



الجمهورية الجزائرية الديمقراطية الشعبية  
République Algérienne Démocratique et Populaire  
وزارة التعليم العالي والبحث العلمي



Ministère de l'Enseignement Supérieur et de la Recherche Scientifique  
جامعة الشاذلي بن جديد - الطارف  
Université Chadli Bendjedid – El Tarf  
كلية العلوم و التكنولوجيا  
Faculté des Sciences et de la Technologie  
قسم الرياضيات  
Département de Mathématiques

## Mémoire de fin d'études

En vue de l'obtention du diplôme de Master

**Domaine :** Mathématiques et Informatique

**Filière :** Mathématiques

**Spécialité:** Analyse fonctionnelle et calcul stochastique

### Thème

La théorie des fils d'attente et ses applications

**Présenté par:**

**NOUIKES SOUHAILA**

**Devant le Jury :**

<b>Dr. GRINE RAZIKA</b>	<b>MCB</b>	Univ Chadli Bendjedid El Tarf	Présidente
<b>Dr. ZIDANI NESRINE</b>	<b>MCB</b>	Univ Chadli Bendjedid El Tarf	Rapporteur
<b>Dr. BENSEGHIR RYM</b>	<b>MCA</b>	Univ Chadli Bendjedid El Tarf	Examinatrice

**Année Universitaire 2022-2023**

---

# Remerciements

*Avant tout, je remercie le bon Dieu qui a illuminé mon chemin et qui j'ai armés le courage et de bonne volonté pour réaliser ce travail. je tenu exprimer mes plus profond remerciement à mes chers parents et toute la famille pour leur amour, leur aide et leur présence au quotidien.*

*Je tenu à remercier vivement Mme ZIDANI NESRINE, pour m'avoir proposée le sujet de ce mémoire, pour m' avoir honorée par sa encadrement, pour sa disponibilité, ses orientations, ses conseils, ses encouragements tout au long de la réalisation de ce mémoire.*

*Je tenu également à exprimer notre gratitude au présidente du jury Mme GRINE RAZIKA pour m' avoir consacré de sa temps il m'a fais l'honneur d'accepter de présider le jury.*

*Je remercie très chaleureusement Mme BENSEGHIR RYM d'avoir accepté d'examiner mon travail*

*J'adresse également des remerciements à tous les enseignants de la faculté de l'université de Chadli Benjdid. Ainsi que tous les membres du département de Mathématique, pour toute l'aide qui m'été accordée.*

---

# Dédicace

*Avec ma gratitude et tout mon amour je dédie ce modeste travail :*

*À l'être la plus chère de ma vie ma mère GHANIA, tu trouves toujours les mots qui m'encourageaient et qui me poussaient à aller de l'avant, quoi que je fasse ou je dise, je ne saurais point te remercier comme il se doit. Ton affection me couvre ta bienveillance me guide et ta présence à mes côtés a toujours été ma source de force pour affronter les différents obstacles.*

*À mon très cher père MOHAMED, pour son soutien, son affection et la confiance qu'il m'a accordé. J'espère que par ce modeste travail, je vous rends un peu de ce sentiment de fierté que j'éprouve d'être votre fille. Que dieu leur procure bonne santé et longue vie.*

*À ma grande mère ZAHYA de m'encourager et de prier pour moi.*

*À mes grands parents qui sont décédés : ZHAIRA, BRAHIM et SALAH.*

*À mes chères belles sœurs : ASMA, CHAIMA qui m'avez toujours soutenu et encouragé durant ces années d'études.*

*À mon très cher frère : ISLEM que je souhaite réussite dans ta vie.*

*À mes nièces : DINA, LILYANE, MARIA et JINEN.*

*À mon très cher AHMED pour tout son aide, son soutien et ses encouragements.*

*À mes chères amis : RAYEN, ILHEM, AYA et IBTISSEM qui me sont chers je leur dédie ce mémoire pour tous les moments de joie qu'on a passé ensemble.*

*À tous les membres de ma famille.*

---

# Résumé

Dans ce travail, nous avons réalisé une synthèse sur la théorie des files d'attente classique et avec rappels.

Dans un premier temps, nous avons effectué une étude bibliographique sur les systèmes de files d'attente classiques, et avons présenté les modèles les plus connus ( $M/M/c$ ,  $M/M/1$  et  $M/G/1$ ).

Dans un deuxième temps, nous avons étudié les modèles d'attente avec rappels, en particulier le système  $M/M/c$  et le système  $M/G/1$ . Puis, nous avons présenté quelques exercices de files d'attente.

**Mots clés :** La files d'attente, client, rappels,  $M/M/1$ ,  $M/M/c$ , processus stochastique.

---

## ملخص

في هذا العمل، أجرينا توليفة حول نظرية قوائم الانتظار الكلاسيكية ومع الاسترجاعات.

أولاً، أجرينا دراسة بيبليوغرافية عن أنظمة الاصطفاف الكلاسيكية، و قدمنا النماذج الأكثر شهرة  $M/M/1$ ،  $M/M/c$  و  $M/G/1$ .

ثانياً، درسنا نماذج الانتظار مع الاسترجاعات، ولا سيما نظام  $M/M/c$  و نظام  $M/G/1$  ثم قدمنا بعض التمارين حول أنظمة الطابور الكلاسيكية ومع الإسترجاعات.

الكلمات المفتاحية: قائمة الانتظار، العملاء، عمليات الاسترجاعات،  $M/M/1$ ،  $M/M/c$  عملية عشوائية.

---

# Abstract

In this work, we have produced a synthesis of classical queuing theory, and retrial queue.

First, we reviewed the literature on classical queuing systems, and presented the best-known models ( $M/M/c$ ,  $M/M/1$  and  $M/G/1$ ).

Secondly, we studied retrial queueing models, in particular the  $M/M/c$  and  $M/G/1$  systems. Then, we presented some queueing exercises.

**Key words :** Queuing, customer, retrial,  $M/M/1$ ,  $M/M/c$ , Stochastic process.

---

# Table des matières

<b>Introduction</b>	<b>10</b>
<b>1 Généralités sur la théorie des probabilités</b>	<b>11</b>
1.1 Introduction . . . . .	11
1.2 Notions de base de probabilité . . . . .	11
1.2.1 Probabilité élémentaire . . . . .	11
1.2.2 Variable aléatoire . . . . .	12
1.2.3 Fonction génératrice . . . . .	12
1.2.4 Transformée de Laplace-Stieltjes . . . . .	13
1.3 Processus stochastique . . . . .	13
1.3.1 Processus de Naissance et de Mort . . . . .	15
1.3.2 Processus de Poisson . . . . .	18
<b>2 Systèmes de files d'attente classiques</b>	<b>19</b>
2.1 Introduction . . . . .	19
2.2 Historique de files d'attente . . . . .	19
2.3 Définition et classification des systèmes de files d'attente . . . . .	20
2.4 Analyse mathématique d'un système de files d'attente . . . . .	23
2.5 Système de files d'attente M/M/c . . . . .	24
2.5.1 Description du modèle . . . . .	24
2.5.2 Analyse du modèle . . . . .	24
2.6 Cas particuliers du système M/M/c . . . . .	26
2.6.1 Système M/M/1 . . . . .	26
2.6.2 Système M/M/∞ . . . . .	27
2.7 Système de files d'attente M/G/1 . . . . .	27
2.7.1 Description du modèle . . . . .	28

2.7.2	Analyse du modèle . . . . .	28
2.7.3	Cas particuliers du modèle M/G/1 . . . . .	33
2.8	Conclusion . . . . .	34
<b>3</b>	<b>Système de files d'attentes avec rappels</b>	<b>35</b>
3.1	Introduction . . . . .	35
3.2	Terminologie et notation . . . . .	36
3.3	Politiques d'accès au serveur à partir de l'orbit . . . . .	37
3.4	Quelques exemples modélisés par des systèmes de files d'attente avec rappels	38
3.5	Modèle M/M/c avec rappels . . . . .	39
3.5.1	Description du modèle . . . . .	39
3.6	Modèle d'attente M/M/1 avec rappels . . . . .	41
3.6.1	Description du modèle . . . . .	41
3.7	Modèle d'attente M/G/1 avec rappels . . . . .	43
3.7.1	Description du modèle . . . . .	43
3.7.2	Méthode des variables supplémentaires . . . . .	43
3.7.3	Mesures de performance . . . . .	47
3.8	Conclusion . . . . .	47
<b>4</b>	<b>Exercices</b>	<b>48</b>
	<b>Conclusion générale</b>	<b>55</b>
	<b>Bibliography</b>	<b>56</b>

## Introduction

La modélisation stochastique est l'application de la théorie des probabilités pour la description et l'analyse des phénomènes du monde réel. Ces derniers sont habituellement si complexes que les lois déterministes ne peuvent pas être formulées. Une circonstance qui mène à un usage envahissant du concept stochastique. Les applications en télécommunication ou en assurance apportent des méthodes et des résultats pour la modélisation stochastique aux sciences appliquées telles que l'ingénierie et l'économie.

Le développement technologique donne souvent lieu à de nouvelles questions, qui en revanche ouvrent une nouvelle perspective pour la recherche stochastique, et fournit donc un élan pour l'application des probabilités. La modélisation stochastique est une science à interaction entre la théorie et l'application pratique; elle combine entre la théorie et les concepts clé du monde réel.

L'un des domaines les plus importants dans la modélisation stochastique est la théorie des files d'attente.

La théorie des files d'attente est un outil de modélisation développé dans les années 1930 pour l'analyse statistique des systèmes téléphoniques (évaluation des performances, dimensionnement du système,...). Cette technique se retrouve aujourd'hui la meilleur pour l'analyse des systèmes informatiques, et des réseaux d'ordinateurs. La modélisation mathématique à l'aide de la théorie des files d'attente est un outil de la logistique. Elle relève du calcul des probabilités : les arrivées et les départs des clients de la file sont analysés le plus souvent comme un processus stochastique typique d'un processus de naissance et de mort.

La théorie des files d'attente classique offre deux possibilités pour résoudre le conflit qui apparaît lorsque un client arrive dans le système à serveur occupé : soit le client quitte le système sans recevoir le service (système à demandes refusées, modèle d'Erlang), soit il prend place dans une file d'attente (système de files d'attente).

Une possibilité alternative est de permettre au client de répéter sa demande de service après une durée de temps aléatoire. Entre deux appels successives (rappels), le client

en question est "en orbite". Un tel système est appelé système de files d'attente avec rappels.

Pour identifier un système de files d'attente, on a besoin de spécifier le flux d'entrée, le mécanisme de service, la discipline d'attente et éventuellement le flux des tentatives répétées. La dimension d'une population des clients potentiels (ou le nombre de sources) peut être finie ou infinie. Un modèle avec la population finie est analytiquement plus compliqué parce que le nombre de clients déjà dans le système à n'importe quel instant affecte le nombre de clients potentiels composant la population. Le processus (flux) des arrivées peut être régulier ou aléatoire. Dans le premier cas, les arrivées des clients se suivent à des intervalles de temps déterminés. Dans les systèmes réels, on rencontre rarement de processus des arrivées de ce genre. Pour les systèmes plus typiques, le flux de demandes est aléatoires (la durée de temps entre deux arrivées successives suit une loi de probabilité).

Le mécanisme de service comprend le nombre de serveurs et la distribution des durées de service. En général, on admet que la population des clients est homogène, c'est-à-dire que les durées des services des clients sont identiquement distribuées selon une loi de probabilité commune. Dans les cas plus complexes où les clients sont divisés en classes, chaque classe peut être identifiée par sa propre distribution du temps de service (population hétérogène). Il y a des systèmes où certains clients jouissent d'une priorité de service. La priorité peut être absolue ou relative. Par priorité absolue, on entend qu'un client moins prioritaire est remis en tête de file d'attente lorsqu'un client plus prioritaire se présente devant la file d'attente. Ce dernier venu commence son service immédiatement. Si la priorité est relative, un nouveau client plus prioritaire attend la fin du service avant de pouvoir commencer le sien.

Les clients peuvent être choisis et servis dans l'ordre d'arrivée (FIFO), ou (LIFO), ou choisis au hasard (RANDOM). La capacité de l'espace d'attente peut être illimitée ou non. Dans le deuxième cas, certains clients qui arrivent vers le système n'ont pas la possibilité d'y entrer.

Puisque les instants d'arrivée et les durées de service sont généralement des quantités aléatoires, le processus décrivant le fonctionnement d'un système de files d'attente est un processus aléatoire (stochastique). Par ailleurs, on suppose généralement que toutes les variables aléatoires introduites pour décrire un système d'attente sont mutuellement indépendantes. L'étude mathématique d'un système de files d'attente se fait par l'introduction du processus

en question. En fonction des quantités qui déterminent la structure du système, on cherche à calculer les probabilités d'état définissant le régime stationnaire. La distribution stationnaire de l'état du système permet d'obtenir les mesures de performance qui peuvent être utilisées dans la pratique.

Dans ce travail, on se propose de réaliser une synthèse sur la théorie des files d'attente (classiques et avec rappels) et nous étudions quelques modèles de files d'attente avec rappels et classiques les plus célèbres.

Le mémoire est composé de quatre chapitres. Dans le premier, nous présentons les outils préliminaires que nous allons utiliser tout au long de ce travail. Nous présentons les notions de probabilité et les fonctions génératrices, et aussi Transformée de Laplace-Stieltjes qui jouent un rôle très important par la suite dans la théorie des files d'attente. Nous présentons aussi les processus stochastiques. Le deuxième chapitre est un rappel des résultats sur les systèmes de file d'attente classiques, nous passons en revue quelques systèmes les plus connus. Le troisième chapitre comprend une étude des systèmes de file d'attente avec rappels, en particulier les modèles  $M/M/c$  et  $M/G/1$  avec rappels. Le dernier chapitre est consacré à l'application des résultats théoriques obtenus sur les systèmes de files d'attente classiques et avec rappels.

