ANEXO F
IMPLEMENTACION DE UN SERVICIO DE DISCO VIRTUAL SEGURO

1. DEFINICIÓN DEL PROBLEMA

Otro servicio considerado como esencial dentro de las necesidades de comunicación y trabajo en red identificadas en el grupo objetivo, fue el servicio de Disco Virtual, también conocido como Disco Web (en los casos en los que el acceso al servicio se haga a través de una aplicación Web).

Gran parte de la actividad académica de los miembros de RUDECOLOMBIA, requiere del movimiento de grandes volumenes de papel por medios físicos. Citando como ejemplo el caso de los seminarios dictados en el programa de Doctorado en Ciencias de la Educación, cada profesor encargado de dictar uno, debe hacerle llegar a los estudiantes el material de referencia del mismo, que será estudiado antes del comienzo de las actividaes, y que por lo tanto debe estar a su disposición un par de semanas antes de que éstas se inicien.

Este material se despacha por correo físico, y con frecuencia se tienen problemas con este medio, bien sea porque los documentos tardan mucho tiempo en arribar a su destino, o porque simplemente no llegan. Aquí se ve una oportunidad de aprovechar las ventajas que ofrecen las redes de transmisión de datos, para difundir la información y permitir un acceso casi instantáneo a la misma. La posibilidad de crear, editar y guardar documentos en formato digital permite la duplicación de lo escrito sin perder calidad, la presentación de contenidos en diferentes formatos, la edición hipertextual[1], el almacenamiento de un gran volumen de información en un espacio reducido y la facilidad de buscar algún dato en esa gran cantidad de información.

Por otra parte, se identificaron necesidades similares, en las que los estudiantes debían hacerle llegar de una forma u otra, documentos de su elaboración tanto a compañeros, como profesores y tutores. Esta actividad, conocida con el nombre de “socialización del trabajo adelantado”, tiene lugar a lo largo de varias instancias del desarrollo del programa de doctorado, en especial, cuando se terminan seminarios y deben darse un tiempo para conceptuar sobre el conocimiento adquirido, las ideas e inquietudes que les haya dejado la experiencia, y cómo articularlas o hacerlas a un lado de su proyecto de Investigación.

1 La edición hipertextual tiene que ver con la posibilidad de escribir, borrar y mover cualquier texto, escribir las ideas en desorden y luego organizarlos, iniciar diferentes documentos y avanzar en ellos en la medida en que se generan las ideas.

En otro contexto, se notó además, la tendencia de ciertos miembros de la Red de Investigación Educativa a intercambiar información en línea, que a menudo podría ser de interés general para la comunidad, y para cuya difusión, se utilizó el sistema de envío de mensajes de correo con un archivo adjunto, a un gran volumen de destinatarios.

Así, se consideró pertinente aprovechar estas y otras formas de interacción, para potenciar el fortalecimiento de vínculos que permitieran trabajar en grupo y compartir información a través de medios electrónicos.

Dado que el conjunto de necesidades descrito se podía soportar perfectamente, a través de un sistema de gestión de archivos que permitiera, subir, descargar, modificar, cambiar el nombre y compartir con otros de una manera controlada, información disponible en diversos formatos, se decidió dotar a la Red de Investigación Educativa con el servicio de Disco Virtual.

El servicio consiste entonces en un repositorio de archivos y directorios alojados en un servidor, que pueden ser administrados desde cualquier computador con conexión a Internet a través de un navegador y una aplicación Web que lo permita.

Este servicio le permitiría a cada miembro, además de intercambiar documentos y referencias con personas de la comundad de la Red de Investigación Educativa, y de tener un espacio privado en el cual almacenar información, publicar información, documentos y archivos en general, susceptibles de ser conocidos por cualquier persona a través de Internet. En este punto, convenía contar con un espacio que fuese visible sólo por los miembros de la Red de Investigación Educativa, y otro que pudise ser asequible para cualquier persona en Internet.

El motivo por el cual se optó por este esquema, fue esencialmente el reconocimiento de la necesidad de diferenciar dominios virtuales privados y personales, de aquellos que se comparten con otras personas, e incluso con toda la Red, cuando se trabaja en un medio como Internet.

