Mini-projet système : développement d'un noyau de système d'exploitation

Responsable : Christophe Rippert
Christophe.Rippert@Grenoble-INP.fr

Gestion des processus : notions de base

Introduction

Le but des 4 séances restantes est de programmer un noyau de processus simpliste. On se limitera à un mode de fonctionnement volontairement très simplifié par rapport à un noyau de système réaliste. Les aspects fondamentaux des systèmes actuels, tels que les primitives de synchronisation, la communication inter-processus, ... seront vus en détail dans le Projet Système.

Notions générales

Spécification des processus

On a vu en SEPC que les processus sont des unités d'exécution permettant d'implanter la notion de pseudo-parallélisme sur un système mono-processeur (remarque : en pratique, la grande majorité des processeurs actuels sont équipés de plusieurs coeurs d'exécution, mais on ne les exploitera pas dans ce TP). Ce pseudo-parallélisme est réalisé très simplement en accordant à chaque processus un temps d'exécution sur le processeur, puis une fois ce temps écoulé, en arrêtant le processus actif pour passer la main à un autre processus prêt à s'exécuter : l'entrelacement suffisamment rapide de l'exécution des différents processus donne à l'utilisateur l'illusion qu'ils s'exécutent réellement en parallèle.
Dans le cadre de ce TP, chaque processus sera défini par :

- son numéro, qui servira d'identifiant unique dans le système, et que l'on appelle typiquement `pid` (pour *Process IDentifier*) ;
- son nom, une chaine de caractères qui servira à produire des traces lisibles ;
- son état : un processus peut être élu (c'est à dire que c'est lui que le processeur est en train d'exécuter), activable (c'est à dire qu'il est prêt à être élu et qu'il attend que le processeur se libère), endormi pour une durée donnée, ... ;
- son contexte d'exécution sur le processeur : il s'agit en pratique d'une zone mémoire servant à sauvegarder le contenu des registres important du processeur lorsqu'on arrête le processus pour passer la main à un autre, et à restaurer ce contenu ensuite lorsque le processus reprend la main ;
- sa pile d'exécution : qui est l'espace mémoire dans lequel sont stockés notamment les variables locales, les paramètres passés aux fonctions, ...

Les processus que l'on va implanter partageront leur espace mémoire de données : il s'agit donc de processus légers (ou *threads*), par opposition à des processus lourds qui seraient totalement isolés dans des espaces d'adressage différents.

Rôle de l'ordonnanceur

Dans un noyau de système, l'ordonnanceur de processus est le composant qui gère l'entrelacement de l'exécution des différents processus. Il contient un certain nombre de variables d'état importantes, et notamment la table des processus : dans le noyau que l'on va implanter, il s'agira simplement d'un tableau de taille définie statiquement par une constante et contenant les structures de données représentant les processus activables. Il est aussi nécessaire de gérer un pointeur sur le processus élu (celui en cours d'exécution) pour accéder à sa structure interne si besoin est.
La partie principale de l'ordonnanceur est évidemment la fonction d'ordonnancement, c'est à dire la primitive qui implante la politique d'ordonnancement. Cette fonction est appelée à chaque fois qu'un

processus a épuisé le temps d'exécution qui lui était alloué, et elle se base sur la politique d'ordonnancement pour choisir quel processus va maintenant s'exécuter. Dans ce TP, on commencera par implanter la politique d'ordonnancement la plus simple, celle du tourniquet (*Round-Robin*) : si par exemple le système gère 4 processus de `pid` 0, 1, 2 et 3, alors l'ordonnanceur les fera s'exécuter dans l'ordre suivant : 0, 1, 2, 3, 0, 1, 2, 3, ...
Une fois que la fonction d'ordonnancement a choisi quel processus sera activé, et qu'elle a mis a jour ses variables internes en conséquence, elle doit provoquer l'arrêt de l'exécution du processus actif et démarrer le processus élu. Cette opération critique est réalisée par une primitive appelée traditionnellement changement de contexte (*context switch*).