---

# Généralités sur la théorie des probabilités

## 1.1 Introduction

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

Dans ce chapitre, nous présentons les outils préliminaires que nous allons utiliser tout au long de ce travail. Nous présentons tout d'abord, les notions de base de probabilité (Probabilité élémentaire, variable aléatoire, fonctions génératrices et aussi Transformée de Laplace-Stieltjes, qui joueront un rôle très important par la suite). Ensuite, nous évoquons les processus stochastiques.

## 1.2 Notions de base de probabilité

### 1.2.1 Probabilité élémentaire

Soit  $\Omega$  l'espace de toutes les réalisations possibles d'une expérience aléatoire.  $\Omega$  est l'univers des événements. Chaque élément  $w$  de  $\Omega$  est appelé événement élémentaire. En théorie des probabilités, on s'intéresse souvent aux sous ensembles de  $\Omega$  appelés événements.

Soit  $\mathcal{A} = P(\Omega)$  une tribu d'événements. On définit alors : [9, 20]

1.  $\Omega \in \mathcal{A}$ ;
2.  $\forall A \in \mathcal{A}, \bar{A} \in \mathcal{A}$  avec  $\bar{A}$  le complémentaire de  $A$ ;
3.  $\forall A, B \in \mathcal{A}, A \cup B \in \mathcal{A}$ ;
4.  $\forall (A_i)_{i \in \mathbb{N}}, (\cup_{i \in \mathbb{N}} A_i) \in \mathcal{A}$ .

**Définition 1.2.1** On appelle probabilité sur l'espace probabilisable  $(\Omega, P(\Omega))$  toute application  $P$  de  $P(\Omega)$  dans  $[0, 1]$ , qui vérifie les conditions suivantes :

1.  $P(\Omega) = 1$
2. si  $A_1, A_2, \dots$  est une suite d'éléments de  $\mathcal{A}$  deux-à-deux disjoints (incompatibles), alors  $P(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$ .

## 1.2.2 Variable aléatoire

On appelle variable aléatoire, notée souvent  $X$ , toute variable dont la valeur dépend de l'issue d'une expérience aléatoire. A chaque événement  $w$  de  $\Omega$ , on associe un nombre  $X(w)$ . Soit  $E$  l'espace d'état, définie comme l'ensemble des valeurs possibles de la variable aléatoire  $X$  ( $E$  peut être fini ou infini). En ce sens,  $X$  est une application dite *mesurable* de l'espace  $(\Omega, P)$  dans  $E$ .

L'espérance mathématique et la variance d'une variable aléatoire se notent respectivement  $E[X]$  et  $\sigma^2[X]$ . [16]

## 1.2.3 Fonction génératrice

Soit  $X$  une variable aléatoire discrète non négative, telle que

$$P(X = n) = P_n, \quad n = 0, 1, 2, \dots$$

**Définition 1.2.2** On appelle fonction génératrice de  $X$  la série entière

$$G_X(t) = \sum_{n=0}^{+\infty} p_n t^n. \quad (1.1)$$

Du fait que  $P_n \geq 0$  et que  $\sum_{n=0}^{\infty} P_n = 1$ , la fonction  $G_X(t)$  est définie pour  $t$  telle que  $|t| \leq 1$  ( $t$  est une variable complexe).

$$P_x(0) = P(0), \quad P_x(1) = E[X] \quad (1.2)$$

Et d'une manière générale,

$$P_x^{(k)} = 1 = E(X(X-1)\dots(X-k+1));$$

où l'exposant  $(k)$  désigne la dérivée  $k^{\text{ième}}$ .

**Théorème 1.2.3** Soient  $G_{X_1}, \dots, G_{X_n}$  les fonctions génératrices respectives de  $n$  variables aléatoires indépendantes  $X_1, \dots, X_n$  et à valeurs dans  $\mathbb{N}$ . La variable aléatoire  $X_1 + \dots + X_n$  a une fonction génératrice  $G_S$  donnée par

$$G_S(t) = \prod_{i=1}^n G_{X_i}(t).$$

### 1.2.4 Transformée de Laplace-Stieltjes

La transformée de Laplace-Stieltjes  $\tilde{X}(s)$  d'une variable aléatoire non négative  $X$  de fonction de répartition  $F(\cdot)$  et définie par :

$$\tilde{X}(s) = E[e^{-sX}] = \int_{x=0}^{\infty} e^{-sx} dF(x), \quad s \geq 0.$$

Lorsque la variable aléatoire admet une densité de probabilité  $f(\cdot)$ , la transformée se simplifie à :

$$\tilde{X}(s) = \int_{x=0}^{\infty} e^{-sx} f(x) dx.$$

Notifions que  $|\tilde{X}(s)| \leq 1$  pour tout  $s = a + ib$ , tel que  $a \geq 0$ .

$$\tilde{X}(0) = 1, \quad \tilde{X}'(0) = E[X], \quad \tilde{X}^{(k)}(0) = (-1)^k E[X^k].$$

Où l'exposant  $(k)$  désigne la dérivée  $k^{\text{ième}}$ .

## 1.3 Processus stochastique

Un processus stochastique  $\{X(t), t \in T\}$  est une fonction du temps dont la valeur à chaque instant  $t \in T$ ,  $X(t)$  est donc une variable aléatoire. Un processus stochastique peut être considéré comme une famille de variables généralement non indépendantes. L'ensemble des

temps  $T$  peut être discret ou continu.  $X(t)$  définit l'état du processus à un instant donné  $t$ . L'ensemble noté  $E$  des valeurs que peut prendre le processus à chaque instant est appelé espace d'états et peut, de même que  $T$ , être discret (fini ou infini) ou continu. En fonction des valeurs possibles de  $T$  et  $E$ , on classe les processus stochastiques de la façon suivante [1, 3, 5, 9] :

- Processus à temps discret et à espace d'état discret.
- Processus à temps continu et à espace d'état discret.
- Processus à temps discret et à espace d'état continu.
- Processus à temps continu et à espace d'état continu.

**Définition 1.3.1** *Un processus stochastique est une famille de variables aléatoires, définie sur  $(\Omega, \mathcal{F})$  à valeurs dans  $(E, \Sigma)$ , notée par  $\{X(t), t \in T\}$  où  $T$  est un ensemble ordonné quelconque tel que*

$$\begin{aligned} X &: \Omega \times T \rightarrow E \\ (\omega, t) &\rightarrow X(\omega, t) \end{aligned}$$

Pour  $\omega \in \Omega$ , fixé l'application  $t \mapsto X_t(\omega)$  de  $T$  dans  $E$  est appelée trajectoire du processus.

**Définition 1.3.2** *Nous disons qu'un processus stochastique  $X_t$  est continue si presque tout  $\omega \in \Omega$ , ses trajectoires sont continues, i.e.,*

$$\mathbb{P}\{\omega \in \Omega, t \mapsto X_t(\omega) \text{ est continue}\} = 1. \quad (1.3)$$

**Définition 1.3.3** *Un processus stochastique  $\{X(t), t \in T\}$  à valeurs dans  $\mathbb{R}^n$  est dit séparable, s'il existe une suite dénombrable  $T = \{t_i\}$  dense dans  $I$ , un sous ensemble  $N$  de  $\Omega$ , avec  $\mathbb{P}(N) = 0$  tels que, si  $\omega \notin N$ , on a  $\{X(t, \omega) \in F, \forall t \in J\} = \{X(t_j, \omega) \in F, \forall t_j \in J\}$ , pour tout sous ensemble ouvert  $J$  de  $I$  et pour tout ensemble fermé  $F$  de  $\mathbb{R}^n$ .*

**Définition 1.3.4** *Deux processus stochastiques  $\{X(t), t \in I\}$  et  $\{X'(t), t \in T\}$  définis sur le même espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$  sont stochastiquement équivalents si  $\mathbb{P}\{X(t) \neq X'(t)\} = 0, t \in T$ . Dans ce cas nous disons que  $X'(t)$  est une version de  $X(t)$ .*

**Définition 1.3.5** *Un processus  $X$  est dit à accroissements indépendants si pour tout  $p \geq 1$  et  $0 < t_1 < t_2 \dots < t_p$ , les variables aléatoires  $X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_p} - X_{t_{p-1}}$  sont indépendantes (sont mutuellement indépendants).*

**Définition 1.3.6** *Un processus est dit à accroissements stationnaires si la loi des accroissements  $X_{t+h} - X_t$  ne dépend pas de  $t > 0$ , i.e.  $X_{t+h} - X_t$  à même loi que  $X_h$ .*

### 1.3.1 Processus de Naissance et de Mort

Ces processus permettent de façon générale de décrire l'évolution temporelle de la taille d'une population d'un type donné [14, 17]. Dans le cas d'un système d'attente, on considère par exemple des populations comprenant tous les clients qui sont dans le système à l'instant  $t$ .

Les processus de naissance et de mort sont des processus stochastiques à temps continu et à espace d'états discret ( $S = \{0, 1, \dots\}$ ). Ils sont caractérisés par deux conditions importantes :

1. Sans mémoire ;
2. À partir d'un état donné  $n$ , des transitions ne sont possibles que vers l'un ou l'autre des états voisins  $(n + 1)$  et  $(n - 1)$  pour  $(n \geq 1)$ .

Soit  $\{N(t), t \geq 0\}$ , où  $N(t)$  est le nombre d'individus dans la population à la date  $t$ , avec  $S = \{0, 1, \dots\}$ . Le processus de Naissance et de Mort est caractérisé par l'apparition et la disparition d'un individu au sein de la population. Il est homogène dans le temps si la probabilité :

1. d'apparition d'un individu pendant l'intervalle  $\Delta t$  sachant qu'il existe déjà  $k$  individus au sein de la population, qui est  $\lambda_k \Delta t + o(\Delta t)$ ,
2. d'apparition d'aucun individu pendant l'intervalle  $\Delta t$  sachant qu'il existe déjà  $k$  individus au sein de la population, qui est  $1 - \lambda_k \Delta t + o(\Delta t)$ ,
3. d'apparition de deux ou plus individus pendant l'intervalle  $\Delta t$  sachant qu'il existe déjà  $k$  individus au sein de la population, qui est  $o(\Delta t)$ ,
4. de disparition d'un individu pendant l'intervalle  $\Delta t$  sachant qu'il existe déjà  $k$  individus au sein de la population, qui est  $\mu_k \Delta t + o(\Delta t)$ ,
5. de disparition d'aucun individu pendant l'intervalle  $\Delta t$  sachant qu'il existe déjà  $k$  individus au sein de la population, qui est  $1 - \mu_k \Delta t + o(\Delta t)$ ,
6. de disparition de deux ou plus individus pendant l'intervalle  $\Delta t$  sachant qu'il existe déjà  $k$  individus au sein de la population, qui est  $o(\Delta t)$ ,

est indépendante de la position de  $\Delta t$  sur l'axe des temps. Ici,  $\lambda_k$  est le taux d'apparition (de croissance),  $\mu_k$  est le taux de disparition (de décroissance). Encore,

$P(N(t+s) = j/N(s) = i) = p_{ij}(t)$  ne dépend pas de  $s$ . Alors,

$$p_{i,i+1}(\Delta t) = \lambda_i \Delta t + o(\Delta t), \quad i \geq 0;$$

$$p_{i,i-1}(\Delta t) = \mu_i \Delta t + o(\Delta t), \quad i \geq 1;$$

$$p_{i,i}(\Delta t) = 1 - (\lambda_i + \mu_i) \Delta t + o(\Delta t), \quad i \geq 0;$$

$$p_{i,j} = o(\Delta t), \quad |i - j| \geq 2;$$

$$p_{i,j}(0) = \delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}.$$

$\lambda_i$  et  $\mu_i$  sont appelés taux de transition (taux de naissance et de mort).

### Régime transitoire

Soient  $p_n(t) = P(N(t) = n)$ ,  $n \geq 0$ , les probabilités d'état.

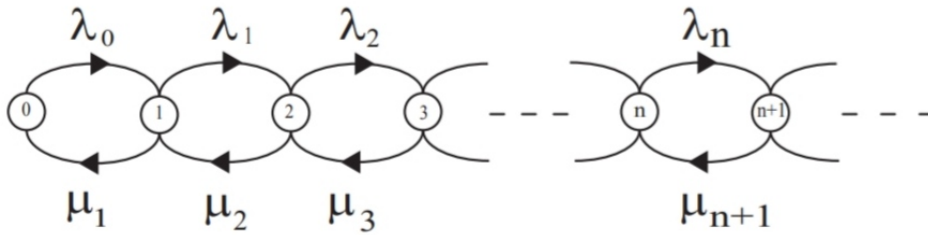


Figure 1 : Graphe de transition

La matrice des transitions correspondante est

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

En appliquant  $P(t + \Delta t) = P(t) \times M$ , on trouve

$$\begin{cases} p_0(t + \Delta t) = (1 - \lambda_0 \Delta t) p_0(t) + \mu_1 \Delta t p_1(t), \\ p_n(t + \Delta t) = \lambda_{n-1} \Delta t p_{n-1}(t) + [1 - (\lambda_n + \mu_n) \Delta t] p_n(t) + \mu_{n+1} \Delta t p_{n+1}(t), \quad n \geq 1. \end{cases} \quad (1.4)$$

On déduit les équations de Kolmogorov

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

**Remarque 1.3.1** Si  $S = \{0, 1, \dots, K\}$ , alors  $\lambda_K = 0$ . D'où

$$p'_K(t) = \lambda_{K-1}p_{K-1}(t) - \mu_K p_K(t).$$

**Remarque 1.3.2** Les équations de Kolmogorov, complétées par des conditions initiales, gouvernent le régime transitoire du processus  $\{N(t), t \geq 0\}$ .

