

# Réseaux de files d'attente

## CHAÎNE DE MARKOV

Nous considérons un système qui peut prendre différents états parmi un ensemble  $E$  :  $E = \{E_i, i = 1, 2, \dots\}$ .

Les changements d'état se produisent dans le cas des chaînes de Markov à des instants déterminés en un nombre dénombrable que nous noterons  $t_1, t_2, \dots, t_k, \dots$ . Nous appellerons  $P_n(t_k)$  la probabilité que, à la date  $t_k$ , le système soit dans l'état  $E_n$ .

On suppose que la transition de l'état  $E_i$  à l'état  $E_j$  ne dépend que de ces deux états. On note  $p_{ij}$  la probabilité qu'une telle transition ait lieu. À l'instant initial  $t_0$  on connaît les différentes probabilités initiales :  $P_n(t_0)$ . L'ensemble  $\{p_{ij}, P_n(t_0) \text{ et } E_i, i = 1, 2, \dots\}$  forme une chaîne de Markov.

Si l'ensemble des états  $E$  est continu, la chaîne de Markov est continue. Si les dates de changement d'état forment un ensemble continu, on définit un processus de Markov. Nous nous intéresserons dans cette partie à un ensemble d'états  $E$ , fini ou dénombrable. Nous avons les équations suivantes :

$$P_n(t_k) = \sum_{i=1}^n p_{in}(t_k - t_{k-1}) P_i(t_{k-1})$$

La matrice des  $p_{ij}$  est une matrice stochastique puisque :

$$\sum_{j=1}^n p_{ij} = 1$$

### Exemple

Soit la chaîne de Markov de matrice de transition  $P$ . Le graphe correspondant est représenté à la figure C.1.



FIGURE C.1 • *Graphe de la matrice de transition P*

On voit sur cet exemple que, partant de n'importe quel point, on est sûr d'y revenir après un nombre de transitions fini. Les états sont dits récurrents. En revanche, dans l'exemple ci-dessous, certains états sont transitoires :



FIGURE C.2 • *Exemple avec état transitoire*

Dans cet exemple, l'état  $E_1$  est transitoire, les trois autres états sont récurrents. Si la matrice  $P$  peut se mettre, éventuellement après une nouvelle numérotation des états, sous la forme suivante :



où  $P_1$  et  $P_2$  sont des sous-matrices carrées, la chaîne de Markov est dite réductible, dans le cas contraire, elle est dite irréductible. Dans une chaîne de Markov réductible, si l'on se trouve dans l'un des états concernés par  $P_1$ , il n'est pas possible d'aller vers  $P_2$  et réciproquement. Dans le premier exemple, la chaîne de Markov était irréductible, dans l'exemple suivant, il y a trois ensembles récurrents.



Si la matrice  $P$  peut se mettre sous la forme :



partant d'un état de  $P_1$ , on peut aller vers un état de  $P_2$ . En revanche, la réciproque est fautive. Les états correspondant à  $P_1$  sont transitoires, ceux correspondant à  $P_2$  sont récurrents. La chaîne de Markov associée est réductible.

Lorsque la matrice  $P$  se présente sous la forme :



la chaîne de Markov est dite périodique. Dans le cas contraire, elle est apériodique.

Si le nombre d'états est fini, une chaîne de Markov apériodique et irréductible est ergodique, c'est-à-dire stable : on peut calculer pour chaque état la probabilité de s'y trouver. On dit encore qu'il existe un régime permanent. Lorsque le nombre d'états est infini, noté, 0, 1, 2... une chaîne de Markov apériodique et irréductible peut ne pas être ergodique, c'est-à-dire s'éloigner de plus en plus de l'état 0 pour ne plus y revenir.

L'équation de Markov que nous avons introduite  $P_j(t_{k+1}) = \sum p_{ij} P_i(t_k)$  peut maintenant s'écrire sous une forme matricielle :  $P_j(t_{k+1}) = P(t_1) ? P$  où  $P(t_k)$  est le vecteur des probabilités d'état.

À partir de l'état initial, nous obtenons :

$$P(t_k) = P(t_0) P_k.$$

La chaîne de Markov est ergodique s'il existe une limite, lorsque  $k$  tend vers l'infini, de  $P_k$ , soit :

$$P^* = \lim P_k$$

Le régime permanent est alors obtenu par :

$$P(t_\infty) = P(t_0) P^*$$

Le régime permanent dépend de l'état initial. Par exemple, dans le cas où la chaîne de Markov est réductible et possède plusieurs sous-chaînes irréductibles, l'état initial est important. En revanche, si la chaîne de Markov est irréductible et apériodique, la matrice  $P$  a toutes ses lignes identiques et la multiplication de matrices du deuxième membre est indépendante de l'état initial et :

$$P(t_\infty) = (P_0, P_1 \dots)$$

où  $P_0, P_1 \dots$  est une ligne de la matrice  $P^*$ .

Grâce aux dérivations précédentes, nous pouvons calculer les probabilités d'état dans le régime permanent. En supposant le nombre d'états fini, le vecteur  $P^\circ = (P_0, P_1 \dots P_k)$  est obtenu comme solution du système :

$$P^\circ = P^\circ P^*$$

Si  $P^{**} = P^* - I$ , il faut obtenir la solution de  $P^\circ P^{**} = 0$

## FILE D'ATTENTE ET RÉSEAU DE FILES D'ATTENTE

Une file d'attente peut être représentée par le schéma suivant :



FIGURE C.3 • Une file d'attente

Un client arrive suivant un certain processus  $A$ . Si le guichet de service est libre, le client y entre et il est servi avec un processus  $S$ . S'il n'est pas seul dans la file d'attente à son arrivée, il lui est appliqué une discipline de service  $DS$ . Les plus classiques sont FIFO (First In First Out) ou PAPS (Premier Arrivé, Premier Sorti) en français. Cette discipline peut être différente de PAPS' (Premier Arrivé, Premier Servi) s'il y a plusieurs guichets de service. Beaucoup d'autres disciplines de service existent : en particulier, LCFS (Last Come, First Served) ou DAPS (Dernier Arrivé, Premier Servi), aléatoire... Le nombre de guichets n'est pas obligatoirement unique, il peut y en avoir  $C$ . La file d'attente peut ne recevoir qu'un nombre fini de clients que nous noterons ici  $K$ . Si un client se présente lorsque la file est pleine, il est refusé. Enfin, notons  $L$  le nombre total de clients qui peuvent emprunter la file d'attente. Par exemple, dans un système informatique transactionnel où  $L$  terminaux sont connectés, il ne peut pas y avoir plus de  $L$  clients dans la file d'attente. Ou encore, si les clients sont dans la file d'attente, il ne peut y avoir d'arrivées.

Nous noterons une file d'attente par :  $A/S/C (DS/K/L)$  ou  $A/S/C/K/L (DS)$ .

