1.034	0.971	0.927	0.894	0.869	0.849	0.833
1.1	1.219	1.151	1.103	1.067	1.040	1.018	1.000
1.3	1.402	1.331	1.281	1.243	1.214	1.192	1.173
1.5	1.583	1.511	1.460	1.422	1.393	1.370	1.351

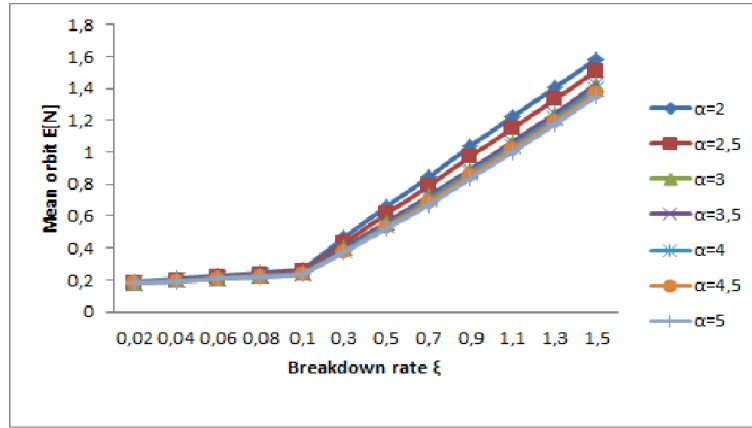


FIGURE 4.5 –

From Figure 4.5, it can be seen that as the Retrial rate ξ increase as the probability that the server is idle decrease and the orbit size increase.

Table5 : Retrial rate ξ and probability q_f against mean number of customers in the orbit for $\lambda = 0.5$, $\mu = 2$, $\alpha = 2$, $\nu = 0.5$, $\eta = 0.5$ $f_e^- = f_e^+ = 0.5$.

q_f	$\xi = 0.2$	$\xi = 0.4$	$\xi = 0.6$	$\xi = 0.8$	$\xi = 1$
0	0.309	0.436	0.550	0.652	0.745
0.1	0.320	0.461	0.590	0.710	0.821
0.2	0.331	0.486	0.631	0.768	0.898
0.3	0.343	0.511	0.671	0.825	0.974
0.4	0.354	0.535	0.712	0.883	1.050
0.5	0.365	0.560	0.752	0.941	1.127
0.6	0.377	0.585	0.793	0.999	1.203
0.7	0.388	0.610	0.833	1.056	1.280
0.8	0.399	0.635	0.874	1.114	1.356
0.9	0.411	0.660	0.914	1.172	1.433
1	0.422	0.685	0.955	1.229	1.509

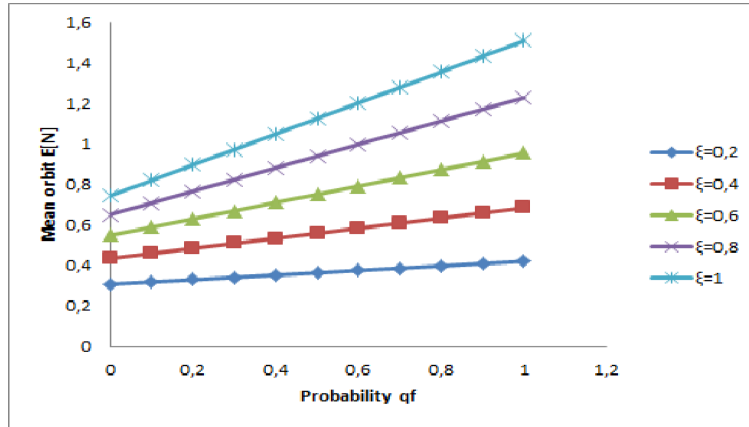


FIGURE 4.6 –

From Figure 4.6, it can be seen that the orbit size length is not affected remarkably by the probability q_f when the breakdown rate ξ is less than 0.2.

Conclusion

- The means number of busy and breakdown servers is independent of retrial rate ν and impatient rate η .
- The orbit size is not affected remarkably by the increase of the breakdown rate ξ when λ is small.

