الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
جامعة الشاذلي بن جديد – الطارف
Université Chadli Bendjedid – El Tarf
كلية العلوم و التكنولوجيا
Faculté des Sciences et de la Technologie
قسم الرياضيات..
Département de Mathématiques

Mémoire de fin d'études

En vue de l'obtention du diplôme de Master

Domaine : .Mathématiques et Informatique

Filière : Mathématiques

Spécialité : Analyse fonctionnel et calcule stochastique

Thème

Analyse des systèmes de files d'attente avec rappels et multiserveurs par la méthode d'Extrapolation de valeur.

Présenté par:

Ben halima Khadidja

Devant le Jury :

Dr. Grabsia Imen	**MCB**	Univ Chadli Bendjedid El Tarf	Président
Dr. Zidani Nesrine	**MCB**	Univ Chadli Bendjedid El Tarf	Rapporteur
Dr. Grine Razika	**MCB**	Univ Chadli Bendjedid El Tarf	Examinatrice

Année Universitaire 2020-2021

Remerciement

Tout d'abord je remercie Dieu de m'avoir donné la force et la capacité de terminer ce mémoire.

Ce travail ne serait pas aussi riche et n'aurait pas pu avoir le jour sans l'aide et l'encadrement de Dr. Zidani Nesrine, je la remercie pour son soutien, sa disponibilité, son assistance, ses orientations et ses conseils.

Je remercier sincèrement les membres du jury :

Grine Razika et Grabsia Imen

Finalement, je remercie mes parents pour m'avoir facilité la vie et donner tout ce que j'en avais besoin pour réussir dans mes études.

Dédicace

Je dédie ce modeste travail à :

Ma chère mère qui m'a encouragé et aider durant le cycle d'études en .lui souhaitant longue vie et que dieu la protéger

Mon père qui n'a jamais cessé de m'encourager dans la poursuite de .mes études en m'apportant tout son soutien, que dieu le garde

Et tous mes frères et ma sœur

Toute la famille

Tous mes voisins, mes amis

Toute personne qui a partagé avec moi un moment de bonheur

Et a mes collègues de promotion de master 2

TABLE DES MATIÈRES

Résumé

Dans ce travail, nous nous sommes intéressés par l'étude d'un système de files d'attente avec rappels et multiserveurs M/M/C. Dans un premier temps, nous avons effectué une étude sur les systèmes de files d'attente classiques. Après nous avons présenté la description du modèle général d'un système de files d'attente avec rappels, en particulier celui M/M/C avec rappels, ainsi les solutions exactes proposées dans la littérature pour ces modèles markoviens.

Dans un deuxième temps, nous avons décrit deux approches d'obtention des solutions, rencontrées dans la littérature, pour les modèles avec rappels et multiserveurs "la troncature finie et l'Extrapolation de Valeur". Et à la fin nous avons examiné la performance de la méthode Extrapolation de Valeur, en termes de précision, en comparant les résultats numériques fournis par cette méthode avec la méthode de Troncature finie.

Mots clés : systèmes de files d'attente, processus stochastique, Extrapolation de Valeur, Troncature Finie.

Abstract

In this work, we were interested to the study of a retrial multiservers queuing system M\M\C. First, we carried out a study on classic queuing systems. Then, we presented the description of the general model of a retrial queuing systems , in particular M\M\C, as well as the exact solutions proposed in the literature for these Markovian models.

Secondly, we have described two approaches for obtaining solutions, encountered in the literature, for models with retrial and multiservers "Finite Truncation and Extrapolation of Values". And at the end we examined the performance of the method of Extrapolation of Values, in terms of precision, by comparing the numerical results provided by this method with the method of Finite Truncation.

Keywords : Queuing Systems, Stochastic Process, Extrapolation of Values, Finite Truncation.

ملخص

في هذا العمل، نحن مهتمون بدراسة نظام الطابور مع عمليات الاسترجاعات والخوادم المتعددة M\M\C. أولاً، أجرينا دراسة حول أنظمة الطوابير التقليدية. بعد ذلك، قدمنا وصفًا للنموذج العام لنظام الانتظار مع عمليات الاسترجاعات، وخاصة M\M\C مع عمليات الاسترجاعات، بالإضافة إلى الحلول الدقيقة المقترحة في الأدبيات الخاصة بنماذج ماركوفيان.

ثانيًا، لقد وصفنا طريقتين للحصول على الحلول، تمت مواجهتهما في الأدبيات، للنماذج ذات الاسترجاعات والخوادم المتعددة "الاقتطاع المحدود واستقراء القيمة". وفي النهاية درسنا أداء طريقة استقراء القيمة، من حيث الدقة، من خلال مقارنة النتائج العددية التي توفرها هذه الطريقة مع طريقة الاقتطاع المحدود.

الكلمات المفتاحية: أنظمة الطابور، العملية العشوائية، استقراء القيمة، الاقتطاع المحدود.

Introduction

La théorie des files d'attente tire son origine des recherches de l'ingénieur danois Agner krarup Erlang entre 1909 et 1920. Ce dernier étudiait le concept de la file d'attente des systèmes téléphoniques, dans les centres d'appels de Copenhague. Depuis, plusieurs mathématiciens se sont intéressés aux files d'attente et ont développé les modèles mathématiques de cette théorie. Plusieurs modèles de files d'attente classiques ont été étudiés depuis Erlang, et plusieurs formules "élégantes" ont été élaborées et proposées comme étant des solutions analytiques de certains types de problèmes.

Dans le modèle des files d'attente classique, il est supposé qu'un client qui ne peut pas obtenir son service immédiatement dès son arrivée, rejoint la file d'attente ou quitte le système définitivement. Les systèmes de file d'attente développés tentent de prendre en considération des phénomènes de répétition de demandes de service, et ceci après une durée du temps aléatoire. Un tel système est connu comme «système de files d'attente avec rappels».

Le système des files d'attente avec rappel a été surtout utilisée pour modéliser les systèmes téléphoniques, centres d'appels et les réseaux informatiques. Elle sert à résoudre des problèmes pratiques, tels que l'analyse du temps d'attente des abonnés dans les réseaux téléphoniques, l'évitement de collision dans les réseaux locaux, l'analyse du temps d'attente pour accéder à la mémoire sur les disques magnétiques.

La théorie analytique des systèmes de files d'attente avec rappels s'avère d'une portée limitée en raison de la complexité des résultats obtenus. En effet, dans la majorité des cas, on se retrouve conforté à des systèmes d'équations dont la résolution est complexe, ou possédant des solutions qui ne sont pas facilement interprétables pour que le praticien puisse en bénéficier. Depuis la publication des premiers résultats dans les années 1950, les files d'attente avec rappels ont été largement utilisés pour modéliser plusieurs problèmes de télécommunication, réseaux d'ordinateurs, et dans la vie quotidienne. L'intérêt le plus grand des files d'attente avec rappels est alors reflété sur l'existence de séries d'ateliers internationnaux sur les files d'attente avec rappels qui ont apparu à Madrid (1998). Par la suite, des sémminaires ont été fait à Minsk (1999), Amsterdam (2000), Cochin (2002), Seoul (2004), Miraflores de la Sierra (2006) et Athens (2008). La nature des résultats obtenus, des méthodes d'analyse et des domaines d'application nous permettent de diviser les files d'attente avec rappels en deux groupes : les systèmes à serveur unique, et à multiserveurs. Les modèles à multiserveurs attirent beaucoup d'attention parce que la conception d'un systèmes de files d'attente avec rappels est déduite de ces importantes applications dans les systèmes de téléphonie.

Du à l'absence des formules analytiques explicites pour les caractéristiques probabilistes principales des modèles de files d'attente avec rappels et multiserveurs, la seule méthode pour obtenir des données numériques précises consiste à résoudre les équations de Kolmogorov numériquement. Mais dès que ce système d'équations est infini, il ne peut être résolu directement même sur ordinateur. Les transformations, qui réduisent ces équations à une solution d'un petit problème fini, dans le cas général ne sont pas disponibles. C'est pourquoi, nous nous adressons aux approches d'approximations.

Cependant, de nombreux chercheurs recourent à des approximations, qui sont souvent basées sur des modèles tronqués finis [30], des modèles tronqués généralises ou sur une homogénéisation de l'espace d'états [26]. Dans les deux premiers cas, le processus stochastique hétérogène, décrivant l'état du système et possédant un espace d'états infini, est remplacé par un autre homogène et ayant l'espace d'états fini ou infini mais résoluble. Le troisième type d'approximation consiste à analyser le système

de files d'attente comme un processus de quasi-naissance et mort dont les probabilités stationnaires peuvent être obtenues en utilisant une méthode matricielle géométrique [23]. Ces approches fournissent une solution numérique à la distribution stationnaire des chaines de Markov à temps continu. Une approche alternative a été proposée dans [22] pour calculer les mesures de performance des processus de Markov à espace d'états infini, appelé Extrapolation de Valeur. L'approche en question consiste à étudier un processus stochastique de Markov en tant qu'un processus de Markov décisionnel : on ne considère plus la probabilité d'être dans un certain état, mais on s'intéresse à une nouvelle métrique, appelée valeur d'état relative. Le but est de trouver la valeur relative moyenne, représentant une mesure de performance du modèle, en résolvant les équations de Howard.

Dans ce mémoire, nous présentons deux approches alternatives : Troncature Finie et Extrapolation de Valeur. Nous présentons des résultats numériques pour examiner la performance de la méthode Extrapolation de Valeur des systèmes de files d'attente avec rappels et multiserveurs $M/M/C$. Nous montrons que la méthode basée sur les outils de la théorie des Processus de Markov Décisionnels est plus appropriée pour la résolution des systèmes $M/M/C$ que celle qui donne la solution pour la distribution stationnaire de la chaine de Markov à temps continu.

Ce mémoire est constitué d'une introduction générale, quatre chapitres, d'une conclusion générale, d'une bibliographie et d'un Annexe.

Dans le premier chapitre, nous présentons les systèmes de files d'attente classiques.

Le second chapitre contient la description du modèle général d'un système de files d'attente avec rappels, en particulier celui $M/M/C$ avec rappels, puis les solutions exactes proposées dans la littérature pour ces modèles markoviens.

Dans le troisième chapitre, nous décrivons le modèle tronqué et l'approche Extrapolation de Valeur.

Dans le quatrième chapitre, nous examinons la performance de la méthode Extrapolation de Valeur, en termes de précision, en comparant les résultats numériques fournis par cette méthode avec ceux de la Troncature Finie.

Dans la conclusion générale, nous présentons un perspective de recherche.

CHAPITRE 1

Systèmes de Files D'attente Classiques

Les systèmes de files d'attente décrivent un aspect de la vie moderne que nous rencontrons à chaque étape de nos activités quotidiennes. Qu'il se produit devant un guichet d'une banque ou en accédant à l'internet, le phénomène de base des files d'attente surgit chaque fois qu'un serveur (guichet, routeur, ...) est consulté pour son service par un grand nombre de taches ou de clients.

Dans ce chapitre, nous présentons les éléments essentiels des systèmes de files d'attente classiques.

1 Rappel historique

La théorie des files d'attente fournit un outil très puissant et efficace pour la modélisation des systèmes admettant un phénomène d'attente. Cette théorie date du début du XXème siècle par les travaux de l'ingénieur danois Agner Krarup Erlang (1878,1929). Ses études sur le trafic téléphonique de Copenhague pour le mieux géré afin de déterminer le nombre de circuits nécessaires pour fournir un service téléphonique acceptable, sont considérées comme la première brique dans cette théorie [5].