Les lettres les plus classiquement utilisées sont :

- M pour markovien (ou poissonien ou exponentiel),
- D pour constant (déterministe),
- $E_k$  pour Erlang k,
- H pour hyperexponentiel.

Nous verrons dans le paragraphe suivant les définitions exactes de ces sigles.

Si les valeurs entre parenthèses ne sont pas explicitées, les valeurs à prendre par défaut sont (FIFO, \_, \_).

Une file M/M/1 est donc une file avec un processus de Markov en entrée et en sortie, un seul serveur, une discipline de service premier arrivé, premier servi, une capacité infinie et un nombre infini de clients qui peuvent entrer dans cette file.

Pour arriver à modéliser des systèmes beaucoup plus complexes, une file d'attente simple n'est pas suffisante, il faut faire appel à des réseaux de files d'attente. Deux catégories de réseaux seront étudiées : les réseaux ouverts qui possèdent une (ou plusieurs) entrées de l'extérieur et les réseaux fermés qui, au contraire, n'ont pas d'entrées de l'extérieur. Le nombre de clients est dans ce cas fixe, c'est-à-dire constant. Lorsque plusieurs classes de clients sont définies, si le réseau est ouvert pour certaines classes et fermé pour d'autres, le réseau est alors mixte.

## Réseau ouvert

---

Soit le réseau représenté figure C.4, contenant N files d'attente. La station 0 est une source et la station N + 1 un puits. Il faut définir les probabilités de routage. Soit  $p_{ij}$  la probabilité qu'un client sortant de la file i se dirige vers la file j. Nous avons :

$$P_{ij} = 0, 1, 2, 3 \dots N ; j = 1, 2, \dots N + 1$$



Les  $p_{ij}$  définissent une chaîne de Markov, c'est-à-dire que la probabilité d'aller vers j en sortant de la file i ne dépend pas des files déjà traversées.

## Réseau fermé

---

Dans le cas d'un réseau fermé, les files 0 et N + 1 n'existent pas et peuvent être supprimées. Lorsqu'il y a plusieurs classes de clients  $r = 1 \dots, R$  il faut définir des probabilités de routage en tenant compte de la classe du client. On définit  $p_{ir, jr}$ , la probabilité d'aller de la file i en classe r dans la file j en classe r'.

Pour les réseaux ouverts, il faut aussi déterminer les arrivées :  $p_{0, jr}$ , la probabilité lorsque l'on arrive de l'extérieur, d'entrer dans la file j en classe r' et  $p_{ir, N+1}$  la probabilité qu'un client sortant de la file i en classe r quitte le système.



FIGURE C.4 • Un réseau de files d'attente ouvert

## LES PROCESSUS D'ENTRÉE ET DE SERVICE

Nous ne nous servirons dans la suite que de processus de renouvellement : les intervalles de temps entre deux arrivées ou les longueurs des services n'ont aucune corrélation entre eux. On retire un nouveau temps à chaque arrivée et à chaque début de service, indépendant de la valeur du précédent tirage.

On note GI, une distribution générale suivant un processus de renouvellement et G le cas le plus général. Par exemple, pour une suite d'intervalles entre arrivées  $A_1, A_2, A_3, \dots$  tirée suivant une distribution  $f$ , nous avons :



C'est un processus GI suivant la distribution  $f$ . En revanche, si l'on prend la suite suivante :



où un intervalle est répété deux fois de suite, de telle sorte qu'il y a évidemment corrélation entre deux arrivées successives  $A_i, A_{i+1}$ , nous avons un processus d'arrivée  $G$  suivant la distribution  $f$ .

## Processus de Poisson et loi exponentielle

---

Un processus de Poisson correspond à des événements aléatoires dans le temps. Soit  $t_i$  les dates de ces événements, et soit :



un intervalle quelconque de durée  $t$ . Nous cherchons à préciser le nombre d'événements aléatoires qui a lieu durant ce laps de temps. Des événements aléatoires sont déterminés par le fait que dans un intervalle  $\Delta t$  infiniment petit, la probabilité qu'il y ait un événement est égale à  $\lambda \Delta t$ , où  $\lambda$  est le taux d'arrivée des événements. On suppose que dans le temps  $\Delta t$ , il ne peut y avoir au plus qu'un événement. Il est clair que les instants d'arrivées forment une chaîne de Markov puisque, à partir d'un temps  $t_i$ , la probabilité d'un nouvel événement ne dépendra pas du passé avant  $t_i$ .

De plus, si l'on décompose le temps en intervalles  $\Delta t$ , les différents points forment aussi une chaîne de Markov puisque le futur ne dépend que de ces points. Nous pouvons écrire l'équation de Markov à l'un de ces instants  $t_n$ , où  $P_j(t_n)$  est la probabilité d'être dans l'état  $j$  au temps  $t_n$  :



qui ici se décompose en :

$$P_j(t_n) = (1 - \lambda \Delta t) P_j(t_{n-1}) + \lambda \Delta t P_{j-1}(t_{n-1})$$

ce qui signifie que soit dans l'intervalle  $\Delta t$  il n'y a pas eu d'événement et la chaîne était déjà dans l'état  $j$  au temps  $t_{n-1}$ , soit il y a eu un événement, avec probabilité  $\lambda \Delta t$  et la chaîne était dans l'état  $j-1$  au temps  $t_{n-1}$ . Cette dernière équation peut encore s'écrire :



La solution de ces équations donne le nombre d'événements dans le laps de temps  $t$ . Soit  $P_n(t)$  la probabilité qu'il y ait  $n$  événements pendant le temps  $t$ . Nous obtenons :

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

C'est la distribution de Poisson de paramètre  $\lambda$ .

Nous pouvons également nous intéresser au temps qui s'écoule entre deux événements successifs. Soit  $A$  cette variable aléatoire, nous avons :

$$\text{Prob}(A = \theta) = \lambda e^{-\lambda \theta}$$

La distribution exponentielle de taux  $\lambda$  se présente sous la forme :

$$f(t) = \lambda e^{-\lambda t}$$

La fonction de répartition associée est  $F(t) = 1 - e^{-\lambda t}$ . La moyenne vaut  $1/\lambda$  et la variance  $1/\lambda^2$  de telle sorte que le carré du coefficient de variation vaut :

$$1$$

## Distribution Erlang $k$

---

La loi d'Erlang peut être représentée par une suite de lois exponentielles. En particulier, la distribution Erlang  $k$  de taux  $\lambda$  est une somme de  $k$  distributions exponentielles de taux  $\lambda/k$ . C'est un produit de convolution de  $k$  distributions exponentielles ou encore le produit des transformées de Laplace. On obtient la distribution :

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

$$F(t) = 1 - \sum_{n=0}^{k-1} \frac{(\lambda t)^n e^{-\lambda t}}{n!}$$

La fonction de répartition associée est :

$$F(t) = 1 - \sum_{n=0}^{k-1} \frac{(\lambda t)^n e^{-\lambda t}}{n!}$$

La variance de la distribution Erlang  $k$  vaut  $1/(k \lambda^2)$  de telle sorte que le carré du coefficient de variation vaut :

## Distribution hyperexponentielle

---

La distribution hyperexponentielle peut être représentée par un serveur parallèle de stations exponentielles.

