

TD 5 : File d'attente

Artificial Intelligence

G.Guérard

Exercice 1

Un système clients-serveur reçoit en moyenne 1000 requêtes par seconde, arrivant selon un processus de Poisson. Il dispose d'un unique serveur pouvant traiter en moyenne 2000 clients par seconde. On suppose que le temps de service d'un client est distribué selon la loi exponentielle. Calculer la probabilité que le temps de service dépasse 2 ms.

Quelle est le pourcentage de clients rejetés pour un système ne comportant pas de file d'attente.

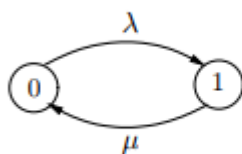
Même question pour un système comportant une file d'attente de 1 place. Calculer le taux d'application du serveur.

Même question pour un système comportant une file d'attente de 2 places.

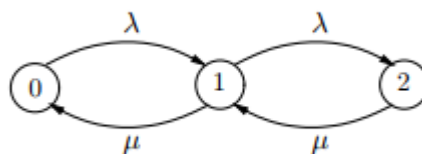
Solution

En choisissant la milliseconde comme unité de temps on obtient un taux de naissance de 1 et un taux de décès de 2. Le temps de service T_S d'un client suit une loi exponentielle de paramètre 2 donc la probabilité que le temps de service dépasse deux secondes est de 0.0183 grâce à la formule suivant :

$$\text{Prob} \{T_S > t\} = \exp(-\mu t).$$



(a) M/M/1/1



(b) M/M/1/2

La chaîne (a) représente une capacité dans la file de 0 place tandis que la chaîne (b) représente une capacité dans la file de 1 place. Pour chaque nouvelle place il suffit de rajouter un état.

Pour la première chaîne, nous allons calculer la distribution stationnaire :

$$\begin{cases} \pi_0 = \frac{\mu}{\lambda + \mu} = \frac{2}{3} \\ \pi_1 = \frac{\lambda}{\lambda + \mu} = \frac{1}{3} \end{cases}$$

Nous avons donc 1/3 de clients rejetés (si on se trouve 1/3 du temps en état 1, alors on ne peut plus recevoir des clients sur 1/3 du fonctionnement de la file). Le taux d'occupation est $\pi_0 = 2/3$.

Pour la deuxième chaîne, nous allons calculer la distribution stationnaire :

$$\begin{cases} \pi_0 = \frac{4}{7} \\ \pi_1 = \frac{2}{7} \\ \pi_2 = \frac{1}{7} \end{cases}$$

Et pour une capacité de 2 dans la file d'attente nous obtenons $\pi_3 = 1/15$ et $\pi_0 = 8/15$.

Exercice 2

Un système clients-serveur reçoit en moyenne 1000 requêtes par seconde, arrivant selon un processus de Poisson. A titre expérimental, on envisage un système sans file d'attente mais comportant plusieurs serveurs de front. Lorsque tous les serveurs sont occupés, les requêtes sont rejetées.

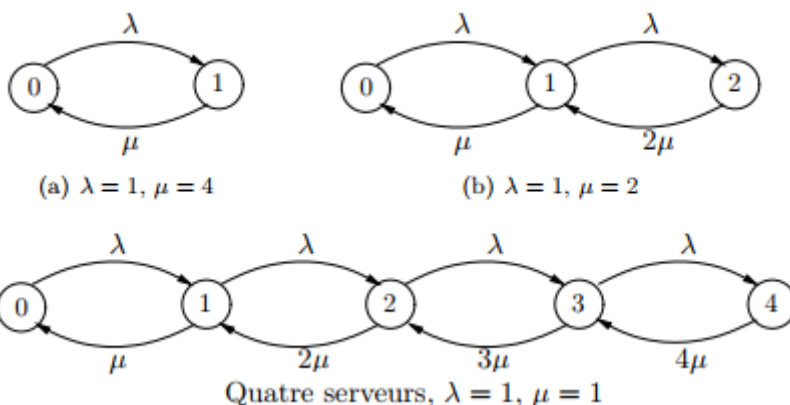
Quelle est le pourcentage de clients rejetés pour un système comportant 1 serveur traitant 4000 requêtes par seconde

Même question pour un système comportant deux serveurs traitant chacun 2000 requêtes par seconde.

Même question pour un système comportant quatre serveurs traitant chacun 1000 requêtes par seconde.

Solution

Peu importe le nombre de serveur, l'idée du calcul parallèle sans file d'attente est toujours construite de la même manière : à chaque nouveau client, un serveur de plus fonctionne.



Il suffit de calculer la distribution stationnaire, le nombre de client rejeté est égale à la distribution du dernier état (le plus à droite ici).