### Régime stationnaire

Soit  $p_n = \lim_{t \rightarrow \infty} p_n(t)$ , qui est la distribution stationnaire de processus étudié. Ces probabilités satisfont le système d'équations de balance suivant :

$$\begin{cases} \lambda_0 p_0 = \mu_1 p_1; \\ (\lambda_n + \mu_n) p_n = \mu_{n+1} p_{n+1} + \lambda_{n-1} p_{n-1}, \quad n \geq 1; \end{cases} \quad (1.5)$$

Avec l'équation de normalisation  $\sum_{n=0}^{\infty} p_n = 1$ . De (1.5), on obtient,

$$p_1 = \frac{\lambda_0}{\mu_1} p_0;$$

Pour  $n = 1$  :

$$(\lambda_1 + \mu_1) p_1 = \mu_2 p_2 + \lambda_0 p_0;$$

d'où

$$\begin{aligned} p_2 &= \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} p_0; \\ &\vdots \\ p_n &= \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{\mu_1 \mu_2 \dots \mu_n} p_0. \end{aligned}$$

Pour déduire  $p_0$ , on utilise l'équation de normalisation. On obtient le résultat suivant

$$p_0 = \left[ 1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} + \dots + \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{\mu_1 \mu_2 \dots \mu_n} + \dots \right]^{-1}.$$

Par conséquent, pour qu'une distribution stationnaire existe, il faut donc que la somme [ ] converge. Ceci a toujours lieu si l'espace d'états du processus à l'étude est fini. Lorsque la somme en question n'est pas convergente,  $p_n = 0$ ,

$\forall n \geq 0$ .

### 1.3.2 Processus de Poisson

Le processus en question est utilisé pour décrire la réalisation dans le temps d'événements aléatoires d'un type donné. La description mathématique d'un flux d'événements aléatoires peut se faire de deux manières différentes :

1. On considère le nombre d'événements  $X(t)$  se produisant dans  $[0, t]$  et on cherche à déterminer la loi de probabilité de cette variable aléatoire discrète. Le processus  $\{X(t), t \geq 0\}$  est appelé *processus de comptage*.
2. On considère les intervalles de temps qui séparent les instants d'apparition de deux événements consécutifs. Ce sont des v.a continues, positives et en général indépendantes et identiquement distribuées.

On dit qu'un processus de comptage  $\{X(t), t \geq 0\}$  est un processus de Poisson s'il satisfait aux trois conditions suivantes :

1. Le processus est homogène dans le temps : la probabilité d'avoir  $k$  événements dans un intervalle de longueur  $t$  ne dépend que de  $t$  et non pas de la position de l'intervalle par rapport à l'axe temporel :  $p_k(t) = P(X(t) = k)$ .
2. Pour tout système d'intervalles disjoints, les nombres des événements s'y produisant sont des variables aléatoires indépendantes.
3. La probabilité

$$p_k \Delta t = \begin{cases} (\Delta t) & \text{si } k \geq 2; \\ \lambda \Delta t + o(\Delta t) & \text{si } k = 1; \\ 1 - \lambda \Delta t + o(\Delta t) & \text{si } k = 0. \end{cases}$$

où  $\lambda$  est la densité ou intensité du processus (le nombre moyen d'événements qui apparaissent par unité de temps).

**Théorème 1.3.7** *Pour un processus de Poisson, on a :*

$$P(X(t) = k) = p_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad \lambda > 0, \quad k \geq 0; \quad (1.6)$$

$$E[X(t)] = \lambda t \text{ et } \text{Var}[X(t)] = \lambda t. \quad (1.7)$$

Ces relations définissent le régime transitoire du processus de Poisson. Aucun régime stationnaire n'existe vu que  $p_k = \lim_{t \rightarrow \infty} p_k(t) = 0, \forall k \geq 0$ .

---

# 2

## Systèmes de files d'attente classiques

### 2.1 Introduction

La théorie des files d'attente est une technique de la recherche opérationnelle qui permet de modéliser un système admettant un phénomène d'attente, de calculer ses performances et déterminer ses caractéristiques pour aider les praticiens dans leurs prises de décisions.

L'objet de ce chapitre est la modélisation des files d'attente par des processus aléatoires. Nous introduisons les concepts généraux des files d'attentes classiques. Nous accordons une attention particulière à la présentation des modèles  $M/M/c$  et  $M/G/1$ .

### 2.2 Historique de files d'attente

Les phénomènes d'attente ont été décrits depuis un siècle déjà pour étudier les fluctuations de la charge des lignes et des centraux téléphoniques. Due à l'importance de ce phénomène dans notre vie quotidienne, plusieurs recherches ont été consacrées pour étudier et trouver les meilleures approximations aux systèmes réels.

Les premiers modèles des systèmes d'attente ont été proposés par l'ingénieur électrique danois, **A.K. Erlang** au début du 20<sup>ième</sup> siècle (1917). Son étude était suivie par celle du **E.C. Molina** 1927 et une année après par **Thornton**. Au début des années trente, **Felix Pollaczek** a effectué un certain travail sur l'entrée de Poisson et les problèmes à un seul et à multiserveurs. Le travail a été effectué à ce moment-là par **Kolmogorov** et **Khintchine** en Russie, par **Crommelin** en France ainsi que par **Palm** en Suède [4, 10].

L'avancement de la recherche dans la théorie des files d'attente a été lente en ses débuts parce que leur utilisation faisait appel à une puissance de calcul non disponible à cet époque

mais a accéléré dans les années cinquante et il y'a eu beaucoup de travail dans ce secteur depuis ces années.

La théorie des files d'attente est une théorie mathématique relevant du domaine des probabilités qui étudie les solutions optimales de gestion de files d'attente (queues). Une queue est nécessaire et se créera d'elle même si elle n'est pas anticipée dans tous les cas où l'offre est inférieure à la demande, même temporairement. L'objectif de la théorie des systèmes de files d'attente est de minimiser le coût total qui équivaut à la somme de deux coûts : le coût associé à la capacité de service mise en place et le coût associé à l'attente des clients.

## 2.3 Définition et classification des systèmes de files d'attente

La formation des files d'attente est un phénomène fréquent qui apparaît chaque fois que les demandes de service dépassent la capacité admise des dispositifs de service.

Le modèle général d'un phénomène d'attente (système d'attente) peut être résumé comme suit : les demandes (clients) arrivent à un certain endroit et réclament un certain service. Si un dispositif de service est libre, la demande qui arrive se dirige immédiatement vers ce dispositif où elle est servie. Dans le cas contraire, on a deux possibilités : soit la demande quitte le système (systèmes à demandes refusées), soit elle prend place dans un file d'attente (système de files d'attente). A un moment donné, la demande est sélectionnée pour service selon une discipline. Après l'achèvement du service, la demande quitte le système.

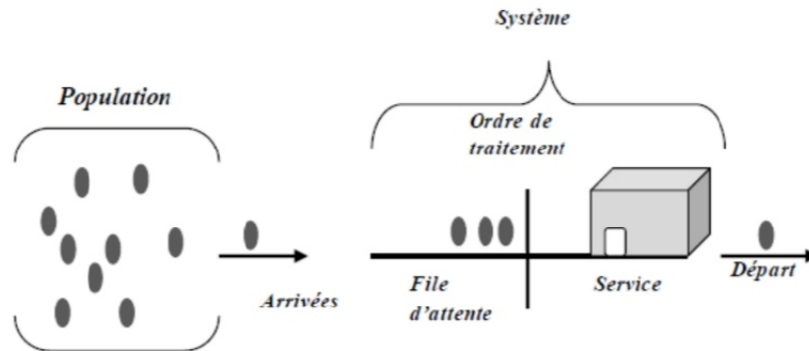
Un système de files d'attente comprend donc un espace de service avec un ou plusieurs dispositifs de service (serveurs) et un espace d'attente dans lequel se forme une éventuelle file d'attente.

Pour identifier un système de files d'attente, on a besoin de spécifier le flux d'entrée, le mécanisme de service et la discipline d'attente.

La dimension d'une population finie des clients potentiels (ou le nombre de sources) peut être finie ou infinie. Un modèle avec la population finie est analytiquement plus compliqué parce que le nombre de clients déjà dans le système à n'importe quel instant affecte le nombre de clients potentiels composant la population. Le processus (flux) des arrivées peut être régulier ou aléatoire. Dans le premier cas, les arrivées des clients se suivent à des intervalles de temps déterminés. Dans les systèmes réels, on rencontre rarement de processus

des arrivées de ce genre. Pour les systèmes plus typiques, le flux de demandes est aléatoire (la durée de temps entre deux arrivées successives suit une loi de probabilité).

Le mécanisme de service comprend le nombre de serveurs et la distribution des durées de service.



**Figure 2 : Système de file d'attente simple**

Pour la classification des systèmes de files d'attente, on a recours à une notation symbolique (notation de Kendall) comprenant six symboles rangés dans l'ordre :

$$A/B/m/K/N/Z$$

où

**A** : Processus d'arrivée des clients.

Les symboles utilisés sont :

- **M** : (loi exponentielle) Inter-arrivées des clients sont identiquement distribuées selon une loi exponentielle. Il correspond à un processus de Poisson ponctuel.
- **D** : (loi déterministe) Les temps inter-arrivées des clients ou les temps de service sont constants et toujours les mêmes.
- **GI** : (général indépendant) Interarrivées de clients ont une distribution générale et peuvent être dépendantes .
- **G** : (loi générale) Inter-arrivées des clients ont une distribution générale (il n'y a aucune hypothèse sur la distribution mais les inter-arrivées sont indépendantes et identiquement distribuées).
- **$E_k$**  : (loi d'Erlang d'ordre k) Ce symbole désigne un processus où les intervalles de temps entre deux arrivées successives sont des variables aléatoires indépendantes et

identiquement distribuées suivant une loi d'Erlang d'ordre  $k$ .

**B** : Processus de service.

**m** : nombre de serveurs.

**K** : capacité du système  $K = (m + \text{nombre de positions d'attente})$ .

**N** : nombre de clients utilisant le système.

**Z** : discipline du système qui décrit la façon dont les clients sont ordonnancés.

Le nombre de serveurs peut varier de 1 à l'infini noté  $(\infty)$ , de même pour  $K$  et  $N$ . En ce qui concerne  $Z$ , les ordonnancements les plus utilisés sont :

- **FIFO (first in, first out)** où **FCFS (first come, first served)** : c'est la file standard dans laquelle les clients sont servis dans leur ordre d'arrivée. Notons que les disciplines FIFO et FCFS ne sont pas équivalentes lorsque la file contient plusieurs serveurs.

Dans la première, le premier client arrivé sera le premier à quitter la file alors que la deuxième, il sera le premier à commencer son service. Rien n'empêche alors qu'un client qui commence son service après lui, dans un autre serveur, termine avant lui.

- **LIFO (last in, first out)** ou **LCFS (last come, first served)** : Cela correspond à une pile, dans laquelle le dernier client arrivé (donc posé sur la pile) sera le premier traité (retiré de la pile). A nouveau, les disciplines LIFO et LCFS ne sont équivalentes que pour une file mono serveur.

- **RANDOM or SIRO (Served In Random Order)** : les prochains clients sont servis aléatoirement.

- **Round-Robin (cyclique avec un quantum Q)** : Tous les clients de la file d'attente entrent en service à tour de rôle, effectuant un quantum  $Q$  de leur temps de service et sont replacés dans la file, jusqu'à ce que leur service soit totalement accompli.

- **PS (Processor Sharing)** : C'est le cas limite de la distribution Round-Robin lorsque le quantum de temps  $Q$  tend vers 0. Tous les clients sont servis en même temps, mais avec une vitesse inversement proportionnelle au nombre de clients simultanément présents. Si le

taux du serveur est égal à  $\mu$  et qu'à un instant donné il y a  $n$  clients à la station, tous les clients sont donc servis simultanément avec un taux  $\frac{\mu}{n}$ .

- **PRIOR (Avec priorité, avec préemption ou sans préemption) or PNP (Priorité service)** : Les clients sont servis selon leur priorité. Tous les clients de la plus haute priorité sont servis premiers, puis les clients de priorité inférieure sont servis, et ainsi de suite.

## 2.4 Analyse mathématique d'un système de files d'attente

L'étude mathématique d'un système de files d'attente se fait par l'introduction d'un processus stochastique qui décrit l'évolution temporelle du système. En fonction des quantités qui déterminent la structure du système, on cherche à calculer les probabilités d'état définissant le régime transitoire du processus en question, puis le régime stationnaire. Les probabilités d'état définissant le régime transitoire dépendent de la distribution initiale du processus étudié. Ainsi, le calcul explicite du régime transitoire s'avère pénible, voir impossible, pour la plupart des modèles.

A partir de la distribution stationnaire, on peut calculer les caractéristiques du système, telles que : le temps d'attente d'un client, le temps de séjour d'un client dans le système, le taux d'occupation des dispositifs de service, la durée de la période d'activité, et également les mesures de performance suivantes (Formules de Little) :

- nombre moyen de clients dans le système  $\bar{n}$ .
- nombre moyen de clients dans la file d'attente  $\bar{n}_f$ .
- temps moyen d'attente d'un client  $\bar{W}$ .
- temps moyen de séjour d'un client dans le système  $\bar{W}_s$ .

