

Chapitre 1

Introduction

1.1 Les systèmes d'exploitation

MAlgré leurs différences du point de vue forme, taille ou type, les ordinateurs se composent de matériel et de logiciels, comme illustré dans la figure 1.1. Un ordinateur sans logiciels n'est que de la ferraille plutôt chère. Le système d'exploitation est une interface logicielle entre le matériel et les autres composants logiciels. Le matériel d'un système informatique est principalement composé d'un ou plusieurs processeurs (système monoprocesseur ou multiprocesseur), de la mémoire centrale (mémoire physique), de périphériques d'entrées/sorties (E/S). Ces composants sont connectés par un bus de communication.

Chaque processeur (CPU pour *Central Processing Unit* ou Unité Centrale de Traitement, UCT en français) se charge d'exécuter les instructions machine¹ d'un code en trois étapes séquentielles ou concurrentes : chargement, décodage et exécution. Sa puissance se mesure en IPS (nombre d'instructions par seconde)², MIPS (méga IPS), GIPS (giga IPS), etc.

La mémoire physique (RAM pour *Random Access Memory*) permet de stocker temporairement les instructions à exécuter et les données à traiter. Elle est l'espace de travail des processeurs. Elle est organisée comme un tableau d'octets. Chaque octet correspond à une adresse physique. Cependant, l'unité d'allocation d'espace physique aux processus est un cadre (frame/page physique). Un cadre est une plage contiguë d'adresses physiques. Sa taille varie d'un système d'exploitation à un autre. Elle est en général égale à 4 096 octets. Les mémoires physiques se distinguent par leurs capacités et leurs temps d'accès.

Les périphériques (clavier, souris, écran tactile, disques, carte réseau, etc.) permettent au(x) processeur(s) de stocker et de récupérer des informations à travers

1. Une instruction machine est une opération élémentaire qu'un processeur est capable de décoder et d'exécuter.

2. *L'instruction par seconde (IPS) est une unité de mesure de la performance d'un système informatique.* Pour plus d'informations voir : https://fr.wikipedia.org/wiki/Instructions_par_seconde

le bus. Ils offrent ainsi la possibilité au système d'exploitation de communiquer et d'interagir avec le monde extérieur (les utilisateurs, l'environnement, les autres ordinateurs, les réseaux Intranet et Internet).

Le bus connecte tous les périphériques d'E/S et la mémoire au(x) processeur(s). Il permet le transfert d'informations par diffusion entre composants connectés. Il comporte principalement des lignes de données, des lignes d'adresse et des lignes de contrôle. Les processeurs ont souvent des bus mémoire dédiés aux accès à la mémoire. La puissance d'un bus se mesure par son taux de transfert (débit) qui s'exprime en général en mégaoctets par seconde.

Les logiciels sont, à leur tour, composés de **programmes système** et de **programmes d'application**. Dans les programmes système, on retrouve le **système d'exploitation** et les différents utilitaires (compilateurs, éditeurs, interpréteurs de commandes, etc.). Les **programmes d'application** sont développés par des utilisateurs afin de répondre à des besoins ou résoudre des problèmes spécifiques.

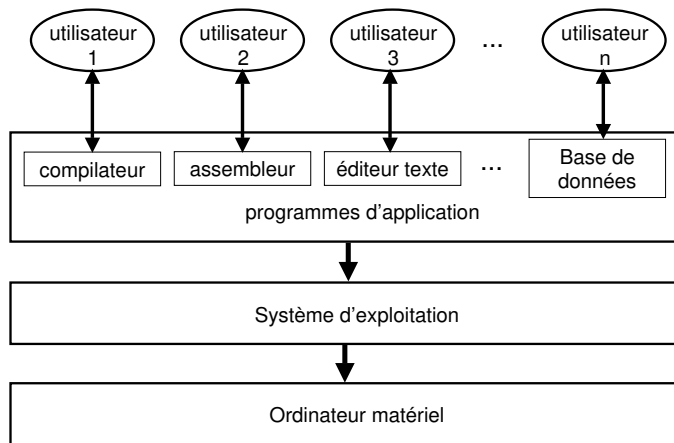


FIGURE 1.1 – Les couches d'un système informatique.

Le **système d'exploitation** gère et contrôle les composants de l'ordinateur tout en fournissant une base, appelée **machine virtuelle** à chaque utilisateur connecté au système. Il met à la disposition de chaque utilisateur un ensemble de services qu'il peut invoquer par de simples **appels système**. Son but principal est de permettre aux utilisateurs de *développer des applications sans se soucier des détails de fonctionnement et de gestion du matériel*. Par exemple, pour les systèmes de la famille Unix, la lecture d'un fichier est réalisée par un simple appel système : `read(fd, buf, 255)`, peu importe le type et l'unité de stockage du fichier (disque dur, SSD, DVD, clé USB, clavier, tube de communication, etc.).