Conclusion

L'objectif principal de cette recherche était de fournir une analyse formelle du système de files d'attente avec rappels $M/M/c$ non fiable. L'étude de la documentation existante sur les files d'attente a révélé qu'il existe très peu de résultats pour les files d'attente de réexécution multiserveurs non fiables. En vertu des hypothèses de notre modèle, il a été démontré que la dérivation de la distribution de probabilité conjointe en régime permanent du nombre de clients sur l'orbite et le statut des serveurs via une approche analytique directe est extrêmement difficile. Cette complexité est due à des transitions qui correspondent à des changements dans le nombre de clients sur l'orbite, c'est-à-dire à des succès de nouvelle tentative, à l'impatience des clients, à un blocage à la première arrivée ou à une préemption par une défaillance du serveur. Par conséquent, nous avons eu recours à une analyse approximative pour obtenir la distribution conjointe en régime permanent du nombre de clients sur l'orbite et le statut des serveurs non fiables.

En appliquant un algorithme de fusion en phase, il a été constaté que le modèle agrégé est analogue à une file d'attente $M/M/\infty$. En résolvant les équations d'équilibre pour le modèle agrégé, nous avons montré que la longueur de l'orbite en régime permanent est approximativement répartie en poisson. En utilisant ce résultat, nous avons approximé la distribution de la probabilité conjointe du nombre de clients sur l'orbite et de l'état des serveurs. Cela nous a permis de dériver une expression approximative de la longueur moyenne de l'orbite en régime permanent, du nombre moyen de clients en service, du nombre moyen de clients dans le système, du temps de séjour moyen du système et du temps de séjour moyen en orbite.

En ce qui concerne les recherches futures, une multitude de variantes de mise en files d'attente peuvent être envisagées dans ce modèle, y compris le cas des distributions de temps des services généraux, en introduisant le concept de deuxième service optionnel. Le modèle peut

également être étendu à un réseau de files d'attente de nouvelle vérification multi-serveurs non fiables dans lesquelles l'algorithme de fusion de phase peut être appliqué.

Bibliographie

- [1] Aguir, S., Karaesmen, F., Aksin, O.Z., Chauvet, F., The impact of retrials on call center performance, *OR Spectr.* **26** (2004) 353-376.
- [2] Aissani, A and J.R. Artalejo, , On the single server retrial queue subject to breakdown, *Queueing Sys.*, **30** (1998), 309-321.
- [3] Arrar.N.K. Problème de convergence, optimisation d'algorithmes et analyse stochastique de systèmes de files d'attente avec rappels. Thèse soutenue Septembre 2012 à l'Université de ANNABA.
- [4] A. Ruegg. Processus stochastique. Presse Polytechnique Rromandes, Suisse, 1985.
- [5] Artalejo, J.R., and Gomez-Corral, A., Steady state solutions of a single server queue with linear repeated requests, *J. Appl. Probab.* 34 (1997), 223-233.
- [6] Artalejo, J.R.(1994).New results in retrial queueing systems with breakdowns of the servers.*StatisticaNeerlandica*, 48(1)23 – 36.
- [7] Artalejo.J.R and V. Pla. On the impact of customer balking, impatience and retrials in telecommunication systems. *Computers Mathematics with Applications*, 57, 217-229, (2009)
- [8] Artalejo, J.R., A classified bibliography of research on retrial queues : Progress in 1990-1999.,*Busin. Econ.*,**7** (1999), 187-211.
- [9] Artalejo, J.R., Accessible bibliography on retrial queues, *Math. Comp. Mod.*, **30** (1999), 1-6.
- [10] Artalejo, J.R. and A. Gomez-Corral, *Retrial Queueing Systems : A Computational Approach*, Springer, Spain, pp : 318. (2008)
- [11] Artalejo, J.R., Accessible bibliography on retrial queues : Progress in 2000-2009., *Math. Comp. Mod.*, **51** (2010), 1071-1081.