Ces mesures permettent de juger le comportement opérationnel d'un système d'attente. Elles ne sont pas indépendantes les unes des autres, mais sont liées par les relations suivantes (Formules de Little) :

$$\begin{aligned}\bar{n} &= \lambda \bar{W}_s; \\ \bar{n}_f &= \lambda \bar{W}; \\ \bar{W}_s &= \bar{W} + \frac{1}{\mu}; \\ \bar{n} &= \bar{n}_f + \frac{\lambda}{\mu}.\end{aligned}$$

où  $\lambda$  et  $\frac{1}{\mu}$  sont respectivement le taux d'arrivées des clients et la durée moyenne de service ( $\mu > 0$ ). Une autre mesure importante d'un système de files d'attente, celle qui mesure le degré de saturation du système, est l'intensité du trafic  $\rho$ . Elle est définie par :

$$\rho = \frac{\text{temps moyen de service}}{\text{temps moyen entre deux arrivées successives}}.$$

## 2.5 Système de files d'attente M/M/c

### 2.5.1 Description du modèle

Les clients arrivent vers le système selon un processus de Poisson de taux  $\lambda > 0$ . Le service est assuré par  $c \geq 1$  serveurs montés en parallèle. A l'arrivée d'un client, si l'un des serveurs est libre, le client commence immédiatement son service. Dans le cas contraire (tous les serveurs sont occupés par le service), le client prend sa place dans la file d'attente, commune pour tous les serveurs. La capacité d'attente est illimitée (le nombre de position d'attente et infini). Lorsqu'un serveurs se libéré, le client en tête de la file occupe ce serveur libéré. Par conséquent, la discipline d'attente est **FIFO**. Les temps de service sont exponentiellement distribués de moyenne finie  $\frac{1}{\mu}$ . Les durées entre deux arrivées consécutives et les durées de service sont mutuellement indépendantes.

### 2.5.2 Analyse du modèle

L'état du système à la date  $t$  peut être décrit par le processus stochastique.

$$\{N(t), t \geq 0\}$$

où  $N(t)$  représente le nombre de clients dans le système à l'instant  $t$ . Ce dernier est un processus de naissance et de mort dont les taux de transitions sont [13, 19] :

$$\begin{cases} \lambda_n = \lambda, & n = 0, 1, 2, \dots; \\ \mu_n = \min\{n, c\}\mu, & n = 1, 2, \dots \end{cases}$$

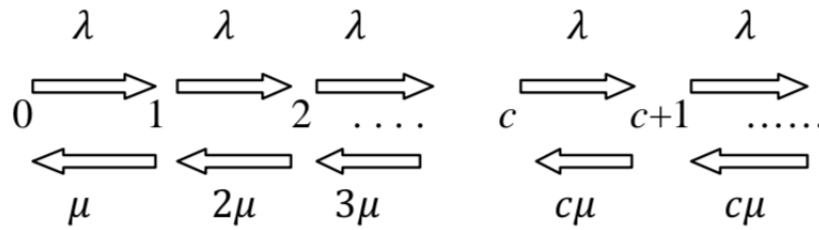


Figure 3 : Evaluation de l'état dans la file d'attente M/M/c.

**Régime transitoire**

Le système d'équations de Kolmogorov pour les probabilités d'état

$p_n(t) = P(N(t) = n)$ ,  $n \geq 0$ , se présente de la manière suivante :

$$p'_0(t) = -\lambda p_0(t) + \mu p_1(t);$$

$$p'_n(t) = \lambda p_{n-1}(t) - (\lambda + n\mu)p_n(t) + (n + 1)\mu p_{n+1}(t), \quad 1 \leq n \leq c;$$

$$p'_n(t) = \lambda p_{n-1}(t) - (\lambda + c\mu)p_n(t) + c\mu p_{n+1}(t), \quad n \geq c.$$

**Régime stationnaire**

Soit  $p_n = \lim_{t \rightarrow \infty} p_n(t)$ ,  $n \geq 0$ . Cette distribution stationnaire satisfait les équations de balance,

$$\lambda p_0 = \mu p_1;$$

$$(\lambda + n\mu)p_n = \lambda p_{n-1} + (n + 1)\mu p_{n+1}, \quad 1 \leq n < c;$$

$$(\lambda + c\mu)p_n = \lambda p_{n-1} + c\mu p_{n+1}, \quad n \geq c;$$

avec

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

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

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

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

où

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

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

A partir de la distribution stationnaire du processus  $\{N(t), t \geq 0\}$ , on peut calculer les caractéristiques du système. En effet,

- Nombre moyen de clients dans le système

$$\bar{n} = \rho + \frac{\rho^{c+1}}{c \cdot c! (1-A)^2} p_0;$$

- Nombre moyen de clients dans la file d'attente

$$\bar{n}_f = \frac{\rho^{c+1}}{c \cdot c! (1-A)^2} p_0;$$

- Temps moyen d'attente d'un client

$$\bar{W} = \frac{c\mu\rho^c}{c!(c\mu - \lambda)^2} p_0;$$

- Temps moyen de séjour d'un client dans la file d'attente

$$\bar{W}_s = \frac{1}{\mu} + \frac{\rho^c}{\mu \cdot c \cdot c! (1-A)^2} p_0.$$

## 2.6 Cas particuliers du système M/M/c

### 2.6.1 Système M/M/1

Le modèle  $M/M/1$  permet d'illustrer les concepts fondamentaux liés à l'attente devant un serveur. Il est décrit par : les clients arrivent dans le système selon un processus de Poisson de taux  $\lambda > 0$ , les durées de service suivent une loi exponentielle de paramètre  $\mu > 0$ , la discipline d'attente est FIFO, la file d'attente est de capacité infinie. Dans ce cas le processus  $\{N(t), t \geq 0\}$  est de naissance et de mort dont les taux de transition sont :

$$\begin{cases} \lambda_n = \lambda, & n \geq 0; \\ \mu_n = \mu, & n \geq 1. \end{cases}$$

Pour trouver la distribution stationnaire et les mesures de performance de ce système, on remplace  $c$  par 1 dans les formules du système  $M/M/c$ ; on obtient :

La distribution stationnaire :  $p_n = (1 - \rho)\rho^n$ .

Les mesures de performance :

- Nombre moyen de clients dans le système :

$$\bar{n} = \frac{\rho}{1 - \rho};$$

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

$$\bar{n}_f = \frac{\lambda^2}{\mu(\mu - \lambda)};$$

- Temps moyen d'attente d'un client :

$$\bar{W} = \frac{\bar{n}_f}{\lambda};$$

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

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

### 2.6.2 Système M/M/ $\infty$

Dans le cas où  $c = \infty$ , nous obtenons le système de files d'attente M/M/ $\infty$ . Il est évident qu'aucune file d'attente ne se forme : chaque client est servi dès son arrivée. La distribution stationnaire est donnée par :

$$p_n = \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} e^{-\frac{\lambda}{\mu}}, \quad n \geq 1.$$

Donc, on a les mesures de performance suivantes :

$$\bar{n} = \rho; \\ \bar{W}_s = \frac{1}{\mu} \text{ tandis que } \bar{n}_f = 0 \text{ et } \bar{W} = 0.$$

## 2.7 Système de files d'attente M/G/1

Pour décrire l'état d'un système de type M/G/1 à la date  $t$ , il faut connaître non seulement le nombre de clients  $N(t)$  qui se trouvent dans le système à la date  $t$ , mais également le temps de service déjà écoulé  $R(t)$  du client qui est en train d'être servi. On peut alors montrer que le processus bidimensionnel  $\{N(t), R(t), t \geq 0\}$  est à nouveau de type markovien. Cependant,

le calcul de son régime transitoire ferait intervenir des équations aux dérivées partielles. Par conséquent, on choisit une autre méthode qui ramène l'étude du processus non markovien  $\{N(t), t \geq 0\}$  à celle d'une chaîne de Markov à temps discret associée dont elle permet de calculer le régime stationnaire [12].

### 2.7.1 Description du modèle

Les clients arrivent dans le système selon un processus de Poisson ( $\lambda > 0$ ). Le service est assuré par un seul serveur. Les durées de service sont des variables aléatoires  $Se$  positives mutuellement indépendantes et distribuées selon une loi générale de fonction de répartition  $B(x)$ , de moyenne finie  $E[Se] = \frac{1}{\mu}$  et de  $E[Se^2]$ . Les durées entre deux arrivées consécutives et les durées de service sont également mutuellement indépendantes.

### 2.7.2 Analyse du modèle

Soit  $\{N(t), t \geq 0\}$ . Montrons que  $\{N(t)\}_{t \geq 0}$  ne définit pas une chaîne de Markov. Soient  $t_d$  et  $t_f$  les dates de début et de fin d'un service,  $t_a$  l'instant d'arrivée d'un nouveau client. Si  $t_d < t_a < t_f$ , la probabilité qu'un départ s'effectue dans l'intervalle  $]t_a, t_a + \Delta t]$  ne dépend pas seulement de  $\Delta t$ , mais de la date  $t_d$  à la quelle le service en cours a commencé. Comme le temps résiduel du service ( $t_f - t_a$ ) dépend du passé, alors le chaîne  $\{N(t)\}_{t \geq 0}$  n'est pas markovienne. Par conséquent, on utilise la méthode de **la chaîne de Markov induite**. A cet effet, on considère  $N(t)$  aux instants  $\xi_1, \xi_2, \dots, \xi_n, \dots$  où les clients terminent leur service et quittent le système. On définit ainsi un processus stochastique à temps discret

$$\{N_n = N(\xi_n), n \geq 1\} \quad (2.1)$$

Pour vérifier que cette suite de variables aléatoires est une chaîne de Markov à temps discret, on considère le nombre  $A_n$  de clients qui entrent dans le système pendant que le n-ème client est servi. Les variables  $A_n$  sont indépendantes entre elles, leur distribution commune est :

$$P(A_n = k) = a_k = \int_0^\infty e^{-\lambda t} \frac{(\lambda t)^k}{k!} dB(t), \text{ où } a_k > 0 \text{ et } k > 0.$$

Alors

$$N_{n+1} = \begin{cases} N_n - 1 + A_{n+1}, & N_n \geq 1; \\ A_n, & N_n = 0; \end{cases}, n \geq 1.$$

L'équation fondamentale de la chaîne vaut donc :

$$N_{n+1} = N_n - \delta_n + A_{n+1}, \quad (2.2)$$

avec

$$\delta_n = \begin{cases} 1, & N_n > 0; \\ 0, & N_n = 0. \end{cases}$$

$N_{n+1}$  ne dépend que de  $N_n$  et de  $A_{n+1}$  et non pas des valeurs prises par  $N_{n-1}, N_{n-2}, \dots$ . La suite  $\{N_n, n \geq 1\}$  est une chaîne de Markov induite du processus  $\{N(t), t \geq 0\}$ . Ses probabilités de transition  $p_{ij} = P(N_{n+1} = j / N_n = i)$  se calcule par

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

La matrice des transitions est

$$M = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & \dots & \dots & \dots \\ a_0 & a_1 & a_2 & a_3 & \dots & \dots & \dots \\ 0 & a_0 & a_1 & a_2 & \dots & \dots & \dots \\ 0 & 0 & a_0 & a_1 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

Vu qu'on peut passer de chaque état vers n'importe quel autre état, il s'agit d'une chaîne de Markov irréductible. De plus, la matrice n'est pas décomposable (est apériodique). La chaîne de Markov est donc ergodique. La distribution stationnaire de la chaîne existe si  $\rho = \frac{\lambda}{\mu} < 1$ .

Pour les variables aléatoires  $A_n$ , nous disposons de quelques résultats importants :

$$E[A_n] = \lambda E[Se] = \frac{\lambda}{\mu} = \rho.$$

La fonction génératrice

$$\begin{aligned} A(z) &= \sum_{k=0}^{\infty} a_k z^k = \sum_{k=0}^{\infty} z^k \int_0^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} dB(t) \\ &= \int_0^{\infty} e^{-\lambda t} \left( \sum_{k=0}^{\infty} \frac{(\lambda t z)^k}{k!} \right) dB(t) \end{aligned}$$

$$= \int_0^\infty e^{-\lambda t} e^{-\lambda t z} dB(t) = \int_0^\infty e^{-(\lambda - \lambda z)t} dB(t).$$

Soit  $\tilde{B}(s) = \int_0^\infty e^{-st} dB(t)$ . Alors  $A(z) = \tilde{B}(\lambda - \lambda z)$ . Encore, la série  $A(z)$  converge pour  $|z| \leq 1$  :

1.  $|z| < 1$ ,  $0 < a_k < 1 \quad \forall k$ , on a  $|a_k z^k| < |z^k|$ ;
2.  $|z| = 1$ ,  $A(1) = 1$ .

**Remarque 2.7.1** *Théorème des probabilités totales :*

- cas discret  $P(A) = \sum_k P(A/Y = y_k) P(Y = y_k)$ ;
- cas continu  $P(A) = \int P(A/Y = y) g(y) dy$ .

**Remarque 2.7.2** *Probabilité que le nombre d'événements  $N$  qui ont lieu pendant un intervalle  $U = u$  dont la densité de probabilité  $f(u)$  est connue, est égal à  $n$  :*

$$P(N = n / U = u) = e^{-\lambda u} \frac{(\lambda u)^n}{n!}.$$

D'où

$$P(N = n) = \int_0^\infty P(N = n / U = u) f(u) du = \frac{1}{n!} \int_0^\infty e^{-\lambda u} (\lambda u)^n f(u) du.$$

$$E[N] = \lambda E[U]; \quad \text{Var}[N] = \lambda^2 \text{Var}[U] + \lambda E[U].$$

Supposons que  $\rho < 1$ . Le système se trouve dans un régime stationnaire. Soit  $\Pi = [\pi_0, \pi_1, \dots]$  la distribution stationnaire de la chaîne de Markov induite ( $\pi_j = \lim_{n \rightarrow \infty} P(N(\xi_n) = j)$ ). Par conséquent,

$$\Pi = \Pi.M, \text{ où } \pi_j = \sum_{i=0}^\infty \pi_i p_{ij}, j \geq 0;$$

$$\pi_j = a_j \pi_0 + \sum_{i=1}^{j+1} a_{j-i+1} \pi_i = a_j \pi_0 + \sum_{i=0}^{j+1} a_{j-i+1} \pi_i - a_{j+1} \pi_0, j \geq 0.$$

A présent, on applique la méthode des fonctions génératrices. En effet,

$$\sum_{j=0}^\infty \pi_j z^j = \pi_0 \sum_{j=0}^\infty a_j z^j + \frac{1}{z} \sum_{j=0}^\infty c_{j+1} z^{j+1} - \frac{\pi_0}{z} \sum_{j=0}^\infty a_{j+1} z^{j+1};$$

où

$$c_{j+1} = \sum_{i=0}^{j+1} a_{j-i+1} \pi_i.$$

En introduisant les fonctions génératrices suivantes :

$$\Pi(z) = \sum_{i=0}^{\infty} \pi_i z^i; A(z) = \sum_{i=0}^{\infty} a_i z^i; C(z) = \sum_{j=0}^{\infty} c_j z^j = \Pi(z)A(z).$$

Finalement, on obtient :

$$\Pi(z) = \pi_0 A(z) + \frac{1}{z} [C(z) - c_0] - \frac{\pi_0}{z} [A(z) - a_0],$$

d'où

$$\Pi(z) = \frac{\pi_0 A(z)(z-1)}{z-A(z)} \text{ Pour } |z| < 1 \text{ et } |z| \neq 0.$$

On a que  $\Pi(1) = 1$ . Cependant,

$$\Pi(1) = \lim_{z \rightarrow 1} \Pi(z) = 0/0.$$

En appliquant la règle de l'Hôpital, on obtient

$$\frac{\pi_0}{1-A'(1)} = 1.$$

Alors,

$$\pi_0 = 1 - A'(1) = 1 - \lambda E[Se] = 1 - \rho.$$

Le résultat final est la première équation de Pollaczek-Khintchine pour le nombre de clients dans le système :

$$\Pi(z) = \frac{(1-\rho)A(z)(z-1)}{z-A(z)} = \frac{(1-\rho)\tilde{B}(\lambda-\lambda z)(z-1)}{z-\tilde{B}(\lambda-\lambda z)}. \quad (2.3)$$

La condition d'existence d'un régime stationnaire est  $\rho = \frac{\lambda}{\mu} < 1$ .

