UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CONTADURÍA Y ADMINISTRACIÓN

AUTOR: L. A. SALVADOR MEZA BADILLO

Sistemas Operativos Multiusuario		Clave: 1268
Plan: 2005		Créditos: 8
Licenciatura: Informática		Semestre: 2°
Área: Informática (Redes y Telecomunicaciones)		Hrs. Asesoría: 2
Requisitos: Ninguna		Hrs. Por semana: 4
Tipo de asignatura:	Obligatoria (x)	Optativa ( )

Objetivo general de la asignatura

Al finalizar el curso, el alumno conocerá los fundamentos de diseño y funcionamiento de un sistema operativo multiusuario, y será capaz de explotar sus servicios.

Temario oficial (64 horas sugeridas)

1. Definición de los conceptos fundamentales (10 horas)
2. Proceso (8 horas)
3. Sincronización y comunicación entre procesos (8 horas)
4. Administración de la memoria (8 horas)
5. Administración de archivos (8 horas)
6. Seguridad (8 horas)
7. Implantación de sistemas operativos (8 horas)
8. Tópicos avanzados de sistemas operativos (6 horas)

Introducción

Sin el software, un equipo de cómputo no es más que un conjunto de dispositivos físicos sin ninguna utilidad, con el software un equipo puede procesar, almacenar, manipular información y realizar diversas actividades para beneficio de las personas e instituciones.

El software para computadoras se clasifica de manera general en dos clases: los **programas de sistema** que controlan la operación de la computadora y los **programas de aplicación**, que resuelven los problemas para los usuarios.

El programa principal de todo el software de una computadora es el sistema operativo, que administra todos los recursos de la computadora y proporciona la base sobre la cual pueden escribirse los programas de aplicación, existen diferentes categorías del sistema operativo; multitareas, monotareas, monousuario, por lotes, en tiempo real, tiempo compartido y multiusuarios.

En los sistemas operativos modernos, la idea de multiusuario guarda el significado original de que éste puede utilizarse por varios usuarios al mismo tiempo, permitiendo la ejecución concurrente de los programas de aplicación, las computadoras modernas utilizan múltiples procesadores y proveen las interfaces de usuario a través de una red de computadoras e inclusive un grupo de computadoras pueden formar un **cluster** (agrupamiento de equipos) logrando altas capacidades de cómputo.

Sistema operativo multiusuario se ha dividido en ocho temas. En el primero se estudian los conceptos fundamentales aplicados a los sistemas operativos multiusuario, en el segundo y tercer tema se estudian los aspectos más relevantes relacionados con los procesos, para entender el funcionamiento de un sistema operativo.

En el cuarto y quinto tema se estudia la administración de un recurso muy importante “la memoria”, y el sistema de archivos, que en conjunto proporcionan el mecanismo para el almacenamiento y el acceso a los datos y programas.

En el sexto tema se describen los conceptos y mecanismos que existen para la protección y seguridad de los sistemas operativos.

En el séptimo y octavo tema se describen las principales técnicas que se utilizan para la implantación de un sistema operativo.

TEMA 1. DEFINICIÓN DE LOS CONCEPTOS FUNDAMENTALES

Objetivo particular

Al culminar el aprendizaje de este tema, el alumno identificará los conceptos más importantes aplicados a los sistemas operativos multiusuario.

Temario detallado

1.1Definición de Sistema Operativo Multiusuario

1.2 Funciones de los Sistemas operativos Multiusuario

Introducción

El núcleo fundamental de una computadora es su sistema operativo, este controla el hardware, carga las aplicaciones en la memoria, ejecuta esas aplicaciones y maneja los dispositivos y periféricos como los discos e impresoras. El objetivo principal de un sistema operativo es hacer que un sistema de cómputo pueda utilizarse de manera cómoda y eficiente.

1.1 Definición de Sistema Operativo Multiusuario

Un sistema de cómputo puede dividirse en cuatro componentes: **el hardware, el sistema operativo, los programas de aplicación y los usuarios.** El sistema operativo es una parte importante de casi todo sistema de cómputo.

Hardware: Máquina y equipo asociados con dispositivos de cómputo; tales como la unidad central de proceso (CPU), memoria, dispositivos periféricos, etc.

Programas de aplicación: Se denomina así al tipo de software que se utiliza para resolver los problemas de cómputo de los usuarios y son los procesadores de texto, hojas de cálculo, manejadores de bases de datos, navegadores de red, etc.

Usuarios: Estos pueden ser equipos de cómputo y usuarios que requieran resolver diversos problemas.

Sistema operativo: Conjunto de programas fundamentales que controlan y coordinan el hardware y los programas de aplicación de los usuarios.

El sistema operativo proporciona los medios para el uso apropiado de los recursos en la operación del sistema de cómputo. Al igual que el gobierno, el sistema operativo por sí mismo no realiza alguna función útil. Simplemente proporciona un ambiente dentro del cual otros programas pueden realizar un trabajo útil.

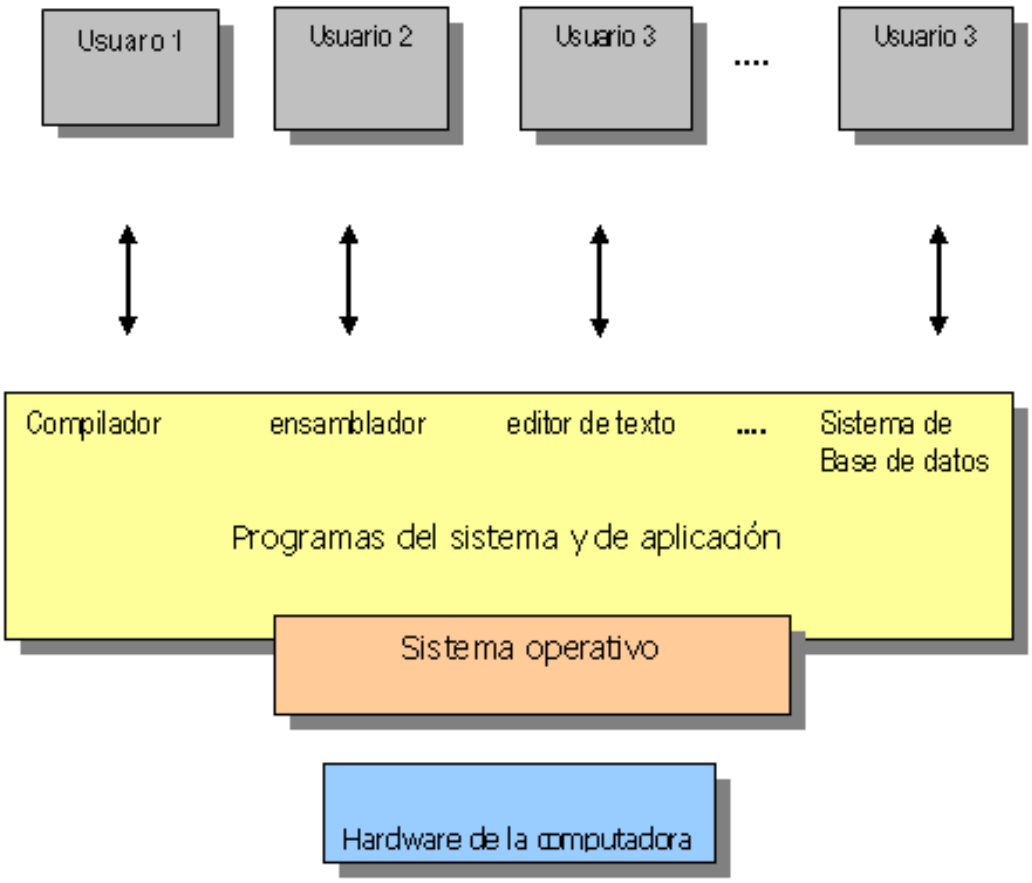


Figura 1. 1. Vista abstracta de los componentes de un sistema de cómputo[1]

Un sistema de cómputo tiene muchos recursos (hardware y software) que se requieren para resolver problemas: tiempo de procesamiento (CPU), espacio de memoria, espacio de almacenamiento para datos, dispositivos de entrada-salida (E/S), etc. El sistema operativo actúa como el administrador de estos recursos y los asigna a programas y usuarios. Debido a que puede haber muchas solicitudes de estos recursos, el sistema operativo debe decidir a qué solicitudes les asignará recursos, de manera que el sistema de cómputo pueda operar de manera eficiente y sin causar conflictos.

[1] Abraham Silberschatz, *Sistemas Operativos*, 6ª ed., México, Limusa Wisley, 2002, p. 4.

No existe una definición completamente adecuada del sistema operativo. Los sistemas operativos existen debido a que son una forma razonable de resolver el problema de crear un sistema de cómputo utilizable. El objetivo fundamental de los sistemas de cómputo es ejecutar los programas del usuario y facilitar la solución de sus problemas.[2]

Es más fácil definir un sistema operativo por lo que hace que por lo que es. El objetivo principal de un sistema operativo es la comodidad para el usuario. Los sistemas operativos existen debido a que se supone que las tareas de cómputo son más sencillas con ellos que sin ellos.

1.2. Funciones de los Sistemas Operativos Multiusuario

Las funciones clásicas del sistema operativo se agrupan en tres áreas

- Gestión de los recursos de la computadora.
- Ejecución de servicios para los programas.
- Ejecución de los mandatos de los usuarios.

- **La gestión de los recursos.-** En una computadora coexisten varios programas de uno o más usuarios que se ejecutan de manera simultánea. Estos programas compiten por los recursos del equipo, y es el sistema operativo el encargado de administrar su asignación y uso. El sistema operativo debe de garantizar la protección de los programas frente a otros y suministrar información sobre el uso de los recursos, es decir, **asigna recursos, ofrece protección entre los usuarios del sistema y lleva la contabilidad sobre el uso de recursos.**[3]

  La ejecución de servicios.- El sistema operativo ofrece a los programas un conjunto de servicios, o llamadas al sistema, que pueden solicitar cuando lo necesiten proporcionando a los programas una visión de máquina extendida.[4]

---

[2] Véase, Heriberto Gabriel Soto: "Sistemas operativos", Monografías, material en línea, disponible en: http://www.monografias.com/trabajos11/oper/oper.shtml, recuperado el 13/01/09.
[3] Slideshare (softsau), "Sistemas operativos", material en línea, diapositiva 3, disponible en: http://www.slideshare.net/softesau/sistemas-operativos-171331/, recuperado el 05/12/08.
[4] Véase material en línea, disponible en: *www.itescam.edu.mx/principal/sylabus/fpdb/recursos/r2305.DOC*, recuperado el 13/01/09.

La ejecución de mandatos.- El módulo del sistema operativo que permite que los usuarios dialoguen de forma interactiva con el sistema operativo es el interprete de comandos conocido como Shell.

A continuación se hace una breve descripción del desarrollo que han tenido los sistemas operativos a lo largo de los últimos 35 años.

Sistemas por lotes

Las primeras computadoras eran grandes máquinas que se operaban desde una consola. Los dispositivos de entrada comunes eran lectores de tarjetas y unidades de cinta, los dispositivos de salida eran impresoras de línea, unidades de cinta y perforadoras de tarjetas. El usuario no interactuaba directamente con este tipo de sistemas de cómputo; más bien, preparaba un trabajo, que consistía en el programa, los datos y la información de control acerca de la naturaleza de trabajo (tarjetas de control) y lo entregaba al operador de la computadora que los organizaba en lotes para su ejecución en el *mainframe*. Sus características más importantes fueron:

- El operador recoge estos programas y los organiza en lotes para su ejecución en el *mainframe*.
- El operador carga el sistema operativo en la memoria mediante el lector de tarjetas.
- El sistema operativo carga y ejecuta cada programa del lote.
- El sistema operativo es muy simple, su único objetivo es cargar y ejecutar cada programa de un lote.
- Los resultados de los programas se generan en papel utilizando una impresora.
- Los programadores implementan sus programas utilizando tarjetas perforadas.

En este tipo de sistemas la unidad central de proceso (CPU) con frecuencia quedaba ociosa, debido a que las velocidades de los dispositivos mecánicos de entada/salida (E/S) son menores que las de los dispositivos electrónicos. Incluso una CPU lenta trabaja en el rango de microsegundos, con miles de instrucciones ejecutadas por segundo, mientras que una lectora de tarjetas rápida leía 1200 tarjetas por minuto (20 tarjetas por segundo).

La introducción de la tecnología de discos permitió al sistema operativo mantener todos los trabajos en un disco, en vez de un lector de tarjetas en serie. Con el acceso directo a varios trabajos, podría realizar una planificación de los trabajos para usar los recursos y ejecutar tareas de manera eficiente.

Sistemas de tiempo compartido

Los sistemas por lotes con multiprogramación proporcionaron un ambiente en el que los diversos recursos del sistema (CPU, memoria, dispositivos periféricos) se utilizaban eficazmente, [5] pero no ofrecían la interacción del usuario con el sistema de cómputo. El tiempo compartido, o multitareas, es una extensión lógica de la multiprogramación. La unidad central de proceso (CPU) ejecuta múltiples trabajos conmutando entre ellos, pero los cambios ocurren de manera tan frecuente que los usuarios pueden interactuar con cada programa mientras está en ejecución. Sus características principales son:

- Cada usuario interactúa con el sistema operativo mediante una terminal
- Son sistemas multiusuario por la capacidad de atender a varios usuarios simultáneamente.
- El sistema mantiene múltiples programas en la memoria y va repartiendo el uso de CPU entre todos ellos.

[5] Véase, Departamento de Electrónica e Informática, Universidad Católica, Paraguay: “Sistema de tiempo compartido”, material en línea, disponible en: http://www.dei.uc.edu.py/tai2003-2/sistemas.operativos/Tiempo%20Compartido.htm, recuperado el 08/12/08.

- Se asigna a los programas un tiempo máximo de ejecución.

Un sistema operativo de tiempo compartido hace uso de la planificación del uso de la unidad central de proceso (CPU) y la multiprogramación para proporcionar a cada usuario una pequeña porción de una computadora de tiempo compartido. Cada usuario tiene por lo menos un programa distinto en la memoria. Un programa que se carga en la memoria y se está ejecutando se le conoce como proceso. Cuando se ejecuta un proceso, éste lo hace sólo por un tiempo breve antes de que termine o necesite realizar operaciones de E/S. Las operaciones E/S pueden ser interactivas, es decir, la entrada es desde el teclado, el ratón u otro dispositivo y la salida es a una terminal.

Los sistemas operativos de tiempo compartido deben administrar y proteger el uso de la memoria para mantener la sincronía de los trabajos y obtener tiempos de respuesta razonables'. Esto se logra mediante el intercambio de memoria principal y el disco, así como la utilización de la memoria virtual que permite la ejecución de un trabajo que puede no estar cargado completamente en la memoria.

El sistema de archivos reside en una colección de discos; por lo tanto, se debe proporcionar una administración de discos. Asimismo, los sistemas de tiempo compartido proporcionan un mecanismo para ejecución concurrente, la cual requiere esquemas complejos de planificación de la unidad central de proceso CPU. Para asegurar una ejecución ordenada, el sistema debe contemplar mecanismo para la sincronización y comunicación de los trabajos, y asegurar que éstos no se atasquen por un bloqueo mutuo, esperando indefinidamente uno a otro.

Sistemas para computadoras personales

Las computadoras personales (*Personal Computer* o PC) aparecieron en la década de los años 70 del siglo pasado, y se referían a las microcomputadoras compatibles con las especificaciones de la empresa IBM. Este tipo de equipos fueron diseñados para ser utilizados por una persona a la vez, carecían de las características necesarias para proteger a un sistema operativo de los programas del usuario; los sistemas operativos para PC, por lo tanto, no eran ni multiusuario ni multitarea. Sin embargo, las metas de estos sistemas operativos han cambiado con el tiempo; en lugar de maximizar la utilización de la CPU y los dispositivos periféricos, los sistemas optan por maximizar la comodidad y grado de respuesta para el usuario, actualmente los sistemas operativos que utilizan los equipos PC son; Microsoft Windows, Macintosh de Apple y Linux.

Los sistemas operativos para estas computadoras se han beneficiado en varias formas con el desarrollo de los utilizados para computadoras grandes (*mainframes*). Las microcomputadoras fueron capaces de adoptar de inmediato parte de la tecnología desarrollada para los sistemas operativos más grandes

Sistemas Paralelos

Los sistemas paralelos son aquellos que tienen más de un procesador (CPU) y están fuertemente acoplados compartiendo el bus, el reloj y en ocasiones la memoria y los dispositivos periféricos, lo que les permite tener una gran capacidad de realizar varias operaciones de manera simultánea y manejar grandes volúmenes de información del orden de los terabytes. Sus características más importantes son:

- Mayor rendimiento.
- Mayor disputa por los recursos compartidos.
- Mayor trabajo en menos tiempo.

- Los procesadores pueden compartir periféricos, almacenamiento masivo y suministro de energía.
- Alta confiabilidad, la falla de un procesador no detendrá el sistema.

Los sistemas de procesadores múltiples utilizan dos tipos de multiprocesamiento; el simétrico que se refiere a que cada uno de los procesadores ejecuta una copia idéntica del sistema operativo y estas copias se comunican entre ellas según se requiera, y el asimétrico, en el que a cada procesador se le asigna una tarea específica. Un procesador maestro controla el sistema, otro procesador solicita instrucciones al maestro para realizar tareas bien definidas. Este esquema se conoce como una relación maestro-esclavo. El procesador maestro programa y asigna el trabajo a los procesadores esclavos. En el multiprocesamiento simétrico (SMP) todos los procesadores están al mismo nivel; no existe una relación maestro-esclavo entre ellos. Cada procesador ejecuta de manera concurrente una copia del sistema operativo.

La diferencia entre el multiprocesamiento simétrico y el asimétrico puede ser el resultado del hardware o del software que se utilice. Un hardware especial puede diferenciar a los multiprocesadores múltiples, o puede escribirse el software de manera que sólo se permita un maestro y múltiples esclavos. Por ejemplo, el sistema operativo SunOs Versión 4 de Sun proporciona multiprocesamiento asimétrico, en tanto que la versión 5 (Solaris 2) es simétrica en el mismo hardware.

Sistemas de Tiempo Real

El sistema operativo de tiempo real es aquel que se caracteriza porque su parámetro clave es el tiempo y ha sido desarrollado para aplicaciones que requieren ser ejecutadas bajo ciertas restricciones de tiempo sobre la operación

de un procesador o flujo de datos y en ocasiones se emplea como dispositivo de control en aplicaciones dedicadas. Por ejemplo, en los sistemas de control de procesos industriales. Los sistemas de control de experimentos científicos, de imágenes médicas, de control industrial y ciertos sistemas de despliegue son sistemas de tiempo real. Un sistema de tiempo real tiene restricciones de tiempo bien definidas y fijas. El procesamiento debe realizarse dentro de los límites definidos, o el sistema fallará. Por ejemplo, no funcionará si a un brazo de robot se le programa para detenerse después de haber chocado con el auto que está construyendo. Un sistema de tiempo real se considera que funciona correctamente sólo si entrega el resultado correcto dentro de las restricciones de tiempo establecidos.

Hay dos clases de sistemas de tiempo real; El sistema **riguroso** que garantiza que las tareas críticas se terminen a tiempo, es decir, es indispensable que la acción se efectué en cierto momento o intervalo. Otro sistema de tiempo real es el **no riguroso**, en el que es aceptable no cumplir estrictamente con el plazo programado. Los sistemas de audio digital, multimedia y realidad virtual pertenecen a este tipo de sistemas. Sus características más importantes son:

- Tiene restricciones de tiempo bien definidas.
- Se utilizan para aplicaciones integrales.
- No utilizan mucha memoria.
- Son sistemas Multi-arquitectura (puertos de código para otras CPU).

Sistemas distribuidos

El desarrollo tecnológico de los procesadores, el crecimiento de las redes de área local LAN y de las telecomunicaciones permitieron conectar computadoras para la transferencia de datos a alta velocidad. Esto dio origen al concepto de “Sistemas distribuidos” y que tiene como ámbito el estudio de redes como por ejemplo:

Internet, redes corporativas, redes de teléfonos móviles, etc. Hoy día todas las computadoras personales PC y estaciones de trabajo modernas son capaces de ejecutar un navegador de red para tener acceso a documentos de hipertexto dentro de la red. Los sistemas operativos actuales como Windows, Linux, etc. incluyen el software del protocolo (TCP/IP y PPP) que permite a la computadora tener acceso a Internet mediante una red de área local o una conexión telefónica.

Las redes de computadoras usadas en este tipo de aplicaciones están compuestas de un conjunto de procesadores que no comparten memoria o un reloj. En su lugar, cada procesador tiene su propia memoria local. Los procesadores se comunican entre ellos mediante varias líneas de comunicación, como buses de alta velocidad o líneas telefónicas estos sistemas generalmente se conocen como sistemas débilmente acoplados o sistemas distribuidos.[6] Sus características más importantes son:

- Concurrencia (los recursos en la red pueden ser usados simultáneamente).
- Carencia de reloj global (la realización de una tarea es distribuida a los componentes).
- Fallas independientes (si un componente falla, los demás siguen funcionando).

Bibliografía del tema 1

Carretero, Jesús, *Sistemas Operativos,* Madrid, McGraw Hill, 2001.

Silbertschatz, Abraham. *Sistemas Operativos*. 6ª ed., México, Limusa Wisley, 2002.

[6] Véase, Universidad de las Palmas de Gran Canaria, Facultad de Informática: “Sistemas operativos, Soluciones, examen parcial, 29/04/06”, material en línea, disponible en: http://sopa.dis.ulpgc.es/so/examenes/2006/soluciones-20060429-primer_parcial.pdf, recuperado el 08/12/08.

Actividades de aprendizaje:

A.1.1. Elabora en un documento lo siguiente:

- Describe la función de los cuatro componentes de un sistema de cómputo.
- ¿Cuáles son las tres funciones de un sistema operativo?
- ¿Cuáles son las principales diferencias entre los sistemas por lotes y los de tiempo compartido?
- Describe las diferencias del multiprocesamiento simétrico y asimétrico.
- Con tus propias palabras define qué es un sistema operativo.

A.1.2. Elabora un cuadro sinóptico sobre las características más importantes de los tipos de sistemas operativos tratados en este documento.

A.1.3 Describe en un documento las ventajas de utilizar sistemas distribuidos.

Cuestionario de autoevaluación

1.- ¿Cuál es el objetivo principal de un sistema operativo?

2.- ¿Cuál es la función de la gestión de recursos en el sistema operativo?

3.- Es la función de la ejecución de servicios en el sistema operativo.

4.- ¿Cuál es la función de la ejecución de mandatos en el sistema operativo?

5.- Menciona tres ejemplos de un programa de aplicación

6.- ¿Cuál es la diferencia entre el multiprocesamiento simétrico y asimétrico?

7.- Describe tres características del sistema de tiempo compartido

8.- Describe la función del sistema riguroso en tiempo real y de dos ejemplos de su aplicación.

9.- Describe tres características de los sistemas distribuidos.

10.- ¿Cuál es la ventaja principal de la multiprogramación?

Examen de autoevaluación:

1. ¿Cuáles son componentes de un sistema de cómputo?

    a) hardware, programas de aplicación y usuario, sistema operativo.

    b) compiladores, traductores, hardware

    c) memoria, archivos, sistema operativo

2. El módulo del sistema operativo que permite que los usuarios dialoguen de forma interactiva con el sistema operativo es:

    a) unidad central de proceso (CPU).

    b) programas de usuario.

    c) interprete de comandos (Shell).

3. ¿Cuál es la función del sistema operativo que proporciona protección entre los usuarios?

    a) ejecución de servicios.

    b) gestión de recursos.

    c) ejecución de mandatos.

4. ¿Cuál es el tipo de sistema operativo que mantiene múltiples programas en la memoria y va repartiendo el uso de CPU entre todos ellos?

    a) sistemas de tiempo real.

    b) sistemas de tiempo compartido.

    c) sistemas distribuidos.

5. ¿A qué se refiere el multiprocesamiento simétrico?

    a) cada procesador ejecuta una copia idéntica del sistema operativo.

    b) a cada procesador se le asigna una tarea específica.

    c) a cada procesador se le asignan diferentes tareas.

6. ¿Cuál es la característica del sistema operativo de tiempo real?

a) mayor rendimiento.

b) mayor trabajo en menos tiempo.

c) tiene restricciones de tiempo bien definidas.

7. ¿Cuál es la característica del sistema operativo de tiempo compartido?

a) se utilizan para aplicaciones integrales.

b) el usuario interactúa mediante una terminal.

c) no utilizan mucha memoria.

8. ¿Cuál es la característica del sistema operativo por lotes?

a) el sistema operativo es muy complejo.

b) el sistema operativo es muy simple.

c) el sistema operativo es multi-arquitectura.

9. ¿Cuál es el objetivo principal de un sistema operativo?

a) resolver problemas.

b) administrar un sistema.

c) comodidad para el usuario.

10. ¿Cuál es el componente que proporciona los recursos básicos de cómputo?

a) software de aplicación.

b) hardware.

c) procesadores.

TEMA 2. PROCESOS

Objetivo particular

Al culminar el aprendizaje de la unidad, el alumno identificará los conceptos más importantes aplicados al diseño y construcción de los sistemas operativos.

Temario detallado

2.1 Definición

2.2 Modelos de procesos e identificadores de procesos

2.3 Jerarquía de procesos, prioridades y colas

2.4 Arrancador o despachador de procesos

Introducción

Los primeros sistemas de cómputo permitían la ejecución de un programa a la vez, las computadoras modernas de hoy día pueden ejecutar varios programas al mismo tiempo. Mientras ejecutan un programa, también pueden leer un disco, leer un dispositivo externo, abrir un navegador, mandar un archivo a la impresora, etc. Para lograr esto se requirió de mayor control y una mayor división de los diferentes programas y dio por resultado el concepto de **proceso**. A continuación se presentan los conceptos más importantes aplicados al proceso en un sistema de cómputo.

2.1 Definición

El proceso se puede definir como:

> un programa en ejecución y es la unidad de procesamiento gestionada por el sistema operativo. El sistema operativo mantiene por cada proceso una serie de estructuras de información para identificar las características de

este, así como los recursos que tienen asignados tales como: segmentos de memoria, puertos de comunicaciones, archivos abiertos, etc. El sistema operativo mantiene una tabla de procesos con todos los **bloques de control de proceso (BCP)**. Por razones de eficiencia, la tabla de procesos se construye normalmente como una estructura estática, que tiene un determinado número de BCP. El proceso no incluye información de entrada-salida E/S ya que esto está reservado al sistema operativo. [7]

2.2 Modelos de procesos e identificadores de procesos

En estos modelos todo el software que se ejecuta en una computadora está organizado en procesos secuenciales, cada proceso tiene su propia unidad central de proceso virtual, es decir, el verdadero CPU cambia en forma continua de un proceso a otro, a esta conmutación se le llama multiprogramación.

Modelo de dos estados:

Es el modelo más simple. En este modelo, un proceso puede estar ejecutándose o no. Cuando se crea un nuevo proceso, se pone en estado de *No ejecución*. En algún momento el proceso que se está ejecutando pasará al estado *No ejecución* y otro proceso se elegirá de la lista de procesos listos para ejecutar para ponerlo en estado *Ejecución*.

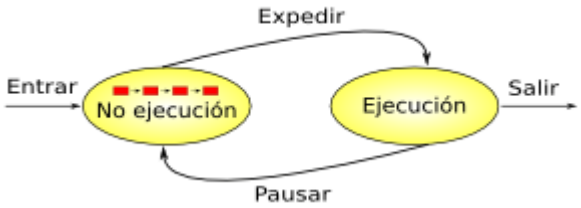


Figura 2.1 Modelo de dos estados[8]

Modelo de cinco estados:

En este modelo se necesita un estado en donde los procesos permanezcan esperando la realización de la operación de Entrada /Salida por parte del Sistema Operativo hasta que puedan proseguir. Se divide entonces al estado *No ejecución* en dos estados: *Listo* y *Espera* y se agregan además un estado *Nuevo* y otro *Terminado*.

---

[7] Jesús Carretero, *Sistemas Operativo*, Madrid, McGraw Hill, 2001, p. 78.
[8] Wikipedia: "Proceso", disponible en: http://upload.wikimedia.org/wikipedia/commons/e/e3/Procesos-2estados.png, 13/01/09.

Los cinco estados de este diagrama son los siguientes:

- **Ejecución:** el proceso está actualmente en ejecución.
- **Listo:** el proceso está listo para ser ejecutado.
- **Espera:** el proceso no puede ejecutar hasta que no se produzca cierto suceso, como la finalización de una operación de Entrada/Salida solicitada por una llamada al sistema operativo.
- **Nuevo:** El proceso fue creado recientemente y todavía no fue admitido por el sistema operativo. En general los procesos que se encuentran en este estado todavía no fueron cargados en la memoria principal.
- **Terminado:** El proceso fue expulsado del grupo de procesos ejecutables, ya sea porque terminó o por alguna falla, como un error de protección, aritmético, etc.

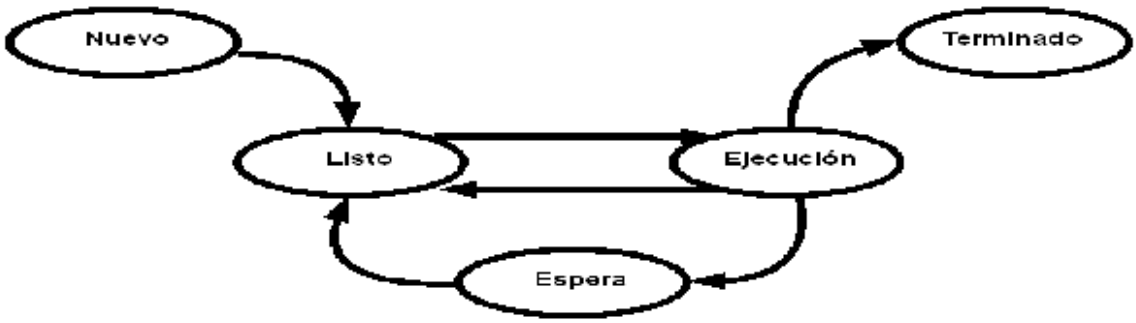


Figura 2.2 Modelo de los cinco estados[9]

También los procesos suspendidos (Hold) en el que dos o más procesos pueden cooperar mediante señales de forma que uno obliga a detenerse a los otros hasta que reciban una señal para continuar.

- Se usa una variable de tipo Semáforo para sincronizar los procesos.
- Si un proceso está esperando una señal, se suspende (Hold) hasta que la señal se envíe (SIGNAL).
- Se mantiene una cola de procesos en espera en el semáforo.
- La forma de elegir los procesos de la cola en ESPERA es mediante una política FIFO (First In First Out) también llamada FCFS (First Come First Served), Round Robin, etc.

[9] Wikipedia: “Proceso”, material en línea, disponible en: http://upload.wikimedia.org/wikipedia/commons/8/8b/Diagrama_de_estados5.PNG, 13/01/09.

> La sincronización explícita entre procesos es un caso particular del estado "bloqueado". En este caso, el suceso que permite desbloquear un proceso no es una operación de entrada/salida, sino una señal generada a propósito por el programador desde otro proceso.[10]

Identificadores de procesos

Los procesos se identifican mediante su **identificador de proceso**, un proceso nuevo se crea por la llamada al sistema fork (bifurcar) y puede tener procesos hijos, el proceso creador se le denomina proceso padre y los nuevos se les denominan procesos hijos. Los procesos nuevos pueden crear otros y formar un árbol procesos.

Un proceso nuevo se puede ejecutar de la siguiente forma:

1. El padre continúa ejecutándose de manera simultánea con sus hijos.
2. El padre espera hasta que alguno o todos sus hijos hayan terminado.

También hay dos posibilidades en términos del espacio de direcciones del nuevo proceso:

> En UNIX. El proceso nuevo consiste en una copia del espacio de direcciones del proceso original; este mecanismo permite que el proceso padre se comunique fácilmente con su proceso hijo. Ambos procesos (el padre y el hijo) continúan su ejecución en la instrucción que va después de la llamada fork, con una diferencia: el código de retorno par la llamada fork es cero para el proceso nuevo (hijo), en tanto que el identificador de proceso (distinto de cero) del hijo se devuelve al padre.[11]

---

[10] Wikipedia: “Proceso (informática)”, material en línea, disponible en: http://es.wikipedia.org/wiki/Proceso_(inform%C3%A1tica), recuperado el 4 de noviembre de 2008.
[11] Depto. de Electrónica e Informática, UC de Paraguay (A. Salazar y P. Coronel): “Sistemas operativos” http://www.dei.uc.edu.py/tai2003-2/sistemas.operativos/sistemas%20operativos.htm, recuperado el 13/01/09.

2.3 Jerarquía de procesos, prioridades y colas

Como se mencionó al inicio de este tema un proceso se representa por un conjunto de datos denominado **bloque de control de procesos** (BCP), estos datos permiten al sistema operativo localizar información sobre el proceso y mantenerlo registrado por si hay que suspender la ejecución temporalmente.

La información contenida es:

- Estado del proceso.
- Estadísticas de tiempo y uso de recursos.
- Ocupación de memoria interna y externa (swapping).
- Recursos en uso.
- Archivos en uso.
- Privilegios.

Los BCP se almacenan en colas y hay una por cada estado posible, se dividen en:

- Activos
- Inactivos

Los *activos* son aquellos que compiten por el procesador o están en condiciones de hacerlo:

- *Ejecución*.- cuando un proceso tiene el control del procesador.
- *Preparado*.- son aquellos procesos que están dispuestos a ser ejecutados.
- *Bloqueado*.- no pueden ejecutarse por que requieren algún recurso no disponible o están en condiciones de hacerlo.

Los *inactivos* son aquellos que no pueden competir por el procesador, pero pueden volver a hacerlo si se soluciona el problema que los ha dejado en “suspenso” (falla de un dispositivo de entrada-salida (E/S):

- *Suspendido bloqueado*.- proceso que ha sido suspendido y que además está a la espera de un evento para desbloquearse.

- *Suspendido preparado*.- proceso que ha sido simplemente suspendido.[12]

Prioridades

A cada proceso se le asigna una prioridad en función de la urgencia y de los recursos que disponga, lo cual determina la frecuencia de acceso al procesador, las prioridades se clasifican en:

- *Asignadas por el sistema operativo*: dependiendo de los privilegios de su propietario y del modo de ejecución.
- *Asignadas por el propietario*: antes de comenzar la ejecución.
- *Estáticas*: no pueden ser modificadas durante la ejecución.
- *Dinámicas*: pueden ser modificadas en función de los eventos que se produzcan.

Los procesos, en los diferentes estados que tienen, son agrupados en listas o colas:

- **Lista de procesos del sistema (job queue):** En esta lista estan todos los procesos del sistema, al crearse un proceso nuevo se agrega el PCB a esta lista, cuando el proceso termina su ejecución es borrado.
- **Cola de procesos listos (ready queue):** Esta cola se compondrá de los procesos que estén en estado listo, la estructura de esta cola dependerá de la estrategia de planificación utilizada.
- **Cola de espera de dispositivos (device queue):** Los procesos que esperan por un dispositivo de E/S en particular, son agrupados en una lista específica al dispositivo. Cada dispositivo de E/S tendrá su cola de espera.[13]

[12] Cf., Luis Panizo Alonso, Universidad de León, *Área de arquitectura de computadores. Sistemas operativos: Procesos*, material en línea, disponible en: aquí tema 3, o bien, http://torio.unileon.es/~dielpa/asig/shannon/SO/teoria/so03.pdf, pp. 1-12, recuperado el 13/01/09.
[13] Facultad de Ingeniería, Univ. de la República, Uruguay, curso en línea de *Sistemas operativos*, "Procesos", disponible en: http://www.fing.edu.uy/inco/cursos/sistoper/recursosTeoricos/SO-Teo-Procesos.pdf , p. 18, recuperado el 13/01/09.

2.4 Arrancador o despachador de procesos

Un componente implicado en la función de la planificación de la unidad central de proceso (CPU) es el **despachador**.[14] Este componente es el módulo que da el control del CPU al proceso seleccionado por el planificador de corto plazo y comprende las siguientes funciones:

- Conmutación de contexto.
- Conmutación a modo de usuario.-
- Saltar a la localidad apropiada en el programa del usuario para reiniciar el programa.

