

STI 4A simulation à événements discrets Projet 2018 – 2019

Dans ce projet nous voulons représenter un serveur Web par un réseau de files d'attente ouvert d'une vue simple et de haut niveau. L'objectif est de créer un modèle qui résume les détails matériels et logiciels pour produire des résultats de performance significatifs en ce qui concerne la relation entre la vitesse du serveur et celle du réseau. Dans ce modèle, nous ignorons les détails de bas niveau des protocoles HTTP et TCP / IP. De même, le modèle actuel agit comme un simple serveur de fichiers sur Internet et ignore à la fois les applications CGI et son comportement.

Ce réseau est composé de quatre files d'attente à serveur unique : deux pour la modélisation du serveur Web lui-même et pour deux la modélisation du réseau de communication Internet. Les demandes de fichier arrivent sur le serveur Web avec le taux d'arrivée du réseau (A) correspondant au nombre moyen de demandes de fichiers HTTP ("hits") reçues par le serveur Web à chaque seconde. La taille moyenne des fichiers (F) est la taille moyenne (en octets) des fichiers servis. De toute évidence, cette valeur variera considérablement d'un site Web à l'autre. Tous les traitements "d'initialisation" uniques sont effectués dans la file d'attente SI. Le travail se poursuit ensuite vers la file d'attente SR où une seule partie du fichier demandé est lue, traitée et transmise au réseau.

Le taux de service de SR est $1 / [Y + (B / R)]$ où Y représente le temps passé à traiter une mémoire tampon indépendante de la taille de cette mémoire tampon (B) et R représente le débit (octets / seconde) auquel le serveur traite le tampon.

A la file d'attente SS, ce bloc de données est transmis à Internet à débit de transfert du serveur (par exemple, 1,5 Mbits sur une ligne T1). S indique la vitesse à laquelle le serveur envoie un tampon de données à Internet. Les valeurs typiques pour S sont 128 Kbits / s - RNIS, 1,5 Mbits / s - T1 et 6 Mbits / s - T3. C désigne la vitesse moyenne à laquelle le logiciel client reçoit un tampon.

Ces données sont transités par Internet et sont reçues par le navigateur du client, représenté par la file d'attente SC. Si le fichier n'a pas été entièrement transmis, le "travail" se branche et revient à la file d'attente SR pour traitement ultérieurement. Sinon, le travail est terminé et quitte le réseau de files d'attente. Pour une taille de fichier moyenne F et une taille de tampon B , la probabilité que le fichier ait été entièrement transmis est de $p = B / F$.

Donc pour notre modèle les huit paramètres sont:

- Taux d'arrivée du réseau (A)
- Taille moyenne du fichier (F)
- Taille du tampon (B)
- Temps d'initialisation (I)
- Heure du serveur statique (Y)
- Taux de serveur dynamique (R)
- Bande passante réseau du serveur (S)
- Bande passante réseau client (C)

Avant d'analyser le modèle, il est important de comprendre la signification des huit paramètres du modèle et leur mode de calcul.

Toutefois, après avoir visité 1 000 pages Web de manière aléatoire en utilisant la fonction de "lien aléatoire" disponible auprès de plusieurs moteurs de recherche et en notant la taille de chaque fichier reçu, y compris les graphiques, la moyenne ainsi obtenue était de 5 275 octets. Cette valeur a été utilisée pour générer certains des graphiques ci-dessous, lorsqu'une valeur fixe de F était requise.

La taille de la mémoire tampon (B) correspond à la taille des fragments de fichier envoyés par le serveur via Internet au navigateur du client. Souvent, cette valeur correspond à la taille du bloc de disque de la machine serveur. L'analyse de notre modèle montre que cette valeur joue un rôle insignifiant sur les performances globales du serveur. Une valeur arbitraire de 2000 octets a été utilisée pour générer tous les graphiques ci-dessous.

La durée d'initialisation (I), la durée de service statique (Y) et la vitesse du serveur dynamique (R) décrivent de manière collective la vitesse à laquelle le serveur Web traite les demandes. Nous avons représenté le temps moyen nécessaire

pour effectuer diverses tâches d'initialisation uniques pour chaque travail (par exemple, mappage de suffixe). La durée de service du nœud SI est I . Y représente le temps passé à traiter une mémoire tampon indépendante de la taille de cette mémoire tampon. Enfin, R représente le débit (octets / seconde) auquel le serveur traite le tampon.

Les serveurs Web fonctionnant sur des ordinateurs modernes peuvent généralement traiter les données beaucoup plus rapidement que les réseaux actuels ne peuvent le transmettre. La bande passante réseau du serveur (S) et la bande passante réseau du client (C) représentent collectivement la vitesse de transmission d'Internet. S indique la vitesse à laquelle le serveur envoie un tampon de données à Internet. Les valeurs typiques pour S sont (128 Kbits / s - RNIS, 1,5 Mbits / s - T1 et 6 Mbits / s - T3). C désigne la vitesse moyenne à laquelle le logiciel client reçoit un tampon. En faisant la moyenne des résultats d'une enquête réalisée auprès d'un utilisateur d'Internet actuel, une valeur raisonnable pour C est de 707 Kbits / s.

Exercices

1) Modéliser le réseau de files d'attente.

2) A l'aide de Ciw, calculer le temps moyen de réponse (la durée de séjour) de chaque demande selon la configuration suivante :

- Temps d'initialisation (I) : 0 s
- Durée du service statique (Y) : 0 s
- Taux de serveur dynamique (R) : 10 Mbits / s
- Bande passante réseau du serveur (S) : 1,5 Mbits / s
- Taux d'arrivée du réseau (A) : entre 0/s et 35/s

Tracer la courbe du temps moyen de réponse en fonction de taux d'arrivée A .

Afin de diminuer le temps de réponse, trois méthodes ont été proposées :

- Doubler la vitesse du serveur (R)
- Double la bande passante réseau du serveur (S)
- Double le nombre de serveur (dans la file d'attente SR)

3) Evaluer et classer les trois propositions selon leurs efficacités de manières théorique et de simulation.