

Principios de Sistemas Operativos

Introducción

Armando Arce, arce@itcr.ac.cr

Tecnológico de Costa Rica

Un sistema operativo es un programa que administra el hardware de una computadora.

- También proporciona la base para los programas de aplicación y actúa como un intermediario entre el usuario y el hardware de la computadora.
- El propósito del sistema operativo es crear un entorno en el que un usuario pueda ejecutar programas.
- Un objetivo secundario es utilizar el hardware del computador de forma eficiente.

Funciones y objetivos del sistema

Un sistema informático puede dividirse en tres componentes:

- El hardware: proporciona los recursos básicos de cómputo (p. ej. CPU, memoria, y dispositivos de E/S).
- Los programas de aplicación: definen las formas en que los recursos se emplean para resolver los problemas informáticos de los usuarios (p.ej. hojas de cálculo, procesadores de palabras, visualizador de web)
- El sistema operativo: controla y coordina el uso del hardware entre los diversos programas de aplicación por parte de los usuarios

Componentes de un sistema informático

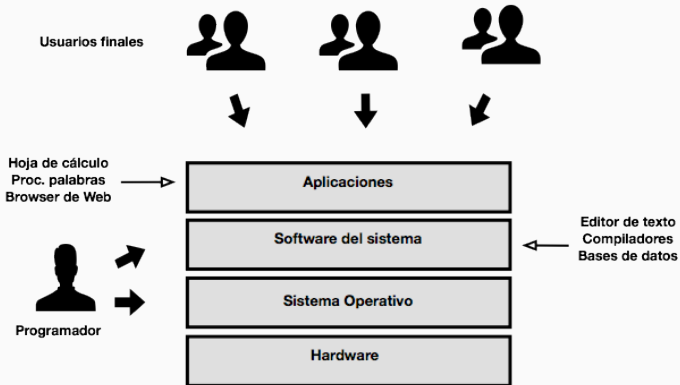


Figure 1: Componentes de un sistema informático

En una computadora personal el sistema operativo se diseña principalmente para que sea de fácil uso, prestando mayor atención al rendimiento y menos a la utilización de recursos.

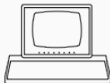
- En un mainframe el sistema operativo se diseña para maximizar la utilización de recursos, asegurando que el CPU, memoria y E/S se usen de forma eficiente entre todos los usuarios del sistema.

Punto de vista del usuario

- En una estación de trabajo, conectada a una red en la que se comparten algunos recursos, el sistema operativo se diseña para llegar a un compromiso entre la usabilidad individual y la utilización de recursos compartidos.
- En los dispositivos móviles el sistema operativo se diseña en función de la usabilidad individual, aunque el rendimiento, medido según la duración de la batería, es también importante.
- En dispositivos empotrados (p.ej. electrodomésticos, automóviles) el sistema operativo se diseña para funcionar sin intervención del usuario.

Punto de vista del usuario

Computador personal



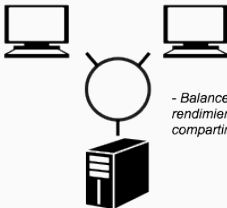
- Facilidad de uso
- Rendimiento

Mainframe



- Maximizar los recursos compartidos

Estaciones de trabajo



- Balance entre rendimiento y compartir recursos

Dispositivo móvil



- Facilidad de uso
- Rendimiento de energía

Dispositivo empotrado



- Funcionamiento sin intervención humana

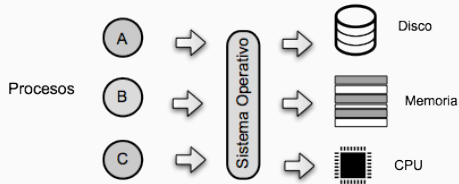
Figure 2: Punto de vista del usuario

El S.O. administra la asignación de recursos (tiempo de CPU, espacio en memoria, espacio en disco, etc.) garantizando una asignación ordenada, controlada y equitativa entre los diversos programas.

- El S.O. también se comporta como un programa de control que administra la ejecución de programas de usuario para evitar errores y mejorar el uso de la computadora.

