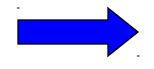




Techniques de modélisation: Méthodes analytiques

1- INTRODUCTION

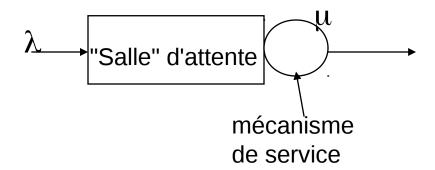
- Systèmes informatiques
 - Matériel, logiciel
 - Jobs
 - partagent des ressources (CPU, Disques, fichiers ...)
 - Ressources utilisées de façon unique à un instant donné, autres demandeurs attendent.
 - On s'intéresse essentiellement:
 - débits du système(Job/s, IO/s...)
 - Temps de réponse sous une charge donnée.



Théorie des files d'attente permet une formalisation et une étude quantitative.

2.1 DEFINITION

- Une file d'attente est caractérisée par :
 - Un flot d'arrivées
 - Un mécanisme de service
 - Une salle d'attente
 - Une discipline de service



- ☐ Flot d'arrivée
 - suite stochastique
 - T_n: est le temps du neme client
 - s_n: la charge apportée par le neme client, service nécessaire.
 - Les clients arrivent successivement, un à la fois, il n'y a pas d'accumulation: $0<T_1<T_2<T_3<.....<T_n<T_{n+1}<...$
- Mécanisme de service
 - nombre de serveurs et leur vitesse, σ_n unités de temps par service.

- Capacité de la file d'attente:
 - nombre de places possibles : limité ou illimité.
 - Si capacité limitée: les clients supplémentaires sont perdus ou rejoignent une autre file d'attente.
 - Le nombre de clients dans le système est différent du nombre de clients dans la file d'attente

- Discipline de service:
 - règle d'ordonnancement des clients au service.
 - FIFO: first in first out
 - LIFO : Last in first out
 - PS : processor sharing, un serveur donne à chaque client en attente une 'tranche' de service.
 - -ALEA un serveur libre choisit un client au hasard dans la file
 - Priorité: on ajoute une suite {U_n}, n appartient à N^{+,} au flot des arrivées où U_n est une variable aléatoire prenant ses valeurs dans l'ensemble des classes de priorités P. U_n=i, signifie que le n^{eme} client, arrivant au temps T_n est de la classe i.
 - Priorité préemptive

- Notation de KENDALL A/B/C/D/E
 - A: statistique du processus d'arrivée (M = markovien;
 D=déterministe; G=générale)
 - B: statistique des lois de service (M = markovien;
 D=déterministe; G=générale)
 - C: nombre de serveurs
 - D: nombre de clients dans le système
 - E: discipline du service

- Application aux réseaux de communications et aux systèmes informatiques:
 - 🔍 clients = tâches, programmes, paquets, ...
 - temps de service = durée de tâche
 - serveur = processeur
 - salle d'attente = tampon

2-2 LOI DE LITTLE

HYPOTHESES

- Lorsqu'un client, ayant terminé son service, quitte le système, il laisse, en moyenne, derrière lui, un nombre de clients égal à E(k). Ce client a trouvé en arrivant E(k) clients déjà
 - Ce client a trouvé en arrivant E(k) clients déjà présents et a passé dans le système un temps, E(T).
- □ Nous supposons que:
 - Le nombre moyen des arrivées est égal au nombre moyen des départs du système.
 - La longueur moyenne de la file lors des arrivées est égale à la longueur moyenne de la file lors des départs

2-2 LOI DE LITTLE

ENONCE

- Si on appelle, λ , le taux moyen des arrivées on a:
 - {Nombre moyen de clients arrivés pendant le séjour du client dans le système} =
- $\square \lambda$ E(t) = {nombre moyen de clients qu'il laisse}
- Et en régime permanent, si T temps passé dans la file :

$$E(k)=\lambda E(T)$$

$$\bar{N} = \lambda \bar{T}$$

2.2 LOI DE LITTLE

VALIDITE

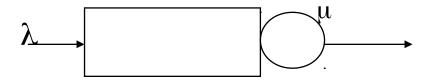
- Régime permanent
- Les formules de Little sont valides pour les files G/G/S. Elles ont un caractère très général. En effet, il n' y a aucune restriction quant à :

la loi d'arrivée, la loi des services, le nombre de serveurs.

Elles peuvent prendre en compte le cas où il existe plusieurs classes de clients mais la discipline de service doit être définie, nous avons considéré la discipline FIFO.

DEFINITION

- \square Arrivées: loi de Poisson de taux λ
- Un serveur
 - Loi de service: exponentielle de paramètre μ
 - Capacité de file: infinie
 - discipline de service : FIFO
 - Si serveur saturé (ρ -->1):
 - La distribution des intervalles de temps séparant deux départ consécutifs de la file saturée tend vers la distribution des temps de service



- PRINCIPAUX RESULTATS
 - État: nombre de clients, k, dans le système
 - probabilité de l'état E_k :

$$p_{k} = \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^{k} = (1 - \rho) \rho^{k}$$

$$p_0 = \left(1 - \frac{\lambda}{\mu}\right) = 1 - \rho$$

 \blacksquare Taux de trafic, ρ (charge, activité du serveur):

$$\rho = \frac{\lambda}{\mu}$$

Condition de stabilité:activité du serveur <1

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

 $\square \lambda < \mu$

Débit d'entrée < débit du serveur

temps moyen inter-arrivée > temps de service moyen

Nombre moyen de clients dans le système, N

$$\bar{N} = \sum_{k=0}^{\infty} k p_{k} = \sum_{k=0}^{\infty} k (1 - \rho) \rho^{k} = (1 - \rho) \rho \sum_{k=1}^{\infty} k \rho^{k-1}$$

$$\bar{N} = (1 - \rho) \rho \sum_{k=0}^{\infty} (\rho^{k})' = (1 - \rho) \rho \left(\frac{1}{(1 - \rho)}\right)'$$

$$\bar{N} = (1 - \rho) \rho \frac{1}{(1 - \rho)^{2}} = \frac{\rho}{1 - \rho}$$

$$\bar{N} = \frac{\rho}{1 - \rho}$$

 \square Nombre moyen de clients dans la file, E(v)

$$E(v) = \sum_{k=1}^{\infty} (k-1) \rho_k = \sum_{k=1}^{\infty} (k-1) (1-\rho) \rho^k$$