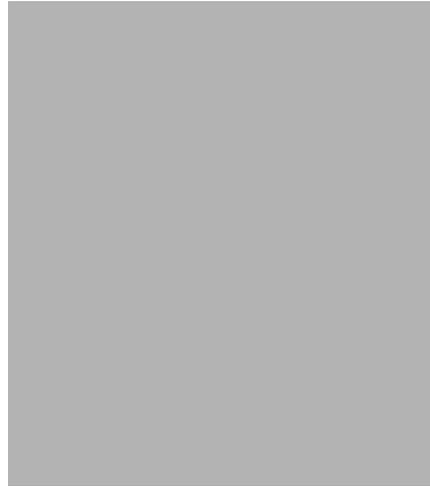


FIGURE C.5 • *La loi hyperexponentielle*

Avec la probabilité  $a_i$  l'intervalle de temps entre deux événements sera une loi exponentielle de paramètre  $\lambda_i$ . La distribution se présente donc sous la forme :



Le taux de la loi hyperexponentielle vaut :



La variance :



Le carré du coefficient de variation que nous noterons  $CCV$  vaut dans ce cas :



Le carré du coefficient de variation vérifie l'inégalité  $CCV \leq 1$ .

Il faut noter que pour une valeur du carré du coefficient de variation supérieure à 1, on peut trouver plusieurs distributions hyperexponentielles correspondantes. Une distribution hyperexponentielle doit être déterminée par tous ses moments où l'ensemble des serveurs parallèles.

## Comparaison des lois exponentielles, Erlang $k$ et hyperexponentielles

---

Cette comparaison peut s'effectuer en examinant les carrés de coefficient de variation qui peuvent être schématisés selon la figure C.6.



FIGURE C.6 • *Comparaison des lois*

Il faut encore noter que les lois Erlang  $k$  ne permettent pas de recouvrir l'ensemble des carrés de coefficient de variation entre 0 et 1. De plus, les distributions que nous avons vues ne sont pas denses dans l'ensemble des distributions, c'est-à-dire que l'on ne peut pas approcher une loi quelconque aussi près que l'on veut. Pour cela, il existe les lois de Cox qui nous serviront ultérieurement.

## Loi de Cox

---

Le guichet de service se compose d'une suite de serveurs exponentiels, disposés comme dans la figure C.7.

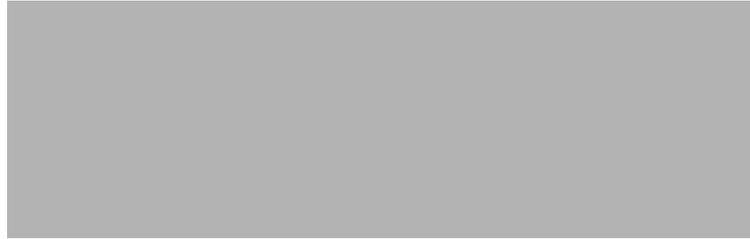


FIGURE C.7 • La loi de Cox

$L$  est le nombre de serveurs de la loi de Cox. Le temps de service de la  $m$ -ième étape, c'est-à-dire du  $m$ -ième serveur, est exponentiellement distribué de taux  $\mu_m$ . À chaque étape un client a la probabilité  $b_m$ ,  $m = 0 \dots, L - 1$  de continuer vers le serveur suivant et  $1 - b_m$  de sortir, c'est-à-dire de terminer le service global.

Si  $A_m = b_0 b_1 \dots b_{m-1}$ , la transformée de Laplace de la loi de Cox est :



On peut exprimer cela par :  $f(t)$  est la convolution des différentes lois exponentielles avec un poids correspondant aux probabilités vers la sortie du serveur. Si le client passe directement de l'entrée vers la sortie avec la probabilité  $1 - b_0$ , la densité de service est une fonction de Dirac.

Les distributions de Cox sont identiques aux distributions qui possèdent une transformée de Laplace rationnelle. En particulier, les distributions d'Erlang et les distributions hyperexponentielles sont des distributions de Cox. De plus, une somme de distributions de Cox est une distribution de Cox.

Une des propriétés importantes des distributions de Cox est d'être dense dans l'ensemble des distributions, c'est-à-dire que toute distribution peut être approchée d'autant près que l'on veut par une distribution de Cox. Ce résultat permettra, dans tous les cas que nous pourrions trouver en modélisation, d'étendre un théorème en distribution générale indépendante lorsqu'il est vrai pour les distributions de Cox. Ce sera en particulier le cas pour les théorèmes BCMP que nous verrons plus loin.

## PROCESSUS DE NAISSANCE ET DE MORT ET FILES ÉLÉMENTAIRES

Un processus de naissance et de mort est une chaîne de Markov pour laquelle, à partir d'un état  $n$ , seuls deux états sont accessibles : les états  $n + 1$  et  $n - 1$ . Soit  $\lambda_n$  et  $\mu_n$ , les

taux de naissance et de mort à partir de l'état  $n$ , c'est-à-dire que la chaîne de Markov peut se présenter comme dans la figure C.8.



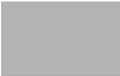
FIGURE C.8 • *La chaîne de Markov*

Nous avons l'équation de Markov pour  $n > 0$  :

$$(\lambda_n + \mu_n) P(n) = \lambda_{n-1} P(n-1) + \mu_{n+1} P(n+1)$$

$$\text{et } \lambda_0 P(0) = \mu_1 P(1)$$

Cette chaîne est ergodique si :



c'est-à-dire si au moins à partir d'un certain état, la chaîne a tendance à revenir vers le point 0. Si cette condition limite est vérifiée, on obtient :



Après normalisation, nous avons la solution générale :



## La file M/M/1

La file M/M/1 est un processus de naissance et de mort mais avec des coefficients constants :

$$\lambda_i = \lambda \quad i = 0, 1, \dots$$

$$\mu_j = \mu \quad j = 1, 2, \dots$$

Nous obtenons la solution générale en posant  $\rho = \lambda/\mu$ ,  $\rho$  étant l'intensité de trafic, et en supposant que  $\rho < 1$  pour que la chaîne de Markov soit ergodique :  $P(n) = \rho^n(1 - \rho)$ .