El despachador es uno de los módulos del administrador de procesos y decide a qué procesador asignar el proceso que tiene que ser ejecutado, este deberá ser muy rápido, ya que es invocado en cada conmutación de procesos. El tiempo que le lleva al despachador detener un proceso e iniciar la ejecución de otros se conoce como latencia de despacho.

Es muy importante distinguir los conceptos de servicio de archivos y servidor de archivos, por ejemplo en los sistemas distribuidos la función es la siguiente:

- **Servicio de archivos:**
  - Es la especificación de los servicios que el sistema de archivos ofrece a sus clientes
  - Describe los parámetros que utilizan y las acciones que se llevan a cabo.
  - Define el servicio con el que pueden contar los clientes.

- **Despachador (servidor) de archivos:**
  - Es un proceso que ejecuta alguna maquina y ayuda con la implantación del servicio de archivos.
  - Pueden existir uno o varios en el sistema
  - Generalmente un servidor de archivos es un proceso de usuario
  - Un sistema puede contener varios servidores de archivos, cada uno con un servicio distinto, por ejemplo servidores con sistema

[14] Abraham Silbertschatz, op. cit., pp. 139-141.

operativo Unix y otros con sistema operativo Windows, en el que cada proceso usuario utilizara el servidor adecuado.[15]

Bibliografía del tema 2

Carretero, Jesús. *Sistemas Operativos*. Madrid, McGraw-Hill, 2001.

Silbertschatz, Abraham. *Sistemas Operativos*. 6ª ed., México, Limusa Wisley, 2002.

Actividades de aprendizaje

A.2.1. Realiza en un documento lo siguiente:

- diagrama de los estados de un proceso
- diagrama del modelo de dos estados
- diagrama del modelo de cinco estados
- describe las principales diferencias ente los modelos de dos y cinco estados.

A.2.2. Realiza un cuadro comparativo sobre la función de los estados activos e inactivos de los BCP.

A.2.3. Investiga y elabora en un documento en Word una descripción de tres funciones de un despachador de archivos en un sistema distribuido.

A.2.4. Describe en un documento la función del servicio de archivos en un sistema distribuido.

---

[15] Nora, 12/06/06: "Windows 95: El despachador de procesos. Conceptos", material en línea disponible en: http://www.wikilearning.com/apuntes/windows_95-el_despachador_de_procesos_conceptos/13976-3, recuperado el 13/01/09.

Cuestionario de autoevaluación

1. ¿Qué es un proceso?
2. ¿Cuál es la función de los bloque de control de procesos (BCP)?
3. Explica el modelo de procesos de dos estados
4. ¿Cuál es la función de los procesos suspendidos?
5. ¿Cómo se identifican los procesos?
6. ¿Qué es una prioridad estática?
7. ¿Qué es una cola de espera de dispositivos (*device queue*)?.
8. ¿Cuál es la función del despachador de procesos?
9. ¿Cuál es la diferencia entre un servicio de archivos y un despachador de archivos en un sistema distribuido?
10. ¿Cuál es la causa de que un proceso quede inactivo?

Examen de autoevaluación

1. ¿Qué es un proceso?
   a. Un recurso de hardware
   c. Un programa en compilación
   c. Un programa en ejecución

2. El estado en **espera** de un proceso es:
   a. Cuando el proceso es expulsado
   b. Cuando el proceso espera a ser creado
   c. Cuando el proceso espera cierto suceso

3. ¿Cuál es la característica de un proceso inactivo?
   a. No compite con el procesador
   b. Si compite con el procesador
   c. No tiene que ver con el procesador

4. ¿Cuál es el componente de hardware que está relacionado con el despachador de procesos?

a. Memoria
b. Disco
c. CPU

5. ¿Cuál es la prioridad del proceso que puede ser modificado durante su ejecución?

a. Dinámica
b. Estática
c. Mutua

6. El proceso de un servidor de archivos consiste en que:

a. Es un proceso que ejecuta alguna maquina
b. Es un proceso exclusivo del sistema operativo
c. Es un proceso exclusivo del cliente

7. ¿Cómo se identifica un proceso?

a. Mediante el identificador de la dirección de memoria
b. Mediante el identificador del proceso
c. Mediante el identificador del servidor

8. Menciona una función del despachador de procesos

a. Conmutación de contexto
b. Conmutación de memoria
c. Conmutación de programas

9. ¿Por qué causa un proceso puede quedar inactivo?

   a. Falla de un programa de aplicación

   b. Falla de un dispositivo de entrada-salida

   c. Falla de la memoria

10. El modelo de dos estados consiste en:

   a. Entrar, salir

   b. Ejecución, no ejecución

   c. Entrar, pausar

TEMA 3. SINCRONIZACIÓN Y COMUNICACIÓN ENTRE PROCESOS

Objetivo particular

Al culminar el aprendizaje del tema, el alumno reconocerá la importancia que tienen los procesos cooperativos y su efecto en el diseño y construcción de los sistemas operativos.

Temario detallado

3.1 Paralelismo y competencia entre procesos
3.2 Estado de procesos
3.3 Transición de estados
3.4 Comunicación entre procesos
3.5 Interrupciones
3.6 Interbloqueos de procesos
3.7 Algoritmos de administración de procesos

Introducción

Los distintos procesos que se ejecutan en una computadora no actúan de forma aislada, por un lado algunos procesos cooperan para lograr un objetivo común; por otro lado, los procesos compiten por el uso de recursos limitados, tales como: el uso del procesador, la memoria y los archivos. Estas actividades de cooperación y competencia llevan asociada la necesidad de que exista alguna comunicación entre estos. En este tema se estudiará cómo se realiza la sincronización y comunicación de los procesos.

3.1 Paralelismo y competencia entre procesos

El paralelismo implica que existen varios procesadores en un sistema que se da entre la unidad central de proceso (CPU) y los dispositivos de entrada/salida.

> Un programa concurrente es visto como una colección de procesos secuenciales autónomos que se ejecutan (lógicamente) en paralelo. La ejecución de procesos toma una de las siguientes formas:
>
> - **Multiprogramación:** ejecución de múltiples procesos en un solo procesador.
> - **Multiprocesamiento:** ejecución de múltiples procesos en un sistema multiprocesador donde hay acceso a memoria compartida.
> - **Programación distribuida:** ejecución de múltiples procesos en varios procesadores los cuales no comparten memoria.[16]

La competencia entre procesos se puede dar en alguna de las siguientes formas:

- **Compatibles:** pueden ser utilizados por varios procesos de forma concurrente.
- **No compatibles:** su uso se restringe a un solo proceso solamente

La naturaleza física del recurso hace que sea imposible compartirlo, por ejemplo si una impresora fuera utilizada por varios procesos simultáneamente, sus salidas se mezclarían en el papel.

Si el recurso es usado en forma concurrente, la acción de uno de ellos, puede interferir con la de otro, por ejemplo un archivo que contiene datos accesibles desde más de un proceso y modificables por uno de ellos.[17]

Dentro de la categoría de los recursos no compatibles se encuentran la mayoría de los periféricos, los archivos de escritura y las zonas de memoria que se pueden modificar. En los recursos compatibles se encuentra la unidad central de

---

[16] Pedro Mejía Álvarez: *Procesos concurrentes*, material electrónico disponible en: http://delta.cs.cinvestav.mx/~pmejia/capi5tr.ppt, diapositiva 6/19. Recuperado el 13/01/09.

[17] Información basada en: Lina García, Universidad de Jaén, "concurrencia", material disponible en: http://wwwdi.ujaen.es/~lina/TemasSO/CONCURRENCIA/1ComunicacionySincronizacion.htm, recuperado el 13/01/09.

proceso (CPU), archivos de lectura zonas de memoria que no esté sujeta a modificación.[18]

3.2 Estado de procesos

A medida que se ejecuta un proceso, cambia su estado. El estado de un proceso se define en parte por la actividad actual de dicho proceso. Cada proceso puede estar en alguno de los siguientes estados:[19]

- **Nuevo**: el proceso se está creando.
- **Ejecución**: se están ejecutando instrucciones
- **En espera**: el proceso está esperando que ocurra algún evento (como la terminación de una operación de E/S o la recepción de una señal).
- **Listo**: El proceso está en espera de ser asignado a un procesador
- **Terminado**: el proceso ha terminado su ejecución.[20]

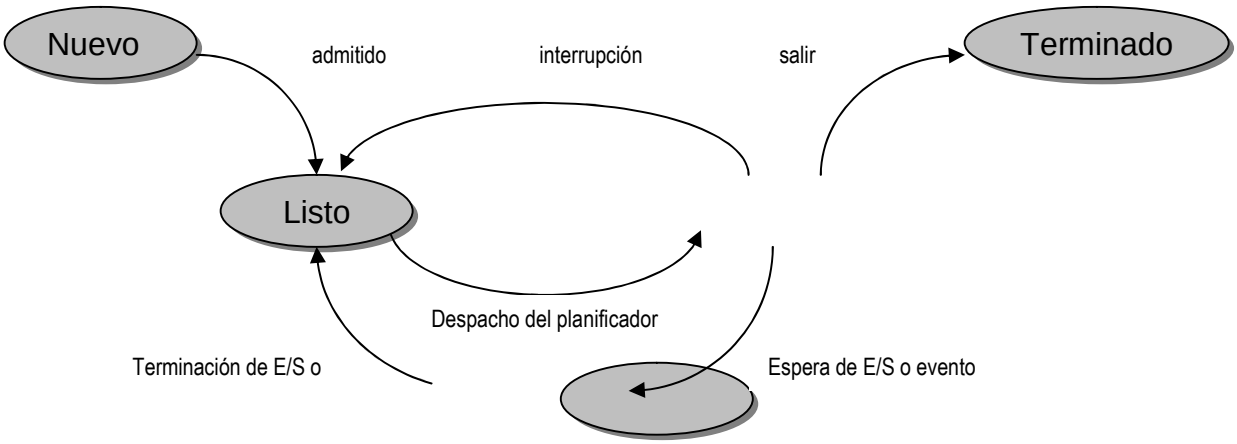


Figura 3.1Diagrama de estados de un proceso

[18] Información basada en: Lina García, Universidad de Jaén, “Concurrencia”, material disponible en: http://wwwdi.ujaen.es/~lina/TemasSO/CONCURRENCIA/1ComunicacionySincronizacion.htm, recuperado el 13/01/09.
[19] Silbertschatz, Abraham. Sistemas Operativos. 6ª ed., Limusa Wiley, México, 2002. Pág. 88-89.
[20] Pedro Mejía Álvarez: *Procesos concurrentes*, material electrónico disponible en:http://delta.cs.cinvestav.mx/~pmejia/capi5tr.ppt, diapositiva 8/19. Recuperado el 13/01/09.

3.3 Transición de estados[21]

Cuando un trabajo es admitido se crea un proceso equivalente, y es insertado en la última parte de la cola de listos. Cuando un proceso pasa de un estado a otro se dice que hace una transición de estado, las posibles transiciones se describen a continuación.

Transición	Proceso	Estado	Descripción
Nulo-Nuevo		Se crea un nuevo proceso para ejecutar un programa.	
Nuevo-Listo	Admitido	El sistema operativo pasará un proceso del estado Nuevo al estado Listo cuando esté preparado para aceptar un proceso más.	Cuando un proceso se ha creado y se le es permitido para competir por la CPU.
Listo-Ejecución	Despacho	Cuando es el momento de seleccionar un nuevo proceso para ejecutar, el sistema operativo uno de los procesos del estado Listo	La asignación de la CPU al primer proceso de la lista de listos se llama despacho y se ejecuta por la entidad de sistema llamada despachador. Mientras que el proceso tenga la CPU se dice que está en ejecución.
Ejecución-Terminado	Salir	El proceso que se está ejecutando es finalizado por el sistema operativo si indica que termino o si se abandona.	Esta transición ocurre cuando el proceso se ha terminado de ejecutarse, y pasa a un estado de terminado.
Ejecución-Listo	Tiempo Excedido	El proceso que se está ejecutando ha alcanzado el tiempo máximo permitido de ejecución interrumpida.	El S.O. cuando un proceso se le expira el intervalo de tiempo asignado para estar en ejecución (CUANTO), hace que este proceso que se hallaba en

[21] Cf. William Stallings, *Sistemas Operativos*. 4ª. ed., Pearson Educación, Madrid, 2001, p. 114. Y sobre todo, Rincón del vago, "Sistema operativo. Procesos. Transición estados. Bloque control proceso", material en línea, disponible en: http://html.rincondelvago.com/sistemas-operativos_26.html, recuperado el 13/01/09.

			estado de ejecución pase al estado de listo y inmediatamente el despachador hace que el primer proceso de la lista pase a estado de ejecución.
Ejecución-Bloqueado	Bloqueo	Un proceso se pone en estado Bloqueado si solicita algo por lo que debe esperar.	Si un proceso que se encuentra en estado de ejecución inicia una operación de E/s antes que termine su cuanto, el proceso voluntariamente abandona la CPU, es decir, el proceso se bloquea a sí mismo.
Bloqueado-Listo	Despertar	Un proceso que está en el estado Bloqueado pasara al estado Listo cuando se produzca el suceso que estaba esperando.	La única transición posible en nuestro modelo básico ocurre cuando acaba una operación de E/S (o alguna otra causa por la que esté esperando el proceso), y esta termina pasa al estado de listo.
Ejecución-Terminado	Salir		Esta transición ocurre cuando el proceso se ha terminado de ejecutarse, y pasa a un estado de terminado.

Tabla 3. 1. Estados de transición[22]

Para prevenir que un proceso monopolice el sistema, el sistema operativo ajusta un reloj de interrupción del hardware para permitir al usuario ejecutar su proceso durante un intervalo de tiempo específico. Cuando este tiempo expira el reloj genera una interrupción, haciendo que el sistema operativo recupere el control.

Cuando hay demasiada carga en el sistema se puede hacer uso de suspensión y reanudación por el S.O., para equilibrar la carga del sistema.

[22] "Sistema operativo. Procesos. Transición estados. Bloque control proceso", material en línea, disponible en: http://html.rincondelvago.com/sistemas-operativos_26.html, recuperado el 13/01/09.

Para la reanudación y la suspensión será necesario anexar otros dos estados los cuales son: suspendido listo y suspendido bloqueado, con las siguientes definiciones de transiciones:

- Suspende _ ejecución (Proceso): En ejecución Suspendido listo.
- Suspende _ bloqueado (Proceso): Bloqueado Suspendido
- Reanuda (Proceso): Suspendido listo.
- Término E/S (Proceso): Suspendido bloqueado Suspendido listo.[23]

3.4 Comunicación entre procesos

Los procesos cooperativos pueden comunicarse en un ambiente de memoria compartida. El esquema requiere que estos procesos compartan una reserva común de buffers y que el programador de la aplicación escriba de manera explícita el código para implementar el buffer. Otra forma de lograr el mismo efecto es que el sistema operativo proporcione los medios para que los procesos cooperativos se comuniquen entre ellos a través de un servicio de comunicación entre procesos (IPC).

El IPC proporciona un mecanismo tanto para que los procesos se comuniquen como para sincronizar sus acciones sin compartir el mismo espacio de direcciones. El IPC es útil en un ambiente distribuido, en donde los procesos que se comunican pueden residir en diferentes computadoras conectadas en una red. Un ejemplo es un programa para chat (conversación) empleado en la World Wide Web.

El servicio de IPC se proporciona mejor mediante un sistema de paso de mensajes. Los sistemas de mensajes pueden definirse de diferentes formas.

---

[23] "Sistema operativo. Procesos. Transición estados. Bloque control proceso", material en línea, disponible en: http://html.rincondelvago.com/sistemas-operativos_26.html, recuperado el 13/01/09.

Los procesos[24] que desean comunicarse necesitan una forma de hacer referencia entre ellos. Pueden usar comunicación directa, o bien, comunicación indirecta.

Comunicación directa

En este tipo de comunicación cada proceso que quiere comunicarse debe nombrar explícitamente al receptor o al emisor de la comunicación. En este esquema, las primitivas send y receive se definen como:

- Send (P, mensaje): Enviar un mensaje al proceso P.
- Receive (Q, mensaje): Recibir un mensaje del proceso Q.

Un enlace de comunicación en este esquema tiene las siguientes propiedades:

- Un enlace se establece automáticamente entre cada par de procesos que desean comunicarse. Los procesos sólo necesitan conocer la identidad de los demás para comunicarse.
- Un enlace está asociado exactamente con dos procesos.
- Entre cada par de procesos, existe exactamente un enlace.

Este esquema exhibe una simetría en el direccionamiento; es decir, tanto el proceso emisor como el receptor tienen que nombrar al otro para comunicarse. Una variante de este esquema emplea la asimetría en el direccionamiento. Sólo el emisor nombra al receptor; no se requiere que éste nombre al emisor. En este esquema, las primitivas send y receive se definen como sigue:

- Send (P, mensaje) –Enviar un mensaje al proceso P

- Receive (ID, mensaje) –Recibir un mensaje de cualquier proceso; el valor de la variable id se fija en el nombre del proceso con el que ha tenido lugar la comunicación.

La desventaja en ambos esquemas (simétrico y asimétrico) es la modularidad limitada de las definiciones de procesos resultantes. Cambiar el nombre de un proceso puede requerir examinar todas las demás definiciones de procesos. Se deben encontrar todas las referencias al nombre anterior, de modo que puedan modificarse con el nuevo nombre. Esta situación no es deseable desde el punto de vista de una compilación separada.

[24] Véase, Abraham Silbertschatz, op. cit., pp. 101-106.

Comunicación indirecta

Con la comunicación indirecta, los mensajes se envían y se reciben de buzones, o puertos. Un buzón puede verse, de manera abstracta, como un objeto en donde los procesos pueden colocar y remover mensajes. Cada buzón tiene una identificación única. En este esquema, un proceso se puede comunicar con otro utilizando varios buzones diferentes. Dos procesos sólo se pueden comunicar si tienen un buzón compartido. Las primitivas send y receive se definen de la siguiente manera:

- Send (A, mensaje) –Enviar un mensaje al buzón A.
- Receive (A, mensaje) –Recibir un mensaje del buzón A.

En este esquema, un enlace de comunicación tiene las siguientes propiedades:

- Se establece un enlace entre un par de procesos sólo si ambos miembros tienen un buzón compartido.
- Un enlace puede estar asociado con más de dos procesos.
- Entre cada par de procesos de comunicación, puede haber varios enlaces diferentes, y cada enlace corresponde un buzón.

Un buzón puede ser propiedad ya sea de un proceso o del sistema operativo. Si el buzón es propiedad de un proceso (es decir, forma parte del espacio de direcciones del proceso), entonces distinguimos entre el propietario (quien sólo puede recibir mensajes a través de este buzón) y el usuario del buzón (quien sólo puede enviar mensajes al buzón). Debido a que cada buzón tiene un propietario único, no puede haber confusión acerca de quien recibirá un mensaje enviado a este buzón. Cuando un proceso propietario de un buzón termina, el buzón desaparece. Cualquier proceso que subsecuentemente envíe un mensaje a este buzón debe ser notificado de que ya no existe.

Por otra parte, un buzón cuyo propietario es el sistema operativo tiene una existencia propia. Es independiente y no está obligado a un proceso en particular. El sistema operativo debe entonces proporcionar un mecanismo que permita a los procesos:

- Crear un nuevo buzón.
- Enviar y recibir mensajes a través del buzón.
- Borrar un buzón.

El proceso que crea un nuevo buzón es, por omisión, el propietario de dicho buzón. Inicialmente, el propietario es el único proceso que puede recibir mensajes a través de este buzón. Sin embargo, el privilegio de propiedad y recepción puede transferirse a otros procesos mediante

llamadas al sistema apropiadas. Por supuesto, esto puede dar como resultado varios receptores para cada buzón.[25]

3.5 Interrupciones

El sistema operativo ocupa una posición intermedia entre los programas de aplicación y el hardware. No se limita a utilizar el hardware a petición de las aplicaciones ya que hay situaciones en las que es el hardware es el que necesita que se ejecute código del sistema operativo. En este caso el hardware debe poder llamar al sistema y existen dos condiciones:

- Algún dispositivo de E/S necesita atención.
- Se ha producido una situación de error al intentar ejecutar una instrucción del programa (normalmente de la aplicación).

En ambos casos, la acción realizada no está ordenada por el programa de aplicación, es decir, no figura en el programa.
Según los dos casos anteriores tenemos las interrupciones y las excepciones:

- Interrupción: señal que envía un dispositivo de E/S a la unidad central de proceso (CPU) para indicar que la operación de la que se estaba ocupando, ya ha terminado.
- Excepción: una situación de error detectada por la unidad central de proceso (CPU) mientras ejecutaba una instrucción, que requiere tratamiento por parte del sistema operativo.[26]

A nivel físico, una interrupción se solicita activando una señal que llega a la unidad de control. El generador de la interrupción, debe de activar la señal cuando necesite que se le atienda (que se ejecute un programa que lo atienda). Ante la solicitud de una interrupción, siempre y cuando este habilitado este tipo de interrupción, la unidad de control realiza un ciclo de aceptación de interrupción.[27] Este ciclo se lleva a cabo en cuanto termina la ejecución de la instrucción máquina que se esté ejecutando y consiste en las siguientes operaciones:

---

[25] "Sistemas operativos (Resumen: 6. Comunicación entre procesos)", material en línea, disponible en: http://ar.geocities.com/clubdealumnos/soperat/Peterson.htm, recuperado el 13/01/09.
[26] Wikipedia: "Sistema operativo: Interrupciones y excepciones", actualizado el 12/01/09, material en línea, disponible en: http://es.wikipedia.org/wiki/Sistema_operativo#Interrupciones_y_excepciones, recuperado el 13/01/09.
[27] Jesús Carretero, op. cit., pp. 7-9.

- Salva algunos registros del procesador (estado y contador de programa).
- Eleva el nivel de ejecución del procesador, pasándolo al núcleo.
- Carga un nuevo valor en el contador de programa, por lo que pasa a ejecutar otro programa.

Las interrupciones se pueden generar por las siguientes causas:

- Excepciones de programa. Hay determinadas causas que hacen que un programa presente un problema en su ejecución, por lo que deberá generarse una interrupción, para que el sistema operativo trate esta causa. Por ejemplo: el desbordamiento en las operaciones aritméticas, la división por cero, el intento de ejecutar una operación con código de operación incorrecto o de direccionar una posición de memoria prohibida.
- Interrupciones de reloj
- Interrupciones de entrada/salida (E/S).Los controladores de de los dispositivos de E/S necesitan interrumpir para indicar que han terminado una operación o conjunto de ellas.
- Excepciones del hardware. La detección de un error de paridad en la memoria o un corte de corriente se avisan mediante la correspondiente interrupción.
- Instrucciones de TRAP. Estas instrucciones permiten que un programa genere una interrupción.

El término reloj se aplica a las computadoras con tres significados diferentes y que están relacionas entre sí:

- Señal que gobierna el ritmo de ejecución de las instrucciones de máquina.
- Generador de interrupciones periódicas
- Contador de fecha y hora

El oscilador que gobierna las fases de ejecución de las instrucciones máquina se denomina reloj. Cuando se dice que un microprocesador es de 600 MHz, lo que se

está especificando es que el oscilador que gobierna el ritmo de su funcionamiento interno produce una onda cuadrada con una frecuencia de 600 MHz.

La señal producida por el oscilador se divide mediante un divisor de frecuencia para generar una interrupción cada cierto intervalo de tiempo. Estas interrupciones que se producen constantemente se denominan interrupciones de reloj o tics, dando lugar al segundo concepto de reloj. El objetivo de estas interrupciones es hacer que el sistema operativo entre a ejecutar de forma sistemática cada cierto intervalo de tiempo y evita que un programa monopolice el uso de la computadora y puede hacer que entren a ejecutarse programas en determinados instantes de tiempo. El tercer significado de reloj se aplica a un contador que permite conocer la fecha y la hora. Este contador se va incrementando con cada interrupción de reloj de forma que, tomando como referencia un determinado instante, se puede calcular la fecha y hora en que estamos.

3.6 Interbloqueos de procesos

En los sistemas operativos, el interbloqueo de procesos (también llamado bloqueo mutuo o abrazo mortal) es el bloqueo permanente de un conjunto de procesos o hilos de ejecución en un sistema concurrente que compiten por un número finito de recursos. Cuando varios procesos compiten por un número finito de recursos puede surgir una situación en la que un proceso solicite un recurso y éste no se encuentre disponible en este momento, en cuyo caso el proceso pasará a un estado de espera.[28] Tal vez suceda que algunos procesos en espera nunca cambien nuevamente su estado, debido a que los recursos que han solicitado están retenidos por otros procesos también en espera. Esta situación se denomina bloqueo mutuo.

[28]Cf. “Sistoper bloqueos mutuos”, material en línea, disponible en: http://www.slideshare.net/cesar2007/sistoper-bloqueos-mutuos/n, consultado el 13/01/09.

A diferencia de otros problemas de concurrencia de procesos, no existe una solución general para los interbloqueos. Todos los interbloqueos surgen de necesidades que no pueden ser satisfechas, por parte de dos o más procesos.

Como se mencionó anteriormente un sistema consta de un número finito de recursos que se van a distribuir entre los procesos que compiten por ellos. Los recursos están divididos en varios tipos: espacio de memoria, ciclos de CPU, archivos, dispositivos de E/S, etc. Si un sistema tiene dos CPU entonces el recurso del tipo CPU tiene dos instancias, de manera similar el recurso del tipo impresora puede tener cinco instancias.

"Si un proceso solicita una instancia de un tipo de recurso, la asignación de cualquier instancia de dicho tipo deberá satisfacer la solicitud. Si no fuera así, entonces las instancias no son idénticas y las clases de tipos de recursos no han sido definidas correctamente". Por ejemplo, un sistema puede tener dos impresoras. Estas dos impresoras pueden definirse para la misma clase de recurso si a nadie le importa qué impresora genera qué salida. Sin embargo, si una impresora está en el noveno piso y la otra en el sótano, entonces las personas del noveno piso tal vez no vean a las dos impresoras como equivalente, y quizá sea necesario definir clases de recursos distintas para cada impresora.

> Un proceso debe solicitar un recurso antes de usarlo y liberarlo después de usarlo. Un proceso puede solicitar tantos recursos como requiere para llevar a cabo su tarea asignada. Obviamente, el número de recursos solicitados no puede exceder el número total de recursos disponibles en el sistema.[29]

Es decir, un proceso no puede solicitar tres impresoras si el sistema sólo tiene dos. Si se hace tal solicitud, será rechazada.

---

[29] Instituto Tecnológico de Celaya: "Sistemas operativos II", disponible en: http://sisinfo.itc.mx/ITC-APIRGG/Materias/Mat4/SistOp-II_Unid2.php, recuperado el 13/01/09.

En el modo normal de operación, un proceso puede usar un recurso sólo en la siguiente secuencia[30]:

1. *Solicitud*: si la solicitud no puede ser atendida inmediatamente (porque el recurso está siendo utilizado por otro proceso), entonces el proceso solicitante debe esperar hasta que pueda adquirir el recurso.

2. *Uso*: el proceso puede operar sobre el recurso (por ejemplo, si el recurso es una impresora, el proceso puede imprimir en ella).

3. *Liberación*: el proceso libera al recurso.

La *solicitud* y *liberación* de recursos son llamadas al sistema. Como ejemplos se tienen las llamadas al sistema request y release device, open y close file, allocate y free memory. La solicitud y liberación de recursos que no son administrados por el sistema operativo pueden realizarse mediante las operaciones P y V en semáforos, o mediante la adquisición y liberación de una cerradura para un objeto Java vía la palabra clave synchronized. Por cada uso de un recurso administrado por el kernel, por parte de un proceso hilo, el sistema operativo hace una verificación para asegurarse de que el proceso solicitó y se le asignó el recurso. Una tabla del sistema registra si cada recurso está libre o asignado y, para cada recurso asignado, a qué proceso. Si un proceso solicita un recurso que actualmente está asignado a otro proceso, puede ser agregado a una cola de procesos que están en espera de dicho recurso.

Un conjunto de procesos se encuentra en un estado de bloqueo mutuo cuando cada proceso del conjunto está esperando un evento que sólo puede ser provocado por otro proceso en el conjunto. La característica de los bloqueos

[30] Abraham Silbertschatz, op. cit., pp. 227-247.

mutuos es que los procesos nunca terminan de ejecutarse y los recursos están inmovilizados, lo que impide que otros trabajos puedan iniciar.

El bloqueo mutuo puede surgir si se presentan simultáneamente las siguientes cuatro condiciones en un sistema:

1. *Exclusión mutua*: al menos un recurso debe estar retenido en un modo no compartido; es decir, solo un proceso a la vez pude usar el recurso. Si otro proceso solicita dicho recurso, el proceso solicitante debe esperar hasta que el recurso haya sido liberado.

2. *Retención y espera*: debe existir un proceso que esté retenido por lo menos un recurso y esté esperando adquirir recursos adicionales que en ese momento estén siendo retenidos por otros procesos.

3. *No apropiación*: Los recursos no pueden ser apropiados; es decir, un recurso sólo puede ser liberado voluntariamente por el proceso que lo está reteniendo, una vez que dicho proceso ha completado su tarea.

4. *Espera circular*: debe existir un conjunto $\{P_0, P_1, P_2 \ldots P_n\}$ de procesos en espera, tal que P_0 esté esperando un recurso que está retenido por P_1, P_1 y estos esperen por un recurso retenido por $P_2, \ldots, P_{n-1}$ espere un recurso retenido por P_n y P_n este esperando un recurso retenido por P_0.

Existen principalmente tres métodos diferentes para manejar el problema de bloqueos mutuos:

1. Utilizar un protocolo para asegurar que el sistema nunca entrará en un estado de bloqueo mutuo.

2. Permitir que el sistema entre en un estado de bloqueo mutuo y luego hacer una recuperación.
3. Ignorar el problema y pretender que los bloqueos mutuos nunca ocurren en el sistema.

La tercera solución es la que utilizan la mayoría de los sistemas operativos, incluyendo UNIX. Para asegurar que nunca ocurran bloqueos mutuos, el sistema puede usar un esquema de prevención de bloqueos o bien un esquema para evitar bloqueos. La prevención de bloqueos mutuos es un conjunto de métodos para asegurar que por lo menos una de las condiciones necesarias no se cumpla. Estos métodos previenen los bloqueos mutuos restringiendo la forma en que pueden hacerse las solicitudes de recursos.

La evitación de bloqueos mutuos, por otra parte, requiere que al sistema operativo se le dé por adelantado información adicional relacionada con los recursos que solicitará y usará un proceso durante la vida de éste. Con esta información adicional, el sistema operativo puede decidir en cada solicitud si el proceso debe o no esperar.

Si un sistema no emplea un algoritmo de prevención de bloqueos mutuos ni un algoritmo para evitar bloqueos mutuos, entonces puede presentarse la situación de bloqueo mutuo. En este ambiente, el sistema pude proporcionar un algoritmo que examine el estado del sistema para determinar si ha ocurrido un bloqueo mutuo, y un algoritmo para recuperarse de éste (si efectivamente ha ocurrido).[31]

Asimismo, si un sistema no asegura que jamás ocurrirá un bloqueo mutuo y tampoco proporciona un mecanismo para detección y recuperación de bloqueos

[31] Cf. Departamento de Informática, “Sistemas operativos y redes”, 23/03/07, material en línea, disponible en: http://www.dirinfo.unsl.edu.ar/~sonet/teorias/SO-clase5-pagina.pdf, pássim. Recuperado el 13/01/09.

mutuos, entonces el sistema puede llegar a un estado de bloqueo mutuo y ni siquiera tener una forma de reconocer lo que ha ocurrido. En este caso, el bloqueo mutuo no detectado dará por resultado un deterioro del desempeño del sistema, debido a que los procesos que no pueden ejecutarse están reteniendo recursos y porque al solicitar recursos, más y más procesos entran en un estado de bloqueo mutuo. Con el tiempo, el sistema dejará de funcionar y tendrá que reiniciarse manualmente.

No obstante que este método no parece ser una forma viable de manejar el problema de los bloqueos mutuos, se utiliza en algunos sistemas operativos. En muchos sistemas, los bloqueos mutuos ocurren con poca frecuencia (por ejemplo, una vez al año); por esta razón, es más barato utilizar este método que emplear los costosos métodos de prevención, evitación, o detección y recuperación de bloqueos mutuos que deben ser usados constantemente

3.7 Algoritmos de administración de procesos

La planificación de la unidad central de proceso (CPU) es la base de los sistemas operativos con multiprogramación. El objetivo de la multiprogramación es tener algún proceso en ejecución en todo momento, para maximizar la utilización de la CPU. En el caso de un sistema con un solo procesador, nunca habrá más de un proceso en ejecución. Si hay más procesos, el resto tendrá que esperar hasta que la CPU esté libre y pueda volver a planificarse.

Con la multiprogramación, se trata de usar este tiempo de manera productiva. Se tienen varios procesos en la memoria a la vez. Cuando alguno de ellos tiene que esperar, el sistema operativo le retira la CPU a dicho proceso y se asigna a otro. Este patrón continúa. Cada vez que un proceso tiene que esperar, otro puede hacer uso de la CPU.

La planificación es una función fundamental en el diseño de un sistema operativo. El éxito de la planificación de la CPU depende de la siguiente propiedad observada en los procesos: la ejecución de procesos consta de un ciclo de ejecución de la CPU y espera de entrada salida (E/S).

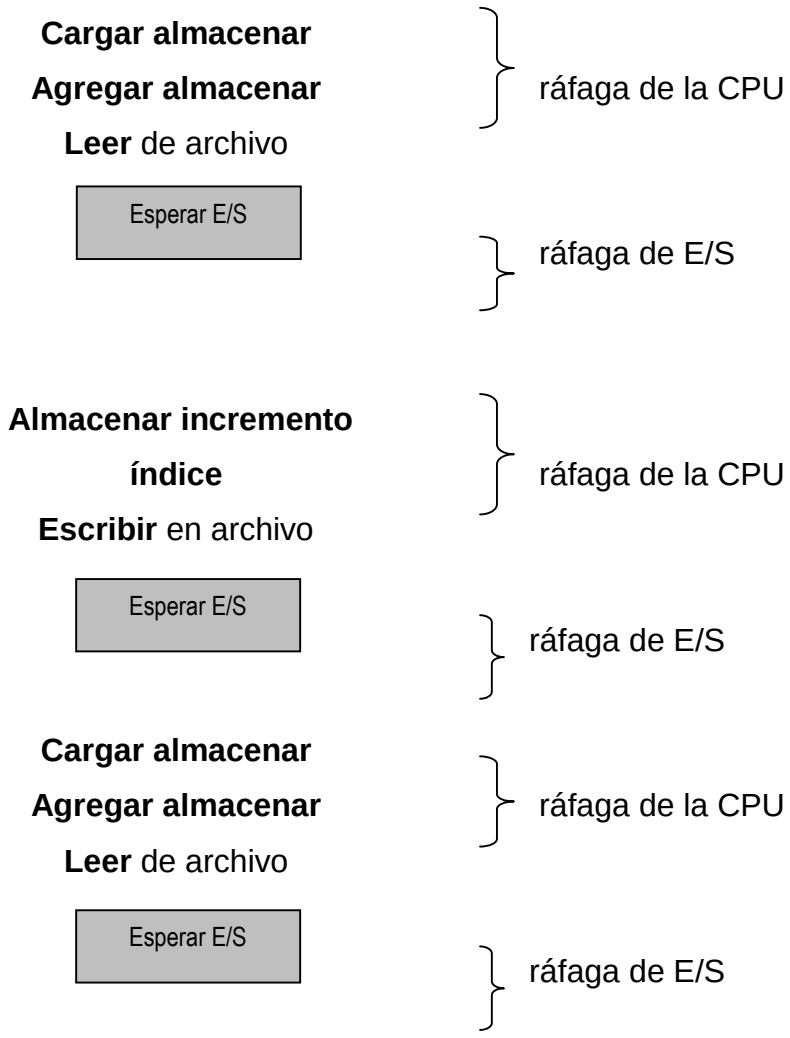


Figura 3.2 Secuencia alternante de ráfaga de CPU y E/S

Los objetivos de la planificación son los siguientes:

- Reparto equitativo del procesador

- Eficiencia, optimizar el uso del procesador (mantener ocupado el 100% de tiempo el CPU).
- Tiempo de respuesta, minimizar el tiempo de respuesta al usuario
- Tiempo de regreso, minimizar el tiempo que deben esperar los usuarios por lotes (batch) para obtener sus resultados
- Rendimiento, maximizar el número de tareas procesadas por hora.[32]

Las decisiones de planificación de la CPU tienen lugar en las siguientes cuatro circunstancias:

1. Cuando un proceso conmuta del estado de ejecución al estado de espera (por ejemplo en una solicitud de E/S, o al invocar una espera para la terminación de uno de los procesos hijos).
2. Cuando un proceso cambia del estado de ejecución al estado listo (por ejemplo, cuando ocurre una interrupción).
3. Cuando un proceso pasa del estado de espera al estado de listo (por ejemplo, en la terminación de una operación de E/S).
4. Cuando un proceso termina.[33]

Para las circunstancias 1 y 4 no existe opción en términos de planificación. Un nuevo proceso (si existe alguno en la cola de listos) debe ser seleccionado para su ejecución. Sin embargo, para las circunstancias 2 y 3, sí existe una opción.

