

1. Introducción a los Sistemas Operativos

Objetivos. El objetivo de este tema es definir el concepto de sistema operativo que se manejará a lo largo de la asignatura, dando a conocer los distintos tipos de sistemas existentes, así como presentar a través de una perspectiva histórica los problemas que ha de resolver un sistema operativo moderno.

1.1 Concepto de Sistema Operativo. Objetivos.

Un sistema operativo es el software que hace de interfaz tanto entre los usuarios y la máquina, como entre esta última y los programas. Dicho de otra forma, el sistema operativo es el software que proporciona un entorno que permite a los usuarios ejecutar programas.

De esta forma, los objetivos fundamentales de un sistema operativo son:

1. Hacer que la máquina sea fácil de utilizar, tanto por los usuarios finales de ésta, como por los desarrolladores y administradores del sistema.
2. Hacer que el funcionamiento de la máquina sea lo más eficiente posible.

Nótese que bajo ambos objetivos subyace el propósito de reducir los costes de explotación del sistema.

En la Ilustración 1-1 se muestra la organización típica de un sistema informático. En ésta, la capa inferior está compuesta por el hardware del sistema (CPU, dispositivos de E/S y memoria, así como otros dispositivos o coprocesadores auxiliares), siendo la interfaz que proporciona a la capa superior de muy bajo nivel: la interfaz de programación es exclusivamente en código máquina, y los dispositivos se manejan directamente a través de sus registros de control, lo cual, evidentemente, resulta extremadamente complejo. Sobre esta interfaz de tan bajo nivel, el desarrollo de un programa sería extremadamente costoso, pues operaciones tan simples como leer un carácter de un archivo conllevaría una gran complejidad. Y una vez desarrollado el programa, cargarlo y ejecutarlo sería también muy complejo y costoso.

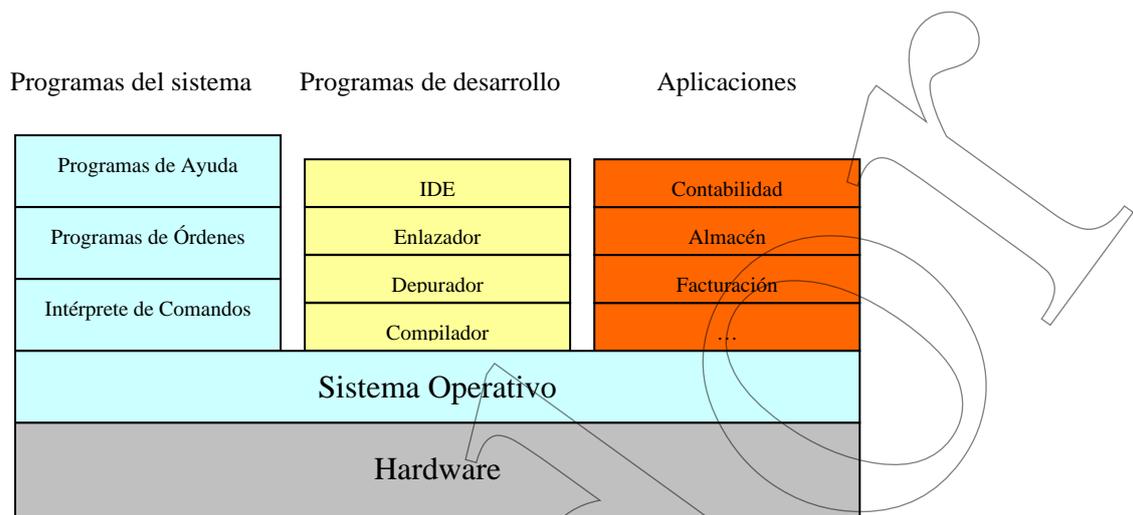


Ilustración 1-1: Estructura típica de un sistema informático

Con la complejidad de esta interfaz hardware se enfrenta el sistema operativo. Éste proporciona a las capas superiores una interfaz con un mayor grado de abstracción, en la que los programadores no necesitan tener en cuenta las particularidades del hardware a la hora de construir sus programas, pues dichas particularidades quedan ocultas tras la interfaz del sistema operativo. De esta forma, no sólo se simplifica la tarea de construcción de programas, sino que además dichos programas se hacen independientes de las particularidades del hardware. Además de ello, tareas del usuario final como organizar la información o ejecutar los programas se simplifican enormemente. Será en el estudio de esta capa en lo que nos centraremos a lo largo de esta asignatura.

