



Gestion des Stocks et File d'attente

Guy Aimé TANONKOU

Industrial Engineering & Computer Science
INRIA-Lorraine

Ile du Saulcy, Bât.A ISGMP
57045 Metz Cedex 01 France

Phone (office) : 00 33 3 87 54 72 97

Email : tanonkou@loria.fr / tanonkou@acsal-science.org

Table des matières

1. La Gestion des Stocks	3
1.1 Les politiques de Gestion des stocks	3
1.2 Coûts associés aux Stocks.....	3
1.3 Gestion calendaire des Stocks à rotation nulle	5
1.3.1 Cas où la Demande D suit une loi discrète	5
1.3.2 Cas où la Demande D suit une loi continue.....	7
1.3.3 Cas d'une loi de demande continue	8
1.3.4 Cas d'une loi de demande discrète	9
2. File d'attente	10
2.1 Evaluation de performance	11
2.2 Etat du système	13
2.3 Quelques cas particuliers	15
3. Référence	21

1. La Gestion des Stocks

Une production sans stock est quasi inconcevable vu les nombreuses fonctions que remplissent les stocks. Nous présentons dans les lignes qui suivent, l'étude moins détaillée que nous avons faite en ce qui concerne la gestion des stocks. Nous avons étudié les problèmes liés aux stockages et nécessité de constituer un stock et puis les politiques de Gestion des stocks qui feront l'objet de notre étude.

1.1 Les politiques de Gestion des stocks

Cette politique vise à répondre à deux grandes questions : Quand déclencher l'approvisionnement du stock? Combien commander ? La réponse à ces deux questions dépend de la politique de gestion adoptée. Nous nous sommes intéressés à deux politiques :

1. **La politique de gestion de stock par point de commande** : dans ce cas, l'approvisionnement du stock est déclenché lorsque l'on observe que le stock descend en dessous d'un niveau s appelé *point de commande*. On commande une quantité fixe notée Q et appelée « *quantité économique de commande* ». Sa détermination résulte d'un calcul d'optimisation.
2. **La politique de gestion calendaire** : dans ce cas, l'approvisionnement du stock est déclenché à intervalles réguliers T , par exemple, chaque jour ou chaque semaine. La quantité commandée est égale à la différence entre le stock résiduel observé R et le niveau de rechargement du stock S c'est à dire le niveau voulu du stock en début de période T .

Nous avons fait une étude plus détaillée sur ces deux politiques de gestion adoptée, suivant que la loi de demande est discrète ou continue.

1.2 Coûts associés aux Stocks

Tout d'abord, nous présentons ici, les différents coûts observés et associés aux stocks :

On pose x = **demande observée au cours d'une période**

- Un stock est constitué pour satisfaire une demande future. En cas de demande aléatoire, il peut y avoir non coïncidence entre la demande et le stock. Dans cette situation, deux cas de figure se présente.
 - soit la demande $x > stock$ alors on parle de *Rupture de stock*
 - soit la demande $x < stock$ alors il y'a un *stock résiduel R*

Dans le cas où $x > stock$, on a le **coût de rupture** $C_r = c_r I_r$ où

$I_r =$ nombre moyen de demandes non satisfaites au cours d'une période
 $c_r =$ coût unitaire de rupture

- Le **coût de commande** C_c (en cas de stock de fabrication, il s'agit du coût de lancement de la production. Par exemple, réglages des machines, etc. En cas de stock d'approvisionnement, il s'agit des coûts administratifs. Par exemple, établissement d'un bordereau, contrôle de livraison, etc.)

$C_c = c_c I_c$ où

$I_c =$ nombre moyen de commande passée au cours d'une période
 $c_c =$ coût unitaire de commande

- Le **coût de possession** C_p (il s'agit des coûts de détention d'un article en stock durant une certaine période en fonction des conditions financières d'acquisition et des éventuelles conditions de reprise, des coûts de stockage qui sont les dépenses logistiques, conservation du stock, coût de location d'entrepôt, coût de manutention, etc.). On a $C_p = c_p I_p$ où

$I_p =$ nombre de stock possédé au cours d'une période
 $c_p =$ coût unitaire de possession

Le coût total de gestion C peut s'écrire comme fonction de ces trois variables d'état,

$$C = C_p + C_r + C_c \quad \text{soit} \quad C = c_p I_p + c_r I_r + c_c I_c$$

Le critère généralement retenu en gestion des Stocks est celui de l'optimisation des coûts qui sont fonction des variables de commande du système.