**Remarque 2.7.3** La probabilité  $\pi_0$  peut être trouvée d'une autre manière. De l'équation fondamentale de la chaîne de Markov induite, on a

$$E[N_{n+1}] = E[N_n] - E[\delta_n] + E[A_{n+1}].$$

Vu que

$$E[N_{n+1}] = E[N_n], E[A_{n+1}] = E[\delta_n] = P(\delta_n > 0) = 1 - P(\delta_n = 0);$$

d'où

$$\pi_0 = 1 - \rho.$$

Considérons les probabilités suivantes :

$$p_j = \lim_{t \rightarrow \infty} P(N(t) = j), \quad j \geq 0;$$

$$\pi_j = \lim_{n \rightarrow \infty} P(N(\xi_n) = j), \quad j \geq 0;$$

$$r_j = \lim_{n \rightarrow \infty} P(N(\zeta_n) = j), \quad j \geq 0, \quad \zeta_n \text{ est l'instant d'arrivée de } n\text{-ème client.}$$

Vu que le processus des arrivées est poissonien, et le nombre  $N(t)$  subit des changements discontinus de taille 1 ( $\pm 1$ ), on obtient  $p_j = r_j = \pi_j$ . Comme suite logique, la distribution stationnaire du processus à temps continu  $\{N(t), t \geq 0\}$  est identique à celle de la chaîne de Markov induite. Par conséquent  $\mathcal{Q}(z) = \sum_{j=0}^{\infty} p_j z^j = \Pi(z)$ .

### Caractéristiques de performance

Formule de Pollaczek-Khintchine pour le nombre moyen de clients dans le système :  
Considérons l'équation fondamentale(2.2). Vu que  $\delta_n^2 = \delta_n$  et  $\delta_n N_n = N_n$ , on trouve,

$$N_{n+1}^2 = N_n^2 + \delta_n + A_{n+1}^2 - 2N_n - 2\delta_n A_{n+1} + 2N_n A_{n+1}.$$

On a que :

$$A_{n+1} \text{ est indépendante de } N_n \text{ et de } \delta_n; \quad E[N_{n+1}^2] = E[N_n^2]; \quad E[A_n] = \rho = \frac{\lambda}{\mu}.$$

Alors,

$$E[N_{n+1}^2] = E[N_n^2] + E[\delta_n] + E[A_{n+1}^2] - 2E[N_n] + 2E[A_{n+1}]E[N_n - \delta_n].$$

Ou bien,

$$0 = \rho + E[A_{n+1}^2] - 2E[N_n] + 2\rho E[N_n - \rho].$$

D'où,

$$E[N_n] = \frac{\rho + E[A_{n+1}^2] - 2\rho^2}{2(1 - \rho)}. \quad (2.4)$$

Pour trouver  $E[A_{n+1}^2]$ , considérons la régime stationnaire.

$$\begin{aligned}
\lim_{n \rightarrow \infty} E[A_{n+1}^2] &= E[A^2] = \int_0^\infty E[A^2 / T = t] dB(t) \\
&= \lambda \int_0^\infty t dB(t) + \lambda^2 \int_0^\infty t^2 dB(t) \\
&= \frac{\lambda}{\mu} + \lambda^2 \left( \text{Var}[Se] + \left(1/\mu\right)^2 \right).
\end{aligned}$$

Enfin, la formule(2.4) devient,

$$\lim_{n \rightarrow \infty} E[N_n] = E[N] = \rho + \frac{\rho^2 + \lambda^2 \text{Var}[Se]}{2(1 - \rho)}. \quad (2.5)$$

Le nombre moyen de client dans le système peut être également trouvé à partir de la fonction génératrice  $\Pi(z)$  :  $E[N] = \bar{n} = \lim_{z \rightarrow 1} \Pi'(z)$ . Ici, le calcul de la limite donne une indétermination. Par conséquent, il est nécessaire d'appliquer la règle de l'Hôpital deux fois.

### 2.7.3 Cas particuliers du modèle M/G/1

• **Modèle  $M/E_k/1$**  : Dans ce système, la durée de service suit une loi d'Erlang d'ordre  $k$  et de moyenne finie  $\frac{1}{\mu}$ . Les fonctions de densité de probabilités et de répartition sont données par :

$$b(t) = \frac{k\mu(k\mu t)^{k-1}}{(k-1)!} e^{-\lambda\mu t},$$

et

$$B(t) = 1 - e^{k\mu t} \sum_{j=0}^{k-1} \frac{(k\mu t)^j}{j}, t \geq 0.$$

On démontre que  $A(z) = \tilde{B}(\lambda - \lambda z) = \left[1 + \frac{\rho(1-z)}{k}\right]^{-k}$ .

Alors l'équation (2.3) devient,

$$\Pi(z) = \frac{(1-\rho)(z-1)}{z \left[1 + \frac{\rho(1-z)}{k}\right]^k} - 1. \quad (2.6)$$

• **Modèle  $M/H_2/1$**  : La durée de service suit une loi hyperexponentielle d'ordre 2 dont la fonction de répartition est donnée par

$$B(t) = 1 - P_1 e^{-\mu_1 t} - P_2 e^{-\mu_2 t},$$

où  $P_1, P_2, \mu_1, \mu_2$  vérifient  $P_1 + P_2 = 1$  et  $\frac{1}{\mu_1} + \frac{1}{\mu_2} = \frac{1}{\mu}$  (ici,  $\frac{1}{\mu}$  est la durée moyenne de service).

Pour une telle distribution, on a,

$$A(z) = \tilde{B}(\lambda - \lambda z) = \frac{P_1}{1 + \rho_1(1 - z)} + \frac{P_2}{1 + \rho_2(1 - z)}.$$

Par conséquent, l'équation (2.3) devient

$$\Pi(z) = \frac{(1 - \rho)[1 + (\rho_1 + \rho_2 - \rho)(1 - z)]}{\rho_1 \rho_2 z^2 - (\rho_1 + \rho_2 + \rho_1 \rho_2)z + 1\rho_1 + \rho_2 - \rho}, \quad (2.7)$$

où  $\rho_i = \frac{\lambda}{\mu_i}$ ,  $i = 1, 2$  et  $\rho = \frac{\lambda}{\mu}$ .

## 2.8 Conclusion

Dans ce chapitre, nous avons présenté les systèmes de files d'attente classiques les plus connues. Il faut noter que ces derniers sont utilisés pour modéliser et évaluer les performances des différents systèmes réels (informatiques, de téléphone, ...).

Les modèles d'attente développés ces dernières décennies tentent de prendre en considération des phénomènes de répétition de demandes de service. Il s'agit donc des systèmes de files d'attente avec rappels.

---

## Système de files d'attentes avec rappels

### 3.1 Introduction

Un système de files d'attente où un client arrivant dans le système et trouvant tous les serveurs et, éventuellement, positions d'attente occupés tente de nouveau son service après une durée de temps, est appelé système de files d'attente avec rappels. Son étude est motivée par diverses application pratiques dans le domaine des télécommunications.

Pour identifier un système de files d'attente avec rappels, on a besoin des spécifications suivantes : la nature stochastique du processus des arrivées, la distribution du temps de service, le nombre de serveurs qui composent l'espace de service, la capacité et discipline d'attente ainsi que la spécification concernant le processus de répétition d'appels.

En général, un système de files d'attente avec rappels contient un espace de service composé de  $c \geq 1$  dispositifs de service et un espace d'attente de  $m - c$  ( $m \geq c$ ) positions d'attente. Les clients arrivent dans le système selon un processus aléatoire avec une loi de probabilité donnée, et forment un « flux de clients primaires ». A l'arrivée d'un client primaire, s'il y a un ou plusieurs serveurs libres, le client sera immédiatement pris en charge. Sinon, s'il y a une position d'attente libres, le client va rejoint la file d'attente selon la discipline de service adaptée dans le système. D'autre part, si un client arrive et trouve tous les serveurs et positions d'attentes occupés, il quitte le système définitivement avec une probabilité  $1 - H_0$  ou fait des tentatives de service après une durée de temps aléatoire avec une probabilité  $H_0$ . Entre les tentatives, le client est en « orbite » et devient « source d'appels secondaires »(rappels). La capacité de l'orbite  $O$  peut être finie ou infinie. Dans le cas où  $O$  est finie et si l'orbite est pleine, le client quitte le système pour toujours. Lorsqu'un client secondaire est rappelé de l'orbite, il est traité de la même manière qu'un client primaire avec une probabilité  $H_k$  (s'il s'agit de la  $k^{\text{ème}}$  tentative échouée) ou bien entre en orbite avec une

probabilité  $H_k$  si l'orbite n'est pas plein.

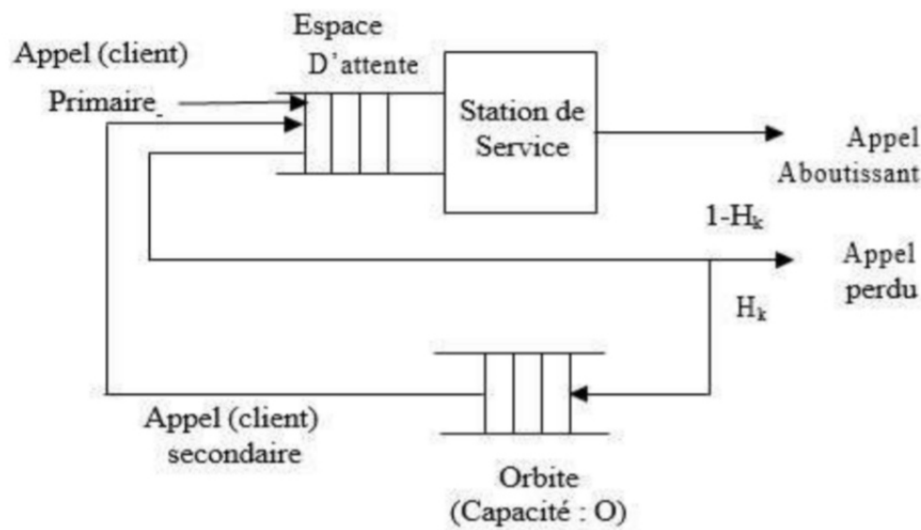


Figure 4 : Système de file d'attente avec rappel

### Remarque 3.1.1

- Le modèle de files d'attente avec rappels décrit 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,...

### Remarque 3.1.2

- Les clients primaires ou secondaires qui arrivent durant un temps de service, entrent en orbite sans aucune influence sur le processus de service.

## 3.2 Terminologie et notation

En utilisant la notation de Kendall établie pour la file d'attente classique, celle du modèle de file d'attente avec rappels s'écrit en ajoutant deux autres symboles aux six symboles précédemment définis :

$$A/B/m/K/N/Z/O/H$$

Avec :

1. **O** : Capacité de l'orbite.

2. **H** : La fonction de persistance qui permet de définir le comportement du client devant une situation de blocage (serveurs occupés).

### Remarque 3.2.1

• *O* peut être supprimé si elle est infinie. *H* peut être également supprimé dans le cas d'un système sans perte ( $c$  à  $d$   $H = 1$ ).

### Remarque 3.2.2

• *Le temps de rappel est défini comme l'intervalle du temps entre deux rappels consécutifs du même client secondaire.*

• *La distribution du temps inter-rappels est supposée généralement exponentielle de taux  $\theta$ .*

• *Lorsque  $\theta \rightarrow \infty$  le système d'attente avec rappels se rapporte à un système classique.*

• *Lorsque  $\theta \rightarrow 0$  le système d'attente avec rappels est un système d'Erlang avec perte.*

## 3.3 Politiques d'accès au serveur à partir de l'orbit

La définition du protocole de rappels est en effet un sujet de controverses, et concerne l'aspect de modélisation du système sous étude.