Existen diferentes tipos de algoritmos asociados a la unidad central de proceso (CPU) los cuales tienen diversas propiedades que favorecen a una clase de procesos sobre otros.

Para seleccionar el algoritmo a utilizar en una situación particular se debe considerar las diferentes propiedades de estos.

---

[32] Jesús Carretero, op. cit., p. 104.
[33] Abraham Silbertschatz, op. cit., 136-156.

Los criterios son los siguientes:

- ***Utilización de CPU*:** mantener la CPU tan ocupada como sea posible. La utilización puede estar entre 0 y 100 por ciento, esto depende el uso del sistema ya sea ligero (40%) o pesado (hasta 90%).

- ***Rendimiento (throughput)**:* se mide por el número de procesos que se terminan por unidad de tiempo. En el caso de procesos de larga duración, esta tasa podría ser un proceso por hora, para transacciones breves el rendimiento podría ser de 10 procesos por segundo.

- ***Tiempo de entrega***: tiempo que se requiere para ejecutar un proceso, desde el momento en que se presenta un proceso hasta su terminación. El tiempo de entrega es la suma de los periodos que se consumen esperando llegar a la memoria, esperando en la cola de listos, en ejecución en el CPU y realizando operaciones de E/S.
- ***Tiempo de espera***: el tiempo de espera es la suma de los periodos esperando en la cola de listos.

- ***Tiempo de respuesta***: esta medida es el tiempo que requiere para empezar a responder y no cuánto se requiere para producir la salida de dicha respuesta. El tiempo de entrega por lo general está limitado por la velocidad del dispositivo de salida.

Algoritmos de planificación:

ALGORITMO	CONOCIDO COMO	DESCRIPCIÓN
Planificación del primero en llegar, primero en ser atendido	first come, fist-served, FCFS (Primero en llegar, primero en ser atendido)	La CPU se asigna al primer proceso que la solicite. La implementación de la política del FCFS se maneja fácilmente con una cola tipo FIFO (first input, first output). Cuando un proceso entra a la cola de los listos, su PCB se enlaza al final de la cola. Cuando la CPU está libre, se asigna al proceso que se encuentra a la cabeza de la cola; el proceso que está en ejecución se remueve entonces de dicha cola.
Planificación del primero el trabajo más corto	shortest-job-first, SJF (Es el algoritmo de primero el trabajo más corto)	Este algoritmo asocia con cada proceso la longitud de su siguiente ráfaga de la CPU. Cuando la CPU está disponible, se le asigna al proceso que tiene la ráfaga siguiente más pequeña de la CPU. Si dos procesos tienen la misma longitud de ráfaga siguiente, se emplea la planificación FCFS para tomar la decisión.
Planificación con prioridad	SJF	Caso especial del algoritmo general de planificación con prioridad. Una prioridad está asociada a cada proceso, y la CPU se asigna al proceso con la prioridad más

		alta. Los procesos con igual prioridad se planifican en un orden tipo FCFS. Es simplemente un algoritmo con prioridad en donde la prioridad (p) es el inverso de la siguiente ráfaga (predicha) de la CPU. Entre mayor sea la ráfaga, menor será la prioridad, y viceversa.
Planificación Round-Robin	Round-Robin (RR) (El algoritmo de planificación por turnos)	Diseñado especialmente para sistemas de tiempo compartido. Es similar a la planificación FCFS, pero se añade apropiación para conmutar entre procesos. Aquí se define una pequeña cantidad de tiempo, denominada quantum (porción de tiempo). Un quantum es por lo general de 10 a 100 milisegundos. La cola de listos es tratada como una cola circular. El planificador de la CPU da vueltas sobre la cola de listos, asignando la CPU a cada proceso durante un intervalo de tiempo de hasta 1 quantum.
Planificación de colas de niveles múltiples	Se ha creado otra clase de algoritmos de planificación para situaciones en las cuales los procesos se clasifican	Por ejemplo, se hace una división común entre procesos de primer plano (interactivos) y procesos de segundo plano (en lotes). Estos dos tipos de procesos tienen distintos requerimientos de tiempo de respuesta, y por lo tanto podrían tener diversas necesidades de planificación. Además los procesos de

	fácilmente en grupos diferentes.	primer plano tal vez tengan prioridad (definida externamente) sobre los procesos de segundo plano. Un algoritmo de planificación con colas de niveles múltiples divide la cola de listos en varias colas separadas. Los procesos se asignan de forma permanente a una cola, por lo general con base en alguna propiedad del proceso como; tamaño de memoria, prioridad del proceso, o el tipo de proceso.
Planificación con colas de niveles múltiples y retroalimentación	Planificación con colas de niveles múltiples y retroalimentación	Permite que un proceso se mueva entre colas, separar los procesos con diferentes características de ráfaga de CPU. Si un proceso emplea demasiado tiempo de CPU, será movido a una cola de menor prioridad. Este esquema deja a los procesos limitados por E/S y a los procesos interactivos en las colas de prioridad alta. De manera similar, un proceso que espera demasiado tiempo en una cola de prioridad baja puede ser movido a una cola de mayor prioridad. Esta forma de envejecimiento impide la inanición (bloqueo indefinido).

Tabla 3.2 Algoritmos de planificación

Bibliografía del tema 3

Carretero, Jesús. *Sistemas Operativos*, Madrid, McGraw-Hill, 2001.

Silbertschatz, Abraham. *Sistemas Operativos*. 6ª ed., México, Limusa/Wisley, 2002.

Stallings, William. *Sistemas Operativos*. 4ª ed., Madrid, Pearson Educación, 2001.

Actividades de aprendizaje

A.3.1. Describe los estados de los procesos y realiza su diagrama.

A.3.2. Realiza un cuadro sinóptico sobre las diferencias de la comunicación directa e indirecta.

A.3.3 Describe en un documento las principales causas por las que se generan las interrupciones.

Cuestionario de autoevaluación

1.- ¿Qué es un programa concurrente?

2.- Dentro de la categoría de recursos no compatibles, ¿qué se encuentra?

3.- ¿Para prevenir que un proceso monopolice el sistema, qué hace el S.O?

4.- ¿Qué significa IPC?

5.- ¿Cuál es la desventaja en los esquemas (simétrico y asimétrico)?

6.- ¿Qué es un interbloqueo de procesos?

7.- ¿Cómo está formado un sistema de cómputo?

8.- ¿A qué se refiere la prevención de bloqueos mutuos?

9.- ¿Cuáles son las 4 condiciones para que surja un bloqueo mutuo en un sistema?

10.- ¿Cuál es el objetivo de la multiprogramación?

Examen de autoevaluación

1.- ¿Cuáles son las operaciones para la ejecución de un programa?

a) ejecución de un programa, atención de un programa

b) salvar registros, cargar un nuevo valor

c) interrupciones de reloj, excepciones del hardware

2.- ¿Qué función tiene el reloj al ejecutarse un programa?

a) divide mediante un divisor de frecuencia

b) aplica un contador que permite conocer la fecha y hora

c) ejecuta el sistema operativo en forma sistemática

3.- ¿Cuál es un ejemplo de un recurso compartido?

a) teléfono con impresora

b) impresora con impresora

c) CPU con teléfono

4.- ¿Cuáles son las llamadas que hace un sistema?

a) request y release device

b) semáforos y sincronización

c) java y kernel

5.- ¿Cuál es un ejemplo de un IPC en diferentes computadoras conectadas en una red?

a) programa de video

b) programa de televisión

c) programa de chat (conversación)

6.- ¿Cómo se conoce al algoritmo de planificación del primero en llegar?

a) FIFO

b) FCFS

c) CPU

7.- Si dos procesos tienen la misma longitud de ráfaga siguiente, ¿qué algoritmo se emplea?

a) SJF

b) RR

c) FCFS

8.- Para qué fue diseñado especialmente el algoritmo Round- Robin (RR):

a) para sistemas abiertos

b) para sistemas colapsados

c) para sistemas de tiempo compartido

9.- Si un proceso emplea demasiado tiempo de CPU, ¿a dónde es movido?

a) a la cola de E/S

b) a una cola de menor prioridad

c) a una cola de alta prioridad

10.- Cuando los procesos se envían a colas de alta y baja prioridad, ¿qué es lo que se impide?

a) inanición

b) rendimiento

c) secuencia

TEMA 4. ADMINISTRACIÓN DE MEMORIA

Objetivo particular

Al culminar el aprendizaje de este tema el alumno identificará las características más importantes de uno de los recursos más críticos de los sistemas operativos “la memoria”.

Temario detallado

4.1. Administración de la memoria
4.2 Particiones fijas y dinámicas
4.3 Asignación estática de la memoria
4.4 Asignación dinámica de la memoria
4.5 Paginación
4.6 Políticas de reemplazo de páginas
4.7 Memoria virtual
4.8 Memoria escondida (caché)

Introducción

En los temas dos y tres se estudiaron los procesos y la importancia que tiene la planificación de la unidad central de proceso (CPU) para mejorar la utilización de una computadora, como por ejemplo; el desempeño (velocidad de respuesta). Sin embargo, para mejorar este desempeño se tienen que mantener varios procesos en la memoria y poder compartirla. En este tema se estudiarán las características más importantes sobre la administración de la memoria, su relación con el hardware y su impacto en los sistemas operativos.

4.1. Administración de la memoria

La administración de la memoria se refiere a los métodos y operaciones para obtener la máxima utilidad de la memoria, a través de la organización de procesos y programas que se ejecutan en una computadora. La parte del sistema operativo que administra la jerarquía de memoria se llama administrador de memoria. Los sistemas de administración de memoria se dividen en dos clases; los que traen y llevan procesos entre la memoria principal y el disco duro durante la ejecución (intercambio y paginación), y los que no lo hacen. El intercambio y paginación son mecanismos artificiales obligados por falta de suficiente memoria principal para correr todos los programas a la vez.

La memoria consiste en un arreglo de bytes, cada uno con su propia dirección. La unidad central de proceso (CPU) acude por instrucciones a la memoria de acuerdo con el valor del contador de programa. Estas instrucciones pueden originar, a su vez, carga y almacenamiento de y hacia direcciones específicas de memoria.

La unidad de memoria sólo ve un flujo de direcciones de memoria; no sabe cómo se generan (el contador de instrucción, por índice, indirección, direcciones literales, etc.) ni lo que son (instrucciones o datos). Por lo tanto, podemos ignorar cómo un programa genera una dirección de memoria. Solo nos interesa la secuencia de direcciones de memoria generadas por el programa que está en ejecución.

Esquemas de administración de memoria[34]

1. **Multiprogramación sin intercambio ni paginación.** Consiste en ejecutar solo un programa a la vez, repartiendo ese programa y el sistema operativo. En este caso el sistema operativo podría estar en la parte más baja de la memoria RAM (memoria de acceso aleatorio; *random acces memory*) o podría estar en ROM (memoria de solo lectura; *read-only memory*) en la parte más alta de la memoria, o los controladores de dispositivos podrían estar en la parte más alta de ROM y el resto en RAM más abajo.

2. **Multiprogramación con particiones fijas.** La multiprogramación eleva el aprovechamiento de la CPU. Esto se logra dividiendo la memoria en **n** particiones. También puede lograrse de forma manual, por ejemplo cuando se pone en marcha el sistema. Cuando llega un trabajo se puede colocar en la cola de entrada de la partición más pequeña en la que quepa, el esquema de las particiones son fijas y cualquier espacio de una partición no ocupada se desperdicia. Una de las desventajas de este esquema es que al repartir los trabajos que llegan entre las distintas colas se hace evidente cuando la cola de una partición grande está vacía pero la de una partición pequeña está llena.

3. **Modelado de la multiprogramación.** Cuando se utiliza la multiprogramación es posible mejorar el aprovechamiento de la CPU, un mejor modelo es ver el aprovechamiento de la CPU desde un punto de vista probabilístico. Un proceso pasa una fracción p de su tiempo esperando a que terminen operaciones de E/S. Si hay n procesos en la memoria a la vez la probabilidad de que todos estén esperando E/S (el CPU estará inactivo) es p^n. Entonces el aprovechamiento de la CPU está dado por la fórmula. **Aprovechamiento de** $CPU = 1 - p^n$**.**

El modelo probabilístico no es más que una aproximación, se basa en la suposición implícita de que los **n** procesos son independientes, por ejemplo; en el caso que se tengan cinco procesos en memoria, tres estén

---

[34] Véase, Andrew S. Tanenbaum, *Sistemas Operativos Modernos*, 2ª ed., México, Pearson Education, 2003. pp. 189-194. Por cierto, esta obra está disponible en versión digital en línea, misma de la cual se ha citado a lo largo de este texto, y corresponde al catálogo de Google Books: http://books.google.com.mx/books?id=g88A4rxPH3wC&pg=RA1-PA193&lpg=RA1-PA193&dq=El+modelo+probabil%C3%ADstico+no+es+mas+que+una+aproximaci%C3%B3n,+se+basa+en+la+suposici%C3%B3n+impl%C3%ADcita+de+que+los+n+procesos+son+independientes&source=web&ots=yrVyRAgL_Q&sig=ucwTFwjEzeSgpVyggm67BKYM7Ek&hl=es&sa=X&oi=book_result&resnum=1&ct=result#PPR7,M1. Fecha de recuperación: 08 de enero de 2009.

en ejecución y dos esperando. Sin embargo, con una sola CPU no es posible tener tres procesos ejecutándose al mismo tiempo, por lo que un proceso que pase al estado listo mientras la CPU está ocupada, tendrá que esperar, por lo que los procesos no son independientes.[35]

4.2. Particiones fijas y dinámicas

Generalmente, "la memoria está dividida en dos particiones: una para el sistema operativo residente y otra para los procesos de usuario".[36] Es posible colocar al sistema operativo en la memoria baja o en la memoria alta. El principal factor que afectará a esta decisión es la ubicación del vector de interrupción. Debido a que este vector generalmente se encuentra en la memoria baja, es más común colocar al sistema operativo en dicha memoria.

> Uno de los métodos más sencillos para la asignación de memoria consiste en dividirla en un número de particiones de tamaño fijo. Cada partición puede contener exactamente un proceso. De esta forma, el grado de multiprogramación está limitado por el número de particiones. Cuando una partición está libre, se selecciona un proceso de la cola de entrada y se carga en ella. Cuando un proceso termina, la partición queda disponible para otro proceso. Este método fue empleado originalmente por el sistema operativo OS/360 de IB (denominado MFT); en la actualidad ya no se usa.[37]

"El método de particiones dinámicas utiliza toda la memoria al cargar las primeras tareas en el sistema que no son del mismo tamaño de las que acaban de salir de la memoria y se acomodan en los espacios disponibles de acuerdo a su prioridad."[38]

---

[35] Cf. Andrew Stuart Tanenbaum, *Sistemas operativos modernos*, ed. cit., p. 193.
[36] Véase, F. Fernández, "Introducción a tecnología de objeto: memoria", en especial al diapositiva 14, material en línea de junio de 1999, disponible en: http://torio.unileon.es/~dielpa/asig/shannon/SO/teoria/memoria.ppt Fecha de recuperación: 08 de enero de 2009.
[37] Salvador Meza Badillo. *Sistemas operativos multiusuarios*, material en línea, disponible en: http://fcasua.contad.unam.mx/apuntes/interiores/docs/98/2/sis_operativos.pdf, p. 27. Fecha de recuperación: 08 de enero de 2009.
[38] Cf., "Administrador de la memoria: particiones dinámicas", material en línea, disponible en: http://www.mitecnologico.com/Main/AdministradorDeLaMemoria Fecha de recuperación: 08 de enero de 2009.

4.3. Asignación estática de la memoria

La asignación estática de memoria consiste en el proceso de asignar memoria en tiempo de compilación antes de que un programa pueda ser ejecutado, es decir se asigna la memoria a medida que se necesita.

Una aplicación de esta técnica conlleva que un módulo de programa (por ejemplo función o subrutina) declare datos estáticos de forma local, de tal forma que estos datos son inaccesibles desde otros módulos a menos que se les pasen referencias como parámetros o que les sean devueltos por la función.

El uso de variables estáticas dentro de una clase en la programación orientada a objetos permite que una copia individual de los datos se comparta entre todos los objetos de esa clase.

Las constantes conocidas en tiempo de compilación, como literales de tipo cadena, se asignan normalmente de forma estática. En programación orientada a objetos, el método usual para las tablas de clases también es la asignación estática de memoria.[39]

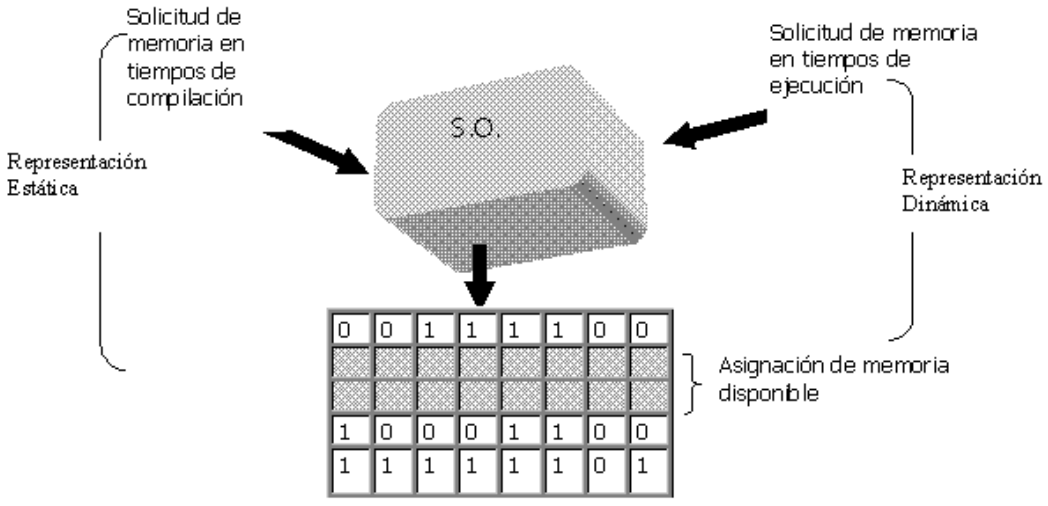


Figura 4.1 Asignación estática de la memoria

[39] Wikipedia: "Asignación de memoria", 25/11/08, disponible en: http://es.wikipedia.org/wiki/Asignaci%C3%B3n_din%C3%A1mica_de_memoria. Fecha de recuperación: 08 de enero de 2009.

4.4. Asignación dinámica de la memoria

Se refiere a la asignación de almacenamiento de memoria para su utilización por parte de un programa de cómputo durante el tiempo de ejecución de ese programa. Cuando el programa se va ejecutando, las direcciones relativas se pasan a absolutas tal como las líneas de código se ejecutan. (Instrucción por instrucción), de esta forma distribuye la propiedad de recursos de memoria limitada entre muchas piezas de código y datos. Un objeto asignado dinámicamente permanece así hasta que es designado explícitamente, o por el programador o por un recolector de datos.

4.5. Paginación[40]

Tanto las particiones de tamaño fijo como las de tamaño variable hacen un uso ineficiente de la memoria; las primeras generan fragmentación interna y las segundas originan fragmentación externa. La paginación[41] es un esquema que permite que el espacio de direcciones lógicas de un proceso no sea contiguo y esto evita el problema de ajustar las porciones de memoria de tamaño variable en el almacén de respaldo, del cual sufrían la mayoría de los esquemas anteriores de administración de la memoria, ya que dividen los programas en pequeñas partes o páginas y de esta forma la cantidad de memoria desperdiciada. Cuando algunos fragmentos de código de datos que residen en la memoria principal necesitan ser intercambiados, se debe encontrar estación en el almacén de respaldo.

La memoria física se descompone en bloques de tamaño fijo denominados marcos. La memoria lógica también se descompone en bloques del mismo tamaño denominados páginas. Cuando se va a ejecutar un proceso, sus páginas se cargan desde el almacén de respaldo en cualquier marco de memoria disponible.

---

[40] William Stallings, *Sistemas Operativos*, 4ª ed., Madrid, Pearson Education, 2001, pp. 306-309.
[41] Abraham Silbertschatz, *Sistemas Operativos,* 6ª ed., México, Limusa Wilsey, 2002, pp. 202-214.

El almacén de respaldo se divide en bloque de tamaño fijo que son del mismo tamaño que los marcos de la memoria.

En un cualquier momento, la memoria se encuentra ocupada con páginas de diferentes procesos, mientras que algunos marcos están disponibles para su uso. El sistema operativo mantiene una lista de estos últimos marcos, y una tabla por cada proceso, donde consta en qué marco se encuentra cada página del proceso. De esta forma, las páginas de un proceso pueden no estar contiguamente ubicadas en memoria, y pueden intercalarse con las páginas de otros procesos.

Cada dirección generada por la CPU se divide en dos partes: un número de página **(p)** y un desplazamiento de página **(d).** El número de página se emplea como un índice en una tabla de páginas. La tabla de páginas contiene la dirección base de cada página en la memoria física. Esta dirección base se combina con el desplazamiento de página para definir la dirección física de la memoria que se envía a la unidad de memoria.

La paginación misma es una forma de relocalización dinámica. El hardware de paginación vincula cada dirección lógica con alguna dirección física. La paginación es similar al uso de una tabla de registros base (de relocalización), uno para cada marco de la memoria.

Cuando utilizamos un esquema de paginación no tenemos fragmentación externa: cualquier marco libre puede ser asignado a un proceso que lo necesite. Sin embargo, podemos tener cierta fragmentación interna.

Cada sistema operativo tiene sus propios métodos para almacenar tablas de páginas. La mayoría asigna una tabla de páginas por cada proceso. Un apuntador a la tabla de páginas se almacena con los demás valores de registros (como el contador de instrucciones) en el bloque de control del proceso. Cuando se le dice al despachador que inicie un proceso, debe recargar los registros del usuario y

definir los valores correctos de la tabla de páginas de hardware a partir de la tabla de páginas del usuario que está almacenada.

La protección de memoria en un ambiente con paginación se realiza mediante bits de protección que están asociados con cada marco. Estos bits normalmente se mantienen en la tabla de páginas. Un bit puede definir que una página sea de lectura y escritura, o sólo de lectura. Cada referencia a la memoria pasa por la tabla de páginas para encontrar el número correcto de marco. La mayoría de los sistemas de cómputo modernos soportan un espacio grande de direcciones lógicas (2^{32} a 2^{64}). En tales ambientes, la tabla de páginas se vuelve excesivamente grande. Una solución sencilla a este problema consiste en dividir la tabla de páginas en piezas más pequeñas (paginación con niveles múltiples).

Otra ventaja de la paginación es la posibilidad de compartir un código común (páginas compartidas). Esta consideración es importante en un ambiente de tiempo compartido. El código reentrante (también denominado código puro) es un código que no puede modificarse a sí mismo. Si el código es reentrante, entonces nunca cambia durante la ejecución. Por lo tanto, dos o más procesos pueden ejecutar el mismo código al mismo tiempo. Cada proceso tiene su propia copia de registros y almacenamiento de datos para contener los datos para la ejecución del proceso. También se pueden compartir otros programas que se utilizan intensamente: compiladores, sistemas de ventanas, sistemas de bases de datos, etc.

4.6. Políticas de reemplazo de páginas

Cuando se presenta una falla de página, el sistema operativo tiene que escoger la página que desalojará de la memoria para hacer espacio y colocar la página que traerá del disco. Si la página a desalojar fue modificada mientras estaba en la memoria, deberá reescribirse en el disco para actualizar la copia. En cambio, si la

página no se ha modificado no será necesario reescribirla. La nueva página sobrescribe la que está desalojando. Aunque es posible escoger una página al azar para desalojarla cuando existe una falla de página, el desempeño del sistema mejora mucho si se escoge una página que no se usa con frecuencia. Si se desaloja una página muy utilizada, lo más seguro es que pronto se tenga que volver a traer a la memoria, con el consiguiente gasto extra. Entre las políticas de los algoritmos de reemplazo que se han considerado se incluyen los siguientes:

1. Óptima.
2. Usada menos recientemente (LRU, Least Recently Used).
3. Primera en entrar, primera en salir (FIFO).
4. Reloj.[42]

1. La política **óptima "**selecciona para reemplazar la página que tiene que esperar una mayor cantidad de tiempo hasta que se produzca la referencia siguiente". Este algoritmo es imposible de implementar, ya que requiere que el sistema operativo tenga un conocimiento exacto de los sucesos futuros. Sin embargo sirve como un estándar para comparar otros algoritmos.

2. La política **usada menos recientemente (LRU)** "reemplaza la página de memoria que no ha sido referenciada desde hace más tiempo. Debido al principio de cercanía, esta debería ser la página con menor probabilidad de ser referenciada en un futuro cercano", el problema de esta política es su método de implementación. "Una solución sería etiquetar cada página con el momento de su última referencia"[43]; esto tendría que hacerse con cada referencia a la memoria, tanto para instrucciones como datos. Incluso si el hardware respaldara este esquema, la sobrecarga sería muy fuerte.

---

[42] William Stallings, op. cit., pp. 344-351.
[43] Véase, ITESO, "Sistemas operativos: Memoria virtual", disponible en http://iteso.mx/~jluis/sopdf/material-oto-04/11-memoria-virtual.pdf Fecha de recuperación: 08 de enero de 2009.

3. La política de **primera en entrar primero en salir (FIFO)** “trata los marcos asignados a un proceso como un buffer circular y las páginas se suprimen de la memoria según la técnica de turno rotatorio (round-robin)”. Todo lo que se requiere es un apuntador que circule a través de los marcos del proceso. Esta es una de las políticas de reemplazo más sencillas de implementar, ya que su lógica es sencilla y consiste en reemplazar la página que ha estado más tiempo en la memoria.

4. La política de **reloj** es aquella que requiere “asociar un bit adicional a cada marco, denominado bit de uso. Cuando se carga una página por primera vez en un marco de memoria, el bit de uso de dicho marco se pone en cero. Cuando se hace referencia a la página posteriormente (después de la referencia que genero la falla en la página), el bit de uso se pone en 1”.[44] El algoritmo de reloj puede hacerse más potente incrementando el número de bits empleados. En todos los procesadores que ofrecen paginación se asocia un bit de modificación con cada página en la memoria principal y, por lo tanto con cada marco. Este bit es necesario para que, cuando se modifique una página, no se reemplace hasta volverla a escribir en la memoria secundaria.

4.7. Memoria virtual

“La memoria virtual es un concepto que permite al software usar más memoria principal” y permitir la ejecución de procesos que pueden no estar completamente en memoria.

> La mayoría de las computadoras tienen cuatro tipos de memoria: registros en la CPU, la memoria cache (dentro y fuera de la CPU), la memoria física (generalmente en forma de RAM, donde la CPU puede escribir y leer directa

[44] ITESO: “Sistemas operativos: Memorial virtual”, material en línea, disponible en: http://iteso.mx/~jluis/sopdf/material-oto-04/11-memoria-virtual.pdf. Fecha de recuperación: 08 de enero de 2009.

y razonablemente rápido) y el disco duro que es mucho más lento, pero más grande.[45]

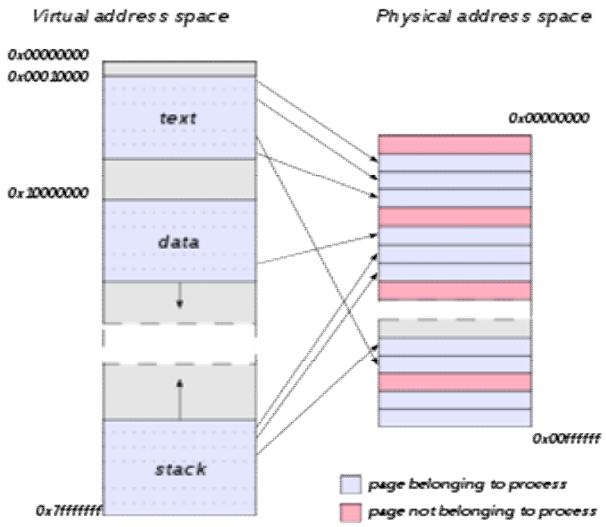


Figura 4.2 Memoria virtual[46]

"La memoria virtual es la separación de la memoria lógica del usuario y de la memoria física [del equipo]. Esta separación permite ofrecer a los programadores una memoria virtual extremadamente grande cuando sólo se dispone de una memoria física más pequeña".[47] Sin embargo, la memoria virtual no es fácil de implementar y puede reducir considerablemente el rendimiento si no se emplea con cuidado.

---

[45] Wikipedia: "Memoria virtual", actualizado por última vez el 05/01/09, disponible en: http://es.wikipedia.org/wiki/Memoria_virtual. Fecha de recuperación: 09 de enero de 2009.

[46] Wikipedia: "Memoria virtual", disponible en: http://upload.wikimedia.org/wikipedia/commons/thumb/3/32/Virtual_address_space_and_physical_address_space_relationship.svg/300px-Virtual_address_space_and_physical_address_space_relationship.svg.png. Recuperado el 08/01/09.

[47] Véase, Universidad de Jaén, Departamento de Informática: "Tema 7: Memoria virtual; 7.1." material electrónico disponible en: http://wwwdi.ujaen.es/~lina/TemasSO/MEMORIAVIRTUAL/1y2Motivaciones,ventajasyEstrategiasdeadministracion.htm Fecha de recuperación: 08 de enero de 2009.

La memoria virtual se implementa comúnmente mediante la paginación por demanda. También puede implementarse en un sistema con segmentación. Varios sistemas proporcionan un esquema de segmentación con paginación, en donde los segmentos descomponen las páginas. Así la visión del usuario es la segmentación, pero el sistema operativo puede implementar esta visión con paginación por demanda. La segmentación por demanda también puede proporcionar la memoria virtual.

En un sistema de paginación por demanda los procesos residen en la memoria secundaria (generalmente un disco). Cuando queremos ejecutar un proceso, lo intercambiamos a la memoria. Sin embargo, en lugar de intercambiar todo el proceso, empleamos un intercambiador perezoso. Un intercambiador perezoso nunca intercambia una página en la memoria a menos que dicha página se vaya a necesitar.

4.8. Memoria escondida (caché)

El término **memoria caché**[48] se aplica normalmente a una memoria más pequeña y más rápida que la memoria principal y que se sitúa entre ésta y el procesador. Este tipo de memoria reduce el tiempo medio de acceso a la memoria aprovechando el principio de cercanía. Este principio también puede aplicarse a la memoria de disco, es decir un caché de disco es una memoria intermedia (buffer), situada en la memoria principal para sectores de disco. La memoria contiene una copia de algunos sectores de disco. Cuando se hace una solicitud de E/S para un sector específico se comprueba si el sector está en la caché del disco. Si es así, la solicitud se satisface con la caché. Si no, se lee el sector solicitado del disco y se coloca en el caché. Debido al concepto de cercanía de referencias, cuando se

[48] Véase, Wikipedia: “Caché”, disponible en: http://es.wikipedia.org/wiki/Cach%C3%A9, recuperado el 08/01/09.

traiga un bloque de datos a la memoria caché para satisfacer una sola solicitud de E/S, será probable que se produzcan referencias futuras en el mismo bloque.

Consideraciones sobre el diseño:
Se deben considerar varios factores que influyen directamente en el rendimiento de la memoria y por lo tanto en su objetivo de aumentar la velocidad de respuesta de la jerarquía de memoria. Estos factores son las políticas de ubicación, extracción, reemplazo, escritura y el tamaño de la caché y de sus bloques.

Política de ubicación
Decide dónde debe colocarse un bloque de memoria principal que entra en la memoria caché. Las más utilizadas son:

- *Directa*: Al bloque $n-ésimo$ de memoria principal le corresponde la posición n *módulo* k donde k es el número de bloques de la memoria caché.
- *Asociativa*: Cualquier bloque de memoria principal puede ir en cualquier lado del bloque de memoria caché.
- *Asociativa por conjuntos*: La memoria caché se divide en n conjuntos de bloques, así al bloque $i-ésimo$ de memoria principal le corresponde el conjunto i *módulo* $\left(k/n\right)$ donde k es el número de bloques de memoria caché. Dicho bloque de memoria podrá ubicarse en cualquier posición dentro del conjunto asociado de la memoria caché.

Política de extracción
Esta política determina cuándo y qué bloque de memoria principal hay que traer a memoria caché.

- *Por demanda*: Un bloque sólo se trae a memoria caché cuando ha sido referenciado y se produzca una falla.
- *Con prebúsqueda*: Cuando se referencia el bloque $i-ésimo$ de memoria principal, se trae además el bloque $(i+1)-ésimo$. Ésta política se basa en la propiedad de localidad espacial de los programas.

Política de reemplazo
Esta política determina qué bloque de memoria caché debe abandonarla cuando no existe espacio disponible para un bloque entrante. Existen cuatro políticas:

- *Aleatoria*: El bloque es reemplazado de forma aleatoria.

- *FIFO*: Se usa un algoritmo *First In First Out* FIFO (PEPS, primero entrado primero salido en español) para determinar qué bloque debe abandonar la caché. Este algoritmo generalmente es poco eficiente.
- *Menos recientemente usado (LRU)*: Se sustituye el bloque que hace más tiempo no se ha utilizado.
- *Menos frecuentemente usado (LFU)*: Se reemplaza el bloque que se ha usado con menos frecuencia.

Siendo la aleatoria y la LRU las de mejor rendimiento.

Política de escritura

Esta política determina cuándo se actualiza la información en memoria principal cuando se ha escrito en memoria caché. Existen dos políticas principales:

- *Escritura inmediata* o *escritura directa* (*Write Through)*. Cuando se escribe en un bloque que se encuentra en memoria caché, la información se modifica también simultáneamente en memoria principal, manteniendo así la coherencia en todo momento. Suele combinarse con la técnica de "No carga en escritura" *(No Write Allocation)* que significa que, cuando haya que escribir en un bloque que no se encuentra en la caché, la modificación se realizará únicamente en memoria principal, sin traer dicho bloque a caché, y además sólo se actualizará la palabra concreta que haya cambiado.

- *Escritura aplazada* o *post-escritura*: En inglés *Write Back*. Cuando se escribe en un bloque que se encuentra en memoria caché, queda marcado como *basura* usando un BIT especial llamado normalmente *dirty BIT* o *BIT de basura*. Cuando el bloque sea desalojado de memoria caché (mediante la correspondiente política de reemplazo), se comprueba el BIT de basura, y si está activado se escribe la información de dicho bloque en memoria principal. Esta política suele combinarse con la técnica de "Carga en escritura" *(Write Allocation)*, que significa que, cuando haya que escribir en un bloque que no se encuentra en la caché, traeremos a caché el bloque en cuestión y lo modificaremos ahí. [49]

[49] Wikipedia: “Caché”, material en línea, disponible en: http://es.wikipedia.org/wiki/Cach%C3%A9. Fecha de recuperación: 08 de junio de 2009.

Bibliografía del tema 4

Silbertschatz, Abraham. *Sistemas Operativos*. 6ª ed., México. Limusa Wiley, 2002.

Stallings, William. Sistemas Operativos. 4ª ed., Madrid, Pearson Educación, 2001.

Tanenbaum, Andrew S. Sistemas Operativos Modernos. 2ª ed., México, Pearson Educación, 2003.

