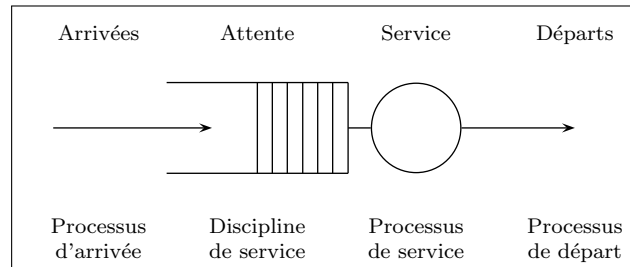


Résumé sur les files d'attente et les réseaux

Généralités

Une file d'attente (ou queue) simple est caractérisée par

- un processus d'arrivée (A),
- un processus de service (S),
- le nombre m de serveurs (tous identiques),
- la capacité d'accueil K de la file (nombre de places d'attente + de service) (valeur par défaut : ∞),
- la taille P de la population (nombre total de clients susceptibles d'entrer dans le système) (valeur par défaut : ∞),
- une discipline de service D (FIFO, LIFO, SIRO, PS, ...) (valeur par défaut : FIFO).



Notation de Kendall

Forme abrégée : $A/S/m$

Forme complète : $A/S/m/K/P/D$

Symboles courants pour les processus d'arrivée et de service :

- M : loi exponentielle (*memoryless*),
- D : loi dégénérée (constante),
- G : loi générale (vrai pour toute loi raisonnable).

Mesures de performance

- \bar{N} : nombre moyen de clients présents (en attente ou en service)
- \bar{Q} : nombre moyen de clients en attente
- \bar{S} : temps moyen de service ($= 1/\mu$)
- \bar{T} : temps moyen de séjour ou de réponse (attente et service)
- \bar{W} : temps moyen d'attente
- U : taux d'utilisation de chaque serveur

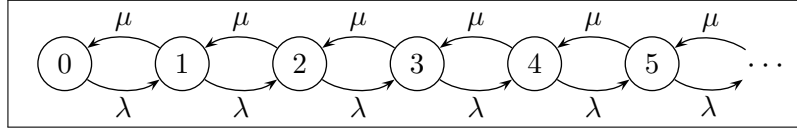
Relations générales

- Une file $G/G/m$ (non déterministe) est stable $\iff \rho = \frac{\lambda}{m\mu} < 1$ où
 - ρ est appelé l'intensité du trafic,
 - λ est le nombre moyen d'arrivées par unité de temps,
 - μ est le nombre moyen de clients traités par chaque serveur par unité de temps.
- Pour un système stable en régime stationnaire on a : $\bar{T} = \bar{W} + \bar{S}$ et $\bar{N} = \bar{Q} + m\bar{U}$.
- Formule de Little (valable pour un système stable en régime stationnaire) : $\bar{N} = \lambda\bar{T}$.

La relation est également valable pour un sous-système, par exemple, les places d'attente : $\bar{Q} = \lambda\bar{W}$.

La file d'attente $M/M/1$

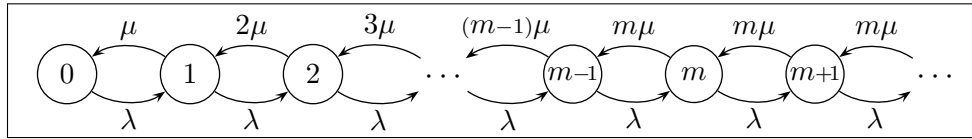
Graphe représentatif du processus de naissance et mort.



- intensité du trafic : $\rho = \lambda/\mu$
- condition de stabilité ; $\rho < 1$
- distribution stationnaire : $\pi_k^* = (1 - \rho)\rho^k$, $k = 0, 1, 2, \dots$
- taux d'utilisation du serveur : $U = 1 - \pi_0^* = \rho = \lambda/\mu$
- nombre moyen de clients présents : $\bar{N} = \rho/(1 - \rho)$
- nombre moyen de clients en attente : $\bar{Q} = \bar{N} - U = \rho^2/(1 - \rho)$
- temps moyen de séjour : $\bar{T} = \frac{\bar{N}}{\lambda} = \frac{\rho}{\lambda(1-\rho)} = \frac{1}{\mu(1-\rho)} = \frac{\bar{S}}{1-\rho}$
- temps moyen d'attente : $\bar{W} = \frac{\bar{Q}}{\lambda} = \frac{\rho^2}{\lambda(1-\rho)} = \frac{\rho}{\mu(1-\rho)} = \frac{\rho \bar{S}}{1-\rho}$
- fonction de distribution du temps de séjour (discipline FIFO) : $F_T(t) = 1 - e^{-\mu(1-\rho)t}$
- fonction de distribution du temps d'attente (discipline FIFO) : $F_W(t) = 1 - \rho e^{-\mu(1-\rho)t}$

La file d'attente $M/M/m$

Graphe représentatif du processus de naissance et mort.



