

Escuela de Ingeniería en Computación

Principio de Sistemas Operativos

**Programa sincronizador de archivos en C**

**Estudiante:**

José Mario Naranjo Leiva

Gabriel Ramírez Ramírez

**Carnet:**

2013034348

2010

**Profesor:**

Armando Arce Orozco

Primer Semestre, 2017

**Tabla de contenido**

[**Tabla de contenido** 2](#_Toc485654271)

[**Tabla de ilustraciones** 2](#_Toc485654272)

[1. Introducción 3](#_Toc485654273)

[2. Descripción del problema 3](#_Toc485654274)

[2.1. Programa Sincronizador 3](#_Toc485654275)

[3. Definición de estructuras de datos 4](#_Toc485654276)

[4. Descripción detallada 5](#_Toc485654277)

[4.1. Manejo de sockets 5](#_Toc485654278)

[4.2. Manejo de solicitudes 5](#_Toc485654279)

[4.3. Rutina de comparación 5](#_Toc485654280)

[4.3.1. Identificación de archivos 6](#_Toc485654281)

[4.3.2. Comparación 6](#_Toc485654282)

[4.4. Manejo de solicitudes 7](#_Toc485654283)

[5. Descripción del protocolos y formatos 7](#_Toc485654284)

[6. Análisis de resultados de pruebas 7](#_Toc485654285)

[6.1. Definición de pruebas 7](#_Toc485654286)

[7. Conclusiones sobre rendimiento y funcionamiento 7](#_Toc485654287)

**Tabla de ilustraciones**

[Ilustración 1: Estructura para almacenar los datos del archivo 4](#_Toc485652520)

[Ilustración 2: Estructura para un arreglo dinámico de archivos 4](#_Toc485652521)

[Ilustración 3 Vista general de la estructura 5](#_Toc485652522)

[Ilustración 4 - Definición de la función para el escaneo del directorio 6](#_Toc485652523)

[Ilustración 5 - Definición de función de comparación 7](#_Toc485652524)

[Ilustración 6 - Definición de función de comparación (modificados) 7](#_Toc485652525)

1. Introducción
2. Descripción del problema

Un problema que presentan los diferentes sistemas de sincronización de archivos (como Dropbox, LiveDrive y SugarSync) es que solo funcionan conectados al Internet. Desde antes existían aplicaciones sencillas que permitían sincronizar máquinas en una red sin conectarse a la Web (por ejemplo, una laptop y una computadora de escritorio).

* 1. Programa Sincronizador

El programa sincronizador se debe ejecutar en dos máquinas al mismo tiempo. En la primera máquina el programa recibirá como parámetro únicamente el nombre del directorio a sincronizar y quedará esperando conexiones en el puerto 8889. En la otra máquina se debe pasar como parámetros el nombre del directorio local a sincronizar y el IP de la primera máquina.

Al iniciarse el programa deberá identificar los archivos en su directorio local, almacenando su tamaño, hora y fecha de la última actualización. Debe comparar dicha información con un listado de su corrida anterior, y determinar cuáles archivos han cambiado. Si algún archivo ha cambiado, se eliminó, o se creó, se debe enviar esa información a la otra máquina para que realice los cambios adecuados. Dentro de dichos cambios está solicitar el archivo actualizado a la otra máquina.

* Una vez que se realiza la sincronización, los programas que corren en las máquinas deben terminar. La siguiente vez que se desee sincronizar se debe volver a ejecutar el programa en ambas máquinas.
* Tome en cuenta que los relojes de las máquinas involucradas pueden estar de sincronizados. Para determinar cuáles archivos son más recientes se debe realizar un cálculo de diferencias entre las horas de ambos relojes.
* NO se deben escribir dos programas diferentes. Es el mismo programa que actúa como cliente o servidor dependiendo de los parámetros que se le pasen.
* Debe utilizar sockets para que se comuniquen los programas. Además, debe definir un formato y protocolo adecuado para enviar y recibir las solicitudes de archivos.
* Debe tomar en cuenta la posibilidad que un archivo haya sido modificado en ambas máquinas, provocando el caso de “copias en conflicto”. Aquí lo mejor es dejar las dos copias, pero con nombres ligeramente diferentes.

1. Definición de estructuras de datos

Para la implementación adecuada de este programa sincronizador, se implementaron las siguientes estructuras de datos.

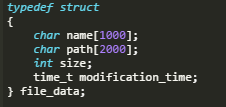


Ilustración 1: Estructura para almacenar los datos del archivo

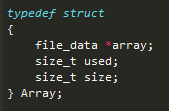


Ilustración 2: Estructura para un arreglo dinámico de archivos

* **Lista**

Para manejar los archivos contenidos en el directorio, se implementó una lista dinámica enlazada, para almacenar la estructura *file\_data* (ver Ilustración 1) que contiene los datos relevantes de un archivo: nombre, carpeta, tamaño y fecha de modificación.

El nodo de la lista enlazada está definido en la estructura *Array* (ver Ilustración 2). Un diagrama general de esta estructura está en Ilustración 3.

