République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Faculté des Sciences Exactes
Département de Recherche Opérationnelle

Mémoire de fin de cycle en Recherche Opérationnelle

Option : Modélisation Mathématique et Technique de Décision

Thème

Analyse et simulation de la communication RF dans les réseaux de capteurs via les files d'attente avec priorité

Présenté par :

BELKACEMI Hamza
BOUZIDI Kamel

Devant le jury composé de :

Promotrice :	Mme Lekadir Ouiza
Président :	Mr Khimoum Nouredine
Examinatrice :	Mme Hocine Safia
Examinatrice :	Melle Outamazirt Assia

Juin 2016-2017

Remerciements

Louange A Dieu, le miséricordieux, sans Lui rien de tout cela n'aurait pu être.

Nous tenons tout d'abord *à* remercier **Mme LEKADIR Ouiza** , pour l'honneur qu'elle nous a fait en acceptant de nous encadrer. Ces conseils précieux ont permis une bonne orientation dans la réalisation de ce modeste travail.
Nous tenons également à remercier les membres du jury **Mr Khimoume Nouredine, Mme Hocine Safia et Melle Outamazirt Assia** pour l'honneur qu'il nous ont fait en acceptant de juger ce travail, et d'avoir consacrer leurs temps pour sa lecture. Un grand remerciement à Mme Hakmi Sedda pour ces conseils et son aide.
Nous tenons à exprimer notre profonde gratitude à l'ensemble du corps enseignant qui a contribué à notre formation.
Enfin nous tenons à rendre hommage à toutes nos familles et tous nos amis pour le soutien qu'ils nous ont apportés durant toutes ces années d'études.

Dédicaces

A cœur veillant, rien d'impossible ;
A conscience tranquille, tout est accessible ;
Quand il y a la soif d'apprendre.
Tout vient à point à qui sait attendre.
Les études sont avant tout notre unique et seul atout.
Souhaitant que le fruit de nos efforts fournis jour et nuit
Nous mènera vers le bonheur fleuri.

Je dédie ce modeste travail :

A celle qui m'a donné la vie, le symbole de tendresse, qui s'est sacrifiée pour mon bonheur et ma réussite, à ma mère.

A mon père, école de mon enfance, qui a été mon ombre durant toutes les années des études, Que Dieu les garde et les protége.

A ma très chére sœur.

A mes très chers frères.

A toute ma famille.

BOUZIDI Kamel

Table des matières

Table des figures

Introduction générale

Les origines de la théorie des files d'attente remontent à 1909 à l'époque où A. K. Erlang a posé les bases dans ses recherches sur le trafic téléphonique. Ses travaux ont par la suite été intégrés à la recherche opérationnelle. Malheureusement, les publications sur la théorie des files d'attente ont adopté un langage de plus en plus mathématique, ce qui a freiné son utilisation.

La plupart des systèmes rencontrés dans différents domaines technologique peuvent être représentés par des modèles de files d'attente. C'est le cas par exemple des réseaux de télécommunications. En effet, afin d'analyser le comportement de ces systèmes, évaluer et optimiser leurs performances, il faut d'abord les représenter par des modèles mathématiques qu'on peut obtenir avec des outils de modélisation tels que la théorie des files d'attente. Un modèle typique de files d'attente nécessite la définition des processus d'inter-arrivées et les durées de service des clients, la taille de la file, ainsi que la discipline de service. Tous ces paramètres sont indiqués dans la notation de Kendall.

Dans certains systèmes, on est souvent amené à imposer des priorités d'utilisation du service, d'ailleurs c'est se qu'on constatera dans ce mémoire qui portera sur l'analyse de la communication RF (radio Fréquence) dans les réseaux de capteurs. En effet, le modèle adéquat qu'on a proposé pour la modélisation de la communication RF (radio Fréquence) dans les réseaux de capteurs est un système de files d'attente avec rappels et vacances avec une nouvelle généralisation de la priorité relative et absolue .

Dans un système d'attente avec priorité, l'admission des unités (clients, appels, messages, pannes,...) au service se fait selon deux possibilités, soit la priorité est de nature relative ou de nature absolue. Dans le premier cas, un client de priorité supérieure ne peut en aucun cas interrompre le service du client de priorité inférieure, il doit attendre jusqu'à ce que le service du client de priorité inférieure soit fini pour qu'il soit servi. Par contre, dans le second cas, un client de priorité supérieure a le droit d'interrompre le service d'un client de priorité inférieure pour être servi, et ce dernier sera réadmis lorsqu'il n'y a pas en attente d'unités de priorités supérieures. Le client interrompu ou bien conserve le bénéfice de la portion du service déjà effectué (priorité absolue conservatrice), ou bien reprend le service dès le début (priorité absolue non conservatrice) [41].

Les deux catégories présentent toutefois plusieurs inconvénients dans les applications

pratiques. Sous la catégorie non préemptive, les clients de priorité supérieure peuvent avoir Attendre même lorsque le service d'un client de priorité inférieure vient de commencer, alors que dans les disciplines préventives, le service de client non prioritaire presque terminé mais il peut être interrompu au raison de l'arrivée d'un client de priorité élevé(causant un très grand retard supplémentaire). Pour que les deux situations sont évitées autant que possible, nous proposons une discipline de planification prioritaire dans laquelle nous introduisons un paramètre γ qui est défini comme la fraction du temps de service qui doit être écoulé pour que le service d'un client à priorité inférieure ne sera plus interrompu lorsqu'un client de priorité supérieure arrive au système.
Autrement dit, le rapport du temps de service écoulé d'un client de priorité bas à l'arrivée d'un client de priorité supérieure à son temps total de service est comparé avec le paramètre introduit γ.

Si ce rapport est inférieur à γ, alors le service de client de priorité bas est interrompu , sinon, son service est complété avant de commencer avec un client de haut priorité. Notez que les deux disciplines préemptives et non préemptives sont deux cas spéciaux de la nouvelle discipline définie, à savoir :
Si $\gamma = 1$, alors en retrouve dans le cas préemptive,
Si $\gamma = 0$, alors en retrouve dans la priorité non préemptive.

Plusieurs situations peuvent être modélisés par des systèmes de file d'attente avec rappels et vacances . Le premier classe est se caractérise par le fait qu'un client qui arrive est trouve le service occupé quitte le système, soit définitivement ou rappeler ultérieurement. Le modèle de file d'attente avec rappels occupe une situation intermédiaire entre le modèle d'Erlang et le modèle classique d'attente FIFO. Les systèmes de file d'attente avec rappels ont trouvé leur première application lors de la modélisation du service d'abonnés dans un centrale téléphonique.
Pour la deuxième classe qui est très intéressant que le premier au sens l'importance, il traite les systèmes dans lesquels le serveur reste oisif quand la file d'attente est vide, en dit dans ce cas que le serveur est en vacance. Durant cette période, le serveur occupé pour les taches supplémentaire. La notion de vacance utilisée dans le but d'améliorer les performances du système.
La première publication concernant les systèmes d'attente avec priorité absolue est l'article de White et Christie [14]. Par la suite sont apparus les articles de Stephan [12], Miller [13] et Jaiswal [17].
La priorité relative a été introduite par Cobham [3]. Cette discipline a été étudiée par plusieurs chercheurs : Holley[16], Dressin et Reich [20], ... Une bibliographie sur le sujet peut être trouvée dans le livre de Jaiswal [18].
Les systèmes d'attente avec priorité peuvent être considérés comme des systèmes d'attente

à serveur non fiable. Les points communs entre un système d'attente avec priorité absolue et un système d'attente à serveur non fiable ont été étudiés pour la première fois par White et christie [14], Keilson [15], Gaver [19] et autres chercheurs. Dans les systèmes d'attente avec priorité relative, la panne est prise en considération après que le client interrompu ait terminé son service. Ce modèle est connu sous le non "Breakdown postponable interrompu". Il a été étudié par Keilson et Gaver [15]. D'un autre coté D.Aissani dans l'article [11] a considéré le système $M_2/G_2/1$ avec priorité relative comme un modèle de refus pour lequel, la réparation de l'élément défaillant, peut être reporté jusqu'au moment de la fin de service de la demande sur l'appareil.
Des résultats analytiques ont pu être obtenus pour certains systèmes particuliers (le système $M_2/M_2/1$ avec priorité absolue [4] et $M_2/M_2/1$ avec priorité relative [3]). Cependant, même dans ces cas, la complexité des formules analytiques ne permet pas de les exploiter dans la pratique. C'est le cas de la transformée de Laplace ou de la fonction génératrice qui ne sont pas disponibles sous formes explicites. C'est pour cela que, lors de la modélisation d'un système réel, on est souvent amené à remplacer les éléments stochastiques réels mais compliqués gouvernant le système, par d'autres éléments plus simples. Ces derniers sont supposés être, dans un certain sens, proches des éléments réels, Le modèle ainsi utilisé représente une "idéalisation" du système réel.
Ce mémoire est structuré comme suit : Après cette introduction, nous donnons certaines définitions et concepts relatifs à la théorie des files d'attente markoviennes et semi markoviennes, avec leurs caractéristiques et propriétés. Nous présentons dans le deuxième chapitre une description des systèmes des files d'attente avec priorité, rappels et vacance.
Le troisième chapitre, contient une présentation de la technique de simulation a événements discrets permettant I' imitation artificielle d'un phénomène d'attente sur un ordinateur, l'avantage mis en application ici est celui d'un outil d'aide a la validation d'une approche analytique d'évaluation des résultats. Les résultats de notre mémoire sont présentés dans le quatrième chapitre, où nous implémentons des algorithmes sur les priorités (relative, absolue et $\gamma-$priorité), afin de l'utiliser pour la construction d'un programme approprié qui peut décrire notre problème fidèlement.

Chapitre 1

Notions générales sur les files d'attente

1.1 Introduction

Les files d'attente peuvent modéliser différents aspects de la vie moderne que nous rencontrons durant nos activités quotidiennes, par exemple l'attente des clients devant un guichet d'une banque, l'attente des malades aux différents services des hôpitaux, l'attente des requêtes dans les réseaux de télécommunication, ...
Ce chapitre comporte deux parties relatives à la théorie des files d'attente. Dans la première, nous allons survoler certains concepts de base de cette théorie, alors que dans la seconde nous allons présenter quelques systèmes de file d'attente classiques avec leurs caractéristiques et propriétés. Le cas particulier du système avec priorité (relative et/ou absolue) sera détaillé pour qu'il puisse être utilisé dans les autres chapitres.

1.2 Le formalisme des files d'attentes

La théorie des files d'attente s'attache à modéliser et analyser de nombreuses situations différentes en apparences, mais qui relèvent néanmoins du schéma descriptif général suivant : des clients arrivent aléatoirement à un système comportant un ou plusieurs serveurs auxquels ils vont demander un service. Ce service sera exécuté durant une durée de temps aléatoire, dite 'durée de service', après avoir été servis (ce qui suppose un arrêt chez un ou plusieurs serveurs selon le cas), les clients quittent le système. Illustrons cette description générale par quelques exemples spécifiques :

Exemple : (Agence bancaire) :
Ici, les serveurs sont les guichets de l'agence. Typiquement, tous les guichets offrent le même service et chaque client ne devra donc visiter qu'un seul guichet.

Exemple : (Parking) :
Les clients sont les véhicules qui cherchent à stationner (plutôt que les occupants, de ces véhicules), les serveurs sont les emplacements de parking, et la durée de service est la durée pendant laquelle chaque véhicule reste stationné.

Définition 1.1. Un système de files d'attente général peut être vu comme un boite noire dans laquelle les clients arrivent suivant un processus quelconque, séjournent pour recevoir un ou plusieurs service et finalement quittent le système. Ce système pourra être composé d'une file simple ou d'un ensemble de files appelé réseau de files d'attente.

Donc, un système de files d'attente est l'abstraction mathématique d'un sujet qu'on peut décrire par les éléments suivants :

- Le flot des arrivées des clients,
- La source des clients,
- Le comportement du client,
- La loi de la durée de service de chaque client,
- La discipline de service,
- Le nombre de serveurs,
- La capacité de la file.

Une représentation graphique d'une file d'attente classique est donnée par la figure (1.1) suivante :

FIG. 1.1 – **Système de file d'attente.**

► Le flot des arrivées des clients :
D'habitude, on suppose que les temps inter-arrivées sont indépendants et identiquement distribués. En général, le flot des arrivées des clients est poissonnien, ce qui revient à dire que la distribution du temps des inter-arrivées est exponentielle. Les clients peuvent arriver individuellement ou par groupes, un exemple d'arrivée par groupes est un poste police au niveau d'une frontière où les passagers ainsi que leurs bagages sont soumis au contrôle.
► La source des clients :
La source peut être finie ou infinie selon le modèle réelle qu'elle modélise. Cependant, les

modèles réels sont souvent à source limitée mais on suppose souvent qu'elle est infinie quand sa capacité est assez grande.

► **Le comportement du client :**

Certains clients peuvent être et vouloir attendre pendant longtemps. Par contre d'autre s'impatientent et quittent après un bout de temps. C'est le cas par exemple d'une centrale téléphonique où les clients raccrochent quand ils ont à attendre longtemps avant qu'une ligne ne soit disponible pour rappeler ultérieurement.

► **La durée de service :**

En général, on suppose que les durées de service sont indépendant, identiquement distribuées et indépendantes des temps des inter arrivées, ce qui n'est toujours pas le cas, Par exemple, le temps de traitement des machines au niveau d'un système de production peut s'élever une fois le nombre de taches à exécuter devient trop grand.

► **La discipline de service :**

Les clients peuvent être servis individuellement ou par groupe. Cependant, plusieurs possibilités existent quant à l'ordre selon lequel ils seront servis.

Les principales disciplines de service sont :

1. **FIFO (First In First Out) :** les entités sortent dans l'ordre suivant lequel elles sont entrées, cette discipline est la plus utilisée.
2. **LIFO (Last In First Out) :** la dernière entité dans la file est la première à être servie, c'est le cas de la pile au niveau des ordinateurs.
3. **Random :** toutes les entités ont la même probabilité d'être servies en premier.
4. **Prioritaire :** les entités sont servies suivant un attribut qui leur est associé, par exemple l'entité ayant le plus court temps de traitement d'abord.

► **Le nombre de serveurs :**

Il peut être égale à l'unité ou plus selon la nature du service à fournir.

► **La capacité de la file :**

Dans pas mal de cas, la file est supposée infinie. Cependant, il n'est pas rare de rencontre des situations dans les quelles elle est finie (par exemple le cas d'une salle d'attente).

Pour la classification des systèmes d'attente, on a recours à la notation symbolique introduit par Kendall au début des années cinquante. Cette notation comprend quatre symboles rangés dans l'ordre suivant :

$$A/B/s/N$$

où :

- A = distribution des temps entre deux arrivées successives,

- B = distribution des durées de service,
- s = nombre de postes de service en parallèle,
- N = capacité du système.
- On peut toutefois faire abstraction du dernier symbole lorsque $N = \infty$.

Pour spécifier les distributions A et B, les symboles suivants sont utilisés :

$\surd$ M : Distribution Markovienne (exponentielle),
$\surd$ E_k : Distribution d'Erlang d'ordre k,
$\surd$ H_k : Distribution hyper-exponentielle,
$\surd$ G : Distribution générale,
$\surd$ D : Distribution déterministe,
⋮

1.2.1 Notion de classes de clients

Une file d'attente peut être parcourue par différentes classes de clients. Ces différentes classes se distinguent par le fait d'avoir :

- des processus d'arrivées différents,
- des temps de service différents,
- un ordonnancement dans la file d'attente en fonction de leurs classes.

Ainsi, pour définir une file multi-classes, il y a lieu de préciser pour chaque classe de clients : son le processus d'arrivée, la distribution du temps de service associée, ainsi que l'ordre suivant le quel les clients des différentes classes sont servis.
Avant de passer à l'analyse mathématique des systèmes de file d'attente, définissons les deux notions fondamentales en théorie des files d'attente qui sont : "Processus stochastique et la chaîne de Markov".

Définition 1.2. [**Processus stochastique**]
Un processus stochastique $\{X_t\}_{t\in T}$ est une famille de variables aléatoires avec, le temps T peut être discret ou continue, $X(t)$ est une variable aléatoire définisse l'état du processus à chaque instant$t \in T$ donné. L'ensemble noté E des valeurs que peut prendre le processus à chaque instant est appelé espace d'états et peut de même que T, soit discret ou continue, en fonction de T et E, on peut classifier les processus stochastique comme suit :

1. Processus à temps discret et à espace d'états discret,
2. Processus à temps continue et à espace d'états discret,
3. Processus à temps continue et à espace d'états discret,
4. Processus à temps continue et à espace d'états continue.

Définition 1.3. [Chaîne de Markov] Une suite $\{X_t, t = 0, 1, 2, ...\}$ est dite possédant une propriété de Markov, ou est une chaîne de Markov, si pour chaque instant t et pour chaque états de X on a :

$$P[X_{t+1} = j/X_0 = a_0, X_1 = a_1, ..., X_t = i] = P[X_{t+1} = j/X_t = i.]$$

Cette propriété signifie qu'étant donné l'ensemble des états passés et présent du système, la probabilité d'un états futur quelconque de ce système est indépendante de son états de passé et dépend seulement de son états actuel. On parlera dans ce cas de processus sans mémoire.
Les probabilités conditionnelles $P[X_{t+1} = j/X_t = i]$ sont dites des probabilités de transitions et sont notés $P^t_{i,j}$. Elle ne dépendra plus de temps si on a :$P^t_{i,j} = P^s_{i,j}$, $\forall t$ et $\forall s$ et dans ce cas les $P^t_{i,j}$ seront notés tout simplement $P_{i,j}$ constituent ainsi une matrice carrée d'ordre n, si n est le nombre des états possibles dans lequel peut se trouver le processus.

Propriété 1.1. Les éléments $P_{i,j}$ vérifient les propriétés suivantes :

$$\begin{aligned} &1) \quad 0 \leq P_{i,j} \leq 1, \quad \forall i, j \\ &2) \quad \sum_{j=1}^{n} P_{i,j} = 1, \quad i = 1, 2, ... \end{aligned}$$

Si les deux conditions sont vérifier dans une matrice, elle est dite matrice stochastique.

1.3 Analyse mathématique des systèmes de files d'attente

L'étude mathématique d'un système d'attente se fait le plus souvent par l'introduction d'un processus stochastique défini de façon appropriée. En général, on s'intéresse au nombre X(t) de clients se trouvant dans le système à l'instant t ($t \geq 0$) en fonction des quantités qui définissent la structure du système, on cherche à calculer :
• Les probabilités d'états $P_n(t) = P(X(t) = n)$ qui définissent le régime transitoire du processus $\{X(t)\}_{t\geq 0}$, les probabilités $P_n(t)$ doivent évidemment dépendre de l'état initial ou de la distribution initiale du processus.
• le régime stationnaire du processus stochastique, défini par

$$P_n = \lim P_n(t) = P(X(+\infty) = n), \quad n = 0, 1, 2, ...$$

A partir de la distribution stationnaire du processus $\{X(t)\}_{t\geq 0}$, il est possible d'obtenir d'autres caractéristiques d'exploitation du système.

1.3.1 Modélisation des systèmes de files d'attente

Plusieurs variantes existent pour la modélisation selon la nature et le comportement du système, on distingue deux catégories de modèles en files d'attente, les modèles Markoviens et non Markoviens.
Si pour les premiers, la propriété d'absence de mémoire permet une grande facilité dans l'étude, il n'en est pas de même pour les modèles non markoviens. Cependant, on dispose de plusieurs méthodes, qui permettent de rendre ces derniers markoviens moyennant certaines transformation.

1.3.2 Modèles markoviens

Ils caractérisent les systèmes dans lesquels les deux quantités stochastiques principales le temps des inter-arrivées et la durée de service sont des variables indépendantes exponentiellement distribuées(les deux distributions sont Markoviens).
La propriété d'absence de mémoire de la loi exponentielle facilite l'étude de ces modèles, l'étude mathématique de tels systèmes se fait par l'introduction d'un processus stochastique approprié, ce processus est souvent le processus de naissance et de mort $\{X(t)\}_{t\geq 0}$ défini par le nombre de clients dans le système à l'instant t.
l'évolution temporelle du processus Markovien $\{X(t)\}_{t\geq 0}$ est complètement défini grâce à la propriété d'absence de mémoire.

► **Processus de naissance et de mort :**
Le processus d'état stochastique $\{n(t)\}_{t\geq 0}$ est un processus de naissance et de mort si, pour chaque $n = 0, 1, 2, \ldots$ il existe des paramètres λ_n et μ_n (avec $\mu_0 = 0$) tels que, lorsque le système est dans l'état n, le processus d'arrivée est poissonnien de taux λ_n et le processus de sortie est poissonnien de taux μ_n.Pour pouvoir étudier ce processus, on devrait d'abord envisager l'étude des processus suivants

► **Processus de naissance :**
On qualifier un processus de naissance,si ce dernier est caractérisé par l'apparition d'un individu au sein d'une population, selon une certaine loi.
un processus de naissance est dit homogène, si la probabilité d'apparition d'un individu pendant l'intervalle Δt, sachant qu'il existe déjà k individus au sein du système, est donnée par :

$$\lambda_k \Delta t + o(\Delta t) \;\; avec \;\; o(\Delta t) \;\; est \;\; infiniment \;\; petit.$$

Elle est indépendante de la position de Δt sur l'axe de temps.
Si on écarte le cas où deux apparitions simultanées ou plus peuvent avoir lieu en même temps, alors on pourrait dire que la probabilité qu'il n'y aucune apparition dans l'intervalle

Δt sachant qu'il y en a déjà k, est égale à :

$$1 - \lambda_k \Delta t + o(\Delta t)$$

Notons à présent par N_t le nombre d'apparitions durant l'intervalle $[0, t]$, et soit $P_n(t)$ la probabilité qu'à l'instant n l'effectif soit n.
D'où $P_{i,j}(t)$ est la probabilité qu'à l'instant t l'effectif est j sachant qu'il y avait déjà i individus dans le système.
$P_{i,j}(t) = 0$ si $j < i$ (on suppose qu'il n'y a eu aucune disparition).
D'autre part : $P_{i,j}(t) = P_{0,n}(t) = P_{1,n+1}(t) = ...$
Le graphe et la matrice de transition décrivant le processus pourrait être schématisés respectivement comme suit :

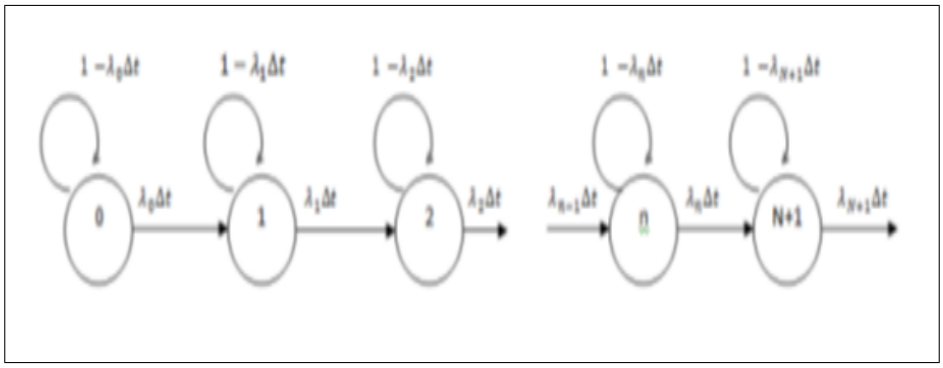

FIG. 1.2 – **Le graphe de transition du processus de naissance.**

$$M = \begin{pmatrix} 1-\lambda_0\Delta t & \lambda_0\Delta t & 0 & 0 & ... & ... & . & . & . \\ 0 & 1-\lambda_1\Delta t & \lambda_1\Delta t & 0 & ... & ... & . & . & . \\ ... & ... & ... & ... & ... & ... & . & . & . \\ 0 & 0 & 0 & ... & 1-\lambda_{n-1}\Delta t & \lambda_{n-1}\Delta t & . & . & . \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \end{pmatrix}$$

Remarquons que les états dans cette matrice(en lignes ou colonne) sont $0, 1, 2, ..., n-1, n, ...$ etc.
Considérons le vecteur des probabilités d'états P_n aux instant t et $t + \Delta t$, d'après la définition de la matrice de transition des chaines de Markov, on peut écrire :

$$[P_0(t + \Delta t), P_1(t + \Delta t), ..., P_n(t + \Delta t)] = [P_0(t), P_0(t), ..., P_0(t)] * M$$

Ou de façon explicite :

$$\begin{aligned} P_0(t+\Delta t) &= P_0(t)(1-\lambda_0\Delta) \\ P_1(t+\Delta t) &= \lambda_0\Delta t P_0(t) + (1-\lambda_1\Delta t)P_1(t) \\ &\vdots \\ P_n(t+\Delta t) &= \lambda_{n-1}\Delta t P_{n-1}(t) + (1-\lambda_n\Delta t)P_n(t) \end{aligned}$$

Considérons cette dernière relation, on remarque qu'elle pourrait être écrite sous la forme :

$$\frac{P_n(t+\Delta t) - P_n(t)}{\Delta t} = \lambda_{n-1}P_{n-1}(t) - \lambda_n P_n(t)$$

Si Δt tend vers 0, le nombre gauche tend vers $P_n^{'}$, ce qui nous donne :

$$P_n^{'}(t) = \lambda_{n-1}P_{n-1}(t) - \lambda_n P_n(t) \quad \forall n = 1, 2, ...$$

Pour $n = 0$, on peut déduire que :

$$\frac{P_0(t+\Delta t) - P_0(t)}{\Delta t} = -\lambda_0 P_0(t)$$

Si on suppose que le système était vide, alors $P_0(0) = 1$ et $P_n(0) = 0$, $\forall n \geq 1$.
Les équations différentielles ainsi définie sont dites des équations de Chapman kolmogorov.

