

POLYTECHNIQUE MONTRÉAL

Département de génie informatique et génie logiciel

Cours INF8480: Systèmes répartis et infonuagique (Automne 2024)

3 crédits (3-1.5-4.5)

CORRIGÉ DE L'EXAMEN FINAL

DATE: Mercredi le 11 décembre 2024

HEURE: 9h30 à 12h00

DUREE: 2H30

NOTE: Aucune documentation permise sauf un aide-memoire, préparé par l'étudiant, qui consiste en une feuille de format lettre manuscrite recto verso, calculatrice non programmable permise

Ce questionnaire comprend 5 questions pour 20 points

Question 1 (5 points)

- a) Un ordinateur A envoie un message à B à 9h30m00.000s pour obtenir le temps et reçoit une réponse à 9h30m00.800s, ces deux temps étant mesurés avec l'horloge de A. L'ordinateur B reçoit la requête de A à 9h30m02.250s et retourne sa réponse à A à 9h30m02.650s, ces deux temps étant mesurés avec l'horloge de B. Quel est le décalage à appliquer sur A? Quel est l'intervalle d'incertitude associé? (2 points)

$$a = 9h30m02.250s - 9h30m00.000s = 2.250s$$

$$b = 9h30m02.650s - 9h30m00.800s = 1.850s$$

$$Ajustement = (a+b)/2 = (2.250s + 1.850s)/2 = 2.050s$$

$$Précision = (a-b)/2 = (2.250 - 1.850s)/2 = 0.200s$$

Le décalage à appliquer à A est de 2.050s et l'incertitude est de +/- 0.200s.

- b) Un groupe de 36 processus sont reliés entre eux sous la forme d'un anneau. Si on déclenche une élection en anneau, combien de messages doivent être envoyés autour de l'anneau i) au minimum et ii) au maximum avant d'identifier le processus élu? iii) Combien de messages s'ajoutent pour que l'élection soit complétée et le résultat communiqué à tous les processus? (2 points)

Si le processus qui déclenche l'élection est celui le plus prioritaire, au bout de 36 messages le vote sera conclu et il saura qu'il a gagné. Par contre, si le processus qui suit le plus prioritaire est celui qui déclenche l'élection, il faudra 35 messages avant que celui plus prioritaire ne vote et ensuite 36 messages pour qu'il se sache élu. Le minimum est donc i) 36 et le maximum ii) 35 + 36 = 71. Ensuite, il faut iii) 36 messages pour faire le tour et annoncer à tous le résultat de l'élection.

- c) Un système réparti avec plusieurs processus utilise des horloges logiques. A chaque événement dans un processus, son horloge est incrémentée de 1. Lorsqu'un message est envoyé, la valeur de l'horloge est jointe au message. Lorsqu'un message est reçu, l'horloge logique du processus en réception prend le maximum de son horloge et de la valeur jointe avec le message. Si on compare la valeur de l'horloge logique pour deux événements sur deux processus différents, on se demande ce qu'on peut dire de l'ordre relatif de ces événements. i) Si un événement est postérieur à un autre avec un lien de causalité qui démontre la postériorité, quelle sera la relation entre les valeurs d'horloges logiques? ii) Si un événement est postérieur à un autre sans lien de causalité qui démontre la postériorité, est-ce que son horloge logique sera nécessairement plus grande? (1 point)

Pour démontrer causalement qu'un événement d'un premier processus est postérieur à un autre sur un second processus, il faut qu'il soit postérieur sur le même processus, ou qu'il soit sur un deuxième processus après la réception d'un message venant du premier processus. L'incréméntation des horloges logiques suit ces liens de causalité (incrément à chaque événement sur le même processus et incrément par maximum lors d'un envoi de message). Ainsi, pour un message qui a un lien de causalité qui en démontre la postériorité, l'horloge de l'événement postérieur sera plus grande que celle du premier événement. S'il n'y a pas de lien de causalité entre deux événements, la valeur de l'horloge logique de l'événement postérieur peut être inférieure, égale ou plus grande que celle du premier événement, selon comment l'horloge logique est incrémentée dans chaque processus séparé.