En **Gestion calendaire des stocks**, la seule variable de commande du système est S , le niveau de rechargement du stock, c'est à dire le stock que l'on cherche à atteindre en fin de période. Dans ce cas, le coût total de gestion est

$$C(S) = c_p I_p(S) + c_r I_r(S) + c_c I_c(S)$$

En **Gestion par point de commande**, les deux variables de commande du système sont Q , la quantité économique de commande et s le point de commande. Dans ce cas, le coût total de gestion est

$$C(Q, s) = c_p I_p(Q, s) + c_r I_r(Q, s) + c_c I_c(Q, s)$$

Comme nous l'avons mentionné plus haut, une étude plus détaillée sur ces deux politiques de gestion adoptées, a été étudiée suivant que la loi de demande est discrète ou continue.

On a vu que

$$\begin{cases} x > stock \Rightarrow \text{rupture de stock} \\ x < stock \Rightarrow \text{présence d'un stock résiduel} \end{cases}$$

Le cas $x = stock$ ne pose aucun problème

Dans le cas où on a un **stock résiduel non nulle**, deux cas de figures se présentent :

- Soit il n'y a pas de report possible des invendus aux périodes suivantes c'est à dire que la vente est perdue. C'est le cas des pâtisseries et des marchands de journaux. On parle de **stock à rotation nulle**
- Soit les invendus d'une période sont vendus aux périodes suivantes. C'est le cas le plus répandu. On parle de **stock à rotation non nulle**

1.3 Gestion calendaire des Stocks à rotation nulle

L'approvisionnement du stock est déclenché à intervalles réguliers T

Problème : Déterminer le niveau de stock initial qui minimise le coût de gestion total. Soit la variable de décision S :

$$S = \text{niveau de stock voulu ou niveau de rechargement du stock}$$

1.3.1 Cas où la Demande D suit une loi discrète

On pose $x =$ demande observée au cours d'une période

$I_c(S) = 1$ puisqu'on passe une commande sur une période

On a donc

$$C(S) = c_p I_p(S) + c_r I_r(S) + c_c \quad (1)$$

Le nombre moyen de rupture est

$$I_r(S) = \sum_{x=S+1}^{\infty} (x-S)P(D=x) \quad (2)$$

Le stock moyen possédé est $I_p(S) = \sum_{x=0}^{S-1} (S-x)P(D=x)$ car le stock moyen possédé en fin de période correspond à l'inventu. On montre que

$$I_p(S) = S - \bar{D} + I_r(S) \quad (3)$$

Condition d'optimalité

Le coût de gestion étant une fonction convexe, le stock optimal S^* est celui pour lequel le coût de gestion est inférieur à celui des stocks immédiatement inférieur ou supérieur c'est à dire que

$$\begin{cases} C(S^*) < C(S^* + 1) \\ C(S^*) < C(S^* - 1) \end{cases} \quad (4)$$

En remplaçant (3) dans l'expression du coût de gestion (1), on observe que

$$\begin{cases} C(S) = c_p (S - \bar{D}) + (c_r + c_p) I_r(S) + c_c \\ I_r(S+1) - I_r(S) = -P(D > S) \end{cases} \quad (5)$$

On fait donc une étude de la différence des coûts de stocks successifs

$$C(S+1) - C(S),$$

on trouve

$$\begin{cases} C(S+1) - C(S) = c_p - (c_r + c_p) P(D > S) \\ C(S-1) - C(S) = c_p - (c_r + c_p) P(D > S-1) \end{cases} \quad (6)$$

En tenant compte des conditions d'optimalité (4)

$$\text{On cherche } S^* \text{ telle } P(D > S^*) < \frac{c_p}{c_p + c_r} < P(D > S^* - 1) \quad (7)$$

1.3.2 Cas où la Demande D suit une loi continue

Dans ce cas, on considère f la densité de probabilité de D . On fait un raisonnement analogue au cas discret en remplaçant le signe somme par l'intégrale.

On montre que

$$\frac{dI_r(S)}{dS} = -\int_S^{\infty} f(x)dx = -P(D > S)$$

On peut maintenant passer à la détermination de la solution optimale S^* .

On cherche ainsi S^* telle que $\frac{dC(S)}{dS} = 0$ et

$$S^* \text{ est optimal si } P(D > S^*) = \frac{c_p}{c_p + c_r}$$

Gestion calendaire des Stocks à Rotation non nulle

Nous rappelons dans cette situation que, les invendus sont vendus à une période suivante.

L'approvisionnement du stock est déclenché à intervalles réguliers T .

La commande à passer pendant la période T n'est plus fixe.

Problème : Déterminer le niveau de stock initial qui minimise le coût de gestion total.

Variable de décision S :

$S =$ niveau de stock voulu ou niveau de rechargement du stock

On a deux cas de figures :

- Si le stock résiduel est positif, on commande la différence entre S et le stock résiduel
- Si le stock résiduel est nul, on commande S augmenté des demandes non satisfaites de la période précédente qui ont pu être reportées.