► **Processus de mort :**

Ce processus caractérise le phénomène de disparition au seine d'une population.
Remarquons que les hypothèses qui peuvent présenter ce processus sont très voisines de celles déjà vues dans l'étude du processus de naissance.
Particulièrement on pose $N_0 = N > 0$ et on suppose que la probabilité pour qu'une disparition ait lieu entre t et $t + \Delta t$ sachant qu'à t, il y avait k individus au sein de système, est donnée par $\mu_k\Delta t$, où μ_k est le taux de disparition.
D'où, on peut écrire :

$$\begin{aligned} P(N_{t+\Delta t-N_t} = -1/N_t = k) &= \mu_k\Delta t + o(\Delta t) \\ P(N_{t+\Delta t-N_t} = 0/N_t = k) &= 1 - \mu_k\Delta t + o(\Delta t) \end{aligned}$$

On constate que μ_k doit être positif, sauf pour $k = 0$ où $\mu_0 = 0$.
Le graphe décrivant le processus de mort, pourrait être schématisé comme suit :

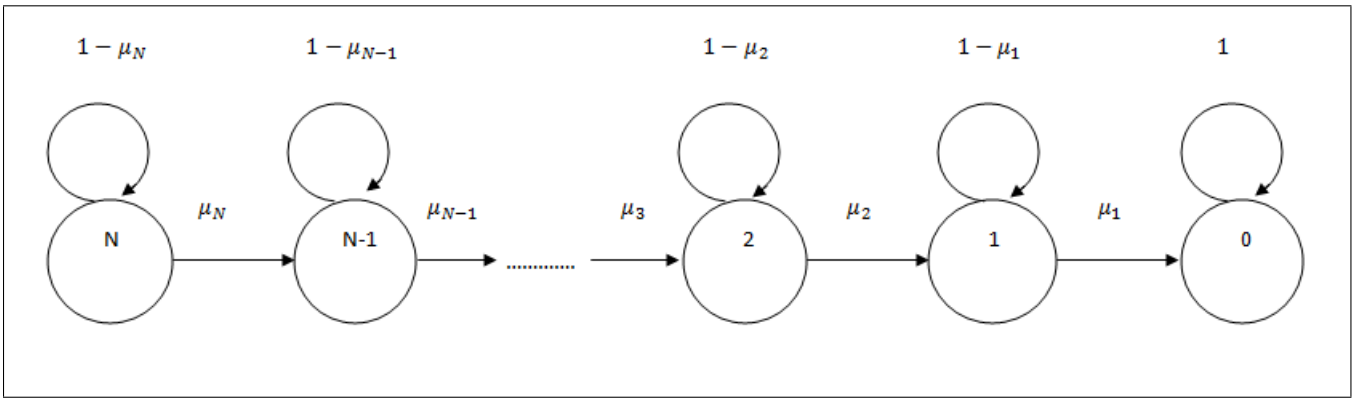

FIG. 1.3 – **Graphe de transition du processus de mort.**

Les équations différentielles associes à ce processus, dont les états de transitions sont $S = (0, 1, ..., N)$ sont donnés par :

1. $P_{i,i-1}(\Delta t) = \mu_i \Delta t + o(\Delta t)$ *avec* $i \in 1, 2, ..., N$
2. $P_{i,i}(\Delta t) = 1 - \mu_i \Delta t + o(\Delta t)$ *avec* $i \in 0, 1, 2, ..., N$
3. $P_{i,j}(\Delta t) = 0 \ \ \forall \mid i - j \mid > 1, \ \ \forall (i, j) \in S^2.$

La troisième équation pour dire qu'il y a pas de possibilité d'avoir plus d'une apparition dans une population. En peut associer la matrice des transitions suivantes :

$$M = \begin{pmatrix} 1 & 0 & 0 & . & . & . & . \\ \mu_1\Delta t & 1-\mu_1\Delta t & 0 & . & . & . & . \\ 0 & \mu_2\Delta t & 1-\mu_2\Delta t & 0 & . & . & . \\ 0 & 0 & \mu_3\Delta t & 1-\mu_3\Delta t & 0 & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & \mu_{N-1}\Delta t & 1-\mu_{N-1}\Delta t & 0 \\ . & . & . & . & . & \mu_N\Delta t & 1-\mu_N\Delta t \end{pmatrix}$$

1.3.3 Processus de naissance et de mort

Ce processus est la fusion des deux processus précédents : le processus de naissance avec le processus de mort.

On doit supposer que pendant l'intervalle de temps élémentaire Δt, il ne peut avoir q'un seul évènement, soit une naissance ou bien un mort.
Désignons par N_t l'effectif au sein de système à la date t. Les probabilités de transitions

$P_{i,j}(\Delta t) = P_{N_{t+\Delta t}=j/N_t=i}$ doivent vérifier les conditions suivantes :

$$\begin{aligned}
P_{i,i+1}(\Delta t) &= \lambda_i \Delta t + o(\Delta t), \quad \forall i \in 0, 1, ... \\
P_{i,i-1}(\Delta t) &= \mu_i \Delta t + o(\Delta t), \quad si \quad i \in 1, 2, ... \\
P_{i,i}(\Delta t) &= 1 - (\lambda_i + \mu_i)\Delta t + o(\Delta t), \quad \forall i \in 0, 1, ... \\
P_{i,j}(\Delta t) &= 0 \quad si \quad \mid i - j \mid > 1 \quad \forall (i,j) \in 0, 1, ...
\end{aligned}$$

Les λ_i et μ_i sont strictement positifs, sauf $\mu_0 = 0$, où λ_i est le taux de naissance si i individus existent déjà dans le système, et μ_i est le taux de mort au sein du même système.

Le graphe de transition correspond à ce processus sera :

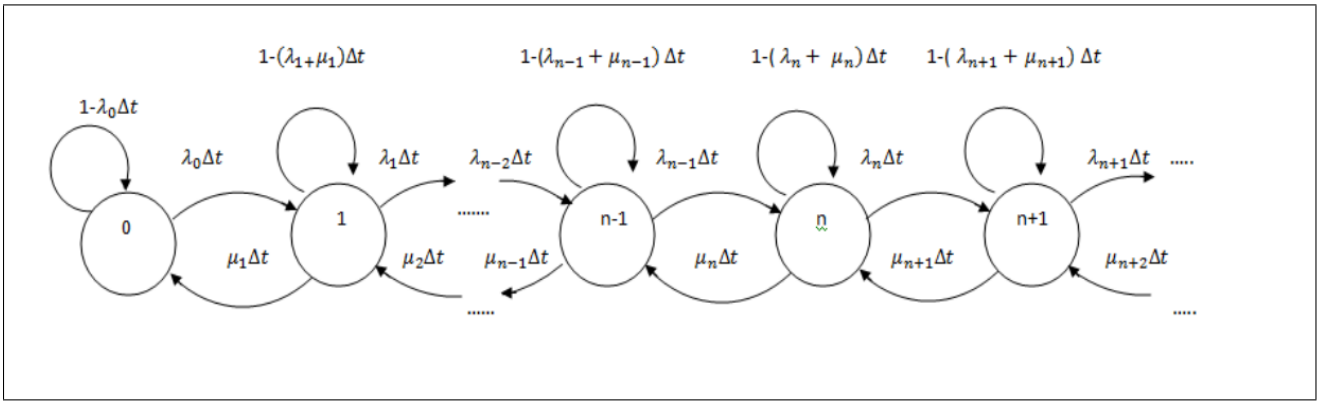

FIG. 1.4 – **Le graphe de transition du processus de naissance et de mort.**

Auquel en peut associer la matrice de transition suivante, notée M dont l'espace d'états est $S = 0, 1, ..., n, n+1, ...$

$$M = \begin{pmatrix}
1 - \lambda_0\Delta t & \lambda_0\Delta t & 0 & . & . & . & . & . & . \\
\mu_1\Delta t & 1 - (\lambda_1 + \mu_1)\Delta t & 0 & . & . & . & . & . & . \\
0 & \mu_2\Delta t & 1 - (\lambda_2 + \mu_2)\Delta t & \lambda_2\Delta t & 0 & . & . & . & . \\
. & 0 & \mu_3\Delta t & 1 - (\lambda_3 + \mu_3)\Delta t & \lambda_3\Delta t & 0 & . & . & . \\
. & . & . & . & . & . & . & . & . \\
. & . & . & . & . & . & . & . & .
\end{pmatrix}$$

En égalisant entre le flux entrant dans une états avec le flux sortant, on trouve les équations différentielle associe à ce processus de naissance et de mort :

$$\begin{aligned}
P_0'(t) &= -\lambda_0 P_0(t) + \mu_1 P_1(t) \\
P_n'(t) &= \lambda_{n-1} P_{n-1}(t) - (\lambda_n + \mu_n) P_n(t) + \mu_{n+1} P_{n+1}(t), \qquad \forall n \geq 1.
\end{aligned}$$

Remarque 1.1. - Si le nombre des états est limité, ce qui revient à dire que l'effectif total est limité à un nombre N, alors dans ce cas $\lambda_n = 0$ pour tout $n > N$ et la dernière équation différentielle devient :

$$P'_N(t) = \lambda_{N-1} P_{N-1}(t) - \mu_N P_N(t).$$

Dans certaines étude on s'intéresse au régime stationnaire, c'est-à-dire quand $t \to \infty$, dans ce cas la distribution des probabilités d'états ne dépend plus de temps, donc leurs dérivés s'annulent. D'où le système des équations différentielles précédentes aura la configuration suivante :

$$0 = -\lambda_0 P_0 + \mu_1 P_1, \tag{1.1}$$
$$0 = \lambda_{n-1} P_{n-1} - (\lambda_n + \mu_n) P_n + \mu_{n+1} P_{n+1}, \quad \forall n \geq 1 \tag{1.2}$$

De la première équations, on tire :

$$P_1 = \tfrac{\lambda_0}{\mu_1} P_0$$

Pour $n = 1$, la deuxième équation donne :

$$P_1 = \tfrac{\lambda_0}{\mu_1} P_0$$

$0 = \lambda_0 P_0 - (\lambda_1 + \mu_1) P_1 + \mu_2 P_2$ D'où $P_2 = \dfrac{\lambda_0 \lambda_1}{\mu_1 \mu_2} P_0$ et de proche en proche(en utilise la récurrence), on déduit :

$$P_n = \frac{\lambda_0 \lambda_1 \lambda_2 ... \lambda_{n-1}}{\mu_1 \mu_2 \mu_3 .... \mu_n} P_0.$$

Pour la déduction de la valeur P_0, on se base sur la propriété $\sum_{i=0}^{n} P_i = 1$, en remplace tout les P_i qui sont en fonction de P_0, on tire :

$$P_0 = \frac{1}{1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} + ... + \frac{\lambda_0 \lambda_1 \lambda_2 ... \lambda_{n-1}}{\mu_1 \mu_2 \mu_3 ... \mu_n} + ...} = \frac{1}{1 + \sum_{n=1}^{\infty} \prod_{i=1}^{n} \frac{\lambda_{i-1}}{\mu_i}}$$

D'où en pourrait déduire les expressions des P_i

1.3.4 Modèles non markoviens

En l'absence de mémoire ou plutôt lorsque l'on s'écarte de l'hypothèse d'exponentialité de l'un des deux quantités stochastiques le temps des inter-arrivées et la durée de service, ou en prenant en compte certaines spécificités des problèmes par introduction des

paramètres supplémentaires, on aboutit à un modèle non Markovien.
La combinaison de tous ces facteurs rend l'étude mathématique du modèle très délicate, voire impossible, on essaye alors de se ramener à un processus de Markov judicieusement choisi à l'aide de l'une des méthodes d'analyse suivantes :

- méthode de la chaîne de Markov induite

Cette méthode, élaborée par Kendall, est souvent utilisée, elle consiste à choisir une séquence d'instants $(1, 2, \ldots, n)$, (déterministes ou aléatoire) telle que la chaîne induite $\{X_n, n \geq 0\}$ où $X_n = X(n)$, soit markovien et homogène.

- Méthode des variables auxiliaires

Cette méthode consiste à compléter l'information sur le processus $\{X(t)\}_{t\geq 0}$ de telle manière à lui donner le caractère markovien. Ainsi, on se ramène à l'étude du processus $\{X(t), A(t_1), A(t_2), ..., A(t_n)\}$.
Les variables $A(t_k), k \in \{1, 2, 3, ..., n\}$ sont des variables aléatoire supplémentaire.

- Méthode des événements fictifs

Le principe de cette méthode est introduire des événements fictifs qui permettent de donner une interprétation probabiliste aux transformées de Laplace et aux variables aléatoires décrivant le système étudié.

-Méthode des étapes d'Erlang

Son principe est d'approximer toute loi de probabilité ayant une transformée de Laplace rationnelle par une loi de Cox, cette dernière possède la propriété d'absence de mémoire.

-Méthode d'approximation

Dans ce cas, on caractérise l'état du système étudie par :
1. des méthodes asymptotiques décrivant l'état du système (chargé, non chargé,...),
2. l'estimation par borne de certaines de ces caractéristiques.

- Simulation

C'est un procédé d'imitation artificielle d'un processus réel donnée, comme résultat de cette imitation, on obtient des approximations des caractéristiques du système étudié, permettant ainsi de mesure ses performances.

1.3.5 Analyse opérationnelle des systèmes de files d'attente

Cette analyse, plus connue sous le nom de l'évaluation de performance, consiste au calcul des caractéristiques de performance d'un système, cette opération s'impose dès lors où l'on souhaite connaître les performances d'un système réel et que l'on ne peut effectuer de mesure directe sur celui-ci.
Les paramètres de performances que l'on souhaite obtenir sont de différents ordre en fonction des systèmes considérés, c'est ainsi dans les systèmes de production, un paramètre de performance important est le débit en produits finis. Tandis que pour le cas d'un guichet, le paramètre de performance qui intéresse au nombre de clients en attente au guichet.

Les caractéristiques de performance
Les caractéristiques d'exploitation du système auxquels on s'intéresse le plus souvent sont :
• Le nombre moyen de clients dans le système,
• La durée de séjour d'un client dans le système,
• La durée d'attente d'un client,
• Le taux d'occupation des postes de service.
La formule de Little
La formule de Little donne une relation très importante entre :"le nombre moyen de clients dans le système $L = E[X]$, La durée moyenne de séjour E_S et le nombre moyen de clients entrant dans le système par unité de temps λ", entre aussi "le nombre moyen de client dans la file $L_q = E[X_q]$, la durée moyen d'attente dans la file E_w, le nombre moyen de client entrant dans le système par unité de temps λ.
Cette formule stipule que

$$L = \lambda E_S$$
$$L_q = \lambda E_w$$

Il est à noter que cette formule est valable sous la vérification de la condition d'ergodicité géométrique du système.

1.4 Quelques systèmes de files d'attente

1.4.1 Le système M/M/1

Pour ce système, le plus simple dans la théorie des files d'attente, le flux des arrivées est poissonnien de paramètre λ et la durée de service est exponentielle de paramètre μ.

1.4.2 Lien avec la théorie générale du processus de naissance et de mort

Soit $X(t)$ est le nombre de cliens dans le système à la date t (nombre de clients en attentes + celle en cors de service).
On pose :

$$\lambda_n(t)\partial t = P(X(t+\partial t) = n+1/X(t) = n)$$
$$\mu_n(t)\partial t = P(X(t+\partial t) = n-1/X(t) = n)$$

Montrons que $\lambda_n(t)$ et $\mu_n(t)$ sont indépendants de t.

$$\lambda_n(t) = P(1\ arrive\ pendant\ le\ temps\ \partial t/X(t) = n) = \lambda_n = \lambda, \forall n \geq 0.$$

Car le processus d'arrives est poissonien.

$$\mu_n(t)\partial t = P(le\ service\ en\ cours\ a\ la\ date\ t\ cesse\ pendant\ l'intervalle\ \partial t / X(t) = n)$$

Soit $D(t)$ la variable aléatoire " durée de service de client au guichet à la date t ", comme la durée de service est poissonien de paramètre μ.
Soit U la durée de service de ce client à la date t.

$$\begin{aligned}\mu_n(t)\partial t &= P(D(t) \leq \mu + \partial t / D(t) > \mu) \\ &= 1 - P(D(t > \mu + \partial t / D(t) > \mu) \\ &= 1 - P(D(t) > \partial t) = P(D(t) \leq \partial t) \\ &= 1 - e^{-\mu \partial t} = \mu \partial t + o(\partial t)\end{aligned}$$

D'où

$$\mu_n(t) = \mu, \quad \forall n > 0$$

Pour $n = 0$ (pas de clients dans le système à la date t), donc pas de départ, alors :
Notre processus est bien un processus de naissance et de mort, avec :

$$\begin{cases} \mu_n = \mu, & \forall n > 0 \\ \mu_0 = 0 & \\ \lambda_n = \lambda, & \forall n \geq 0 \end{cases}$$

Ètude de régime permanent

Revenons au système d'équations différentielles vu en processus de naissance et de mort, on peut montrer que la condition pour qu'un régime permanent s'établisse s'écrit :

$$\sum_{n=1}^{\infty} \prod_{i=1}^{n} \frac{\lambda_{i-1}}{\mu_i} < \infty$$

Nous avons $\lambda_n = \lambda$ et $\mu_n = \mu$, $\forall n > 0$ d'où la condition précédente s'écrit :

$$\sum_{n=1}^{\infty} \prod_{i=1}^{n} \frac{\lambda}{\mu} < \infty \Rightarrow \sum_{n=1}^{\infty} (\frac{\lambda}{\mu})^n < \infty$$

On pose $\rho = \frac{\lambda}{\mu}$(qui s'appelle intensité de trafic) donc : $\sum_{n=1}^{\infty} \rho^n < \infty$
D'où le régime permanent s'établit si : $\frac{\lambda}{\mu} < 1 \Rightarrow \frac{1}{\mu} < \frac{1}{\lambda}$ qui veut dire que la durée moyenne des services est inférieure à l'intervalle moyen entre deux arrivées.

Dans ce cas :

$$P_n = \rho^n P_0 \quad et \quad \sum_{n=0}^{\infty} \rho^n = \frac{1}{1-\rho}$$
$$= \rho^n (1-\rho).$$

Régime stationnaire

Lorsque t tend vers l'infini dans le système d'équation différentielles donnés dans le régime transitoire, on peut montrer que les limites $\lim_{n\to\infty} P_n(t) = \pi_n$ existent et sont indépendantes de l'état initial du processus et que :

$$\lim_{n\to\infty} P_n^{'}(t) = 0, \quad \forall n = 0, 1, ...$$

Ainsi, à la place d'un système d'équations différentielles, on obtient un système d'équations linéaires et homogènes :

$$\begin{cases} \mu\pi_1 & = -\lambda\pi_0, \\ \lambda\pi_{n-1} + \mu\pi_{n+1} & = (\lambda+\mu)\pi_n \quad n = 1, 2, ... \end{cases}$$

De plus,nous avons la condition $\sum_{n=0}^{\infty} \pi_n = 1$ car $(\pi_n)_n$ est une distribution de probabilité.

La solution de ces équations est donnée par :

$$\pi_n = \pi_0(\tfrac{\lambda}{\mu})^n = (1 - \tfrac{\lambda}{\mu})(\tfrac{\lambda}{\mu})^n, \quad n = 1, 2, ...$$

A condition que $\lambda \leq \mu$ (condition d'ergodicité géométrique du système), on montre que le régime stationnaire du système M/M/1 est gouverné par la loi géométrique.

1.4.3 Quelques caractéristiques

• **Le nombre moyen de client dans le système :** si on note cette caractéristique par L, alors :

$$\begin{aligned} L = E(X) &= \sum_{n=0}^{\infty} n\pi_n \\ &= (1-\rho)\sum_{n=0}^{\infty} n\rho^n \\ &= \rho(1-\rho)\sum_{n=1}^{\infty} n\rho^{n-1} \\ &= \rho(1-\rho)\sum_{n=1}^{\infty}(\rho^n)' \\ &= \rho(1-\rho)(\frac{1}{1-\rho})' \\ &= \frac{\lambda}{\lambda-\mu} \end{aligned}$$

$\rho = \dfrac{\lambda}{\mu}$ est appelé l'intensité du trafic et aussi, la charge du système.

• **Le nombre moyen de client dans la file :** notons cette caractéristique par L_q, soit X_q le nombre de client dans la file d'attente à la date t, on aura donc :

$$X_q = \begin{cases} 0 & si\ X = 0 \\ \text{X-1} & siX \geq 1 \end{cases}$$

Alors,

$$\begin{aligned} L_q = E(X_q) &= \sum_{n=1}^{\infty}(n-1)\pi_n \\ &= (1-\rho)\sum_{n=1}^{\infty}(n-1)\rho^n \\ &= \rho^2(1-\rho)\sum_{n=1}^{\infty}(n-1)\rho^{n-2} \end{aligned}$$

$$
\begin{aligned}
&= \rho^2(1-\rho)\sum_{n=1}^{\infty}(\rho^{n-1})' \\
&= \rho^2(1-\rho)(\frac{1}{1-\rho})' \\
&= \frac{\lambda^2}{\mu(\mu-\lambda)}
\end{aligned}
$$

D'autres caractéristiques de ce système peuvent être calculées, soit directement à partir de la distribution stationnaire, soit d'après la formule de Little.