Le nombre moyen de clients est obtenu facilement en calculant l'espérance de cette probabilité :

$$\sum_{n=0}^{\infty} n P(n) = \sum_{n=0}^{\infty} n \rho^n (1 - \rho) = \rho \sum_{n=1}^{\infty} n \rho^{n-1} (1 - \rho) = \rho \sum_{n=1}^{\infty} n \rho^{n-1} (1 - \rho) = \rho \frac{1}{(1 - \rho)^2} (1 - \rho) = \frac{\rho}{1 - \rho}$$

Pour obtenir le temps d'attente moyen, il faut passer par l'intermédiaire des transformées de Laplace. La transformée de Laplace de la loi du temps de service s'écrit :

$$\int_0^{\infty} e^{-st} \mu e^{-\mu t} dt = \frac{\mu}{s + \mu}$$

La transformée de Laplace de la convolution de  $j$  telles distributions est :

$$\left( \frac{\mu}{s + \mu} \right)^j$$

La transformée de Laplace du temps de réponse est donc :

$$\frac{\rho}{1 - \rho} \left( \frac{\mu}{s + \mu} \right)^j + \frac{1 - \rho}{s + \mu}$$

Mais cela n'est pas autre chose que la transformée de Laplace de la distribution  $(\mu - \lambda) e^{-(\mu - \lambda)t}$ . La distribution du temps de réponse est exponentielle de taux  $\mu - \lambda$ . La moyenne est obtenue par :

$$\frac{\rho}{1 - \rho} \frac{1}{\mu - \lambda} + \frac{1 - \rho}{\mu - \lambda}$$

Il faut noter l'égalité entre  $N$  et  $T$  :

$$\frac{\rho}{1 - \rho} \frac{1}{\mu - \lambda} + \frac{1 - \rho}{\mu - \lambda} = \frac{\rho}{1 - \rho} \frac{1}{\mu - \lambda} + \frac{1 - \rho}{\mu - \lambda}$$

qui est une formule qui est toujours vraie dans les files d'attente et qui a été démontrée par Little (formule de Little).

## La file M/M/1 avec dépendance de l'état

---

Nous examinerons le cas général où le processus de naissance et de mort n'est pas dégénéré. La file d'attente possède des arrivées de Poisson de paramètre  $\lambda_n$  lorsque le nombre de clients dans la file est de  $n$  et des temps de service exponentiellement distribués de taux  $\mu_n$  lorsqu'il y a  $n$  clients dans la file d'attente.

La solution est celle donnée par le processus de naissance et de mort que nous avons étudié précédemment :



Il faut noter que ce résultat, assez simple pour une file complexe, est obtenu grâce à la propriété sans mémoire de la loi exponentielle. En effet, si nous regardons le processus de service au moment d'une arrivée, le temps de service moyen change et passe de  $\mu_{n-1}$  à  $\mu_n$ . Comme le service est exponentiellement distribué, le temps de service qui s'est déjà écoulé n'a aucune influence sur le futur ; on peut donc recommencer le service avec un nouveau taux sans modifier le comportement de la file. Pour une autre loi que la loi exponentielle, il est très difficile d'exprimer le comportement d'une file avec dépendance de l'état : lorsqu'un nouveau client arrive, il faut décompter la partie de service déjà faite.

## La file M/M/c

---

Nous avons toujours un processus de naissance et de mort dans lequel les taux sont les suivants :

$$\begin{aligned} \lambda_i &= \lambda & i = 0, 1, \dots \\ \mu_1 &= \mu & \mu_2 = 2\mu \dots \mu_c = c\mu \\ \mu_i &= c\mu & \text{si } i > c \end{aligned}$$

La condition d'ergodicité de la chaîne de Markov est ici  $\lambda < c\mu$ . Si cette condition est vérifiée, nous obtenons :



où  $P(0)$  est obtenue par normalisation.

## La file M/M/ $\infty$

---

La file M/M/ $\infty$  possède un nombre de serveurs toujours supérieur au nombre de clients, de telle sorte qu'aucun client n'a d'attente. Le client rentre directement dans le service

et se présente à l'entrée de la file suivant un processus de Poisson de paramètre  $\lambda$ . Les temps de service sont tous indépendants et exponentiellement distribués de taux  $\mu$ . Le nombre de clients dans la file est un processus de naissance et de mort, de taux :

$$\lambda(i) = \lambda \text{ pour tout } i,$$

$$\mu(i) = i\mu \text{ pour tout } i.$$

La condition d'ergodicité du processus de naissance et de mort est toujours satisfaite puisqu'il n'y a pas d'attente et la solution à l'état stationnaire se calcule facilement :



ce qui donne :



On peut montrer que la file plus générale  $M/G/\infty$  a exactement la même solution. Le processus de service ne dépend que de son premier moment. Cette file d'attente  $M/G/\infty$  sera très utile pour modéliser les stations dont le nombre de guichets est suffisamment grand pour qu'il n'y ait jamais d'attente dans une file devant les serveurs.

Le nombre moyen de clients dans la file d'attente ou encore ce qui est équivalent, le nombre de serveurs occupés, est :

$$N = \rho$$

Il faut noter dans cette file que, étant toujours stable,  $\lambda$  peut être supérieur à  $\mu$ . Le nombre moyen de clients dans la file  $\rho$  peut bien sûr être supérieur à 1.

## La file M/M/c/c

---

Ce système est une file d'attente avec un processus d'arrivée de Poisson, un temps de service exponentiellement distribué, un nombre de guichets égal à  $c$  et une capacité limitée à  $c$ . C'est un système avec perte : lorsqu'un client arrive et trouve la file pleine, ce client est perdu. Le nombre de clients au temps  $t$  est un processus de naissance et de mort de paramètre :

$$\lambda(k) = \lambda \text{ si } k \leq c \text{ et } \lambda(k) = 0 \text{ si } k > c$$

C'est encore une file d'attente qui est toujours stable puisque la file ne peut dépasser  $c$  clients. On obtient :



d'où, si  $\rho = \lambda/\mu$  :



Cette formule a été démontrée par Erlang pour des réseaux téléphoniques. On a, en particulier, la formule qui donne la probabilité que les  $c$  serveurs soient occupés :



Il s'agit de la formule d'Erlang. C'est encore la probabilité qu'un commutateur téléphonique qui possède  $c$ -joncteurs, rejette des appels arrivant des abonnés.

### File M/M/c/m/K

---

Cette file d'attente contient  $c$  serveurs, la file est globalement limitée à  $m$  et le nombre de clients qui peuvent l'emprunter est  $K$ . Nous supposons que  $c \leq m \leq K$ . Dans ce cas, le taux des arrivées dépend du nombre de clients dans la file. En effet, les clients qui sont déjà en attente dans la file ne peuvent plus y arriver par l'extérieur. Nous avons donc un processus de Poisson de paramètre  $\lambda(K - n)$  où  $n$  est le nombre de clients présents dans le système. C'est de nouveau un processus de naissance et de mort de paramètre :



Ce système est toujours stable. On obtient la solution suivante :



si  $n > m$   $P(n) = 0$ .

La valeur de  $P(0)$  est obtenue par normalisation.

Quelques cas particuliers donnent lieu à des formules classiques que nous allons décrire.