Les fonctions principales d'un système d'exploitation

Les systèmes d'exploitation modernes ont comme fonctions principales :

- chargement et lancement des programmes,
- gestion du matériel (processeurs, mémoire physique, bus et périphériques),
- gestion des processus (programmes en cours d'exécution),
- gestion des fichiers,
- protection contre les erreurs et détection des erreurs, etc.

Exemples de systèmes d'exploitation

Parmi les systèmes d'exploitation qui ont été ou sont actuellement populaires, on peut citer :

- Windows (et MS-DOS),
- OS/2,
- Mac-OS,
- Unix (AIX, Xenix, Ultrix, Solaris, BSD, etc.)
- GNU/Linux

On retrouve également le système d'exploitation éducatif Minix³ développé par Andrew S. Tanenbaum et le simulateur de systèmes d'exploitation Nachos⁴.

1.2 Évolution des systèmes d'exploitation

Exploitation porte ouverte : 1945-1955

Les machines à calculer étaient construites au moyen de tubes électroniques. Ces machines étaient énormes, coûteuses, très peu fiables et beaucoup moins rapides car le temps de cycle se mesurait en secondes. Elles n'avaient pas de systèmes d'exploitation. Les programmes étaient écrits directement en langage machine : ils étaient chargés en mémoire, exécutés et mis au point à partir d'un pupitre de commande. Le mode de fonctionnement consistait à réserver une tranche de temps pour chaque programmeur (un seul programme en mémoire). Au début de 1950, la procédure s'est améliorée grâce à l'introduction de cartes perforées. Les problèmes évidents étaient dus au mode d'exploitation : peu économique, le matériel coûteux, la longue durée d'entrée des programmes et une mise au point pénible.

3. <http://www.cs.vu.nl/~ast/minix.html>

4. <http://www.cs.washington.edu/homes/tom/nachos/>

Traitement par lots : 1955-1965

Les machines étaient construites au moyen de transistors et dotées d'unités de bandes magnétiques. Elles étaient plus fiables mais toujours énormes (enfermées dans des salles climatisées). Les programmes étaient écrits en Fortran ou en assembleur sur des cartes perforées. Le mode de fonctionnement utilisé, montré à la figure 1.2, était le traitement par lots qui consiste à :

- transférer les travaux sur une bande magnétique,
- monter la bande sur le lecteur de bandes,
- charger en mémoire un programme spécial (l'ancêtre des systèmes d'exploitation appelé moniteur d'enchaînement de travaux) qui lit puis exécute, l'un à la suite de l'autre, les programmes de la bande (les résultats sont récupérés sur une autre bande, à la fin de l'exécution de tout le lot) et
- imprimer les résultats.

Ce mode d'exploitation nécessitait deux machines dont la plus puissante était réservée aux calculs et l'autre, moins chère, s'occupait des périphériques lents.

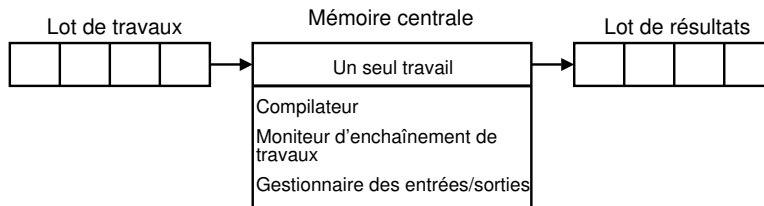


FIGURE 1.2 – Traitement par lots et organisation de la mémoire.

Le problème est que le processeur restait inutilisé pendant toutes les opérations d'entrée/sortie (E/S), comme on le voit dans la figure 1.3. Le temps d'attente des résultats était trop long et en plus il n'y avait pas d'interaction avec l'utilisateur.

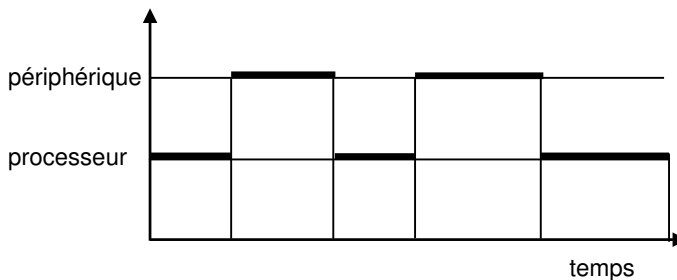


FIGURE 1.3 – Utilisation du processeur et des périphériques.

Multiprogrammation et traitement par lots : 1965-1980

L'introduction des circuits intégrés dans la construction des machines a permis d'offrir un meilleur rapport coût/performance. L'arrivée sur le marché des unités de disques, qui offrent l'accès aléatoire et des capacités de stockage importantes, a contribué à une série de développements des systèmes, notamment la possibilité de transférer les travaux vers le disque dès leur arrivée dans la salle machine. Cette technique s'appelle le *spool* (*Simultaneous Peripheral Operations On Line*) et est également utilisée pour les sorties. Cette notion de *spooling* ou traitement par lots subsiste dans les systèmes d'exploitation modernes.