1.4.4 Le système de files d'attente M/G/1

ce système possède un processus d'arrivée de poisson de paramètre λ et une loi de service quelconque H, de moyenne $\frac{1}{\mu}$. La propriété du Markov du processus $\{X_t\}_{t\geq 0}$ facilitant l'analyse de système M/M/1 n'est plus vérifiée pour le système M/G/1, ce qui rend son analyse plus délicate. Les méthodes des systèmes non markoviens citées précédemment peuvent être utilisées pour l'analyse de ce système. Nous nous limiterons à la méthode de la chaîne de Markov induite.

La chaîne de Markov induite

La méthode des variables auxiliaires s'applique aux systèmes M/G/1 complétant l'information sur X(t) :"le nombre de clients dans le système" par la variable $A(t1)$ qui représente le temps de service déjà écoulé d'un client à l'instant t, le processus bidimensionnel $\{X(t), A_1(t)\}_{t\geq 0}$ décrit complètement le système M/G/1. le calcul de son régime transitoire serait intervenir des équation aux dérivées partielles. Pour éviter cela, la méthode de la chaîne de Markov induite ramène l'étude ce processus au cas discret. En effet, en considérant les instants d_n de départ du n^{ieme} clients, le processus $\{X(d_n), A_1(d_n)\}$ sera équivalent à $X_n = X(d_n)$ puisque $A_1(d_n) = 0$. La variable aléatoire X_n représentant le nombre de clients dans le système juste après l'instant d_n est une chaîne de Markov à temps discret. On considère le processus E_n "le nombre de clients qui entrent pendant le service de n^{ieme} client". Les variables E_n sont indépendantes entre elles, leur distribution commune est :

$$
\begin{gathered}
P(E_n = k) = a_k = \int_0^{\infty} e^{-\lambda t}\frac{(\lambda t)^k}{k!}\partial H(t), \quad \forall k = 0, 1, 2.... \\
X_{n+1} = X_n - \delta + E_{n+1}, \forall \ \ n \in N
\end{gathered}
$$

Avec

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

X_{n+1} ne dépend que de X_n et de E_{n+1} et non des valeurs de X_{n-1}, X_{n-2},...
La variable X_n ainsi définie est la chaîne de Markov induite du processus $\{X(t), t \geq 0\}$.

Régime transitoire

Il est aisé de vérifier que les probabilités de transition $P_{i,j} = P(X_{n+1} = j/X_n = i)$ sont données par :

$$P_{i,j} = \begin{cases} a_j & si\ j \geq 0, i = 0 \\ a_{j-i+1} & si 1 \leq i \leq j \\ 0 & sinon \end{cases}$$

Ainsi, la matrice de transition prend la forme suivante :

$$P_{i,j} = \begin{pmatrix} a_0 & a_1 & a_2 & ... & ... \\ a_0 & a_1 & a_2 & ... & ... \\ 0 & a_0 & a_1 & ... & ... \\ 0 & 0 & a_2 & ... & ... \\ 0 & 0 & a_0 & ... & ... \\ . & . & . & \ddots & \ddots \end{pmatrix}$$

Régime stationnaire

Supposant que $\rho \leq 1$ et soit $\pi = (\pi_0, \pi_1, ..., \pi_n)$ la distribution stationnaire de la chaîne X_n, il ne sera généralement pas possible de calculer π elle-même, mais nous pouvons calculer la fonction génératrice correspondant $\Pi(z)$:

$$\Pi(z) = \bar{H}(\lambda - \lambda z)\frac{(1-\rho)(z-1)}{z - \bar{H}(\lambda z - \lambda)} \tag{1.3}$$

Où $\bar{H}$ est la transformée de Laplace de la densité de probabilité du temps de service, et $z \in C$ tel que $\mid z \mid < 1$. La formule (1.1) est connue sous le nom de première formule de Pollaczek-Khinchine.

Quelques caractéristiques

• **Le nombre de clients dans le système :** Pour le système d'attente M/G/1, le nombre moyen de client dans le système au temps t quelconque égale au nombre moyen de clients dans le système au temps d_n, égale à $E(X)$, qui vaut :

$$E(X) = \frac{Q + E(A_n^2 + 1 - 2Q)}{2(1 - Q)}$$

où

$$Q = E(A_n)$$

Conclusion

Dans ce chapitre, nous avons introduit les résultats classiques concernant les systèmes de files d'attente en générale. En première lieu, nous avons parlé a propos du formalisme des files d'attente, après nous avons traité quelques systèmes markovien et non markovien. Cette étude théorique nous permet d'introduire les approches d'étude des systèmes de files d'attente avec priorité et système avec rappels et vacance.

Chapitre 2

Systèmes prioritaires et système avec rappels et vacance

introduction

Dans le cadre général, les unités (clients, appels, message,...) qui s'accumulent dans une file, sont traités dans l'ordre de leur arrivée. C'est ce qu'on appelle la procédure PAPS (Premier Arrivé Premier Servi). D'autres procédures peuvent être mise en œuvre, parmi lesquelles, on présente l'admission avec priorité [41].

Les programmes dans un ordinateur ou les paquets dans un réseau informatique peuvent ne pas être traités de la même façon : quelques uns peuvent recevoir un traitement préférentiel. Dans une salle d'urgence des hôpitaux, une personne inconscients ou ayant une crise cardiaque aura la priorité au service par rapport aux autres qui ont subi une blessure mineure. Les systèmes d'attente dans lesquels des clients reçoivent des traitements préférentiels sont dits systèmes d'attente prioritaires.

donc, dans ce chapitre, il nous a paru opportun de décrire le comportement de types de file d'attente, à savoir, les systèmes prioritaires et les systèmes avec rappels et vacance.

Les systèmes avec rappels apparaissent dans la modélisation des phénomènes du genre :" service des avions dans un aéroport, comportement des processus(tâches, programme,...) dans un réseau informatique constitué d'un ordinateur central et d'un ensemble de périphériques(terminaux)[41].

Les systèmes d'attente avec vacance est introduite en générale dont le but est d'exploiter le temps inoccupé du serveur, dont le but est d'améliorer les caractéristiques de performances, ce système est analyser par plusieurs chercheurs parmi eux, en peut citer Doshi(1986), Teghem (1986), Takagi (1991), Tian et Zhang (2006)...

2.1 Les systèmes de files d'attente avec priorité

Il n'est pas rare, dans la vie courante de rencontrer des systèmes d'attente avec plusieurs types de clients, où certains sont prioritaires par rapport à d'autres. Pour mieux illustrer cette situation, considérons deux systèmes de files d'attente (M/M/1 et M/G/1) avec m types de clients. Les clients de type i arrivent indépendamment des autres types de clients suivant un taux $\lambda_i, i = 1, 2, ..., m$, Le client de type i à la i^{ieme} priorité et la i^{ieme} priorité est supérieur à la k^{ieme} priorité pour tout $k > i$.

On distingue deux types de priorités : **priorité relative et priorité absolue.**

Dans ce qui suit nous allons montrer comment obtenir certaines caractéristiques d'un système prioritaire dans les deux cas.

Priorité relative

La priorité relative se caractérise par le fait qu'un client de priorité supérieur ne peut en aucun cas interrompre le service d'un client priorité inférieure. Il doit attendre (s'il le désire bien sûr) jusqu'à ce que le service du client de priorité inférieure soit fini avant d'être servi. Cette discipline a été introduite par Cobham [3] et étudié par plusieurs chercheurs (Holley [2], Dressin et Reich [4], Morse [5] et autres). Le système d'attente avec priorité relative peut aussi être considéré comme un système d'attente à serveur non fiable, la panne est prise en considération après que le client en service ait terminé son service. Ce modèle est connu sous le non " **Breakdown postponable interruption** ". Il a été étudié par Keilson [6], Gaver [7] et Hodgson [8].

2.1.1 Système $M_R/G_R/1$ avec priorité relative

Dans ce système, il existe R classes indépendants de clients qui arrivent suivant des flux poissonniens de taux $\lambda_k, k = 1, 2, ..., R$, correspondant à chaque classe. Le taux arrivée total est $\sum_{k=1}^{R} \lambda_k$.

La distribution de temps de service est générale, de fonction de réparation $B_k(.)$ et de moyenne $E(B_k) = \dfrac{1}{\mu_k}$ pour les clients de classe k. On défini $\rho_k = \dfrac{\lambda_k}{\mu_k}$, La condition $\sum_{i=1}^{R} \rho_i < 1$, dite d'ergodicité géométrique du système, est supposée être vérifiée. On s'intéresse principalement au calcul du temps moyen d'attente d'un client dans la file (c'est-à-dire, le temps qu'il passe dans la file avant le début de son service).
Le temps d'attente W_k de dernier client arrivé C_k, de classe k contient trois composantes

[11]

-Le temps résiduel W_0 de client en cours de service,

-Le temps de service $W_{1,k}$ de tous les clients qui sont déjà présents, à l'arrivée du client C_k et qui ont une priorité égale ou supérieure à celle de C_k,

-Le temps de service $W_{2,k}$ de tous les clients qui sont arrivés pendant le temps d'attente du client C_k et qui ont une priorité supérieure à celle de C_k.
Donc, le temps d'attente du client C_k est donnée par :

$$W_k = W_0 + W_{1,k} + W_{2,k},\ k = 1, 2, ..., R$$

Calcul de temps moyen résiduel du client en cours de service

soit α ="le temps de service résiduel", et β ="le client en cours de service est de classe i ",donc, le temps moyen résiduel du client en cours de service prend cette forme :

$$\begin{aligned} E[W_0] &= \sum_{i=1}^{R} E(\alpha/\beta).P(\beta) \\ &= \sum_{i=1}^{R} \frac{E(B_i^2)}{2E(B_i)}.P(\beta) \\ &= \sum_{i=1}^{R} \frac{E(B_i^2)}{2E(B_i)}.\rho_i \end{aligned}$$

Si on prend l'avantage que $\rho_i = \lambda_i E(B_i)$, on arrive à

$$E(W_0) = \sum_{i=1}^{R} \frac{\lambda_i E(B_i^2)}{2} \tag{2.1}$$

Par conséquent, le temps résiduel moyen de service d'un client dépend des moments d'ordre 2 des distributions des temps de service.

Calcul de temps moyen de service de tous les clients qui sont déjà présents à l'arrivée du client C_k et qui ont une priorité égale ou supérieure à celle de C_k

Soit $N_{1,k}$ le nombre de clients de classe i qui sont déjà présents dans la file à l'arrivée du client C_k (de classe k) et qui sont service avant C_k, On a donc :

$$E[W_{1,k}] = \sum_{i=1}^{k} E[N_{i,k}].E[B_i]$$

Puisque les temps de service sont indépendants. On peut maintenant appliquer la formule de Little :

$$E[N_{i,k}] = \lambda_i E[W_i]$$

Où W_i dénote le temps d'attente dans la file pour les clients de classe i. on arrive à :

$$E[W_{1,k}] = \sum_{i}^{k} E[N_{i,k}].E[B_i] = \sum_{i=1}^{k} \lambda_i E[W_i]E[B_i]$$

D'où

$$E[W_{1,k}] = \sum_{i=1}^{k} \rho_i E[W_i], \quad (k = 1, 2, ..., R) \tag{2.2}$$

Calcul de temps moyen de service de tous les clients qui sont arrivés pendant le temps d'attente du client C_k et qui ont une priorité supérieure à celle de C_k

Soit $M_{i,k}$ le nombre de clients de classe i arrivant durant le temps d'attente W_k du client C_k et qui reçoivent leur service avant le client C_k (c'est-à-dire, ils ont une priorité supérieure).

A l'aide de la formule du Little, on exprime $E[M_{i,k}]$ par :

$$E[M_{i,k}] = \lambda_i E[W_i]$$

Donc on a :

$$E[W_{2,k}] = \sum_{i=1}^{k-1} E[M_{i,k}].E[B_i] = \sum_{i=1}^{k-1} \lambda_i E[W_k]E[B_i]$$

D'où

$$E[W_{2,k}] = E[W_k] \sum_{i=1}^{k-1} \rho_i, \quad (k = 1, 2, ..., R) \tag{2.3}$$

Temps moyen d'attente $E[W_k]$, d'un client de classe k

On va inclure les trois résultats trouvés (2.1),(2.2) et (2.3), dans l'expression de base de $E[W_k]$:

$$\begin{aligned} E[W_k] &= E[W_0] + E[W_{1,k}] + E[W_{2,k}] \\ &= E[W_0] + \sum_{i=1}^{k} \rho_i E[W_i] + E[W_k] \sum_{i=1}^{k-1} \rho_i \end{aligned}$$

De cette dernière équation, on trouve :

$$E[W_k] = \frac{1}{1 - \sum_{i=1}^{k} \rho_i} (E[W_0] \sum_{i=1}^{k-1} \rho_i E[W_i]),$$

pour $(k = 1, 2, ..., R)$ on obtient un système d'équations linéaires, qu'on peut résoudre d'une manière récursive. On utilisera cette courte notation :

$$\sigma_k = \sum_{i=1}^{k} \rho_i,$$

Par cette natation on peut écrire pour $k = 1$:

$$E[W_1] = \frac{E[W_0]}{1 - \sigma_1}$$

Pour $k = 2$ on obtient :

$$\begin{aligned} E[W_2] &= \frac{1}{1 - \sigma_2} (E[W_0] + \rho_1 \frac{E[W_0]}{1 - \sigma_1}) \\ &= \frac{E[w_0]}{1 - \sigma_2} (\frac{1 - \rho_1 + \rho_1}{1 - \sigma_1}) \end{aligned}$$

On aura donc,

$$E[W_2] = \frac{E[W_0]}{(1 - \sigma_2)(1 - \sigma_1)}$$

Il est facile à présent de prouver d'une manière récursive que pour $k < R$:

$$E[W_k] = \frac{E[w_0]}{(1 - \sigma_k)(1 - \sigma_{k-1})}$$

Temps moyen d'attente $E[W_k]$ est donné par :

$$E[W_k] = \frac{\sum_{i=1}^{R} \lambda_i E[B_i^2]}{2(1 - \sum_{i=1}^{k} \rho_i)(1 - \sum_{i=1}^{k-1} \rho_i)}, \quad (k = 1, 2, ..., R)$$

Temps moyen de séjour d'un client de classe k dans le système

$$E[S_k] = E[W_k] + E[B_k] = \frac{\sum_{i=1}^{R} \lambda_i E[B_i^2]}{2(1-\sum_{i=1}^{k} \rho_i)(1-\sum_{i=1}^{k-1} \rho_i)} + E[B_k],\ (k = 1, 2, ..., R)$$

A l'aide de la formule de Little, on trouve aussi :

Nombre moyen de clients de classe k dans la file

$$Q_k = \lambda_k E[W_k] = \lambda_k \frac{\sum_{i=1}^{R} \lambda_i E[B_i^2]}{2(1-\sum_{i=1}^{k} \rho_i)(1-\sum_{i=1}^{k-1} \rho_i)},\ (k = 1, 2, ..., R)$$

Nombre moyen de clients de classe k dans le système

$$L_k = \lambda_k E[S_k] = \lambda_k \frac{\sum_{i=1}^{R} \lambda_i E[B_i^2]}{2(1-\sum_{i=1}^{k} \rho_i)(1-\sum_{i=1}^{k-1} \rho_i)} + \rho_k, \quad (k = 1, 2, ..., R)$$

2.1.2 Système $M_2/G_2/1$ avec priorité relative

Ce système a été étudié dans l'article [31], dans ce système, les clients arrivent en deux calasse, selon un processus de poisson de paramètre λ_1 et λ_2, La distribution de service est générale de fonction B_i, $i = 1, 2$ et de moyenne $\frac{1}{\mu_1}$ et $\frac{1}{\mu_2}$.

Le temps moyen d'attente d'un client dans la file

Client prioritaire :

$$E[W_1] = \frac{(\lambda_1 E[B_1^2] + \lambda_2 E[B_2^2])}{2(1-\rho_1)}$$

Client non prioritaire :

$$E[W_2] = \frac{(\lambda_1 E[B_1^2] + \lambda_2 E[B_2^2])}{2(1-\rho_1-\rho_2)(1-\rho_1)}$$

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

Client prioritaire :

$$E[S_1] = \frac{(\lambda_1 E[B_1^2] + \lambda_2 E[B_2^2])}{2(1-\rho_1)} + E[B_1]$$

Client non prioritaire :

$$E[S_2] = \frac{(\lambda_1 E[B_1^2] + \lambda_2 E[B_2^2])}{2(1-\rho_1-\rho_2)(1-\rho_1)} + E[B_2]$$

2.2 Le système $M_2/M_2/1$ avec priorité relative

Ce système a été étudié par plusieurs chercheurs, parmi eux Rupert dans l'article [28], Guy dans [29] et D.Gross dans le livre [30].

Considérons un système $M_2/M_2/1/3$ avec priorité relative dans lequel arrivent deux classes de clients que nous appelons :
Classe 1 : **Clients non prioritaires.**
Classe 2 : **Clients prioritaires.**

Les deux types des clients arrivent indépendamment l'un de l'autre suivant un processus poissonnien avec respectivement les taux λ_1 et λ_2. Le service des clients prioritaires et non prioritaires se fait suivant la loi exponentielle μ_1 et μ_2 respectivement.

Soit $P_{n,m,r}(t)$ = P(au temps t,avoir n clients de classe 1, m clients de classe 2,r) avec :

$$r = \begin{cases} 1 & si \text{ le client en service est de classe 1} \\ 2 & si \text{le client en service est de classe 2} \end{cases}$$

A l'état d'équilibre, nous obtenons pour $\rho = \sum_{i=1}^{2} \frac{\lambda_i}{\mu_i}$, le système d'équation suivant :

$$\begin{aligned}
P_{0,0}(\lambda_1+\lambda_2) &= P_{0,1,2}\mu_2 + P_{1,0,1}\mu_1 \\
P_{0,1,2}(\lambda_1+\lambda_2+\mu_2) &= P_{0,0}\lambda_2 + P_{0,2,2}\mu_2 + P_{1,1,1}\mu_1 \\
P_{1,0,1}(\lambda_1+\lambda_2+\mu_1) &= P_{0,0}\lambda_1 + P_{2,0,1}\mu_1 + P_{1,1,2}\mu_2 \\
P_{1,1,2}(\lambda_1+\lambda_2+\mu_2) &= P_{1,2,2}\mu_2 + P_{2,1,1}\mu_1 + P_{0,1,2}\lambda_2 \\
P_{0,2,2}(\lambda_1+\lambda_2+\mu_2) &= P_{0,1,2}\lambda_2 + P_{1,2,1}\mu_1 + P_{0,3,2}\mu_2 \\
P_{2,0,1}(\lambda_1+\lambda_2+\mu_1) &= P_{1,0,1}\lambda_1 + P_{2,1,2}\mu_2 + P_{3,0,1}\mu_1
\end{aligned}$$

Ces probabilités de transition sont illustrées dans la figure suivante :

Pour les systèmes $M_2/M_2/1/\infty$, si on considère un système d'attente avec priorité relative,

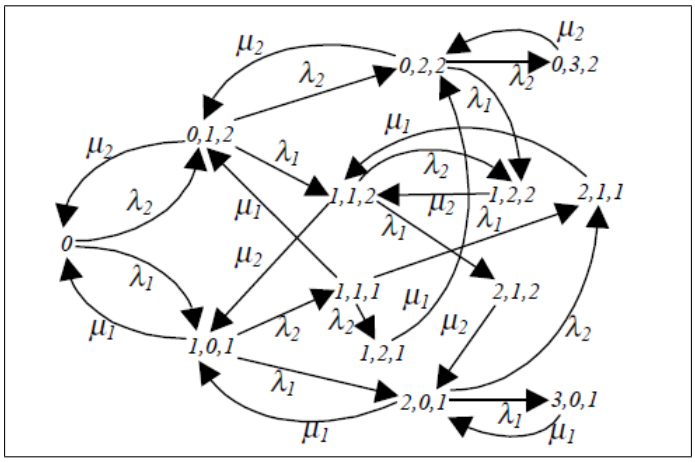

FIG. 2.1 – **Les probabilité de transition du système $M_2/M_2/1/3$ avec priorité relative.**

les clients arrivent selon un processus de poisson, le taux des arrivées de la première classe est λ_1 (respectivement λ_2 pour la deuxième classe). Les services des clients prioritaires et non prioritaires se fait suivant la loi exponentielle de taux μ_1 et μ_2 respectivement. Notons $P_{n,m,r}$ la probabilité d'avoir n clients non prioritaires et m clients prioritaires au temps t et r indique le client en service. Alors $P_{n,m,r}$ est donnée par le système d'équation suivant :

$$\begin{aligned}
(\lambda_1+\lambda_2)P_{0,0} &= P_{0,1,2}\mu_2 + P_{1,0,1}\mu_1 \\
(\lambda_1+\lambda_2+\mu_1)P_{1,0,1} &= P_{0,0}\lambda 1 + P_{2,0,1}\mu_1 + P_{1,1,2}\mu_2 \\
(\lambda_1+\lambda_2+\mu_1)P_{0,1,2} &= P_{0,0}\lambda 2 + P_{0,2,2}\mu_2 + P_{1,1,1}\mu_1 \\
(\lambda_1+\lambda_2+\mu_1)P_{m,0,1} &= P_{m-1,0,1}\lambda 1 + P_{m,1,2}\mu_2 + P_{m+1,0,1}\mu_1 \\
(\lambda_1+\lambda_2+\mu_2)P_{0,n,2} &= P_{0,n-1,2}\lambda 1 + P_{0,n+1,2}\mu_2 + P_{1,n,1}\mu_1 \\
(\lambda_1+\lambda_2+\mu_2)P_{1,n,1} &= P_{1,n-1,1}\lambda 2 + P_{1,n+1,2}\mu_2 + P_{2,n,1}\mu_1 \\
(\lambda_1+\lambda_2+\mu_1)P_{m,n,1} &= P_{m-1,n,1}\lambda 1 + P_{m,n-1,1}\lambda_2 + P_{m+1,n,1}\mu_1 + P_{m,n+1,2}\mu_2, m>1, n>0 \\
(\lambda_1+\mu_2)P_{m,n,2} &= P_{m,n-1,2}\lambda_1, m>0, n>1
\end{aligned}$$

Nous obtenons à l'état d'équilibre la probabilité $P_n(t)$. La probabilité d'avoir au temps t, n clients de classe 1, m clients de classe 2.

$$P_n = \sum_{m=0}^{n-1}(P_{n-m,m,1} + P_{n,n-m,2}) = (1-\rho)\rho,\ (n>1)$$

En utilisant les fonctions génératrices marginales :

$$\begin{aligned}
P_{m,1}(z) &= \sum_{n=0}^{\infty} P_{n,m,1}z^n, \\
P_{m,2}(z) &= \sum_{m=0}^{\infty} P_{n,m,2}z^m,
\end{aligned}$$

$$H_1(y,z) = \sum_{m=1}^{\infty} y^m P_{m,1}(z), \text{avec}, H_1(1,1) = \frac{\lambda_1}{\mu_1}$$
$$H_2(y,z) = \sum_{m=0}^{\infty} y^m P_{m,2}(z), \text{avec}, H_2(1,1) = \frac{\lambda_2}{\mu_2}$$

On obtient la fonction génératrice jointe $H(y,z)$

$$\begin{aligned} H(y,z) &= H_1(y,z) + H_2(y,z) + P_0 \\ &= \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} y^m z^n (P_{m,n,1} + P_{m,n,2}) \sum_{m=1}^{\infty} y^m P_{m,0,1} + \sum_{n=1}^{\infty} z^n P_{0,n,2} + P_0 \end{aligned}$$

D'après le calcule de le fonction génératrice jointe $H(y,z)$, en peut l'utiliser afin de trouver le nombre moyen des clients prioritaires dans le système L_1, le nombre moyen de clients non prioritaires dans le système L_2, alors :

$$\begin{aligned} L_1 &= \frac{\partial H(y,z)}{\partial y}_{/z=y=1} = L_{q1} + \frac{\lambda_1}{\mu_1} \\ &= \frac{\rho_1(1+\rho_2)}{1-\rho_1} \end{aligned} \tag{2.4}$$

$$\begin{aligned} L_2 &= \frac{\partial H(y,z)}{\partial z}_{/z=y=1} = L_{q2} + \frac{\lambda_2}{\mu_2} \\ &= \frac{\rho_2(1-\rho_1(1-\rho_1-\rho_2))}{(1-\rho_1)(1-\rho_1-\rho_2)} \end{aligned} \tag{2.5}$$

Au moyen des formules (2.4),(2.5) et de la formule de Little, on aura, le temps moyen de séjour des clients prioritaires et non prioritaires donné par les deux formules suivantes :

$$E[S_1] = \frac{L_1}{\lambda_1} = \frac{1+\rho_2}{\mu_1(1-\rho_1)} \tag{2.6}$$
$$E[S_2] = \frac{1-\rho_1(1-\rho_1-\rho_2)}{\mu_2(1-\rho_1)(1-\rho_1-\rho_2)} \tag{2.7}$$

2.2.1 Priorité absolue

Contrairement au cas précédent, ici, un client de priorité supérieure à la droit d'interrompre le service d'un client de priorité inférieure pour se faire servir. Le service interrompu sera alors repris à partir du point où il était suspendu.