Le changement de contexte d'exécution

Lorsqu'on veut arrêter l'exécution d'un processus pour donner la main à un autre, il faut sauvegarder le contexte d'exécution du processus actif, de façon à pouvoir reprendre son exécution exactement là où il en était lorsqu'il prendra à nouveau la main. En pratique, cela consiste à sauvegarder les valeurs des registres importants du processeur dans une zone mémoire réservée pour cela (dans la structure du processus telle que présentée plus haut).
Une fois le contexte du processus actif sauvegardé, on doit restaurer le contexte du processus élu (celui qu'on veut activer) : en effet, ce processus avait lui-même été arrêté auparavant dans un certain état et on veut reprendre son exécution dans cet état précis (le problème de la première exécution du processus sera expliqué plus bas).
Parmi tous les registres à restaurer, le pointeur de pile (`%esp` sur x86) est certainement le plus important : en effet, c'est ce registre qui contient l'adresse de la pile d'exécution du processus élu. Cette pile contient elle-même non-seulement les valeurs des variables locales, paramètres, etc. du processus, mais aussi et surtout, l'adresse de retour à la fonction qui était en cours d'exécution lorsque le processus a été arrêté. Une fois le pointeur de pile restauré à la valeur sauvegardé, il suffit donc de terminer la fonction de changement de contexte par l'instruction `ret` habituelle pour que le fil d'exécution redémarre dans le processus élu (remarque : ce point peut vous paraitre difficile à comprendre en lisant cette explication formelle, mais il s'éclaircira lorsque vous le mettrez en pratique sur un exemple concret).
Comme noté plus haut, la première exécution d'un processus pose un problème particulier. En effet, si le processus à activer n'a jamais été exécuté, son contexte n'a pas été sauvegardé (plus qu'il n'en a pas encore). Mais il faut tout de même s'assurer qu'après l'exécution de l'instruction `ret` du *context switch*, l'exécution démarrera bien au début de la fonction principale du processus. Une façon simple de garantir cela est de préparer la pile d'exécution du processus en stockant en sommet de pile l'adresse de la fonction principale du processus, et de placer l'adresse du sommet de pile dans la zone mémoire correspondant à la sauvegarde du registre `%esp`.

Exemple avec deux processus

Pour commencer, on va travailler avec 2 processus qui vont se passer la main explicitement (dans les parties suivantes, le changement de processus sera déclenché par l'interruption de l'horloge, comme dans les systèmes classiques). Pour garantir qu'il n'y a pas d'interférence (et pour permettre le blocage du système), vous devez désactiver les interruptions. Le plus simple pour cela est d'enlever l'appel à la fonction `sti()` que vous aviez ajouté à la fin de votre fonction `kernel_start`.
Pour commencer, on va mettre au point le mécanisme de passage d'un processus à l'autre. Dans le système réalisé, il y aura donc seulement 2 processus :

- le processus `idle` (qui a par convention le `pid` 0) est le processus par défaut : en effet, dans un système, il doit toujours y avoir au moins un processus actif, car le système ne doit jamais s'arrêter ;
- le processus `proc1` (de `pid` 1) qui représentera un processus de calcul quelconque.

Passage d'un processus à l'autre

Dans une première étape, on va implanter simplement le passage du processus `idle` au processus `proc1`
Plus précisément, le code de ces processus sera :

```
void idle(void)
```

```
{
    printf("[idle] je tente de passer la main a proc1...\n");
    ctx_sw(..., ...);
}

void proc1(void)
{
    printf("[proc1] idle m'a donne la main\n");
    printf("[proc1] j'arrete le systeme\n");
    hlt();
}
```

Vous devrez donc voir apparaitre à l'écran la trace de `idle` suivi des deux traces de `proc1`, puis le système se bloquera grâce au `hlt()`.
Pour implanter ce petit test, vous aurez besoin de définir la structure des processus décrite plus haut. Précisément, le type structure de processus que vous devez définir comprendra :