1. **Politique de rappels classiques** : Le protocole le plus décrit dans la théorie des files d'attente avec rappels est la politique de rappels classiques dans laquelle chaque source dans l'orbite rappelle après un temps exponentiellement distribué avec un paramètre  $\theta$ . Donc il y a une probabilité  $n\theta dt + o(dt)$  d'un nouveau rappel dans le prochain intervalle  $(t; t + dt)$ , sachant que  $n$  clients sont en orbite à l'instant  $t$ . Une telle politique a été motivée naturellement par des applications dans la modélisation du comportement des abonnés dans les réseaux téléphoniques depuis les années 1940.
2. **Politique de rappels constants** : Durant les dernières années, les technologies ont considérablement évolué. La littérature des files d'attente avec rappels décrit différents protocoles de rappels spécifiques à certains réseaux informatiques et de communications modernes. Dans les protocoles en question, le temps inter rappelle peut-être contrôler par un dispositif électronique et par conséquent, est indépendant du nombre d'unités demandant le service. Ici, la probabilité d'un rappel durant  $(t; t + dt)$ , sachant que l'orbite n'est pas vide, est  $\beta dt + o(dt)$ . Ce type de discipline de rappels est appelé politique de rappels constants. Le premier travail dans cette

direction, considère. une file d'attente  $M/M/1$ , où uniquement le client en tête de la file d'attente en orbite peut demander un service après un temps de rappels exponentiellement distribué avec un taux constant. Cette sorte de politique de contrôle de rappels est bien connue pour protocole ALOHA dans le système de communication.

3. **Politique de rappels linéaires** : Artalejo et Gomez-Corral traitent les deux cas d'une manière unifiée en définissant une politique de rappels linéaires pour laquelle la probabilité d'un rappel durant  $(t; t + dt)$ , sachant que  $n$  clients sont en orbite à l'instant  $t$ , est  $(\beta(1 - \delta_{0n}) + n\theta) + o(dt)$ . On mentionne aussi l'existence d'une autre politique dite politique de rappels quadratique.

### 3.4 Quelques exemples modélisés par des systèmes de files d'attente avec rappels

1. **Problème de réservation** : C'est le plus simple exemple, soit un client qui veut prendre une réservation par téléphone dans un restaurant, il y a une seule ligne qui est employée à répondre aux réservations des clients. Ainsi, si un client appelle et trouve la ligne téléphonique occupée il rappellera après une certaine période de temps aléatoire avec la probabilité  $H_k$  ( $H_k < 1$ ) car le client ne peut rappeler indéfiniment. Cet exemple peut être modélisé par une file d'attente  $M/G/1$  avec rappel.
2. **Réseaux de communication par paquet** : Considérons un réseau de communications d'ordinateurs dans lequel on trouve un ensemble d'interfaces « Interface Message Processors : IMP » (C'est le nœud de commutation de paquets utilisé pour connecter les ordinateurs à l'ARPANET) reliées entre elles par des câbles. Un ordinateur principal est connecté à l'une de ces interfaces. Si l'ordinateur envoie un message à un autre ordinateur il doit en premier lieu envoyer le message avec l'adresse de destination à l'interface à laquelle il est connecté. L'interface à son tour envoie le message à l'ordinateur destinataire directement si elle  $y$  est connectée, ou indirectement via d'autres interfaces. Considérons une interface à laquelle un ordinateur principal est connecté. Les messages arrivent de l'extérieur selon un processus aléatoire. Après la réception du message, l'ordinateur l'envoie immédiatement à son interface. S'il  $y$  a un tampon libre, le message est accepté. Dans le cas contraire, le message est rejeté et l'ordinateur doit réessayer une autre fois après une période de

temps. S'il existe des tampons libres, le message rejeté sera stocké dans un tampon de l'ordinateur principal. Dans le cas contraire, le message est considéré comme perdu. Ce problème peut être modéliser comme un système de file d'attente avec rappels à serveur unique (interface IMP) possédant des tampons (positions d'attente). Le nombre de tampons de l'ordinateur principal constitue la capacité de l'orbite.

## 3.5 Modèle M/M/c avec rappels

### 3.5.1 Description du modèle

Nous considérons un système de files d'attente avec rappels où l'espace de service comprend  $c > 1$  serveurs. Les clients primaires arrivent selon un processus de Poisson de taux  $\lambda > 0$ . Si un client primaire arrivant trouve au moins un serveur libre, il commence son service. Sinon, il entre en orbite. Nous admettons que la durée de service et la durée entre deux rappels consécutives sont exponentiellement distribuées de moyenne finies, respectivement  $\frac{1}{\mu}$  et  $\frac{1}{\theta}$ .

Ce modèle a été étudié pour la première fois par Wilkinson (1956) [18]. L'état du système est décrit par le processus stochastique de Markov

$$\{c(t), N_\theta(t), t \geq 0\}$$

où  $c(t)$  représente le nombre de client dans l'espace de service, et  $N_\theta(t)$  est le nombre de client en orbite à l'instant  $t$ . Son espace d'états est donc  $S = \{0, 1, \dots, c\} \times N$ . Les probabilités d'état sont :

$$P_{ij}(t) = P(c(t) = i, N_\theta(t) = j), \quad (i, j) \in s.$$

Les taux de transition en régime stationnaire sont données par

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

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

2. Pour  $i = c$

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

Les caractéristiques importantes de la qualité de service sont :

- probabilité que tous les serveurs sont occupés  $P_c = \lim_{t \rightarrow \infty} P(c(t) = c)$ ;
- nombre moyen de clients en orbite  $\bar{N}_0 = \lim_{t \rightarrow \infty} E[N_0(t)]$ ;
- nombre moyen de serveurs occupés  $\bar{c} = \lim_{t \rightarrow \infty} E[c(t)]$ .

De [7], la condition suffisante et nécessaire d'existence d'un régime stationnaire du système est  $\lambda < c\mu$ . La distribution stationnaire  $P_{ij} = \lim_{t \rightarrow \infty} P_{ij}(t)$  satisfait le système d'équations de Kolmogorov suivant :

$$(\lambda + i\mu + j\theta)p_{ij} = \lambda p_{i-1,j} + (j+1)\theta p_{i-1,j+1} + (i+1)\mu p_{i+1,j}, \quad (3.1)$$

*si*  $0 \leq i \leq c-1$ , *et*  $j \geq 0$ ;

$$(\lambda + c\mu)p_{cj} = \lambda p_{c-1,j} + (j+1)\theta p_{c-1,j+1} + p_{c,j-1}, \quad (3.2)$$

*si*  $i = c$  *et*  $j \geq 0$ .

Pour les fonctions génératrices

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

Ces équations deviennent

$$(\lambda + i\mu)p_i(z) + \theta z p'_i(z) = \lambda p_{i-1}(z) + \theta p'_{i-1}(z) + (i+1)\mu p_{i+1}(z), \quad (3.3)$$

*si*  $0 \leq i \leq c-1$ ,

$$(\lambda + c\mu)p_c(z) = \lambda p_{c-1}(z) + \theta p'_{c-1}(z) + \lambda z p_c(z), \quad (3.4)$$

*si*  $i = c$ .

Maintenant, introduisons la fonction génératrice bivariée

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

Nous supposons que le taux de service  $\mu = 1$ , donc les équations (3.3), (3.4) deviennent :

$$\begin{aligned} \lambda(1-x)p(x, z) + \theta(z-x)p'_z(x, z) + (x-1)p'_x(x, z) \\ + \lambda x^c(x-z)p_c(z) + \theta x^c(x-z)p'_c(z) = 0. \end{aligned} \quad (3.5)$$

En différenciant l'équation (3.5) par rapport à  $z, x, xx, zz$  au point  $x = 1, z = 1$ , on obtient les équations suivantes :

$$\begin{aligned} \theta \bar{N}_0 - \lambda P_c - \theta \bar{N}_c &= 0, \\ \lambda + \theta \bar{N}_0 - \bar{c} - \lambda P_c - \theta \bar{N}_c &= 0, \\ \lambda \bar{c} + \theta p''_{xz} - p''_{xx} - \lambda c P_c - \theta c \bar{N}_c &= 0, \\ -\lambda \bar{N}_0 - \theta p''_{zz} + (1 + \theta) p''_{xz} + \lambda - \theta c \bar{N}_c - \lambda c P_c + \theta p''_{czz} &= 0, \\ \theta p''_{zz} - \lambda \bar{N}_c - \theta p''_{czz} &= 0, \end{aligned} \quad (3.6)$$

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

$$\bar{c} = \lambda. \quad (3.7)$$

$$\bar{N}_0 = \frac{1 + \theta}{\theta} \cdot \frac{\lambda - \lim_{t \rightarrow \infty} Var[c(t)]}{c - \lambda}. \quad (3.8)$$

L'équation (3.8) peut être réécrite sous une forme équivalente telle que :

$$\bar{N}_0 = \frac{1 + \theta}{\theta} \cdot \frac{\lambda + \lambda^2 - \lim_{t \rightarrow \infty} E[(c(t))^2]}{c - \lambda}.$$

## 3.6 Modèle d'attente M/M/1 avec rappels

### 3.6.1 Description du modèle

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

que entre deux arrivées primaires successives sont mutuellement indépendantes. L'état du système peut être décrit par le processus :

$$\{c(t), N_o(t), t \geq 0\}, \tag{3.9}$$

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

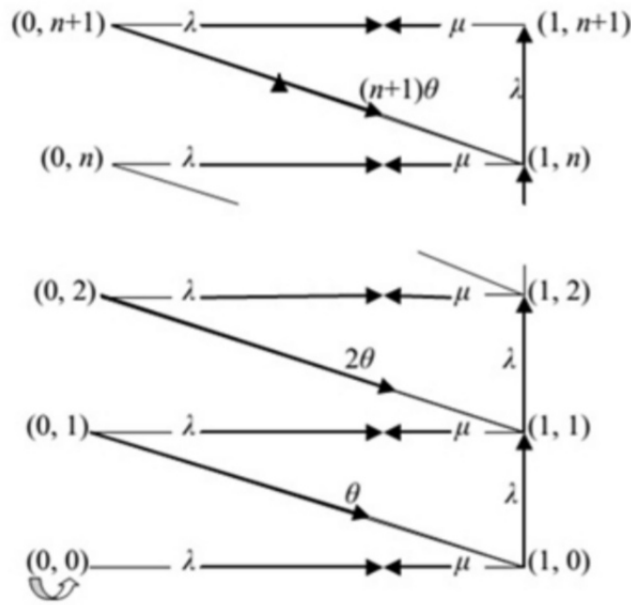


Figure 5 : La structure générale du modèle M/M/1 avec rappels.

Les équations d'équilibre statistique sont :

$$(\lambda + j\theta)p_{0j} = \mu p_{0j}, \tag{3.10}$$

$$(\lambda + \mu)p_{1j} = \lambda p_{0j} + (j + 1)\theta p_{0,j+1} + \lambda p_{0,j-1}. \tag{3.11}$$

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

$$p_0(z) = \sum_{j=0}^{\infty} Z^j p_{0j},$$

$$p_1(z) = \sum_{j=0}^{\infty} Z^j p_{1j}.$$

A l'aide de ses fonctions et à partir des équations (3.10) et (3.11), on obtient :

$$p_0(z) = (1 - \rho) \left( \frac{1 - \rho}{1 - z\rho} \right)^{\frac{\lambda}{\theta}} \quad (3.12)$$

$$p_1(z) = \rho \left( \frac{1 - \rho}{1 - z\rho} \right)^{\frac{\lambda}{\theta} + 1} \quad (3.13)$$

## 3.7 Modèle d'attente M/G/1 avec rappels

### 3.7.1 Description du modèle

Les clients arrivent dans le système selon un processus homogène de Poisson de taux  $\lambda > 0$  :  $P(\tau_n^e \leq x) = 1 - e^{-\lambda x}$ . Le service des clients est assuré par un seul serveur. La durée de service suit une loi générale  $P(\tau_n^e \leq x) = B(x)$ , éventuellement exponentielle, de transformée de Laplace-Stieltjes  $\tilde{B}(s)$ ,  $Re(s) > 0$ . Soient les moments d'ordre  $k$ ,  $\beta_k = (-1)^k \tilde{B}(0)$ , l'intensité du trafic  $\rho = \lambda\beta_1$  et  $\mu = \frac{1}{\beta_1}$ . La durée entre deux rappels successifs d'une même source secondaire est exponentiellement distribuée de paramètre  $\theta > 0$  :  $T(x) = P(\tau_n^e \leq x) = 1 - e^{-\theta x}$ .

Le système évolue de la manière suivante : On suppose que le  $(n + 1)^{ième}$  client termine son service à l'instant  $\xi_{n-1}$  (les clients sont numérotés dans l'ordre de service) et le serveur devient libre. Même s'il y a des clients dans le système, ils ne peuvent pas occuper le dispositif de service immédiatement à cause de leur ignorance de ce dernier. C'est pourquoi le  $n^{ième}$  client suivant n'entre en service qu'après un intervalle de temps  $R_n$  durant lequel le serveur est libre. A l'instant  $\eta_n = \xi_{n-1} + R_n$ , le  $n^{ième}$  client débute le service durant un temps  $\tau_s^n$ . Les clients primaires arrivent dans le système pendant ce temps deviennent sources de clients secondaires. Tous les clients qui arrivent durant ce temps de service n'influent pas sur le processus. A l'instant  $\xi_n = \eta_n + \tau_s^n$ , le  $n^{ième}$  client achève son service, le serveur devient libre et ainsi de suite.

### 3.7.2 Méthode des variables supplémentaires

Le premier résultat sur le système M/G/1 avec rappels a été obtenu par Keilson, basé sur la méthode des variables supplémentaires.

L'état du système peut être décrit par le processus

$$X(t) = \begin{cases} N_0(t) & \text{si } C(t) = 0 \\ \{c(t), N_0(t), \xi(t)\} & \text{si } c(t) = 1 \end{cases},$$

où  $c(t) = 0$  ou  $1$  selon que le serveur est libre ou actif,  $N_0(t)$  représente le nombre de clients en orbite et  $\xi(t)$  est une variable aléatoire supplémentaire à valeurs dans  $\mathbb{R}^+$ , et désignant la durée de service écoulée à la date  $t$ , si  $C(t) = 1$ .