Les premiers articles publiés concernant les systèmes d'attente avec priorité absolue sont les articles de White et Christie [34]. Par la suite sont apparus les articles Stephan [32],

Miller [33], Jaiswall [38] et Welch [39]. Takacs et Chang [36] ont étudié la priorité absolue avec une source infinie et différentes supposition à propos de la distribution de service. Les points communs entre les systèmes d'attente avec priorité absolue et les systèmes à serveurs non fiable (Pannes) ont été étudiés pour la première fois par White et Christie [34]. Keilson [35], Gaver [40], Avi-Itzhak et Naor [27], ont utilisé cette ressemblance pour étudier la priorité absolue à partir des systèmes à serveur non faible (breakdown models). Les systèmes prioritaires avec source finie (dans laquelle au moins un type de client à source finie) ont été étudiés par Avi-tzhak et Naor [27].

2.2.2 Système $M_R/G_R/1$ avec priorité absolue

Gelenbe dans l'article [21] à étudié le système d'attente de type M/G/1,dans ce système les clients arrivent en k classes indépendantes suivant un processus de Poisson. Le taux d'arrivée pour chaque classe est λ_i et le taux d'arrivée totale est $\lambda = \lambda_1 + \lambda_2 + ... + \lambda_r$. la distribution de service est générale de moyenne $\frac{1}{\mu}$, dans un système avec priorité absolue conservatrice, Le client interrompu reprend son service au point où il est interrompu.

Dans ce cas, les clients de priorités inférieures sont totalement "invisible" et n'affectent en aucun cas la file des clients de hautes priorités. Alors, pour un client de classe k, on peut procéder comme si $\lambda_{k+1} = \lambda_{k+2} = ... = \lambda_R = 0$.

Analysons, à présent, un système dans lequel les classes $k+1$ jusqu'à R n'existe pas et comparons les temps d'attente d'un client. Le temps moyen d'attente W_k, du dernier client arrivé C_k de classe k, avant qu'il entre en service pour la premier fois, est le même que dans le cas de priorité relative avec k classe de priorité.
on a

$$E[W_k] = \frac{\sum_{i=1}^{k} \lambda_i E[B_i^2]}{2(1-\sigma_k)(1-\sigma_{k-1})}$$

avec $\sigma_k = \sum_{i=1}^{k} \rho_i$

Le temps moyen de séjour d'un client dans le système S_j, contient trois parties :

$$E[S_j] = E[W_j] + E[B_j] + E[I_j], j = 1, 2, ..., k \tag{2.8}$$

Le premier terme $E[W_j]$, est le temps moyen d'attente avant que le client entre en service pour la première fois, le second terme $E[B_j]$, est le temps moyen de service de client et le troisième terme $E[I_j]$, est l'espérance de la variable aléatoire I_j, représentant

le temps total d'interruption du client durant son service le temps total d'interruption consiste en deux parties : la somme des temps de service des clients qui ont interrompu le service de client C_k, et la somme des temps de service des clients qui sont arrivés durant les périodes dans lesquelles le client est déjà interrompu, On obtient alors :

$$\begin{aligned} E[I_j] &= E[B_j]\sum_{i=1}^{k-1}\lambda_i E[B_i] + E[I_j]\sum_{i=1}^{k-1}\lambda_i E[B_i] \\ &= E[B_j]\sum_{i=1}^{k-1}\rho_i + E[I_j]\sum_{i=1}^{k-1}\rho_i \\ &= \frac{E[B_j]\sum_{i=1}^{k-1}\rho_i}{1-\sum_{i=1}^{k-1}\rho_i}, j = 1, ..., k \end{aligned}$$

La somme du seconde et troisième terme de 2.4, le temps moyen effectif de service et les temps totaux moyens d'interruption, est ce qu'on peut appeler le temps généralisé de service.

$$E[B_j] + E[I_j] = \frac{E[B_j]}{1-\sum_{i=1}^{k-1}\rho_i}$$

A partir les résultats précédents, on trouve :

$$E[S_j] = \frac{\sum_{i=1}^{k-1}\lambda_i E[B_i^2]}{2(1-\sigma_k)(1-\sigma_{k-1})} + \frac{E[B_j]}{1-\sum_{i=1}^{k-1}\rho_i}, j = 1, ..., k$$

A l'aide de la formule de Little, on trouve aussi :

$$Q_j = \lambda_j E[W_j] = \lambda_j \frac{\sum_{i=1}^{k}\lambda_i E[B_i^2]}{2(1-\sum_{i=1}^{k}\rho_i)(1-\sum_{i=1}^{k-1}\rho_i)}, k = 1, ..., R$$

Nombre moyen de clients de classe k dans le système

$$L_j = \lambda_j E[S_j] = \lambda_j E[W_j] = \lambda_j \frac{\sum_{i=1}^{k}\lambda_i E[B_i^2]}{2(1-\sum_{i=1}^{k}\rho_i)(1-\sum_{i=1}^{k-1}\rho_i)} + \frac{\rho_j}{1-\sum_{i=1}^{k-1}\rho_i}, j = 1, 2, ..., k$$

2.2.3 Système $M_2/G_2/1$ avec priorité absolue

Dans ce système, les clients arrivent en deux calasse, selon un processus de poisson de paramètre λ_1 et λ_2, La distribution de service des clients prioritaires et non prioritaires

suivent des lois générales, de fonction de répartition $B_i(.)$,$i = 1, 2$ et de moyenne $\frac{1}{\mu_1}$ et $\frac{1}{\mu_2}$.

Le temps moyen d'attente d'un client dans la file
Client prioritaire :

$$E[W_1] = \frac{\lambda_1 E[B_1^2]}{2(1-\rho_1)}$$

Client non prioritaire :

$$E[W_2] = \frac{(\lambda_1 E[B_1^2] + \lambda_2 E[B_2^2])}{2(1-\rho_1)(1-\rho_1-\rho_2)}$$

Le temps moyen de séjour d'un client dans le système
Client prioritaire :

$$E[S_1] = \frac{\lambda_1 E[B_1^2]}{2(1-\rho_1)} + E[B_1]$$

Client non prioritaire :

$$E[S_2] = \frac{(\lambda_1 E[B_1^2] + \lambda_2 E[B_2^2])}{2(1-\rho_1)(1-\rho_1\rho_2)} + \frac{E[B_{12}]}{1-\rho_1}$$

2.2.4 Le système $M_2/M_2/1$ avec priorité absolue

Ce système a été étudié Avi-Itzhak dans l'article [37], il a considéré un système $M_2/M_2/1/3$ avec priorité absolue dans lequel arrivent deux classes :
Classe 1 : **Clients non prioritaires.**
Classe 2 : **Clients prioritaires.**

Les deux types des clients arrivent indépendamment l'un de l'autre suivant un processus poissonnien avec respectivement les taux λ_1 et λ_2, Le service des clients prioritaires et non prioritaires se fait suivant la loi exponentielle μ_1 et μ_2 respectivement.

Écrivons les équations de Chapman Kolmogrov à l'état d'équilibre. Notons P_{n_1,n_2} la probabilité d'avoir n_1 clients de classe 1 et n_2 clients de classe 2 dans le système et $P_{0,0}$ la probabilité d'avoir 0 clients de classe 1, 0 clients de classe 2 dans le système.

$$
\begin{aligned}
P_{0,0}(\lambda_1+\lambda_2) &= P_{0,1}\mu_2 + P_{1,0}\mu_1,\\
P_{0,1}(\lambda_1+\lambda_2+\mu_2) &= P_{0,0}\lambda_2 + P_{0,2}\mu_2\\
P_{1,0}(\lambda_1+\lambda_2+\mu_1) &= P_{0,0}\lambda_1 + P_{2,0}\mu_1 + P_{1,1}\mu_2\\
P_{1,1}(\lambda_1+\lambda_2+\mu_2) &= P_{1,2}\mu_2 + P_{1,0}\lambda_2 + P_{0,1}\lambda_1\\
P_{0,2}(\lambda_1+\lambda_2+\mu_2) &= \lambda_2 P_{0,1} + \mu_2 P_{0,3} + \mu_1 P_{1,2}\\
P_{2,0}(\lambda_1+\lambda_2+\mu_1) &= \lambda_1 P_{1,0} + \mu_2 P_{2,1} + \mu_1 P_{3,0}\\
\mu_2 P_{1,2} &= \lambda_1 P_{0,2} + \lambda_2 P_{1,1}\\
\mu_2 P_{2,1} &= \lambda_1 P_{1,1} + \lambda_2 P_{2,0}\\
\mu_2 P_{0,3} &= \lambda_2 P_{0,2}\\
\mu_1 P_{3,0} &= \lambda_1 P_{2,0}
\end{aligned}
$$

Les transitions de la chaîne sont illustrées dans la figure suivante :

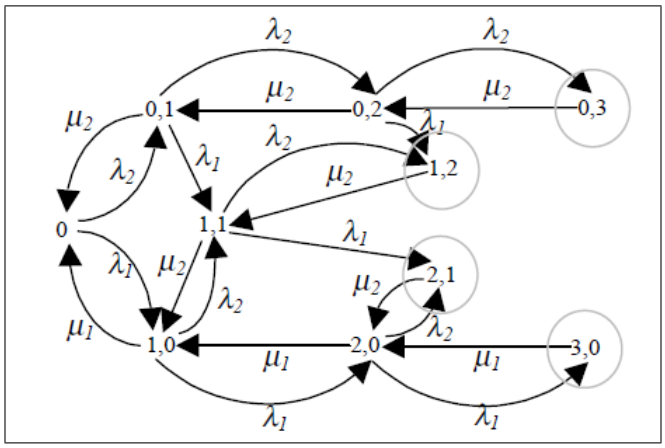

FIG. 2.2 – **Les probabilité de transition du système $M_2/M_2/1/3$ avec priorité absolue.**

Pour le système $M_2/M_2/1/\infty$, si on considère un système d'attente avec priorité absolue, les clients arrivent selon un processus de poisson, le taux des arrivées de la première classe est λ_1 (respectivement λ_2 pour la deuxième classe). Le service des clients prioritaires et non prioritaires se fait suivant la loi exponentielle μ_1 et μ_2 respectivement.

Notons $P_{n_1,n_2,r}$ la probabilité d'avoir n_1 clients prioritaires et n_2 clients non prioritaires au temps t et r indique le client en service, alors $P_{n_1,n_2,r}$ est donnée par le système d'équation suivant :

$$P_{0,0}(\lambda_1+\lambda_2) = P_{0,1}\mu_2 + P_{1,0}\mu_1, \quad si \;\; n_1 = 0 \;\; et \;\; n_2 = 0$$
$$P_{0,n_2}(\lambda_1+\lambda_2+\mu_2) = P_{0,n_2}\lambda_2 + P_{0,n_2+1}\mu_2, \quad si \;\; n_1 = 0 \;\; n_2 > 0$$
$$P_{n_1,0}(\lambda_1+\lambda_2+\mu_1) = P_{n_1-1,0}\lambda_1 + P_{n_1+1,0\mu_1} + P_{n_1+1,1}\mu_2, \quad si \;\; n_1 > 0 \;\; et \;\; n_2 = 0$$
$$P_{n_1,n_2}(\lambda_1+\lambda_2+\mu_1) = P_{n_1,n_2-1}\lambda_2 + P_{n_1-1,n_2}\lambda_1 + P_{n_1+1,n_2}\mu_1 + P_{n_1,n_2+1}\mu_2, si \quad n_1 > 0, \quad n_2 > 0$$

Les calcules étant longs et fastidieux, mais nous nous limiterons ici au la présentation de la démarche d'obtention des résultats. Il faut tout d'abord calculer la fonction génératrice :

$$H(y,z) = \sum_{n_1,n_2} P_{n_1,n_2} y^{n_1} z^{n_2}$$

A partir des équations de Chapman Kolmogorov on trouve :

$$H(y,z) = (\frac{1-\rho_1-\rho_2}{1-\eta-z\rho_2})(\frac{1-\eta}{1-\eta y})$$

Où
$\eta = \dfrac{1}{2\mu_1}[\mu_1+\lambda_1+\lambda_2(1-z)-\sqrt{[\mu_1+_1+\lambda_2(1-z)]^2-4\mu_1\lambda_1}]$ De la fonction génératrice $H(y,z)$ nous déduisons :
Le nombre moyen de clients prioritaire dans le système est

$$E[N_1] = (\frac{\partial H(y,z)}{\partial y})_{y=z=1} = \frac{\rho_1}{1-\rho_1} \tag{2.9}$$

Le nombre de clients non prioritaires dans le système est

$$\begin{aligned} E[N_2] &= (\frac{\partial H(y,z)}{\partial z})_{y=z=1} = \frac{\rho_2+E[N_1]\rho_2}{1-\rho_1-\rho_2} \\ &= \frac{\rho_2}{(1-\rho_1)(1-\rho_1-\rho_2)} \end{aligned} \tag{2.10}$$

D'après les formules (2.5) et (2.6), et au moyen de la formule de Little, en peut tirer le temps moyen de séjour dans le système des clients prioritaires et non prioritaires respectivement, donné par les deux formules suivantes :

$$E[S_1] = \frac{E[N_1]}{\lambda_1} = \frac{1}{\mu_1(1-\rho_1)} \tag{2.11}$$

$$E[S_2] = \frac{E[N_2]}{\lambda_2} = \frac{1}{\mu_2(1-\rho_1)(1-\rho_1-\rho_2)} \tag{2.12}$$

Pour l'état d'équilibre, le calcule des probabilités d'état s'effectuerait de façon suivante :

$$P_{n_1,n_2} = \frac{1}{n_1!n_2!}[\frac{\partial^{n_1+n_2}H(y,z)}{\partial y^{n_1}\partial z^{n_2}}]_{y=z=0}$$

En particulier, $P_{0,0} = 1-\rho_1-\rho_2$.

La condition de stabilité est alors donnée par $P_{0,0} > 0$, c'est à dire $\rho_1 + \rho_2 < 1$.

2.3 Les systèmes de file d'attente avec rappels

Les systèmes de files d'attente avec rappels se caractérisent par le fait que :" un client qui arrive et trouve tous les serveurs occupés, quitte le système définitivement, ou rappelle ultérieurement à des instant aléatoires, dans cette dernière on dit que le client entre dans l'*orbite* ".

le système de file d'attente avec rappel est défini complètement par la connaissance de ces éléments principaux : le processus des inter-arrivés, le mécanisme de service(la disponibilité et nombre de serveurs, la discipline de passage au service),en ajoutant, un élément décrivant la loi des répétition d'appels.

Le modèle de file d'attente avec rappels occupe une situation intermédiaire entre le modèle d'Erlang et le modèle classique d'attente FIFO, qui en constituent les modèles limites dans les cas de faible et forte intensité de rappels.

Les systèmes de file d'attente avec rappels ont trouvé leur première application lors de la modélisation du service d'abonnés dans un central téléphonique.

Plusieurs situations d'attente ont la caractéristique que les clients doivent rappeler, pour une certaine raison, pour être servis. Quand le service d'un client est insatisfait, il doit rappeler jusqu'à a l'accomplissement de son service. Ces modèles d'attente apparaissent dans la modélisation stochastique de plusieurs situations réelles. Par exemple, dans la transmission de données, un paquet transmis de la source 'a la destination peut être retournée et le processus doit se répéter jusqu'à ce que le paquet soit finalement transmis [25].

2.3.1 Description des systèmes de file d'attente avec rappels

considérons un système de file d'attente, de flot des arrivées primaires poissonien de paramètre λ, et ayant $s(s \geq 1)$ serveurs identique et indépendants, de distribution de service générale, de fonction de répartition $B(x)$. Il y a dans ce système $m - s$ position d'attente. A l'arrivée d'un client, s'il y a un ou plusieurs serveurs libres, celui-ci sera immédiatement pris en charge. sinon, s'il y a une position libre, le client rejoint la file d'attente. Lorsque tous les serveurs et toutes les positions d'attentes sont occupées, le client est obligatoirement quitte le système, soit définitivement avec une probabilité $1 - H_0$, soit temporairement avec une probabilité H_0 et rappelle ultérieurement, après un temps aléatoirement.

La capacité O de l'orbite peut être finie ou infinie. Dans le cas où O est finie, si l'orbite est pleine alors le client quitte le système définitivement.

Yang et Templeton introduisent, pour désigner les systèmes avec rappels, la notation suivante : $A/B/s/m/O/H$, où A et B décrivent respectivement la distribution du temps des inter-arrivés et la distribution de temps de service, s représente le nombre de serveurs, m est le nombre de position d'attente, et O est la capacité de l'orbite, ce système peut ne pas apparaitre dans la lorsque la capacité de l'orbite est infinie. La séquence $H = \{H_i, i \geq 0\}$ est la fonction de persévérance, où H_i est la probabilité qu'un client fasse une $(i+1)^{eme}$ tentative de rappel, après une i^{eme} tentative échouée. Si tous les clients sont persévérants $H_i = 1$, pour tout j, le symbole H pourra être également supprimé.

Aujourd'hui, il y a un centaines d'exemples qui peuvent être modélisé par un modèle de file d'attente avec rappels, en peut citer quelques exemples pour bien comprendre ce modèle.

1. Problème de réservation

C'est l'exemple le plus simple d'un client qui souhaite réserver dans un restaurant par téléphone, mais, il y a une unique ligne qui consacré à rependre aux requêtes des réservations. Ainsi, si le client appelle est trouve la ligne occupée, il renouvellera sa tentative après une certaine période de temps aléatoire avec la probabilité H_k qui est en pratique inférieure à 1, car le client ne peut pas 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'arrivée 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, ...

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 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.

3. Réseaux locaux CSMA

Dans les réseaux locaux se partageant un bus unique, l'un des protocoles de communication le plus généralement utilisé est appelé protocole non-persistant CSMA (Carrier Sense Multiple Access), c'est une méthode d'accès à un réseau local.

Un réseau local simple est composé de stations ou de terminaux inter-connectés 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. Une telle architecture de réseau d'ordinateurs local est appelée architecture en bus.

Des messages de longueurs variables arrivent aux stations du monde extérieur. En recevant le message, la station le découpe en un nombre fini de paquets de longueur fixe, et consulte immédiatement le bus pour voir s'il est occupé ou bien libre. Si le bus 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. Par contre, si le bus n'est pas libre, tous les paquets sont stockés dans le tampon (positions d'attente) et la station peut reconnecter le bus après une certaine période aléatoire.

Ce problème peut être modélisé comme un système d'attente avec rappels à un seul

serveur, qui est le bus, et les tampons des stations représentent l'orbite.

Le modèle d'attente avec rappels d'écrit ci-dessus est un modèle général. Plusieurs systèmes de files d'attente avec rappels peuvent être considérés comme des cas particuliers tels que : les systèmes sans buffer, les systèmes à un seul serveur, ...

par Le schéma général d'un système avec rappels est donné par la figue suivante.

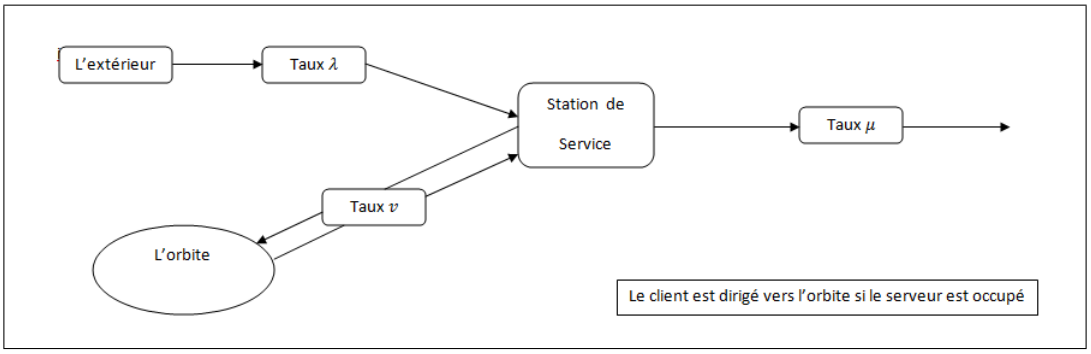

FIG. 2.3 – **Système générale d'une file d'attente avec rappel**

2.3.2 Quelques variantes des modèles de files d'attente avec rappels

Modèle markoviens

Les modèles markoviens sont utilisés pour décrire les systèmes dans lesquels les temps des inter-arrivés primaires, les durées de service et les temps inter rappels sont des variables aléatoires indépendantes et exponentiellement distribués.

Pour décrire un système de file d'attente avec rappels, dont le nombre de serveurs est m, le flux des inter-arrivés primaires est poissonien de paramètre λ, la durée de service des clients est exponentielle de paramètre μ et la durée entre deux rappels consécutifs d'une même source secondaire est aussi exponentielle de paramètre θ.

Soit le processus markovien $X(t) = \{C(t), N(t)\}_{t\geq 0}$, dont l'espace d'états est $S = \{0, 1, ..., m\} \times N$, où $C(t)$ est le nombre de client en cours de service à la date t, $N(t)$ est le nombre de client en régime de rappels à la date t.

Modèles semi-markovien

Lorsqu'on peut pas décrire le système d'attente avec rappels avec un processus markovien, en parle de modèle appelle semi-markovien. Donc, ce dernier peut être définie comme étant un modèle où l'un de ces processus n'est pas markovien(flux des inter-arrivés des clients primaires, la distribution de temps de service ou la distribution de temps de rappels).