Question 2 (4 points)

- a) Un nouveau conte de Noël, intitulé Le sapin de Pollux, vient d'être rendu disponible à l'adresse www.pollux.org. L'adresse de www.pollux.org est disponible du serveur de noms pour le domaine pollux.org avec un TTL (Time To Live) de 2 minutes. L'adresse du serveur de noms pour pollux.org est disponible sur les 13 serveurs racine de l'Internet avec un TTL de 24h. De nombreux clients des 39 fournisseurs Internet officiels de Noël, hohoho.ca, hohoho.fr, hohoho.it..., veulent accéder le conte à cette adresse, www.pollux.org. Ces clients passent toujours par le serveur de noms de leur fournisseur Internet pour obtenir l'adresse cherchée, ils n'ont pas de cache local. Le serveur de noms de chaque fournisseur Internet fait des accès récursifs, de manière à maintenir en cache, pour toute leur période de validité, toutes les entrées recherchées par ses clients. On suppose que les requêtes de ces fournisseurs Internet sont uniformément réparties entre les serveurs de noms racine. Chacun de ces 39 fournisseurs Internet possède exactement 2880 clients, et chaque client accède au conte une fois le jour de sa sortie. Les requêtes sont réparties uniformément sur les 24 heures du jour de sortie de ce conte. i) Combien de ces requêtes pour le serveur

de noms de pollux.org est-ce que chaque serveur racine recevra en moyenne pour ce jour de sortie du conte? ii) Combien de requêtes par minute est-ce que chaque serveur de nom d'un de ces fournisseurs Internet recevra pour www.pollux.org pendant cette journée de sortie du conte? iii) Combien de requêtes pour www.pollux.org est-ce que le serveur de noms du domaine pollux.org recevra pendant cette journée de sortie du conte? **(2 points)**

Le serveur de noms de chacun des 39 fournisseurs Internet demandera une fois l'adresse du serveur de noms pour le domaine pollux.org. Ces requêtes seront réparties entre les 13 serveurs racine, ce qui donnera $39 / 13 = 3$ requêtes pour chaque serveur racine.

Puisque chacun des 2880 clients d'un fournisseur fera cette requête une fois pendant la journée, cela donne $2880 / (24 \text{ heures} \times 60 \text{ minutes}) = 2$ requêtes par minute.

Chaque serveur de noms d'un fournisseur Internet fera une requête par 2 minutes, soit 30 requêtes / heure au serveur de noms du domaine pollux.org. Le nombre de requêtes pour la journée sera donc de $39 \times 24 \text{ heures} \times 30 \text{ requêtes} / \text{heure} = 28080$ requêtes.

- b) Une entreprise veut installer un service LDAP redondant et très sécuritaire. Pour ce faire, chaque client peut interroger un ou plusieurs de 6 serveurs redondants. Supposons que chaque serveur a une probabilité d'être disponible de 0.45. i) Si ces serveurs présentent des pannes par omission, combien de serveurs en panne peut-on tolérer et quelle est la probabilité que le service soit non disponible au client? ii) Si les pannes sont de type réponse aléatoire et le client prend un vote parmi les réponses des 6 serveurs, combien de serveurs en panne peut-on tolérer, et que devient la probabilité que le service soit non disponible? **(2 points)**

i) Avec des pannes par omission, on peut tolérer jusqu'à 5 serveurs en panne. La probabilité que les 6 serveurs soient en panne est de $(1 - 0.45)^6 = 0.027680641$. ii) Avec des pannes de réponse aléatoire, il faut avoir au moins 2 serveurs disponibles. Nous avons une probabilité de $(1 - 0.45)^6 = 0.027680641$ d'avoir les 6 serveurs en panne, et de $6 \times (1 - 0.45)^5 \times 0.45^1 = 0.135886781$ d'avoir exactement 5 serveurs en panne et 1 disponible. Ceci représente les cas où le service sera non disponible, une probabilité totale de $0.027680641 + 0.135886781 = 0.163567422$. Les autres cas, 2 à 6 serveurs disponibles, font que nous avons au moins 2 serveurs disponibles et le service fonctionne.

