

Systeme.

McROSS "Beat" DEKOUALITAT

5 octobre 2014

Introduction

Basé sur *Programmation système en C sous Linux* par CHRISTOPHE BLAESS et le cours de CARLOS.
C'est un *WIP*, donc encore (et certainement pour toujours) imparfait.

Chapitre 1

Processus

"[moi], va t'acheter une vie"

une GROSSE...

1.1 La notion de processus

Un processus, c'est une *tâche en cours d'exécution*. C'est en même temps du code et des données. Pour généraliser grossièrement, et pas de façon absolument vraie, un processus peut être représenté comme étant un programme en cours de fonctionnement.

1.2 Création d'un processus

Un processus est créé par le biais de l'appel système *fork()*. *fork()* fait la copie du processus père en un nouveau processus. Ce nouveau processus, qu'on appelle *processus fils*, exécutera le même code que le processus père¹ et *héritera* des données du père (variables par exemple) mais aussi de l'environnement du processus père. Si, par exemple, le processus père modifie des *variables d'environnement*, le fils héritera de ces variables modifiées. Bien sur, qui dit nouveau processus, dit nouveaux identifiants pour le nouveau processus.

Cependant, et c'est un détail important, les données dupliquées entre le père et le fils ne sont pas immédiatement copiées; c'est à dire que tant que le fils et le père partagent des informations communes, cette information existe en *un seul exemplaire* dans le système. Par contre, si jamais l'un des deux décide de modifier cette information, c'est à ce moment que cette information est dupliquée. C'est la méthode de *copie sur écriture*.

Enfin, dernière note : il faut savoir qu'il n'y **pas** d'ordre d'exécution prédéfini entre un processus fils et un processus père. Cela veut dire que le fils peut prendre la main à tout moment et le père peut en faire de même.

1.3 Les états

Un processus peut avoir plusieurs états (un seul à la fois) :

- **RUNNING** : le processus est actuellement en train de faire son job ;
- **WAITING/SLEEPING** : le processus est en attente de ressources ou d'événements extérieurs ;

1. Bien sur, le code qui se trouve après le *fork*.

- STOPPED : le processus est stoppé par un *signal* et ne reprendra que par un *signal de redémarrage* ;
- TERMINATED : le processus a fini son travail.²

2. Dans le cas où le processus a fini son travail mais que son père n'a pas encore lu son code de retour avec *wait()*, il reste alors présent dans la table des processus tant qu'on a pas lu son code de retour ; on parle d'état ZOMBIE dans ce cas.

Chapitre 2

Exécution de programmes

"je vais me chékou, j'ai les yeux rouges"

une GROSSE...

2.1 Lancement d'un programme

On a vu que la création d'un processus se fait par le biais de *fork*. Pour un programme, on utilise l'appel système *exec()* (ou du moins l'une des fonctions de la famille *exec()*).

Il existe deux fonctions : *execl* et *execlp*. Ces deux fonctions fonctionnent de la manière suivante :

- *execl*("chemin absolu", "nom de l'application", "argument", ..., NULL)
- *execlp*("nom de l'executable", "nom de l'application", "argument", ..., NULL)

La différence entre les deux tient dans le chemin que l'on va indiquer à la fonction : *execl* prend en paramètre le chemin absolu (par exemple /usr/bin/vim) alors que *execlp* prend en paramètre le nom direct de l'application et cherchera directement dans le \$PATH.

Le lancement d'un nouveau programme *remplace totalement l'ancien*, c'est à dire que tout ce qui est *code segment*, *segment mémoire*, etc est réinitialisé pour le nouveau programme. Par contre, il n'y a **pas** de **nouveau processus** créé. Le processus qui lance un programme garde **le même** PID, PPID, etc.