Parmi les modèles semi-markovien avec rappels en peut citer celle qui est plus utilisé, le modèle $M/G/1$ avec rappels.

La description du modèle $M/G/1$ est faite par :

introduisant un processus des inter-arrivés des clients primaires qui est poissonien de paramètre λ, la durée de service τ est de loi générale de distribution $B(x)$. La durée entre deux rappels est successifs d'une même source secondaire est exponentielle de paramètre θ.
On décrit le système de la manière suivante :
On suppose que le $(i-1)^{ieme}$ appel termine son service à l'instant η_{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 $(i)^{ieme}$ rappels suivant n'entre en service qu'après un intervalle de temps R_i durant lequel le serveur est libre.
En peut citer d'autres systèmes de fille d'attente avec rappels qui sont :

- modèle avec rappels et serveur non fiable,

- modèle avec rappels et arrivées par groupes,

- modèle avec multi-classe de client,

- modèle avec deux types des clients impatients,

- les modèles avec priorité.

Quelques caractéristiques du système $M/G/1$ avec rappels

Dans cette partie en s'intéresse pas au détail de calcule, c'est-à-dire, comment on a trouver les mesures de performance, le nombre moyen de clients dans le système, le nombre moyen de clients dans l'orbite, mais, juste de donner ces caractéristiques afin

de l'utiliser dans le dernier chapitre, exactement pour la validation de simulateur approprie.

Nombre moyen de clients dans le système

$$\bar{n} = \rho + \frac{\lambda^2 E(\tau^2)}{2(1-\rho)} + \frac{\lambda\rho}{\theta(1-\rho)} \tag{2.13}$$

Nombre moyen de clients dans l'orbite

D'après la formule de Little, on a

$$\bar{n_0} = \bar{n} - \rho = \frac{\lambda^2 E(\tau^2)}{2(1-\rho)} + \frac{\lambda\rho}{\theta(1-\rho)} \tag{2.14}$$

Temps moyen de séjour

D'après la formule de Little $\bar{n} = \bar{w}\lambda$. On aura

$$\bar{w} = \frac{\rho}{\lambda} + \frac{\lambda E(\tau^2)}{2(1-\rho)} + \frac{\rho}{\theta(1-\rho)} \tag{2.15}$$

Quelques caractéristiques du système $M/G/1$ avec rappels et priorité

Dans cette section aussi, on doit juste donné quelques caractéristiques de performances pour qu'on puisse utiliser dans les sections qui se suit.

Dans ce système, on considère notre file de capacité infinie, et constituer d'un seul serveur. Le processus d'arrivé est poissonien de paramètre λ_1 et λ_2 respectivement, Les taux μ_1 et μ_2 sont des taux de services des clients de type 1 et de type 2 , et ν est le taux de rappel dans l'orbite, alors :

Le nombre moyen de client de type 1 dans le système

$$N_1 = \frac{\lambda_1(\lambda_1\beta_{1,2} + \lambda_2\beta_{2,2})}{2(1-\rho_1)} \tag{2.16}$$

Le nombre moyen de client de type 2 dans le système

$$N_2 = \frac{\lambda_1(\lambda_1\beta_{1,2} + \lambda_2\beta_{2,2})}{2(1-\rho_1)(1-\rho_1-\rho_2)} + \frac{\lambda_2(\rho_1+\rho_2)}{\nu(1-\rho_1-\rho_2)} \tag{2.17}$$

Avec $\beta_{1,2}$ et $\beta_{2,2}$ représente le moment d'ordre deux de la durée de service des clients de types 1 et 2 respectivement, c'est-à-dire :

$$\beta_{1,2} = E(\tau_1^2)$$
$$\beta_{2,2} = E(\tau_2^2)$$

avec τ_1 (τ_2) est les variables qui décrivent la distribution de service des clients de type 1 (type 2) respectivement.

2.4 Système de file d'attente avec vacance

La notion de vacances est introduite en général pour exploiter le temps inoccupé du serveur pour un autre travail secondaire dans le but d'améliorer la performance du système.

L'analyse de la modélisation par les systèmes d'attente avec vacances a été faite par un nombre considérable de travaux dans le passé et a été utilisée de manière réussie dans différents problèmes pratiques comme les systèmes de production, systèmes de communication et systèmes informatiques (voir la monographie de Doshi (1986) [21]). Une excellente étude compréhensive sur les modèles d'attente avec vacances peut être trouvée dans Teghem (1986) [23], Takagi (1991) [22] et Tian et Zhang (2006) [24].

2.4.1 Quelques cas modélisés par des systèmes de files d'attente

Exemple :
Dans le modèle opérationnel du serveur WWW, les requêtes HTTP arrivent au serveur WWW suivant un flux de Poisson et peuvent être interrompues par un utilisateur avant d'arriver au serveur WWW. Lorsque les requêtes arrivent dans le serveur WWW, une requêter est sélectionnée pour être servie et les autres entreront dans le tampon situé à l'intérieur du serveur WWW. Dans le tampon, chaque requêter attend un certain temps puis demande le service de nouveau. Un programme " daemon " est mis en application dans le serveur WWW pour diriger les requêtes de service à partir du tampon. À chaque fois qu'elle essaye mais échoue, elle attend un autre moment avant de réessayer de nouveau. Si

la page web cible est située dans le même serveur WWW, la requêter peut retourner au serveur. Pour garder le serveur WWW en bon fonctionnement, des activités de maintenance telles que scanner les virus peuvent être réalisées lorsque le serveur WWW est inactif. Ce type de maintenance peut être programmé pour fonctionner sur une base régulière. Cependant, ces activités de maintenance ne se répètent pas continuellement. Lorsque ces activités sont achevées, le serveur WWW entrera de nouveau en état d'inactivité et attendra l'arrivée de nouvelles requêtes.
Dans ce scénario, le tampon dans le serveur WWW, le serveur WWW, la politique de retransmission et les activités de maintenance en période d'inactivité correspondent respectivement à l'orbite, le serveur, la discipline de rappels et la politique de vacances, dans la terminologie de files d'attente.
Exemple :
Dans le modèle de transfert d'un système e-mail, le système e-mail utilise le protocole SMTP (Simple Mail Transfer Protocol) pour délivrer les messages entre les serveurs e-mail.
Quand un programme de transfert de e-mail contacte un serveur sur une machine éloignée, il forme une connections TCP (Transfer Connection Program) à travers laquelle il communique. Une fois la connections est en place, les deux programmes suivent SMTP qui permet à l'expéditeur de s'identifier, spécifier un destinataire, et transférer un message e-mail. Ce dernier peut essayer de façon répétée d'envoyer le message de contact au serveur cible jusqu'à ce qu'il devienne opérationnel. Typiquement, les messages de contact arrivent au serveur e-mail suivant un flot de Poisson. Ces messages de contact peuvent être interrompus par un expéditeur avant 'arriver au serveur e-mail. Quand les messages arrivent au serveur mail, un message est sélectionné pour service et les autres joindront le buffer. Dans le buffer, chaque message attend un certain temps pour demander encore une fois le service. Il y a un programme mis en application et implémenté au serveur mail pour diriger les requêtes de service du buffer. À chaque fois qu'une requêter essaye mais échoue, elle attendra un autre moment avant de recommencer de nouveau. Le serveur cible est le même que le serveur mail de l'expéditeur et le message envoyé peut revenir au serveur pour demander le service. Pour garder le serveur mail en bon fonctionnement, scanner les virus est une importante activité de maintenance pour le serveur e-mail. Elle peut être réalisée lorsque le serveur e-mail est inactif. Cependant, ces activités de maintenance ne se répètent pas continuellement.
Quand ces activités sont finies, le serveur e-mail entrera encore une fois en état d'inactivité (oisiveté) et attendra l'arrivée des messages de contact. Parce qu'il n'y a pas de mécanisme pour enregistrer le nombre de messages de contact parvenant couramment des différents expéditeurs, il est approprié de concevoir un programme pour collectionner l'information des messages de contact pour des raisons d'efficacité.
Dans ce scénario, le buffer dans le serveur mail de l'expéditeur, le serveur e-mail destina-

taire, la politique de retransmission, et les activités de maintenance correspondent respectivement à l'orbite, le serveur, la discipline de rappels, et la politique de vacances dans la terminologie de files d'attente.

Classification des modèles d'attente avec vacances

Les files d'attente avec vacances peuvent être classifiées de différentes façons. Les disciplines de service les plus connues sont :
- La discipline de service exhaustif : Dans un système avec vacances et service exhaustif, chaque fois que le serveur revient d'une vacance, il servira tous les clients en attente dans le système avant de commencer une autre vacance.
- La discipline de service avec barrière : Dans le cas du service avec barrière, quand le serveur revient d'une vacance, il sert seulement les clients qui étaient en attente dans la file 'a son arrivée. Autrement dit, d'es l'arrivée du serveur, il met une barrière fictive derrière les clients en attente dans la file et ne prend une autre vacance qu'une fois que tous les clients qui étaient présents à son arrivée soient servis.
- La discipline de service limité : Dans un système avec service limité, on se fixe un nombre k. 'A son retour de la vacance, le serveur servira au plus k clients et commencera ensuite une autre vacance. Ainsi, le serveur sert jusqu'à ce que la file d'attente soit vide ou bien jusqu'à ce que k clients soient servis, ensuite il prend une autre vacance. Lors de l'analyse de ces système, on a remarquer des difficultés lors de l'application et aussi même pour l'application des méthodes appropriés. Pour bien palier à ces difficultés plusieurs méthodes approximatives d'analyse ont été développés. Parmi les principales approches, en trouve la méthode de simulation à temps discret qui nous permet de conduire à des estimations quantitatives des caractéristiques de système étudie.

Conclusion

Dans ce chapitre, nous avons présenter quelques résultats analytique sur les systèmes de file d'attente, à savoir, les systèmes prioritaires, les systèmes avec rappels et priorité et celle avec vacance, dont le but est de l'utiliser dans le chapitre quatre pour pouvoir valider les modèles appropriés.

Chapitre 3

Systèmes prioritaires et système avec rappels et vacance

introduction

Dans le cadre général, les unités (clients, appels, message,...) qui s'accumulent dans une file, sont traités dans l'ordre de leur arrivée. C'est ce qu'on appelle la procédure PAPS (Premier Arrivé Premier Servi). D'autres procédures peuvent être mise en œuvre, parmi lesquelles, on présente l'admission avec priorité [41].

Les programmes dans un ordinateur ou les paquets dans un réseau informatique peuvent ne pas être traités de la même façon : quelques uns peuvent recevoir un traitement préférentiel. Dans une salle d'urgence des hôpitaux, une personne inconscients ou ayant une crise cardiaque aura la priorité au service par rapport aux autres qui ont subi une blessure mineure. Les systèmes d'attente dans lesquels des clients reçoivent des traitements préférentiels sont dits systèmes d'attente prioritaires.

donc, dans ce chapitre, il nous a paru opportun de décrire le comportement de types de file d'attente, à savoir, les systèmes prioritaires et les systèmes avec rappels et vacance.

Les systèmes avec rappels apparaissent dans la modélisation des phénomènes du genre :" service des avions dans un aéroport, comportement des processus(tâches, programme,...) dans un réseau informatique constitué d'un ordinateur central et d'un ensemble de périphériques(terminaux)[41].

Les systèmes d'attente avec vacance est introduite en générale dont le but est d'exploiter le temps inoccupé du serveur, dont le but est d'améliorer les caractéristiques de performances, ce système est analyser par plusieurs chercheurs parmi eux, en peut citer Doshi(1986), Teghem (1986), Takagi (1991), Tian et Zhang (2006)...

3.1 Les systèmes de files d'attente avec priorité

Il n'est pas rare, dans la vie courante de rencontrer des systèmes d'attente avec plusieurs types de clients, où certains sont prioritaires par rapport à d'autres. Pour mieux illustrer cette situation, considérons deux systèmes de files d'attente (M/M/1 et M/G/1) avec m types de clients. Les clients de type i arrivent indépendamment des autres types de clients suivant un taux $\lambda_i, i = 1, 2, ..., m$, Le client de type i à la i^{ieme} priorité et la i^{ieme} priorité est supérieur à la k^{ieme} priorité pour tout $k > i$.

On distingue deux types de priorités : **priorité relative et priorité absolue.**

Dans ce qui suit nous allons montrer comment obtenir certaines caractéristiques d'un système prioritaire dans les deux cas.

Priorité relative

La priorité relative se caractérise par le fait qu'un client de priorité supérieur ne peut en aucun cas interrompre le service d'un client priorité inférieure. Il doit attendre (s'il le désire bien sûr) jusqu'à ce que le service du client de priorité inférieure soit fini avant d'être servi. Cette discipline a été introduite par Cobham [3] et étudié par plusieurs chercheurs (Holley [2], Dressin et Reich [4], Morse [5] et autres). Le système d'attente avec priorité relative peut aussi être considéré comme un système d'attente à serveur non fiable, la panne est prise en considération après que le client en service ait terminé son service. Ce modèle est connu sous le non " **Breakdown postponable interruption** ". Il a été étudié par Keilson [6], Gaver [7] et Hodgson [8].

3.1.1 Système $M_R/G_R/1$ avec priorité relative

Dans ce système, il existe R classes indépendants de clients qui arrivent suivant des flux poissonniens de taux $\lambda_k, k = 1, 2, ..., R$, correspondant à chaque classe. Le taux arrivée total est $\sum_{k=1}^{R} \lambda_k$.

La distribution de temps de service est générale, de fonction de réparation $B_k(.)$ et de moyenne $E(B_k) = \dfrac{1}{\mu_k}$ pour les clients de classe k. On défini $\rho_k = \dfrac{\lambda_k}{\mu_k}$, La condition $\sum_{i=1}^{R} \rho_i < 1$, dite d'ergodicité géométrique du système, est supposée être vérifiée. On s'intéresse principalement au calcul du temps moyen d'attente d'un client dans la file (c'est-à-dire, le temps qu'il passe dans la file avant le début de son service).
Le temps d'attente W_k de dernier client arrivé C_k, de classe k contient trois composantes

[11]

-Le temps résiduel W_0 de client en cours de service,

-Le temps de service $W_{1,k}$ de tous les clients qui sont déjà présents, à l'arrivée du client C_k et qui ont une priorité égale ou supérieure à celle de C_k,

-Le temps de service $W_{2,k}$ de tous les clients qui sont arrivés pendant le temps d'attente du client C_k et qui ont une priorité supérieure à celle de C_k.
Donc, le temps d'attente du client C_k est donnée par :

$$W_k = W_0 + W_{1,k} + W_{2,k},\ k = 1, 2, ..., R$$

Calcul de temps moyen résiduel du client en cours de service

soit α ="le temps de service résiduel", et β ="le client en cours de service est de classe i ",donc, le temps moyen résiduel du client en cours de service prend cette forme :

$$\begin{aligned} E[W_0] &= \sum_{i=1}^{R} E(\alpha/\beta).P(\beta) \\ &= \sum_{i=1}^{R} \frac{E(B_i^2)}{2E(B_i)}.P(\beta) \\ &= \sum_{i=1}^{R} \frac{E(B_i^2)}{2E(B_i)}.\rho_i \end{aligned}$$

Si on prend l'avantage que $\rho_i = \lambda_i E(B_i)$, on arrive à

$$E(W_0) = \sum_{i=1}^{R} \frac{\lambda_i E(B_i^2)}{2} \tag{3.1}$$

Par conséquent, le temps résiduel moyen de service d'un client dépend des moments d'ordre 2 des distributions des temps de service.

Calcul de temps moyen de service de tous les clients qui sont déjà présents à l'arrivée du client C_k et qui ont une priorité égale ou supérieure à celle de C_k

Soit $N_{1,k}$ le nombre de clients de classe i qui sont déjà présents dans la file à l'arrivée du client C_k (de classe k) et qui sont service avant C_k, On a donc :

$$E[W_{1,k}] = \sum_{i=1}^{k} E[N_{i,k}].E[B_i]$$

Puisque les temps de service sont indépendants. On peut maintenant appliquer la formule de Little :

$$E[N_{i,k}] = \lambda_i E[W_i]$$

Où W_i dénote le temps d'attente dans la file pour les clients de classe i. on arrive à :

$$E[W_{1,k}] = \sum_{i}^{k} E[N_{i,k}].E[B_i] = \sum_{i=1}^{k} \lambda_i E[W_i]E[B_i]$$

D'où

$$E[W_{1,k}] = \sum_{i=1}^{k} \rho_i E[W_i], \quad (k = 1, 2, ..., R) \tag{3.2}$$

Calcul de temps moyen de service de tous les clients qui sont arrivés pendant le temps d'attente du client C_k et qui ont une priorité supérieure à celle de C_k

Soit $M_{i,k}$ le nombre de clients de classe i arrivant durant le temps d'attente W_k du client C_k et qui reçoivent leur service avant le client C_k (c'est-à-dire, ils ont une priorité supérieure).

A l'aide de la formule du Little, on exprime $E[M_{i,k}]$ par :

$$E[M_{i,k}] = \lambda_i E[W_i]$$

Donc on a :

$$E[W_{2,k}] = \sum_{i=1}^{k-1} E[M_{i,k}].E[B_i] = \sum_{i=1}^{k-1} \lambda_i E[W_k]E[B_i]$$

D'où

$$E[W_{2,k}] = E[W_k] \sum_{i=1}^{k-1} \rho_i, \quad (k = 1, 2, ..., R) \tag{3.3}$$

Temps moyen d'attente $E[W_k]$, d'un client de classe k

On va inclure les trois résultats trouvés (2.1),(2.2) et (2.3), dans l'expression de base de $E[W_k]$:

$$\begin{aligned} E[W_k] &= E[W_0] + E[W_{1,k}] + E[W_{2,k}] \\ &= E[W_0] + \sum_{i=1}^{k} \rho_i E[W_i] + E[W_k] \sum_{i=1}^{k-1} \rho_i \end{aligned}$$

De cette dernière équation, on trouve :

$$E[W_k] = \frac{1}{1 - \sum_{i=1}^{k} \rho_i} (E[W_0] \sum_{i=1}^{k-1} \rho_i E[W_i]),$$

pour $(k = 1, 2, ..., R)$ on obtient un système d'équations linéaires, qu'on peut résoudre d'une manière récursive. On utilisera cette courte notation :

$$\sigma_k = \sum_{i=1}^{k} \rho_i,$$

Par cette natation on peut écrire pour $k = 1$:

$$E[W_1] = \frac{E[W_0]}{1 - \sigma_1}$$

Pour $k = 2$ on obtient :

$$\begin{aligned} E[W_2] &= \frac{1}{1 - \sigma_2} (E[W_0] + \rho_1 \frac{E[W_0]}{1 - \sigma_1}) \\ &= \frac{E[w_0]}{1 - \sigma_2} (\frac{1 - \rho_1 + \rho_1}{1 - \sigma_1}) \end{aligned}$$

On aura donc,

$$E[W_2] = \frac{E[W_0]}{(1 - \sigma_2)(1 - \sigma_1)}$$

Il est facile à présent de prouver d'une manière récursive que pour $k < R$:

$$E[W_k] = \frac{E[w_0]}{(1 - \sigma_k)(1 - \sigma_{k-1})}$$

Temps moyen d'attente $E[W_k]$ est donné par :

$$E[W_k] = \frac{\sum_{i=1}^{R} \lambda_i E[B_i^2]}{2(1 - \sum_{i=1}^{k} \rho_i)(1 - \sum_{i=1}^{k-1} \rho_i)}, \quad (k = 1, 2, ..., R)$$

Temps moyen de séjour d'un client de classe k dans le système

$$E[S_k] = E[W_k] + E[B_k] = \frac{\sum_{i=1}^{R} \lambda_i E[B_i^2]}{2(1-\sum_{i=1}^{k} \rho_i)(1-\sum_{i=1}^{k-1} \rho_i)} + E[B_k],\ (k = 1, 2, ..., R)$$

A l'aide de la formule de Little, on trouve aussi :

Nombre moyen de clients de classe k dans la file

$$Q_k = \lambda_k E[W_k] = \lambda_k \frac{\sum_{i=1}^{R} \lambda_i E[B_i^2]}{2(1-\sum_{i=1}^{k} \rho_i)(1-\sum_{i=1}^{k-1} \rho_i)},\ (k = 1, 2, ..., R)$$

Nombre moyen de clients de classe k dans le système

$$L_k = \lambda_k E[S_k] = \lambda_k \frac{\sum_{i=1}^{R} \lambda_i E[B_i^2]}{2(1-\sum_{i=1}^{k} \rho_i)(1-\sum_{i=1}^{k-1} \rho_i)} + \rho_k, \quad (k = 1, 2, ..., R)$$

3.1.2 Système $M_2/G_2/1$ avec priorité relative

Ce système a été étudié dans l'article [31], dans ce système, les clients arrivent en deux calasse, selon un processus de poisson de paramètre λ_1 et λ_2, La distribution de service est générale de fonction B_i, $i = 1, 2$ et de moyenne $\frac{1}{\mu_1}$ et $\frac{1}{\mu_2}$.

Le temps moyen d'attente d'un client dans la file

Client prioritaire :

$$E[W_1] = \frac{(\lambda_1 E[B_1^2] + \lambda_2 E[B_2^2])}{2(1-\rho_1)}$$

Client non prioritaire :

$$E[W_2] = \frac{(\lambda_1 E[B_1^2] + \lambda_2 E[B_2^2])}{2(1-\rho_1-\rho_2)(1-\rho_1)}$$

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

Client prioritaire :

$$E[S_1] = \frac{(\lambda_1 E[B_1^2] + \lambda_2 E[B_2^2])}{2(1-\rho_1)} + E[B_1]$$

Client non prioritaire :

$$E[S_2] = \frac{(\lambda_1 E[B_1^2] + \lambda_2 E[B_2^2])}{2(1-\rho_1-\rho_2)(1-\rho_1)} + E[B_2]$$

3.2 Le système $M_2/M_2/1$ avec priorité relative

Ce système a été étudié par plusieurs chercheurs, parmi eux Rupert dans l'article [28], Guy dans [29] et D.Gross dans le livre [30].

Considérons un système $M_2/M_2/1/3$ avec priorité relative dans lequel arrivent deux classes de clients que nous appelons :
Classe 1 : **Clients non prioritaires.**
Classe 2 : **Clients prioritaires.**

Les deux types des clients arrivent indépendamment l'un de l'autre suivant un processus poissonnien avec respectivement les taux λ_1 et λ_2. Le service des clients prioritaires et non prioritaires se fait suivant la loi exponentielle μ_1 et μ_2 respectivement.

