
Table des matières

1	Préliminaire	9
1.1	Processus stochastiques	9
1.1.1	Processus de comptage	9
1.1.2	Processus de renouvellement	10
1.1.3	Processus de Poisson	10
1.1.4	Loi exponentielle	11
1.1.5	Processus de naissance et de mort	12
1.2	Analyse mathématique d'un système de files d'attente	14
1.2.1	File d'Attente Simple	15
1.2.2	Structure et discipline de la file	15
1.2.3	Discipline de service	17
1.2.4	Notation de Kendall	17
1.2.5	Loi de Little	18
1.2.6	Mesures de performance d'une file d'attente	19
2	Systèmes de files d'attente classiques	20
2.1	Système de files d'attente M/M/1	20
2.2	Système de files d'attente M/M/c	24
2.3	Système de files d'attente M/M/c/K	29
2.4	Files d'attente avec impatience	31
2.4.1	Files d'attente avec impatience : Une classe de clients	31
2.4.2	Disciplines de service	33
2.4.3	Les mesures de performances	34

2.5	Le système M/M/1+M	34
2.5.1	Calcul des mesures de performances	36
3	Etude du système M/M/s/(s + N) avec clients impatientes	37
3.1	Description du modèle	37
3.2	Les mesures de performances	40
3.2.1	La probabilité de blocage	40
3.2.2	Nombre moyen de clients dans la file	40
3.2.3	La probabilité d'abandon	40
3.2.4	Nombre moyen de clients en service	41
3.2.5	Probabilité d'atteinte du service	42
3.2.6	Le rapport de blocage, d'abandon et de clients servis	43
3.2.7	Nombre de clients dans le système vu par une arrivée acceptée	44
3.2.8	Probabilité d'attente	44
3.2.9	La durée moyenne d'attente	45

Table des figures

1.1	Représentation d'une file d'attente.	15
1.2	File d'attente avec plusieurs serveurs.	16
1.3	Capacité d'une file d'attente.	17
2.1	File d'attente M/M/1.	20
2.2	Diagramme de transition d'état M/M/1.	21
2.3	File d'attente M/M/c.	25
2.4	Diagramme de transition d'état M/M/c.	26
2.5	Diagramme de transition d'état M/M/c/K.	29
2.6	Graphe des premiers états de la file M/M/1+M.	35
3.1	Shéma du système M/M/s/(s + N) avec impatience.	38
3.2	Graphe de transition du système M/M/s/(s+N) avec impatience.	38

Remerciement

En premier lieu, je tiens à remercier mon **Dieu** qui m'a donné le courage et la volonté pour réaliser ce travail.

On tient à remercier vivement notre encadreur **Mdm B.wahiba** pour le soutien et l'aide qu'il n'a jamais manqué de nous apporter, pour ses conseils et ses orientations durant l'élaboration du travail.

On remercie aussi les membres du jury qui ont accepté l'évaluation de cet ouvrage et de nous faire part de leurs remarques surement pertinentes, qui avec un peu de recul contribuerons sans doute au perfectionnement du présent travail

Dédicaces

A ma très chère mère

Qui n'a jamais dit non à mes exigences et qui n'a épargné aucun effort pour me rendre heureuse. Quoi que je fasse ou que je dise ; je ne saurai point te remercier comme il se doit.

Ta présence à mes côtés a toujours été ma source de force

A mon très cher père

Pour ta confiance ; ton soutien ; tes sacrifices et toutes les valeurs que tu as su m'inculquer

.Tu as été toujours à mes côtés malgré les distances qui nous séparent.

"Veuillez trouver ici le témoignage de mon amour éternel"

A mes chères frères **Toufik,Hichem,Mohamed**

Que Dieu vous procure santé ; prospérité et bonheur...

A tous mes amis ; mes collègues et tous ceux ou celles qui ont cru en moi et m'ont chaleureusement supporté et encouragé tout au long de mon parcours. Pour plusieurs raisons ; je veux remercier toutes les personnes qui ont aidé du près ou du loin pour réaliser ce modeste travail .

Introduction générale

LA théorie des files d'attente est l'un des outils analytiques les plus puissants pour la modélisation de systèmes de logistiques et de communications. Cette théorie a pour objet l'étude de systèmes et réseaux ou des entités, appelées clients, cherchent à accéder à des ressources, généralement limitées, afin d'en obtenir un service.

Prenons par exemple le cas d'un opérateur téléphonique et qui veut mettre en œuvre un service d'assistance technique pour sa clientèle. Bien évidemment, cet opérateur cherche à acquérir / garder une bonne réputation en résolvant les problèmes rencontrés par ses clients. Par contre, ses clients sont dans la majorité des cas impatients car ils ne peuvent pas attendre longtemps pour être mis en relation avec un téléconseiller. Une solution naïve serait d'embaucher autant de télé conseillers que de clients mais on voit rapidement que cette solution n'est ni envisageable ni optimale.

Dans notre travail, nous avons refait et détaillé quelques résultats pour les files d'attente avec des clients impatients dans le cas markovien, avec une seule classe de clients. Dans le cas d'une file d'attente avec une seule classe de clients, avec un temps d'attente maximum déterministe pour chaque client, nous avons refait le cadre de probabilité de perte, qui a la particularité d'être général à toute arrivée de distribution et temps de service. Nous donnons quelques applications de cette inégalité pour des files d'attente particulières. Ce cadre permet d'estimer à l'avance la probabilité de perte qui conditionne le choix de la bonne configuration du système. Par la suite, nous nous sommes intéressés au système Markovien multiserveurs, aux capacités limitées et aux clients impatients en tenant compte de son importance dans la modélisation des téléphones des centres commerciaux. Une étude détaillée des mesures de performance. Par contre il est clair qu'en intégrant une possibilité d'abandon dans un système à deux classes de clients, le calcul se complique d'autant. Cependant, dans le cas

particulier où l'on a le même taux d'abandon γ et avec une priorité stricte, nous avons développé une approche très simple, qui permet de généraliser les résultats trouvés avec une seule classe à n classes de clients tout en offrant la possibilité de programmation récurrente qui est plutôt efficace pour ce genre de problème.

Ce mémoire est composé d'une introduction générale, de trois chapitres et une conclusion générale.

Le chapitre 1 résume les résultats importants de la théorie de file d'attente classique. Le deuxième chapitre traite les files d'attente avec impatience dans le cas markovien ainsi que le cas des délais déterministe une classe et deux classes de clients. Dans le chapitre 3 on étudie le système markovien à plusieurs serveurs, capacité limitée et avec clients impatients.

Chapitre 1

Préliminaire

1.1 Processus stochastiques

Définition 1.1.0.1. (*Processus stochastiques*)[36]

Un processus stochastique $\{X(t), t \in T\}$ est une collection de variables aléatoires définies sur un même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. Le paramètre t est généralement interprété comme le temps et appartient à un ensemble ordonné T .

Généralement $X(t)$ représente l'état du processus stochastique au temps t .

- Si T est dénombrable, i.e $T \subseteq \mathbb{N}$, alors nous disons que $\{X(t), t \in T\}$ est un processus à temps discret. On le dénote par $\{X_n, n \geq 0\}$.
- Si T est un intervalle de $[0; \infty)$, alors le processus stochastique est dit un processus à temps continu. On le dénote par $\{X(t), t \geq 0\}$.

L'ensemble des valeurs de $X(t)$ est appelé l'espace d'état, qui peut également être soit discret (fini ou infini dénombrable) ou continu (un sous-ensemble de \mathbb{R} ou \mathbb{R}^n).

1.1.1 Processus de comptage

Définition 1.1.1.1. (*Processus de comptage*)[33]

Un processus stochastiques $\{N(t), t \geq 0\}$ est dit processus de comptage ou processus de dénombrement si

$N(t)$ représente le nombre d'événements se produisant dans l'intervalle $[0, t]$ vérifiant :

- $N(t) \geq 0, \forall t \geq 0$.

- $\forall t > s, N(t) \geq N(s)$.
- Pour $s < t$, $N(t) - N(s)$ représente le nombre d'événements se produisant dans l'intervalle $(s, t]$.

1.1.2 Processus de renouvellement

Un processus de renouvellement a pour fonction le dénombrement des occurrences d'un phénomène donné, lorsque les délais entre deux occurrences consécutives sont des variables aléatoires indépendantes et identiquement distribuées.

Définition 1.1.2.1. [17]

Un processus de comptage pour lequel les temps entre deux arrivées consécutives sont des variables aléatoires i.i.d, s'appelle processus de renouvellement. Les temps de renouvellement (ou les temps de la n -ième arrivée) sont :

$$A_n = \sum_{i=1}^n a_i, \quad n = 1, 2, \dots$$

avec $a_i, i = 1, 2, \dots$ est le temps entre deux arrivées consécutives. Il est facile de voir que le nombre d'arrivées avant le temps t , i.e. le processus

$$\{N(t), t \geq 0\} = \sup_k \{k \in \mathbb{N} : A_k \leq t\}$$

est un processus de comptage.

1.1.3 Processus de Poisson

Le processus de poisson est le plus utilisé dans la théorie des files d'attente. Il modélisera généralement le processus d'arrivée des clients dans un système.

Définition 1.1.3.1. (Processus de poisson)[11]

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

C1 ► Le processus est homogène dans le temps : La probabilité d'avoir k événements dans

un intervalle de longueur donné t ne dépend que de t et non pas de la position de l'intervalle par rapport l'axe temporel :

$$p_k(t) = \mathbb{P}(N(t) = k) = \mathbb{P}(N(t+s) - N(s) = k) \text{ pour tout } s > 0, t > 0.$$

C2 ► Le processus $N(t)$ est à accroissement indépendants :

$$\begin{aligned} \mathbb{P}(N(t+s) - N(s) = k, N(s) = j) &= \mathbb{P}(N(t+s) - N(s) = k) \mathbb{P}(N(s) = j) \\ &= p_k(t) p_j(s) \end{aligned}$$

pour tout $s > 0, t > 0$.

C3 ► La probabilité $p_k(\Delta t)$

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

λ est appelé densité ou intensité du processus. C'est le nombre d'évènements qui apparaissent par unité de temps.