- le `pid` du processus, sous la forme d'un entier (signé, car traditionnellement la fonction de création d'un processus renvoie -1 en cas d'erreur) ;
- le nom du processus : un chaine de caractères avec une taille maximum fixée ;
- l'état du processus : dans ce premier exemple, il n'y a que 2 états possibles : élu ou activable, vous pouvez toutefois déjà définir proprement les états comme une énumération de constantes qui sera facile à étendre par la suite ;
- la zone de sauvegarde des registres du processeur : il s'agira tout simplement d'un tableau d'entiers, puisqu'on manipule des registres 32 bits, et vous avez besoin de 5 cases car il n'y a que 5 registres importants à sauvegarder sur x86 (voir plus bas l'explication du changement de contexte) ;
- la pile d'exécution du processus, qui sera définie comme un tableau d'entiers d'une taille fixée par une constante, par exemple 512.

Ensuite, il faut définir la table des processus, c'est à dire tout simplement un tableau de structure de processus, dont la taille est là encore une constante prédéfinie dans le système. Dans cette première étape, cette constante sera fixée à 2.
Lors du démarrage du système, il faut initialiser les structures de processus de `idle` et `proc1` avec des valeurs pertinentes, c'est à dire :

- pour les deux processus, son `pid`, son nom et son état : par défaut, c'est le processus `idle` qui est élu le premier ;
- pour `proc1` la case de la zone de sauvegarde des registres correspondant à `%esp` doit pointer sur le sommet de pile, et la case en sommet de pile doit contenir l'adresse de la fonction `proc1` : c'est nécessaire comme expliqué plus haut pour gérer la première exécution de `proc1`.

Il n'est pas nécessaire d'initialiser la pile d'exécution du processus `idle` : en fait, ce processus n'utilisera pas la pile allouée dans sa structure de processus mais directement la pile du noyau (celle qui est utilisée par la fonction `kernel_start` notamment). De même, il n'est pas nécessaire d'initialiser la zone de sauvegarde de `%esp` pour `idle` puisque ce processus sera exécuté directement par la fonction `kernel_start`, qui devrait donc ressembler au squelette ci-dessous :

```
void kernel_start(void)
{
    // initialisation des structures de processus
    ...
    // demarrage du processus par defaut
    idle();
}
```

On vous fourni l'implantation du changement de contexte dans le fichier `ctx_sw.S`. Cette fonction est cruciale dans le mécanisme d'ordonnancement car c'est elle qui va effectuer le passage d'un processus à l'autre. On détaille son code ci-dessous :

```
    .globl ctx_sw
# Structure de la pile en entree :
#   %esp + 4 : adresse de l'ancien contexte
```

```
#   %esp + 8 : adresse du nouveau contexte
ctx_sw:
    # sauvegarde du contexte de l'ancien processus
    movl 4(%esp), %eax
    movl %ebx, (%eax)
    movl %esp, 4(%eax)
    movl %ebp, 8(%eax)
    movl %esi, 12(%eax)
    movl %edi, 16(%eax)
    # restauration du contexte du nouveau processus
    movl 8(%esp), %eax
    movl (%eax), %ebx
    movl 4(%eax), %esp
    movl 8(%eax), %ebp
    movl 12(%eax), %esi
    movl 16(%eax), %edi
    # on passe la main au nouveau processus
    ret
```

Le changement de contexte est une fonction particulière : on note déjà qu'il n'y a pas de mise en place d'un contexte d'exécution (ce qui serait contre-productif puisque la fonction doit justement manipuler les contextes de l'appelant et de celui à qui elle va passer la main).
Elle prend 2 paramètres de types pointeurs sur des entiers : il s'agit en fait des adresses des champs `regs` des contextes de l'ancien processus et du nouveau, c'est à dire des zones où on sauvegarde les registres du processeur au moment où on change de processus actif.
Le principe de la fonction est simple :

1. on commence par sauvegarder les registres dans la zone appropriée du contexte de l'ancien processus (celui qu'on veut arrêter) ;
2. on restaure ensuite les registres depuis la zone de sauvegarde du nouveau processus (celui qu'on veut re-démarrer) ;
3. on passe enfin la main au nouveau processus avec l'instruction `ret`.

