

UNIVERSITE SAAD DAHLAB DE BLIDA

Faculté des Sciences

Département de mathématiques

THESE DE DOCTORAT

en Mathématiques

Spécialité: Modélisation Mathématique et Statistique

MODELES D'ATTENTE AVEC RAPPELS

À INTERRUPTIONS DE SERVICE

Présentée par

Mlle Lila RIAH

Devant le jury composé de :

Mr M.Blidia	Professeur, U. de Blida	Président
Mme N.Oukid	Maître de conférences A, U. de Blida	Directrice de thèse
Mr A.Latreche	Professeur, ENSSEA.Koléa	Examineur
Mme G.Saidi	Maître de conférences A, ENSSEA.Koléa	Examinatrice
Mr H.Ould Rouis	Maître de conférences A, U. de Blida	Examineur

Blida, Juillet 2017

RESUME

L'étude des modèles de files d'attente avec rappels est motivée par diverses applications notamment l'émergence de nouveaux systèmes de communication (Mobile, ATM, Réseaux locaux, Internet...) et de production (ateliers flexibles,...). On les rencontre également en intelligence artificielle pour la modélisation de réseaux de neurones et autres systèmes intelligents. Avec le développement technologique des réseaux de communication, la demande d'attribuer la priorité d'une tâche par rapport à d'autres tâches a augmenté rapidement, ainsi qu'en pratique, on rencontre souvent des cas où les serveurs sont sujets à des pannes aléatoires, par conséquent il est plus réaliste de considérer les files d'attente avec des répétitions d'appels et le serveur exposé à des interruptions aléatoires qui empêchent la continuité du service des clients.

Dans cette thèse, nous étudions la fiabilité des modèles d'attente avec rappels à interruptions de service (interruptions dues aux pannes aléatoires ou à des priorités).

En premier lieu, nous actualisons les synthèses existantes sur les systèmes avec rappels, avec rappels et pannes, avec rappels, pannes et priorité.

En second lieu, nous considérons un système $M/M/2$ avec rappels et pannes aléatoires et en appliquant « phase merging algorithm », nous obtenons des approximations de certains indices de fiabilité. Ensuite, nous évaluons ces approximations par quelques illustrations numériques.

En dernier lieu, nous analysons mathématiquement le système de files d'attente $M/G/1$ avec rappels, pannes et priorité absolue. Notre analyse inclut : description de la chaîne de Markov induite, condition d'érgodicité, distribution stationnaire de l'état du système, obtention de mesures de performance et de quelques indices de fiabilité. Des illustrations numériques sont également introduites.

ABSTRACT

The study of the retrial queues models is motivated by various applications including the emergence of new communication systems (Mobile, ATM, Local networks, Internet ...), Production (flexible workshops,...). They are also encountered in artificial intelligence for the modeling of networks of neurons and other intelligent systems. With ever-fast growth in communication networks technology, the demand to assign priority to one type of jobs over other type of jobs has been growing rapidly, and in practice, there are often cases where servers are subject to random failures, so it is more realistic to consider queues with repeated calls and the server exposed to random interruptions that prevent the continuity of customer service.

In this thesis, we study the reliability of retrial retrial queues models with interruptions of service (interruptions due to random failures or priorities).

First, we update the work on retrial queues with breakdowns.

Secondly, we consider an $M/M/2$ retrial queue with active and idle breakdowns and by applying "phase merging algorithm", we obtain some approximate reliability indexes. Next, we evaluate these approximations with some numerical illustrations.

Finally, we analyze mathematically the $M/G/1$ retrial queuing system with failures and priority. Our analysis includes : description of the embedded Markov chain, ergodicity condition, stationary distribution of the system state, obtaining some performance measures and some reliability indexes. Digital illustrations are also introduced.

تلخيص

دراسة نظم طوابير الانتظار التي لها خاصية إعادة التنبية بطلب الخدمة محفزة بمجموعة متنوعة من التطبيقات خاصة بظهور أنظمة الاتصالات الجديدة (موبايل، ATM، الشبكات المحلية و الإنترنت) و أنظمة الإنتاج (نظم التصنيع المرنة...)، كما أن لها تطبيقات في مجال الذكاء الاصطناعي لنمذجة الشبكات العصبية والأنظمة الذكية الأخرى . مع التطور التكنولوجي لشبكات الاتصالات، طلب تعيين أولوية زبون بالنسبة إلي زبون آخر يزداد يوميا و في المجال التطبيقي غالبا ما نواجه الحالات التي يكون فيها الخوادم تخضع لأعطال عشوائية، لذلك هو أكثر واقعية للنظر في طوابير الانتظار التي لها خاصية إعادة التنبية بطلب الخدمة و التي تتعرض لانقطاعات عشوائية التي تحول دون استمرارية خدمة الزبائن.

في هذا العمل نقوم بدراسة موثوقية نظم طوابير الانتظار التي لها خاصية إعادة التنبية بطلب الخدمة مع انقطاع الخدمة الذي يكون سببه قدوم زبون يتمتع بأولوية مطلقة أو أعطال عشوائية.

أولا، نقوم بتحديث بعض النتائج المكرسة لأنظمة الطابور غير الموثوق بهامع إعادة التنبية بطلب الخدمة.

ثانيا، نعتبر نظام $M / M / 2$ بخاصية إعادة التنبية بطلب الخدمة و أعطال عشوائية وبتطبيق "phase merging algorithm"، نتحصل على تقديرات لبعض مؤشرات الموثوقية . ثم نقيم هذه النتائج بواسطة أمثلة عددية.

في الأخير نقوم بتحليل العشوائية لنظام $M/G/1$ بخاصية إعادة التنبية بطلب الخدمة والأولوية المطلقة. تحليل العشوائية الذي أجريناه يتضمن مايلي: وصف سلسلة ماركوف المضمنة، إعطاء شرط احتمال حدوث الخدمة مرة أخرى، التوزيع الثابت لحالة النظام، الحصول على بعض مقاييس الأداء وبعض مؤشرات الموثوقية.

REMERCIEMENTS

Je remercie vivement mon encadreur Mme N. OUKID (Maitre de conférence A à l'Université de Blida) pour l'aide précieuse qu'elle m'a apportée et pour sa disponibilité. Qu'elle trouve ici toute ma reconnaissance.

Je remercie également Mr M.BLIDIA (Professeur à l'Université de Blida) pour l'honneur qu'il me fait en acceptant de présider le jury de soutenance de cette thèse.

Mes remerciements s'adressent également à Mme G.SAIDI (M.C.A à ENSSEA. Koléa) , Mr A. LATRECH (Professeur à ENSSEA. Koléa) et Mr H. OULD ROUIS (M.C.A à l'Université de Blida) de m'avoir honoré en acceptant de faire partie du jury pour l'évaluation de ce travail.

Je remercie vivement mes parents pour leur soutien et leurs encouragements au cours de mes années d'études. Qu'ils trouvent ici l'expression de ma profonde gratitude et de mon plus grand respect.

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

TABLE DES MATIERES

RESUME	1
REMERCIEMENTS	4
TABLE DES MATIERES	5
Liste des illustrations, graphiques et tableaux	6
INTRODUCTION	9
1 SYSTEMES D'ATTENTE AVEC RAPPELS	14
1.1 Introduction	14
1.2 Modèles d'attente avec rappels	15
1.3 Modèles d'attente avec rappels et interruptions de service	20
1.4 Autres modèles	28
1.5 Exemples de systèmes modélisés par des files d'attente avec rappels	28
1.6 Exemples de systèmes modélisés par des files d'attente avec rappels et interruptions de service	30
1.7 Conclusion	32
2 NOTIONS PRELIMINAIRES DE FIABILITE	33
2.1 Introduction	33
2.2 Vocabulaire de la fiabilité	33
2.3 Indices de fiabilité	35
2.4 Principales lois de probabilité utilisées en fiabilité	38
2.5 Fiabilité d'un système	40
2.6 Méthodologies dynamiques pour l'étude de la fiabilité d'un système	42

2.7	Les applications de la fiabilité	45
2.8	Conclusion	51
3	ANALYSE APPROXIMATIVE DE LA FIABILITE DU SYSTEME M/M/2 AVEC RAPPELS	52
3.1	Introduction	52
3.2	description du modèle	53
3.3	Indices de fiabilité des serveurs	55
3.4	Illustrations numériques	61
3.5	Conclusion	64
4	MODELE D'ATTENTE AVEC RAPPELS, PANNES ET PRIORITE	66
4.1	Introduction	66
4.2	Modèle d'attente avec rappels et priorité absolue	66
4.3	Modèle M/G/1 avec rappels, pannes et priorité absolue	69
4.4	Illustrations numériques	87
4.5	Conclusion	88
	CONCLUSION ET PERSPECTIVES	89
	ANNEXES	91

LISTE DES ILLUSTRATIONS, GRAPHIQUES ET TABLEAUX

Figure 1.1	La structure du système	16
Figure 2.1	Evolution temporelle de l'état d'un système.	34
Figure 2.2	Courbe de survie ou de fiabilité.	35
Figure 2.3	Fonctions de distribution $F(t)$ et de fiabilité $R(t)$	36
Figure 2.4	Courbe en baignoire.	37
Figure 2.5	Diagramme de fiabilité pour un système en série.	40
Figure 2.6	Diagramme de fiabilité pour un système en parallèle.	42
Figure 2.7	Un modèle Markovien.	43
Figure 2.8	Réseau de Petri.	44
Figure 2.9	Réducteur de vitesse.	46
Figure 2.10	Exemple d'une modélisation du réducteur en schéma bloc.	47
Figure 2.11	Présentation graphique de la fiabilité du système et de ses composants.	48
Figure 3.1	La structure du système.	54
Figure 3.2	Graphe de transitions du système M/M/2 avec rappels et pannes.	55
Figure 3.3	une classe du modèle M/M/2 avec rappels et pannes.	57
Figure 3.4	Disponibilité et fréquence de panne vs. α	62
Figure 3.5	Disponibilité et fréquence de panne vs. β	62
Figure 3.6	Disponibilité et fréquence de panne vs. λ	63
Figure 3.7	Disponibilité et fréquence de panne vs. μ	64
Figure 3.8	Disponibilité et fréquence de panne vs. β	64

Figure 4.1	Disponibilité et fréquence de panne vs. α_2	87
Figure 4.2	Le comportement des probabilités p_4, p_5 et p_6 vs. α_1 and α_2	88

LISTE DES TABLEAUX

Tableau 2.1	Paramètres de la loi de défaillance des éléments du réducteur (distribution de Weibull).	47
Tableau 3.1	Les substitutions des états de serveurs.	57

INTRODUCTION

La théorie des files d'attente s'attache à modéliser et à analyser de nombreuses situations en apparence très diverses, mais qui relèvent néanmoins toutes du schéma descriptif général suivant. Des clients arrivent à intervalles aléatoires dans un système comportant plusieurs serveurs auxquels ils vont adresser une requête. La durée du service auprès de chaque serveur est elle-même aléatoire. Après avoir été servis (ce qui suppose un arrêt chez un ou plusieurs serveurs selon le cas), les clients quittent le système.

L'origine des études sur les phénomènes d'attente remonte aux années 1909-1920 avec les travaux de A.K. Erlang concernant le réseau téléphonique de Copenhague. La théorie mathématique s'est ensuite développée notamment grâce aux contributions de Palm, Kolmogorov, Khintchine, Pollaczek,... et fait actuellement toujours l'objet de nombreuses publications scientifiques. Cette théorie s'est ensuite étendue à de nombreux champs d'application comme la gestion de stocks, les télécommunications en général, la fiabilité de systèmes complexes,...

Cette théorie de files d'attente classique s'est très vite montrée inefficace face à des systèmes réels de plus en plus complexes, tels que les systèmes téléphoniques où les abonnés répétaient leurs appels en recomposant le numéro plusieurs fois jusqu'à l'obtention de la communication. Dès la fin des années 1940, des chercheurs tels que KOSTEN [1] et WILKINSON ont mis en évidence les limites de la théorie classique des files d'attente qui ne permettait pas d'expliquer le comportement stochastique de ce type de système.

Ce phénomène de répétition de demandes du service a poussé certains chercheurs à étendre le modèle d'attente classique à celui dit avec rappels. Ce type de systèmes de files d'attente avec rappels peut être appliqué pour résoudre les problèmes pratiques, tels que l'analyse du comportement des abonnés dans les réseaux téléphoniques et l'analyse du temps d'attente pour accéder à la mémoire sur les disques magnétiques. Ce type de modèles se rencontre également dans la modélisation de protocoles spécifiques de communication, tels que CSMA

(Carrier Sens Multiple Access) ou encore les disciplines Auto-Repeat, Ring-Back-When-Free, Repeat-Last-Number. Les progrès récents dans ce domaine sont résumés dans les articles de synthèse de AISSANI (1988) [2], FALIN (1990) [3], KULKARNI et LIANG (1997) [4], TEMPLETON (1999) [5] et dans les monographies de FALIN et TEMPLETON (1997) [6] et ARTALEJO et GOMEZ (2008) [7].

Dans le but d'étudier mathématiquement un système de files d'attente avec rappels, en premier lieu, on doit introduire un processus stochastique décrivant l'état du système à un instant donné. En général, on distingue deux catégories de processus stochastique décrivant l'état du système de file d'attente : les processus stochastiques Markoviens et ceux non-Markoviens. Cependant, on dispose de plusieurs méthodes qui permettent de rendre ces derniers Markoviens moyennant certaines transformations (méthode de la chaîne de Markov induite, méthode des variables supplémentaires). Puis, en fonction des quantités qui déterminent la structure du système, on cherche à calculer le régime transitoire. Il est constaté que le calcul explicite de ce dernier est pénible (voire impossible) pour la majorité des modèles. C'est pourquoi, il est souvent préférable de se contenter par la détermination du régime stationnaire. La distribution stationnaire du processus stochastique introduit permet d'obtenir les indices de performance du système : le temps d'attente d'un client, le nombre moyen de client dans le système, le taux d'occupation des dispositifs de service,...

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 le service d'un client est interrompu à cause d'arrivée d'un client prioritaire. La priorité peut être absolue ou relative. Par priorité absolue, on entend que le service d'un client moins prioritaire sera suspendu lorsqu'un client plus prioritaire se présente devant le serveur. Ce dernier venu commence son service immédiatement. Si la priorité est celle relative, un nouveau client plus prioritaire attend la fin du service avant de commencer le sien. Dans le cas de priorité absolue, deux possibilités se présentent : soit le client suspendu reprend son service là où il a été interrompu, soit il le reprend depuis le début. Ainsi qu'on rencontre des cas où les serveurs sont sujets à des pannes aléatoires (serveurs non fiables).

La théorie de fiabilité est une branche de la théorie des probabilités et de la statistique mathématique. C'est la discipline scientifique qui étudie les méthodes générales et les procédés à suivre lors de la réception, le transport, le stockage et/ou l'exploitation du système pour

garantir une efficacité maximale.

- Elle étudie la loi d'exploitation des défaillances et les méthodes de leur prévision.
- Elle établit et étudie les caractéristiques quantitatives (critères) de sûreté de fonctionnement.
- Elle élabore les méthodes de contrôle de la fiabilité et les procédures d'essais de fiabilité.
- Elle étudie les méthodes d'évaluation de la fiabilité à l'étape de la conception, par exemple, le choix de la structure optimale du système qui garantit une fiabilité donnée.
- Elle établit les méthodes de maintenance préventive (choix optimal des dates de renouvellement prophylactique) ainsi que les méthodes permettant de conserver la fiabilité à son niveau maximal au cours de l'exploitation.

L'objet principal des études de la fiabilité est donc de choisir des méthodes de décision, des techniques de fabrication... qui assurent l'incidence la plus faible de l'intervention humaine sur la fiabilité intrinsèque du système. Les modèles de fiabilité servent à comprendre comment, pourquoi et quand un système tombe en panne.

La théorie analytique des modèles d'attente avec rappels non fiables s'avère d'une portée limitée en raison de la complexité des résultats connus. En effet, dans la majorité des cas, on se retrouve confronté à des systèmes d'équations dont la résolution est complexe ou possédant des solutions qui ne sont pas facilement interprétables afin que le praticien puisse en bénéficier. Cette difficulté réside essentiellement dans l'utilisation des inverses des transformées de Laplace-Stieljes et des distributions marginales. Pour pallier à toutes ces difficultés, les chercheurs ont recouru aux méthodes d'approximation. Parmi les principales approches développées ces dernières années, on trouve les méthodes de comparaison stochastique [8], qui permettent d'avoir des estimations qualitatives pour certaines mesures de performance, la méthode de stabilité forte [9] qui permet de réaliser aussi bien une analyse quantitative qu'une analyse qualitative des systèmes complexes et l'algorithme de phase merging [10], [11] qui est utile pour l'analyse de chaînes de Markov bidimensionnelles à temps continu et qui permet de déterminer les distributions approximatives de probabilité à l'état stationnaire.

L'objectif de notre travail est d'élargir le champs d'application de la fiabilité sur des systèmes d'attente avec rappels non fiables, en particulier, nous nous intéressons à l'étude de l'influence de la non-fiabilité des serveurs sur les caractéristiques du système. Nous avons choisi deux types de modèles dont l'étude est sans aucun doute très importante pour les applications pratiques, notamment avec l'évolution rapide des systèmes informatiques et des réseaux de

télécommunication.

Le premier est un modèle markovien avec deux serveurs qui sont sujets à des pannes aléatoires qui peuvent se produire aussi bien lorsqu'ils sont libres que lorsqu'ils sont actifs. Les clients qui sont dans l'incapacité d'accéder à un serveur à cause de la panne ou de la congestion et les clients dont le service est interrompu par une panne doivent décider de rejoindre l'orbite ou de quitter le système définitivement. En général, les systèmes d'attente avec rappels à multiserveurs non fiables sont difficiles à analyser d'un point de vue mathématique et nous constatons qu'il n'y a pas de solutions exactes lorsque le nombre de serveurs dépasse un. Par conséquent nous cherchons à obtenir des approximations de certains indices de fiabilité en utilisant les probabilités approximatives d'état déterminées en appliquant l'algorithme de phase merging.

Le deuxième est un modèle semi markovien avec rappels à un seul serveur non fiable et priorité absolue. Nous considérons deux types de clients (plus prioritaire type 1 et moins prioritaire type 2). Un client type 1, arrivant dans le système et trouvant un autre client type 1 dans le serveur ou trouvant le serveur sous réparation, rejoint l'orbite et devient source de tentatives répétées. Un client type 1, arrivant dans le système et trouvant un client type 2 dans le serveur, interrompt le service en cours et commence le sien. Le client dont le service a été interrompu, reste auprès du serveur pour terminer son service ultérieurement. En outre, les clients type 2, arrivant dans le système et trouvant le serveur occupé, quittent le système sans entrer dans l'espace de service.

Cette étude est structurée comme suit :

- Le premier chapitre comprend une synthèse des principaux résultats analytiques sur les systèmes d'attente avec rappels et les systèmes avec rappels à interruptions de service, notamment ceux relatifs aux systèmes d'attente de type $M/G/1$.

- Dans le deuxième chapitre, nous introduisons des notions préliminaires de la fiabilité. Ainsi, on a axé notre présentation principalement sur les éléments essentiels nous permettant d'aborder le problème de la fiabilité des systèmes d'attente considérés dans ce travail.

- Le troisième chapitre est consacré à l'application de l'algorithme de phase merging au système de files d'attente $M/M/2$ avec rappels et pannes. En effet, après clarification de la manière d'application de l'algorithme pour avoir les probabilités approximatives d'état, nous obtenons des approximations de certains indices de fiabilité et nous évaluons ces approximations par des illustrations numériques sous MATLAB.

- Une partie du quatrième chapitre, est consacrée aux travaux de BOUTERFA et DJELLAB [12] concernant l'obtention de la condition d'ergodicité et quelques mesures de performance de modèle M/G/1 avec rappels, priorité et serveur fiable. Dans l'autre partie, nous présentons notre contribution à l'étude du modèle M/G/1 avec rappels, priorité et serveur non fiable en appliquant la méthode des variables supplémentaires. Notre analyse inclut : description de la chaîne de Markov induite, condition d'ergodicité, distribution stationnaire de l'état du système, obtention de mesures de performance et de quelques indices de fiabilité. Ces études théoriques sont soutenues par des illustrations numériques.

- Le travail s'achève par une conclusion mettant l'accent sur les perspectives et les directions de recherche induites par les résultats obtenus.

- Dans l'annexe, nous présentons les programmes développés dans le logiciel MATLAB permettant de réaliser les illustrations numériques fournis dans cette étude.

CHAPITRE 1

SYSTEMES D'ATTENTE AVEC RAPPELS

1.1 Introduction

La théorie des files d'attente, ou des queues, est un des outils analytiques les plus puissants pour la modélisation de systèmes de logistiques et de communication. Le phénomène de base de files d'attente surgit chaque fois qu'un serveur est consulté pour son service par un grand nombre de tâche ou de clients. La théorie des files d'attente classique offre deux possibilités pour résoudre le conflit qui apparaît lorsqu'un client arrive dans le système à serveur unique et trouve le serveur occupé :

- Le client arrivé se place dans la file d'attente et attend son tour selon l'une des disciplines : FIFO, LIFO ou Random.
- Le client peut quitter le système définitivement. C'est le système classique d'Erlang (avec refus ou avec perte).

Une possibilité alternative est que le client qui ne peut être servi, libère l'espace du serveur mais après une durée aléatoire revient au système pour répéter sa demande de service. Entre les appels successifs, le client en question se trouve en "orbite". Un tel système est appelé système de files d'attente avec "rappels".

Les systèmes de files d'attente avec rappels sont identifiés par la nature stochastique du processus des arrivées, la distribution du temps de service, le nombre de serveurs qui composent l'espace de service, la capacité et la discipline d'attente ainsi que la spécification concernant le processus de répétition d'appels.

Lors de l'étude des problèmes de la théorie de files d'attente on supposait que les serveurs

étaient absolument fiables. Cependant en pratique, on rencontre souvent des cas où les serveurs sont sujets à des pannes aléatoires et par conséquent, durant un certain intervalle de temps, le service des clients est interrompu.

Dans ce chapitre, nous allons présenter une synthèse bibliographique de certains résultats concernant les modèles de files d'attente avec rappels, ainsi que les différentes approches utilisées pour l'analyse de ce type de modèles. Cette synthèse actualise celle de OUKID [thèse de doctorat] [13].

1.2 Modèles d'attente avec rappels

Le modèle général d'un système de files d'attente avec répétition d'appels peut être décrit comme suit : Le système est composé de C ($C \geq 1$) serveurs identiques et indépendants, d'un espace d'attente de capacité $K - C$ ($K \geq 1$) et d'un orbite de capacité finie ou infinie. A l'arrivée d'un client, si tous les serveurs et positions d'attentes sont occupés le client rejoint l'orbite et devient source d'appels secondaires et il rappellera ultérieurement. Un client est admis au serveur uniquement lorsque le serveur est trouvé inactif et en bon état. Les clients dont le service est interrompu par une panne peuvent avoir la possibilité de rester sur le serveur jusqu'à ce que la réparation soit terminée, de laisser le système entièrement ou de rejoindre l'orbite pour répéter ou reprendre le service.

Tout client en orbite est supposé rappeler pour le service à des intervalles de temps suivant une loi de probabilité. Chacun de ces clients secondaires est traité comme un nouveau client qui arrive de l'extérieur du système (client primaire). La structure générale d'un système avec rappels est présentée dans la figure 1.1.

A notre connaissance, l'étude des modèles de files d'attente avec rappels remonte déjà aux travaux de KOSTEN (1947) [1] et CLOS (1948)[14]. Il faut attendre 1957 pour voir publier l'article de COHEN [15] lors de la modélisation d'un central téléphonique. Cet article contient une analyse systématique du modèle Markovien $M/M/m$. Une décennie plus tard, KEILSON, COZZOLINO et YOUNG [16] ont examiné le système $M/G/1$ et le système $M/M/2$ avec rappels. Pour le $M/G/1$, les clients qui trouvent le serveur occupé entrent en orbite et deviennent source d'appel secondaire. Puisque les temps de service sont supposés être généralement distribués, les auteurs utilisent la méthode des variables supplémentaires pour transformer le processus stochastique en un processus markovien. Le cas de $M/M/2$ est étudié par la résolution des

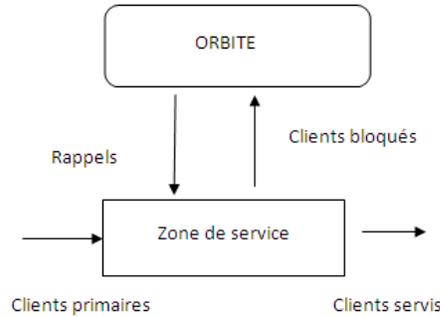


FIGURE 1.1 – La structure du système

équations de balance en utilisant l'équation de normalisation.

Ce n'est que vers les années 1970-1980 qu'on constate un regain d'intérêt pour cette catégorie de modèles, en raison de nouveaux développements technologiques des systèmes de télécommunications. Les systèmes avec rappels permettent en effet de mieux modéliser des protocoles spécifiques de communication, tels que CSMA (Carrier Sense Multiple Access) ou encore les disciplines Ring-back-when-free, Repeatlastnumber, Auto-repeat ALEXANDROV [17]. Une description complète de situations de systèmes de files d'attente avec rappels peut être trouvée dans la monographie de FALIN et TEMPLETON (1997) [6] et dans [4; 18–20]. Une classification bibliographique est donnée dans les articles de ARTALEJO (1999) [21] et (2010) [22].

1.2.1 Modèle d'attente M/G/1 avec rappels

Le modèle M/G/1 avec rappels est le modèle le plus étudié par les spécialistes. Il existe une littérature abondante sur ses diverses propriétés [23–27].

Description du modèle

Les clients arrivent dans le système selon un processus de Poisson de taux $\lambda > 0$. Le service des clients est assuré par un seul serveur. La durée de service τ suit une loi générale de fonction de répartition $B(x)$ et de transformée de Laplace-Stieltjes $\tilde{B}(s)$, $Re(s) > 0$. Soient les moments $\beta_k = (-1)^k \tilde{B}_i^{(k)}(0)$, l'intensité du trafic $\rho = \lambda\beta_1$ et $\mu = \frac{1}{\beta_1}$. La durée entre deux rappels successifs

d'une même source secondaire est exponentiellement distribuée de paramètre $\theta > 0$. Le système évolue de la manière suivante : On suppose que le $(i - 1)^{\text{ème}}$ appel termine son service à l'instant $t_{d_{i-1}}$ (les appels sont numérotés dans l'ordre de service) et le serveur devient libre. Même s'il y a des clients dans le système, ils ne peuvent occuper le service immédiatement. Donc le $i^{\text{ème}}$ appel suivant, n'entre en service qu'après un intervalle de temps R_i durant lequel le canal est libre, bien qu'en général il y ait des clients qui attendent. A l'instant $\xi_i = t_{d_{i-1}} + R_i$, le $i^{\text{ème}}$ client débute le service durant un temps S_i . Tous les rappels qui arrivent durant ce temps de service n'influent pas sur le processus. Alors à l'instant $t_{d_i} = \xi_i + S_i$, le $i^{\text{ème}}$ client achève son service et le canal devient encore libre et ainsi de suite.

Chaîne de Markov induite

L'état du système peut-être décrit par le processus

$$X(t) = \begin{cases} N(t), & \text{si } C(t) = 0, \\ \{C(t), N(t), \xi(t)\}, & \text{si } C(t) = 1. \end{cases} \quad (1.1)$$

où $C(t)$ est le nombre de serveurs occupés, $N(t)$ est le nombre de clients en orbite, $\xi(t)$ est une variable supplémentaire à valeurs dans \mathbb{R}^+ et désignant la durée de service écoulée à la date t .

ce processus n'est pas un processus de Markov, mais il possède une chaîne de Markov induite. Cette chaîne a été décrite pour la première fois par Choo et Conolly (1979) [28] . Soit (X_i) la chaîne de Markov induite aux instants de départs, où $X_i = N(t_{d_i})$ représente le nombre de clients en orbite après le $i^{\text{ème}}$ départ, dont l'équation fondamentale est :

$$X_{i+1} = X_i - B_i + v_{i+1}$$

où v_{i+1} est le nombre de clients primaires arrivant dans le système durant le service du $(i + 1)^{\text{ème}}$ client. Elle ne dépend pas des événements qui se sont produits avant l'instant ξ_{i+1} du début de service du $(i + 1)^{\text{ème}}$ client. La distribution de v_{i+1} est la suivante :

$$P(v_i = n) = k_n = \int_0^{\infty} \exp(-\lambda x) \frac{(\lambda x)^n}{n!} dB(x)$$

avec une fonction génératrice

$$K(z) = \sum_{n=0}^{\infty} z^n k_n = \tilde{B}(\lambda - \lambda z),$$

et une moyenne

$$E[v_i] = K'(1) = \sum_{n=0}^{\infty} nk_n = \rho.$$

La variable aléatoire B_i est une variable de Bernoulli définie par

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

Elle dépend de X_{i+1} et sa distribution conditionnelle est donnée par

$$P(B_i = 1/X_i = n) = \frac{n\theta}{\lambda + n\theta},$$

$$P(B_i = 0/X_i = n) = \frac{\lambda}{\lambda + n\theta}.$$

Les probabilités de transition de l'état n à l'état m ($m \geq 0$ et $0 \leq n \leq m$) sont

$$r_{nm} = \frac{\lambda}{\lambda + n\theta} k_{m-n} + \frac{n\theta}{\lambda + n\theta} k_{m-n+1}$$

La condition d'existence du régime stationnaire peut être obtenue comme suit :

L'accroissement moyen de la chaîne vaut

$$E[X_{i+1} - X_i/X_i = n] = E[v_{i+1}] - E[B_i = 1/X_i = n] = \rho - \frac{n\theta}{\lambda + n\theta}.$$

Si $\rho < 1$, alors $\lim_{n \rightarrow \infty} E[X_{i+1} - X_i/X_i = n] = \rho - 1$ et la chaîne est donc ergodique. Par contre, si $\rho \geq 1$, alors $\lim_{n \rightarrow \infty} E[X_{i+1} - X_i/X_i = n] = \rho - \frac{n\theta}{\lambda + n\theta} \geq 1 - \frac{n\theta}{\lambda + n\theta} = \frac{\lambda}{\lambda + n\theta} > 0$. Puisque la chaîne est bornée inférieurement par la chaîne induite du système M/G/1 classique, donc la chaîne n'est pas ergodique (elle est transitoire).