1.1.4 Loi exponentielle

Définition 1.1.4.1. (loi exponentielle)[2]

Soit $\mu > 0$. On dit qu'une variable aléatoire réelle continue T suit la loi exponentielle de paramètre μ ($T \sim \text{Exp}(\mu)$) si

$$f(t) = \begin{cases} \mu e^{-\mu t} & t \geq 0, \\ 0 & \text{sinon.} \end{cases}$$

Distribution exponentielle :

La fonction de répartition de cette loi est :

$$F(t) = \begin{cases} 1 - e^{-\mu t} & t \geq 0, \\ 0 & \text{sinon.} \end{cases}$$

L'espérance et la variance d'une loi exponentielle sont :

$$\mathbb{E}(T) = \frac{1}{\mu}.$$

$$\text{Var}(T) = \frac{1}{\mu^2}.$$

1.1.5 Processus de naissance et de mort

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

Les files d'attente de type Markovien (M/M) sont des cas particuliers très importants de processus de naissance et de mort. Leur étude complète sera effectuée dans le chapitre 2.

Définition 1.1.5.1. [39]

Soit un processus stochastique $\{N(t), t \geq 0\}$ à états discrets $n \in \mathbb{N}$, et homogène dans le temps, c'est à dire :

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

Le processus $\{N(t), t \geq 0\}$ est un processus de naissance et de mort s'il satisfait les conditions suivantes :

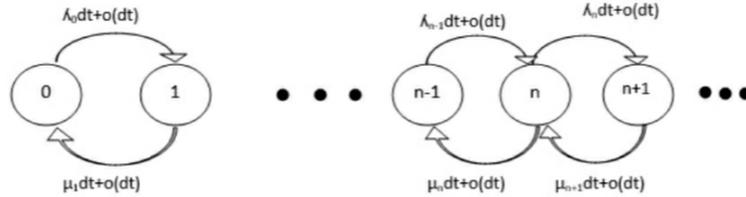
$$\begin{cases} p_{i,i+1}(\Delta t) = \lambda_i \Delta t + o(\Delta t), & i \geq 0, \\ p_{i,i-1}(\Delta t) = \mu_i \Delta t + o(\Delta t), & i \geq 1, \\ p_{i,i}(\Delta t) = 1 - (\lambda_i + \mu_i) \Delta t + o(\Delta t), & i \geq 0, \\ p_{i,j}(\Delta t) = o(\Delta t) & \text{si } |i - j| \geq 2, \end{cases}$$

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

Les coefficients positifs $\lambda_i > 0$ et $\mu_i > 0$, ($\mu_0 = 0$), sont appelés taux de transition, plus particulièrement taux de naissance (ou de croissance) pour λ_i et taux de mort (ou de décroissance) pour μ_i .

Régime transitoire :

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



La matrice des transitions correspondante est :

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

En appliquant $\mathbb{P}(t + \Delta t) = \mathbb{P}(t) \times Q$, on trouve

$$p_0(t + \Delta t) = (1 - \lambda_0 \Delta t)p_0(t) + \mu_1 \Delta t p_1(t); \quad (1.1)$$

$$p_n(t + \Delta t) = \lambda_{n-1} \Delta t p_{n-1}(t) + (1 - (\lambda_n + \mu_n) \Delta t)p_n(t) + \mu_{n+1} \Delta t p_{n+1}(t), \quad n \geq 1.$$

$$\frac{p_n(t + \Delta t) - p_n(t)}{\Delta t} = \lambda_{n-1} p_{n-1}(t) - (\lambda_n + \mu_n) p_n(t) + \mu_{n+1} p_{n+1}(t).$$

On faisant tendre Δt vers 0, on trouve :

$$p'_n(t) = \lambda_{n-1} p_{n-1}(t) - (\lambda_n + \mu_n) p_n(t) + \mu_{n+1} p_{n+1}(t), \quad n \geq 1. \quad (1.2)$$

Pour $n = 0$:

$$p'_0(t) = -\lambda_0 p_0(t) + \mu_1 p_1(t). \quad (1.3)$$

Donc :

$$\begin{cases} p'_0(t) = -\lambda_0 p_0(t) + \mu_1 p_1(t); \\ p'_n(t) = \lambda_{n-1} p_{n-1}(t) - (\lambda_n + \mu_n) p_n(t) + \mu_{n+1} p_{n+1}(t), \quad n \geq 1. \end{cases} \quad (1.4)$$

Les équations (1.4) sont connues sous le nom "équations différentielles de Kolmogorov " elles permettent de calculer les probabilités d'état $p_n(t)$ si l'on connaît les conditions initiales du processus.

Régime stationnaire :

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

$$\lambda_0 p_0 = \mu_1 p_1; \tag{1.5}$$

$$(\lambda_n + \mu_n) p_n = \lambda_{n-1} p_{n-1} + \mu_{n+1} p_{n+1}, \quad n \geq 1$$

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 $n = 1$:

$$(\lambda_1 + \mu_1) p_1 = \lambda_0 p_0 + \mu_2 p_2 \Rightarrow p_2 = \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} p_0$$

⋮

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

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

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

Pour que le régime existe il faut que la somme ci-dessus converge.

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

► L'étude mathématique d'un système de files d'attente se fait le plus souvent par l'introduction d'un processus stochastique, défini de façon appropriée.

En premier lieu, on s'intéresse principalement au nombre de clients $N(t)$, se trouvant dans

le système à l'instant t ($t \geq 0$).

En fonction des quantités qui définissent le système, on cherche à calculer :

★ Les probabilités d'état $p_n(t) = \mathbb{P}(N(t) = n)$, qui définissent le **régime transitoire** du processus stochastique $\{N(t), t \geq 0\}$. Il est évident que les fonctions $p_n(t)$ dépendent de l'état initial ou de la distribution initiale du processus.

★ Le **régime stationnaire** du processus stochastique est défini par :

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

Où, $\{p_n\}_{n \geq 0}$ est appelée distribution stationnaire du processus $\{N(t), t \geq 0\}$.

1.2.1 File d'Attente Simple

Le modèle général d'un système de files d'attente peut être résumé comme suit.

Les demandes de service (clients) arrivent à un certain endroit et réclament un certain service. Si un dispositif de service (serveur) est libre, le client qui arrive se dirige vers ce dernier où il est servi. Dans le cas contraire, on a deux possibilités : soit le client quitte le système, soit il prend une place dans une file d'attente. A un moment donné, le client est sélectionné pour le service selon une discipline donnée. Une représentation graphique est donnée par la figure (1.1)

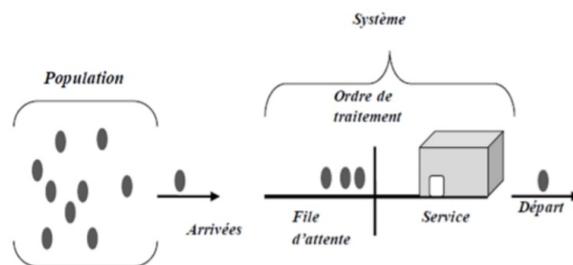


FIGURE 1.1 – Représentation d'une file d'attente.

1.2.2 Structure et discipline de la file

File d'attente : "Lieu" où les clients font la queue avant d'être servis.[4]

Système d'attente : la file d'attente + service en cours.

- **Le processus des arrivées des clients :**

Les arrivées des clients sont caractérisées par l'ensemble des instants d'arrivées de chaque client ou d'un groupe de clients dans le système. La collection de ces instants forment un processus des arrivées. Souvent, on suppose que les temps entre deux arrivées consécutives sont indépendants et identiquement distribués.

- **La source des clients :**

La population source, d'où proviennent les clients, peut être finie ou infinie, unique ou multiples.

- **Nombre de serveurs :**

Une station peut disposer de plusieurs serveurs en parallèle. Soit C (Voire FIG (1.2)) le nombre de serveurs. Dès qu'un client arrive à la station, soit il y a un serveur de libre et le client entre instantanément en service, soit tous les serveurs sont occupés et le client se place dans la file en attente de libération d'un des serveurs. La plupart du temps, les serveurs sont supposés identiques (ils possèdent donc la même distribution) et indépendants les uns des autres.

Une station particulière est la station IS (infini servers) dans laquelle le nombre de serveurs est infini. Cette station ne comporte donc pas de file d'attente.

Dès qu'un client s'y présente, il trouve en effet instantanément un serveur disponible et entre donc directement en service. Elle permet de représenter des systèmes pour lesquels le nombre de serveurs est toujours supérieur au nombre de clients qui peuvent s'y trouver.

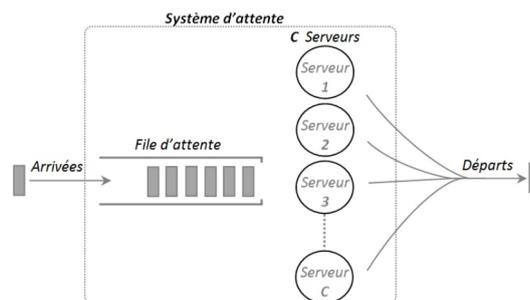


FIGURE 1.2 – File d'attente avec plusieurs serveurs.

- **Capacité de la file :**

Elle représente le nombre maximal de clients dans le système. Un client arrivant et trouvant ce nombre de clients présents dans le système sera perdu.

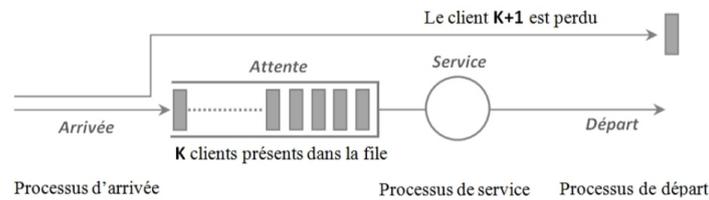


FIGURE 1.3 – Capacité d'une file d'attente.

1.2.3 Discipline de service

Elle spécifie la manière avec laquelle le serveur sélectionne le prochain client à servir. Cependant, plusieurs possibilités existent quant à l'ordre selon lequel les clients seront servis. Les principales disciplines de service sont :

FIFO (first in, first out) : Cette discipline est la plus usuelle. Les clients quittent le système dans l'ordre suivant lequel ils sont entrés.

LIFO (last in, first out) : Le dernier client dans la file est le premier à être servi.

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

Prioritaire : Les clients sont servis suivant un attribut qui leur est associé.

PS (Processor Sharing) : les clients sont servis de manière égale. La capacité du système est partagée entre les clients.

1.2.4 Notation de Kendall

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

$$\mathbf{A/B/c/m/n/Z}$$

où **A** et **B** décrivent respectivement la distribution des temps entre deux arrivées successives

et la distribution des temps de service, \mathbf{c} est le nombre de serveurs (montés en parallèle), \mathbf{m} est la capacité du système. Le dernier symbole peut être supprimé si $\mathbf{m} = \infty$.