Sobre el sistema operativo, se pueden ejecutar tres tipos de programas:

- **Programas del sistema:** Son necesarios para utilizar el sistema operativo, por lo que habitualmente se distribuyen juntos y proceden del mismo fabricante. Al ser externos al sistema operativo, se pueden sustituir por otros con formas de utilización diferentes. En este grupo de programas encontramos por ejemplo los intérpretes de comandos (que se estudiarán en el próximo capítulo) los programas de órdenes, como son `ls`, `cp`, o `mv` en el caso de los sistemas UNIX, o `format` o `xcopy` en el caso de los sistemas de Microsoft, o los programas de ayuda o documentación en línea.
- **Programas de desarrollo:** Sirven para construir otros programas. A este grupo pertenecen los compiladores, depuradores, montadores de enlace, gestores de bibliotecas o los entornos de desarrollo integrados. Si la máquina no se va a utilizar para desarrollar programas sino que sólo se utilizará para explotarlos, no son necesarios, por lo que no todos los sistemas operativos los incluyen como parte de sí mismos.
- **Programas de aplicación:** También conocidos simplemente como aplicaciones, son los programas que realizan el trabajo útil, verdadero objetivo final del sistema informático. Dado que normalmente sus usuarios no son informáticos, deben estar contruidos de forma que no se requieran conocimientos en informática para manejarlos.

A la vista de lo expuesto hasta el momento, se pueden entender las funciones del sistema operativo desde dos puntos de vistas complementarios, que se describen en los siguientes epígrafes.

1.1.1 El sistema operativo como máquina ampliada

Como se ha indicado anteriormente, el sistema operativo presenta a las capas superiores una interfaz de programación de más alto nivel que la proporcionada por el hardware desnudo. Esto equivaldría a considerar la máquina sobre la que se coloca el sistema operativo como una máquina ampliada, más fácil de utilizar que la real, que utiliza criterios homogéneos, es decir, que problemas similares se resuelven de formas similares.

Por ejemplo, una solicitud del tipo “escribir estos 20 caracteres en este archivo”, se resuelve con tan sólo tres parámetros: archivo, dirección de los caracteres y número de caracteres. No es necesario por tanto que el programado tenga en cuenta el tipo de dispositivo sobre el que se encuentra el archivo, que determine la ubicación física del bloque o bloques en los que debe escribir los datos, que programe la transferencia de los datos atendiendo al tipo de dispositivo, que compruebe si la transmisión se ha efectuado correctamente, etcétera.

El objetivo es por tanto que los programas no necesiten conocer sobre qué máquina se están ejecutando. Esto es posible puesto que el sistema operativo ofrece la misma interfaz incluso entre máquinas muy distintas.

1.1.2 El sistema operativo como gestor de recursos

Este otro enfoque, complementario con el anterior, considera al sistema operativo como un administrador al que los programas de usuario solicitan recursos o piden que realice para ellos determinadas actividades. El sistema operativo analiza cada petición, decidiendo si los programas tienen derecho a efectuarlas, y decide cuándo atenderá cada petición en función de las condiciones de trabajo para obtener el mejor rendimiento posible.

El sistema debe por tanto gestionar los recursos de forma eficiente, tratando de aprovechar el equipo lo mejor posible. Para ello, el sistema operativo asigna coordinadamente los recursos a los procesos, de manera que se reduzcan los desperdicios al mínimo.

1.2 Evolución Histórica

A continuación, se presenta a vuelo de pluma una perspectiva histórica de la evolución de los sistemas operativos en paralelo a la evolución del hardware sobre el cual se han ido ejecutando, desde las primeras máquinas construidas a mediados del siglo pasado hasta las máquinas actuales. El estudio se presenta por generaciones, siendo cada generación consecuencia de una o más innovaciones tecnológicas. En cada generación, las características de los sistemas operativos serán consecuencia de las necesidades impuestas por el hardware.

1.2.1 Primera generación: 1945-1954

1.2.1.1 Innovaciones tecnológicas

Las máquinas de cálculo (como así se les llamaba) construidas con anterioridad a esta fecha se basaban en componentes electromecánicos. No fue hasta la aparición de la lámpara termoiónica, dispositivo de incandescencia similar a una bombilla en apariencia, pero funcionalmente comparable a un transistor, cuando se construyeron las primeras máquinas de cálculo totalmente electrónicas.

Otras innovaciones aparecidas en este periodo fueron el tambor magnético (similar funcionalmente al disco magnético, pero físicamente en forma de cilindro, estando la información almacenada en el lateral del mismo). Los dispositivos de entrada/salida más habituales eran las tarjetas perforadas y las cintas de papel perforado.

