Systèmes avec rappels -

Introduction :

La théorie classique des files d'attente donne deux principales méthodes pour résoudre le conflit qui se produit lorsqu'un client arrivant dans le système, trouve le(s) serveur(s) occupé(s) :

1) Le client peut quitter le système pour toujours sans être servi, ceci correspond au « système d'ERLANG avec refus » appelé aussi « modèle à appels perdus ».

2) Le client peut attendre, en file d'attente, pour être servi après la libération du serveur, ceci correspond au « système de file d'attente classique ».

Une situation intermédiaire envisage la possibilité pour un client qui trouve le(s) serveur(s) occupé(s) de rappeler ultérieurement pour le service autant de fois que nécessaire, et à des intervalles de temps aléatoires, jusqu'à ce qu'il trouve un serveur libre et que son service puisse commencer, on parle alors de systèmes de files d'attente avec rappels ou encore de systèmes de files d'attente avec appels répétés. (« Retrial queueing)
Syst...

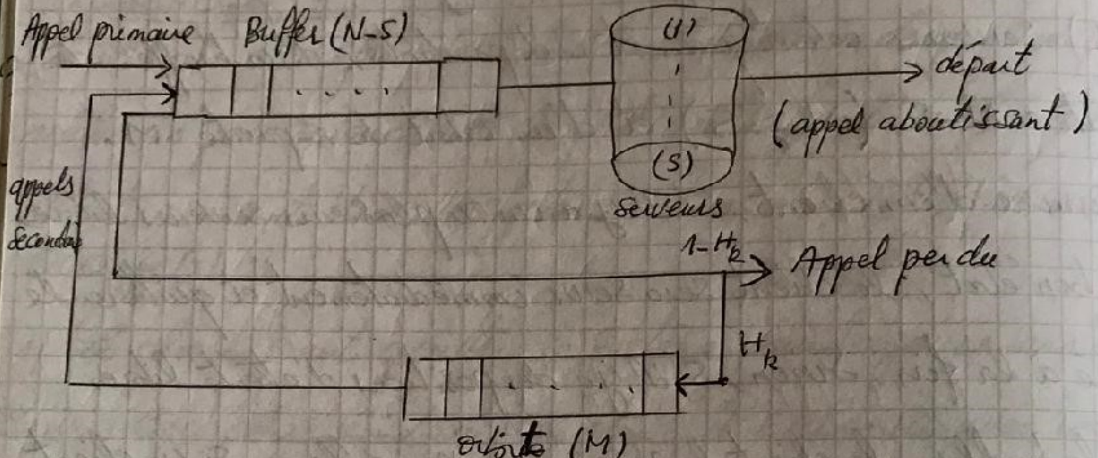
Description du modèle de file d'attente avec rappels :

Un système d'attente avec rappels est un système composé de s ($s \geq 1$) serveurs identiques et indépendants, d'un buffer de capacité $N-S$ ($N \geq S$) et d'un orbit de capacité M.

A l'arrivée d'un client, s'il y a un ou plusieurs serveurs libre et en bon état, le client sera servi immédiatement et quittera le système à la fin. Sinon, s'il y a des positions d'attente libres dans le buffer, le client le rejoindra. par ailleurs, si un client arrive et trouve tous les serveurs et toutes les positions d'attente du buffer occupées, il quittera le système définitivement avec la probabilité $1-H_0$ ou bien entre en orbite avec la probabilité H_0 et devient une source d'appels répétés et il tentera sa chance après une durée de temps aléatoire.

Les clients qui reviendront et rappelleront pour le service sont dits « en orbit ».

Chaque client en orbit appelé aussi client secondaire, est supposé rappeler pour le service à des intervalles de temps suivant une loi de probabilité. chacun de ces clients secondaire est traité comme un client primaire càd un nouveau client qui arrive de l'extérieur du système.



« Modèle de file d'attente avec rappels »

Remarques :

1) Le modèle d'attente avec rappels décrit ci-dessus est un modèle général. plusieurs systèmes peuvent être considérés comme des cas particuliers tels que : les systèmes sans buffer, les système à un serveur, ...

2) Les appels primaires ou secondaires qui arrivent durant un temps de service, entrent en orbit sans aucune influence sur le processus de service.

Notation :

En utilisant la notation de Kendall, un modèle de file d'attente

[redacted]

avec rappels est noté comme suit : A/B/S/N/M/H où

A décrit la distribution des temps des inter-arrivées des clients

B // // // du temps de service de chaque client

S est le nombre de serveurs dans le système.

N st la capacité du système

M st la capacité de l'orbit

H st la fonction de persévérance qui permet de définir le comportement du client devant une situation de blocage (serveurs occupés).

Remarques :

1/ M peut être supprimé s'il est infini et H peut être également supprimée si tous les clients sont persévérants c.à d. $H = 1$.

2) Le temps de rappel st défini comme l'intervalle de temps entre deux rappels consécutifs du même client secondaire.

3) La distribution des temps de rappels est supposée généralement exponentielle de taux θ : autrement dit, $1/\theta$ st la durée moyenne des intervalles de rappels. C'st la raison pour laquelle elle est omise de la notation.