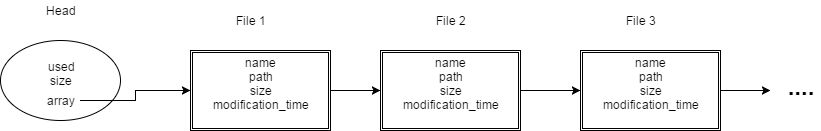


Ilustración Vista general de la estructura

* **Estructuras de sincronización**

Las estructuras de sincronización son definidas para el uso adecuado de la comunicación de las computadoras y la sincronización de los archivos. Estas se ven con más detalle en la sección

1. Descripción detallada

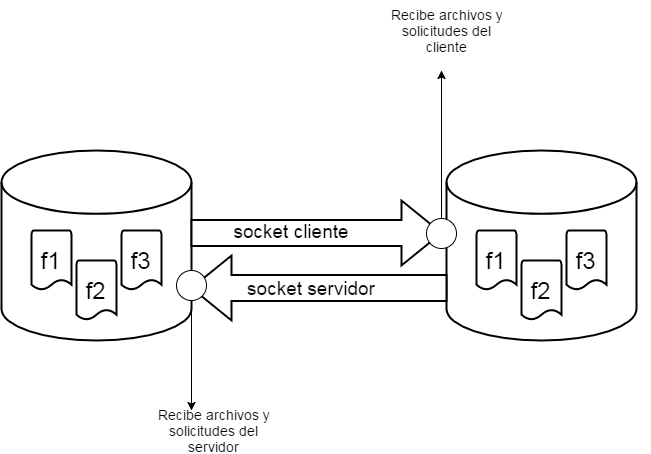
A continuación, se muestra la explicación de los componentes principales del servidor y el cliente del sistema sincronizador.

* 1. Manejo de sockets

Los sockets son parte importante para la implementación del proyecto. Los sockets Berkeley son usados para la comunicación de ambos programas.

Estos sockets son un API para el uso de sockets de internet y de dominio UNIX, usados para la intercomunicación de procesos. Por sus orígenes, estos usan el estándar de componentes POSIX.

Para la implementación del proyecto, se decidió el uso de dos sockets principales: el socket cliente y el socket servidor; entre ambos se realiza la comunicación de ambas máquinas.



Computadora A

Computadora B

Ilustración - Diagrama comunicación de sockets

Los programas ejecutándose en la computadora A y B tiene instancias de los sockets servidor y cliente para la comunicación entre ambas computadoras. Claramente, la computadora que esté sirviendo como cliente usará el socket cliente y lo mismo para el servidor. La vista general de los programas con los sockets se puede ver en Ilustración 4.

Los sockets son manejados durante la ejecución por medio que variables (variables identificador del socket). En el rol de servidor, el identificador de socket servidor es usando para las operaciones de envío de paquetes al cliente, y el socket del cliente, se usará para la lectura o recibimiento de paquetes del cliente. En el rol cliente, ocurre el modo contrario de los casos.

* 1. Manejo de solicitudes
  2. Rutina de comparación

A continuación, se explica el algoritmo de comparación implementado para determinar las diferencias en los archivos anteriores y actuales del directorio por sincronizar.

### Identificación de archivos

Para la lectura de los archivos del directorio se usa la funcionalidad “*scandir*”, definida en el encabezado **#include <dirent.h>**. Esta permite escanear el directorio y cada archivo es almacenado en una estructura dinámica, para el procesamiento posterior de los archivos.

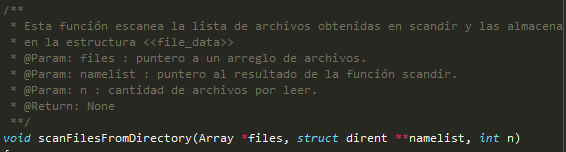


Ilustración - Definición de la función para el escaneo del directorio

Luego, se usa la función (ver Ilustración 5) para obtener los datos de los archivos del directorio y almacenarlos en una lista dinámica. Posteriormente, esta lista de archivos actuales es almacenada en un fichero en el directorio <<*.meta>>*, con el objetivo de realizar la comparación con los archivos locales actuales y los archivos de la corrida anterior.

### Comparación

El algoritmo de comparación implementado (ver Ilustración 6) está basado en la diferencia de conjuntos; así, siendo el listado actual y el listado anterior los conjuntos a cuáles se les van comparar.

La comparación se realiza sobre el listado del directorio anterior contra el listado del directorio actual, para determinar los archivos eliminados y agregados. El algoritmo genera un listado con los ficheros agregados y otro listado para ficheros eliminados.

No obstante; existe una extensión del algoritmo propio (ver Ilustración 7) para determinar los archivos modificados, esto porque se necesitan realizar una comparación especial de fechas, que no se necesitan para archivos eliminados o agregados.

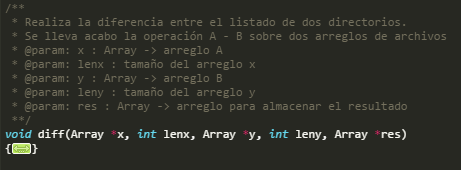


Ilustración - Definición de función de comparación

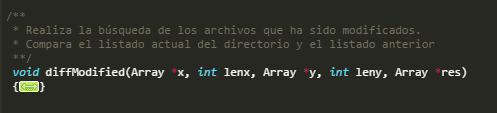


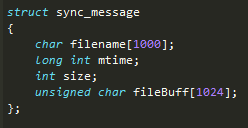
Ilustración - Definición de función de comparación (modificados)

El algoritmo funciona de la siguiente manera:

* Siendo A = lista de archivos anterior y B = lista de archivos actuales, para un directorio dado.
* La función diff mencionada anteriormente, obtiene la diferencia de los conjuntos A y B.
* Cuando se invoca la función *diff (A, B, C)*, (A – B = C), se obtiene en C los archivos eliminados en el directorio.
* Cuando se invoca la función *diff (B, A, C)*, (B – a = C), se obtiene en C los archivos agregados al directorio.
  1. Mecanismo de bloqueo

1. Descripción del protocolos y formatos

A continuación, se explican el formato y protocolo definidos para el envió y recibimiento de solicitudes de archivos.



1. Análisis de resultados de pruebas

## Definición de pruebas

A continuación, se presenta la definición de pruebas para el sistema implementado.

1. Conclusiones sobre rendimiento y funcionamiento