\mathbf{n} : population des usagers.

\mathbf{Z} : discipline de service c'est la façon dont les clients sont ordonnés pour être servi.

• Pour spécifier les distributions \mathbf{A} et \mathbf{B} , on introduit les symboles suivants :

\mathbf{M} : inter-arrivées des clients sont indépendamment, identiquement distribuées selon une loi exponentielle. Il correspond à un processus de Poisson ponctuel (propriété sans mémoire).

E_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 .

H_k : distribution hyperexponentielle de degré k .

\mathbf{D} : les temps inter-arrivées des clients ou les temps de service sont constants et toujours les mêmes.

\mathbf{G} : Inter-arrivées de clients ont une distribution générale et peuvent être dépendantes.

1.2.5 Loi de Little

La loi de Little est une relation très générale qui s'applique à une grande classe de systèmes.

Elle ne concerne que le régime permanent du système. Aucune hypothèse sur les variables aléatoires qui caractérisent le système (temps d'inter - arrivées, temps de service, ...) n'est nécessaire. La seule condition d'application de la loi de Little est que le système soit stable. Le débit du système est alors indifféremment, soit le débit d'entrée, soit le débit de sortie : $\lambda_s = \lambda_e = \lambda$. La loi de Little s'exprime telle que dans la propriété suivante :

Théorème 1.2.5.0.1. [4](Formule de Little)

Le nombre moyen de clients, le temps moyen passé dans le système et le débit moyen d'un système stable en régime permanent se relient de la façon suivante :

$$\bar{N} = \lambda_e \bar{T}_s .$$

Où λ_e est le taux d'entrée dans le système ($\lambda_e = \lambda$ pour une file $(M/M/1)$).

1.2.6 Mesures de performance d'une file d'attente

On note λ le taux d'arrivée des clients. Cela signifie que l'espérance mathématique de la durée séparant deux arrivées successives est $\mathbb{E}(A) = \frac{1}{\lambda}$.

On note μ le taux de service des clients. Cela signifie que l'espérance de la durée de service est $\mathbb{E}(S) = \frac{1}{\mu}$.

L'intensité du trafic s'exprime de la manière suivante :

$$\rho = \frac{\lambda}{\mu} = \frac{\mathbb{E}(S)}{\mathbb{E}(A)} = \text{temps moyen de service} / \text{temps moyen entre deux arrivées successives.}$$

La distribution stationnaire du processus stochastique introduit permet d'obtenir les caractéristiques d'exploitation du système, telles que :

Le temps d'attente d'un client W , le temps de séjour d'un client dans le système W_s , le taux d'occupation des dispositifs de service, la durée de la période d'activité, le nombre de clients dans le système N , nombre de clients dans la files d'attente N_f .

Les mesures de performance sont :

- Le nombre moyen de clients dans le système \bar{N} ;
- Le nombre moyen de clients dans la file d'attente \bar{N}_s ;
- Le temps moyen d'attente d'un client \bar{T}_c ;
- Le temps moyen de séjour d'un client dans le système \bar{T}_s ;

Soient encore des relations (formules de Little) :

$$\bar{N} = \lambda \bar{T}_s; \quad \bar{N}_s = \lambda \bar{T}_c; \quad \bar{T}_s = \bar{T}_c + 1/\mu; \quad \bar{T}_c = \frac{\bar{N}_s}{\lambda}; \quad \bar{N} = \bar{N}_s + \frac{\lambda}{\mu}.$$

Chapitre 2

Systèmes de files d'attente classiques

2.1 Système de files d'attente M/M/1

Description du modèle :

Le système d'attente M/M/1 est un système formé d'une file de capacité infinie, d'un unique serveur et la discipline d'attente est **FIFO**. Les clients arrivent vers le système selon un processus de Poisson de taux $\lambda > 0$ (nombre moyen de clients arrivant pendant une unité de temps), le taux de service est μ (nombre moyen de clients servis pendant une unité de temps)

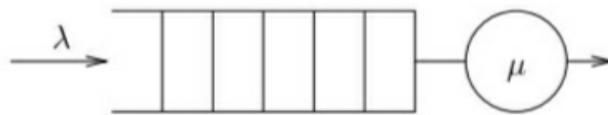


FIGURE 2.1 – File d'attente M/M/1.

La file peut être considérée comme un processus de naissance et de mort, pour lequel :

Les taux des arrivés λ_n et de service μ_n sont :

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

$$\mu_n = \begin{cases} \mu & n \neq 0 \\ 0 & n = 0 \end{cases}.$$

Le système est stable si :

$$\rho = \frac{\lambda}{\mu} < 1.$$

- Si $\rho > 1$ le nombre de client tend vers l'infini donc le système n'est pas stable.

Analyse du modèle :

L'état du système à la date t peut être décrit par le processus stochastique (??). Grâce aux propriétés fondamentales du processus de Poisson et de la loi exponentielle, on a pour un petit intervalle du temps Δt les probabilités suivantes :

$$\mathbb{P}(\text{exactement une arrivée pendant } \Delta t) = \lambda \Delta t + o(\Delta t);$$

$$\mathbb{P}(\text{aucune arrivée pendant } \Delta t) = 1 - \lambda \Delta t + o(\Delta t);$$

$$\mathbb{P}(\text{deux arrivées ou plus pendant } \Delta t) = o(\Delta t);$$

$$\mathbb{P}(\text{exactement un départ pendant } \Delta t / N(t) > 0) = \mu \Delta t + o(\Delta t);$$

$$\mathbb{P}(\text{aucun départ pendant } \Delta t / N(t) > 0) = 1 - \mu \Delta t + o(\Delta t);$$

$$\mathbb{P}(\text{deux départs ou plus pendant } \Delta t) = o(\Delta t).$$

- Ces probabilités ne dépendent ni de temps t ni de l'état $N(t)$ dans lequel le système se trouve.

Soient $p_{ij}(\Delta t) = \mathbb{P}(N(t + \Delta t) = j / N(t) = i); i, j = 0, 1, 2, \dots$. Ces *probabilités de transition* ne dépendent pas de l'instant t . On suppose que les arrivées et les départs sont mutuellement indépendants.

Régime transitoire :

Soit $p_n(t) = \mathbb{P}(N(t) = n)$. Le graphe des transitions se présente de la manière suivante

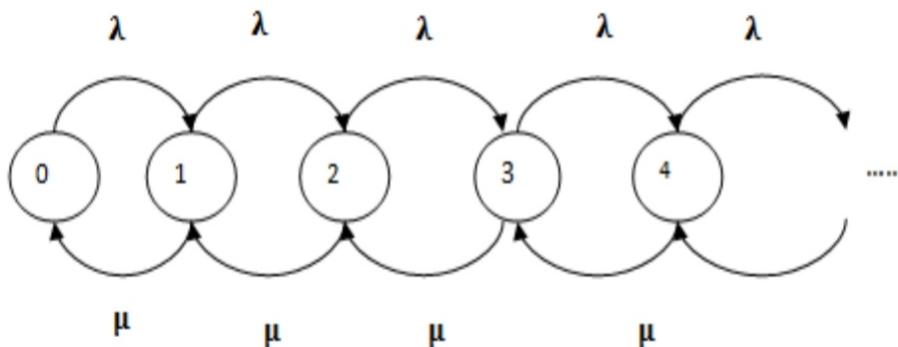


FIGURE 2.2 – Diagramme de transition d'état M/M/1.

A partir du graphe des transitions, on obtient :

$$\begin{cases} p_0(t + \Delta t) = \mu(\Delta t)p_1(t) + (1 - \lambda\Delta t)p_0(t); \\ p_n(t + \Delta t) = \mu(\Delta t)p_{n+1}(t) + \lambda(\Delta t)p_{n-1}(t) + (1 - (\lambda + \mu)\Delta t)p_n(t), \quad n \geq 1. \end{cases}$$

Puis, les équations de Kolmogorov :

$$\begin{cases} p'_0(t) = -\lambda p_0(t) + \mu p_1(t); \\ p'_n(t) = -(\lambda + \mu)p_n(t) + \lambda p_{n-1}(t) + \mu p_{n+1}(t), \quad n \geq 1. \end{cases} \quad (2.1)$$

Ces équations permettent, en principe, de calculer les probabilités d'état $p_n(t)$, si l'on connaît en plus les conditions initiales du processus, c'est-à-dire la distribution de $N(0)$.

Régime stationnaire :

Il est démontré que $\lim_{t \rightarrow \infty} p_n(t) = p_n$, $n \geq 0$, existent et sont indépendantes de l'état initial du processus $\{N(t), t \geq 0\}$, et $\lim_{t \rightarrow \infty} p'_n(t) = 0$, $n \geq 0$. De (2.1), on obtient le système d'équations de balance suivant :

$$\begin{cases} \mu p_1 = \lambda p_0, \\ \lambda p_{n-1} + \mu p_{n+1} = (\lambda + \mu)p_n, \quad n \geq 1, \end{cases} \quad (2.2)$$

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

La résolution du système (2.2) (la résolution du modèle) s'effectue de la manière suivante :

$$p_1 = \frac{\lambda}{\mu} p_0.$$

Pour $n = 1$:

$$\lambda p_0 + \mu p_2 = (\lambda + \mu)p_1 \implies p_2 = \left(\frac{\lambda}{\mu}\right)^2 p_0.$$

Pour $n > 1$:

$$p_n = \left(\frac{\lambda}{\mu}\right)^n p_0.$$

Pour trouver la probabilité p_0 , on utilise l'équation de normalisation.

En effet : $\sum_{n=0}^{\infty} p_n = 1 \implies p_0 + \frac{\lambda}{\mu} p_0 + \left(\frac{\lambda}{\mu}\right)^2 p_0 + \dots = 1;$

$$p_0 = \frac{1}{1 + \frac{\lambda}{\mu} + \left(\frac{\lambda}{\mu}\right)^2 + \dots},$$

où $1 + \frac{\lambda}{\mu} + \left(\frac{\lambda}{\mu}\right)^2 + \dots$ est une progression géométrique de raison $\frac{\lambda}{\mu}$. Elle converge si $\frac{\lambda}{\mu} < 1$, et est égale à $\frac{1}{1 - \frac{\lambda}{\mu}}$. Alors

$$p_0 = 1 - \frac{\lambda}{\mu}.$$

D'où

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

$\rho = \frac{\lambda}{\mu}$ est l'intensité du trafic. $\rho < 1$ est la condition d'existence du régime stationnaire.

