

CAHIER DU LAMSADE

Laboratoire d'Analyse et Modélisation de Systèmes pour l'Aide à la Décision
(Université Paris-Dauphine)
Unité de Recherche Associée au CNRS n° 825

RÉSOLUTION ANALYTIQUE DE FILE D'ATTENTE AVEC ARRIVÉES GÉNÉRÉES PAR PARALLÉLISATION

CAHIER N° 148
septembre 1997

Danielle MAILLES ¹
Jalal KADAoui-ABBASSI ¹

reçu : juin 1997.

¹ LAMSADE, Université Paris-Dauphine, Place du Maréchal De Lattre de Tassigny, 75775 Paris Cedex 16 (e-mail: {maillès,kadaoui}@lamsade.dauphine.fr).

TABLE DES MATIÈRES

	<u>Pages</u>
Abstract	1
Résumé	1
1. Introduction	2
2. Etat de l'art	3
2.1 Réseaux parallèles	3
2.1.1 Modèle de dimension 2	3
2.1.2 Modèle de dimension K (bornes)	4
2.2 Réseaux séries-parallèles	5
2.3 Cas général	5
3. Modèle de base	6
3.1 Introduction	6
3.2 Etude du processus d'arrivée dans la file 0	7
3.3 Périodicité du processus d'arrivée dans la file commune	8
4. Résolution de la file commune	9
4.1 Distribution du nombre de clients aux instants de la chaîne incluse	10
4.2 Distribution du temps de réponse de la file 0	11
4.3 Distribution stationnaire de la file 0	11
5. Conclusion	12
Annexe	13
Bibliographie	15

ANALYTICAL SOLUTION OF A QUEING SYSTEM

WITH ARRIVAL GENERATED FROM A PARALLEL SYSTEM

ABSTRACT

This paper considers the fork-join problem which is very usefull in modelling parallel activities. Several works published on this subject are presented. Then we study a new model in which customers are created simultaneously and shared a common queue. We give the queue characteristics and the analytic expression of the response time distributions for the model under study. We conclude with the possible generalisation of the method when the states probabilities of the upstream network are known.

RESOLUTION ANALYTIQUE DE FILE D'ATTENTE

AVEC ARRIVEES GENEREES PAR PARALLELISATION

RESUME

Nous nous intéressons ici au problème du rendez-vous propre à modéliser des activités parallèles. Nous présentons différents travaux publiés sur ce sujet. Nous étudions ensuite un nouveau modèle dans lequel les clients générés en parallèle se retrouvent dans une file commune. Nous étudions les caractéristiques de cette file et en donnons la distribution du temps de réponse. Nous concluons en montrant que la méthode se généralise pour peu que soient connues les distributions stationnaires du réseau de files d'attente en amont de la file commune.

Mots clés: Files d'attente synchronisées, processus semi-markoviens, caractère périodique des modèles de « Fork ».

1- INTRODUCTION

Sous sa forme la plus générale, on peut définir le problème de rendez-vous de la façon suivante. Chaque client accédant au système, modélisé par un réseau de files d'attente, génère un certain nombre de fils qui sont dirigés vers le réseau. Les différents frères se synchroniseront à la sortie du réseau de manière à quitter le système simultanément.

Le modèle de files d'attente "descriptives" de la figure 1 décrit ce comportement. Outre le réseau de files d'attente, il comporte un objet dit de "dédoublément" et un objet dit de "regroupement". Ce symbolisme est basé sur l'exploitation des transitions utilisées dans les réseaux de Petri et a été introduit dans [Fdi 86], [Mai 87].

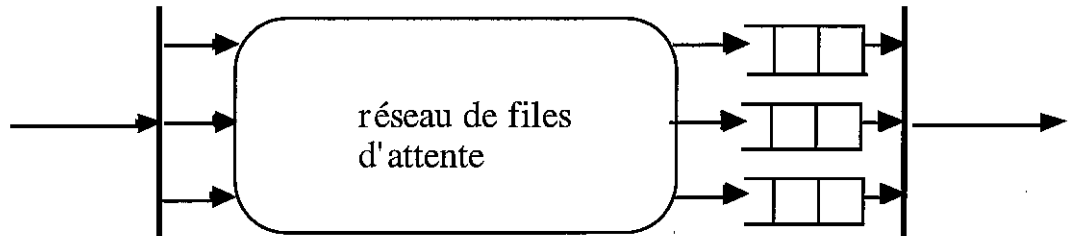


Figure 1 - Le modèle de rendez-vous

L'étude de ce modèle vise généralement à calculer le temps de réponse de bout en bout. Il se formalise de la façon suivante :

Notons : C_n le $n^{\text{ième}}$ père accédant au système
 C_n^k , $k=1, \dots, K$ les K fils générés (on considérera que chaque père engendre exactement K fils sauf spécification contraire)
 t_n^k de temps de traversée du réseau du client C_n^k

Dès lors le temps de traversée de C_n (noté t_n), qui comprend le temps de synchronisation, sera le temps de traversée du plus lent de ses fils, soit :

$$t_n = \text{Max} \{ t_n^k \mid 1 \leq k \leq K \}$$

Cette détermination n'est pratiquement pas possible dans la mesure où

1°) les distributions des temps de traversée des différents frères à travers un réseau (de type B.C.M.P par exemple) ne sont généralement pas connues

2°) les temps de traversée des frères ne sont pas indépendants, du fait de la simultanéité des arrivées dans le réseau.

