

ECOLE POLYTECHNIQUE DE MONTREAL

Département de génie informatique et génie logiciel

Cours INF4410: Systèmes répartis et infonuagique (Automne 2014)

3 crédits (3-1.5-4.5)

Contrôle périodique

DATE: Vendredi le 7 novembre 2014

HEURE: 9h30 à 11h20

DUREE: 1H50

NOTE: Toute documentation permise, calculatrice non programmable permise

Ce questionnaire comprend 4 questions pour 20 points

Question 1 (5 points)

- a) Les techniciens du département doivent copier une nouvelle version de l'image des postes de laboratoire, à partir d'un serveur, vers chacun des 150 postes. L'image occupe 45×10^9 octets. Ils utilisent le protocole UDP en multi-diffusion et chaque poste envoie un accusé de réception négatif pour chaque paquet manquant. Ainsi, chaque paquet est numéroté et les stations redemandent les paquets manquants lorsqu'ils réalisent qu'il leur manque un paquet dans la séquence. On néglige l'espace requis pour les entêtes et la numérotation des paquets. La probabilité de perdre un paquet, (qu'il ne se rende pas à un poste donné), est proportionnelle à sa longueur $po \times n$ où po est la probabilité par octet et vaut 10^{-10} , et n est la longueur des paquets et doit être compris entre 1 et 10^8 . Si la longueur des paquets est de 10^7 octets, combien d'accusés de réception négatifs seront reçus suite à l'envoi initial de l'image, (avant la retransmission des paquets manquants qui pourraient à leur tour être perdus). **(2 points)**
- b) L'Internet, comme la plupart des réseaux, est considéré non fiable. Des messages peuvent être interceptés, lus, enlevés, ajoutés ou modifiés. Pour pallier à cela, divers mécanismes comme la double encryption, avec des clés privées et publiques, peuvent permettre d'authentifier l'expéditeur et de s'assurer que seul le destinataire prévu peut lire le message. Est-ce suffisant dans tous les cas de s'assurer que si le message se rend, il est authentique et lisible seulement par son destinataire? Sinon, quels sont les problèmes possibles et quelles autres précautions seraient requises? **(2 points)**
- c) Un nouvel atelier de fabrication doit être installé et contiendra une fraiseuse à commande numérique, équipement dispendieux, sujet à l'usure, et avec une longue durée de vie. Cette fraiseuse constitue un petit système réparti avec un ordinateur central et quatre ordinateurs de commande, un pour le moteur de chacun des axes X, Y et Z, ainsi qu'un pour le moteur de la fraiseuse qui fait tourner l'outil de coupe. Un des administrateurs de l'atelier insiste pour que les ordinateurs de commande soient ouverts et leur logiciel à code source libre, alors qu'un deuxième insiste pour que simplement le protocole utilisé pour parler à ceux-ci soit ouvert. Finalement, le troisième préfère que tout soit fermé et secret pour éviter la tentation de jouer dedans et pour minimiser les risques de sécurité informatique. Quelle est l'utilité d'avoir un système ouvert à code source libre? Un système avec un protocole ouvert? **(1 point)**

Question 2 (5 points)

- a) Vous devez réaliser un service de fichiers simple à l'aide de Java RMI. Le client peut faire les appels `fp->lire(nom)` qui retourne `contenu`, et `fp->ecrire(nom, contenu)`. L'argument `nom` (nom du fichier) est de type `String`, et `contenu` (contenu du fichier) est de type `byte[]`, alors que `fp` est une instance qui offre l'interface `Fichier`. Le nom du fichier est par rapport au répertoire courant du processus qui offre ce service. Écrivez l'interface `Fichier`, et la classe `FichierServant` (qui fournit les méthodes `lire` et `écrire` sur le serveur). La classe `FichierServant` lit ou crée et écrit en conséquence les fichiers demandés dans le répertoire courant. Pour simplifier le problème, vous n'avez pas à gérer les erreurs et les exceptions, montrer les inclusions (`import`) ou fournir les destructeurs ou constructeurs. Vous ne

- devez écrire que l'interface `Fichier` et la classe `FichierServant`, sans fournir le code qui utilise ces dernières pour instancier et exporter le service, ni le code pour implémenter le client. **(2 points)**
- b) Un des premiers systèmes d'appel à distance largement utilisés est SUN RPC. Il est maintenant disponible avec TCP ou UDP. En supposant qu'avec TCP la connexion est maintenue pour plusieurs appels à distance, les coûts de création et de destruction de la connexion peuvent facilement être amortis sur de nombreux appels. Par contre, pour chaque paquet envoyé, on suppose que TCP envoie un accusé de réception. Un appel RPC simple est effectué entre deux programmes et aucun paquet n'est perdu. Combien de paquets seront transmis si UDP est utilisé? Si TCP est utilisé? Les SUN RPC utilisent une sémantique *au moins une fois*, quel comportement auquel il faut faire attention cela peut-il causer sur UDP? sur TCP? **(2 points)**
- c) Quels sont les avantages respectifs des systèmes d'appels à distance CORBA et Java RMI? Expliquez? **(1 point)**

Question 3 (5 points)