Encore, $p_n = (1 - \rho)\rho^n$, $n \geq 0$, est la distribution stationnaire du nombre de clients dans le système M/M/1.

Caractéristiques du système M/M/1 :

► **Le nombre moyen de clients dans le système :**

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

$$\begin{aligned} \bar{n} &= \mathbb{E}(N) = \sum_{n=0}^{\infty} n p_n \\ &= (1 - \rho) \sum_{n=0}^{\infty} n \rho^n \\ &= (1 - \rho) \rho \sum_{n=1}^{\infty} \frac{d}{d\rho} (\rho)^n \\ &= (1 - \rho) \rho \frac{d}{d\rho} \sum_{n=1}^{\infty} (\rho)^n \\ &= (1 - \rho) \rho \frac{d}{d\rho} \left(\frac{1}{1 - \rho} \right) \\ &= (1 - \rho) \rho \left(\frac{1}{(1 - \rho)^2} \right) \\ &= \frac{\rho}{1 - \rho}. \end{aligned}$$

Donc

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

► **Le nombre moyen de clients dans la file d'attente \bar{n}_f :**

Soit $N_f = \lim_{t \rightarrow \infty} N_f(t)$, où $N_f(t)$ est le nombre de clients dans la file d'attente à la date t . La

variable N_f est définie de la manière suivante :
$$N_f = \begin{cases} 0 & N = 0 \\ N - 1 & N \geq 1 \end{cases}.$$

$$\begin{aligned} \bar{n}_f &= \mathbb{E}(N_f) = \sum_{n=1}^{\infty} (n-1)p_n \\ &= \frac{\rho^2}{1 - \rho} \\ &= \frac{\lambda^2}{\mu(\mu - \lambda)}. \end{aligned}$$

Ou bien

$$\bar{n}_f = \bar{n} - \rho.$$

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

$$\begin{aligned} \bar{T}_s &= \frac{\lambda}{\mu - \lambda} \cdot \frac{1}{\lambda} \\ &= \frac{1}{\mu - \lambda}. \end{aligned}$$

► **Le temps moyen d'attente d'un client \bar{W} :**

$$\bar{T}_c = \frac{\lambda}{\mu(\mu - \lambda)}$$

2.2 Système de files d'attente M/M/c

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 place dans la file d'attente, commune pour tous les serveurs. La capacité d'attente est illimitée (le nombre de positions d'attente est infini). Lorsqu'un serveur se libère, le client en tête de la file d'attente occupe le serveur libéré. Par conséquent, la discipline d'attente est FIFO. Les temps de service sont exponentiellement distribués de moyenne finie $1/\mu$. Les durées entre deux arrivées consécutives et les durées de service sont mutuellement indépendantes.

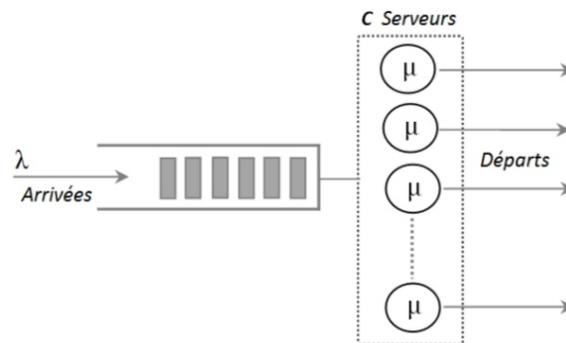


FIGURE 2.3 – File d'attente M/M/c.

Nous avons donc un modèle de file d'attente où les arrivées et les départs sont modélisés par un processus de naissance et de mort où :

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

$$\mu_n = \begin{cases} 0 & \text{si } n = 0; \\ n\mu & \text{si } n < c; \\ c\mu & \text{si } n \geq c. \end{cases}$$

Stabilité du système :

La condition de stabilité de cette file est $\lambda < c\mu$, et exprime le fait que le nombre moyen de clients qui arrivent à la file par unité de temps doit être inférieur au nombre moyen de clients que les serveurs de la file sont capables de traiter par unité de temps.

$$\rho = \frac{\lambda}{c\mu} < 1.$$

Graphe de transition :

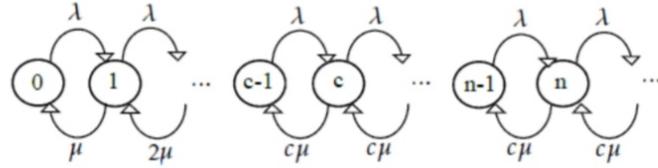


FIGURE 2.4 – Diagramme de transition d'état M/M/c.

Régime transitoire :

Le système d'équations de Kolmogorov pour les probabilités d'état $p_n(t) = \mathbb{P}(N(t) = n)$, $n \geq 0$, se présente de la manière suivante :

$$\begin{cases} p'_0(t) = -\lambda p_0(t) + \mu p_1(t); \\ p'_n(t) = \lambda p_{n-1}(t) - (\lambda + n\mu)p_n(t) + (n+1)\mu p_{n+1}(t), & 1 \leq n < c; \\ p'_n(t) = \lambda p_{n-1}(t) - (\lambda + c\mu)p_n(t) + c\mu p_{n+1}(t), & n \geq c. \end{cases}$$

Régime stationnaire :

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

Si : $n < c$

$$\begin{cases} n = 0 : & \lambda p_0 = \mu p_1; \\ n = 1 : & \lambda p_1 + \mu p_1 = \lambda p_0 + 2\mu p_2; \\ & \vdots \\ n = (c-1) : & \lambda p_{c-1} + (c-1)\mu p_{c-1} = \lambda p_{c-2} + c\mu p_c. \end{cases}$$

Donc :

$$\left\{ \begin{array}{l} p_1 = \left(\frac{\lambda}{\mu}\right)p_0 \\ p_2 = \left(\frac{\lambda}{2\mu}\right)p_1 = \frac{1}{2} \left(\frac{\lambda}{\mu}\right)^2 p_0 \\ p_3 = \left(\frac{\lambda}{3\mu}\right)p_2 = \frac{1}{3 \times 2} \left(\frac{\lambda}{\mu}\right)^3 p_0 \\ \vdots \\ p_c = \frac{1}{c!} \left(\frac{\lambda}{\mu}\right)^c p_0. \end{array} \right.$$

Si : $n \geq c$

$$\left\{ \begin{array}{l} \lambda p_c + c\mu p_c = \lambda p_{c-1} + c\mu p_{c+1} \\ \lambda p_{c+1} + c\mu p_{c+1} = \lambda p_c + c\mu p_{c+2} \\ \vdots \\ \vdots \end{array} \right.$$

$$\left\{ \begin{array}{l} p_c = \frac{1}{c!} \left(\frac{\lambda}{\mu}\right)^c p_0 \\ p_{c+1} = \frac{1}{c!c} \left(\frac{\lambda}{\mu}\right)^{c+1} p_0 \\ p_{c+2} = \frac{1}{c!c^2} \left(\frac{\lambda}{\mu}\right)^{c+2} p_0 \\ \vdots \\ p_n = \frac{1}{c!c^{n-c}} \left(\frac{\lambda}{\mu}\right)^n p_0. \end{array} \right.$$

D'où

$$p_n = \begin{cases} \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n p_0, & \text{si } 1 \leq n \leq c; \\ \frac{1}{c!c^{n-c}} \left(\frac{\lambda}{\mu}\right)^n p_0, & \text{si } n \geq c. \end{cases}$$

On remarque que pour $n = c$, les deux formules donnent la même valeur.

Pour calculer la probabilité pour que le système est vide p_0 , on applique l'équation de

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

En effet :

$$p_0 = \left(\sum_{n=0}^{c-1} \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n + \underbrace{\sum_{k=0}^{\infty} \frac{1}{c!c^k} \left(\frac{\lambda}{\mu}\right)^{c+k}}_{\star} \right)^{-1}.$$

La deuxième somme (★) peut être réécrite de la manière suivante :

$$\frac{1}{c!} \left(\frac{\lambda}{\mu}\right)^c \left(\underbrace{1 + \frac{\lambda}{\mu c} + \left(\frac{\lambda}{\mu c}\right)^2 + \dots}_{\star\star} \right).$$

La somme (★★) possède une limite égale à $\frac{1}{1 - \frac{\lambda}{\mu c}}$ si $\frac{\lambda}{\mu c} < 1$. Par conséquent, le système considéré est en régime stationnaire si $\rho = \frac{\lambda}{\mu c} < 1$, ρ est l'intensité globale du trafic. On obtient ainsi

$$p_0 = \left(\sum_{n=0}^{c-1} \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n + \frac{(\lambda/\mu)^c}{c!(1 - \frac{\lambda}{\mu c})} \right)^{-1}.$$

Encors,

$$p_0 = \left(\sum_{n=0}^{c-1} \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n + \frac{1}{c!} \left(\frac{\lambda}{\mu}\right)^c \sum_{n=c}^{\infty} \rho^{n-c} \right)^{-1},$$

et

$$p_n = \frac{1}{c!} \left(\frac{\lambda}{\mu}\right)^c \left(\frac{\lambda}{\mu c}\right)^{n-c} p_0 = \rho^{n-c} p_c.$$

Caractéristiques du système M/M/c :

► **Le nombre moyen de clients dans le système \bar{n} :**

$$\bar{n} = \sum_{n=0}^{\infty} n p_n = \sum_{n=1}^{c-1} \frac{n(\lambda/\mu)^n}{n!} p_0 + \sum_{n=c}^{\infty} \frac{n(\lambda/\mu)^n}{c! c^{n-c}} p_0$$

$$\bar{n} = \frac{\lambda}{\mu} + \frac{(\lambda/\mu)^{c+1}}{c! c \left(1 - \frac{\lambda}{\mu c}\right)^2} p_0.$$

► **Le nombre moyen de clients dans la file d'attente \bar{n}_f :**

$$\bar{n}_f = \sum_{k=0}^{\infty} k p_{c+k} = \frac{(\lambda/\mu)^c}{c!} \sum_{k=0}^{\infty} k \left(\frac{\lambda}{\mu c}\right)^k p_0$$

$$\bar{n}_f = \frac{(\lambda/\mu)^{c+1}}{c! c \left(1 - \frac{\lambda}{\mu c}\right)^2} p_0.$$

► Le temps moyen de séjour d'un client dans le système \bar{W}_s :

$$\bar{T}_s = \frac{\bar{n}}{\lambda} = \frac{1}{\mu} + \frac{(\lambda/\mu)^c}{c!c\mu \left(1 - \frac{\lambda}{\mu c}\right)^2} p_0.$$

► Le temps moyen d'attente d'un client \bar{W} :

$$\bar{T}_c = \frac{\bar{n}_f}{\lambda} = \frac{c\mu(\lambda/\mu)^c}{c!(c\mu - \lambda)^2} p_0.$$

2.3 Système de files d'attente M/M/c/K

A présent, supposons que dans le système M/M/c, le nombre de positions d'attente est limité (égal à K). A l'arrivée d'un client, si tous les serveurs et toutes les positions d'attente sont occupées, le client quitte le système définitivement sans recevoir le service.