Soit $P_{n,m,r}(t)$ = P(au temps t,avoir n clients de classe 1, m clients de classe 2,r) avec :

$$r = \begin{cases} 1 & si \text{ le client en service est de classe 1} \\ 2 & si \text{le client en service est de classe 2} \end{cases}$$

A l'état d'équilibre, nous obtenons pour $\rho = \sum_{i=1}^{2} \frac{\lambda_i}{\mu_i}$, le système d'équation suivant :

$$\begin{aligned}
P_{0,0}(\lambda_1+\lambda_2) &= P_{0,1,2}\mu_2 + P_{1,0,1}\mu_1 \\
P_{0,1,2}(\lambda_1+\lambda_2+\mu_2) &= P_{0,0}\lambda_2 + P_{0,2,2}\mu_2 + P_{1,1,1}\mu_1 \\
P_{1,0,1}(\lambda_1+\lambda_2+\mu_1) &= P_{0,0}\lambda_1 + P_{2,0,1}\mu_1 + P_{1,1,2}\mu_2 \\
P_{1,1,2}(\lambda_1+\lambda_2+\mu_2) &= P_{1,2,2}\mu_2 + P_{2,1,1}\mu_1 + P_{0,1,2}\lambda_2 \\
P_{0,2,2}(\lambda_1+\lambda_2+\mu_2) &= P_{0,1,2}\lambda_2 + P_{1,2,1}\mu_1 + P_{0,3,2}\mu_2 \\
P_{2,0,1}(\lambda_1+\lambda_2+\mu_1) &= P_{1,0,1}\lambda_1 + P_{2,1,2}\mu_2 + P_{3,0,1}\mu_1
\end{aligned}$$

Ces probabilités de transition sont illustrées dans la figure suivante :

Pour les systèmes $M_2/M_2/1/\infty$, si on considère un système d'attente avec priorité relative,

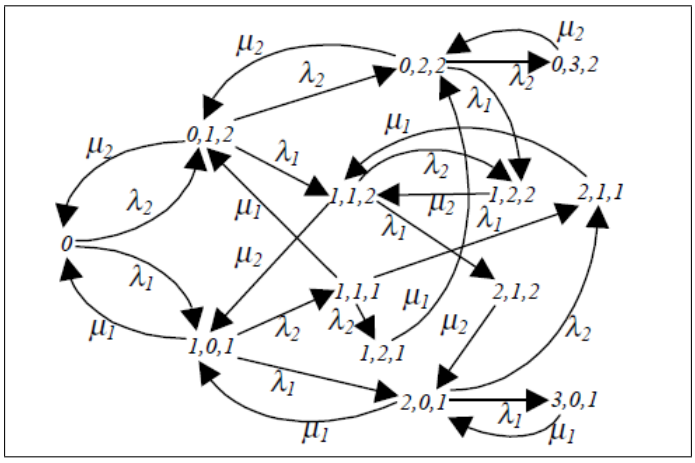

FIG. 3.1 – **Les probabilité de transition du système $M_2/M_2/1/3$ avec priorité relative.**

les clients arrivent selon un processus de poisson, le taux des arrivées de la première classe est λ_1 (respectivement λ_2 pour la deuxième classe). Les services des clients prioritaires et non prioritaires se fait suivant la loi exponentielle de taux μ_1 et μ_2 respectivement. Notons $P_{n,m,r}$ la probabilité d'avoir n clients non prioritaires et m clients prioritaires au temps t et r indique le client en service. Alors $P_{n,m,r}$ est donnée par le système d'équation suivant :

$$\begin{aligned}
(\lambda_1+\lambda_2)P_{0,0} &= P_{0,1,2}\mu_2 + P_{1,0,1}\mu_1 \\
(\lambda_1+\lambda_2+\mu_1)P_{1,0,1} &= P_{0,0}\lambda 1 + P_{2,0,1}\mu_1 + P_{1,1,2}\mu_2 \\
(\lambda_1+\lambda_2+\mu_1)P_{0,1,2} &= P_{0,0}\lambda 2 + P_{0,2,2}\mu_2 + P_{1,1,1}\mu_1 \\
(\lambda_1+\lambda_2+\mu_1)P_{m,0,1} &= P_{m-1,0,1}\lambda 1 + P_{m,1,2}\mu_2 + P_{m+1,0,1}\mu_1 \\
(\lambda_1+\lambda_2+\mu_2)P_{0,n,2} &= P_{0,n-1,2}\lambda 1 + P_{0,n+1,2}\mu_2 + P_{1,n,1}\mu_1 \\
(\lambda_1+\lambda_2+\mu_2)P_{1,n,1} &= P_{1,n-1,1}\lambda 2 + P_{1,n+1,2}\mu_2 + P_{2,n,1}\mu_1 \\
(\lambda_1+\lambda_2+\mu_1)P_{m,n,1} &= P_{m-1,n,1}\lambda 1 + P_{m,n-1,1}\lambda_2 + P_{m+1,n,1}\mu_1 + P_{m,n+1,2}\mu_2, m>1, n>0 \\
(\lambda_1+\mu_2)P_{m,n,2} &= P_{m,n-1,2}\lambda_1, m>0, n>1
\end{aligned}$$

Nous obtenons à l'état d'équilibre la probabilité $P_n(t)$. La probabilité d'avoir au temps t, n clients de classe 1, m clients de classe 2.

$$P_n = \sum_{m=0}^{n-1}(P_{n-m,m,1} + P_{n,n-m,2}) = (1-\rho)\rho,\ (n>1)$$

En utilisant les fonctions génératrices marginales :

$$\begin{aligned}
P_{m,1}(z) &= \sum_{n=0}^{\infty} P_{n,m,1}z^n, \\
P_{m,2}(z) &= \sum_{m=0}^{\infty} P_{n,m,2}z^m,
\end{aligned}$$

$$H_1(y,z) = \sum_{m=1}^{\infty} y^m P_{m,1}(z), \text{avec}, H_1(1,1) = \frac{\lambda_1}{\mu_1}$$
$$H_2(y,z) = \sum_{m=0}^{\infty} y^m P_{m,2}(z), \text{avec}, H_2(1,1) = \frac{\lambda_2}{\mu_2}$$

On obtient la fonction génératrice jointe $H(y,z)$

$$\begin{aligned} H(y,z) &= H_1(y,z) + H_2(y,z) + P_0 \\ &= \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} y^m z^n (P_{m,n,1} + P_{m,n,2}) \sum_{m=1}^{\infty} y^m P_{m,0,1} + \sum_{n=1}^{\infty} z^n P_{0,n,2} + P_0 \end{aligned}$$

D'après le calcule de le fonction génératrice jointe $H(y,z)$, en peut l'utiliser afin de trouver le nombre moyen des clients prioritaires dans le système L_1, le nombre moyen de clients non prioritaires dans le système L_2, alors :

$$\begin{aligned} L_1 &= \frac{\partial H(y,z)}{\partial y}_{/z=y=1} = L_{q1} + \frac{\lambda_1}{\mu_1} \\ &= \frac{\rho_1(1+\rho_2)}{1-\rho_1} \end{aligned} \tag{3.4}$$

$$\begin{aligned} L_2 &= \frac{\partial H(y,z)}{\partial z}_{/z=y=1} = L_{q2} + \frac{\lambda_2}{\mu_2} \\ &= \frac{\rho_2(1-\rho_1(1-\rho_1-\rho_2))}{(1-\rho_1)(1-\rho_1-\rho_2)} \end{aligned} \tag{3.5}$$

Au moyen des formules (2.4),(2.5) et de la formule de Little, on aura, le temps moyen de séjour des clients prioritaires et non prioritaires donné par les deux formules suivantes :

$$E[S_1] = \frac{L_1}{\lambda_1} = \frac{1+\rho_2}{\mu_1(1-\rho_1)} \tag{3.6}$$
$$E[S_2] = \frac{1-\rho_1(1-\rho_1-\rho_2)}{\mu_2(1-\rho_1)(1-\rho_1-\rho_2)} \tag{3.7}$$

3.2.1 Priorité absolue

Contrairement au cas précédent, ici, un client de priorité supérieure à la droit d'interrompre le service d'un client de priorité inférieure pour se faire servir. Le service interrompu sera alors repris à partir du point où il était suspendu.

Les premiers articles publiés concernant les systèmes d'attente avec priorité absolue sont les articles de White et Christie [34]. Par la suite sont apparus les articles Stephan [32],

Miller [33], Jaiswall [38] et Welch [39]. Takacs et Chang [36] ont étudié la priorité absolue avec une source infinie et différentes supposition à propos de la distribution de service. Les points communs entre les systèmes d'attente avec priorité absolue et les systèmes à serveurs non fiable (Pannes) ont été étudiés pour la première fois par White et Christie [34]. Keilson [35], Gaver [40], Avi-Itzhak et Naor [27], ont utilisé cette ressemblance pour étudier la priorité absolue à partir des systèmes à serveur non faible (breakdown models). Les systèmes prioritaires avec source finie (dans laquelle au moins un type de client à source finie) ont été étudiés par Avi-tzhak et Naor [27].

3.2.2 Système $M_R/G_R/1$ avec priorité absolue

Gelenbe dans l'article [21] à étudié le système d'attente de type M/G/1,dans ce système les clients arrivent en k classes indépendantes suivant un processus de Poisson. Le taux d'arrivée pour chaque classe est λ_i et le taux d'arrivée totale est $\lambda = \lambda_1 + \lambda_2 + ... + \lambda_r$. la distribution de service est générale de moyenne $\frac{1}{\mu}$, dans un système avec priorité absolue conservatrice, Le client interrompu reprend son service au point où il est interrompu.

Dans ce cas, les clients de priorités inférieures sont totalement "invisible" et n'affectent en aucun cas la file des clients de hautes priorités. Alors, pour un client de classe k, on peut procéder comme si $\lambda_{k+1} = \lambda_{k+2} = ... = \lambda_R = 0$.

Analysons, à présent, un système dans lequel les classes $k+1$ jusqu'à R n'existe pas et comparons les temps d'attente d'un client. Le temps moyen d'attente W_k, du dernier client arrivé C_k de classe k, avant qu'il entre en service pour la premier fois, est le même que dans le cas de priorité relative avec k classe de priorité.
on a

$$E[W_k] = \frac{\sum_{i=1}^{k} \lambda_i E[B_i^2]}{2(1-\sigma_k)(1-\sigma_{k-1})}$$

avec $\sigma_k = \sum_{i=1}^{k} \rho_i$

Le temps moyen de séjour d'un client dans le système S_j, contient trois parties :

$$E[S_j] = E[W_j] + E[B_j] + E[I_j], j = 1, 2, ..., k \tag{3.8}$$

Le premier terme $E[W_j]$, est le temps moyen d'attente avant que le client entre en service pour la première fois, le second terme $E[B_j]$, est le temps moyen de service de client et le troisième terme $E[I_j]$, est l'espérance de la variable aléatoire I_j, représentant

le temps total d'interruption du client durant son service le temps total d'interruption consiste en deux parties : la somme des temps de service des clients qui ont interrompu le service de client C_k, et la somme des temps de service des clients qui sont arrivés durant les périodes dans lesquelles le client est déjà interrompu, On obtient alors :

$$\begin{aligned} E[I_j] &= E[B_j]\sum_{i=1}^{k-1}\lambda_i E[B_i] + E[I_j]\sum_{i=1}^{k-1}\lambda_i E[B_i] \\ &= E[B_j]\sum_{i=1}^{k-1}\rho_i + E[I_j]\sum_{i=1}^{k-1}\rho_i \\ &= \frac{E[B_j]\sum_{i=1}^{k-1}\rho_i}{1-\sum_{i=1}^{k-1}\rho_i}, j = 1, ..., k \end{aligned}$$

La somme du seconde et troisième terme de 2.4, le temps moyen effectif de service et les temps totaux moyens d'interruption, est ce qu'on peut appeler le temps généralisé de service.

$$E[B_j] + E[I_j] = \frac{E[B_j]}{1-\sum_{i=1}^{k-1}\rho_i}$$

A partir les résultats précédents, on trouve :

$$E[S_j] = \frac{\sum_{i=1}^{k-1}\lambda_i E[B_i^2]}{2(1-\sigma_k)(1-\sigma_{k-1})} + \frac{E[B_j]}{1-\sum_{i=1}^{k-1}\rho_i}, j = 1, ..., k$$

A l'aide de la formule de Little, on trouve aussi :

$$Q_j = \lambda_j E[W_j] = \lambda_j \frac{\sum_{i=1}^{k}\lambda_i E[B_i^2]}{2(1-\sum_{i=1}^{k}\rho_i)(1-\sum_{i=1}^{k-1}\rho_i)}, k = 1, ..., R$$

Nombre moyen de clients de classe k dans le système

$$L_j = \lambda_j E[S_j] = \lambda_j E[W_j] = \lambda_j \frac{\sum_{i=1}^{k}\lambda_i E[B_i^2]}{2(1-\sum_{i=1}^{k}\rho_i)(1-\sum_{i=1}^{k-1}\rho_i)} + \frac{\rho_j}{1-\sum_{i=1}^{k-1}\rho_i}, j = 1, 2, ..., k$$

3.2.3 Système $M_2/G_2/1$ avec priorité absolue

Dans ce système, les clients arrivent en deux calasse, selon un processus de poisson de paramètre λ_1 et λ_2, La distribution de service des clients prioritaires et non prioritaires

suivent des lois générales, de fonction de répartition $B_i(.)$,$i = 1, 2$ et de moyenne $\frac{1}{\mu_1}$ et $\frac{1}{\mu_2}$.

Le temps moyen d'attente d'un client dans la file
Client prioritaire :

$$E[W_1] = \frac{\lambda_1 E[B_1^2]}{2(1-\rho_1)}$$

Client non prioritaire :

$$E[W_2] = \frac{(\lambda_1 E[B_1^2] + \lambda_2 E[B_2^2])}{2(1-\rho_1)(1-\rho_1-\rho_2)}$$

Le temps moyen de séjour d'un client dans le système
Client prioritaire :

$$E[S_1] = \frac{\lambda_1 E[B_1^2]}{2(1-\rho_1)} + E[B_1]$$

Client non prioritaire :

$$E[S_2] = \frac{(\lambda_1 E[B_1^2] + \lambda_2 E[B_2^2])}{2(1-\rho_1)(1-\rho_1\rho_2)} + \frac{E[B_{12}]}{1-\rho_1}$$

3.2.4 Le système $M_2/M_2/1$ avec priorité absolue

Ce système a été étudié Avi-Itzhak dans l'article [37], il a considéré un système $M_2/M_2/1/3$ avec priorité absolue dans lequel arrivent deux classes :
Classe 1 : **Clients non prioritaires.**
Classe 2 : **Clients prioritaires.**

Les deux types des clients arrivent indépendamment l'un de l'autre suivant un processus poissonnien avec respectivement les taux λ_1 et λ_2, Le service des clients prioritaires et non prioritaires se fait suivant la loi exponentielle μ_1 et μ_2 respectivement.

Écrivons les équations de Chapman Kolmogrov à l'état d'équilibre. Notons P_{n_1,n_2} la probabilité d'avoir n_1 clients de classe 1 et n_2 clients de classe 2 dans le système et $P_{0,0}$ la probabilité d'avoir 0 clients de classe 1, 0 clients de classe 2 dans le système.

$$
\begin{aligned}
P_{0,0}(\lambda_1+\lambda_2) &= P_{0,1}\mu_2 + P_{1,0}\mu_1,\\
P_{0,1}(\lambda_1+\lambda_2+\mu_2) &= P_{0,0}\lambda_2 + P_{0,2}\mu_2\\
P_{1,0}(\lambda_1+\lambda_2+\mu_1) &= P_{0,0}\lambda_1 + P_{2,0}\mu_1 + P_{1,1}\mu_2\\
P_{1,1}(\lambda_1+\lambda_2+\mu_2) &= P_{1,2}\mu_2 + P_{1,0}\lambda_2 + P_{0,1}\lambda_1\\
P_{0,2}(\lambda_1+\lambda_2+\mu_2) &= \lambda_2 P_{0,1} + \mu_2 P_{0,3} + \mu_1 P_{1,2}\\
P_{2,0}(\lambda_1+\lambda_2+\mu_1) &= \lambda_1 P_{1,0} + \mu_2 P_{2,1} + \mu_1 P_{3,0}\\
\mu_2 P_{1,2} &= \lambda_1 P_{0,2} + \lambda_2 P_{1,1}\\
\mu_2 P_{2,1} &= \lambda_1 P_{1,1} + \lambda_2 P_{2,0}\\
\mu_2 P_{0,3} &= \lambda_2 P_{0,2}\\
\mu_1 P_{3,0} &= \lambda_1 P_{2,0}
\end{aligned}
$$

Les transitions de la chaîne sont illustrées dans la figure suivante :

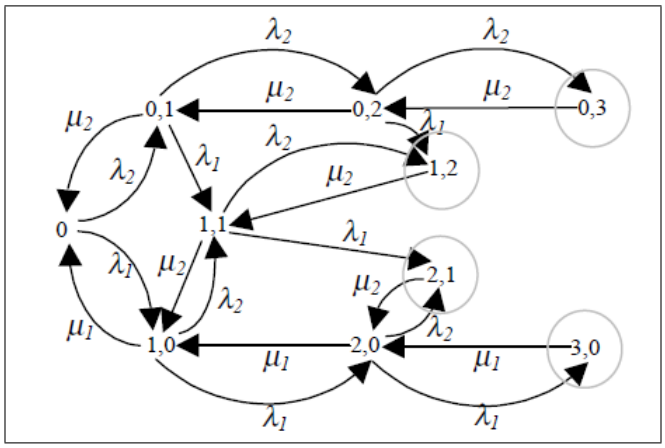

FIG. 3.2 – **Les probabilité de transition du système $M_2/M_2/1/3$ avec priorité absolue.**

Pour le système $M_2/M_2/1/\infty$, si on considère un système d'attente avec priorité absolue, les clients arrivent selon un processus de poisson, le taux des arrivées de la première classe est λ_1 (respectivement λ_2 pour la deuxième classe). Le service des clients prioritaires et non prioritaires se fait suivant la loi exponentielle μ_1 et μ_2 respectivement.

Notons $P_{n_1,n_2,r}$ la probabilité d'avoir n_1 clients prioritaires et n_2 clients non prioritaires au temps t et r indique le client en service, alors $P_{n_1,n_2,r}$ est donnée par le système d'équation suivant :

$$
\begin{aligned}
P_{0,0}(\lambda_1+\lambda_2) &= P_{0,1}\mu_2 + P_{1,0}\mu_1, \quad si \;\; n_1=0 \;\; et \;\; n_2=0\\
P_{0,n_2}(\lambda_1+\lambda_2+\mu_2) &= P_{0,n_2}\lambda_2 + P_{0,n_2+1}\mu_2, \quad si \;\; n_1=0 \;\; n_2>0\\
P_{n_1,0}(\lambda_1+\lambda_2+\mu_1) &= P_{n_1-1,0}\lambda_1 + P_{n_1+1,0\mu_1} + P_{n_1+1,1}\mu_2, \quad si \;\; n_1>0 \;\; et \;\; n_2=0\\
P_{n_1,n_2}(\lambda_1+\lambda_2+\mu_1) &= P_{n_1,n_2-1}\lambda_2 + P_{n_1-1,n_2}\lambda_1 + P_{n_1+1,n_2}\mu_1 + P_{n_1,n_2+1}\mu_2, si \quad n_1>0, \quad n_2>0
\end{aligned}
$$

Les calcules étant longs et fastidieux, mais nous nous limiterons ici au la présentation de la démarche d'obtention des résultats. Il faut tout d'abord calculer la fonction génératrice :

$$H(y,z) = \sum_{n_1,n_2} P_{n_1,n_2} y^{n_1} z^{n_2}$$

A partir des équations de Chapman Kolmogorov on trouve :

$$H(y,z) = (\frac{1-\rho_1-\rho_2}{1-\eta-z\rho_2})(\frac{1-\eta}{1-\eta y})$$

Où
$\eta = \dfrac{1}{2\mu_1}[\mu_1+\lambda_1+\lambda_2(1-z)-\sqrt{[\mu_1+_1+\lambda_2(1-z)]^2-4\mu_1\lambda_1}]$ De la fonction génératrice $H(y,z)$ nous déduisons :

Le nombre moyen de clients prioritaire dans le système est

$$E[N_1] = (\frac{\partial H(y,z)}{\partial y})_{y=z=1} = \frac{\rho_1}{1-\rho_1} \tag{3.9}$$

Le nombre de clients non prioritaires dans le système est

$$
\begin{aligned}
E[N_2] = (\frac{\partial H(y,z)}{\partial z})_{y=z=1} &= \frac{\rho_2+E[N_1]\rho_2}{1-\rho_1-\rho_2}\\
&= \frac{\rho_2}{(1-\rho_1)(1-\rho_1-\rho_2)}
\end{aligned} \tag{3.10}
$$

D'après les formules (2.5) et (2.6), et au moyen de la formule de Little, en peut tirer le temps moyen de séjour dans le système des clients prioritaires et non prioritaires respectivement, donné par les deux formules suivantes :

$$E[S_1] = \frac{E[N_1]}{\lambda_1} = \frac{1}{\mu_1(1-\rho_1)} \tag{3.11}$$

$$E[S_2] = \frac{E[N_2]}{\lambda_2} = \frac{1}{\mu_2(1-\rho_1)(1-\rho_1-\rho_2)} \tag{3.12}$$

Pour l'état d'équilibre, le calcule des probabilités d'état s'effectuerait de façon suivante :

$$P_{n_1,n_2} = \frac{1}{n_1!n_2!}[\frac{\partial^{n_1+n_2}H(y,z)}{\partial y^{n_1}\partial z^{n_2}}]_{y=z=0}$$

En particulier, $P_{0,0} = 1-\rho_1-\rho_2$.

La condition de stabilité est alors donnée par $P_{0,0} > 0$, c'est à dire $\rho_1 + \rho_2 < 1$.

3.3 Les systèmes de file d'attente avec rappels

Les systèmes de files d'attente avec rappels se caractérisent par le fait que :" un client qui arrive et trouve tous les serveurs occupés, quitte le système définitivement, ou rappelle ultérieurement à des instant aléatoires, dans cette dernière on dit que le client entre dans l'*orbite* ".

le système de file d'attente avec rappel est défini complètement par la connaissance de ces éléments principaux : le processus des inter-arrivés, le mécanisme de service(la disponibilité et nombre de serveurs, la discipline de passage au service),en ajoutant, un élément décrivant la loi des répétition d'appels.

Le modèle de file d'attente avec rappels occupe une situation intermédiaire entre le modèle d'Erlang et le modèle classique d'attente FIFO, qui en constituent les modèles limites dans les cas de faible et forte intensité de rappels.

Les systèmes de file d'attente avec rappels ont trouvé leur première application lors de la modélisation du service d'abonnés dans un central téléphonique.

Plusieurs situations d'attente ont la caractéristique que les clients doivent rappeler, pour une certaine raison, pour être servis. Quand le service d'un client est insatisfait, il doit rappeler jusqu'à a l'accomplissement de son service. Ces modèles d'attente apparaissent dans la modélisation stochastique de plusieurs situations réelles. Par exemple, dans la transmission de données, un paquet transmis de la source 'a la destination peut être retournée et le processus doit se répéter jusqu'à ce que le paquet soit finalement transmis [25].

3.3.1 Description des systèmes de file d'attente avec rappels

considérons un système de file d'attente, de flot des arrivées primaires poissonien de paramètre λ, et ayant $s(s \geq 1)$ serveurs identique et indépendants, de distribution de service générale, de fonction de répartition $B(x)$. Il y a dans ce système $m - s$ position d'attente. A l'arrivée d'un client, s'il y a un ou plusieurs serveurs libres, celui-ci sera immédiatement pris en charge. sinon, s'il y a une position libre, le client rejoint la file d'attente. Lorsque tous les serveurs et toutes les positions d'attentes sont occupées, le client est obligatoirement quitte le système, soit définitivement avec une probabilité $1 - H_0$, soit temporairement avec une probabilité H_0 et rappelle ultérieurement, après un temps aléatoirement.

La capacité O de l'orbite peut être finie ou infinie. Dans le cas où O est finie, si l'orbite est pleine alors le client quitte le système définitivement.