Ensuite, les files d'attente ont été intégrés dans la modélisation des systèmes in-

formatiques et aux réseaux de communication. Cette intégration dans ces domaines et d'autres a permis une évolution de cette théorie surtout dans l'évaluation des paramètres de performances des systèmes informatiques et aux réseaux de communication.

Actuellement ce sont les applications dans le domaine de l'analyse de performance des réseaux (téléphone mobile, Internet, multimédia,...) qui suscitent le plus de travaux.

L'étude d'un système de files d'attente consiste à calculer ces paramètres de performances afin d'évaluer son rendement, et améliorer son fonctionnement (minimiser le temps d'attente et le temps d'inactivité de l'installation) de savoir par exemple si le nombre de serveurs dans le système et est adéquate pour gérer le flux de demandes ou encore d'appréhender les effets d'une modification des conditions de fonctionnement, et ainsi prendre des décisions sur le nombre minimum de ressources nécessaires.

Depuis les travaux d'Erlang un grand nombre d'applications dans tous les domaines ont été mis en œuvre et publiées. En 1953, David G. Kendall a introduit la notation de Kendall pour décrire les caractéristiques d'un système de file d'attente. En 1957 d'une manière particulièrement élégante et efficace Jackson a traité certains réseaux de files d'attente. En 1961, Thomas L. Saaty [3], auteur de l'un des premiers livres complets sur la théorie des files d'attente. Ensuite c'est les contributions des mathématiciens Khintchine, Palm, Pollaczek et Kolmogorov qui ont vraiment poussé la théorie des files d'attente.

2 Description des files d'attente

Une file d'attente ou queue est un système stochastique composé d'un certain nombre(fini ou non) de places d'attente d'un ou plusieurs serveurs et bien sûr de clients qui arrivent, attendent, se font servir selon des règles de priorité données et quittent le système. La description précédente d'une file d'attente, dont une représentation schématique est donnée en figure 1.1, ne saurait capturer toutes les caractéristiques des différents modèles que comptent la littérature, mais elle identifie les éléments principaux permettant la classification de la grande majorité des files

d'attente simple.

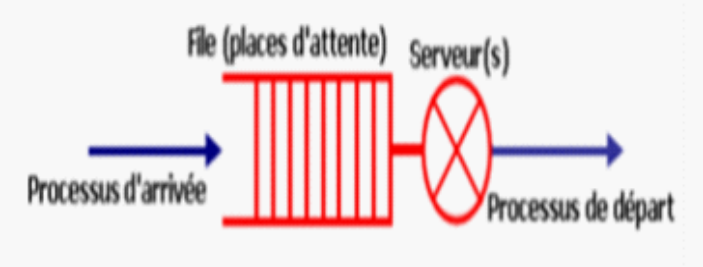

FIG. 1.1 – Système de files d'attente classique

3 caractéristiques d'un système de files d'attente

Pour identifier un système de files d'attente, on doit spécifier le processus d'arrivée, le processus de service, le nombre de serveurs, la discipline de service, la capacité de la file d'attente et la taille de la population.

3.1 Processus d'arrivée

Dans des situations de files d'attente habituelles, le processus d'arrivée est stochastique. Il est nécessaire donc de :

– Définir la probabilité décrivant les temps entre deux arrivés successifs des clients (les inters arrivés). Dans la théorie classique des files d'attente, on fait le plus souvent l'hypothèse que les clients arrivent de manière isolée et indépendamment les uns des autres.

– Connaitre la réaction du client lors d'entrer au système : s'il decide d'attente dans la file ou bien, si cette dernière est très long, il decide de quitter le système sans recevoir le service. Lorsque'un client decide de ne pas entrer dans la file dès son arrive, il est appelé client rejeté. Cependant s'il decide d'entrer dans la file d'attente, mais perd sa patience après quelque temps, il est appelé client découragé.

3.2 Processus de service

Le temps de service est la durée nécessaire au traitement d'un client (c'est le temps écoulé depuis l'arrivée du client dans le serveur jusqu'au point de départ), il est supposé être une variable aléatoire indépendante et identiquement distribuée.

3.3 Nombre de serveur

Le nombre de serveurs correspond au nombre maximal de clients qui peuvent être traités simultanément. Tous les serveurs sont supposés identiques, en particulier les temps de service sont indépendants d'un serveur à l'autre et distribués selon une même loi de probabilité.

3.4 Discipline de service

La discipline de service détermine l'ordre dans lequel les clients sont rangés dans la file pour pouvoir déterminer le tour de chaque client pour effectuer son service. Les disciplines les plus courantes sont [8] :

– FIFO (first in, first out) ou (premier arrivé, premier servi) : c'est la file standard dans laquelle les clients sont servis dans leur ordre d'arrivée. Le premier client arrivé est servi premier.

– LIFO (last in, first out) ou LCFS (last come, first served) ou DAPS (dernier arrivé, premier servi) : cela correspond à une file, dans laquelle le dernier client arrivé sera le premier traité, les disciplines LIFO et LCFS ne sont équivalentes que pour une file mono serveur.

– RANDOM (aléatoire) : Le prochain client qui sera servi est choisi aléatoirement dans la file d'attente.

– Round-Robin (cyclique) : Tous les clients de la file d'attente entrent en service à tour de rôle, effectuant un quantum Q de leur temps de service et sont replacés dans la file, jusqu'à ce que leur service soit totalement accompli. Cette discipline de service a été introduite afin de modéliser les systèmes informatiques.

3.5 Capacité de la file

Dans certains systèmes de file d'attente, des contraintes physiques ou organisationnelles peuvent exister et limitent la longueur maximale de la file. Dans ces types de cas, la capacité du système indique le nombre maximal des clients qui peuvent se retrouver dans le système (en attente de service et en service). Dans un système de production, cette capacité peut être liée à une limite de l'espace de stockage.

3.6 Taille de la population

Dans la théorie des files d'attente, la population est la source de clients potentiels. Il y a deux situations possibles. Dans le premier cas, la population est infinie, c'est-à-dire que le nombre potentiel de clients est infiniment grand en tout temps. C'est le cas des clients des supermarchés, des banques, des restaurants, des cinémas, des centres d'appels, trafic urbain, etc. De plus, les clients proviennent de toutes les régions possibles.

Dans la deuxième situation, la population est finie, ce qui signifie que le nombre de clients potentiels est limité. Un bon exemple est le nombre de machines, d'avions, etc., en réparation dans le centre de maintenance d'une entreprise.

4 Notation de Kendalle

Pour la classification des systèmes d'attente, on recourt à une "notation symbolique" dite notation de Kendall. Cette notation comprend des symboles rangés dans l'ordre, elle est noté par $A\backslash B\backslash C\backslash m\backslash O\backslash H$ tel que [8] :

– A indique la loi des temps interarrivées des clients.

– B indique la loi des temps de services.

– C indique le nombre de serveurs.

– m indique la capacité de la salle d'attente, ce dernier sera supprimé si $m = \infty$.

– O indique la discipline du service.

– H la source des clients potentiels, qui peut être fini ou infini, homogène ou hétérogène.

Les lois des deux premiers symboles sont données par :

– M : loi exponentielle,

– G : loi générale,

– D : loi constante (déterministe),

– E_k : loi d'Erlang d'ordre k,

– H_k : loi hyperexponentielle.

5 Mesures de performance

Les mesures de performance sont :

• Le nombre moyen de clients dans le système $\overline{N}$.

• Le nombre moyen de clients dans la file d'attente $\overline{N}_f$.

• Le temps moyen d'attente d'un client dans la file d'attente $\overline{W}$.

• Le temps moyen de séjour d'un client dans le système $\overline{W}_S$.

Ces valeurs ne sont pas indépendantes les unes des autres, mais sont liées par les relations suivantes (Formules de Little) :

$$\begin{aligned}
\overline{N} &= \lambda \overline{W}_S, \\
\overline{N}_f &= \lambda \overline{W}, \\
\overline{W}_s &= \overline{W} + \frac{1}{\mu}, \\
\overline{W} &= \frac{\overline{N}_f}{\lambda}, \\
\overline{N} &= \overline{N}_f + \frac{\lambda}{\mu}.
\end{aligned}$$

Où λ est le taux d'entrée des clients dans le système, et $\frac{1}{\mu}$ la durée moyenne du service ($\mu > 0$). Une autre mesure importante d'un système de files d'attente, celle qui mesure le degré de saturation du système, est l'intensité du trafic ρ

$$\rho = \frac{temps\ moyen\ du\ service}{temps\ moyen\ entre\ deux\ arrivées\ succesives}$$

CHAPITRE 2

Systèmes de files d'attente avec rappel

Les systèmes de files d'attente avec rappels ou avec répétition d'appels se caractérisent par le fait suivant : un client qui arrive dans le système et trouve tous les serveurs occupés, quitte le système définitivement, ou rappelle ultérieurement à des instants aléatoires. Un client qui attend pour rappeler est dit en orbite et devient source d'appels secondaires. Ces modèles d'attente apparaissent dans la modélisation stochastique de plusieurs situations réelles et des réseaux de télécommunications. Par exemple dans la transmission de données, un paquet transmis de la source à la destination peut être retourné, et le processus doit être répété jusqu'à ce que le paquet soit finalement transmis. Les premières tentatives rigoureuses sur les systèmes de files d'attente avec rappels remontent aux travaux de Kosten (1947) [21], Wilkinson (1956) [28], Cohen (1957) [6], et ceci lors de la modélisation du service d'abonnés dans un central téléphonique. Les progrès dans ce domaine sont résumés dans [2, 10, 11],....., et dans les monographies de Falin et Templeton (1997) [12] et A rtalejo et Gomez-Corral (2008) [4].

1 Modèle général

Le modèle général d'un système de files d'attente avec répétition de demandes peut être décrit comme suit : le système contient un espace de service composé de $C \geq 1$ dispositifs de service et d'un espace d'attente ayant $m - c(m \geq c)$ positions d'attente. À l'arrivée d'un client primaire, s'il y a un ou plusieurs serveurs libres, le client sera immédiatement pris en charge. Sinon, s'il y a une position d'attente libre, le client rejoint la file d'attente. Lorsque tous les serveurs et positions d'attente sont occupés, le client quitte le système, soit définitivement avec une probabilité $1 - H_0$ soit temporairement avec une probabilité H_0 et rappelle ultérieurement, après un temps aléatoire. La capacité O de l'orbite peut-être finie ou infinie. Dans le cas où O est finie, et si l'orbite est pleine, le client quitte le système définitivement. Chaque client de l'orbite forme un processus d'arrivées secondaires de taux θ et il est traité de la même manière qu'un client primaire avec une probabilité H_K (s'il s'agit de la K ième tentative échouée). Le schéma général d'un système avec rappels est donné dans la figure suivante :

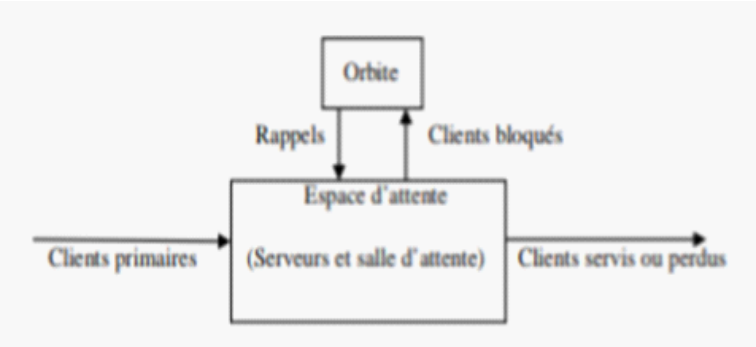

FIG. 2.1 – Schéma général d'un système de files d'attente avec rappels

2 Modèles markoviens

2.1 Modèle M\M\1 avec rappels

On considère un système de files d'attente sans positions d'attente. Le service est assuré par un seul serveur. Les clients primaires arrivent selon un processus de Poisson de taux $\lambda > 0$. Les durées de service suivent une loi exponentielle de fonction de répartition $B(x) = 1 - \exp(-\mu x)$, $x \geq 0$ et de moyenne finie $\frac{1}{\mu}$. Les temps entre deux rappels consécutifs sont également exponentiels de paramètre $\theta > 0$ (la fonction de répartition $T(x) = 1 - \exp(-\theta x)$, $x \geq 0$). Nous admettons que les durées de service, les durées entre deux rappels consécutifs ainsi qu'entre deux arrivées primaires successives sont mutuellement indépendantes. L'état du système peut être décrit par le processus

$$\{C(t), N_0(t); t \geq 0\}, \tag{2.1}$$

Où $C(t)$ est égale à 0 ou 1 selon le fait que le serveur est libre ou non, $N_0(t)$ est le nombre de clients en orbite l'instant t. Supposons que le régime stationnaire existe ($\rho = \frac{\lambda}{\mu} < 1$). Le processus (2.1) est de Markov d'espace d'états $S = \{0, 1\} \times \mathbb{N}$.

