Parallélisme imbriqué en OpenMP

1 Le problème du voyageur de commerce

On cherche à optimiser une tournée d'un commercial, tournée passant par un ensemble de villes et revenant à son point départ. Ici on considère un espace euclidien dans lequel les villes sont toutes connectées deux à deux.

1.1 Quelques mots sur le code

Le nombre de villes est contenu dans `NrTowns` et la variable `minimun` contient la longueur de la plus petite tournée connue.

Lors d'un appel `void tsp (hops, len, path)`, le paramètre `path` contiendra un chemin de `hops` entiers (villes) tous distincts ; la longueur de ce chemin est `len`.

La variable `grain` contient le niveau de parallélisme imbriqué demandé (0 - pas de parallélisme ; 1 - une seule équipe de thread est créée au premier niveau de l'arbre de recherche ; 2 - des équipes de threads sont en plus créées au niveau 2 de l'arbre, etc).

1.2 Version séquentielle

Etudiez rapidement l'implémentation fournie. Essayez-la pour vérifier qu'elle fonctionne (avec 12 villes et une `seed` 1234, on trouve un chemin minimal 278). Vous pouvez décommenter l'appel à `printPath` pour observer la solution, mais pour les mesures de performances on ne gardera pas l'affichage. A des fins de calcul d'accélération, mesurer le temps nécessaire pour le cas 13 villes et une `seed` 1234.

Notez que l'analyse de l'arbre de recherche est exhaustive et donc l'analyse engendrée par deux débuts de chemins de k villes nécessitent le même nombre d'opérations (leur complexité ne dépend finalement que k et du nombre de villes).

2 Parallélisation en créant de nombreux threads

Dupliquer le répertoire source. Puis insérer le pragma suivant :
`#pragma omp parallel for if (hops <= grain)`
juste avant la boucle qui lance l'exploration des sous arbres :
`for (i=1; i < NrTowns; i++)`
de la fonction `tsp`.

Poursuivez la parallélisation du code en faisant attention aux variables partagées ou privées. Par exemple il s'agit d'éviter aux threads de tous travailler sur un unique et même tableau. Notons qu'un tableau ne peut être rendu privé, il est donc nécessaire de recopier le tableau `path` dans un nouveau tableau (alloué dans la pile). Protégez également les accès concurents à la variable `minimun`.

Observez les performances obtenues en faisant varier le paramètre graine (3ième paramètres). Notons que pour créer des threads récursivement il faut positionner la variable d'environnement `OMP_NESTED` à `true` (ou bien faire l'appel `omp_set_nested(1)`) car, par défaut, le support d'exécution d'OpenMP empêche la création récursive de threads.

Les perfomances obtenues ne devraient pas être terrible du tout car ce programme recopie beaucoup trop de chemins et, de plus, l'utilisation du pragma parallel a un surcoût même lorsque qu'une clause `if` désactive le parallélisme.

2.1 Optimisations de la parallélisation

Tout d'abord il s'agit d'éliminer les surcoûts inutiles en dupliquant ainsi le code :

```
if (hops <= grain) { // version parallèle
#pragma omp parallel for ...
    for (i=1; i < NrTowns; i++) {
        ...
    }
} else { // version séquentielle
    for (i=1; i < NrTowns; i++) {
        ...
    }
}
```

Ensuite il faut faire en sorte de ne créer que le nombre nécessaire de threads et, par conséquent, d'utiliser une politique de répartition dynamique.

Enfin on observe qu'il n'est pas utile de protéger par une section critique tous les accès à la variable minimum : seules doivent se faire en section critique les comparaisons susceptible d'entrainer une modification du minimum.

Noter les performances obtenus pour le cas 13 villes et pour les grains allant de 0 à 5. Calculer les accélérations obtenues par rapport à la version séquentielle.

3 Parallélisation à l'aide de la directive collapse

Dupliquer le répertoire source initial. Puis insérer la fonction suivante et l'appeler directement dans le main() :

```
void par_tsp ()
{
  int i,j,k;
#pragma omp parallel for collapse(3) schedule(runtime)
 for (i=1; i < NrTowns; i++)
   for(j=1; j < NrTowns; j++)
     for(k=1; k < NrTowns; k++)
       if(i != j && i != k && j != k)
         {
          int chemin[NrTowns];
          chemin[0] = 0;
          chemin[1] = i;
          chemin[2] = j;
          chemin[3] = k;
          int dist = distance[0][i] + distance[i][j] + distance[j][k];
          tsp (4, dist, chemin) ;
         }
}
```

Calculer les accélérations obtenues pour le cas (13 villes et seed 1234) pour les codes suivants : collapse + distribution dynamique ; collapse + distribution statique ; équipes imbriquées.

Noter les performances obtenus pour le cas 13 villes et collapse(3). Calculer l'accélération obtenue par rapport à la version séquentielle.

4 Optimisation de nature algorithmique

Clairement, il est inutile de poursuivre l'évaluation d'un début de chemin lorsque sa longueur est supérieure au minimum courant (correspondant à la longueur du chemin complet le plus petit qu'on a déjà trouvé). Pour mettre en oeuvre cette optimisation, insérez le test suivant au début des trois versions du tsp.

```
if (len +  distance[0][path[hops-1]]>= minimum)
    return;
```

En considérant le cas « 15 villes, seed 1234 », calculer à nouveau les accélérations obtenues dans cas suivante :

1. équipes imbriquées.
2. collapse(3) + distribution dynamique ;
3. collapse(3) + distribution statique ;
4. collapse(4) + distribution dynamique ;
5. collapse(4) + distribution statique ;

Les résultats sont-ils surprenant ?

Commentaires : Un des effets de cette optimisation est de désiquilibrer le calcul car l'analyse des chemins n'est plus exhaustive : certaines branches sont coupées très haut, d'autres très bas. Le code optimisé à un comportement peu prévisible car la complexité du calcul dépend des résultats intermédiaires. On dit que le code optimisé a un comportement *irrégulier* : en séquentiel et à nombre de villes égales, deux configurations différentes auront des temps de calcul différents. Renforcé par le caractère non déterministe du parallélisme le comportement de l'application parallèle est maintenant peu prédictible : pour un même cas on observe des variations significatives (eg. 25%) du temps de calcul entre différentes exécutions parallèles.

5 Parallélisation à l'aide de tâches OpenMP

Dupliquer le répertoire source initial pour paralléliser l'application à l'aide de tâches. Au niveau du main() il s'agit de créer une équipe de threads et de faire en sorte qu'un seul thread démarre l'analyse. Au niveau de la fonction tsp lancer l'analyse en faisant en sorte de ne créer des tâches parallèles que jusqu'au niveau `grain`. Deux techniques d'allocation mémoire sont à comparer :

1. allocation dynamique : un tableau est alloué dynamiquement et initialisé avant la création de la tâche - ce tableau sera libéré à la fin de la tâche ;
2. allocation automatique : le tableau est une variable locale allouée et initialisée dans la tâche - il est alors nécessaire d'utiliser la directive taskwait après avoir créé toutes les tâches filles.

Comparer les performance obtenues par les deux approches sur le cas 15 villes et seed 1234 pour des grains variant de 1 à 9. Comparer à celles obtenues à l'aide des techniques *imbriquées* et *collapse*.

Relever ensuite le(s) meilleur(s) grain(s) pour 2, 4, 6, 12 et 24 threads. Calculer les accélérations obtenues.