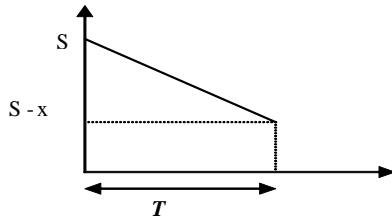
$x =$ demande observée au cours d'une période

On a toujours $C(S) = c_p I_p(S) + c_r I_r(S) + c_c$

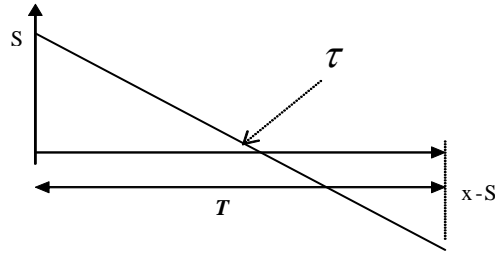
Pour le calcul du stock moyen possédé, il faut distinguer deux cas :

Le cas où $x < S$ et le cas $x > S$

Cas où $x < S$



Cas où $x > S$



Dans le cas où $x > S$ on peut écrire l'équation d'évolution du stock $S(t) = S - \frac{x}{T}t$

$$\text{Donc } \tau = \frac{S}{x/T}$$

Le stock moyen possédé sur τ est $\frac{S+0}{2}$, dont sur la période T on a en moyenne

$$\frac{S+0}{2} \frac{\tau}{T}, \text{ soit } \frac{S^2}{2x}$$

Dans le cas où $x < S$, on a en moyenne en stock $\frac{S+S-x}{2}$, soit $S - \frac{x}{2}$

1.3.3 Cas d'une loi de demande continue

$$I_p(S) = \int_{x=0}^S (S - \frac{x}{2})f(x)dx + \int_{x=S}^{\infty} \frac{S^2}{2x}f(x)dx \quad (8)$$

$$I_r(S) = \int_{x=S}^{\infty} (x - S)f(x)dx$$

Le calcul du coût de Gestion est extrêmement *compliqué* (Le résultat analytique de ce calcul reste un challenge). C'est la raison pour laquelle, on fait une hypothèse simplificatrice à savoir que la rupture se fait en fin de période. Dans ce

cas on a en stock en moyenne $\frac{S}{2}$ et on a

$$I_p(S) = \int_{x=0}^S (S - \frac{x}{2})f(x)dx + \int_{x=S}^{\infty} \frac{S}{2}f(x)dx$$

On montre que $I_p(S) = S - \frac{\bar{D}}{2} + \frac{I_r(S)}{2}$ et

$$C(S) = c_p(S - \frac{\bar{D}}{2}) + (c_r + \frac{c_p}{2})I_r(S) + c_c$$

$$S^* \text{ est optimal si } P(D > S^*) = \frac{c_p}{\frac{c_p}{2} + c_r} \quad (9)$$

1.3.4 Cas d'une loi de demande discrète

On fait un raisonnement analogue en changeant le signe intégral par le signe somme.

Gestion par point de commande

On rappelle ici que *la politique de gestion de stock par point de commande* est adoptée lorsque l'on observe que le stock descend en dessous d'un niveau s appelé **point de commande**. On commande une quantité fixe notée Q appelé « *quantité économique de commande* ».

Variable de décision (Q, s)

On a deux cas de figures :

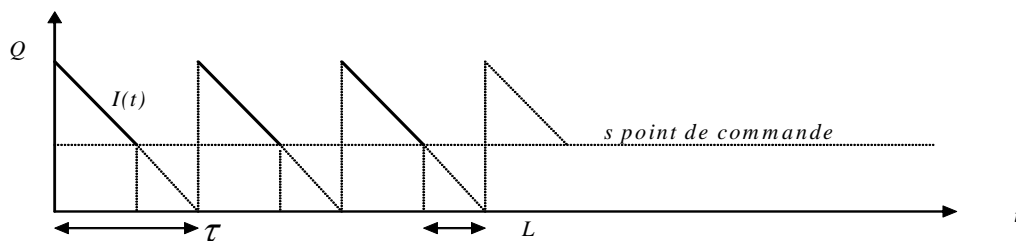
- **Demande certaine** : dans ce cas, on commande avant rupture de stock et il y'a pas de coût de rupture. Cela suppose que le point de commande s est connu à l'avance. La seule variable de décision est donc Q quantité de la commande, déterminer de manière à minimiser le coût total de gestion

$$C(Q) = c_p I_p(Q) + c_c I_c(Q)$$

Soit D la demande sur une année par exemple

L = délai de réapprovisionnement supposé connu

$I(t)$ = évolution du stock pendant la période



$\tau = \frac{Q}{D}$ et le stock moyen possédé pendant la période τ est

$$\int_0^{\tau} I(t) dt = \frac{\tau Q}{2} = \frac{Q^2}{2D}$$

On a en déduit le stock moyen possédé annuel $I_p(Q) = \frac{Q^2}{2D} \times \frac{D}{Q} = \frac{Q}{2}$ où $\frac{D}{Q}$ est le nombre de commande par an. On a donc

$$C(Q) = c_p \frac{Q}{2} + c_c \frac{D}{Q} \Rightarrow C'(Q) = \frac{c_p}{2} - c_c \frac{D}{Q^2}$$

$$Q^* \text{ est optimal si } Q^* = \sqrt{\frac{2Dc_c}{c_p}} \text{ Formule de Wilson} \quad (10)$$