### **La file M/M/c/K**

Nous obtenons :



### **La file M/M/1/K**



### **La file M/M/c/K**



## LES RÉSEAUX DE FILES D'ATTENTE

### Les réseaux de Jackson

---

Nous avons étudié dans les paragraphes précédents des systèmes simples où il n'y a qu'une seule file d'attente. Dans les réseaux de Jackson, nous prenons un modèle plus général constitué de plusieurs files d'attente. Les réseaux de Jackson sont soit ouverts, soit fermés. Ils sont formés uniquement de files d'attente avec des serveurs exponentiellement distribués, et dans le cas ouvert, de clients arrivant de l'extérieur suivant un processus de Poisson. Nous allons, dans un premier temps, nous intéresser aux réseaux ouverts et plus particulièrement à un réseau constitué de deux files d'attente en série comme dessiné dans la figure suivante :



FIGURE C.9 • Deux files d'attente en tandem

Les arrivées forment un processus de Poisson de paramètre  $\lambda$  et les temps de service sont exponentiellement distribués de taux  $\mu_1$  et  $\mu_2$ .

Nous pouvons écrire l'équation de balance globale pour ce système. Soit  $n_1$  et  $n_2$  les nombres de clients dans les files 1 et 2.

Si  $n_1$  et  $n_2 > 0$

$$(\lambda + \mu_1 + \mu_2) P(n_1, n_2) = \lambda P(n_1 - 1, n_2) + \mu_1 P(n_1 + 1, n_2 - 1) + \mu_2 P(n_1, n_2 + 1)$$

Cette équation peut être décomposée en sous-équations que nous appellerons équations de balance locale :

$$\lambda P(n_1, n_2) = \mu_2 P(n_1, n_2 + 1)$$

$$\mu_1 P(n_1, n_2) = \lambda P(n_1 - 1, n_2)$$

$$\mu_2 P(n_1, n_2) = \mu_1 P(n_1 + 1, n_2 - 1)$$

Les cas frontières s'écrivent :

$$(\lambda + \mu_1) P(n_1, 0) = \lambda P(n_1 - 1, 0) + \mu_2 P(n_1, 1)$$

$$(\lambda + \mu_1) P(0, n_2) = \mu_1 P(1, n_2 - 1) + \mu_2 P(0, n_2 + 1)$$

$$\lambda P(0,0) = \mu_2 P(0, 1).$$

La première de ces trois équations frontières peut se décomposer en deux équations de balance locale :

$$\lambda P(n_1, 0) = \mu_2 P(n_1, 1)$$

$$\mu_1 P(n_1, 0) = \lambda P(n_1 - 1, 0)$$

La seconde se décompose en :

$$\lambda P(0, n_2) = \mu_2 P(0, n_2 + 1)$$

$$\mu_2 P(0, n_2 - 1) = \mu_1 P(1, n_2 - 1)$$

Définissons plus formellement les équations de balance locale. Elles proviennent de l'égalité suivante : le flot de clients arrivant dans une station contenant  $n$  clients est égal au flot des clients sortant de cette même station en y laissant  $n$  clients. Ces équations de balance locale ne font intervenir que l'état d'une seule file d'attente à la fois.

Dans l'exemple précédent et dans le cas général,

$$\mu_1 P(n_1, n_2) = \lambda P(n_1 - 1, n_2)$$

correspond à l'équation de balance locale de la première file ;

$$\mu_2 P(n_1, n_2) = \mu_1 P(n_1 + 1, n_2 - 1)$$

correspond à l'équation de balance locale de la deuxième file ;

$$\lambda P(n_1, n_2) = \mu_2 P(n_1, n_2 + 1)$$

correspond à l'équation de balance locale d'une file externe que nous noterons 0 et qui peut être interprétée comme la source des clients et le puits des clients.

Si l'on trouve une solution à partir des équations de balance locale, cette solution sera également vraie pour les équations globales. Par contre, si en les utilisant aucune solution n'est trouvée, il faut revenir aux équations de balance globale.

À partir des équations de balance locale, on trouve facilement la solution :

$$P(n_1, n_2) = P_1(n_1) P_2(n_2)$$



C'est une solution en forme produit, c'est-à-dire le produit des solutions marginales ou encore le produit de la solution de chaque file d'attente.

Passons à un exemple de réseau de Jackson plus compliqué avant d'aborder le cas général. Soit le modèle décrit dans la figure C.10.



FIGURE C.10 • *Le modèle à l'étude*

Les clients entrent par la station 1 qui possède un guichet de service exponentiel de taux  $\mu_1$ . Lorsqu'un client sort, avec la probabilité  $(1 - p_1)$  il va vers l'extérieur (le puits) et avec la probabilité  $p_1$  vers la seconde file d'attente dotée également d'un serveur exponentiellement distribué mais de taux  $\mu_2$ . Lorsqu'un client finit son service, avec la probabilité  $(1 - p_2)$ , il part vers l'extérieur et avec la probabilité  $p_2$  il reboucle sur la première station.

Écrivons les équations de balance globale et de balance locale dans le cas où  $n_1$  et  $n_2$ , les nombres de clients dans les files 1 et 2 sont strictement supérieurs à 0.

$$\begin{aligned}
 (\lambda + \mu_1 + \mu_2) P(n_1, n_2) &= \lambda P(n_1 - 1, n_2) \\
 &\quad + \mu_1 p_1 P(n_1 + 1, n_2 - 1) \\
 &\quad + \mu_1 (1 - p_1) P(n_1 + 1, n_2) \\
 &\quad + \mu_2 p_2 P(n_1 - 1, n_2 + 1) \\
 &\quad + \mu_2 (1 - p_2) P(n_1, n_2 + 1)
 \end{aligned}$$

Les trois équations de balance locale s'écrivent :

$$\begin{aligned}
 \lambda P(n_1, n_2) &= \mu_1 (1 - p_1) P(n_1 + 1, n_2) + \mu_2 (1 - p_2) P(n_1, n_2 + 1) \\
 \mu_1 P(n_1, n_2) &= \lambda P(n_1 - 1, n_2) + \mu_2 p_2 P(n_1 - 1, n_2 + 1) \\
 \mu_2 P(n_1, n_2) &= \mu_1 p_1 P(n_1 + 1, n_2 - 1).
 \end{aligned}$$

La première équation correspond à la source/puits, la deuxième à la première file et la troisième à la deuxième file d'attente.

On peut de nouveau découper les équations de balance locale en des équations appelées balance de station et qui permettent d'obtenir des équations encore plus simples. Ici, cette décomposition supplémentaire n'est pas possible.

Si nous supposons que la solution est le produit des deux solutions marginales, il nous faut calculer les flux d'entrée  $\lambda_1$  et  $\lambda_2$  dans les deux stations. Nous avons les équations :

$$\begin{aligned}
 \lambda_1 &= \lambda + \lambda_2 p_2 \\
 \lambda_2 &= \lambda_1 p_1
 \end{aligned}$$



Si la solution est en forme produit, elle se présente sous la forme :

$$P(n_1, n_2) = \rho_1 n_1 (1 - \rho_1) \cdot \rho_2 n_2 (1 - \rho_2)$$





On peut vérifier facilement sur les équations de balance locale que cette solution est exacte.