Théoreme 1.2.1. [29] Soit $\rho < 1$. La distribution stationnaire de la chaîne de Markov induite possède la fonction génératrice suivante :

$$\varphi(z) = \sum_{i=0}^{\infty} z^i \pi_i = \frac{(1-\rho)(1-z)\tilde{B}(\lambda-\lambda z)}{\tilde{B}(\lambda-\lambda z) - z} \exp \left\{ \frac{\lambda}{\theta} \int_1^z \frac{1 - \tilde{B}(\lambda - \lambda u)}{\tilde{B}(\lambda - \lambda u) - u} du \right\}.$$

Distribution stationnaire de l'état du système

L'état du système au temps t est décrit par le processus (1.1). Lorsque $\rho < 1$, le système est en régime stationnaire. Dénotons par

$$P_{0i} = \lim_{t \rightarrow \infty} P(C(t) = 0, N(t) = i);$$

$$P_{1i}(x) = \lim_{t \rightarrow \infty} \frac{d}{dx} P(C(t) = 1, \xi(t) \leq x, N(t) = i);$$

$$P_{1i} = \int_0^{\infty} P_{1i}(x) dx.$$

Théoreme 1.2.2. *Les fonctions génératrices de la distribution stationnaire conjointe de l'état du serveur et de la taille de l'orbite en régime stationnaire ($\rho < 1$) sont données par*

$$P_0(z) = \sum_{i=0}^{\infty} z^i p_{0i} = (1 - \rho) \exp \left\{ \frac{\lambda}{\theta} \int_1^z \frac{1 - \tilde{B}(\lambda - \lambda u)}{\tilde{B}(\lambda - \lambda u) - u} du \right\}, P_1(z) = \sum_{i=0}^{\infty} z^i p_{1i} = \frac{1 - \tilde{B}(\lambda - \lambda z)}{\tilde{B}(\lambda - \lambda z) - z} P_0(z).$$

Par conséquent, la distribution marginale du nombre de serveurs occupés s'exprime de la manière suivante

$$P_0 = \lim_{t \rightarrow \infty} P(C(t) = 0) = P_0(1) = 1 - \rho,$$

$$P_1 = \lim_{t \rightarrow \infty} P(C(t) = 1) = P_1(1) = \rho.$$

La fonction génératrice de la distribution marginale de la taille de l'orbite est définie par

$$P(z) = P_0(z) + P_1(z) = \frac{1 - z}{\tilde{B}(\lambda - \lambda z) - z} P_0(z),$$

et la fonction génératrice de la distribution de l'état stationnaire du nombre de clients dans le système est

$$Q(z) = \frac{(1 - \rho)(1 - z)\tilde{B}(\lambda - \lambda z)}{\tilde{B}(\lambda - \lambda z) - z} \exp \left\{ \frac{\lambda}{\theta} \int_1^z \frac{1 - \tilde{B}(\lambda - \lambda u)}{\tilde{B}(\lambda - \lambda u) - u} du \right\}.$$

Mesures de performance

Les caractéristiques du modèle sont [30]

– Nombre moyen de clients dans le système

$$\bar{n} = Q'(1) = \rho + \frac{\lambda^2 \beta_2}{2(1 - \rho)} + \frac{\lambda \rho}{\theta(1 - \rho)}.$$

– Nombre moyen de clients en orbite

$$\bar{n}_0 = P'(1) = \bar{n} - \rho = \frac{\lambda^2 \beta_2}{2(1 - \rho)} + \frac{\lambda \rho}{\theta(1 - \rho)}.$$

– Temps moyen d'attente d'un client

$$\bar{W} = \frac{\bar{n}_0}{\lambda} = \frac{\lambda \beta_2}{2(1 - \rho)} + \frac{\rho}{\theta(1 - \rho)}.$$

– Nombre moyen de rappels par client (d'après la formule de Little)

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

1.3 Modèles d'attente avec rappels et interruptions de service

Les systèmes de files d'attente avec interruptions du service sont courants dans la modélisation des systèmes informatiques et dans les réseaux de télécommunication [31]. D'un point de vue pratique, il est plus réaliste de considérer les files d'attente avec des répétitions d'appels et le serveur exposé à des interruptions aléatoires qui empêchent la continuité du service des clients. Les modèles d'attente avec des interruptions de service se sont révélés très utiles dans les situations où un serveur est partagé par plusieurs files d'attente, lorsque le serveur est sujet à des pannes ou l'arrivée de clients de haute priorité.

Ces systèmes ont été étudiés par de nombreux auteurs. FIEMS et al (2008) [32] ont considéré le système M/G/1 avec divers types d'interruptions et KERNANE (2013) [33] a généralisé ce travail au cas d'un système avec rappels. GRAVER [34] considère le cas d'interruption avec préemption (l'interruption intervient pendant le service d'un client) où le client reprend le service à partir du point d'interruption ou dès le début.

1.3.1 Modèles d'attente avec rappels et pannes

Les systèmes d'attente avec rappels et interruptions dues aux pannes aléatoires des serveurs sont d'une grande importance et ils ont les mêmes caractéristiques que les modèles précédemment étudiés dans la section (1.2). Ce qui les distingue, est le risque des pannes qui présentent un moyen supplémentaire pour qu'un client arrivant rejoint l'orbite. Les types de pannes de serveur les plus rencontrés en littérature sont ARTALEJO [35] :

- Pannes avec perte momentanée du client : Dès que la panne se produit, le client entre en orbite.
- Pannes avec perte définitive du client : Dès que la panne se produit, le client quitte le système définitivement.
- Pannes conservatrices : Dès que la panne se produit, le client attend que la réparation soit terminée et le service reprend en tenant compte de la partie de service déjà acquise.
- Pannes non conservatrices : Le client attend que la réparation soit terminée, mais le service doit-être repris à zéro.

Ces types de systèmes ont été étudiés indépendamment par Aissani [2] en 1988 et par KULKARNI et CHOI [36] en 1989. AISSANI [2] a considéré un système de file d'attente M/G/1/1 avec rappels et un serveur non fiable. KULKARNI et CHOI ont considéré deux modèles différents

de types M/G/1. Dans le premier modèle, un client dont le service est interrompu par une défaillance du serveur rejoint l'orbite avec une probabilité P ou quitte le système avec une probabilité $1 - P$. Le deuxième modèle permet au client de rester à la station service pendant que le serveur est en réparation et le service est redémarré une fois la réparation terminée.

AISSANI et ARTALEJO [37] ont fait une extension des résultats de KULKARNI et CHOI [36] en se concentrant sur la fiabilité d'un système M/G/1 avec rappels et pannes. En 1993 AISSANI [38] a continué son travail sur le système M/G/1 avec rappels et cette fois en considérant des hypothèses plus générales et il a obtenu la distribution du nombre de clients en orbite et du nombre de clients dans le système à l'état stationnaire en appliquant la méthode des variables supplémentaires.

Une variété importante de système M/G/1 avec rappels non fiable existent. YANG et LI ont écrit deux ouvrages [39] et [40] dans lesquels ils ont étudié davantage le système M/G/1 avec rappels et pannes et le système avec rappels et vacances du serveur. De leur côté, J.WANG, J.CAO et Q.LI [41] en 2001, considèrent le modèle M/G/1 avec rappels, pannes et réparations, ils donnent une analyse détaillée de la fiabilité du serveur en utilisant la méthode des variables supplémentaires, ils ont obtenus les expressions explicites de certains indices de fiabilité comme la disponibilité et la fréquence de défaillance. DJELLAB [42–44] en 2002 a étudié la propriété de décomposition stochastique du système de files d'attente M/G/1 avec rappels et serveur non fiable dans le cas de la distribution générale du temps des inter-rappels, et elle a aussi analysé les effets du type de distribution du temps des inter-rappels, de l'intensité des rappels ainsi que des pannes sur la performance du modèle.

Certains auteurs ont étudié les systèmes M/M/1 avec rappels et pannes. En 2005, AL-MASI, ROSZIK et SZTRIK [45] considéraient un système de source finie avec des pannes et des réparations, le serveur sujet à des pannes actives et passives. Les durées de pannes sont généralement distribuées et les durées de réparation sont exponentielles. Les clients préemptés par une panne peuvent choisir de rester sur le serveur et de reprendre le service une fois la réparation terminée, ou de rejoindre l'orbite. Les durées de rappels sont de taux exponentiel. Les auteurs ont obtenu certaines mesures de tels systèmes à l'état stationnaire. LI et ZHAO [46] ont assumé un système M/M/1 avec deux files d'attente à la fois pour les clients en attente ou préemptés. Toutes les durées sont distribuées exponentiellement et seules les pannes actives sont considérées. En 2006 SHERMAN et KHAROUFEH [47] ont étudié le système M/M/1 avec rappels et pannes actives et passives, un orbite de capacité infinie et une file d'attente. Ils ont

fournit la condition de stabilité ainsi que plusieurs résultats de décomposition stochastique.

Plus récemment, JAIN et al. [48] en 2007 ont étudié le système d'attente M/G/1 avec rappels, pannes et lancement. D'autre part, FALIN [49] en 2010 a étudié un modèle d'attente avec rappels et pannes du type M/G/1, il a trouvé une distribution conjointe de l'état de serveur, le nombre de client dans la file et le nombre de client dans le groupe de rappel dans l'état stable quand le temps de service et le temps de réparation ont une distribution générale en utilisant la méthode de variables supplémentaires.

Modèle d'attente M/G/1 avec rappels et pannes

Description du modèle

Le système considéré est M/G/1 avec rappels et pannes. Le flot des arrivées primaires est poissonnien de taux λ . Le temps de service est distribué suivant une loi arbitraire, de fonction de répartition $B(x)$ et de transformée de Laplace-Stieltjes $\tilde{B}(s)$, $Re(s) > 0$. Nous supposons qu'il n'y a pas d'espace d'attente et donc si un client primaire trouve le serveur libre, il est immédiatement pris en charge. Sinon, il entre en orbite et devient source d'appels secondaires. La durée entre deux rappels successifs d'une même source secondaire est exponentielle de paramètre θ . Le flux d'entrée des arrivées primaires, les temps de service et les intervalles entre les essais répétés sont supposés mutuellement indépendants. En outre, le système est soumis à deux types d'interruptions poissonniennes, actives et inactives. Dans ce modèle, Le client qui est incapable d'accéder à un serveur à cause de la panne ou la congestion peut choisir d'entrer en orbite avec une probabilité P_1 ou de quitter le système avec une probabilité $(1 - P_1)$. On suppose aussi que les clients dont le service est interrompu par une panne active, doivent décider, soit de rejoindre le groupe de rappels avec une probabilité P_a ou de quitter définitivement le système après interruption avec une probabilité $(1 - P_a)$. Le flot des pannes du serveur lorsqu'il est inactif ($C(t) = 0$) est poissonnien de taux α_0 (respectivement α_1 , lorsque le serveur est actif). Les temps de réparation sont des variables aléatoires d_0 de fonction de répartition $R_0(x)$ et de transformée de Laplace-Stieltjes $\tilde{R}_0(s)$ dans le cas d'une panne passive et des variables aléatoires d_1 de fonction de répartition $R_1(x)$ et de transformée de Laplace-Stieltjes $\tilde{R}_1(s)$ dans le cas d'une panne active. Ces nouvelles variables aléatoires et tous les processus vérifient l'hypothèse d'indépendance mutuelle.

L'état du système peut-être décrit par le processus

$$X(t) = \{C(t), N(t), t \geq 0\},$$

où $N(t)$ est le nombre de clients en orbite à l'instant t et $C(t)$ (l'état du serveur à l'instant t) est défini comme suit :

$$X(t) = \begin{cases} 0 & \text{si le serveur est libre et en bon état à l'instant } t \\ 1 & \text{si le serveur est occupé et en bon état à l'instant } t \\ 2 & \text{si le serveur est en panne à l'instant } t \end{cases} \quad (1.2)$$

ce processus n'est pas un processus de Markov, Pour le rendre markovien, on introduit une variable supplémentaire $\xi(t)$ à valeurs dans \mathfrak{R}^+ avec

$$\xi(t) = \begin{cases} \text{est la durée de service écoulée si } C(t) = 1 \\ \text{est la durée de réparation écoulée si } C(t) = 2 \end{cases} \quad (1.3)$$

Soit la chaîne de Markov induite $\{N_i, i \geq 0\}$ où $N_i = N(\xi_i)$ est le nombre de clients en orbite à la date ξ_i lorsque le serveur devient libre et en bon état pour la $i^{\text{ème}}$ fois. La fonction génératrice de la distribution stationnaire de la chaîne de Markov induite N_i est donnée par [36]

$$\varphi(z) = K \frac{\alpha(z)}{\beta(z)} \exp \left\{ \frac{-1}{\theta} \int_1^z \frac{1}{u} \frac{(\lambda + \alpha_0)\beta(u) - \alpha(u)}{\beta(u)} du \right\},$$

où

$$\beta(z) = (1 - \tilde{\beta}(z))$$

$$\alpha(z) = \tilde{\alpha}(z) - (\alpha_0 + \lambda)\tilde{\beta}(z),$$

$$\tilde{\alpha}(z) = \alpha_0 \tilde{R}_0(\lambda - \lambda z) + \lambda \alpha_1 \frac{1 - \tilde{\beta}(\lambda - \lambda z + \alpha_1) \tilde{R}_1(\lambda - \lambda z)(1 - P_a(1 - z))}{\lambda - \lambda z + \alpha_1} + \lambda \tilde{\beta}(\lambda - \lambda z + \alpha_1)$$

$$\tilde{\beta}(z) = \alpha_1 \frac{(1 - \tilde{\beta}(\lambda - \lambda z + \alpha_1)) \tilde{R}_1(\lambda - \lambda z)}{\lambda - \lambda z + \alpha_1} (P_a + \frac{1 - P_a}{z}) + \frac{1}{z} \tilde{\beta}(\lambda - \lambda z + \alpha_1),$$

$$K = \frac{1 - \rho}{\lambda(1 + \alpha_0 E(d_0)) + \alpha_0(1 - \rho)}$$

$$\rho = \lambda \frac{1 - \tilde{\beta}(\alpha_1)}{\alpha_1} \left[1 + \alpha_1 (E(d_1) + \frac{P_a}{\lambda}) \right].$$

Distributions stationnaires

Soit $N_t; t \geq 0$ un processus stochastique, qui représente le nombre de clients en orbite à la date t . La fonction génératrice de la distribution stationnaire de $N_t, t \rightarrow \infty$ est donnée par [36]

$$Q(z) = P_0(z) + P_1(z) + P_2(z),$$

$$\begin{aligned}
P_0(z) &= \frac{1}{A}M(z), \\
P_1(z) &= \frac{1}{A} \left(\frac{1 - \tilde{\beta}(\lambda - \lambda z + \alpha_1)}{\lambda - \lambda z + \alpha_1} \right) (\lambda M(z) + \theta M'(z)), \\
P_2(z) &= \frac{1}{A} \left[\frac{1 - \tilde{R}_0(\lambda - \lambda z)}{\lambda - \lambda z} \alpha_0 M(z) + M(z)N(z)(\lambda(1 - P_a) + \lambda P_a z) + M'(z)N(z)(\theta(1 - P_a) + \theta P_a z) \right], \\
N(z) &= \alpha_1 \frac{\tilde{\beta}(\lambda - \lambda z + \alpha_1)}{\lambda - \lambda z + \alpha_1} \frac{1 - \tilde{R}_1(\lambda - \lambda z)}{\lambda - \lambda z}, \\
M(z) &= \sum_{j=0}^{\infty} \frac{\pi_j z^j}{\lambda + \alpha_0 + j\theta}, \\
A &= K(1 + \alpha_0 E(d_0)) + (1 - \tilde{\beta}(\alpha_1)) (1 - \alpha_0 K) \left(E(d_1) + \frac{1}{\alpha_1} \right),
\end{aligned}$$

Donc

$$\begin{aligned}
Q(z) &= \frac{1}{A} [M(z) + \frac{1 - \tilde{\beta}(\lambda - \lambda z + \alpha_1)}{\lambda - \lambda z + \alpha_1} (\lambda M(z) + \theta M'(z)) + \frac{1 - \tilde{R}_0(\lambda - \lambda z)}{\lambda - \lambda z} \alpha_0 M(z) \\
&\quad + M(z)N(z)(\lambda(1 - P_a) + \lambda P_a z) + M'(z)N(z)(\theta(1 - P_a) + \theta P_a z)].
\end{aligned}$$

1.3.2 Modèles d'attente avec rappels et priorité

De nos jours, tout en analysant les scénarios de files d'attente, presque toutes les zones sont affectées par le service prioritaire d'un client par rapport à un autre client. La priorité peut être préemptive ou non préemptive. Dans le cas de priorité préemptive, le service d'un client moins prioritaire sera suspendu lorsqu'un client plus prioritaire se présente devant le serveur et le client suspendu reprend son service là où il a été interrompu, ou il le reprend depuis le début. Si la priorité est non préemptive, un nouveau client plus prioritaire attend la fin du service avant de commencer le sien. Plusieurs auteurs ont étudié les systèmes d'attente avec rappels et priorités. CHOI et PARK (1990) [50] optent pour la méthode de la variable supplémentaire pour étudier une file d'attente $M_1, M_2/G/1$ avec rappels à deux types d'appels, avec une file d'attente prioritaire infinie pour le type d'appels I et un orbite infini de type d'appels II et ils ont obtenu la fonction génératrice commune des longueurs de la file d'attente. Ces même mesures de performance ont été obtenues par MOUTZOUKIS et LANGARIS (1995) [51] pour un modèle avec deux types d'appels et priorité préemptive, alors que FALIN et al. (1993) [52] ont étendu les résultats du CHOI et PARK [50] du cas où deux types d'appels peuvent avoir différentes distributions du temps de service. CHOI et CHANG (1999) [53] ont analysé divers systèmes d'attente avec rappels et priorité et d'autres caractéristiques comme la perte géométrique, les

pannes, etc. ARTALEJO et al. (2001) [54] ont effectué une analyse stationnaire d'un système d'attente avec rappels préemptifs. Les modèles prioritaires M/M/1 et M/M/1/1 ont été utilisés dans [54] pour discuter l'influence de divers critères d'accès sur le débit des clients de type 2. Récemment, ARTALEJO et PHUNG-DUC (2011) [55] ont effectué une étude détaillée de la file d'attente M/M/1 avec rappels à communication bidirectionnelle. Ils ont obtenu des expressions explicites pour la distribution stationnaire du nombre de clients en orbite, ainsi que les moments partiels. En (2013), ARTALEJO et al [56] ont étudié le comportement d'un état d'équilibre d'une file M/G/1 avec rappels dans laquelle il y'a deux flux d'arrivée entrants à savoir les appels effectués par des clients réguliers et les appels sortants effectués par le serveur lorsqu'il est inactif. Ils ont effectué une analyse stationnaire du système, y compris la condition de stabilité, la chaine de Markov induite, la distribution stationnaire de l'état du serveur, le nombre de clients en orbite et le calcul des premiers moments. Ils ont aussi obtenu les résultats asymptotiques pour le nombre de clients en orbite. DIMITRIOU (2013) [57] a étudié le système d'attente avec rappels, priorités, arrivées négatives, un serveur non fiable et des vacances multiples et a exploré les conditions de stabilité ainsi que les probabilités d'état du système. BOUTERFA (2015) et al. [12] ont réalisé l'analyse stochastique de fles d'attente $M_1, M_2/G_1, G_2/1$ avec rappels et priorité préemptive.

Modèle d'attente avec rappels de type M/G/1 et priorité non préemptive

Description du modèle

Le système considéré est formé d'une file de capacité infinie, et d'un unique serveur. Deux types de clients (type1-prioritaire et type2-non prioritaire) arrivent dans le système selon un processus de Poisson de taux λ_1 et λ_2 respectivement. Les durées de services des deux types des clients sont des variables aléatoires positives et de lois générales différentes de fonction de répartition $B_i(x)$ et de transformée de Laplace-Stieltjes $\tilde{B}_i(s)$, $Re(s) > 0$ et $i = 1, 2$. Soient les moments $\beta_{ik} = (-1)^k \tilde{B}_i^{(k)}(0)$ et l'intensité instantanée du service est

$$b_i(x) = \frac{B'_i(x)}{1 - B_i(x)}, i = 1, 2.$$

La durée entre deux rappels d'une même source secondaire est exponentielle de taux θ . Le comportement d'un client, qui trouve à son arrivée le serveur occupé par le service d'un client, dépend de son type de priorité. S'il est prioritaire, il prend place dans la file d'attente. S'il est

non prioritaire, il entre en orbite et devient source d'appels secondaires (rappels). Toutes les variables définies précédemment sont supposées mutuellement indépendantes.

Soit $\rho = \rho_1 + \rho_2$, où $\rho_i = \lambda_i \beta_{i,1}$, $i = 1, 2$ qui est la charge contribuéée par des clients de type i . De Choi et Chang [53], la condition pour que le système se trouve en régime stationnaire est $\rho < 1$. On a également

$$k_{i,m,n} = \int_0^\infty \frac{(\lambda_1 x)^m}{m!} e^{-\lambda_1 x} \frac{(\lambda_2 x)^n}{n!} e^{-\lambda_2 x} dB_i(x)$$

la distribution conjointe du nombre de clients primaires de deux types arrivant dans le système durant le temps de service d'un client de type $i = 1, 2$. Sa fonction génératrice est donnée par :

$$K_i(z_1, z_2) = \tilde{B}_i(\lambda_1 - \lambda_1 z_1 + \lambda_2 - \lambda_2 z_2).$$

Distributions stationnaires

L'état du système à la date t peut être décrit à l'aide d'un processus de Markov à temps continu, tel que

$$X(t) = \{C(t), N_1(t), N_2(t), \xi(t), t \geq 0\},$$

$C(t)$: La variable aléatoire indiquant l'état du serveur à l'instant t .

$C(t) = 0$ si le serveur est libre à l'instant t .

$C(t) = 1$ si le serveur est occupé à l'instant t par un client de type 1.

$C(t) = 2$ si le serveur est occupé à l'instant t par un client de type 2.

$N_1(t)$: Le nombre de clients de type 1 dans la file d'attente à l'instant t .

$N_2(t)$: Le nombre de clients de type 2 dans l'orbite à l'instant t .

$\xi(t)$ est une variable supplémentaire à valeurs dans \mathfrak{R}^+ et désignant la durée de service écoulée à la date t .

Supposons $\rho < 1$, nous présentons le système d'équations de balance comme suit

$$(\lambda_1 + \lambda_2 + j\theta)P_{00j} = \int_0^\infty P_{10j}(x)b_1(x)dx + \int_0^\infty P_{20j}(x)b_2(x)dx, i = 0, j \geq 0;$$

$$P'_{1ij}(x) = -(\lambda_1 + \lambda_2 + b_1(x))P_{1ij}(x) + \lambda_1 P_{1i-1j}(x) + \lambda_2 P_{1ij-1}(x);$$

$$P_{1ij}(0) = \lambda_1 P_{00j} \delta_{i0} + \int_0^\infty P_{1i+1j}(x)b_1(x)dx + \int_0^\infty P_{2i+1j}(x)b_2(x)dx;$$

$$P'_{2ij}(x) = -(\lambda_1 + \lambda_2 + b_2(x))P_{2ij}(x) + \lambda_1 P_{2i-1j}(x) + \lambda_2 P_{2ij-1}(x);$$

$$P_{2ij}(0) = \lambda_2 P_{00j} + (j+1)\theta P_{00j}, i=0, j \geq 0;$$

où δ_{i0} est la fonction de Kronecker.

En appliquant la méthode des fonctions génératrices nous obtenons les probabilités stationnaires P_{0nm} , P_{1nm} , et P_{2nm} ,

$$P_0(z_2) = (1 - \rho_1 - \rho_2) \exp \left\{ \frac{1}{\theta} \int_1^{z_2} \frac{\lambda_1 - \lambda_1 h(u) + \lambda_2 - \lambda_2 K_2(h(u), u)}{K_2(h(u), u) - u} du \right\};$$

$$\begin{aligned} P_1(z_1, z_2) &= [(\lambda_1 - \lambda_1 h(z_2) + \lambda_2 - \lambda_2 K_2(h(z_2), z_2))(K_2(z_1, z_2) - z_2) \\ &\quad - (\lambda_1 - \lambda_1 z_1 + \lambda_2 - \lambda_2 K_2(z_1, z_2))(K_2(h(z_2), z_2) - z_2)] \\ &\quad \times \frac{1}{(K_2(h(z_2), z_2) - z_2)(z_1 - K_1(z_1, z_2))} \times \frac{1 - K_1(z_1, z_2)}{\lambda_1 - \lambda_1 z_1 + \lambda_2 - \lambda_2 z_2} P_0(z_2); \end{aligned}$$

$$P_2(z_1, z_2) = \frac{\lambda_1 - \lambda_1 h(z_2) + \lambda_2 - \lambda_2 z_2}{K_2(h(z_2), z_2) - z_2} \times \frac{1 - K_2(z_1, z_2)}{\lambda_1 - \lambda_1 z_1 + \lambda_2 - \lambda_2 z_2} P_0(z_2);$$

Mesures de performance

A l'aide des fonctions génératrices $P_0(z_2)$, $P_1(z_1, z_2)$ et $P_2(z_1, z_2)$ on obtient les mesures de performance suivantes :

- Le nombre moyen de clients de type 1 dans le système

$$N_1 = \frac{\lambda_1(\lambda_1 \beta_{12} + \lambda_2 \beta_{22})}{2(1 - \rho_1)}.$$

- Le nombre moyen de clients de type 2 dans le système

$$N_2 = \frac{\lambda_2(\lambda_1 \beta_{12} + \lambda_2 \beta_{22})}{2(1 - \rho_1)(1 - \rho_1 - \rho_2)} \frac{\lambda_2(\rho_1 + \rho_2)}{\theta(1 - \rho_1 - \rho_2)}.$$

- La probabilité que le serveur soit occupé

$$P(C(t) = 1) = \rho_1 + \rho_2.$$

1.4 Autres modèles

Les systèmes avec rappels à source finie sont étudiés par FALIN et ARTALEJO [58] qui ont fait l'analyse de la distribution des arrivées, la période d'occupation et le temps d'attente. Les périodes d'activité et d'inactivité de l'orbite d'un système $M/G/1$ ont fait l'objet de l'article de ARTALEJO et FALIN [59] où ils obtiennent la probabilité conjointe des deux périodes et le nombre correspondant de clients servis pendant ces périodes ainsi que la relation entre la période d'activité de l'orbite et celle du système.

Les systèmes à arrivées par groupes ont été étudiés par FALIN [60] qui a considéré le cas des clients persistants. Plus tard plusieurs auteurs ont présenté un algorithme d'approximation pour un système avec arrivées par lots et clients non persistants. Une analyse détaillée de ce modèle a été donnée par FALIN [61] qui a étudié le régime transitoire et la période d'activité. Une autre approche a été proposée par YANG et TEMPLETON [62] où des expressions analytiques des caractéristiques probabilistes du modèle ont été obtenues.

Les systèmes à multiserveurs sont décrits par BOOTS et TIJMS [63] ainsi que par HAN-SCHKE [64]. Ce dernier a montré que les fonctions génératrices partielles des probabilités d'états stationnaires peuvent-être exprimées en termes de fonctions hypergéométriques généralisées. Dans son article, il propose un test simple pour obtenir les conditions suffisantes pour l'ergodicité des systèmes avec rappels et multi-serveurs. L'investigation sur les files d'attente à multi-serveurs est plus difficile que celle des files d'attente à un seul serveur. Néanmoins, les files d'attente avec rappels et multi-serveurs sont plus flexibles et applicables en pratique que les files d'attente avec rappels et un seul serveur. Pour les files d'attente avec rappels et multi-serveurs, les lecteurs peuvent se référer à NEUTS et RAO [65], CHOI et al. [66], ARTALEJO et POZO [67], KRISHNA KUMAR et RAJA [68], KIM et al. [69] et d'autres.

1.5 Exemples de systèmes modélisés par des files d'attente avec rappels

1.5.1 Problème de réservation

C'est l'exemple le plus simple d'un client qui sollicite une réservation par téléphone dans un restaurant. Il y a une ligne unique qui est consacrée à répondre aux requêtes des réservations. Ainsi, si un client appelle et trouve la ligne occupée, il renouvellera sa tentative après une certaine période de temps aléatoire avec la probabilité p qui, en pratique, est strictement

inférieure à 1 car le client ne peut rappeler indéfiniment. Cet exemple peut être modélisé par une file d'attente $M/G/1$ avec rappels et avec perte en considérant que le processus d'arrivées des appels est poissonnien. L'étude de ce genre de problèmes permet de prédire le temps d'attente du client, le nombre de clients perdus dû à ce blocage,...

1.5.2 Système informatique à temps réel

Dans un système informatique à temps réel, on trouve M terminaux et S canaux de transmission tels que $M > S$. Pour qu'un terminal soit connecté à l'ordinateur, il suffit d'un canal de transmission libre. L'illustration de ce genre de système est le centre de calcul où arrive un étudiant pour utiliser l'ordinateur pendant une période de temps aléatoire. Celui-ci doit d'abord trouver un terminal libre pour se connecter. S'il n'y a aucun terminal disponible, il retentera sa chance après un temps aléatoire. Sinon, il envoie sa demande au commutateur central pour se connecter à l'ordinateur. Le terminal est alors connecté selon que le canal serait disponible ou pas. Dans ce dernier cas, la demande est mise dans la file par le commutateur en attente de libération d'un canal. Ce système peut être modélisé par une file $G/G/S$ avec rappels, avec un tampon (espace d'attente) de capacité M et une orbite de taille infinie, où les canaux de transmission correspondent aux serveurs et les terminaux au tampon.

1.5.3 Le protocole de communication CSMA

Un réseau local simple est composé de stations interconnectées par un bus unique qui est le canal de communication. Ainsi, les stations communiquent les unes avec les autres via le bus qui peut être utilisé par une seule station à la fois. Un des protocoles de communication le plus généralement utilisé dans les réseaux locaux est le protocole non-persistant CSMA qui est une méthode d'accès à un réseau local fondée sur la reconnaissance par les différentes stations de l'état de liaison (libre ou occupé) à un instant donné. Sous ce protocole, des messages de longueur variable arrivent aux stations du monde extérieur. A la réception d'un message, la station le découpe en un nombre de paquets (de longueur fixe) et vérifie immédiatement si le bus est occupé ou libre. S'il est libre, l'un de ces paquets est transmis via ce bus à la station de destination, et les autres paquets sont stockés dans le tampon pour une transmission ultérieure. Si par contre, le bus est occupé, tous les paquets sont stockés dans le buffer, et la station peut réessayer la transmission après une période de temps dite temps de rappels. Ce problème peut être modélisé comme un système de files d'attente avec rappels et un seul serveur, qui est le

bus, et les buffers des stations représentent l'orbite.