Dans le cas où on cherche à lancer un nouveau programme sans remplacer pour autant le processus en cours, il existe 2 fonctions qu'on ne détaillera pas¹ : *system()* et la paire *popen()/pclose()*. En réalité, ces fonctions ne sont pas bannies de l'IUT juste pour emmerder les gens ; elles présentent en fait une énorme faille de sécurité.²

2.2 Fin d'un programme

Un programme peut mettre fin à son execution soit de manière *normale* (abandon par l'utilisateur, tâche finie) ou soit, vous l'aurez deviné, de manière *anormale* (bug, crash, tout ce que vous voulez).

1. leur utilisation amène à une lapidation par petits cailloux pointus rouillés atteints du sida

2. programme compilé en root, script *rm -r* / déguisé = boum

2.2.1 Arrêt normal

Comme vous le savez, le moyen *propre* d'arrêter un programme se fait par le biais de `exit()` ou `return()`. La différence entre les deux est que `exit()` peut être utilisé partout, alors que `return()` ne quitte le programme que si on l'appelle depuis le *main*.

2.2.2 Arrêt anormal

Lorsque un programme fait une boulette (par exemple, il tente d'accéder au contenu d'un pointeur non initialisé), un *signal* est émis et arrête le programme tout en créant un dump mémoire.³

Mis à part ça, on peut arrêter proprement mais *anormalement* (heh) un programme en utilisant `abort()` et `assert()`. Ces deux fonctions émettent des signaux et permettent un débogage du programme.

2.2.3 Dans le cas d'un processus fils

Imaginons que nous avons un processus fils. Que ce processus se soit arrêté normalement ou pas, son processus père se doit de lire son code de retour, afin que le processus fils n'erre indéfiniment dans les limbes des processus.

Pour cela, on utilise `wait()`. Cette fonction (ou l'une des fonctions de la famille `wait()`) permet au processus père de lire le code de retour d'un proc fils. En fait, le processus père *attend* que l'un des processus fils se termine. `wait()` peut prendre en paramètre l'adresse d'un int afin de connaître les circonstances de la mort du processus fils. On le fait de manière suivante :

```
int status; /* le int qui nous permet d'avoir le code retour */

/* ici on met le code avec le fork() et tout le tintamarre */

wait(&status); /*on attend la mort du fils*/

if (WIFEXITED(status)) /* on vérifie maintenant comment le fils est mort */
    printf("code_%d",WEXITSTATUS(status));
else if ... /* bla bla bla */
```

Voilà. Il existe 3 macros qui permettent de savoir comment le fils est mort :

- WIFEXITED, qui est vraie si le programme s'est quitté tout seul (va avec WEXITSTATUS afin de connaître le code retour)
 - WIFSIGNALED, si le programme a été tué par un *signal* (exemple : SIGKILL aka kill -9). Utiliser WTERMSIG afin de connaître quel signal a tué le processus.
 - WIFSTOPPED, si le programme est temporairement stoppé par un signal (bien souvent SIGSTOP).
- Enfin, notons qu'il faut autant de `wait()` que de processus fils lancés !

3. On verra tout ça dans le prochain chapitre youpi super cool génial [P-REC] Bookmark.

Chapitre 3

Les signaux

"mais mec, y a le mifa chez ouam"

une GROSSE...

3.1 Signal ? c'est un dentifrice non ???????

Un *signal* est une sorte de message envoyé par un processus à un autre processus. Face à un signal, un processus peut effectuer une des actions au choix :

- *Ignorer* le signal, mais attention, ce n'est pas possible pour tous les signaux (par exemple, il est impossible d'ignorer SIGKILL) ;
- *Capturer* le signal, c'est à dire exécuter une procédure particulière quand le programme reçoit un signal particulier ;
- *Laisser le kernel s'en occuper*, c'est à dire laisser l'action par défaut définie par chaque signal (exemple : se terminer anormalement pour SIGINT, ou ne rien faire pour SIGCHILD)

Un signal est émis soit :

- *Par le système*, lorsque il détecte quelque chose (instruction illégale, fin d'un processus fils, ~~rencontre avec ma GROSSE...~~)
- *Par l'utilisateur* lui même, par le biais de commandes (CTRL+C, CTRL+U, etc), par la fonction *kill()*, ou encore avec des fonctions (*raise()*, *kill()*, etc)

