Architecture Interne des Systèmes d'Exploitation (AISE)

Master CHPS

Jean-Baptiste BESNARD jbbesnard@paratools.fr

Organisation du Cours

Cour MATIN, TD après-midi

- → 10/01: Généralités sur les OS et Utilisation de base
- **→** 15/01 : La Chaine de Compilation et l'exécution d'un programme
- **→** 16/01 : Les I/Os POSIX et Introduction aux Sockets
- 01/02 : Méthodes de communication Inter-Processus
 - → 06/02 : Programmation et reverse et Q/A projets (journée TD)
 - →14/02: Mémoire avancée (mmap, madvise, pages, TLB, ...)
 - → 13/02 : TD Déboggage (gdb, valgrind) && TD mesure du temps et profilage (perf, kcachegrind) et Q/A projets (journée TD)
 - →Un examen final (25 Mars matin)

Généralités sur les IPC System V

Apparus dans Unix en 1983 ils permettent des communication inter-inter-processus (Inter-Process Communications, IPC)

- Files de messages
- Segment de mémoire partagée
- Sémaphores

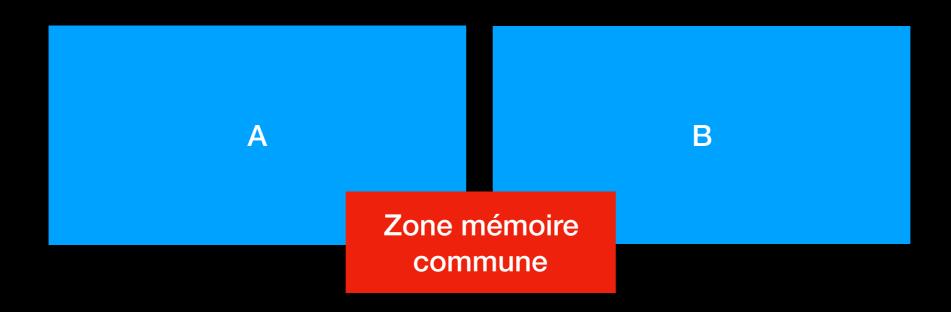
Le noyau est chargé de la gestion des ressources associées via des commandes

Files de Messages



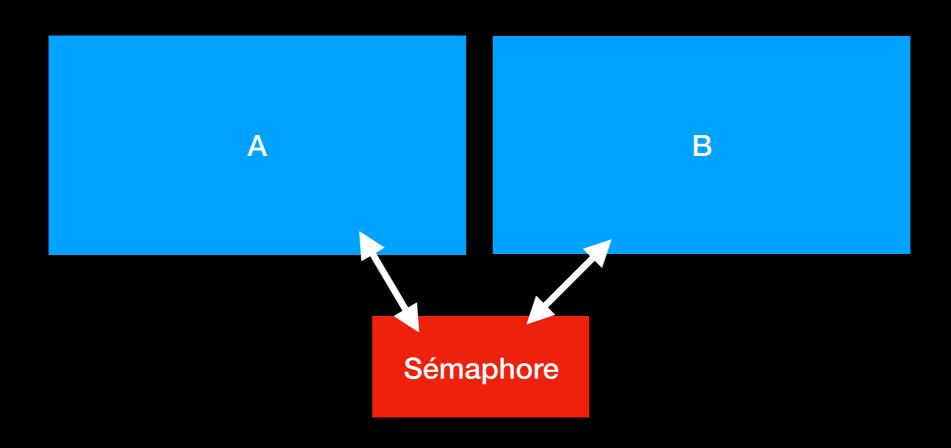
- ftok: génération d'une clef IPC
- · msgget: Récupère un identificateur de file de message
- msgrecv: Réception d'un message depuis une file
- msgsend: Envoi d'un message dans une file
- msgct/: Contrôle de la file de messages

Segment de mémoire partagée



- ftok: génération d'une clef IPC
- shmget: Récupère un identificateur de segment shm
- shmat: Projection d'un segment SHM
- shmdt: Supression d'un segment she
- shmct/: Contrôle du segment SHM