En général $s = DL$

- **Demande incertaine** : dans ce cas, la quantité de commande et le point de commande ne sont pas connus, de ce fait, il y'a possibilité de rupture de stock. Le coût total de gestion est donc une fonction à deux variables (Q, s) définie par

$$C(Q, s) = c_p I_p(Q, s) + c_c I_c(Q, s) + c_r I_r(Q, s)$$

Une étude analytique de ce cas s'avère très compliqué. C'est la raison pour laquelle l'on procède par des approximations et par simulation. Il faut distinguer le cas où la demande est aléatoire et L certain et celui où la demande est aléatoire et L aléatoire. Nous étudierons prochainement les approximations utilisées pour approcher le problème et les résultats obtenus par simulation.

Ceci achève l'étude que nous avons effectué en Gestion des Stocks.

2. File d'attente

Nous avons fait une étude sur les notions de files d'attente et leurs applications aux systèmes industriels. On s'intéresse ici au phénomène d'attente qui peut être ramené de façon générale au problème suivant : des clients/pièces se présentent dans un lieu donné pour obtenir un service. La figure 1 présente le modèle de base.

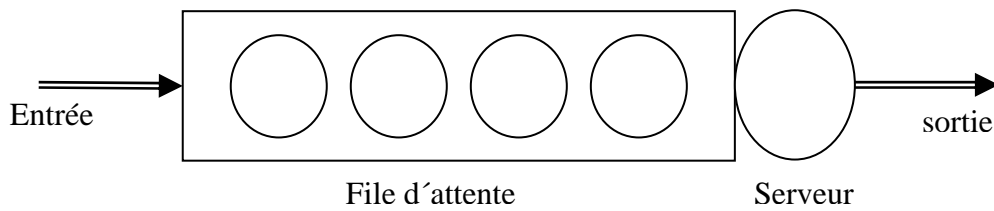


Figure 1. Model de

2.1 Evaluation de performance

Nous nous sommes préoccupés aux questions suivantes (*Mesures de performances*) :

- Combien y a t'il couramment de clients (ou pièces) dans le système (en attente ou train d'être servis) ?
- Combien y a t'il de clients ou pièces dans la queue ?
- Quel est, en moyenne, le temps passé par un client ou pièce dans le système ?
- Quel est, en moyenne, le temps passé par un client à attendre dans la queue ?
- Quel est le taux d'occupation du serveur ?
- Quel est le taux de sortie du système ?

La réponse à ces questions nécessite un plan d'étude qui suit :

1. Description des Processus.

- processus des arrivées (les clients ou pièces arrivent de manière déterministe selon un planning donné, ou bien arrivent de manière aléatoire)
- processus de service (le temps de service peut être déterministe ou aléatoire, c'est à dire le temps de service est indépendant de l'historique passé et peut ainsi être décrit par la distribution exponentielle négative)

Les cas couramment rencontrés sont ceux où l'arrivée des pièces suit une distribution de Poisson et le temps de service une distribution Exponentielle.

2. Gestion de la file d'attente

- La capacité de la file d'attente peut être limitée (par exemple la salle d'attente d'un cabinet médical) ou illimité (par exemple les péages de l'autoroute)
- Discipline de service. Les clients peuvent être servi suivant les règles de gestion.
 - o FIFO (Premier entrée – Premier sortie)
 - o LIFO (Dernier entrée – Premier sortie)
 - o Choix suivant les règles de priorité et etc.
- Comportement des clients
 - o Les clients peuvent être *patients* (s'ils attendent patiemment leur tour)
 - o Les clients peuvent être *impatiens* (s'ils refusent à rejoindre la file longue)
 - o Retraitement des pièces non conformes

3. Organisation du système

On peut avoir les cas suivant :

- o Une file d'attente et un serveur
- o Une file d'attente et plusieurs serveurs
- o Plusieurs files d'attente et un serveur
- o Un Réseau de files d'attente

Conventionnellement, on représente une file d'attente par la *notation de Kendall* $A/B/m/n$

Où A décrit le processus des arrivées ; B décrit le processus de service ; m le nombre de serveur ; et n le nombre de clients dans la file. On peut avoir $n, m = \infty$

Couramment, lorsque le processus est Exponentielle, on utilise la lettre M , et lorsqu'il est déterministe on utilise la lettre D , et enfin lorsque le processus suit une loi de distribution générale, on utilise la lettre G

2.2 Etat du système

Il est important de décrire l'état du système. En général, c'est le nombre de clients ou de pièces dans le système, en service et en attente.