De nombreux auteurs se sont intéressés à ce problème. Des résultats plus ou moins riches ont été obtenus selon la structure du réseau traversé. Ceux-ci sont résumés dans le paragraphe 2.

2 - ETAT DE L'ART

Le problème du rendez-vous fait l'objet de nombreux articles. Notre but n'est pas d'en faire une présentation exhaustive; nous nous sommes plutôt attaché à présenter, en fonction des structures possibles du réseau de files d'attente, les techniques utilisées face aux difficultés théoriques rencontrées.

2.1 - Réseaux parallèles

Le cas le plus simple de réseau est le réseau réduit à K files en parallèle tel que représenté par la figure 2

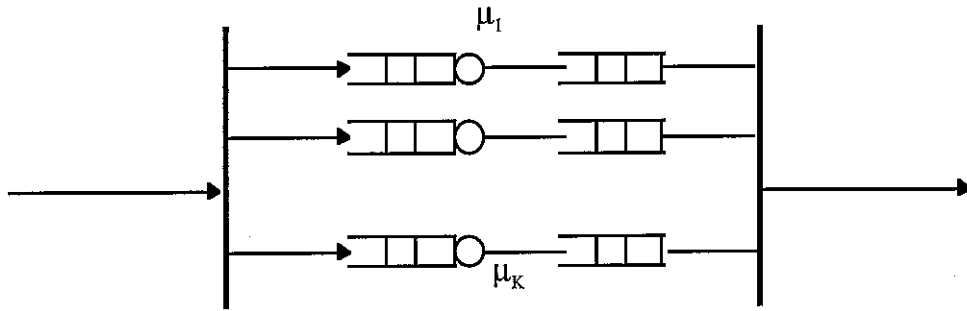


Figure 2 - Réseau parallèle

Dans ce cas, la détermination des temps de traversée des différents frères est classique.

On note que les t_n^k peuvent se décomposer en un temps d'attente w_n^k et un temps de service s_n^k

$$t_n^k = w_n^k + s_n^k$$

et que les w_n^k vérifient :

$$w_{n+1}^k = (w_n^k + s_n^k - \tau_{n+1})^+$$

où τ_n représente l'inter-arrivée entre C_{n-1} et C_n et $(x)^+ = \text{Max}(0, x)$.

Sous l'hypothèse que chaque file se comporte comme une file GI/GI/1, la solution de ce problème s'obtient par résolution de l'équation intégrale de Lindley.

Néanmoins, du fait de la simultanéité des arrivées (ce qui se traduit par le fait que l'inter-arrivée τ_{n+1} est identique pour les K files), les variables t_n^k ne sont pas indépendantes et le calcul des caractéristiques de t_n est pratiquement impossible.

2.1.1 - Modèle de dimension 2

Le modèle de dimension 2, réduit au dédoublement suivi de 2 files en parallèle, a été résolu analytiquement par Flatto et Hahn [Fla 84] dans le cas particulier où les arrivées extérieures sont poissonniennes et les serveurs exponentiels. Ils analysent la fonction génératrice $\Pi(x,y)$ des probabilités stationnaires $\pi(i,j)$ où i et j représentent les nombre de clients dans les files 1 et 2. Elle est définie par :

$$\Pi(x,y) = \frac{\mu_1 y(1-x)\Pi(0,y) + \mu_2 x(1-y)\Pi(x,0)}{\lambda xy(xy-1) + \mu_1 y(1-x) + \mu_2 x(1-y)} \quad (1)$$

La détermination de $\Pi(x,0)$ et $\Pi(0,y)$ résulte de l'analyse des pôles de $\Pi(x,y)$ dans le domaine de convergence de la série. Ceux-ci sont définis par la portion

$$\{|x|, |y| \leq 1\} \text{ de } S = \{(x,y) \mid D(x,y) = 0\}$$

où $D(x,y)$ est le dénominateur de $\Pi(x,y)$. S est une surface de Riemann d'ordre 1 qui peut être paramétrée par deux fonctions elliptiques (de Weierstrasse) $x(z)$ et $y(z)$. Elles induisent des conditions sur $\Pi(x(z),0)$ et $\Pi(0,y(z))$ qui permettent de déterminer $\Pi(x,0)$ et $\Pi(0,y)$ par prolongement analytique.

Ils montrent ainsi que si $\rho_1 = \lambda/\mu_1 \geq \rho_2 = \lambda/\mu_2$, $\Pi(x,0)$ vérifie

$$\Pi(x,0) = (1-\rho_2) \frac{\Phi(x)}{\Phi(1)} \quad (2)$$

où

$$\Phi(x) = \frac{\sqrt{a-x} + \sqrt{a-1}}{(\sqrt{a-x} + \sqrt{a\rho_1 - \rho_2})(\sqrt{a-x} + \sqrt{a\rho_1 - 1})}$$

et où a est la plus grande racine de l'équation :

$$[(\rho_1 + \rho_2 + \rho_1\rho_2)x - \rho_2]^2 - 4\rho_1^2\rho_2x^3 = 0$$

L'étude du modèle complet, incluant le regroupement, et l'algorithme de calcul des caractéristiques sont présentés dans [Mai 87]. Ces résultats seront utilisés dans le paragraphe 3.2.