Pour un serveur : $1/5$

Pour deux serveurs : 1/13

Pour quatre serveurs : 1/65

Exercice 3

Des camions arrivent dans une station-service pour passer des tests de sécurité, suivant un processus de Poisson de taux de 6 / jour. La durée des tests pour chaque camion est une v.a exponentielle d'espérance mathématique de 1h 30 mn. On suppose que le processus d'arrivée ne s'interrompt pas et que la station travaille 24 heures sur 24.

Le système admet-il une distribution stationnaire ?

Si oui la calculer et donner le nombre moyen d'utilisateurs dans le système, le temps moyen passé dans le système, la longueur moyenne de la file d'attente et le temps moyen passé dans la file (en régime stationnaire).

Solution

Nous avons $\lambda = \frac{1}{4}$ et $\mu = \frac{2}{3}$ (en heure). Comme $\rho = \frac{\lambda}{\mu} = \frac{3}{8} < 1$, le système est ergodique.

D'après les formules du calcul des paramètres des systèmes $M/M/1$, nous aurons :

$$P_0 = 1 - \rho = \frac{5}{8}, \quad P_k = \frac{5}{8} \left(\frac{3}{8}\right)^k, \quad k \in \mathbb{N}$$

$$\bar{N} = \frac{\rho}{1 - \rho} = \frac{3}{5}, \quad \bar{T} = \frac{\bar{N}}{\lambda} = 2,4 \text{ h.}$$

La longueur moyenne de la file d'attente et la durée d'attente dans la file sont données par :

$$\bar{N}_f = \bar{N} - 1 + P_0 = \frac{9}{40}$$

$$\text{et } \bar{T}_f = \bar{N}_f \times \frac{1}{\mu} = \frac{9}{10} \text{ h} = 54 \text{ mn.} \quad (\text{On a } \bar{T}_f = \bar{T} - \frac{1}{\mu})$$

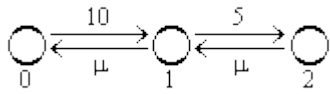
Exercice 4

Une entreprise de construction possède deux engins identiques, chacun pouvant tomber en panne indépendamment de l'autre suivant un processus de Poisson de taux 5 fois par mois. On suppose que la durée de réparation est une v.a. qui suit une loi exponentielle de paramètre μ (taux de service).

Pour quelle valeur de μ , les deux engins seront-ils simultanément en état de marche au moins la moitié du temps ?

Solution

Le système est modélisé par un processus de naissance et de mort à 3 états (0, 1 ou 2 machines en panne) de diagramme de transition suivant :



Nous avons :

$$p_1 = \frac{10}{\mu} p_0, \quad p_2 = \frac{50}{\mu} p_0$$

$$\text{D'où } p_0 = \left[1 + \frac{10}{\mu} + \frac{50}{\mu^2} \right]^{-1} = \frac{\mu^2}{\mu^2 + 10\mu + 50}$$

p_0 étant la probabilité stationnaire pour que les deux engins soient en état de marche, on devra avoir :

$$\begin{aligned} p_0 \geq \frac{1}{2} &\implies \frac{\mu^2}{\mu^2 + 10\mu + 50} \geq \frac{1}{2} \\ &\implies \mu^2 - 10\mu - 50 \geq 0 \end{aligned}$$

Tenant compte de la positivité de μ , on doit avoir

$$\mu \geq 5 + \sqrt{75} = 5(1 + \sqrt{3}) \approx 13,660254$$

Il faudra donc un taux de service au moins égal à **13,660254** pour que les deux engins soient en état de marche au moins une fois sur deux.

Exercice 5

Une blanchisserie possède 3 machines identiques et indépendantes. Chaque machine a une durée de fonctionnement indépendante du passé, suivant une loi exponentielle d'espérance de deux jours.

Une machine qui tombe en panne est réparée par un technicien ; la durée de réparation est une v.a. d'espérance d'un jour. La blanchisserie n'a qu'un technicien.

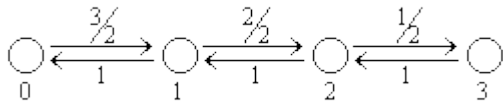
Dessiner le diagramme de transition.

Quelle est la fraction de temps où toutes les machines fonctionnent et celle où toutes les machines sont hors service ?

Quel est le nombre moyen de machines en état de marche ? Quel est le nombre moyen de machines immobilisées ?