Sémaphore IPC



- ftok: génération d'une clef IPC
- · semget: Récupère un identificateur de sémaphore
- semop: Fait une opération sur le sémaphore
- semct/: Contrôle du sémaphore

```
$ ipcs
----- Files de messages -----
clef
        msqid propriétaire perms
                                     octets utilisés messages
----- Segment de mémoire partagée -----
        shmid propriétaire perms
                                              nattch
                                                       états
clef
                                     octets
0x00000000 42729472 jbbesnard 600 1048576
                                                      dest
0x00000000 39616513 jbbesnard 600
                                    524288
                                                      dest
----- Tableaux de sémaphores -----
clef semid propriétaire perms
                                     nsems
```

```
$ ipcrm -h
Utilisation:
 ipcrm [options]
 ipcrm shm|msq|sem <id>...
Supprimer certaines ressources IPC.
Options:
 -m, --shmem-id <ident.>
                            retirer le segment de mémoire partagée par ident.
 -M, --shmem-key <clef>
                            retirer le segment de mémoire partagée par clef
 -q, --queue-id <ident.>
                            retirer la file de messages par identifiant
 -Q, --queue-key <clef>
                            retirer la file de messages par clef
 -s, --semaphore-id <id.>
                            retirer le sémaphore par identifiant
 -s, --semaphore-key <clef> retirer le sémaphore par clef
                            tout retirer (dans la catégorie indiquée)
 -a, --all[=shm|msg|sem]
 -v, --verbose
                            expliquer les actions en cours
 -h, --help afficher cette aide et quitter
 -V, --version afficher les informations de version et quitter
Consultez ipcrm(1) pour obtenir des précisions complémentaires.
```

```
$ ipcmk -h
Utilisation:
 ipcmk [options]
Créer diverses ressources IPC.
Options:
 -M, --shmem <taille>
                          créer un segment de mémoire partagée de taille <taille>
 -S, --semaphore <nsems>
                         créer un tableau de sémaphores à <nsems> éléments
 -Q, --queue
                          créer une file de messages
                          droits de la ressource (0644 par défaut)
 -p, --mode <mode>
               afficher cette aide et quitter
 -h, --help
 -V, --version afficher les informations de version et quitter
Consultez ipcmk(1) pour obtenir des précisions complémentaires.
```

Resources

Les resources IPC sont indépendante des processus

- Il est possible de laisser des scories si l'on ne fait pas attention
- Un processus peut se « ratacher » à un segment lors de son redémarrage par exemple
- Les processus partagent des segments avec un mécanisme de clef qui est un secret « a priori » pour la sécurité

Clefs pour les IPCs System V

La Clef

Un IPC (de tout type) est partagé par une clef:

- C'est un entier qui doit être le même entre tous les processus partageant la resource;
- On peut la connaitre a priori avec risque de conflit (un peut comme un port TCP);
- Une clef spéciale IPC_PRIVATE crée une file limité à un processus et l'ensemble de ses descendants;
- On peut la créer avec une fonction « ftok » qui repose sur un fichier et un nom de projet.

Ftok

SYNOPSIS

#include <sys/types.h>
#include <sys/ipc.h>

key_t ftok(const char *pathname, int proj_id);

DESCRIPTION

The ftok() function uses the identity of the file named by the given pathname (which must refer to an existing, accessible file) and the least significant 8 bits of proj_id (which must be nonzero) to generate a key_t type System V IPC key, suitable for use with msgget(2), semget(2), or shmget(2).

The resulting value is the same for all pathnames that name the same file, when the same value of proj_id is used. The value returned should be different when the (simultaneously existing) files or the project IDs differ.

RETURN VALUE

On success, the generated key_t value is returned. On failure -1 is returned, with errno indicating the error as for the stat(2) system call.

	Projet	Device		Inode	
31	24	23	16 15		0

Création / Récupération de ressources

Une fois que l'on a une clef de type *key_t* on peut retrouver/créer une resource:

- File de message : *msgget*
- Segment de mémoire partagée: shmget
- Sémaphore: semget

Les Files de Messages IPC SYSTEM V

Files de Messages pour une Communication entre Processus sur un Même Noeud.

Le message sera toujours de la forme:

```
Struct XXX {
    long id; // Toujours > 0 !
    ... DATA ....
    // Taille max sans le long MSGMAX (8192 Octets)
};
```

Lors de l'envoi et de la réception d'un message la taille et TOUJOURS sans le long qui définit le type de message. Cette même valeur (ici id) doit TOUJOURS être supérieure à 0.

En pratique on crée une struct statique sur la pile car l'allocation d'un objet avec piggybacking demande plus de code.

Créer Une File de Messages

```
#include <sys/types.h>
#include <sys/ipc.h>
#include <sys/msg.h>
int msgget(key_t key, int msgflg);
```

- Key: Une clef, soit manuelle, soit via ftok ou bien IPC_PRIVATE
- msgflg: mode de création de la file et ses droits UNIX
 - → IPC_CREAT crée une file s'il y en a aucune associée à cette clef
 - →IPC_EXCL échoue s'il existe déjà une file sur la clef indiqué (toujours combiné avec IPC_CREAT!)
 - → **0**600 droit UNIX en octal (important car si omis **0**000 et la file et moins pratique!)