1.6 Exemples de systèmes modélisés par des files d'attente avec rappels et interruptions de service

1.6.1 Systèmes réels modélisables comme un système de files d'attente avec rappels et priorité

Systèmes téléphoniques modernes

Dans les échanges téléphoniques modernes, les lignes des abonnés sont connectées aux modules qui traitent les appels arrivants et ceux sortants. Dans le cas du blocage (les canaux sont occupés), les appels sortants sont placés en attente dans un tampon (de capacité infinie) tandis que les appels arrivants sont refusés et plus tard réinitialisés pour établir la connexion. Lorsque l'un des canaux se libère, un appel sortant (s'il se trouve dans le tampon) l'occupe immédiatement. De ce fait, les appels arrivants ne réussissent pas à établir la connexion aussi longtemps qu'il y a des appels dans le tampon. Ce comportement implique que les appels sortants possèdent une priorité non préemptive sur les appels arrivants.

Réseaux de communication

Dans des réseaux de communication, des messages de longueur variables arrivent à une station de service qui les traite dans deux étapes par un seul serveur. A l'étape de prétraitement (le service préliminaire), le serveur peut interrompre le message actuel du processus en cours pour satisfaire le nouveau message prioritaire en arrivée. Après l'achèvement de prétraitement, si le message est qualifié, le traitement principal commencera et si autrement, il doit quitter le système. Le serveur est bloqué pendant le traitement principal pour qu'un message arrivant soit acheminé à l'orbite pour faire des rappels. Ce système a été modélisé par un système M/G/1, service à deux phases et priorité préemptive dans[70].

1.6.2 Systèmes réels modélisables comme un système de files d'attente avec rappels et serveurs non fiables

Transmission des messages à travers des réseaux fac-similes

Le modèle M/G/1 avec rappels et pannes peut être utilisé pour étudier le problème provenant des possibilités "autorepeat" pour la transmission des messages à travers des réseaux fac-similes. Considérons une entreprise de communication. Cette dernière réserve une machine fac-simile, spécifique pour envoyer des messages aux destinataires se trouvant à l'extérieur de l'entreprise en question. A l'arrivée d'un utilisateur désirant envoyer un message, si le medium de transmission est libre, la transmission commence. Dans le cas contraire, le message est soit stocké dans un tampon et la demande sera réinitialisée après une certaine durée aléatoire ou quitte le medium pour toujours. Si la transmission d'un message ne se termine pas avec succès pour des raisons telles qu'un blocage dans les lignes extérieurs ou une erreur de transmission, le message en question quitte le medium et, également, rejoint le tampon. Il est utile d'indiquer que l'équipement fac-simile reste opérationnel et les interruptions du service sont dues aux facteurs extérieurs. La correspondance avec le modèle est la suivante : les clients en orbite, la distribution du temps inter-rappels et la distribution du temps de service sont, respectivement, les messages bloqués/interrompus et stockés dans le tampon, le protocole de retransmission et le temps de transmission.

Réseaux informatiques avec infection virale

Quand il n'y a aucun virus, les réseaux informatiques peuvent être modélisés et analysés en utilisant des réseaux de files d'attente conventionnels avec ou sans rappels. Les clients représentent l'exécution des opérations algorithmiques et logiques, et les noeuds représentent les dispositifs CPUS, I/O, . . . etc. Quand un virus entre dans un noeud, un ou plusieurs fichiers peuvent être infectés et l'administrateur système devrait passer un certain nombre de sauvegardes afin de récupérer les fichiers infectés. Dans certains cas, ils ne peuvent pas être récupérables. Un virus peut émaner de l'extérieure du réseau, par exemple, disquette, ou peut venir d'un autre noeud dans le réseau, par exemple, un courrier électronique. Le temps de récupération des fichiers infectés peut être considéré comme le temps de réparation du serveur (le noeud) par suite d'une arrivée de client négatif (le virus). Ce cas réel a été modélisé par un système Géo/Géo/1 avec rappels, clients négatifs et serveur non fiable dans [71] où les auteurs ont étudié l'impact des clients négatifs et le facteur de manque de fiabilité sur les performances du système.

Interaction hôte-parasite (application biologique)

Considérons le cas d'un seul hôte (un organisme qui héberge un parasite) qui est né, sans parasite, au temps $t=0$ et dont sa durée de vie est exposée aux parasites de temps en temps dans une zone infectée. A un instant d'exposition, l'hôte acquiert un seul parasite indépendamment d'une exposition à une autre. Le nombre de parasites dans l'hôte peut augmenter en raison de la reproduction de parasites et peut provoquer la mortalité de l'hôte (quand il atteint un certain nombre m) à l'âge t_0 . L'intérêt du biologiste est au nombre de parasites acquis par l'hôte jusqu'à l'instant t quand l'hôte a été déplacé vers une zone non infectée à un certain âge $t_0 < t$. L'hôte entre dans la zone non infectée seulement s'il est vivant et infecté à l'âge t_0 . Dans la zone non infectée, l'hôte vit dans des conditions non infectieuses. Cela signifie que l'hôte n'acquiert pas de nouveaux parasites et subit un traitement concret pour diminuer le nombre de parasites acquis jusqu'à l'instant de déplacement. L'interaction précédemment décrite peut être modélisé par un système de files d'attente avec rappels [72] (les tentatives faites par un parasite à fin de s'héberger dans l'hôte qui lui résiste), un orbite (la zone infectée) de capacité infinie et un seul serveur (l'hôte) sujet à des pannes (infection par m ou plus parasites) réparables (le traitement de diminution de parasites représente la réparation de l'hôte).

1.7 Conclusion

Dans ce chapitre nous avons effectué une synthèse bibliographique sur les modèles d'attente non fiables et présenté les résultats existants ainsi que les différentes approches utilisées pour l'analyse de ce type de modèles. Une attention particulière a été accordée au modèle d'attente $M/G/1$ avec rappels et $M/G/1$ avec rappels et interruptions de services (interruptions dues aux pannes aléatoires de serveurs ou à des priorités). En passant en revue certaines recherches effectuées sur les modèles avec rappels non fiables, nous avons constaté que leur analyse présente de grandes difficultés analytiques qui augmentent avec le nombre de serveurs ou lorsqu'on considère une structure complexe des pannes de serveurs. Pour pallier à ces difficultés, plusieurs auteurs ont tenté de développer des méthodes approximatives d'analyse de ce type de systèmes. Nous avons constaté aussi que peu de travaux traitent l'analyse de fiabilité des systèmes avec rappels non fiables.

CHAPITRE 2

NOTIONS PRELIMINAIRES DE FIABILITE

2.1 Introduction

La fiabilité est l'une des composantes essentielles de la qualité d'un produit et elle est retenue en tant que critère fondamental pour leur élaboration. Elle est prise en considération dès le stade de la conception. La fiabilité est la caractéristique d'un dispositif exprimée par la probabilité que ce dispositif accomplisse une fonction requise dans des conditions d'utilisation et pour une période de temps déterminée. Les calculs de fiabilité font désormais partie intégrante de la démarche de qualité dans la production industrielle.

Dans ce chapitre, nous rappelons quelques notions que nous utiliserons dans ce travail. Le lecteur peut trouver plus de détails dans les ouvrages suivants : [73–76]

2.2 Vocabulaire de la fiabilité

On définit les notions de base :

- **Fiabilité** (Reliability) : Comme nous l'avons vu précédemment, la fiabilité " R " est la probabilité qu'un dispositif prélevé au hasard dans la population considérée accomplisse de manière satisfaisante, une fonction requise, sous des conditions données et pendant une période de temps donnée.

- **Disponibilité** (Availability) La disponibilité est l'aptitude d'un équipement ou d'un système à être en état d'accomplir les fonctions requises dans les conditions données et à un instant donné. Elle est caractérisée par la probabilité $A(t)$ que l'équipement E soit en état, à

l'instant t , d'accomplir les fonctions requises dans des conditions données.

$$A(t) = P[E \text{ non défaillant à l'instant } t]$$

- **Maintenabilité** (Maintainability) : La maintenabilité est l'aptitude d'une entité à être maintenue ou rétablie dans un état dans lequel elle peut accomplir une fonction requise, lorsque la maintenance est réalisée dans des conditions données avec des procédures et des moyens prescrits. Elle est caractérisée par la probabilité $M(t)$ que l'entité E soit en état, à l'instant t , d'accomplir ses fonctions, sachant que l'entité était en panne à l'instant 0.

$$M(t) = P[E \text{ est réparée sur } [0, t]]$$

- **Sécurité** (Safety) : C'est l'aptitude d'une entité à éviter de faire apparaître, dans des conditions données, des événements critiques ou catastrophiques. Elle est caractérisée par la probabilité $S(t)$ que l'entité E ne laisse pas apparaître dans des conditions données, des événements critiques ou catastrophiques.

$$S(t) = P[E \text{ évite des événements critiques ou catastrophiques sur } [0, t]]$$

- **Défaillance** (Failure) : Une défaillance est la cessation du système à accomplir la fonction pour laquelle il a été conçu. Le système passe alors à l'état de panne. Pour un système donné, on peut alors observer le schéma suivant :

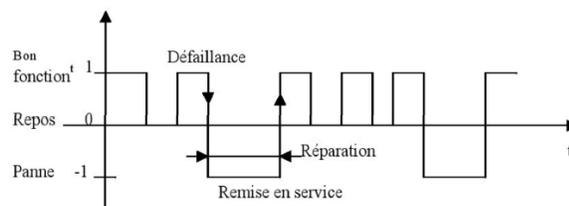


FIGURE 2.1 – Evolution temporelle de l'état d'un système.

- **Durée de réparation** : L'intervalle de temps qui s'écoule entre la manifestation d'une panne et la remise en service du système.

- **Durée de vie** : La durée de vie X d'un système est une variable aléatoire non négative qui représente l'intervalle de temps entre la mise en exploitation jusqu'à la première défaillance.

- **Durée de bon fonctionnement** : c'est la période pendant laquelle le dispositif, en activité ou en service, est exposé à des défaillances.

2.3 Indices de fiabilité

On considère un équipement qui commence à fonctionner à l'instant t_0 . Soit X une variable aléatoire positive et possédant une densité de probabilité f , représentant la durée de vie de cet équipement.

La fiabilité peut être évaluée à l'aide de plusieurs indices, parmi lesquels on peut citer :

2.3.1 Fonction de défaillance. Fonction de Fiabilité

Définition 2.3.1. *La fonction de défaillance de l'équipement est définie par :*

$$F(t) = P(X \leq t) = \int_0^t f(x) dx, \text{ pour } x \geq 0.$$

$F(t)$ représente la probabilité que l'équipement tombe en panne avant l'instant t .

Définition 2.3.2. *La fonction de fiabilité est définie par :*

$$R(t) = P(X > t) = 1 - F(t), \text{ pour } t \geq 0.$$

$R(t)$ est la probabilité que l'équipement fonctionne au-delà de l'instant t . On l'appelle aussi fonction de survie. C'est une fonction continue, strictement positive et décroissante, qui vaut 1 en $t = 0$ et qui tend vers 0 quand t tend vers l'infini.

La Figure 2.2 présente une allure de la fonction $R(t)$ en fonction du temps.

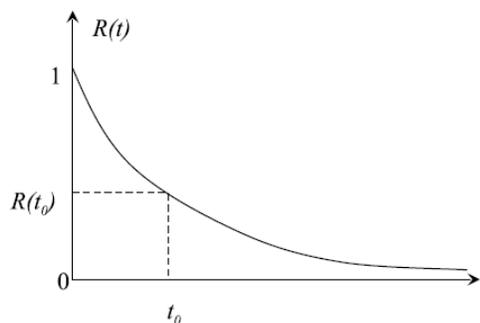


FIGURE 2.2 – Courbe de survie ou de fiabilité.

- F est la fonction de distribution (répartition) de la variable aléatoire X .

Les fonctions R et F ont des variations complémentaires (figure 2.3).

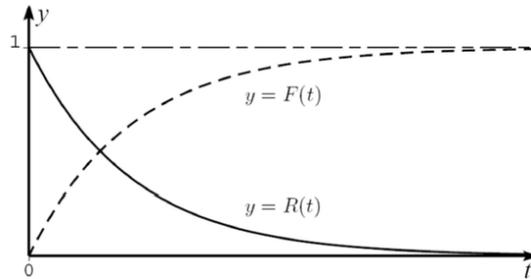


FIGURE 2.3 – Fonctions de distribution $F(t)$ et de fiabilité $R(t)$.

2.3.2 Taux de défaillance

Le taux de défaillance $\lambda(t)$ est l'un des principaux indicateurs de la fiabilité utilisés industriellement. Il caractérise la vitesse de variation de la fiabilité au cours du temps.

$$\lambda(t) = \frac{f(t)}{1 - F(t)},$$

Il est facile d'établir le lien entre le taux de défaillance et la fonction de fiabilité :

$$\forall t \geq 0, \lambda(t) = \lim_{x \rightarrow 0} \frac{P(X \leq t + x / X > t)}{x}$$

$\lambda(t)$ représente la probabilité que le système tombe en panne entre t et $t + x$ sachant qu'il a bien fonctionné entre 0 et t .

La fiabilité s'exprime à l'aide du taux de défaillance comme suit :

$$R(t) = \exp \left\{ - \int_0^t \lambda(y) dy \right\}.$$

Il est fréquent de représenter l'évolution du taux de défaillance $\lambda(t)$ au cours du temps t selon une courbe caractéristique dite en "baignoire" (voir figure 2.4).

Cette courbe se décompose en trois parties :

- La première partie est la période de défaillance précoce (ou période de jeunesse) : c'est le début de la vie du produit et les défaillances sont dites " de jeunesse " (composant neufs présentant des défauts de fabrication ...). Le taux de défaillance λ décroît rapidement au cours du temps.

- la deuxième partie est caractérisée par un taux de défaillance approximativement constant. c'est la zone de maturité ou de pleine activité du produit. C'est également le domaine des défaillances imprévisibles se produisant de façon aléatoire.

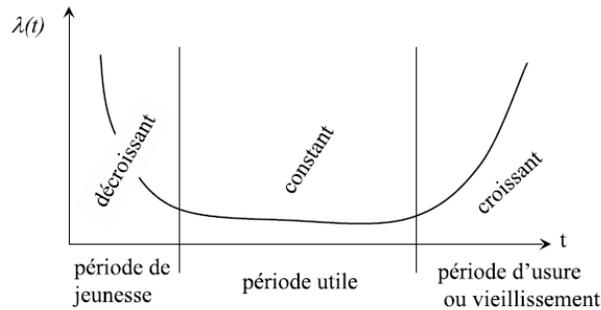


FIGURE 2.4 – Courbe en baignoire.

– la troisième partie correspond aux défaillances dues à des phénomènes d'usure, de vieillissement, ... C'est la période de fin de vie du produit. Le taux de défaillance λ croît rapidement avec le temps, du fait de la dégradation du matériel (usures mécaniques, phénomènes de fatigue, dérive des composants électroniques, ...)

2.3.3 Taux de défaillance cumulé

La fonction Λ définie par :

$$\Lambda(t) = \int_0^t \lambda(x) dx$$

s'appelle fonction de hasard et représente le taux de défaillance cumulé jusqu'à l'instant t .

2.3.4 MTTF (mean time to failure)

Le temps moyen de fonctionnement jusqu'à la première défaillance, noté M.T.T.F (" Mean Time To Failure ") est une autre grandeur associée à la fiabilité, elle permet notamment de visualiser plus concrètement la fiabilité d'un matériel. Le MTTF correspond à l'espérance de la durée de vie X

$$\text{MTTF} = E[X] = \int_0^{\infty} f(x) dx = \int_0^{\infty} R(t) dt.$$

2.3.5 Durée de vie résiduelle d'un élément d'âge t

Lorsqu'un élément a fonctionné jusqu'à la date t , le temps d'attente de la panne est appelé la durée de survie de l'élément au temps t ou la durée de vie résiduelle, elle représente la variable aléatoire $X_t = X - t$ conditionnée par l'événement $X > t$.

La fonction de répartition de X_t notée F_t est définie par :

$$\begin{aligned} F_t(x) &= P(X_t \leq x / X > t) = P(X - t \leq x / X > t) \\ &= \frac{P(t < X \leq t + x)}{P(X > t)} \\ &= \frac{F(t + x) - F(t)}{R(t)}. \end{aligned}$$

Sa fonction de fiabilité est alors R_t donnée par :

$$R_t(x) = \frac{R(t + x)}{R(t)}.$$

2.4 Principales lois de probabilité utilisées en fiabilité

Nous présentons dans cette section quelques distributions de vie qui interviennent le plus fréquemment dans l'analyse de la fiabilité. Nous parlerons en particuliers des lois continues. Nous énoncerons les principales propriétés de ces lois, les fonctions de survie (Fiabilité) associées ainsi que les taux de défaillance.

2.4.1 La loi Exponentielle

Cette loi a de nombreuses applications dans plusieurs domaines. C'est une loi simple, très utilisée en fiabilité dont le taux de défaillance est constant. Elle décrit la vie des matériels qui subissent des défaillances brutales.

La fonction de survie d'une loi exponentielle de paramètre θ est :

$$R(t) = \exp\{-\theta t\} \text{ pour } t > 0$$

Le temps moyen de bon fonctionnement est :

$$\text{MTTF} = \frac{1}{\theta}.$$

Le taux de défaillance est :

$$\lambda(t) = \theta.$$

2.4.2 Loi de Weibull

La plus populaire des lois, souvent utilisée aussi bien en électronique qu'en mécanique ; elle caractérise mieux le comportement du produit dans les trois phases de vie : période de jeunesse, période de vie utile et période d'usure ou vieillissement.

- La fonction de survie d'une loi de Weibull de paramètres η et β est :

$$R(t) = \exp \left\{ - \left(\frac{t}{\eta} \right)^\beta \right\} \text{ pour } t > 0.$$

- Le taux de défaillance :

$$\lambda(t) = \frac{\beta}{\eta} \left(\frac{t}{\eta} \right)^{\beta-1}.$$

2.4.3 La loi normale

La loi normale est très répandue parmi les lois de probabilité car elle s'applique à de nombreux phénomènes. En fiabilité, la distribution normale est utilisée pour représenter la distribution des durées de vie de dispositifs en fin de vie (usure) car le taux de défaillance est toujours croissant. On ne l'utilisera que si la moyenne des durées de vie est supérieur à 3 fois l'écart type.

La fonction de densité est définie par la moyenne μ et l'écart type σ :

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} \text{ pour } t \in \mathfrak{R}$$

La fonction de répartition s'écrit :

$$F(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx.$$

La fiabilité est donnée par :

$$R(t) = 1 - \Phi \left(\frac{t - \mu}{\sigma} \right).$$

Où Φ est la fonction de répartition de la loi normale centrée ($\mu = 0$) réduite ($\sigma = 1$) :

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t \exp \left\{ -\frac{x^2}{2} \right\} dx.$$

2.4.4 La loi lognormale

Une variable aléatoire continue et positive t est distribuée selon une loi lognormale si son logarithme est distribué suivant une loi normale. Cette distribution est utilisée en fiabilité pour modéliser les défaillances par fatigue.

- La fonction de survie d'une loi lognormale de paramètres μ et σ est :

$$R(t) = 1 - \Phi \left(\frac{\log t - \mu}{\sigma} \right) \text{ pour } t > 0.$$

- Le taux de défaillance est donné par :

$$\lambda(t) = \frac{\exp\left\{-\frac{1}{2}\left(\frac{\log t - \mu}{\sigma}\right)^2\right\}}{t \int_0^\infty \sigma \sqrt{2\pi} f(t) dt}.$$

2.4.5 La loi Gamma

Elle représente la loi de probabilité d'occurrence de α événements dans un processus poissonien. Par exemple si t_i est le temps entre les défaillances successives d'un système, et que t_i suit une distribution exponentielle, le temps cumulé d'apparition de α défaillances suit une loi Gamma de densité de probabilité :

$$f(t) = \frac{\beta^\alpha t^{\alpha-1} e^{-\beta t}}{\Gamma(\alpha)}, t > 0.$$

- Le taux de défaillance est donné par :

$$\lambda(t) = \frac{\beta^\alpha t^{\alpha-1} e^{-\beta t}}{\int_t^\infty \Gamma(\alpha) f(\mu) d\mu}.$$

2.5 Fiabilité d'un système

Les systèmes sont généralement constitués de plusieurs composants élémentaires, et leurs fiabilité dépend à la fois de la fiabilité de ses composants et de la façon dont le bon fonctionnement ou la panne de chaque composant influe sur le bon fonctionnement ou la panne du système tout entier. Dans la suite, nous allons calculer la fiabilité de quelques systèmes simples fréquemment rencontrés dans la pratique. Soit alors X la durée de fonctionnement du système et x_i la durée de fonctionnement du composant i .

2.5.1 Système en série

Un système en série est un système qui ne fonctionne que si tous ses composants fonctionnent. Le diagramme de fiabilité est représenté dans la figure 2.5.

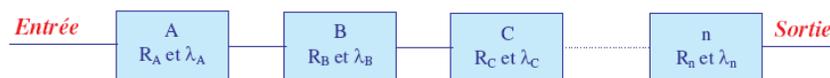


FIGURE 2.5 – Diagramme de fiabilité pour un système en série.

Dans ce cas, la durée de vie du système est définie par :

$$X = \min(X_1, X_2, \dots, X_n),$$

La fiabilité du système est alors :

$$R(t) = P(X > t) = P(\min(X_1, X_2, \dots, X_n) > t) = P\left(\bigcap_{i=1}^n [X_i > t]\right),$$

comme les X_i sont indépendantes alors :

$$R(t) = \prod_{i=1}^n R_i(t),$$

on a donc :

$$R(t) = \prod_{i=1}^n \exp\left\{-\int_0^t \lambda_i(x) dx\right\} = \exp\left\{-\sum_{i=1}^n \int_0^t \lambda_i(x) dx\right\} = \exp\left\{-\int_0^t \sum_{i=1}^n \lambda_i(x) dx\right\}.$$

Le taux de défaillance du système est défini par :

$$\lambda(t) = \sum_{i=1}^n \lambda_i(t).$$

2.5.2 Système en parallèle

Un système en parallèle est un système tel qu'il suffit qu'un seul de ses composants fonctionne pour qu'il fonctionne. Autrement dit, la défaillance du système survient quand tous ses composants sont en panne. Dans les systèmes parallèles, on distingue deux cas :

- La redondance passive ou stand-by : un seul composant fonctionne à la fois. Quand le composant qui fonctionne tombe en panne, il est instantanément remplacé par un des composants en attente. Dans ce cas, $X = \sum_{i=1}^n nX_i$.

- La redondance active : les n composants fonctionnent en même temps.

On se place dans la suite de cette section dans le cas de la redondance active. Le diagramme de fiabilité est donné dans la figure 2.6.

On a évidemment :

$$X = \max(X_1, X_2, \dots, X_n).$$

La fiabilité de ce système est :

$$R(t) = P(\max(X_1, X_2, \dots, X_n) > t) = 1 - P(\max(X_1, X_2, \dots, X_n) \leq t) = 1 - \left(\bigcap_{i=1}^n [X_i \leq t]\right),$$

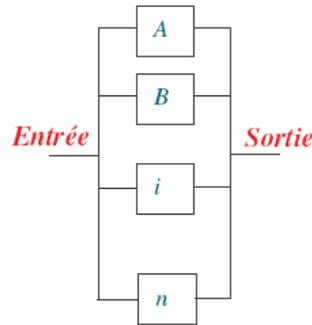


FIGURE 2.6 – Diagramme de fiabilité pour un système en parallèle.

comme les X_i sont indépendantes, alors

$$R(t) = 1 - \prod_{i=1}^n (1 - R_i(t)),$$

en écrivant :

$$R(t) = 1 - \prod_{i=1}^n \left(1 - \exp \left\{ - \int_0^t \lambda_i(x) dx \right\} \right).$$

Puis :

$$\lambda(t) = - \frac{R'(t)}{R(t)},$$

on obtient le taux de défaillance :

$$\lambda(t) = \frac{\sum_{i=1}^n \lambda_i(t) \exp \left\{ - \int_0^t \lambda_i(x) dx \right\} \prod_j (1 - \exp \left\{ - \int_0^t \lambda_j(x) dx \right\})}{1 - \prod_{i=1}^n (1 - \exp \left\{ - \int_0^t \lambda_i(x) dx \right\})}.$$

2.6 Méthodologies dynamiques pour l'étude de la fiabilité d'un système

2.6.1 Chaînes de Markov

Les chaînes de Markov -ou Méthode de l'Espace des Etats (MEE)- ont été développées dans les années 1950 pour l'analyse de la fiabilité des systèmes réparables [77; 78].

Cette méthode consiste à représenter le fonctionnement d'un système par un ensemble de composants pouvant se trouver dans un nombre fini d'états de fonctionnement et de panne. Un support graphique (le graphe des états) permet de visualiser les différents états d'un système qui sont représentés par des cercles et relier entre eux par des arcs orientés qui correspondent aux transitions (pannes et réparations) entre états. Pour un système à n composants, si chaque

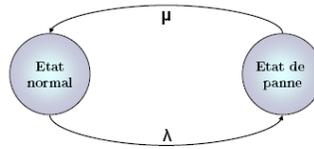


FIGURE 2.7 – Un modèle Markovien.

composant a deux états (fonctionnement et panne), le nombre maximum d'états est 2^n . Un modèle Markovien est présenté sur la Figure 2.7.

Pour réaliser cette analyse, il faut tout d'abord recenser tous les états du système, les classer en états de fonctionnement ou en états de panne. Ensuite, il est nécessaire de chercher comment passer d'un état à un autre lors d'un dysfonctionnement ou d'une réparation. A chaque transition de l'état E_i vers l'état E_j , un taux de transition L_{ij} est associé qui est défini de telle façon que $L_{ij}.dt$ est égal à la probabilité de passer de E_i vers E_j entre deux instants très proches t et $t + dt$ sachant que l'on est à l'instant t en E_i . Enfin, la dernière étape consiste à calculer les probabilités de se trouver dans les différents états au cours d'une période de vie du système ainsi que de calculer les caractéristiques de sûreté de fonctionnement (MTTE, MTBE, MTTR, etc.) [78].

2.6.2 Réseaux Bayésiens

Le problème de la modélisation et de l'analyse de la fiabilité dynamique se pose dès lors que l'état de fonctionnement du système et l'état des variables fonctionnelles du système s'influencent mutuellement [79]. Les Réseaux Bayésiens Dynamiques (DBN) constituent un outil mathématique intéressant pour modéliser ce problème en permettant une représentation graphique des processus stochastiques. Ces DBN sont utilisés pour représenter l'interaction complexe entre l'état du système et l'état des variables processus, d'une part, et le processus et la perturbation externe, d'autre part.

2.6.3 Méthode des réseaux de Petri

Les réseaux de Petri ont été inventés en 1962 par CARL ADAM PETRI [80]. Ils sont basés sur la théorie des automates. Ces réseaux permettent de représenter le comportement des systèmes dans les conditions de fonctionnement normal ainsi que leur comportement en cas de défaillance de leurs composants [81–84].

Le principal avantage des réseaux de Petri est la possibilité d'analyser le comportement dynamique d'un système en présence de défaillances. Cette modélisation dynamique permet d'obtenir des mesures en termes de fiabilité. Un réseau de Petri permet de modéliser d'une part le fonctionnement d'un système et d'autre part les occurrences de défaillances.

Les réseaux de Petri sont décrits par un 7-tuplé, $(P, T, A, W, M_0; Pre; Post)$, défini par :

- l'ensemble des places p_i de P , tel que $p_i \in P$, est fini et non vide ;
- l'ensemble des transitions t_i de T , tel que $t_i \in T$, est fini et non vide ;
- l'ensemble des arcs a_i de A , tel que $a_i \in A$;
- L'ensemble des poids w_i affectés aux arcs a_i , tel que $w_i \in W$, souvent égale à 1 pour les réseaux déterministes, évalué à partir de probabilité pour les réseaux stochastiques ;
- Le marquage initial M_0 avec ses jetons J . C'est le caractère dynamique du réseau de Petri et sa capacité à supplanter les chaînes de Markov, dans son utilisation en fiabilité des systèmes ;
- $Pre(p_i; t_i)$ est l'application d'incidence avant, de type $P \times T \rightarrow N$ correspond aux arcs allant d'une place vers une transition ;
- $Post(p_i; t_i)$ est l'application d'incidence arrière, de type $T \times P \rightarrow N$ correspond aux arcs allant d'une transition vers une place.

La figure 2.8 illustre un exemple d'un réseau de Petri.

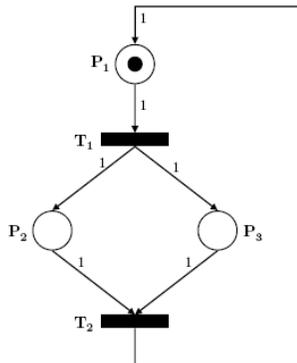


FIGURE 2.8 – Réseau de Petri.

Les réseaux de Petri sont très utilisés dans la modélisation des systèmes à événements discrets et dans les études de sûreté de fonctionnement des systèmes dynamiques. Ils sont caractérisés par une évolution asynchrone dans laquelle les transitions des composantes parallèles sont franchies les unes après les autres, et par une représentation explicite des synchronisations et des mécanismes d'allocation. Plusieurs extensions des réseaux de Petri ont

été élaborées pour répondre à la modélisation des problèmes spécifiques et pour maîtriser la taille et la lisibilité des modèles.

2.7 Les applications de la fiabilité

La fiabilité est un concept qui intéresse de nombreux domaines de l'activité humaine : économique, scientifique, technique, industriel et technologique. Elle est liée à des notions de sécurité de fonctionnement, de qualité, d'efficacité et de performance. Son objectif principal est l'analyse de la probabilité de défaillance d'un système et concerne plus particulièrement le domaine mathématique. Dans cette section, nous allons présenter quelques travaux concernant l'étude de fiabilité de quelques systèmes dans des différents domaines : mécanique, file d'attente...

2.7.1 Modélisation de la fiabilité des équipements mécaniques

Procédure générale

La détermination de la fiabilité d'un système électronique, mécanique ou autre nécessite tout d'abord de connaître la loi de la fiabilité (ou la loi de défaillance) de chacun des composants intervenant dans le système [77; 85; 86]. Ceci est simple pour certains types de systèmes tels que les systèmes électroniques, or ce n'est pas le cas pour des systèmes mécaniques à cause de la complexité de la structure du système étudié. Les systèmes mécaniques sont des ensembles d'éléments technologiques liés par des relations statiques et dynamiques assez complexes.