2.1.2 - Modèle de dimension K (bornes)

Ce modèle a été étudié par Baccelli [Bac 85], [Bac 85b]. Il fait toujours l'hypothèse que les files se comportent comme des files GI/GI/1 et que les différents services $\{s_n^k \mid 1 \leq k \leq K\}$ sont indépendants.

Par comparaison stochastique avec deux réseaux à forme produit qui possèdent la même condition de stabilité, il obtient des bornes du temps de réponse, tant en régime transitoire qu'en régime permanent.

- la borne inférieure est obtenue en considérant des arrivées déterministes (i.e. $\tau_{n+1} = 1/\lambda$)
- la borne supérieure est obtenue en supposant que les files sont indépendantes, c'est-à-dire en négligeant la simultanéité des arrivées (i.e. on remplace τ_{n+1} par τ_{n+1}^k avec $\tau_{n+1} = (1/K) \sum \{ \tau_{n+1}^k \mid 1 \leq k \leq K \}$).

Dans le cas où les K serveurs sont exponentiels de même taux de service μ , Nelson et Tantawi [Nel 87] proposent une évaluation empirique du temps de réponse du modèle à K dimensions en supposant que le temps de réponse a une croissance harmonique.

2.2 - Réseaux séries-parallèles

Le second type de réseau étudié concerne les réseaux séries-parallèles. Cette classe de réseau a été définie récursivement par Baccelli [Bac 85a]: elle contient tous les réseaux formés d'une file FIFO unique ainsi que tout réseau formé par concaténation ou parallélisation de réseaux de même type. On note en particulier que ce type de réseau se caractérise par le fait qu'il n'y a pas de dépassement possible des frères successifs.

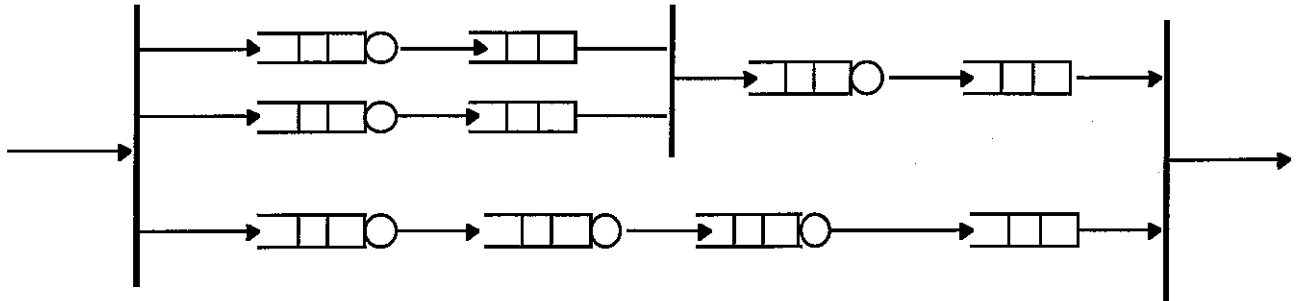


Figure 4 - Exemple de réseau série-parallèle

Les résultats cités au § 2.1.2 concernant les bornes du temps de traversée se généralisent dans ce cas. Ce type de réseau a par ailleurs été étudié de façon approchée par A. Duda [Dud 85] en utilisant une technique d'agrégation. Celle-ci consiste à remplacer le système par une file unique à taux de service dépendant de l'état obtenu classiquement par l'étude du réseau fermé. La complexité numérique limite la méthode à des réseaux de petite taille. Citons enfin les travaux de Liu et Perros [Liu 88] permettant une exploitation numérique plus raisonnable dans certains cas particuliers.

2.3 - Cas général

On note que si, dans le cas des réseaux envisagés, on peut obtenir des estimations sur le temps de traversée des différents frères, l'hypothèse d'indépendance entre ceux-ci reste un obstacle majeur pour la détermination du temps de réponse de bout en bout, hormis en ce qui concerne les bornes que l'on peut obtenir. Dans le cas général, deux problèmes supplémentaires se posent. Premièrement, le phénomène de déséquencement possible des clients dû, en particulier, aux rebouclages possibles; secondement aux phénomènes d'"interblocage" qui se produisent du fait que des frères peuvent emprunter une même file du réseau. Ces derniers problèmes ne sont pas abordés dans la littérature. Nous nous proposons d'analyser ce deuxième phénomène dans le paragraphe suivant.

3 - MODELE DE BASE

3.1 - Introduction

Nous avons vu que le problème du rendez-vous était toujours abordé pour des réseaux "série-parallèle". Nous nous proposons ici d'étudier le problème posé par le partage d'une ressource commune par les différents frères avant synchronisation.

Soit le modèle de base décrit par la figure 5.