Analyse du modèle : Le processus (2.1) décrivant l'état du système à l'étude à la date t est celui de naissance et de mort avec $\lambda_n = \lambda$ si $0 \leq n < K$, et $\mu_n = \mu \times \min\{n, c\}$ si $1 \leq n \leq K$. L'espace des états est $S = \{0, 1, 2, \dots, K\}$.

Graphe de transition :

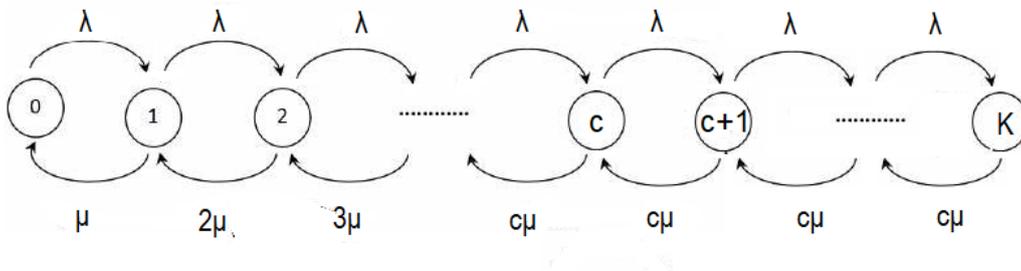


FIGURE 2.5 – Diagramme de transition d'état M/M/c/K.

Régime transitoire

Soient $p_n(t) = P(N(t) = n)$, $0 \leq n \leq K$. Le système d'équations de Kolmogorov pour les probabilités s'états s'obtient à partir du graphe des transitions ci-dessus. En effet,

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

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

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

$$p'_k(t) = \lambda p_{k-1}(t) - c\mu p_k(t).$$

Régime stationnaire

Soient $p_n = \lim_{t \rightarrow \infty} p_n(t)$, $0 \leq n \leq K$. La distribution stationnaire p_n satisfait le système d'équations de balance suivant

$$0 = -\lambda p_0 + \mu p_1;$$

$$0 = \lambda p_{n-1} - (\lambda + n\mu)p_n + (n+1)\mu p_{n+1}, \quad 1 \leq n < c;$$

$$0 = \lambda p_{n-1} - (\lambda + c\mu)p_n + c\mu p_{n+1}, \quad c \leq n < K;$$

$$0 = \lambda p_{k-1} - c\mu p_k.$$

La résolution de ce système, nous donne

$$p_n = \begin{cases} \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n p_0, & 1 \leq n < c; \\ \frac{1}{c!c^{n-c}} \left(\frac{\lambda}{\mu}\right)^n p_0 = \frac{1}{c!} \left(\frac{\lambda}{c\mu}\right)^{n-c} \left(\frac{\lambda}{\mu}\right)^c p_0, & c \leq n < K. \end{cases}$$

La mesure importante de ce système est la probabilité de perte, qui est la probabilité pour que le système se trouve dans l'état K :

$$p_k = \frac{1}{c!} \left(\frac{\lambda}{c\mu}\right)^{k-c} \left(\frac{\lambda}{\mu}\right)^c p_0.$$

La probabilité p_0 s'obtient à partir de l'équation de normalisation $\sum_{n=0}^k p_n = 1$:

$$p_0 = \left(\sum_{n=1}^{c-1} \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n + \sum_{n=c}^k \frac{1}{c!} \left(\frac{\lambda}{\mu}\right)^c \left(\frac{\lambda}{c\mu}\right)^{n-c} \right)^{-1}.$$

Dans le cas particulier où $K = c$ (système à demandes refusées), la distribution stationnaire du processus $(N(t), t \geq 0)$ correspondant (formule d'Erlang) est

$$p_n = \frac{1}{n!} \frac{\lambda^n}{\mu^n} p_0, \quad 0 \leq n \leq c, \quad \text{où } p_0 = \left(\sum_{n=0}^c \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n \right)^{-1}.$$

Mesures de performance : On démontre que

$$\begin{aligned}\bar{n}_f &= \sum_{n=1}^{K-c} n p_{c+n} = \frac{\lambda}{\mu} \frac{(\lambda/\mu)^c}{c!c} p_0 \left(1 + 2 \left(\frac{\lambda}{\mu c} \right) + 3 \left(\frac{\lambda}{\mu c} \right)^2 + \dots + (K-c) \left(\frac{\lambda}{\mu c} \right)^{K-c-1} \right) \\ &= \frac{(\lambda/\mu)^{c+1}}{(c-1)!} \frac{1 - \left(\frac{\lambda}{\mu c} \right)^{K-c} \left(1 + (K-c) \left(1 - \frac{\lambda}{\mu c} \right) \right)}{\left(c - \frac{\lambda}{\mu} \right)^2} p_0.\end{aligned}$$

L'application des relations de Little fournit d'autres mesures de performance

$$\bar{W}_c = \frac{\bar{n}_f}{\lambda}, \bar{W}_s = \frac{\bar{n}}{\lambda}, \bar{n} = \bar{n}_f + \frac{\lambda}{\mu}.$$

2.4 Files d'attente avec impatience

2.4.1 Files d'attente avec impatience : Une classe de clients

Pour modéliser un système de file d'attente avec impatience, on doit ajouter une contrainte au système en spécifiant que les clients sont perdus si le temps qu'ils passent dans le système est plus grand qu'un délai qui leur est alloué. Le modèle que nous considérons est donc une file d'attente où les clients ont un délai (ou patience) pour être servis, au delà duquel ils sortent du système sans avoir été satisfait. Cette section, décrit les modèles d'attente avec impatience et une classe de clients où plusieurs mesures de performances, à savoir la probabilité de perte, le temps d'attente moyen et le nombre moyen de clients dans le système seront étudiés.

Notations

L'impatience :

- L'impatience est une caractéristique très importante de la théorie des files d'attente. Les modèles de files d'attente de vacances avec l'impatience des clients sont considérés comme des outils très appropriés pour analyser divers systèmes de services complexes et industries importantes.
- Dans la littérature traditionnelle sur les files d'attente de vacances avec des clients impatients, les études sur le comportement des clients ont toujours été basées sur l'hypothèse que l'impatience des clients ne se produit que lorsque le serveur est en vacances.

• C'est le cas où les clients peuvent voir l'état du serveur. Cependant, dans de nombreuses situations réelles, y compris les centres d'appels et les systèmes de production, il peut ne pas être possible d'obtenir des informations sur l'état du serveur. De plus, une longue attente dans la file d'attente est un autre facteur qui conduit à l'impatience des clients quel que soit l'état du système (actif ou en vacances).

Considérons le n ème client C_n entrant dans la file d'attente est affecté du délai D_n , qui est une variable aléatoire p.s strictement positive. S'il n'a pas pu atteindre le serveur à la fin de sa patience (i.e à l'instant $T_n + D_n$), il est éliminé du système. Les délais seront donc considérés éliminatoires et ce jusqu'au début de service. Par ailleurs, la suite $D_n, n \in \mathbb{N}^*$ sera toujours supposée stationnaire de loi générique la loi de D . On supposera de plus que,

$$\mathbb{E}[D] < \infty.$$

Et on notera que,

$$\gamma = (\mathbb{E}[D])^{-1}.$$

Par ailleurs, si le client C_n est toujours dans le système à l'instant t , on appellera son délai résiduel à t $DR_n(t)$ le temps restant à t avant l'expiration de son délai. On a alors,

$$DR_n(t) = D_n - (t - T_n).$$

On peut d'écrire la file d'attente avec clients impatients avec la nomenclature suivante, introduite par Barrer [2] **A/B/s + D-DS**, où :

♣ **A, B** et s représentent comme dans la nomenclature de Kendall, respectivement la loi du processus des arrivées, des temps de services et le nombre de serveurs.

♣ **D** représente la loi générique des délais des clients, éliminatoires et jusqu'au début du services.

♣ **DS** est la discipline de service.

Remarque 2.4.1.1. ✓ La notation de Barrer indique que l'on suppose toujours que le buffer (la file) est de taille infinie. Le paramètre aléatoire D (le délai) remplace alors dans la nomenclature de Kendall classique le paramètre N comme facteur limitant la capacité du

ystème. La file d'attente avec clients impatients est donc un modèle à perte, au même titre par exemple qu'une file à buffer fini $G/G/s/N$. (où un client est perdu s'il entre dans le système alors que le buffer est déjà plein).

✓ *Il est par ailleurs à noter que dans la littérature, les délais peuvent aussi ne pas être éliminatoires (auquel cas ils ne servent qu'à donner des priorités aux clients). Un client reste alors dans le système après expiration de sa patience, même s'il n'a pas atteint le serveur. Enfin, les délais peuvent aussi être éliminatoire, mais jusqu'à la fin du service, dans ce cas un client est éliminé s'il n'a pas terminé son service à l'instant $T_n + D_n$*

2.4.2 Disciplines de service

Les disciplines de services mentionnées dans le chapitre 1 sont bien sûr applicable à une file d'attente avec clients impatients. La dynamique de la file d'attente est alors celle d'une file d'attente classique, modifiée par les éliminations de clients, qui ne sont plus à servir. Ceci étant, aucune d'entre elles ne tient à priori compte des délais des clients. Nous présentons ici deux disciplines de services, définies dans le cas où les clients sont affectés de délais et qui joueront un rôle central dans cette étude, les disciplines EDF et LDF.

Définition

♠ La discipline de service d'une file d'attente avec clients impatients est dite EDF (Earliest Dead line First) non-préemptive (notée EDF) si à chaque fin de service, le serveur sert le client dont le délai résiduel à cet instant est le plus court. Il sert alors ce client jusqu'à complétion de son service.

♠ La discipline de service est dite EDF préemptive (notée EDF_p) si le serveur sert à chaque instant le client dont le délai résiduel à cet instant est le plus court.