Pour un système électronique chaque composant a un poids important dans la fiabilité du système, la fiabilité du système est donc calculée en fonction de la fiabilité de tous ses composants. Les calculs sont effectués sous l'hypothèse que les taux de défaillance sont constants dans le temps, une hypothèse acceptable pour la plupart des composants, ce qui rend les calculs beaucoup plus simple. La détermination des taux de défaillance des composants est effectuée soit à partir des modèles développés dans des bases de données disponibles, soit à partir d'essais effectués sur les composants ou bien à partir des résultats d'exploitation des produits.

La fiabilité d'un système mécanique, contrairement à l'électronique, repose sur la fiabilité de quelques composants élémentaires responsables de son dysfonctionnement, dits composants " responsables " ou " critiques " (parfois un seul), contribuant presque totalement à la probabilité

de défaillance de l'ensemble. Voir [87; 88]. Les autres composants pouvant être considérés de probabilité de défaillance pratiquement nulle. L'identification de ces composants se fait en effectuant des analyses qualitatives, telles que l'APR (Analyse Préliminaire des Risques) et l'AMDEC (Analyse de Modes de défaillance, de leurs effets et de leurs criticité), et des analyses quantitatives telles que l'analyse par arbres de défaillance. Ensuite, nous en créerons un modèle à partir d'un diagramme de fiabilité, par exemple sous forme d'un schéma bloc. Le taux de défaillance du système sera donc calculé en fonction de l'architecture du système donnée par ce schéma.

Illustration

À l'aide d'un exemple tiré de [76], nous éclaircissons la procédure. Soit le réducteur de vitesse, représenté par la figure 2.9.

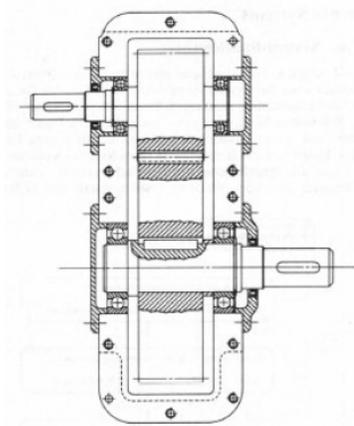


FIGURE 2.9 – Réducteur de vitesse.

1) Données

Les différents composants critiques vis-à-vis de la fiabilité sont les suivants : deux roues dentées, un engrenage, deux roulements et deux joints à lèvres. Bertsche et Lechner proposent de leurs associer comme loi de défaillance la distribution de Weibull dont les paramètres figurent dans le tableau 2.1 (η s'exprime en tours d'arbre d'entrée), en supposant les paramètres de décalage nuls ($\gamma = 0$).

2) Modélisation du système

En général, un système se représente sous forme d'un schéma bloc. Ceci sous-entend qu'une analyse topologique préalable du système doit être réalisée. La défaillance de tel compo-

Composant		Mode de défaillance	η	β
Composant 1	Roue d'entrée 1	Rupture de la roue dentée 1	38.000	1,4
Composant 2	Roue d'entrée 2	Rupture de la roue dentée 2	70.500	1,8
Composant3	Engrounage	Pitting de l'engrenage	1.966.600	1,3
composant 4	Roulement 1	Défaillance du roulement 1	9.100.000	1,11
Composant 5	Roulement 2	Défaillance du roulement 2	15.200.000	1,11
Composant 6	Joint à lèvres radial 1	Défaillance du joint à lèvres radial 1	66.000.000	1,0
Composant 7	Joint à lèvres radial 2	Défaillance du joint à lèvres radial 2	6.000.000	1,0

TABLEAU 2.1 – Paramètres de la loi de défaillance des éléments du réducteur (distribution de Weibull).

sant entraîne-t-elle la défaillance du système ? Si la réponse est affirmative alors ce composant doit être associé en série.

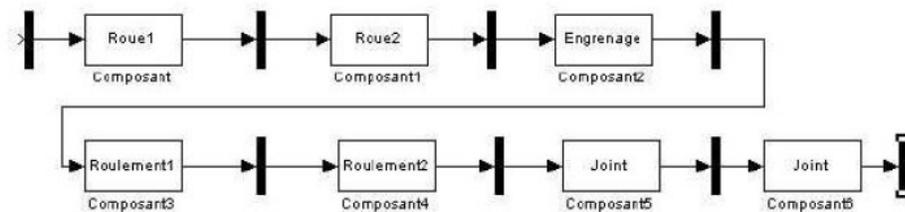


FIGURE 2.10 – Exemple d'une modélisation du réducteur en schéma bloc.

3) Calcul et résultats

La fiabilité du système est calculée en fonction des fiabilités des sept composants qui le constituent : $R(t) = \prod_{i=1}^7 R_i(t)$ où $R_i(t) = \exp\left\{-\left(\frac{t}{\eta_i}\right)^\beta\right\}$ $t \geq 0$ $i = 1, \dots, 7$. Les résultats sont représentés par un seul graphique (voir figure 2.11).

On remarque que la courbe des deux premiers composants (les deux roues dentées) sont prépondérantes. Il en résulte que la fiabilité du réducteur de vitesse dépend principalement de la fiabilité des deux roues dentées.

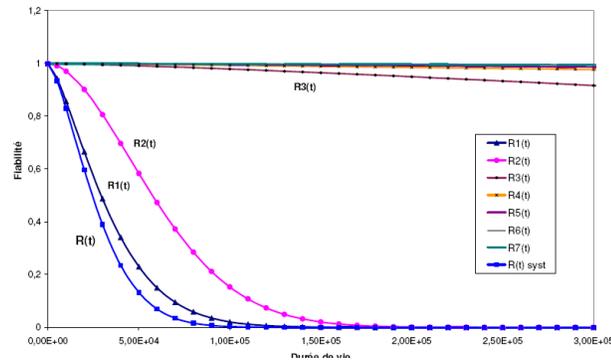


FIGURE 2.11 – Présentation graphique de la fiabilité du système et de ses composants.

2.7.2 File d'attente et fiabilité

l'analyse de fiabilité d'un système avec rappels et un serveur sujet à des pannes et des réparations

[41] Ce travail propose une analyse détaillée de la notion de fiabilité de système M/G/1 avec rappels non fiable. Danc ce qui suit, nous allons présenter les principaux résultats obtenus par les auteurs.

Description du modèle

Un système d'attente à un seul serveur est considéré. Les nouveaux clients (clients primaires) arrivent selon un processus de Poisson de taux $\lambda > 0$. Si un client primaire trouve le serveur libre, il obtient son service immédiatement et quitte le système après la fin du service. Sinon, si le serveur est trouvé occupé ou en panne, le client rejoint l'orbite et il devient une source d'appels secondaires. L'orbite de sources d'appels secondaires peut être considéré comme une sorte de file d'attente avec une capacité infinie. Les clients qui retournent se comportent indépendamment les uns des autres et sont persistants en ce sens qu'ils continuent de faire des rappels jusqu'à ce qu'ils reçoivent leur service demandé. La durée entre deux rappels successifs d'une même source secondaire est exponentiellement distribuée de paramètre $\theta > 0$.

Les temps de service des clients suivent une distribution générale de fonction de répartition $B(x)$, de transformée de Laplace-Stieltjes $b^*(s)$, $Re(s) > 0$, de fonction de densité $b(x)$ définie par

$$b(x) = \mu(x) \exp \left\{ - \int_0^x \mu(t) dt \right\},$$

avec $\mu(x)$ est le taux d'achèvement du service tel que

$$\mu(x) = \frac{b(x)}{1 - B(x)}$$

Soient les moments d'ordre k , $\beta_k = (-1)^k b^{*(k)}(0)$.

le système est soumis à des pannes aléatoires et à des réparations. Les pannes se produisent indépendamment via un processus de Poisson de taux α . Lorsque le serveur tombe en panne, les réparations commencent immédiatement et le temps de réparation est une variables aléatoires Y et suit une distribution générale de fonction de répartition $G(y)$, de transformée de Laplace-Stieltjes $g^*(s)$, $Re(s) > 0$, de fonction de densité $g(y)$ définie par

$$g(y) = \beta(y) \exp \left\{ - \int_0^y \beta(t) dt \right\},$$

avec $\beta(y)$ est le taux d'achèvement de réparation tel que

$$\beta(y) = \frac{g(y)}{1 - G(y)}$$

Soient les moments d'ordre k , $\gamma_k = (-1)^k g^{*(k)}(0)$ et l'intensité du trafic $\rho = \lambda \beta_1$ due aux appels primaires.

Le flux d'entrée des arrivées primaires, les temps de service, les intervalles entre les essais répétés, les temps de panne et les temps de réparation sont supposés mutuellement indépendants.

La condition de stabilité et la distribution stationnaire de l'état du système

L'état du système à la date t peut être décrit par le processus $\{N(t), X(t), Y(t), t \geq 0\}$, où $N(t)$ est le nombre de clients dans le système (en orbite et en service) à l'instant t , $X(t)$ est une variable supplémentaire à valeurs dans \mathfrak{R}^+ désignant la durée de service écoulée à la date t et $Y(t)$ est une variable supplémentaire à valeurs dans \mathfrak{R}^+ désignant la durée de réparation écoulée à la date t . Les probabilités d'états sont définies par :

- $P_{Wi1}(t, x)dx$ est la probabilité conjointe qu'à l'instant t il y a i clients en orbite, le serveur est occupé et le temps de service écoulé entre x et $x + dx$ ($i \geq 0$).

- $P_{Ri1}(t, x, y)dy$ est la probabilité conjointe qu'au moment t il y a i clients en orbite, le temps de service écoulé par le client sous service est égal à x et le serveur est en réparation avec le temps de réparation écoulé entre y et $y + dy$ ($i \geq 0$).

- $P_{Ii0}(t)$ est la probabilité que le serveur est inactif et il y a i clients en orbite.

Ainsi le processus stochastique $\{N(t), X(t), Y(t), t \geq 0\}$ d'espace d'états $\{(I, i, 0), (W, i, 1), (R, i, 1, x, y) | 0 \leq i \leq +\infty, 0, y < +\infty\}$.

La condition nécessaire et suffisante pour que ce système soit stable est :

$$\rho(1 + \alpha\gamma_1) < 1.$$

- à l'état stationnaire, la fonction génératrice de la distribution conjointe de l'état de serveur et le nombre de clients dans le système est :

$$Q_{Ii0}(z) = [1 - \rho(1 + \alpha\gamma_1)] \times \exp \left\{ -\frac{\lambda}{\theta} \int_z^1 \frac{b[\lambda(1 - \mu) + \alpha\bar{g}^*(\lambda(1 - \mu))] - 1}{\mu - b[\lambda(1 - \mu) + \alpha\bar{g}^*(\lambda(1 - \mu))]} d\mu \right\},$$

$$Q_{W1}(z, x) = \frac{\lambda(z - 1)}{z - b^*(\lambda(1 - z) + \alpha\bar{g}^*(\lambda - \lambda z))} \times \exp \left\{ -[\lambda(1 - z) + \alpha\bar{g}^*(\lambda(1 - z))] x \right\} \bar{B}(x) Q_{I0}(z),$$

$$Q_{R1}(z, x, y) = \frac{\alpha\lambda(z - 1)}{z - b^*(\lambda(1 - z) + \alpha\bar{g}^*(\lambda - \lambda z))} \times \exp \left\{ -[\lambda(1 - z) + \alpha\bar{g}^*(\lambda(1 - z))] x \right\} \bar{B}(x) \\ \times \exp \left\{ -\lambda(1 - z) y \right\} \bar{G}(y) Q_{I0}(z).$$

- La probabilité que le serveur est inactif est $I = 1 - \rho(1 + \alpha\gamma_1)$.

- La probabilité que le serveur est occupé est $B = \rho$.

- La probabilité que le serveur est en réparation est $R = \rho\alpha\gamma_1$.

- La fonction génératrice de la distribution du nombre de clients en orbite est :

$$p(z) = Q_{I0}(z) + Q_{W1}(z) + Q_{R1}(z) \\ = [1 - \rho(1 + \alpha\gamma_1)] \frac{1 - z}{b^*[\lambda(1 - z) + \alpha\bar{g}^*(\lambda(1 - z))] - z} \\ \times \exp \left\{ -\frac{\lambda}{\theta} \int_z^1 \frac{b^*[\lambda(1 - u) + \alpha\bar{g}^*(\lambda(1 - u))] - 1}{u - b^*[\lambda(1 - u) + \alpha\bar{g}^*(\lambda(1 - u))]} du \right\}.$$

- Le nombre moyen de clients en orbite $EN(t)$ est donné par

$$EN(t) = \frac{\lambda^2}{1 - \rho(1 + \alpha\gamma_1)} \left[\frac{\alpha\beta_1\gamma_2 + (1 + \alpha\gamma_1)^2\beta_2}{2} + \frac{(1 + \alpha\gamma_1)\beta_1}{\theta} \right].$$

- La fonction génératrice de la distribution du nombre de clients dans le système est :

$$Q(z) = [1 - \rho(1 + \alpha\gamma_1)] \frac{(1 - z)b^*[\lambda(1 - z) + \alpha\bar{g}^*(\lambda(1 - z))]}{b^*[\lambda(1 - z) + \alpha\bar{g}^*(\lambda(1 - z))] - z} \\ \times \exp \left\{ -\frac{\lambda}{\theta} \int_z^1 \frac{b^*[\lambda(1 - u) + \alpha\bar{g}^*(\lambda(1 - u))] - 1}{u - b^*[\lambda(1 - u) + \alpha\bar{g}^*(\lambda(1 - u))]} du \right\}.$$

- Le nombre moyen de clients dans le système $EK(t)$ est donné par

$$EK(t) = \rho(1 + \alpha\gamma_1) + \frac{\lambda^2}{1 - \rho(1 + \alpha\gamma_1)} \left[\frac{\alpha\beta_1\gamma_2 + (1 + \alpha\gamma_1)^2\beta_2}{2} + \frac{(1 + \alpha\gamma_1)\beta_1}{\theta} \right].$$

Les indices de fiabilité du serveur

Les indices de fiabilité obtenus par [41] sont :

- La disponibilité du serveur à l'état stationnaire est

$$A = \sum_{i=0}^{\infty} P_{Ii0} + \sum_{i=0}^{\infty} \int_{i=0}^{\infty} P_{Wi1}(x) dx = \lim_{z \rightarrow 1} \left[Q_{I0}(z) + \int_0^{\infty} Q_{W1}(z, x) dx \right] = 1 - \rho\alpha\gamma_1.$$

- La fréquence de panne du serveur à l'état stationnaire est

$$W_f = \sum_{i=0}^{\infty} \int_{i=0}^{\infty} W_{i1}(x) dx = \lim_{z \rightarrow 1} \int_0^{\infty} W_1(z, x) dx = \alpha\rho.$$

- La transformée de laplace de la fonction de fiabilité $R(t)$ est donnée par

$$R^*(s) = \frac{1}{s + \alpha} + \frac{\alpha}{s + \alpha} \int_1^{\omega} \frac{1}{\theta[b^*(s + \alpha + \lambda(1 - y)) - y](1 - y)} \\ \times \exp \left\{ \frac{1}{\theta} \int_y^1 \frac{s + \lambda - \lambda b^*(s + \alpha + \lambda(1 - x))}{b^*(s + \alpha + \lambda(1 - x)) - x} dx \right\} dy,$$

où ω est la racine de l'équation $z = b^*(s + \alpha + \lambda(1 - z))$, $\text{Re}(s) > 0$.

- Le temps moyen de bon fonctionnement est donné par

$$\text{MTTF} = \frac{1}{\alpha} + \int_1^{\omega} \frac{1}{\theta[b^*(\alpha + \lambda(1 - y)) - y](1 - y)} \\ \times \exp \left\{ \frac{1}{\theta} \int_y^1 \frac{\lambda - \lambda b^*(\alpha + \lambda(1 - x))}{b^*(\alpha + \lambda(1 - x)) - x} dx \right\} dy,$$

où ω est la racine de l'équation $z = b^*(s + \alpha + \lambda(1 - z))$, $|z| = 1$.

2.8 Conclusion

Au cours de ce chapitre, nous avons passé en revue les notions de base de la fiabilité, nous avons présenté aussi les méthodologies dynamiques pour déterminer la fiabilité d'un système complexe et quelques exemples d'applications de la fiabilité dans divers domaines.

Dans les prochains chapitres, nous allons nous intéresser à l'étude de la fiabilité de certains systèmes d'attente avec rappels non fiables en utilisant la méthode des chaînes de Markov.

CHAPITRE 3

ANALYSE APPROXIMATIVE DE LA FIABILITE DU SYSTEME M/M/2 AVEC RAPPELS

3.1 Introduction

Les systèmes d'attente avec rappels à multiserveurs non fiables ont été considérés pour la première fois par ARTALEJO [35] en 1994. L'auteur a examiné un système avec rappels M/M/c/k dans lequel le serveur est soumis uniquement à des pannes actives. En 2005, GHARBI et IOUALALEN [89] ont étudié un système avec rappels à source finie avec plusieurs serveurs soumis à des pannes (passives et actives) et des réparations. Les auteurs incluent un scénario de décomposition dépendant dans lequel la probabilité de défaillance dépend de l'état du serveur. Les clients dont le service est interrompu par une panne rentrent en orbite avec une mémoire de leur temps de service écoulé. Les auteurs ont utilisé un modèle généralisé de réseau de Petri stochastique (GSPN) pour obtenir plusieurs mesures de performance et indices de fiabilité. Ces systèmes sont encore à l'heure actuelle peu étudiés et les modèles proposés sont d'ailleurs étudiés sous des hypothèses très restrictives. La théorie analytique actuelle de ces systèmes a une portée limitée en raison de la complexité des résultats connus. En effet, plusieurs formules analytiques (complexes) sont difficilement exploitables en pratique. C'est pour cela que, lors de l'étude des systèmes concrets on est souvent amené à remplacer le système réel (généralement complexe), par un système plus simple et pour lequel il existe des résultats analytiques

exploitables, ce qui a poussé les auteurs à tenter de développer des méthodes approximatives d'analyse des phénomènes de répétition d'appels. Des méthodes d'approximation pour ce type de modèles ont été développées ces dernières années. Dans [13; 90], il est obtenu des inégalités et bornes pour certaines caractéristiques du système moyennant les techniques de comparaisons stochastiques. De leur côté, CRAWFORD [91] a utilisé l'algorithme de "phase-merging" pour analyser approximativement un modèle généralisé du modèle d'attente M/M/1 avec pannes et rappels introduit initialement par SHERMAN [47].

Ce chapitre traite l'analyse de la fiabilité de système M/M/2 avec rappels pour lequel les deux serveurs sont sujets à des pannes. Nous obtenons des expressions approximatives de certains indices de fiabilité comme la disponibilité, la fréquence de panne des serveurs en utilisant l'algorithme de "phase merging". Nous fournissons également quelques illustrations numériques pour évaluer les résultats théoriques obtenus. Ces résultats font l'objet de l'étude réalisée dans [92].

3.2 description du modèle

Nous considérons le système M/M/2 avec rappels et pannes dans lequel les clients arrivent selon un processus de Poisson de taux λ ($\lambda > 0$). Les temps de service sont distribués suivant une loi exponentielle, avec un taux μ . Nous supposons qu'il n'y a pas d'espace d'attente et donc si un client primaire trouve au moins un serveur libre et en bon état, il est immédiatement pris en charge. Sinon, il peut choisir d'entrer en orbite avec une probabilité P_1 et devenir une source d'appels secondaires ou de quitter le système avec une probabilité $1 - P_1$. Il est également supposé que le client dont le service est interrompu par une panne active, doit décider, soit de rejoindre le groupe de rappels avec une probabilité P_A ou de quitter définitivement le système après interruption avec une probabilité $1 - P_A$. La structure générale du système est illustrée dans la figure 3.1.

La durée entre deux rappels successifs d'une même source secondaire est exponentielle de paramètre ν . Le flux d'entrée des arrivées primaires, les temps de service et les intervalles entre les essais répétés sont supposés mutuellement indépendants. Les pannes pour les deux serveurs se produisent indépendamment via un processus de Poisson de taux α et les temps de réparation pour chaque serveur sont des variables aléatoires distribuées exponentiellement avec un taux β . On suppose que les réparations commencent immédiatement après un échec.

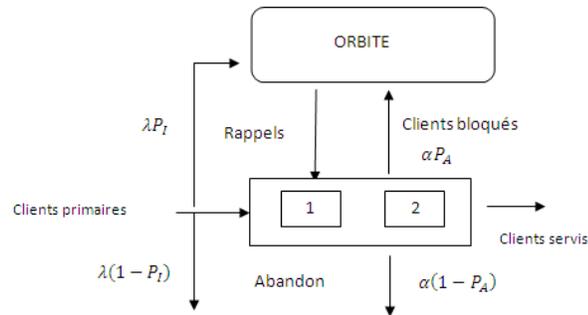


FIGURE 3.1 – La structure du système.

Ces nouvelles variables aléatoires et tous les processus vérifient l'hypothèse d'indépendance mutuelle.

L'état du système à l'instant t peut être décrit par le processus Markovien

$$X(t) = \{(N(t), C(t), R(t))\}, \quad (3.1)$$

d'espace d'états $S = \{(i, j, k) : i \geq 0, j + k \leq 2, j, k \in \{0, 1, 2\}\}$, où $N(t)$ est le nombre de clients en orbite à l'instant t , $C(t)$ est le nombre de serveurs occupés à l'instant t et $R(t)$ est le nombre de serveurs en panne à l'instant t .

Soit :

$$P_{ijk} = \lim_{t \rightarrow \infty} P(N(t) = i, C(t) = j, R(t) = k), \quad (3.2)$$

avec $i \geq 0$, et $(j, k) \in E$ tel que : $E = \{(0, 0), (0, 1), (0, 2), (1, 0), (1, 1), (2, 0)\}$, la distribution stationnaire du processus (3.1). La figure 3.2 représente le diagramme de transition pour la CMTC.

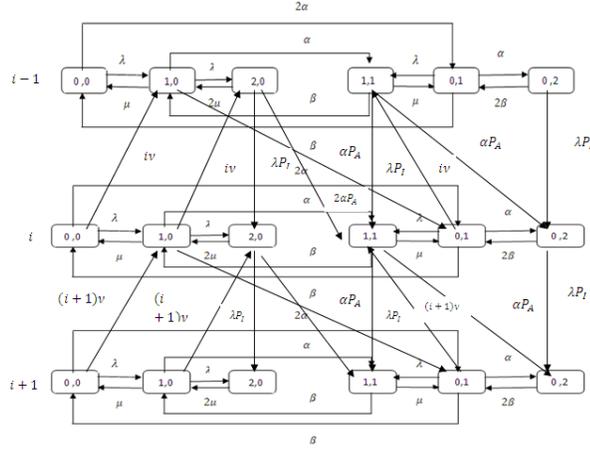


FIGURE 3.2 – Graphe de transitions du système M/M/2 avec rappels et pannes.

3.3 Indices de fiabilité des serveurs

Dans cette section, nous fournissons des approximations pour la disponibilité et la fréquence de pannes de serveurs.

Théorème 3.3.1. Soit $A(t)$ la probabilité que le serveur soit fiable à l'instant t , qui est définie comme la disponibilité des serveurs et définissons la disponibilité des serveurs comme

$$A = \lim_{t \rightarrow \infty} A(t).$$

La disponibilité approximative des serveurs est

$$A \approx \frac{\beta^2}{H} [(\beta + \lambda + \mu)(2\mu^2 + 2\lambda\mu + \lambda^2) + 2\alpha(\lambda + \mu)^2], \quad (3.3)$$

Où la constante H est donnée par,

$$\begin{aligned} H = & \mu^2 \alpha^2 \lambda + 6\mu^2 \alpha^2 \beta + 2\mu^3 \alpha^2 + 2\mu^2 \alpha^3 + 6\beta^2 \mu \alpha \lambda + 2\mu \beta^3 \lambda \\ & + 4\lambda \beta^2 \mu^2 + \lambda^2 \beta^3 + 3\lambda^2 \beta^2 \mu + \lambda^3 \beta^2 + 2\alpha \lambda^2 \beta^2 + 4\mu^3 \alpha \beta \\ & + 6\beta^2 \mu^2 \alpha + 2\mu^2 \beta^3 + 2\mu^3 \beta^2 + 2\mu \alpha \lambda^2 \beta + 6\mu^2 \alpha \lambda \beta + 4\beta \mu \alpha^2 \lambda. \end{aligned} \quad (3.4)$$

Preuve 3.3.1. A est obtenue en considérant l'équation suivante

$$A = \sum_{i=0}^{\infty} P_{i00} + \sum_{i=0}^{\infty} P_{i10} + \sum_{i=0}^{\infty} P_{i20}. \quad (3.5)$$

A partir du graphe de transitions (figure 3.2), les probabilités P_{ijk} vérifient le système d'équations de balance :

Pour $i = 0$

$$(\lambda + 2\alpha)P_{000} = \mu P_{010} + \beta P_{001}$$

Pour $i \geq 1$

$$(\lambda + i\nu + 2\alpha)P_{i00} = \mu p_{i10} + \beta P_{i01}$$

$$(\lambda + i\nu + \alpha + \alpha P_A + \mu)P_{i10} = \lambda P_{i00} + \beta P_{i11} + 2\mu P_{i20} + (i + 1)\nu P_{i+100}$$

$$(\lambda P_I + 2\mu + 2\alpha P_A)P_{i20} = \lambda P_I P_{i-120} + \lambda P_{i10} + (i + 1)\nu P_{i+110}$$

$$(\lambda P_I + \mu + \alpha P_A + \beta)P_{i11} = \lambda P_I P_{i-111} + \lambda P_{i01} + (i + 1)\nu P_{i+101} + \alpha P_{i10} + 2\alpha P_A P_{i-120}$$

$$(\lambda + i\nu + \alpha + \beta)P_{i01} = \mu P_{i11} + \alpha P_A P_{i-110} + 2\alpha P_{i00} + 2\beta P_{i02}$$

$$(\lambda P_I + 2\beta)P_{i02} = \lambda P_I P_{i-102} + \alpha P_{i01} + \alpha P_A P_{i-111}$$

$$\sum_{i=0}^{\infty} [P_{i00} + P_{i10} + P_{i20} + P_{i11} + P_{i01} + P_{i02}] = 1.$$

La résolution de ce système de manière récursive est très compliquée. Une autre façon de calculer les probabilités d'état stationnaire est la méthode des fonctions génératrices. En utilisant les fonctions génératrices définies par : $p_{jk}(z) = \sum_{i=0}^{\infty} z^i P_{ijk}$, $(j, k) \in E$, le système d'équations de balance devient :

$$(\lambda + 2\alpha)p_{00}(z) + \nu z p'_{00}(z) = \mu p_{10}(z) + \beta p_{01}(z)$$

$$(\lambda + \alpha + \alpha P_A + \mu)p_{10}(z) + \nu z p'_{10}(z) = \lambda p_{00}(z) + \beta p_{11}(z) + 2\mu p_{20}(z) + \nu p'_{00}(z)$$

$$(\lambda P_I + 2\mu + 2\alpha P_A)p_{20}(z) = \lambda P_I z p_{20}(z) + \lambda p_{10}(z) + \nu p'_{10}(z)$$

$$(\lambda P_I + \mu + \alpha P_I + \beta)p_{11}(z) = \lambda P_I z p_{11}(z) + \lambda p_{01} + \nu p'_{01}(z) + \alpha p_{10}(z) + 2\alpha P_A z p_{20}(z)$$

$$(\lambda + \alpha + \beta)p_{01}(z) + \nu z p'_{01}(z) = \mu p_{11}(z) + \alpha P_A z p_{10}(z) + 2\alpha p_{00}(z) + 2\beta p_{02}(z)$$

$$(\lambda P_I + 2\beta)p_{02}(z) = \lambda z p_{02}(z) + \alpha p_{01}(z) + \alpha P_A z p_{11}(z)$$

$$\sum_{(j,k) \in E} p_{jk}(1) = 1$$

La résolution de ce système d'équations pour avoir les probabilités d'état est aussi compliquée, par conséquent, nous ferons recours à une analyse approximative du système. Pour déterminer les distributions approximatives de probabilité à l'état stationnaire, nous appliquons l'algorithme de "phase merging" qui a été développé dans [10] et [11] et nous procédons de la manière suivante. la première étape consiste à déterminer la distribution de probabilité conditionnelle de l'état des

serveurs à l'instant t étant donné le nombre de clients en orbite à l'instant t , puis on rapproche la distribution de probabilité marginale du nombre de clients en orbite.

Afin de rapprocher avec précision la distribution de probabilité conjointe de l'état des serveurs et du nombre de clients en orbite, nous supposons que les intensités de transition entre les états du même niveau sont significativement plus grande que les intensités de transition circulant entre les niveaux. Chaque niveau est analysée comme une CTMC à partir de laquelle on obtient les probabilités conditionnelles approximatives.

1) Pour trouver les $P_{l|i}$ nous réduisons la dimension de l'espace d'état en définissant $Y(t)$ comme l'état des serveurs à l'instant t , tel que :

$$Y(t) = \{C(t), R(t) : t \geq 0\}$$

où $C(t)$ et $R(t)$ sont définis dans (3.1). Soit

$$P_{l|i} = \lim_{t \rightarrow \infty} P\{Y(t) = l \mid N(t) = i, l = 1, 2, \dots, 6\}$$

avec l'indice l comme défini :

<i>l'état (j,k)</i>	(0,0)	(1,0)	(2,0)	(1,1)	(0,1)	(0,2)
<i>l'indice l</i>	1	2	3	4	5	6

TABLEAU 3.1 – Les substitutions des états de serveurs.

Cette étape aboutit à un nombre infini de niveaux qui peuvent être analysés individuellement. La figure (3.3) représente le niveau du système.

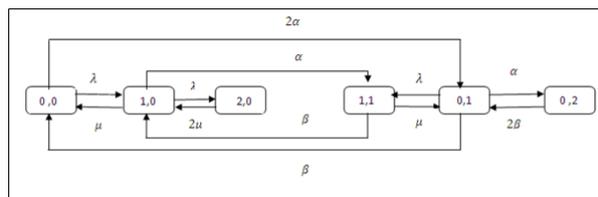


FIGURE 3.3 – une classe du modèle M/M/2 avec rappels et pannes.