Actividades de aprendizaje

A.4.1. Describe y explica por medio de una gráfica la asignación estática de la memoria.

A.4.2. Realiza un cuadro comparativo sobre los esquemas de administración de memoria y explica brevemente la función que realizan cada uno de estos esquemas.

A.4.3. Realiza un cuadro comparativo sobre las políticas de reemplazo de páginas y explica brevemente la función de cada una de estas políticas.

Cuestionario de autoevaluación

1. ¿Qué es la paginación?
2. ¿En qué consiste la política de reemplazo de páginas “reloj”?
3. ¿En qué consiste la política de reemplazo de páginas “optima”?
4. ¿Cuáles son los factores que influyen directamente en el rendimiento de la memoria?
5. ¿Qué significa FIFO?
6. ¿Cuál es la función de la memoria cache?
7. ¿Que es la memoria virtual?

8. ¿Cómo se implementa la memoria virtual?

9. ¿Qué es la política de ubicación asociativa?

10. ¿Qué es la memoria ROM?

Examen de autoevaluación

1. ¿Qué permite que una página sea de lectura y escritura o solo de lectura?

a) Byte

b) Bit

c) Macros

2. La memoria física se descompone en bloques de tamaño fijo, ¿denominados?

a) páginas

b) marcos

c) diagramas

3. ¿Cuáles son las dos partes que divide la CPU por cada dirección generada?

a) un número de página (p) y un desplazamiento de páginas (d)

b) una memoria física(a) y una memoria lógica (b)

c) una dirección lógica (c) y una dirección física (d)

4. ¿Cuál es una ventaja de la paginación en un sistema operativo?

a) compartir un código común

b) compartir un dispositivo

c) compartir direcciones lógicas

5. ¿La paginación es un esquema que permite el espacio de direcciones?

a) lógicas

b) dinámicas

c) físicas

6. ¿Cuáles son los cuatro tipos de memoria que tiene una computadora?

a) memoria secundaria, memoria ROM, memoria RAM, memoria extendida

b) registros en la CPU, memoria cache, memoria física, disco duro

c) procesador, memoria física, disco duro, memoria RAM

7. ¿La memoria virtual puede implementarse por?

a) paginación

b) memoria caché

c) multiproceso

8. ¿En la paginación por demanda los procesos residen en?

a) memoria primeria

b) memoria secundaria

c) memoria extendida

9. ¿Qué reduce la memoria virtual cuando no se implementa con cuidado en un programa?

a) los bits

b) el rendimiento

c) los procesos

10. ¿Quién decide dónde debe colocarse un bloque de memoria principal que entra por la memoria caché?

a) la política de reemplazo

b) la política de extracción

c) la política de ubicación

TEMA 5. ADMINISTRACIÓN DE ARCHIVOS

Objetivo particular

Al finalizar el aprendizaje de este tema, el alumno reconocerá las características más importantes del sistema de administración de archivos y su impacto en las aplicaciones de un sistema de cómputo.

Temario detallado

5.1 Conceptos básicos de archivos
5.2 Directorios y nombres de archivos
5.3 Permisos
5.4 Los nodos-i de UNIX
5.5 Jerarquía de directorios
5.6 Administración de dispositivos de entrada y salida (E/S)
5.7 Copias de respaldo y compresión de archivos
5.8 Mantenimiento al sistema de archivos

Introducción

En la mayoría de las aplicaciones, el archivo es el elemento central. Proporciona el mecanismo para el almacenamiento y el acceso en línea a datos y programas que pertenecen al sistema operativo y a todos los usuarios del sistema de cómputo. El sistema de archivos consta de dos partes distintas: una colección de archivos, cada uno para el almacenamiento de datos relacionados, y una estructura de directorios, que organiza y proporciona información acerca de todos los archivos en el sistema. En este tema se abordarán los conceptos de archivo, su estructura y la forma en que los organiza el sistema operativo.

5.1 Conceptos básicos de archivos

El “Sistema de Archivos” es la parte del sistema de administración del almacenamiento responsable, principalmente, de la administración de los archivos del almacenamiento secundario. Los archivos (fuente) “son un mecanismo de abstracción que permite almacenar información en el disco y leerla después”. [50] Esto debe hacerse de modo que el usuario no tenga que enterarse de los detalles de cómo y dónde está almacenada la información y de cómo funciona en los discos de una computadora. Los archivos se pueden estructurar de varias maneras, las más comunes son:

1. Secuencia de bytes:
    a. El archivo es una serie no estructurada de bytes.
    b. Posee máxima flexibilidad.
    c. El sistema operativo no sabe que contiene el archivo.

2. Secuencia de registros:
    a. El archivo es una secuencia de registros de longitud fija, cada uno con su propia estructura interna.

3. Árbol:
    a. El archivo consta de un árbol de registros, no necesariamente de la misma longitud.
    b. Cada registro tiene un campo llamado key (llave o clave) en una posición fija del registro.
    c. El árbol se ordena mediante el campo de clave para permitir una rápida búsqueda de una clave particular.

Desde la perspectiva de un usuario, un archivo es la porción más pequeña de almacenamiento secundario lógico; es decir, no pueden escribirse datos en

[50] A.S. Tanenbaum, op. cit., pp. 382-383.

almacenamiento secundario a menos que se encuentren dentro de un archivo. La información de un archivo es definida por su creador. En un archivo se pueden almacenar diferentes tipos de información: programas fuente, programas objeto, programas ejecutables, datos numéricos, texto, registros de nómina, imágenes, grabaciones de sonido, etc.

Un archivo recibe un nombre, para conveniencia de sus usuarios, y se hace referencia a él por dicho nombre. Un nombre es generalmente una cadena de caracteres. Algunos sistemas distinguen entre mayúsculas y minúsculas en los nombres, en tanto que otros sistemas consideran los dos casos como equivalentes. Cuando se asigna un nombre a un archivo, éste se vuelve independiente del proceso del usuario, e incluso del sistema que lo creó.

Un archivo tiene generalmente los siguientes atributos:

1. **Nombre**: el nombre simbólico del archivo es la única información que se mantiene en forma legible para los humanos.

2. **Tipo**: esta información es necesaria para aquellos sistemas que soportan diferentes tipos.

3. **Ubicación**: esta información es un apuntador a un dispositivo y a la ubicación del archivo en dicho dispositivo.

4. **Tamaño**: en este atributo se incluyen el tamaño actual del archivo (en bytes, palabras o bloques) y, posiblemente, el tamaño máximo permitido.

5. **Protección**: información de control de acceso que determina quién puede leer, escribir, ejecutar, etc., el archivo.

6. **Hora, fecha e identificación del usuario**: esta información puede mantenerse para 1) la creación, 2) la última modificación y 3) el último uso. Estos datos pueden ser útiles para protección, seguridad y control de uso.[51]

[51] A. Silbertschatz, op. cit., pp. 346-351.

Para definir adecuadamente a los archivos, necesitamos considerar las operaciones que se pueden realizar sobre ellos. “El sistema operativo, -según Meza Badillo- proporciona llamadas al sistema para crear, escribir, leer, reposicionar, borrar y truncar archivos”. A continuación se describen las seis operaciones básicas sobre archivos:

I. Crear un archivo: se debe encontrar espacio para el archivo en el sistema de archivos y posteriormente se debe hacer una entrada en el directorio para el nuevo archivo. La entrada en el directorio registra el nombre del archivo y su ubicación en el sistema de archivos.

II. Escribir un archivo: se hace una llamada al sistema especificando tanto el nombre del archivo como la información que se va a escribir en él. El sistema debe mantener un apuntador de escritura a la ubicación en el archivo donde va a tener lugar la siguiente escritura. El apuntador de escritura debe actualizarse siempre que ocurre una escritura.

III. Leer un archivo: se hace una llamada al sistema que especifica el nombre del archivo y el lugar (en la memoria) donde deberá colocarse el siguiente bloque del mismo. Nuevamente, se busca en el directorio la entrada asociada, y el sistema mantiene un apuntador de lectura a la ubicación en el archivo en donde va a tener lugar la siguiente lectura. Una vez que se ha realizado la operación, el apuntador de lectura se actualiza. Tanto la operación de lectura como la de escritura emplean este mismo apuntador, ahorrando espacio y reduciendo la complejidad del sistema.

IV. Reposicionarse dentro de un archivo: se busca en el directorio la entrada apropiada, y se asigna un valor dado a la posición actual del archivo. El reposicionamiento dentro de un archivo no necesita incluir una operación real de E/S. Esta operación sobre el archivo también se conoce como búsqueda en archivo.

V. Borrar un archivo: se busca en el directorio el archivo designado. Una vez que se ha encontrado la entrada asociada, se libera todo el espacio del archivo (para que pueda ser reutilizado por otros archivos) y se borra la entrada del directorio.

VI. Truncar un archivo: hay ocasiones en que el usuario desea que los atributos de un archivo permanezcan iguales, pero quiere borrar el contenido del archivo. En lugar de obligar al usuario a borrar el archivo y después volver a crearlo, esta función permite que todos los atributos

permanezcan sin modificación (excepto la longitud del archivo), pero restableciendo el archivo a una longitud cero.[52]

Tipo de archivo	*Extensión usual*	*Función*
Ejecutable	exe, com, bin ó ninguna	Programa en lenguaje de máquina listo para correr
Objeto	obj, o	Compilado, en lenguaje de máquina, no enlazado
Código fuente	c, cc, pas, java, asm, a	Código fuente en varios lenguajes
Por lotes	bat, sh	Comandos al intérprete de comandos
Texto	txt, doc	Datos textuales, documentos
Procesador de palabras	wpd, tex, doc, etcétera	Varios formatos de procesador de palabras
Biblioteca	lib, a, DLL	Bibliotecas de rutinas para programadores
Impresión o vista	ps, dvi, gif	Archivo ASCII o binario en un formato para impresión o vista
Archivos	arc, zip, tar	Archivos relacionados agrupados en un archivo, a veces comprimido, para archivarlo o almacenarlo

Figura 5.1 Tipos comunes de archivos

5.2 Directorios y nombres de archivos

Para llevar el control de los archivos, el sistema de archivos comúnmente tiene directorios o carpetas. La forma más sencilla del sistema de directorios es que un directorio contenga todos los archivos, a veces se le llama directorio raíz.

[52] Salvador Meza Badillo. Sistemas operativos multiusuarios en: http://fcasua.contad.unam.mx/apuntes/interiores/docs/98/2/sis_operativos.pdf. Fecha de recuperación: 08 de enero de 2009.

> Los sistemas de archivos de las computadoras pueden ser extensos. Algunos sistemas almacenan miles de archivos en cientos de gigabytes de disco. Para manejar todos estos datos, se necesitan organizar. Esta organización generalmente se realiza en dos partes.[53]

1. El sistema de archivos se descompone en particiones, también conocida como minidiscos, cada disco en un sistema contiene por lo menos una partición, que es una estructura de bajo nivel en la que residen archivos y directorios. Algunos sistemas utilizan particiones para proporcionar varias áreas separadas dentro de un disco, tratando a cada una como un dispositivo de almacenamiento distinto, y otros sistemas permiten que las particiones sean más grandes que un disco de manera que puedan agrupar discos en una estructura lógica. De esta forma, el usuario sólo necesita preocuparse de la estructura lógica de directorios y archivos; puede ignorar completamente los problemas de la asignación física de espacio para los archivos. Por esta razón, las particiones pueden ser consideradas como discos virtuales.

2. Cada partición contiene información acerca de los archivos dentro de ella. Esta información se mantiene en entradas en un directorio del dispositivo o tabla de contenido del volumen. El directorio del dispositivo (comúnmente conocido sólo como directorio) registra información –como nombre, ubicación, tamaño y tipo– de todos los archivos en dicha partición.

Operaciones que se realizan a los directorios:

- **Buscar un archivo:** consiste en hacer una búsqueda de una estructura de directorios para encontrar la entrada para un archivo particular.
- **Crear un archivo:** consiste en crear nuevos archivos y agregarlos al directorio

[53] A. Silbertschatz, op. cit., pp.357-359.

- **Borrar un archivo:** cuando un archivo ya no se necesita, se requiere poder removerlo del directorio

- **Listar un directorio:** consiste en obtener una lista de los archivos en un directorio y el contenido de la entrada del directorio para cada archivo de la lista.

- **Renombrar un archivo:** el nombre de un archivo representa su contenido, es necesario poder cambiar tal nombre cuando cambia el contenido o el uso del archivo. Renombrar un archivo también puede permitir que se modifique su posición dentro de la estructura de directorios.

- **Recorrer el sistema de archivos:** consiste en tener acceso a cada directorio y a cada archivo dentro de una estructura de directorios. Es recomendable realizar una copia de todos los archivos en cinta magnética. Esta técnica proporciona una copia de respaldo en caso de una falla del sistema o si el archivo simplemente ya no está en uso. En este caso, el archivo puede copiarse en una cinta y liberar el espacio en disco de dicho archivo para que pueda ser utilizado por otro.

5.3 Permisos[54]

La necesidad de proteger archivos es un resultado directo de la capacidad para acceder a archivos. En los sistemas que no permiten el acceso a archivos de otros usuarios, la protección no es necesaria.

Los mecanismos de protección proporcionan un acceso controlado limitando los tipos de acceso que pueden hacerse a los archivos. El acceso se permite o se niega dependiendo de varios factores, uno de los cuales es el tipo de acceso solicitado. A continuación se describen los diversos tipos de operaciones:

- *Leer*: leer un archivo.
- *Escribir*: escribir o volver a escribir el archivo.
- *Ejecutar*: cargar el archivo en memoria y ejecutarlo.
- *Anexar*: escribir nueva información al final del archivo.
- *Borrar*: borrar el archivo y liberar su espacio para una posible reutilización.
- *Listar*: listar el nombre y los atributos del archivo.

También se pueden controlar otras operaciones, tales como; renombrar, copiar o editar el archivo. Sin embargo, en el caso de muchos sistemas, estas funciones de alto nivel (como copiar) pueden implantarse mediante un programa de sistema que realice llamadas al sistema de bajo nivel. La protección sólo se proporciona en el nivel inferior.

Se han propuesto muchos mecanismos de protección diferentes. Cada esquema tiene sus ventajas y desventajas, por lo que cada quien debe seleccionar el apropiado para la aplicación deseada. Esto depende del tipo de protección que requiera cada sistema de cómputo en particular.

---

[54] ibid., 369-370.

El enfoque más común para el problema de la protección consiste en hacer que el acceso dependa de la identidad del usuario. Varios usuarios pueden necesitar diferentes tipos de acceso a un archivo o directorio. El esquema más general para implementar un acceso que dependa de la identidad consiste en asociar una lista de acceso con cada archivo y directorio, especificando para cada usuario de la lista el nombre y los tipos de acceso permitidos.

El principal problema con las listas de acceso es su longitud. Si se permite que todos puedan leer el archivo, debemos listar a todos los usuarios y concederles acceso a lectura. Esta técnica tiene dos consecuencias no deseables:

1. La construcción de la lista puede ser una tarea tediosa que no ofrece alguna utilidad, especialmente si no se conoce por adelantado la lista de usuarios del sistema.
2. La entrada del directorio que anteriormente era de tamaño fijo ahora necesita ser de tamaño variable, haciendo que la administración de espacio sea más compleja.

Este problema se resuelve empleando una versión condensada de la lista de acceso.

Para condensar la longitud de la lista de acceso, muchos sistemas reconocen tres clasificaciones de usuarios con relación a cada archivo:

- **Propietario:** el usuario que creó el archivo es el propietario.
- **Grupo:** un conjunto de usuarios que están compartiendo el archivo y necesitan acceso similar es un grupo, o grupo de trabajo.
- **Universo:** todos los demás usuarios del sistema constituyen el universo.

5.4 Los nodos-i de UNIX[55]

El método para llevar el control de qué bloques pertenecen a qué archivos consiste en asociar a cada archivo una estructura de datos llamada **nodo-i (nodo índice)**. La ventaja principal de este esquema es que el nodo-i solo tiene que estar en la memoria cuando el archivo correspondiente está abierto. El sistema de archivos UNIX tiene la forma de un árbol que nace en el directorio raíz, con la adición de enlaces para formar una grafica acíclica dirigida. Una entrada de directorio UNIX contiene una entrada para cada archivo de ese directorio, estas entradas utilizan el esquema de nodos-i. Una entrada de directorio contiene dos campos: el nombre de archivo (14 bytes) y el nodo $-i$ correspondiente a ese archivo (2 bytes). Estos parámetros limitan el número de archivos por sistema de archivos a $64k$. Los nodos $-i$ de UNIX contienen atributos tales como; tamaño del archivo, hora de creación, último acceso, última modificación, dueño, grupo, información de protección y una cuenta del número de entradas de directorio que apuntan al nodo-i.

5.5 Jerarquía de directorios

El *número y organización de directorios* varía según el sistema, a continuación se describen las principales jerarquías:

Directorio de un solo nivel

Todos los archivos están contenidos en el mismo directorio, el cual es fácil de soportar y entender. Sin embargo, cuando aumenta el número de archivos o cuando hay más de un usuario un directorio de un solo nivel tiene limitaciones considerables. Debido a que todos los archivos están en el mismo directorio, deben tener nombres únicos.

---

[55] A.S. Tanenbaum, op. cit., pp. 445-448.

La principal desventaja de un directorio de un solo nivel es la confusión de los nombres de archivos creados por usuarios diferentes. La solución estándar consiste en crear un directorio distinto para cada usuario.

Directorio de dos niveles

Cada usuario tiene su propio directorio de archivos de usuario (user file directory, UFD). Cada UFD tiene una estructura similar, pero lista sólo los archivos de un usuario. Cuando comienza un trabajo de usuario o se conecta un usuario, se hace una búsqueda en el directorio de archivos maestro (master file directory, MFD). El MFD está indexado por el nombre de usuario o el número de cuenta, y cada entrada apunta al UFD para dicho usuario.

Directorios con estructura de árbol

Se puede visualizar como un árbol de dos niveles, la generalización natural consiste en extender la estructura del directorio a un árbol de altura arbitraria. Esta generalización permite a los usuarios crear sus propios subdirectorios y organizar sus archivos con base en esto. El sistema MS-DOS, por ejemplo, está estructurado como un árbol. El árbol tiene un directorio raíz. Cada archivo del sistema tiene un nombre de ruta único.

Un directorio (o subdirectorio) contiene un conjunto de archivos o subdirectorios. Un directorio es simplemente otro archivo, pero es tratado en una forma especial. Todos los directorios tienen el mismo formato interno. Un bit en cada entrada del directorio define la entrada como un archivo (0) o como un subdirectorio (1). Ciertas llamadas especiales al sistema crean y borran directorios.

Con un sistema de directorios con estructura de árbol, los usuarios pueden tener acceso a los archivos de otros usuarios, además de sus propios archivos.

Directorios de gráfica acíclica

Una estructura de árbol prohíbe el compartimiento de archivos o directorios. Una gráfica acíclica (gráfica sin ciclos) permite que los directorios tengan subdirectorios y archivos compartidos. El mismo archivo o subdirectorio puede estar en dos directorios diferentes. Una gráfica acíclica es una generalización natural del esquema de directorios con estructura de árbol. Cuando varias personas están trabajando como equipo, todos los archivos que se van a compartir pueden colocarse juntos en un directorio. Cada uno de los directorios de archivos de usuario de todos los miembros del equipo contiene este directorio de archivos compartidos como un subdirectorio.

Una estructura de directorios de gráfica acíclica es más flexible que una estructura sencilla de árbol, pero también es más compleja.

5.6 Administración de dispositivos de entrada y salida (E/S)[56]

Uno de los aspectos más confusos en el diseño de los sistemas operativos es la entrada y salida (E/S). Debido a la amplia variedad de dispositivos y aplicaciones de esos dispositivos, es difícil desarrollar una solución general y consistente. Los dispositivos externos que tienen que hacer funcionar la E/S en los sistemas informáticos se clasifican en tres categorías:

- Dispositivos legibles por los humanos: son apropiados para la comunicación con el usuario. Ejemplo: terminales de video, teclados, pantallas, impresoras, etc.
- Dispositivos legibles por la máquina: son adecuados para comunicarse con equipos electrónicos. Ejemplo: discos, unidades de cinta, sensores, controladores e impulsores.

[56] W. Stallings, op. cit., pp. 462-463.

- Dispositivos de comunicaciones: apropiados para comunicarse con dispositivos lejanos. Ejemplo: adaptadores de líneas digitales, módem, etc.
-

Existen grandes diferencias entre las clases de dispositivos y son:

- Velocidad de los datos (teclado, disco duro, modem, ratos, etc.).
- Aplicaciones (utilidad que se le da a un dispositivo; disco de archivos, disco de aplicaciones).
- Complejidad de control (interfaz de impresora, interfaz de disco, etc.).
- Unidad de transferencia (flujo de bytes, bloques de E/S a disco).
- Representación de los datos (codificación de datos, convenios de paridad).
- Condiciones de error (naturaleza de errores, consecuencias, etc.).

5.7 Copias de respaldo y compresión de archivos

Se debe asegurar que los datos no se pierdan en caso de una falla. Para esto podemos emplear programas de sistema para respaldar datos del disco a otro dispositivo de almacenamiento, como un disco flexible, una cinta magnética o un disco óptico. La recuperación de la pérdida de un archivo individual, o de todo un disco, puede implicar simplemente restablecer los datos a partir del respaldo.

Para minimizar el copiado requerido, podemos utilizar la información de cada entrada del archivo en el directorio. Por ejemplo, si el programa de respaldo sabe cuándo se realizó el último respaldo de un archivo, y la fecha de la última escritura del archivo en el directorio indica que el archivo no ha cambiado desde ese momento, entonces el archivo no necesita copiarse nuevamente.

Los tipos de copias de seguridad pueden ser:

- **Normal:** conocido como **respaldo completo**, se copian todos los archivos y carpetas seleccionados. Este tipo de respaldo no toma en cuenta los

marcadores (bits) para determinar qué archivos, elimina el atributo de archivo de todos los archivos que se van a respaldar.

- **Copia:** en este tipo de respaldo se realiza una copia de seguridad de todos los archivos y carpetas seleccionados y no se buscan ni se borran los marcadores.
- **Diferencial:** en este tipo solo se realiza una copia de seguridad de los archivos y carpetas seleccionados que tienen un marcador. Este tipo de respaldo es moderadamente rápida en la copia y restauración de los datos.
- **Incremental:** en este solo se realizan copia de seguridad de los archivos y carpetas que tienen un marcador.
- **Diaria:** realiza copia de seguridad de todos los archivos y carpetas seleccionados que han cambiado durante el día. Una copia de este tipo no busca ni borra los marcadores.

Ejemplo de un plan de respaldos:

a. Día 1: copiar en un medio de respaldo todos los archivos del disco (respaldo completo).

b. Día 2: copiar en otro medio todos los archivos modificados desde el día 1 (respaldo incremental).

c. Día 3: copiar en otro medio todos los archivos modificados desde el día 2.

d. Día N: copiar en otro medio todos los archivos modificados desde el día N – 1. Luego, regresar al día 1.

El respaldo del nuevo ciclo puede escribirse sobre el conjunto anterior, o sobre un nuevo conjunto de medios de respaldo. Una ventaja de este ciclo de respaldo es que se puede restablecer cualquier archivo que se haya borrado accidentalmente durante el ciclo, recuperando el archivo borrado a partir del respaldo del día anterior. También es muy importante almacenar estos respaldos permanentes en un lugar alejado de los respaldos regulares, como protección contra peligros como

un incendio que pudiera destruir el contenido de la computadora y también todos los respaldos.

Si en este proceso se reutilizan los medios, se debe tener mucho cuidado de no reutilizar estos medios demasiadas veces (desgaste físico), ya que puede no recuperarse el respaldo.

La **compresión de datos** consiste en la reducción del volumen de información tratable (procesar, transmitir o grabar), con esto se pretende transportar la misma información, pero empleando la menor cantidad de espacio.

La compresión de datos se basa fundamentalmente en buscar repeticiones en series de datos para después almacenar solo el dato junto al número de veces que se repite. Así, por ejemplo, si en un archivo aparece una secuencia como "AAAAAA", ocupando 6 bytes se podría almacenar simplemente "6A" que ocupa solo 2 bytes.

El proceso de compresión es más complejo, ya que raramente se consigue encontrar patrones de repetición tan exactos (salvo en algunas imágenes). Existen algoritmos de compresión como los siguientes.

- Algoritmos que buscan series largas que luego se codifican en formas más reducidas.
- Algoritmos que examinan los caracteres más repetidos para luego codificar de forma más corta los que más se repiten (algoritmo de Huffman).
- Algoritmos que construyen un diccionario con los patrones encontrados, a los cuales se hace referencia de manera posterior.

En la compresión hay que tomar en cuenta dos conceptos:

- **Redundancia:** Datos que son repetitivos o previsibles
- **Entropía:** La información nueva o esencial que se define como la diferencia entre la cantidad total de datos de un mensaje y su redundancia.

La información que transmiten los datos puede ser de tres tipos:

- **Redundante:** información repetitiva o predecible.

- **Irrelevante:** información que no podemos apreciar y cuya eliminación por tanto no afecta al contenido del mensaje. Por ejemplo, si las frecuencias que es capaz de captar el oído humano están entre 16/20 Hz y 16.000/20.000 Hz s, serían irrelevantes aquellas frecuencias que estuvieran por debajo o por encima de estos valores.
- **Básica:** la relevante. La que no es ni redundante ni irrelevante. La que debe ser transmitida para que se pueda reconstruir la señal.

Teniendo en cuenta estos tres tipos de información, se establecen tres tipos de compresión de la información:

- **Sin pérdidas reales:** transmitiendo toda la entropía del mensaje (toda la información básica e irrelevante, pero eliminando la redundante).
- **Subjetivamente sin pérdidas:** además de eliminar la información redundante se elimina también la irrelevante.
- **Subjetivamente con pérdidas:** se elimina cierta cantidad de información básica, por lo que el mensaje se reconstruirá con errores perceptibles pero tolerables (por ejemplo: la videoconferencia)[57].

5.8 Mantenimiento al sistema de archivos[58]

Con las limitaciones que existen en el espacio de los discos duros es necesario reutilizar este espacio para que sea utilizado por nuevos archivos, existen dispositivos que solo permiten una escritura en cualquier sector, por lo que no es posible su reutilización. En los discos duros el sistema mantiene una lista de espacio libre en la que registra todos los bloques del disco que están libres (no asignados a algún archivo o directorio). Para crear un archivo buscamos en la lista de espacio libre la cantidad de espacio requerido, y asignamos dicho espacio al nuevo archivo. Este espacio se remueve después de la lista de espacio libre. Cuando se borra un archivo, su espacio en disco se agrega a la lista de espacio libre. La lista de espacio libre, a pesar de su nombre, podría no estar

[57] Wikipedia: “Compresión de datos”, material electrónico actualizado por última vez en 14/11/08, disponible en http://es.wikipedia.org/wiki/Compresi%C3%B3n_de_datos. Fecha de recuperación: 08 de enero de 2009.
[58] A. Silbertschatz, op. cit., pp. 386-388.

implementada como una lista. La lista de espacio libre se implementa de la siguiente forma:

1. Vector de bits. Cada bloque se representa mediante 1 bit. Si el bloque está libre, el bit es 1; si el bloque está asignado, el bit es 0. La principal ventaja de este enfoque es que es sencillo y eficiente encontrar el primer bloque libre, o n bloques libres consecutivos en el disco, muchas computadoras incluyen instrucciones para la manipulación de bits que pueden usarse eficazmente para este fin. Por ejemplo la familia Intel, a partir del procesador 80386, y la familia Motorola, desde el procesador 68020.

2. Lista enlazada. Consiste en enlazar todos los bloques libres del disco, manteniendo un apuntador al primer bloque libre en una localidad especial en el disco y colocándolo en caché en memoria. Este primer bloque contiene un apuntador en el siguiente bloque libre en el disco, y así sucesivamente. Sin embargo, este esquema no es eficiente; para recorrer la lista, ya que se debe leer cada bloque, lo cual requiere una cantidad considerable de tiempo de E/S. Por lo general, el sistema operativo necesita sólo un bloque libre para poder asignar dicho bloque a un archivo, por lo que se utiliza el primer bloque en la lista de bloques libres.

3. Agrupación. Una modificación del enfoque de la lista de bloques libres consiste en almacenar las direcciones de n bloques libres en el primer bloque libre. Los primeros $n-1$ de estos bloques están efectivamente libres. El bloque final contiene las direcciones de otros n bloques libres, y así sucesivamente. La importancia de esta implementación es que se pueden encontrar rápidamente las direcciones de un gran número de bloques libres, a diferencia del enfoque estándar de lista enlazada.

4. Conteo. Consiste en aprovechar el hecho de que, generalmente, varios bloques contiguos pueden ser asignados o liberados de manera simultánea, particularmente cuando se asigna espacio con el algoritmo de asignación contigua o mediante agrupamientos. Así, en lugar de mantener una lista de n direcciones de disco libres, puede mantener la dirección del primer bloque libre y el número n de bloques contiguos libres que siguen al primer bloque. Cada entrada en la lista de espacio libre consiste entonces en una dirección de disco y una cuenta. Aunque cada entrada requiere más espacio del que utilizaría una dirección de disco sencilla, la lista global será más corta, siembre y cuando la cuenta sea generalmente mayor que 1.

Se deben de considerar también los siguientes aspectos:

- La forma de almacenamiento de archivos y directorios.
- La administración del espacio en disco.
- La forma de hacerlo de manera eficiente y confiable.

Se deben tener presentes los siguientes problemas que ocasiona la *"fragmentación"* creciente del espacio en el disco duro:

- Ocasiona problemas de performance al hacer que los archivos se desperdiguen a través de bloques muy dispersos.

Una técnica para aliviar el problema de la *"fragmentación"* consiste en realizar periódicamente:

- *"Condensación":* se pueden *"reorganizar"* los archivos expresamente o automáticamente según algún criterio predefinido.
- "*Recolección de basura o residuos*": se puede hacer fuera de línea o en línea, con el sistema activo.

Bibliografía del tema 5

Tanenbaum, Andrew S. *Sistemas Operativos Modernos.* 2ª ed., México, Pearson Educación, 2003.

Silbertschatz, Abraham. *Sistemas Operativos*. 6ª ed., México. Limusa Wiley, 2002.

Stallings, William. Sistemas Operativos. 4ª ed., Madrid, Pearson Educación, 2001.

Actividades de aprendizaje

A.5.1. Realiza un cuadro sinóptico explicando cada uno de los atributos que pueden tener los archivos.

A.5.2. Realiza un cuadro comparativo sobre las 6 operaciones básicas que pueden realizarse a los archivos.

A.5.3. Realiza una tabla sobre los diferentes tipos de archivos descritos en este tema y da un ejemplo de cada uno de ellos.

Cuestionario de autoevaluación

1. ¿Qué es un archivo?
2.- ¿Qué se puede almacenar en un archivo de datos?
3. ¿Qué otro nombre reciben las particiones en los discos?
4. ¿Cuáles son las operaciones que se realizan en los directorios?
5. ¿Cómo se clasifican los usuarios en relación a cada archivo?
6. ¿A qué se refiere el directorio de un solo nivel?
7. ¿Qué significa el término UFD?
8. ¿Qué permite realizar una grafica aciclica?
9. ¿Cuáles son las 3 categorías de la administración de un dispositivo de E/S?
10. ¿En qué consiste la comprensión de datos?

Examen de autoevaluación

1. ¿Cómo se le llama al directorio principal que contiene todos los archivos de un sistema?

a) directorio padre

b) directorio raíz

c) directorio hijo

2. ¿Un sistema de archivos se descompone en particiones llamados?

a) minidiscos

b) memorias

c) directorio

3. ¿Qué es lo que registra el directorio cuando se realiza una partición?

a) nombre, ubicación, tamaño, tipo

b) tamaño, espacio, apellidos, dirección

c) ubicación, tamaño, dirección, tipo

4. ¿Cuáles son los diversos tipos de operaciones que proporcionan la protección de archivos?

a) leer, escribir, ejecutar, anexar, borrar

b) listar, ejecutar, sobre escribir, leer

c) ejecutar, leer, borrar, cambiar nombre

5. ¿Cuáles son las consecuencias NO deseables para conceder el acceso de lectura de un archivo a los usuarios?

a) la construcción de lista, la entrada del directorio

b) la capacidad de memoria, la ejecución del archivo

c) la lista de archivos, el renombre del archivo

6. ¿El respaldo “normal” es considerado como?

a) respaldo completo

b) respaldo incremental

c) respaldo diario

7. ¿Cuáles son los atributos de un nodo i de UNIX?

a) tamaño del archivo, hora de creación, ultimo acceso

b) sesión, grupo, hora de termino, memoria

c) cuenta de usuario, ultimo acceso, tamaño del archivo, disco

8. ¿Cuál es la principal desventaja de un directorio de un solo nivel?

a) confusión de nombres de archivos creados por usuarios diferentes

b) confusión de nombres de directorios

c) confusión de los atributos asociados a un directorio

9. ¿Cuáles son las siglas que identifican el directorio de archivo maestro?

a) MDF

b) MFD

c) FMD

10. ¿Cuáles son tres tipos de respaldo de seguridad?

a) normal, diaria, incremental

b) diferencial, comprensión de datos, fragmentación

c) backup, restore, respaldar

TEMA 6. SEGURIDAD

Objetivo particular

Al finalizar el aprendizaje de este tema, el alumno reconocerá el concepto general de seguridad, así como los mecanismos de protección que el sistema operativo requiere para proteger la información de una computadora.

Temario detallado

6.1 Conceptos básicos de seguridad

6.2 Encriptamiento sencillo con llave secreta

6.3 Encriptamiento con llave pública

6.4 Estándares de criptografía

6.5 Capacidades, derechos y matriz de acceso

6.6 Virus y sus variantes

6.7 Contraseñas de una sola vez

6.8 Amenazas, ataques y vigilancia

6.9 Reconstrucción de un sistema violado

6.10 La bitácora o diario de operaciones

Introducción

El concepto de seguridad informática es muy amplio y se tiene que definir claramente en donde desea implementarse, puede aplicarse a las bases de datos, a las redes de computadoras, a los sistemas operativos, etc. La seguridad de los sistemas operativos es solo una pequeña parte del problema total de la seguridad en los sistemas de computación, pero éste viene incrementándose en gran medida. Si tomamos en cuenta que todo software no está libre de fallas, entonces un software complejo como lo es un sistema operativo es probable que falle y un porcentaje de estas fallas afecte a la seguridad.

La única manera razonable de probar la seguridad de un sistema es realizar evaluaciones de seguridad en él. Sin embargo, cuanto más complejo es el sistema, más difícil se vuelve la evaluación de su seguridad. Un sistema más complejo tendrá más errores relacionados con la seguridad en su análisis, diseño y programación. Y desgraciadamente, el número de errores y la dificultad de evaluación no crecen de acuerdo con la complejidad, crece mucho más rápido.[59]

En este tema se describen los conceptos y mecanismos de seguridad que se aplican a los sistemas operativos multiusuario.

6.1 Conceptos básicos de seguridad

La seguridad tiene muchas facetas. Tres de las más importantes son la naturaleza de las amenazas, la naturaleza de los intrusos y la pérdida accidental de datos.

- **Amenazas**

Desde una perspectiva de seguridad, los sistemas de computación tienen tres metas generales, con sus respectivas amenazas, como se muestra en la figura 6-1. La primera, la confidencialidad de los datos, tiene que ver con mantener en secreto los datos. De manera más específica, si el dueño de ciertos datos ha decidido que sólo deben proporcionarse a ciertas personas y a nadie más, el sistema debe garantizar que los datos no se proporcionarán a personas no autorizadas. Como mínimo, el dueño deberá poder especificar quién puede ver qué, y el sistema deberá hacer que se respeten esas especificaciones.

META	AMENAZA
Confidencialidad de los datos	Revelación de los datos
Integridad de los datos	Alteración de los datos
Disponibilidad del sistema	Negación del servicio

Figura 6.1 Metas de seguridad y sus amenazas[60]

[59] VV. AA., Facultad de Ciencias Exactas y Agromesura, UNNE: *Seguridad en sistemas operativos"*, material electrónico, disponible en: http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MonogSO/SEGSO00.htm, recuperado el 08/01/09.

[60] A. S. Tanenbaum, op. cit., p.584.