3.2 Les principaux signaux

3.2.1 Signaux de terminaison (sigint, sigkill, sigquit, sigterm)

- SIGINT, code 2, appelée aussi avec CTRL+C, correspond à un SIGnal d'INTerruption, donc termine le processus. Peut être ignoré/capturé.
- SIGQUIT, code 3, appelée aussi avec CTRL+\, termine le processus mais crée un fichier core/dump. Peut être ignoré/capturé.
- SIGKILL, code 9, **tue** (pas proprement) le processus à **tout les coups**. À utiliser en dernier recours. Ne peut **pas** être ignoré/capturé.
- SIGTERM, code 15, tue **proprement** le processus. Correspond à l'action par défaut de la fonction *kill()*. Peut être ignoré/capturé.

3.2.2 Signaux de gestion de processus (sigchld, sigstop ou sigstp, sigcont, sigalrm, sigusr1&sigusr2)

- SIGCHLD, code 17, envoyé au processus père lorsque l'un des processus fils s'est terminé. Est ignoré par défaut. Peut être ignoré/capturé.
- SIGSTOP\SIGSTP, code 19 et 20. Stoppe le processus (voir états des processus). La différence entre les deux tient dans le fait que SIGSTOP est pas ignorable ou capturable alors que SIGSTP l'est.
- SIGCONT, code 18, relance le processus si il est stoppé. Ne peut pas être ignoré/capturé.
- SIGALRM, code 14, est envoyé lorsque le temps défini par l'appel système *alarm()* a expiré.
- SIGUSR1&SIGUSR2, code 10 et 12, sont des signaux *personnalisables*. Par défaut, ils terminent le processus, mais on peut les capturer pour faire autre chose avec.

3.2.3 Signaux d'erreur (pas très important)

Par défaut, ces signaux arrêtent le processus.

- SIGBUS & SIGSEGV : adresse mémoire invalide ou erreur de segmentation.
- SIGFPE : erreur de calcul (ex : division par 0).
- SIGXCPU & SIGXFSZ : limite d'utilisation de ressources dépassé.
- SIGILL : instruction assembleur illégale.

3.3 Les fonctions C

Il existe une ribambelle de fonctions C pour émettre/capturer/attendre/~~exploser~~ un ou des signaux.

3.3.1 Émettre un signal

Il existe ici deux fonctions, *kill()* et *raise()*. La différence entre les deux est que *raise()* envoie le signal au processus qui l'appelle, alors que *kill()* permet d'envoyer un signal à n'importe quel processus (suivant les droits bien_entendu).

kill() s'utilise la manière suivante :

```
kill(<pid du proc visé>, <code du signal>);
```

kill() retourne un int suivant sa réussite. Il faut noter qu'on peut utiliser "-1" en tant que pid pour envoyer le signal à tous les processus (a part init, of_course).

Petite note supplémentaire : il existe une fonction *alarm()* qui prend en paramètre un nombre de secondes. Cette fonction permet d'envoyer un SIGALRM au processus en cours (c'est à dire *soi même*) après le nombre de secondes passé en paramètre.

3.3.2 Attendre un signal

La fonction *pause()* permet de mettre le programme en attente d'un signal. Le programme se mettra en pause tant qu'il n'aura pas reçu de signal.

Note : il ne faut **pas** utiliser de *sleep()*, *wait()*, ou autres conneries pour attendre un signal. Il faut **toujours** utiliser *pause()*.¹

pause() s'utilise de la manière suivante :

```
pause( );
```

```
ez non ????????????????
```

1. au risque de se faire taper dessus par CARLOS

3.3.3 Gérer les signaux

Il faut d'abord créer une **procédure** qui prendra en paramètre le signal. Cette procédure s'occupera du traitement du signal.