On peut maintenant comprendre plus précisément le fonctionnement de l'initialisation de la pile de `proc1` que l'on a fait plus haut : lors de la restauration des registres de `proc1`, la valeur de `%esp` restaurée sera un pointeur sur le sommet de la pile de `proc1`, qui contient l'adresse de la fonction `proc1`. Lorsqu'on arrivera sur le `ret` à la fin du changement de contexte, c'est donc cette adresse qui sera dépilée et à laquelle le flot d'exécution sautera.
On remarque qu'on ne sauvegarde pas tous les registres du processus : en effet, il est inutile de sauvegarder les registres `%eax`, `%ecx` et `%edx` dans lequel le compilateur C n'a pas le droit de laisser des valeurs importantes lorsqu'il fait un appel de fonction. Plus tard, lorsqu'on branchement l'ordonnancement sur l'interruption horloge, on devrait formellement sauvegarder ces registres (puisqu'une interruption peut intervenir n'importe-quand), mais on le fait en fait déjà dans le traitant de l'interruption : inutile donc de le refaire dans le changement de contexte. On rappelle enfin que le registre des indicateurs `%eflags` est lui automatiquement sauvegardé par le processeur avant l'appel au traitant d'interruption.

Aller-retour entre les processus

Une fois que le test précédent fonctionne correctement, on va l'étendre pour faire revenir l'exécution à `idle` après être passé par `proc1` (et pour se convaincre que cela marche vraiment, on va faire l'aller-retour 3 fois...). Le code des processus sera donc maintenant :

```
void idle(void)
{
    for (int i = 0; i < 3; i++) {
        printf("[idle] je tente de passer la main a proc1...\n");
        ctx_sw(..., ...);
        printf("[idle] proc1 m'a redonne la main\n");
```

```
    }
    printf("[idle] je bloque le systeme\n");
    hlt();
}

void proc1(void)
{
    for (;;) {
        printf("[proc1] idle m'a donne la main\n");
        printf("[proc1] je tente de lui la redonner...\n");
        ctx_sw(..., ...);
    }
}
```

Ordonnancement selon l'algorithme du tourniquet

On va maintenant implanter un mécanisme d'ordonnancement simple connu sous le nom d'ordonnancement collaboratif : ce sont les processus eux-mêmes qui vont explicitement se passer la main en appelant la fonction d'ordonnancement. On continue à travailler avec les processus `idle` et `proc1` dont on donne le nouveau code ci-dessous :

```
void idle(void)
{
    for (;;) {
        printf("[%s] pid = %i\n", mon_nom(), mon_pid());
        for (int32_t i = 0; i < 100000000; i++);
        ordonnance();
    }
}

void proc1(void) {
    for (;;) {
        printf("[%s] pid = %i\n", mon_nom(), mon_pid());
        for (int32_t i = 0; i < 100000000; i++);
        ordonnance();
    }
}
```

On devra voir s'afficher à l'écran alternativement la trace de chaque processus. Les boucles `for` servent simplement à ralentir l'affichage à l'écran pour le rendre lisible (adaptez la valeur de la constante à la vitesse de votre machine).
Vous devez compléter votre gestion de processus pour implanter ce test :

- la fonction `void ordonnance(void)` a pour rôle d'implanter la politique d'ordonnancement en choisissant le prochain processus à activer (comme il n'y en a que 2 pour l'instant, ça ne devrait pas poser de difficulté) et de provoquer le changement de processus en appelant la fonction `ctx_sw` avec les bons paramètres ;
- l'ordonnanceur a besoin de savoir quel est le processus en cours d'exécution : le plus simple pour cela est de conserver un pointeur vers la structure de processus sous la forme d'une variable globale `actif` (de type pointeur sur une structure de processus par exemple ou de façon équivalente, le `pid` du processus actif sous forme d'un entier signé) ;
- n'oubliez pas de changer l'état des processus dans la fonction `ordonnance` : le processus élu doit prendre l'état `ELU` et l'autre devenir `ACTIVABLE` ;
- les fonctions `int32_t mon_pid(void)` et `char *mon_nom(void)` renvoient simplement le `pid` et le nom du processus en cours d'exécution, elles ne posent pas de difficulté d'implantation.