$$E(v) = (1-\rho) \rho^2 \sum_{k=1}^{\infty} (k-1) \rho^{k-2}$$

$$E(v) = (1-\rho) \rho^2 \frac{1}{(1-\rho)^2} = \frac{\rho^2}{1-\rho}$$

$$E(v) = \frac{\rho^2}{1-\rho}$$

Temps moyen,T, passé par un clients dans système

N : nombre de clients

 λ : débit d'entrée, nombre de clients par seconde

T: durée d'observation

$$\bar{N} = \lambda \bar{T}$$

Le nombre moyen de clients dans le système s'écrit (Little):

$$\overline{N} = E(m) = \lambda E(t)$$

et le nombre moyen de clients dans la file s'écrit:

$$E(v) = \lambda E(t_f)$$

Temps moyen, T, passé par un client dans le système:

$$\overline{T} = \frac{\overline{N}}{\lambda} = \left(\frac{\rho}{1 - \rho}\right) \left(\frac{1}{\lambda}\right)$$

$$\overline{T} = \frac{1}{\mu} \frac{1}{1 - \rho}$$

Nombre moyen de clients dans le système:

$$\bar{N} = \frac{\rho}{1 - \rho}$$

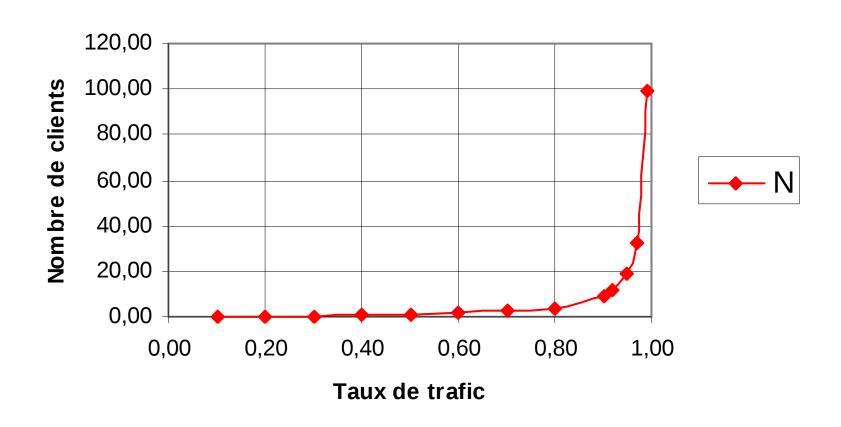
Nombre moyen de clients dans la file:

$$\bar{v} = \frac{\rho^2}{1-\rho}$$

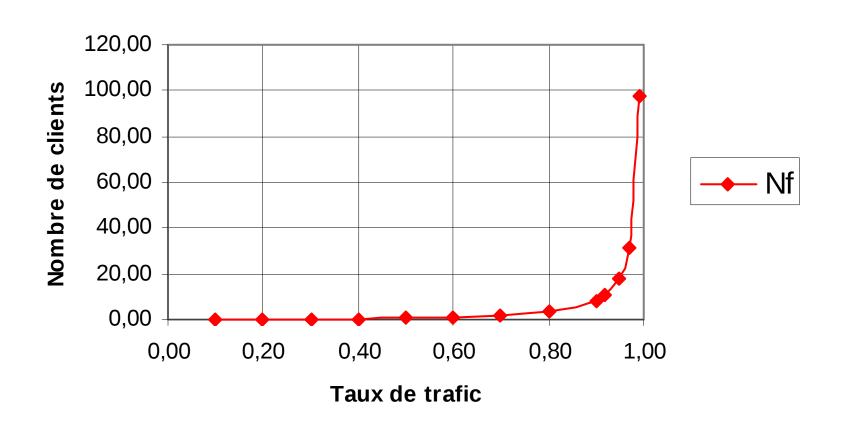
Temps moyen passé par un client dans le système:

$$\overline{T} = \frac{1}{\mu} \frac{1}{1 - \rho}$$

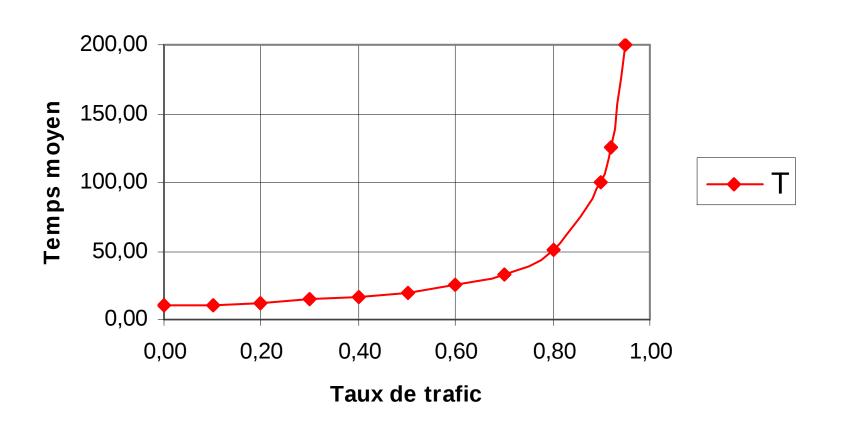
Nombre de clients dans le système en fonction du taux de trafic



Nombre de clients dans la file en fonction du taux de trafic

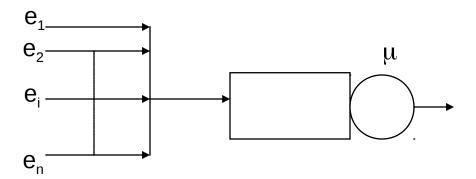


Temps moyen passé par un client dans le système en fonction du taux de trafic



- DISCUSSION DES HYPOTHESES
 - Prob(1 arrivée pdt Δ T)= $\lambda \Delta$ t
 - Prob(1 départ pdt ΔT / serveur occupé)= $\mu \Delta t$
 - Les probabilités d'occurrence des événements ne dépendent que de l'état du système à l'instant d'observation:
 - Indépendantes de l'état antérieur du système
 - propriété markovienne
 - Loi exponentielle

- Processus d'arrivée:
 - temps inter-arrivées: variable aléatoire exponentielle
 - Processus de Poisson
 - Vérification pratique hypothèse de Poisson :
 Bien vérifiée dans un contexte informatique.



Sources indépendantes (terminaux...)