Notez que cette procédure prend exclusivement **1 paramètre**, un entier. Pour les détails (ou pour bluffer 2/3 pequenots), cet entier représente la constante symbolique du signal. Cet entier sera rempli automatiquement par le système d'exploitation.

Voyons un exemple :

```
void gestSig(int sig)
{
    /* on met nos traitements par signal */
    /* par exemple, si le proc reçoit SIGINT */
    if (sig==SIGINT)
        printf("CTRL-C\recu\n");
}
```

Ici, cette procédure gère le signal SIGINT. Notez qu'on est *pas obligé* de faire le test. On peut, par exemple, créer autant de procédures que de signaux que l'on souhaite traiter, ou ne créer qu'une seule procédure qui sera utilisé par tous les signaux.

Ensuite, il faut utiliser *signal()* ou *sigaction()*. On se penchera ici sur *signal()*.

```
/* on imagine qu'on a notre gestionnaire de signal "gestSig" qui affiche "oui"
 * dès qu'on reçoit un SIGINT */
int main()
{
    signal(SIGINT, gestSig);
    pause();
    /* on reçoit CTRL-C aka SIGINT */
    exit 0;
}
```

Ce qui donnera à l'affichage, lorsque on exécute le programme "prog" au dessus :

```
>>> prog
(on tape CTRL-C)
>>> oui
(fin du prog)
>>>
```

ez aussi non?????

À noter : il existe aussi des constantes pour gérer les signaux, comme SIG_IGN par exemple. On utilise ces constantes de la manière suivante :

```
signal(SIGINT, SIG_IGN);
/* SIGINT sera ignoré */
```

Chapitre 4

Gestion de fichiers

"Sous UNIX, tout est fichier."

les randoms gens sur internet

4.1 Fichier ? Ça fi-chier hehehahhrheuhrzuhrezu

Un fichier possède des attributs qui diffèrent entre les systèmes de fichiers (*ext2, ext4, fat32, ...*). Pour ne pas se compliquer la vie, sachez qu'un fichier possède un nom (O RLY ?), le nom de son créateur, sa taille, la date de création, ... Bon c'est pas très important.

Comme disent les *randoms gens sur internet*, **tout est fichier sous Linux/Unix**. Cela ne veut pas forcément dire que *tous* les fichiers sont de *même type*.

Ainsi, il existe plusieurs types de fichier sous Linux/Unix :

- **Les fichiers ordinaires**, c'est à dire les fichiers random comme *merci.c*, *i_et_y_se_font_defoncer.mp4*, *mcross.sh*, ...
- **Les répertoires**, c'est à dire les... répertoires et dossiers qui contiennent des fichiers.
- **Les fichiers spéciaux**, qui eux mêmes existent en deux type : les fichiers caractères, qui représentent les périphériques d'entrée-sorties (par exemple, le clavier est représenté sous la forme d'un fichier, allez voir dans */dev/*), et les fichiers bloc, qui représentent tout ce qui est disque et support de stockage.
- **Les liens**, qui ne comportent qu'un pointeur vers un fichier.