1.2.1.2 Algunas máquinas relevantes

- ENIAC (Electronic Numerical Integrator and Calculator, 1945): Esta fue la primera máquina de cálculo construida exclusivamente con dispositivos electrónicos, si bien hay que aclarar que su organización no era Von-Neumann (no disponía de programa almacenado), pues la programación se llevaba a cabo desde un panel de interconexión, cambiando manualmente conexiones físicas. Esta máquina contaba con 17.468 lámparas y simulaba electrónicamente el funcionamiento de las ruedas de las calculadoras mecánicas. Disponía de 20 registros de 10 lámparas cada uno, de forma que sólo conducía una lámpara, simulando así cada registro un dígito, pudiendo operar con tanto con números de 20 dígitos.
- EDSAC (Electronic Delay Storage Automatic Calculator, 1949): Esta fue la primera máquina con organización Von-Neumann. En 1950 se desarrolló un lenguaje de programación en ensamblador para ella, así como una metodología de desarrollo basada en la construcción de biblioteca de rutinas.
- Primeras máquinas comerciales: A partir de 1951 se construyeron las primeras máquinas de uso comercial: Univac I y Mark I (1951), IBM 701 (1952), IBM 702 e IBM 650 (1953).

1.2.1.3 Explotación

El altísimo coste de estas máquinas tenía como consecuencia que cualquier organización que desembolsara los fondos necesarios para adquirir una, debía emplearla para satisfacer las necesidades de cálculo del mayor número posible de usuarios. Con este fin, la explotación del sistema se podía llevar a cabo de tres maneras:

1. Acceso sin restricciones: Cada usuario del sistema solicita turno para emplear la máquina. Se le asignan un número de horas, y el sufrido usuario llega con sus programas escritos en papel, los pasa a tarjetas perforadas, los ensambla obteniendo un nuevo conjunto de tarjetas, carga el programa ensamblado en memoria, lo ejecuta, corrige los errores que encuentre, volviendo a repetir desde el principio, y si todo va bien, al fi-

nal de su turno de uso tendrá un listado en papel con los resultados de su ejecución.

El problema de esta forma de explotación es que los usuarios del sistema son ingenieros, físicos o matemáticos con ciertos conocimientos de programación, pero posiblemente un tanto patosos en el manejo de los distintos dispositivos que componen la máquina. Además, parte del tiempo que el usuario tiene asignada la máquina se invierte en buscar errores de programación. Esto reduce significativamente el aprovechamiento de la máquina

2. Acceso reservado al operador: Una alternativa a la anterior forma de explotación consiste en contratar un usuario (operador) especializado no necesariamente en la programación de la máquina, sino en su manejo físico, que realice las labores de perforado de tarjetas, ensamblado de programas, ejecución, obtención de listados... los usuarios del sistema entregan a este sus programas, y pasado un tiempo se pasan a recoger o bien el listado con los resultados de su programa... o de sus errores.
3. Procesamiento discontinuo: Esta forma de explotación consiste simplemente en que el operador organiza su trabajo por tandas para mejorar el rendimiento del sistema: en lugar de procesar trabajo por trabajo, ensamblando primero y ejecutando después cada uno de ellos, en primer lugar carga el ensamblador y ensambla todos los programas, y a continuación, los ejecuta uno tras otro. De esta forma, se ahorra tener que cargar una y otra vez el ensamblador y los programas auxiliares de este, mejorando sensiblemente el aprovechamiento de la máquina.

1.2.1.4 Necesidades a satisfacer por parte del sistema operativo

En esta etapa no se puede hablar aún de sistema operativo propiamente dicho. El equivalente a éste sería un "embrión de sistema operativo" consistente en unas cuantas rutinas de entrada/salida que permiten la carga y ejecución de programas, entre ellos, los propios programas del sistema como el ensamblador.

1.2.2 Segunda generación: 1954-1963

1.2.2.1 Innovaciones tecnológicas

La innovación tecnológica que actuó como precursora de esta generación fue el transistor. Los primeros transistores eran componentes discretos (no encapsulados en chips, lo cual ocurrirá en la próxima generación) capaces de sustituir a las lámparas termoiónicas, pero con un tamaño varias decenas de veces inferior y un consumo eléctrico también varios órdenes de magnitud inferior, siendo además sensiblemente más rápidos y baratos.

Además de ello, se construyeron también las primeras memorias de ferrita, que reducían drásticamente el espacio necesario para almacenar un bit, así como las unidades de cinta magnética, con un rendimiento y fiabilidad muy superior a sus equivalentes de papel perforado.

1.2.2.2 Algunas máquinas relevantes

- TRADIC (Transistorized Digital Computer, Bell Laboratories, 1955): Fue el primer equipo totalmente transistorizado. La primera unidad destinada a su venta se fabricó en 1958.
- IBM 7094 (1962): Fue una de las máquinas más potentes de la época. Disponía de una memoria de 32K palabras de 36 bits. Era capaz de alcanzar 0,35 MIPS en operaciones de coma flotante.
- PDP-1 (Digital, 1961): Fue uno de los primeros minicomputadores. Disponía de 4K palabras de 16 bits, si bien su precio era el 5% de un IBM 7094 mientras que para algunas aplicaciones era tan rápido como éste.