Créer Une File de Messages

- Créer une file pour un processus et ses fils
 - → file = msgget(IPC_PRIVATE, 0600);
- Créer une file pour accéder à une file potentiellement existante:
 - → file = msgget(key , IPC_CREAT | 0600);
- Pour être sûr de créer une nouvelle file en lecture écriture pour soi et en lecture seule pour les autres utilisateurs:
 - → file = msgget(key, IPC_CREAT | IPC_EXCL | 0622);
- Utiliser uniquement une file existante précédemment créée par un serveur:
 - → file = msgget(key, 0);

Envoyer un Message

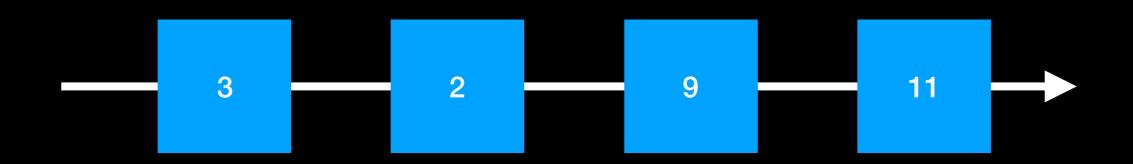
```
#include <sys/types.h>
#include <sys/ipc.h>
#include <sys/msg.h>
int msgsnd(int msqid, const void *msgp, size_t msgsz, int msgflg);
```

- msqid : file de message à utliser, créée avec msgget
- msgp : pointeur vers les donnée à envoyer (comprend forcément un long qui est l'ID de message)
- size : taille du message **SANS** le long qui est l'ID du message
- msgflg: mode d'envoi du message
 - ⇒ IPC_NOWAIT ne pas bloquer si la file est pleine (renvoie EAGAIN dans errno)
 - → 0 en général

- msqid : file de message à utliser, créée avec msgget
- msgp: pointeur vers les donnée à envoyer (comprend forcément un long qui est l'ID de message)
- size : taille du message SANS le long qui est l'ID du message
- msgtyp: type de message à recevoir:

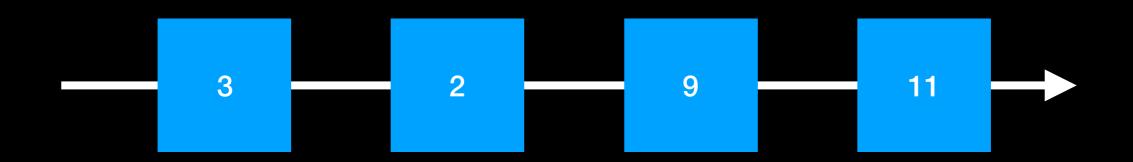
#include <sys/types.h>

- → 0 : prochain message de la file
- → 0 < TYP prochain message avec l'ID donné
- TYP < 0 prochain message avec un ID inférieur ou égal à TYP, utilisé pour gérer des priorité de messages
- msgflg: mode de réception du message:
 - → IPC_NOWAIT ne pas bloquer si pas de message du TYP demain (renvoie ENOMSG dans errno
 - → MSG_EXCEPT renvoie un message d'un TYP différent de celui donné (seulement pour TYP > 0)
 - → MSG_NO_ERROR permettre au message d'être tronqués à la réception (à la différence du comportement de base)



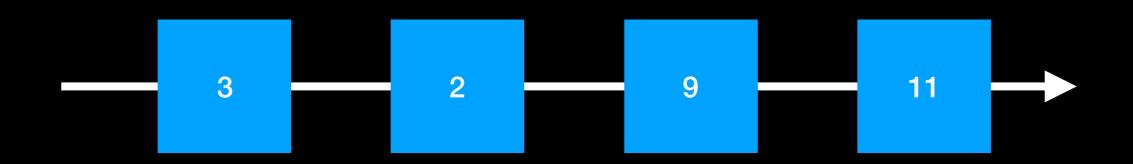
msgrcv(file, &msg, sizeof(int), 2, 0);

Quel message ??



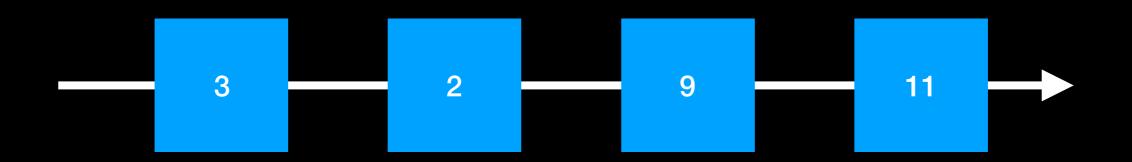
msgrcv(file, &msg, sizeof(int), 2, 0);

Quel message??