A continuación se explicita cada uno de estos “espacios” existentes en el Disco Virtual:

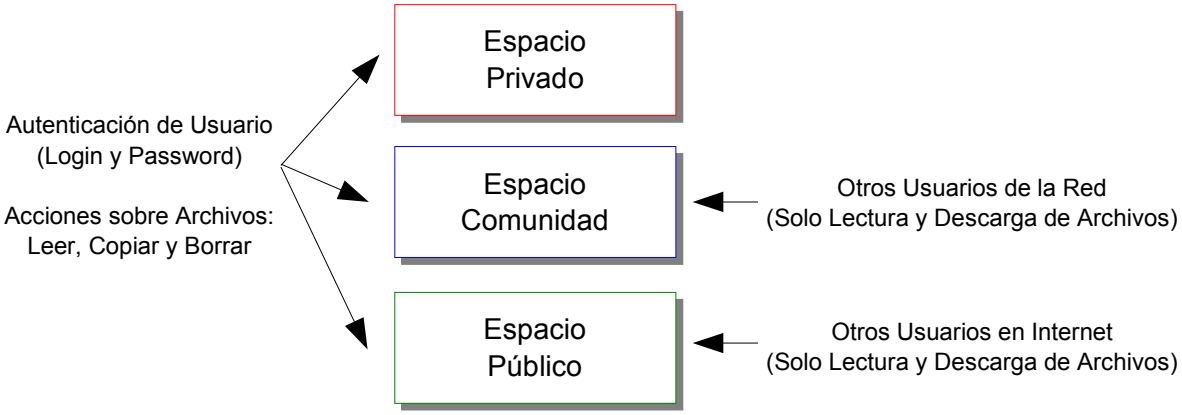
- En el Espacio Privado, los archivos o directorios que ahí se ubiquen, sólo podrán ser accedidos por su propietario, previa autenticación ante el sistema con su nombre de usuario contraseña.

- El Espacio de Comunidad es un directorio en donde los usuarios pueden ubicar documentos que sólo estarán disponibles para las personas de la Red que tengan acceso al servicio de Disco Virtual, lo que implica que si una persona no

tiene un nombre de usuario y una contraseña en el servidor, no podrá ver los documentos en este espacio.

- Finalmente, el Espacio Publico es un directorio en donde los usuarios pueden ubicar documentos (y otra clase de archivos) que son accesibles desde la Web. Con ello, se elimina la restricción de proporcionar un nombre de usuario y una contraseña para acceder a la información, y se le permite a cualquier persona que navegue por Internet, descargar archivos ubicados en este espacio.

Figura 1: Espacios disponibles en el Servicio de Disco Virtual implementado



El servicio de Disco Virtual, le ofrece a los usuarios inscritos en la Red de Investigación Educativa, una capacidad de almacenamiento de información de 50MB, compartidos con el Correo Electrónico.

Al igual que con el correo electrónico, una vez se identificaron el conjunto de necesidades que sugerían la implementación de un servicio de Disco Virtual, para la Red de Investigación Educativa, se examinaron un conjunto de alternativas para su prestación, y se determinó que era necesario garantizar el acceso a un servicio de Disco Virtual Seguro.

En este contexto, se entendió por “seguro”, un servicio que ofreciera un conjunto mínimo de características que tuviera en consideración las primitivas de la seguridad informática mencionadas en el anexo E: Aceptación, Identificación, Autenticación, Autorización, Confidencialidad e Integridad.

En este orden de Ideas, dentro de las consideraciones que los autores tuvieran en cuenta a la hora de implementar el servicio de Disco Virtual, debían figurar:

- Identificación y Autenticación de los usuarios, para permitir el acceso al Disco Virtual sólo a los usuarios que tengan una cuenta válida en el sistema.

- Cifrado de las conexiones entre el cliente y el servidor para evitar que la información de identificación y autenticación que intercambiaran, se transmitiera de forma insegura, es decir, sin el soporte de una capa de transporte que evitara la interceptación, alteración o interrupción de la comunicación y su contenido.

- Definición de políticas de acceso para separar los espacios privados de los espacios compartidos con otros usuarios o los disponibles a través de Internet, para evitar suplantaciones en la publicación de contenidos y el acceso a información que no se deseara hacer pública.

- Definición de esquemas de Autorización, para el control del acceso que cada usuario tiene al espacio compartido con otros, de tal forma que no se pueda modificar ni eliminar la información puesta por otra persona.

- Cifrado del contenido de la información que se intercambiara entre el cliente y el servidor, para garantizar su integridad y confidencialidad, en los casos en que no se desease hacer público su contenido a otras personas (espacio privado del Disco Virtual).