Soit $M(t)$ = nombre de clients / pièces dans le système en service et en attente

Pour décrire le comportement du système, on utilise

$$P_n(t) = P(M(t) = n)$$

On suppose au début que le système est dans un état connu. C'est à dire que à l'instant $t = 0$, il n'y a pas de pièces (clients) dans le système. On a alors

$$\begin{cases} M(0) = 0 \\ P_0(0) = 1 \\ P_i(0) = 0, \quad \forall i \geq 1 \end{cases} \quad (11)$$

Dans les instants immédiats qui suivent, un client peut arriver, **mais sûrement pas deux**. Dans ce cas, si l'on fait une prédiction sur l'état du système en t très petit, on aura

$$\begin{cases} P_0(t) \approx 1 \\ P_1(t) \approx 0 \\ P_i(t) = 0, \quad \forall i > 1 \end{cases} \quad (12)$$

Lorsque t devient grand, les probabilités des états $i = 2, 3, \dots$ deviennent non nulles.

Question : Quel est le comportement de $P_i(t)$ lorsque $t \rightarrow \infty$

Pour répondre à cette question, nous faisons appel à la Théorie sur les Processus aléatoires, en particulier le Processus de Naissance et de Mort.

« Une modélisation et description réaliste d'un système de file d'attente doit évidemment tenir compte à la fois des naissances et des morts des clients/pièces qui la compose » Supposons λ_i le taux d'arrivée des clients/pièces et μ_i le taux de service lorsque le système est à l'état i . Dans un petit intervalle de temps Δt , la transition de $i \rightarrow i+1$ correspond à une naissance avec la probabilité $\lambda_i \Delta t$ et la transition de $i \rightarrow i-1$ correspond à un décès avec la probabilité $\mu_i \Delta t$, et rien ne se produit avec une probabilité $(1 - (\lambda_i + \mu_i)) \Delta t$. On dit que le processus des arrivées

est un processus de naissance. Donc $M(t)$ est un processus de naissance et de mort. On a le diagramme de transition d'état suivant :

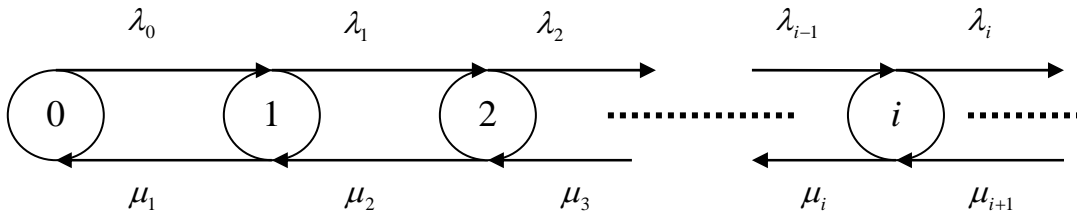
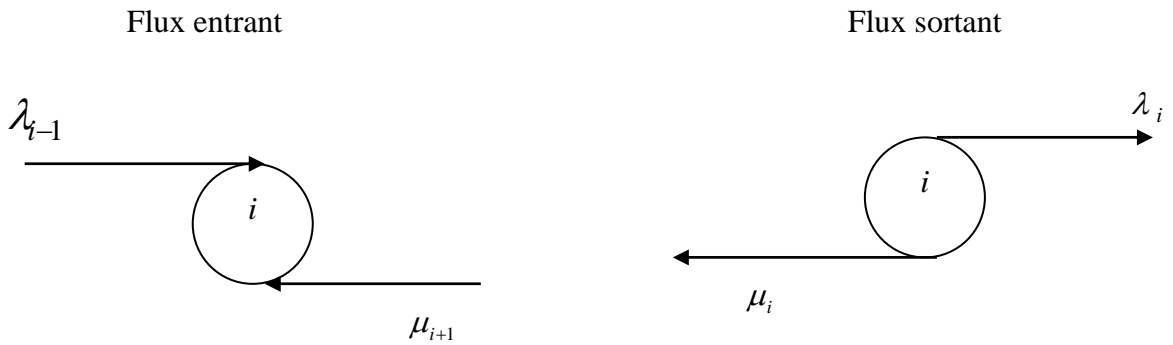


Figure. 2 Diagramme de Transition d'état

Ce diagramme de transition est une chaîne de Markov. Le flux sortant à l'état i est égal au flux entrant à l'état i .



$$\lambda_{i-1}P_{i-1}(t) + \mu_{i+1}P_{i+1}(t) = (\lambda_i + \mu_i)P_i(t) : \text{Equations d'état d'équilibre (*)}$$

Les taux de naissance $\lambda_0, \lambda_1, \dots$ et les taux de mort μ_0, μ_1, \dots sont telles que

$$\lambda_i \geq 0, \mu_i > 0, \mu_0 = 0$$

En faisant varier $i = 0, 1, 2, \dots$

$$\begin{cases} \lambda_0 P_0 - \mu_1 P_1 = 0 \\ \lambda_1 P_1 - \mu_2 P_2 = \lambda_0 P_0 - \mu_1 P_1 = 0 \\ \dots & \dots \\ \lambda_i P_i - \mu_{i+1} P_{i+1} = \lambda_{i-1} P_{i-1} - \mu_i P_i = 0 \end{cases} \quad (14)$$

La résolution de ce système donne $P_i = P_0 \prod_{j=0}^{i-1} \frac{\lambda_j}{\mu_{j+1}}$.