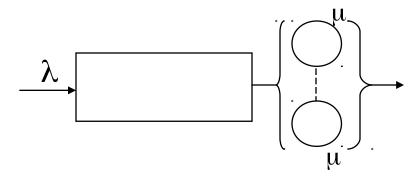
Superposition Poissor

- Durée du service:
 - temps service: variable aléatoire exponentielle
 - Vérification pratique hypothèse exponentielle : souvent non satisfaite dans contexte informatique coefficient de variation, c, élevé.

$$c = \frac{ecart_type}{moyenne} > 1$$

DEFINITION

- \square Arrivées: loi de Poisson de taux λ
- S serveurs
 - S serveurs indépendants
 - Loi de service: exponentielle de paramètre μ
 - Capacité de file: infinie
 - discipline de service : FIFO

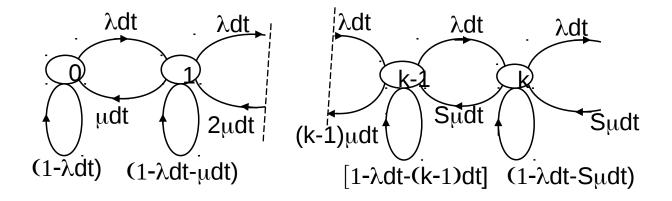


S serveurs

- Modélisation
 - processus de naissance et mort.
 - Etat, Ek, où k unités présentes dans le système
 - file, serveurs.
 - S serveurs
 - Taux de naissance et mort:

$$\begin{cases} \lambda_k = \lambda \\ \mu_k = k\mu & \forall k < S & et \in N \\ \mu_k = S\mu & \forall k \ge S & et \in N \end{cases}$$

Le graphe associé est le suivant:



- PRINCIPAUX RESULTATS
 - État: nombre de clients, k, dans le système
 - probabilité de l'état E_k :

$$\begin{cases} p_k = \frac{1}{k!} \left(\frac{\lambda}{\mu}\right)^k p_0 & \text{Si } k < S \\ p_k = \frac{1}{S^{k-S}S!} \left(\frac{\lambda}{\mu}\right)^k p_0 & \text{Si } k \ge S \end{cases}$$

Calcul de p₀

Calcul de p₀

$$\sum_{k=0}^{\infty} P_k = 1$$

Soit en remplaçant P_k , par les expressions trouvées ci-dessus:

$$P_0 \left[\sum_{k=0}^{S-1} \left(\frac{\lambda}{\mu} \right)^k \frac{1}{k!} + \sum_{k=S}^{\infty} \left(\frac{\lambda}{\mu} \right)^k \frac{1}{S^{k-S}S!} \right] = 1$$

$$P_0 \left[\sum_{k=0}^{S-1} \left(\frac{\lambda}{\mu} \right)^k \frac{1}{k!} + \frac{S^S}{S!} \sum_{k=S}^{\infty} \left(\frac{\lambda}{S\mu} \right)^k \right] = 1$$

Posons
$$r = \frac{\lambda}{S\mu}$$

$$P_0 \left[\sum_{k=0}^{S-1} \left(\frac{\lambda}{\mu} \right)^k \frac{1}{k!} + \frac{S^S}{S!} \sum_{k=S}^{\infty} (r)^k \right] = 1$$

Le second terme entre crochets s'écrit:

$$\frac{S^{S}}{S!} \sum_{k=S}^{\infty} (r)^{k} = \frac{S^{S}}{S!} \left[r^{S} + r^{S+1} + r^{S+2} + \dots \right] = \frac{S^{S}}{S!} r^{S} \left[1 + r + r^{2} + \dots \right]$$

Calcul de p₀

$$[1+r+r^2+....]$$
, une série géométrique de terme général: $r=\frac{\lambda}{S\mu}$

Série est convergente si : $\frac{\lambda}{S\mu} \le 1$ et sa somme est égale à $\frac{1}{1-r}$

On peut alors écrire :

$$\frac{S^{S}}{S!} \sum_{k=S}^{\infty} r^{k} = \frac{S^{S}}{S!} r^{S} \frac{1}{1-r} = \frac{S^{S}}{S!} \frac{r^{S}}{1-r}$$
 en remplaçant r par sa valeur:

$$\frac{S^{s}}{S!} \sum_{k=s}^{\infty} r^{k} = \frac{S^{s}}{S!} \frac{\left(\frac{\lambda}{S\mu}\right)^{s}}{1 - \left(\frac{\lambda}{S\mu}\right)} = \frac{1}{S!} \left(\frac{\lambda}{\mu}\right)^{s} \frac{1}{1 - \left(\frac{\lambda}{S\mu}\right)}$$

Soit pour P₀:

$$P_0 = \frac{1}{\sum_{k=0}^{S-1} \frac{1}{k!} \left(\frac{\lambda}{\mu}\right)^k + \frac{1}{S!} \left(\frac{\lambda}{\mu}\right)^S \frac{1}{1 - \left(\frac{\lambda}{S\mu}\right)}}$$

 \square Taux de trafic, ρ (charge, activité du serveur):

$$\rho = \frac{\lambda}{S\mu}$$

☐ Condition de stabilité:

$$\rho = \frac{\lambda}{\mathsf{S}\mu} < 1$$

Nombre moyen de guichets, g, occupés:

$$\bar{g} = \sum_{k=0}^{S} k P_k + \sum_{k=S+1}^{\infty} S P_k$$

$$\bar{g} = \sum_{k=1}^{S} k \frac{\left(\frac{\lambda}{\mu}\right)^k}{k!} P_0 + \sum_{k=S+1}^{\infty} S \frac{\left(\frac{\lambda}{\mu}\right)^k}{S! S^{k-S}} P_0$$

$$\bar{g} = P_0 \frac{\lambda}{\mu} \left[\sum_{k=1}^{S} \frac{\left(\frac{\lambda}{\mu}\right)^{k-1}}{(k-1)!} + \sum_{k=S+1}^{\infty} \frac{\left(\frac{\lambda}{\mu}\right)^{k-1}}{S! S^{k-S-1}} \right]$$

Posons:j=k-1

□ Nombre moyen de guichets, g, occupés:

Avec:j=k-1

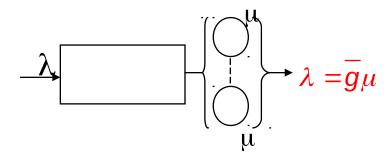
$$\bar{g} = P_0 \frac{\lambda}{\mu} \left[\sum_{j=0}^{S-1} \frac{\left(\frac{\lambda}{\mu}\right)^j}{j!} + \frac{S^S}{S!} \sum_{j=S}^{\infty} \frac{\left(\frac{\lambda}{\mu}\right)^j}{S^j} \right]$$

$$\overline{g} = P_0 \frac{\lambda}{\mu} \left[\sum_{j=0}^{S-1} \left(\frac{\lambda}{\mu} \right)^j \frac{1}{j!} + \frac{S^S}{S!} \sum_{j=S}^{\infty} \left(\frac{\lambda}{S^j \mu} \right)^j \right]$$

Le terme entre crochets représente l'expression $\frac{1}{P_0}$, il reste alors:

$$\overline{g} = \frac{\lambda}{\mu}$$

- **=** g=λ/μ
 - résultat très simple, *principe de conservation:*
 - flux d'arrivées des clients est λ , personne ne disparaît dans le système, le flux de sortie doit être λ , hors saturation.
 - Ce nombre est égal au nombre moyen de guichets actifs g par le taux individuel de chacun μ .