Con estos requisitos en mente, los autores se dieron a la tarea de buscar alternativas que satisfacieran las necesidades de intercambio de información y le permitieran a los miembros de la Red de Investigación Educativa, contar con un repositorio virtual de la misma, que permitiera administrarla de la forma más segura posible.

En este punto, conviene recordar que “un sistema es tan seguró como su eslabón más débil”. En este sentido, aparte de ofrecerle el acceso a un servicio implementado siguiendo un conjunto de recomendaciones mínimas de seguridad, se debe usar a los usuarios en procedimientos básicos tales como la utilización de contraseñas largas, con caracteres alfanuméricos y especiales, el uso de contraseñas diferentes para cada cuenta de la que dispongan en Internet (de correo electrónico, mensajería instantánea, o cualquier otro servicio), y el acceso a servicios que soporten una capa segura de transporte, si se ofrece tanto esta alternativa como la que permite un acceso inseguro.

2. CONCEPTOS BÁSICOS

2.1 El Servicio de Transferencia de Archivos

El Servicio de Transferencia de Archivos ha sido desde los inicios de Internet, uno de los más utilizados, al lado del correo electrónico y la World Wide Web.

Al igual que la mayoría de los servicios que conocemos, se basa en una arquitectura cliente/servidor, en la que el cliente se conecta a una máquina remota a través de un programa, y puede ejecutar comandos sobre ella, que van desde subir y descargar archivos, hasta listar elementos, renombrarlos, eliminarlos, etc.

Este servicio le permite a las personas intercambiar intercambien información entre sí, facilitando que se compartan recursos documentales, imágenes, o cualquier otro tipo de datos, y se puede implementar de muchas maneras. La implementación más antigua y aún ampliamente utilizada, ha fue a través del protocolo FTP.

2.2 Transferencia de archivos con FTP

Antes del adviniento de HTTP, el protocolo de facto para intercambiar archivos era FTP (*File Transfer Protocol*, Protocolo de Transferencia de Archivos), que pese a seguir siendo ampliamente utilizado, ha perdido participación frente a HTTP en ésta área.

El protocolo FTP se describe en el RFC 959 como un protocolo que busca: 1) Promover que se compartan archivos (programas de computador o datos), 2) fomentar la utilización indirecta o implícita (a través de programas) de computadores remotos, 3) Proteger al usuario de variaciones en los sistemas de almacenamiento de archivos entre los computadores, y 4) Transferir datos de manera confiable y eficiente. A pesar de ser poder ser utilizado directamente por usuarios en una terminal, FTP está diseñado especialmente para ser usado por programas.

Usualmente, un programa cliente de FTP, muestra en su interfaz, una visión que corresponde a los archivos existentes en la máquina local por un lado, y el listado de directorios y archivos en el servidor a los que se puede acceder, en el otro. Dependiendo del esquema de control de acceso que se haya definido, el usuario podrá a través del cliente, subir archivos al servidor, descargarlos, ingresar a carpetas en el equipo remoto, renombrar archivos, etc.

A pesar de su eficacia como protocolo para la transferencia de archivos, FTP posee deficiencias elementales en materia de seguridad, que se han tratado de

subsanar a través de la publicación de nuevas especificaciones, la inclusión de extensiones de seguridad y la utilización de nuevos protocolos.

2.3 Transferencia de archivos con HTTP

Si bien la transferencia de archivos a través de redes de datos, precedió a la presentación de contenidos a través de páginas Web utilizando HTTP (*Hiper Text Transfer Protocol*, Protocolo de Transferencia de Hipertexto), hoy en día gran parte del flujo de archivos existente en Internet utiliza este protocolo.

La tendencia a llevar los servicios más utilizados a la World Wide Web, ha llevado a que hoy en día la mayoría de los navegadores soporten navegación en sitios FTP. Si bien la tendencia a poner archivos para su descarga a través de HTTP en lugar de FTP es cada vez mayor, por defecto aquel no permite el uso del navegador como interfaz para subir archivos a un servidor. Para esto se debe recurrir a programas que operen a través de una interfaz Web que incluya esta funcionalidad. Como ejemplo se tienen “los administradores de archivos” de algunas aplicaciones groupware, donde empleando una interfaz Web, se le permite a los usuarios subir y modificar archivos de su cuenta en el servidor donde se encuentra registrado.