Solution

a. Supposons que l'état E_j désigne l'état dans lequel j machines sont immobilisées. Nous avons le diagramme de transition suivant :



b. Nous avons en régime stationnaire les équations suivantes :

$$P_1 = \frac{3}{2}P_0, \quad P_2 = \frac{3}{2}P_0, \quad P_3 = \frac{3}{4}P_0$$

$$\text{D'où } P_0 = \frac{4}{19}, \quad P_1 = P_2 = \frac{6}{19}, \quad P_3 = \frac{3}{19}.$$

Les fractions recherchées sont donc $\frac{4}{19}$ (pour toutes les machines en marche) et $\frac{3}{19}$ (pour toutes les machines immobilisées).

c. Le nombre moyen de machines en état de marche vaut :

$$3 \times \frac{4}{19} + 2 \times \frac{6}{19} + 1 \times \frac{6}{19} = \frac{30}{19}$$

Et le nombre moyen de machines immobilisées vaut :

$$\frac{9}{19} + \frac{12}{19} + \frac{6}{19} = \frac{27}{19}$$

$$\left(\text{Nous avons bien } \frac{30}{19} + \frac{27}{19} = 3 \right)$$

Exercice 6

Un serveur de base de données reçoit en moyenne 100 requêtes par seconde, arrivant selon un processus de Poisson. Le temps de traitement d'une requête suit la loi exponentielle de paramètre μ . Quand le serveur est occupé, les requêtes sont stockées sur un disque de grande taille pour être traitées ultérieurement selon le principe "premier arrivé, premier servi". Il existe, sur le marché, 4 types différents de serveurs pouvant traiter respectivement 100, 150, 200 ou 300 requêtes par seconde quand ils fonctionnent sans interruption.

Que se passera-t-il si l'on prévoit d'installer un serveur pouvant traiter 100 requêtes par seconde ?

On souhaite qu'un client qui émet une requête ait la réponse au bout de 1/100 seconde en moyenne. Quel type de serveur faut-il prévoir ?

On convient que lorsqu'il y a déjà 8 requêtes stockées dans la file d'attente, les nouvelles requêtes arrivantes seraient trop pénalisées au niveau du temps de réponse. Donc on décide que dans ce cas, elles seront redirigées instantanément sur un serveur auxiliaire. Pour simplifier, on supposera que cette nouvelle disposition a une influence négligeable sur l'ancienne distribution stationnaire (calculée sans serveur auxiliaire). Calculer la probabilité qu'une requête qui arrive sur le serveur principal soit redirigée sur le serveur auxiliaire. Combien, en moyenne, le serveur auxiliaire recevra-t-il de requêtes par seconde ?

Solution

Il s'agit d'une file d'attente M/M/1, nous pouvons donc utiliser les lois de Little. Il y a distribution stationnaire si $\lambda/\mu < 1$, si on prend un serveur qui traite 100/seconde alors nous avons $\lambda/\mu = 1$ la chaîne n'est pas ergodique (accumulation infinie de requête) !

Si on veut un temps de réponse de 1/100 seconde alors nous devons résoudre l'équation $1/100 = E(T) = 1/(\mu - \lambda)$ donc $\mu = 200$ requêtes/seconde.

Les requêtes sont redirigées s'il y a déjà 8 requêtes dans le système. Donc il y a redirection lorsque le système atteint 9 requêtes dans le système. Il suffit de calculer π_9 pour connaître la probabilité que la requête soit redirigée. D'après la loi de Little $\pi_9 = (1 - \rho) \rho^9 = 1/1024$.

Puisque le système reçoit 100 requêtes par seconde, le serveur auxiliaire reçoit 100/1024 requête par seconde.

Exercice 7

Dans une usine de production automobile, on a étudié le problème de la détermination du nombre optimal d'employés à placer aux guichets d'un magasin chargé de fournir l'outillage nécessaire aux ouvriers. L'étude du magasin a commencé par la détermination des caractéristiques statistiques des arrivées des ouvriers (1.6 arrivées/minute) et des temps passés par les employés pour fournir les outillages demandés (0.9 service/minute).

Calculer le temps moyen d'attente dans la file (W_q) pour $S=2$, $S=3$, $S=4$. Calculer le nombre moyen de clients dans une journée de 8 heures et le temps de service moyen correspondant. Selon le nombre d'employés (S), calculer le nombre d'heures d'inactivité par jour de 8 heures.

Calculer le temps perdu chaque jour par les ouvriers, du fait de l'attente ? Si le prix de revient horaire d'un employé est 40,00 \$ et celui d'un ouvrier 80,00 \$, quel est le coût total des heures perdues ?

Conclure en donnant le nombre optimal d'employés à prévoir pour le magasin.

Solution

Connaissant l et m , on a calculé $l/m = 1.6/0.9 = 1.77 > 1$.