Yang et Templeton introduisent, pour désigner les systèmes avec rappels, la notation suivante : $A/B/s/m/O/H$, où A et B décrivent respectivement la distribution du temps des inter-arrivés et la distribution de temps de service, s représente le nombre de serveurs, m est le nombre de position d'attente, et O est la capacité de l'orbite, ce système peut ne pas apparaitre dans la lorsque la capacité de l'orbite est infinie. La séquence $H = \{H_i, i \geq 0\}$ est la fonction de persévérance, où H_i est la probabilité qu'un client fasse une $(i+1)^{eme}$ tentative de rappel, après une i^{eme} tentative échouée. Si tous les clients sont persévérants $H_i = 1$, pour tout j, le symbole H pourra être également supprimé.

Aujourd'hui, il y a un centaines d'exemples qui peuvent être modélisé par un modèle de file d'attente avec rappels, en peut citer quelques exemples pour bien comprendre ce modèle.

1. Problème de réservation

C'est l'exemple le plus simple d'un client qui souhaite réserver dans un restaurant par téléphone, mais, il y a une unique ligne qui consacré à rependre aux requêtes des réservations. Ainsi, si le client appelle est trouve la ligne occupée, il renouvellera sa tentative après une certaine période de temps aléatoire avec la probabilité H_k qui est en pratique inférieure à 1, car le client ne peut pas 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'arrivée 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, ...

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 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.

3. Réseaux locaux CSMA

Dans les réseaux locaux se partageant un bus unique, l'un des protocoles de communication le plus généralement utilisé est appelé protocole non-persistant CSMA (Carrier Sense Multiple Access), c'est une méthode d'accès à un réseau local.

Un réseau local simple est composé de stations ou de terminaux inter-connectés 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. Une telle architecture de réseau d'ordinateurs local est appelée architecture en bus.

Des messages de longueurs variables arrivent aux stations du monde extérieur. En recevant le message, la station le découpe en un nombre fini de paquets de longueur fixe, et consulte immédiatement le bus pour voir s'il est occupé ou bien libre. Si le bus 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. Par contre, si le bus n'est pas libre, tous les paquets sont stockés dans le tampon (positions d'attente) et la station peut reconnecter le bus après une certaine période aléatoire.

Ce problème peut être modélisé comme un système d'attente avec rappels à un seul

serveur, qui est le bus, et les tampons des stations représentent l'orbite.

Le modèle d'attente avec rappels d'écrit ci-dessus est un modèle général. Plusieurs systèmes de files d'attente avec rappels peuvent être considérés comme des cas particuliers tels que : les systèmes sans buffer, les systèmes à un seul serveur, ...

par Le schéma général d'un système avec rappels est donné par la figue suivante.

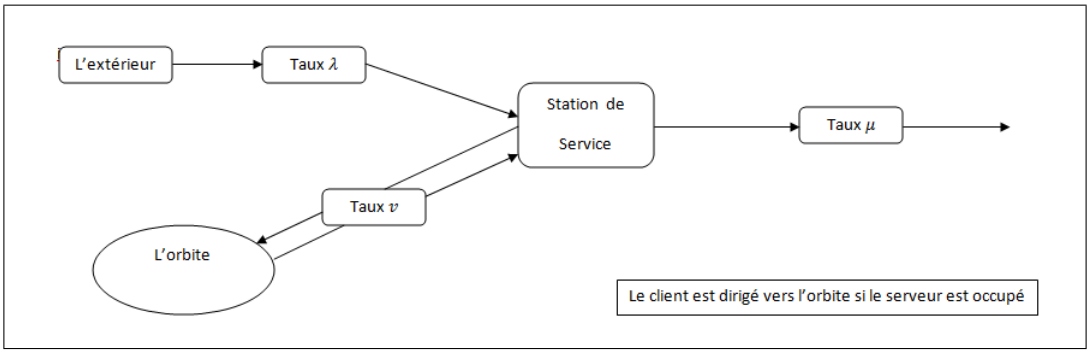

FIG. 3.3 – **Système générale d'une file d'attente avec rappel**

3.3.2 Quelques variantes des modèles de files d'attente avec rappels

Modèle markoviens

Les modèles markoviens sont utilisés pour décrire les systèmes dans lesquels les temps des inter-arrivés primaires, les durées de service et les temps inter rappels sont des variables aléatoires indépendantes et exponentiellement distribués.

Pour décrire un système de file d'attente avec rappels, dont le nombre de serveurs est m, le flux des inter-arrivés primaires est poissonien de paramètre λ, la durée de service des clients est exponentielle de paramètre μ et la durée entre deux rappels consécutifs d'une même source secondaire est aussi exponentielle de paramètre θ.

Soit le processus markovien $X(t) = \{C(t), N(t)\}_{t\geq 0}$, dont l'espace d'états est $S = \{0, 1, ..., m\} \times N$, où $C(t)$ est le nombre de client en cours de service à la date t, $N(t)$ est le nombre de client en régime de rappels à la date t.

Modèles semi-markovien

Lorsqu'on peut pas décrire le système d'attente avec rappels avec un processus markovien, en parle de modèle appelle semi-markovien. Donc, ce dernier peut être définie comme étant un modèle où l'un de ces processus n'est pas markovien(flux des inter-arrivés des clients primaires, la distribution de temps de service ou la distribution de temps de rappels).

Parmi les modèles semi-markovien avec rappels en peut citer celle qui est plus utilisé, le modèle $M/G/1$ avec rappels.

La description du modèle $M/G/1$ est faite par :

introduisant un processus des inter-arrivés des clients primaires qui est poissonien de paramètre λ, la durée de service τ est de loi générale de distribution $B(x)$. La durée entre deux rappels est successifs d'une même source secondaire est exponentielle de paramètre θ.
On décrit le système de la manière suivante :
On suppose que le $(i-1)^{ieme}$ appel termine son service à l'instant η_{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 $(i)^{ieme}$ rappels suivant n'entre en service qu'après un intervalle de temps R_i durant lequel le serveur est libre.
En peut citer d'autres systèmes de fille d'attente avec rappels qui sont :

- modèle avec rappels et serveur non fiable,

- modèle avec rappels et arrivées par groupes,

- modèle avec multi-classe de client,

- modèle avec deux types des clients impatients,

- les modèles avec priorité.

Quelques caractéristiques du système $M/G/1$ avec rappels

Dans cette partie en s'intéresse pas au détail de calcule, c'est-à-dire, comment on a trouver les mesures de performance, le nombre moyen de clients dans le système, le nombre moyen de clients dans l'orbite, mais, juste de donner ces caractéristiques afin

de l'utiliser dans le dernier chapitre, exactement pour la validation de simulateur approprie.

Nombre moyen de clients dans le système

$$\bar{n} = \rho + \frac{\lambda^2 E(\tau^2)}{2(1-\rho)} + \frac{\lambda\rho}{\theta(1-\rho)} \tag{3.13}$$

Nombre moyen de clients dans l'orbite

D'après la formule de Little, on a

$$\bar{n_0} = \bar{n} - \rho = \frac{\lambda^2 E(\tau^2)}{2(1-\rho)} + \frac{\lambda\rho}{\theta(1-\rho)} \tag{3.14}$$

Temps moyen de séjour

D'après la formule de Little $\bar{n} = \bar{w}\lambda$. On aura

$$\bar{w} = \frac{\rho}{\lambda} + \frac{\lambda E(\tau^2)}{2(1-\rho)} + \frac{\rho}{\theta(1-\rho)} \tag{3.15}$$

Quelques caractéristiques du système $M/G/1$ avec rappels et priorité

Dans cette section aussi, on doit juste donné quelques caractéristiques de performances pour qu'on puisse utiliser dans les sections qui se suit.

Dans ce système, on considère notre file de capacité infinie, et constituer d'un seul serveur. Le processus d'arrivé est poissonien de paramètre λ_1 et λ_2 respectivement, Les taux μ_1 et μ_2 sont des taux de services des clients de type 1 et de type 2 , et ν est le taux de rappel dans l'orbite, alors :

Le nombre moyen de client de type 1 dans le système

$$N_1 = \frac{\lambda_1(\lambda_1\beta_{1,2} + \lambda_2\beta_{2,2})}{2(1-\rho_1)} \tag{3.16}$$

Le nombre moyen de client de type 2 dans le système

$$N_2 = \frac{\lambda_1(\lambda_1\beta_{1,2} + \lambda_2\beta_{2,2})}{2(1-\rho_1)(1-\rho_1-\rho_2)} + \frac{\lambda_2(\rho_1+\rho_2)}{\nu(1-\rho_1-\rho_2)} \tag{3.17}$$

Avec $\beta_{1,2}$ et $\beta_{2,2}$ représente le moment d'ordre deux de la durée de service des clients de types 1 et 2 respectivement, c'est-à-dire :

$$\beta_{1,2} = E(\tau_1^2)$$
$$\beta_{2,2} = E(\tau_2^2)$$

avec τ_1 (τ_2) est les variables qui décrivent la distribution de service des clients de type 1 (type 2) respectivement.

3.4 Système de file d'attente avec vacance

La notion de vacances est introduite en général pour exploiter le temps inoccupé du serveur pour un autre travail secondaire dans le but d'améliorer la performance du système.

L'analyse de la modélisation par les systèmes d'attente avec vacances a été faite par un nombre considérable de travaux dans le passé et a été utilisée de manière réussie dans différents problèmes pratiques comme les systèmes de production, systèmes de communication et systèmes informatiques (voir la monographie de Doshi (1986) [21]). Une excellente étude compréhensive sur les modèles d'attente avec vacances peut être trouvée dans Teghem (1986) [23], Takagi (1991) [22] et Tian et Zhang (2006) [24].

3.4.1 Quelques cas modélisés par des systèmes de files d'attente

Exemple :
Dans le modèle opérationnel du serveur WWW, les requêtes HTTP arrivent au serveur WWW suivant un flux de Poisson et peuvent être interrompues par un utilisateur avant d'arriver au serveur WWW. Lorsque les requêtes arrivent dans le serveur WWW, une requêter est sélectionnée pour être servie et les autres entreront dans le tampon situé à l'intérieur du serveur WWW. Dans le tampon, chaque requêter attend un certain temps puis demande le service de nouveau. Un programme " daemon " est mis en application dans le serveur WWW pour diriger les requêtes de service à partir du tampon. À chaque fois qu'elle essaye mais échoue, elle attend un autre moment avant de réessayer de nouveau. Si

la page web cible est située dans le même serveur WWW, la requêter peut retourner au serveur. Pour garder le serveur WWW en bon fonctionnement, des activités de maintenance telles que scanner les virus peuvent être réalisées lorsque le serveur WWW est inactif. Ce type de maintenance peut être programmé pour fonctionner sur une base régulière. Cependant, ces activités de maintenance ne se répètent pas continuellement. Lorsque ces activités sont achevées, le serveur WWW entrera de nouveau en état d'inactivité et attendra l'arrivée de nouvelles requêtes.
Dans ce scénario, le tampon dans le serveur WWW, le serveur WWW, la politique de retransmission et les activités de maintenance en période d'inactivité correspondent respectivement à l'orbite, le serveur, la discipline de rappels et la politique de vacances, dans la terminologie de files d'attente.
Exemple :
Dans le modèle de transfert d'un système e-mail, le système e-mail utilise le protocole SMTP (Simple Mail Transfer Protocol) pour délivrer les messages entre les serveurs e-mail.
Quand un programme de transfert de e-mail contacte un serveur sur une machine éloignée, il forme une connections TCP (Transfer Connection Program) à travers laquelle il communique. Une fois la connections est en place, les deux programmes suivent SMTP qui permet à l'expéditeur de s'identifier, spécifier un destinataire, et transférer un message e-mail. Ce dernier peut essayer de façon répétée d'envoyer le message de contact au serveur cible jusqu'à ce qu'il devienne opérationnel. Typiquement, les messages de contact arrivent au serveur e-mail suivant un flot de Poisson. Ces messages de contact peuvent être interrompus par un expéditeur avant 'arriver au serveur e-mail. Quand les messages arrivent au serveur mail, un message est sélectionné pour service et les autres joindront le buffer. Dans le buffer, chaque message attend un certain temps pour demander encore une fois le service. Il y a un programme mis en application et implémenté au serveur mail pour diriger les requêtes de service du buffer. À chaque fois qu'une requêter essaye mais échoue, elle attendra un autre moment avant de recommencer de nouveau. Le serveur cible est le même que le serveur mail de l'expéditeur et le message envoyé peut revenir au serveur pour demander le service. Pour garder le serveur mail en bon fonctionnement, scanner les virus est une importante activité de maintenance pour le serveur e-mail. Elle peut être réalisée lorsque le serveur e-mail est inactif. Cependant, ces activités de maintenance ne se répètent pas continuellement.
Quand ces activités sont finies, le serveur e-mail entrera encore une fois en état d'inactivité (oisiveté) et attendra l'arrivée des messages de contact. Parce qu'il n'y a pas de mécanisme pour enregistrer le nombre de messages de contact parvenant couramment des différents expéditeurs, il est approprié de concevoir un programme pour collectionner l'information des messages de contact pour des raisons d'efficacité.
Dans ce scénario, le buffer dans le serveur mail de l'expéditeur, le serveur e-mail destina-

taire, la politique de retransmission, et les activités de maintenance correspondent respectivement à l'orbite, le serveur, la discipline de rappels, et la politique de vacances dans la terminologie de files d'attente.

Classification des modèles d'attente avec vacances

Les files d'attente avec vacances peuvent être classifiées de différentes façons. Les disciplines de service les plus connues sont :
- La discipline de service exhaustif : Dans un système avec vacances et service exhaustif, chaque fois que le serveur revient d'une vacance, il servira tous les clients en attente dans le système avant de commencer une autre vacance.
- La discipline de service avec barrière : Dans le cas du service avec barrière, quand le serveur revient d'une vacance, il sert seulement les clients qui étaient en attente dans la file 'a son arrivée. Autrement dit, d'es l'arrivée du serveur, il met une barrière fictive derrière les clients en attente dans la file et ne prend une autre vacance qu'une fois que tous les clients qui étaient présents à son arrivée soient servis.
- La discipline de service limité : Dans un système avec service limité, on se fixe un nombre k. A son retour de la vacance, le serveur servira au plus k clients et commencera ensuite une autre vacance. Ainsi, le serveur sert jusqu'à ce que la file d'attente soit vide ou bien jusqu'à ce que k clients soient servis, ensuite il prend une autre vacance. Lors de l'analyse de ces système, on a remarquer des difficultés lors de l'application et aussi même pour l'application des méthodes appropriés. Pour bien palier à ces difficultés plusieurs méthodes approximatives d'analyse ont été développés. Parmi les principales approches, en trouve la méthode de simulation à temps discret qui nous permet de conduire à des estimations quantitatives des caractéristiques de système étudie.

Conclusion

Dans ce chapitre, nous avons présenter quelques résultats analytique sur les systèmes de file d'attente, à savoir, les systèmes prioritaires, les systèmes avec rappels et priorité et celle avec vacance, dont le but est de l'utiliser dans le chapitre quatre pour pouvoir valider les modèles appropriés.

Chapitre 4

Simulation

Introduction

D'après le verbe "*simuler*" qui veut dire faire semblant de, en peut dire que la technique de simulation est l'un des meilleurs moyens qui nous permettent de connaitre la conception et de simuler le fonctionnement des systèmes dont l'étude analytique directe est assez difficile, ou parfois impossible.

Le comportement transitoire des systèmes simulés, peut être évalué. En effet, la machine génère l'évolution du modèle au cours de temps et estime les critères de performances par des moyennes temporelles alors que les modèles analytiques sont généralement utilisés pour étudier le comportement stationnaire d'un système.

Dans ce chapitre, nous allons présenter les caractéristiques principales de cet outil. Nous commençons par introduire les concepts de base de la simulation et focalisons la présentation sur la simulation à événements discrets qui est la technique que nous avons utilisé pour la réalisation des programmes de simulation.

4.1 Présentation de modèle de simulation

Dans une étude basée sur l'approche de simulation, il nous faut des modèles logiques afin d'établir et employer. Le modèle générale de simulation est illustré dans la figure (3.1) suivante :

Les distributions des variables d'entrées sont supposées connues alors que celle des variables de sorties ne le sont pas. Pendant la phase de simulation, les variables aléatoires d'entrées sont remplacées par des échantillons. Ainsi des expériences sont effectuées sur le modèle

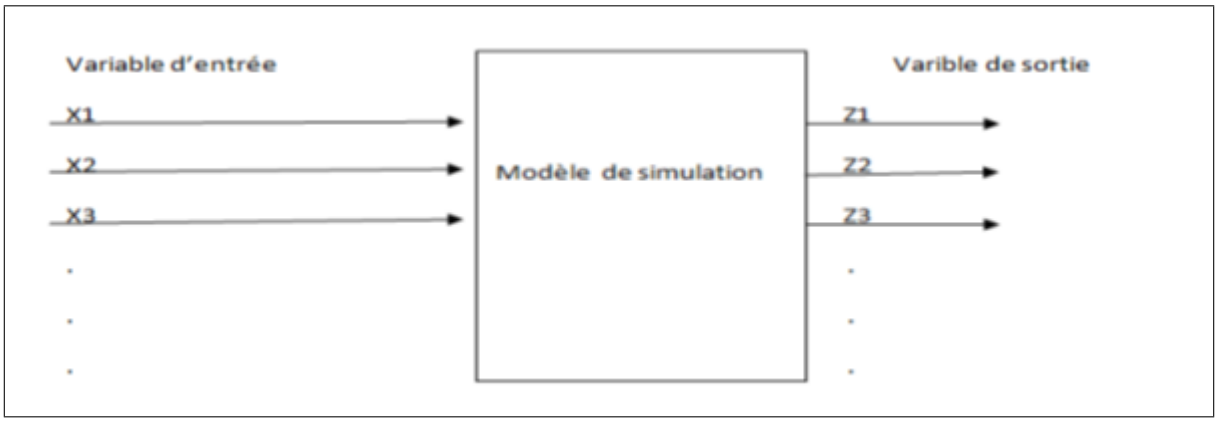

FIG. 4.1 – **Présentation de modèle de simulation**

établi et des paramètres inconnus des variables aléatoires de sortie sont estimés.

4.2 Concepts de base

La simulation est l'imitation de l'opération d'un système réel à travers un modèle. Elle requiert une génération artificielle du comportement dont l'observation permet de déduire les caractéristiques de fonctionnement du système réel.

La simulation permet de réaliser des expérimentations imaginables dans le système réel. Par exemple :

- Elle permet d'estimer le comportement du système sous des conditions d'opération extrêmes sans mettre en cause l'intégrité du système réel.

- Si l'on vaut modifier le système, on peut évaluer et comparer plusieurs scénarios afin d'identifier les comportements les plus sensibles à la modification.

4.2.1 Simulation à événements discrets

La simulation à événements discrets modélise le système à simuler sous forme d'une séquences d'événements qui dirigent révolution du système.

Caractéristiques des modèles de simulation discret

Un système :C'est l'interaction d'une collection d'objets clans un environnement fermé. Il est affecté par les éventuels changements de son environnement.

Un modèle : Un modèle est une représentation simplifiée de la réalité.

Un processus : Un processus est une succession d'un nombre fini d'états d'une ressource.

Une activité : C'est un intervalle de temps pendant lequel l'état de la ressource ne change pas.

Les entités : Ce sont des classes identifiables d'objets qui peuvent varier en nombres. Elles peuvent avoir un nombre variable de caractéristiques identificatrices appelées attributs. Ces derniers sont relies entre eux et A leur environnement.

Un événement : On appelle événement un instant précis de changement d'état de ressource.

Le simulateur :C'est un programme contenant l'algorithme utilisé pour simuler le système étudié. II est constitué d'un ensemble d'entités qui décrivent une composante du système réel.

4.2.2 Principale approches de la simulation à événements discrets

On distingue trois types d'approches de simulation à événements discrets : approche par événements, approche par activités et approche par processus.

Approche par événements

C'est l'approche de base, elle consiste en :
• Identification des différents types d'événements possibles au cours de la durée de vie du système
• La description de la logique de fonctionnement entre événements, déterminer les changements d'état correspondant a chaque événement et les événements qui en résultent
• L'utilisation de calendriers d'événements ou échéancier, liste des événements et leur date d'occurrence.

Approche par activités

Dans cette approche, les principales phases sont :
• L'identification des différents types d'activités possibles,
• La description des caractéristiques des activités : conditions, conséquences,..
• La représentation par diagrammes de cycles : succession d'états actifs (activités opérationnelles) et d'états passifs (attentes).

Approche par activités et évènement (méthode de 3 phases)

Cette approche est représenter par Tocher et décrite en trois phases :
• Phase A : déterminer l'instant d'occurrence de l'évènement suivant et avancer le temps simulé à ce point.
• Phase B : exécuter toutes les activités B qui doivent se produire à l'instant d'occurrence.
• Phase C : exécuter toutes les activités C dont les conditions sont satisfaites à l'instant d'occurrence.

Dans cette méthode, le temps simulé est contrôlé par la technique du prochain évènement, appelé méthode à pas variable. Le temps avance lorsqu'un évènement se produit et provoque un changement d'état du système.

Approche par processus

Cette approche se caractérise par :
• La présence des séquences d'événements ou des activités similaires pour un type d'objets, défini sous forme de processus,
• La description du fonctionnement du système complet par macro-représentation ,
• La gestion des conflits et la synchronisation entre processus par des règles d' interruption et de reprise.

Simulation distribuée

Ce type de simulation consiste à faire coopérer plusieurs processus de simulation pour atteindre un objectif commun.

4.3 Étapes de la simulation

La simulation d'un système passe par les étapes suivantes :

* **Formulation du problème :** Cette étape consiste à identifier et analyser le problème en identifiant ces composantes, leurs relations, les frontières entre le système et son environnement,

* **Élaboration du modèle :** il s'agit ici d'extraire un modèle aussi fidèle que possible du modèle réel Car si le modèle du système est inadéquat, les conclusions que l'on peut en tires seront inadéquates elles aussi,

* **Identification des paramètres et collecte des données :** La phase de l'identification de types des données à introduire dans le modèle est une phase très délicate et essentielle. La collecte des données est indispensable pour l'estimation des paramètres du modèle. Ceci requiert une connaissance des méthodes statistiques et des tests d'hypothèses. Il s'agit d'une étape fondamentale pour le développement du modèle,

* **Validation du modèle :** II s'agit du contrôle de la correspondance entre le modèle et la réalité ; les mesures des variables de sortie du modèle doivent correspondre aux mesures faites sur le système réel,

* **Exécution de la simulation :** Le concepteur doit pouvoir mettre à l'épreuve le modèle en agissant sur les paramètres qui le configurent. II s'agit d'effectuer plusieurs exécutions et de recueillir les résultats obtenus,

* **Analyse et interprétation des résultats :**Une fois les résultats obtenus, le concepteur passe à l'analyse et a l'interprétation de ces résultats pour donner des recommandations et des propositions,

* **Conclusion et exploitation des résultats :** Cette étape consiste à évaluer les perspectives d'exploitation du modèle pour d'autres préoccupations.

4.4 Avantages et inconvénients de la simulation

• **Avantage :**

La grande popularité de la simulation, s'explique par les nombreux avantages offerts par cette technique. On peut citer, entre autres, les suivants :

1. La simulation permet de mettre en œuvre une expérience hautement contrôlée, d'ordinaire mieux contrôlée qu'une expérimentation dans le système réel.