Définition

♠ La discipline est dite LDF (Largest Dead line First) non-préemptive (LDF) si le serveur sert à chaque fois qu'il complète un service le client dont le délai résiduel à cet instant est le plus long. Il sert alors ce client jusqu'à complétion de son service.

♠ La discipline est dite LDF préemptive (LDF_p) si le serveur sert à chaque instant le client dont le délai résiduel à cet instant est le plus long.

2.4.3 Les mesures de performances

Définissons à présent les différentes variables aléatoire caractérisant notre problème et permettant d'évaluer ses performances

- ♣ Le nombre de client moyen L dans le système (file + serveur), avec $L = \mathbb{E}[N]$.
- ♣ Le temps d'attente proposé V_n au client C_n est le temps qu'aura à attendre C_n avant d'être servis. Ce temps peut être supérieur à D_n , auquel cas C_n est perdu (il n'a pas pu atteindre le serveur).
- ♣ Le temps de séjour W_n est le temps que passe effectivement le client C_n dans le système (file + serveur),

$$W_n = (V_n + s_n) \mathbf{1}_{[V_n < D_n]} + D_n \mathbf{1}_{[V_n \geq D_n]}.$$

- ♣ Le temps d'attente W_n^q d'un client C_n est le temps effectivement passé par C_n dans la file,

$$W_n^q = V_n \mathbf{1}_{[V_n < D_n]} + D_n \mathbf{1}_{[V_n \geq D_n]}.$$

Pour tout système de file d'attente à perte, la probabilité de perte π_k de client C_k est la probabilité que ce client soit perdu. Par exemple, pour une file d'attente classique M/M/1/N fini, cette probabilité serait donnée par la probabilité qu'il y ait N clients dans le système à l'entrée de C_k .

Dans le cas d'une file d'attente avec clients impatients, la condition de perte est plus complexe π_k est donnée par la probabilité que C_k se voit proposer un temps d'attente supérieur à son délai.

$$\pi_k = \mathbb{P}[C_k \text{ est perdu}] = \mathbb{P}[V_n > D_n]$$

2.5 Le système M/M/1+M

Description du modèle : Considérons la file M/M/1+M où les arrivées sont markoviens de taux λ , le temps de service est exponentiel de taux μ et le délai de patience est exponentiel de taux γ .

Le processus $N(t)$ est un processus de naissance et de mort, dont les taux de transitions sont voir le graphe ci-dessous,

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

$$\text{et } \mu_n = \mu + n\gamma, \quad \forall n > 1$$

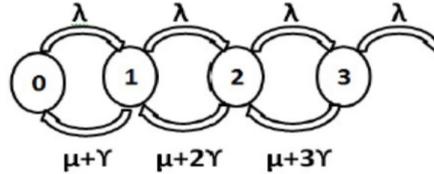


FIGURE 2.6 – Graphe des premiers états de la file M/M/1+M.

Régime stationnaire :

Posons p_n la probabilité associée à l'état n . Ainsi, et par conservation de flux nous obtenons la formule de récurrence suivante :

$$\lambda p_n = [\mu + (n + 1)\gamma] p_{n+1}.$$

Et on montrera aisément par récurrence que,

$$p_n = \left(\prod_{k=1}^n \frac{\lambda}{\mu + k\gamma} \right) p_0.$$

Par souci de simplification posons,

$$\rho_k = \frac{\lambda}{\mu + k\gamma}, \quad \forall k \geq 1.$$

Ainsi,

$$p_n = \left(\prod_{k=1}^n \rho_k \right) p_0.$$

D'autre part, et comme les probabilités somment à 1, alors, on déduit que,

$$p_0 = \frac{1}{1 + \sum_{n=1}^{\infty} \prod_{k=1}^n \rho_k}$$

p_0 existe car $\sum_{n=1}^{\infty} \prod_{k=1}^n \frac{\lambda}{\mu + k\gamma}$ converge.

2.5.1 Calcul des mesures de performances

Les mesures de performances Définissons à présent les différentes variables aléatoire caractérisant notre problème et permettant d'évaluer ses performances

- Le nombre de client moyen \bar{n} dans le système (file + serveur), avec $\bar{n} = \mathbb{E}[N]$.
- Le temps d'attente proposé V_n au client C_n est le temps qu'aura à attendre C_n avant d'être servis. Ce temps peut être supérieur à D_n , auquel cas C_n est perdu (il n'a pas pu atteindre le serveur).

Calcul des mesures de performances

Le nombre moyen de clients dans le système :

$$\bar{n}_s = \sum_{n=1}^{\infty} np_n = n \sum_{n=1}^{\infty} p_0 \prod_{k=1}^n \rho_k.$$

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

En appliquant la loi de Little $\bar{n} = \lambda \times \bar{W}_s$, on trouve.

$$\begin{aligned} \bar{W}_s &= \frac{\bar{n}}{\lambda} \\ &= \frac{p_0}{\lambda} \sum_{n=1}^{\infty} n \prod_{k=1}^n \rho_k. \end{aligned}$$

Probabilité de perte d'un client à l'état stationnaire :

$$\Pi = \mathbb{P}[V > D] = \frac{\gamma}{\lambda} \sum_{n=1}^{\infty} p_0 \prod_{k=1}^n \rho_k = \frac{\gamma}{\lambda} (1 - p_0).$$

Les résultats obtenus, lors du calcul des performances sont purement théoriques. Ainsi, dans la pratique nous utilisons des approches algorithmiques beaucoup plus intéressantes, comme la troncature des sommes infinies, afin de réduire le temps d'exécution des programmes les implémentant.

Chapitre 3

Etude du système $M/M/s/(s + N)$ avec clients impatientes

Les modèles de file d'attente ont été largement utilisés pour modéliser les performances des centres d'appel avec des clients impatientes, ce qui signifie que les clients en attente peuvent quitter la file avant d'obtenir un service. La file d'attente $M/M/s$ avec des clients impatientes est appelé "un modèle d'Erlang-A, "A" pour abandon " par Maudelbaum et Zeltyn en contraste avec le modèle bien connu Erlang-N, $M/M/s$ avec seulement des clients patients. Pour une synthèse sur les modèles de files d'attente des modèles pour les centres d'appels, voir par exemple ,Gans et al. et Koole et Maudelbaum .

3.1 Description du modèle

On considère le système de file d'attente $M/M/s/(s+N)$ avec clients impatientes où les arrivées forment un processus de Poisson de taux λ , les temps de services sont exponentielles de moyenne $\frac{1}{\mu}$. La capacité de la file N est finie, le délai de patience de chaque client est exponentielle de moyenne $\frac{1}{\gamma}$. Une fois le service est commencé, il ne quitte pas le système. Ce système est représenté dans la figure 3.1.

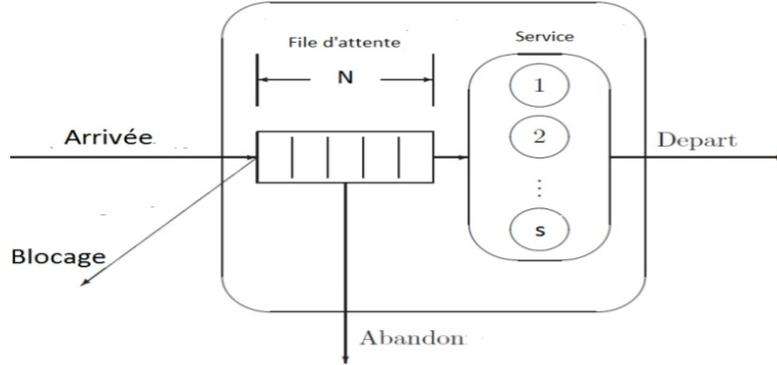


FIGURE 3.1 – Shéma du système M/M/s/(s + N) avec impatience.

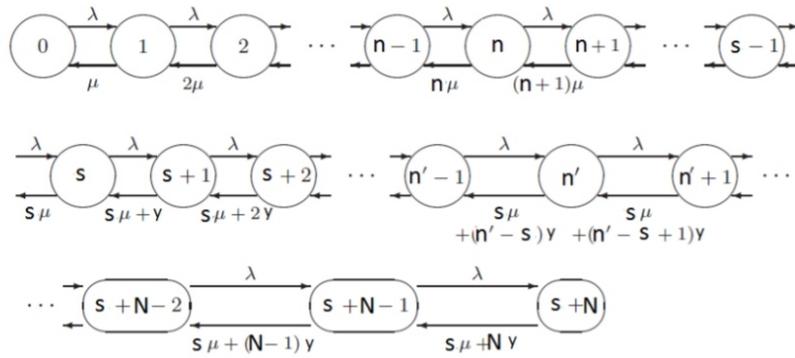


FIGURE 3.2 – Graphe de transition du système M/M/s/(s+N) avec impatience.

Régime stationnaire

Soient $p_n = \lim_{t \rightarrow \infty} P(N(t) = n), 0 \leq n \leq s + N$, La distribution stationnaire p_n satisfait le système d'équations de balance suivant :

$$\left\{ \begin{array}{l} n = 0 : \quad \lambda p_0 = \mu p_1; \\ 1 \leq n \leq s - 1 : \quad (\lambda + n\mu)p_n = \lambda p_{n-1} + (n + 1)\mu p_{n+1}; \\ s \leq n \leq s + N - 1 : \quad (\lambda + s\mu + (n - s)\gamma)p_n = \lambda p_{n-1} + [s\mu + (n - s + 1)\gamma]p_{n+1}; \\ \quad \quad \quad (s\mu + N\gamma)p_{s+N} = \lambda p_{s+N-1}. \end{array} \right.$$

Avec La condition de Normalisation est donnée par :

$$\sum_{n=0}^{s+N} p_n = 1.$$

L'ensemble des équations de balances est équivalent à l'ensemble des équations suivantes :

$$\lambda p_{n-1} = \begin{cases} n\mu p_n, & 1 \leq n \leq s-1; \\ (s\mu + (n-s)\gamma)p_n, & s \leq n \leq s+N. \end{cases}$$

et

$$1 \leq n \leq s : p_n = p_0 \frac{\rho^n}{n!} \quad \text{avec,} \quad \rho = \frac{\lambda}{\mu};$$

Pour $1 \leq n \leq s+N$, nous avons,

$$\begin{aligned} p_{n+s} &= \frac{\lambda}{s\mu + n\gamma} p_{s+n-1} = \frac{\lambda}{s\mu \left[1 + \frac{n\gamma}{s\mu}\right]} p_{s+n-1} \\ &= \frac{\frac{\rho}{s}}{1 + n\frac{\xi}{s}} p_{s+n-1} = \frac{\frac{\rho}{s}}{1 + n\frac{\xi}{s}} p_{s+n-2} \\ &= \dots \\ &= p_0 \frac{\rho^s}{s!} \frac{\left(\frac{\rho}{s}\right)^n}{\prod_{j=1}^n \left(1 + j\frac{\xi}{s}\right)}. \\ p_{n+s} &= \frac{\lambda}{s\mu + n\gamma} p_{s+n-1} = p_0 \frac{\rho^s}{s!} \frac{\left(\frac{\rho}{s}\right)^n}{\prod_{i=1}^n \left(1 + j\frac{\xi}{s}\right)}. \end{aligned}$$

Où $\xi = \frac{\gamma}{\mu}$: est le taux moyen de patience par le temps moyen de service. En utilisant condition de normalisation, on détermine :

$$\frac{1}{p_0} = \sum_{n=0}^{s-1} \frac{\rho^n}{n!} + \frac{\rho^s}{s!} \frac{\left(\frac{\rho}{s}\right)^n}{\prod_{j=1}^n \left(1 + j\frac{\xi}{s}\right)} \quad (3.1)$$

Ainsi, on trouve $\{p_n, 0 \leq n \leq s+N\}$ explicitement.