### Réseau de Jackson ouvert

Soit le réseau de Jackson le plus général possible représenté dans la figure C.11.

Il y a  $N$  stations avec des guichets de service exponentiellement distribués de taux  $\mu(n_i)$ ,  $i = 1, \dots, N$ . Autrement dit, le taux de service dépend du nombre de clients, de telle sorte que des files du type  $\bullet/M/C_1$  peuvent composer le réseau. Le  $\bullet$  indique que le processus d'entrée n'est pas déterminé puisqu'il provient des flux des autres stations s'ajoutant, le cas échéant, à un flux provenant de l'extérieur. Les clients venant de la source 0 se dirigent vers la station  $i$  avec la probabilité  $p_{0i}$ . La source débitant suivant un processus de Poisson, le flot  $\lambda p_{0i}$  est lui-même un processus de Poisson. Nous avons :



Soit  $p_{ij}$ ,  $1 \leq j \leq N$ ,  $1 \leq i \leq N + 1$ , la probabilité constante qu'un client ayant terminé d'être servi par le serveur  $i$  se dirige vers la file  $j$  ou vers l'extérieur (le puits) dénommé station  $N + 1$ . Dans certains cas, on confond 0 et  $N + 1$ .



FIGURE C.11 • Un réseau de Jackson ouvert

Soit  $n = (n_1, \dots, n_N)$  le vecteur formé des nombres entiers  $n_i$  à chacun des serveurs. Soit encore  $a(n, i)$  un vecteur identique au vecteur  $n$  à l'exception de sa composante  $n_i$  qui est remplacée par  $(n_i + 1)$ , et soit  $b(n, i)$  un vecteur identique au vecteur  $n$ , à l'exception

de sa composante  $n_i$  qui est remplacée par  $(n_i - 1)$ . Enfin, le vecteur  $c(n, i, j)$  est identique à  $n$  à l'exception des composantes  $n_i$  et  $n_j$ , lesquelles sont remplacées par  $(n_i + 1)$  et  $(n_j - 1)$ .

Les équations de balance globale à l'état d'équilibre ou équation de Chapman-Kolmogorov s'écrivent facilement :



où  $\mu_i(n_i) = 0$  si l'entier  $n_i$  est nul et  $P(n) = 0$  si le vecteur  $n$  contient une composante négative. Ces équations peuvent se décomposer en équations de balance locale. Pour  $i = 1, \dots, N$ , nous avons :



Si nous supposons toujours que la solution à l'état d'équilibre est en forme produit, il nous faut calculer le taux d'arrivée  $\lambda_i$  dans la file  $i$ . Nous avons, pour  $i = 1, \dots, N$  :



Le flot des arrivées dans la file  $i$  est composé de ce qui arrive de l'extérieur et de ce qui provient des diverses stations. Cela nous donne  $N$  équations à  $N$  inconnues.

Posons  $\lambda_i = \lambda e_i$ . La valeur  $e_i$  peut être interprétée comme le nombre moyen de passages d'un client par la station  $i$ . Nous obtenons le nouveau système :



qui peut s'écrire sous forme matricielle :

$$e = p + eP$$

où  $p = (p_{01}, p_{02}, \dots, p_{0N})$ ,

et  $P = (p_{ij}, i = 1, \dots, N ; j = 1, \dots, N)$ .

Comme le réseau est ouvert, la matrice  $P$  possède des lignes indépendantes et  $e_i$  est complètement déterminé par les équations  $e = p + eP$ .

Nous noterons  $P_i(0)$  la probabilité marginale pour la file  $i$  d'être dans l'état vide et plus généralement  $P_i(n_i)$  la probabilité marginale d'avoir  $n_i$  clients dans la file  $i$ .

**Théorème**

Un réseau de Jackson ouvert possède une distribution de probabilité stationnaire, si elle existe, donnée par :

$$P_i = \frac{e_i}{\sum_{j=1}^N e_j} \prod_{k=1}^i \lambda_k$$

où  $e_i$ ,  $i = 1 \dots N$  est donné par la solution unique des équations :

$$e = p + eP \text{ et}$$

$$e_i = \lambda_i e_{i-1} + \mu_i e_{i+1}$$

La condition nécessaire et suffisante d'existence est  $P_i(0) > 0$  pour  $i = 1 \dots, N$ , c'est-à-dire pour  $i = 1 \dots, N : \limsup [\lambda e_i < \mu_i (\mu)]$ .

Lorsque les taux de service ne dépendent pas de l'état de la file d'attente, la solution est plus simple :

$$P_i = \frac{e_i}{\sum_{j=1}^N e_j} \prod_{k=1}^i \lambda_k$$

$$e_i = \lambda_i e_{i-1}$$

La condition nécessaire et suffisante de stabilité est :

$$\lambda e_i < \mu_i \text{ pour } i = 1 \dots, N.$$

La démonstration consiste à vérifier que la solution donnée dans le théorème est bien vérifiée par les équations de balance locale.

Des extensions sont possibles : par exemple, le taux des arrivées dans le réseau peut dépendre du nombre global de clients dans le réseau :

$$\lambda_i = \lambda_i(N)$$

Dans ce cas, la solution générale est donnée par :

$$P_i = \frac{e_i}{\sum_{j=1}^N e_j} \prod_{k=1}^i \lambda_k(N)$$

Une autre extension intéressante concerne la limitation de capacité du réseau. Soit  $K^*$  le nombre maximal de clients que le réseau peut absorber. Si une nouvelle arrivée se produit lorsque le réseau est saturé, le client arrivant est perdu. Cela revient à prendre  $\lambda(K^*) = 0$  dans la solution précédente ou, ce qui revient au même, à calculer la probabilité  $P(0)$  sur l'ensemble des états  $n$  tel que :

### Réseau de Jackson fermé

Nous considérons toujours un système de  $N$  files d'attente mais les stations 0 et  $N + 1$  sont reliées de telle sorte qu'aucun client ne peut ni entrer ni sortir. Il y a un nombre  $K$  constant de clients qui circulent entre les différentes stations. Comme  $p_{0i} = 0$ , pour calculer  $e_i$  nous avons à résoudre le système de  $N$  équations à  $N$  inconnues :

dont la notation matricielle est :  $e = eP$ .

Ce système n'a pas de solution unique puisque le déterminant de la matrice  $P$  est nul. La valeur de  $e_i$  ne peut être déterminée qu'à une constante près. En effet, le nombre moyen de passages d'un client par une station  $i$  est infini puisqu'il n'y a pas de sortie vers l'extérieur.

Si pour une station déterminée  $m$ , on suppose  $e_m = 1$ , on peut calculer en regard de cette valeur le nombre moyen de passages relatifs dans les autres stations. Ou encore, entre deux passages par la station  $m$ , on passera en moyenne  $e_i$  fois par les stations  $i = 1, \dots, N, i \neq m$ . De toute façon, la solution générale ne dépendra pas de la constante près qui est utilisée dans le calcul des  $e_i$ .