- a) Un serveur de fichiers sert de nombreux clients. Chaque client lit un fichier de code source (toujours le même), l'édite pendant 5 minutes, le sauve (réécrit son contenu), le compile et l'exécute. Le fichier de code source occupe 16Kio. La compilation demande de lire 128Kio de fichiers (les 16Kio du fichier de code source, et les 112Kio de 8 fichiers d'entête ou de bibliothèques qui sont les mêmes d'une compilation à l'autre) et d'écrire un exécutable de 64Kio. Finalement, l'exécution demande de lire l'exécutable qui vient d'être écrit. Toutes ces opérations de compilation et d'exécution, (hormis l'édition par l'utilisateur), prennent un temps très court (i.e., moins d'une seconde). Le protocole NFS 3 est utilisé pour ce service de fichiers. Il n'y a pas d'autres accès que ceux décrits ici (e.g., personne qui modifie les fichiers d'entête ou de bibliothèques), et beaucoup de mémoire est disponible pour conserver localement une copie des fichiers lus ou écrits récemment par NFS. On peut donc déduire de cette description l'état de chaque fichier accédé: absent, en cache localement et validé depuis plus que 3 secondes, ou depuis moins que 3 secondes. Contrairement aux lectures ou aux écritures qui se font un bloc à la fois, la validation fournit la date de dernière modification du fichier au complet, et une seule requête de validation couvre tous les blocs du fichier. Chaque requête au serveur, lecture de 4Kio, écriture de 4Kio ou validation prend 2ms de temps CPU sur le serveur. En plus, chaque lecture prend 8ms de disque 30% du temps, chaque validation prend 8ms de disque 10% du temps, alors que chaque écriture prend 8ms de disque à chaque fois. Combien de clients ce serveur peut-il supporter au maximum s'il contient un CPU et un disque? **(2 points)**
- b) Un centre de recherche se prépare pour une expérience qui générera un très grand débit de données. On compte 5000 capteurs, et chacun génère 100Mbit/s et est connecté sur une prise réseau 1Gbit/s. Toutes les prises réseau sont commutées et connectées à un ensemble de serveurs. Chaque serveur est connecté à l'aide de deux cartes réseau 10Gbit/s sur deux prises qui peuvent soutenir le même débit. Le bus de données de chaque serveur a une capacité de 16Goctet/s. Chaque ordinateur a 16 disques connectés sur son bus, chaque disque ayant une bande passante

de 100Moctets/s en régime continu. L'idée est de bien répartir la charge entre tous les disques sur tous les serveurs, afin de supporter tous ces capteurs avec le nombre minimal possible de serveurs. Quel est ce nombre minimal de serveurs requis? Quel système de service de fichiers pourrait typiquement être utilisé pour une telle organisation? **(2 points)**

- c) Donnez un exemple d'application où BitTorrent est beaucoup plus intéressant que Gnutella et un autre où Gnutella est plus intéressant. Expliquez. **(1 point)**

Question 4 (5 points)

- a) En bourse, il existe des règles sur la divulgation d'information. Un employé qui a de l'information privilégiée sur la situation de sa compagnie ne peut pas utiliser cette information pour jouer à la bourse et avoir un avantage sur les non initiés. Lorsqu'un spéculateur fait un gros coup en bourse, une enquête est souvent effectuée pour voir s'il a reçu de l'information privilégiée illégalement. Vous devez concevoir le système de communication par messages d'une compagnie de manière à aider ces enquêtes. En effet, on veut pouvoir facilement prouver qu'il n'y a pas eu de chaîne de messages liant une information privilégiée et un ordre d'achat. Par exemple, supposons que A reçoit une telle information, envoie ensuite un message à B qui ensuite envoie un message à C. Si C a envoyé un ordre d'achat avant la réception de ce message, il n'y a pas de lien possible, alors que si l'ordre a été envoyé après, il y a un délit potentiel. On suppose que toutes les communications se font de manière électronique par messages, qu'il est possible d'ajouter une étiquette avec de l'information à chaque message, et qu'un journal des messages échangés est sauvegardé, avec les informations pertinentes comme l'étiquette du message et les temps d'envoi et de réception. Toutefois, une estampille de temps sur l'envoi et la réception de chaque message ne suffirait pas dans le journal pour les fins d'enquête, car les horloges des différents ordinateurs ne sont pas synchronisées. Proposez une méthode d'étiquetage ou de gestion des messages qui fera en sorte qu'il soit possible facilement de déterminer si deux événements (information reçue et ordre d'achat) peuvent être liés. **(2 points)**
- b) Un gros fournisseur Internet supporte les requêtes DNS de ses nombreux clients. Chaque client effectue en moyenne 1 requête DNS par seconde. Le serveur DNS peut répondre à i) 95% des requêtes en 1ms de CPU, ii) 4% des requêtes en 2ms de CPU et 10ms de disque, et iii) 1% des requêtes en 4ms CPU, 10ms de disque et 100ms d'attente après le réseau pour faire la requête à un autre serveur dans la hiérarchie. Cet ordinateur possède 8 CPU et 4 disques. Combien de clients peut-il supporter au maximum? Combien de fils d'exécution doit-il utiliser au minimum pour cela? **(2 points)**
- c) Une grosse entreprise veut installer un service de noms pour maintenir et rendre disponible les informations sur ses usagers et ses ordinateurs. Ceci lui permettra d'avoir un grand nombre de postes de travail pratiquement sans information de configuration, et n'importe quel usager pourra ainsi travailler à partir de n'importe quel poste de travail, tout en utilisant son mot de passe usuel et en accédant ses fichiers à partir du serveur. La compagnie hésite entre un service X500, LDAP et Active Directory. Que leur conseillez-vous? Quels sont les avantages et inconvénients de chacun? **(1 point)**

Le professeur: Michel Dagenais