3.2 Les mesures de performances

3.2.1 La probabilité de blocage

Les clients qui arrivent au système quand toute la file est pleine, à savoir il y a $(s+N)$ clients dans le système, sont bloqués. Ils sont perdus à jamais. D'après la propriété de PASTA la distribution de probabilité à l'heure d'arrivée est la même à un moment quelconque. Par conséquent, la probabilité de blocage \mathcal{BLOC} des clients est :

$$\mathbb{P}\{\mathcal{BLOC}\} = p_{s+N} = p_0 \frac{\rho^s}{s!} \frac{\left(\frac{\rho}{s}\right)^n}{\prod_{j=1}^n \left(1 + j \frac{\rho}{s}\right)}.$$

3.2.2 Nombre moyen de clients dans la file

Soit N_q : "le nombre de clients dans la file d'attente à un temps arbitraire". S'il y a $n+s$ clients dans le système, il y a n clients dans la file d'attente. Par conséquent, nous avons :

$$\begin{aligned} \mathbb{P}\{N_q = 0\} &= \sum_{n=0}^s p_n = \sum_{n=0}^s p_0 \frac{\rho^n}{n!} \\ \mathbb{P}\{N_q = n\} &= p_{s+n} = p_0 \frac{\rho^s}{s!} \frac{\left(\frac{\rho}{s}\right)^n}{\prod_{j=1}^n \left(1 + j \frac{\rho}{s}\right)}, \quad 1 \leq n \leq s. \end{aligned}$$

Le nombre moyen de clients dans la file d'attente est donnée par :

$$L_q = \sum_{n=1}^N n p_{s+n} = p_0 \frac{\rho^s}{s!} \sum_{n=1}^N \frac{n \left(\frac{\rho}{s}\right)^n}{\prod_{j=1}^n \left(1 + j \frac{\rho}{s}\right)}.$$

Quand on trace L_q en fonction de ρ , on a les limites suivantes :

$$\lim_{\rho \rightarrow 0} L_q = 0; \quad \lim_{\rho \rightarrow \infty} L_q = N.$$

3.2.3 La probabilité d'abandon

Etant donné que chaque client quitte la file d'attente avec un taux γ , la fraction des clients qui abandonnent est donné par :

$$\mathbb{P}\{Ab\} = \frac{1}{\lambda} \sum_{n=1}^N (n\gamma) p_{s+n} = \frac{\gamma}{\lambda} p_0 \sum_{n=1}^N \frac{n \left(\frac{\rho}{s}\right)^n}{\prod_{j=1}^n \left(1 + j \frac{\rho}{s}\right)}$$

Ainsi, nous avons la relation suivante :

$$\gamma L_q = \lambda \mathbb{P}\{Ab\}.$$

Après, nous dérivons la relation :

$$\lambda(1 - \mathbb{P}\{\mathcal{B}\}) = \gamma L_q + \mu \mathbb{E}[Z].$$

avec $s + N$ est le nombre de client dans le système. La probabilité qu'un client arbitraire abandonne est donnée par :

$$\mathbb{P}\{Ab \mid Nb\} = \frac{\mathbb{P}\{Ab\}}{1 - \mathbb{P}\{\mathcal{B}\}} = \frac{\gamma L_q}{\gamma L_q + \mu \mathbb{E}[N]}. \quad (3.2)$$

où l'événement Nb est qu'un client qui arrive soit acceptée.

Depuis,

$$\begin{aligned} \lim_{\rho \rightarrow +\infty} \mathbb{P}\{Ab \mid Nb\} &= \frac{\gamma N}{\gamma N + \mu s} \\ &= \frac{1}{1 + \frac{s}{(N\xi)}}. \end{aligned}$$

3.2.4 Nombre moyen de clients en service

Soit Z : "le nombre de clients en service à l'instant t ". Après on a :

$$\begin{aligned} \mathbb{P}\{Z = n\} &= p_n = p_0 \frac{\rho^n}{n!}, \quad 1 \leq n \leq s-1 \\ \mathbb{P}\{Z = s\} &= \sum_{n=0}^N p_{s+n} = p_0 \frac{\rho^s}{s!} \sum_{n=0}^N \frac{\left(\frac{\rho}{s}\right)^n}{\prod_{j=1}^n \left(1 + j \frac{\xi}{s}\right)}. \end{aligned}$$

Le nombre moyen de clients en service est égale au nombre moyen de serveurs occupés, qui est donné par :

$$\mathbb{E}[Z] = \sum_{n=1}^s np_{n+s} \sum_{n=1}^N p_{s+n}. \quad (3.3)$$

Quand :

$\lambda p_{n-1} = n\mu p_n$ pour $1 \leq n \leq s$, on a :

$$\mu \sum_{n=1}^s np_n = \lambda \sum_{n=1}^s p_{n-1} = \lambda \sum_{n=0}^{s-1} p_n.$$

par conséquent, on aura :

$$\mathbb{E}[Z] = \rho \sum_{n=1}^s p_{n-1} = \lambda \sum_{n=0}^{s-1} p_n. \quad (3.4)$$

Quand on trace $\mathbb{E}[Z]$ en fonction de ρ , on a les limites suivantes :

$$\lim_{\rho \rightarrow 0} \mathbb{E}[Z] = 0; \quad \lim_{\rho \rightarrow \infty} \mathbb{E}[Z] = s.$$

3.2.5 Probabilité d'atteinte du service

Le débit du système noté λ' est défini comme étant le nombre moyen de clients servis par unité de temps. D'autre part, la probabilité qu'un service soit réussi, noté $\mathbb{P}\{Sr\}$ est donnée par la fraction d'arriver des clients que le client atteint son service. Par conséquent, on aura la relation suivante :

$$\lambda' = \lambda \mathbb{P}\{Sr\} = \mu \mathbb{E}[Z].$$

Par conséquent, on obtient :

$$\mathbb{P}\{Sr\} = \frac{\mathbb{E}[Z]}{\rho} = \sum_{n=0}^{s-1} p_n + \frac{s}{\rho} \sum_{n=1}^N p_{s+n}. \quad (3.5)$$

La probabilité qu'un client arbitraire soit servi est donnée par :

$$\mathbb{P}\{Sr/Nb\} = \frac{\mathbb{P}\{Sr\}}{1 - \mathbb{P}\{\mathcal{B}\}} = \frac{\mu \mathbb{E}[Z]}{\gamma L_q + \mu \mathbb{E}[Z]} \quad (3.6)$$

Qui a pour limite :

$$\lim_{\rho \rightarrow \infty} \mathbb{P}\{Sr \mid Nb\} = \frac{\mu s}{\gamma N + \mu s} = \frac{1}{1 + N \frac{\xi}{s}}.$$

3.2.6 Le rapport de blocage, d'abandon et de clients servis

Un client qui arrive trouve trois cas devant lui, soit il sera bloqué à l'arrivée avec une probabilité $\mathbb{P}\{\mathcal{BLOC}\}$, il entre mais abandonne avec probabilité $\mathbb{P}\{Ab\}$, où il sera servi avec succès avec la probabilité $\mathbb{P}\{Sr\}$:

$$\mathbb{P}\{\mathcal{BLOC}\} + \mathbb{P}\{Ab\} + \mathbb{P}\{Sr\} = 1.$$

On multiplie les deux côtés par λ , on obtient :

$$\lambda \mathbb{P}\{\mathcal{BLOC}\} + \gamma L_q + \mu \mathbb{E}[Z] = \lambda.$$

Sur le côté gauche, $\lambda \mathbb{P}\{\mathcal{BLOC}\}$ est le nombre moyen de clients qui sont bloqué à l'arrivée par unité de temps, γL_q le nombre moyen de clients qui quittent la file par unité de temps, et $\lambda' = \mu \mathbb{E}[Z]$ est le nombre moyen de clients qui sont servis avec succès par unité de temps.

En utilisant la condition de normalisation $\sum_{n=0}^{s+N} p_n = 1$, on obtient,

$$\begin{aligned} \lambda \mathbb{P}\{\mathcal{BLOC}\} + \mathbb{P}\{Ab\} &= 1 - \mathbb{P}\{Sr\} = 1 - \sum_{n=0}^{s-1} p_n - \frac{s}{\rho} \sum_{n=0}^N p_{s+n} \\ &= 1 - \sum_{n=0}^{s-1} p_n - \frac{s}{\rho} \left(1 - \sum_{n=0}^s p_n \right) \\ &= \left(1 - \frac{s}{\rho} \right) \left(1 - \sum_{n=0}^s p_n \right) + \frac{s}{\rho} p_s \\ &= \left(1 - \frac{s}{\rho} \right) \left(1 - \sum_{n=0}^s p_n \right) + p_{s-1} \end{aligned}$$

$$\lambda \mathbb{P}\{\mathcal{BLOC}\} + \mathbb{P}\{Ab\} = 1 - \mathbb{P}\{Sr\} = \left(1 - \frac{s}{\rho} \right) \left(1 - \sum_{n=0}^s p_n \right) + p_{s-1}. \quad (3.7)$$

qui est la probabilité qu'un client arbitraire est perdu, soit en étant bloqué à l'arrivée ou par l'abandon dans la file d'attente.