1.2.2.3 Explotación

La forma en la que se lleva a cabo la explotación de los equipos es una evolución natural del procesamiento discontinuo que se describe en el apartado Explotación 1.2.1.3, si bien aparece el acceso interactivo.

- Ejecución continua de trabajos: Se reserva una parte de la memoria para mantener en ella un programa de control, que automatiza el tránsito de un programa al siguiente, y unos subprogramas de lectura y escritura, que llevan a cabo estas operaciones, y que enlazan con el anterior cuando detectan algún fallo¹. Así surge el concepto de trabajo, como una sucesión de operaciones que forman una unidad a efectos de ejecución y contabilidad de recursos del sistema. En definitiva, lo que se hace es automatizar el trabajo realizado por el operador que se describe en epígrafe 1.2.1.3.
- Procesamiento por lotes: Las máquinas construidas a lo largo de la anterior generación, ya anticuadas, resultaron no obstante demasiado caras como para deshacerse de ellas sin más. Estas máquinas encontraron utilidad como equipos auxiliares en la ejecución continua de trabajos descrita anteriormente. En la Ilustración 1-2 se muestra una configuración típica, basado un IBM 7094, usando como equipo auxiliar un IBM 1401. Esta máquina se usaba en primer lugar para pasar los programas de tarjetas perforadas a cinta magnética, formando así un lote de trabajo que se pasaba a la máquina principal, el IBM 7094. Esta máquina lee los trabajos de la cinta y los ejecuta uno tras otro, guardando los resultados en una cinta de salida, que posteriormente se llevaba de vuelta al IBM 1401 y se imprimían en papel.

¹ Los programas informaban al programa de control sobre la ocurrencia o no de errores mediante un código de terminación numérico. Esta es una práctica que se mantiene aún hoy en día.



Ilustración 1-2: Procesamiento por lotes usando equipos auxiliares

- Acceso interactivo: El principal inconveniente de los sistemas por lotes es que un programador tenía que esperar, a veces durante horas, desde que entregaba un trabajo hasta que recogía los resultados, resultados que perfectamente podrían consistir en un código de error. Para eliminar esta espera, a finales de esta generación se construyeron los primeros sistemas interactivos. Estos sistemas están dotados de un terminal en el que el usuario puede teclear una orden y esperar un resultado, corrigiendo de inmediato cualquier error cometido. No obstante, no será hasta la próxima generación cuando este tipo de sistema se implante definitivamente.

1.2.2.4 Necesidades a satisfacer por parte del sistema operativo

Es en esta generación cuando aparece por primera vez el término Sistema Operativo para hacer referencia al software que facilita el uso de la máquina. Se constituye un grupo de usuarios SHARE que en 1955 aporta ideas y sugerencia para un sistema operativo común, denominado SOS (SHARE Operating System).

Las funciones llevadas cabo por estos elementales sistemas operativos se concentran en la automatización de la ejecución continuada o por lotes de trabajos. El sistema operativo proporciona la infraestructura necesaria para la ejecución de un trabajo tras otro, garantizando que la máquina esté en un estado inicial determinado al principio de cada trabajo, con independencia de lo que haya hecho el trabajo anterior. El elemento central del sistema operativo se denomina supervisor o también monitor de encadenamiento.

1.2.3 Tercera generación: 1963-1971

1.2.3.1 Innovaciones tecnológicas

El punto de inflexión tecnológico que desencadenó esta nueva etapa fue la aparición de los circuitos integrados. A partir de ahora se pueden integrar decenas (SSI, Small Scale Integration) o centenares (MSI, Medium Scale Integration) de transistores en un circuito integrado, lo que reduce enormemente tanto el tamaño de los circuitos como su coste y su consumo, aumentando a su vez la velocidad de proceso.

Al aumentar drásticamente la velocidad de los procesadores, empezó a resultar prohibitivo que estos esperen a los dispositivos, lo que tuvo como consecuencia la necesidad de la multiprogramación desde el punto de vista software, y la aparición de coprocesadores para gestionar el control de los diversos subsistemas. Por ejemplo, técnicas

como el uso de interrupciones para el control de los dispositivos aparecieron en esta generación.

Además de ello, hicieron su aparición los discos magnéticos, los cuales relegaron las cintas magnéticas a un segundo plano. Por su parte, las memorias de ferrita se vieron también desplazadas por las memorias en circuito integrado, mucho más rápidas y de mayor capacidad, aunque de tamaño y consumo mucho más reducido.

1.2.3.2 Algunas máquinas relevantes

Una característica importante de esta generación es la aparición de las primeras familias de ordenadores. Hasta ahora, cada máquina se construía con una configuración básica y un reducido (o inexistente) número de opciones, construyéndose cada dispositivo de forma específica para cada modelo de máquina. En 1960, IBM lanza la serie IBM 360, que se constituyó la primera familia de ordenadores y productos que cubría tanto el mercado de gestión como científico. Esta familia disponía de una gama de modelos y periféricos compatibles, compartiendo el mismo lenguaje máquina y el mismo sistema operativo, OS/360. Como lenguaje de programación de alto nivel, compartían también el lenguaje PL/1 (Programming Language 1) y como lenguaje de control de trabajos, también compartían el lenguaje JCL (Job Control Language).