Punto de vista del sistema

Asignador de recursos



Controlador de procesos

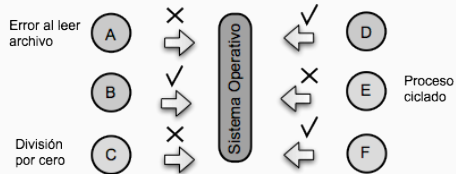


Figure 3: Punto de vista del sistema

Definición de sistemas operativos

El objetivo fundamental de una computadora es ejecutar programas de usuario y resolver problemas fácilmente.

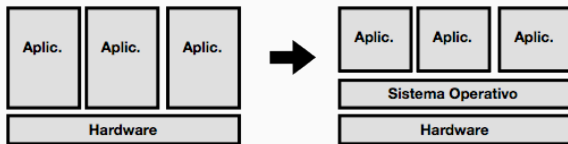
- Con este fin se construye el hardware de la computadora.
- Como el hardware por sí solo no es fácil de utilizar, se desarrollan programas de aplicación.
- Estos programas requieren ciertas operaciones comunes, como las que controlan los dispositivos de E/S.
- Así, las funciones comunes de controlar y asignar recursos se reúnen en un solo programa: el sistema operativo.

Definición de sistemas operativos

Una definición común es: un sistema operativo es aquel programa que se ejecuta continuamente en la computadora (llamado también kernel), todos los demás son programas de aplicación.

Definición de sistemas operativos

Agrupar funciones comunes



El programa que corre todo el tiempo



Arrow icon by
Kayla Petrich, from Noun Project

Figure 4: Definición de sistema operativo

Para ver qué son y qué hacen los sistemas operativos, es necesario hacer una recopilación de su evolución a través del tiempo.

- Esta evolución permite identificar los elementos comunes de los sistemas operativos y entender por qué se han desarrollado en la forma en que lo han hecho.

Sistemas por lotes sencillos

Los primeros computadores eran máquinas enormes (físicamente) que se controlaban desde una consola.

- El sistema operativo en estos primeros computadores era sencillo.
- Su principal obligación era transferir el control automáticamente de un trabajo al siguiente.
- El sistema operativo siempre estaba (residente) en memoria.

A fin de agilizar el procesamiento, los programas con necesidades similares se agrupaban *en lotes* y se introducían en el computador (por parte del operador) como un grupo. Este leía cada trabajo individual con sus propias tarjetas de control y una vez terminado el trabajo su salida era impresa.

La característica definitiva de un sistema por lotes es la falta de interacción entre el usuario y el trabajo mientras éste se ejecuta.

- El trabajo se prepara y se entrega, y cierto tiempo después aparecía la salida.
- El retardo entre la presentación del trabajo y su terminación (llamado *tiempo de retorno*) se debía al volumen de cómputo requerido, o de retardos antes de que el sistema operativo comenzara a procesar el trabajo.

Sistemas por lotes sencillos

Sistema por lotes



Spooling

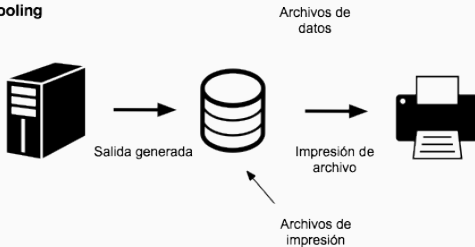


Figure 5: Sistema por lotes sencillo

En un sistema por lotes sencillo la CPU con frecuencia está ociosa.

- Esta inactividad ocurre porque los dispositivos de E/S mecánicos son intrínsecamente más lentos que los dispositivos electrónicos.
- La introducción de la tecnología de discos fue útil en este sentido.
- En vez de leer las tarjetas directamente del lector a la memoria, para después procesar el trabajo, las tarjetas pueden leerse del lector al disco.

Sistemas por lotes sencillos

De igual forma cuando un trabajo escribe su salida cada línea se copia en un *buffer* (área de memoria) del sistema y se escribe al disco. Una vez que el trabajo termina, la salida se imprime realmente.

- Esta forma de procesamiento, se llama *spooling* (simultaneous peripheral operation on-line).
- En resumen, el disco se utiliza como buffer de gran tamaño, para *leer por adelantado* hasta donde sea posible de los dispositivos de entrada, y para guardar los archivos de salida hasta que los dispositivos de salida puedan aceptarlos.