La segunda meta, la integridad de los datos, implica que los usuarios no autorizados no podrán modificar ningún dato sin permiso del dueño. En este contexto, la modificación de los datos no sólo incluye alterarlos, sino también eliminar datos y añadir datos falsos. Si un sistema no puede garantizar que los datos almacenados en él permanecerán sin cambios hasta que el usuario decida modificarlos, no sirve de mucho como sistema de información.

La tercera meta es la disponibilidad del sistema que implica que nadie podrá alterar el sistema de modo que no pueda usarse. Los ataques de negación del servicio son cada vez, más comunes.

Por ejemplo, si una computadora es un servidor de Internet, el envío de una fuerte cantidad de solicitudes a ese servidor podría bloquearlo, obligándolo a gastar todos los recursos de CPU en examinar y desechar las solicitudes que llegan. Actualmente se cuenta con modelos y tecnologías adecuadas para detener los ataques contra la confidencialidad y la integridad; detener los ataques de negación de servicio es mucho más difícil.

- **Intrusos**

Por desgracia existen personas que causan ataques y daños a las redes y a los sistemas. En la bibliografía sobre seguridad, a quienes realizan este tipo de actividades se les denomina intrusos o, a veces, adversarios. Los intrusos actúan de dos maneras. Los intrusos pasivos que sólo quieren leer archivos que no están autorizados para leer. Los intrusos activos realizan cambios no autorizados a los datos. Al diseñar un sistema es muy importante considerar la posibilidad de sufrir este tipo de ataques.

Los ataques más comunes se clasifican de la siguiente forma:

a. Curiosidad de usuarios no técnicos. Se refiere a personas que no tienen conocimientos técnicos pero que cuentan con una computadora que puede conectarse a las redes como Internet, por ejemplo, que les permite bajar programas que pueden utilizar para conectarse a computadoras que están desprotegidas y esto les permita leer el correo electrónico, archivos, etc. También pueden deberse a la forma en que trabajan los sistemas operativos. En casi todos los sistemas UNIX los archivos recién creados son públicos de manera predeterminada.

b. Ataques por parte de personal interno. Los estudiantes, programadores de sistemas, operadores y demás personal técnico muchas veces consideran como un desafío personal encontrar la forma de violar la seguridad de un sistema de computación local. En muchos casos estas personas tienen

grandes habilidades y están dispuestas a dedicar un tiempo considerable a superar ese reto.

c. Intentos decididos por hacer dinero. Algunos programadores de bancos han intentado robar el banco para el que trabajaban. Las técnicas han variado desde modificar el software hasta truncar en vez de redondear las cifras de intereses, quedándose con la fracción de centavo; desde desfalcar cuentas que no se han usado en varios años hasta chantaje ("Páguenme o destruiré todos los registros del banco").[61]

d. Espionaje comercial o militar. El espionaje se refiere a un intento serio y bien financiado por robar programas, secretos industriales, ideas, patentes, tecnología, diseños de circuitos, planes de negocios, etc., Esto lo puede realizar un competidor e inclusive un país extranjero. En muchos casos el intento implica intervención de líneas o incluso levantar antenas dirigidas a la computadora para capturar su radiación electromagnética.

Otra categoría son los virus que cada ves son mas sofisticados e inclusive difícil de detectar. Un virus es un fragmento de código que se reproduce por su cuenta y (por lo regular) causa algún daño. En cierto sentido, el creador de un virus también es un intruso, y a menudo posee habilidades técnicas considerables. La diferencia entre un intruso convencional y un virus es que el primero es alguien que está tratando de violar un sistema en forma personal para causar daño, mientras que el segundo es un programa escrito por una persona mal intencionada que lo difunde por el mundo para también causar daños. Los intrusos tratan de penetrar en sistemas específicos para robar o destruir datos específicos, mientras que un virus casi siempre causa daños más generales.

[61] ibid., 583 y ss.

Pérdida accidental de datos

Además de las amenazas provenientes de intrusos mal intencionados, es posible perder por accidente datos valiosos. Entre las causas más comunes de la pérdida accidental de datos están:

1. *Actos fortuitos*: incendios, inundaciones, terremotos, guerras, motines o roedores que dañan: cintas, disquetes, cables.

2. *Errores de hardware o software*: fallas de CPU, discos o cintas ilegibles, errores de telecomunicaciones, errores de programación.

3. *Errores humanos*: captura incorrecta de datos, montaje de una cinta o disco equivocado, ejecución del programa equivocado, extravío de un disco o una cinta, o alguna otra equivocación.

Casi todos estos problemas pueden evitarse manteniendo respaldos adecuados, de preferencia lejos de los datos originales. La perdida accidental de datos puede llegar a causar más daños que los intrusos.

6.2. Encriptamiento sencillo con llave secreta[62]

Criptografía es el arte de crear y usar un criptosistema (método para ocultar mensajes), es el arte y la ciencia de desarrollar y usar mecanismos para transformar los datos en registros de información ilegibles para cualquiera, excepto para el destinatario quien lo puede descifrar.

La criptografía comprende un conjunto de técnicas que proporcionan los siguientes servicios:

[62] ibid., p. 588.

- Cifrado, que transforma los datos a una forma ilegible, para asegurar la privacidad o confidencialidad de los mismos.
- Descifrado, que es el proceso inverso al cifrado. Transforma datos cifrados a su forma original.
- Autentificación, que identifica una entidad, como una persona, una máquina en la red, una organización, un documento un software, etc.
- Firmas digitales, que ligan un documento con el propietario de una clave particular y es el equivalente de las firmas de papel.

La **criptografía con llave secreta** es aquella que utiliza una misma clave para cifrar y para descifrar mensajes, Las dos partes que se comunican han de ponerse de acuerdo de antemano sobre la clave a usar. Una vez que ambas tienen acceso a esta clave, el remitente cifra un mensaje usándola, lo envía al destinatario, y éste lo descifra con la misma.

Este tipo de criptografía utiliza los principios del cifrado convencional moderno representado fundamentalmente por el algoritmo DATA ENCRYPTION STANDARD (DES) este es un esquema de llave privada o secreta que cifra bloques de 64 bits, mediante llaves de 56, esto hace que existan **2^{56} = 7.2 x10 ^{16} llaves**.

Un ejemplo de este tipo de cifrado es en el que cada letra se sustituye por una letra distinta; por ejemplo, todas las A se sustituyen por Q, todas las B se sustituyen por W, todas las C se sustituyen por E, etcétera, como sigue:

Texto simple: ABCDEFGHIJKLMNÑOPQRSTUVWXYZ
Texto cifrado: QWERTYUIOPASDFGHJKLÑZXCVBNM

Figura 6. 1 Encriptamiento sencillo con llave secreta

Este sistema general se denomina sustitución mono alfabética, y la clave es la cadena de 27 letras que corresponde al alfabeto completo. La clave de cifrado en este ejemplo es *QWERTYUIOPASDFGHJKLÑZXCVBNM.* Con esta llave, el texto simple ATAQUE se transformaría en el texto cifrado *QZQKXT.* La clave de descifrado indica cómo volver del texto cifrado al texto simple. En este ejemplo, la clave de descifrado es *KXVMCNÑOHPQRZYSUADLEGWBUFT* porque una A en el texto cifrado es una K en el texto simple, una B en el texto cifrado es una X en el texto simple, y así en forma sucesiva.

Muchos sistemas criptográficos, entre ellos éste, tienen la propiedad de que dada la clave de cifrado, es fácil deducir la clave de descifrado, y viceversa. Aunque los sistemas de sustitución mono alfabética no sirven de mucho, se conocen otros algoritmos de clave simétrica que son relativamente seguros si las claves tienen la longitud suficiente. Para lograr una seguridad razonable es necesario utilizar claves de mayor tamaño como por ejemplo de 1024 bits.

6.3. Encriptamiento con llave pública[63]

En el esquema de llave pública todos los usuarios tienen una llave pública y una privada. Si alguien quiere enviar un mensaje, obtiene una copia de la llave pública con la cual cifra el mensaje que sólo podrá descifrarse con la llave secreta. Los mensajes cifrados con la llave pública no se pueden descifrar con la misma llave pública.

El cifrado de llave pública está basado en funciones matemáticas cuya complejidad hace poco posible que con un tiempo y potencia de cómputo razonable, conociendo solo el menaje cifrado y la llave pública pueda deducirse la llave privada y con ella obtener el mensaje original.

---

[63] Véase, William Stallings, op. cit., p. 675.

Es casi imposible descubrir a partir de estas, la clave de descifrado correspondiente. En estas circunstancias, la clave de cifrado puede hacerse pública, manteniendo en secreto sólo la clave de descifrado privada. El problema principal de la criptografía de clave pública es que es mucho más lenta que la criptografía simétrica[31].

6.4. Estándares de criptografía[64]

I. Cifrado Clásico.- El cifrado clásico, que también se conoce como cifrado simétrico o cifrado de clave única, era el único tipo de cifrado de datos que se utilizaba antes de la introducción del cifrado de clave pública a finales de los setenta.

Un esquema de cifrado de datos clásico consta de cinco elementos:

- **Texto sin cifrar:** es el mensaje original o los datos que se introducen en el algoritmo como datos de entrada.
- **Algoritmo de cifrado:** el algoritmo de cifrado realiza diversas sustituciones y transformaciones sobre el texto sin cifrar.
- **Clave secreta**: la clave secreta también se introduce en el algoritmo de cifrado. Las transformaciones y sustituciones exactas que realiza el algoritmo dependen de esta clave.
- **Texto cifrado:** es el mensaje ya codificado que se da como salida. Depende del texto sin cifrar y de la clave secreta. Para un mismo mensaje, dos claves diferentes producirían dos textos cifrados diferentes.
- **Algoritmo de descifrado:** este es esencialmente el algoritmo de cifrado ejecutado a la inversa. Toma el texto cifrado y la clave secreta y produce el texto sin cifrar original.

Se necesitan dos requisitos para un uso seguro de un cifrado clásico:

---

[64] William Stallings, op. cit., p. 676.

1. Se necesita un algoritmo robusto de cifrado. Se requiere de un algoritmo tal que una persona que lo conociera y que tuviera acceso a uno o más textos cifrados, fuera incapaz de descifrarlos o de averiguar la clave.

2. El emisor y el receptor deben haber recibido copias de la clave secreta de una forma segura y deben mantener la clave secreta de una forma segura. Si alguien fuera capaz de descubrir la clave y llegar a conocer el algoritmo, toda la comunicación que utilice esta clave se convierte en legible.

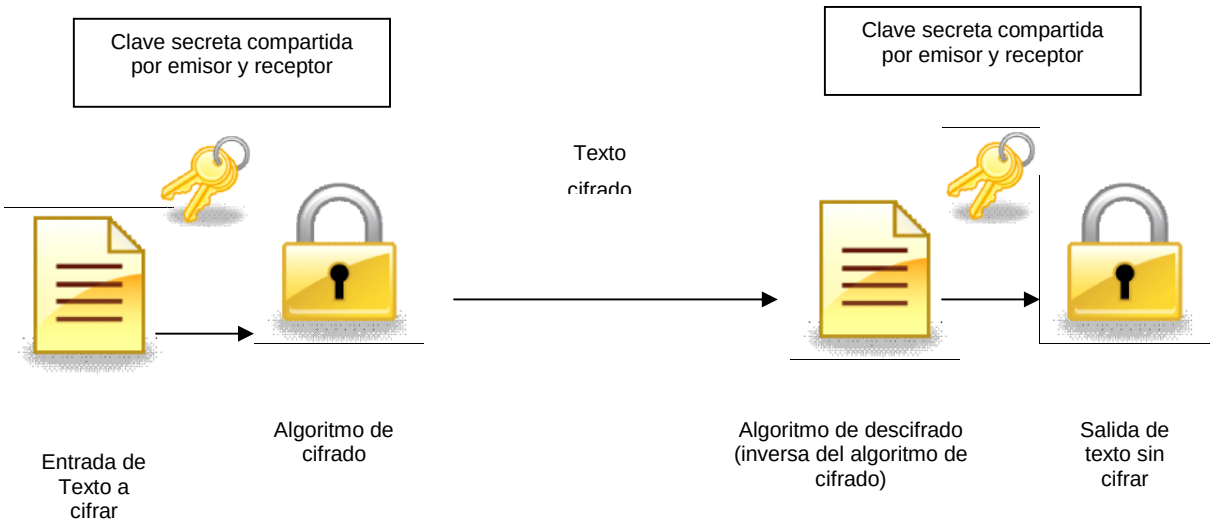


Figura 6.2 Cifrado clásico

II. El Estándar de Cifrado de Datos (DES) [65]**.-** El esquema de cifrado que más se utiliza viene definido por el Estándar de Cifrado de Datos *(DES, Data Encryption Standard)* adoptado en 1977 por la Oficina Nacional de Estándares *(NBS, National Bureau of Standards),* ahora Instituto Nacional de Estándares y Tecnologías *(NIST, National Instituto of Standards and Technology).* Como en cualquier esquema de cifrado, hay dos entradas para la función del DES; el texto sin cifrar que ha de ser cifrado y la clave. Con DES el texto sin cifrar debe ser de 64bits y la

[65] William Stallings, op. cit., p. 678.

clave de 56 bits. Los bloques de texto sin cifrar mayores de 64 bits se cifran en porciones 64 bits.

Esencialmente, DES procesa el texto sin cifrar haciendo pasar cada entrada de 64 bits a través de 16 iteraciones, produciendo un valor intermedio de 64 bits al final de cada iteración. Cada iteración es, en esencia, la misma función compleja que engloba una permutación de bits y la sustitución de un patrón de bits por otro. La entrada de cada fase está formada por la salida de la fase anterior y por una permutación de los bits de la clave, donde la permutación se denomina subclave.

El proceso de descifrado con DES es esencialmente el mismo que el de cifrado: utiliza el texto cifrado como entrada para el algoritmo DES, y utiliza las subclaves generadas para cada iteración en orden inverso (es decir utiliza la subclave decimosexta para la primera iteración, la subclave decimocuarta para la segunda iteración, etc.).

III. Triple DEA[66]**.-** Dada la potencial vulnerabilidad del DES ante un ataque de fuerza bruta, ha habido un considerable interés por encontrar una alternativa. Un método, que preserva la inversión existente en software y equipos, consiste en utilizar cifrado múltiple con DES y múltiples claves. El triple DEA (TDEA) se convirtió por primera vez en un estándar para programas financieros en el estándar ANSI X9.17, en 1985. TDEA fue incorporado como parte del DES en 1999, con la publicación del F1PS PUB 46-3.

TDEA utiliza tres claves y tres ejecuciones del algoritmo DES. La función sigue una secuencia de cifrado-descifrado-cifrado (EDE, *Encrypt-Decrypt-Encrypt).* Con tres claves distintas, TDKA tiene una longitud de clave efectiva de 168 bits. FIPS 46-3 también permite el uso de dos claves haciendo Kl = K3; esto ofrece una

[66] Ibid., p. 679.

longitud de clave de 112 bits. FIPS 46-3 incluye las siguientes directrices para TDEA:

- TDEA es la mejor elección de algoritmo de cifrado clásico aprobado por FIPS.
- El DEA original, que utiliza una clave de 56 bits, está permitido bajo el estándar sólo para sistemas heredados. Las nuevas leyes deberían admitir TDEA.
- A las organizaciones con sistemas heredados DEA se les anima a pasarse a TDEA.
- Es previsible que TDEA y el Estándar de cifrado avanzado *(AKS, Advanced Encryption Standard)* coexistan como algoritmos aprobados por la FIPS, permitiendo así una transición gradual a AES.

Es fácil darse cuenta de que TDEA es un algoritmo formidable. Dado que el algoritmo criptográfico subyacente es DEA, TDEA puede presumir de la misma resistencia al criptoanálisis basado en algoritmos que se presume para DEA. Además con una longitud de clave de 168 bits, los ataques por fuerza bruta son prácticamente inviables.

IV. Estándar de cifrado avanzado[67]**.-** TDEA tiene dos atractivos que aseguran su extensa utilización en los próximos años. Primero, con la longitud de clave de 168 bits, supera la vulnerabilidad ante los ataques de la fuerza bruta del DEA. Segundo, el algoritmo de cifrado subyacente en TDEA es el mismo que en DEA. Este algoritmo ha estado sujeto a un mayor examen que cualquier otro algoritmo de cifrado durante mucho tiempo, y no se ha hallado ningún ataque criptoanalítico basado en algoritmo más eficaz que el de fuerza bruta. En concordancia, hay un alto grado de confianza en que TDEA es muy resistente al criptoanálisis. Si lo único a tener en cuenta fuera la seguridad, entonces TDEA constituiría la elección

[67].Ibid., p. 675.

apropiada para un algoritmo estandarizado de cifrado durante las décadas venideras.

La principal desventaja del TDEA es que el algoritmo es relativamente lento en software. El DEA original fue diseñado para implementaciones hardware de mediados de los setenta y no produce código de software eficiente. TDEA, que realiza tres veces más iteraciones que DEA, en correspondencia más lento. Una desventaja secundaria es que tanto DEA como TDEA utilizan un tamaño de bloque de 64 bits. Por razones tanto de eficacia como de seguridad, sería deseable un tamaño de bloque mayor.

Debido a esta desventaja, TDEA no es un candidato razonable para un uso a largo plazo. Como sustituto, NIST emitió en 1997 una convocatoria para propuestas de un nuevo Estándar de Cifrado Avanzado *(AES, Advanced Encryption Standard)* que debía tener una capacidad de seguridad igual o mejor que TDEA y una mejora significativa en la eficiencia. Además de estos requisitos generales, NIST especificó que AES debe ser un cifrador de bloques simétrico con una longitud de bloques de 128 bits y soportar longitud de claves de 128, 192 y 256 bits. Los criterios de evaluación incluyen seguridad, eficiencia computacional, necesidades de memoria, adaptación al software y hardware y flexibilidad.

V. Algoritmo de Rivest-Shamir-Adleman (RSA)[68]**.-** Uno de los primeros esquemas de clave pública fue el desarrollado en 1977 por Ron Rivest, Adi Shamir y Len Adleman en el MIT. El esquema RSA tiene la supremacía desde entonces como el único enfoque ampliamente aceptado e implementado para el cifrado de clave pública. RSA es un cifrador en el que el texto sin cifrar y el texto cifrado son enteros entre 0 y $n-1$ para algún n. El cifrado implica aritmética modular. El

[68] loc. cit.

punto fuerte del algoritmo se basa en la dificultad que supone factorizar números en sus factores primos.

VI. Firmas digitales.- En muchos casos es necesario firmar un documento de manera digital. Las firmas digitales hacen que sea posible firmar mensajes de correo electrónico y otros documentos digitales de modo tal que no puedan ser negados después por quien los envió. Un método común consiste en aplicar primero al documento un algoritmo de *hash* unidireccional que sea muy difícil invertir. La función de *hash* por lo regular produce un resultado de longitud fija independiente del tamaño del documento original. Las funciones de *hash* más utilizadas son MD5 (sinopsis de mensaje; Message Digest) que produce un resultado de 16 bytes (Rivest, 1992) y SHA (algoritmo de hash seguro; Secare Hash Algorithm), que produce un resultado de 20 bytes (NIST, 1995).

Si el receptor quiere usar este esquema de firma, necesita conocer la clave pública del transmisor. Algunos usuarios publican su clave pública en su sitio Web. Otros no lo hacen por temor a que un intruso ingrese al sitio y altere su clave. En este caso se requiere un mecanismo alterno para distribuir claves públicas. Un método común es que quienes transmiten mensajes anexen un certificado al mensaje, con el nombre y la clave pública del usuario, y firmado en forma digital por un tercero confiable. Cuando el usuario tenga la clave pública del tercero confiable, podrá aceptar certificados de todos los transmisores que utilicen ese tercero confiable para generar sus certificados.

Es importante mencionar que también existen esquemas en los que no se utiliza criptografía de clave pública.[69]

6.5 Capacidades, derechos y matriz de acceso[70]

Los derechos de acceso definen qué acceso tienen los sujetos sobre los objetos. Los objetos son entidades que contienen información, pueden ser físicos o abstractos. Los sujetos acceden a los objetos, y pueden ser usuarios, procesos, programas u otras entidades.

[69] Andrew S. Tanenbaum, op. cit. pp. 590-91.
[70] Véase, William Stallings, op. cit., pp. 638-641.

> Los derechos de accesos más comunes son: acceso de lectura, acceso de escritura y acceso de ejecución. Estos derechos pueden implementarse usando una matriz de control de acceso.[71]

Las medidas tomadas para controlar el acceso en los sistemas de proceso de datos pueden encuadrarse en dos categorías: las asociadas con el usuario y las asociadas con los datos.

Control de acceso orientado al usuario

> La técnica más habitual de control de acceso al usuario en un sistema de tiempo compartido o en un servidor es la conexión del usuario, que requiere un identificador de usuario (ID) y una contraseña. El sistema permitirá a un usuario conectarse sólo si el ID es conocido por el sistema y si el usuario sabe la contraseña asociada por el sistema a dicho ID. Este esquema ID/contraseña es un método notablemente poco fiable de control de acceso al usuario.[72]

Los usuarios pueden olvidar sus contraseñas y pueden revelarlas accidental o deliberadamente. Los piratas informáticos *(hackers)* son muy habilidosos en adivinar los ID tic usuarios especiales, como el personal de control o de administración del sistema. Por último, el esquema ID/contraseña está sujeto a los intentos de penetración.

El control de acceso descentralizado considera la red como un enlace transparente de comunicaciones y el procedimiento usual de conexión lo lleva a cabo el servidor de destino. Desde luego, debe seguir considerándose la seguridad concerniente a la transmisión de contraseñas por la red.

---

[71] Salvador Meza Badillo, *Sistemas Operativos Multiusuarios*, SUA-FCA-UNAM, pp. 43-4, material disponible en línea: http://fcasua.contad.unam.mx/apuntes/interiores/docs/98/2/sis_operativos.pdf, recuperado el 08/01/09.

[72] Departamento de Áreas Informáticas: *Sistemas operativos: control de acceso orientado a* usuarios, material electrónico disponible en: http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MonogSO/SEG02.html#IIB, recuperado el 08/01/09.

En muchas redes, pueden emplearse dos niveles de control de acceso. Los servidores individuales pueden estar provistos de un servicio de conexión que proteja las aplicaciones y los recursos específicos del servidor. Además, la red en su conjunto puede ofrecer una protección para restringir el acceso a los usuarios no autorizados.

Control de acceso orientado a los datos

> Después de una conexión con éxito, al usuario se le habrá concedido el acceso a uno o más servidores y aplicaciones. Esto no suele ser suficiente en un sistema que incluya datos sensibles en su base de datos. Mediante el procedimiento de control de acceso al usuario, el sistema puede identificar a un usuario. Asociado con cada usuario, puede haber un perfil de usuario que especifique las operaciones y los accesos a archivos permisibles.[73]

El sistema operativo puede hacer valer reglas en función del perfil del usuario. El sistema gestor de la base de datos, sin embargo, debe controlar el acceso a registros específicos o incluso partes de un registro. Por ejemplo, puede permitirse que cualquier administrador obtenga una lista del personal de una organización, pero solamente unos individuos elegidos pueden tener acceso a la información de sueldos y salarios. La cuestión es más importante de lo que parece. Mientras que el sistema operativo puede otorgar a un usuario permiso para acceder a un archivo o utilizar una aplicación, tras lo cual no se producen más controles de seguridad, el sistema gestor de la base de datos debe tomar decisiones sobre cada intento de acceso individual. Dicha decisión dependerá no sólo de la identidad del usuario, sino también de las partes específicas de datos a las que se accede e, incluso, de la información ya divulgada al usuario.

> Un modelo general de control de acceso ejercido por un sistema gestor de archivos o bases de datos es el de una matriz de acceso. Los elementos básicos del modelo son los siguientes:

---

[73] Ver cita anterior.

- **Matriz de acceso.** El modelo de protección del sistema se puede ver en forma abstracta como una matriz, la *matriz de acceso (Figura 6-3).*
- **Sujeto:** una entidad capaz de acceder a los objetos. En general, el concepto de sujeto es equiparable con el de proceso. Cualquier usuario o aplicación consigue acceder en realidad a un objeto por medio de un proceso que representa al usuario o a la aplicación.
- **Objeto:** cualquier cosa cuyo acceso debe controlarse. Como ejemplos se incluyen los archivos, partes de archivos, programas y segmentos de memoria.
- **Derecho de acceso:** la manera en que un sujeto accede a un objeto. Como ejemplos está Leer, Escribir y Ejecutar.

Una dimensión de la matriz consta de los sujetos identificados que pueden intentar acceder a los datos. Normalmente, esta lista está formada por usuarios individuales o grupos de usuarios, aunque se puede controlar el acceso para terminales, servidores o aplicaciones, en lugar de o además de usuarios. La otra dimensión enumera los objetos a los que se puede acceder. Más concretamente, los objetos pueden ser campos de datos individuales. También pueden ser objetos de la matriz agrupaciones más globales, como registros, archivos o incluso la base de datos entera. Cada entrada de la matriz indica los derechos de acceso de ese sujeto a ese objeto.

En la práctica, las matrices de acceso suelen estar dispersas y se implementan por descomposiciones en una de las dos dimensiones. La matriz se puede descomponer en columna, para obtener listas de control de acceso (Figura 6-4). Así pues, para cada objeto, una lista de control de acceso enumera los usuarios y sus derechos de acceso permitidos.[74]

La lista de control de acceso puede contener una entrada por defecto o pública. Se permite que los usuarios a los que no se les haya concedido explícitamente unos derechos especiales dispongan de un conjunto de derechos por omisión. Los elementos de la lista por omisión pueden incluir a usuarios individuales, así como grupos de usuarios.

[74] Departamento de Informática UNNE: *Sistema operativo: Seguridad*. “Control de acceso orientado a datos”. Material en línea, disponible en: http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MonogSO/SEG02.html#IIC, recuperado el 08/01/09.

	Archivo 1	Archivo 2	Archivo 3	Archivo 4	Cuenta 1	Cuenta 2
Usuario A	Propietario		Propietario		Solicitar	
	R		R		crédito	
	W		W			
Usuario B		Propietario			Solicitar	Solicitar
	R	R	W	R	débito	crédito
		W				
Usuario C	R			Propietario		Solicitar
	W	R		R		débito
				W		

Figura 6.3. Matriz de acceso

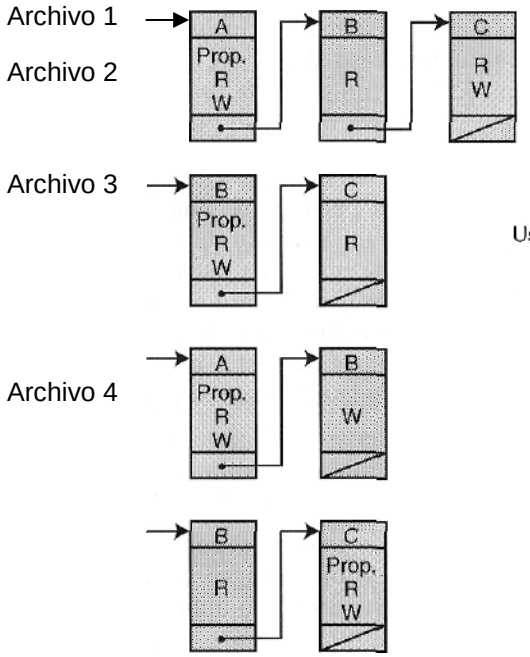


Figura 6.4 Lista de capacidades de acceso para archivos de parte (a)

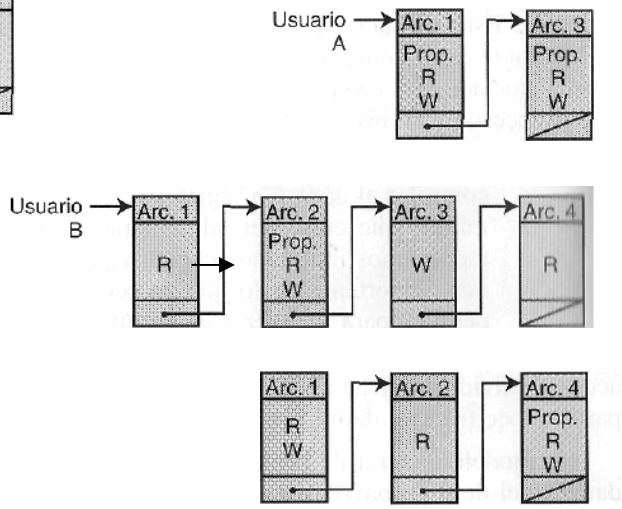


Figura 6.5 Lista de capacidades de acceso para usuarios de parte (a)

"Con la descomposición por filas se obtienen etiquetas de capacidad de acceso (Figura 6.5). Una etiqueta de Capacidades especifica los objetos y las operaciones autorizadas para un usuario. Cada usuario tiene un número de etiquetas y puede estar autorizado para prestarlas o concederlas a los otros." Como las etiquetas pueden estar dispersas por el sistema, presentan un problema de seguridad mayor que el de las listas de control de acceso. "En concreto, las etiquetas no pueden ser falsificadas. Una manera de conseguirlo es que el sistema operativo

guarde todas las etiquetas en vez de los usuarios. Las etiquetas deben guardarse en una zona de la memoria inaccesible para los usuarios."[75]

6.6 Virus y sus variantes[76]

Un virus es un programa que puede «infectar» a otros programas, alterándolos; la alteración incluye una copia del programa de virus, que puede entonces seguir infectando a otros programas.

Un virus informático porta en su código las instrucciones para hacer copias perfectas de sí mismo. Una vez alojado en una computadora el virus toma el control temporalmente del sistema operativo situado en el disco de la computadora. Entonces, cuando la computadora infectada entre en contacto con un elemento del software no infectado, se pasa una nueva copia del virus al programa. Así pues, la infección puede extenderse de una computadora a otra, a través de usuarios que intercambian sus discos o bien se envían programas a través de la red. En un entorno de red, la capacidad de acceder a las aplicaciones y los servicios del sistema de otras computadoras que no estén protegidas hacen posible que la propagación de virus sea muy severa y afecte considerablemente a los equipos causando graves daños a la información, programas de aplicación y al sistema operativo.

La naturaleza de los virus

Un virus puede hacer cualquier cosa que hagan otros programas. La única diferencia es que se acopla a otro programa y se ejecuta de forma oculta cada vez que se ejecuta el programa anfitrión. Una vez que un virus se está ejecutando, puede realizar cualquier función, como borrar archivos y programas.

Durante su vida, un virus típico pasa por las siguientes cuatro etapas:

1. Fase latente: el virus está inactivo. El virus se activará finalmente por algún suceso, como una fecha, la presencia de otro programa o archivo o que la capacidad del disco exceda de cierto límite. No todos los virus pasan por esta etapa.

---

[75] Departamento de Informática, UNNE, *Sistema operativo: seguridad*, material en línea, disponible en: http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MonogSO/SEG02.html. Recuperado el 08 de enero de 2009.

[76] Cf., William Stallings, op. cit., pp. 655 - 658.

2. Fase de propagación: el virus hace copias idénticas a él en otros programas o en ciertas áreas del sistema del disco. Cada programa infectado contendrá ahora un clon del virus, que entrará a su vez en la fase de propagación.

3. Fase de activación: el virus se activa para llevar a cabo la función para la que está pensado. Como en la fase latente, la fase de activación puede producirse por múltiples sucesos del sistema, incluyendo una cuenta del número de veces que esta copia del virus ha hecho copias de sí mismo.

4. Fase de ejecución: se lleva a cabo la función. La función puede ser no dañina, como dar un mensaje por la pantalla, o dañina, como la destrucción de archivos de programas y datos.[77]

La mayoría de los virus llevan a cabo su trabajo de manera específica para un sistema operativo concreto y, en algunos casos, específicamente para una plataforma de hardware en particular. Así pues, están diseñados para atacar y tomar ventaja de las debilidades de sistemas concretos.

Tipos de virus

Desde que los virus aparecieron por vez primera, se ha producido una carrera de armamento entre los escritores de virus y los escritores de software antivirus. A medida que se han desarrollado contramedidas eficaces para los tipos de virus existentes, se han desarrollado nuevos tipos de virus.

Tipos de virus más significativos:

- Virus parásitos: un virus parásito se anexa a los archivos ejecutables y se reproduce, al ejecutar el programa infectado, buscando otros archivos ejecutables que infectar.

[77] Departamento de Informática, Universidad Nacional del Nordeste, Argentina, *Sistemas operativos, Seguridad,* material en línea, disponible en: http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MonogSO/SEG02.html Fecha de recuperación: 09 de enero de 2009.

- Virus residentes en la memoria: se alojan en la memoria principal como parte de un programa del sistema residente. De este modo infecta a todos los programas que se ejecutan.

- Virus del sector de arranque: infecta el sector principal de arranque *(master boot record)* o sector de arranque y se propaga cuando el sistema arranca desde el disco que contiene el virus.

- Virus clandestino: una forma de virus diseñado explícitamente para esconderse de la detección mediante un software antivirus.

- Virus polimorfo: un virus que muta con cada infección, haciendo imposible la detección por la «firma» del virus.

Un ejemplo de **virus clandestino** es el que utiliza compresión para que el programa infectado tenga exactamente la misma longitud que una versión no infectada. Por ejemplo, un virus puede poner alguna lógica de interceptación en las rutinas de E/S a disco, de modo que cuando haya un intento de leer partes sospechosas del disco mediante estas rutinas, el virus presente el programa original no infectado Así pues, el término clandestino no se aplica a los virus como tales sino que, más bien, es una técnica empleada por los virus para evitar su detección.

Un **virus polimorfo** “crea copias durante la reproducción que son funcionalmente equivalentes pero que tienen diferentes patrones de bits”. Como con los virus clandestinos, la finalidad es vencer a los programas que buscan virus. En tal caso, la «firma» del virus varía con cada copia. Para lograr esta variación, el virus puede insertar aleatoriamente instrucciones superfinas o intercambiar el orden de las instrucciones independientes. Un método más eficaz es usar técnicas de cifrado.

> Una parte del virus, generalmente llamada *motor de mutación,* crea una clave de cifrado aleatoria para cifrar el resto del virus. Dicha clave se almacena junto con el virus y se modifica el propio motor de mutación. Cuando se invoca a un programa infectado, el virus utiliza la clave aleatoria almacenada para descifrar el virus. Cuando el virus se reproduce, se escoge una clave aleatoria diferente.

> Otra arma del armamento de los escritores de virus es un juego de utilidades para la creación de virus. [78]

Dicho juego permite que un novato cree rápidamente una serie de virus diferentes. Aunque los virus creados con estas utilidades tienden a ser menos sofisticados que los virus diseñados desde cero, el número absoluto de nuevos virus que pueden generarse crea un problema para los procedimientos antivirus.

6.7. Contraseñas de una sola vez[79]

La forma más extrema de la política de cambiar las contraseñas en forma continua es la **contraseña para usarse sólo una vez**. Cuando se usa este tipo de contraseñas, se entrega al usuario una lista de contraseñas. En cada inicio de sesión se usa la siguiente contraseña de la lista. Si un intruso llega a descubrir una contraseña, no le servirá de nada, pues en la siguiente ocasión deberá usarse una contraseña distinta. Aquí el problema pude ser que el usuario pierda o le roben la lista de contraseñas.

Existe el método “Lamport” ideado por Lesile Lamport que permite al usuario iniciar una sesión de forma segura en una red insegura, empleando contraseñas para usarse sólo una vez. Este método puede servir para que un usuario inicie sesión en un servidor a través de Internet desde su computadora personal, aunque los intrusos puedan ver y copiar todo el tráfico en ambas direcciones. Además, no es necesario almacenar secretos en el sistema de archivos del servidor ni en el de la computadora del usuario.

---

[78] Departamento de Informática, Universidad Nacional del Nordeste, Argentina, *Sistemas operativos, Seguridad,* material en línea, disponible en: http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MonogSO/SEG02.html Fecha de recuperación: 09 de enero de 2009.

[79] A. S. Tanenbaum, op. cit., pp. 599-600.

El algoritmo se basa en una función unidireccional, es decir, una función $y = f(x)$ con la propiedad de que, dado x, es fácil calcular y, pero dado y no es factible determinar x desde el punto de vista computacional. La entrada y la salida deben tener la misma longitud, por ejemplo, 128 bits.