4.2 Système de fichiers sous Unix/Linux

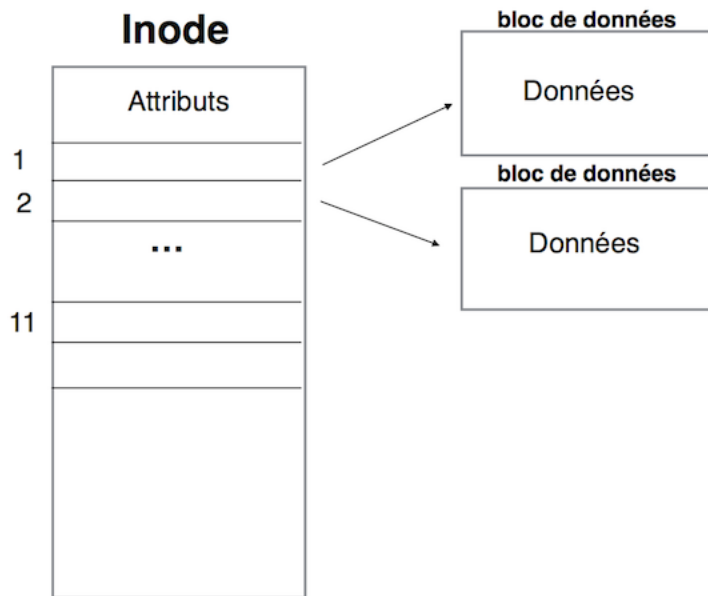
4.2.1 Les fichiers ordinaires

La gestion des fichiers sous U/L¹ repose sur la notion d'**inode** (ou i-noeud). Un inode est un *index* avec des cases. Pour des raisons de commodité, on va dire que la première case comporte les attributs du fichier. Ensuite, il y a 10 cases (ou 12 cases, sous Linux) qui comportent chacune une adresse vers un bloc de données. Après ces 10 (ou 12 cases), il y a 3 cases *spéciales*. Ces 3 cases sont spéciales car elles permettent ce qu'on appelle l'*adressage indirect*.

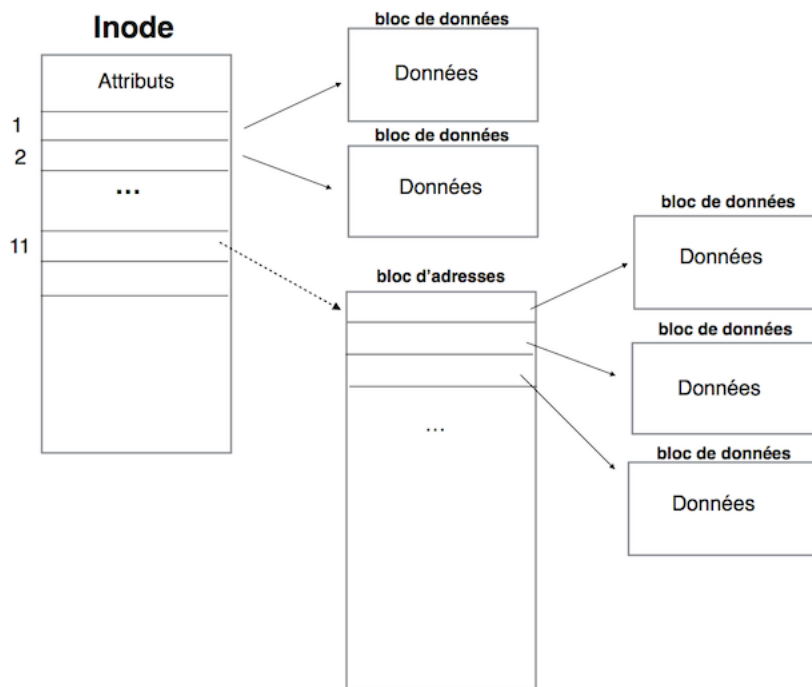
Mr. Toutlemonde : "waaaa c quoi l'adressage ??????"

Du calme, Mr. Toutlemonde. L'adressage signifie que la case donne l'adresse d'un bloc de données :

1. vu que j'ai la flemme de réécrire Unix/Linux partout, je vais abréger ça par U/L



On a ici de l'*adressage direct*. Voyons maintenant l'*adressage indirect* :



Comme vous le voyez, l'adressage indirect s'appelle comme ça car l'adresse contenue dans la case pointe vers *un bloc d'adresses*, et non pas **directement** un bloc de données. Chacune des adresses contenue dans le bloc d'adresses pointe vers un bloc de données.

Ainsi, la case 11 pointe vers un bloc d'adresse, la case 12 pointe vers un bloc d'adresses dont chaque adresse pointe vers **un autre** bloc d'adresses, et dans ce bloc d'adresse chaque adresse pointe vers un bloc de données

(adressage indirect double), et enfin pour la case 13 on applique l'adressage indirect triple².