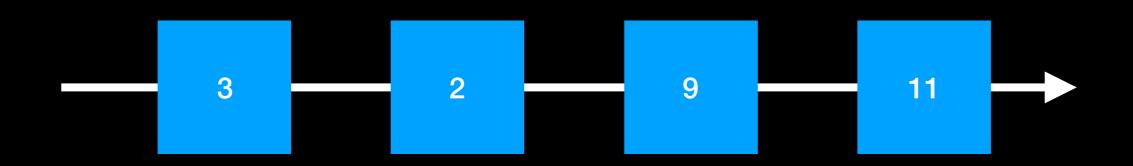
msgrcv(file, &msg, sizeof(int), -10, 0);

Quel message ??



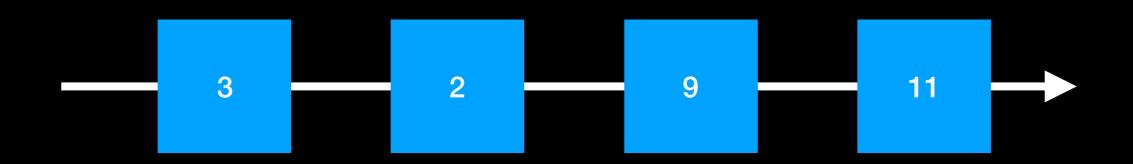
msgrcv(file, &msg, sizeof(int), -10, 0);

Quel message??



msgrcv(file, &msg, sizeof(int), 99, 0);

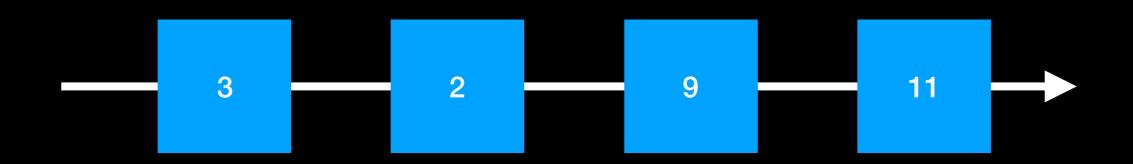
Quel message ??



msgrcv(file, &msg, sizeof(int), 99, 0);

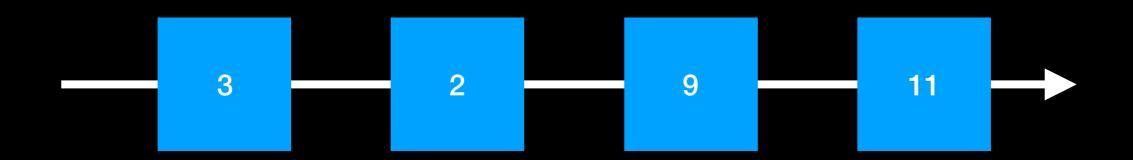
Quel message??

L'appel reste bloqué indéfiniment si un message 99 n'est jamais posté.



msgrcv(file, &msg, sizeof(int), 99, IPC_NOWAIT);

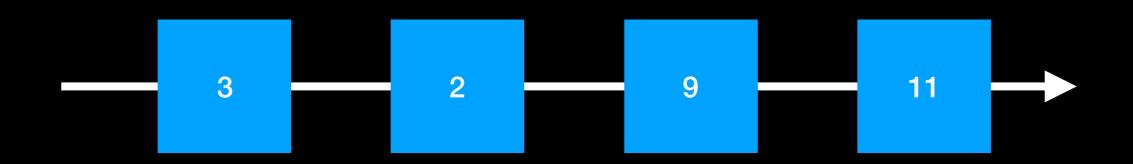
Quel message ??



msgrcv(file, &msg, sizeof(int), 99, IPC_NOWAIT);

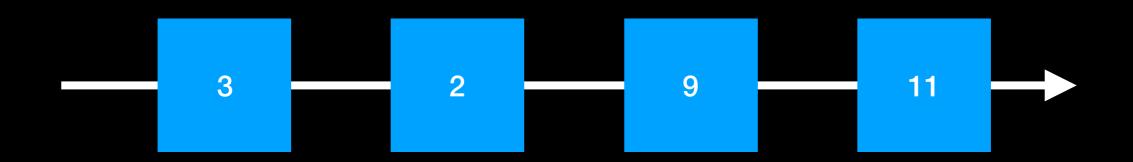
Quel message??