6.8 Amenazas, ataques y vigilancia[80]

Tipos de amenazas

Los tipos de amenazas a la seguridad de un sistema de computadoras o una red pueden ser de varios tipos, de manera general se produce un flujo de información desde un origen hacia un destino, ya sea enviando un archivo o conectándose a otro equipo como un usuario utilizando los recursos de la memoria principal, discos, etc. Este flujo está representado en la Figura 6.6. El resto de la figura muestra cuatro categorías generales de amenazas:

- **Interrupción**: se destruye un elemento del sistema o se hace inaccesible o inútil. Este es un ataque a la disponibilidad. Como ejemplos se incluyen la destrucción de una pieza del hardware, como un disco duro, el corte de una línea de comunicaciones o la inutilización del sistema de administrador de archivos.

- **Interceptación**: una parte no autorizada consigue acceder a un elemento. Este es un ataque al secreto. La parte no autorizada puede ser una persona, un programa o una computadora. Como ejemplos se incluyen la intervención de las conexiones telefónicas para conseguir datos de una red y la copia ilícita de archivos o programas.

- **Modificación**: una parte no autorizada no sólo consigue acceder, sino que falsifica un elemento. Este es un ataque a la integridad. Como ejemplos se incluyen el cambio de valores en un archivo de datos, la alteración de un programa para que se comporte de manera diferente y la modificación del contenido de los mensajes transmitidos en una red.

[80] loc. cit.

- **Invención:** una parte no autorizada inserta objetos falsos en el sistema. Este es también un ataque a la autenticidad**.** Como ejemplos se incluyen la inserción de mensajes falsos en una red o la adición de registros a un archivo.[81]

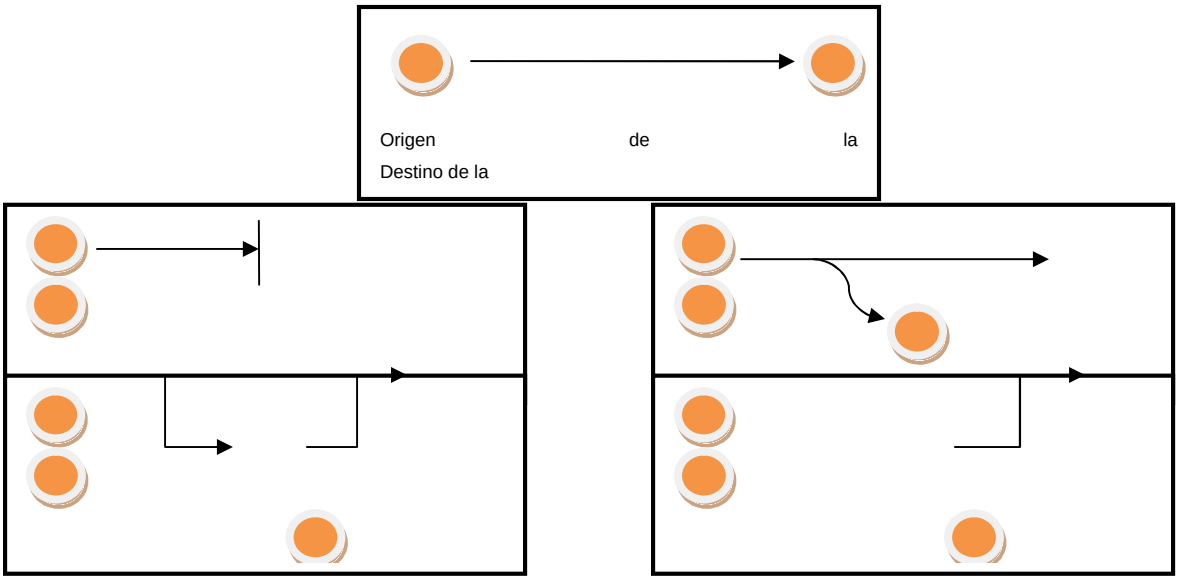


Figura 6. 6. Amenazas a la seguridad

Ataques

Un "**ataque**" consiste en aprovechar una vulnerabilidad de un sistema informático (sistema operativo, programa de software o sistema del usuario) con propósitos desconocidos por el operador del sistema y que, por lo general, causan un daño.

Los ataques siempre se producen en Internet, a razón de varios ataques por minuto en cada equipo conectado. En su mayoría, se lanzan automáticamente desde equipos infectados (a través de virus, troyanos, gusanos, etc.) sin que el propietario se dé cuenta de lo que está ocurriendo.

Para bloquear estos ataques es importante estar familiarizado con los principales tipos de ataques y tomar medidas preventivas.

---

[81] Departamento de Informática, UNNE, *Sistemas operativos: Seguridad; tipos de amenazas.*, Material en línea, disponible en: http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MonogSO/SEG02.html#IA, recuperado el 08/01/09.

Los ataques pueden ejecutarse por diversos motivos:

- Para obtener acceso al sistema.
- Para robar información, como secretos industriales o propiedad intelectual.
- Para recopilar información personal acerca de un usuario.
- Para obtener información de cuentas bancarias.
- Para obtener información acerca de una organización
- Para afectar el funcionamiento normal de un servicio.
- Para utilizar las computadoras como un medio de ataque a otros equipos y redes.
- Para usar los recursos del sistema del usuario (recursos de hardware y software, ancho de banda)

Tipos de Ataques

Los sistemas informáticos utilizan una gran cantidad de recursos, tales como la energía eléctrica, redes, software, hardware, sistemas operativos, etc. Y en cada uno de estos pueden existir riesgos que limiten o detengan su funcionamiento.

Los riesgos se pueden clasificar de la siguiente manera:

1. Acceso físico: en este caso, el atacante tiene acceso a las instalaciones e incluso a los equipos:

- Interrupción del suministro eléctrico.
- Apagado manual del equipo.
- Vandalismo.
- Acceso a los componentes del equipo (discos, memoria, etc.).
- Monitoreo del tráfico de red.

2. Intercepción de comunicaciones:

- Secuestro de sesión.
- Falsificación de identidad.
- Re-direccionamiento o alteración de mensajes.

3. Denegación del servicio: el objetivo de estos ataques reside en interrumpir el funcionamiento normal de un servicio. La denegación de servicio se divide de la siguiente manera:

- Explotación de las debilidades del protocolo TCP/IP.
- Explotación de las vulnerabilidades del software del servidor.

4. Intrusiones:

- Análisis de puertos.
- Elevación de privilegios: este tipo de ataque consiste en aprovechar una vulnerabilidad en una aplicación al enviar una solicitud específica (no planeada por su diseñador). En ciertos casos, esto genera que se pueda acceder al sistema con derechos de aplicación. Los ataques de desbordamiento de la memoria intermedia (búfer) usan este principio.
- Ataques malintencionados (virus, gusanos, troyanos).

5. Ingeniería social: en la mayoría de los casos, el eslabón más débil es el mismo usuario. Muchas veces es él quien, por ignorancia o a causa de un engaño, genera una vulnerabilidad en el sistema al brindar información, por ejemplo la contraseña. Es necesario que se aplique la educación a los usuarios y las buenas prácticas para evitar este tipo de ataques.

6. Puertas trampa: son puertas traseras ocultas en un programa de software que brindan acceso a su diseñador en todo momento.

Es muy importante que los errores de programación sean corregidos con rapidez cuando se conozca una vulnerabilidad, también es que los administradores y usuarios se mantengan informados sobre las actualizaciones de los programas y aplicaciones que utilizan con el fin de limitar los riesgos de ataques. Además, existen diversos dispositivos (firewalls, sistemas de detección de intrusiones, antivirus) que brindan la posibilidad de aumentar el nivel de seguridad.[82]

Vigilancia[83]

La vigilancia tiene que ver con:

- La verificación y la auditoría del sistema.
- La autentificación de los usuarios.
- Los sistemas sofisticados de autentificación de usuarios resultan muy difíciles de evitar por parte de los intrusos.

---

[82] Kioskea: *Introducción a los ataques*, actualizado el 16/10/08, material en línea, disponible en: http://es.kioskea.net/contents/attaques/attaques.php3 Fecha de recuperación: 09 de enero de 2009.

[83] H. M. Deitel. *Introducción a los Sistemas Operativos*. México, Addison-Wesley Iberoamericana, 1987.

Un problema existente es la posibilidad de que el sistema rechace a usuarios legítimos:

- Un sistema de reconocimiento de voz podría rechazar a un usuario legítimo con alguna enfermedad en la garganta.
- Un sistema de huellas digitales podría rechazar a un usuario legítimo que tenga una cortadura o una quemadura.

Verificación de Amenazas

Es una técnica según la cual los usuarios *no pueden tener acceso directo a un recurso*:

- Solo lo tienen las rutinas de sistema operativo llamadas *programas de vigilancia.*
- El usuario solicita el acceso al sistema operativo.
- El sistema operativo niega o permite el acceso.
- El acceso lo hace un programa de vigilancia que luego pasa los resultados al programa del usuario.
- Permite:
  - Detectar los intentos de penetración en el momento en que se producen.
  - Advertir en consecuencia.

Amplificación

La *amplificación* se produce cuando:

- Un *programa de vigilancia* necesita para cumplir su cometido mayores derechos de acceso de los que disponen los usuarios:
  - Ej.: se requiere calcular un promedio para lo cual es necesario leer un conjunto de registros a los que el usuario no tiene acceso individualmente.[84]

6.9. Reconstrucción de un sistema violado

Se define como violación de acceso (*access violation* o *segmentation fault* en inglés) al intento fallido de acceso a información o a programas a los

[84] David Luis la Red Martínez. *Sistemas operativos: Vigilancia, Verificación de Amenazas y Amplificación*, material en línea, disponible en: http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/SO14.htm. Fecha de recuperación: 09 de enero de 2009.

que no se tiene autorización para ver o modificar. Este mensaje puede ser causado por la configuración de software o por los programadores.

Con los sistemas operativos actuales, cada proceso tiene uno o más segmentos de la memoria del sistema donde puede almacenar y recuperar la información. Cada proceso puede solicitar más o menos memoria (según lo requerido), y la petición será reconocida por el sistema operativo y comparada con la sección de memoria concedida para el proceso. Generalmente, el proceso que solicitó la memoria es el único que puede leerla o modificarla.

Una violación de acceso ocurre cuando un proceso trata de acceder a una parte de la memoria asignada a otra aplicación, o a un área no usada de la memoria, no teniendo los permisos para hacerlo. Normalmente se produce como resultado de un error de programación, por ejemplo, un apuntador perdido.[85]

6.10 La bitácora o diario de operaciones

Uso de bitácora de transacciones

Es importante que cada vez que un usuario ingrese o modifique datos se grabe un campo de auditoría en un registro identificándolo, debe de asegurarse de que los registros no puedan ser modificados ni borrados.

Las pistas de auditoría están diseñadas para permitir el rastreo de cualquier registro de entrada o proceso llevado a cabo en un sistema. Los detalles de cada transacción se registran en un archivo de transacciones. El estudio de transacciones puede proporcionar información de cómo se modificó el archivo. El almacenamiento de estos detalles es automático e invisible para el usuario; también se debe almacenar la información relativa al usuario, de forma que sea claro saber quién llevó a cabo la transacción. Si el sistema tiene un reloj interno, también se marca cada transacción con la hora exacta para saber cuándo ocurrió. Si surge la necesidad de revisar un registro particular en un archivo, es relativamente fácil determinar quién hizo la transacción, cuándo ocurrió, cuáles datos contenía la transacción y cómo se modificaron la base de datos o el registro del archivo maestro.

---

[85] Wikipedia:"Violación de acceso", actualizado el 03/01/09, disponible en: http://es.wikipedia.org/wiki/Access_violation Fecha de recuperación: 09 de enero de 2009.

En caso de las aplicaciones integradas que corren en redes, es casi imprescindible el establecimiento de autorizaciones individuales para cada consulta de datos y para cada tipo de modificación, pues de otro modo, se haría difícil evitar que cualquier persona consulte o modifique datos sin autorización de la gerencia. Para esto:

a) Debería existir una tabla que indique, para cada potencial usuario: su contraseña; a que aplicaciones puede acceder; qué actividades de consulta o modificación de datos está autorizado a realizar.

b) El sistema de control interno debería prever: quién tendrá la responsabilidad de manejar dicha tabla; quién puede autorizar modificaciones a ella; qué constancias escritas deben dejarse de cada modificación.[86]

Bibliografía del tema 6

Stallings, William. Sistemas Operativos. 4ª ed., Madrid, Pearson Educación, 2001.

Tanenbaum, Andrew S. Sistemas Operativos Modernos. 2ª ed., México, Pearson Educación, 2003.

Actividades de aprendizaje

A.6.1. Realiza un diagrama sobre el funcionamiento general del cifrado clásico y explica cada uno de sus elementos.

A.6.2. Realiza un cuadro sinóptico sobre los estándares de criptografía tratados en este tema (objetivo, tipo de criptografía, ventajas y desventajas).

A.6.3. Realiza un cuadro comparativo sobre la función de los tipos de virus tratados en este tema.

---

[86] Leonardo Sena Mayans *Seguridad informática.* Julio 2000. Apud. Elianna Cassinelli, et al. *Criterios de evaluación para la selección de paquetes confiables aplicables a empresas nacionales*, material en línea, disponible en: http://www.ccee.edu.uy/ensenian/catcomp/material/Inform_%20II/Criterios%20evaluacion.pdf, pp. 40-41. Fecha de recuperación: 09 de enero de 2009.

A.6.4. Describe en qué consisten los ataques de “puertas trampa” y como protegerse de esto.

Cuestionario de autoevaluación

1. ¿Qué importancia tiene la seguridad en un sistema operativo?
2. ¿La seguridad ocupa recursos del sistema operativo?
3. Menciona tres errores de hardware que ocasionen la pérdida accidental de los datos.
4. ¿Cuál es el tipo de amenaza en el que un elemento del sistema se hace inaccesible?
5. Menciona tres ejemplos de un tipo de criptografía.
6. ¿Qué es un ataque de ingeniería social?
7. Describe tres características de las amenazas de seguridad.
8. ¿Qué es una firma digital?
9. Describe tres características de la criptografía con llave secreta.
10. ¿Cuáles son las desventajas de utilizar contraseñas de una sola vez?

Examen de autoevaluación

1. ¿Cuál es una amenaza para un sistema operativo?

a) Versión del sistema.
b) Autenticación del usuario
c) Revelación de los datos

2. Es causa de la pérdida accidental de datos:

a) Unidad central de proceso (CPU)
b) Errores de hardware o software
c) encriptación md5

3. ¿Cuál es la función de la seguridad en un sistema operativo que proporciona protección entre los usuarios?

a) Proteger el correo electrónico

b) Proteger el servidor de datos

c) Proteger los derechos de acceso

4. ¿Cuál es la función de una firma digital?

a) Aplica al documento el algoritmo md5

b) Aplica al documento un algoritmo de hash unidireccional

c) Aplica al documento criptografía de llave privada

5. ¿A qué se refiere una interceptación en la seguridad?

a) Una parte autorizada accede al documento

b) Una parte no autorizada consigue acceder al documento

c) Una parte reside en el receptor del documento

6. ¿Cuál es la naturaleza de los virus?

a) Fases: nacen, se reproducen, ataque, mueren

b) Fases: latente, propagación, activación, ejecución

c) Fases: tiempo definido, ejecución, ataque, terminación

7. ¿El cifrado clásico también se conoce como?

a) cifrado integral

b) simétrico o clave única

c) funcione hash

8) ¿Cuál es la característica del estándar de cifrado de datos (DES)?

a) El texto debe contener caracteres alfanuméricos

b) El texto sin cifrar debe de ser de 64 bits.

c) El texto sin cifrar debe de ser de 256 bits

9) ¿Cuál es la función de un antivirus?

a) Proteger el núcleo del sistema operativo

b) Proteger el sistema operativo de programas no deseados

c) Proteger los datos del usuario

10) ¿Cuál es la amenaza a la disponibilidad de los datos?

a) alteración de los datos

b) negación del servicio

c) revelación de los datos

TEMA 7. IMPLANTACIÓN DE SISTEMAS OPERATIVOS

Objetivo particular

Al finalizar el aprendizaje de este tema, el alumno identificará las diferentes consideraciones que se requieren para realizar la implantación y administración de un sistema operativo multiusuario.

Temario detallado

7.1 El superusuario o administrador del sistema
7.2 Selección del SO (Linux vs Windows NT)
7.3 Preparación de discos de arranque
7.4 Planeación de la utilización de los discos
7.5 Creación del Sistema de Archivos
7.6 Administración del espacio libre
7.7 Instalación de Shells, herramientas y compiladores
7.8 Creación de usuarios y grupos

Introducción

El desarrollo de la industria de la computación ha generado que exista una gran variedad de sistemas operativos, cada uno de estos ha sido diseñado para cumplir diversos objetivos. Existen sistemas basados en software libre y otros bajo licencias específicas en las que hay que pagar por hacer uso de ellas, todos los sistemas deben cumplir con los principios de diseño como lo son: la compatibilidad, escalabilidad, seguridad, eficiencia, etc. En este tema conocerás la importancia que tiene la cuenta de administrador, el sistema de archivos, el manejo de los discos, etc. Para poder realizar de manera correcta la implantación de un sistema operativo multiusuario.

7.1 El Súper usuario o administrador del sistema

La cuenta del Superusuario

La cuenta superusuario o administrador del sistema, normalmente llamada root (en UNIX-Linux), viene pre configurada para facilitar la administración del sistema, y no debería ser utilizada para tareas cotidianas como enviar o recibir correo, exploración general del sistema, o programación. Esto es así porque el superusuario, a diferencia de las cuentas de usuario, puede operar sin límites, y un mal uso de esta cuenta puede causar un daño considerable al sistema, se pueden asignar privilegios a las cuentas de usuario cuando así se requiera.

Se debe comprobar siempre dos o tres veces los comandos que se ejecutan como superusuario, ya que un espacio de más o un carácter omitido pueden significar una pérdida de datos irreparable.[87]

Ejemplos de operaciones restringidas al superusuario:

- Montar y desmontar sistemas de archivos.
- Cambiar el directorio raíz de un proceso.
- Crear archivos de dispositivos.
- Fijar la hora del sistema
- Cambiar propiedad de archivos.
- Fijar límites al uso de recursos; fijar prioridades de procesos.
- Fijar el nombre propio de la máquina donde reside el sistema.
- Configurar interfaces de red.
- Cerrar el sistema.[88]

Un Administrador de sistema es aquella persona que se dedica a mantener y operar un sistema de cómputo o una red. Los administradores de sistemas pueden ser miembros de un departamento de tecnologías de información.

Las responsabilidades de un administrador de sistemas son muy amplias, y varían enormemente de una organización a otra. Por lo general se les encomienda la instalación, soporte y mantenimiento de los servidores u

---

[87] Manual de Free BSD: “La cuenta superusuario”, material en línea, disponible en: http://www.freebsd.org/doc/es_ES.ISO8859-1/books/handbook/users-superuser.html. Fecha de recuperación 09 de enero de 2009.

[88] Víctor A. González Barbone, *Administración UNIX: Superusuario y usuarios especiales*, Instituto de Ingeniería Eléctrica, Fac. de Ingienería, Montevideo material en línea, disponible en: http://iie.fing.edu.uy/ense/asign/admunix/superusu.htm Fecha de recuperación: 09 de enero de 2009.

otros sistemas de cómputo, la planeación de respuesta a contingencias y otros problemas. Algunas otras responsabilidades pudieran incluir la programación de scripts o programación (en distintos niveles), manejo de proyectos relacionados con el sistema, supervisión o entrenamiento de operadores de cómputo y ser el consultor para los problemas que se encuentran más allá del conocimiento técnico del personal de soporte. Un administrador de sistemas debe mostrar una mezcla de habilidades técnicas y responsabilidad.[89]

7.2 Selección del SO (Linux vs. Windows NT)

Ventajas de Linux

Hay tres puntos a resaltar de Linux de manera general:

- Linux es muy robusto, estable y rápido: Ideal para servidores y aplicaciones distribuidas.
- Puede funcionar en máquinas de capacidades reducidas: Linux puede correr servicios en un procesador x86 a 200 MHz con buena calidad.
- Linux es libre: Esto implica no sólo la libertad del uso del software, sino también que Linux es de código abierto (modificable) y que Linux tiene una gran cantidad de aplicaciones libres en Internet.
- Linux ya no está restringido a personas con grandes conocimientos de informática: Los desarrolladores de Linux han hecho un gran esfuerzo por dotar al sistema de asistentes de configuración y ayuda, además de un sistema gráfico muy potente. Distribuciones Linux como Red Hat/Fedora, Ubuntu, centOS, etc., tienen aplicaciones de configuración similares a las de Windows.

Desventajas de Linux

- Linux no cuenta con una empresa que lo respalde, por lo que no existe un verdadero soporte como el de otros sistemas operativos.

---

[89] Wikipedia: Administración de sistemas", actualizado el 07/01/09, disponible en: http://es.wikipedia.org/wiki/Administrador_de_sistemas Fecha de recuperación: 09 de enero de 2009.

- La curva de aprendizaje es lenta.
- Requiere consulta, lectura e investigación en lista, foros o en bibliografía dedicada al tema.
- Instalar controladores de Hardware y programas resulta ser más complicado que en Windows. Esto se debe a que las empresas creadoras de controladores crean sus productos con base en Windows, el sistema operativo más usado a nivel mundial.

Ventajas de Windows NT

- La instalación es muy sencilla y no requiere de mucha experiencia.
- Permite realizar diferentes tipos de auditorías, tales como del acceso a archivos, conexión y desconexión, encendido y apagado del sistema, errores del sistema, información de archivos y directorios, etc.
- Muestra estadísticas de errores del sistema, caché, información del disco duro,
- Información de Manejadores, Nº de archivos abiertos, porcentaje de uso del CPU,
- Información general del servidor y de las estaciones de trabajo, etc.

Desventajas de Windows NT

- Tiene ciertas limitaciones por RAM, como: Nº Máximo de archivos abiertos.
- Requiere como mínimo 16MB en RAM y un procesador Pentium de 133 MHz o superior.
- El usuario no puede limitar la cantidad de espacio en el disco duro.
- No soporta archivos de NFS.
- No ofrece el bloqueo de intrusos.
- No soporta la ejecución de algunas aplicaciones para DOS.

7.3 Preparación de discos de arranque

Hoy en día la mayoría de los sistemas operativos cuentan para su instalación con uno o más CD-ROMs y/o DVD autoejecutables, solamente hay que configurar el equipo para que arranque desde la unidad donde leerá el disco.

Windows es un sistema cerrado o de paga, su licencia tiene un costo dependiendo de la versión que se vaya a utilizar y en el ambiente de servidores las licencias tiene un costo adicional por usuario, por el otro lado Linux tiene licencia GNU General Public License o simplemente su acrónimo del inglés GNU GPL, que significa que es libre y no tiene costo la licencia de uso.

La manera más sencilla de crear un disco de arranque de Linux es descargando la imagen de la distribución deseada desde el siguiente enlace: http://www.linux.org/dist/list.html.

Posteriormente hay que grabar la Imagen en un CD-ROM/DVD para poder utilizarlo en la instalación del Sistema Operativo.

Para preparar un disco de arranque de Windows es necesario comprar la versión que se requiera en los diferentes distribuidores con los que cuenta Microsoft, Sun Microsystems, etc.

Live CD o Live DVD

> Live distro (CD autónomo) es un sistema operativo (normalmente acompañado de un conjunto de aplicaciones) almacenado en un medio extraíble, tradicionalmente un CD o un DVD que puede ejecutarse desde éste sin necesidad de instalarlo en el disco duro de una computadora, para lo cual usa la memoria RAM como disco duro virtual y el propio medio como sistema de archivos.
>
> Algunos Live CD incluyen una herramienta que permite instalarlos en el disco duro. Otra característica es que por lo general no se efectúan

cambios en la computadora utilizada, aunque algunos pueden almacenar preferencias si así se desea.

Para usar un Live CD es necesario obtener uno (muchos de ellos distribuyen libremente una imagen ISO que puede bajarse de Internet y grabarse en disco) y configurar la computadora para que arranque desde la unidad lectora, reiniciando luego la computadora con el disco en la lectora, con lo que el Live CD se iniciará automáticamente.[90]

7.4 Planeación de la utilización de los discos

En la planeación para la utilización de los discos se debe de considerar qué tipo de sistema operativo se va a instalar, se asigna la partición mediante algún sistema de archivos como: FAT, NTFS, ext3, ext2, FAT32, ReiserFS, Reiser4 u otro. En Windows, las particiones reconocidas son identificadas con una letra seguida por un signo de doble punto (p.e C:\). En sistemas basados en linux, se le asigna un archivo especial en la carpeta /dev a cada partición (p.e. hda1, sda2, etc.); el archivo recibe un nombre compuesto de tres letras seguidas de un número. Estos archivos especiales representan la partición, y gracias a estos archivos, una partición puede montarse en cualquier archivo del sistema.

Un único disco físico puede contener hasta cuatro particiones primarias; prácticamente todo tipo de discos magnéticos y memorias flash pueden particionarse. Sin embargo, para tener la posibilidad de más particiones en un solo disco, se utilizan las particiones extendidas, las cuales pueden contener un número ilimitado de particiones lógicas en su interior. Para este último tipo de particiones, no es recomendado su uso para instalar ciertos sistemas operativos, sino que son más útiles para guardar documentos o ejecutables no indispensables para el sistema. Los discos ópticos (DVD, CD) no soportan particiones.

Existen 3 tipos de particiones:

1. *Partición primaria*: Son las divisiones crudas o primarias del disco, solo puede haber 4 de éstas. Depende de una tabla de particiones. Un disco físico completamente formateado, consiste en realidad de una partición primaria que ocupa todo el espacio del disco, y posee un sistema de archivos. A este tipo de particiones, prácticamente cualquier sistema operativo puede

[90] Wikipedia, “Live CD o Live DVD”, actualizado por última vez el 11/01/09, disponible en, http://es.wikipedia.org/wiki/CD_aut%C3%B3nomo Fecha de recuperación: 09 de enero de 2009.

detectarlas y asignarles una unidad, siempre y cuando el sistema operativo reconozca su formato (sistema de archivos).

2. *Partición extendida*: Es otro tipo de partición que actúa como una partición primaria; sirve para contener infinidad de unidades lógicas en su interior. Fue ideada para romper la limitación de 4 particiones primarias en un solo disco físico. Solo puede existir una partición de este tipo por disco, y solo sirve para contener particiones lógicas. Por lo tanto, es el único tipo de partición que no soporta un sistema de archivos directamente.
3. *Partición lógica*: Ocupa un trozo de partición extendida o la totalidad de la misma, la cual se ha formateado con un tipo específico de sistema de archivos (FAT32, NTFS, ext2,...) y se le ha asignado una unidad, si el sistema operativo reconoce las particiones lógicas o su sistema de archivos.[91]

Razones para el uso de particiones.

- Algunos sistemas de archivos (p.ej. versiones antiguas de sistemas FAT de Microsoft) tienen tamaños máximos más pequeños que los que el tamaño que proporciona un disco, siendo necesaria una partición de tamaño pequeño, para que sea posible el adecuado funcionamiento de este antiguo sistema de archivos.
- Se puede guardar una copia de seguridad de los datos del usuario en otra partición del mismo disco, para evitar la pérdida de información importante. Esto es similar a un RAID, excepto en que está en el mismo disco.
- En algunos sistemas operativos aconsejan más de una partición para funcionar, como por ejemplo, la partición de intercambio (swap) en los sistemas operativos basados en Linux.
- A menudo, dos sistemas operativos no pueden coexistir en la misma partición, o usar diferentes formatos de disco “nativo”. La unidad se particiona para diferentes sistemas operativos.
- Uno de los principales usos que se le suele dar a las particiones (principalmente a la extendida) es la de almacenar toda la información del usuario (música, fotos, vídeos, documentos), para que al momento de reinstalar algún sistema operativo se formatee únicamente la unidad que lo contiene sin perder el resto de la información del usuario.

A lo largo de los años han aparecido numerosos sistemas de particionamiento, para casi todas las arquitecturas de computadoras existentes. Muchos son relativamente transparentes y permiten la

[91] Blog: Arquitectura de PCs: “Las particiones”, 20/08/08, disponible en: http://arquitecturapcs.blogspot.com/2008_08_20_archive.html. Fecha de recuperación: 09 de enero de 2009.

manipulación conveniente de las particiones de disco; algunos, sin embargo, son obsoletos.

Este esquema se considera obsoleto, porque sólo admite discos duros de más de 8 gigabytes de espacio. Como la arquitectura IBM PC es muy común, las tablas de partición probablemente subsistirán cierto tiempo. Sin embargo, un proyecto reciente de Intel y Microsoft llamado Extensible Firmware Initiative (EFI) tiene un componente llamado GUID Partition Table.[92]

Gparted (Linux/Unix)

Es el editor de particiones de GNOME. Esta aplicación es usada para crear, destruir, redimensionar, inspeccionar y copiar particiones, como también sistemas de archivos.

Esto es útil para crear espacio para nuevos sistemas operativos, para reorganizar el uso del disco y para crear imágenes de un disco en una partición. QtParted, es la contraparte de GParted pero para entornos de escritorios KDE.

Gparted se encuentra disponible en un LiveCD, basado en Slackware y construido sobre la última rama estable núcleo de Linux (2.6). El LiveCD es actualizado con cada lanzamiento de GParted. El LiveCD de Ubuntu incluye esta aplicación entre sus utilidades. También se encuentra disponible en una versión LiveUSB.[93]

FDisk (DOS)

Fdisk es un programa de computadora disponible en varios sistemas operativos que permite editar las particiones de un disco duro.

7.5 Creación del Sistema de Archivos

Aunque los discos duros pueden ser muy chicos, aún así contienen millones de bits, y por lo tanto necesitan organizarse para poder ubicar la información. Éste es el propósito del sistema de archivos. Un disco duro

---

[92] SENA, *Instalación de redes*, p. 5/18, material en línea, disponible en: http://www.scribd.com/doc/5997438/E5 Fecha de recuperación: 09 de enero de 2009.

[93] Wikipedia, "Partición de disco: Gparted", actualizado el 09/12/08, disponible en: http://es.wikipedia.org/wiki/Partici%C3%B3n_de_disco Fecha de recuperación: 09 de enero de 2009.

se conforma de varios discos circulares que giran en torno a un eje. Las pistas (áreas concéntricas escritas a ambos lados del disco) se dividen en piezas llamadas sectores (cada uno de los cuales contiene 512 bytes). El formateado lógico de un disco permite que se cree un sistema de archivos en el disco, lo cual, a su vez, permitirá que un sistema operativo (DOS, Windows 9x, UNIX,...) use el espacio disponible en disco para almacenar y utilizar archivos. El sistema de archivos se basa en la administración de clústers, la unidad de disco más chica que el sistema operativo puede administrar.

Un clúster consiste en uno o más sectores. Por esta razón, cuanto más grande sea el tamaño del clúster, menores utilidades tendrá que administrar el sistema operativo.

Por el otro lado, ya que un sistema operativo sólo sabe administrar unidades enteras de asignación (es decir que un archivo ocupa un número entero de clústers), cuantos más sectores haya por clúster, más espacio desperdiciado habrá. Por esta razón, la elección de un sistema de archivos es importante.[94]

Sistema Operativo	Tipos De Sistemas De Archivos Admitidos
Dos	FAT16
Windows 95	FAT16
Windows 95 OSR2	FAT16, FAT32
Windows 98	FAT16, FAT32
Windows NT4	FAT, NTFS (versión 4)
Windows 2000/XP	FAT, FAT16, FAT32, NTFS (versiones 4 y 5)
Linux	Ext2, Ext3, ReiserFS, Linux Swap (FAT16, FAT32, NTFS)
MacOS	HFS (Sistema de Archivos Jerárquico), MFS (Sistemas de Archivos Macintosh)

[94] Kioskea, “El sistema de archivos”, actualizado el 16/10/08, material en línea, disponible en: http://es.kioskea.net/contents/repar/filesys.php3 Fecha de recuperación: 09 de enero de 2009.

OS/2	HPFS (Sistema de Archivos de Alto Rendimiento)
SGI IRIX	XFS
FreeBSD, OpenBSD	UFS (Sistema de Archivos Unix)
Sun Solaris	UFS (Sistema de Archivos Unix)
IBM AIX	JFS (Sistema Diario de Archivos)

Funciones del sistema de archivos

Los usuarios deben poder crear, modificar y borrar archivos. Se deben poder compartir los archivos de una manera controlada.

El mecanismo encargado de compartir los archivos debe proporcionar varios tipos de acceso controlado:

- Ej.:*"Acceso de Lectura"*, *"Acceso de Escritura"*, *"Acceso de Ejecución"*, varias combinaciones de estos, etc.

Se debe poder estructurar los archivos de la manera más apropiada a cada aplicación.

Los usuarios deben poder ordenar la transferencia de información entre archivos.

Se deben proporcionar posibilidades de "*respaldo*" y "*recuperación*" para prevenirse contra:

- La pérdida accidental de información.
- La destrucción maliciosa de información.

Se debe poder referenciar a los archivos mediante "*Nombres Simbólicos*", brindando "*Independencia de Dispositivos*".

En ambientes sensibles, el sistema de archivos debe proporcionar posibilidades de "*Cifrado*" y "*Descifrado*".

El sistema de archivos debe brindar una interfaz favorable al usuario:

- Debe suministrar una "*visión lógica*" de los datos y de las funciones que serán ejecutadas, en vez de una "*visión física*".
- El usuario no debe tener que preocuparse por:
  - Los dispositivos particulares.

- Dónde serán almacenados los datos.
- El formato de los datos en los dispositivos.
- Los medios físicos de la transferencia de datos hacia y desde los dispositivos. [95]

7.6 Administración del espacio libre

En la implantación del sistema de archivos y su relación con la asignación y liberación de espacio se consideran aspectos tales como:

- La forma de almacenamiento de archivos y directorios.
- La administración del espacio en disco.
- La forma de hacerlo de manera eficiente y confiable.

Se deben tener presentes problemas tales como la "fragmentación" creciente del espacio en disco:

- Ocasiona problemas de desempeño al hacer que los archivos se dispersen.
- Una técnica para solucionar el problema de la "fragmentación" consiste en realizar periódicamente:
  - "*Condensación*": se pueden "reorganizar" los archivos expresamente o automáticamente según algún criterio predefinido.
  - "*Recolección de basura o residuos*": se puede hacer fuera de línea o en línea, con el sistema activo, según la implementación.[96]

Administración del espacio en disco

Existen dos estrategias generales para almacenar un archivo de "n" bytes:

1. Asignar "n" bytes consecutivos de espacio en el disco:
   Tiene el problema de que si un archivo crece será muy probable que deba desplazarse en el disco, lo que puede afectar seriamente al rendimiento.

2. Dividir el archivo en cierto número de bloques (no necesariamente) adyacentes:

---

[95] S/A, *Sistemas operativos*, 17/11/02, en diapositivas, #22, material electrónico, disponible en: http://arantxa.ii.uam.es/~siguenza/Sistemas%20operativos.ppt.rRecuperado el 12/01/09.

[96] Departamento de Informática, UNNE, *Sistemas operativos*, material en línea, disponible en: http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/SO4.htm. Fecha de recuperación: 09 de enero de 2009.

Generalmente los sistemas de archivos utilizan esta estrategia con bloques de tamaño fijo. [97]

Tamaño del bloque

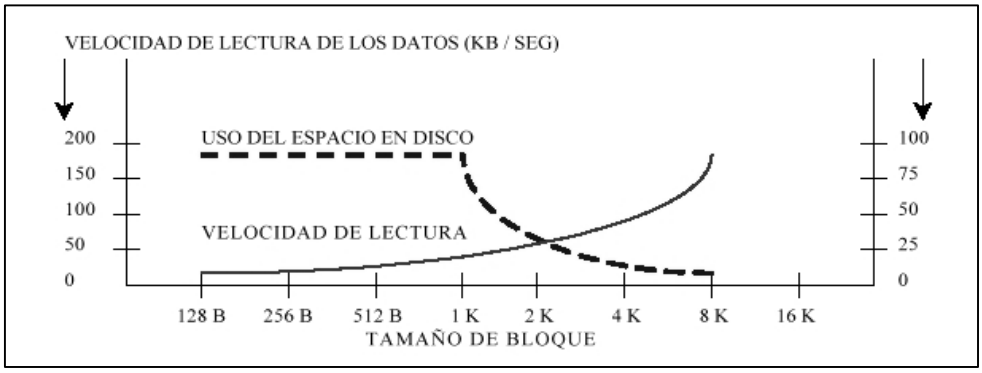
Dada la forma en que están organizados los bloques, el sector, la pista y el cilindro son los candidatos obvios como unidades de asignación.