Como principal problema de esta familia, hay que citar su complejidad desde el punto de vista software. El sistema operativo OS/360 estaba escrito a base de millones de líneas de código, conteniendo gran cantidad de errores, lo que obligaba a producir continuas revisiones que subsanaban determinados errores, e introducían otros.

1.2.3.3 Explotación

La característica más importante en esta etapa es que los procesadores comienzan a ser muchos órdenes de magnitud más rápidos que los dispositivos. Esto hace que mantener en espera al procesador durante las operaciones de entrada/salida sea prohibitivo, especialmente teniendo en cuenta el aún elevadísimo precio de los equipos. Para evitar que el procesador tenga que esperar a los dispositivos, aparece el concepto de multiprogramación, consistente en tener varios procesos en ejecución, de forma que cuando un proceso efectúa una operación que no se pueda satisfacer inmediatamente, otro proceso puede hacer uso de la CPU mientras dicha operación se lleva a cabo.

Dado que al final de la anterior generación ya se desarrollaron los primeros sistemas interactivos, las formas de explotación que nos encontramos en esta generación son:

- **Procesamiento por lotes con multiprogramación:** Es la evolución natural del procesamiento por lotes descrito en el apartado 1.2.2.3. La diferencia fundamental es que ahora se tienen varios procesos cargados simultáneamente en memoria, eliminándose las máquinas auxiliares empleadas en la anterior generación mediante el uso de la técnica conocida como SPOOL (Simultaneous Peripheral Operation On Line). Esta técnica consiste en que los trabajos se entregan al sistema en tarjetas o en cinta magnética, siendo leídos por un proceso auxiliar que los pasa a disco. Una vez en disco, estos se ejecutan, produciendo su salida en un archivo en disco, que es a continuación volcado a impresora por otro proceso

auxiliar. Nótese que el trabajo realizado por los equipos auxiliares en la anterior generación, es ahora realizado por procesos.

- Sistemas interactivos en tiempo compartido: Si bien el acceso interactivo tuvo su origen a finales de la anterior generación, la aparición de la multiprogramación hace posible que una sola máquina pueda proporcionar acceso interactivo a múltiples usuarios (mediante múltiples terminales conectados a ella) de forma simultánea. Esto es posible mediante las técnicas de tiempo compartido, que reparten el tiempo de CPU entre los procesos de los distintos usuarios, aún cuando estos procesos no efectúen operaciones bloqueantes.

Sobre este último punto, cabe destacar la creación del sistema operativo MULTICS (MULTIplexed Information and Computing Service). Este fue un proyecto conjunto llevado a cabo por el MIT (Massachusetts Institute of Technology), Bell Laboratories y General Electric. La idea era distribuir servicio de computación a toda una ciudad, de la misma forma que se distribuye la energía eléctrica... usando como soporte una máquina con dos procesadores y 256K palabras de 36 bits (algo más de 1Mb). MULTICS no consiguió sus objetivos, aunque se ha llegado a usar en distintas organizaciones (especialmente universidades). Sí que es importante destacar que uno de los científicos de Bell Laboratories, Ken Thompson, reescribió una versión monousuario de MULTICS a la que socarronamente llamó UNICS (de UNIplexed ☺). Posteriormente, Dennis Ritchie lo reescribió en el lenguaje C que acababa de desarrollar: había nacido UNIX.

1.2.3.4 Necesidades a satisfacer por parte del sistema operativo

La aparición de la multiprogramación introduce un número considerable de complejidades en el sistema operativo:

- El tiempo de CPU ha de repartirse, de acuerdo con determinados criterios, entre todos los procesos en ejecución. Aparece así el concepto de planificación. Este reparto puede ser especialmente delicado en los sistemas interactivos, pues todos los usuarios deberían tener la percepción de recibir un trato justo por parte del sistema.
- Desde el momento en que varios procesos se encuentran simultáneamente en ejecución, es necesario que el sistema operativo establezca mecanismos de protección entre estos, para evitar que un proceso erróneo o malintencionado pueda afectar la correcta ejecución de otros procesos, o del propio sistema operativo. Estos mecanismos de protección afectan tanto al espacio de memoria como al espacio en disco y otros recursos del sistema.
- La gestión de recursos se complica también, pues el sistema ha de controlar qué recursos tiene asignado cada proceso, y si dicha asignación es en exclusividad o compartida con otros procesos. Aparece también el riesgo de interbloqueos.
- Los dispositivos deben ser capaces de operar requiriendo la mínima atención posible por parte de la CPU. Esto se soluciona mediante el uso de

circuitería especializada en la gestión de entrada/salida, circuitería que debe ser controlada por el sistema operativo.