Les équations d'équilibre stationnaire sont :

$$(\lambda + j\theta)p_{0j} = \mu p_{1j}; \tag{2.2}$$

$$(\lambda + \mu)p_{1j} = \lambda p_{0j} + (j+1)\theta p_{0,j+1} + \lambda p_{1,j-1}. \tag{2.3}$$

Ici, $p_{ij} = \lim_{t\to\infty} P(C(t) = i, N_0(t) = j), i = 0, 1$ et $j \geq 0$, représentent la distribution stationnaire conjointe de l'état du serveur et du nombre de clients en orbite. Introduisons les fonctions génératrices suivantes :

$$\begin{aligned} P_0(z) &= \sum_{j=0}^{\infty} z^j p_{0j}; \\ P_1(z) &= \sum_{j=0}^{\infty} z^j p_{1j}; \end{aligned}$$

À l'aide de ses fonctions et à partir des équations (2.2) et (2.3), on obtient :

$$P_0(z) = (1-\rho)(\frac{1-\rho}{1-z\rho})^{\frac{\lambda}{\theta}}; \tag{2.4}$$

$$P_1(z) = \rho(\frac{1-\rho}{1-z\rho})^{\frac{\lambda}{\theta}+1}. \tag{2.5}$$

Les transformées inverses des (2.4) et (2.5) nous donnent les formules analytiques explicites [13] :

$$\begin{aligned} p_{0j} &= \frac{\rho}{j!\theta^j}\Pi_{k=0}^{j-1}(1+k\theta)(1-\rho)^{\frac{\lambda}{\theta}+1}; \\ p_{1j} &= \frac{\rho^{j+1}}{j!\theta^j}\Pi_{k=1}^{j}(\lambda+k\theta)(1-\rho)^{\frac{\lambda}{\theta}+1}; \end{aligned}$$

2.2 Modèle M\M\C avec rappels

Les modèles à multiserveurs attirent beaucoup d'attention parce que la conception d'un système de files d'attente avec rappels est déduite d'importantes applications dans les systèmes de téléphonie. Une étude d'un système d'attente avec rappel peut être considérée comme complète si la distribution stationnaire du processus stochastique en question, ou au moins une dépendance analytique explicite entre les caractéristiques essentielles du système et les paramètres initiaux λ, C, θ, μ est établie. À ce jour, seuls quelques résultats peuvent être considérés comme valides pour les modèles avec rappels et multiserveurs, et ceci dans le cas d'un service exponentiel.

Considérons un groupe de C serveurs entièrement disponibles dans lequel un flux des arrivées primaires poissonnien arrive avec un taux $\lambda > 0$. Si un client primaire arrivant trouve un serveur libre, il occupe immédiatement un serveur et quitte le système après le service. Sinon, si tous les serveurs sont occupés, il entre on orbite.

Nous supposons que la durée entre deux rappels consécutifs est distribuée exponentiellement avec le paramètre θ, et les temps de service sont distribués exponentiellement avec le paramètre μ.

Le fonctionnement du système peut être décrit par le processus stochastique de Markov $\{C(t), N_0(t); t \geq 0\}$, Où $C(t)$ est le nombre de clients dans l'espace de service et $N_0(t)$ est le nombre de clients en orbite à l'instant t. Son espace d'états est donc $S = \{0, 1..., C\} \times N$. Ses taux de transition infinitésimaux $q_{(ij)(nm)}$ sont donnés par :

1. Pour $0 \leq i \leq C - 1$

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

2. Pour $i = C$

$$q_{(Cj)(n,m)} = \begin{cases} \lambda, & \text{si } (n,m) = (C, j+1), \\ C\mu, & \text{si } (n,m) = (C-1, j), \\ -(\lambda + C\mu), & \text{si } (n,m) = (i,j), \\ 0 & \text{sinon} \end{cases}$$

Les caractéristiques importantes de la qualité de service sont :

- probabilité que tous les serveurs occupés $P_C = \lim_{t \to \infty} P\ (C(t) = C)$;
- nombre moyen de clients en orbite $\overline{N}_0 = \lim_{t \to \infty} E[N_0(t)]$;
- nombre moyen de serveurs occupés $\overline{C} = \lim_{t \to \infty} E[C(t)]$.

La condition suffisante et nécessaire d'existence d'un régime stationnaire du

système est $\lambda < C\mu$. La distribution stationnaire $p_{ij} = \lim_{t \to \infty} p_{ij}(t)$ satisfait le système d'équations de Kolmogorov suivant :

$$\begin{aligned} (\lambda + i\mu + j\theta)p_{ij} &= \lambda p_{i-1,j} + (j+1)\theta p_{i-1,j+1} + (i+1)\mu p_{i+1,j}, \\ \text{si } 0 &\leq i \leq C-1, \text{ et } j \geq 0; \end{aligned} \tag{2.6}$$

$$\begin{aligned} (\lambda + C\mu)p_{Cj} &= \lambda p_{C-1,j} + (j+1)\theta p_{C-1,j+1} + \lambda p_{C,j-1}, \\ \text{si } i &= C \text{ et } j \geq 0. \end{aligned} \tag{2.7}$$

Pour les fonctions génératrices

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

ces équations deviennent

$$\begin{aligned} (\lambda + i\mu)p_i(z) + \theta z p_i'(z) &= \lambda p_{i-1}(z) + \theta p_{i-1}'(z) + (i+1)\mu p_{i+1}(z), \\ \text{si } 0 &\leq i \leq C-1, \end{aligned} \tag{2.8}$$

$$\begin{aligned} (\lambda + C\mu)p_C(z) &= \lambda p_{C-1}(z) + \theta p_{C-1}'(z) + \lambda z p_C(z), \\ \text{si } i &= C. \end{aligned} \tag{2.9}$$

Maintenant, introduisons la fonction génératrice bivariée

$$p(x,z) = \sum_{i=0}^{C} x^i p_i(z).$$

Nous supposons que le taux de service $\mu = 1$, donc les équations (2.8), (2.9)

deviennent :

$$\lambda(1-x)p(x,z) + \theta(z-x)p'_z(x,z) + (x-1)p'_x(x,z) \tag{2.10}$$
$$+\lambda x^C(x-z)p_C(z) + \theta x^C(x-z)p'_C(z) = 0.$$

En différenciant l'équation (2.10) par rapport à z, x, xx, xz, zz au point $x = 1, z = 1$, on obtient les equations suivantes :

$$\begin{aligned}
\theta\overline{N}_0 - \lambda P_C - \theta\overline{N}_C &= 0, \qquad (2.11)\\
\lambda + \theta\overline{N}_0 - \overline{C} - \lambda P_C - \theta\overline{N}_C &= 0,\\
\lambda\overline{C} + \theta p''_{xz} - p''_{xx} - \lambda C P_C - \theta C\overline{N}_C &= 0,\\
-\lambda\overline{N}_0 - \theta p''_{zz} + (1+\theta)p''_{xz} + \lambda - \theta C\overline{N}_C - \lambda C P_C + \theta p''_{czz} &= 0,\\
\theta p''_{zz} - \lambda\overline{N}_C - \theta p''_{czz} &= 0,
\end{aligned}$$

Où $\overline{N}_0 = P'_z\,(1.1), P_C = P_C(1), \overline{C} = P'_x(1.1), \overline{N}_C = \lim_{t\,\to\infty} E[N_0(t), C(t) = C] = P'_C(1)$. En éliminant de ces équations les variables $\overline{N}_C$, P''_{xz}, P''_{zz}, P''_{czz} et en tenant compte du fait que $P''_{xx}(1.1) = (Var[C(t)] + (E[C(t)]^2) - E[C(t)])$, nous obtenons

$$\overline{C} = \lambda. \tag{2.12}$$

$$\overline{N}_0 = \frac{1+\theta}{\theta}.\frac{\lambda - \lim_{t\,\to\infty} Var[C(t)]}{C-\lambda}. \tag{2.13}$$

L'équation (2.13) peut être réécrite sous une forme équivalente telle que :

$$\overline{N}_0 = \frac{1+\theta}{\theta}.\frac{\lambda + \lambda^2 - \lim_{t\,\to\infty} E[(C(t))^2]}{C-\lambda}. \tag{2.14}$$

L'équation (2.12) peut être considérée comme une variante de la formule de Little et représente l'égalité du trafic offert et de celui transporté. L'équation (2.13) donne une description partielle de la dépendance de la taille moyenne de l'orbite des paramètres du système étudié et réduit le calcul de cette mesure de performance

($\overline{N}_0$) au calcul des caractéristiques du nombre de serveurs occupés, ce qui est un problème plus simple.