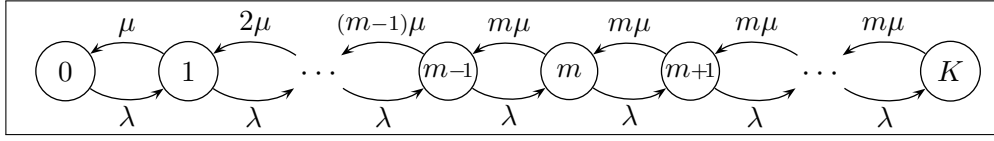
- intensité du trafic : $\rho = \lambda/(m\mu)$
 - condition de stabilité : $\rho < 1$
 - distribution stationnaire : $\pi_k^* = \begin{cases} \frac{(m\rho)^k}{k!} \pi_0^* & k = 0, \dots, m-1 \\ \frac{\rho^k m^m}{m!} \pi_0^* & k = m, m+1, \dots \end{cases}$
- avec $\pi_0^* = \left[1 + \frac{(m\rho)^m}{m!(1-\rho)} + \sum_{k=1}^{m-1} \frac{(m\rho)^k}{k!} \right]^{-1}$
- formule C d'Erlang : $\zeta = P[\geq m \text{ clients présents}] = \sum_{k=m}^{\infty} \pi_k^* = \frac{(m\rho)^m}{m!(1-\rho)} \pi_0^*$

| | $m = 1$ | $m = 2$ | $m = 3$ |
|-----------|------------|-----------------------------|---|
| π_0^* | $1 - \rho$ | $\frac{1 - \rho}{1 + \rho}$ | $\frac{2 - 2\rho}{3\rho^2 + 4\rho + 2}$ |
| ζ | ρ | $\frac{2\rho^2}{1 + \rho}$ | $\frac{9\rho^3}{3\rho^2 + 4\rho + 2}$ |

- taux d'utilisation de chaque serveur $U = \rho = \lambda/(m\mu)$
- nombre moyen de clients présents : $\bar{N} = m\rho + \frac{\rho \zeta}{1-\rho}$
- nombre moyen de clients en attente : $\bar{Q} = \frac{\rho \zeta}{1-\rho}$
- temps moyen de séjour : $\bar{T} = \frac{1}{\mu} \left(1 + \frac{\zeta}{m(1-\rho)} \right) = \bar{S} + \frac{\zeta \bar{S}}{m(1-\rho)}$
- temps moyen d'attente : $\bar{W} = \frac{\zeta}{m\mu(1-\rho)} = \frac{\zeta \bar{S}}{m(1-\rho)}$

La file d'attente $M/M/m/K$ ($K \geq m$)

Graphe représentatif du processus de naissance et mort.



- intensité du trafic : $\rho = \lambda/(m\mu)$
- condition de stabilité : la file est toujours stable
- distribution stationnaire :

$$\pi_k^* = \begin{cases} \frac{(m\rho)^k}{k!} \pi_0^* & k = 1, \dots, m-1 \\ \frac{\rho^k m^m}{m!} \pi_0^* & k = m, m+1, \dots, K \end{cases}$$

avec

$$\pi_0^* = \left[1 + \frac{(1 - \rho^{K-m+1})(m\rho)^m}{m!(1 - \rho)} + \sum_{k=1}^{m-1} \frac{(m\rho)^k}{k!} \right]^{-1} \quad \text{si } \rho \neq 1$$

et

$$\pi_0^* = \left[1 + \frac{(m)^m}{m!}(K - m + 1) + \sum_{k=1}^{m-1} \frac{(m)^k}{k!} \right]^{-1} \quad \text{si } \rho = 1$$

- **Attention !** le taux effectif d'arrivée n'est plus λ mais λ' :

$$\lambda' = \sum_{k=0}^{K-1} \lambda \pi_k^* = \lambda(1 - \pi_K^*)$$

- taux d'utilisation de chaque serveur : $U = \left(\sum_{i=0}^m i \pi_i^* + \sum_{i=m+1}^K m \pi_i^* \right) / m = \lambda' / m\mu$
- nombre moyen de clients présents : $\bar{N} = \sum_{i=0}^K i \pi_i^*$
- nombre moyen de clients en attente : $\bar{Q} = \sum_{i=m+1}^K (i - m) \pi_i^*$
- temps moyen de séjour : $\bar{T} = \bar{N} / \lambda'$
- temps moyen d'attente : $\bar{W} = \bar{Q} / \lambda'$