Les équations d'équilibre global à l'état stationnaire s'écrivent :

### *Théorème*

Un réseau de Jackson fermé possède une distribution de probabilité stationnaire donnée par :



La démonstration s'effectue toujours par vérification des équations de balance locale ou globale.

### Exemple d'application

Soit le réseau fermé de trois stations, représenté dans la figure C.12 :



FIGURE C.12 • Exemple de trois files en réseau fermé

Les temps de service sont exponentiellement distribués de taux  $\mu_1$ ,  $\mu_2$  et  $\mu_3$ . Soit encore  $K$  le nombre total de clients dans le réseau.

Appelons  $p_2$  (respectivement  $p_3$ ) la probabilité qu'un client qui vient d'être servi par le serveur 1 se dirige vers le serveur 2 (respectivement 3), et  $p_1 = 1 - (p_2 + p_3)$  la probabilité que ce client retourne directement dans la file d'attente de la station 1. La matrice  $P = (p_{ij})$  devient :



La solution du système  $e = eP$ , en prenant  $e_1 = 1$ , est :

$$e_1 = 1,$$

$$e_2 = p_2,$$

$$e_3 = p_3.$$

Comme le système est fermé, les hypothèses du théorème sont toujours satisfaites et nous obtenons :

$$\mu_1 n_1 = \mu_2 n_2 = \mu_3 n_3 = \dots = \mu_N n_N = G.$$

En multipliant l'expression de droite par  $\mu_1 n_1$ , nous avons :

$$\mu_1^2 n_1^2 = \mu_1 \mu_2 n_1 n_2 = \mu_1 \mu_3 n_1 n_3 = \dots = \mu_1 \mu_N n_1 n_N = G \mu_1 n_1.$$

La constante de normalisation  $G'$  ( $K$ ) s'obtient très aisément :

$$G' = \sum_{r=1}^R \sum_{n_1=0}^{\infty} \dots \sum_{n_N=0}^{\infty} \mu_1^{n_1} \mu_2^{n_2} \dots \mu_N^{n_N}.$$

Le taux d'occupation du serveur 1 peut encore s'écrire de la façon suivante :

$$\rho_1 = \frac{\mu_1 n_1}{G} = \frac{\mu_1 n_1}{\sum_{r=1}^R \sum_{n_1=0}^{\infty} \dots \sum_{n_N=0}^{\infty} \mu_1^{n_1} \mu_2^{n_2} \dots \mu_N^{n_N}}.$$

## RÉSEAUX BCMP (BASKETT, CHANDY, MUNTZ ET PALACIOS)

Les résultats obtenus pour les réseaux de Jackson peuvent être étendus dans des cadres plus généraux, et en particulier sur les réseaux BCMP (d'autres extensions sont développées dans le tome 2 : réseaux de Kelly, réseaux "insensibles").

Nous considérons toujours le réseau général représenté figure C.11. Mais, il peut maintenant y avoir des classes de clients. Soit  $r = 1 \dots, R$  les  $R$  classes de clients. Les probabilités d'un acheminement de clients à travers le réseau sont données par l'intermédiaire d'une chaîne de Markov de probabilité de transition :

$$P = (p_{ir, jr}, i = 0 \dots, N \text{ et } j = 1 \dots, N + 1, r = 1 \dots, R \text{ et } r' = 1 \dots, R).$$

La quantité  $p_{ir, jr}$  exprime la probabilité qu'un client de classe  $r$  dans la station  $i$  soit dirigé vers la file  $j$  dans la classe  $r'$ . Les valeurs  $p_{0r, jr}$ , et  $p_{ir, (N+1)r}$  sont respectivement les probabilités qu'un client de source  $r$  entre dans la file  $j$  en classe  $r'$  et qu'un client de la file  $i$  sorte vers l'extérieur dans le puits avec la classe  $r'$ . Il est également possible d'avoir une source et un puits unique pour l'ensemble des classes de clients

qui s'exprimeraient par des probabilités de transition  $p_{0, jr}$  et  $p_{ir, N+1}$ , mais cette présentation est moins générale, car elle ne permet pas de différencier des flots d'arrivée du type  $\lambda_r$  qui eux-mêmes peuvent dépendre du nombre de clients de classe  $r$  à l'intérieur du réseau. Nous allons y revenir.

Les clients peuvent changer de classe au fur et à mesure qu'ils avancent dans le réseau. Nous allons, dans un premier temps, différencier les classes de clients qui appartiennent à une même chaîne : soit la relation d'équivalence  $r \sim r'$  :  $r$  et  $r'$  appartiennent à la même classe d'équivalence appelée chaîne si un client de classe  $r$  peut à un instant ultérieur quelconque, devenir un client de la classe  $r'$ . Nous allons avoir une partition des classes de clients en  $I$  chaînes.

### Exemple

Les clients de classe 1 entrent dans la première file d'attente, ils rebouclent en classe 2 et ils vont vers la deuxième file en classe 3. Les clients de classe 3 sortent ensuite vers l'extérieur. Ce modèle permet de faire passer exactement deux fois chaque client dans une file d'attente. Les classes 1, 2 et 3 forment une chaîne. Les clients de classe 4 entrent dans la première file d'attente et ces clients sont envoyés automatiquement vers la deuxième file dans laquelle ils rentrent en classe 5. (Il aurait été équivalent de supposer que les clients de classe 4 entrent dans la deuxième file en classe 3, si l'on n'a pas besoin de distinguer les clients de cette deuxième file). Les clients de classe 4 et 5 forment une deuxième chaîne.



FIGURE C.13 • Exemple de chaîne

Dans le cas d'un réseau BCMP ouvert, le processus des arrivées de nouveaux clients dans le réseau suit une loi de Poisson avec deux types de paramètres possibles :

- $\lambda(K)$  lorsqu'il y a globalement  $K$  clients dans le réseau,
- $\lambda_i(K_i)$  lorsqu'il y a  $K_i$  clients de la chaîne  $i$  dans le réseau.

Si l'on définit des arrivées par classes de la forme  $\lambda_r(K_r)$ , il faut nécessairement se ramener à des chaînes ; si l'on ne peut pas, il n'y aura pas de solution du type BCMP.

Dans la suite, lorsque l'on écrit  $p_{0r, jr}$ , cela désigne un client provenant de la source de la chaîne à laquelle la classe  $r$  appartient. On définit également des chaînes pour les réseaux fermés. Lorsque certaines chaînes sont ouvertes et d'autres fermées, le réseau

est dit mixte.