- El hecho de que un sistema disponga de múltiples terminales, complica el modelo de entrada/salida, pues los procesos interactivos deben poder ejecutarse independientemente del terminal con el que operan.

1.2.4 Cuarta generación: 1971-1984

1.2.4.1 Innovaciones tecnológicas

En esta ocasión, no hay una innovación tecnológica concreta que claramente actúe como precursora de esta nueva generación. Se considera que esta generación arranca con la aparición de los primeros circuitos integrados con tecnología LSI (Large Scale Integration, conteniendo miles de transistores) y posteriormente VLSI (Very Large Scale Integration, conteniendo cientos de miles de transistores). La característica fundamental es que, por primera vez, el procesador suele ocupar un único circuito integrado. Las memorias de ferrita, por su parte, quedan definitivamente reemplazadas por las memorias en circuito integrado.

Si bien no son una innovación específica de esta generación, se hace habitual el uso de redes locales (en 1973 Xerox presenta Ethernet). En 1984 nace Internet.

1.2.4.2 Algunas máquinas relevantes

La industria informática está plenamente consolidada, y se construye gran cantidad de ordenadores, llegando estos a convertirse incluso en artículos de consumo a principios de los ochenta, gracias a su espectacular abaratamiento. En líneas generales, cabe destacar:

- El concepto de familia de ordenadores está plenamente consolidado. La mayoría de los ordenadores construidos se integran en alguna familia.
- Los miniordenadores alcanzan e incluso superan prestaciones a los ordenadores corporativos de la etapa anterior.
- Aparecen los procesadores en cadena, o procesadores pipeline, capaces de comenzar la ejecución de una instrucción antes de terminar las anteriores.
- Hacen también su aparición los procesadores vectoriales, capaces de trabajar con vectores numéricos de la misma forma que los procesadores normales (escalares) trabajan con números.
- Si bien ya se habían construido con anterioridad ordenadores con varios procesadores, se hace frecuente que los supercomputadores dispongan de varios procesadores.
- Aparece el ordenador personal, o microordenador, cuyo propósito es dar servicio a una sola persona. Aparecen incluso los primeros ordenadores domésticos, a principios de los ochenta.

1.2.4.3 Explotación

En líneas generales, se mantienen los métodos de explotación de la anterior generación expuestos en el epígrafe 1.2.3.3, si bien en la mayoría de los casos en los sistemas no interactivos todo el proceso de SPOOL se hace sobre disco, desapareciendo las tarjetas perforadas y relegando las cintas magnéticas a un segundo plano, especialmente como dispositivos de copia de seguridad o de almacenamiento masivo. Con respecto a los sistemas interactivos, aparecen las primeras interfaces gráficas de usuario (GUI) a finales de esta etapa.

Con respecto a los microordenadores, estos son siempre interactivos, y mono-programados en la mayoría de los casos, pues su propósito es dar servicio a un único usuario.

1.2.4.4 Necesidades a satisfacer por parte del sistema operativo

Sobre las necesidades planteadas en la etapa anterior (epígrafe 1.2.3.4) se añaden las siguientes:

- Proporcionar soporte de red, proporcionando acceso a dispositivos remotos, como impresoras, sistemas de archivos, etcétera.
- En el caso de los multiprocesadores, los métodos de planificación han de darle soporte adecuado.
- Especialmente en el caso de los ordenadores personales, la facilidad de uso se convierte en un factor crítico, pues la mayoría de sus usuarios no ni de lejos son especialistas en informática.

1.2.5 Quinta generación: 1984 en adelante

1.2.5.1 Innovaciones tecnológicas

En esta ocasión, nuevamente no hay ninguna innovación concreta que actúe nuevamente como precursora de esta generación. De hecho, hay autores [Tanenbaum03] que ni siquiera considera esta generación como tal. Se podría considerar como característica más representativa la construcción de máquinas con un elevado número de procesadores, llegando a construir máquinas con cientos de procesadores.

Por su parte, se consolida el uso de Internet, llegando un elevado porcentaje de la población a tener acceso a la red², convirtiéndose esta en un bien de consumo y en un espacio comercial de primer orden.

² Según cifras oficiales, entre los años 2002 y 2004 se ha pasado de un 20% a un 37% de la población española con acceso a la red. No obstante, estas cifras están aún por debajo de la media europea, en la que en el mismo periodo se ha pasado del 40% al 50%.