- [12] Badja.L and ABBAS.L Analyse stochastique des performances du système M/G/1 avec rappels ayant plusieurs types d'appels, Mémoire de Master, Université de Bjaïa, soutenue (20/06/2017).
- [13] Boutarfa.L. Modèles d'attente avec rappels et priorité et leurs applications. Thèse soutenue à l'Université de ANNABA (2014/2015).
- [14] Brian.P, Crawfor.D.A, *Approximate analysis of an unreliable M/M/2 retrial queue*, thesis, (2012)
- [15] Chunyan Li and Dequan Yue,A Queuing Model of the N-design Multi-skill Call Center with Impatient Customers ; International Journal of u-and-e-Service, Science and Technology, Vol.8,No.4 (2015),pp.51-60.
- [16] Cohen, J. W., Basic problems of telephone traffic theory and the influence of repeated calls, Philips Telecommunication Review, 18, No. 2, 1957.
- [17] Courtois, P.J., Decomposability, instabilities, and saturation in multiprogramming systems, *Communications of the ACM*, **18** (7) (1975), 371-377.
- [18] D.P.Gaver, P.A.Jacobs, G Latouche, Finite Birth-and-Death Models in Randomly Changing Environments, Adv.Appl.Prob.16, 715-731 (1984) Printed in N.Ireland.
- [19] D.P. Gaver, A Waiting line with interrupted service including priorities, J. Roy.Stat. Soc, B24(1962), pp.73-90
- [20] Elhaddad, M., Belarbi, F., Approximate analysis of an unreliable $M/M/c$ retrial queue with phase merging algorithm, *New Trends in Mathematical Sciences*, **4**, (2016), 9-21.
- [21] Falin, G., A survey of retrial queues, *Queueing Sys.*, **7**, (1990), 127-167.
- [22] G.I.Falin,J.G.C. Templeton, Retrial Queues-Springer US(1997)Monographs on Statistics and Applied Probability 75
- [23] Falin, G., Artalejo, J.R., Approximations for multiserver queues with balking/retrial discipline, *Or Spektrum*, **17**, (1995), 239-244.
- [24] Falin, G.I. and Templeton, J.G.C., *Retrial queues*, Champman and Hall, London, pp : 328. (1997)
- [25] Fakhari.F, Application de la théorie des files d'attente au niveau des Banques commerciales Algériennes, Université de Mohamed Boudiaf. M'sila.Algérie (2017).
- [26] Gomez-Corral.A,Retrial Queueing systems,A Computational Approach 2008
- [27] Kernane.T, Stabilité de Modèles de Files d'Attente, Thèse soutenue Avril 2007, USTHB.

- [28] Korolyuk, V.S. and Korolyuk, V.V., *Stochastic models of systems*. Kluwer Academic Publishers, Boston (1999).
- [29] Korolyuk, V. S., Melikov, A. Z., Ponomarenko, L. A., Rustamov, A. M., Methods for analysis of multi-channel queueing system with instantaneous and delayed feedbacks, *Cybernetics and Systems Analysis*, Vol. 52, No. 1 (2016), 58-70.
- [30] Vladimir S. Korolyuk, Vladimir V., *Stochastic Models of Systems*-Springer Netherlands(1999)- Mathematics and its Applications 469
- [31] Vladimir S., Korolyuk.V.S and Iasnianskiy.N, *Stochastic Systems in Merging phase space*, Singapore
- [32] Kosten, L., On the influence of repeated calls in the theory of probabilities of blocking, *De Ingenieur*, 59,(1947)
- [33] Kulkarni, V.G. and Choi, B.D., Retrial queues with server subject to breakdown and repairs, *Queueing Sys.*, **7** (1990), 191-208.
- [34] Mehdaoui.O, and Messaoui.F, Approximation fonctionnelle des systèmes d'attente à serveur non fiable, Mémoire de Master, Université A/Mira de Béjaïa (2016)
- [35] Neuts,M.F. and Rao, B.M.(1990) Numerical investigation of multiserver retrial model. *Queueing Systems*, 7, No2, 169-190.
- [36] Roszik,J.Sztrik(2004).Performance analysis of finite-source retrial queue with non-reliable heterogeneous servers, in : Proceedings of the XXIV Seminar on Stability Problems of Stochastic Models,Jurmala,Latvia.
- [37] Sztrik2016, GlobeEdit *Basic Queueing Theory*
- [38] Shin, Y.W., Choo, T.S., *M/M/s queue with impatient customers and retrials*, *Applied Mathematical Modelling*, **33** (2009), 2596-2606.
- [39] Subramanian, M.G., Ayyappan, G. and Sekar, G., *M/M/c Retrial queueing system with breakdown and repair of services*, *Asian Journal of Mathematics and Statistics* **4** (4) (2011), 214-223.
- [40] Thiruvengadam,K.(1963). Queueing with breakdowns. *Operations Research*, 11,62-71.
- [41] Titouah.A. Étude comparative de quelques techniques de troncature des chaînes de Markov . Mémoire de Master. Université A. Mira de Béjaïa (2012).
- [42] V. Naumov, U .Krieger, and D. Wagner. Analysis of a multi-server delay-loss system with a general Markovian arrival process. In S. Chakravathy and A. Alfa, editors.