Pour chaque $i \geq 0$, les taux de transition du processus $Y(t)$ sont décrits par la matrice :

$$Q_i = \begin{bmatrix} -(\lambda + 2\alpha) & \lambda & 0 & 0 & 2\alpha & 0 \\ \mu & -(\lambda + \alpha + \mu) & \lambda & \alpha & 0 & 0 \\ 0 & 2\mu & -2\mu & 0 & 0 & 0 \\ 0 & \beta & 0 & -(\beta + \mu) & \mu & 0 \\ \beta & 0 & 0 & \lambda & -(\lambda + \alpha + \beta) & \alpha \\ 0 & 0 & 0 & 0 & 2\beta & -2\beta \end{bmatrix}$$

Soit : $P_i = P_{i \setminus i}$

La résolution de l'équation $p_i Q_i = 0$ nécessite la résolution du système suivant

$$\left\{ \begin{array}{l} (\lambda + 2\alpha)P_{1 \setminus i} = \mu P_{2 \setminus i} + \beta P_{5 \setminus i} \\ (\lambda + \alpha + \mu)P_{2 \setminus i} = \lambda P_{1 \setminus i} + 2\mu P_{3 \setminus i} + \beta P_{4 \setminus i} \\ 2\mu P_{3 \setminus i} = \lambda P_{2 \setminus i} \\ (\beta + \mu)P_{4 \setminus i} = \alpha P_{2 \setminus i} + \lambda P_{5 \setminus i} \\ (\lambda + \alpha + \beta)P_{5 \setminus i} = 2\alpha P_{1 \setminus i} + \mu P_{4 \setminus i} + 2\beta P_{6 \setminus i} \\ 2\beta P_{6 \setminus i} = \alpha P_{5 \setminus i} \\ \sum_{l=1}^6 P_{l \setminus i} = 1 \end{array} \right. \quad (3.6)$$

En résolvant le système ci dessus, On obtient les formules des probabilités conditionnelles pour tout $i \geq 0$:

$$P_{1 \setminus i} = \frac{1}{H} (2(\beta + \lambda + \alpha + \mu)\beta^2 \mu^2); \quad (3.7)$$

$$P_{2 \setminus i} = \frac{1}{H} (2(\beta + \mu + \lambda + 2\alpha)\beta^2 \mu \lambda); \quad (3.8)$$

$$P_{3 \setminus i} = \frac{1}{H} ((\beta + \mu + \lambda + 2\alpha)\beta^2 \lambda^2); \quad (3.9)$$

$$P_{4 \setminus i} = \frac{1}{H} (2(\beta + \lambda + 2\mu + 2\alpha)\beta \mu \alpha \lambda); \quad (3.10)$$

$$P_{5 \setminus i} = \frac{1}{H} (2\alpha \beta \mu^2 (\lambda + 2\beta + 2\mu + 2\alpha)); \quad (3.11)$$

$$P_{6 \setminus i} = \frac{1}{H} (\mu^2 \alpha^2 (\lambda + 2\beta + 2\mu + 2\alpha)). \quad (3.12)$$

Où la constante H est donnée par(3.4).

2) Chaque niveau est ensuite considéré comme un état d'une CTMC regroupée, où les taux de transition entre les niveaux correspondent aux clients entrant ou sortant de l'orbite. L'analyse de ce système des états regroupés donne la distribution de probabilité marginale approximative du nombre de clients en orbite.

Les taux de transition entre les états regroupés peuvent être exprimés par

$$q_{ij} = \begin{cases} \alpha P_A P_{2\setminus i} + (\lambda P_I + 2\alpha P_A) P_{3\setminus i} + (\lambda P_I + \alpha P_A) P_{4\setminus i} + \lambda P_I P_{6\setminus i}, & i \geq 0, j = i + 1; \\ i\nu(P_{1\setminus i} + P_{2\setminus i} + P_{5\setminus i}), & i \geq 1, j = i - 1; \\ -[\alpha P_A P_{2\setminus i} + (\lambda P_I + 2\alpha P_A) P_{3\setminus i} + (\lambda P_I + \alpha P_A) P_{4\setminus i} \\ + \lambda P_I P_{6\setminus i} + i\nu(P_{1\setminus i} + P_{2\setminus i} + P_{5\setminus i})], & i = j; \\ 0 & \text{elsewhere.} \end{cases} \quad (3.13)$$

Posons

$$\Lambda = \alpha P_A P_{2\setminus i} + (\lambda P_I + 2\alpha P_A) P_{3\setminus i} + (\lambda P_I + \alpha P_A) P_{4\setminus i} + \lambda P_I P_{6\setminus i};$$

$$\nu = i\nu(P_{1\setminus i} + P_{2\setminus i} + P_{5\setminus i});$$

L'analyse de ce nouveau modèle est analogue à celle du système $M/M/\infty$, donc on obtient la distribution de probabilité marginale

$$\pi_i = \frac{1}{i!} \left(\frac{\Lambda}{\nu}\right)^i \pi_0, \quad i \geq 0; \quad (3.14)$$

En utilisant l'équation de normalisation $\sum_{j=0}^{\infty} \pi_j = 1$ on obtient

$$\pi_0 = e^{-\frac{\Lambda}{\nu}}. \quad (3.15)$$

3) une approximation de la distribution de probabilité conjointe de l'état des serveurs et du nombre de client en orbite à l'état stationnaire peut être obtenue en prenant le produit de la probabilité conditionnelle approximative et la probabilité marginale.

$$P_{il} \approx \hat{P}_{il};$$

Avec

$$\hat{P}_{il} = P_{l\setminus i} \times \pi_i = \frac{P_{l\setminus i}}{i!} \left(\frac{\Lambda}{\nu}\right)^i e^{-\frac{\Lambda}{\nu}}. \quad (3.16)$$

À l'aide de (3.3), la disponibilité approximative des serveurs devient

$$A \approx \sum_{i=0}^{\infty} \hat{P}_{i1} + \sum_{i=0}^{\infty} \hat{P}_{i2} + \sum_{i=0}^{\infty} \hat{P}_{i3}; \quad (3.17)$$

De (3.16), nous avons

$$\sum_{i=0}^{\infty} \hat{P}_{i1} = \sum_{i=0}^{\infty} P_{1\setminus i} \times \pi_i; \quad (3.18)$$

$$\sum_{i=0}^{\infty} \hat{P}_{i2} = \sum_{i=0}^{\infty} P_{2\setminus i} \times \pi_i; \quad (3.19)$$

$$\sum_{i=0}^{\infty} \hat{P}_{i3} = \sum_{i=0}^{\infty} P_{3\setminus i} \times \pi_i. \quad (3.20)$$

En substituant (3.18), (3.19), (3.20) dans (3.17) on obtient :

$$\begin{aligned}
A &\approx \sum_{i=0}^{\infty} P_{1\setminus i} \times \pi_i + \sum_{i=0}^{\infty} P_{2\setminus i} \times \pi_i + \sum_{i=0}^{\infty} P_{3\setminus i} \times \pi_i \\
&\approx \sum_{i=0}^{\infty} \left[\frac{P_{1\setminus i}}{i!} \left(\frac{\Lambda}{\nu}\right)^i e^{-\frac{\Lambda}{\nu}} + \frac{P_{2\setminus i}}{i!} \left(\frac{\Lambda}{\nu}\right)^i e^{-\frac{\Lambda}{\nu}} + \frac{P_{3\setminus i}}{i!} \left(\frac{\Lambda}{\nu}\right)^i e^{-\frac{\Lambda}{\nu}} \right] \\
&\approx \sum_{i=0}^{\infty} \left[(P_{1\setminus i} + P_{2\setminus i} + P_{3\setminus i}) \left(\frac{1}{i!}\right) \left(\frac{\Lambda}{\nu}\right)^i e^{-\frac{\Lambda}{\nu}} \right] \\
&\approx (P_{1\setminus i} + P_{2\setminus i} + P_{3\setminus i}) e^{-\frac{\Lambda}{\nu}} \sum_{i=0}^{\infty} \left(\frac{1}{i!}\right) \left(\frac{\Lambda}{\nu}\right)^i \\
&\approx (P_{1\setminus i} + P_{2\setminus i} + P_{3\setminus i}) e^{-\frac{\Lambda}{\nu}} e^{\frac{\Lambda}{\nu}}
\end{aligned}$$

Donc

$$A \approx P_{1\setminus i} + P_{2\setminus i} + P_{3\setminus i} \quad (3.21)$$

En substituant (3.7), (3.8), (3.9) dans (3.21) et en effectuant tous les calculs nécessaires, on aboutit à la formule (3.3).

Théoreme 3.3.2. Soit F la fréquence de panne des serveurs. La fréquence de défaillance approximative des serveurs est donnée par

$$F \approx \frac{\mu\alpha}{H} [(\beta + \lambda + 2\mu + 2\alpha)(2\beta\lambda + 2\beta\mu + \mu\alpha) + \beta\mu(2\beta + \alpha)]; \quad (3.22)$$

Où la constante H est donnée par(3.4).

Preuve 3.3.2. F est donnée par

$$F = \sum_{i=0}^{\infty} P_{i11} + \sum_{i=0}^{\infty} P_{i01} + \sum_{i=0}^{\infty} P_{i02}; \quad (3.23)$$

En utilisant la table 1 et (3.16), nous obtenons

$$\sum_{i=0}^{\infty} P_{i11} \approx \sum_{i=0}^{\infty} \hat{P}_{i4} = \sum_{i=0}^{\infty} P_{4\setminus i} \times \pi_i; \quad (3.24)$$

$$\sum_{i=0}^{\infty} P_{i01} \approx \sum_{i=0}^{\infty} \hat{P}_{i5} = \sum_{i=0}^{\infty} P_{5\setminus i} \times \pi_i; \quad (3.25)$$

$$\sum_{i=0}^{\infty} P_{i02} \approx \sum_{i=0}^{\infty} \hat{P}_{i6} = \sum_{i=0}^{\infty} P_{6\setminus i} \times \pi_i. \quad (3.26)$$

En substituant (3.24), (3.25), (3.26) dans (3.23) on obtient :

$$\begin{aligned}
F &\approx \sum_{i=0}^{\infty} P_{4\setminus i} \times \pi_i + \sum_{i=0}^{\infty} P_{5\setminus i} \times \pi_i + \sum_{i=0}^{\infty} P_{6\setminus i} \times \pi_i \\
&\approx \sum_{i=0}^{\infty} \left[\frac{P_{4\setminus i}}{i!} \left(\frac{\Lambda}{\nu}\right)^i e^{-\frac{\Lambda}{\nu}} + \frac{P_{5\setminus i}}{i!} \left(\frac{\Lambda}{\nu}\right)^i e^{-\frac{\Lambda}{\nu}} + \frac{P_{6\setminus i}}{i!} \left(\frac{\Lambda}{\nu}\right)^i e^{-\frac{\Lambda}{\nu}} \right] \\
&\approx \sum_{i=0}^{\infty} \left[(P_{4\setminus i} + P_{5\setminus i} + P_{6\setminus i}) \left(\frac{1}{i!}\right) \left(\frac{\Lambda}{\nu}\right)^i e^{-\frac{\Lambda}{\nu}} \right] \\
&\approx (P_{4\setminus i} + P_{5\setminus i} + P_{6\setminus i}) e^{-\frac{\Lambda}{\nu}} \sum_{i=0}^{\infty} \left(\frac{1}{i!}\right) \left(\frac{\Lambda}{\nu}\right)^i \\
&\approx (P_{4\setminus i} + P_{5\setminus i} + P_{6\setminus i}) e^{-\frac{\Lambda}{\nu}} e^{\frac{\Lambda}{\nu}}
\end{aligned}$$

Donc

$$F \approx P_{4\setminus i} + P_{5\setminus i} + P_{6\setminus i} \quad (3.27)$$

En substituant (3.10), (3.11), (3.12) dans (3.27) et après quelques calculs algébrique on obtient la formule (3.22).

3.4 Illustrations numériques

Dans la première partie, nous présentons des illustrations numériques pour évaluer la qualité de l'approximation de "phase merging algorithm" utilisée pour rapprocher les indices de fiabilité du système M/M/2 avec rappels et pannes. (Section (3.4.1)). Dans la deuxième partie, nous étudions l'impact des paramètres sur la disponibilité et la fréquence de panne des serveurs sans tenir compte du taux de rappels. Puisque ces indices sont indépendants de ce paramètre, nous ne varierons donc que λ , μ , α , β . Rappelons que pour que l'algorithme produise des résultats efficaces, les intensités de transition entre les états du même niveau sont significativement plus grande que les intensités de transition circulant entre les niveaux. (Section (3.4.2))

3.4.1 Evaluation de la qualité de l'approximation de l'algorithme "phase merging"

Pour évaluer la qualité des approximations que nous avons obtenu, nous présentons deux exemples.

Example 1. Dans cet exemple, nous fixons $\mu = 6$, $\lambda = 4$, $\beta = 0.1$ et varions α de 0 à 2 par incréments de 0.01.

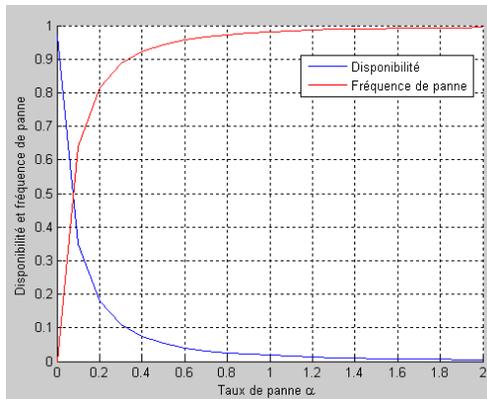


FIGURE 3.4 – Disponibilité et fréquence de panne vs. α

La figure 3.4 décrit l'influence du paramètre α sur la disponibilité et la fréquence de panne des serveurs. Comme prévu (intuitivement), avec l'augmentation du taux de pannes α , la probabilité de trouver un serveur en bon état est presque nulle ce qui conduit à diminuer la disponibilité des serveurs et augmenter la fréquence de pannes des serveurs.

Exemple 2. Dans cet exemple, nous fixons $\mu = 5$, $\alpha = 0.1$, $\lambda = 2$ et varions β de 0 à 1 par pas de 0.01.

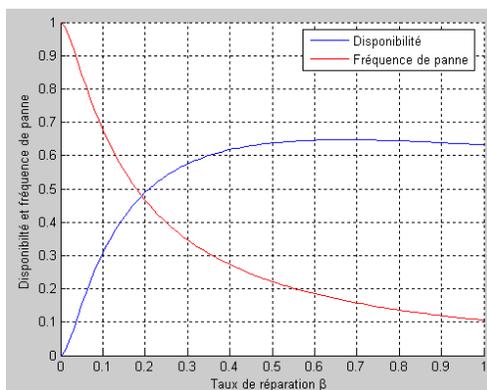


FIGURE 3.5 – Disponibilité et fréquence de panne vs. β

Dans la figure 3.5, nous décrivons les variations de la disponibilité et de la fréquence de panne des serveurs en fonction du taux de réparation. Il est également intuitif que si le taux de service est supérieur au taux d'arrivée et avec l'augmentation du taux de réparation, les serveurs seront souvent disponibles ce qui conduit à diminuer la fréquence de panne.

3.4.2 Impact des paramètres sur les indices de fiabilité approximatifs calculés

Premièrement, nous présentons un exemple où l'hypothèse selon laquelle les taux à l'intérieur des niveaux de l'orbite sont significativement supérieurs à ceux entre les niveaux est violée. Puis nous donnons deux exemples qui nous montrent comment choisir les paramètres λ et μ pour avoir une bonne disponibilité de serveur.

Exemple 3. Nous fixons $\mu = 10$, $\beta = 5$, $\alpha = 30$ et varions λ de 0 à 200 par pas de 0.1.

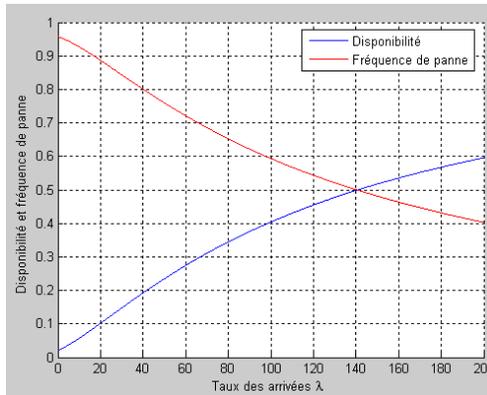


FIGURE 3.6 – Disponibilité et fréquence de panne vs. λ

Dans cet exemple, nous pouvons voir que l'hypothèse est violée. Dans la figure 3.6, nous montrons que la disponibilité augmente rapidement et que la fréquence des défaillances diminue, alors que le taux de réparation est très faible comparé au taux de panne. En outre, le taux de service est faible et le taux d'arrivée est en augmentation. Donc, si les hypothèses mentionnées ci-dessus, sont violées, la méthode peut fonctionner très mal.

Exemple 4. Dans cet exemple, dans un premier temps, nous fixons $\lambda = 10$, $\alpha = 0.1$, $\beta = 2$ et varions μ de 1 à 200 par pas de 0.1, ensuite nous mettons $\lambda = 70$ et gardons les mêmes valeurs pour les autres paramètres.

En regardant la figure 3.7, nous remarquons que la disponibilité diminue avec des valeurs croissantes de μ , d'autre part la fréquence de panne connaît une petite augmentation puis elle diminue. Cependant, la disponibilité sera plus grande lorsque le taux d'arrivée est élevé, donc pour une bonne disponibilité du serveur et une fréquence de panne plus petite, les paramètres λ et μ doivent se rapprocher.

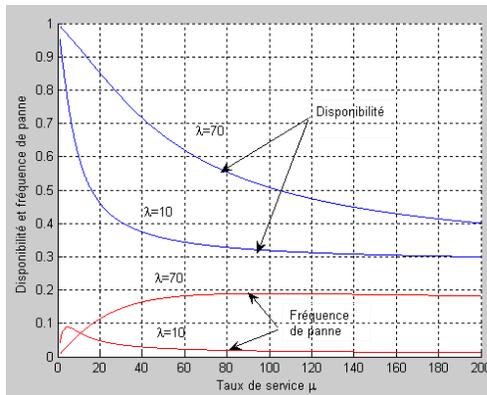


FIGURE 3.7 – Disponibilité et fréquence de panne vs. μ

Example 5. Finalement, nous fixons $\lambda = 4$, $\mu = 6$, $\alpha = 0.1$ et varions β de 0 à 5 par pas de 0.01.

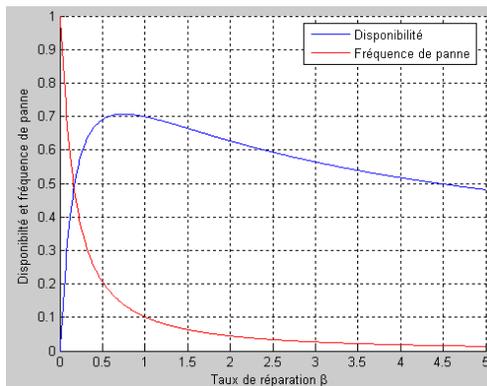


FIGURE 3.8 – Disponibilité et fréquence de panne vs. β

La figure 3.8 décrit l'influence du paramètre β sur la disponibilité et la fréquence de panne.

Nous observons qu'avec des valeurs croissantes de taux de réparation, la fréquence de panne diminue et la disponibilité du serveur augmente (ce qui est très logique). Mais à partir d'un certain rang des taux de réparation, la disponibilité commence à diminuer, donc un petit taux de panne a besoin d'un petit taux de réparation.

3.5 Conclusion

Le système M/M/2 avec rappels et pannes a été étudié. Des approximations de quelques indices de fiabilité du modèle sont obtenues en utilisant les probabilités approximatives d'état

déterminées en appliquant l'algorithme de phase merging. Ces investigations théoriques sont illustrées par des illustrations numériques. Le modèle présenté dans ce chapitre est nouveau et il est réaliste car il peut être applicable à diverses situations de congestion rencontrées dans les systèmes de télécommunication, les réseaux informatiques et les banques qui impliquent l'utilisation de plusieurs serveurs en même temps.

CHAPITRE 4

MODELE D'ATTENTE AVEC RAPPELS, PANNES ET PRIORITE

4.1 Introduction

Dans ce chapitre, nous considérons le système d'attente de type $M/G/1$ avec rappels, pannes et priorité. Premièrement, nous discutons la condition d'ergodicité, nous obtenons les fonctions génératrices partielles de la distribution stationnaire de l'état du système et quelques mesures de performances en utilisant la méthode des variables supplémentaires. Dans un deuxième temps, nous obtenons quelques indices de fiabilité d'un tel système. Ces résultats font l'objet de l'étude réalisée dans [93] et ils sont des extensions de ceux de BOUTERFA et DJELLAB [12], obtenus pour le système $M/G/1$ avec rappels et priorité, pour lequel le serveur est absolument fiable. La section suivante sera consacrée aux résultats de BOUTERFA et DJELLAB [12] que nous utilisons pour la suite du travail.

4.2 Modèle d'attente avec rappels et priorité absolue

4.2.1 Description du modèle

Considérons un système de files d'attente à un seul serveur, auquel deux types différents de clients primaires arrivent selon des processus de Poisson indépendants de taux λ_1 et λ_2 respectivement. Les clients du premier flux de taux $\lambda_1 > 0$ ont une priorité absolue sur les clients du deuxième flux de taux $\lambda_2 > 0$. Ainsi, les règles suivantes gouvernent la dynamique

des clients :

- Tout client primaire arrivant et trouvant le serveur inoccupé, occupe immédiatement le serveur et quitte le système après l'achèvement du service.

- Tout client prioritaire arrivant et trouvant le serveur occupé par le service d'un autre client prioritaire, rejoint l'orbite. La durée entre deux rappels successifs d'un même client en orbite suit une loi exponentielle de taux $\theta > 0$.

- Tout client prioritaire arrivant (primaire ou de l'orbite) et trouvant le serveur occupé par le service d'un client non prioritaire, entre directement dans la station de service. Le client non prioritaire, dont le service a été interrompu, persiste dans la zone de service jusqu'à l'achèvement du service du client prioritaire pour reprendre son service d'où il a été interrompu (la politique "preemptive resume").

- Tout client primaire non prioritaire arrivant et trouvant le serveur occupé quitte le système sans service.

Les durées de service des deux types de clients suivent une distribution générale de fonction de répartition $B_i(x)$, $i \in \{1, 2\}$, de transformée de Laplace- Stieltjes

$\tilde{B}_i(s) = \int_0^\infty \exp(-sx) dB_i(x)$ et $\text{Re}(s) > 0$. Soient les moments d'ordre k , $\beta_{i,k} = (-1)^k \tilde{B}_i^{(k)}(0)$ de la durée de service à partir de l'origine ; $b_i(x) = \frac{B_i'(x)}{1-B_i(x)}$, est l'intensité instantanée du service du client de type i , étant donné que la durée de service écoulée est égale à x . $K_i(z_1, z_2) = \tilde{B}_i(\lambda_1 - \lambda_1 z_1 + \lambda_2 - \lambda_2 z_2)$ est la fonction génératrice du nombre de clients primaires de deux types arrivant dans le système durant le temps de service d'un client de type i ($i \in \{1, 2\}$). Finalement, nous admettons l'hypothèse de l'indépendance mutuelle entre toutes les variables aléatoires définies précédemment.

4.2.2 La condition de stabilité et la distribution stationnaire de l'état du système

L'état du système à la date t peut être décrit à l'aide d'un processus de Markov à temps continu, tel que

$$\{C(t), N_o(t), \xi_1(t), \xi_2(t), t \geq 0\}, \quad (4.1)$$

Où $N_o(t)$ est le nombre de clients prioritaires en orbite et $C(t)$ représente l'état de station-service à la date t . $C(t)$ est égal à 0, 1, 2 ou 3 selon le fait que le serveur est inoccupé, un client prioritaire est servi et il n'y a aucun client non prioritaire dont le service a été interrompu dans la zone de service, un client non prioritaire est servi ou un client prioritaire est servi et il y a un

client non prioritaire dont le service a été interrompu dans la zone de service. Si $C(t) \in \{1, 3\}$ (If $C(t) = 2$), $\xi_1(t)$, $(\xi_2(t))$ représente le temps de service écoulé du client prioritaire (du client non prioritaire) dans le serveur à la date t .

Les probabilités d'états sont définies par :

$$P_{0i} = \lim_{t \rightarrow \infty} P(C(t) = 0, N_o(t) = i), i \geq 0;$$

$$P_{1i}(x) = \lim_{t \rightarrow \infty} \frac{d}{dx} P(C(t) = 1, \xi_1(t) \leq x, N_o(t) = i), i \geq 0 \text{ et } x \geq 0;$$

$$P_{2i}(y) = \lim_{t \rightarrow \infty} \frac{d}{dy} P(C(t) = 2, \xi_2(t) \leq y, N_o(t) = i), i \geq 0 \text{ et } y \geq 0;$$

$$P_{3i}(x, y) = \lim_{t \rightarrow \infty} \frac{d}{dx} P(C(t) = 3, \xi_1(t) \leq x, \xi_2(t) \leq y, N_o(t) = i), i \geq 0, x \geq 0 \text{ et } y \geq 0.$$

Soit $\rho_1 = \lambda_1 \beta_{1,1}$, la condition nécessaire et suffisante pour que ce système soit stable est : $\rho_1 < 1$.

- à l'état stationnaire, la fonction génératrice du nombre de clients en orbite est

$$Q(z_1) = \left(\frac{(1-z)(1+\lambda_2(\int_0^\infty (1-B_2(y))dy))}{2(1+\lambda_1+\lambda_2)\lambda_2(K_1(z_1,1)-z_1)} \exp \left\{ \int_1^z \frac{\lambda_1 - \lambda_1 K_1(z_1,1)}{\theta(K_1(z_1,1))} dz_1 \right\} \right).$$

- La fonction génératrice du nombre de clients prioritaires dans le système

$$H(z_1) = \frac{K_1(z_1,1)(1-z_1)(1+\lambda_2(\int_0^\infty (1-B_2(y))dy))}{2(1+\lambda_1+\lambda_2)\lambda_2(K_1(z_1,1)-z_1)} \exp \left\{ \int_1^z \frac{\lambda_1 - \lambda_1 K_1(z_1,1)}{\theta(K_1(z_1,1))} dz_1 \right\}.$$

- La probabilité p_1 que le serveur est occupé par le service d'un client prioritaire et il n'y a aucun client non prioritaire dont le service a été interrompu dans la station-service

$$p_1 = \frac{\rho_1}{2\lambda_2(1-\rho_1)(1+\lambda_1+\lambda_2)}.$$

- La probabilité p_2 que le serveur est occupé par le service d'un client non prioritaire

$$p_2 = \frac{1}{2(1+\lambda_1+\lambda_2)} \int_0^\infty (1-B_2(y))dy.$$

- La probabilité p_3 que le serveur est occupé par le service d'un client prioritaire et il y a un client non prioritaire dont le service a été interrompu dans la station service

$$p_3 = \frac{\rho_1}{2(1+\lambda_1+\lambda_2)(1-\rho_1)} \int_0^\infty (1-B_2(y))dy$$

- Nombre moyen de clients prioritaires en orbite \bar{n}_o

$$\bar{n}_o = \frac{2\rho_1\lambda_1 + \theta(1-\rho_1)\beta_{1,2} + 4\rho_1\lambda_1\beta_{1,1} + 2\lambda_1\beta_{1,1}^2}{4\theta(1-\rho_1)(1+\lambda_1+\lambda_2)\lambda_2(1+\beta_{1,1})^2} \left(1 + \lambda_2 \int_0^\infty (1-B_2(y))dy \right)$$

- Nombre moyen de clients prioritaires dans le système \bar{n}

$$\bar{n} = \frac{2\rho_1\lambda_1 + \theta(1 - \rho_1\beta_{1,2} + 4\rho_1\lambda_1\beta_{1,1} + 2\lambda_1\beta_{1,1}^2)}{4\theta(1 - \rho_1)(1 + \lambda_1 + \lambda_2)\lambda_2(1 + \beta_{1,1})^2} \times \left(1 + \lambda_2 \int_0^\infty (1 - B_2(y)) dy \right) + \frac{\rho_1 (1 + \lambda_2 \int_0^\infty (1 - B_2(y)) dy)}{2(1 - \rho_1)\lambda_2(1 + \lambda_1 + \lambda_2)}.$$

Dans le reste de ce chapitre, nous allons présenter notre contribution à ce sujet, en faisant l'extension à un système M/G/1 avec rappels, priorité et serveur non fiable.

4.3 Modèle M/G/1 avec rappels, pannes et priorité absolue

4.3.1 description du modèle

Nous considérons un système de files d'attente à un seul serveur avec deux types différents de clients primaires qui arrivent selon des processus de poisson indépendants de taux λ_1 et λ_2 . Les durées de services des deux types de clients suivent une distribution générale de fonction de répartition $B_i(x)$, $i \in \{1, 2\}$, de transformée de Laplace-Stieltjes $\tilde{B}_i(s) = \int_0^\infty \exp(-sx) dB_i(x)$ et $\text{Re}(s) > 0$. Soient les moments d'ordre k , $\beta_{i,k} = (-1)^k \tilde{B}_i^{(k)}(0)$ de la durée de service à partir de l'origine ; $\beta_{i,k} = (-1)^k \tilde{B}_i^{(k)}(0)$ est l'intensité instantanée du service du client de type i , étant donné que la durée de service écoulée est égale à x . $K_i(z_1, z_2) = \tilde{B}_i(\lambda_1 - \lambda_1 z_1 + \lambda_2 - \lambda_2 z_2)$ est la fonction génératrice du nombre de clients primaires des deux types de clients arrivant dans le système durant le temps de service d'un client de type i . Les clients du premier flux de taux $\lambda_1 > 0$ ont une priorité absolue sur les clients du deuxième flux de taux $\lambda_2 > 0$. Si un client primaire trouve le serveur libre, il occupe immédiatement le serveur et il quitte le système après l'achèvement du service.

Nous supposons que le serveur est sujet à des pannes actives. Les pannes se produisent indépendamment via un processus de poisson de taux α_1 et α_2 respectivement. Une fois le serveur occupé tombe en panne, la réparation commence immédiatement. Si le serveur tombe en panne au moment de servir les clients de type i , ($i = \{1; 2\}$) les durées de réparation sont des variables aléatoires suivent une distribution générale de fonction de répartition $G_i(x)$, de fonction de densité $g_i(x)$, de transformée de Laplace-Stieltjes $\tilde{G}_i(s) = \int_0^\infty \exp(-sx) dG_i(x)$ et $\text{Re}(s) > 0$. Soit $\gamma_i(x) = \frac{G_i'(x)}{1 - G_i(x)}$ l'intensité instantanée de réparation lorsque la rupture du type i se produit, étant donné que la durée de réparation écoulée est égale à y . $A_i(z) = \tilde{G}_i(\lambda_1 - \lambda_1 z)$ est la fonction génératrice du nombre de clients primaires de deux types de clients arrivant dans le système durant le temps de réparation d'un client de type i .

