

TD 9 : Analyse de files d'attente

lionel.rieg@ens-lyon.fr

Exercice 1 (Distribution du temps d'attente)

On considère une file d'attente M/M/1 de taux d'utilisation $\rho < 1$. On rappelle que sous l'hypothèse de régime permanent, on a les résultats suivants (vu au TD précédent) :

- la distribution du nombre de personnes dans le système suit une loi géométrique de paramètre ρ (décalée de 1 car elle peut prendre la valeur 0) ;
- l'espérance du nombre de personnes dans le système est $L = \frac{\rho}{1-\rho}$;
- l'espérance du nombre de personnes en attente est $L_q = L - (1 - \rho) = \frac{\rho^2}{1-\rho}$;
- pour obtenir la distribution du temps d'attente, il nous faut choisir une politique de service, ici PAPS ;
- dans le cas d'arrivée Poissonienne (c'est faux sinon), la probabilité q_n qu'un client arrivant dans le système trouve n personnes devant lui est égale à la probabilité p_n qu'il y ait n personnes dans le système dans l'absolu ;
- la variable aléatoire T_q représentant le temps d'attente est à densité pour $t > 0$ et possède un point de masse en $t = 0$;
- sa fonction de répartition $W_q(t) = F_{T_q}(t)$ vérifie $W_q(0) = q_0 = 1 - \rho$.

Le temps de service de n clients suit une distribution d'Erlang de paramètres n (paramètre de forme) et μ (paramètre d'intensité) dont la fonction de densité est

$$E(n, \mu, x) = \frac{\mu^n x^{n-1} e^{-\mu x}}{(n-1)!}.$$

4. En utilisant la formule des probabilités totales, donner une expression pour $W_q(t)$.
5. Donner l'espérance W_q de T_q .
6. Adapter les deux questions précédentes pour traiter le cas de T , le temps de séjour total dans le système.
7. Quelle relation y a-t-il entre L_q et W_q d'une part, et L et W d'autre part ?

Exercice 2 (Serveur et miroir)

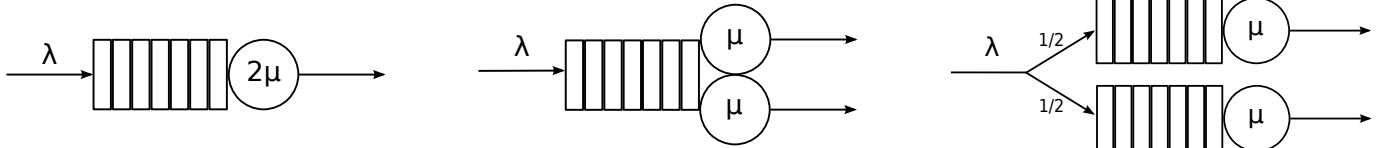
On dispose de deux serveurs Web : un serveur principal et un serveur miroir. Les requêtes arrivent à un routeur qui doit essayer de répartir le travail entre le serveur principal ou son miroir, qui est supposé moins performant. Le routeur ne connaissant pas la charge respective des deux serveurs, il décide d'envoyer chaque requête au serveur principal avec probabilité p ou au serveur miroir avec probabilité $1 - p$. On cherche à optimiser le choix de p .

On suppose que les arrivées suivent une loi de Poisson de paramètre λ . Les temps de service des serveurs sont exponentiels de paramètres respectifs μ pour le serveur principal et ν pour le miroir.

1. Montrer que les arrivées des paquets dans les files du serveur principal et de son miroir sont des processus de Poisson de paramètres respectifs $p\lambda$ et $(1-p)\lambda$.
2. Quelles sont les conditions de stabilité du système ?
3. Calculer le nombre de requêtes dans le serveur principal et dans le miroir à l'état stationnaire.
4. Donnez le temps de réponse moyen dans le système global.
5. Déterminez la valeur optimale de p .

Exercice 3 (Comparaison de trois modèles de serveur)

On souhaite comparer les trois architectures suivantes de file d'attente :



1. Intuitivement, quelle configuration semble la meilleure ?

Pour chacune des trois configurations :

2. Donnez sa condition de stabilité.
3. Écrire son générateur infinitésimal (en précisant les cas limites).
4. Calculer le nombre moyen de clients dans le système.
5. En déduire le temps moyen de réponse.

Où se trouve la différence entre les différentes configurations ? Laquelle préférer ?

Exercice 4 (Dimensionnement de buffer)

On considère un réseau d'entreprise composé de deux LANs, l'un constituant le réseau de production (LAN 1) et l'autre hébergeant la base de données de l'entreprise (LAN 2). Ces deux sous-réseaux sont connectés par un lien à 1,5 Mbits/seconde. On suppose que :

- Les requêtes provenant du réseau de production pour la base de données arrivent au routeur d'interconnexion selon un processus de Poisson d'intensité égale à 15 requêtes par seconde.
- Les documents hébergés par la base de données ont une taille aléatoire de distribution exponentielle, de moyenne 100 kbits.
- Les délais de propagation et de traitement sont négligés, de sorte que les temps de transferts de ces documents entre les deux sites sont proportionnels à la taille des documents demandés.
- Les temps de traitement par la base de données sont négligés.
- La taille des requêtes est négligeable devant la taille des documents demandés.

1. En supposant les buffers de taille infinie, montrer que le processus d'arrivée des documents au routeur d'interconnexion du LAN 2 vers le LAN 1 peut être modélisé par une file M/M/1. Calculer ses taux d'arrivée et de service.
2. Le système est-il stable ? Décrivez l'évolution asymptotique du temps de réponse.
3. On suppose que l'on augmente le débit du lien d'interconnexion à 3 Mbits/seconde. Calculer la taille du buffer du routeur (mise en attente des requêtes) afin de garantir une probabilité de perte inférieure à 10^{-3} .