Question 3 (4 points)

- a) Les transactions T, U et V s'exécutent en même temps et leurs opérations de lecture et d'écriture sur des variables (a, b, c, d, e) sont entrelacées. Les lectures d'une transaction sont effectuées sur les versions courantes des variables, et les écritures d'une transaction sont effectuées sur une version provisoire des variables pour la transaction. Lorsque la transaction se termine et est acceptée, la version provisoire des variables écrites par la transaction devient la version courante. Une validation de la cohérence par contrôle optimiste de la concurrence est effectuée pour accepter ou non chaque transaction. Il faut tenir compte des transactions précédentes qui ont été validées (et ignorer celles qui ne l'ont pas été) pour savoir si chacune des transactions est acceptée ou non. Lesquelles des transactions T, U et V pourraient être validées, i) si une validation en reculant était utilisée pour vérifier la cohérence des transactions? ii) Une validation en avançant? Pour chaque transaction non validée, donnez la ou les variables en conflit. **(2 points)**

- | | |
|------------------|-------------------|
| 1 T: Début | 10 T: Compléter |
| 2 T: Read(a) | 11 U: Read(b) |
| 3 T: Write(b, 1) | 12 V: Read(a) |
| 4 T: Write(d, 3) | 13 V: Read(d) |
| 5 U: Début | 14 V: Read(c) |
| 6 U: Read(a) | 15 V: Write(c, 3) |
| 7 U: Read(e) | 16 V: Write(e, 7) |
| 8 U: Write(c, 2) | 17 V: Compléter |
| 9 V: Début | 18 U: Compléter |

Pour la validation en reculant i), au moment où T termine, il ne peut y avoir de problème et T est accepté. Pour compléter V, on examine ce qui est lu (a, c, d) versus ce qui a été écrit par T (b, d) et V ne peut compléter (conflit sur d). Rendu à compléter U, ce qui est lu (a, b, e) intersecte avec T (b, d) et U doit aussi être abandonné (conflit sur b). Pour la validation en avançant ii), au moment où T termine, ce que T a écrit (b, d) n'intersecte pas avec ce que U a lu (a, e) ni V (rien encore) et T peut compléter. Au moment de compléter V, les écritures (c, e) intersectent

avec les lectures de $U(a, b, e)$ et V ne peut compléter (conflit sur e). La dernière transaction, U , peut nécessairement compléter.

- b) Une transaction répartie T , gérée par le coordonnateur C , veut écrire les variables $a=2$ et $b=3$ sur le serveur s_1 , $c=22$ et $d=33$ sur le serveur s_2 , et $e=12$ et $f=24$ sur le serveur s_3 . Cette transaction répartie est commise en utilisant le protocole de fin de transaction atomique à deux phases. i) Quels seront les messages échangés entre le coordonnateur C et le serveur s_3 pour ce protocole de fin de transaction à 2 phases? ii) Quelles sont les entrées qui seront ajoutées au journal du serveur s_3 pour cette transaction, et à quel moment par rapport aux messages échangés avec le coordonnateur? **(2 points)**

Le coordonnateur C va envoyer un message à tous les participants, incluant s_3 , pour savoir s'ils sont prêts à accepter la transaction (canCommit?). Le serveur s_3 répondra qu'il est prêt après avoir écrit dans son journal les nouvelles valeurs écrites et le fait qu'il est prêt à commettre cette transaction (P1: write($e, 12$); P2 write($f, 24$); P3 Prepare T(P1, P2) P0;). Si tous les serveurs répondent positivement au coordonnateur, celui-ci procédera avec la seconde phase et annoncera à tous, incluant s_3 , que la transaction est acceptée (doCommit). Le serveur s_3 écrira dans son journal que la transaction est acceptée (P4: Completer T, P3) et répondra au coordonnateur C que c'est fait (haveCommitted).