- Matrix-Analytic Methods in Stochastic Models 183, of Lecture Note in Pure and Applied Mathematics. Marcel Dekker, 1996.
- [43] M. Neuts. Matrix-geometric solutions in stochastic models : An Algorithmic Approach. The Johns Hopkins University Press, Baltimore, 1981.
- [44] V. Naumov. Modified matrix-geometric solution for finite QBD processes, in advances in algorithmic methods for stochastic models. Notable Publications Inc, 2000.
- [45] G. Latouche and V. Ramaswami. Introduction to matrix analytic methods in stochastic modelling. SIAM, Philadelphia PA, 1999.
- [46] G.A. Tanonkou. Gestion de stocks et File d'attente, Industrial Engineering and Computer Science, INRIA-Lorraine.
- [47] Gomez-Corral, A. Stochastic analysis of a single server retrial queue with general retrial times, Nav. Res. Logistics. 1999, 46, 561-581.
- [48] G. Fayolle and M.A. Brun. On a system with impatience and repeated calls. In : Queueing theory and applications. Liber Amicorum for J.W. Cohen, pp. 283-305. North Holland, Amsterdam, (1988).
- [49] Wüchner.P, Sztrik.J, Hermann.M, Finite source M/M/S retrial queue with search for balking and impatient customers from the orbit. Faculty of Mathematics,University of Passau, Bavaria, Germany, Computer Networks 53 (2009) 1264-1273.

Bibliographie Générale

- Abou El-Ata, M. O., Hariri, A.M.A. (1992). The M/M/c/N queue with balking and reneging. *Comput. Oper. Res.*, 19(13) : 713-716.
- Abramov, V. M.(2006). Analysis of multiserver retrial queueing system : A martingale approach and an algorithm of solution. *Annals of Operations Research*, **141**, 19-50.
- Aguir, M.S., Karaesmen, F., Aksin, O.Z., Chauvet, F., (2004). The impact of retrials on call center performance, *OR Spectr.* **26** (2004) 353-376.
- Aissani, A and J.R. Artalejo, (1998). On the single server retrial queue subject to breakdown, *Queueing Sys.*, **30** (1998), 309-321.
- Ancker, Jr., C. J. and Gafarian, A. V. Some queueing problems with balking and reneging : I. *Operations Research*. 11 (1963), No. 1, 88-100.
- Artalejo, J.R. and Resing J.A.C., (2010). Mean value analysis of single server retrial queues. *Asia-Pacific Journal of Operational Research.*, 27 :335-345.
- Artalejo, J.R., (1994). New results in retrial queueing systems with breakdowns of the servers. *Statistica Neerlandica*, 48 (1)23-36.
- Artalejo, J.R., and Gomez-Corral, A., (1997). Steady state solutions of a single server queue with linear repeated requests, *J. Appl. Probab.* 34 , 223-233.
- Artalejo, J.R. (1999a). A classified bibliography of research on retrial queues : Progress in 1990-1999., *Busin. Econ.*, **7** , 187-211.
- Artalejo, J.R., (1999). Accessible bibliography on retrial queues, *Math. Comp. Mod.*, **30**, 1-6.
- Artalejo, J.R. and A. Gomez-Corral, (2008). *Retrial Queueing Systems : A Computational Approach*, Springer, Spain, pp : 318.
- Artalejo, J.R and V. Pla., (2009). On the impact of customer balking, impatience and retrials in telecommunication systems. *Computers Mathematics with Applications*, 57, 217-229.
- Artalejo, J.R., (2010). Accessible bibliography on retrial queues : Progress in 2000-2009, *Math. Comp. Mod.*, **51** , 1071-1081.
- Baccelli F. and Hebuterne G. (1981). On queues with impatient customers, in *Perforamnce'* 81 (F. Kylstra, ed.), North-Holland, Amsterdam, 159-179.
- Baccelli F., P. Boyer and G. Hebuterne (1984). Single-server queues with impatient customers, *Advances in Applied Probability*, 16 , 887-905.
- Bhulai, S. and Koole, G. A., (2003). Queueing model for call blending in call centers. *IEEE*