Sistemas por lotes multiprogramados

El *spooling* da origen a una importante estructura de datos, la *cola de trabajos*. Como resultado del *spooling*, puede haber varios trabajos ya leídos esperando en el disco, listos para ejecutarse.

- La cola de trabajos contiene todos los procesos que residen en disco esperando la asignación de la memoria principal.
- Si hay varios trabajos preparados para pasar a memoria y no hay espacio suficiente para ellos, entonces el sistema debe hacer una selección de los mismos.
- La toma de esta decisión es lo que se conoce como *planificación de trabajos*.

Sistemas por lotes multiprogramados

Uno de los avances más importantes de los sistemas operativos es la multiprogramación. La operación *fuera de línea* (no interactiva) y el *spooling* para traslapar la E/S tienen sus limitaciones.

- En general, un solo proceso no puede mantener ni la CPU ni los dispositivos de E/S ocupados todo el tiempo.
- Por ello, el sistema operativo mantiene en memoria simultáneamente varios trabajos.
- Este conjunto de trabajos puede ser un subconjunto de los trabajos guardados en la cola de trabajos, la cual contiene todos los trabajos que entran al sistema.

Sistemas por lotes multiprogramados

Cola de trabajos



Multiprogramación

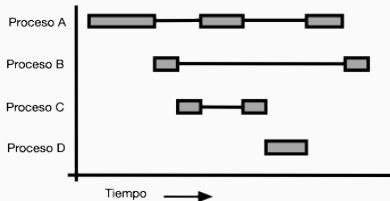


Figure 6: Sistema multiprogramados

Sistemas por lotes multiprogramados

En la *multiprogramación* el sistema operativo toma y comienza a ejecutar uno de los trabajos que está en memoria.

- Eventualmente, el trabajo puede tener que esperar a que se complete alguna otra tarea, como por ejemplo una operación de E/S.
- En este momento el S.O. cambia de trabajo y ejecuta otro.
- Cuando el nuevo trabajo tiene que esperar, la CPU, conmuta al trabajo anterior (o bien, a otro trabajo si el primero aún se encuentra esperando) y así sucesivamente.

Sistemas de tiempo compartido

El *tiempo compartido* (o *multitarea*) fue una extensión lógica de la multiprogramación, que proporciona la interacción del usuario.

- En los sistemas de tiempo compartido la CPU ejecuta múltiples trabajos conmutando entre ellos, pero las conmutaciones se producen tan frecuentemente que múltiples usuarios pueden interactuar con cada programa mientras éste está en ejecución.
- Un sistema de este tipo se le conoce como *multiusuario*.

Sistemas de tiempo compartido

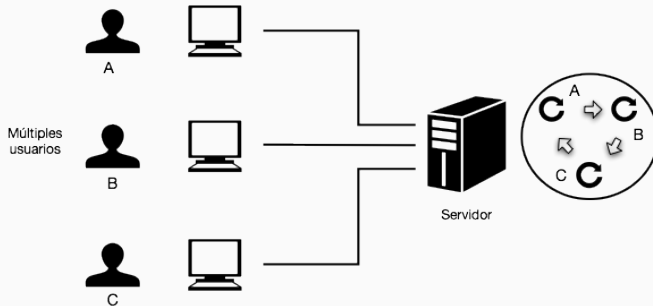


Figure 7: Sistema de tiempo compartido

Sistemas de tiempo compartido

Un sistema operativo de tiempo compartido permite que muchos usuarios compartan en forma interactiva la computadora.

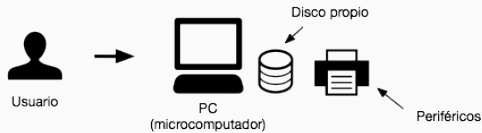
- Dado que el tiempo de ejecución de cada comando del usuario es pequeño, sólo es necesario un tiempo pequeño de CPU para cada usuario.
- Puesto que el sistema cambia rápidamente entre usuarios, cada usuario tiene la impresión que el sistema está dedicado sólo a él.

La caída en los costos de hardware volvió factible tener un sistema de computación dedicado a un solo usuario (*monousuario*).