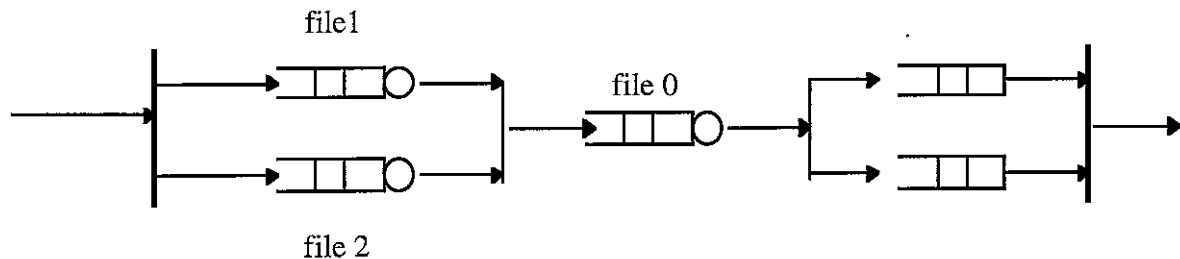


Figure 5 - Modèle de base

Dans ce système, les fils générés transitent tous les deux par une même file (file 0). Néanmoins, nous sommes toujours dans le cas où il n'y a pas de déséquencement possible.

Les temps de traversée des différents frères (t_n^k) ne peuvent plus, dans ce cas, s'obtenir sans prendre en compte la simultanéité des arrivées. En effet, en ignorant le phénomène de dépendance introduit, on est amené à sous-estimer le temps de transit de la file 0. On note en particulier que si on suppose que l'une des files 1 et 2 est saturée, le processus d'arrivée dans la file 0 est bien la superposition de deux processus indépendants et le résultat est alors évident. Si par contre ces deux files ont un taux d'occupation nul, le processus d'arrivée dans la file commune est un processus groupé de taille 2, ce qui aura pour effet d'augmenter le temps de réponse de la file 0. C'est pourquoi nous étudions par la suite le comportement de cette file.

On introduit les variables aléatoires suivantes :

I, J : nombres de clients dans les files 1 et 2,
E : inter-arrivées dans le système,
B_i (i=1,2) : durées de service des files 1 et 2
A : inter-arrivées dans la file 0.

Le processus d'arrivée dans le système est poissonnien de taux λ et les serveurs sont tous exponentiels, de taux respectifs μ_1 , μ_2 et ν . La suite des inter-arrivées et des services successifs sont supposés stationnaires et indépendants entre eux. On notera par la suite :

$$\rho_1 = \lambda / \mu_1 \quad \rho_2 = \lambda / \mu_2 \quad \rho_0 = 2\lambda / \nu$$

Nous supposons, sans perte de généralité, que $\rho_1 \leq \rho_2$.

3.2 - Etude du processus d'arrivée dans la file 0 [MAIL 90]

Soit $t_0, t_1, \dots, t_n, \dots$ les instants d'occurrence d'événements (départs de clients de l'une ou l'autre des files en amont de la file commune).

On note $A_n = t_n - t_{n-1}$ l'intervalle entre deux événements successifs ($n^{\text{ième}}$ inter-arrivée dans la file 0).

Considérons le sous-système composé des files 1 et 2 aux instants d'arrivée des clients dans la file 0. On note :

$$P_{00} = \text{Prob}(I = 0 \text{ et } J = 0) \quad P_{0J} = \text{Prob}(I = 0 \text{ et } J > 0) \quad P_{I0} = \text{Prob}(I > 0 \text{ et } J = 0) \quad P_{IJ} = \text{Prob}(I > 0 \text{ et } J > 0) \quad (3)$$

les probabilités d'état des files 1 et 2 à ces instants et $A_{00}, A_{I0}, A_{0J}, A_{IJ}$ les variables aléatoires conditionnelles associées à A_n . Dans ces quatre cas, on peut estimer la distribution des inter-arrivées dans la file 0 (définie par sa transformée de Laplace) de la façon suivante :

- Si les deux premières files sont vides, $A_{00} = E + \min\{B1, B2\}$ On a donc, en raison de l'indépendance de ces variables :

$$A_{00}^*(s) = \frac{\lambda}{\lambda + s} \frac{\mu_1 + \mu_2}{\mu_1 + \mu_2 + s}$$

- Si la file 2 est vide mais pas la file 1, $A_{I0} = \min\{B1, E + B2\}$, soit :

$$A_{I0}^*(s) = \frac{\mu_2}{\mu_2 - \lambda} \frac{\mu_1 + \lambda}{\mu_1 + \lambda + s} - \frac{\lambda}{\mu_2 - \lambda} \frac{\mu_1 + \mu_2}{\mu_1 + \mu_2 + s}$$

- Par symétrie, on déduit que :

$$A_{0J}^*(s) = \frac{\mu_1}{\mu_1 - \lambda} \frac{\mu_2 + \lambda}{\mu_2 + \lambda + s} - \frac{\lambda}{\mu_1 - \lambda} \frac{\mu_1 + \mu_2}{\mu_1 + \mu_2 + s}$$

- Dans le dernier cas, on a simplement $A_{IJ} = \min\{B1, B2\}$:

$$A_{IJ}^*(s) = \frac{\mu_1 + \mu_2}{\mu_1 + \mu_2 + s}$$

Il en résulte que la distribution des inter-arrivées dans la file 0 a pour transformée de Laplace :

$$A^*(s) = P_{00}A_{00}^*(s) + P_{I0}A_{I0}^*(s) + P_{0J}A_{0J}^*(s) + P_{IJ}A_{IJ}^*(s) \quad (4)$$

Les probabilités intervenant dans le calcul de $A^*(s)$ et définies par (3) se déduisent des probabilités stationnaires notées $\pi(i,j)$ du modèle constitué des files 1 et 2 établies par Flatto et présentées dans le § 2.1.1.