{g guichets occupés}

- Nombre moyen de clients , v, dans la file:
 - Il existe, des clients dans la file d'attente, si k>S

Soient P_0, P_{S+i} les probabilités qu'il y ait respectivement 0 et (S+i) clients dans le système.

$$v = \sum_{i=0}^{\infty} i P_{S+i} = \sum_{i=0}^{\infty} i \left(\frac{\lambda}{S\mu} \right)^{S+i} \frac{S^{S}}{S!} P_{0}$$

$$v = P_0 \frac{S^{S}}{S!} \left(\frac{\lambda}{S\mu} \right)^{S+1} \sum_{i=1}^{\infty} i \left(\frac{\lambda}{S\mu} \right)^{i-1}$$

Posons:
$$\frac{\lambda}{Su} = x$$

$$ix^{i-1} = dérivée de x^i$$

- \square Nombre moyen de clients , ν , dans la file:
 - Il existe un file d'attente si k>S

$$\sum_{i=0}^{\infty} x^{i} = \frac{1}{1-x}$$

$$\sum_{i=0}^{\infty} ix^{i} = \frac{d}{dx} \left(\frac{1}{1-x} \right) = \frac{1}{(1-x)^{2}}$$
d'où:
$$v = P_{0} \frac{S^{S}}{S!} \left(\frac{\lambda}{S\mu} \right)^{S+1} \frac{1}{\left(1 - \frac{\lambda}{S\mu} \right)^{2}}$$

$$v = \frac{1}{S!S} \left(\frac{\lambda}{\mu} \right)^{S+1} \frac{1}{\left(1 - \frac{\lambda}{S\mu} \right)^{2}} P_{0}$$

3-2 FILE M/M/S

- \square Temps d'attente moyen, t_f , dans la file:
 - On utilise la formule de LITTLE

Appelons: \overline{v} , le nombre moyen de clients dans la file

 λ , le taux d'arrivée des clients

 $t_{\rm f}$, le temps d'attente moyen des clients dans la file

$$\overline{v} = \lambda t_{f}$$

$$t_{f} = \frac{\overline{v}}{\lambda}$$

$$t_{f} = \frac{1}{\lambda} \frac{1}{S!S} \left(\frac{\lambda}{\mu}\right)^{S+1} \frac{1}{\left(1 - \frac{\lambda}{S\mu}\right)^{2}} P_{0}$$

$$t_{f} = \frac{1}{S!} \left(\frac{\lambda}{\mu}\right)^{S} \frac{1}{S\mu} \frac{1}{\left(1 - \frac{\lambda}{S\mu}\right)^{2}} P_{0}$$

3-2 FILE M/M/S

- Nombre moyen de clients dans le système:
 - Somme du nombre moyen de clients dans la file, n, et du nombre moyen de clients occupant les guichets

$$\overline{N} = \overline{v} + \overline{g}$$

$$\overline{N} = \frac{\lambda}{\mu} (\mu \overline{t_f} + 1)$$

DEFINITION

- \square Arrivées: loi de Poisson de taux λ
- Un serveur
 - Loi de service: générale de moyenne μ
 - Capacité de file: infinie
 - discipline de service : FIFO
 - serveur n'est plus sans mémoire
 - Analyse plus complexe
 - Formule de Pollacek-Kintchine

- DESCRIPTION (instant t --->)
 - Description du système: <u>Vecteur dim. 2</u> $[N(t),X_0(t)]$
 - N(t): nombre de clients dans système
 - $X_0(t)$: durée de service déjà reçu
 - service : mémoire

- CHAINE DE MARKOV IMMERGEE
 - Description état système:
 - Vecteur DEUX dimensions --> <u>UNE dimension N(t)</u>
 - Prévision du comportement du système: nécessité de tenir compte de la durée du service déjà reçu.
 - POINTS D'OBSERVATION
 - sélectionnés dans le temps
 - opour lesquels la connaissance à t --->(N(t), loi d'arrivée)
 - 5 -->suffisant prévision N(t) au point sélectionné suivant
 - On tient ainsi compte implicitement du temps passé au service.
 - Si on connaît le nombre de clients laissés après le départ d'un client servi, on peut calculer le N(t) à l'instant du départ du client suivant.

- POINTS DE REGENERATION
 - SYSTEME MARKOVIEN
 - Espace d' état discret
 - Distribution du temps entre ces points est celle du temps de service chaque fois qu' un client quitte le serveur.
 - Si les arrivées sont *poissonniennes* le système décrit une chaîne de Markov si on ne l' observe qu' aux points de régénération

- v_{n+1} nb de clients arrivés pendant le service de C_{n+1} qui laisse la station à t_{n+1}
- Relations
 - Instant t_n -->q_n unités après départ du client C_n
 - Instant t_{n+1} --> q_{n+1} unités après départ du client C_{n+1}

- Le comportement de N(t) nombre de clients entre [t,t+dt] chaîne de Markov
 - La connaissance à t_n de N(t) et des transitions entre t_n et t_{n+1} suffit à prédire $N(t_{n+1})$

PROPRIETES DE LA FILE M/G/1

- ightharpoonup p_n(t) : Prob d'avoir n clients dans le système à t
- q_{n(}t) : Prob pour que lors de son arrivée le client en trouve n déjà présents dans le système à t
- $\Gamma_n(t)$: Prob un client partant, laisse n clients dans le système à t
- Observation du système à des instants aléatoires ou au moment d' une nouvelle arrivée ---->
 <u>même résultat</u>
- Propriété vraie car arrivées poissonniennes

PROPRIETES DE LA FILE M/G/1

- En régime permanent :

 - $q_n = \lim_{t \to -\infty} q_n(t)$
 - $= \lim_{t \to \inf} r_n(t)$
 - \bullet $p_n = q_n = r_n$
- En régime permanent, si on prend la statistique particulière du nombre de clients obtenu aux instants de départ des clients qui terminent leur service, on obtient le même résultat que si on avait pris les instants d'arrivées ou des instants aléatoires.