3.2.7 Nombre de clients dans le système vu par une arrivée acceptée

Soit N Nombre de clients dans le système vu par une arrivée acceptée. La distribution de \hat{N} est donnée par :

$$\hat{P}_n = P\hat{N} = n = \frac{P_n}{1 - P\{\mathbb{B}\}}, \quad 0 \leq n \leq s + N - 1.$$

Ce résultat peut s'écrire comme suit :

$$\hat{P}_n = \hat{P}_0 \frac{\rho^n}{n!}, \quad 1 \leq n \leq s.$$

$$\hat{P}_{s+n} = \hat{P}_0 \frac{\rho^s}{s!} \frac{\binom{\rho}{s}^n}{\prod_{j=1}^n (1 + j \frac{\rho}{s})}, \quad 1 \leq n \leq N - 1.$$

Avec,

$$\frac{1}{\hat{P}_0} = \frac{1 - P\{\mathbb{B}\}}{p_0} = \sum_{n=0}^{s-1} \frac{\rho^n}{n!} + \frac{\rho^s}{s!} \sum_{n=0}^{N-1} \frac{\binom{\rho}{s}^n}{\prod_{j=1}^n (1 + j \frac{\rho}{s})}.$$

3.2.8 Probabilité d'attente

On note W' le temps d'attente des clients qui sont acceptés soit servi ou abandonné dans la file d'attente. Plus tard, nous considérons séparément les temps d'attente pour les clients servis, les clients abandonnés et les clients restés dans la file d'attente. Un client arrivant qui est accepté doit attendre si tous les serveurs sont occupés à savoir s'il y a s clients ou plus dans le système immédiatement avant l'arrivée. Par conséquent, la probabilité d'attente d'un client arbitraire accepté est donné par :

$$\mathbb{P}\{W' > 0\} = \sum_{n=0}^{s-1} \hat{P}_{s+n} = \hat{P}_0 \frac{\rho^s}{s!} \sum_{n=0}^{N-1} \frac{\binom{\rho}{s}^n}{\prod_{j=0}^n (1 + j \frac{\rho}{s})}. \quad (3.8)$$

3.2.9 La durée moyenne d'attente

Il résulte des formules de Little appliquées aux clients admis que,

$$W = \frac{L_q}{\lambda(1 - \mathbb{P}\{\mathcal{B}\})} = \frac{\hat{P}_0 \rho^s}{\lambda s!} \sum_{n=1}^N \frac{n \left(\frac{\rho}{s}\right)^n}{\prod_{j=0}^n \left(1 + j \frac{\xi}{s}\right)}. \quad (3.9)$$

Par conséquent, nous avons la relation suivante :

$$\mathbb{P}\{Ab \mid Nb\} = \gamma W$$

$$\lim_{\rho \rightarrow \infty} W = \frac{N}{\gamma N + s\mu}.$$

Conclusion

Dans ce travail, nous avons étudié un modèle qu'est largement présent dans la vie courante. Dans le cas d'un impatient de clients, nous avons refait le cadre de la probabilité de perte, qui a la particularité d'être général, c'est-à-dire indépendant de la distribution qui caractérise le moment des arrivées, ce qui peut aider les administrateurs des systèmes d'information à choisir le bon type qui correspond au besoin du client. D'autre part, dans le cas de deux classes de clients strictement prioritaires, c'est-à-dire que la première classe a plus de priorité que la seconde, on s'intéresse au cas particulier où les temps de dépose et de service sont les mêmes pour les clients des deux classes, et on permet de profiter des résultats trouvés dans le cas d'une seule classe pour conclure ceux de la seconde classe, ce qui se répète par induction, afin de généraliser les résultats à n classes de clients. Signalons enfin que le cas sans préemption reste difficile à traiter car il nécessite une étude très profonde afin d'en explorer toutes les possibilités.

Bibliographie

- [1] A. Gomez-Corral and M.F. Ramalhoto. On the waiting time distribution and the busy period of a retrial queue with constant retrial rate. *Stochastic Modelling and Applications*, 3, 37-47, (2000).
- [2] Abdel-Karim Aboul-Hassan, Sherif I. Rabia and Ahmed Kadry, Analytical study of a discrete time retrial queue with balking customers and early arrival scheme, *Alexandria Engineering Journal*, 44 (2005), No. 6, 911-91.
- [3] Allen, A.O. 1990. *Probability, Statistics, and Queueing Theory with Computer Science Applications*. Second edition, Academic Press, New York (First edition :1978).
- [4] Agnès Lagnoux, Claudie Hassenforder :Processus stochastiques modélisation Responsable UE (lagnoux@univ-tlse2.fr) Conception polycopié.
- [5] Anisimov, V. V, Zakusilo, O. K, and Donchenko, V. S. 1987. *Elements of Queueing Theory and Asymptotic System Analysis*. Vishcha Shkola, Kiev (in Russian).
- [6] Baynat, B. *Théorie des files d'attente-des chaînes de Markove aux réseaux à forme produit*. Paris, Hermès Science Publications , 2000.
- [7] B. T. Doshi. *Queueing systems with vacations a survey*. *Queueing Systems Theory and Applications*, 1986.
- [8] B. Doshi, "Single server queues with vacations," in *Stochastic Analysis of Computer and Communication Systems*, H. Takag, Ed., pp. 217-265, Elsevier, 1990.
- [9] Boxma, O.J., Schlegel, S. and Yechiali, U., "A Note on the M/G/1 Queue with a Waiting Server, Timer and Vacations," *American Mathematical Society Translations, Series 2*, 207 (2002) 25 :35.

-
- [10] Cheprasov, V. P. Elements of Queueing Theory. Kazan Aviation Institute (in Russian)1985.
- [11] Chretienne. P. And Faure. R. Processus stochastique, leurs graphes, leurs usages. Gauthier villars, Paris, 1957.
- [12] Claudie Haseforder CHABRIAC , Eléments de Théorie des files d'attente, page05, Janvier 2008.
- [13] Chabriac. C. Processus stochastiques et modélisation. Université de Toulouse le Mirail, Master 2, Année 2012-2013.
- [14] D. Gross, J.F. Shortle, J.M. Thompson and C.M. Harris. Fundamentals of Queueing Theory. John Wiley and Sons, 2008.
- [15] D. Baum and L. Breuer . An Introduction to Queueing Theory and Matrix-Analytic Methods. Springer, 2005.
- [16] Eitan. Altman - Uri. Yechiali. Analysis of customers' impatience in queues with server vacations, Queueing Syst (2006) 52 :261-279.
- [17] Florin Avran, "Processus de Markov, de Levy, Files d'attente, Actuariat et Fiabilité", 2009-2010.
- [18] H. Takagi, Queueing Analysis : A Foundation of Performance Analysis, vol. 1 of Vacation and Priority Systems, part 1, Elsevier Science Publishers B.V., Amsterdam, The Netherlands, 1991.
- [19] Khintchine, A. Y 1969. Mathematical Methods in the Theory of Queueing. Second edition, Hafner Publishing Company, New York (First edition : Griffin,London, 1960 ; Russian original : 1955). 859-866.
- [20] Kashyap, B. R. K., and Chaudhry, M. L. An Introduction to Queueing Theory. Aarkay, Calcutta, India. 1988.
- [21] Lionel. B. Processus stochastique : Processus de poisson et chaîne de Markov. 2004.
- [22] M. Boualem, N. Djellab, D. Aissani, "An M/G/1 retrial queue with exhaustive service and server vacations", Journal of Communication and Computer 8 (2011) 720-726.
- [23] Moshe Zukerman. Introduction to Queueing Theory and Stochastic Teletraffic Models. EE Department. City University of Hong Kong.

- [24] Newell, G. F. 1982. Applications of Queueing Theory. Second edition, Chapman and Hall, London (First edition : 1971).
- [25] Nicolas Fournier, d'après Philippe Bougerol. Processus de Sauts et Files d'Attente. Université Pierre et Marie Curie. Master 1 de Mathématiques.
- [26] N. Tian and Z. G. Zhang, Vacation Queueing Models : Theory and Applications, Springer, New York, NY, USA, 2006.
- [27] Pn Philippe. N. Basic elements of queueing theory : application to medeling of computer systems. Tech report, The University of Massachusetts, 2004.
- [28] P. Robert. Réseaux et Files d'Attente : Méthodes Probabilistes. Springer-Verlag Berlin Heidelberg, 2000.
- [29] Philippe Malbos. Analyse Matricielle et Algèbre Linéaire Appliquée. université Claude Bernard Lyon 1.Licence Sciences, Technologies, Santé.
- [30] Raphaël. Danchin. Equations différentielles L3 de Mathématiques. Année 2010 - 2011.
- [31] Rugg. R. Processus stochastique. Presses Polytechniques Romandes, 1989.
- [32] Ruegg.A : processus stochastique, Presses polytechnique romandes, Lausanne Suisse,(1989).
- [33] S. Ross. Stochastic Processes. John-Wiley and Sons, New York, 2e éd. 1996.
- [34] Sheldon M. Ross. Introduction to Probability Models Tenth Edition. University of Southern California Los Angeles, California.
- [35] Sanjay K. Bose, "Analyzing the M/G/1 queue using the Methode of Supplementary Variables", 2002.
- [36] Sébastien Loustau. Chaînes de Markov et Processus markoviens de sauts. Applications aux files d'attente. Ecole Centrale de Marseille, Année 2008-2009.
- [37] T.L. Saaty. Elements of Queueing Theory. M.C. Graw-Hill, New York, 1961.
- [38] Takacs, L. "Introduction to the Theory of Queues," Oxford University Press, New York (1962).
- [39] U. Narayan Bhat. An Introduction to Queueing Theory, Modeling and Analysis in Applications. Second Edition.

-
- [40] Watson, G. S. and Leadbetter, M. R. Hazard Analysis I. *Biometrika*, 51, 175-184 (1964 a).
- [41] Watson, G. S. and Leadbetter, M. R. Hazard Analysis II. *Sankhya Ser. A*, 26, 101-116 (1964 b).
- [42] W.Grassmann, Modeling Markovien Queues and Similar Process, Departement of Computer Science, University of Sakatchewan, 2000.
- [43] Y. Levy and U. Yechiali, "Utilization of idle time in an M/G/1 queueing system," *Management Science*, vol. 22, no. 2, pp. 202- 211, 1975.
- [44] Zakhar Kabluchko, *Stochastic Processes (Stochastik II)*, University of Ulm Institute of Stochastics, (2013-2014).
- [45] Zhang, M., Hou, Z. : Performance analysis of M/G/1 queue with working vacations and vacation interruption. *J. Comput. Appl. Math.* 234(10), 2977-2985 (2010).