Et tenant compte de la condition $\sum_{i \geq 0} P_i = 1$,

On trouve

$$P_0 = \frac{1}{1 + \sum_{i \geq 1} \prod_{j=0}^{i-1} \frac{\lambda_j}{\mu_{j+1}}}, \quad (15)$$

Qui existe si la série $\sum_{i \geq 1} \prod_{j=0}^{i-1} \frac{\lambda_j}{\mu_{j+1}} < \infty$ (condition de stabilité)

A partir de ces résultats, on peut ainsi définir l'état du système et, calculer les mesures de performances.

2.3 Quelques cas particuliers

- **File d'attente $M/M/1/\infty$**

λ : Le taux d'arrivée des clients/pièces (par unité de temps). $1/\lambda$ représente donc l'intervalle de temps moyen entre deux arrivées dans le système

μ : Le taux de service (par unité de temps). $1/\mu$ représente donc la durée moyenne de service.

On a un serveur et une infinité de clients dans le système.

On a donc $\lambda_i = \lambda, \forall i \geq 0$ et $\mu_i = \mu, \forall i \geq 1$ En posant $\rho = \frac{\lambda}{\mu}$, on a

$$P_i = \rho^i P_0 \quad \text{et} \quad P_0 = \frac{1}{1 + \sum_{i \geq 1} \rho^i},$$

La série du dénominateur converge si $\rho < 1$ (Qui est une condition nécessaire pour que l'équation d'équilibre soit asymptotiquement stable. Ce paramètre est appelé **Intensité de trafic**).

Dans cette condition, on a

$$\begin{cases} P_0 = 1 - \rho \\ P_i = (1 - \rho)\rho^i, \quad i \geq 1 \end{cases} \quad (16)$$

Afin de calculer les **mesures de performances du système**, définissons les grandeurs suivantes :

$S = \text{temps de service}$

$W = \text{temps mis par le client / pièce dans la queue avant le service}$

$R = \text{temps mis par le client / pièce dans le système (service inclus)}$

$N_s = \text{nombre de client / pièce dans le service}$

$N_Q = \text{nombre de client / pièce dans la queue (exclus les clients entrain d'être servis)}$

On peut alors calculer le *Nombre moyen de clients/pièces dans le système* en

posant :

$$E[M] = \sum_{i=0}^{\infty} i \cdot P_i$$

On utilise la formule de Little pour trouver le *Temps passé par un client/pièce dans le système* $E[R]$.

$$\text{Formule de Little : } E[M] = \lambda E[R]$$

D'où
$$E[R] = \lambda^{-1} E[M]$$

$$W = R - S \Rightarrow E[W] = E[R] - E[S].$$

Donc le *Temps moyen mis par le client/pièce avant le service* est

$$E[W] = \lambda^{-1} E[M] - \frac{1}{\mu}, \quad \text{où } E[S] = \frac{1}{\mu},$$

D'après la formule de Little appliquée au serveur, $E[N_s] = \lambda \cdot E[S] = \rho$ qui est le *Nombre moyen de clients/pièces dans le service*.

D'après la formule de Little appliquée à la queue, $E[N_Q] = \lambda \cdot E[W]$ qui est le *Nombre moyen de clients/pièces dans la queue*

- **File d'attente $M / M / 1 / K$**

Comme précédemment, on définit

λ : Le taux d'arrivée des clients/pièces (par unité de temps). $\frac{1}{\lambda}$ représente donc l'intervalle de temps moyen entre deux arrivées dans le système

μ : Le taux de service (par unité de temps). $\frac{1}{\mu}$ représente donc la durée moyenne de service. On a un serveur et une capacité finie K de clients/pièces dans la file. C'est à dire que lorsqu'un client est entrain d'être servi, $K - 1$

clients/pièces attendent dans la file et toute arrivée dans la file est interdite (refusée).