- NOMBRE MOYEN DE CLIENTS DANS LE SYSTEME
 - FORMULE P-K PAR UNE METHODE DIRECTE
 - Tout nouveau client, lors de son arrivée dans le système, trouve une des situations suivantes:
 - a- so: aucune unité dans le système
 - § --->le serveur est libre
 - § ----> occupation immédiate
 - \S ----> temps d' attente est nul: $E(T_f/0)=0$
 - § qo probabilité de cette situation

- b- S1: une unité dans le système
 - ---> le serveur est occupé.
 - ►(T_f/1)=E(R) qui représente l'espérance du temps d' attente dans la file si une unité est dans le système, celle-ci est égale au temps de service restant encore à donner.

$$E(T_f/1) = 1/\mu$$
 (si loi exponentielle)

Si la loi est générale :

$$E(R) = \frac{E(X^2)}{2E(X)} = \frac{\mu}{2}E(X^2)$$

X: variable aléatoire "durée de service"

q₁ probabilité de cette situation

- c- S2 : <u>deux clients dans le système</u>
 - ---> l'un occupe le serveur, fin de son service
 - \circ ---->une attente E(R).
 - \circ ---> L' autre est dans la file et subira un service complet de durée $(1/\mu)$.
 - Attente moyenne pour le nouvel arrivant:

$$E(T_f/2) = E(R) + 1/\mu$$

 \circ q_2 probabilité d'une telle situation

- □ d- S_n : <u>n clients dans le système</u>
 - ---> l'un occupe le serveur, fin son service
 - ---> attente E(R).
 - ---> (n-1) autres sont dans la file et subiront chacun un service complet de durée (1/μ).
 - L'attente moyenne pour le nouvel arrivant sera:

$$E(T_f/n) = E(R) + (n-1)/\mu$$

- \bullet q_n probabilité d'une telle situation
- Espérance du temps d' attente

$$E(T_f) = \sum_{n=0}^{\infty} E(T_f/n)q_n$$

Espérance du temps d' attente

$$\begin{split} & E(T_f) = \sum_{n=1}^{\infty} E(T_f/n) q_n \quad \text{car si } n = 0 \to T_f/0 = 0 \quad \text{et } p_n = q_n \\ & E(T_f) = \sum_{n=1}^{\infty} p_n \left[E(R) + \frac{n-1}{\mu} \right] \\ & E(T_f) = E(R) \sum_{n=1}^{\infty} p_n + \frac{1}{\mu} \sum_{n=1}^{\infty} (n-1) p_n \\ & E(T_f) = (1 - p_0) E(R) + \frac{1}{\mu} E(\nu) \end{split}$$

- \square Calcul de (1-p₀) en appliquant le principe de conservation des clients:
 - le taux d'arrivée est : λ
 - taux de service: $\mu=1/E(X)$
 - Taux de sortie = Taux service*Pr(guichet travaille) = $\mu(1-p_0)$
 - Taux d'arrivée = Taux de sortie(régime permanent)

$$\mu(1-p_0)$$

$$(1-p_0) = \lambda / \mu$$

$$p_0 = 1-\lambda / \mu$$

 \square E(v)?: en utilisant LITTLE

- \Box Le calcul E(\lor) s'effectue en utilisant LITTLE
 - $E(v) = \lambda E(T_f)$ en remplaçant dans l'expression de $E(T_f)$

$$\begin{split} & \mathsf{E}(\mathsf{T}_\mathsf{f}) = \frac{\lambda}{\mu} \mathsf{E}(\mathsf{R}) + \frac{\lambda}{\mu} \mathsf{E}(\mathsf{T}_\mathsf{f}) \\ & \left(1 - \frac{\lambda}{\mu}\right) \mathsf{E}(\mathsf{T}_\mathsf{f}) = \frac{\lambda}{\mu} \mathsf{E}(\mathsf{R}) \\ & \left(1 - \frac{\lambda}{\mu}\right) \mathsf{E}(\mathsf{T}_\mathsf{f}) = \frac{\lambda}{\mu} \frac{\mu}{2} \mathsf{E}(\mathsf{X}^2) = \frac{\lambda}{2} \mathsf{E}(\mathsf{X}^2) \\ & \mathsf{E}(\mathsf{T}_\mathsf{f}) = \frac{\lambda}{2} \frac{\mathsf{E}(\mathsf{X}^2)}{1 - \frac{\lambda}{\mu}} & \mathsf{Formule Pollaczek et Khintchine} \end{split}$$

- \square Application: Temps d'attente, T_f , dans la file M/M/1
 - Loi d'arrivée: exponentielle de moyenne 1/l
 - Loi de service : exponentielle
 - durée moyenne du service =1/μ
 - Calcul de E(X²)
 - Var(X)= E(X2) [E(X)]2= E(X2) $1/\mu^2$
 - $Var(X)=1/\mu^2$ Variance du service exponentiel
 - \circ E(X²) = 2/ μ ²
 - En remplaçant dans la formule P-K

$$E(T_f) = \frac{\lambda}{\mu(\mu - \lambda)} = \frac{\rho}{\mu(1 - \rho)}$$

- \square Application:Temps d'attente, T_f , dans la file M/D/1
 - Temps de service constant=1/μ
 - Calcul de $E(X^2)$

$$\circ$$
 $\sigma^2 = E(X^2) - [E(X)]^2 = E(X^2) - 1/\mu^2$

- \circ $\sigma^2=0$ (temps de service constant)
- \circ E(X2) = 1/ μ 2

$$E(T_f) = \frac{\lambda}{2\mu(\mu - \lambda)} = \frac{\rho}{2\mu(1 - \rho)}$$

Remarque: Pour une même cadence d'arrivées et de services la dispersion du service (sa nature aléatoire) entraîne le doublement du temps d'attente.