Si se tiene una unidad de asignación grande, como un cilindro, esto significa que cada archivo, inclusive uno pequeño, ocupará todo un cilindro; con esto se desperdicia espacio de almacenamiento en disco.

Si se utiliza una unidad de asignación pequeña, como un sector, implica que cada archivo constará de muchos bloques; con esto su lectura generará muchas operaciones de e/s afectando el desempeño.

Lo anterior indica que la eficiencia en tiempo y espacio tienen un conflicto inherente.

Generalmente se utilizan como solución de compromiso bloques de 1/2 k, 1k, 2k o 4k. (Ver Figura 7.1).



Hay que recordar que el tiempo de lectura de un bloque de disco es la suma de los tiempos de:

- Búsqueda.
- Demora rotacional.
- Transferencia.

Registro de los bloques libres

Se utilizan por lo general dos métodos:

---

[97] Andrew S. Tanenbaum: *Sistemas operativos modernos*, México, Pearson Education, 2003. Apud: http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/SO4.htm Por cierto, cuando el lector quiera saber más datos de ese Departamento de Informática, puede visitar este sitio: http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/SO0.htm

1. La lista de bloques libres como lista ligada.
2. Un mapa de bits.

Lista ligada de bloques de disco:

- Cada bloque contiene tantos números de bloques libres como pueda.
- Los bloques libres se utilizan para contener a la lista de bloques libres.

Mapa de bits:

- Un disco con "n" bloques necesita un mapa de bits con "n" bits.
- Los bloques libres se representa con "1" y los asignados con "0" (o viceversa).
- Generalmente este método es preferible cuando existe espacio suficiente en la memoria principal para contener completo el mapa de bits.

Disk quotas

Para evitar que los usuarios se apropien de un espacio excesivo en disco, los sistemas operativos multiusuario proporcionan generalmente un mecanismo para establecer las cuotas en el disco.

La idea es que:

- Un administrador del sistema asigne a cada usuario una proporción máxima de archivos y bloques.
- El sistema operativo garantice que los usuarios no excedan sus cuotas.

Un mecanismo utilizado es el siguiente:

- Cuando un usuario **abre un archivo**:
  - Se localizan los atributos y direcciones en disco.
  - Se colocan en una tabla de archivos abiertos en la memoria principal.
  - Uno de los atributos indica el propietario del archivo; cualquier aumento del tamaño del archivo se carga a la cuota del propietario.
  - Una segunda tabla contiene el registro de las cuotas para cada uno de los usuarios que tengan un archivo abierto en ese momento, aún cuando el archivo lo haya abierto otro usuario.

- Cuando se escribe una **nueva entrada en la tabla de archivos abiertos:**
  - 'Se introduce un apuntador al registro de la cuota del propietario para localizar los límites.

- Cuando se **añade un bloque a un archivo:**
  - Se incrementa el total de bloques cargados al propietario.
  - Se verifica este valor contra los límites estricto y flexible (el primero no se puede superar, el segundo sí).
  - También se verifica el número de archivos.[98]

7.7 Instalación de Shells, herramientas y compiladores

"El intérprete de comandos es la interfaz entre el usuario y el sistema operativo, por este motivo se le da el nombre "shell", que en castellano significa 'caparazón'".[99]

Linux

El núcleo es la parte del sistema operativo que sirve para interactuar con el hardware. Proporciona una serie de servicios que pueden ser utilizados por los programas, sin que éstos tengan que preocuparse de cómo se administra el hardware.

En general, el núcleo es el encargado de administrar la memoria, mantener el sistema de archivos, manejo de las interrupciones, manejo de errores, realización de los servicios de entrada/salida, asignación de los recursos de la CPU, gestión de periféricos de entrada/salida, etcétera.

Cada programa se relaciona con la máquina a través del núcleo. Un programa realizará al núcleo las llamadas al sistema. Con estas el programa indicará, por ejemplo, que le abra un archivo, que escriba en otro, que utilice la impresora, que cambie la prioridad de ejecución de otro proceso, etcétera.

El núcleo del sistema operativo Unix/Linux, que recibe el nombre de KERNEL, actúa directamente con los elementos físicos de la computadora, y se carga en memoria al arrancar la máquina. Permanece en ella hasta que ésta se apaga. Recordemos que en DOS, el núcleo estaba formado por dos programas MSDOS.SYS y IO.SYS.

---

[98] Departamento de Informática, UNNE: *Sistemas operativos:* "Administración del espacio en disco", material en línea, disponible en: http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/SO4.htm, recuperado el 08/01/09.

[99] Kioskea: *Linux Shell*, actualizado el 16/10/08, material en línea, disponible en: http://es.kioskea.net/contents/linux/linshell.php3, recuperado el 08/01/09.

El Shell es el intérprete de mandatos o de comandos con el que cuenta este sistema operativo. En DOS es el Command. Com el que se encarga de realizar esta función.

El Shell actúa como interfaz de comunicación entre el usuario y la computadora, y cuando un usuario se conecta con el servidor Unix/Linux, automáticamente se arranca un Shell para que pueda trabajar. Cada usuario conectado al servidor tendrá un Shell para su uso.

Al contrario que en DOS, en el que el intérprete de comandos es único, en Unix/Linux existen varios. Éstos son los siguientes:

- Shell Bourne (sh). Creado por S. Bourne, es el más utilizado en la actualidad. El prompt del sistema queda representado por el símbolo «$». Este shell es el estándar de AT&T y el que se monta en casi todos los sistemas Unix/Linux.
- C-Shell (csh). Procedente del sistema BSD, proporciona características tales como control de trabajos, historia de comandos (como el doskey en DOS), capacidades de edición, etc. Ofrece importantes características para los programadores que trabajan en lenguaje C. Su prompt de sistema queda representado con el símbolo «%».
- Shell job (jsh). Incorpora algunas características de control al shell estándar del sistema.
- Shell Korn (ksh). Escrito por David Korn, amplia el shell del sistema añadiendo historia de comandos, edición de la línea de órdenes y características ampliadas de programación.
- Bourne Again shell (Bash). Fue creado para usarlo en el proyecto GNU. BASH, por lo tanto, es un shell o intérprete de comandos GNU; éste es compatible con el shell sh. Además, incorpora algunas características útiles de ksh y csh, y otras propias, como la edición de línea de comandos, tamaño ilimitado del histórico de comandos, control de trabajos y procesos, funciones y alias, cálculos aritméticos con números enteros, etcétera.[100]

Windows

Windows Shell es el aspecto más visible de la línea de Microsoft Windows de los sistemas operativos. Es el contenedor dentro de la que toda la interfaz de usuario se presenta, incluyendo la barra de tareas,

[100] F.J. Muñoz: *Linux –Unix. El núcleo y el* Shell, MailxMax, material electrónico en línea, disponible en: http://www.mailxmail.com/curso/informatica/linux-unix/capitulo5.htm. Fecha de recuperación: 09 de enero de 2009.

el escritorio, el explorador de Windows, así como muchos de los cuadros de diálogo y controles de interfaz.

El valor por defecto se llama shell de Windows Explorer - este es el programa que determina el aspecto del escritorio (se crea la barra de tareas, el área de notificación, el menú de inicio, etc.).

A través del tiempo Windows se ha ido desarrollando generando versiones como: Windows 95, Windows 95C, 98, Windows 2000 (NT 5.0), XP, Server 2003, Vista. Cada una ha mejorado su interfaz con el usuario.

La interfaz de usuario Vista-Aero en la actualidad está incluido en todas las versiones de Windows Vista (con excepción de la version Home Basic), así como Windows Server 2008.[101]

7.8 Creación de usuarios y grupos

Windows NT

El concepto de usuario

Windows NT es un sistema operativo que administra sesiones. Esto significa que cuando se inicia un sistema, es necesario registrarse con un nombre de usuario y una contraseña.

Cuando se instala Windows NT, la cuenta de administrador se crea de forma predeterminada, así como también una cuenta denominada invitado. Es posible (y se recomienda) modificar los permisos de los usuarios (qué acciones pueden realizar) y también agregar usuarios mediante el Administrador de usuarios. Una cuenta de usuario es una identificación asignada de manera única al usuario para permitirle:

- Iniciar sesión en un dominio para acceder a los recursos de red.
- Iniciar sesión en un equipo local para acceder a los recursos locales.

Por lo tanto, todos los usuarios que utilizan habitualmente la red deben tener una cuenta.

Administración de usuarios

El Administrador de usuarios es la utilidad estándar que ofrece Windows NT. Como su nombre indica, se encarga de la administración de los

[101] Wikipedia: "Windows Shell", actualizado el 07/10/08, material en línea, disponible en: http://es.wikipedia.org/wiki/Windows_Shell Fecha de recuperación: 09 de enero de 2009.

usuarios. Se encuentra en el Menú Inicio (Programas/Herramientas de administración). Figura 7.1

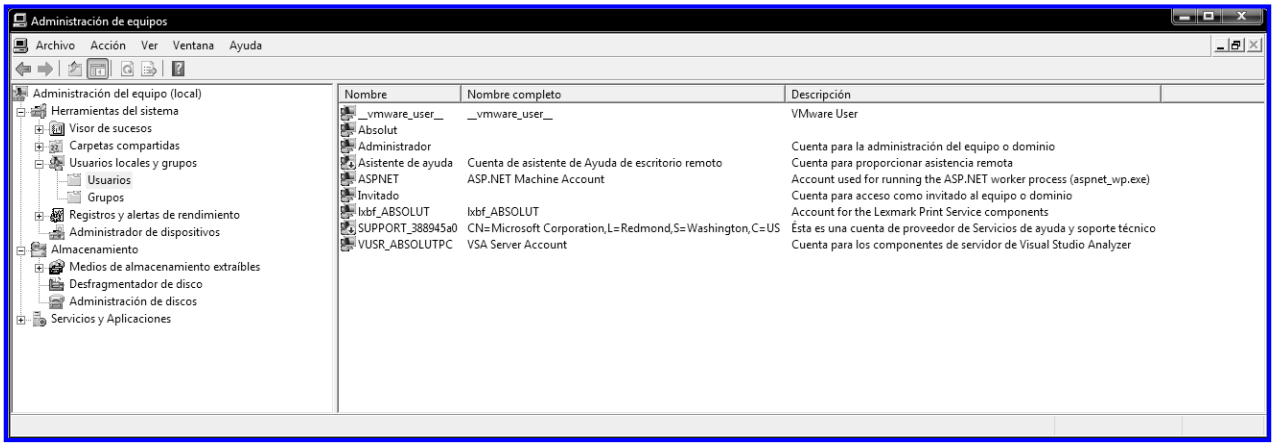


Figura 7.1 Administración de usuarios

Para la creación de una cuenta nueva, Aparecerá un cuadro de diálogo para especificar la información acerca del nuevo usuario (Figura 7.2):

- Usuario: Nombre de inicio de sesión del usuario.
- Nombre completo: Información opcional del usuario.
- Descripción: Campo opcional.
- Los campos para la Contraseña son opcionales. Aún así, se recomienda llenarlos y marcar la casilla con la etiqueta "El usuario debe cambiar la contraseña" por razones de seguridad.

Figura 7.2 Creación de un nuevo usuario

Convenciones para el nombre de usuario

El administrador identifica a los usuarios por medio de convenciones para el nombre del usuario. Se debe tener en cuenta la siguiente información:

- Los nombres de usuario deben ser únicos (dentro de un dominio o en un equipo local)
- Los nombres de usuario pueden contener cualquier letra mayúscula o minúscula, pero no debe contener los siguientes caracteres: / \ [ ]: . | = , . | = , + * ? < >
- Hay que evitar crear nombres de usuario similares.

Cuentas y seguridad de usuario

Los diferentes tipos de cuentas son:

- *Cuentas de usuarios*: Creadas para iniciar sesión en una red y acceder a sus recursos. Estas cuentas poseen información acerca del usuario, en particular su nombre y contraseña.
- *Invitado*: Permite que, en ocasiones, los usuarios inicien sesión y tengan acceso al equipo local. Esta opción está desactivada de forma predeterminada.
- *Administrador*: Se utiliza para administrar la configuración global de equipos y dominios. Esta cuenta puede llevar a cabo cualquier tarea.

Es fundamental:

- En primer lugar, desactivar la cuenta de invitado, que permitiría que cualquier usuario inicie sesión en el sistema.
- En segundo lugar, cambiar el nombre de la cuenta de administrador para reducir el riesgo de intrusión mediante esta cuenta. Debido a que la cuenta de administrador posee todos los permisos, es un objetivo prioritario de los posibles intrusos.

Administración del entorno de trabajo del usuario

Cuando un usuario inicia sesión por primera vez desde un cliente que ejecuta Windows NT, se crea un perfil de usuario predeterminado para ese usuario. Este perfil configura elementos como, por ejemplo, el entorno de trabajo del usuario, y las conexiones de red y de impresoras. Este perfil se puede personalizar para restringir ciertos elementos en el escritorio u ocultar algunas herramientas del equipo.

Administración de grupos

Windows NT también permite que la administración de usuario mediante grupo. Esto significa que puede definir grupos de usuarios con el mismo tipo de permisos, organizándolos en categorías.

Un grupo es un conjunto de cuentas de usuario. Un usuario que se agrega a un grupo obtiene todos los permisos y derechos de ese grupo. Los grupos de usuarios hacen más sencilla la administración, ya que es posible otorgar permisos a varios usuarios a la vez. Hay dos tipos diferentes de grupos:

- *Grupos locales*: Otorga a los usuarios permisos para que accedan a un recurso de red. También sirven para conceder a los usuarios privilegios para abrir tareas de sistema (como cambiar la hora, hacer copias de seguridad, recuperar archivos, etc.). Existen grupos locales pre-configurados.

- *Grupos globales*: Se usan para organizar las cuentas de usuario de dominio. También se usan en redes de varios dominios, cuando los usuarios de un dominio necesitan tener acceso a recursos de otro dominio.

Cuando se inicia Windows NT por primera vez, se crean seis grupos de forma predeterminada:

A. Administradores.
B. Operadores de copia.
C. Duplicadores.
D. Usuarios Avanzados.
E. Usuarios.
F. Invitados.

Estos grupos predeterminados se pueden eliminar, y se pueden añadir grupos personalizados de usuarios con permisos especiales, de acuerdo con las operaciones que vayan a realizar en el sistema. (Figura 7.3).

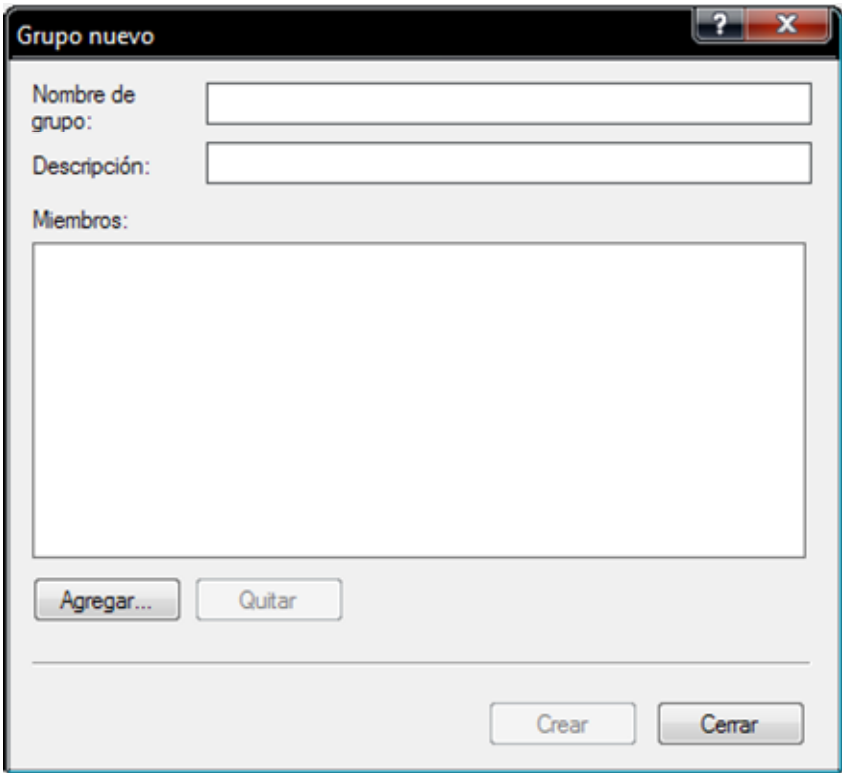


Figura 7.3 Creación de un nuevo grupo de usuarios

Implementación de grupos incorporados

Los grupos incorporados son aquellos que tienen privilegios de usuario predeterminados. Los privilegios de usuario determinan qué tareas puede ejecutar un usuario o miembro de un grupo incorporado. Estos son los tres grupos incorporados en Windows NT:

I. *Grupos locales incorporados*: Otorgan a los usuarios privilegios que les permiten ejecutar tareas de sistema como realizar copias de seguridad y restaurar datos, cambiar la hora y administrar los recursos del sistema. Se encuentran en todas los equipos que ejecutan Windows NT.

II. *Grupos globales incorporados*: Proporcionan a los administradores una forma sencilla de controlar a todos los usuarios del dominio. Los grupos globales se encuentran únicamente en los controladores de dominio.

III. *Grupos de sistema*: Organizan a los usuarios automáticamente en función del uso del sistema. Los administradores no agregan usuarios a estos grupos. Los usuarios pueden ser miembros de estos grupos de forma predeterminada, o pueden convertirse en miembros a través de su actividad en la red. Se encuentran en todos los equipos que ejecutan Windows NT.

No se puede cambiar el nombre ni eliminar ninguno de los grupos incorporados. Los grupos locales incorporados son los siguientes:

a) **Usuarios:** Pueden ejecutar las tareas para las que tienen privilegios de acceso, y pueden tener acceso a recursos para los que han obtenido permiso. El grupo local Usuarios avanzados reside únicamente en servidores miembro y equipos que ejecutan NT Workstation. Los miembros de este grupo pueden crear y modificar cuentas, y también compartir recursos.

b) **Administradores:** Pueden ejecutar todas las tareas administrativas en el equipo local. Si el equipo es un controlador de dominio, los miembros también pueden administrar el dominio entero.

c) **Invitados:** Pueden ejecutar cualquier tarea para la que tienen privilegios de acceso, y pueden obtener acceso a recursos por los que han obtenido permiso. Sus miembros no pueden modificar permanentemente sus entornos locales.

d) **Operadores de copia:** Pueden utilizar el programa de seguridad de Windows NT para hacer copias de seguridad y restaurar los equipos que ejecutan Windows NT[2]

e) **Duplicadores:** Utilizados por el servicio Duplicador de directorio. Este grupo no se utiliza para la administración.

Los siguientes grupos solo se definen en controladores de dominio:

a) **Operadores de cuentas:** Pueden crear, eliminar y modificar usuarios, grupos locales y grupos globales. No pueden modificar Administradores ni Operadores de servidores

b) **Operadores de servidores:** Pueden compartir recursos de disco, hacer copias de seguridad de los datos que están en los servidores y restaurarlos

c) **Operadores de impresión:** Pueden configurar y administrar impresoras de red. Cuando se instala Windows NT Server como controlador de dominio, se crean tres grupos globales en la SAM. De forma predeterminada, estos grupos no tienen ningún privilegio inherente. Adquieren privilegios cuando se agregan a grupos locales o cuando se les concede privilegios o permisos de usuario.

d) **Usuarios del dominio:** Se agrega automáticamente al grupo local Usuarios. De forma predeterminada, una cuenta de Administrador es miembro de este grupo.

e) **Administrador del dominio:** Se agrega automáticamente al grupo local Usuarios. Estos miembros pueden ejecutar tareas administrativas en el equipo local. De forma predeterminada, una cuenta de Administrador es miembro de este grupo.

f) **Invitados de dominio:** Se agrega automáticamente al grupo local Usuarios. De forma predeterminada, una cuenta de Invitado es miembro de este grupo.

Por último, los grupos incorporados del sistema residen en todas los equipos que ejecutan Windows NT. Los usuarios se vuelven miembros de estos grupos de forma predeterminada a medida que la red está en funcionamiento. El estado de los miembros no se puede modificar.

- **Todos** incluye a todos los usuarios remotos y locales con acceso al equipo. También contiene todas las cuentas, excepto las que el Administrador del dominio ha creado.
- **Creado por/Propietario** incluye al usuario que creado o ha obtenido la propiedad de un recurso. Este grupo se puede utilizar para administrar el acceso a los archivos y las carpetas sólo en volúmenes NTFS.
- **Red incluye** a cualquier usuario que esté conectado a un recurso compartido en su equipo desde otro equipo de la red.
- **Interactivo** incluye automáticamente a cualquier usuario que esté conectado localmente al equipo. Los miembros interactivos pueden acceder a recursos en el equipo al que están conectados.[102]

Linux

Creando una cuenta en el modo de texto/consola/shell: useradd y passwd

Este procedimiento puede realizarse de forma segura tanto fuera de X Window como desde una ventana terminal en el entorno gráfico del que se disponga.

El primer paso para crear una nueva cuenta consiste en utilizar el mandato **useradd** del siguiente modo:

```
useradd nombre_del_usuario
```

Ejemplo:

```
useradd unam
```

El Comando: passwd

Después de crear la nueva cuenta con useradd hay que especificar una contraseña para el usuario. Debe de ser fácil de recordar, que contenga

[102] Kioskea: *Administración de usuarios en Windows NT*, actualizado el 16/10/08, disponible en: http://es.kioskea.net/contents/winnt/ntusers.php3 Fecha de recuperación 09 de enero de 2009.

números, mayúsculas y minúsculas y que, preferentemente, no contenga palabras que se encontrarían fácilmente en el diccionario.

Aunque el sistema siempre tratará de prevenir la asignación de una mala contraseña, no impedirá que se haga. Especificar una nueva contraseña para un usuario, o bien cambiar la existente, se puede realizar utilizando el comando passwd del siguiente modo:

passwd *nombre_del_usuario*

Ejemplo:

passwd *usr_un4m*

El sistema solicitará entonces que proceda a teclear la nueva contraseña para el usuario y que repita ésta para confirmar. El sistema no mostrará los caracteres tecleados, por lo que debe hacerse con cuidado, el sistema informará si coincide o no lo tecleado. Si todo salió bien recibirá como respuesta del sistema code 0. Si en cambio recibe code 1, significa que deberá repetir el procedimiento, ya que ocurrió un error.

Este procedimiento también puede utilizarse para cambiar una contraseña existente.

Opciones avanzadas

Esto se utiliza para crear una cuenta con mayores restricciones, atributos y/o permisos, pueden utilizarse las siguientes opciones de **useradd**:

- C comment
  Se utiliza para especificar el archivo de comentario de campo para la nueva cuenta.

- D home dir
  Se utiliza para establecer el directorio de trabajo del usuario. Es conveniente, a fin de tener un sistema bien organizado, que este se localice dentro del directorio*/home*.

- E expire date
  Se utiliza para establecerla fecha de expiración de una cuenta de usuario. Esta debe ingresarse en el siguiente formato: AAAA-MM-DD.

- G initial group

Se utiliza para establecer el grupo inicial al que pertenecerá el usuario. De forma predeterminada se establece como único grupo *1*. Nota: el grupo asignado debe de existir.

- G group,[...]
  Se utiliza para establecer grupos adicionales a los que pertenecerá el usuario. Estos deben separarse utilizando una coma y sin espacios. Esto es muy conveniente cuando se desea que el usuario tenga acceso a determinados recursos del sistema, como acceso a la unidad de disquetes, administración de cuentas PPP y POP. Nota: los grupos asignado deben de existir.

- m
  Se utiliza para especificar que el directorio de trabajo del usuario debe ser creado si acaso este no existiese, y se copiaran dentro de este los archivos especificados en*/etc/skel*.

- S shell
  Se utiliza para establecer el *Shel l*que podrá utilizar el usuario. De forma predeterminada, en Red Hat Linux y Fedora Core, se establece *bash* como *Shell* predefinido.

- U uid
  Se utiliza para establecer el UID, es decir, la ID del usuario. Este debe ser único. De forma predeterminada se establece como UID el número mínimo mayor a 99 y mayor que el de otro usuario existente. Cuando se crea una cuenta de usuario por primera vez, como ocurre en Red Hat Linux y Fedora Core generalmente se asignará *500* como UID del usuario. Los UID entre 0 y 99 son reservados para las cuentas de los servicios del sistema.

Ejemplo:

```
useradd -u 500 -d /home/unam -G
respaldo,desarrollo,correo unam
```

Manejo de Grupos

I. Alta de grupos

```
groupadd desarrollo
```

II. Alta de grupos de sistema
Un grupo de sistema es aquel que tiene un número de identidad de grupo (GID) por debajo del 500. Regularmente se asigna automáticamente el número de identidad de grupo más bajo disponible.

```
groupadd -r desarrollo
```

III. Baja de grupos

```
groupdel respaldo
```

Asignación de usuario existente a grupos existentes

```
gpasswd -a unam desarrollo[103]
```

El archivo /**etc/group** contiene una lista de los usuarios que pertenecen a los diferentes grupos. De hecho, cuando un gran número de usuarios tiene acceso al sistema, generalmente están ubicados en grupos diferentes, cada uno de los cuales posee sus propios derechos de acceso a los archivos y a los directorios.

Tiene diferentes campos separados por "**:**":

```
nombre_de_grupo : campo_especial : numero_de_grupo : miembro1,
miembro2
```

Con frecuencia, el campo especial está vacío.

El número de grupo corresponde al número del vínculo entre los archivos /etc/group y los archivos /etc/passwd.

[103] Joel Barrios Dueñas: *Cómo crear cuentas de usuario*, Linux para todos, material en línea, CC, disponible en: http://www.linuxparatodos.net/portal/staticpages/index.php?page=02-cuentas-usuario&mode=print, recuperado el 13/01/09. Ejemplos ligeramente alterados.

Ejemplo de un archivo ***/etc/group***:

root:x:0:root
bin:x:1:root,bin,daemon
daemon:x:2:
tty:x:5:
disk:x:6:
lp:x:7:
wwwadmin:x:8:
kmem:x:9:
wheel:x:10:
mail:x:12:cyrus
news:x:13:news

* Cuando el comando ls se utiliza con la opción -l, el número de grupo se muestra junto con el del usuario al que pertenece el archivo (o directorio). Este número único corresponde al nombre de grupo único (a menudo tiene un máximo de 8 caracteres).
* El mismo usuario puede aparecer en varios grupos. Cuando se conecta al sistema, el usuario pertenece a un grupo especificado en /etc/passwd (en el campo GID). Puede modificarlo usando el comando newgrp. Luego se definen los derechos de acceso de archivo.
* Las protecciones del archivo deben impedir que los usuarios sin privilegios puedan modificar los archivos.
* Para añadir un grupo, el administrador puede cambiar el archivo /etc/group con un editor de texto. También puede utilizar el comando addgroup o groupadd (Como se explico anteriormente). En primer lugar, sólo tendrá que añadir las líneas relacionadas con los grupos. Por ejemplo, la línea:

```
admin : : 56 : cf
```

* Para agregar un usuario a un grupo, sólo debe editar el archivo /etc/group y agregar el nombre al final de la línea, separando los nombres de los miembros por medio de una coma.
* Para borrar un grupo, edite el archivo /etc/group y elimine la línea correspondiente. Importante: no debe olvidar cambiar los números de los grupos eliminados (GID) en el archivo /etc/passwd, si los usuarios pertenecen a ese grupo. También es importante buscar los archivos y directorios de este grupo para cambiarlos (de lo contrario, los archivos y directorios pueden volverse inaccesibles).[104]

[104] Kioskea: *Linux, gestión de usuarios*, actualizado el 16/10/08, disponible en: http://es.kioskea.net/contents/unix/unix-users.php3 Fecha de recuperación: 09 de enero de 2009.

Bibliografía del tema 7

Tanenbaum, Andrew S. *Sistemas Operativos Modernos*. 2ª ed., México, Pearson Educación, 2003.

H. M. Deitel. *Introducción a los Sistemas Operativos*. México, Addison-Wesley Iberoamericana, 1987.

Actividades de aprendizaje

A.7.1. Describe con tus propias palabras cual es la diferencia principal entre una cuenta de administrador y un administrador de un sistema, asimismo menciona las operaciones que están restringidas a la cuenta del administrador.

A.7.2. Realiza un cuadro comparativo sobre las ventajas y desventajas de utilizar los sistemas operativos Windows y Linux.

A.7.3. Describe las consideraciones más importantes sobre la planeación de la utilización de discos.

Cuestionario de autoevaluación

1. ¿Qué es un Live CD?
2. ¿Qué es una partición extendida?
3. ¿Qué es una partición primaria?
4. ¿Cuál es la función de la cuenta de administrador o súper usuario?
5. ¿Qué es el kernel?
6. ¿Cuál es el propósito de un sistema de archivos?
7. ¿Para qué se utiliza la cuota de disco?
8. ¿Cuál es la función del programa fdisk?
9. ¿Cuál es la función de un Shell?
10. ¿Qué es la condensación de archivos?

Examen de autoevaluación

1. ¿Cuál es el “Shell” en el sistema operativo Windows?

a) KSH

b) MS-DOS

c) CSH

2. ¿Cuál es la partición que corresponde al sistema operativo Linux?

a) FAT, NTFS

b) Ext2, Ext3

c) XFS, NXS

3. ¿Cuál es la función de un sistema de archivos?

a) permitir acceso a la red

b) permitir acceso a los archivos

c) permitir acceso a los permisos

4. ¿Cuál es la característica del usuario invitado en Windows?

a) Capacidad para iniciar solo servicios

b) Capacidad limitada y controlada por el administrador

c) Capacidad ilimitada para ejecutar programas

5. ¿Cuál es el comando para crear un usuario en Linux?

a) adduser usuario

b) SUDO usuario

c) useradd usuario

6. ¿Cuál es el tipo de licencia de Linux?

a) BSD

b) GNU/GPL

c) EULA

7. ¿Cuál es el comando para crear un grupo en Linux?

a) newgroup -r grupo

b) sudo group + grupo

c) groupadd –r grupo

8. ¿Cuál es la función de “fdisk” en ms-dos?

a) permite editar particiones en el disco duro

b) permite corregir sectores en el disco duro

c) permite comprimir los datos en el disco duro

9. ¿Qué permite realizar “disk quota”?

a) permite utilizar discos duros restringidos

b) permite el acceso solo al administrador del sistema

c) permite asignar tamaños de uso de disco

10. ¿Cuál es la filosofía de Linux?

a) es un software de código restringido

b) es un software de código abierto

c) es un software muy seguro

TEMA 8. TÓPICOS AVANZADOS DE SISTEMAS OPERATIVOS

Objetivo particular

Al culminar el aprendizaje de este tema, el alumno diferenciará las principales características de los sistemas operativos distribuidos y de red, asimismo reconocerá la importancia que tienen los controladores (drivers), para medir su eficiencia y desempeño.

Temario detallado

8.1 Eficiencia y rendimiento o desempeño del SO
8.2 Escritura de drivers (controladores)
8.3 Sistemas Operativos de Red
8.4 Sistemas Operativos Distribuidos
8.5 Servicios remotos en Internet

Introducción

La industria de la computación, la informática, la electrónica y las telecomunicaciones han tenido un desarrollo muy importante logrando sistemas con mayores capacidades como lo son los sistemas distribuidos y de red, una de las características más importantes de estos sistemas es que están basados en varios procesadores y están fuertemente acoplados entres sí. Los nodos de un sistema distribuido son una computadora completa y cuentan con una unidad central de proceso CPU, tarjeta de red y posiblemente un disco duro que requieren para realizar la paginación, así como sus periféricos de entrada-salida E/S. Estos sistemas se comunican a través de redes de alta velocidad utilizando diversos protocolos y servicios de red, además los nodos de un sistema distribuido pueden

estar dispersos por todo el mundo. En el presente tema se describen las principales características de estos sistemas operativos.

8.1 Eficiencia y rendimiento o desempeño del SO[105]

Recordemos que el sistema operativo es un administrador de recursos, por ello es importante poder determinar con qué efectividad este sistema los administra.

Aunque existe una gran capacidad para que los sistemas operativos mejoren el uso de los recursos, en muchas ocasiones por falta de conocimiento técnico no se implementa el análisis de su rendimiento y desempeño. Por lo que no se lleva a cabo lo siguiente:

- No existe un control y evaluación del uso de los recursos.
- Cuando se hacen controles específicos se generan grandes cantidades de datos que muchas veces no se sabe cómo interpretar.[106]

Durante los primeros años del desarrollo de las computadoras, el hardware representaba el costo dominante de los sistemas, y debido a ello los estudios de rendimiento se concentraban en el hardware.

Actualmente y según la tendencia:

- El software representa una importancia cada vez mayor de los presupuestos en el área de informática.
- El software incluye el sistema operativo de multiprogramación, de multiproceso, sistemas de comunicaciones de datos, sistemas de administración de bases de datos, sistemas de apoyo a varias aplicaciones, etc.

---

[105] Universidad Nacional del Nordeste, Departamento de Informática: “S.O. Rendimiento”, en línea, disponible en: http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/SO12.htm, recuperado el 16/02/09.
[106] H. M. Deitel, *Introducción a los Sistemas Operativos*. México, Addison-Wesley Iberoamericana, 1987.

- El software frecuentemente oculta el hardware al usuario creando una máquina virtual, que está definida por las características operativas del software.

Un software deficiente y / o mal utilizado puede ser causa de un rendimiento pobre del hardware, por lo tanto es importante controlar y evaluar el rendimiento del hardware y del software.

Tendencias importantes que afectan a los aspectos del rendimiento:
Con los avances en la tecnología de hardware los costos del mismo se han reducido drásticamente y todo hace suponer que esta tendencia continuará.

Los costos de trabajo (personal) han ido aumentando:

- Significan un porcentaje importante del costo de los sistemas informáticos.
- Se debe reformular el aspecto del rendimiento del hardware base y medirlo de manera más adaptada a la productividad humana.

El advenimiento del microprocesador en la década de los años 70 ha permitido:

- Bajar considerablemente el costo de los ciclos de CPU.
- Ha desplazado el foco de atención de la evaluación del rendimiento a otras áreas donde los costos no disminuyeron proporcionalmente; ej.: utilización de dispositivos de entrada / salida.

También influyen en la evaluación del rendimiento aspectos como:

- La construcción de redes
- El procesamiento distribuido

Las conexiones se hacen con redes y no solo con computadoras específicas:

- Se puede disponer de cientos o miles de sistemas de computación.

- Se puede acceder a complejos sistemas de comunicaciones de datos.[107]

Mediciones del rendimiento

El rendimiento expresa la manera o la eficiencia con que un sistema de computación cumple sus metas.

El rendimiento es una cantidad relativa más que absoluta, aunque suele hablarse de medidas absolutas de rendimiento, ej.: número de trabajos atendidos por unidad de tiempo.

Algunas mediciones son difíciles de cuantificar, Ej.: facilidad de uso.

Otras mediciones son fáciles de cuantificar, Ej.: accesos a un disco en la unidad de tiempo.

Las mediciones del rendimiento pueden estar:

- Orientadas hacia el usuario, Ej.: tiempos de respuesta.
- Orientadas hacia el sistema, ej.: utilización de la unidad central de proceso CPU.

Algunas mediciones del rendimiento más comunes son:

- Tiempo de regreso
- Tiempo desde la entrega del trabajo hasta su regreso al usuario (para procesamiento por lotes)
- Tiempo de respuesta
- Tiempo de regreso de un sistema interactivo
- Tiempo de reacción del sistema
- Tiempo desde que el usuario presiona la tecla “enter” hasta que se da la primera sección de tiempo de servicio

Las anteriores son cantidades probabilísticas y se consideran como variables aleatorias en los estudios de:

- Simulación.
- Modelado de sistemas.