Transactions on Automatic Control., 48 :1434-1438.

Bouchentouf, A. A., Yahiaoui, L., (2017). On feedback queueing system with reneging and retention of reneged customers, multiple working vacations and Bernoulli schedule vacation interruption. *Arabian Journal of Mathematics.*, 6(1) : 1-11.

Boots, N.K. and H. Tijms, A multiserver queueing system with impatient customers, *Management Science*, 45 (1999), 444-448.

Boyer, P., Dupuis, A. and Khelladi, A. (1988). A simple model for repeated calls due to time-outs. *Teletraffic Science for New Cost Effective Systems, Networks and Services. ITC-12*, 356-363.

Brian.P, Crawford,B.A, (2007). *Approximate analysis of an unreliable M/M/2 retrial queue*, 07-05.

Chen, P. S., Zhu, Y. J. and Geng, X. (2008). M/M/m/k queue with preemptive resume and impatience of the prioritized customers. *Systems Engineering and Electronics.*, 30(6) : 1069-1073.

Chunyan Li and Dequan Yue, (2015). A Queueing Model of the N-design Multi-skill Call Center with Impatient Customers ; *International Journal of u-and-e-Service, Science and Technology*, Vol.8,No.4 ,pp.51-60.

Clos, C., (1948). An aspect of the dialing behaviour of subscribers and its effect on the trunk plant. *Bell systems technical journal*,27 , 424-445.

Cohen, J. W., (1957). Basic problems of telephone traffic theory and the influence of repeated calls, *Philips Telecommunication Review*, 18, No. 2.

Courtois, P.J., (1975). Decomposability, instabilities, and saturation in multiprogramming systems, *Communications of the ACM*, **18** (7) , 371-377.

Deslauriers, A., L., Ecuyer, P., Pichitlamken, J., Ingolfsson, A. and Avramidis, A.N., (2007). Markov chain models of a telephone call center with call blending. *Computers Operations Research.*, 34 :1616-1645.

Deul, N., (1980). Stationary conditions for multi-server queueing systems with repeated calls. (*Journal of Information Processing and Cybernetics*), 16, No.10-12, 607-613.

Elhaddad, M., Belarbi, F., (2016). Approximate analysis of an unreliable $M/M/c$ retrial queue with phase merging algorithm, *New Trends in Mathematical Sciences*, **4**, 9-21.

E. Altman and U. Yechiali, (2006). Analysis of customers' impatience in queues with server vacations, *Queueing Systems*, 52 , 261-279.