- Expression du temps d'attente total, T, dans le système
 - On a avec E(X) temps du guichet= 1/μ
 - Temps total:

$$E(T)=E(T_f)+E(X)$$

$$E(T) = \frac{1}{\mu} + \frac{\lambda E(X^2)}{2(1 - \frac{\lambda}{\mu})}$$

- Une nouvelle unité trouve lors de son arrivée:
 - guichet occupé avec une probabilité (1-p₀) et attend E(R)
 - -Une longueur de file d'attente, E(v) et il lui faut attendre, de ce fait, la résorption de cette file soit un temps $(1/\mu)E(v)$

Forme condensée du raisonnement "PK"

PRINCIPAUX RESULTATS

- État: nombre de clients, k, dans le système.
- Temps moyen inter-arrivée: 1/λ
- Temps moyen de service :1/μ
- \square Taux de trafic, ρ (charge, activité du serveur):

$$\rho = \frac{\lambda}{\mu}$$

Condition de stabilité:

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

Débit d'entrée < débit du serveur

Nombre moyen de clients dans le système:

$$\bar{N} = \rho + \rho^{2} \frac{1 + c^{2}}{2(1 - \rho)}$$

$$c^{2} = \frac{Var[temps_service]}{[moy(temps_service)]^{2}}$$

☐ N : fonction croissante de c^2 Si c^2 = 0 ---> service constant Si c^2 = 1 ---> service exponentiel

Temps moyen de réponse: $T_r=N/\lambda$

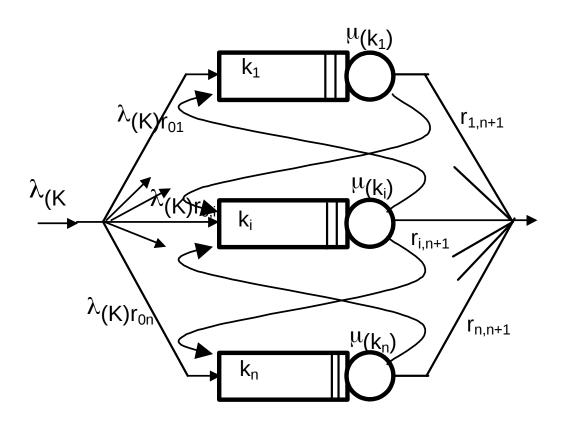
4- RESEAUX DE FILES D'ATTENTE

DEFINITION

- Ensemble de files d'attente interconnectées.
- Topologie arbitraire:
 - Ouvert
 - Fermé
 - Mixte
- Clients
 - Répartis en classes
 - Peuvent changer de classes lors d'une transition entre stations
- Arrivées(ouvert)
 - Poissonniennes de taux λ_i
 - Taux peut être fonction de l'état global du réseau:λ_i(K)

4- RESEAUX DE FILES D'ATTENTE

Réseau ouvert mono-classe



- THOREME DE JACKSON (1963)
 - Précurseur de BCMP (1975)
 - Application:
 - Stations : M/M/m/FCFS
 - Une seule classe de clients
 - Propriétés
 - Tous les types de stations sont tels que si le flot des entrées dans la station est poissonnien la sortie l'est aussi. Condition nécessaire de la forme produit. (Vrai pour BCMP).
 - Dans un réseau de Jackson ouvert, toutes les stations se comportent comme des stations isolées, M/M/m, soumises à un flot poissonnien d'intensité égale.

- Paramètres du réseau
 - $k=(k_1,k_2,...k_N)$ vecteur définissant l'état du système
 - ki: nombre de clients dans la station i.
 - λ(K)= processus d'arrivées des clients.
 - $\mu(n,k)$ =processus de service du serveur n
 - r(m,n) = probabilité de circulation des clients de la station, m, à la station, n, après avoir reçu le service de m.
 - r(0,n): probabilité pour que le serveur, n, soit le premier visité
 - r(m,N+1): probabilité pour que le serveur, m, dirige ses clients servis vers l'extérieur.

- Nous ferons deux hypothèses:
 - 1- Pour chaque valeur de m[[0,N], l'ensemble {r(m,n) sachant que n[[1,N+1]} est une distribution de probabilités.
 - **2-** Le système d'équations :

$$e(n) = r(0,n) + \sum_{m=1}^{N} e(m)r(m,n)$$
 avec $n \in [1,N]$ (1)

a une solution unique et e(n) est non négatif.

L'hypothèse (1) est nécessaire à l'interprétation de r(m,n). L'hypothèse (2) signifie que chaque circulation de client est finie. En effet l'espérance mathématique du nombre de serveurs visités par un client du système doit être finie, e(n) est l'espérance du nombre de passage d'un usager par la file d'attente du serveur n pendant son existence dans le système.(e(n) nombre de visites moyen d'un client à la station n).

- □ Solution stationnaire(1)
 - notations supplémentaires

$$W(k) = \prod_{i=0}^{K-1} \lambda(i)$$

$$K = 0,1,2,...$$

$$W(k) = \prod_{n=1}^{N} \prod_{i=1}^{k_n} \frac{e(n)}{\mu(n,i)}$$
pour le vecteur d'état $k = (k_0, k_1,..., k_N)$

$$T(K) = \sum_{i=0}^{N} w(k)$$

Somme effectuée sur tous les vecteurs d'état k avec S(k)=K

$$y = \frac{1}{\sum_{K=0}^{\infty} W(K)T(K)}$$
 Si la somme converge

y=0 Si la somme ne converge pas

□ Solution stationnaire(2)

Théorème:

La solution stationnaire p(k), de la distribution de probabilité satisfaisant le système d'équations différentielles du réseau existe et est égale à:

$$p(k) = \gamma w(k) W(S(k))$$

si γ >0 et si la solution e(n) de (1) est unique.

- Solution stationnaire des probabilités d'états Réseaux ouverts à taux de service constants:
 - Chaque file du réseau admet isolément la même solution que M/M/1:

$$\begin{split} & \rho_i = & \frac{\lambda_i}{\mu_i} = \frac{\lambda e_i}{\mu_i} < 1 \qquad \forall i, \text{ condition d'ergodicite} \\ & \begin{cases} P_i(k_i) = & (1 - \rho_i) \rho_i^{k_i} \\ \hline \overline{n}_i = & \sum_j^\infty j. P_i(j) = \frac{\rho_i}{1 - \rho_i} \end{cases} \end{split}$$

Vecteur d'etat du reseau:

$$P(k) = \prod_{i} P_{i}(k_{i}) = \prod_{i} (1 - \rho_{i}) \rho_{i}^{k_{i}}$$

- Inconvénients du modèle de Jackson
 - Réseau monoclasse de clients.
 - Restriction des temps de service à des lois exponentielles et lois d'arrivées à des lois de Poisson.
 - Discipline de service FIFO.
 - Ne peut prendre en compte des lois et des disciplines de service dépendant de l'état total du système.