- Durante su primer década, la CPU de los PC carecía de las funciones necesarias para proteger el sistema operativo de los programas de usuario.
- Por ello, los sistemas operativos de PC no eran ni *multiusuario* ni *multitarea*, aunque sí eran *interactivos*.

Sistemas de computador personal

Sistema monousuario



Sistema monotarea

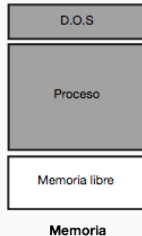


Figure 8: Sistema de computador personal

Los objetivos de estos sistemas operativos han cambiado con el tiempo; en lugar de maximizar la aprovechamiento de la CPU y los periféricos, hoy en día estos sistemas optan por maximizar la comodidad del usuario y la rapidez con la que responden a sus necesidades.

- Además, con el tiempo las funciones creadas para un sistema de *macrocomputadoras* grandes han pasado a los *microcomputadoras*.

Anteriormente la mayor cantidad de sistemas eran mono procesador; es decir, tenían una sola CPU.

- Sin embargo, la tendencia actual es hacia sistemas multiprocesador.
- Tales sistemas tienen más de un procesador en comunicación íntima, los cuales comparten el bus del computador, el reloj, la memoria y los dispositivos periféricos.
- Se dice que estos sistemas están *fuertemente acoplados*.

Sistemas multiprocesador

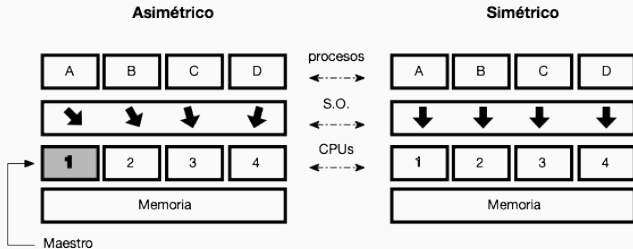


Figure 9: Sistemas multiprocesador

Sistemas multiprocesador

Los sistemas de múltiple procesador más comunes en la actualidad sigue el modelo de *multiprocesamiento simétrico*, en el que cada procesador ejecuta una copia idéntica el sistema operativo, y estas copias se comunican entre sí cuando es necesario.

- Algunos sistemas utilizan *multiprocesamiento asimétrico*, en el que a cada procesador se asigna una tarea específica.
- Un procesador maestro controla el sistema; los demás procesadores obtienen sus instrucciones del maestro o bien tienen tareas pre definidas.
- Este esquema define una relación *maestro-esclavo*.
- El procesador maestro planifica y asigna trabajo a los procesadores esclavos.

Un sistema distribuido es una colección de computadoras físicamente separadas y posiblemente heterogéneas que están conectadas en red para proporcionar a los usuarios acceso a los diversos recursos que el sistema mantiene.

- Se dice que estos temas están *débilmente acoplados*.

En un sistema distribuido las diferentes computadoras se conectan por medio una red de comunicaciones, que puede configurarse de varias maneras distintas.

- La red puede estar total o parcialmente conectada, y su diseño debe considerar estrategias de encaminamiento y conexión, y también los problemas de contención y seguridad.

Sistemas distribuidos

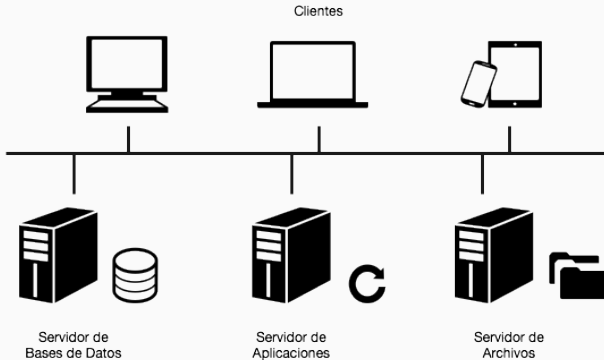


Figure 10: Sistemas distribuidos

El acceso a recursos compartidos permite acelerar los cálculos, incrementar la funcionalidad, ampliar la disponibilidad de los datos y mejorar la confiabilidad.

- Bajo un sistema operativo distribuido las máquinas pueden realizar sus tareas mientras estén conectadas a las demás.
- Las diferentes máquinas se comunican de modo que se crea la ilusión de que un único sistema controla la red.