Otras medidas del rendimiento utilizadas son:

[107] Stallings, Willings. *Sistemas Operativos*. 4ª ed., Pearson Educación, Madrid, 2001. 800 pp.

- Varianza de los tiempos de respuesta (o de otra de las variables aleatorias consideradas):
  - Es una medida de dispersión.
  - Si es pequeña indica tiempos próximos a la media.
  - Si es grande indica tiempos alejados de la media.
  - Es una medida de la predecibilidad (capacidad de prever el comportamiento)
- Capacidad de ejecución:
  - Es la medida de la ejecución de trabajo por unidad de tiempo.
- Carga de trabajo:
  - Es la medida de la cantidad de trabajo que:
    - Ha sido introducida en el sistema.
    - El sistema debe procesar normalmente para funcionar de manera aceptable.
- Capacidad:
  - Es la medida de la capacidad de rendimiento máxima que un sistema puede tener siempre que:
    - El sistema esté listo para aceptar más trabajos.
    - Haya alguno inmediatamente disponible.
- Utilización:
  - Es la fracción de tiempo que un recurso está en uso.
  - Es deseable un gran porcentaje de utilización, pero éste puede ser el resultado de un uso ineficiente.
  - Cuando se aplica a la CPU se debe distinguir entre:
    - Uso en trabajos productivos de aplicación.
    - Uso en sobrecarga del sistema.

8.2 Escritura de drives

El drive es una capa de código ubicada entre el dispositivo de hardware y la aplicación. Un driver usa los privilegios con los que se ejecuta su código para definir cómo se quiere que un dispositivo sea visto por una aplicación. Pueden existir diferentes drivers para un mismo dispositivo.

El driver se ubica en el **núcleo** (kernel) y realiza las siguientes tareas:

- Manejo de procesos: Creación y destrucción de procesos, comunicación entre procesos, asignación de CPU.
- Manejo de memoria: La memoria es un recurso crítico, y el núcleo administra su asignación.
- Sistemas de archivos
- Control de dispositivos (drivers)

- Networking (integración de redes)

El núcleo diferencia tres tipos de drivers:
  - Drivers de carácter (char devices)
  - Drivers de bloque (block devices)
  - Drivers de red (network devices)

No todos los drivers son de dispositivos. Algunos son de software. Por ejemplo, el driver de un sistema de archivos como [[ext3]] o [[reiserfs]] son drivers de software, que mapea estructuras de datos de bajo nivel a estructuras de datos de más alto nivel.

¿Por qué escribir un driver?
Existen muchas razones para querer escribir un driver.

- Para dar soporte a nuevo hardware
- Para mantener un producto propio
- Se está creando hardware a un ritmo muy rápido, y los programadores de drivers requieren estar programándolos.

Recomendaciones
I. Proveer mecanismos y no políticas

La idea es que un driver permita acceder a los dispositivos de hardware, sin imponer restricciones arbitrarias a los que usan el driver. En UNIX esto es una regla de diseño, que se conoce como separación de mecanismos y políticas.
Para no escribir políticas en el driver, se puede hacer una aplicación de usuario que se encargue de configurar el dispositivo, y/o una librería para acceder a él. La idea es que el driver cambie muy poco.

II. Uso de módulos
Los módulos tienen la ventaja de permitir adicionar y remover funcionalidades del núcleo.

III. Módulo para un driver de carácter
Este módulo:
- Usa un mayor número dinámico.
- Recibe parámetros cuando es insertado, y en un parámetro se puede especificar el mayor número.
- Usa memoria dinámica.
- Funciona llseek (implementa llamadas al sistema)
- Soporta llamadas ioctl (llamada a la función de control de resultados)

- Tiene un programa en el espacio del usuario para su configuración (ioctl).

8.3 Sistemas Operativos de Red

A un sistema operativo de red se le conoce como NOS. Es un software que rige y administra los recursos, archivos, periféricos, usuarios, etc., en una red y lleva el control de seguridad de los mismos.

Un NOS maneja los servicios necesarios para asegurar que el usuario final tenga o esté libre de error al ingresar a la red, normalmente este provee una interfaz de usuario para reducir la complejidad y conflictos al momento de usar la red. Dentro del contexto del NOS, se pueden escribir aplicaciones tales como un sistema de correo electrónico, este sistema pueden ser escrito para que se permitan "conexiones virtuales" entre entidades de red, sin la intervención humana directa.[108]

Figura 8.1 Esquema de un sistema operativo de red

En un entorno NOS (figura 8.1) cada nodo, mostrado aquí como círculo, es administrado por su sistema operativo local representado con un triángulo. Los sistemas operativos de red respectivos, denotados con un cuadro entran en funciones sólo cuando un sistema de un sitio necesita trabajar con el sistema del otro sitio.[109]

---

[108] Monografías: “Arquitectura del SO Windows NT”, en línea, disponible en: http://www.monografias.com/trabajos7/arso/arso.shtml, §2, recuperado el 16/02/09.
[109] Véase Ida Flynn, *Sistemas operativos*. 3ª ed., México, Thomson Learning, 2001, p. 236.

Análisis y Comparación de Sistemas Operativos de Red

Los sistemas operativos de red han ido mucho más allá del servidor de archivos y del servidor de impresión. Otras funciones como comunicaciones, bases de datos, aplicaciones y servicio de administración, son igualmente importantes.

Existe una organización llamada IDC (International Data Corporation, Corporación Internacional de Datos), la cual frecuentemente se encarga de comparar las características y funciones suministradas por cada uno de los mayores competidores en el área.

Las funciones de IDC son implementadas de dos formas:

- Como un ambiente operativo aislado, que puede o no permitir el soporte de servicios adicionales, como bases de datos o mensajes electrónicos.
- Como una capa de servicio adicional sobre un sistema operativo de propósito general, como UNIX u OpenMVS.

La mejor parte de los ambientes NOS es que proveen el más alto desempeño, cuando los requerimientos sólo son servicios de archivos y de impresión. Ellos frecuentemente son menos estables y robustos cuando soportan otras funciones. Cuando se requieren los servidores multifuncionales, la capa de NOS que funciona sobre un sistema operativo de propósito general se convierte en la mejor opción.

Los mejores NOS residen en el segundo o tercer nivel de la arquitectura distribuida. Los sistemas operativos de propósito general permiten mejorar las arquitecturas multinivel más que los sistemas de propósito general.

IDC construyó un modelo con las características y las funciones que se requieren para el soporte exitoso de funciones NOS. El modelo de requerimientos para NOS de IDC rompe con la complejidad, dividiéndose en las siguientes categorías:

a) Arquitectura.- Las organizaciones tienen cargas de trabajo crecientes y dinámicas que necesitan ser concernientes a la arquitectura de un NOS. Para lo cual se necesita considerar lo siguiente:

- ¿El NOS soporta sistemas de proceso múltiple (multiproceso)?
- ¿El NOS Soporta multiprocesamiento asimétrico o simétrico?
- ¿Pueden las funciones de NOS ser particionadas para correr en más de un procesador simultáneamente?

* ¿El NOS soporta múltiples arquitecturas de microprocesadores?

b) Escalabilidad.- Las compañías esperan un NOS que soporte pequeños, medianos y grandes ambientes usando el mismo NOS, para esto se necesitan considerar las siguientes preguntas:

* ¿Cuál es la memoria de caché de disco máxima y mínima soportada por cada NOS?

* ¿Cuál es el máximo número de archivos abiertos, clientes concurrentes, servidores en cualquier dominio y dominios soportados por cada NOS?

c) Disponibilidad y características de fiabilidad.- Un servidor debe de ser tolerante a fallas y debe tener diferentes tipos de recuperación en su configuración de hardware y software Esto es muy importante por la función crítica que realizan en las organizaciones.

d) Soporte a clientes.- Los dispositivos típicos que se deben soportar son sistemas que corren en DOS, DOS/Windows, Windows para Trabajo en Grupo, Windows 95, Windows NT, Macintosh, OS/2 y UNIX.

e) Impresión en red.- La impresión es una de las funciones primarias de los NOS. Para la elección de un NOS se debe de considerar lo siguiente:

-¿Cuántas impresoras son soportadas por el servidor?
-¿Puede una cola de impresión ser manejada por múltiples impresoras?
- ¿Pueden múltiples colas manejar una impresión?
-¿El NOS mandará un mensaje de alarma al operador si se originan problemas de impresión?
-¿El NOS informara al usuario cuando un trabajo de impresión se termina?
-¿Puede la función de impresión ser manejada remotamente?

f) Medios de transmisión.- Actualmente existen diferentes medios de transmisión para el transporte de la información. Un NOS debe ser capaz de soportar una amplia gama de medios de transmisión, y estándares de redes de área local y amplia (LAN, WAN), tales como:

* Medios de transmisión Ethernet.
* Medios de transmisión Token-Ring.
* Líneas telefónicas.
* PSDN como X.25, fibra óptica, Servicios Integrados de Red Digital (ISDN).

Sin este soporte será muy difícil construir una infraestructura óptima de red, para satisfacer las necesidades de las empresas.

Los protocolos que debe soportar un NOS son:

- Emulación de terminal asíncrona.
- AFP AppleTalk.
- TCP/IP.
- Telnet.
- SMTP (Protocolo Simple de Transferencia de Correo).
- SNMP (Protocolo Simple Manejador de Red).
- SNA (Permite comunicaciones punto a punto).
- APPC (Comunicación Avanzada Programa a Programa).
- FTP (Protocolo de Transferencia de Archivos).
- Servicios de impresión NetWare.
- NETBios (Protocolo para compartir recursos).
- NETBeui (Protocolo e interfaz para PC-IBM)

Esta categoría también considera que los clientes pueden ingresar al servidor a través de líneas asíncronas, PSDN, Internet o ISDN.

g) Servicios de red.- Esta categoría evalúa las plataformas NOS para determinar si se proveen las funciones necesarias para soportar ambientes de grandes corporaciones, incluyendo soporte para servicio de directorio que permite a los usuarios acceder a los servicios de la red sin el conocimiento de la dirección de la red.

h) Administración del servidor.- Esta categoría revisa las herramientas disponibles para manejar plataformas NOS, incluyendo funciones de revisión de pista y también las siguientes:

* Administración de archivos.
* Administración de cuentas de usuario.
* Reporte de errores.
* Reporte de desempeño del servidor.

i) Seguridad.- Las empresas necesitan sentir la confianza de que los datos están seguros. Esta categoría revisa si soporta listas de control de acceso, cuotas de disco, administración de intrusos, soporte a sistemas de administración, como Kerberos (protocolo de autenticación de redes). También revisa si están disponibles o no, los servicios de encriptación.

j) Funcionalidad/Utilidad.- Un sistema operativo de propósito único puede proveer un alto desempeño para servicios de impresión o servidor de archivos, pero pueden existir empresas u organizaciones que requieran

de otros sistemas operativos para soportar servidores con funciones múltiples y con esto se puede minimizar el costo y la administración del sistema. Una plataforma multifuncional puede ser capaz de soportar aplicaciones multiusuario de tiempo compartido así como soporte a clientes.[110]

8.4 Sistemas operativos distribuidos

Desde el inicio de la era de la computadora moderna (1945), hasta cerca de 1985, solo se conocía la computación centralizada.[111]

A partir de la mitad de la década de los ochentas aparecen dos avances tecnológicos fundamentales:

- Desarrollo de microprocesadores poderosos y económicos con arquitecturas de 8, 16, 32 y 64 bits.
- Desarrollo de redes de área local (LAN) de alta velocidad, con posibilidad de conectar cientos de máquinas a velocidades de transferencia de millones de bits por segundo (mb/seg).

Aparecen los sistemas distribuidos, en contraste con los *sistemas centralizados.* Los sistemas distribuidos necesitan un software distinto al de los sistemas centralizados.

Los sistemas operativos para sistemas distribuidos han tenido importantes desarrollos pero todavía existe un largo camino por recorrer. Los usuarios pueden acceder a una gran variedad de recursos computacionales:

- De hardware y de software.
- Distribuidos entre un gran número de sistemas computacionales conectados.

Un importante antecedente de las redes de computadoras lo constituye Arpanet, iniciada en 1968 en los EE. UU.[112]

---

[110] Instituto Tecnológico del Itsmo: “Sistemas operativos de Red”, § 2.5, en línea, disponible en: http://www.itistmo.edu.mx/Desarrollo%20de%20Proyectos/consisor.html, recuperado el 16/02/09.
[111] Véase, A. S. Tanenbaum. *Sistemas Operativos Distribuidos*. México, Prentice Hall Hispanoamericana, 1996.
[112] Efraín Acosta Arzola: “Tendencias de sistemas distribuidos”, Introducción, en línea, disponible en: http://sisefrain.blogdiario.com, recuperado el 16/02/09.

Ventajas de los sistemas distribuidos con respecto a los centralizados

Una razón para la tendencia hacia la descentralización es la economía. Herb Grosch formuló la que se llamaría "Ley de Grosch".[113]

- El poder de cómputo de una CPU es proporcional al cuadrado de su precio:
  - Si se paga el doble se obtiene el cuádruple del desempeño.
- Fue aplicable en los años setentas y ochentas a la tecnología mainframe.
- No es aplicable a la tecnología del microprocesador:
  - La solución más eficaz en cuanto a costo es limitarse a un gran número de CPU baratos reunidos en un mismo sistema.

Los sistemas distribuidos generalmente tienen en potencia una proporción precio / desempeño mucho mejor que la de un único sistema centralizado.

Algunos autores distinguen entre:

- **Sistemas distribuidos:** están diseñados para que muchos usuarios trabajen en forma conjunta.
- **Sistemas paralelos:** están diseñados para lograr la máxima rapidez en un único problema.

En general se consideran sistemas distribuidos, en sentido amplio, a los sistemas en que:

- Existen varias CPU conectadas entre sí.
- Las distintas CPU trabajan de manera conjunta.

Ciertas aplicaciones son distribuidas en forma inherente:

- Ej.: sistema de automatización de una fábrica:
  - Controla los robots y máquinas en la línea de montaje.

[113] A. S. Tanenbaum. Sistemas Operativos Distribuidos. Prentice Hall Hispanoamericana, S.A., México, 1996.

' Cada robot o máquina es controlado por su propia computadora.
' Las distintas computadoras están interconectadas.

Una ventaja potencial de un sistema distribuido es una mayor confiabilidad:

- Al distribuir la carga de trabajo en muchas máquinas, la falla de una de ellas no afectara a las demás:

' La carga de trabajo podría distribuirse.

- Si una máquina se descompone:

' Sobrevive el sistema como un todo.

Otra ventaja importante es la posibilidad del crecimiento incremental o por incrementos:

- Podrían añadirse procesadores al sistema, permitiendo un desarrollo gradual según las necesidades.
- No son necesarios grandes incrementos de potencia en breves lapsos de tiempo.
- Se puede añadir poder de cómputo en pequeños incrementos.

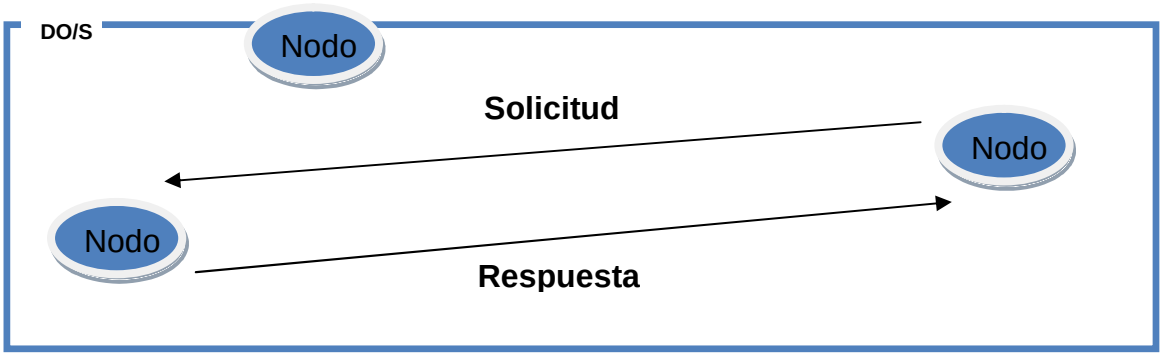


Figura 8.2 Esquema de un sistema distribuido (DO/S)[114]

[114] Flynn, Ida. Sistemas operativos. 3ª ed., México, Thomson Learning, 2001, p. 237.

En un entorno DO/S (fig. 8.2) los nodos forman parte de un sistema operativo globalmente administrado, diseñado para optimizar los recursos del sistema. El DOS/S maneja las solicitudes y operaciones entre nodos.

Desventajas de los sistemas distribuidos

El principal problema es el software, ya que el diseño, implantación y uso del software distribuido presenta numerosos inconvenientes.[115]

Los principales interrogantes son las siguientes:

- ¿Qué tipo de S. O., lenguaje de programación y aplicaciones son adecuados para estos sistemas?
- ¿Cuánto deben saber los usuarios de la distribución?
- ¿Qué tanto debe hacer el sistema y qué tanto deben hacer los usuarios?

La respuesta a estos interrogantes no es uniforme entre los especialistas, pues existe una gran diversidad de criterios y de interpretaciones al respecto.

Otro problema potencial tiene que ver con las redes de comunicaciones, ya que se deben considerar problemas debidos a pérdidas de mensajes, saturación en el tráfico, expansión, etc.

El hecho de que sea fácil compartir los datos es una ventaja pero se puede convertir en un gran problema, por lo que la seguridad debe organizarse adecuadamente.

En general se considera que las ventajas superan a las desventajas, si estas últimas se administran seriamente.[116]

---

[115] A. S. Tanenbaum. Sistemas Operativos Distribuidos. Prentice Hall Hispanoamericana, S.A., México, 1996.

[116] Universidad Nacional del Nordeste, departamento de Informática: “S.O. Introducción a los Sistemas Distribuidos”, en línea, disponible en: http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/SO7.htm, recuperado el 16/02/09.

Sistema Operativo de red (NOS)	Sistema Operativo distribuido (DO/S)
Recursos propiedad de los nodos locales	Recursos propiedad del sistema global
Recursos Locales administrados por el sistema operativo local	Recursos locales administrados por un DO/S global
Acceso ejecutado mediante un sistema operativo local	Acceso ejecutado por el DO/S
Solicitudes pasadas de un sistema operativo local a otro vía el NOS	Solicitudes pasadas directamente de un nodo a otro vía el DO/S

Cuadro 8.1. Cuadro comparativo de sistemas operativos.

8.5. Servicios remotos en Internet

> **Internet** es un conjunto descentralizado de redes de comunicación interconectadas, que utilizan la familia de protocolos TCP/IP, garantizando que las redes físicas heterogéneas que la componen funcionen como una red lógica única, de alcance mundial. [117]

De acuerdo con la definición anterior se entiende que Internet es una forma de comunicación de alcance mundial, además el desarrollo tecnológico ha impulsado que se logren grandes anchos de banda y un gran número de aplicaciones. Los sistemas operativos juegan un papel muy importante, según lo visto en los siete temas anteriores se puede afirmar que el corazón de Internet son los sistemas operativos, ya que sin ellos sería imposible poder acceder a todas las ventajas que ofrece a los cientos y aun miles de computadoras conectadas entre sí.

[117] Wikipedia: “Internet”, material en línea, disponible en: http://es.wikipedia.org/wiki/Acceso_a_internet, recuperado el 21/01/09.

La comunicación de computadoras en Internet se realiza mediante la transmisión de paquetes. Cada paquete contiene en su estructura la **dirección IP** (Protocolo Internet) de origen y destino de la computadora conectada a la Red. Cuando el paquete llega al **ruteador** (equipo de interconexión de redes) este extrae la dirección de destino utilizada para decidir por qué ruta de Internet debe enviarlo a la computadora destino. Los ruteadores contienen tablas de ruteo en las que se almacena la información sobre las direcciones IP de las computadoras conectadas a la red y se actualizan constantemente.

Servicios de red

Las redes de computadoras ofrecen servicios y procesos a los host que lo requieran, de manera general los protocolos de conectividad de redes y el tráfico de datos que soportan se clasifican en:

- Servicios orientados a la conexión.
- Servicios NO orientados a la conexión.

Servicios orientados a la conexión: Implica el uso de una trayectoria específica que se establece de manera permanente durante el tiempo que dura la conexión. Tiene tres fases: el establecimiento de la conexión, la transferencia de datos y la terminación de la conexión. Los servicios orientados a la conexión son muy útiles para la transmisión de aplicaciones que no toleran retardos y secuenciación de paquetes, **los servicios de voz y video se suelen basar en servicios orientados a la conexión.**

Servicios NO orientados a la conexión: Este servicio se basa en la selección dinámica de la trayectoria y del ancho de banda, logrando con esto que el tráfico sea ruteado y evite su paso por fallas en la red. Este tipo de servicio es muy útil en la transmisión de aplicaciones que pueden tolerar ciertos retados y re-secuenciación de paquetes. **Las aplicaciones de datos se basan en servicios no orientados a la conexión.**

Servicios más comunes en Internet

Aunque los servicios de Internet más populares son: la navegación por web, correo electrónico, chat, etc., no se limitan a ello, ya que en la realidad se concibe a Internet como un sistema distribuido con alcance mundial.

- Correo electrónico (e-mail). Utiliza el protocolo **SMTP** ("Simple Mail Transfer Protocol"), para la recepción y envío.
- Emulación de terminal **TELNET**. Se utiliza para conectar a equipos remotos mediante la red emulando un terminal del equipo al que se realiza la conexión.
- Transferencia de archivos. Utiliza el protocolo **FTP** ("File Transfer Protocol"), se usa para enviar o recibir ficheros (de cualquier tipo) entre dos equipos conectados a la red.
- Servicio de nombres de dominio **DNS** ("Domain Name Service").
- **Gopher**. Un servicio de información basado en servidores y que sirve de interfaz para otros servicios de información.
- **WAIS** ("Wide Area Information Service"). Es otro servicio de información basado en bases de datos de archivos que permiten su rápida localización.
- **finger**. Servicio de identificación de usuarios.
- La **Web**, WWW, W3. Servicio basado en **HTTP** (Hyper Text Transfer Protocol) para la navegación en Internet.
- **NFS** ("Network File System"). Sistema que permite a equipos físicamente distantes, compartir discos y directorios mediante la técnica denominada **RPC** ("Remote Procedure Call"), que hace que los recursos aparezcan como si estuvieran en el propio sistema.
- Servicios de Información de Red, **NIS** ("Network Information Services"). También basados en **RPC**, permite que varios sistemas puedan compartir una misma base de datos situada remotamente; por ejemplo, varios sistemas pueden compartir bases de datos con el mismo archivo de seguridad (password file), lo que facilita su gestión centralizada.
- Servicios "**R**". Tales como **rlogin, rsh** y otros. Utilizan la idea de acuerdos entre sistemas (hosts trusting), que permite ejecutar comandos y otras órdenes en equipos remotos sin requerir un password.[118]

[118] Zator Systems: “Servicios de Internet”, en línea, disponible en: http://www.zator.com/Internet/A8.htm, recuperado el 22-enereo-2009.

Bibliografía del tema 8

Stallings, William. *Sistemas Operativos*. 4ª ed., Madrid, Pearson Educación, 2001.

H. M. Deitel. *Introducción a los Sistemas Operativos*. México, Addison-Wesley Iberoamericana, 1987.

Flynn, Ida. *Sistemas operativos*. 3ª ed., México, Thomson Learning, 2001.

Tanenbaum, Andrew S. *Sistemas Operativos Distribuidos*. México, Prentice Hall Hispanoamericana, 1996.

Referencias electrónicas

http://es.wikipedia.org/wiki/Acceso_a_internet, 21/01/09.

http://www.zator.com/Internet/A8.htm, 22-enereo-2009.

Actividades de aprendizaje

A.8.1. Describe en un documento cuáles son las mediciones más comunes que se utilizan para evaluar el rendimiento de un sistema operativo.

A.8.2. Realiza un cuadro sinóptico donde menciones las tareas que realiza el núcleo (kernel) y sus tipos de drivers en un sistema operativo.

A.8.3. Grafica y explica el funcionamiento del esquema de un sistema distribuido (DOS/S).

Cuestionario de autoevaluación

1. ¿Para qué se utiliza la medición del rendimiento en un sistema operativo?
2. ¿Qué parámetros mide la función de rendimiento “utilización”?
3. ¿Cuáles son las funciones que realiza el núcleo (kernel)?
4. ¿Cuáles son los tres tipos de drivers que utiliza el núcleo (kernel)?
5. ¿Por qué causa se requiere la instalación de un driver?
6. ¿Cuál es la función de un sistema operativo de red (NOS)?
7. ¿Qué es Internet?
8. ¿Qué es un servicio de red orientado a la conexión?
9. ¿Qué es un servicio de red NO orientado a la conexión?
10. ¿Cuál es la función del servicio NFS (Network File System)?

Examen de autoevaluación

1. ¿Cuál es el driver de software en un sistema de archivos?
    a) ext3 o reiserfs
    b) ext2 o reiserXFS
    c) ext1 o NFS

2. ¿En dónde se ubica el driver?
    a) en el núcleo
    b) en el disco
    c) en la memoria

3. ¿Cuál es la ventaja del uso de módulos?

a) permite adicionar y remover funcionalidades del kernel

b) permite utilizar parámetros avanzados del administrador

c) permite soportar sistemas distribuidos

4. ¿Cuál es una característica que contiene un módulo, para un driver de carácter?

a) implementa llamadas al sistema “llsek”

b) implementa sistema de archivos “nfs”

c) implementa políticas de uso “rules”

5. ¿Cuál es el sistema operativo que permite que muchos usuarios trabajen en forma conjunta?

a) red

b) paralelo

c) distribuido

6. Los servicios de voz y video se basan en:

a) servicios orientados a conexión

b) servicios no orientados a conexión

c) servicios híbridos

7. ¿Cuál es una característica de un sistema distribuido?

a) varias redes conectadas entre sí

b) varias CPU conectadas entre sí

c) varios usuarios conectados entre sí

8. ¿Cuál es el servicio que realiza la identificación de usuarios?

a) WAIS

b) DNS

c) finger

9. ¿Cuál es la función que realiza un ruteador?

a) interconexión de sistemas operativos

b) interconexión de redes

c) interconexión de datos

10. ¿Los sistemas paralelos están diseñados para lograr?

a) mayor rapidez
b) mayor conectividad de nodos
c) mayor cantidad de sistemas operativos

Bibliografía básica

Carretero, Jesús. *Sistemas Operativos*. Madrid, McGraw Hill, 2001.

Silbertschatz, Abraham. *Sistemas Operativos*. 6ª ed., México. Limusa Wiley, 2002.

Stallings, William. *Sistemas Operativos*. 4ª ed., Madrid, Pearson Educación, 2001.

Tanenbaum, Andrew S., *Sistemas Operativos Modernos*. 2ª ed., México, Pearson Educación, 2003.

Flynn, Ida. *Sistemas operativos*. 3ª ed., México, Thomson Learning, 2001.

H. M. Deitel. *Introducción a los Sistemas Operativos*. México, Addison-Wesley Iberoamericana, 1987.

Sitios electrónicos

Tema 1

http://www.monografias.com/trabajos11/oper/oper.shtml, recuperado el 13/01/09

http://www.slideshare.net/softesau/sistemas-operativos-171331/, recuperado el 05/12/08.

www.itescam.edu.mx/principal/sylabus/fpdb/recursos/r2305.DOC, recuperado el 13/01/09.

http://www.dei.uc.edu.py/tai2003-2/sistemas.operativos/Tiempo%20Compartido.htm, recuperado el 08/12/08.

http://sopa.dis.ulpgc.es/so/examenes/2006/soluciones-20060429-primer_parcial.pdf, recuperado el 08/12/08.

Tema 2.

http://es.wikipedia.org/wiki/Proceso_(inform%C3%A1tica), recuperado el 4 de noviembre de 2008.

http://upload.wikimedia.org/wikipedia/commons/e/e3/Procesos-2estados.png, 13/01/09.

http://upload.wikimedia.org/wikipedia/commons/8/8b/Diagrama_de_estados5.PNG, 13/01/09.

http://www.dei.uc.edu.py/tai2003-2/sistemas.operativos/sistemas%20operativos.htm, recuperado el 13/01/09.

http://torio.unileon.es/~dielpa/asig/shannon/SO/teoria/so03.pdf, pp. 1-12, recuperado el 13/01/09.

http://www.wikilearning.com/apuntes/windows_95-el_despachador_de_procesos_conceptos/13976-3, recuperado el 13/01/09.

Tema 3.

http://wwwdi.ujaen.es/~lina/TemasSO/CONCURRENCIA/1ComunicacionySincronizacion.htm, recuperado el 13/01/09.

http://wwwdi.ujaen.es/~lina/TemasSO/CONCURRENCIA/1ComunicacionySincronizacion.htm, recuperado el 13/01/09.

http://delta.cs.cinvestav.mx/~pmejia/capi5tr.ppt, diapositiva 8/19. Recuperado el 13/01/09.

http://html.rincondelvago.com/sistemas-operativos_26.html, recuperado el 13/01/09.

http://html.rincondelvago.com/sistemas-operativos_26.html, recuperado el 13/01/09.

http://html.rincondelvago.com/sistemas-operativos_26.html, recuperado el 13/01/09.

http://ar.geocities.com/clubdealumnos/soperat/Peterson.htm, recuperado el 13/01/09.

http://es.wikipedia.org/wiki/Sistema_operativo#Interrupciones_y_excepciones, recuperado el 13/01/09.

http://www.slideshare.net/cesar2007/sistoper-bloqueos-mutuos/n, consultado el 13/01/09.

http://sisinfo.itc.mx/ITC-APIRGG/Materias/Mat4/SistOp-II_Unid2.php, recuperado el 13/01/09.

http://www.dirinfo.unsl.edu.ar/~sonet/teorias/SO-clase5-pagina.pdf, pássim. Recuperado el 13/01/09.

Tema 4.

http://books.google.com.mx/books?id=g88A4rxPH3wC&pg=RA1-PA193&lpg=RA1-PA193&dq=El+modelo+probabil%C3%ADstico+no+es+mas+que+una+aproximaci%C3%B3n,+se+basa+en+la+suposici%C3%B3n+impl%C3%ADcita+de+que+los+n+procesos+son+independientes&source=web&ots=yrVyRAgL_Q&sig=ucwTFwjEzeSgpVyggm67BKYM7Ek&hl=es&sa=X&oi=book_result&resnum=1&ct=result#PPR7,M1. Fecha de recuperación: 08 de enero de 2009.

http://torio.unileon.es/~dielpa/asig/shannon/SO/teoria/memoria.ppt Fecha de recuperación: 08 de enero de 2009.

http://fcasua.contad.unam.mx/apuntes/interiores/docs/98/2/sis_operativos.pdf, p. 27. Fecha de recuperación: 08 de enero de 2009.

http://www.mitecnologico.com/Main/AdministradorDeLaMemoria Fecha de recuperación: 08 de enero de 2009.

http://es.wikipedia.org/wiki/Asignaci%C3%B3n_din%C3%A1mica_de_memoria. Fecha de recuperación: 08 de enero de 2009.

http://iteso.mx/~jluis/sopdf/material-oto-04/11-memoria-virtual.pdf Fecha de recuperación: 08 de enero de 2009.

http://iteso.mx/~jluis/sopdf/material-oto-04/11-memoria-virtual.pdf. Fecha de recuperación: 08 de enero de 2009.

http://es.wikipedia.org/wiki/Memoria_virtual. Fecha de recuperación: 09 de enero de 2009.

http://upload.wikimedia.org/wikipedia/commons/thumb/3/32/Virtual_address_space_and_physical_address_space_relationship.svg/300px-Virtual_address_space_and_physical_address_space_relationship.svg.png. Recuperado el 08/01/09.

http://wwwdi.ujaen.es/~lina/TemasSO/MEMORIAVIRTUAL/1y2Motivaciones,ventajasyEstrategiasdeadministracion.htm Fecha de recuperación: 08 de enero de 2009.

http://es.wikipedia.org/wiki/Cach%C3%A9, recuperado el 08/01/09.

http://es.wikipedia.org/wiki/Cach%C3%A9. Fecha de recuperación: 08 de junio de 2009.

Tema 5.

http://fcasua.contad.unam.mx/apuntes/interiores/docs/98/2/sis_operativos.pdf . Fecha de recuperación: 08 de enero de 2009.

http://es.wikipedia.org/wiki/Compresi%C3%B3n_de_datos. Fecha de recuperación: 08 de enero de 2009.

Tema 6.

http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MonogSO/SEGSO00.htm, recuperado el 08/01/09.

http://fcasua.contad.unam.mx/apuntes/interiores/docs/98/2/sis_operativos.pdf, recuperado el 08/01/09.

http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MonogSO/SEG02.html#IIB, recuperado el 08/01/09.
http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MonogSO/SEG02.html#IIC, recuperado el 08/01/09.

http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MonogSO/SEG02.html. Recuperado el 08 de enero de 2009.

http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MonogSO/SEG02.html Fecha de recuperación: 09 de enero de 2009.

http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MonogSO/SEG02.html Fecha de recuperación: 09 de enero de 2009.

http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MonogSO/SEG02.html#IA, recuperado el 08/01/09.

http://es.kioskea.net/contents/attaques/attaques.php3 Fecha de recuperación: 09 de enero de 2009.

http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/SO14.htm. Fecha de recuperación: 09 de enero de 2009.

http://es.wikipedia.org/wiki/Access_violation Fecha de recuperación: 09 de enero de 2009.

http://www.ccee.edu.uy/ensenian/catcomp/material/Inform_%20II/Criterios%20evaluacion.pdf, pp. 40-41. Fecha de recuperación: 09 de enero de 2009

Tema 7

http://www.freebsd.org/doc/es_ES.ISO8859-1/books/handbook/users-superuser.html. Fecha de recuperación 09 de enero de 2009.

http://iie.fing.edu.uy/ense/asign/admunix/superusu.htm Fecha de recuperación: 09 de enero de 2009.

http://es.wikipedia.org/wiki/Administrador_de_sistemas Fecha de recuperación: 09 de enero de 2009.

http://es.wikipedia.org/wiki/CD_aut%C3%B3nomo Fecha de recuperación: 09 de enero de 2009.
http://arquitecturapcs.blogspot.com/2008_08_20_archive.html. Fecha de recuperación: 09 de enero de 2009.

http://www.scribd.com/doc/5997438/E5 Fecha de recuperación: 09 de enero de 2009.

http://es.wikipedia.org/wiki/Partici%C3%B3n_de_disco Fecha de recuperación: 09 de enero de 2009.

http://es.kioskea.net/contents/repar/filesys.php3 Fecha de recuperación: 09 de enero de 2009

http://arantxa.ii.uam.es/~siguenza/Sistemas%20operativos.ppt.rRecuperado el 12/01/09

http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/SO4.htm.
Fecha de recuperación: 09 de enero de 2009.

http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/SO0.htm
http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/SO4.htm, recuperado el 08/01/09.

http://es.kioskea.net/contents/linux/linshell.php3, recuperado el 08/01/09.

http://www.mailxmail.com/curso/informatica/linux-unix/capitulo5.htm. Fecha de recuperación: 09 de enero de 2009.

http://es.wikipedia.org/wiki/Windows_Shell Fecha de recuperación: 09 de enero de 2009.

http://es.kioskea.net/contents/winnt/ntusers.php3 Fecha de recuperación 09 de enero de 2009

http://www.linuxparatodos.net/portal/staticpages/index.php?page=02-cuentas-usuario&mode=print, recuperado el 13/01/09

http://es.kioskea.net/contents/unix/unix-users.php3 Fecha de recuperación: 09 de enero de 2009.

Respuestas a los exámenes de autoevaluación

SISTEMAS OPERATIVOS MULTIUSUARIOS

	Tema 1	**Tema 2**	**Tema 3**	**Tema 4**	**Tema 5**	**Tema 6**	**Tema 7**	**Tema 8**
1.	a	c	b	b	b	c	b	a
2.	c	c	b	b	a	b	b	a
3.	b	a	b	a	a	c	b	a
4.	b	c	a	a	a	b	b	a
5.	a	a	c	a	a	b	c	c
6.	c	a	b	b	a	b	b	a
7.	b	b	c	a	a	b	c	b
8.	b	a	c	b	a	b	a	c
9.	c	b	b	b	b	b	c	b
10.	b	b	a	c	a	b	b	a