Tous les travaux sur le disque en attente d'exécution sont conservés dans le *pool* des travaux en entrée. La mémoire est organisée en un ensemble de n partitions (figure 1.4). Chaque partition peut contenir au plus un travail. Le système d'exploitation réside aussi dans une partition.

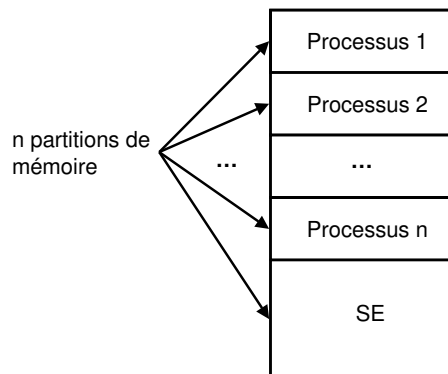


FIGURE 1.4 – Partitions de mémoire.

S'il y a une partition libre et des travaux dans le *pool* d'entrée, le système d'exploitation choisit un travail puis lance son chargement en mémoire. Il conserve en mémoire plusieurs travaux et gère le partage du processeur entre les différents travaux chargés en mémoire (la multiprogrammation). Le processeur est alloué à un travail jusqu'à ce qu'il demande une E/S (premier arrivé, premier servi) ou se termine. Lorsqu'un travail demande une E/S (en attente de la fin d'une E/S) ou se termine, le processeur est alloué à un autre travail en mémoire (le suivant). À la fin d'une E/S, une interruption se produit et le système d'exploitation reprend le contrôle pour traiter l'interruption et lancer ou poursuivre l'exécution d'un travail. Dès qu'un travail se termine, le système d'exploitation peut lancer le chargement, à partir du disque, d'un nouveau travail dans la partition qui vient de se libérer.

Supposons trois travaux **A**, **B** et **C**. Dans un système multiprogrammé, trois activités peuvent être donc menées en parallèle : 1. le chargement du travail **C** en mémoire, 2. l'exécution du travail **B** et 3. l'édition des résultats du travail **A**.

La multiprogrammation nécessite des circuits de contrôle pour protéger chaque travail contre les intrusions et les erreurs des autres. Les ordinateurs de cette époque possédaient ces circuits.

► **Exemple 1.** La figure 1.5 montre la gestion de l'exécution des travaux **A**, **B** et **C** supposés chargés en mémoire dans cet ordre.

L'exécution de **A** est entamée en premier puis lorsqu'il demande une entrée/sortie (E/S), le processeur commute sur **B**. A la fin de l'E/S demandée par **A**, le processeur suspend l'exécution de **B** pour commuter sur **A**. Conformément à leur ordre d'arrivée, on suppose que le travail **A** est plus prioritaire que **B** qui, à son tour, plus prioritaire que **C**. Après un certain temps de calcul, **A** demande de nouveau une E/S; ce qui provoque la commutation du processeur sur **B**. Durant l'exécution de l'E/S de **A**, le travail **B** demande une E/S. Il se met donc en attente car le périphérique est occupé. Le processeur commute alors sur le travail **C**. Lorsque l'exécution de l'E/S demandée par **A** est terminée, le processeur commute sur **A** et le traitement de la demande d'E/S du travail **B** est entamé par le périphérique d'E/S, etc.

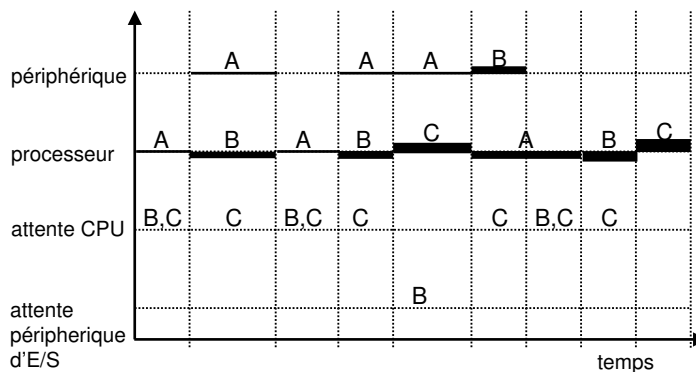


FIGURE 1.5 – Commutation du processeur.

Multiprogrammation et partage de temps : 1965-1980

Le désir d'un temps de réponse plus rapide et d'interactivité de l'exploitation a introduit la technique de partage de temps (systèmes temps partagé et/ou multi-utilisateurs) : plusieurs utilisateurs peuvent se connecter à la machine par l'intermédiaire de leurs terminaux et travailler en même temps.