Le taux de processus de naissance et de mort peut être définie dans ce cas par :

$$\lambda_i = \begin{cases} \lambda & i < K \\ 0 & i \geq K \end{cases}$$

$$\mu_i = \mu, \forall i \geq 1$$

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

On a les probabilités de transition

$$\begin{cases} P_i = P_0 \prod_{j=0}^{i-1} \rho = P_0 \rho^i & i \leq K \\ P_i = 0 & i > K \end{cases} \quad (17)$$

En tenant compte de la condition de normalisation (théorie de la probabilité) on voit que

$$P_0 = \frac{1-\rho}{1-\rho^{K+1}}, \quad (18)$$

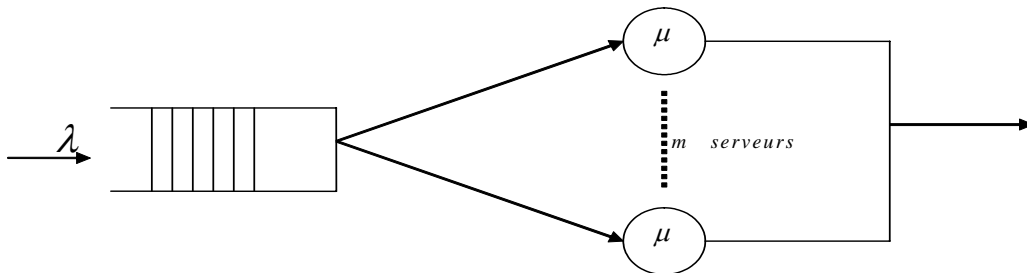
Et on trouve donc que pour une File d'attente de capacité finie

$$\begin{cases} P_i = \frac{(1-\rho)\rho^i}{1-\rho^{K+1}} & i \leq K \\ P_i = 0 & i > K \end{cases} \quad (19)$$

On peut donc calculer les mesures de performances du système $M/M/1/K$.

- **File d'attente $M/M/m/\infty$**

Le taux d'arrivée des clients/pièces est une constante λ et le taux de service μ dépend de m . On a m serveurs et une capacité infinie de clients/pièces dans la file. Lorsqu'un serveur est vide, il est immédiatement occupé par le client en attente.



Le taux de processus de naissance et de mort peut être définie par :

$$\mu_i = \begin{cases} i\mu & 0 < i < m \\ m\mu & i \geq m \end{cases}$$

$$\lambda_i = \lambda, \quad \forall i \geq 0$$

On a les probabilités de transition

$$\begin{cases} P_i = P_0 \prod_{j=0}^{i-1} \frac{\lambda}{(j+1)\mu} = P_0 \left(\frac{\lambda}{\mu} \right)^i \frac{1}{i!}, & i < m \\ P_i = P_0 \prod_{j=0}^{m-1} \frac{\lambda}{(j+1)\mu} \cdot \prod_{k=m}^{i-1} \frac{\lambda}{\mu m} = P_0 \left(\frac{\lambda}{\mu} \right)^i \frac{1}{m! m^{i-m}}, & i \geq m \end{cases} \quad (20)$$

On pose $\rho = \frac{\lambda}{m\mu}$ (*Intensité de trafic*)

$$P_i = \begin{cases} P_0 \frac{(m\rho)^i}{i!}, & i < m \\ P_0 \frac{\rho^i m^m}{m!}, & i \geq m \end{cases} \quad (21)$$

En tenant compte de la condition de normalisation (théorie de la probabilité) et la condition de stabilité de l'équation d'équilibre $\rho < 1$, on a

$$P_0 = \left(\sum_{i=0}^{m-1} \frac{(m\rho)^i}{i!} + \sum_{i=m}^{\infty} \frac{\rho^i m^m}{m!} \right)^{-1}$$

En effectuant un changement de variable $k = i - m$, on observe que

$$P_0 = \left(\sum_{i=0}^{m-1} \frac{(m\rho)^i}{i!} + \frac{(m\rho)^m}{m!} \frac{1}{1-\rho} \right)^{-1} \quad (22)$$

On peut calculer les mesures de performance du système $M / M / m / \infty$.

- **File d'attente $M / M / \infty$**

Le taux d'arrivée des clients/pièces est une constante λ et le taux de service μ dépend du nombre de clients dans le système. On a une infinité de clients/pièces dans la file et une infinité de serveurs.

Le taux de processus de naissance et de mort peut être définie par :

$$\begin{cases} \lambda_i = \lambda & i \geq 0 \\ \mu_i = i\mu & i \geq 0 \end{cases}$$

On a les probabilités de transition

$$P_i = P_0 \prod_{j=1}^{i-1} \frac{\lambda}{(j+1)\mu} = P_0 \frac{1}{i!} \left(\frac{\lambda}{\mu} \right)^i \quad (23)$$

En tenant compte de la condition de normalisation (théorie de la probabilité) et la condition de stabilité de l'équation d'équilibre $\rho = \frac{\lambda}{\mu} < 1$, on a

$$P_0 = \frac{1}{1 + \sum_{i=1}^{\infty} \frac{\rho^i}{i!}} = e^{-\rho} \quad \text{et} \quad P_i = \frac{\rho^i}{i!} e^{-\rho} \quad (24)$$

Qui n'est rien d'autre que la Loi de probabilité d'une Distribution de Poisson de paramètre ρ

On peut calculer les mesures de performance du système $M / M / \infty$.