Notons :

$$p_{0n} = \lim_{t \rightarrow \infty} P(c(t) = 0, N_0(t) = n);$$

$$p_{1n}(x) = \lim_{t \rightarrow \infty} \frac{d}{dx} P(c(t) = 1, \xi(t) \leq x, N_0(t) = n).$$

**Théorème 3.7.1** *Les fonctions génératrices de la distribution conjointe de l'état du serveur et de la taille de l'orbite en régime stationnaire ( $\rho = \lambda\beta_1 < 1$ ) sont données par :*

$$P_0(z) = \sum_{n=0}^{\infty} p_{0n} z^n = (1 - \rho) \exp \left\{ \frac{\lambda}{\theta} \int_1^z \frac{1 - \tilde{B}(\lambda - \lambda u)}{\tilde{B}(\lambda - \lambda u) - u} du \right\}; \quad (3.14)$$

$$P_1(z) = \sum_{n=0}^{\infty} p_{1n} z^n = \frac{1 - \tilde{B}(\lambda - \lambda z)}{\tilde{B}(\lambda - \lambda z) - z} P_0(z). \quad (3.15)$$

Les probabilités  $p_{0n}$  et  $p_{1n}(x)$  vérifient le système d'équations de balance :

$$(\lambda + n\theta)p_{0n} = \int_0^{\infty} p_{1n}(x)b(x)dx;$$

$$p'_{1n}(x) = -(\lambda + b(x))p_{1n}(x) + \lambda p_{1,n-1}(x);$$

$$p_{1n}(0) = \lambda p_{0n} + (n + 1)\theta p_{0,n+1}.$$

Où  $b(x) = B'(x)/(1 - B(x))$  est l'intensité instantanée du service étant donné que la durée écoulée est égale à  $x$ .

On introduit les fonctions génératrices, telles que  $P_0(z) = \sum_{n=0}^{\infty} z^n p_{0n}$  et  $p_1(z, x) = \sum_{n=0}^{\infty} z^n p_{1n}(x)$ .

Le système d'équations de balance devient,

$$\begin{cases} \lambda \sum_{n=0}^{\infty} z^n p_{0n} + \theta \sum_{n=0}^{\infty} z^n n p_{0n} = \int_0^{\infty} \sum_{n=0}^{\infty} z^n p_{1n}(x) dx; \\ \sum_{n=0}^{\infty} z^n p'_{1n}(x) = -(\lambda + b(x)) \sum_{n=0}^{\infty} z^n p_{1n}(x) + \lambda \sum_{n=0}^{\infty} z^n p_{1n-1}(x); \\ \sum_{n=0}^{\infty} z^n p_{1n}(0) = \lambda \sum_{n=0}^{\infty} z^n p_{0n} + \theta \sum_{n=0}^{\infty} z^n (n + 1) p_{0n}. \end{cases}$$

D'où

$$\begin{cases} \lambda P_0(z) + \theta z P_0'(z) = \int_0^\infty P_1(z, x) b(x) dx; \\ P_1'(z, x) = (\lambda z - \lambda - b(x)) P_1(z, x); \\ P_1(z, 0) = \lambda P_0(z) + \theta P_0'(z). \end{cases} \quad (3.16)$$

De la deuxième équation de (3.16), on a

$$P_1(z, x) = P_1(z, 0) [1 - B(x)] \exp(-(\lambda - \lambda z)x). \quad (3.17)$$

Donc, la première équation de (3.16) devient

$$\begin{aligned} \lambda P_0(z) + \theta z P_0'(z) &= \int_0^\infty P_1(z, 0) [1 - B(x)] \exp(-(\lambda - \lambda z)x) b(x) dx \\ &= P_1(z, 0) \int_0^\infty \tilde{B}'(x) \exp(-(\lambda - \lambda z)x) dx \\ &= P_1(z, 0) \int_0^\infty \exp(-(\lambda - \lambda z)x) dBx \\ &= P_1(z, 0) \tilde{B}(\lambda - \lambda z) = P_1(z, 0) K(z). \end{aligned} \quad (3.18)$$

A partir des équations (3.16) et (3.18), on a

$$\begin{aligned} P_1(z, 0) K(z) &= \lambda P_0(z) + \theta z \left( \frac{P_1(z, 0)}{\theta} - \frac{\lambda}{\theta} P_0(z) \right); \\ P_1(z, 0) (K(z) - z) &= (\lambda - \lambda z) P_0(z); \\ P_1(z, 0) &= \frac{\lambda - \lambda z}{K(z) - z} P_0(z). \end{aligned}$$

A l'aide de l'équation (3.17), on trouve

$$P_1(z, 0) = \frac{\lambda - \lambda z}{K(z) - z} P_0(z) [1 - B(x)] \exp(-(\lambda - \lambda z)x). \quad (3.19)$$

En intégrant cette équation, et en utilisant la formule suivante :

$$\int_0^\infty \exp(-sx) [1 - B(x)] dx = (1 - \tilde{B}(s))/s.$$

On obtient :

$$P_1(z) = \int_0^\infty P_1(z, x) dx = \frac{\lambda - \lambda z}{K(z) - z} P_0(z) \int_0^\infty [1 - B(x)] \exp(-(\lambda - \lambda z)x) dx.$$

$$= \frac{\lambda - \lambda z}{K(z) - z} P_0(z) \frac{1 - \tilde{B}(\lambda - \lambda z)}{\lambda - \lambda z} = P_0(z) \frac{1 - K(z)}{K(z) - z}.$$

Pour trouver  $P_0(z)$ , on utilise les équations (3.16) et (3.18) :

$$\begin{aligned} \lambda P_0(z) + \theta z P_0'(z) &= K(z)[\lambda P_0(z) + \theta P_0'(z)]; \\ \theta[K(z) - z]P_0'(z) &= \lambda[1 - K(z)]P_0(z). \end{aligned} \quad (3.20)$$

Considérons  $f(z) = K(z) - z$ . On a :

$$f(1) = \tilde{B}(0) - 1 = 1 - 1 = 0; \quad f'(z) = -\lambda \tilde{B}'(\lambda - \lambda z) - 1.$$

Par conséquent,  $f'(1) = -\lambda \tilde{B}'(0) - 1 = \rho - 1 < 0$  et  $f''(z) = \lambda^2 \tilde{B}''(\lambda - \lambda z) \geq 0$ . La fonction  $f(z)$  est décroissante sur  $[0, 1]$  et positive :

pour  $\rho < 1$  et  $z \in [0, 1]$ ,  $z < K(z) < 1$ . En plus,

$$\lim_{z \rightarrow 1} \frac{1 - K(z)}{K(z) - z} = \frac{K'(1)}{1 - K'(1)} = \frac{\rho}{1 - \rho} < \infty.$$

De ce fait, la fonction  $\frac{1 - K(z)}{K(z) - z}$  peut être définie au point  $z = 1$  comme  $\frac{\rho}{1 - \rho}$ . Donc

$$P_0'(z) = \frac{\lambda}{\theta} \frac{1 - K(z)}{K(z) - z} P_0(z).$$

La résolution de cette équation nous donne l'équation (3.14). Encore, à partir de l'équation (3.15) nous obtenons  $P_1(1) = P_0(1) = \frac{\rho}{1 - \rho}$ . De plus, vu que  $P_0(1) + P_1(1) = 1$ , on trouve  $P_1(1) = \rho$  et  $P_0(1) = 1 - \rho$ .

### Conséquences

1. La distribution marginale du nombre de serveurs occupés s'exprime de la manière suivante :

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

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

2. Soit  $P_n = \lim_{t \rightarrow \infty} P(N_o(t) = n)$ . Alors  $P_n = P_{0n} + P_{1n}$ . En terme des fonctions génératrice la dernière équation s'exprime de la manière suivante :

$$\sum_{n=0}^{\infty} z^n P_n = \sum_{n=0}^{\infty} z^n P_{0n} + \sum_{n=0}^{\infty} z^n P_{1n};$$

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

3. Soit  $q_n = \lim_{t \rightarrow \infty} P(c(t) + N_o(t) = n)$ . Alors  $q_n = P_{0n} + P_{1,n-1}$ . Par conséquent,

$$\sum_{n=0}^{\infty} z^n q_n = \sum_{n=0}^{\infty} z^n P_{0n} + \sum_{n=0}^{\infty} z^n P_{1,n-1};$$

$$\mathcal{Q}(z) = P_0(z) + zP_1(z) = \frac{(1-\rho)(1-z)\tilde{B}(\lambda - \lambda z)}{\tilde{B}(\lambda - \lambda z) - z} \exp\left\{ \frac{\lambda}{\theta} \int_1^z \frac{1 - \tilde{B}(\lambda - \lambda u)}{\tilde{B}(\lambda - \lambda u) - u} du \right\}.$$

### 3.7.3 Mesures de performance

Le système M/G/1 peut être caractérisés par les mesures de performance, telles que :

- Nombre moyen de clients dans le système

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

- Nombre moyen de clients en orbite

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

- Temps moyen d'attente d'un client

$$\bar{W} = \frac{\bar{n}_0}{\lambda} = \frac{\lambda \beta_2}{2(1-\rho)} + \frac{\lambda \beta_1}{\theta(1-\rho)}.$$

- Nombre moyen de rappels par client

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

## 3.8 Conclusion

Dans ce chapitre, nous avons présenté un aperçu sur les systèmes de files d'attente avec rappels. Une attention particulière a été accordée aux modèles  $M/M/c$ ,  $M/M/1$  et  $M/G/1$  avec rappels. Une étude poussée des ce genre de systèmes est nécessaire pour améliorer et mieux évaluer les performances des systèmes informatiques, des réseaux de communications et systèmes complexes dans nombreux domaines.

# 4 Exercices

**Exercice 1** *SLOBOVIAN SCIENTIFIC* dispose d'un système interactif à temps partagé composé de 20 terminaux actifs, qui peut être étudié à l'aide du modèle de réparation des machines. Le temps de service moyen de l'unité centrale, y compris la permutation, est de 2 secondes, tandis que le temps de réflexion moyen est de 20 secondes.

1. Trouver  $p_0$ ,  $\rho$ ,  $\lambda$  et temp moyen d'attente  $\bar{W}$ .
2. Notons que  $\lambda$  est le débit moyen en interactions par seconde. Quel serait l'effet de l'ajout de cinq terminaux ?

## Solution 1

1. Pour 20 terminaux

$$p_0 = \left[ \sum_{n=0}^{20} \frac{20!}{(20-n)!} \left(\frac{2}{20}\right)^n \right]^{-1} = 0.001869.$$

Ensuite,

$$\rho = 1 - p_0 = 0.998131, \quad \lambda = 0.49907 \text{ interactions/seconde},$$

et

$$\bar{W} = (20/0.49907) - 20 = 20.075 \text{ seconds}.$$

2. Avec 25 terminaux

$$p_0 = 0.00002927, \quad \rho = 1 - p_0 = 0.99997073,$$

$$\lambda = \rho/E[s] = 0.499985365,$$

et

$$\bar{W} = (25/0.49907) - 20 = 30 \text{ seconds} .$$

**Exercice 2** Une compagnie aérienne envisage d'ouvrir un point de vente dans un nouveau centre commercial. Elle compte y faire travailler un agent qui sera responsable des réservations et de la vente des billets. On prévoit un achalandage de 15 clients à l'heure en moyenne ; on estime aussi que la distribution des arrivées peut être calculée selon la loi de Poisson et que le temps de service sera de 3 minutes en moyenne par client (distribution exponentielle). Déterminez les mesures de performance suivantes :

1. Taux d'utilisation du système.
2. Pourcentage d'inactivité de l'agent.
3. Nombre moyen de clients qui attendent pour être servis.
4. Temps moyen passé par un client dans le système.

**Solution 2**

$\lambda = 15$  clients à l'heure et  $\frac{3 \text{ minutes}}{\text{client}} = \text{temps de service } \frac{1}{\mu}$ .

Donc,  $\mu = \frac{1 \text{ client}}{3 \text{ minutes}} * 60 \text{ minutes par heure} = 20$  clients à l'heure.

1. Taux d'utilisation du système

$$\rho = \frac{\lambda}{c\mu} = \frac{15}{1(20)} = 0.75.$$

2. Pourcentage d'inactivité de l'agent. Inactivité de l'agent

$$1 - \rho = 1 - 0.75 = 0.25 \text{ c'est à dire } 25\% \text{ du temps.}$$

3. Nombre moyen de clients qui attendent pour être servis. Signifie, que l'on souhaite calculer le nombre de clients dans la file

$$\bar{n}_f = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{15^2}{20(20 - 15)} = 2.25 \text{ clients.}$$

4. Temps moyen passé par un client dans le système.

$$\bar{W} = \frac{1}{\mu - \lambda} = \frac{1}{20 - 15} = 0.20 \text{ heure.}$$

**Exercice 3** Des camions arrivent dans une station service pour passer des tests de sécurité, suivant un processus de Poisson de taux de 6 / jour. La durée des tests pour chaque camion est une v.a exponentielle d'espérance mathématique de 1h 30 mn. On suppose que le processus d'arrivée ne s'interrompt pas et que la station travaille 24 heures sur 24.

1. Le système admet-il une distribution stationnaire ?
2. Si oui la calculer et donner le nombre moyen d'utilisateurs dans le système, le temps moyen passé dans le système.