Le processeur est alloué, à tour de rôle, pendant un certain temps à chacun des travaux en attente d'exécution. Au bout de ce temps, même si le travail en cours ne s'est pas terminé, son exécution est suspendue. Le processeur est alloué à un autre travail. Si plusieurs utilisateurs lancent à partir de leurs terminaux leurs programmes simultanément, ce mode d'exploitation donne l'impression que les

programmes s'exécutent en parallèle. Le cas de trois travaux A, B et C est montré dans la figure 1.6. Le temps de réponse pour chaque utilisateur est acceptable.

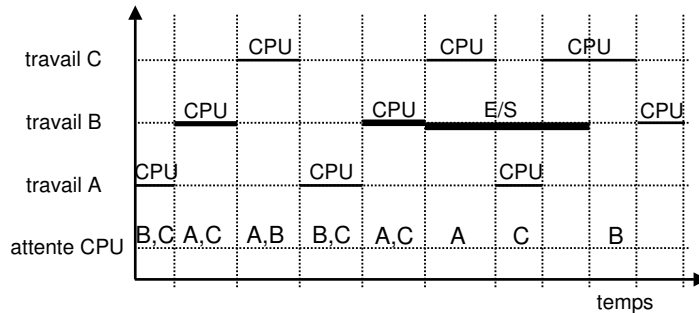


FIGURE 1.6 – Temps partagé.

Systèmes d'exploitation d'ordinateurs personnels

Ces systèmes d'exploitation mono-utilisateur ne mettaient pas l'accent sur l'usage du processeur, ni sur la protection. Leur but était de fournir une interface conviviale et une rapidité de réaction. Ensuite, les fonctionnalités des gros systèmes ont été peu à peu transférées vers les micro-ordinateurs.

Exploitation en réseau

Les réseaux d'ordinateurs personnels qui fonctionnent sous des systèmes d'exploitation en réseau permettent de se connecter sur une machine distante et de transférer des fichiers d'une machine à une autre. Ces systèmes d'exploitation requièrent une interface réseau, un logiciel de contrôle de bas niveau, ainsi que des programmes qui permettent une connexion distante et un accès aux fichiers qui se trouvent sur les différentes machines.

Exploitation en distribué

Les réseaux d'ordinateurs qui fonctionnent sous des systèmes d'exploitation distribués apparaissent aux yeux des utilisateurs comme une machine monoprocesseur, même lorsque ce n'est pas le cas. Le système d'exploitation distribué gère et contrôle l'ensemble des composants de tous les ordinateurs connectés (les processeurs, les mémoires, les disques, etc.).

Systèmes multiprocesseurs

Ces systèmes sont composés de plusieurs processeurs reliés au bus de l'ordinateur. Ils se caractérisent par leur capacité de traitement et leur fiabilité : la panne d'un processeur n'arrêtera pas le système

Système d'exploitation temps réel

Ce sont des systèmes spécialisés dans la conduite d'appareillage industriel ou dans la commande de processus où le temps joue un rôle critique (des contraintes temporelles strictes ou souples à respecter).

1.3 Interactions utilisateur/système

Pour un utilisateur, le système d'exploitation apparaît comme un ensemble de procédures, trop complexes pour qu'il les écrive lui-même. Les bibliothèques des **appels système** sont alors des procédures mises à la disposition des programmeurs. Sous un système GNU/Linux, un programme C/C++ peut utiliser des appels système comme `open()`, `write()` et `read()` pour effectuer des entrées/sorties de bas niveau.

L'**interpréteur de commandes** constitue une interface utilisateur/système et est disponible dans tous les systèmes d'exploitation. Pour les systèmes d'exploitation de la famille Unix ou GNU/Linux, il est lancé dès la connexion au système et invite l'utilisateur à introduire une commande. L'**interpréteur de commandes** récupère puis exécute la commande par combinaison d'appels système et d'outils (compilateurs, éditeurs de lien, etc.). Il affiche les résultats ou les erreurs, puis se met en attente de la commande suivante. Par exemple, la commande de l'interpréteur (shell) d'Unix suivante permet d'afficher à l'écran le contenu du fichier appelé `archive` :
`cat archive.`

L'introduction du graphisme dans les interfaces utilisateur a révolutionné le monde de l'informatique (celle du son est déjà en cours !). L'interface graphique a été rendue populaire par le Macintosh de Apple. Elle est maintenant proposée dans toutes les machines.

1.4 Appels système

En général, les processeurs ont deux modes de fonctionnement :

- Le **mode superviseur** (noyau, privilégié ou maître) réservé au système d'exploitation, où toutes les instructions sont autorisées.
- Le **mode utilisateur** (esclave) réservé aux programmes des utilisateurs et les utilitaires, où certaines instructions ne sont pas permises.