2. Elle permet de tester aisément des hypothèses sur le fonctionnement du système et de mieux le comprendre.

3. Elle permet d'étudier les conditions d'opération extrêmes du système et d'en évaluer les conséquences sans mettre en danger ni le système ni son environnement.

4. Mieux comprendre les anciens problèmes et possibilité de détecter les autres problèmes insoupçonnés.

5. Excellente méthode de comparaison, et de vérification des solutions obtenues par d'autres méthodes de mettre à épreuve des modèles théoriques.

7. Permet d'adresser des systèmes très compliqués.

8. Répétition d'expérience.

• **Inconvénients :**

Ce n'est qu'une expérimentation sur le modèle, elle :

— ne donne pas la solution optimale,

— n'est pas précise par nature (a cause des simplifications et extrapolations),

— demande une mise au point toujours longue, et ses solutions générales s'obtiennent seulement par une induction à partir de cas numériques.

Le problème majeur d'une expérience de simulation est celui de la correspondance entre le modèle simulable et la réalité qu'il représente. Lorsque le phénomène que l'on veut reproduire artificiellement est bien connu, c'est-a-dire lorsqu'il fait l'objet d'une théorie complète, cohérente et valide, il est possible d'élaborer un modèle qui sort une représentation très fidèle de ce phénomène. Il est en général, très difficile de valider qualitativement les simulations. En effet, même si on peut borner l'erreur statistique de la simulation, il n'est pas évident de distinguer si les résultats obtenus correspondent aux spécifications du projet ou s'ils sont dérives d'un modèle inadéquat du système.

4.5 Types de simulation

4.5.1 Simulation finie

La simulation finie démarre d'un certain état jusqu'à ce que une condition d'arrêt soit vérifiée, Le système n'est pas sensé d'atteindre un régime stationnaire et les paramètres estimés seront transitifs dans le sens où ils dépendent des conditions initiales [10].

Exemple :

On peut citer a titre d'exemple la méthode des réplications indépendantes qui consiste à exécuter k réplications indépendantes de la simulation, si chaque histoire est de taille $m(n = mk)$, alors on obtient un échantillon $x_{i1}, x_{i2}, ..., x_{in}$, pour l'histoire i Les moyennes $Y_i = \frac{1}{m}\sum_{j=1}^{m} x_{ij}$ seront indépendantes et suivent approximativement une loi normale.
Donc, $\bar{Y} = \frac{1}{k}\sum_{i=1}^{k} Y_i$ est un estimateur sans biais de la moyenne et $S^2 = \frac{1}{k-1}\sum_{i=1}^{k}(Y_i - \bar{Y})^2$, est la variance empirique.
Quand k est suffisamment grand, un intervalle de confiance à $(1|\alpha)$ pour la moyenne est donné par :

$$\bar{Y} \pm t_{(k-1,1-\frac{\alpha}{2})}\sqrt{\frac{S^2Y}{k}}$$

Où $t_{(k-1,1-\frac{\alpha}{2})}$ est le quantile d'ordre $(1-\alpha)$ de la loi de Student à $(k-1)$ degré de liberté.

4.5.2 Simulation stationnaire

Le but dans ce cas est d'étudier le système à large terme, on désire estimer au régime stationnaire les paramètres d'un système.

Remarque 4.1. Dans ce type de la simulation on s'intéresse à éliminer les conditions initiales.

Exemple :

II existe plusieurs méthodes pour ce type de simulation, on se limite à donner le principe de trois méthodes.

Méthode des blocs

Elle a pour principe l'exécution d'une seule simulation du taille $n = km$ On divise les observations $x_{i1}, x_{i2}, ..., x_{in}$ en k blocs disjoints et contiguës,donc, le i^{ieme} bloc est compose des observations suivantes : $x_{i-1\ m+1}, ..., x_{im}$, $i = 1,\bar{}k$, la moyenne empirique sera :

$$y_i = \frac{1}{m}\sum_{j=1}^{m} X_{i-1\ m+j}$$

La variance empirique est :

$$S^2 = \frac{1}{k-1}\sum_{i=1}^{k}(Y_i - \bar{Y}^2),$$
$$\bar{Y} = \frac{1}{k}\sum_{i=1}^{k} Y_i$$

avec $\bar{Y} = \frac{1}{k}\sum_{i=1}^{k} Y_i$ est un estimateur de la moyenne, et l'intervalle de confiance à $1-\alpha$, pour la moyen donné par la formule suivante :
$\bar{Y} \pm t_{k-1,1-\frac{\alpha}{2}}\sqrt{\frac{S^2Y}{k}}$.

Méthode es réplications(suppression)

Cet approche propose de faire k réplications indépendantes de langueur $(L+m)$ observations on élimine en suite les L premières observations de chaque réplication.

Méthode de la régénération

Cette méthode consiste à chercher une séquence de points aléatoires dans lesquels le processus se régénère dans le sens où les variables aléatoires représentant le système dans ces points sont indépendantes et identiquement distribuées.

4.6 Génération des variables aléatoires

Génération des nombres aléatoires

Dans la simulation d'un phénomène stochastique, la génération des nombres aléatoires est primordiale. Elle sera incluse dans le modèle et fournira, au fur et à mesure, des échantillons artificiels d'entrée au simulateur. Pour que ce dernier reproduise fidèlement le phénomène réel, il est absolument nécessaire que ces échantillons d'entrée suivant la même loi de probabilité qu'un échantillon construit d'observations faites sur le phénomène réel.

L'objectif de la génération des nombres aléatoires par ordinateur et de produire une suite de nombres statistiquement indépendantes et réparties de manière uniforme sur l'intervalle $[0, 1]$.

4.7 Technique de génération des variables aléatoires

Nous allons présenter ici la technique de transformation inverse, utilisée pour générer des variables aléatoires suivant différentes lois de probabilités.

Propriété

Si U une variable aléatoire uniforme sur $[0,1]$, Si X une variable aléatoire de fonction de répartition F_x.
alors, $X = F_X^{-1}(U)$
Si U suit une loi uniforme $U_{[0,1]}$ alors $U^{'} = 1 - U$ suit aussi une loi uniforme $U^{'}_{[0,1]}$.

La loi de Bernoulli

On dit que la variable aléatoire X est suit la loi de Bernoulli, de paramètre P et en écrit, $X \rightsquigarrow B_P$, si sa densité de probabilité est de la forme suivante :

$$\begin{cases} P(X=1) = P \\ P(X=0) = 1-P \end{cases}$$

Et sa fonction de répartition est donnée par

$$F(x) = \begin{cases} 0 & si \;\; x < 0 \\ 1-P & si \;\; 0 \leq x < 1 \\ 1 & si \;\; x \geq 1 \end{cases}$$

Donc

$$\begin{cases} x_i = 0 & si \;\; u_i < 1-P \\ x_i = 1 & si \;\; u_i \geq 1-P \end{cases}$$

Donc

$$\begin{cases} x_i = 0 & si \;\; u_i > P \\ x_i = 1 & si \;\; u_i \leq P \end{cases}$$

La loi de Binomiale

On utilise le fait que la somme de n variables aléatoires de Bernoulli de paramètre P, suit une loi de Binomiale de (n, P).

Il suffit donc de simuler n variables aléatoires de Bernoulli et de faire la somme.
Pour simuler $X \rightsquigarrow B_{(n,P)}$, il suffit de générer des nombres aléatoires u_i de la loi uniforme sur $[0,1]$ et en déduire les réalisations x_i tel que :

$$x_i = \begin{cases} x_{i-1} + 1 & si \;\; u_i > P \\ x & si \;_{i-1} \; u_i < P \end{cases}$$

La loi de poisson

La loi de nombre d'occurrence d'évènement sur l'intervalle $[0,t]$ est de poisson de paramètre λt si :

$$P(X=n)=\frac{(\lambda t)^n}{n!}e^{-\lambda t}$$

Si $\lambda=1$, les données des intervalles suivent l'exponentielle de paramètre t alors :

$$P(X=n)=\frac{t^n}{n!}e^{-\lambda t}$$

Soit A l'évènement " avoir n occurrences pendant $[0,\lambda]$ ", alors :

$$P(A)=\frac{\lambda^n}{n!}e^{-\lambda}$$

Considérons les relations suivantes :

$0=t_0<t_1<t_{2<....<t_n}$ de variables aléatoires $exp(1)$ alors :

$$\exists n\in N^*\ \ tel\ \ que:\ \ \sum_{i=0}^{n}t_i<\lambda<\sum_{i=0}^{n+1}t_i$$

n peut être considéré comme le nombre d'occurrences d'évènement pendant $[0,\lambda]$, lorsque les intervalles de temps entre deux occurrences sont $exp(1)$ et est bien une réalisation d'une variable aléatoire qui suit la loi de $P(\lambda)$.
Comme $t_i=-\log u_i$

$$-\sum_{i=0}^{n}\log u_i<\lambda<-\sum_{i=0}^{n+1}\log u_i$$
$$-\log(\prod_{i=0}^{n}u_i)\leq\lambda\leq-\log(\prod_{i=0}^{n+1}u_i)$$
$$\log(\prod_{i=0}^{n+1}u_i)<-\lambda\leq\log(\prod_{i=0}^{n+1}u_i)$$
$$\prod_{i=0}^{n+1}u_i<e^{-\lambda}\leq\prod_{i=0}^{n}u_i$$

Pour $t_0=0\ \ \Rightarrow\ \ -\log u_0=0\ \ \Rightarrow\ \ u_0=1$

Pour simuler une loi de poisson de paramètre λ, il suffit de générer des nombres aléatoires u_i de loi uniforme sur [0,1], $i\geq 1$ et puis en multiple entre elle,(les observations obtenues) et on arrête le processus dès que le produit obtenu devient inférieur à $e^{-\lambda}$, et on déduire une réalisation n, et on recommence pour une autre réalisation.

La loi uniforme

La fonction de densité d'une variable aléatoire suivant une loi uniforme sur $[a, b]$ est donnée par

$$f(x) = \begin{cases} \frac{1}{b-a}, & si \quad a \leq x \leq b \\ 0 & \text{sinon} \end{cases}$$

La fonction de répartition est donnée par

$$F(x) = \begin{cases} 0, & si \quad x \leq 0 \\ \dfrac{x-a}{b-a}, & a \leq x \leq b \\ 1 & \text{sinon} \end{cases}$$

pour générer des variables aléatoires qui suivent la loi uniforme $U_{[0,1]}$, en pose $F_X = \dfrac{X-a}{b-a} = U$, en trouve $X = a + (b-a)U$

La loi exponentielle
La fonction de densité de probabilité est donné par

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & si \quad x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

Et sa fonction de répartition est donnée par

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

Le paramètre λ peut être interprété comme le nombre moyen d'occurrences par unité de temps.

L'objectif ici est de développer une procédure pour générer les variables $X_1, X_2, ...$ suivant une distribution exponentielle. La technique de transformation inverse est utilisée pour n'importe quelle distribution, mais elle est préférable pour des distributions dont $F(x)$ est simple. Voici les étapes de cette technique illustre pour la loi exponentielle :
1. Calculer $F(x)$ Pour la distribution exponentielle, $F_x = 1 - e^{-\lambda x}$, $x \geq 0$,
2. Poser $F(x) = U$, où U est variable aléatoire uniformément distribué sur l'intervalle $[0, 1]$, Pour la loi exponentielle on a $1 - e^{-\lambda x} = U$,
3. Résoudre l'équation $F(x) = U$

$$1 - e^{-\lambda x} = U$$
$$e^{-\lambda x} = 1 - U$$

$$X = -\frac{1}{\lambda}\ln(1-U) \tag{4.1}$$

L'équation (3.1) est appelée générateur aléatoire pour la distribution exponentielle.
4. Générer des nombres aléatoires $U_1, U_2, U_3, ...$ et calculer $X_i = F^{-1}(U_i)$

$$X_i = -\frac{1}{\lambda}\ln(1-U_i) \tag{4.2}$$

L'équation (3.2) peut-être remplacée par $X_i = -\frac{1}{\lambda}\ln(U_i)$

4.7.1 Validation du simulateur

Soit un échantillon $(X_i)_{i=1,2,...,n}$ d'une v. a. X en utilisant cet échantillon, on s'intéresse à vérifier :
∘ l'indépendance des $(X_i)_{i=1,2,...,n}$,
∘ L'adaptation de la loi de l'échantillon par la loi théorique (test de χ^2 et kolmogorov-Smirnov).

L'une des méthodes les plus utilisée dans cette étape consiste à comparer les résultats issus de la simulation à des données issues du système réel. Si par exemple, on dispose d'une quantité suffisante de données réelles, alors une partie sera utilisée pour la validation de modèle (estimation des paramètres, loi...), tandis qu'une autre partie sera exploitée pour la validation, toute en la comparant aux résultats de la simulation (Graphiques, intervalle de confiance, Tests,...).

Conclusion

Dans ce chapitre, nous avons introduit les concepts de base de la simulation par événements discrets. Nous avons souligné les caractéristiques et les outils principaux de cette technique, ce qui nous permettra dans la suite de notre travail, de construire un simulateur qui répondra à nos exigences.

Conclusion générale

La complexité des phénomènes et des processus technologiques croit sans cesse. Ceci à pousse les chercheurs à élaborer des théories et des techniques d'analyse et d'approximation qui sont elles mêmes aussi complexes. Parmi ces principales approches, nous nous sommes intéresses à l'analyse des systèmes d'attentes, à l'exploitabilité de ses résultats ainsi qu'à sa mesure de performance par rapport a la technique de simulation, pour le cas des systèmes prioritaires.

Pour évaluer les performances d'un système complexe, il est parfois nécessaire de recourir à des approximations de modèles très compliques par d'autres plus simples ou pour lesquels des résultats analytiques existent.

Pour se faire, nous nous sommes intéresses a la mesure de performance de la méthode analytiques (théorique) par rapport a la technique de simulation. Dans ce mémoire, nous avons prouve l'analyse et la simulation des systèmes de files d'attente avec priorité (absolue et relative). Nous avons commencé par l'actualisation des résultats connus sur les systèmes de files d'attente avec priorité. Lors de l'étude de ces systèmes on a constaté les difficultés de leur analyse. Plusieurs méthodes d'approximations [37, 93, 94, 73, 81, 52] ont été utilises pour approximer ces systèmes d'attente par d'autres dont les caractéristiques sont déjà calcules. Notre contribution consiste à simuler et analyser les systèmes $M_2/G_2/1$, $M_2/M_2/1$, $M_2/M_2/1/(N_1,N_2)$ avec rappels (source fini) et $M_2/M_2/1/(N_1,N_2)$ avec rappels et vacance (source fini). Nous nous sommes intéresses a l'effet de présence des clients prioritaires dans les systèmes déjà citées.

Nous nous somme intéresses aussi à la mesure de performance (nombre moyen de client prioritaire dans la file, nombre moyen de client non prioritaire dans la file, nombre moyen de client non prioritaire dans l'orbite, taux moyen de séjour dans le système, ...), puis la simulation. Nous avons compare les résultats obtenus par la simulation et les valeurs analytique (théorique) et calculer l'erreur entre les résultats. Nous avons montre que l'erreur simulé à chaque fois diminué (pour un grand temps de simulation). Ce qui prouve que l'approximation est possible entre les résultats.

Les résultats obtenus par ce travail ouvre quelques perspectives, comme l'analyse d'autres systemes de files d'attente plus complexes, par exemple :

- aux systemes d'attente prioritaire avec la conservation de service.
- aux systemes d'attente prioritaire et arrivées par groupe.
- aux systemes d'attente prioritaire et pannes.
- aux réseaux de file d'attente avec priorité.

Bibliographie

[1] A.A. Borvokov. Théorème d'ergodicité pour une classe d'équations stochastiques et leurs applications. *Theory of Probability and thier Applications*, 2 (23), 1978.

[2] A.N.Dudin, V. I. Klinerok. The M_1 ;$M_2/G_1(1)$;$G_1(2)$;$G_2/1$ model with the controlled service of the waiting flow and the low-priority retrying flow. in : G. Latouche, P,G. Taylor (Eds.), Advances in Algorithmic Methods for Stochastic Models, proceding of the Tird Intenational Conference on Matrix Analytic Methods. *Leuven12-14*, pages 99-114, july 2000

[3] A.Cobham. Priority assignement in waiting line problems. *Operations Res*, 2 :70-76, 1954

[4] Avi-Itzhak and P.Naor. Some queueing probleme with the service station subject to breakdown. *Operations Res*, 11 :303-320, 1963.

[5] B.Baynat. *Théorie des files d'attente des chaînes de Markov aux réseaux à forme produit.* Hernés . Science paris, 2000.

[6] B.D. Choi , Chang B. MAP_1,MAP_2/ M / c retrial queue with guard channe and its applications to cellular networks. Top 7., pages 231-248, 1999.

[7] B.D. Choi , J. W.Kim. Discrete-time Geo_1/ Geo_2/G /1 retrial queueing systems with two types of calls,. Copmut. Math, Appl, 33 :79-88, 1997.

[8] B.D. Choi , K.K .Park. The M/ G/1 retrial queue with Bernoulli schedule, Queueing Systems. 7 :219-227, 1990

[9] B.D. Choi, D.H. Han, G.I.Flin. On the virtual waiting time for an M/G/1 retrial queue with two types of calls,. *J. Appl. Maath Stochastic Anal.*, 6 :11-23, 1993.

[10] B.D.Choi, D.H.Ham, G.I.Falin. On The virtual waiting time for an M/G/1 retrial queue withtwo type of calls. J.Appl .Math.Statistic , Anal .6 :11-23, 1993.

[11] D.Aissani. Application of the operator methods to obtain inequalities of stability in the $M_2/G_2/1$ system with a relative priority. *Annales Maghrébienes de l'Ingénieur, Numéro hors série*, 2 :pp 790-795, 1991.

[12] F.F.Stephan. Two Queues inder pre-emptive priority with poisson arrival and service rates. *J.Opens.Res.Soc. Amer*, 4, 213-220.

[13] G.Rupert, Miller. jr. Prioity Queues. *The Annals of Mathematical Statistics*, Vol.31(N°1) :86-103, Mars 1960.

[14] H.White and L.S. Christie. Queueing with preemptive or with breakdown.*Operations Res*, 6 :79-95, 1958.

[15] J.Keilson. Queues subject to service interruption. Ann . *Math Statis* , 33 :1314-1322, 1962.

[16] J.L. Holley. Waiting line subject to priorities. *Operations Res*, 2 :341-343, 1954.

[17] N.K. Jaiswal. Time -dependent solution of the head of the line priority queue. *J Roy Stat.Soc*, series B 24 :91-101, 1962.

[18] N.K.Jaiswal. Prioity Queues. Acadimic Press New York and London, University of Southern Califournia, 1968.

[19] P.Gaver. Jr. A Waiting line with interrupted service, including priority. *J ,Roy -Stat. Soc*, Serie B 24 :73-90, 1962.

[20] S.A. Dressin and E.Reich . Priority assignement on a waiting line. *Quart Appel Math*, 15 :208-211, 1957.

[21] B.T.DOSHI, Queueing systems with vacations, a survey. *Queueing Systems-Theory and Applications* 1, 29–66, 1986.

[22] H.TAKAGI, Queueing analysis : A foundation of performance evaluation, Vol. I, vacation and priority, Part I. North-Holland, Amsterdam, 1991.

[23] J.TEGHEM , Control of the service processing a queueing system. European Journal of Operational Research 23, 141–158, 1986.

[24] N.TIAN et Z.G.ZHANG, Vacation queueing models : Theory and Applications. Ed. Springer Science, Business Media, LLC, 2006.

[25] B.Mohamed,Systemes d'attente avec rappelles et vacance, Thèse doctora,2009.

[26] S.Hakim, Analysis of RF communication, 2009.

[27] D,Aissni, N.V.Kartchov, Strong stability of an in beded Markov chin in an M/G/1 system. theory of probability and mathematical statistcs, Americain mathematical, sosiety, 29 :1-5.

[28] A.S.Alfa, Matrix-geometric solution of descret time MAP/PH/1,*priorety queue Naval Research logistics*, 45(1) :23-50.

[29] V.V.Anisimov, Swiching processes in queueing models, *Wiley ISTE. Glaxo Smith Kline.UK*.

[30] J.R.Artalejo, Accessible bibliography on retrial queues, *Mathematical and Computer Modeling*, 30 :1-6, 1999.

[31] G.ayyappan, M/M/1 Retrial queueing system with N-policy multiplt vacation under non premitive priority servise by matrix geomitric method. *Applied Mathematical Sciences*. Vol, 4(23),1141-1154

[32] B.D.Choi, K.K.Park, The M/G/1 retrial queue with bernoulli schedule queueing system,7 :219-227,1990.

[33] B.D.Choi, K.K.Park, The M/G/1 retrial queue with bernoulli schedule queueing system, 7 :219-227, 1990.

[34] B.D.Choi, J.Chung, M/M/1 queueing with impctient customer of higher priority Queueting system, 38 :49-66.

[35] A.N.Dudin et V.I.Klimenok, The M_1,M_2/G_1,G_2 /1 model with the controlled service of the waiting flow and the low priority retrying flow.*The tird internationnal confrance on Matrix Analytic methods*,99-114.

[36] J.A.Hooke, A priority queue with low priority arrivals general, *Operation Research*, 20 :373-380, 1972.

[37] A.Itzhak,A.Naor, Some queueing probleme with the service station subject to break down, *OperationsRes*, 11, 303-320,1963 .

[38] W.Grassmann, inding transient solutions in markovian event systems through randomization, Numerical solutions of Markov chains , *Marcel Dekker, New yourk*, 357-371, 1991.

[39] V.V.Kalachinkov, G.S.Tistsiavihi, On the stability of queueing systems with respect to disturbances of thier distrubution functions, *Journal Izu AN USSR Techniques Cybernetiques*,2,41-94,1972.

[40] N.V.Kartashov, Cristeria for uniform ergodicity and strang stable Markov chains ,*VSPUtrech TbiMC Scientific Publishers*, 1985.

[41] N.HAMADOUCHE, Approximation dans les systèmes En prioritaires, Mémoire de magister,Université A. Mira- Béjaia, 2004.

Résumé

Dans ce travail, nous donnons une brève analyse des systèmes de files d'attente avec différentes priorités (relative, absolue et on introduit le mixage de ces deux priorités) et ce pour l'étude étudier un système de file d'attente avec rappels et vacances avec ces différents priorités avec lequel on modélisera le sytème de communication radio fréquence dans les réseaux de capteurs.
Après, cette analyse et cette modélisation, une simultion des modèles obtenus est réalisée. Cette simulation permet le calcul des mesures de performance des différents modèles considérés. Une interprétation des résultats de la simulation illuste l'intérêt de l'introdution de la politique de mixage des deux priorités, relative et absolue.

Mots-clés
Système de file d'attente, priorité relative, priorité absolue, Rappels, vacances, chaîne de Markov, algorithme.

Abstract :

In this work, we give a brief analysis of the queuing systems with different priorities (relative, absolute and we introduce the mixing of these two priorities) and this in order to study a system of retrial queues with vacancy and these different priorities. With this system the radio frequency communication in sensors networks will be modeled.
After this analysis and modeling, a simulation of the models obtained is carried out. This simulation permits to compute the performance measures of the different considered models. An interpretation of the simulation results illustrates the necessary of introducing the policy of mixing the two priorities, relative and absolute.

keywords :
Queueing system, preemptive priority, non preemptive priority, retrials, vacancy, Markov Chain, Algorithm.