La file d'attente $M/G/1$

- intensité du trafic : $\rho = \lambda/\mu = \lambda\bar{S}$
- condition de stabilité : $\rho < 1$

Formule de Pollaczek-Khinchin :

$$\bar{Q}_{M/G/1} = \left(\frac{1 + C_S^2}{2} \right) \frac{\rho^2}{1 - \rho} = \left(\frac{1 + C_S^2}{2} \right) \bar{Q}_{M/M/1}$$

$$\bar{W}_{M/G/1} = \left(\frac{1 + C_S^2}{2} \right) \frac{\rho \bar{S}}{1 - \rho} = \left(\frac{1 + C_S^2}{2} \right) \bar{W}_{M/M/1}$$

où C_S^2 est le coefficient de variation au carré du temps de service : $C_S^2 = \sigma_S^2 / \bar{S}^2 = \mu^2 \sigma_S^2$.

Pour le calcul de \bar{N} et \bar{T} on utilise les relations :

$$\bar{N} = \bar{Q} + U \quad \text{et} \quad \bar{T} = \bar{W} + \bar{S}$$

avec $U = \rho$.

La file d'attente $M/G/\infty$ (centre de délai)

- intensité du trafic : $\rho = \lambda/\mu = \lambda\bar{S}$
- condition de stabilité : la file est toujours stable
- distribution stationnaire : $\pi_k^* = \frac{e^{-\rho}}{k!} \rho^k$, $k = 0, 1, \dots$
- nombre moyen de clients présents : $\bar{N} = \rho = \lambda/\mu$
- nombre moyen de clients en attente : $\bar{Q} = 0$
- temps moyen de séjour : $\bar{T} = \bar{S} = 1/\mu$
- temps moyen d'attente : $\bar{W} = 0$

La file d'attente $G/G/m$

Approximation basée sur la formule de Pollaczek-Khinchin (exacte pour les files $M/M/m$ et $M/G/1$) :

$$\bar{Q}_{G/G/m} \simeq \left(\frac{C_A^2 + C_S^2}{2} \right) \bar{Q}_{M/M/m}$$

Les réseaux de Jackson

Équations de conservation (pour un réseau formé de J files $*/M/m_j$) :

$$\lambda_j = \gamma_j + \sum_{i=1}^J \lambda_i r_{ij} \quad j = 1, \dots, J.$$

où γ_i est le taux d'arrivée externe dans la file i , λ_i est le taux effectif d'arrivée dans la file i et r_{ij} est la probabilité de routage vers la station j d'un client quittant la station i .

Forme matricielle :

$$\boldsymbol{\lambda} = \boldsymbol{\gamma} + \boldsymbol{\lambda} \mathbf{R} \quad \Longleftrightarrow \quad \boldsymbol{\lambda}(\mathbf{I} - \mathbf{R}) = \boldsymbol{\gamma}.$$

Pour un réseau ouvert

$$\boldsymbol{\lambda} = \boldsymbol{\gamma}(\mathbf{I} - \mathbf{R})^{-1}$$

et le réseau est stable si et seulement si chacune des files le composant l'est, c'est-à-dire si et seulement si

$$\rho_j = \frac{\lambda_j}{m_j \mu_j} < 1 \quad \forall j = 1, \dots, J.$$

Distribution stationnaire d'un réseau de Jackson ouvert et stable (distribution à forme produit) :

$$\pi^*(\mathbf{x}) = \pi^*(x_1, \dots, x_J) = \prod_{j=1}^J \pi_j^*(x_j).$$

où l'état \mathbf{x} du réseau est le vecteur dont la j^e composante représente le nombre de clients présents dans la station j et $\pi_j^*(x_j)$ est la probabilité stationnaire d'avoir x_j clients dans une file $M/M/m_j$ ($\pi_j^*(x_j) = (1 - \rho_j) \rho_j^{x_j}$ si $m_j = 1$).

Tout se passe comme si le système était formé de J files $M/M/m_j$ indépendantes d'intensité respective $\rho_1 = \lambda_1/(m_1 \mu_1), \dots, \rho_J = \lambda_J/(m_J \mu_J)$.