L'appel renvoie -1 et met errno à ENOMSG



msgrcv(file, &msg, sizeof(int), 11, MSG_EXCEPT);

Quel message ??



msgrcv(file, &msg, sizeof(int), 11, MSG_EXCEPT);

Quel message??

Contrôler une File

```
#include <sys/types.h>
#include <sys/ipc.h>
#include <sys/msg.h>
int msgctl(int msqid, int cmd, struct msqid_ds *buf);
```

- msqid : ID de la file à contrôler
- cmd: commande à appliquer à la file
 - → IPC_STAT récupères les informations sur la file dans la struct msqid_ds (voir man)
 - →IPC_SET permet de régler certains attributs en passant une struct msqid_ds
 - ➡ IPC_RMID supprime la file toute les opérations courantes ou future échouent (avec la possibilité non gérée qu'une nouvelle file soit créée avec la même clef). La synchronisation et à la charge du programmeur.

PENSEZ à SUPPRIMER VOS FILES !!!

```
#include <sys/ipc.h>
#include <sys/msq.h>
#include <sys/wait.h>
double get time(){
    struct timeval val;
     gettimeofday(&val, NULL);
     return (double) val.tv sec + 1e-6 * val.tv usec;
#define SIZE 16
struct msg t{
    long type;
    int data[SIZE];
};
#define NUM MSG 65536
int main( int argc, char ** argv ){
    int file = MSqqet(IPC PRIVATE, IPC CREAT | 0600);
    if( file < 0 ){
          perror("msgget");
          return 1;
    int i;
    struct msg_t m;
    m.type = 1;
    int pid = fork();
    if( pid == 0 )
          int stop = 0;
          while(!stop)
              msgrcv(file, &m, size*sizeof(int), 0, 0);
               /* Notify end */
               if( m.data[0] == 0 )
                    stop = 1;
              msgsnd(file, &m, SIZE*sizeof(int), 0);
```

Suite ...

```
else
    double total time = 0.0;
    for( i = 1 ; i <= NUM MSG ; i++)</pre>
         m.data[0] = i;
         double start = get time();
         int ret = msgsnd(file, &m, SIZE*sizeof(int), 0);
         if( ret < 0 )
             perror("msgsend");
             return 1;
         double end = get time();
         total time += end - start;
         msgrcv(file, &m, size*sizeof(int), 0, 0);
    m.data[0] = 0;
    msgsnd(file, &m, SIZE*sizeof(int), 0);
    wait( NULL );
     msgctl( file, IPC_RMID, NULL);
    fprintf(stderr, "Pingpong takes %g usec Bandwidth is %g MB/s
                     total time/NUM MSG*1e6,
                     (double)(SIZE*NUM MSG*sizeof(int))/
                              (total time*1024.0*1024.0));
return 0;
```

Les Segments SHM IPC SYSTEM V

Partager une Zone Mémoire entre Deux Processus

SHM = SHared Memory

Les avantages:

- Communication directe sans recopie mémoire;
- Pas de passage par l'espace noyau à la différence des files messages (context switch et recopie);
- Latences plus faible (même mémoire)

Les inconvénients:

- Il faut manuellement synchroniser les communications (lock ou sémaphore)
 - → Comprenez qu'il est possible de mettre un lock dans cette zone mémoire, un spin lock directement, un mutex avec le bon attribut (PTHREAD_PROCESS_SHARED). Ou bien un sémaphore des IPC.
- La structuration des données est à la charge du programme

Créer le Segment SHM

```
#include <sys/ipc.h>
#include <sys/shm.h>
int shmget(key_t key, size_t size, int shmflg);
```

- key: Une clef, soit manuelle, soit via ftok ou bien IPC_PRIVATE
- Size: taille su segment SHM en octet (arrondie à la page supérieure).
 Donc mapper un int est un gros gâchis de mémoire (une page fait 4 KB).
- shmflg: mode de création de la file et ses droits UNIX
 - → IPC_CREAT crée une file s'il y en a aucune associée à cette clef
 - →IPC_EXCL échoue s'il existe déjà une file sur la clef indiqué (toujours combiné avec IPC_CREAT!)
 - → 0600 droit UNIX en octal (important car si omis 0000 et la file et moins pratique!)