L'équation (2) donne la valeur de $\Pi_{00} = \pi(0,0)$ dont on déduit toutes les autres, à savoir

$$\Pi_{10} = \sum_{i \geq 1} \pi(i, 0) = 1 - \rho_2 - \Pi_{00} \quad \text{et} \quad \Pi_{0J} = \sum_{j \geq 1} \pi(0, j) = 1 - \rho_1 - \Pi_{00}$$

et $\Pi_{IJ} = \sum_{i \geq 1} \sum_{j \geq 1} \pi(i, j)$ par normalisation.

A partir des probabilités stationnaires, on obtient les probabilités correspondantes aux instants de la chaîne incluse, que l'on note $p(i, j)$.

Il est clair que si un client laisse le système amont vide après son départ (avec la probabilité $P_{00} = p(0, 0)$), celui-ci le restera tant qu'il ne se produira aucune arrivée extérieure. On sait de plus que l'espérance mathématique des inter-arrivées dans la file 0 vaut $1/2\lambda$. On peut en déduire, par application du théorème de renouvellement markovien [Cin 75], que :

$$\pi(0, 0) = 2\lambda p(0, 0) \int_0^{\infty} e^{-\lambda t} dt$$

De la même manière, on établit que

$$\pi(i, 0) = 2\lambda p(i, 0) \int_0^{\infty} e^{-(\lambda + \mu_1)t} dt$$

et par conséquent on a :

$$P_{00} = \frac{\Pi_{00}}{2} \quad P_{10} = \frac{\Pi_{10}}{2} \frac{\rho_1}{1 + \rho_1} \quad P_{0J} = \frac{\Pi_{0J}}{2} \frac{\rho_2}{1 + \rho_2} \quad (5)$$

3.3 - Périodicité du processus d'arrivée dans la file commune

La simultanéité des arrivées extérieures induit un comportement périodique sur la nature des inter-arrivées. Il suffit pour s'en convaincre de constater que les files 1 et 2 ne peuvent être simultanément vides qu'à des instants d'indices pairs (en supposant que le système soit vide à l'état initial). De façon générale, en notant X_n l'état global des files 1 et 2 (nombre total de clients dans les deux files), X_n change de parité à chaque occurrence d'événement, conserve sa parité sur une arrivée extérieure (qui ne constitue pas un événement).

Si les variables aléatoires A_{00} , A_{10} , A_{0J} et A_{IJ} définies dans le paragraphe précédent sont bien indépendantes entre elles, le processus $\{A_n | n \in \mathbb{N}\}$ n'est donc pas un processus de renouvellement.

Nous noterons A_p et A_i les variables aléatoires correspondant aux cas pairs et impairs et $A_p^*(s)$ et $A_i^*(s)$ les transformées de Laplace des distributions correspondantes. Ces deux derniers processus ont des distributions de type GI (distributions générales indépendantes). Le processus A est un processus GI alterné qui constitue un cas particulier de processus semi-markovien. On a évidemment :

$$A^*(s) = 1/2 [A_p^*(s) + A_i^*(s)]$$

où

$$\begin{aligned} A_p^*(s) &= 2[P_{00}A_{00}^*(s) + P_{10p}A_{10}^*(s) + P_{0Jp}A_{0J}^*(s) + P_{IJp}A_{IJ}^*(s)] \\ A_i^*(s) &= 2[P_{10i}A_{10}^*(s) + P_{0Ji}A_{0J}^*(s) + P_{IJi}A_{IJ}^*(s)] \end{aligned}$$

les probabilités d'état des files 1 et 2 étant calculées aux instants pairs et impairs.

Ces probabilités s'obtiennent à partir des équations (5) en remarquant que Π_{I0i} se déduit de (2) car $\Pi_{I0i} = \frac{1}{2}[\Pi(1,0) - \Pi(-1,0)]$ et, par suite, $\Pi_{I0p} = \Pi_{I0} - \Pi_{I0i}$.

Les caractéristiques de A s'expriment en fonction de celles de A_p et A_i . On note m_p et σ_p (resp. m_i et σ_i) l'espérance mathématique et l'écart-type de A_p (resp. de A_i). On a donc [Cox 69]:

$$\begin{aligned} E[A_n] &= E[A] = \frac{1}{2}(m_p + m_i) = \frac{1}{2\lambda} \\ \text{Var}[A_n] &= \text{Var}[A] = \frac{1}{2}(\sigma_p^2 + \sigma_i^2) - \frac{1}{4}(m_p - m_i)^2 \\ \text{Co var}[A_n, A_{n+k}] &= \gamma_k[A] = \frac{1}{4}(m_p - m_i)^2 (-1)^k \end{aligned}$$

ne dépendant donc que des caractéristiques au premier ordre des processus A_p et A_i .

4 - RESOLUTION DE LA FILE COMMUNE

La résolution de la file 0 (supposée à serveur exponentiel) est basée sur le principe développé par Cinlar [Cin 67] concernant les files à serveur exponentiel et à arrivées semi-markoviennes. Le caractère périodique induit par le « fork » nécessite cependant d'adapter la méthode. Il convient en effet d'étudier le modèle sur un intervalle double $[t_{2n}, t_{2n+2}]$ correspondant à la suite (A_p, A_i) ou $[t_{2n+1}, t_{2n+3}]$ correspondant à la suite (A_i, A_p) .