Donc l'état du serveur peut être libre, occupé ou sous réparation. Si le serveur se trouve occupé ou sous réparation, le comportement d'un client primaire arrivant dépend de son type :

- Tout client primaire prioritaire arrivant et trouvant le serveur occupé par un autre client prioritaire ou sous réparation rejoint l'orbite. La durée entre deux rappels successifs d'un même client en orbite suit une loi exponentielle de taux $\theta > 0$.

- Tout client prioritaire arrivant et trouvant le serveur occupé par le service d'un client non prioritaire occupe immédiatement le serveur. Le client dont le service a été interrompu, persiste dans la zone de service jusqu'à ce que le serveur devient disponible pour reprendre son service d'où il a été interrompu.

- Tout client primaire non prioritaire arrivant et trouvant le serveur occupé quitte le système sans service.

Toutes les variables aléatoires définies précédemment vérifient l'hypothèse d'indépendance mutuelle.

L'état du système à la date t peut être décrit à l'aide d'un processus de Markov à temps continu, tel que

$$X(t) = \{C(t), N(t), \xi_1(t), \xi_2(t), \xi_3(t), t \geq 0\}, \quad (4.2)$$

Où

$N(t)$ est le nombre de clients prioritaires en orbite.

$C(t)$ représente l'état du serveur à la date t .

$C(t) = 0$ si le serveur est inoccupé.

$C(t) = 1$ si un client prioritaire est servi et il n'y a aucun client non prioritaire dont le service a été interrompu dans la station-service.

$C(t) = 2$ si un client non prioritaire est servi.

$C(t) = 3$ si un client prioritaire est servi et il y a un client non prioritaire dont le service a été interrompu dans la station-service.

$C(t) = 4$ si le serveur tombe en panne lorsqu'un client prioritaire est servi et il n'y a aucun client non prioritaire dont le service a été interrompu dans la station-service.

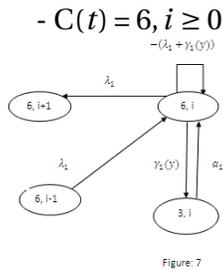
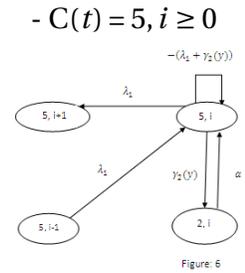
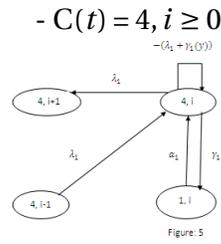
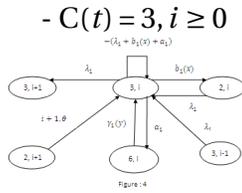
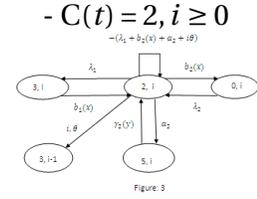
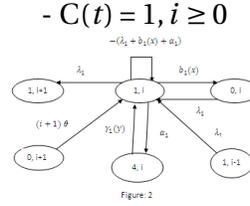
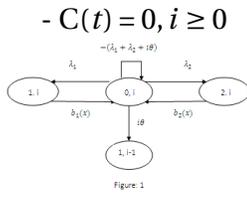
$C(t) = 5$ si le serveur tombe en panne lorsqu'un client non prioritaire est servi.

$C(t) = 6$ si le serveur tombe en panne lorsqu'un client prioritaire est servi et il y a un client non prioritaire dont le service a été interrompu dans la station-service.

Si $C(t) \in \{1, 3, 4, 6\}$ (If $C(t) \in \{2, 5\}$), $\xi_1(t)$, $(\xi_2(t))$ représente le temps de service écoulé du

client prioritaire (du client non prioritaire) dans le serveur à la date t . Si $C(t) \in \{4, 5, 6\}$ ($\xi_3(t)$) représente le temps de réparation écoule du serveur en cours de réparation.

Les transitions entre les états sont définies comme suit :



4.3.2 Condition d'ergodicité

Pour prouver l'ergodicité, nous définissons T_{in} , le temps de service généralisé de n ème client de type i , c'est la durée du début de service de n ème client de type i jusqu'à fin de son service. Ainsi, T_{in} inclut certains temps d'arrêt possibles du serveur en raison des pannes et des réparations du serveur pendant la période de service du n ème client de type i . Il a été montré dans [94] que les durées de service généralisées successives T_{in} sont des variables aléatoires identiques et généralement distribuées avec une fonction de distribution

$$D_i(t) \triangleq \Pr\{T_{in} \leq t\} = \sum_{i=0}^{\infty} \int_0^t G_i^{(l)}(t-u) e^{-\alpha_i u} \frac{(\alpha_i u)^l}{l!} dB_i(u), i = 1, 2,$$

qui est indépendante de n .

une transformée de Laplace-Stieltjes

$$\tilde{D}_i(s) = \int_0^\infty \exp(-st) dD_i(t) = \tilde{B}_i(s + \alpha_i - \alpha_i \tilde{G}_i(s));$$

et une valeur moyenne donnée par

$$E(T_{in}) = -\frac{dD_i(s)}{ds} \Big|_{s=0} = \beta_{i1}(1 + \alpha_i \gamma_{i1}). \quad (4.3)$$

Soit t_d la date de d ème depart. La suite de vecteurs aléatoires $X_d = (U_d, N_{1,d}, N_{2,d})$, $d \geq 1$, forme une chaîne de Markov à espace d'états $S = \{1, 2\} \times Z_+ \times \{0, 1\}$, où $N_{1,d}$ ($N_{2,d}$) est le nombre de clients prioritaires (non prioritaires) en orbite resp. (dans la station service) juste avant l'instant t_d et U_d est le type de d ème client servis qui peut être 1 ou 2 selon que le client en question est un client prioritaire ou non prioritaire. Nous avons les équations fondamentales suivantes :

$$N_{1,d} = N_{1,d-1} - V_d + Y_{1,d}; \quad (4.4)$$

$$N_{2,d} = N_{2,d-1} \times B_d, Y_{2,d} = 0. \quad (4.5)$$

Où V_d est égale à 0 ou 1, selon que le d ème client servi est client prioritaire primaire ou secondaire ; $Y_{1,d}$ ($Y_{2,d}$) est le nombre de clients prioritaire (non prioritaire) arrivant au système pendant le temps de service généralisé du d ème client ; B_d est égale à 0 ou 1 selon que le d ème client servi est non prioritaire ou prioritaire. Dénontons par U_d le type du d ème client servi (qui peut être 1 ou 2 selon que le client en question est un client prioritaire ou non prioritaire).

Le vecteur aléatoire (U_d, V_d) dépend de l'histoire du système avant le temps t_d seulement par le vecteur $(N_{1,d-1}, N_{2,d-1})$. Sa distribution conditionnelle est définie par

$$P((U_d = 1, V_d = 1) / ((N_{1,d-1}, N_{2,d-1}) = (m, n))) = \frac{m\theta}{\lambda_1 + \lambda_2 + m\theta};$$

$$P((U_d = 1, V_d = 0) / ((N_{1,d-1}, N_{2,d-1}) = (m, n))) = \frac{\lambda_1}{\lambda_1 + \lambda_2 + m\theta};$$

$$P((U_d = 2, V_d = 1) / ((N_{1,d-1}, N_{2,d-1}) = (m, n))) = 0;$$

$$P((U_d = 2, V_d = 0) / ((N_{1,d-1}, N_{2,d-1}) = (m, n))) = \frac{\lambda_2}{\lambda_1 + \lambda_2 + m\theta}.$$

Soit

$$k_{i,m,n} = \int_0^\infty \frac{(\lambda_1 x)^m}{m!} e^{-\lambda_1 x} \frac{(\lambda_2 x)^n}{n!} e^{-\lambda_2 x} dD_i(x)$$

la distribution conjointe du nombre de clients primaires des deux types qui arrivent pendant le temps de service généralisé de i ème client.

Les probabilités de transition à une étape de la chaîne de Markov induite $r_{(l,m,n)(c,i,j)}$ = $P(X_d = (c, i, j) / X_{d-1} = (l, m, n))$, où $c \in \{1, 2\}$, sont données par :

$$r_{(l,m,n)(2,i,j)} = \frac{\lambda_2}{\lambda_1 + \lambda_2 + m\theta} k_{2,i-m,0}, n \in \{0, 1\} m = 1; \quad (4.6)$$

$$r_{(l,m,n)(1,i,j)} = \frac{\lambda_1}{\lambda_1 + \lambda_2 + m\theta} k_{1,i-m,0} + \frac{m\theta}{\lambda_1 + \lambda_2 + m\theta} k_{1,i-m+1,0}. \quad (4.7)$$

Le problème à étudier par la suite est celui de l'ergodicité de notre chaîne. Comme les équations fondamentales ont une structure récursive, nous utilisons le critère de Foster [95] et nous considérons la fonction de Lyapounov $f(s)$, $s \in S$ tel que $f(l, m, n) = m + (1 - \rho_1)n$, où $\rho_1 = \lambda_1 \beta_{11} (1 + \alpha_1 \gamma_{11})$, l'accroissement moyen de la chaîne induite $x_{l,m,n} = E[(f(X_d) - f(X_{d-1})) / X_{d-1} = (l, m, n)]$ peut être obtenu de la manière suivante :

- Si $n = 0$, les équations (4.4), (4.5) deviennent $N_{1,d} = N_{1,d-1} - V_d + Y_{1,d} = m - V_d + Y_{1,d}$; $N_{2,d} = 0$. Donc $f(X_{d-1}) = m$ et $f(X_d) = m - V_d + Y_{1,d}$, ce qui implique que

$$\begin{aligned} x_{l,m,n}^{(1)} &= E[(-V_d + Y_{1,d}) / ((N_{1,d-1}, N_{2,d-1}) = (m, 0))] \\ &= E[Y_{1,d} / (N_{1,d-1}, N_{2,d-1}) = (m, 0)] - E[V_d / (N_{1,d-1}, N_{2,d-1}) = (m, 0)] \\ &= \frac{\lambda_1 + m\theta}{\lambda_1 + \lambda_2 + m\theta} \lambda_1 \beta_{11} (1 + \alpha_1 \gamma_{11}) - \frac{m\theta}{\lambda_1 + \lambda_2 + m\theta}. \end{aligned}$$

- Si $n = 1$, les équations (4.4), (4.5) deviennent $N_{1,d} = N_{1,d-1} - V_d + Y_{1,d} = m - V_d + Y_{1,d}$; $N_{2,d} = B_d$. Donc $f(X_d) - f(X_{d-1}) = -V_d + Y_{1,d} + (1 - \rho_1)(B_d - 1)$. Ainsi

$$\begin{aligned} x_{l,m,n}^{(2)} &= E[(-V_d + Y_{1,d} + (1 - \rho_1)(B_d - 1)) / (N_{1,d-1}, N_{2,d-1}) = (m, 1)] \\ &= E[(-V_d + Y_{1,d}) / (N_{1,d-1}, N_{2,d-1}) = (m, 1)] + (1 - \rho_1) E[U_d / (N_{1,d-1}, N_{2,d-1}) = (m, 1)] - (1 - \rho_1) \\ &= -\frac{m\theta}{\lambda_1 + \lambda_2 + m\theta} + \frac{\lambda_1 + m\theta}{\lambda_1 + \lambda_2 + m\theta} \lambda_1 \beta_{11} (1 + \alpha_1 \gamma_{11}) + (1 - \rho_1) \left(\frac{\lambda_1 + m\theta}{\lambda_1 + \lambda_2 + m\theta} - 1 \right) \end{aligned}$$

$\lim_{m \rightarrow \infty} x_{l,m,n}^{(1)} = \lim_{m \rightarrow \infty} x_{l,m,n}^{(2)} = \rho_1 - 1$. Cette limite est négative si $\rho_1 < 1$. Donc, la condition suffisante est $\rho_1 < 1$. Puisque $r_{(l,m,n)(1,i,j)} = r_{(l,m,n)(2,i,j)} = 0$ pour $i < m - 1$, $\rho_1 < 1$ est aussi une condition nécessaire pour l'ergodicité de notre chaîne de Markov induite. Finalement, la condition nécessaire et suffisante pour que le système soit stable est $\rho_1 < 1$.

4.3.3 Distribution stationnaire de l'état du système

Dans cette section, nous étudions le système à l'état stationnaire. Par conséquent, nous supposons que la condition $\rho_1 < 1$ est réalisée et introduisons les fonctions suivantes qui décrivent la distribution conjointe de l'état du serveur et le nombre de clients en orbite :

$$\begin{aligned}
P_{0i} &= \lim_{t \rightarrow \infty} P(C(t) = 0, N(t) = i); \\
P_{1i}(x_1) &= \lim_{t \rightarrow \infty} \frac{d}{dx_1} P(C(t) = 1, \xi_1(t) \leq x_1, N(t) = i); \\
P_{2i}(x_2) &= \lim_{t \rightarrow \infty} \frac{d}{dx_2} P(C(t) = 2, \xi_2(t) \leq x_2, N(t) = i); \\
P_{3i}(x_1, x_2) &= \lim_{t \rightarrow \infty} \frac{d}{dx_1} P(C(t) = 3, \xi_1(t) \leq x_1, \xi_2(t) \leq x_2, N(t) = i); \\
P_{4i}(x_1, y) &= \lim_{t \rightarrow \infty} \frac{d}{dy} P(C(t) = 4, \xi_1(t) \leq x_1, \xi_3(t) \leq y, N(t) = i); \\
P_{5i}(x_2, y) &= \lim_{t \rightarrow \infty} \frac{d}{dy} P(C(t) = 5, \xi_2(t) \leq x_2, \xi_3(t) \leq y, N(t) = i); \\
P_{6i}(x_1, x_2, y) &= \lim_{t \rightarrow \infty} \frac{d}{dy} P(C(t) = 6, \xi_1(t) \leq x_1, \xi_2(t) \leq x_2, \xi_3(t) \leq y, N(t) = i).
\end{aligned}$$

Alors, en suivant la méthode des variables supplémentaires, nous constatons que les probabilités considérées satisfont les équations d'équilibre statistique suivantes :

$$(\lambda_1 + \lambda_2 + i\theta)P_{i0} = \int_0^\infty P_{1i}b_1(x_1)dx_1 + \int_0^\infty P_{2i}b_2(x_2)dx_2; \quad (4.8)$$

$$P'_{1i} = -(\lambda_1 + b_1(x_1) + \alpha_1)P_{1i}(x_1) + \lambda_1 P_{1i-1}(x_1) + \int_0^\infty P_{4i}(x_1, y)\gamma_1(y)dy; \quad (4.9)$$

$$P_{1i}(0) = \lambda_1 P_{0i} + (i+1)\theta P_{0i+1}; \quad (4.10)$$

$$\begin{aligned}
P'_{2i} &= -(\lambda_1 + b_2(x_2) + \alpha_2 + i\theta)P_{2i}(x_2) + \int_0^\infty P_{3i}(x_1, x_2)b_1(x_1)dx_1 \\
&\quad + \int_0^\infty P_{5i}(x_2, y)\gamma_2(y)dy;
\end{aligned} \quad (4.11)$$

$$P_{2i}(0) = \lambda_2 P_{0i}; \quad (4.12)$$

$$P'_{3i} = -(\lambda_1 + b_1(x_1) + \alpha_1)P_{3i}(x_1, x_2) + \lambda_1 P_{3i-1}(x_1, x_2) + \int_0^\infty P_{6i}(x_1, x_2, y) \gamma_1(y) dy; \quad (4.13)$$

$$P_{3i}(0, x_2) = \lambda_1 P_{2i}(x_2) + (i+1)\theta P_{2i+1}(x_2); \quad (4.14)$$

$$\frac{\partial P_{4i}(x_1, y)}{\partial y} = -(\lambda_1 + \gamma_1(y))P_{4i}(x_1, y) + \lambda_1 P_{4i-1}(x_1, y); \quad (4.15)$$

$$P_{4i}(x_1, 0) = \alpha_1 P_{1i}(x_1); \quad (4.16)$$

$$\frac{\partial P_{5i}(x_2, y)}{\partial y} = -(\lambda_1 + \gamma_2(y))P_{5i}(x_2, y) + \lambda_1 P_{5i-1}(x_2, y); \quad (4.17)$$

$$P_{5i}(x_2, 0) = \alpha_2 P_{2i}(x_2); \quad (4.18)$$

$$\frac{\partial P_{6i}(x_1, x_2, y)}{\partial y} = -(\lambda_1 + \gamma_1(y))P_{6i}(x_1, x_2, y) + \lambda_1 P_{6i-1}(x_1, x_2, y); \quad (4.19)$$

$$P_{6i}(x_1, x_2, 0) = \alpha_1 P_{3i}(x_1, x_2). \quad (4.20)$$

La résolution du système (4.8)-(4.20) est très compliquée. A l'aide des fonctions génératrices $P_1(z, x_1) = \sum_{i=0}^{\infty} z^i P_{1i}(x_1)$, $P_2(z, x_2) = \sum_{i=0}^{\infty} z^i P_{2i}(x_2)$, $P_3(z, x_1, x_2) = \sum_{i=0}^{\infty} z^i P_{3i}(x_1, x_2)$, $P_4(z, x_1, y) = \sum_{i=0}^{\infty} z^i P_{4i}(x_1, y)$, $P_5(z, x_2, y) = \sum_{i=0}^{\infty} z^i P_{5i}(x_2, y)$, $P_6(z, x_1, x_2, y) = \sum_{i=0}^{\infty} z^i P_{6i}(x_1, x_2, y)$, le système (4.8)-(4.20) devient

$$z\theta \frac{dP_0(z)}{dz} = -(\lambda_1 + \lambda_2)P_0(z) + \int_0^\infty P_1(z, x_1)b_1(x_1)dx_1 + \int_0^\infty P_2(z, x_2)b_2(x_2)dx_2; \quad (4.21)$$

$$\frac{\partial P_1(z, x_1)}{\partial x_1} = -(\lambda_1 + b_1(x_1) + \alpha_1 - \lambda_1 z)P_1(z, x_1) + \int_0^\infty P_4(z, x_1, y)\gamma_1(y)dy; \quad (4.22)$$

$$P_1(z, 0) = \lambda_1 P_0(z) + \theta \frac{dP_0(z)}{dz}; \quad (4.23)$$

$$\begin{aligned} \frac{\partial P_2(z, x_2)}{\partial x_2} &= -(\lambda_1 + b_2(x_2) + \alpha_2)P_2(z, x_2) - \theta z \frac{\partial P_2(z, x_2)}{\partial z} + \int_0^\infty P_3(z, x_1, x_2)b_1(x_1)dx_1 \\ &+ \int_0^\infty P_5(z, x_2, y)\gamma_2(y)dy; \end{aligned} \quad (4.24)$$

$$P_2(z, 0) = \lambda_2 P_0(z); \quad (4.25)$$

$$\frac{\partial P_3(z, x_1, x_2)}{\partial x_1} = -(\lambda_1 + b_1(x_1) + \alpha_1 - \lambda_1 z)P_3(z, x_1, x_2) + \int_0^\infty P_6(z, x_1, x_2, y)\gamma_1(y)dy; \quad (4.26)$$

$$P_3(z, 0, x_2) = \lambda_1 P_2(z, x_2) + \theta \frac{\partial P_2(z, x_2)}{\partial z}; \quad (4.27)$$

$$\frac{\partial P_4(z, x_1, y)}{\partial y} = -(\lambda_1 + \gamma_1(y) - \lambda_1 z)P_4(z, x_1, y); \quad (4.28)$$

$$P_4(z, x_1, 0) = \alpha_1 P_1(z, x_1); \quad (4.29)$$

$$\frac{\partial P_5(z, x_2, y)}{\partial y} = -(\lambda_1 + \gamma_2(y) - \lambda_1 z)P_5(z, x_2, y); \quad (4.30)$$

$$P_5(z, x_2, 0) = \alpha_2 P_2(z, x_2); \quad (4.31)$$

$$\frac{\partial P_6(z, x_1, x_2, y)}{\partial y} = -(\lambda_1 + \gamma_1(y) - \lambda_1 z) P_6(z, x_1, x_2, y); \quad (4.32)$$

$$P_6(z, x_1, x_2, 0) = \alpha_1 P_3(z, x_1, x_2). \quad (4.33)$$

La résolution du système (4.21)-(4.33) est moins compliquée que celle du système (4.8)-(4.20) et peut être faite en procédant de la manière suivante : des équations (4.28), (4.30) et (4.32) on trouve $P_4(z, x_1, y)$, $P_5(z, x_2, y)$ et $P_6(z, x_1, x_2, y)$ en fonction de $P_1(z, x_1)$, $P_2(z, x_2)$ et $P_3(z, x_1, x_2)$ (respectivement) comme suit :

- l'intégration des deux parties de l'équation (4.28) donne

$$\begin{aligned} \int_0^y \frac{dP_4(z, x_1, y)}{P_4(z, x_1, y)} &= \int_0^y (-\lambda_1 - \gamma_1(y) + \lambda_1 z) dy; \\ \int_0^y \frac{dP_4(z, x_1, y)}{P_4(z, x_1, y)} &= \int_0^y \frac{-G_1'(y)}{1 - G_1(y)} dy + \int_0^y (-\lambda_1 + \lambda_1 z) dy; \\ \ln(P_4(z, x_1, y)) - \ln(P_4(z, x_1, 0)) &= [\ln(1 - G_1(y)) - \ln(1 - G_1(0))] + (-\lambda_1 + \lambda_1 z)y; \\ \ln\left(\frac{P_4(z, x_1, y)}{P_4(z, x_1, 0)}\right) &= \ln(1 - G_1(y)) + (-\lambda_1 + \lambda_1 z)y, ((G_1(0) = 0, (\text{par convention}) \Rightarrow (\ln(1 - G_1(0)) = 0)); \end{aligned}$$

D'où

$$\begin{aligned} P_4(z, x_1, y) &= P_4(z, x_1, 0) (1 - G_1(y)) \exp\{(-\lambda_1 + \lambda_1 z) y\}; \\ &= \alpha_1 P_1(z, x_1) (1 - G_1(y)) \exp\{(-\lambda_1 + \lambda_1 z) y\}. \end{aligned} \quad (4.34)$$

- l'intégration des deux parties de l'équation (4.30) donne

$$\begin{aligned} \int_0^y \frac{dP_5(z, x_2, y)}{P_5(z, x_2, y)} &= \int_0^y (-\lambda_1 - \gamma_2(y) + \lambda_1 z) dy; \\ \int_0^y \frac{dP_5(z, x_2, y)}{P_5(z, x_2, y)} &= \int_0^y \frac{-G_2'(y)}{1 - G_2(y)} dy + \int_0^y (-\lambda_1 + \lambda_1 z) dy; \\ \ln(P_5(z, x_2, y)) - \ln(P_5(z, x_2, 0)) &= [\ln(1 - G_2(y)) - \ln(1 - G_2(0))] + (-\lambda_1 + \lambda_1 z)y; \\ \ln\left(\frac{P_5(z, x_2, y)}{P_5(z, x_2, 0)}\right) &= \ln(1 - G_2(y)) + (-\lambda_1 + \lambda_1 z)y, ((G_2(0) = 0, (\text{par convention}) \Rightarrow (\ln(1 - G_2(0)) = 0)); \end{aligned}$$

D'où

$$\begin{aligned} P_5(z, x_2, y) &= P_5(z, x_2, 0) (1 - G_2(y)) \exp\{(-\lambda_1 + \lambda_1 z) y\}; \\ &= \alpha_2 P_2(z, x_2) (1 - G_2(y)) \exp\{(-\lambda_1 + \lambda_1 z) y\}. \end{aligned} \quad (4.35)$$

- l'intégration des deux parties de l'équation (4.32) donne

$$\begin{aligned} \int_0^y \frac{dP_6(z, x_1, x_2, y)}{P_6(z, x_1, x_2, y)} &= \int_0^y (-\lambda_1 - \gamma_1(y) + \lambda_1 z) dy; \\ \int_0^y \frac{dP_6(z, x_1, x_2, y)}{P_6(z, x_1, x_2, y)} &= \int_0^y \frac{-G_1'(y)}{1 - G_1(y)} dy + \int_0^y (-\lambda_1 + \lambda_1 z) dy; \\ \ln(P_6(z, x_1, x_2, y)) - \ln(P_6(z, x_1, x_2, 0)) &= [\ln(1 - G_1(y)) - \ln(1 - G_1(0))] + (-\lambda_1 + \lambda_1 z)y; \\ \ln\left(\frac{P_6(z, x_1, x_2, y)}{P_6(z, x_1, x_2, 0)}\right) &= \ln(1 - G_1(y)) + (-\lambda_1 + \lambda_1 z)y, ((G_1(0) = 0, (\text{par convention}) \Rightarrow (\ln(1 - G_1(0)) = 0)); \end{aligned}$$

D'où

$$\begin{aligned} P_6(z, x_1, x_2, y) &= P_6(z, x_1, x_2, 0) (1 - G_1(y)) \exp\{(-\lambda_1 + \lambda_1 z) y\}; \\ &= \alpha_1 P_3(z, x_1, x_2) (1 - G_1(y)) \exp\{(-\lambda_1 + \lambda_1 z) y\}. \end{aligned} \quad (4.36)$$

- En remplaçant $P_4(z, x_1, y)$ par son expression donnée par (4.34) dans l'équation (4.22), on obtient

$$\begin{aligned} \frac{\partial P_1(z, x_1)}{\partial x_1} &= -(\lambda_1 + b_1(x_1) + \alpha_1 - \lambda_1 z)P_1(z, x_1) + \int_0^\infty P_4(z, x_1, 0) (1 - G_1(y)) \exp\{(-\lambda_1 + \lambda_1 z) y\} \gamma_1(y) dy; \\ &= -(\lambda_1 + b_1(x_1) + \alpha_1 - \lambda_1 z)P_1(z, x_1) + P_4(z, x_1, 0) \int_0^\infty \exp\{(-\lambda_1 + \lambda_1 z) y\} dG_1(y); \\ &= -(\lambda_1 + b_1(x_1) + \alpha_1 - \lambda_1 z)P_1(z, x_1) + \alpha_1 P_1(z, x_1) A_1(z); \end{aligned}$$

$$\frac{\partial P_1(z, x_1)}{\partial x_1} = (-\lambda_1 + \lambda_1 z - \alpha_1 - b_1(x_1) + \alpha_1 A_1(z)) P_1(z, x_1);$$

$$\frac{dP_1(z, x_1)}{P_1(z, x_1)} = (-\alpha_1 - \lambda_1 + \lambda_1 z) - b_1(x_1) + \alpha_1 A_1(z) \quad (4.37)$$

- L'intégration des deux parties de l'équation (4.37) donne

$$\begin{aligned} \int_0^{x_1} \frac{dP_1(z, x_1)}{P_1(z, x_1)} &= \int_0^{x_1} ((-\alpha_1 - \lambda_1 + \lambda_1 z) - b_1(x_1) + \alpha_1 A_1(z)) dx_1; \\ [\ln(P_1(z, x_1)) - \ln(P_1(z, 0))] &= \ln(1 - B_1(x_1)) + (-\alpha_1 - \lambda_1 + \lambda_1 z)x_1 + \alpha_1 x_1 A_1(z); \\ \ln\left(\frac{P_1(z, x_1)}{P_1(z, 0)}\right) &= \ln(1 - B_1(x_1)) + (-\alpha_1 - \lambda_1 + \lambda_1 z)x_1 + \alpha_1 x_1 A_1(z); \end{aligned}$$

D'où

$$P_1(z, x_1) = P_1(z, 0) (1 - B_1(x_1)) \exp\{(-\lambda_1 + \lambda_1 z - \alpha_1 + \alpha_1 A_1(z)) x_1\}. \quad (4.38)$$

- En remplaçant $P_6(z, x_1, x_2, y)$ par son expression donnée par (4.36) dans l'équation (4.26), on obtient

$$\begin{aligned} \frac{\partial P_3(z, x_1, x_2)}{\partial x_1} &= -(\lambda_1 + b_1(x_1) + \alpha_1 - \lambda_1 z) P_3(z, x_1, x_2) \\ &+ \int_0^\infty P_6(z, x_1, x_2, 0) (1 - G_1(y)) \exp\{(-\lambda_1 + \lambda_1 z) y\} \gamma_1(y) dy; \\ &= -(\lambda_1 + b_1(x_1) + \alpha_1 - \lambda_1 z) P_3(z, x_1, x_2) \\ &+ \int_0^\infty \alpha_1 P_3(z, x_1, x_2) (1 - G_1(y)) \exp\{(-\lambda_1 + \lambda_1 z) y\} \gamma_1(y) dy; \\ &= -(\lambda_1 + b_1(x_1) + \alpha_1 - \lambda_1 z) P_3(z, x_1, x_2) + \alpha_1 P_3(z, x_1, x_2) \int_0^\infty \exp\{(-\lambda_1 + \lambda_1 z) y\} dG_1(y); \end{aligned}$$

$$\frac{\partial P_3(z, x_1, x_2)}{\partial x_1} = (-\lambda_1 - b_1(x_1) - \alpha_1 + \lambda_1 z) P_3(z, x_1, x_2) + \alpha_1 P_3(z, x_1, x_2) A_1(z); \quad (4.39)$$

- L'intégration des deux parties de l'équation (4.39) donne

$$\begin{aligned} \int_0^{x_1} \frac{dP_3(z, x_1, x_2)}{P_3(z, x_1, x_2)} &= \int_0^{x_1} ((-\alpha_1 - \lambda_1 + \lambda_1 z) - b_1(x_1) + \alpha_1 A_1(z)) dx_1; \\ \ln(P_3(z, x_1, x_2)) - \ln(P_3(z, 0, x_2)) &= \ln(1 - B_1(x_1)) + (-\alpha_1 - \lambda_1 + \lambda_1 z) x_1 + \alpha_1 x_1 A_1(z); \\ \ln\left(\frac{P_3(z, x_1, x_2)}{P_3(z, 0, x_2)}\right) &= \ln(1 - B_1(x_1)) + (-\alpha_1 - \lambda_1 + \lambda_1 z) x_1 + \alpha_1 x_1 A_1(z); \end{aligned} \quad \text{dotted}$$

D'où

$$P_3(z, x_1, x_2) = P_3(z, 0, x_2) (1 - B_1(x_1)) \exp\{(-\lambda_1 + \lambda_1 z - \alpha_1 + \alpha_1 A_1(z)) x_1\}. \quad (4.40)$$