Fayolle, G. and Iasnogorodski, R., (1987). Criteria for the non-ergodicity of stochastic pro-

- cesses : application to the exponential Back-protocol, J. Appl. Probab. 24 , 347-354.
- Fayolle, G. and Brun, M.A. (1988). On a system with impatience and repeated calls. Queueing Theory and its Applications. Liber Amicorum for J. W. Cohen, (eds O.J.Boxma and R.Syski), Elsevier Science Publishers B.V., Amsterdam, 283-305.
- Falin, G.(1990). A survey of retrial queues, *Queueing Sys.*, **7**, 127-167.
- Falin, J.G.C. Templeton, (1997). Retrial Queues-Springer US Monographs on Statistics and Applied Probability 75.
- Falin, G., Artalejo, J.R. (1995). Approximations for multiserver queues with balking/retrial discipline, *Or Spektrum*, **17**, 239-244.
- Falin, G.I. and Templeton, J.G.C.(1997). *Retrial queues*, Champman and Hall, London, pp : 328.
- Falin, G.I. (1984a) On sufficient conditions for ergodicity of multi channel queueing systems with repeated calls. Advances in Applied Probabability, 16, 447-448.
- Falin, G.I. (1984b). Two-channel queueing system with repeated calls. Paper 4221-84, All-Union Institute for Scientific and Technical Information, Moscow (in Russian).
- Falin, G.I. (1984c). Asymptotic investigation of fully available switching systems with high repetition intensity of blocked calls. Moscow University Mathematics Bulletin, 39, No.6, 72-77.
- Falin, G.I. (1987a). On ergodicity of multilinear queueing systems with repeated calls. Soviet Journal of Computer and Systems Sciences, 25, No.1, 60-65. chains associated with repeated calls models. Moscow University.
- Fayolle G. and M.A. Brun, (1988). On a system with impatience and repeated calls. In : Queueing theory and applications. Liber Amicorum for J.W. Cohen, pp. 283-305. North Holland, Amsterdam.
- Gaver,D.P., Jacobs P.A., Latouche G., (1984). Finite Birth-and-Death Models in Randomly Chnging Environments, Adv.Appl.Prob.16, 715-731 Printed in N.Ireland.
- Gaver,D.P. (1962). A Waiting line with interrupted service including priorities, J. Roy.Stat. Soc, B24, pp.73-90.
- Gomez-Corral, A.(1999). Stochastic analysis of a single server retrial queue with general retrial times, Nav. Res. Logistics, 46, 561-581.
- Greenberg, B.S. (1986) Queueing systems with returning customers and the optimal order of tandem queues. Ph.D. Thesis, University of Cal ifornia, Berkeley.
- Gupta,N., Garg,R . (2012).A view of queue analysis with customer behaviour, balking and

- reneging, Proceedings of the National Conference on Trends and Advances in Mechanical Engineering, YMCA University of Science Technology. Faridabad. Haryana., 19-20.
- Hanschke, T., (1987). Explicit formulas for the characteristics of the M/M/2/2 queue with repeated attempts. *Journal of Applied Probability*, 24, 486-494.
- Hanschke, T. Die von Bretschneider, Cohen und Schwarzbart /Puri entwickelte Warteschlangenmodelle mit wiederholten Versuchen : (1978). eine Methode zur Berechnung der ergodischen Projection ihrer Markovschen Warteprozesse und die Simulation der Wartezeit, Fakultät für Mathematik der Universität Karlsruhe.
- Haight, F. A., (1957). Queuing with balking, I, *Biometrika*, Vol. 44, pp. 360-369.
- Haight, F. A., (1959), Queueing with Reneging, *Metrika*. Vol.2, pp. 186- 197.
- Jonin, G.L. and Sedol, J.J., (1970a). Telephone systems with repeated calls. A paper presented at the 6th International Teletraffic Congress, Munich.
- Jonin, G.L. and Sedol, Ya.Ya. (1970b). Investigation of telephone systems in presence of repeated calls. *Latvian Mathematical Yearbook*, 7, Riga, 71-83 (in Russian).
- Jonin, G.L. and Sedol, J.J. (1970c). Tables of Probabilistic Characteristics of Fully Available Trunk Group in the Case of Repeated Calls, Nauka, Moscow, (in Russian).
- Kapodistria, S. (2011) The M/M/1 Queue with Synchronized Abandonments, *Queueing Systems*, 68, 79109.
- Kernane, T. (2007) Stabilité de Modèles de Files d'Attente, Thèse soutenue Avril, USTHB.
- Korolyuk, V.S. and Korolyuk, V.V., (1999). *Stochastic models of systems*. Kluwer Academic Publishers, Boston .
- Korolyuk, V. S., Melikov, A. Z., Ponomarenko, L. A., Rustamov, A. M., (2016). Methods for analysis of multi-channel queueing system with instantaneous and delayed feedbacks, *Cybernetics and Systems Analysis*, Vol. 52, No. 1 58-70.
- Kosten, L., (1947). On the influence of repeated calls in the theory of probabilities of blocking, *De Ingenieur*, 59.
- Kulkarni, V.G. and Choi, B.D., (1990). Retrial queues with server subject to breakdown and repairs, *Queueing Sys.*, 7, 191-208.
- Kumar and Sharma (2014). *A multi-server Markovian Feedback Queue with Balking Reneging and Retention of reneged Customers*.
- Latouche G. and V. Ramaswami, (1999). *Introduction to matrix analytic methods in stochastic modelling*. SIAM, Philadelphia PA.
- Mehdaoui, O., and Messaoui, F., (2016). *Approximation fonctionnelle des systèmes d'attente*