1.2.5.2 Algunas máquinas relevantes

La principal característica de esta etapa es la drástica reducción de precios, a la vez que las prestaciones aumentan de forma inversa:

- En el segmento de mercado de los ordenadores personales, estos alcanzan en prestaciones a los mainframes de hace algunos años, mientras que sus reducidos precios los convierten en un bien de consumo doméstico.
- El segmento de los supercomputadores se ve dominado por los grandes multiprocesadores. Por ejemplo, el Cray XT4 es un sistema escalable (que permite al usuario añadir tantos procesadores como necesite para a sus necesidades), pudiendo gestionar más de 30.000 procesadores. Por ejemplo, con la configuración máxima de 30.508 procesadores AMD Opteron Dual Core a 2.6 GHz, proporciona según el fabricante unas prestaciones máximas de 318 TFLOPS³ (la versión Cray XT3, dos años antes, sólo declaraba 147 TFLOPS), pudiendo manejar hasta 239 TBytes de memoria.

1.2.5.3 Explotación

Se siguen manteniendo lo expuesto en el epígrafe 1.2.4.3 de la anterior generación, si bien los sistemas interactivos hacen uso casi siempre de interfaz gráfica de usuario, que tiende a convertirse en la forma de uso interactiva por excelencia.

Gracias a la conectividad mundial proporcionada por Internet, los sistemas distribuidos toman también una especial relevancia. Se lleva a cabo un gran esfuerzo de investigación por desarrollar sistemas que hagan transparente el acceso a recursos distribuidos tanto hardware como software.

Por su parte, los ordenadores personales se vuelven multiprogramados.

1.2.5.4 Necesidades a satisfacer por parte del sistema operativo

A lo expuesto en el epígrafe 1.2.4.4 de la anterior etapa, habría que añadir:

- En el campo de los multiprocesadores, la gestión y planificación de estos ha de contemplar la posibilidad de escalabilidad, así como la tolerancia a fallos (capacidad del sistema para seguir funcionando en caso de que alguno de los procesadores se averíe).
- En el campo de los sistemas distribuidos, el principal objetivo es hacer transparente al usuario el hecho de que el sistema con el que trabaja no es un sistema único sino que es un sistema distribuido a través de una red de comunicaciones.
- En el campo de los ordenadores personales, la facilidad de uso y la “amigabilidad” de la interfaz se convierte definitivamente en norma sagrada.

³ 1 TFLOP= 1 billón de operaciones en coma flotante por segundo.

Los sistemas operativos modernos son capaces de autoconfigurarse en respuesta a cambios en el hardware. En algunos casos, se permite incluso la conexión de periféricos en caliente (sin necesidad de desconectar la máquina).

1.3 Tipos de Sistemas Operativos

Realizar una clasificación de los distintos tipos de sistemas operativos no es fácil, pues esta clasificación se puede realizar atendiendo a diversos criterios. En nuestra opinión, la clasificación propuesta en [Tanenbaum03] refleja satisfactoriamente la realidad actual. Según este texto, los sistemas operativos se pueden clasificar en las siguientes categorías no excluyentes entre sí:

- Sistemas operativos de mainframe
- Sistemas operativos de servidor
- Sistemas operativos multiprocesador
- Sistemas operativos de ordenador personal
- Sistemas operativos en tiempo real
- Sistemas operativos integrados
- Sistemas operativos de tarjeta inteligente

A continuación se detallan cada una de ellas.

1.3.1 Sistemas operativos de mainframe

En el extremo superior de la clasificación están los sistemas operativos para los mainframes. Estas máquinas se caracterizan por su capacidad de E/S, sin resultar raro sistemas con 1000 discos y miles de gigabytes. Los mainframes también están renaciendo un poco como servidores Web avanzados, servidores para sitios de comercio electrónico a gran escala y servidores para transacciones de negocio a negocio.

Los sistemas operativos para mainframe están claramente orientados al procesamiento de varios trabajos a la vez, casi todos los cuales necesitan cantidades enormes de E/S. Los servicios que ofrecen suelen ser de tres tipos:

- **Por lotes:** Un sistema por lotes procesa trabajos rutinarios sin que haya un usuario interactivo presente. Por ejemplo, el procesamiento de reclamaciones en una compañía de seguros o los informes de ventas de una cadena de tiendas por lo regular se hacen por lotes.
- **Procesamiento de transacciones:** Los sistemas procesadores de transacciones manejan numerosas solicitudes pequeñas, como procesamiento de cheques en un banco o reservaciones de pasajes aéreos. Cada unidad de trabajo es pequeña, pero el sistema debe manejar cientos o miles de ellas por segundo.

- **Tiempo compartido:** Los sistemas de tiempo compartido permiten a múltiples usuarios remotos ejecutar trabajos en la computadora de forma simultánea, como las consultas de una base de datos grande.

Estas funciones están íntimamente relacionadas, por lo que muchos sistemas operativos para mainframe las realizan todas. Un ejemplo de sistema operativo para mainframe son z/OS y OS/390 de IBM.