Mr. Toutlemonde : "mé là y a CARLOS ki me demande 2 calculé la taille maxi d'1 féchier, cmnt je fé????"
Bon, j'imagine que *Mr. Toutlemonde* parle de l'exercice 1 du TD4. On va le faire ensemble avec les calculs.

Exercice 1 : Un système de gestion de fichiers sous Unix utilise des **blocs de 4096 octets** et code les **adresses sur 32 bits**.

Oui merci tg l'exercice. Tout d'abord, notez bien les informations en gras, ce sont les infos les plus importantes. Vu que 1 octet = 8 bits, une adresse tient sur **4 octets**.

Ensuite, on sait qu'un bloc tient sur 4096 octets, ce qui correspond à **4 ko**. On en déduit qu'un bloc contient donc **1024 adresses**.

Maintenant, on peut calculer la taille selon le type d'adressage :

- Direct : on a 10 blocs de 1024 adresses, donc 4096 octets
- Indirect simple : on a 1024 blocs, donc 4 Mo
- Indirect double : on a 1024*1024 blocs, donc 4 Go
- Indirect triple : on a 1024*1024*1024 blocs, donc 4 To

On fait maintenant la somme ; la taille max correspond à un fichier de 4 To et 4 Go et 4 Mo et 4096 octets.

4.2.2 Les répertoires

Les répertoires possèdent une structure différente entre UNIX et Linux.

Unix Sous Unix, un dossier est une table avec 2 entrées ; le numéro de l'inode du fichier sur 2 octets et son nom sur 14 octets. Cela se présente de la manière suivante :

N°Inode (2 oc)	Nom du fichier (14 oc)
432432	merci.c
9888889	ofc1
7893427489	sysfiche.tex
...	...

2. j'ai la flemme de détailler, mdr

Linux Sous Linux, c'est un peu plus compliqué :

Chaque case doit faire un multiple de 4 octets.					
N°Inoeud (2oc)	Taille de l'entrée (2oc)	Longueur du nom fichier (2oc)	Type du fichier (2oc)	Nom (chaque partie doit faire 4 oc)	
21	12 (8+4)		1 2 (repertoire)	. \0 \0 \0	
41	12 (8+4)		2 2 (repertoire)	. . \0 \0	
32	16 (8+8)		6 1 (fichier ordinaire)	E x o 1	. c \0 \0
39	16 (8+8)		8 1 (fichier ordinaire)	m i k u	. p n g
1337	12 (8+4)		3 2 (repertoire)	r e p \0	

La taille de l'entrée correspond à la somme des caractéristiques (8 octets) plus le nom du fichier. Vous devez vous demander à quoi correspondent les `\0` ; en fait, chaque case dans les noms de fichier doit faire **4 octets** ; si on ne remplit pas la case, on ajoute des `\0`. C'est tout.

4.3 Primitives systèmes en C (tou)

Parce oui, il faut bien jouer avec les fichiers en C.

Ce qu'il faut savoir, c'est que chaque processus possède une *File Descriptor Table*, soit, en bon français, une *Table de Descripteurs de fichiers*. En fait, un fichier est identifié au niveau système par un entier que l'on appelle *file descriptor*. La table contient donc des descripteurs de fichiers qui pointent chacune sur un fichier. Attention : les descripteurs 0, 1 et 2 sont réservés pour 3 fichiers très spéciaux, respectivement *stdin*, *stdout* et enfin *stderr*.

Enfin, le système d'exploitation possède une table commune à tous les processus ; c'est la *Table des fichiers ouverts*. Cette table permet au système de savoir quel fichier est ouvert par qui et avec quels droits. On ne va pas s'étendre dessus, c'est pas très intéressant.

Sur U/L, il faut passer par plusieurs étapes afin d'utiliser un fichier dans un programme C : il faut **ouvrir** le fichier, faire son petit traitement (**lecture**, **écriture**, etc) puis **fermer** le fichier. Tout cela se fait par le biais d'appels systèmes que nous allons détailler.