On note $a_i(k)$ ($a_p(k)$) la probabilité de k départs de la file 0 pendant une inter-arrivée de durée A_i (resp., A_p). Les fonctions génératrices de ces distributions vérifient :

$$A_r(z) = \sum_{k=0}^{\infty} a_r(k) z^k = A_r^*[v(1-z)] \quad r \in \{p, i\}$$

On note $\mathcal{A}_p$ (resp., $\mathcal{A}_i$) la matrice de transition du processus décrivant la file 0 sur un intervalle de durée A_p (resp., A_i). Pour tout $r \in \{p, i\}$, on a :

$$\begin{aligned} \mathcal{A}_r(i, j) &= a_r(i+1-j) \quad j \geq 1 \\ \mathcal{A}_r(i, 0) &= \bar{a}_r(i) = 1 - \sum_{k=0}^i a_r(k) \end{aligned}$$

Ces matrices ont la structure de la matrice présentée en annexe pour $N=1$.

La matrice de transition, notée C_p , sur un intervalle de type $[t_{2n}, t_{2n+2}]$ vérifie $C_p = \mathcal{A}_p \mathcal{A}_i$. Elle a aussi la même structure pour $N=2$. Les coefficients $c(k)$ ont pour fonction génératrice:

$$C(z) = A_p(z)A_i(z)$$

Il en est de même pour la matrice $C_i = \mathcal{A}_i \mathcal{A}_p$. Seules les premières colonnes diffèrent et vérifient :

$$C_p(j, 1) = \bar{a}_p(j) a_i(0) + \sum_{k=0}^j a_p(j-k) a_i(k+1) = \bar{a}_p(j+1) a_i(0) + c(j+1)$$

de fonction génératrice, notée $C_p(z,1)$:

$$C_p(z,1) = \frac{1}{z} \left[a_i(0) \frac{z - A_p(z)}{1 - z} + C(z) \right]$$

Par symétrie on a également :

$$C_i(z,1) = \frac{1}{z} \left[a_p(0) \frac{z - A_i(z)}{1 - z} + C(z) \right]$$

4.1 - Distribution du nombre de clients aux instants de la chaîne incluse

Les résultats présentés dans l'annexe permettent alors de calculer les probabilités stationnaires de la file 0, $q_r(j)$ pour tout $r \in \{p,i\}$ aux instants de la chaîne incluse. On calcule les deux seules racines σ_1 et σ_2 de module inférieur à 1 de l'équation :

$$C(\sigma) = \sigma^2$$

On en déduit que

$$q_r(j) = q_r(0) \left[\alpha_r \sigma_1^j + (1 - \alpha_r) \sigma_2^j \right]$$

et par normalisation

$$q_r(0) = \left[\frac{\alpha_r}{1 - \sigma_1} + \frac{1 - \alpha_r}{1 - \sigma_2} \right]^{-1}$$

Les coefficients α_r vérifiant les équations :

$$\frac{\alpha_r}{\sigma_1} \frac{\sigma_1 - A_r(\sigma_1)}{1 - \sigma_1} + \frac{1 - \alpha_r}{\sigma_2} \frac{\sigma_2 - A_r(\sigma_2)}{1 - \sigma_2} = 0$$

Plus simplement en notant :

$$\beta_r = \frac{q_r(0) \alpha_r}{1 - \sigma_1} \quad \text{soit} \quad 1 - \beta_r = \frac{q_r(0) (1 - \alpha_r)}{1 - \sigma_2}$$

les probabilités stationnaires de la chaîne incluse s'écrivent alors sous la forme :

$$q_r(j) = \beta_r (1 - \sigma_1) \sigma_1^j + (1 - \beta_r) (1 - \sigma_2) \sigma_2^j$$

de fonctions génératrices :

$$Q_r(z) = \beta_r \frac{1 - \sigma_1}{1 - \sigma_1 z} + (1 - \beta_r) \frac{1 - \sigma_2}{1 - \sigma_2 z}$$

dont on vérifie que les coefficients β_r se déduisent immédiatement des équations :

$$\beta_r \left[1 - \frac{A_r(\sigma_1)}{\sigma_1} \right] + (1 - \beta_r) \left[1 - \frac{A_r(\sigma_2)}{\sigma_2} \right] = 0$$

4.2 - Distribution du temps de réponse de la file 0

La connaissance des probabilités d'état aux instants de la chaîne incluse permettent alors de déterminer la distribution des temps de réponse pour les clients pairs et impairs. On a en effet

$$T_r^*(s) = \frac{v}{v+s} Q_r \left(\frac{v}{v+s} \right)$$

où $T_r^*(s)$ est la transformée de Laplace du temps de réponse pour un client de type r et $Q_r(z)$ la fonction génératrice des probabilités $q_r(j)$.

