TP 5 : Découverte du réseaux *Myrinet*

Programmation avec MX

Alexandre DENIS, Olivier AUMAGE
denis@labri.fr, aumage@labri.fr
http://dept-info.labri.fr/~denis/

Systèmes et supports d'exécution — 2007–2008

L'objectif de ce TP est de découvrir le réseau haute performance *Myrinet* à travers son interface de programmation MX (*Myrinet Express*).

1 Infrastructures

Salle 007 Le TP a lieu en salle 007, la salle « expérimentale » — seule salle équipée d'un réseau *Myrinet* : cartes dans les PC, fibres optiques (câbles orange) et commutateur MYRICOM.

Live CD Les machines de la salle 007 n'ont pas de système d'exploitation sur le disque dur local. Pour travailler, nous utiliserons donc un système d'exploitation sur CD *bootable* (souvent appelé *live CD* en anglais) de la distribution LINUX UBUNTU v7.10. Le CD originel de la UBUNTU a été partiellement modifié[1] pour les besoins du TP, notamment pour ajouter le support logiciel MX du réseau *Myrinet*. Vous êtes automatiquement connecté avec le compte utilisateur `ubuntu` et vous avez les droits nécessaires pour utiliser la commande `sudo`.

Clavier Le système est par défaut configuré pour un clavier américain. Pensez à adapter la configuration au matériel de la salle.

Accès réseau, authentification NOCAT Les machines sont quasiment isolées du réseau du CREMI et vous n'aurez donc pas un accès direct à votre compte par NFS ni même aux autres machines du CREMI. En revanche, une authentification NOCAT[2] est disponible.

Pour autoriser la connexion, ouvrez un navigateur vers n'importe quelle adresse web. La fenêtre d'authentification s'affiche. Répondez par votre login et mot de passe CREMI. **Gardez cette fenêtre ouverte** pour garder la connexion internet active. Après authentification, l'accès réseau est ouvert pour les services HTTP et SSH. Vous utiliserez `scp` pour sauvegarder votre travail sur votre compte.

[1] Cf `http://forum.ubuntu-fr.org/viewtopic.php?id=132294` pour ceux qui veulent savoir comment modifier un live CD.

[2] `http://nocat.net/`

Sauvegardes

Pensez à faire des sauvegardes fréquentes et régulières !

Comme nous travaillons avec un live CD, **rien** n'est enregistré sur le disque dûr de la machine. Donc, en cas d'arrêt, de redémarrage ou de plantage de la machine, toutes les données non sauvegardées par `scp` vers votre compte CREMI seront perdues.

Pensez à faire des sauvegardes fréquentes et régulières !

2 Présentation de *Myrinet*

Myrinet est un réseau haute performance basé sur des cartes de communication équipées d'un processeur programmable (le *LANai*) et de mémoire SRAM embarquée. Le réseau est entièrement commuté, *full duplex*, et les communications sont fiables (sans perte). Plutôt que de programmer directement la carte *Myrinet* et son *LANai* (ce qui est faisable, mais complexe), nous l'utiliserons au travers de son pilote MX qui contient un *firmware* complet pour le *LANai*, un module noyau, et une bibliothèque utilisateur appelée `myriexpress`.

La documentation complète de MX est disponible en ligne sur `http://www.myri.com/scs/MX/doc/`

Une copie de la documentation MX ainsi qu'un squelette de code pour initialiser la carte sont disponibles à l'adresse suivante : `http://dept-info.labri.fr/~aumage/SSECPD_07/MX/`

Myrinet en salle 007

Le logiciel MX est installé dans le répertoire `/opt/mx`. En revanche, les modules noyau MX ne sont pas chargés par défaut. Il faut donc les charger à la main.

1. Toutes les commandes suivantes doivent être exécutées en tant que `root`. Vous utiliserez donc la commande `sudo`.
2. Par défaut, une machine démarrée avec UBUNTU s'appelle `ubuntu`. Pour éviter que toute la salle 007 porte le même nom, changez le nom de votre machine avec la commande `hostname` (donnez-lui par exemple votre nom, pour s'y retrouver...)
3. Chargez le pilote MX en lançant la commande :

   ```
   /etc/init.d/mx start
   ```

4. Contrôlez la configuration de MX avec la commande :

   ```
   /opt/mx/bin/mx_info -v
   ```

   Votre machine ainsi que toutes celles qui ont déjà chargé le pilote MX doivent apparaître, avec leur nom correct (qui est de la forme « *nom de machine* :0 », le « :0 » signifiant qu'il s'agit de la carte Myrinet numéro 0 de la machine donnée.
5. En cas de problème de configuration, vous pouvez :
   – Corriger le nom de votre machine dans un pilote MX déjà chargé. Utilisez pour celà la commande `/opt/mx/bin/mx_hostname -n` *nom*

– Mettre à jour le cache des tables de routage (s'il vous manque des machines dans `mx_info` ou si les noms ne sont pas à jour). Utilisez la commande `/opt/mx/bin/mx_hostname -c`
– En cas de problème insoluble, rechargez le pilote MX : `/etc/init.d/mx restart`

Vous n'avez ensuite plus besoin de commandes `root`. Tous vos travaux se font sous le nom d'utilisateur `ubuntu`.

3 Travail à faire

Commencer par lire attentivement au moins les sections 1 (*overview*) et 2 (*Concepts*) de la documentation de MX 1.0.

Note :

– pour compiler un code qui utilise MX, n'oubliez l'argument `-I/opt/mx/include/` sur la ligne de commande de compilation. L'en-tête à inclure dans vos programmes est `myriexpress.h`.
– pour l'édition de lien, n'oubliez pas les arguments `-L/opt/mx/lib -lmyriexpress` sur la ligne de commande d'édition de liens.
– vous pourriez être tentés de faire des tests en local. Cependant, le support du *loopback* ne fonctionne pas dans la version de MX présente sur les CD. Il faut obligatoirement utiliser deux machines distinctes pour vos tests.

3.1 Envoi simple

Ecrire un programme émetteur et un programme récepteur s'échangeant un message de typr « *Hello world* » (taille fixe et connue par les deux parties).

3.2 Ping-pong

Étendre le programme précédent pour réaliser un ping-pong, de façon à pouvoir mesurer les performances.

3.3 Benchmark

– Modifier le programme de façon à envoyer des messages de taille croissante (les puissances de 2, de 4 octets à 4 Mo par exemple), et mesurer le temps de transfert pour chacune des tailles. Tracer ensuite la courbe de débit en fonction de la taille des messages à l'aide de `gnuplot`.
– Comparer les performances obtenues en utilisant les primitives de communication synchrones et celles de communication asynchrones.