- à serveur non fiable, *Mémoire de Master, Université A/Mira de Béjaïa*.
- Neuts, M. (1981). *Matrix-geometric solutions in stochastic models : An Algorithmic Approach*. The Johns Hopkins University Press, Baltimore.
- Neuts, M.F. and Rao, B.M., (1990). Numerical investigation of multi-server retrial model. *Queueing Systems*, 7, No2, 169-190.
- Naumov, V., Krieger U. , and D. Wagner, (1996). Analysis of a multi-server delay-loss system with a general Markovian arrival process. In S. Chakravathy and A. Alfa, editors. *Matrix-Analytic Methods in Stochastic Models* 183, of *Lecture Note in Pure and Applied Mathematics*. Marcel Dekker.
- Naumov, V. (2000). Modified matrix-geometric solution for finite QBD processes, in *advances in algorithmic methods for stochastic models*. Notable Publications Inc.
- Neuts, M.F. and Rao, B.M. (1990) Numerical investigation of multi server retrial model. *Queueing Systems*, 7, No.2, 169-190.
- Perel, N. and Yechiali, U. (2010). Queues with slow servers and impatient customers. *European Journal of Operational Research.*, 201(1) : 247-258.
- Roszik, J. Sztrik, (2004). Performance analysis of finite-source retrial queue with non-reliable heterogeneous servers, in : *Proceedings of the XXIV Seminar on Stability Problems of Stochastic Models*, Jurmala, Latvia.
- Ruegg, A. (1985). *Processus stochastique*. Presse Polytechnique Rromandes, Suisse.
- Seedya, El-Sherbinya and Ammar (2009). Transient solution of the M/M/c queue with balking and reneging Volume 57, Issue 8, April 2009, Pages 1280-1285.
- Sztrik, (2016). *Globe Edit Basic Queueing Theory*.
- Shin, Y.W., Choo, T.S. (2009). M/M/s queue with impatient customers and retrials, *Applied Mathematical Modelling*, **33** , 2596-2606.
- Shekhar, C., Raina, A.A and Kumar, A., (2016). A brief review on retrial queue : Progress in 2010-2015. *Int. Journal of Applied Sciences and Engineering Research.*, 5(4) : 508-536.
- Shin Y.W, T.S. Choo, (2009). M/M/s queue with impatient customers and retrial , *Applied Mathematical Modelling* 33 (6) 2596-2606.
- Subramanian, M.G., Ayyappan, G. and Sekar, G., (2011). M/M/c Retrial queueing system with breakdown and repair of services, *Asian Journal of Mthematics and Statistics* **4**, 214-223.
- Tanonkou, G.A. *Gestion de stocks et File d'attente*, *Industrial Engineering and Computer Science*, INRIA-Lorraine.