Question 4 (4 points)

- a) Un service de base de données réparti est offert par 3 serveurs redondants. Au moins un serveur doit être disponible pour que le service soit disponible. Chaque serveur est constitué d'un boîtier et son électronique, avec une probabilité de disponibilité de 0.6, ainsi que d'un ensemble de disques en RAID, 5 disques dont au moins 3 doivent être fonctionnels. La probabilité d'être fonctionnel pour un disque est de 0.7. i) Quelle est la probabilité qu'un ensemble de disques en RAID soit fonctionnel? ii) Un serveur de base de données? iii) Le service de base de données réparti? **(2 points)**

L'ensemble de disques RAID a une probabilité de fonctionner de $0.7^5 = 0.16807$ (5 disques fonctionnels) plus $5!/((5-4)!4!) \times 0.7^4 \times (1-0.7)^{5-4} = 0.36015$ (exactement 4 disques fonctionnels) plus $5!/((5-3)!3!) \times 0.7^3 \times (1-0.7)^{5-3} = 0.3087$ (exactement 3 disques fonctionnels), pour un total de $0.16807 + 0.36015 + 0.3087 = 0.83692$. Un serveur est opérationnel si le boîtier est fonctionnel de même que son ensemble de disques RAID, ce qui donne $0.6 \times 0.83692 = 0.502152$. Le service sera disponible sauf si les 3 serveurs sont en panne, une probabilité de $1 - (1 - 0.502152)^3 = 0.876607063$.

- b) Sur un réseau qui ne supporte pas la multi-diffusion, un message de groupe atomique est envoyé d'un membre M aux autres. Le groupe compte 28 membres. i) Si aucun message n'est perdu, combien de messages envoyés au total sur le réseau seront requis pour compléter ce message atomique? ii) Si un des messages envoyés par M à un autre participant est perdu, comment est-ce que cela peut être détecté? **(2 points)**

M envoie un message aux 27 autres membres et attend les 27 réponses positives. Ensuite, il envoie la confirmation aux 27 autres membres. Le total est donc de $3 \times 27 = 81$ messages. Si un message de préparation est perdu, M détectera un problème lorsqu'il ne recevra pas de réponse positive ou négative. Si la réponse est perdue, la détection se fera de même. Si un message de confirmation est perdu, le membre qui ne reçoit pas détectera qu'il a reçu un message de préparation, mais pas de message de confirmation correspondant.

Question 5 (3 points)

Vous devez planifier l'installation d'un nouveau centre de données au Québec. Ce centre regroupera des milliers de serveurs afin d'offrir des services infonuagiques.

- a) En tant qu'ingénieur, quelles sont les lois applicables pour un tel projet, en ce qui concerne les aspects de développement durable? **(1 point)**

L'ingénieur doit tenir compte des conséquences de l'exécution de ses travaux sur l'environnement dans une perspective de développement durable. Ceci est prescrit par la loi canadienne de 2008 sur le développement durable, la loi québécoise de 2006 sur le développement durable, et le code de déontologie de l'Ordre des Ingénieurs du Québec. La procédure d'évaluation environnementale du BAPE ne s'applique normalement pas à un centre de données.

- b) Outre le volet purement financier, dans une approche de développement durable et d'analyse du cycle de vie, quels sont les quatre niveaux pour lesquels on doit évaluer l'impact environnemental d'un tel projet? **(1 point)**

Pour toutes les activités qui se retrouvent dans les différentes phases du cycle de vie d'un projet, il faut examiner l'impact sur la santé humaine, sur l'écologie, sur les changements climatiques et sur l'appauvrissement des ressources.

- c) Pour un projet de centre de données, quels sont typiquement les impacts environnementaux les plus importants pour chacune des phases du cycle de vie? **(1 point)**

Les phases du cycle de vie sont la fabrication (construction du site et fabrication des équipements), le transport (des matériaux pour la construction et des équipements), l'opération du site (entretien, alimentation électrique...) et finalement la fin de vie (la démolition ou conversion du bâtiment, la restauration du site, et le recyclage ou la mise aux rebuts des matériaux et équipements). Pour un centre de données typique, une grande partie de l'impact est reliée à la consommation d'énergie pendant l'opération de l'équipement informatique et de la climatisation. Pour un centre au Québec, étant donné la disponibilité de l'énergie hydroélectrique, l'impact de la consommation d'électricité, sans être négligeable, est moindre que dans la plupart des autres endroits. Un autre impact important est la fabrication des équipement informatiques et électriques, en particulier le raffinage de l'or et du cuivre requis pour ces équipements. La construction du site, le transport et le démantèlement du site à la fin ont normalement un impact beaucoup moins grand que l'opération et que la fabrication des équipements.

Le professeur: Michel Dagenais