- En remplaçant $P_5(z, x_2, y)$ par son expression donnée par (4.35) dans l'équation (4.24), on obtient

$$\begin{aligned}
\frac{\partial P_2(z, x_2)}{\partial x_2} &= -(\lambda_1 + b_2(x_2) + \alpha_2)P_2(z, x_2) - \theta z \frac{\partial P_2(z, x_2)}{\partial z} + \int_0^\infty P_3(z, x_1, x_2) b_1(x_1) dx_1 \\
&+ \int_0^\infty P_5(z, x_2, 0) (1 - G_2(y)) \exp\{(-\lambda_1 + \lambda_1 z) y\} \Upsilon_2(y) dy; \\
&= -(\lambda_1 + b_2(x_2) + \alpha_2)P_2(z, x_2) - \theta z \frac{\partial P_2(z, x_2)}{\partial z} + \int_0^\infty P_3(z, x_1, x_2) b_1(x_1) dx_1 \\
&+ \int_0^\infty \alpha_2 P_2(z, x_2) (1 - G_2(y)) \exp\{(-\lambda_1 + \lambda_1 z) y\} \Upsilon_2(y) dy; \\
&= -(\lambda_1 + b_2(x_2) + \alpha_2)P_2(z, x_2) - \theta z \frac{\partial P_2(z, x_2)}{\partial z} + \int_0^\infty P_3(z, x_1, x_2) b_1(x_1) dx_1 \\
&+ \alpha_2 P_2(z, x_2) \int_0^\infty \exp\{(-\lambda_1 + \lambda_1 z) y\} \Upsilon_2(z);
\end{aligned}$$

$$\begin{aligned}
\frac{\partial P_2(z, x_2)}{\partial x_2} &= -(\lambda_1 + b_2(x_2) + \alpha_2)P_2(z, x_2) - \theta z \frac{\partial P_2(z, x_2)}{\partial z} \\
&+ \int_0^\infty P_3(z, x_1, x_2) b_1(x_1) dx_1 + \alpha_2 P_2(z, x_2) A_2(z);
\end{aligned} \tag{4.41}$$

En remplaçant $P_3(z, x_1, x_2)$ par son expression donnée par (4.40) dans l'équation (4.41), on obtient :

$$\begin{aligned}
\frac{\partial P_2(z, x_2)}{\partial x_2} &= -(\lambda_1 + b_2(x_2) + \alpha_2)P_2(z, x_2) - \theta z \frac{\partial P_2(z, x_2)}{\partial z} + \alpha_2 P_2(z, x_2) A_2(z) \\
&+ \int_0^\infty P_3(z, 0, x_2) (1 - B_1(x_1)) \exp\{(-\lambda_1 + \lambda_1 z - \alpha_1 + \alpha_1 A_1(z)) x_1\} b_1(x_1) dx_1; \\
&= -(\lambda_1 + b_2(x_2) + \alpha_2)P_2(z, x_2) - \theta z \frac{\partial P_2(z, x_2)}{\partial z} + \alpha_2 P_2(z, x_2) A_2(z) \\
&+ P_3(z, 0, x_2) \int_0^\infty \exp\{(-\lambda_1 + \lambda_1 z - \alpha_1 + \alpha_1 A_1(z)) x_1\} dB_1(x_1);
\end{aligned}$$

$$\begin{aligned}
\frac{\partial P_2(z, x_2)}{\partial x_2} &= -(\lambda_1 + b_2(x_2) + \alpha_2)P_2(z, x_2) - \theta z \frac{\partial P_2(z, x_2)}{\partial z} + \alpha_2 P_2(z, x_2) A_2(z) \\
&+ P_3(z, 0, x_2) K_1(z, A_1(z))
\end{aligned} \tag{4.42}$$

En remplaçant $P_3(z, 0, x_2)$ par son expression donnée par (4.27) dans l'équation (4.42), on obtient :

$$\begin{aligned}
\frac{\partial P_2(z, x_2)}{\partial x_2} &= -(\lambda_1 + b_2(x_2) + \alpha_2)P_2(z, x_2) - \theta z \frac{\partial P_2(z, x_2)}{\partial z} + \alpha_2 P_2(z, x_2) A_2(z) \\
&+ \left(\lambda_1 P_2(z, x_2) + \theta \frac{\partial P_2(z, y)}{\partial z} \right) K_1(z, A_1(z)).
\end{aligned}$$

$$\begin{aligned} \frac{\partial P_2(z, x_2)}{\partial x_2} &= (-\lambda_1 - b_2(x_2) - \alpha_2 + \alpha_2 A_2(z) + \lambda_1 K_1(z, A_1(z))) \\ &+ \theta(-z + K_1(z, A_1(z))) \frac{\partial P_2(z, y)}{\partial z}. \end{aligned} \quad (4.43)$$

L'équation (4.43) admet comme solution

$$\begin{aligned} P_2(z, x_2) &= \text{const.} \exp \left\{ \int_0^{x_2} b_2(x_2) dx_2 \right\} \cdot \exp \left\{ \int_1^z \frac{\lambda_1 - \lambda_1 K_1(z, A_1(z)) + \alpha_2 - \alpha_2 A_2(z)}{\theta(K_1(z, A_1(z)) - z)} dz \right\}; \\ &= \text{const.} (1 - B_2(x_2)) \cdot \exp \left\{ \int_1^z \frac{\lambda_1 - \lambda_1 K_1(z, A_1(z)) + \alpha_2 - \alpha_2 A_2(z)}{\theta(K_1(z, A_1(z)) - z)} dz \right\}. \end{aligned} \quad (4.44)$$

En prenant $x_2 = 0$ dans l'équation (4.44), on obtient une deuxième expression de $P_2(z, 0)$, la première est donnée par l'équation (4.25)

$$P_2(z, 0) = \text{const.} \exp \left\{ \int_1^z \frac{\lambda_1 - \lambda_1 K_1(z, A_1(z)) + \alpha_2 - \alpha_2 A_2(z)}{\theta(K_1(z, A_1(z)) - z)} dz \right\}. \quad (4.45)$$

Pour avoir l'expression de $P_2(z)$, il suffit d'intégrer l'équation (4.44) par rapport à x_2

$$\begin{aligned} P_2(z) &= \int_0^\infty P_2(z, x_2) dx_2 \\ &= \text{const.} \left(\int_0^\infty (1 - B_2(x_2)) dx_2 \right) \exp \left\{ \int_1^z \frac{\lambda_1 - \lambda_1 K_1(z, A_1(z)) + \alpha_2 - \alpha_2 A_2(z)}{\theta(K_1(z, A_1(z)) - z)} dz \right\}; \end{aligned} \quad (4.46)$$

À l'aide de l'équation (4.25) on obtient

$$P_0(z) = \frac{\text{const}}{\lambda_2} \exp \left\{ \int_1^z \frac{\lambda_1 - \lambda_1 K_1(z, A_1(z)) + \alpha_2 - \alpha_2 A_2(z)}{\theta(K_1(z, A_1(z)) - z)} dz \right\}. \quad (4.47)$$

À présent, en remplaçant $P_0(z)$ par son expression donnée par (4.47) dans l'équation (4.23), on obtient

$$\begin{aligned} P_1(z, 0) &= \frac{\text{const}}{\lambda_2} \cdot \left[\frac{\lambda_1 - \lambda_1 z + \alpha_2 - \alpha_2 A_2(z)}{K_1(z, A_1(z)) - z} \right] \\ &\times \exp \left\{ \int_1^z \frac{\lambda_1 - \lambda_1 K_1(z, A_1(z)) + \alpha_2 - \alpha_2 A_2(z)}{\theta(K_1(z, A_1(z)) - z)} dz \right\}. \end{aligned} \quad (4.48)$$

En remplaçant $P_1(z, 0)$ par son expression donnée par (4.48) dans l'équation (4.38), on obtient

$$\begin{aligned} P_1(z, x_1) &= \frac{\text{const}}{\lambda_2} \left[\frac{\lambda_1 - \lambda_1 z + \alpha_2 - \alpha_2 A_2(z)}{K_1(z, A_1(z)) - z} \right] (1 - B_1(x_1)) \exp \{ (\lambda_1 z - \lambda_1 + \alpha_1 A_1(z) - \alpha_1) x_1 \} \\ &\times \exp \left\{ \int_1^z \frac{\lambda_1 - \lambda_1 K_1(z, A_1(z)) + \alpha_2 - \alpha_2 A_2(z)}{\theta(K_1(z, A_1(z)) - z)} dz \right\}; \end{aligned} \quad (4.49)$$

Pour avoir l'expression de $P_1(z)$, il suffit d'intégrer l'équation (4.49) par rapport à x_1

$$\begin{aligned}
P_1(z) &= \int_0^\infty P_1(z, x_1) dx_1 \\
&= \frac{const}{\lambda_2} \left[\frac{\lambda_1 - \lambda_1 z + \alpha_2 - \alpha_2 A_2(z)}{K_1(z, A_1(z)) - z} \right] \int_0^\infty (1 - B_1(x_1)) \exp\{(\lambda_1 z - \lambda_1 + \alpha_1 A_1(z) - \alpha_1) x_1\} dx_1 \\
&\quad \times \exp \left\{ \int_1^z \frac{\lambda_1 - \lambda_1 K_1(z, A_1(z)) + \alpha_2 - \alpha_2 A_2(z)}{\theta(K_1(z, A_1(z)) - z)} dz \right\};
\end{aligned}$$

Nous savons que :

$$\int_0^\infty (1 - B_1(x_1)) \exp\{(\lambda_1 z - \lambda_1 + \alpha_1 A_1(z) - \alpha_1) x_1\} dx_1 = \frac{1 - K_1(z, A_1(z))}{(\lambda_1 - \lambda_1 z + \alpha_1 - \alpha_1 A_1(z))};$$

Alors

$$\begin{aligned}
P_1(z) &= \frac{const}{\lambda_2} \left[\frac{\lambda_1 - \lambda_1 z + \alpha_2 - \alpha_2 A_2(z)}{\lambda_1 - \lambda_1 z + \alpha_1 - \alpha_1 A_1(z)} \times \frac{1 - K_1(z, A_1(z))}{K_1(z, A_1(z)) - z} \right] \\
&\quad \times \exp \left\{ \int_1^z \frac{\lambda_1 - \lambda_1 K_1(z, A_1(z)) + \alpha_2 - \alpha_2 A_2(z)}{\theta(K_1(z, A_1(z)) - z)} dz \right\}.
\end{aligned} \tag{4.50}$$

En remplaçant $P_2(z, x_2)$ par son expression donnée par (4.44) dans l'équation (4.27), on obtient :

$$\begin{aligned}
P_3(z, 0, x_2) &= \lambda_1 const (1 - B_2(x_2)) \exp \left\{ \int_1^z \frac{\lambda_1 - \lambda_1 K_1(z, A_1(z)) + \alpha_2 - \alpha_2 A_2(z)}{\theta(K_1(z, A_1(z)) - z)} dz \right\} \\
&\quad + \theta \left[(1 - B_2(x_2)) \times \frac{\lambda_1 - \lambda_1 K_1(z, A_1(z)) + \alpha_2 - \alpha_2 A_2(z)}{\theta(K_1(z, A_1(z)) - z)} \right. \\
&\quad \left. \times \exp \left\{ \int_1^z \frac{\lambda_1 - \lambda_1 K_1(z, A_1(z)) + \alpha_2 - \alpha_2 A_2(z)}{\theta(K_1(z, A_1(z)) - z)} dz \right\} \right] \\
&= \lambda_1 const (1 - B_2(x_2)) \left[\frac{\lambda_1 - \lambda_1 z + \alpha_2 - \alpha_2 A_2(z)}{K_1(z, A_1(z)) - z} \right] \\
&\quad \exp \left\{ \int_1^z \frac{\lambda_1 - \lambda_1 K_1(z, A_1(z)) + \alpha_2 - \alpha_2 A_2(z)}{\theta(K_1(z, A_1(z)) - z)} dz \right\};
\end{aligned}$$

Donc

$$\begin{aligned}
P_3(z, x_1, x_2) &= const \left[\frac{\lambda_1 - \lambda_1 z + \alpha_2 - \alpha_2 A_2(z)}{K_1(z, A_1(z)) - z} \right] (1 - B_2(x_2)) (1 - B_1(x_1)) \\
&\quad \times \exp\{(\lambda_1 z - \lambda_1 + \alpha_1 A_1(z) - \alpha_1) x_1\} \\
&\quad \times \exp \left\{ \int_1^z \frac{\lambda_1 - \lambda_1 K_1(z, A_1(z)) + \alpha_2 - \alpha_2 A_2(z)}{\theta(K_1(z, A_1(z)) - z)} dz \right\};
\end{aligned} \tag{4.51}$$

Une intégration double de (4.51) par rapport à x_1 et x_2 donne l'expression de $P_3(z)$

$$\begin{aligned}
P_3(z) &= \int_0^\infty \int_0^\infty P_3(z, x_1, x_2) dx_1 dx_2 \\
&= \text{const} \left(\int_0^\infty (1 - B_2(x_2)) dx_2 \right) \left[\frac{\lambda_1 - \lambda_1 z + \alpha_2 - \alpha_2 A_2(z)}{\lambda_1 - \lambda_1 z + \alpha_1 - \alpha_1 A_1(z)} \times \frac{1 - K_1(z, A_1(z))}{K_1(z, A_1(z)) - z} \right] \\
&\quad \times \exp \left\{ \int_1^z \frac{\lambda_1 - \lambda_1 K_1(z, A_1(z)) + \alpha_2 - \alpha_2 A_2(z)}{\theta(K_1(z, A_1(z)) - z)} dz \right\}.
\end{aligned} \tag{4.52}$$

Pour avoir l'expression de $P_4(z)$, il suffit de calculer l'intégrale double de l'équation (4.34) par rapport à x_1 et y

$$\begin{aligned}
P_4(z) &= \int_0^\infty \int_0^\infty P_4(z, x_1, y) dx_1 dy \\
&= \alpha_1 P_1(z) \frac{1 - A_1(z)}{\lambda_1 - \lambda_1 z} \\
&= \alpha_1 \frac{1 - A_1(z)}{\lambda_1 - \lambda_1 z} \times \frac{\text{const}}{\lambda_2} \left[\frac{\lambda_1 - \lambda_1 z + \alpha_2 - \alpha_2 A_2(z)}{\lambda_1 - \lambda_1 z + \alpha_1 - \alpha_1 A_1(z)} \times \frac{1 - K_1(z, A_1(z))}{K_1(z, A_1(z)) - z} \right] \\
&\quad \times \exp \left\{ \int_1^z \frac{\lambda_1 - \lambda_1 K_1(z, A_1(z)) + \alpha_2 - \alpha_2 A_2(z)}{\theta(K_1(z, A_1(z)) - z)} dz \right\}.
\end{aligned} \tag{4.53}$$

Pour avoir l'expression de $P_5(z)$, il suffit de calculer l'intégrale double de l'équation (4.35) par rapport à x_2 et y

$$\begin{aligned}
P_5(z) &= \int_0^\infty \int_0^\infty P_5(z, x_2, y) dx_2 dy \\
&= \alpha_2 P_2(z) \frac{1 - A_2(z)}{\lambda_1 - \lambda_1 z}; \\
&= \alpha_2 \frac{1 - A_2(z)}{\lambda_1 - \lambda_1 z} \times \text{const} \left(\int_0^\infty (1 - B_2(x_2)) dx_2 \right) \\
&\quad \times \exp \left\{ \int_1^z \frac{\lambda_1 - \lambda_1 K_1(z, A_1(z)) + \alpha_2 - \alpha_2 A_2(z)}{\theta(K_1(z, A_1(z)) - z)} dz \right\}.
\end{aligned} \tag{4.54}$$

Pour avoir l'expression de $P_6(z)$, il suffit de calculer l'intégrale triple de l'équation (4.36) par rapport à x_1 , x_2 et y

$$\begin{aligned}
P_6(z) &= \int_0^\infty \int_0^\infty \int_0^\infty P_6(z, x_1, x_2, y) dx_1 dx_2 dy \\
&= \alpha_1 P_3(z) \frac{1 - A_1(z)}{\lambda_1 - \lambda_1 z} \\
&= \text{const} \cdot \alpha_1 \frac{1 - A_1(z)}{\lambda_1 - \lambda_1 z} \left(\int_0^\infty (1 - B_2(x_2)) dx_2 \right) \left[\frac{\lambda_1 - \lambda_1 z + \alpha_2 - \alpha_2 A_2(z)}{\lambda_1 - \lambda_1 z + \alpha_1 - \alpha_1 A_1(z)} \times \frac{1 - K_1(z, A_1(z))}{K_1(z, A_1(z)) - z} \right] \\
&\quad \times \exp \left\{ \int_1^z \frac{\lambda_1 - \lambda_1 K_1(z, A_1(z)) + \alpha_2 - \alpha_2 A_2(z)}{\theta(K_1(z, A_1(z)) - z)} dz \right\};
\end{aligned} \tag{4.55}$$

Pour rendre les expressions de $P_0(z)$, $P_1(z)$, $P_2(z)$, $P_3(z)$, $P_4(z)$, $P_5(z)$ et $P_6(z)$ définies et continues au point $z = 1$, il suffit prolonger la fonction $\left[\frac{\lambda_1 - \lambda_1 z + \alpha_2 - \alpha_2 A_2(z)}{\lambda_1 - \lambda_1 z + \alpha_1 - \alpha_1 A_1(z)} \times \frac{1 - K_1(z, A_1(z))}{K_1(z, A_1(z)) - z} \right]$ par continuité au point $z = 1$. Alors, on doit calculer la limite suivante

$$\begin{aligned}
\lim_{z \rightarrow 1} \left[\frac{\lambda_1 - \lambda_1 z + \alpha_2 - \alpha_2 A_2(z)}{\lambda_1 - \lambda_1 z + \alpha_1 - \alpha_1 A_1(z)} \times \frac{1 - K_1(z, A_1(z))}{K_1(z, A_1(z)) - z} \right] &= \lim_{z \rightarrow 1} \frac{1 - K_1(z, A_1(z))}{K_1(z, A_1(z)) - z} \times \lim_{z \rightarrow 1} \frac{\lambda_1 - \lambda_1 z + \alpha_2 - \alpha_2 A_2(z)}{\lambda_1 - \lambda_1 z + \alpha_1 - \alpha_1 A_1(z)} \\
&= \lim_{z \rightarrow 1} \frac{1 - \tilde{B}(-\lambda_1 + \lambda_1 z - \alpha_1 + \alpha_1 A_1(z))}{\tilde{B}(-\lambda_1 + \lambda_1 z - \alpha_1 + \alpha_1 A_1(z)) - z} \\
&\quad \times \lim_{z \rightarrow 1} \frac{\lambda_1 - \lambda_1 z + \alpha_2 - \alpha_2 \tilde{G}_2(\lambda_1 - \lambda_1 z)}{\lambda_1 - \lambda_1 z + \alpha_1 - \alpha_1 \tilde{G}_1(\lambda_1 - \lambda_1 z)} \\
&= \lim_{z \rightarrow 1} \frac{\left(-\lambda_1 - \alpha_1 \lambda_1 \tilde{G}_1^{(1)}(-\lambda_1 + \lambda_1 z) \right) \tilde{B}_1^{(1)}(-\lambda_1 + \lambda_1 z - \alpha_1 + \alpha_1 A_1(z))}{\left(\lambda_1 + \alpha_1 \lambda_1 \tilde{G}_1^{(1)}(-\lambda_1 + \lambda_1 z) \right) \tilde{B}_1^{(1)}(-\lambda_1 + \lambda_1 z - \alpha_1 + \alpha_1 A_1(z)) - 1} \\
&\quad \times \lim_{z \rightarrow 1} \frac{-\lambda_1 - \lambda_1 \alpha_2 \tilde{G}_2^{(1)}(\lambda_1 - \lambda_1 z)}{-\lambda_1 - \lambda_1 \alpha_2 \tilde{G}_1^{(1)}(\lambda_1 - \lambda_1 z)} \\
&= \frac{(-\lambda_1 - \alpha_1 \lambda_1 \gamma_{11}) \beta_{11}}{(\lambda_1 + \alpha_1 \lambda_1 \gamma_{11}) \beta_{11}} \times \frac{\lambda_1 + \alpha_2 \lambda_1 \gamma_{21}}{\lambda_1 + \alpha_1 \lambda_1 \gamma_{11}} \\
&= \frac{(1 + \alpha_1 \gamma_{11}) \lambda_1 \beta_{11}}{1 - (1 + \alpha_1 \gamma_{11}) \lambda_1 \beta_{11}} \times \frac{1 + \alpha_2 \gamma_{21}}{1 + \alpha_1 \gamma_{11}} \\
&= \frac{\lambda_1 \beta_{11} (1 + \alpha_2 \gamma_{21})}{1 - \rho_1}.
\end{aligned}$$

Par suite, les expressions (4.46), (4.47), (4.46), (4.50), (4.52), (4.53), (4.54) et (4.55) deviennent au point $z = 1$

$$P_0(1) = \frac{const}{\lambda_2},$$

$$P_1(1) = \frac{const. \lambda_1 \beta_{11} (1 + \alpha_2 \gamma_{21})}{\lambda_2 (1 - \rho_1)},$$

$$P_2(1) = const \int_0^\infty (1 - B_2(x_2)) dx_2,$$

$$P_3(1) = \frac{const. \lambda_1 \beta_{11} (1 + \alpha_2 \gamma_{21})}{1 - \rho_1} \int_0^\infty (1 - B_2(x_2)) dx_2,$$

$$P_4(1) = \alpha_1 \gamma_{11} \frac{const. \lambda_1 \beta_{11} (1 + \alpha_2 \gamma_{21})}{\lambda_2 (1 - \rho_1)},$$

$$P_5(1) = \alpha_2 \gamma_{21} const \int_0^\infty (1 - B_2(x_2)) dx_2,$$

$$P_6(1) = \alpha_1 \gamma_{11} \frac{\text{const.} \lambda_1 \beta_{11} (1 + \alpha_2 \gamma_{21})}{1 - \rho_1} \int_0^\infty (1 - B_2(x_2)) dx_2.$$

Posons $Const = H$, à l'aide de l'équation de normalisation

$$P_0(1) + P_1(1) + P_2(2) + P_3(1) + P_4(1) + P_5(1) + P_6(1) = 1,$$

on trouve que

$$H = \frac{\lambda_2 (1 - \rho_1)}{1 + \lambda_2 \int_0^\infty (1 - B_2(x_2)) dx_2 + \alpha_2 \gamma_{21} (\rho_1 + \lambda_2 \int_0^\infty (1 - B_2(x_2)) dx_2)}.$$

A présent, nous pouvons trouver la fonction génératrice du nombre de clients en orbite

$$\begin{aligned} Q_o(z) &= P_0(z) + P_1(z) + P_2(z) + P_3(z) + P_4(z) + P_5(z) + P_6(z) \\ &= \left(\frac{(1-z)(1 + \lambda_2 \int_0^\infty (1 - B_2(x_2)) dx_2)}{\lambda_2 (K_1(z, A_1(z)) - z)} + \frac{(\alpha_2 - \alpha_2 A_2(z)) (\int_0^\infty (1 - B_2(x_2)) dx_2)}{\lambda_1 (K_1(z, A_1(z)) - z)} \right. \\ &\quad \left. + \frac{(\alpha_2 - \alpha_2 A_2(z)) (1 - K_1(z, A_1(z)))}{\lambda_2 (\lambda_1 - \lambda_1 z) (K_1(z, A_1(z)) - z)} \right) \text{Hexp} \left\{ \int_1^z \frac{\lambda_1 - \lambda_1 K_1(z, A_1(z)) + \alpha_2 - \alpha_2 A_2(z)}{\theta(K_1(z, A_1(z)) - z)} dz \right\}. \end{aligned}$$

Aussi, la fonction génératrice du nombre de clients prioritaires dans le système

$$\begin{aligned} Q_s(z) &= P_0(z) + zP_1(z) + P_2(z) + zP_3(z) + zP_4(z) + P_5(z) + zP_6(z) \\ &= \left(\frac{K_1(z, A_1(z)) (1-z)(1 + \lambda_2 \int_0^\infty (1 - B_2(x_2)) dx_2)}{\lambda_2 (K_1(z, A_1(z)) - z)} + \frac{K_1(z, A_1(z)) (\alpha_2 - \alpha_2 A_2(z)) (\int_0^\infty (1 - B_2(x_2)) dx_2)}{\lambda_1 (K_1(z, A_1(z)) - z)} \right. \\ &\quad \left. + \frac{(\alpha_2 - \alpha_2 A_2(z)) (1 - K_1(z, A_1(z)))}{\lambda_2 (\lambda_1 - \lambda_1 z) (K_1(z, A_1(z)) - z)} \right) \text{Hexp} \left\{ \int_1^z \frac{\lambda_1 - \lambda_1 K_1(z, A_1(z)) + \alpha_2 - \alpha_2 A_2(z)}{\theta(K_1(z, A_1(z)) - z)} dz \right\}. \end{aligned}$$

4.3.4 Mesures de performance

À l'aide des fonctions génératrices obtenues, nous pouvons calculer plusieurs mesures de performance du système, telles que :

- La probabilité p_1 , un client prioritaire est servi et il n'y a aucun client non prioritaire dont le service a été interrompu dans la station-service

$$p_1 = P_1(1) = \frac{H \lambda_1 \beta_{11} (1 + \alpha_2 \gamma_{21})}{\lambda_2 (1 - \rho_1)}.$$

- La probabilité p_2 , qu'un client non prioritaire est servi

$$p_2 = P_2(1) = H \int_0^\infty (1 - B_2(x_2)) dx_2.$$

- La probabilité p_3 , un client prioritaire est servi et il y a un client non prioritaire dont le service a été interrompu dans la station-service

$$p_3 = P_3(1) = \frac{H \lambda_1 \beta_{11} (1 + \alpha_2 \gamma_{21})}{1 - \rho_1} \int_0^\infty (1 - B_2(x_2)) dx_2.$$

- La probabilité p_4 , que le serveur tombe en panne lorsqu'un client prioritaire est servi et il n'y a aucun client non prioritaire dont le service a été interrompu dans la station-service

$$p_4 = P_4(1) = \alpha_1 \gamma_{11} \frac{H \lambda_1 \beta_{11} (1 + \alpha_2 \gamma_{21})}{\lambda_2 (1 - \rho_1)}.$$

- La probabilité p_5 , le serveur tombe en panne lorsqu'un client non prioritaire est servi

$$p_5 = P_5(1) = \alpha_2 \gamma_{21} H \int_0^{\infty} (1 - B_2(x_2)) dx_2.$$

- La probabilité p_6 , le serveur tombe en panne lorsqu'un client prioritaire est servi et il y a un client non prioritaire dont le service a été interrompu dans la station-service

$$p_6 = P_6(1) = \alpha_1 \gamma_{11} \frac{H \lambda_1 \beta_{11} (1 + \alpha_2 \gamma_{21})}{1 - \rho_1} \int_0^{\infty} (1 - B_2(x_2)) dx_2.$$

4.3.5 Indices de fiabilité du serveur

Dans cette section, nous fournissons la disponibilité et la fréquence de panne du serveur.

Soit $A(t)$ la probabilité que le serveur soit disponible à l'instant t et définissons la disponibilité du serveur comme

$$A = \lim_{t \rightarrow \infty} A(t).$$

A est obtenue en considérant l'équation suivante

$$\begin{aligned} A &= \sum_{i=0}^{\infty} P_{0i} + \sum_{i=0}^{\infty} \int_0^{\infty} P_{1i}(x_1) dx_1 + \sum_{i=0}^{\infty} \int_0^{\infty} P_{2i}(x_2) dx_2 + \sum_{i=0}^{\infty} \int_0^{\infty} \int_0^{\infty} P_{3i}(x_1, x_2) dx_1 dx_2 \\ &= \lim_{z \rightarrow 1} \left[P_0(z) + \int_0^{\infty} P_1(z, x_1) dx_1 + \int_0^{\infty} P_2(z, x_2) dx_2 + \int_0^{\infty} \int_0^{\infty} P_3(z, x_1, x_2) dx_1 dx_2 \right]. \end{aligned}$$

Alors

$$A = \frac{(1 + \lambda_1 \beta_{11} (\alpha_2 \gamma_{21} - \alpha_1 \gamma_{11})) (1 + \lambda_2 \int_0^{\infty} (1 - B_2(x_2)) dx_2)}{(1 + \lambda_2 \int_0^{\infty} (1 - B_2(x_2)) dx_2) + \alpha_2 \gamma_{21} (\rho_1 + \lambda_2 \int_0^{\infty} (1 - B_2(x_2)) dx_2)}.$$

Soit F la fréquence de panne du serveur. Étant donné que la fréquence de panne à l'état stationnaire du service est

$$\begin{aligned} F &= \sum_{i=0}^{\infty} \int_0^{\infty} \alpha_1 P_{1i}(x_1) dx_1 + \sum_{i=0}^{\infty} \int_0^{\infty} \alpha_2 P_{2i}(x_2) dx_2 + \sum_{i=0}^{\infty} \int_0^{\infty} \int_0^{\infty} \alpha_1 P_{3i}(x_1, x_2) dx_1 dx_2 \\ &= \lim_{z \rightarrow 1} \left[\int_0^{\infty} \alpha_1 P_1(z, x_1) dx_1 + \int_0^{\infty} \alpha_2 P_2(z, x_2) dx_2 + \int_0^{\infty} \int_0^{\infty} \alpha_1 P_3(z, x_1, x_2) dx_1 dx_2 \right]. \end{aligned}$$

On obtient

$$F = \frac{\alpha_1 \lambda_1 \beta_{11} (1 + \alpha_2 \gamma_{21}) (1 + \lambda_2 \int_0^{\infty} (1 - B_2(x_2)) dx_2)}{(1 + \lambda_2 \int_0^{\infty} (1 - B_2(x_2)) dx_2) + \alpha_2 \gamma_{21} (\rho_1 + \lambda_2 \int_0^{\infty} (1 - B_2(x_2)) dx_2)} + \alpha_2 H \int_0^{\infty} (1 - B_2(x_2)) dx_2.$$

4.4 Illustrations numériques

Dans cette section, nous fournissons des illustrations numériques qui sont réalisées en développant un programme dans le logiciel MATLAB pour valider la traçabilité et l'applicabilité de l'approche de calcul proposée pour le modèle avec rappels concerné. Nous étudions l'effet des pannes sur la sensibilité du système et sur la disponibilité et la fréquence de panne du serveur. Les valeurs des paramètres par défaut sont choisies comme suit : $\lambda_1 = 0.3$, $\lambda_2 = 0.2$ et $\alpha_1 = 0.1$. Les temps de service et les temps de réparation sont supposés distribués exponentiellement avec les moments $\beta_{11} = \beta_{21} = \gamma_{11} = \gamma_{21} = 1$. Il faut noter qu'avec ce choix, la condition d'ergodicité ($\rho_1 < 1$) est respectée. Dans un premier temps, nous analysons l'impact du taux de panne α_2 sur les indices de fiabilité obtenus précédemment. L'impact en question est illustré dans la figure 4.1 en considérant l'ensemble des paramètres par défaut et en variant α_2 de 0 à 2 par incréments de 0,01.