RESEAUX BCMP

- Réseaux ouverts, fermés ou mixtes.
- Clients: répartis en classes avec changements possibles.
- Stations:
 - capacités infinies
 - type définis
- Routage:
 - {R_{ic;jc'}}= Probabilité (statique) pour un client de classe =C de quitter la station, i, pour aller à la station j avec la classe C'.
- Arrivées(ouvert)
 - Poissonniennes de taux λ_i
 - Taux peut être fonction de l'état global du réseau:λ_i(K)

Probabilité d'états

- P(k) = probabilité stationnaire de l'état k.
- \Box C = Constante de normalisation.
- d(K) = ne dépend que du processus d'arrivée(d(K)=1 si réseau fermé)
- \Box g_i(k_i) = ne dépend que de la station
- \Box La solution stationnaire P(k) n'existe que si

$$C^{-1} < +\infty$$
 (non saturation)

$$P(k) = C.d(K).\prod_{i} g_{i}(k_{i})$$

- Types de stations
 - Type1:
 - FIFO
 - Temps de service exponentiel
 - Taux de service identique pour toutes les classes, peut dépendre du nombre total de clients présents dans la file: $\mu_i(k_i)$.
 - Type1:
 - PS: processor sharing
 - Temps de service général(Loi de Cox)
 - Taux de service différents par classe, peut dépendre du nombre de clients de la classe présents et ou du nombre total de clients dans la file: μ_{ic}(k_i,k_{ic}).

- Types de stations
 - Type3:
 - Infinie
 - Nombre de serveurs ≥ au nombre instantané de clients présents.
 - identique au type 2.
 - Type4:
 - LIFO-PR: Last Come First Served-Préemptive Resume)
 - identique au type 2.
 - Stations fictives
 - Station "0": Source exponentielle de clients taux d'émission fonction du nombre total de clients dans le réseau: λ(K)
 - Station N+1: "Puits" de clients = sortie

CALCUL DES g_i(k_i)

\Box {e _{ic} }	= nombre de visites d'un client de classe, c, à la
	station, i.

$$\mathbf{e}_{jc} = \mathbf{r}_{0;jc} + \sum_{c'} \sum_{i} \mathbf{e}_{ic'} \cdot \mathbf{r}_{ic';jc} \qquad \forall j, c$$

- ☐ k_{ic}=nombre instantané de clients de classe, c, dans i
- \square $k_i = (k_{i1}, k_{i2}, ..., k_{in})$: vecteur d'état de la station i
- \square $K_i = \sum k_{ic}$: nombre total des clients de la station i
- = k=(k₁, k₂,...): matrice d'état du réseau
- \sqsubseteq K= \sum k_i: nombre total des clients du réseau
- μ_{ic} = taux de service de la station, i, pour les clients de classe, c.
- La solution stationnaire n'existe que si $C^{-1} < +\infty$.

4-2 RESEAUX DE FILES D'ATTENTE THEOREME BCMP

CALCUL DES g_i(k_i)

Pour les services indépendants de l'état:

$$g_{i}(k_{i})! \prod_{c} \frac{1}{k_{ic}!} \left(\frac{e_{ic}}{\mu_{i}}\right)^{k_{ic}} \quad \text{station de type1}$$

$$g_{i}(k_{i}) = \begin{cases} (K_{i})! \prod_{c} \frac{1}{k_{ic}!} \left(\frac{e_{ic}}{\mu_{ic}}\right)^{k_{ic}} & \text{stations de type 2 et 4} \\ \prod_{c} \frac{1}{k_{ic}!} \left(\frac{e_{ic}}{\mu_{ic}}\right)^{k_{ic}} & \text{stations de type3} \end{cases}$$

$$d(K) = \begin{cases} 1 & \text{reseau ferme} \\ \prod_{m=0}^{K-1} \lambda(m) & \text{reseau ouvert} \end{cases}$$

$$C = \left[\sum_{k} d(K) \prod_{i} g_{i}(k_{i})\right]^{-1}$$

4-2 RESEAUX DE FILES D'ATTENTE THEOREME BCMP

REMARQUES

- 1- La forme des distributions de temps de service des stations de type2, type3 et type4 est indifférente, seules interviennent les moyennes des distributions, μ_{ic}-1
- 2- La topologie du réseau et le fractionnement des demandes service sont indifférentes quant à l'équilibre stationnaire, seules comptent les demandes globales de service, (e_{ic}/ μ_{ic}).
- 3- Numériquement, les formules du BCMP ne sont applicables directement que sur des cas simples.

4-2 RESEAUX DE FILES D'ATTENTE THEOREME BCMP

REMARQUES

4- En cas de taux de service dépendants, substituer :

 $\square \mu_i$ ----> μ_i $\alpha_i(K_i)$ pour station de type 1

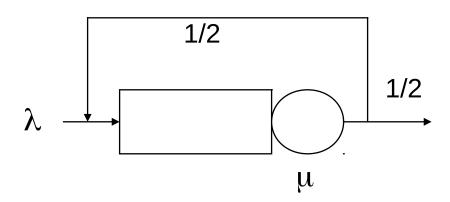
 $\square \mu_{ic}$ ----> μ_{ic} $^{0}a_{i}(K_{i})$ b_{ic} (k_{ic}) pour stations de type 2,3,4 $\{a_{i}\}$ et $\{b_{ic}\}$ sont des fonctions positives quelconques.

5- Pour les réseaux fermés, les $\{e_{ic}\}$ ne sont déterminés qu'à une constante multiplicative près. On les rapporte à un passage unique en un point distingué du réseau e_{i0c0} =1.

- RESEAUX BCMP "EQUIVALENTS"
 - Trois réseaux apparemment différents
 - Le théorème BCMP s'applique
 - λ: taux d'arrivée des clients suivant un processus "général"
 - μ: taux de service du serveur exponentiel
 - La topologie précise des réseaux BCMP est indifférente quant à l'état stationnaire: P(k).
 - Temps de séjour moyens identiques (Little).
 - Seules comptent les demandes de service cumulées(=e_i/μ_i).
 - La distribution complète des temps de séjour n'est pas indifférente aux distributions.

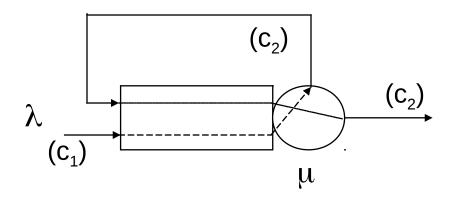
RESEAUX BCMP "EQUIVALENTS"

- File:G/M/1
- Une seule classe de clients
- nombre de passages au serveur suit une loi géométrique: $n=1+1/2+(1/2)^2+...=2$
- Temps=2/μ

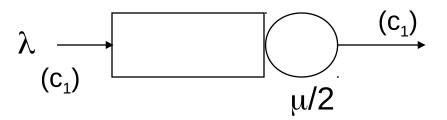


RESEAUX BCMP "EQUIVALENTS"

- File :G/M/1
- Une seule classe initiale de clients
- Les clients changent de classe au premier service. Il y a deux classes de clients après le premier service. Le client quitte la station au deuxième service reçu.
- nombre de passages au serveur n=2
- Temps=2/μ



- RESEAUX BCMP "EQUIVALENTS"
 - File :G/M/1
 - Une seule classe initiale de clients
 - Temps de service = $2/\mu$



4-3 RESEAUX DE FILES D'ATTENTE APPROXIMATION

METHODES ITERATIVES

- PRINCIPE
 - Chaque file est étudiée séparément, en présence de son "complémentaire": $\lambda(n)/G/1/N$.
 - Un réseau auxiliaire BCMP est utilisé pour le calcul du complémentaire, résultats approchés.
 - Les caractéristiques du réseau auxiliaire sont modifiées à chaque itération pour réduire l'erreur sur les invariants.
- Utilité
 - réseaux fermés mono-classe non BCMP.
- Précision
 - inconnue sauf comparaison à une méthode exacte.
 - Réputée bonne en général.