Autre approche

On a vu plus haut que dans un intervalle de temps petit Δt la transition de $i \rightarrow i+1$ correspond à une naissance avec la probabilité $\lambda_i \Delta t$ et la transition de $i \rightarrow i-1$ correspond à un décès avec la probabilité $\mu_i \Delta t$, et rien ne se produit avec une probabilité $(1 - (\lambda_i + \mu_i)) \Delta t$. Donc par le théorème des probabilités totales, on peut écrire

$$\begin{aligned} P(M(t + \Delta t) = i \mid M(t) = i - 1) &= \lambda_{i-1} \Delta t + o(\Delta t) \\ P(M(t + \Delta t) = i \mid M(t) = i + 1) &= \mu_{i+1} \Delta t + o(\Delta t) \\ P(M(t + \Delta t) = i \mid M(t) = i) &= 1 - \lambda_i \Delta t - \mu_i \Delta t + o(\Delta t) \end{aligned} \quad (25)$$

Où $\lim_{\Delta t \rightarrow 0} \frac{o(\Delta t)}{\Delta t} = 0$

Posons $P_i(t) = P(M(t) = i)$

On a les relations

$$\begin{cases} P_0(t + \Delta t) = \mu_1 \Delta t P_1(t) + (1 - \lambda_0 \Delta t) P_0(t) & i = 0 \\ P_i(t + \Delta t) = \lambda_{i-1} \Delta t P_{i-1}(t) + \mu_{i+1} \Delta t P_{i+1}(t) + (1 - \lambda_i \Delta t - \mu_i \Delta t) P_i(t) & i > 0 \end{cases}$$

soit

$$\begin{cases} \frac{P_0(t + \Delta t) - P_0(t)}{\Delta t} = -\lambda_0 P_0(t) + \mu_1 P_1(t) & i = 0 \\ \frac{P_i(t + \Delta t) - P_i(t)}{\Delta t} = -(\lambda_i + \mu_i) P_i(t) + \mu_{i+1} P_{i+1}(t) + \lambda_{i-1} P_{i-1}(t) & i > 0 \end{cases}$$

En prenant la limite quand $\Delta t \rightarrow 0$, on obtient le système d'équation différentielle linéaire,

$$\begin{cases} \frac{dP_0(t)}{dt} = -\lambda_0 P_0(t) + \mu_1 P_1(t) & i = 0 \\ \frac{dP_i(t)}{dt} = -(\lambda_i + \mu_i) P_i(t) + \mu_{i+1} P_{i+1}(t) + \lambda_{i-1} P_{i-1}(t) & i > 0 \end{cases}$$

et compte tenu du fait que la variation de flux à l'infini dans un Processus de naissance et de mort est nulle, c'est à dire que $\lim_{t \rightarrow \infty} \frac{dP_i(t)}{dt} = 0, \quad i = 0, 1, 2, \dots$

On retrouve bien les conditions d'état d'équilibre (*)

Remarque : Si $\rho \geq 1$ alors le serveur ne peut satisfaire la demande de service. Et la file est condamnée à croître définitivement.

La règle de service étudiée dans ce rapport est FIFO c'est à dire le premier client/pièce arrivé est le premier servi. Les mesures de performance du système changent si l'on utilise une autre règle de service et si l'on tient compte du comportement des clients (patients ou impatientes). Nous pouvons utiliser le Processus de naissance et de mort pour évaluer les performances des systèmes $M/G/m, G/G/m$ où M est le Processus des arrivées suivant une Loi Exponentielle ou une loi de Distribution général, et G une Loi de Distribution général. Les calculs s'avèrent très compliquer et n'est pas pris en charge dans le présent rapport. L'une des raisons de cette difficulté provient du fait que le principe d'équilibre n'est plus applicable car le processus de durée de service n'est plus sans mémoire.

Un processus T est sans mémoire si $P(T > s + t | T > s) = P(T > t)$. C'est à dire que les occurrences des événements sont indépendantes les unes des autres ou que l'occurrence d'événements avant la date t n'influe en rien sur l'occurrence

d'événements après t . Seule la loi Exponentielle à la propriété d'absence de mémoire).

Une étude plus poussée des chaînes de Markov et des réseaux de file d'attente est nécessaire pour améliorer et mieux évaluer les performances des systèmes informatiques, des réseaux de communication et des systèmes de production donc ce dernier cas est l'essentielle de notre thèse. Nous poursuivrons nos travaux par l'application des files d'attente pour la simulation des cas simples des systèmes de production.

3. Référence

- [1]. « Foundations of Inventory Management » by Paul H. Zipkin
- [2]. « Introduction à la Gestion des Stocks » par Daniel Wolf
- [3]. « An Introduction to Queueing system » by Sanjay K. Bose