De même, les moments supérieurs du nombre de clients en orbite peuvent être exprimés en termes de la distribution du nombre de serveurs occupés. Par exemple, le deuxième moment factoriel $\Phi_2 \equiv E[N_0(t).(N_0(t)-1)] = p''_{zz}(1.1)$ est donné par [9] :

$$\begin{aligned}\Phi_2 &= \frac{1+2\theta}{2\theta^2(C-\lambda)^2}\{-(2+\theta)(C-\lambda)p'''_{xxx}(1.1) \\ &\quad +[\lambda(4+\theta)(C-\lambda)+2C(1+\theta)(C-\lambda-1)]p''_{xx}(1.1) \\ &\quad -2\lambda^2C(1+\theta)(C-\lambda-1)-2\lambda^3(C-\lambda)\}.\end{aligned}$$

Une autre information intéressante sur la distribution stationnaire du nombre de serveurs occupés $P_i = \lim_{t\to\infty} P(C(t)=i)$ peut être obtenu comme suit. Mettons dans les équations (2.8), (2.9) $z=1$:

$$\begin{aligned}\lambda P_i + \theta\overline{N}_i - (i+1)P_{i+1} &= \lambda P_{i-1} + \theta\overline{N}_{i-1} - iP_i,\ 0 \le i \le C-1 \\ \lambda P_{C-1} + \theta\overline{N}_{C-1} - Cp_C &= 0,\end{aligned}$$

Où $\overline{N}_i = \lim_{t\to\infty} E[N_0(t),\, C(t)=i]$. Les équations précédentes donnent que

$$\lambda P_i + \theta\overline{N}_i - (i+1)P_{i+1} = 0,\ 0 \le i \le C-1. \tag{2.15}$$

Définissons le taux du flux des clients secondaires étant donné que le nombre de serveurs occupés est égal à i par $r_i = \frac{\theta\overline{N}_i}{P_i}$. Alors, l'équation (2.15) devient :

$$P_{i+1} = \frac{\lambda + r_i}{i+1}P_i,\ 0 \le i \le C-1.$$

Enfin

$$P_i = \frac{(\lambda + r_{i-1})...(\lambda + r_0)}{i!}P_0,\ 0 \le i \le C, \tag{2.16}$$

La probabilité P_0 s'obtient à partir de l'équation de normalisation $\sum_{i=0}^{C} P_i = 1$. En effet,

$$P_0 = [\sum_{i=0}^{C} \frac{(\lambda + r_{i-1})...(\lambda + r_0)}{i!}]^{-1}. \tag{2.17}$$

Bien que les taux r_i, $0 \leq i \leq C-1$, soient inconnus et que les équations (2.16)-(2.17) ne présentent pas de solution sous forme « close », ces dernières fournissent un aperçu du problème et seront utilisées plus tard. En outre, l'équation (2.12) permet d'avoir une relation, telle que

$$\sum_{i=0}^{C} i\frac{(\lambda + r_{i-1})...(\lambda + r_0)}{i!} = \lambda \sum_{i=0}^{C} \frac{(\lambda + r_{i-1})...(\lambda + r_0)}{i!}. \tag{2.18}$$

CHAPITRE 3

Les approches d'analyses des systémes à multiserveurs

1 Introduction

L'hétérogénies spatiale des chaines de Markov associés aux systèmes des files d'attente avec rappels et multiserveurs, et l'espace d'états infini conduit à l'absence des formules analytiques explicites pour les caractéristiques probabilistes principales de ces systèmes. La seule méthode pour obtenir des données numériques précises consiste à résoudre les équations de Kolmogorov numériquement. Mais dès que ce système d'équations est infini, il ne peut pas être résolu directement même sur ordinateur. Les transformations qui réduisent ces équations à une solution d'un petit problème fini dans le cas général ne sont pas disponibles. C'est pourquoi, nous nous adressons aux approches d'approximation.

Dans ce chapitre, nous présentons deux approches d'obtention des solutions, rencontrées dans la littérature, pour les modèles avec rappels et multiserveurs.

2 Modèles tronqués

L'approche de troncation finie de l'espace d'états est celle la plus connue. Introduite par Wilkinson [28], elle consiste à remplacer l'espace d'états infini par un autre

fini, de telle manière que la distribution stationnaire de l'état du système puisse être obtenue.

Autrement dit, le système de files d'attente avec rappels et multiserveurs initial S, où la capacité de l'orbite est illimitée, est remplacé par un système similaire $S^{(M)}$, où le nombre de clients en orbite est limité par une constante M suffisamment large. Cependant, l'implémentation d'un tel schéma peut conduire à prendre en considération un grand nombre d'états ayant des probabilités négligeables. À l'aide de la méthodologie proposée par Stepanov (1999) [27], cet inconvénient pourrait être atténué en explorant toutes les directions dans l'espace d'états considérés où les probabilités d'état diminuent puis en déterminant les bornes de troncation. Ceci nous donne la possibilité de prévoir l'erreur relative donnée des estimations des mesures de performance [31].

2.1 Description du modèle

Dans ce modèle, par opposition au modèle principal, la taille de l'orbite est bornée par une constante donnée M (choisie de manière appropriée). De ce fait, la dynamique stochastique du système peut être décrite au moyen d'un processus bivarité $\{C^{(M)}(t), N_0^{(M)}(t); t \geq 0\}$ markovien qui possède l'espace d'état $S^{(M)} = \{0, 1, ..., C\} \times \{0, 1, ..., M\}$. Ses taux de transition infinitésimaux $q^{(M)}_{(ij)(nm)}$ sont donnés par :

1. Pour $0 \leq i \leq C-1$, $0 \leq j \leq M$

$$q^{(M)}_{(ij)(nm)} = \begin{cases} \lambda, & \text{si } (n,m) = (i+1, j), \\ i\mu, & \text{si } (n,m) = (i-1, j), \\ j\theta, & \text{si } (n,m) = (i+1, j-1), \\ -(\lambda + i + j\theta), & \text{si } (n,m) = (i,j), \\ 0 & \text{sinon} \end{cases}$$

2. Pour $i = C$, $0 \leq j \leq M-1$

$$q^{(M)}_{(ij)(nm)} = \begin{cases} \lambda, & \text{si } (n,m) = (C, j+1), \\ C\mu, & \text{si } (n,m) = (C-1, j), \\ -(\lambda + C\mu), & \text{si } (n,m) = (i,j), \\ 0 & \text{sinon} \end{cases}$$

3. Pour $i = C, j = M$

$$q^{(M)}_{(ij)(nm)} = \begin{cases} C\mu, & \text{si } (n,m) = (C-1, M), \\ -C\mu, & \text{si } (n,m) = (C, M), \\ 0 & \text{sinon} \end{cases}$$

La distribution stationnaire $P^{(M)}_{ij} = \lim_{t\to\infty} P(C^{(M)}(t) = i, N^{(M)}_0(t) = j)$ satisfait le système d'équations suivant :

$$\begin{aligned} (\lambda + i\mu + j\theta)p^{(M)}_{ij} &= \lambda p^{(M)}_{i-1,j} + (j+1)\theta p^{(M)}_{i-1,j+1} + (i+1)\mu p^{(M)}_{i+1,j}, \qquad (3.1) \\ 0 &\leq i \leq C-1,\ 0 \leq j \leq M-1; \end{aligned}$$

$$\begin{aligned} (\lambda + i\mu + M\theta)p^{(M)}_{i,M} &= \lambda p^{(M)}_{i-1,M} + (i+1)\mu p^{(M)}_{i+1,M}\ , \qquad (3.2) \\ 0 &\leq i \leq C-1\ ,\ j = M; \end{aligned}$$

$$\begin{aligned} (\lambda + C\mu)p^{(M)}_{ij} &= \lambda p^{(M)}_{i-1,j} + (j+1)\theta p^{(M)}_{i-1,j+1} + \lambda p^{(M)}_{i,,j-1}\ , \qquad (3.3) \\ i &= C\ ,\ 0 \leq j \leq M-1; \end{aligned}$$

$$\begin{aligned} C\mu p^{(M)}_{CM} &= \lambda p^{(M)}_{C-1,M} + \lambda p^{(M)}_{CM-1}, \qquad (3.4) \\ i &= C\ ,\ j = M; \end{aligned}$$

qui satisfait à la condition de normalisation :

$$\sum_{i=0}^{C}\sum_{j=0}^{M} p_{ij}^{(M)} = 1. \tag{3.5}$$

2.2 Formules explicites pour les principales caractéristiques de performance :

Pour les fonctions génératrices

$$p_i^{(M)}(z) = \sum_{j=0}^{M} z^i p_{ij}^{(M)}, 0 \leq i \leq C$$

les équations (3.1)-(3.4) deviennent

$$(\lambda+i\mu)p_i^{(M)}(z)+\theta z\frac{dp_i^{(M)}(z)}{dz} = \lambda p_{i-1}^{(M)}(z)+\theta\frac{dp_{i-1}^{(M)}(z)}{dz}+p_{i+1}^{(M)}(z),\ 0 \leq i \leq C-1, \tag{3.6}$$

$$(\lambda+C\mu)p_C^{(M)}(z)-\lambda z^M p_{CM}^{(M)} = \lambda p_{C-1}^{(M)}(z)+\theta\frac{dp_{C-1}^{(M)}(z)}{dz}+\lambda z^{(M)}p_C(z)-\lambda z^{M+1}p_{CM}^{(M)}. \tag{3.7}$$

Maintenant, introduisons la fonction génératrice bivariée

$$p^{(M)}(x,z) = \sum_{i=0}^{C} x^i p_i^{(M)}(z).$$

Nous supposons toujours que le taux de service $\mu = 1$. Alors les équations (3.6), (3.7) deviennent :

$$\lambda(1-x)p^{(M)}(x,z) + \theta(z-x)\frac{\partial p^{(M)}(x,z)}{\partial z}$$

$$+(x-1)\frac{\partial p^{(M)}(x,z)}{\partial x}+\lambda x^C(x-z)p_C^{(M)}(z)$$

$$+\theta x^C(x-z)\frac{dp_C^{(M)}(z)}{dz}+\lambda z^M x^C(z-1)p_{CM}^{(M)}=0.$$

En différenciant cette équation par rapport à z, x, xx, xz, zz au point $x=1, z=1$; on obtient les équations suivantes :

$$\theta\overline{N}_0^{(M)}-\lambda P_C^{(M)}-\theta\overline{N}_C^{(M)}+\lambda p_{CM}^{(M)}=0,$$

$$\lambda+\theta\overline{N}_0^{(M)}-\overline{c}^{(M)}-\lambda P_C^{(M)}-\theta\overline{N}_C^{(M)}=0,$$

$$\theta\frac{\partial^2 p^{(M)}(1,1)}{\partial z^2}-\lambda\overline{N}_C^{(M)}-\theta\frac{d^2 p_C^{(M)}(1)}{dz^2}+\lambda M p_{CM}^{(M)}=0,$$

$$-\lambda\overline{N}_0^{(M)}-\theta\frac{\partial^2 p^{(M)}(1,1)}{\partial z^2}+(1+\theta)\frac{\partial^2 p^{(M)}(1,1)}{\partial x\partial z}$$

$$+\lambda\overline{N}_C^{(M)}-\theta C\overline{N}_0^{(M)}-\lambda C P_C^{(M)}+\theta\frac{d^2 p_C^{(M)}(1)}{dz^2}+\lambda C p_{CM}^{(M)}=0,$$

$$\lambda\overline{C}^{(M)}+\theta\frac{\partial^2 p^{(M)}(1,1)}{\partial x\partial z}-\frac{\partial^2 p^{(M)}(1,1)}{\partial x^2}-\lambda C P_C^{(M)}=\theta C\overline{N}_C^{(M)},$$

où

$$\overline{N}_0^{(M)}\equiv\lim_{t\to\infty}E[N_0^{(M)}](t)=\frac{\partial p^{(M)}(1,1)}{\partial z},$$

$$P_C^{(M)}\equiv\lim_{t\to\infty}P(C^{(M)}(t)=C)=p_C^{(M)}(1),$$

$$\overline{C}^{(M)}\equiv\lim_{t\to\infty}E[C^{(M)}(t)]=\frac{\partial p^{(M)}(1,1)}{\partial x},$$

$$\overline{N}_C^{(M)}\equiv\lim_{t\to\infty}E[N_0^{(M)}(t),C^{(M)}(t)=C]=\frac{dp_C^{(M)}(1)}{dz}.$$

En éliminant de ces équations les variables :

$$\overline{N}_C^{(M)},\frac{\partial^2 p^{(M)}(1,1)}{\partial x\partial z},\frac{\partial^2 p^{(M)}(1,1)}{\partial z^2},\frac{d^2 p_C^{(M)}(1)}{dz^2}$$

et en tenant compte du fait que

$$\frac{\partial^2 p^{(M)}(1,1)}{\partial x^2} = VarC^{(M)}(t) + (EC^{(M)}(t))^2 - EC^{(M)}(t)$$

nous obtenons :

$$\overline{C}^{(M)} = \lambda - \lambda P_{CM}^{(M)}, \tag{3.8}$$

$$\begin{aligned} \overline{N}_0^{(M)} &= \frac{1+\theta}{\theta}.\frac{\lambda + \lambda^2 - E[(C^M(t))^2]}{C-\lambda} \\ &\quad -\frac{\lambda}{\theta}.\frac{(C+1+\lambda)(1+\theta)+M\theta}{C-\lambda} P_{CM}^{(M)}. \end{aligned} \tag{3.9}$$

2.3 Un algorithme de calcul numérique de la stationnaire distribution dans le système tronqué

En utilisant la forme spécifique des équations (3.1)-(3.4), il est possible de proposer un algorithme récursif pour le calcul des probabilités $P_{ij}^{(M)}$.