De esta forma tenemos que, pese a la compatibilidad de los navegadores modernos tanto con el protocolo HTTP como con FTP, el primero es cada vez más utilizado para realizar descargas, pero es limitado en funcionalidades, para cuya extensión se necesita recurrir a al menos un aplicación Web que cumpla esta función.

2.4 Transferencia de archivos con SFTP

SFTP (*Secure File Transfer Program*, Programa para la transferencia segura de archivos) es un programa interactivo para la transferencia de archivos, muy similar a la mayoría de los clientes conocidos de FTP, que se diferencia en que efectúa todas las operaciones a través de una capa de transporte cifrada, utilizando SSH (*Secure Shell*, Consola de Comandos Segura). SFTP puede hacer uso de varias de las características de SSH, tales como autenticación de llave pública y compresión.

Existe otra acepción de la sigla SFTP, que corresponde a: *Simple File Transfer Protocol* (Protocolo Simple de Transferencia de Archivos). Este protocolo, descrito en el RFC 913, fue diseñado para satisfacer las necesidades de personas que buscaran un protocolo más funcional que TFTP (*Trivial File Transfer Protocol*, Protocolo Trivial de Transferencia de Archivos), pero más sencillo de implementar y menos potente que FTP. SFTP soporta control de acceso de usuarios,

transferencias de archivos, listado de directorios, cambios de directorio, renombrado de archivos y eliminación de archivos.

No obstante, en el resto del documento se hará referencia a la primera acepción vista de SFTP cuando aparezca en el texto.

3. ALTERNATIVAS DE SOLUCIÓN CONSIDERADAS Y JUSTIFICACIÓN DE LAS SOLUCIÓN ESCOGIDA

3.1 Disco Virtual utilizando FTP

La implementación del disco virtual utilizando FTP, requería la instalación y configuración de este servicio en una máquina, especificando políticas de acceso bajo el esquema usuario/contraseña. Se debía especificar muy bien la clase de permisos que cada usuario podría tener sobre directorios cuya propiedad compartiera con otros, y se debería permitir el acceso anónimo al servicio, para el caso de los documentos que se desearan hacer públicos para toda la Red.

Implementar un servicio de esta forma, a parte de los riesgos de seguridad que implicaría, obligaría a los usuarios a utilizar programas para el acceso a servicios de FTP, y con ello, se limitaría el número de computadores desde los cuales podría acceder sin necesidad de la instalación de software adicional. A diferencia de los navegadores Web, es raro encontrar clientes para FTP en la mayoría de los equipos de cómputo.

Volviendo al tema de la seguridad, dentro de los principales problemas que encierra FTP (Anónimo 2000), figuran la utilización de autenticación estándar de nombres de usuario y contraseñas, con lo que el servidor no puede determinar de manera fidedigna si un determinado usuario es quien dice ser, que las contraseñas se transmiten en texto plano sin formato, y que las sesiones no están cifradas, y por lo tanto carecen de seguridad.

Existen varias formas de implementar un servicio de transferencia de archivo, que oscilan entre la utilización de diferentes protocolos, y el diseño del servicio para la utilizacion de determinados clientes. Este conjunto de alternativas son las que examinaremos más adelante.

3.2 Disco Virtual utilizando SSH

SSH se refiere tanto al servicio como al protocolo que permite el acceso remoto a un equipo, y la ejecución de comandos a través del uso de una consola. Fue concebido como un reemplazo para telnet, rlogin y rsh, y proporciona

comunicaciones cifradas en forma segua, entre dos equipos sin relaciones de confianza, sobre una red insegura.

Para implementar este servicio se debe instalar y configurar un servidor SSH, e instalar y configurar clientes SSH en los equipos de los usuarios para permitirles el acceso a aquel. Los clientes SSH son muy comunes en equipos corriendo el sistema operativo Unix/Linux y sus afines, pero existen también clientes disponibles en versiones libres, gratuitas y comerciales, para Microsoft Windows y Mac OS. La última versión liberada del protocolo, se denominó SSH2. En la actualidad, la IETF (Internet Engineering Task Force, Grupo de Trabajo en Ingeniería de Internet)[2] está trabajando en la estandarización del protocolo, a través del grupo "secsh".

En sistemas GNU/Linux, la aplicación más utilizada para implementar este servicio es OpenSSH, una versión libre del protocolo SSH disponible bajo la licencia BSD, que en la actualidad soporta las versiones 1.3, 1.5, y 2.0 del protocolo SSH. Existe también una implementación del proyeco GNU para SSH, denominada lsh.