GENERALITES

- Outil flexible et utile dans l'évaluation des performances.
- Pallie certaines limitations de la modélisation analytique.
- Modèle "simulable" avec niveau arbitraire de détails peut être décrit.
 - Cela permet de prendre en compte des situations très complexes.
 - ATTENTION aux erreurs.
- Peut être utilisé pour valider un modèle analytique.

TYPES DE SIMULATEURS

- Simulateurs de trace:
 - Simulation d'un modèle déterministe commandé par une trace obtenue sur un système existant.
 - Le modèle n'a pas obligatoirement une structure en file d'attente.
 - Si ce modèle est de mise en œuvre facile, il est possible de vérifier l'influence de paramètres déterminants.

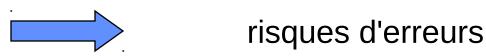
- ☐ Simulateurs à événements stochastiques discrets
 - Simulation à événements discrets du modèle
 - Simulation du comportement dynamique des éléments constituants le modèle(stations, clients, ressources,...)
 - Le comportement est analysé aux instants où un événement a lieu dans le modèle(transition de clients, début de service...)
 - Générateurs de nombres aléatoires utilisés pour reproduire le comportement stochastique des composants du modèle
 - odistribution dans le temps des services, probabilités de transitions,...

ANALYSE DES RESULTATS

- Modèles commandés par des séquences d'entrées aléatoires.(Temps de service, transitions,...)
- Obtient des séquences de sorties aléatoires(flux de clients, temps de réponse,..) ---> débits, temps de réponse
 - Précision des résultats, temps de simulation.
 - Intervalle de confiance

CONCLUSION

- Si les systèmes à analyser sont complexes et que des simplifications ne sont pas possibles...
- Nécessité de beaucoup de temps calcul.
- Attention à la complexité du modèle



6- COMPARAISON MODELE ANALYTIQUE, MODELE SIMULABLE

- MODELE ANALYTIQUE
 - Avantages
 - Simplicité conceptuelle et programmation(+++)
 - Temps de calcul faible(++)
 - Solution stationnaire garantie(+)
 - Inconvénients
 - Hypothèses de résolution contraignantes(---)
 - Résultats limités aux moyennes(-)

6- COMPARAISON MODELE ANALYTIQUE, MODELE SIMULABLE

- MODELE SIMULABLE
 - Avantages
 - Absence de contrainte de modélisation(+++)
 Attention! Absence de contrainte
 - Inconvénients
 - Difficultés de validation si complexe (---)
 - Temps de calcul élevé (-)
 - Précision statistique à évaluer(-)
 - Résultats limités aux moyennes(-)

7- CONCLUSION

- BUT DES MODELES
 - Aider le designer à répondre aux questions:
 - Quelles performances à quel coût ?
- MODELES
 - (file d'attente, réseaux de Pétri,) ne sont pas parfaits...
 - ce qu'il y a de mieux pour évaluer les performances des systèmes
 - Succès dû à deux facteurs
 - puissance de représentation
 - pouvoir de décision
 - détermination des propriétés du système par analyse du modèle.

7- CONCLUSION

- Choix entre modèle analytique ou simulable:
 - circonstances
 - compétences
 - de goûts
 - des outils
 - du prix.....
- BILAN
 - Précision des évaluations produites est du même ordre de grandeur que les données d'entrée
 - Nécessité absolue d'effectuer des mesures(logiciel, matériel)

8- BIBLIOGRAPHIE

PROCESSUS DE MARKOV ET FILES D'ATTENTE

KLEINROCK, L
 Queuing Systems: Volme 1 Theory; Volume 2: Computer Applications
 John Wiley and Sons

RAJ JAIN

The art of computer systems performance analysis

John Wiley and Sons

FELLER,W.

An Introduction To Probability Theory And Its Applications 3e ed 1968 J. Whiley & Sons

CINLAR, E

An introduction to stochastique processes

Prentice-Hall

COX, D.R

A use of complex probabilities in the theory of stochastic processes *Proceedings Cambridge Philosophical society, 51, 1955, [313-319]*

R.FAURE, B. LEMAIRE, C PICOULEAU

Précis de Recherche Opérationnelle

5ème Edition Dunod

8- BIBLIOGRAPHIE

- RESAUX DE FILES D'ATTENTE
- JACKSON J.R. Jobshop-like queuing systems Management science Vol. 10 No. 1 october 1963
- F.BASKETT, K.M. CHANDY, R.R. MUNTZ, F.G. PALACIOS

 Open, Closed, and Mixed Networks of Queues with différent Classess of Customers

 J.A.C.M. Vol.22 no.2, April 75 pp248-260
- [E. GEL 82] E. GELENBE

 Résultats Classiques Forme Produit [12-39]

 Eyrolles-82
- LEMAIRE, B

Contribution A L'étude Des Systèmes d'Attente Par Les Méthodes De Conservation. Application à des Problèmes de R.O. et d'Informatique .

Thèse doctorat d'état -78

[ROB.90] T.G., ROBERTAZZI

Computer Networks And Systems

Queuing Theory and Performance Evaluation

Telecommunication Network and computer systems

8- BIBLIOGRAPHIE

- RESEAUX DE FILES D'ATTENTE
- [CHAN 72] K.M.CHANDY

 The analysis and solutions for general queueing network

 Proceedings 6th annual Princeton conference on information sciences and systems, Princeton University(N.J.) Mars 72 p 224,228
- A.E. CONWAY, N.D. GEORGANAS

 Queuing Network Exact Computational Algorithms

 A Unified Theory based on Decomposition and aggregation

 The M.I.T press
- [COX 55] COX, D.R

 A use of complex probabilities in the theory of stochastic processes

 Proceedings Cambridge Philosophical society, 51, 1955, [313-319]
- [WHIT 68] P. WHITTLE

 Equilibrum for an open migration process

 Jo1. Appl. Prob. 5, 1968 p.567-571