On introduit de nouvelles variables $r_{ij}^{(M)}, 0 \leq i \leq C$, $0 \leq j \leq M$, comme suit :

$$r_{ij}^{(M)} = \frac{P_{ij}^{(M)}}{P_{0M}^{(M)}}.$$

Il est clair, si on trouve des variables $r_{ij}^{(M)}$, alors on peut calculer les probabilités $P_{ij}^{(M)}$ comme suit :

$$P_{ij}^{(M)} = \frac{r_{ij}^{(M)}}{\sum_{i=0}^{C}\sum_{j=0}^{M} r_{ij}^{(M)}}.$$

Les variables $r_{ij}^{(M)}$ satisfont l'ensemble d'équations suivant (obtenu à partir des équations (3.1)-(3.4) pour les probabilités $P_{ij}^{(M)}$) [30] :

$$r_{0M}^{(M)} = 1, \tag{3.10}$$

$$\begin{aligned}(\lambda + i\mu + j\theta)r_{ij}^{(M)} &= \lambda r_{i-1,j}^{(M)} + (j+1)\theta r_{i-1,j+1}^{(M)} + (i+1)\mu r_{i+1,j}^{(M)}\ , \\ \text{si } 0 &\leq i \leq C-1,\ 0 \leq j \leq M-1,\end{aligned} \tag{3.11}$$

$$\begin{aligned}(\lambda + i\mu + M\theta)r_{iM}^{(M)} &= \lambda r_{i-1,M}^{(M)} + (i+1)\mu r_{i+1,M}^{(M)}\ , \\ \text{si } 0 &\leq i \leq C-1,\ j = M,\end{aligned} \tag{3.12}$$

$$\begin{aligned}(\lambda + C\mu)r_{Cj}^{(M)} &= \lambda r_{C-1,j}^{(M)} + (j+1)\theta r_{C-1,j+1}^{(M)} + \lambda r_{C,j-1}^{(M)}, \\ \text{si } i &= C,\ \ 0 \leq j \leq M-1,\end{aligned} \tag{3.13}$$

$$C\mu r_{CM}^{(M)} = \lambda r_{C-1,M}^{(M)} + \lambda r_{C,M-1}^{(M)}. \tag{3.14}$$

Calculons les variables $r_{ij}^{(M)}$par groupes, chacun de taille $C+1$; Au début on calcule $r_{0M}^{(M)}, ..., r_{CM}^{(M)}$, ensuite $r_{0,M-1}^{(M)}, ..., r_{C,M-1}^{(M)}$ et ainsi de suite, jusqu'à ce qu'on trouve $r_{00}^{(M)}, ..., r_{C,0}^{(M)}$.

1. Mettre $j = M$.

1.2. Pour trouver le groupe $r_{0M}^{(M)}, ...r_{CM}^{(M)}$ réécrire l'équation (3.12) comme suit respectivement :

$$r_{i+1,M}^{(M)} = \frac{(\lambda + i\mu + M\theta)r_{iM}^{(M)} - \lambda r_{i-1,M}^{(M)}}{(i+1)\mu},\ 0 \leq i \leq C-1,$$

Ce qui est équivalent à

$$r_{i,M}^{(M)} = \frac{(\lambda + (i-1)\mu + M\theta)r_{i-1,M}^{(M)} - \lambda r_{i-2,M}^{(M)}}{i\mu},\ 1 \leq i \leq C.$$

D'après (3.10) $r_{0M}^{(M)} = 1$, nous pouvons calculer récursivement des variables $r_{1M}^{(M)}, ..., r_{CM}^{(M)}$.

2. mettre $j = j-1$. Calculer les variables $r_{0j}^{(M)}, ..., r_{Cj}^{(M)}$.

2.1. la dernière variable, $r_{Cj}^{(M)}$ peut être trouvé à partir l'équation (3.14) (si $j = M-1$) :

$$r_{C,M-1}^{(M)} = \frac{C\mu r_{CM}^{(M)} - \lambda r_{C-1,M}^{(M)}}{\lambda} \tag{3.15}$$

Où à partir de l'équation (3.13) avec j remplacé par $j+1$ (si $j < M-1$) :

$$r_{Cj}^{(M)} = \frac{(\lambda + C\mu) r_{C,j+1}^{(M)} - \lambda r_{C-1,j+1}^{(M)} - (j+2)\theta r_{C-1,j+2}^{(M)}}{\lambda}. \tag{3.16}$$

2.2 pour trouver les variables $r_{0j}^{(M)}, ..., r_{C-1,j}^{(M)}$, nous utilisons l'équation (3.11) pour $i = 0, ..., C-1$. Cet ensemble d'équations à la forme :

$$\alpha_i x_{i-1} + \beta_i x_i + \gamma_i x_{i+1} = \delta_i, \ 0 \leq i \leq C-1, \tag{3.17}$$

où

$$\begin{aligned} x_i &= r_{ij}^{(M)}, \\ \alpha_i &= -\lambda, \\ \beta_i &= \lambda + i\mu + j\theta, \\ \gamma_i &= -(i+1)\mu, \\ \delta_i &= (j+1)\theta r_{i-1,j+1}^{(M)}, \end{aligned}$$

et les valeurs

$$x_{-1} = 0, \ x_C = r_{C,j}^{(M)}$$

sont connues.

Ces équations, qui sont des équations aux différences, produisent des solutions numériques comme les équations différentielles de second ordre. L'algorithme le plus efficace pour leur résolution peut être trouvé dans plusieurs livres sur les méthodes numériques (voir, par exemple [16]). D'après cet algorithme, on calcule en premier lieu les variables B_i, D_i, $0 \leq i \leq C-1$, par des formules récursives

$$\begin{aligned} B_0 = \beta_0, \qquad & D_0 = \delta_0, \\ B_i = \beta_i - \tfrac{\alpha_i \gamma_{i-1}}{\beta_{i-1}}, \qquad & D_i = \delta_i - \tfrac{\alpha_i D_{I-1}}{B_{i-1}}, \ 1 \leq i \leq C-1, \end{aligned}$$

puis à partir de l'équation

$$B_i x_i + \gamma_i x_{i+1} = D_i,\ 0 \leq i \leq C-1,$$

On calcule récursivement (dans l'ordre inverse) les inconnues $x_{C-1}, ..., x_0$. Dans notre cas, cela donne la procédure suivante :

– calculer les variables B_{ij}, D_{ij}, $0 \leq i \leq C-1$, à l'aide des équations

$$\begin{aligned} B_{0j} &= \lambda + j\theta, \\ B_{ij} &= \lambda + i\mu + j\theta - \frac{\lambda i}{B_{i-1,j}},\ \text{pour } 1 \leq i \leq C-1; \\ D_{0_j} &= 0, \\ D_{ij} &= (j+1)\theta r^{(M)}_{i-1,j+1} + \frac{\lambda D_{i-1,j}}{B_{i-1,j}},\ \text{pour } 1 \leq i \leq C-1. \end{aligned} \tag{3.18}$$

– puis calculer récursivement $r^{(M)}_{ij}$, $0 \leq i \leq C-1$, (dans l'ordre inverse, commencer par $r^{(M)}_{Cj}$ connu à partir de l'étape 2.1) à l'aide d'équation

$$r^{(M)}_{ij} = \frac{D_{ij} + (i+1)\mu r^{(M)}_{i+1,j}}{B_{ij}},\ i = C-1, C-2, ...1, 0.$$

3. La répétition de l'étape 2 tant que $j \geq 0$ permet d'obtenir toutes les variables $r^{(M)}_{ij}$(ce qui est, successivement $j = M-2,\ M-3, ..., 0$).

Puisque $P^{(M)}_{ij} = r^{(M)}_{ij}.P^{(M)}_{0M}$, alors on aura :

$$P^{(M)}_{0M} = \frac{1}{\sum_{i=0}^{C}\sum_{j=0}^{M} r^{(M)}_{ij}}.$$

A présent, on peut calculer les probabilités $P^{(M)}_{ij} = r^{(M)}_{ij}.P^{(M)}_{0M}$ et les principales caractéristiques probabilistes du système tronqué $S^{(M)}$ [29] :

(a) La probabilité de blocage :

$$P_C^{(M)} = \sum_{j=0}^{M} r_{Cj}^{(M)}.P_{0M}^{(M)};$$

(b) Le nombre moyen de serveurs occupés :

$$\overline{C}^M = \sum_{i=0}^{C}\sum_{j=0}^{M} ir_{ij}^{(M)}.P_{0M}^{(M)};$$

(c) Le nombre moyen de clients en orbite :

$$\overline{N}_0^{(M)} = \sum_{i=0}^{C}\sum_{j=0}^{M} jr_{ij}^{(M)}.P_{0M}^{(M)}.$$

3 Approche d'extrapolation de valeur

L'approche présentée ci-dessus fournit une solution numérique à la distribution stationnaire P_{ij} des chaînes de Markov à temps continu. Elles utilisent les équations de Kolmogorov pour calculer les caractéristiques de performance souhaitées. Le probléme à résoudre est de trouver la solution du système d'équations suivant :

$$P_{ij}\sum_{s'\neq s} q_{s's} = \sum_{s'\neq s} q_{s's}P_{ij},\ \forall s \in S;$$

$$\sum_{s} P_s = 1,$$

où $q_{s's}$ représente le taux de transition de l'état $s = (i, j)$ à $s' = (k, l)$.

Dans ce qui suit, nous considérons une approche alternative, appelée Extrapolation de Valeur. L'approche en question est basée sur les outils de la théorie des processus de Markov Décisionnels [25] et également sur le principe de troncation. L'idée est de trouver la valeur relative moyenne en résolvant les équations de Ho-

ward, écrites pour un espace d'états tronqué. Sa particularité consiste dans le fait qu'au lieu d'une simple troncation, les valeurs d'état relatives à l'extérieur de l'espace d'états tronqué sont estimées par l'extrapolation polynomiale des valeurs d'état relatives à l'intérieur de l'espace tronqué en question. De cette manière, nous obtenons un système fermé ainsi que les résultats précis avec de faibles niveaux de troncation.

3.1 Processus de Markov Décisionnels

Un processus de Markov Décisionnel (PMD) [25] est défini comme $\{S, A, P, R\}$ où S est un ensemble d'états, A est un ensemble d'action, P est une fonction de transition d'état et R est une fonction de revenu. L'état du système peut être contrôlé, en choisissant les actions a de A, influençant ainsi les transitions d'état. La fonction de transition $P : S \times S \times A \rightarrow R_+$ spécifie le taux de transition entre deux états lorsqu'une certaine action est prise à l'état d'origine [19].