El paquete de OpenSSH incluye el programa ssh (cliente SSH), scp (Secure Copy de Secure Shell, que permite copiar archivos entre equipos), sftp (una implementación segura de ftp), y otra utilidades como ssh-add (que permite registrar llaves nuevas al agente de autenticación ssh-agent), ssh-agent (que permite autenticar equipos usando RSA[3]), sshd (el servicio SSH, que por defecto escucha en el puerto 22), ssh-keysign (para acceder a las llaves locales para generar firmas digitales en la autenticacón de equipos remotos a través del protocolo SSH2), ssh-keyscan (que permite recuperar las llaves públicas de los equipos en una red), ssh-keygen (utilizado generar y gestionar las llaves) y sftp-server (que se ocupa del lado del servidor, de las peticiones de transferencia de archivos efectuadas por clienets sftp).

SSH soporta varios algoritmos entre los que se incluyen (Anónimo 2000): Blowfish[4], Triple DES[5], IDEA[6] y RSA.

2 La IETF es una organización abierta, de voluntarios sin membresía formal o requerimientos especiales de pertenencia, que se encarga de desarrollar y promover estándares para Internet.

3 RSA, es el nombre de un algoritmo que utiliza criptografía asimétrica. Su nombre se debe a las iniciales de sus creadores: Rivest, Shamir y Adelman.

4 Blowfish es un esquema de cifrado de la información de llave simétrica y secreta. Utiliza un tamaño de bloque de 64 bits y es uno de los bloques de cifrado más rápidos existentes.

5 Triple DES (Data Encryption Standard, Estándar de Datos), también conocido como 3 DES, es un esquema de cifrado e bloque formado a partir de DESdesarrollado por IBM. Tiene una longitud de clave de 165 bits (tres claves DES de 56 bits), pero el tamaño efectivo de llave es de 112 bits.

La implementación de la transferencia de archivos con SSH para el grupo, requería que las personas aprendieran a instalar y configurar uno cualquiera de los programas disponibles en Internet para administrar su cuenta de disco virtual en el servidor. Si bien, en una primera implementación, este fue el esquema que se siguió, muchos usuarios experimentaron dificultades en ello, y se trataron de buscar alternativas para facilitar la utilización de este servicio.

Como permitirle el acceso por consola a un gran conjunto de usuarios, puede ocasionar problemas graves de seguridad, el acceso al servidor se hacía únicamente a través del programa de transferencia de archivos (el cliente sftp o el SSH file tranfer program, programa SSH para la transferencia de archivos[7]), deshabilitando el acceso directo por consola. Ninguna de las persona que dispusiera de una cuenta de correo electrónico en el sistema, podría acceder en forma remota y ejecutar comandos en el servidor.

La implementación del servicio de Disco Virtual con SSH, protege tanto el intercambio de la información de identificación del usuario y su contraseña, como los archivos que transmita y reciba desde el servidor.

3.3 Disco Virtual a través de una aplicación Web utilizando HTTP

Tal como se explicó con anterioridad, el protocolo HTTP es cada vez más utilizado en la publicación de información, susceptible de ser descargada desde cualquier ubicación en Internet, sin necesidad de contar con algo distinto a un navegador Web.

Debido a las características exigidas en la implementación que se haría del Servicio de Disco Virtual, y puntualmente, el requerimiento de los tres espacios (privado, comunitario y público) para las cuentas de los usuarios, se vió la necesidad de buscar aplicaciones Web que en lo posible, soportaran no sólo la transferencia de archivos hacia el servidor, sino que permitieran definir estos modos de acceso de una manera eficiente e intuitiva para el usuario.

En la exploración que se hizo de alternativas para la implementación de este servicio a través de la Web, se encontró que en su mayoría, las soluciones sugerían la utilización de un componente de gestión de archivos, disponible en varias aplicaciones groupware.

6 IDEA (International Data Encryption Algorithm, Algoritmo de Encripción de Datos Internacional), es un eficaz algoritmo de cifrado de bloques que funciona con una clave de 128 bits. Idea cifra los datos de forma más rápida que 3 DES y es mucho más seguro.