La distributions du temps de réponse est alors simplement la moyenne des 2 distributions précédente :

$$T^*(s) = \frac{1}{2} (T_p^*(s) + T_i^*(s))$$

et par conséquent on en déduit sans difficulté que :

$$T^*(s) = \beta \frac{v(1-\sigma_1)}{v(1-\sigma_1)+s} + (1-\beta) \frac{v(1-\sigma_2)}{v(1-\sigma_2)+s}$$

où

$$\beta = (\beta_p + \beta_i)/2$$

Les caractéristiques du temps de réponse de la file commune s'en déduisent par dérivation de $T^*(s)$. On a ainsi :

$$E(T) = \frac{1}{v} \left[\frac{\beta}{1-\sigma_1} + \frac{1-\beta}{1-\sigma_2} \right]$$

$$E(T^2) = \frac{2}{v^2} \left[\frac{\beta}{(1-\sigma_1)^2} + \frac{1-\beta}{(1-\sigma_2)^2} \right]$$

4.3 - Distribution stationnaire de la file 0

Notons $\gamma_r(j)$ les probabilités stationnaires de la file commune à un instant quelconque. Elles se déduisent des probabilités $q_r(j)$ de la chaîne incluse. Nous ne détaillons pas ici la méthode qui est classique [Cin 75]. On obtient

$$\gamma_r(j) = \frac{1}{vE(A_r)} \left[\alpha_r \sigma_1^{j-1} (1 - A_r(\sigma_1)) + (1 - \alpha_r) \sigma_2^{j-1} (1 - A_r(\sigma_2)) \right]$$

correspondant respectivement aux intervalles de type A_p et A_i . La solution générale s'obtient alors en faisant la moyenne pondérée par l'espérance mathématique des durées de ces intervalles.

$$\gamma(j) = \frac{E(A_p)}{E(A_p) + E(A_i)} \gamma_p(j) + \frac{E(A_i)}{E(A_p) + E(A_i)} \gamma_i(j)$$

conduisant, après simplification, à la solution générale :

$$\gamma(j) = \frac{2\lambda}{v} \left[\beta(1-\sigma_1)\sigma_1^{j-1} + (1-\beta)(1-\sigma_2)\sigma_2^{j-1} \right] \quad \text{pour } j \geq 1$$

$$\gamma(0) = 1 - \frac{2\lambda}{v} = 1 - \rho_0$$

5 - CONCLUSION

La méthode se généralise dans nombre de cas où on connaît les distributions stationnaires du réseau amont. On note en particulier que des modèles de « Fork » de dimension k (génération de k fils en parallèle) sont susceptibles de la même méthode de résolution; le processus d'arrivée dans la file commune est alors de périodicité k . Il n'en reste pas moins que la résolution d'un réseau parallèle à k files n'est pas connue, ce qui limite l'intérêt de la méthode.

Par contre, le modèle de la figure 6 peut être résolu dans le cas où le réseau R1 est lui-même constitué des files 1 et 2 tel que décrit dans la figure 5. Le problème est alors périodique d'ordre 3.

On peut également résoudre le problème si le réseau R1 est de type BCMP. En effet, la distribution des probabilités d'état de ce réseau est connue aux instants de la chaîne incluse (correspondant par ailleurs aux distributions stationnaires). La méthode présentée ici permet ainsi de résoudre la file 0 sous réserve que l'on puisse définir les distributions de probabilités des interarrivées conditionnelles dans la file 0.

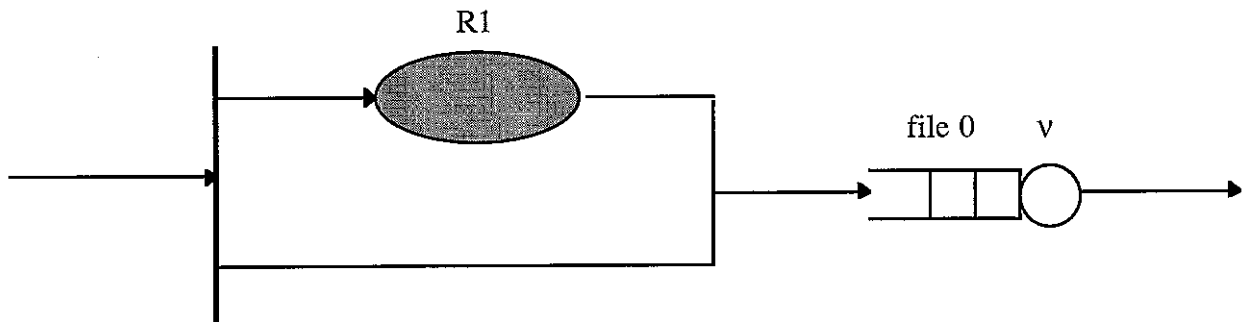


Figure 6 : généralisation du modèle de base

Hormis ces cas, nous ne pouvons obtenir de solution analytique. Nous pouvons néanmoins envisager de développer des heuristiques basées sur les propriétés de périodicité mises en évidence dans ce papier.

ANNEXE

On considère un processus de Markov dont la matrice de transition a la structure suivante :

$$C = \begin{pmatrix} 0 & 1 & \dots & N & N+1 & \dots \\ & & & c_0 & 0 & \dots \\ & & & c_1 & c_0 & 0 \\ & & c_{ij} & c_2 & c_1 & c_0 & \ddots \\ & & & \vdots & c_2 & c_1 & \ddots \end{pmatrix}$$

Ce type de processus généralise les modèles relatifs aux files de type $GI^{[N]}/M/1$ (files à durée de service exponentiel et à arrivées générales indépendantes par groupe de taille N) ou encore les modèles équivalents pour les files $GI/Er-N/1$ (file à arrivées générales indépendantes et à durée de service suivant une loi d'Erlang- N). Dans ce cas particulier, les éléments c_{ij} de la matrice de transition C vérifient aussi (pour tout $1 \leq j \leq N$) : $c_{ij} = c_{N+i-j}$.