Désignons par  $e_{ir}$  la solution du système :



La quantité  $e_{ir}$  donne le nombre moyen de passages d'un client de classe  $r$  dans la file  $i$ . Définissons encore les types de files d'attente qui sont acceptables pour le réseau BCMP. Il y en a quatre :

- *type 1* : la discipline de service est premier arrivé – premier sorti (FIFO), la station a un seul serveur et le temps de service est exponentiellement distribué de même moyenne pour toutes les classes de clients. Si la station  $i$  possède un tel serveur, nous noterons par  $\mu_i$  ( $n_i$ ) le taux de service, s'il y a  $n_i$  clients dans la station ( $y$  compris celui qui est au service).
- *type 2* : la discipline est temps partagé (PS processor sharing) ; un client reçoit  $\theta$  fraction de seconde de service par seconde et revient à la dernière place de la file d'attente si le service n'est pas terminé durant cet intervalle de temps. Ce quantum  $\theta$  doit tendre vers 0 pour que la discipline de service temps partagé soit effective. La distribution du temps de service peut être une loi de Cox, distincte pour chaque classe de client.
- *type 3* : le nombre de serveurs de la file d'attente est suffisant pour qu'il y en ait toujours un de libre. Cela se traduit par la disponibilité immédiate d'un serveur lorsqu'un client se présente dans la file. Les distributions de service sont des lois de Cox, distinctes pour chaque classe de clients.
- *type 4* : la discipline de service est dernier arrivé – premier servi avec une priorité préemptive : lorsqu'un client arrive dans la station, il va directement au service et celui qui y était revient en tête de la file d'attente et reprendra son service là où il l'a laissé. De nouveau, la distribution de service est une loi de Cox, distincte pour les différentes classes de clients.

Pour définir l'état du système, il faut non seulement connaître le nombre de clients de chaque classe, mais également la position des clients dans la file d'attente. Soit très généralement :

$$n = (n_1, \dots, n_N)$$

le vecteur d'état du réseau définissant chaque station.

Lorsque la solution ne nécessite pas de connaître les classes de clients,  $n_i$  sera le nombre de clients dans la file  $i$  et pourra être décomposé en  $n_{ir}$  avec  $n_i = \sum n_{ir}$ , où  $n_{ir}$  est le nombre de clients en classe  $r$  dans la station  $i$ . En revanche, pour avoir une connaissance plus précise de l'état de la station, nous définissons :

- *type 1*  
 $n_i = (r_{i1}, \dots, r_{in_i})$   
 où  $r_{ij}$  est la classe du  $j$ -ième client en attente dans la station  $i$  dans l'ordre premier arrivé - premier servi (FIFO).
- *type 2 ou 3*  
 $n_i = ((r_{i1}, s_{i1}), \dots, (r_{in_i}, s_{in_i}))$

où  $r_{ij}$  est la classe du  $j$ -ième client en attente suivant l'ordre d'arrivée et  $s_{ij}$  l'étape du modèle de Cox.

- *type 4*

$$n_i = ((r_{i1}, s_{i1}), \dots, (r_{in_i}, s_{in_i}))$$

où  $r_{ij}$  et  $s_{ij}$  sont identiques à ceux définis dans les types 2 et 3 mais l'ordre des  $r_{ij}$  est défini par la discipline de service dernier arrivé - premier servi avec priorité absolue.

Soit encore :



ou bien, dans le cas de taux d'entrée dépendant de la chaîne :



Définissons les paramètres de la loi de Cox. La quantité  $l_{ir}$  est le nombre total de serveurs de la loi de Cox pour un client de classe  $r$  dans la station  $i$  ; la  $m$ -ième étape sera notée  $\mu_{irm}$  et  $A_{irm}$  sera la probabilité pour qu'un client de classe  $r$  dans la station  $i$  atteigne la  $m$ -ième étape. Enfin,  $n_{irm}$  est le nombre de clients à la  $m$ -ième étape de classe  $r$  dans le serveur  $i$ .

*Théorème*

La solution à l'état stationnaire (s'il existe), d'un réseau BCMP est donnée par :



Le théorème est trop général pour être applicable facilement. Si l'on s'intéresse aux nombres de clients dans chaque file, nous obtenons le théorème simplifié :



Cela peut encore se simplifier si nous ne nous intéressons qu'au nombre global de clients dans chaque file, sans dépendance du taux des arrivées.



où  $\rho_i$  est le taux d'utilisation du serveur de la file  $i$ , donné par :



La démonstration de ces théorèmes s'effectue en vérifiant que les équations de balance locales sont bien vérifiées.



## RÉFÉRENCES

Livre d'introduction aux techniques de modélisation.

S. FDIDA, G. PUJOLLE – *Modèles de systèmes et de réseaux*, tome I – *Performance*, Eyrolles, Paris, avril 1989.

Livre d'introduction à l'évaluation de performance.

D. FERRARI – *Computer systems performance evaluation*, Prentice-Hall, Englewood Cliffs, 1978.

Livre introductif sur les réseaux de files d'attente.

E. GELENBE, G. PUJOLLE – *Introduction aux réseaux de files d'attente*, Eyrolles,

Paris, 1982.

Livre assez semblable au précédent mais fortement complété par les «G-Networks» et les réseaux de files d'attente à temps discret.

E. GELENBE, G. PUJOLLE – *Introduction to queueing networks*, Wiley, 1999.

Livre sur l'évaluation de performance des réseaux locaux.

J. HAMMOND, R. O'REILLY – *Performance analysis of local computer networks*, Addison-Wesley, Reading, MA, 1986.

Article décrivant la simulation d'un réseau Ethernet.

H.D. HUGHES, L. LI – *Simulation model of an Ethernet computer performance*, vol. 3, 4, pp. 210-217, décembre 1982.

Livre en deux tomes qui a été durant de nombreuses années la référence sur les systèmes de files d'attente.

L. KLEINROCK – *Queueing systems*. Volume 1 : *Theory*, John Wiley, New York, 1975.

L. KLEINROCK – *Queueing systems*. Volume 2 : *Computer applications*, John Wiley, New York, 1976.

Excellente synthèse des systèmes de files d'attente pour la modélisation des systèmes de communication.

H. KOBAYASHI, A.C. KONHEIM – *Queueing models for computer communications system analysis*, *IEEE Transactions on Communications*, vol. 25, pp. 2-29, 1977.

Livre très complet sur l'analyse de performance.

E. LAZOWSKA, J. ZAHORJAN, G. GRAHAM, K. SEVCIK – *Quantitative system performance : computer system analysis using queueing network models*, Prentice-Hall, Englewood Cliffs, 1984.

Livre sur les réseaux de files d'attente avec des applications au dimensionnement des systèmes de communication.

G. PUJOLLE, S. FDIDA – *Modèles de systèmes et de réseaux*, tome II – *Files d'attente*, Eyrolles, Paris, avril 1989.

Un des livres de référence sur l'évaluation des réseaux de communication.

M. SCHWARTZ – *Computer communication network design and analysis*, Prentice-Hall, Englewood Cliffs, 1977.

Exemple de mesures de performance d'un réseau local Ethernet.

J.F. SHOCH, J.A. HUPP – *Measured performance of an Ethernet local network*, *C. ACM*, vol. 23, 12, pp. 711-721, décembre 1980.