Projeter le Segment SHM

```
#include <sys/types.h>
#include <sys/shm.h>

void *shmat(int shmid, const void *shmaddr, int shmflg);
```

- shmid : le descripteur du segment SHM
- shmaddr: une addressee où mapper le segment, alignée sur une frontière de page. NULL si indifférent.
- shmflg: options relative à la projection du segment
 - → SHM_RND arrondis l'adresse passée par *shmaddr* à une frontière de page
 - →SHM_RDONLY partager le segment en lecture seule
 - → ... il existe d'autre flags voir man

Retirer le Segment SHM

```
#include <sys/types.h>
#include <sys/shm.h>
int shdt(const void *shmaddr);
```

• shmaddr: adresse <u>renvoyée par shmat</u>

Tous les processus doivent retirer le segment de leur mémoire autrement la suppression avec shmctl n'est pas effective. Si un processus se termine il détache la mémoire mais cela ne marque pas le segment pour suppression.

Supprimer le Segment SHM

```
#include <sys/ipc.h>
#include <sys/shm.h>
int shmctl(int shmid, int cmd, struct shmid_ds *buf);
```

- shmid : ID du segment à contrôler
- cmd: commande à appliquer à la file
 - → IPC_STAT récupères les informations sur la file dans la struct shmid_ds (voir man)
 - →IPC_SET permet de régler certains attributs en passant une struct shmid_ds
 - ➡ IPC_RMID marque le segment SHM pour destruction cela ne se produira que quand tout les processus l'ayant projeté se seront détachés
 - ... il existe d'autre flags voir man particulièrement IPC_INFO et SHM_INFO utiles pour connaitre les limites sur le système cible

PENSEZ à SUPPRIMER VOS Segments !!!

Totalement arbitraire

```
#include <sys/ipc.h>
#include <sys/shm.h>
#include <unistd.h>
#include <stdio.h>
int main(int argc, char **argx
    int shm = shmget(19999, 2 * sizeof(int),
                     IPC CREAT | IPC EXCL | 0600 );
    if( shm < 0)
         perror("shmget");
         return 1;
    int *val = (int*) shmat(shm, NULL, 0);
    if( !val )
         perror("shmat");
         return 1;
                                      Serveur
    /* valeur de départ */
    val[0] = 1;
    val[1] = 0;
    while(val[0])
         sleep(1);
         val[1]++;
     /* Unmap segment */
    shmdt(val);
    /* Server marks the segment for deletion */
    shmctl(shm, IPC RMID, NULL);
    return 0;
```

```
#include <sys/ipc.h>
#include <unistd.h>
#include <stdio.h>
int main(int argc, char **argv)
     int shm = shmget(19999, 2 * sizeof(int), 0 );
     if(shm < 0)
         perror("shmget");
         return 1;
     int *val = (int*) shmat(shm, NULL, 0);
     if( !val )
                                    Client
         perror("shmat");
          return 1;
     /* valeur de départ */
     int last val = -1;
    while(1)
         if( val[1] != last val ){
               printf("Val is %d max is 60\n", val[1]);
               last_val = val[1];
               /* Stop condition */
               if( 60 <= val[1] )
                    val[0] = 0;
                   break;
         else
               usleep(100);
     /* Unmap segment */
     shmdt(val);
     return 0;
```

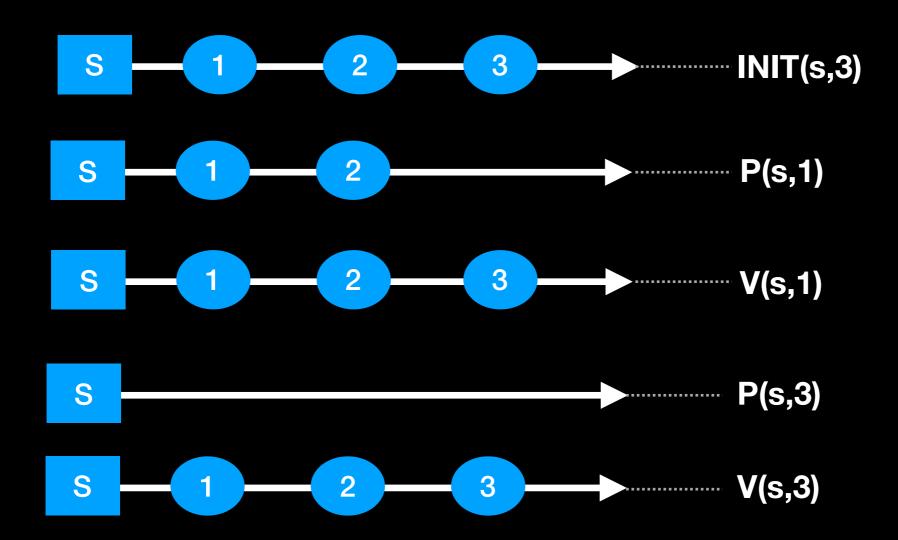
```
$ ./serveur &
$ ipcs -m
----- Segment de mémoire partagée -----
          shmid
clef
                  propriétaire perms
                                                                états
                                           octets
                                                     nattch
0x00004e1f 42827778
                    jbbesnard 600
$ ./client
Val is 0 max is 60
Val is 1 max is 60
(...)
Val is 7 max is 60
Val is 8 max is 60
Val is 60 max is 60
[2]+ Fini
                            ./server
$ ipcs -m
----- Segment de mémoire partagée -----
                    propriétaire perms
clef
          shmid
                                                                états
                                           octets
                                                     nattch
```