La première caractéristique de la technique Extrapolation de Valeur est la nécessité de la définition d'une fonction de revenu qui doit être une fonction de l'état du système, qui est $r(s)$. À la suite de la définition de la fonction de revenu pour chaque état, en régime stationnaire, on peut introduire un taux de revenu moyen du processus entier comme $\overline{r} = \sum_{(i,j)\epsilon S} P_{ij} r(i,j)$. Dans la technique Extrapolation de Valeur, la fonction de revenu R doit être définie de sorte que le revenu moyen résultant $\overline{r}$ coïncide avec la métrique de performance souhaitée. Une fois que nous avons défini le cadre PMD ainsi que la fonction de revenu, nous sommes en mesure de définir les valeurs d'état relatives. Il est évident qu'après avoir exécuté une action dans un état $s \epsilon S$, le système collectera un revenu pour cette action $(r(s))$, mais à mesure que le nombre de transitions augmente, le revenu moyen collecté converge vers $\overline{r}$ [22]. La valeur d'état relative $v(s)$ montre la différence entre le revenu total cumulé lorsque le processus débute à l'état s et le revenu total cumulé par le processus, le taux moyen de revenu étant $\overline{r}$:

$$v(s) = E[\int_0^\infty (r(S(t) - r)dt / S(0) = s].$$

Les équations de Howard mettent en relation les revenus. les valeurs d'état relatives et les taux de transition de la manière suivante [18] :

$$r(s) - r + \sum_{s'} q_{ss'}(v(s') - v(s)) = 0;\ \forall s \epsilon S.$$

Les équations de Howard qui correspondent au système étudié M/M/C avec rappels sont :

$$r(i,j) - \overline{r} + \lambda(v(i+1,j) - v(i,j)) + i\mu(v(i-1,j) - v(i,j)) + j\theta(v(i+1,j-1) - v(i,j)) = 0,$$

$$0 \leq i \leq C-1;$$

$$r(i,j) - \overline{r} + \lambda(v(C,j+1) - v(C,j)) + C\mu(v(C-1,j) - v(C,j)) = 0,$$

$$i = C.$$

Comme nous pouvons observer le nombre d'états est infinie parce que j peut prendre n'importe quelle valeur dans Z_+. Nous devons donc tronquer l'espace d'état S à $S^{(M)}$:

$$S^{(M)} = \{s = (i,j);\ 0 \leq j \leq C;\ 0 \leq j \leq M\}. \tag{3.21}$$

En effet, il y aura autant d'équations d'Howard que le nombre d'états, $\left|S^{(M)}\right|$. Le nombre d'inconnus sera les $|S^{(M)}|$ valeurs d'état relatives plus le revenu attendu $\overline{r}$, c'est-à-dire $|S^{(M)}| + 1$ inconnues. Cependant, comme seules les différences dans les valeurs relatives apparaissent dans les équations de Howard, nous pouvons poser $v(0) = 0$. Par conséquent on aura un système linéaire d'équations résoluble ayant le même nombre d'équations que le nombre d'inconnues.

3.2 Ajustement polynomial

L'Extrapolation de Valeur considère les valeurs d'état relatives en dehors de $S^{(M)}$ qui apparaissent dans les équations de Howard comme une extrapolation de certaines valeurs relatives correspondant aux états se trouvant à l'intérieur de $S^{(M)}$, En résumè,

l'objectif de l'Extrapolation de Valeur est de trouver une fonction d'extrapolation qui s'adapte à certains points de $S^{(M)}$ de telle sorte qu'elle se rapproche également de points en dehors de $S^{(M)}$, Il est important de choisir une fonction d'extrapolation qui fait que les équations de Howard forment un système fermé d'équations linéaires [22]. Les fonctions les plus courantes qui remplissent cette condition sont les polynômes. Nous pouvons utiliser tous les états de $S^{(M)}$ dans la procédure d'implantation globale ou ce qui est le plus couramment utilisé, seulement un sous-ensemble (S_f) de leur localisation. Par souci de simplicité, dans la description qui suit, on suppose qu'il existe une application W à partir de l'ensemble bidimensionnel d'états vers un ensemble à une dimension, par exemple les nombres réels : $W : \widehat{S}_f \rightarrow R$. Done, l'application W traite des états comme s'ils étaient des valeurs réelles $w = W(s)$. Le choix de W dépendra fortement des états dans lesquels nous voulons extrapoler sa valeur d'état relative. Notons également que la fonction d'extrapolation $f(w)$ et l'ensemble $S_f^{(M)}$ doivent être choisis de façon à ce que les paramètres de $f(w)$ aient des valeurs non ambigües, c'est-à-dire que dans le cas du choix d'un polynôme comme fonction d'extrapolation, le nombre de points différents dans $S_f^{(M)}$ doit être égal ou supérieur au nombre de coefficients dans le polynôme. En général, la procédure du calcul des coefficients du polynôme consiste à minimiser l'erreur quadratique minimale [15] :

$$E = \sum_{w \epsilon W} (f(w) - v(w))^2. \tag{3.22}$$

Alors les valeurs optimales pour les coefficients du polynôme α_i peuvent être calculées en résolvant les équations :

$$\frac{\partial E}{\partial a_i} = 0, \quad \forall i. \tag{3.23}$$

Dans notre cas, nous utilisons autant de points que le nombre de paramètres du polynôme d'interpolation, de sorte que la procédure d'ajustement est une interpolation polynomiale ordinaire et $E = 0$, c'est-à-dire que tous les points considérés se trouveront dans la courbe du polynôme. Dans ce cas, le problème peut être formulé comme suit : Ètant donné un ensemble de $n = (W(S_f^{(M)})) = \left|S_f^{(M)}\right|$ points

$(w_0, v(w_0), ..., (w_{n-1}, v(w_{n-1}))$ on peut déterminer un polynôme de degré $(n-1)$ de sorte que $f(w_i) = v(w_i)$, pour $i = 0, ..., n-1$, où

$$f(w) = a_0 + a_1 w + a_2 w^2 + ... + a_{n-1} w^{n-1}.$$

Le polynôme d'interpolation satisfait aux n équation linéaires suivantes :

$$f(w_i) = a_0 + a_1 w_i + a_2 w_i^2 + ... + a_{n-1} w_i^{n-1} = v(w_i),\ i = 0, ..., n-1.$$

Ces dernières, sous forme matricielle, se présentent de la manière suivante :

$$Aa = \begin{bmatrix} 1 & w_0 & \cdots & w_0^{n-1} \\ 1 & w_1 & \cdots & w_1^{n-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & w_{n-1} & \cdots & w_{n-1}^{n-1} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{bmatrix} = \begin{bmatrix} v(w_0) \\ v(w_1) \\ \vdots \\ v(w_{n-1}) \end{bmatrix} = b.$$

La matrice des coefficients de ce système A est une matrice de Vendermonde, dont le déterminant est non nul et donc A est inversible. Ainsi, il existe toujours une solution unique au système linéaire d'équations considéré ou, de façon équivalente, il existe un polynôme unique qui passe par tous les n points. Cependant, les matrices de Vandermonde sont souvent mal conditionnées, surtout si certaines w_i sont très proches, donc la procédure pour calculer le polynôme est également mal conditionnée [17]. Il est important de noter que l'unicité du polynôme d'interpolation ne signifie pas qu'il ne peut pas être écrit dans une base différente de la base standard [20]. Plus concrètement dans ce travail, nous avons utilisé la base de Lagrange. Pour le problème d'interpolation considéré, le polynôme est une combinaison linéaire

$$L(w) = \sum_{j=0}^{n-1} v(w_i) l_j(w),$$

où

$$l_j(w) = \prod_{\substack{i=0 \\ i \neq j}}^{n-1} \frac{w - w_i}{w_j - w_i} = \frac{w - w_0}{w_j - w_0} ... \frac{w - w_{j-1}}{w_j - w_{j-1}} \frac{w - w_{j+1}}{w - w_{j+1}} ... \frac{w - w_{n-1}}{w - w_{n-1}}.$$

Pour le problème étudié, nous aurons une équation de Howard dans laquelle apparaît $v(C, M+1)$, c'est-à-dire une valeur relative d'un état qui n'appartient pas à $S^{(M)}$, Nous devons donc approximer la valeur relative $v(C, M+1)$ à l'aide de certaines valeurs d'état relatives des états appartenant à $S^{(M)}$. Il est important de souligner que pour l'extrapolation de $v(C, M+1)$, nous n'utilisons que des états de la forme $s = (C, j)$ avec j variables. De ce fait, on définit l'application w comme $w((C, j)) = j$. De plus, on utilise un polynôme de degré $(n-1)$ qui interpole les n points de$S_f = \{s_i = (C, Q-i) / i = 0, ..., n-1\}$ et alors $W(S_f) = \{w_i = Q - i / i = 0, ..., n-1\}$. En effet,

$$\begin{aligned}
w_0 &= M \rightarrow v(w_0) = v(C, M); \\
w_1 &= M - 1 \rightarrow v(w_1) = v(C, M-1); \\
&\vdots \\
w_j &= M - j \rightarrow v(w_j) = v(C, M-j); \\
w_{n-1} &= M - (n-1) \rightarrow v(w_{n-1}) = v(C, M-(n-1)).
\end{aligned}$$

De cette façon, la forme générale d'extrapoltion lors de l'utilisation d'un polynôme de degré $(n-1)$ est :

$$v^{(n)}(C, M+1) = L^{(n)}(M+1) = \sum_{j=0}^{n-1} v(C, M-j) l_j(M+1).$$

Par exemple, dans le cas de l'extrapolation linéaire $n = 2$, on utilise $(M, v(C, M))$

et $(M-1, v(C, M-1))$:

$$\begin{aligned} v^{(2)}(C, M+1) &= L^{(2)}(M+1) = v(C, M)l_0(M+1) + v(C, M-1)l_1(M+1) \\ &= v(C, M)\frac{(M+1)-(M-1)}{M-(M-1)} + v(C, M-1)\frac{(M+1)-M}{(M-1)-M} \\ &= 2v(C, M) - v(C, M-1). \end{aligned}$$

Suivant la procèdure similaire, nous obtenons les relations suivantes pour $n = 3$ et $n = 4$:

$$\begin{aligned} v^{(3)}(C, M+1) &= 3v(C, M) - 3v(C, M-1) + v(C, M-2); \\ v^{(4)}(C, M+1) &= 4v(C, M) - 6v(C, M-1) + 4v(C, M-2) - v(C, M-3). \end{aligned}$$

Finalement

$$v^{(n)}(C, M+1) = \sum_{k=0}^{n-1}(-1)^k \binom{n}{k+1} v(C, M-k),$$

où n est le nombre de coefficients pris pour les polynômes de Lagrange.

3.3 Fonction Revenu

Par définition, $r(s)$ est le taux de revenu obtenu lorsque le système est dans l'état s. Par conséquent, nous devons définir le revenu comme la caractéristique de performance que nous voulons calculer. De plus, les entrées $r(s)$ dans les équations de Howard doivent être correctement définies. Le tableau 3.1 donne plusieurs exemples sur la façon dont $r(s)$ peuvent être déterminés pour obtenir certaines caractéristiques de performance. Par exemple, nous choisissons la probabilité de blocage et nous définissons la fonction de revenu comme étant 1 dans les états où un client est bloqué, c'est-à-dire $r(C, j) = 1$, pour tout $0 \leq j \leq M$, et 0 dans le reste des états $r_{(i,j)} = 0$, pour $0 \leq i \leq C-1$ et pour tout $0 \leq j \leq M$.