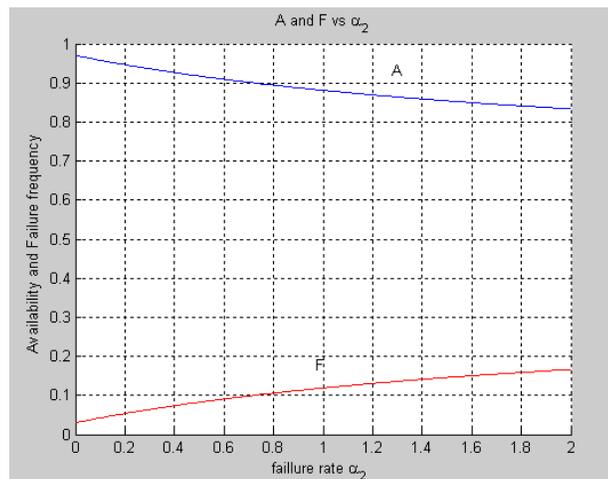


FIGURE 4.1 – Disponibilité et fréquence de panne vs. α_2

Comme prévu (intuitivement), avec l'augmentation du taux de panne de serveur lorsque des clients non prioritaires sont servi α_2 , la probabilité de trouver un serveur en bon état est faible, ce qui augmente la fréquence de panne du serveur.

Maintenant, nous étudions l'influence des taux de pannes α_1 et α_2 sur les mesures de performance comme les probabilités p_4 , p_5 et p_6 . Pour les mêmes valeurs de paramètres par défaut et en variant α_1 et α_2 de 0 à 2 par incréments de 0,01, nous avons obtenu un ensemble de valeurs numériques pour les probabilités mentionnées ci-dessus. L'ensemble en question est présenté dans la figure 4.2.

intuitivement, avec l'augmentation de taux de panne du serveur au moment de servir les clients non prioritaires α_2 , la probabilité que le serveur tombe en panne lorsqu'un client non prioritaire est servi

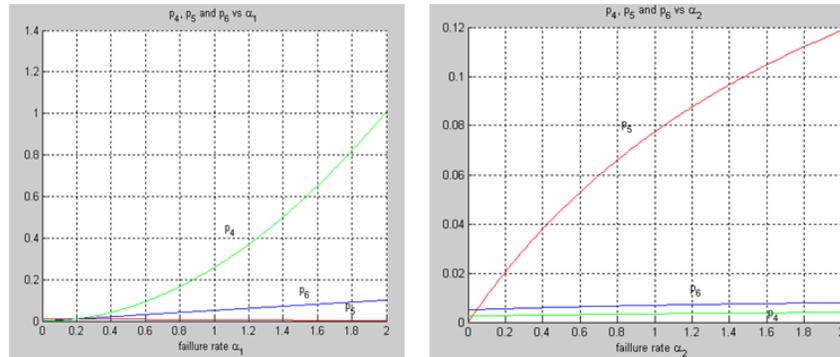


FIGURE 4.2 – Le comportement des probabilités p_4, p_5 et p_6 vs. α_1 and α_2

augmente alors que la probabilité que le serveur tombe en panne lorsque le client prioritaire est servi soit presque nulle et vice versa avec l'augmentation du taux de panne de serveur au moment de servir les clients prioritaires α_1 .

4.5 Conclusion

Dans ce chapitre, nous avons considéré un système d'attente avec rappels, pannes et priorité absolue. La méthode des variables supplémentaires est proposée pour déterminer la condition d'ergodicité et les fonctions génératrices partielles de la distribution stationnaire de l'état du système. Diverses mesures de performance et certains indices de fiabilité sont obtenus en utilisant les probabilités d'état. Ces études théoriques sont soutenues par des illustrations numériques. Le modèle présenté dans ce travail est nouveau et il est réaliste car il peut s'appliquer à diverses situations de congestion rencontrées dans les systèmes de télécommunication, les hôpitaux, les banques, les systèmes de fabrication qui impliquent l'entretien de deux types de clients selon certaines règles de priorité.

CONCLUSION ET PERSPECTIVES

Dans ce travail, nous nous sommes intéressés aux modèles d'attente avec rappels lorsque le service est sujet à des interruptions dues aux pannes aléatoires des serveurs ou à l'arrivée d'un client prioritaire.

Dans un premier temps, nous avons passé en revue certains résultats consacrés aux systèmes d'attente avec rappels. Ainsi, nous avons actualisé la synthèse des résultats connus sur les systèmes d'attente avec rappels à interruptions de service : les modèles avec pannes aléatoires et les modèles avec priorité, où une attention particulière a été accordée au modèle M/G/1 fiable, non fiable et avec priorité. Nous avons constaté que la théorie analytique des systèmes d'attente avec rappels non fiables a une portée limitée en raison de la complexité des résultats connus, particulièrement, lorsque le nombre de serveurs dépasse un ou lorsque le système a une structure plus complexe telle que les systèmes d'attente avec rappels, les pannes et arrivées négatives, les systèmes d'attente avec rappels, pannes et priorité...

Dans un deuxième temps, nous avons réalisé une analyse approximative de la fiabilité du modèle M/M/2 avec rappels et pannes en appliquant « phase merging algorithm ». Nous avons obtenu des approximations de quelques indices de fiabilité et nous avons évalué ces résultats par des illustrations numériques. Enfin, Nous avons réalisé une analyse mathématique du système d'attente M/G/1 avec rappels panne et priorité absolue. A cet effet, la chaîne de Markov induite est décrite, la condition d'ergodicité est établie à l'aide du critère de Foster et condition de Kaplan, la distribution stationnaire de l'état du système est trouvée en appliquant la méthode des variables supplémentaires et les mesures de performance sont obtenues. En outre, nous avons étudié la fiabilité de tel système en calculant certains indices de fiabilité et en présentant des illustrations numériques indiquant l'influence du taux de pannes sur les mesures de performance.

Les résultats obtenus dans cette thèse, permettent d'envisager de nouvelles perspectives de recherche :

1. Extension des résultats obtenus dans le chapitre 3 à des systèmes M/M/C ($C > 2$) qui sont plus complexes.
2. Extension des résultats obtenus dans le chapitre 4 à des systèmes à plusieurs serveurs non fiables avec d'autres politiques de priorité.

3. Application des résultats obtenus pour évaluer les performances des systèmes réels.


```

v1=B1.*k1+o1;
% la disponibilité de serveurs
A=C.*v1
% la fréquence de pannes
B2=(b+l+2*m+2*a);
k2=(2*b*l+2*b*m+m*a);
o2=(a+2*b)*m*b;
v2=B2.*k2+o2;
E=(m*a)./H;
F=E.*v2
% les graphes
figure
hold on
plot(1,A(:),'b')
plot(1,F(:),'r')
xlabel('Taux des arrivées {\lambda} ')
ylabel('Disponibilité et fréquence de panne ')
grid
hold off
H
A
F

```

Disponibilité et fréquence de panne vs. α

```

m=06; l=04; b=0.1;
A=[];
F=[];
C=[];
D=[];
E=[];
H=[];
a=0:0.1:2;
H=m^2*a.^2*l+6*m^2*a.^2*b+2*m^3*a.^2+2*m^2*a.^3+6*b^2*m*a*l+2*m*b^3*l+...
+4*b^2*m^2*l+b^3*l^2+3*b^2*m*l^2+b^2*l^3+2*a*b^2*l^2+4*m^3*a*b+...

```

```

+6*b^2*m^2*a+2*m^2*b^3+2*m^3*b^3+2*m^3*b^2+2*m*a*b*l^2+6*m^2*a*b*l+4*b*m*a.^2*l;
C=b^2./H;
B1=(b+l+m);
k1=(2*m^2+2*m*l+l^2);
o1=2*(l+m)^2*a;
v1=B1.*k1+o1;
% la disponibilité de serveurs
A=C.*v1
% la fréquence de pannes
B2=(b+l+2*m+2*a);
k2=(2*b*l+2*b*m+m*a);
o2=(b*m*(2*b+a));
v2=B2.*k2+o2;
E=(m*a)./H;
F=E.*v2
% les graphes
figure
hold on
plot(a,A(:),'b')
plot(a,F(:),'r')
xlabel('Taux de panne {\alpha} ')
ylabel('Disponibilité et fréquence de panne ')
grid
hold off
H
A
F

```

Disponibilité et fréquence de panne vs. β

```

a=0.1; l=2; m=5;
A=[];
F=[];
C=[];
D=[];

```

```

E=[];
H=[];
b=0:0.01:1;
H=m^2*a^2*l+6*m^2*a^2*b+2*m^3*a^2+2*m^2*a^3+6*b.^2*m*a*l+2*m*b.^3*l+...
    +4*b.^2*m^2*l+b.^3*l^2+3*b.^2*m*l^2+b.^2*l^3+2*a*b.^2*l^2+4*m^3*a*b+...
    +6*b.^2*m^2*a+2*m^2*b.^3+2*m^3*b.^3+2*m^3*b.^2+2*m*a*b*l^2+6*m^2*a*b*l+4*b*m*a^2*l;
C=b.^2./H;
B1=(b+l+m);
k1=(2*m^2+2*m*l+l^2);
o1=2*(l+m)^2*a;
v1=B1.*k1+o1;
% la disponibilité de serveurs
A=C.*v1
% la fréquence de pannes
B2=(b+l+2*m+2*a);
k2=(2*b*l+2*b*m+m*a);
o2=(a+2*b)*m.*b;
v2=B2.*k2+o2;
E=(m*a)./H;
F=E.*v2
% les graphes
figure
hold on
plot(b,A(:), 'b')
plot(b,F(:), 'r')
xlabel('Taux de réparation {\beta} ')
ylabel('Disponibilité et fréquence de panne ')
grid
hold off
H
A
F

```

Disponibilité et fréquence de panne vs. μ

```

a=0.1; l=70; b=2;l1=10;
A=[];
F=[];
C=[];
D=[];
E=[];
H=[];
m=1:0.1:200;
H=m.^2*a^2*l+6*m.^2*a^2*b+2*m.^3*a^2+2*m.^2*a^3+6*b^2*m*a*l+2*m*b^3*l+...
    +4*b^2*m.^2*l+b^3*l^2+3*b^2*m*l^2+b^2*l^3+2*a*b^2*l^2+4*m.^3*a*b+...
    +6*b^2*m.^2*a+2*m.^2*b^3+2*m.^3*b^3+2*m.^3*b^2+2*m*a*b*l^2+6*m.^2*a*b*l+4*b*m*a^2*l;
C=b^2./H;
B1=(b+l+m);
k1=(2*m.^2+2*m*l+l^2);
o1=2*(l+m).^2*a;
v1=B1.*k1+o1;
% la disponibilité de serveurs
A=C.*v1
%calculer A pour une autre valeur de lambda
H1=m.^2*a^2*l1+6*m.^2*a^2*b+2*m.^3*a^2+2*m.^2*a^3+6*b^2*m*a*l1+2*m*b^3*l1+...
    +4*b^2*m.^2*l1+b^3*l1^2+3*b^2*m*l1^2+b^2*l1^3+2*a*b^2*l1^2+4*m.^3*a*b+...
    +6*b^2*m.^2*a+2*m.^2*b^3+2*m.^3*b^3+2*m.^3*b^2+2*m*a*b*l1^2+6*m.^2*a*b*l1+4*b*m*a^2*l1
C1=b^2./H1;
B3=(b+l1+m);
k3=(2*m.^2+2*m*l1+l1^2);
o3=2*(l1+m).^2*a;
v3=B3.*k3+o3;
% la disponibilité de serveurs
A1=C1.*v3
% la fréquence de pannes
B2=(b+l+2*m+2*a);
k2=(2*b*l+2*b*m+m*a);
o2=(2*b+a)*b*m;
v2=B2.*k2+o2;

```

```

E=(m*a)./H;
F=E.*v2
% calculer la fréquence de pannes pour une autre valeur de lambda
B5=(b+l1+m+2*a);
k5=(2*b*l1+2*b+m*a);
o2=(2*a^2);
v5=B5.*k5+o2;
E1=(m*a)./H1;
F1=E1.*v5
% les graphes
figure
hold on
plot(m,A(:),m,A1(:),'b')
plot(m,F(:),'r',m,F1(:),'r')
xlabel('Taux de service {\mu} ')
ylabel('Disponibilité et fréquence de panne ')
grid
gtext('\lambda=70 ')
gtext('\lambda=10 ')
gtext('\lambda=70 ')
gtext('\lambda=10 ')
hold off
H
A
F

```

Code MATLAB : Cas du modèle M/G/1 avec rappels, pannes et priorité

Disponibilité et fréquence de panne vs. α_2

```

% les paramètres*
% l1 taux des arrivées des clients de type1
% l2 taux des arrivées des clients de type2
% m1 taux de service des clients de type1
% m2 taux de service des clients de type2

```



```

hold on
plot(a2,F(:),'r')
plot(a2,A(:),'b')
title('A and F vs {\alpha_2}')
xlabel('faillure rate {\alpha_2} ')
gtext('A ')
gtext(' F ')
ylabel('Availability and Failure frequency ')
grid
hold off
H
A
F

```

Le comportement des probabilités p_4, p_5 et p_6 vs. α_1 et α_2

```

l1=0.3; l2=0.2; m1=0.1;m2=0.3;B11=1;B21=1;G11=1;G21=1;
B2=inline('1-exp(-1*x)', 'x')
a2=0.1;
I=1;
R1=[];
  a1=0:0.01:2;
R1=l1*B11*(1+G11.*a1);
H=[];
P4=[];
P5=[];
P6=[];
%calcul de H
L1=l2*(1-R1);
L22=R1+l2*I;
L2=(1+l2*I)+L22*G21.*a2;
H=L1./L2;
%calcul de P1
P11=(l1*B11)*(1+G21*a2);
P13=l2*(1-R1)

```

```

P12=1./P13;
P15=P11.*P12;
P1=H.*P15;
%calcul de P2
P2=I.*H;
%calcul de P3
P31=I./(1-R1);
P32=P31.*P11;
P3=P32.*H;
%calcul de P4
P41=G11.*a1.*P1;
P4=P41.*a1;
%calcul de P5
P51=G21.*P2;
P5=P51*a2;
%calcul de P6
P61=G11.*P3;
P6=P61.*a1;
%Graphe
figure
hold on
plot(a1,P5(:),'r')
plot(a1,P6(:),'b')
plot(a1,P4(:),'g')
xlabel('faillure rate {\alpha_1} ')
ylabel(' ')
gtext('p_5 ')
gtext(' p_6 ')
gtext(' p_4 ')
title('p_4, p_5 and p_6 vs {\alpha_1} ')
grid
hold off

```

REFERENCES

1. L. Kosten. On the influence of repeated calls in the theory of probabilities of blocking. *De Ingenieur (dutch)*, 59, 1947.
2. A. Aissani. On the $M/G/1/1$ queueing system with repeated orders and unreliable server. *Journal of Technology*, 6 :98–123, 1988.
3. G.I. Falin. a survey of retrial queues. *Queueing Systems*, 7 :127– 168, 1990.
4. V. G. Kulkarni and H. M Liang. Retrial queues revisited, in : Frontiers in queueing, ed.r. *J.H. Dshalalow (CRC Press, Boca Raton, FL, 1997)*, pages 19–34, 1997.
5. J.G.C. Templeton. Retrial queues. *Top*, 7 :351–353, 1999.
6. G.I. Falin and J.G.C. Templeton. Retrial queues. *Chapman and Hall, London*, 1997.
7. J.R. Artalejo and A. Gomez-Corral. Retrial queueing systems : A computational approach, systems with customer retrials. *Springer, Berlin, Germany*, 2008.
8. D. Stoyan. Comparison methods for queues and other stochastic models. *New York*, 1983.
9. L Berjoudj and D Aissani. Strong stability in retrial queues. *International journal in the theory of probability and mathematical statistics*, 2003.
10. P.J.S. Courtois. Decomposability, instabilities, and saturation in multi-programming systems. *Communications of the ACM*, 18(7) :371–377, 1975.
11. V.S. Korolyuk and V.V. Korolyuk. Stochastic models of systems. *Kluwer Academic Publishers, Boston*, 1999.
12. L. Bouterfa and N. Djellab. On the performance of the $M_1, M_2/G_1, G_2/1$ retrial queue with preemptive resume policy. *Yugoslav journal of operations research*, 25(1) :153–164, 2015.

13. N. Oukid. Comparaisons stochastiques et applications. *Thèse de doctorat d'état, Université de Blida*, 2009.
14. C. Clos. An aspect of the dialing behaviour of subscribers and its effect on the trunk plant. *Bell Syst. Tech. J.*, 27 :424–445, 1948.
15. J.W. Cohen. Basic problem of telephone and influence of repeated calls. *Telecommunication Review*, 2 :42–100, 1957.
16. J. Keilson, V.A Cozzolino, and H. Young. A service system with unfiled requests repeated. *Operations Research*, 16 :1126–1137, 1968.
17. A.M. Alexandrov. A queueing system with repeated orders. *Engenneering Cybernetics Review*, 12(3) :1–4, 1974.
18. J. Amador and J. R. Artalejo. Transient analysis of the successful and blocked events in retrial queues. *Telecommunication Systems*, 41 :255–265, 2009.
19. J. R Artalejo, G. I Falin, and M. J Lòpez-Herrero. A second order analysis of the waiting time in the M/G/1 retrial queue. *Asia-Pacific Journal of Operational Research*, 19 :131–148, 2002.
20. N. Oukid. Comparaisons stochastiques de files d'attente. *Thèse de magister, Université de Blida*, juin 1995.
21. J. R. Artalejo. Accessible bibliography on retrial queues. *Mathematical and computer modelling*, 30 :1–6, 1999.
22. J. R. Artalejo. Accessible bibliography on retrial queues : Progress in 2000-2009. *Mathematical and computer modelling*, 51 :1071–1081, 2010.
23. A. Aissani. Influence des pannes des serveurs sur la distribution de nombre de clients en orbite et dans un système M/G/1/0 avec rappels. *Technologies Avancées*, 2 :23–38, 1991.
24. J. R. Artalejo and G. Choudhoury. Steady state analysis of an M/G/1 queue with repeated attempts and two phase service. *Quality Technologi and Quantitative Management*, 1(2) :189–199, 2004.
25. J. R. Artalejo and M. J. Lòpez-Herrero. The M/G/1 retrial queue : An information theoretic approach. *Statistics and Operations Association*, 2 :1–17, 2005.
26. M. J. Lòpez-Herrero. Distribution of the number of customers served in an M/G/1 retrial queue. *Journal of applied Probability*, 39 :407–412, 2002.

27. B. Krishna Kumar and D. Arivudainambi. The M/G/1 retrial queue with bernoulli schedule and general retrial times. *Computers and Mathematics with Applications*, 43 :251–264, 2002.
28. Q.H. Choo and B. Conolly. new results in the theory of repeated orders queueing systems. *J. Appl. Prob*, 16 :631–640, 1979.
29. G.I. Falin. On heavily loaded systems with repeated calls. *Sov. J. Comput. Syst. Sci*, 24(4), 1986.
30. T. Yang and J.G.C. Templeton. The $Ca/M/s/m$ retrial queue : a computational approach. *ORSA Journal on Computing*, 4(2) :182–191, 1992.
31. V.I. Meykshan and I.G. Fidelman. The design of communications networks with bypass routings when there are repeated calls and lines with undetected breakdowns. *Telecommunications and Radio Engineering*, 49(7) :40 – 44, 1995.
32. D. Fiems, T. Maertens, and H. Bruneel. Queueing systems with different types of server interruptions. *European journal of operational research*, 188 :838–845, 2008.
33. T. Kernane. A single server retrial queue with different types of server interruptions. *Math.PR*, 2013.
34. D.P. Gaver. A waiting line with interrupted service including priorities. *Journal of the royal statistical society*, B24 :73–90, 1962.
35. J. R. Artalejo. new results in retrial queueing systems with breakdown of the servers. *Statistica Neerlandica*, 48 :23– 36, 1994.
36. V. G. Kulkarni and B. D. Choi. Retrial queues with server subject to breakdowns and repairs. *Queueing Systems*, 7(2) :191–208, 1990.
37. A. Aissani and J. R. Artalejo. On the single server retrial queue subject to breakdowns. *Queueing Systems*, 30(3-4) :309–321, 1998.
38. A. Aissani. Unreliable queueing with repeated orders. *Microelectronics and Reliability*, 33(14) :2093–2106, 1993.
39. T. Yang and H. Li. The M/G/1 retrial queue with the server subject to starting failures. *Queueing Systems*, 16(1-2) :83–96, 1994.
40. H. Li and T. Yang. A single-server retrial queue with server vacations and a finite number of input sources. *European Journal of Operations Research*, 85(1) :149–160, 1995.

41. J. Wang, J Cao, and Q. Li. Reliability analysis of the retrial queue with server breakdowns and repairs. *Queueing Systems*, 38(4) :363–380, 2001.
42. N.V Djellab. retrial queue with random breakdowns and general retrial times. *Proceedings of the Int. Conference "Modern Mathematical Methods of Investigating of the Information Networks", Minsk*, page 73 – 78, 2001.
43. N.V Djellab. Systèmes de files d'attente avec rappels : Méthode d'approximation pour un système M/G/1 avec rappels et pannes. *Thèse de Doctorat, Université de Annaba*, 2003.
44. N.V Djellab. On the M/G/1 retrial queue subjected to breakdowns. *RAIRO : Operations Research*, 36 :299–310, 2002.
45. B. Almasi, J Roszik, and J Sztrik. Homogeneous finite-source retrial queues with server subject to breakdowns and repairs. *Mathematical and Computer Modelling*, 42 :673–682, 2005.
46. H. Li and Y. Q. Zhao. A retrial queue with a constant retrial rate, server breakdowns and impatient customers. *Stochastic Models*, 21(2-3) :531–550, 2005.
47. N. Sherman and J. Kharoufeh. An M/M/1 retrial queue with unreliable server. *Operations Research Letters*, 34 :697–705, 2006.
48. M. Jain, G.C. Sharma, and R.S Pundhir. The M/G/1 retrial queueing system with setup, server breakdown and repair. *Ganita*, 58(2) :137–155, 2007.
49. G.I Falin. An M/G/1 retrial queue with an unreliable server and general repairs times. *Performance Evaluation*, 67 :569–582, 2010.
50. B.D. Choi and K.K. Park. The M/G/1 retrial queue with bernoulli schedule. *Queueing systems*, 7(2) :219–228, 1990.
51. C. Langaris and E. Moutzoukis. A retrial queue with structured batch arrivals priorities and server vacation. *Queueing systems*, 20 :341–361, 1995.
52. G.I. Falin, J.R. Artalejo, and M. Martin. On the single retrial queue with priority customers. *Queueing systems*, 14(3-4) :439–455, 1993.
53. B.D. Choi and Y. Chang. single retrial queue with priority calls. *Mathematical and computer modeling*, 30(3-4) :7–32, 1999.

54. J.R Artalejo, A.N. Dudin, and V.I. Klimenok. Stationary analysis of a retrial queue with preemptive repeated attempts. *Oper. Res. Lett*, 28 :173–180, 2001.
55. G.I. Falin, J.R. Artalejo, and M. Martin. Markovian single server retrial queues with two way communication. *in : Proceedings of the 6th International Conference on Queueing Theory and Network Applications, Seoul*, pages 1–7, 2011.
56. J. R Artalejo and T. Phung-Duc. Single server retrial queues with two way communication. *Applied Mathematical Modelling*, 37 :1811–1822, 2013.
57. I. Dimitriou. A mixed priority retrial queue with negative arrivals, unreliable server and multiple vacations. *Appl. Math. Model*, 37 :1295–1309, 2013.
58. G.I. Falin and J.R. Artalejo. A finite source retrial queue. *European Journal of Operational Research*, 108 :409–424, 1998.
59. J.R Artalejo and G.I. Falin. On the orbit characteristics of the M/G/1 retrial queue. *Naval Research Logistics*, 43 :1147–1161, 1996.
60. G.I. Falin. Aggregate arrival of customers in one-line system with repeated calls. *Ukrainian Math. Journal*, 28 : :437– 440, 1976.
61. G.I. Falin. Functioning under nonsteady conditions of a single-channel system with group arrival of requests and repeated calls. *Ukrainian Math. Journal*, 33(4) :429–432, 1981.
62. T. Yang and J.G.C. Templeton. A survey on retrial queues. *Queueing systems*, 2 :203–233, 1987.
63. N.K. Boots and H. Tijms. An M/M/c queue with impatient customers. *TOP*, 7 :213–220, 1999.
64. T. Hanschke. Explicit formulas for the characteristics of the M/M/2/ queue with repeated attempts. *J. Appl. Prob*, 24 :486–494, 1987.
65. M.F. Neuts and B.M. Rao. Numerical investigation of a multiserver retrial model. *Queueing systems*, 7 :169–190, 1990.
66. B.D. Choi, Y.C. Kim, and Y.W. Lee. The M/M/c retrial queue with geometric loss and feedback. *Comput. Math. Appl*, 36 :41–52, 1998.
67. J.R. Artalejo and M. Pozo. Numerical calculation of the stationary distribution of the main multi-server retrial queue. *Ann. Oper. Res*, 116 :41–56, 2002.

68. B. Krishna Kumar and J. Raja. On multiserver feedback retrial queues with balking and control retrial rate. *Ann. Oper. Res.*, 141 :211–232, 2006.
69. J. Kim and B. Kim. Tail asymptotics of the queue size distribution in the $M/M/m$ retrial queue. *Comput. Appl. Math.*, 236 :3445–3460, 2012.
70. B. K. Kuma, A. Vijayakumar, and D. Arivudainambi. An $m/g/1$ retrial queueing system with two-phase service and preemptive resume. *Annals of Operations Research*, (113) :61–79, 2002.
71. J. Wang and P. Zhang. A discrete-time retrial queue with negative customers and unreliable server. *Computer and Industrial Engineering*, (57) :433–443, 2009.
72. A. GÓMEZ-CORRAL and M. L. GARCÍA. Control strategies for a stochastic model of host–parasite interaction in a seasonal environment. *Journal of Theoretical Biology*, 354 :1–11, 2014.
73. F. Guérin. Estimation de la fiabilité par les essais. *Thèse de doctorat. Université d'Angers*, 2004.
74. M. Mouy. Estimation de la fiabilité d'un système et de ses composants. *Thèse de Magister. Université de Mentouri*, 2009.
75. G. Saidi. Lois non paramétrique de service et leurs applications. *Thèse de doctorat. Ecole nationale supérieure de statistique et d'économie appliquée*, 2011.
76. O. Tebbi. Estimation de lois de fiabilité en mécanique par les essais accélérés. *Thèse de doctorat. Université d'Angers*, 2005.
77. A. Pages and M. Gondran. Fiabilité des systèmes. *Ed Eyrolles*, 1980.
78. A. Villemeur. Sûreté de fonctionnement des systèmes industriels : fiabilité, facteurs humains, informatisation. *Eyrolles, Paris*, 1988.
79. A Tchangani and D. Noyes. Attempt to modeling dynamic reliability using dynamic bayesian networks. *Proc. QUALITA*.
80. C. Petri. Kommunikation mit automaten. *PhD thesis, Universitat Bonn*, 1962.
81. O. Daniel. Les réseaux de petri stochastiques pour l'évaluation des attributs de la sûreté de fonctionnement des systèmes manufacturiers. *PhD thesis, LAG Grenoble*, 1995.
82. A. Demri. Contribution à l'évaluation de la fiabilité d'un système mécatronique par modélisation fonctionnelle et dysfonctionnelle. *PhD thesis, Université d'Angers*, 2009.

83. C. Girault and R. Valk. Petri nets for systems engineering. *guide to modelling verification and application*. Springer, 2003.
84. P. Lyonnet. Ingénierie de la fiabilité. *Lavoisier*, 2006.
85. J.L. Bon. On the fiabilité des systèmes : méthodes mathématiques. *Ed Masson*, 1995.
86. C. Coccozza-Thivent. Processus stochastiques et fiabilité des systèmes.
87. R. Doyle. Mechanical reliability. *In IEEE Proceedings Annual Reliability and Maintainability Symposium, Tutorial notes, USA*, 1991.
88. J. Debray. Analyse de fiabilité et essais accélérés pour la validation de pinces saisie pour véhicules à câbles. *Actes de la journée d'étude Société des ingénieurs automobile (SIA) du 07/04/2000. Fiabilité expérimentale : essais accélérés et autres techniques pour démontrer un niveau de fiabilité au moindre coût*. Paris, 2000.
89. N. Gharbi and M. Ioualalen. Gspn analysis of retrial systems with servers breakdowns and repairs. *Applied Mathematics and Computation*, 174 :1151–1168, 2006.
90. N. Oukid and A. Aissani. Bounds on busy period for queues with breakdowns. *Advances and Applications in Statistics*, 11 :137–156, 2009.
91. B.P. Crawford. Approximate analysis of an unreliable M/M/2 retrial queue. *Masters Thesis, Department of the Air Force, Graduate School of Engineering and Management, Air Force Institute of Technology, Air University, Ohio, USA*, 2007.
92. L. Raiah and N. Oukid. An M/M/2 retrial queue with breakdowns and repairs. *Romanian journal of mathematics and computer science*, 7(1) :11–20, 2017.
93. L. Raiah, N. Oukid, and N. Djellab. On the retrial queue with priority calls and server breakdowns. *Afrika matematika, article soumis*.
94. J. Cao and K. Cheng. Analysis of M/G/1 queueing system with repairable service station. *Acta-Math.Appl.Sinica*, (5) :113–127, 1982.
95. E. Cinlar. Introduction to stochastic processes. *New york*, 1975.
96. G.I. Falin. Single-line repeated orders queueing systems. *Engineering Cybernetics Review*, 21(6) :21–25, 1983.

97. L. Raiah and N. Oukid. Approximate analysis for reliability of an unreliable M/M/2 retrial queue. *La première conférence nationale sur les Systèmes Dynamiques, Equations Différentielles et Applications, Oum El Bouaghi*, 10-11 Mars 2015.
98. L. Raiah and N. Oukid. An M/M/2 retrial queue with breakdowns and repairs. *3rd International Conference on Recent Advances in Pure and Applied Mathematics, ICRAPAM, Turquie*, 19-23 Mai 2016.
99. L. Raiah and N. Oukid. On the $M_1, M_2/G_1, G_2/1$ retrial queue with priority calls and server breakdowns. *4th International Conference on Recent Advances in Pure and Applied Mathematics, ICRAPAM, Turquie*, 11-15 Mai 2017.