Les Sémaphores IPC SYSTEM V

Notion de Sémaphore

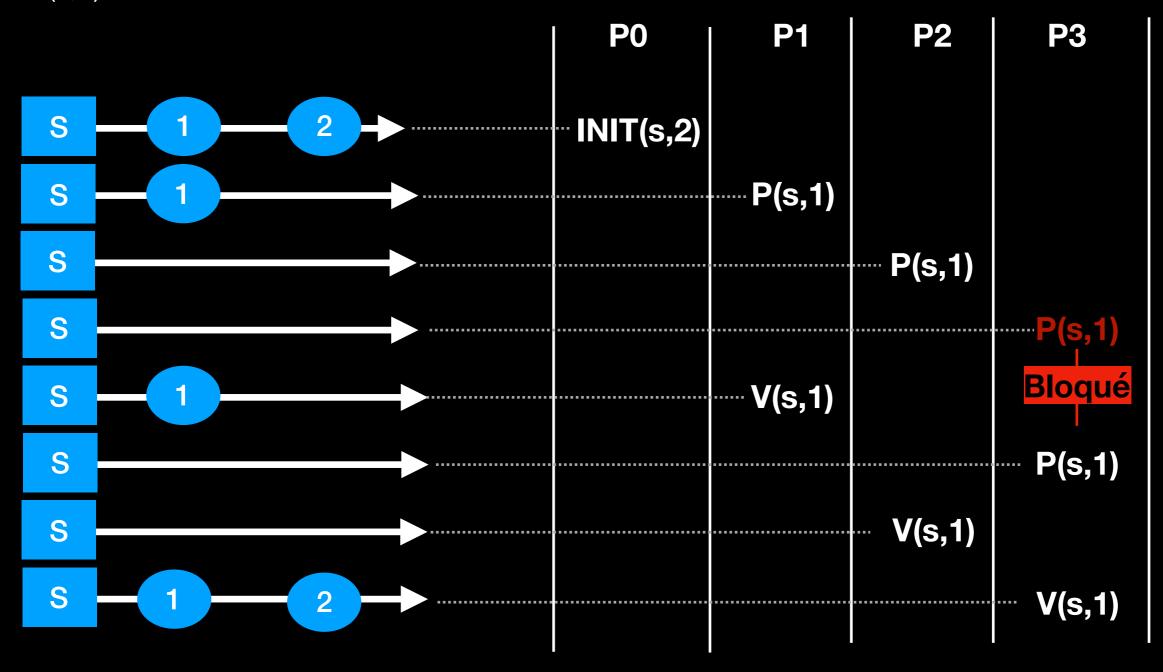
Un sémaphore est un élément de synchronisation qui permet de partager un ensemble de resources. Il existe des sémaphores pour la programmation en mémoire partagée. Ici les sémaphore System V sont inter-processus. On définit classiquement deux opérations:

- P(s,n) : « Tester » (de l'allemand passering du fait de Dijkstra)
- V(s,n) : « Relâcher » (de l'allemand vrijgave du fait de Dijkstra)



Synchronisation avec des Sémaphores

- P(s,n): « Tester »
- V(s,n): « Relâcher »



Créer des Sémaphores

```
#include <sys/types.h>
#include <sys/ipc.h>
#include <sys/sem.h>
int semget(key_t key, int nsems, int semflg);
```

- key: Une clef, soit manuelle, soit via ftok ou bien IPC_PRIVATE
- nsem: nombre de sémaphores à créer
- shmflg: mode de création de la file et ses droits UNIX
 - → IPC_CREAT crée une file s'il y en a aucune associée à cette clef
 - ⇒IPC_EXCL échoue s'il existe déjà une file sur la clef indiqué (toujours combiné avec IPC_CREAT!)
 - → 0600 droit UNIX en octal (important car si omis 0000 et la file et moins pratique!)