TAB. 3.1 – Définition de la fonction revenu.

Probabilité de blocage	P_C	$r(i,j) = 1$, pour $i = C$ et $j \geq 0$ $r(i,j) = 0$ sinon
Nombre moyen de clients en orbite	$\overline{N}_0$	$r(i,j) = j$, pour $0 \leq i \leq C$ et $0 \leq j \leq M$ $r(i,j) = 0$ sion
Nombre moyen de serveurs occupés	$\overline{C}$	$r(i,j) = i$, pour $0 \leq i \leq C$ et $0 \leq j \leq M$ $r(i,j) = 0$ sinon

3.4 Effet de l'extrapolation de la valeur dans les équations d'Howard

Dans notre problème, nous n'aurons qu'à remplacer $v(C, M+1)$ par sa valeur approximative dans l'équation de Howard qui correspond à l'état $(C, M+1)$. Par exemple, si nous utilisons l'extrapolation linéaire $n = 2$, cette équation devient

$$\begin{aligned} & r(C,M) - r + v(C,M)(-\lambda - C\mu) + \lambda v(C,M+1) + C\mu\nu(C-1,M) \\ = \; & r(C,M) - r + v(C,M)(\lambda - C\mu) + C\mu v(C-1,M) + v(C,M-1) = 0. \end{aligned}$$

Comme $v(C, M+1)$ n'apparaît plus dans les équations de Howard, nous avons un système linéaire de $(C+1) \times (M+1)$ équations ayant le même nombre d'inconnues. Ce système peut être exprimé sous forme matricielle pour des raisons de simplicité. Par conséquent, le système peut être réécrit comme $xT = b$, où x est un vecteur avec les inconnues $(C+1) \times (M+1)$ ($\overline{r}$ et les valeurs d'état relatives $v(s)$) et b comporte les taux de revenu négatifs pour les différents états [14] :

$$\begin{aligned} x &= [\overline{r}, v(0,1), ...v(0,M), v(1,0), ...v(C,M)); \\ b &= [-r(0,0), -r(0.1), ..., -r(C,0), -r(C,M)]; \end{aligned}$$

La matrice T représente la matrice des coefficients et peut être construite en mettant tous les éléments de la première rangée de la matrice T_0 égale à -1. La

matrice T_0 est donnée par :

$$T_0 = \begin{bmatrix} A_1^0 & A_0^0 & \cdots & O & O \\ A_2^1 & A_1^1 & \cdots & O & O \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & \cdots & A_1^{C-1} & A_0^{C-1} \\ O & O & \cdots & A_2^C & A_1^C \end{bmatrix},$$

où les sous-matrices sont définies comme

$$A_0^i = (i+1)\mu I,\ 0 \leq i \leq C-1;$$

$$A_2^i = \begin{bmatrix} \lambda & \theta & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 2\theta & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & M\theta \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{bmatrix},\ 1 \leq i \leq C;$$

$$A_1^i = \begin{bmatrix} \alpha & 0 & 0 & \cdots & 0 \\ 0 & \alpha-\theta & 0 & \cdots & 0 \\ 0 & 0 & \alpha-2\theta & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \alpha-M\theta \end{bmatrix}, \text{ pour } \alpha = -\lambda - i\mu,\ 0 \leq i \leq C-1.$$

Lorsque $i = C$, et s'il s'agit de l'extrapolation linéaire $n = 2$ et quadratique $n = 3$, nous obtenons respectivement

$$A_1^C = \begin{bmatrix} \beta & 0 & \cdots & 0 & 0 \\ \lambda & \beta & \cdots & 0 & 0 \\ 0 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & \beta & -\lambda \\ 0 & 0 & \cdots & \lambda & \lambda - C\mu \end{bmatrix},$$

$$A_1^C = \begin{bmatrix} \beta & 0 & \cdots & 0 & 0 \\ \lambda & \beta & \cdots & 0 & 0 \\ 0 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \lambda \\ 0 & 0 & \cdots & \beta & -3\lambda \\ 0 & 0 & \cdots & \lambda & 2\lambda - C\mu \end{bmatrix},$$

où $\beta = -\lambda - C\mu$. En génerwal, si l'extrapolation est fait avec $n \leq M+1$ points, la matrice A_1^C est donnée par :

$$A_1^C = \begin{bmatrix} \beta & 0 & \cdots & 0 & \lambda c_M^{(n)} \\ \lambda & \beta & \cdots & 0 & \lambda c_{M-1}^{(n)} \\ 0 & \lambda & \cdots & 0 & \lambda c_{M-2}^{(n)} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \lambda c_2^{(n)} \\ 0 & 0 & \cdots & \beta & \lambda c_1^{(n)} \\ 0 & 0 & \cdots & \lambda & -\lambda - C\mu + \lambda c_0^{(n)} \end{bmatrix},$$

où

$$c_l^{(n)} = \begin{cases} (-1)^l C_n^{l+1}, & \text{si } l < n \\ 0, & \text{si } l \geq n \end{cases}.$$

Nous soulignons que la taille de la matrice T ne dépend pas du degré de polynôme utilisé pour effectuer l'extrapolation ; seule la dernière colonne de la matrice A dépend de l'ajustement polynômial. Cette caractéristique à l'avantage suivant : il n'y aura pas de différence dans le coût de calcul lorsqu'on utilise un degré d'extrapolation plus élevé.

L'inconvénient principal de la technique d'extrapolation de la valeur est que cette technique est seulement capable de calculer une mesure de performance chaque fois que nous résolvons le système. Néanmoins, nous pouvons surmonter cet inconvénient de la suivante. D'une manière générale, la solution du système $xT = b$ peut être obtenue en utilisant la matrice inverse de T en faisant $x = bT^{-1}$. Notons également

que le choix d'une autre mesure de performance n'affectera que les valeurs dans b. Par conséquent, le calcul d'une seconde mesure de performance augmente seulement les dépenses de calcul par le coût du produit bT^{-1}, car le reste du processus de calcul de la matrice inverse T^{-1} n'est résolu qu'une seule fois [14].

CHAPITRE 4

Performance de la méthode extrapolation de valeur

1 Introduction

Dans ce chapitre, nous examinons efficacité et effectivité de l'approche Extrapolation de Valeur en comparant avec une autre approche basée sur la troncature directe de l'espace d'états infini. En outre, nous étudions l'influence des paramètres du modèle considéré sur sa performance.

2 Comparaison de EV et TF

Ce paragraphe traite des résultats numériques rendus disponibles par les procédures données en 3.2 et 3.3 : Troncature Finie et Extrapolation de Valeur. Nous comparons la valeur minimal du niveau de troncature M de la probabilités de blocage et le nombre moyen de clients on orbite (lors de l'application de EV et par les résultats obtenue par Falin on appliquons la TF [9]).

Dans [9], pour $C = 5$, $\lambda = 4$ et $\theta = 20$ à calculer la probabilité de blocage et le nombre moyen de client en orbite pour différente valeurs de M. Avec les mêmes paramètres ($C = 5$, $\lambda = 4$ et $\theta = 20$) ont a calculer la probabilité de blocage et le nombre moyen de client en orbite avec la méthode d'Extrapolation de la Valeur.

TAB. 4.1 – Comparaison entre TF et EV

M	Méthode	P_C	$\overline{N}_0$
5	VE	0.5331	0.9797
	TF	0.4539	1.6240
10	VE	0.5345	1.7057
	TF	0.5102	2.1396
15	VE	0.5347	2.0893
	TF	0.5268	2.3119
20	VE	0.5347	2.2687
	TF	0.5321	2.3692
25	VE	0.5347	2.3456
	TF	0.5338	2.3881
30	VE	0.5347	2.3773
	TF	0.5344	2.3944

A partir de la Table 4.1, on voit que la méthode d'Extrapolation de Valeur présente des performances beaucoup plus élevées en termes de précision

Remarque 2.1 *Dans ce paragraphe, pour résoudre le système d'équations de Howard, le logiciel abordable et personnalisable Matlab a été utilisé.*

3 Etude de la performance du système

Maintenant, nous étudions l'effet de taux de rappels sur les mesures de performance $\overline{N}_0$, P_C et $\overline{C}$. Les résultats numériques sont obtenus en appliquant la méthode Extrapolation de Valeur et à l'aide du logiciel Matlab.

A partir de la Table 4.2, on peut voir que l'augmentation du taux de rappels θ entraîne une amélioration sensible de $\overline{N}_0$ mais elle déteriore la probabilité de blocage P_C.

Le scénario considéré est :

– Le nombre de serveurs $C = 20$;

– Le taux d'arrivée $\lambda = 15$;

TAB. 4.2 – Nombre moyen de client en orbite et Probabilité de blocage par rapport à theta

θ	P_C	$\overline{N}_0$
1	0,0821	1,6387
10	0,1229	0,6247
100	0,1527	0,4930

– Le taux de service $\mu = 1$;
– Le niveau de troncature $M = 15$;

4 Conclusion

Nous concluons que l'approche, basée sur les outils de la théorie de Processus de Markov Décisionnels (EV), peut être utilisée avec succès lors de la résolution du modèle à multiserveurs avec rappels. En outre, l'approche en question permet d'obtenir des résultats précis avec de faibles niveaux de troncature, ceci réduit considérablement le coût de calcul.

Conclusion générale

Dans ce travail, nous nous sommes intéressés à l'étude des modèles de type M/M/C avec rappels. Vu l'impossibilité de fournir une solution exacte aux caractéristiques essentielles du système, nous avons fait appel aux approximations. En effet, nous avons considéré deux approches alternative : la première permet d'obtenir une solution numérique approchée à la distribution stationnaire de l'état du système, puis, à l'aide des outils fondamentaux de la théorie des probabilités, calculer les mesures de performance ; tandis que la seconde utilise les moyens fournis par la théorie des Processus de Markov Décisionnels et permet d'exprimer et par la suite de déterminer les indices de performance en question en termes d'une certaine métrique, appelée valeur d'état relative. Toutefois, les deux approches sont basées sur le principe de troncature de l'espace d'états infini.

Notre apport est théorique et pratique. Dans un premier temps, nous avons décrit le modèle général d'un système de files d'attente classique ainsi que d'un systèm de files d'attene avec rappels et d'un modèle markovien à multiserveurs. Nous avons discuté les approches les plus intéressantes et les plus performantes pouvant être appliquées lors de l'analyse des modèles avec rappels et multiserveurs.

Dans un deuxième temps nous avons obtenu des solutions pour le modèle M/M/C avec rappels à l'aide de la méthode de Troncature Finie et de la méthode d'Extrapolation de Valeur. En comparant les performances des méthodes en question, nous

pouvons conclure que la seconde méthode est plus appropriée pour la résolution du modèle étudié, en particulier sous les niveaux élevés de la congestion. Elle permet d'obtenir les résultats précis avec de faibles niveaux de troncature.

Les résultats obtenus dans ce travail permettent d'envisager un nouveaux perspective de recherchee qu'est :

- Définir une nouvelle approche de résolution des modèles à multiserveurs basée sur les notions et principes issus de la théorie des jeux.

Annexe

Variable aléatoire

Une variable aléatoire X est une application qui, à chaque événement élémentaire de l'univers, associe un nombre réel. Définir une loi de probabilité de X, c'est associer à chaque résultat x_i un nombre p_i positif tel que la somme des p_i soit égale à 1.