**Solution 3**

1. Nous avons,

$$\lambda = \frac{6}{24} = \frac{1}{4}, \quad \mu = \frac{2}{3}, \quad \text{comme } \rho = \frac{3}{8} < 1 \text{ le système est ergodique.}$$

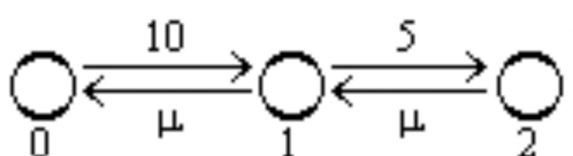
2. D'après les formules du calcul des paramètres des systèmes  $M/M/1$ , nous aurons :

$$p_0 = 1 - \rho = 1 - \frac{3}{8} = \frac{5}{8}, \quad p_k = \frac{5}{8} \left(\frac{3}{8}\right)^k, \quad k \in \mathbb{N}.$$

$$\bar{n} = \frac{\rho}{1 - \rho} = \frac{3}{5}, \quad \bar{W}_s = \frac{\bar{n}}{\lambda} = 2.4h.$$

**Exercice 4** Une entreprise de construction possède deux engins identiques, chacun pouvant tomber en panne indépendamment de l'autre suivant un processus de Poisson de taux 5 fois par mois. On suppose que la durée de réparation est une v.a. qui suit une loi exponentielle de paramètre  $\mu$  (taux de service). Pour quelle valeur de  $\mu$ , les deux engins seront-ils simultanément en état de marche au moins la moitié du temps ?

**Solution 4** Le système est modélisé par un processus de naissance et de mort à 3 états (0, 1 ou 2 machines en panne) de diagramme de transition suivant :



Nous avons :

$$p_1 = \frac{10}{\mu} p_0, \quad p_2 = \frac{50}{\mu} p_0.$$

D'où

$$p_0 = \left[1 + \frac{10}{\mu} + \frac{50}{\mu^2}\right]^{-1} = \frac{\mu^2}{\mu^2 + 10\mu + 50\mu}$$

$p_0$  étant la probabilité stationnaire pour que les deux engins soient en état de marche, on devra avoir :

$$p_0 \geq \frac{1}{2} \Rightarrow \frac{\mu^2}{\mu^2 + 10\mu + 50\mu} \geq \frac{1}{2} \Rightarrow \mu^2 + 10\mu + 50\mu \geq 0.$$

Tenant compte de la positivité de  $\mu$ , on doit avoir

$$\mu \geq 5 + \sqrt{75} = 5(1 + \sqrt{3}) \approx 13.660254.$$

Il faudra donc un taux de service au moins égal à 13,660254 pour que les deux engins soient en état de marche au moins une fois sur deux.

**Exercice 5** Dans une usine de production automobile, on a étudié le problème de la détermination du nombre optimal d'employés à placer aux guichets d'un magasin chargé de fournir l'outillage nécessaire aux ouvriers. L'étude du magasin a commencé par la détermination des caractéristiques statistiques des arrivées des ouvriers (1.6 arrivées/minute) et des temps passés par les employés pour fournir les outillages demandés (0.9 service/minute).

1. Calculer le temps moyen d'attente dans la file  $\bar{W}_s$  pour  $S = 2, S = 3, S = 4$ .
2. Calculer le nombre moyen de clients dans une journée de 8 heures et le temps de service moyen correspondant.

**Solution 5** 1. Connaissant  $\lambda$  et  $\mu$  on a  $\frac{\lambda}{\mu} = \frac{1.6}{0.9} = 1.77 > 1$ .

Puisque  $\frac{\lambda}{\mu} = 1.77 > 1$ , on s'est intéressé uniquement aux valeurs de  $S = 2, S = 3$ , et  $S = 4$ .

Pour calculer le temps moyen d'attente dans la file, on a d'abord cherché  $p_0$  pour chaque valeur respective de  $S$ .

$$S = 2, p_0 = 0.061,$$

$$S = 3, p_0 = 0.152,$$

$$S = 4, p_0 = 0.166,$$

ce qui donne

$$S = 2, \bar{W}_s = 4.00.$$

$$S = 3, \bar{W}_s = 0.31.$$

$$S = 4, \bar{W}_s = 0.06.$$

2. Calculons maintenant le nombre moyen de clients dans une journée de 8 heures

$$\lambda \times 60 \times 8 = 1.6 \times 60 \times 8 = 768,$$

et pour ce nombre d'arrivées, il faudra (avec un temps de service de  $\frac{1}{\mu}$ )

$$\frac{768}{\mu} = \frac{768}{0.9} = 853 \text{ minutes de service par jour.}$$

$$= 14.21 \text{ heures.}$$

**Exercice 6** *Un organisme public est ouvert, chaque jour ouvrable, de 9h à 17h sans interruption. Il accueille, en moyenne, 64 usagers par jour; un guichet unique sert à traiter le dossier de chaque usager, ceci en un temps moyen de 2,5 minutes. Les usagers, si nécessaire, font la queue dans l'ordre de leur arrivée; même si la queue est importante, on ne refuse aucun usager. Une étude statistique a permis de conclure que la durée aléatoire des services suit une loi exponentielle et que le régime des arrivées des usagers forment un processus de Poisson.*

1. Donner la notation de Kendall de cette file et préciser les taux.
2. Quelle est la probabilité qu'il y ait  $k$  clients dans le système à un instant donné ?
3. Quelles sont les probabilités qu'il n'arrive aucun client entre 15H et 16H ? Que 6 clients arrivent entre 16H et 17H ?
4. Quelle est, en moyenne et par heure, la durée pendant laquelle l'employé du guichet ne s'occupe pas des usagers ?

**Solution 6** 1.  $M/M/1/\infty$  avec  $\lambda = 8$  (clients par Heure) et  $\mu = 24$  (clients par Heure).

2. On pose  $\rho = \frac{\lambda}{\mu} = \frac{1}{3}$ .

Comme  $\rho < 1$  donc la distribution stationnaire existe; la probabilité qu'il ait  $k$  clients dans le système à un instant données  $t$ ,

$$\pi_k = (1 - \rho)\rho^k.$$

3. Le temps,  $T$ , passé par un client dans le système (attente + service) suit une loi exponentielle de paramètre  $(\mu - \lambda)$ ,

$$\bar{W} = \frac{1}{\mu - \lambda} = \frac{1}{16} \text{ heures.}$$

Donc le temps moyen d'attente dans la file est

$$\bar{W}_s = \frac{\bar{W} - 1}{\mu} = \frac{\lambda}{\mu(\mu - \lambda)} = \frac{1}{48}.$$

4. Probabilité qu'il n'arrive aucun client entre 15H et 16H. On sait que le nombre  $N$  de clients arrivant dans le système pendant une unité de temps est une variable aléatoire qui suit la loi de Poisson

$$\mathbb{P}(N = K) = \exp(-\lambda) \frac{\lambda^k}{k!}.$$

Pour  $k = 0$  on obtient  $\mathbb{P}(N = 0) = \exp(-8)$ .

Pour  $k = 6$  on obtient

$$\mathbb{P}(N = 6) = \exp(-8) \frac{8^6}{6!}.$$

5. L'employé est inoccupé lorsque le système est dans l'état 0, ce qui correspond à la probabilité  $\pi_0 = (1 - \rho) = \frac{2}{3}$  soit en moyenne 40 minutes par heure.

**Exercice 7** On considère une file d'attente  $M/M/1$  avec rappels de taux  $\lambda = 1$ ,  $\mu = 2$  et  $\theta = 3$ .

• Calculer le nombre moyen de clients dans le système  $\bar{N}$ .

**Solution 7** On a :  $c = 1$ ,  $\lambda = 1$ ,  $\mu = 2$ ,  $\theta = 3$  et  $\rho = \frac{1}{2}$ .

$$\bar{N} = \lim_{t \rightarrow \infty} E[C(t) + N_0(t)] = \varphi'(1),$$

où,

$$\varphi(z) = P_0(Z) + ZP_1(Z).$$

On a aussi

$$P_0(Z) = (1 - \rho) \left( \frac{1 - \rho}{1 - Z\rho} \right)^{\frac{\lambda}{\theta}},$$

$$P_1(Z) = \rho \left( \frac{1 - \rho}{1 - Z\rho} \right)^{\frac{\lambda}{\theta} + 1},$$

d'où

$$\begin{aligned} \varphi(Z) &= (1 - \rho) \left( \frac{1 - \rho}{1 - Z\rho} \right)^{\frac{\lambda}{\theta}} + Z\rho \left( \frac{1 - \rho}{1 - Z\rho} \right)^{\frac{\lambda}{\theta} + 1} \\ &= \left( \frac{1 - \rho}{1 - Z\rho} \right)^{\frac{\lambda}{\theta} + 1} \left( (1 - \rho) \left( \frac{1 - \rho}{1 - Z\rho} \right)^{-1} + Z\rho \right) \\ &= \left( \frac{1 - \rho}{1 - Z\rho} \right)^{\frac{\lambda}{\theta} + 1} (1 - Z\rho + Z\rho) \\ \varphi(Z) &= \left( \frac{1 - \rho}{1 - Z\rho} \right)^{\frac{\lambda}{\theta} + 1}. \end{aligned}$$

Alors

$$\bar{N} = \frac{\rho(\lambda + \theta)}{\theta(1 - \rho)}.$$

D'où on obtient :

$$\bar{N} = \frac{\frac{1}{2}(1+3)}{3(1-\frac{1}{2})} = \frac{4}{3}.$$

**Exercice 8** Un petit magasin possède une seule ligne téléphonique. Les appels vers ce magasin peuvent être modélisés par un système de files d'attente  $M/M/1$  avec rappels, où le taux d'arrivée des appels primaires  $\lambda = 10$  appels/heure, le taux de service  $\gamma = 15$  appels/heure des rappels  $\theta = 6$  tentatives/heure.

1. Sachant que le nombre moyen de clients dans le système est défini par  $\bar{n} = \frac{\rho(\theta + \lambda)}{\theta(1 - \rho)}$ , déterminer :
  - A) le temps moyen qu'un client met pour joindre (par téléphone) un agent de ce magasin.
  - B) le nombre moyen de tentatives fait par un client avant de pouvoir discuter par téléphone avec un agent du magasin.
2. Quelle est la proportion des appels téléphoniques aboutissant au signal "occupé" ?

**Solution 8**

1.A)  $\lambda = 10$  appels/heure ;  $\gamma = 15$  appels/heure.

$$\rho = \frac{\lambda}{\gamma} = \frac{10}{15} = \frac{2}{3}.$$

$$\bar{n} = \frac{\rho(\theta + \lambda)}{\theta(1 - \rho)} = \frac{2/3(6 + 10)}{6(1 - 2/3)} = \frac{16}{3} \text{ appels.}$$

$$\bar{W} = \frac{\bar{n}}{\lambda} - \frac{1}{\gamma} = \frac{(16/3)}{10} - \frac{1}{15} = \frac{7}{15} h.$$

1.B)  $\bar{R} = \theta \bar{W} = 6 \times \frac{7}{15} = \frac{42}{15}.$

2.  $P(\text{occupé}) = \lim_{t \rightarrow \infty} P(C(t) = 1) = P_0(Z = 1) = \rho = \frac{2}{3}.$

## *Conclusion générale*

*Dans ce travail, nous nous sommes intéressés à la théorie des files d'attente classiques et avec rappels.*

*Dans un premier temps, nous avons effectué une étude bibliographique sur les systèmes de files d'attente classiques, nous avons présenter les modèles les plus connus ( $M/M/1$ ,  $M/M/c$  et  $M/G/1$ ).*

*Dans un deuxième temps, nous avons présenté une brève description des systèmes de files d'attente avec rappels en particulier les systèmes à multiserveurs et le système  $M/G/1$ .*

*Ensuite, nous avons présenté quelques exercices sur les systèmes de files d'attente classiques et avec rappels avec leurs corrigés.*

---

## ***Bibliographie***

- [1] A. Attahiru. *Queueing Theory for Telecommunications : Discrete Time Modelling of a Single Node System*, 10.1007/978-1-4419-7314-6, 2010.
- [2] K. Belmadi et A. Diouane. *Analyse d'un système M/M/1 Avec rappels, serveur non fiable et clients impatientes*, 2020.
- [3] R.N. Bhattacharya et E.C. Waymire. *Stochastic Processes with application*, Wiley, N.Y., 1990.
- [4] G. Christos. Cassandras et S. Lafortune. *Introduction to discrete event systems, Second Edition*.
- [5] J.L. Doob. *Stochastic Processes*, Wiley, N.Y., 1953.
- [6] S. Doghmane. *Le comportement asymptotique du système de files d'attente M/G/1 avec rappels*, 8 – 14, 2012.
- [7] G.I. Falin. *Ergodicity and stability of systems with repeated attempts. Ukrainian Mathematical Journal*, 41(7) : 813-817, 1989.
- [8] G.I.Falin et J.G.C. Templeton. *Retrial queues*, 1997.
- [9] W. Feller. *An Introduction to Probability Theory And Its Applications, vol 1*, Wiley, N.Y, 1968.
- [10] E. Geleube et G. Pujolle. *Introduction aux réseaux de files d'attente*, Paris, 1985.
- [11] K.Hafsi. *Étude d'un système de files d'attente avec serveurs hétérogènes, dérobade et abandon des clients*, 24-25, 2018.
- [12] D.P. Heymen. *Optimal operating policies for M/G/1 queueing systems. Oper.Res*, 16 ; 362-382, 1968.
- [13] L. Kleinrock. *Queueing Systems (Theory), Vol 1*, John Wily and sons, Edition, 1975.
- [14] L. Kleinrock. *Queueing Systems (Computer Application), Vol 2*, John Wily and sons, Edition, 1976.

- [15] *M.Labani. Analyse stochastique de file d'attente avec rappel, 36, 39-41, 2016.*
- [16] *M. Lejeune. Statistique, la théorie et ses applications, 2eme Edition. Springer-Verlag France, Paris, 2010.*
- [17] *M. Rachid. Réseaux et Services Mobiles : Conception et développement d'un outil d'aide à l'analyse des indicateurs qualité d'un réseaux GPRS, 2007.*
- [18] *R.I. Wilkinson. Theories for toll traffic engineering in USA, Bell System Tech. J, 35 : 421-507, 1956.*
- [19] *T. Yang et J.G.C Templeton. A survey on retrial queues, Queueing systems, 2 : 201-233, 1987.*
- [20] *R.D. Yates et D.J. Goodman, Probability and Stochastic Processes, Wiley, N.Y., 1999.*