Puisque $l/m > 1$, on s'est intéressé uniquement aux valeurs de $S=2$, $S=3$, et $S=4$.

Pour calculer le temps moyen d'attente dans la file, on a d'abord cherché P_0 , pour chaque valeur respective de S .

$$S=2, P_0 = 0.061$$

$$S=3, P_0 = 0.152$$

$$S=4, P_0 = 0.166$$

ce qui donne

$$S=2, W_q = 4.00$$

$$S=3, W_q = 0.31$$

$$S=4, W_q = 0.06$$

Calculons maintenant le nombre moyen de clients dans une journée de 8 heures

$$l \times 60 \times 8 = 1.6 \times 60 \times 8 = 768$$

et, pour ce nombre d'arrivées, il faudra (avec un temps de service de $1/m$)

$$768/m = 768/0.9 = 853 \text{ minutes de service par jour.}$$

$$= 14.21 \text{ heures}$$

Exercice 8

Un laboratoire contient 2 micro-ordinateurs identiques. Les usagers arrivent selon un processus de Poisson de paramètre $l = 3$ usagers par heure. La durée d'utilisation d'un micro-ordinateur est une v.a. exponentielle de moyenne 0.6 heure ($m = 1 / 0.6$).

- Taux d'utilisation du laboratoire
- Nombre d'usagers dans la file
- Temps d'attente dans la file

Le patron trouve que les gens perdent trop de temps à attendre, et décide d'ajouter des micro-ordinateurs. Il désire que le temps moyen d'attente soit inférieur à 20 minutes. Combien doit-il en ajouter ? Quel est le nouveau taux d'utilisation du laboratoire ?

Qu'advierait-il en ce qui a trait au temps d'attente des usagers si on les séparait en 3 groupes, assignant un micro-ordinateur à chaque groupe ?

Qu'en concluez-vous ?

Solution

On a une file d'attente M/M/2.

$$\rho = \lambda/2\mu = 0.9$$

$$P_0 = 1/19 \quad P_1 = \lambda/\mu P_0 = 1.8/19$$

$$Q = [\rho/1-\rho] (1 - P_0 - P_1) = 7.674$$

$$W_q = 2.56 \text{ h}$$

Le système n'est utilisé qu'à 90%, mais chaque usager doit attendre en moyenne plus de 2 heures 30.

Le patron trouve que les gens perdent trop de temps à attendre et décide d'ajouter des micro-ordinateurs. Il veut que le temps moyen d'attente soit inférieur à 20 minutes. Combien doit-il en ajouter ?

On peut essayer avec plusieurs valeurs de $S = s$, et calculer W_q pour chaque cas :

s	Q	Wq
3	0.532	0.177 h. \approx 10.6 min.

Avec 3 micro-ordinateurs, la contrainte est satisfaite, mais le taux d'utilisation du système est maintenant de $\rho = 0.6$.

Qu'advierait-il si on séparait les usagers en 3 groupes, assignant un micro-ordinateur à chaque groupe ?

On a maintenant 3 files d'attente M/M/1, chacune ayant un taux d'arrivée $\lambda = 3/3 = 1$, un taux de service $\mu = 1/0.6$ et un facteur d'utilisation $\rho = 0.6$.

On obtient :

$$W_q = \rho^2 / [\lambda(1-\rho)] = 0.9 \text{ heure}$$

$$= 54 \text{ min.}$$

Il est donc préférable de partager les ressources.

Ainsi,

$S = 2$ employés auraient $2 \times 8 - 14.21 = 1.79$ heures d'inactivité par jour

S = 3 employés auraient $3 \times 8 - 14.21 = 9.79$ heures d'inactivité par jour

S = 4 employés auraient $4 \times 8 - 14.21 = 17.79$ heures d'inactivité par jour

Cherchons maintenant le temps perdu chaque jour par les ouvriers, du fait de l'attente :

$$S = 2 \quad 768 \times 4 = 3072 \text{ min.} = 51.2 \text{ heures}$$

$$S = 3 \quad 768 \times 0.31 = 238 \text{ min.} = 3.96 \text{ heures}$$

$$S = 4 \quad 768 \times 0.06 = 46 \text{ min.} = 0.76 \text{ heures}$$

Le coût total des heures perdues a pour valeur :

$$S = 2 \quad \$ 4\,167.60 = 1.79 \times 40 + 51.2 \times 80$$

$$S = 3 \quad \$ 708.40 = 9.79 \times 40 + 3.96 \times 80 \leftarrow \text{τηε βεστ ονε !}$$

$$S = 4 \quad \$ 772.40 = 17.79 \times 40 + 0.76 \times 80$$