On note $C(z, j) = \sum_{i=0}^{\infty} c_{ij} z^i$ la fonction génératrice des colonnes de la matrice C et $C(z)$ la fonction génératrice des $\{c_i\}$.

On a donc $C(z, N) = \sum_{i=0}^{\infty} c_i z^i = C(z)$ et $C(z, N+j) = z^j C(z)$ pour $j \geq 1$

Régime stationnaire :

Les probabilités stationnaires (si elles existent) vérifient $q = q C$. On note que, pour toute solution du type $q(j) = a z^j$, les équations de balances sont vérifiées à partir du rang N . En effet, on a $z^{N+j} = z^j C(z)$, soit pour toute valeur de z vérifiant l'équation $z^N = C(z)$.

Elles le seront donc aussi pour toute combinaison linéaire des racines de l'équation précédente.

On montre que (théorème de Rouché) l'équation $z^N = C(z)$ admet exactement N racines de module inférieur à 1 si $|C(z)| < |z^N|$ pour toute valeur de z de module ≤ 1 , soit pour $C'(z)_{z=1} < N$.

On pose dès lors :

$$q(j) = q(0) \left[\sum_{k=1}^N \alpha_k \sigma_k^j \right]$$

où les $\{\sigma_k / k = 1, \dots, N\}$ sont les N racines de module inférieur à 1 de l'équation $z^N = C(z)$

Les coefficient α_k sont déterminés à partir des N-1 premières équations de balance, ce qui implique que :

$$\sum_{k=1}^N \alpha_k \left[C(\sigma_k, j) - \sigma_k^j \right] = 0 \quad \text{quel que soit } j=1, \dots, N-1$$

avec $\sum_{k=1}^N \alpha_k = 1$

Algorithme de calcul des σ_k :

On trouve dans [Ada 96] une méthode efficace de calcul des N racines, de module inférieur à 1, de l'équation $z^N = C(z)$. Les auteurs montrent que si le système est stable et si $C(z)$ ne possède pas de zéros à l'intérieur du cercle unité on peut calculer les N racines par itérations successives:

$$\sigma_k^{(n+1)} = \phi_k \sqrt[N]{C(\sigma_k^{(n)})} \quad \text{où } \phi_k \text{ est la } k^{\text{ième}} \text{ racine de l'équation } \phi^N = 1$$

à partir de la solution initiale $\sigma_k^{(0)} = 0$.

BIBLIOGRAPHIE

- [Ada 96] I.ADAN, Y.ZHAO, "Analyzing GI/Er/1 Queues", Operations Research Letters 19 n°4, Octobre 1996, pp 183-190.
- [Bac 85] F. BACCELLI, A.M. MAKOWSKI, "Simple Computable Bounds for Fork-Join Queue", Rapport INRIA n°394, April 1985.
- [Bac 85a] F. BACCELLI, W.A. MASSEY, "Series-Parallel Fork-Join Queueing Networks and their Stochastic Ordering", Rapport INRIA n°534, June 1986.
- [Bac 85b] F. BACCELLI, A.M. MAKOWSKI, "Stability and Bounds for Single Server Queues in Random Environment", Rapport INRIA n°536, June 1986.
- [Cin 67] E. CINLAR, "Queues with Semi-Markovian Arrivals", J.Appl.Prob.4, 1967, pp 365-379.
- [Cin 75] E. CINLAR, "Introduction to Stochastic Processes", Engelwood Cliffs, N.J. Prentice Hall, 1975.
- [Cox 69] D.R.COX, P.A.W. LEWIS, " L'Analyse Statistique des Séries d'Événements", Dunod, 1966.
- [Dud 85] A. DUDA, T. CZACHORSKI, "Performance Evaluation and Fork-Join Primitives", Rapport de Recherche ISEM, n°45, Université Paris Sud, January 1986.
- [Fdi 86] S. FDIDA, G. PUJOLLE, D. MAILLES, " Queueing Systems with Ressource Sharing", Journal of System and Software, Vol 6, N°1 &2, pp 23-29, May 1986.
- [Fla 84] L. FLATTO, S. HAHN, " Two Parallel Queues Created by Arrivals with Two Demands",SIAM J. Appl. Math., vol 44,n° 5, october 1984
- [Liu 88] Y.C. LIU, H.G. PERROS, " A Decomposition Procedure for the Analysis of a Closed Fork-Join Queueing System", Technical Report- North Carolina State University, April 1988.
- [Mai 87] D. MAILLES, " Files d'Attente Descriptives pour la Modélisation de la Synchronisation dans les Systèmes Informatiques", Thèse d' état, Université P. et M. Curie, Septembre 1987.
- [Mai 90] D. MAILLES, "Un problème de Rendez-Vous",Rapport M.A.S.I 90.06 Institut BLAISE PASCAL, Université P. et M. Curie, Février 1990.
- [Nel 87] R. NELSON, A.N. TANTAWI, "Approximate Analysis of Fork-Join Synchronisation in Parallel Queues", IBM Research Report RC11481, 1987.
- [Neu 66] M.R. NEUTS, "The Single Server-Queue with Poisson Arrivals and Semi-Markov Service Times", J.Appl.Prob.3, 1966, pp 202-230.