Fonctions génératrice

X est une variable aléatoire à valeur entiéres non négative. La fonction génératrice de X est définie par :

$$F(z) = E(z^k) = \sum_{n=0}^{\infty} p_n z^n \quad \text{Où} \quad p_n = p(X = n) \ ,n \ \epsilon \ \mathbb{N}$$

Et z est une variable complexe. On vérifie immédiatement que $f(z)$ est définie au moins pour $|z| \leq 1$ et que

$f(0) = p_0$ et $f(1) = 1$.

Processus stochastique

Un processus stochastique $X(t)_{t\ \epsilon T}$, est une fonction du temps dont la valeur à chaque instant dépend de l'issue d'une expérience aléatoire. A chaque instant $t\ \epsilon T$, $X(t)$ est donc une variable aléatoire.

Un processus stochastique peut donc être considéré comme une famille de variable aléatoires (généralement non indépendantes). L'ensemble des temps T peut être discret ou continu. $X(t)$ définit l'état du processus à un instant donné t. A nouveau, l'ensemble E des valeurs que peut prendre le processus à chaque instant est appelé espace d'état et peut, de même que T, être discret (finit ou infini) ou continu.

Processus de naissance et de mort

Les processus de naissance et de mort sont des processus stochastiques à temps continu et à espace d'états discrets $n = 0, 1, 2.....$ Ils sont sans mémoire, et à partir d'un état donné n, seules les transitions vers l'un des états voisins $(n+1)$ *et* $(n-1)$ avec $n \geq 1$ sont possibles. On parle alors de « naissances » et de « morts ». Ces processus sont utilisés pour modéliser les systèmes d'attente et l'évolution de populations.

Définition 4.1 *soit un processus stochastique* $\{X(t); t \geq 0\}$ *à états discrets* $n\ \epsilon\ \mathbb{N}$, *et homogène dans le temps, c'est a dire :*

$$P(X(t+s) = j \diagup X(s) = i) = p_{ij}(t),\ \textit{ne dépend pas de } s$$

Le processus $\{X(t); t \geq 0\}$ *est un processus de naissance et de mort s'il satisfait les conditions Suivantes :*

$$\begin{cases} p_{i,i+1}(\triangle t) = \lambda_i \triangle t + o(\triangle t) & (i \geq 0) \\ p_{i,i-1}(\triangle t) = \mu_i \triangle t + o(\triangle t) & (i \geq 1) \\ p_{i,i} = 1 - (\lambda_i + \mu_i)\triangle t + o(\triangle t) & (i \geq 0) \\ p_{i,j}(\triangle t) = o(\triangle t) & |i-j| \geq 2 \end{cases}$$

Les coefficients positifs λ_i et μ_i sont appelés taux de transition, plus particulièrement taux de naissance (ou de croissance) pour λ_i et taux de mort (ou de décroissance) pour μ_i.

Régime transitoire

Pour calculer les probabilités d'états $P_n(t) = P(X(t) = n)$ nous pouvons écrire d'aprés le théorème des probabilités totales et pour $n > 1$.

$$\begin{aligned} p_n(\triangle t + t) &= \sum_{t\geq 0} p_i(t)p_{in}(\triangle t) \\ &= p_{n-1}(t)p_{n-1,n}(\triangle t) + p_n(t)p_{n,n}(\triangle t) + p_{n+1}(t)p_{n+1,n}(\triangle t) \\ &= \lambda_{n-1}p_{n-1}(t)\triangle t + (1-(\lambda_n+\mu_n))p_n(t)\triangle t + \mu_{n+1}p_{n+1}(t)\triangle t + o(\triangle t) \\ \Longrightarrow \quad & \frac{p_n(t+\triangle t) - p_n(t)}{\triangle t} = \lambda_{n-1}p_{n-1}(t) - (\lambda_n+\mu_n)p_n(t) + \mu p_{n+1}(t) + \frac{o(\triangle t)}{\triangle t} \end{aligned}$$

On faisant tendre $\triangle t$ vers 0, on trouve :

$$p'_n(t) = \lambda_{n-1}p_{n-1}(t) - (\lambda_n+\mu_n)p_n(t) + \mu_{n+1}p_{n+1}(t),\ \forall n \geq 1..... \tag{1.1}$$

pour $n = 0$, on pose $\mu_0 = 0$,

$$p'_0(t) = -\lambda_0 p_0(t) + \mu_1 p_1(t)..... \tag{1.2}$$

On a $p_{n-1}(t) = 0$ si $n = 0$

$$\begin{aligned} p'_0(t) &= -\lambda_0 p_0(t) + \mu_1 p_1(t) \\ p'_n(t) &= \lambda_{n-1}p_{n-1}(t) - (\lambda_n+\mu_n)p_n(t) + \mu_{n+1}p_{n+1}(t),\ \forall n \geq 1 \end{aligned}$$

Les équations (1.1) et (1.2) sont connues sous le nom "équations différentielles de Kolmogorov "elles permettent de calculer les probabilités d'état $p_n(t)$ si l'on connait les condition initiales du processus.

Régime stationnaire

Lorsque $t \to +\infty$ les limites $p_n = \lim_{t\to\infty} p_n(t)$ existent et sont indépendantes de l'état initial du processus, on a alors $\lim_{t\to\infty} p'_n(t) = 0$. Ceci se traduit par les équations dites de balances.

$$\begin{aligned} 0 &= -(\lambda_n + \mu_n)p_n + \lambda_{n-1}p_{n-1} + \mu_{n+1}p_{n+1} \quad n \geq 1 \\ 0 &= \lambda_0 p_0 + \mu p_1 \qquad\qquad\qquad\qquad\qquad\qquad\quad n = 0 \end{aligned}$$

aux quelles il faut ajouter la condition $\sum_{n=0}^{\infty} p_n = 1$.

En addionnant les $(n+1)$ premières equations, on obtient : $\mu_{n+1} = \lambda_n p_n$

d'où

$$p_n = \frac{\lambda_0\lambda_1....\lambda_{n-1}}{\mu_1\mu_2.....\mu_n}p_0$$

$$\sum_{n\geq 0} p_n = 1 \Longrightarrow p_0 = \frac{1}{\sum_{n\geq 0}\prod_{J\geq 0}\frac{\lambda_{j-1}}{\mu_j}}$$

BIBLIOGRAPHIE

[1] A. Aboul-Hassan, S. Rabia and A. Kadry. Analytical study of a discrete time retrial queue with balking customers and early arrival scheme. *Alexandria Engineering Journal*, 44(6) : 911-917, 2005.

[2] A. Aissani. Asurvey on retrial queueing models. *Actes des Journées Statistiques Appliquées*, U.S.T.H.B., Alger, 1-11, 1994.

[3] A. O. Allen. *Probability, Statistics, and Queueing Theory with Computer Science Applications*. Second edition, Academic Press, New York (First edition : 1978), 1990.

[4] J.R. Artalejo and A. Gomez-Corral. *Retrial Queueing Systems* : A Computational Approach. Springer, 2008.

[5] M. L. Chaudhry and J. G. C Templeton. *A First Course in Bulk Queues*. John Wiley and Sons, New York, 1983.

[6] J.W. Cohen. Basic problems of the telephone traffic theory and the influence of repeated calls. *Philips Telecom*, 18(2) : 49-100, 1957.

[7] M.J. Domenech-Benllach, J.M. Gimenez- Guzman, V. Pla, J. Martinez-Bauset and V. Casares-Giner. On the efficient solution of a multiserver system with two reattempt orbits. *Mathematical and Computer Modelling*, 51 : 1082-1092, 2010.

[8] V. Fabrice. *Les files d'attentes : Modélisation et évaluation de perforamances de réseaux*, Technical report, Université Lyon, France, 2003.

[9] G.I. Falin et J.G.C. Templeton. *Retrial Queues*. Chapman and Hall,1997.

[10] G.I. Falin. A survey of retrial queues. *Queueing Systems*, 7 : 127-168, 1990.

[11] G.I. Falin and J.R. Artalejo. A finite source retrial queue. *Europeen Journal of Operational Research*, 108 : 409-424, 1998.

[12] G.I. Falin and J.G.C Templeton. *Retrial queues*. Chapman and Hall, 1997.

[13] A. Federgruen et L.Green. Queueing systems with service interruptions. *Research Working*. Columbia University,30 : 5-84, 1984.

[14] J.M. Gimenez-Guzman , M.J. Domenech-Belloch, V. Pla, V. Casares-Giner and J. Martinez-Bauset. Value extrapolation technique to solveretrial queues : a comparative perspective. *ETRI Journal*, 30 : 492-494, 2008.

[15] J.M. Gimenez-Guzman, M.J. Domenech-Benlloch, V. Pla, J. Martinez-Bauset and V. Casares-Giner. Efficient method to approximately solve retrial systems with impatience. *Journal of Applied Mathematics*, 2012 : 1- 18, 2012.

[16] S.K. Godunov and V.S. Ryabenkii. *Difference Schemes*. Northe Holland, 47-50, 1987.

[17] P. Goatin. *Analyse Numérique*. Cours de $1^{ére}$ année- Var ISITV, Université du Sud Toulon.

[18] R.A. Howard. Dynamic Programming and Markov Processes, *The Technology Press of MIT*, Cambridge, Mass, USA, 1960.

[19] O. Ibe. *Markov Processess for Stochastic Modeling*. Elseiver Academic Press, 2009.

[20] F. Jedrzejewski, *Introduction aux méthodes numériques*. Springer, 2005.

[21] L. Kosten. *On the influence of repeated calls in the theory of probabilities of blocking*. De Ingenieur, 59 : 1-125, 1947.

[22] J. Leino and J. Virtamo. An approximate method for calculating performance measures of Markov processes. *Preceedings of VALUETOOLS*, 2006.

[23] M.F. Neuts. *Matrix-Geometric Solutions in Stochastic Models*. The Johns Hopkins University Press, Baltimore, MD, 1981.

[24] M.F. Neuts and B.M. Rao. Numerical investigation of a multiserver retrial model . *Queueing Systems*, 7 : 169-190, 1990.

[25] M.L. Puterman. *Markov Decision Processes : Discrete Stochastic Dynamic Programming*. Wiley series in Probability and Statistics, Wiley, 2008.

[26] J. Raj. *Introduction to queueing theory*. Technical report. Washington university, 2008.

[27] S.N. Stepanov. Markov models with retrials : the calculation of stationary performance measures based on the concept or truncation. *Mathematical and Computer Modelling*, 30 : 207- 228, 1999.

[28] R.I. Wilkinson. *Theories for toll traffic engineering in USA*. Bell System Tech.J., 35 : 421-507, 1956.

[29] N. Zidani et N. Djellab. L'approximation de la distribution stationnaire de l'état des systems des files d'attente avec rappels et multiserveurs. *Journées Jeunes Chercheurs*, Université Badji Mokhtar- Annaba, 30 septembre - 1 Octobre 2014.

[30] N. Zidani et N. Djellab. Calcul numerique des caractéristiques stationnaires du modèle tronqué. *Journées Nationales sur les Mathématiques Appliquées "JNMA'14"*, Université 20 août 1955 Skikda, 26-27 Novembre 2014.

[31] N. Zidani et N. Djellab. Approximation of multiserver retrial queues by truncation technique. 6^{th} *Operational Research Practice in Africa Conference*, University of Sciences and Technology Houari Boumediene, 20-22 Avril 2015.