4) Lorsque $\theta \to \infty$, le système d'attente avec rappels st un système d'attente classique.

[redacted]

5) Lorsque $\theta \to 0$, le système d'attente avec rappels est un système d'ERLANG avec perte.

de 4) et 5) On conclu que le modèle d'attente avec rappels occupe une situation intermédiaire entre le modèle d'ERLANG avec refus et le modèle classique avec attente (FIFO), dans le cas de faible et forte intensité de rappels respectivement.

Modèles Markoviens

Les modèles Markoviens sont des systèmes où les temps inter-arrivées primaires, les durées de service et les temps inter-rappels sont des variables aléatoires indépendantes et exponentiellement distribuées.

Etude du système M/M/m/m avec rappels.

Considérons un système avec rappels sans buffer à m serveurs. Le flot des arrivées primaires est poissonnien de taux λ. Si un client primaire trouve au moins un serveur libre, il est immédiatement pris en charge. Sinon, il entre en orbite et devient source d'appels secondaires. La durée de service est exponentielle de paramètre μ. La durée entre deux rappels successifs d'une même source secondaire est exponentielle de paramètre θ.

[redacted]

Le système peut être décrit par le processus markovien $X(t) = \{C(t), N(t)\}$, d'espace d'états $S = \{0,1,\dots,m\} \times \mathbb{N}$, où $C(t)$ est le nombre de clients en cours de service à la date t et $N(t)$ est le nombre de clients en orbite à la date t.

Soit $P_{(i,j)}(t) = \{C(t) = i, N(t) = j\}$ avec $i = 0,\dots m$, $j \in \mathbb{N}$.

Les probabilités de transitions à l'état stationnaire sont données par :

pour $0 \leq i \leq m-1$

$$P_{(i,j)}(k,\ell) = \begin{cases} \lambda & \text{si } (k,\ell) = (i+1,j) \\ i\mu & \text{si } (k,\ell) = (i-1,j) \\ j\theta & \text{si } (k,\ell) = (i+1,j-1) \\ -(\lambda + i\mu + j\theta) & \text{si } (k,\ell) = (i,j) \\ 0 & \text{sinon} \end{cases}$$

pour $i = m$

$$P_{(m,j)}(k,\ell) = \begin{cases} \lambda & \text{si } (k,\ell) = (m,j+1) \\ m\mu & \text{si } (k,\ell) = (m-1,j) \\ -(\lambda + m\mu) & \text{si } (k,\ell) = (m,j) \\ 0 & \text{sinon} \end{cases}$$

Dans le cas où $m = 1$, sous la condition d'ergodicité :

$\rho = \lambda/\mu < 1$, les probabilités stationnaires existent et sont

[redacted]

données par :

$$P_{0j} = \frac{\rho^j}{j!\,\mu^j} \prod_{i=0}^{j-1} (\lambda + i\mu)(1-\rho)^{1+\lambda/\mu}$$

$$P_{1j} = \frac{\rho^{j+1}}{j!\,\mu^j} \prod_{i=1}^{j} (\lambda + i\mu)(1-\rho)^{1+\lambda/\mu}$$

Les fonctions génératrices correspondantes sont données par :

$$P_0(z) = \sum_{j=0}^{\infty} z^j P_{0j} = (1-\rho)\left(\frac{1-\rho}{1-\rho z}\right)^{\lambda/\mu}$$

$$P_1(z) = \sum_{j=0}^{\infty} z^j P_{1j} = \rho\left(\frac{1-\rho}{1-\rho z}\right)^{1+\lambda/\mu}$$

A partir de ces fonctions génératrices, on peut obtenir, en régime stationnaire, les mesures de performance du système.

1) La distribution stationnaire du nombre de clients en orbite $q_j = P(N(t) = j)$ a la fonction génératrice suivante :

$$P(z) = P_0(z) + P_1(z) = (1+\rho-\rho z)\left(\frac{1-\rho}{1-\rho z}\right)^{1+\lambda/\mu}$$

par conséquent : $E(N(t)) = P'(1) = \dfrac{\rho(\lambda+\rho\mu)}{(1-\rho)\mu}$

$$\boxed{E(N(t)) = \frac{\rho(\lambda+\rho\mu)}{\mu(1-\rho)}}$$

[redacted]

$$\boxed{Var(N(t)) = \frac{\rho(\lambda + \rho\mu + \rho^2\mu - \rho^3\mu)}{(1-\rho)^2\mu}}$$

2/ La fonction génératrice associée à la probabilité stationnaire du nombre de clients dans le système $q_n = P(K(t)=n)$ est donnée par :

$$\varphi(z) = P_0(z) + zP_1(z) = \left(\frac{1-\rho}{1-\rho z}\right)^{1+\lambda/\mu}$$

D'où

$$\boxed{E(K(t)) = \frac{\rho(\lambda+\mu)}{(1-\rho)\mu}}$$

$$\boxed{Var(K(t)) = \frac{\rho(\lambda+\mu)}{(1-\rho)^2\mu}}$$

3) La probabilité de blocage q_1, i.e la probabilité que le serveur soit occupé est donnée par : $\boxed{q_1 = P_1(1) = \rho.}$

Modèles semi-markoviens

Etude du système M/G/1.