1.3.2 Sistemas operativos de servidor

Un nivel por debajo en la clasificación están los sistemas operativos de servidor. Éstos se ejecutan en servidores, que son computadoras personales muy grandes, estaciones de trabajo o incluso mainframes, y dan servicio a múltiples usuarios a través de una red, permitiéndoles compartir recursos de hardware y software. Los servidores pueden prestar servicios de impresión, de archivo o de Web. Los proveedores de Internet operan muchas máquinas servidoras para atender a sus clientes, y los sitios Web emplean servidores para almacenar las páginas Web y manejar las solicitudes que llegan. Entre los sistemas operativos de servidor típicos están UNIX y Windows 2000 y 2003 en sus versiones server y advanced server. Linux también está ganando terreno en los servidores.

1.3.3 Sistemas operativos multiprocesador

Una forma cada vez más común de obtener potencia de computación a gran escala es dotar de varias CPUs a los sistemas. Se necesitan sistemas operativos especiales, pero con frecuencia éstos son variaciones de los sistemas operativos de servidor, con funciones especiales para comunicación y conectividad.

1.3.4 Sistemas operativos de computadora personal

La siguiente categoría es el sistema operativo de computadora personal. Su misión consiste en presentar una buena interfaz a un solo usuario. Se les usa en forma amplia para procesamiento de texto, hojas de cálculo y acceso a Internet. Ejemplos comunes son las distintas versiones personales de Windows, el sistema operativo Macintosh y Linux.

1.3.5 Sistemas operativos de tiempo real

Otro tipo de sistema operativo es el sistema de tiempo real. Se caracteriza porque debe garantizar unos tiempos de respuesta suficiente exactos ante estímulos externos. Por ejemplo, en los sistemas de control de procesos industriales, computadoras de tiempo real tienen que capturar datos acerca del proceso de producción y utilizarlos para controlar las máquinas de la fábrica. Con frecuencia deben cumplirse plazos estrictos. Por ejemplo, si un automóvil avanza en una línea de ensamble, deben efectuarse ciertas acciones en ciertos instantes. Si un robot soldador suelda demasiado pronto o demasiado tarde, el automóvil quedará arruinado.

VxWorks y QNX Neutrino son sistemas operativos de tiempo real muy conocidos.

1.3.6 Sistemas operativos integrados

Pasando a sistemas cada vez más pequeños, llegamos a las computadoras de bolsillo (palm-top) y sistemas integrados (también conocidos como empotrados). Una computadora de bolsillo o Asistente Personal Digital (PDA= Personal Digital Assistant) es una computadora pequeña que cabe en el bolsillo de la camisa y realiza unas cuantas funciones como libreta de direcciones electrónica y bloc de notas. Los sistemas integrados operan en las computadoras que controlan dispositivos que por lo general no se consideran computadoras, como televisores, hornos de microondas y teléfonos móviles. Estos sistemas suelen tener algunas características de los sistemas de tiempo real pero también tienen limitaciones de tamaño, memoria y sobre todo consumo de electricidad que los hacen especiales. Algunos ejemplos de tales sistemas operativos son PalmOS y Windows CE (Consumer Electronics).

1.3.7 Sistemas operativos de tarjeta inteligente

Los sistemas operativos más pequeños se ejecutan en tarjetas inteligentes, que son dispositivos del tamaño de una tarjeta de crédito que contienen un chip de CPU. Sus limitaciones en cuanto a potencia de procesamiento y memoria son muy severas. Algunos de ellos sólo pueden desempeñar una función, digamos pagos electrónicos, pero otros pueden realizar varias funciones en la misma tarjeta. Muchos de éstos son sistemas patentados.

Algunas tarjetas inteligentes están orientadas hacia Java. Esto implica que la ROM de la tarjeta inteligente contiene un intérprete de la Máquina Virtual de Java. Los applets (pequeños programas) de Java se descargan a la tarjeta y el intérprete de la misma los procesa. Algunas de estas tarjetas pueden manejar varios applets al mismo tiempo, lo que da pie a multiprogramación y a la necesidad de planificación. La administración y protección de recursos también adquieren importancia cuando dos o más applets están presentes al mismo tiempo. El sistema operativo (por lo regular muy elemental) de la tarjeta debe resolver estas cuestiones.

1.4 Para ampliar conocimientos

En el libro [Silberschatz99], en la página 6 se propone una clasificación de los sistemas operativos ligeramente distinta a la expuesta en [Tanenbaum03], expuesta en el epígrafe 1.3, y que es la que preferimos nosotros, por ser más moderna y atender al propósito del sistema operativo, más que a sus características. Considere simplemente dicha clasificación como una clasificación alternativa atendiendo a las características intrínsecas del sistema. No obstante, en las notas bibliográficas que encontrará en la página 21 [Silberschatz99] encontrará un importante filón de información con el que ampliar los conocimientos expuestos en este tema.