- Thiruvengadam, K. (1963). *Queueing with breakdowns*. *Operations Research*, 11, 62-71.
- Titouah, A., (2012). *Etude comparative de quelques techniques de troncature des chaines de Markov*. Mémoire de Master. Université A. Mira de Béjaia .
- Vladimir S. Korolyuk, Vladimir V., (1999). *Stochastic Models of Systems*-Springer Netherlands-Mathematics and its Applications 469.
- Vladimir S., Korolyuk, V.S and Iasnianski, N, *Stochastic Systems in Merging phase space*, Singapore.
- Wuchner, P., Sztrik, J., Hermann, M., (2009). *Finite source M/M/S retrial queue with search for balking and impatient customers from the orbit*. Faculty of Mathematics, University of Passau, Bavaria, Germany, *Computer Networks* 53 , 1264-1273.
- Wilkinson, R. I., (1956). *Theories for toll traffic engineering in the USA*. *Bell systems technical journal*, 35 (2), 421-507.
- White, H. and L. S. Christie, (1958). *Queueing with breakdowns*. *Operations Research*, 6 79-95.
- Yang, T. and J. G. C. Templeton, (1987). *A survey on retrial queues*. *Queueing systems*, 2 (3), 203-233.
- Yechiali, U. (2007). *Queues with system disasters and impatient customers when system is down*, *Queueing Systems*, 56, 195-202.
- Yechiali, U. and E. Altman, (2008). *Infinite-server queues with systems' additional task and impatient customers*, *Probability in the Engineering and Informational Sciences*, 22, 477-493.
- Yue D., Yue W., Saffer Z. and Chen X. (2014). *Analysis of an M/M/1 queueing system with impatient customers and a variant of multiple vacation policy*, *Journal of Industrial and Management Optimization*, 10, 89-112.

Résumé

Les réseaux de files d'attente sont des outils importants pour modéliser et évaluer les performances de toute une variété de systèmes réels. Cependant, les résultats analytiques ne sont pas toujours disponibles sous forme explicite. Pour la résolution de ces problèmes, on fait souvent recours à des méthodes d'approximation qui consistent à assimiler le système complexe à un système plus simple, analytiquement exploitable, qui lui est proche dans un certain sens.

Dans ce travail, nous nous sommes intéressés à évaluer les performances d'un système de files d'attente avec rappels de source infinie M/M/C avec impatience où tous les serveurs sont sujets à des pannes. En appliquant l'algorithme de phase merging, un modèle simplifié du système est construit sur la base de l'espace de phase fusionné de l'état du système, Cela nous a permis d'approximer l'état stationnaire de la répartition conjointe du nombre de clients en orbite et du nombre de clients dans la zone de service.

Abstract

Queue networks are important tools for modeling and assessing performance of a variety of real systems. However, the analytical results are not always available in explicit form.

To solve these problems, we often use approximation methods which consist in assimilating the Complex system to a simpler, analytically exploitable system, which is close to him in a certain sense. In this work, we are interested in evaluating the performance of an M/M/c Markovian multiserver infinite source retrial queuing with impatience, in which all the servers are subject to breakdown and repairs. By applying the merging phase algorithm, a simplified model of the system is built on the basis of the merged phase space of the system state, This allowed us to approximate the stationary state of the joint distribution of the customers in orbit and of the number of customers in the service area.

ملخص

شبكات قائمة الانتظار هي أدوات مهمة لنمذجة وتقييم الأداء لمجموعة متنوعة من الأنظمة الحقيقية. ومع ذلك ، فإن النتائج التحليلية ليست كذلك متاحة دائما في شكل واضح. لحل هذه المشكلات ، غالبا ما نستخدم طرق التقريب التي تتمثل في استيعاب النظام المعقد لنظام أبسط وقابل للاستغلال التحليلي ، وهو قريب منه بمعنى ما.

في هذا العمل كنا مهتمين بتقييم اداء نظام قائمة الانتظار ذات المصدر غير المحدود حيث تكون جميع الخوادم عرضة للعطل وهذا من خلال تطبيق خوارزمية مرحلة الدمج.

هذا ما سمح لنا بتقريب الحالة الثابتة للتوزيع المشترك لعدد العملاء في المدار و كذا عدد العملاء في منطقة الخدمة.