Opération sur des Sémaphores

```
#include <sys/types.h>
#include <sys/ipc.h>
#include <sys/sem.h>
int semop(int semid, struct sembuf *sops, size_t nsops);
• semid : identifiant du sémaphore
• sembuf: opération(s) à effectuer via un tableau
           struct sembuf {
               unsigned short sem num; /* semaphore number */
                          sem op; /* semaphore operation */
               short
                          sem flg; /* operation flags */
               short
           };
         ⇒ sem_num: numéro du sémaphore
         ⇒ sem_op: opération à effectuer
             Sem op > 0 : V(s)
             Sem op < 0 : P(s)</p>
             > sem op == 0 : attente de la valeur 0 -> utile pour synchroniser les processus
         → Drapeau a utiliser :
```

- ► IPC_NOWAIT: non-bloquant et renvoie EAGAIN si l'opération avait dû bloquer
- ► IPC_UNDO: demande au noyau d'annuler l'opération si le processus se termine en cas d'arrêt intempestif
- nsops: nombre d'opérations à effectuer (elle sont faites de manière atomique)

Contrôle du Sémaphore

- semnum: identifiant du sémaphore
- cmd: commande à appliquer au sémaphore
 - ➡ IPC_STAT récupères les informations sur le sémaphore
 - SETALL définit la valeur du sémaphore (prend un tableau de unsigned short int en paramètre additionnel
 - → IPC_RMID supprime immédiatement le sémaphore et débloque les processus en attente
 - il existe **BEAUCOUP** d'autre flags voir man

```
#include <stdio.h>
#include <sys/types.h>
#include <sys/ipc.h>
#include <sys/sem.h>
#include <errno.h>
#include <unistd.h>
#include <sys/wait.h>
int main( int argc, char ** argv ){
     int sem = semget(IPC PRIVATE, 1, IPC CREAT | 0600);
     if( sem < 0 ){
          perror("msgget");
          return 1;
     unsigned short val = 1;
     if( semctl(sem, 0, SETALL, &val) < 0){</pre>
          perror("semctl");
          return 1;
     int pid = fork();
     struct sembuf p;
     p.sem_num = 0;
     p.sem_op = -1;
     p.sem flg = SEM UNDO;
     struct sembuf v;
     v.sem_num = 0;
     v.sem op = 1;
     v.sem flg = SEM UNDO;
     if( pid == 0 ) { /* Child */
          while(1){
               if ( semop(sem, &p, 1) < 0) {
                         printf("Child: SEM deleted\n");
                    return 0;
               printf("CHILD holding the sem\n");
               sleep(1);
               semop(sem, &v, 1);
```

Suite ...

```
else
{
    /* Parent */
    int i = 0;
    while(i < 5)
    {
        semop(sem, &p, 1);

        printf("PARENT holding the sem\n");
        sleep(1);
        semop(sem, &v, 1);
        i++;
    }

    /* Parent delete the sem and unlock the child */
    semctl(sem, 0, IPC_RMID);
    wait( NULL );
}

return 0;</pre>
```

Sortie du Programme

\$./a.out
PARENT holding the sem
CHILD holding the sem
PARENT holding the sem
CHILD holding the sem
PARENT holding the sem
CHILD holding the sem

IPCs POSIX

IPC POSIX

Le standard POSIX plus récent propose également les même mécanismes:

- Files de messages
- Segment de mémoire partagée
- Sémaphores

- Il sont plus fiables en termes de libération et de partage de la resource;
- Enfin l'ensemble de l'interface est thread-safe;
- Les objets sont demandés par nom et non avec une valeur donnée;
- Ces appels sont un peu moins portable et sont à attendre plus sur des LINUX que des UNIX au sens large;
- On les décrit généralement comme plus simples à utiliser.

Files de Message POSIX

À vous de jouer avec le man:

- mq_open
- mq_close
- mq_send
- mq_receive
- mq_unlink

Portez l'exemple SYS-V

Que pensez-vous de mq_notify?

Segment SHM POSIX

À vous de jouer avec le man:

- shm_open
- shm_unlink

Portez l'exemple SYS-V

Sémaphore IPC POSIX

Aussi « sémaphore nommé » à ne pas confondre avec les sémaphore « anonymes » de la NPTL (libpthread) qui sont dans le même header.

Rappel (ou pas) pour un sémaphore «anonyme »:

- sem_init
- sem_destroy

À vous de jouer avec le man pour un sémaphore nommé:

- sem_open
- sem_close
- sem_post
- sem_wait
- sem_unlink

Portez l'exemple SYS-V

Peut-on l'implémenter avec un sémaphore anonyme et pourquoi?

Préférez vous POSIX ou SYS-V?