4.3.1 Ouvrir un fichier avec *open()*

open() se présente de la manière suivante :

```
int open( const char *fichier, int flag );
```

Commençons par les paramètres. **fichier* représente ici le nom du fichier **et son chemin**. Le *flag* représente en fait une constante définie dans *fcntl.h* qui définit les conditions dans lesquelles on souhaite ouvrir le fichier. Il en existe 4 d'importantes :

- `O_RDONLY`, comme `Open_ReadONLY`, va ouvrir le fichier en lecture uniquement
- `O_WRONLY`, ouvre le fichier en écriture uniquement
- `O_RDWR`, ouvre le fichier en lecture et écriture
- `O_APPEND`, ouvre le fichier en écriture **en fin de fichier**

Enfin, *open()* renvoie un entier (ou -1 si il se plante), qui correspond en fait au fameux **file descriptor**. Il est donc primordial de récupérer la valeur de retour d'*open* !

Exemple Imaginons que l'on souhaite ouvrir un fichier *fic* se trouvant à */home/mooshi/rep* en lecture et écriture, sachant que l'on se trouve dans un autre répertoire que le fichier *fic* ; on doit écrire :

```
int fd = open("/home/mooshi/rep/fic", O_RDWR);
```

Et C'EST GAGNÉ.

Création d'un fichier À noter : *open()* permet aussi de créer un fichier, selon les flags que l'on met. Si on utilise *O_CREAT*, on doit rajouter un paramètre pour les droits. Regardez l'exemple, ça sera plus clair.

```
/* Je souhaite créer un fichier oui avec des droits pour tout le monde */  
int fd_oui = open("oui", O_CREAT, 0777);
```

```
/* Je souhaite créer un fichier merci avec des droits que pour moi  
et seulement si il existe pas déjà */  
int fd_merci = open("merci", O_CREAT | O_EXCL, 0700);
```

Le 3eme paramètre est un octal qui définit les droits. Vous vous rappelez de *chmod*? C'est la même chose.

4.3.2 Lire dans un fichier avec *read()*

Regardons *read()* :

```
size_t read( int fd, void *buf, size_t nbOct )
```

Mr. Toutlemonde : omg c'est quoi ce *size_t*???????

size_t correspond juste à un int. C'est juste pour montrer que ce int désigne une taille en octets.

Maintenant, regardons les paramètres. En premier, on retrouve notre cher *file descriptor* que l'on a récupéré un peu plus tôt.

Pour le deuxième paramètre, il va falloir éclaircir la notion de **buffer**. Un buffer est une sorte de paquet d'une taille définie dans laquelle on va mettre des données. Très généralement, c'est un tableau de caractère qui occupe le rôle de buffer.

Enfin, le dernier paramètre correspond au nombre d'octets que l'on souhaite lire dans le fichier.

read() retourne le nombre d'octets effectivement lu dans le fichier.

Exemple Reprenons notre fichier *fic* du premier exemple. On souhaite lire son contenu sur 20 octets.

```
char buf[20]; // On cree un buffer de taille 20 octets  
int nbOctetsLus = read(fd, buf, 20); //On lit 20 octets du fichier
```

Mr. Toutlemonde : ouais mais il va où le contenu du fichier???

Ben il va ~~dans ton ent~~ ENFIN, non, mais on va voir ça tout de suite avec *write*.

4.3.3 Écrire dans un fichier avec *write()*

En fait, les données lues vont dans notre buffer, c'est à dire, dans l'exemple précédent, *buf*. *write()*, au lieu d'écrire dans le buffer depuis le fichier, va écrire dans le fichier depuis le buffer. Ainsi, on peut noter de nombreuses similitudes entre *write()* et *open()* :

```
size_t write(int fd, void *buf, size_t nbOct)
```

Ici, *nbOct* représente le nombre d'octet que l'on souhaite écrire dans le fichier et cette primitive renvoie ici le nombre d'octets effectivement écrits dans le fichier.

Souvent, on fait travailler *open* et *write* à la suite, puisque on utilise plusieurs fois le buffer.

Exemple