7 Cliente más conocido y utilizado de SSH para entornos Windows. Su uso es gratuito para fines no comerciales.

Conociendo la experiencia de algunas tentativas encaminadas a soportar el trabajo y el intercambio de información de un conjunto de personas, utilizando gropware, se deshechó esta posibilidad, y se buscaron otras alternativas para la implementación del Disco Virtual. La razón para ello, se encuentra en la complejidad que encierra el trabajo de llevar a la aplicación, el conjunto de procedimientos identificados entre el grupo objetivo, pues a menudo, estos programas poseen funcionalidades que las personas no necesitan, y que o no se utilizan o pueden llegar a relentizar el trabajo. Para usar groupware de manera efectiva, se requiere un acuerdo de entendimiento previo sobre cada posibilidad que ofrece el sistema, y para efectos de los que se quería lograr con la Red de Investigación Educativa, esto era difícil de lograr en el corto plazo, y no era realmente lo que se buscaba en esta primera experiencia de acercamiento tecnológico y trabajo en red.

Por otra parte, preocupaba el hecho de contar con dos aplicaciones Web independientes, pues esto podría confundir a los usuarios sobre la utilidad final de cada una, y cuáles eran las condiciones en que convendría más hacer uso de un servicio que del otro (después de todo, al adjuntar archivos, resulta sencillo compartir información con otras personas).

Con esto en mente, se inició una segunda búsqueda de clientes Web para el correo virtual, pero esta vez, procurando que se pudiesen integrar con facilidad con el cliente Web de Correo Electrónico implementado.

Dentro de las alternativas consideradas, se escogió OpenWebmail, que contaba con un servicio de Disco Virtual plenamente integrado en la interfaz, y ofrecía una gran compatibilidad con el correo electrónico (existe por ejemplo, la posibilidad de adjuntar archivos alojados en el Disco Virtual), además de las razones expuestas en el Anexo E.

4. IMPLEMENTACIÓN DE LA SOLUCIÓN

Como se ha mencionado en dos ocasiones, el Servicio de Disco Virtual (Web) implementado para la Red de Investigación Educativa, se diseñó pensando en diferenciar tres espacios complementarios entre sí, a través de los cuales el usuario pudiera movilizarse dependiendo de lo que quisiera hacer con los archivos que fuera a subir a su cuenta. A estos espacios se les denominó: espacio privado, espacio comunitario y espacio público.

Para llevar estos requerimientos a la práctica, se definieron un conjunto de políticas en la creación de los usuarios del sistema, de forma tal, que cada uno contara con un espacio de cuenta en el cual almacenar información, pero también tuviese otro que fuese visible por los demás usuarios, y a la vez, todos contaran con uno que fuese accesible desde Internet.

Lo que se hizo entonces, fue aprovechar las características de la arquitectura de los sistemas operativos GNU/Linux y su Control de Acceso Discrecional DAC (Discretional Access Control), que permiten controlar el grado de dominio sobre archivos y directorios, que pueden tener tantos los usuarios, como las aplicaciones del sistema; y se definieron un conjunto apropiado de permisos[8] para cada usuario, que habrían de permitirle acceso exclusivo a su espacio privado, acceso de lectura al espacio comunitario de los demás, y lectura universal de los directorios de cada usuario listados bajo la carpeta de documentos públicos, a quienes los accedieran desde Internet.

En suma se crearon tres directorios: doc-comunidad y doc-publicos, con subcarpetas en cada uno, que correspondían a los usuarios registrados en la Red de Investigación Educativa y que podrían alojar información ahí, y una carpeta denominada doc-privados, que se ubicaría dentro de la carpeta "home" de cada usuario y almacenaría los documentos y archivos personales de cada uno.

Una vez definida la estructura, sólo restaba instalar y configurar al cliente Web de correo electrónico para que permitiera visualizar y gestionar los archivos del usuario, y se logró sin mayores dificultades.

Por otra parte, aprovechando el soporte para la implementación de Alojamiento Virtual (Virtual Hosts)[9] del servidor Web Apache, se utilizaron enlaces simbólicos para permitir la lectura y la presentación del contenido del espacio público de cada usuario en Internet, empleando para ello simplemente un navegador.

5. RECOMENDACIONES Y TRABAJO FUTURO

Para evitar posibles abusos en la capacidad de almacenamiento de información permitida a cada usuario (50MB compartidos con el correo electrónico), se recomienda fuertemente el establecimiento de un mecanismo de control de quotas.

Dada la poca formación que pueden tener algunos usuarios en la utilización de ciertos servicios telemáticos, se recomienda explicarles las ventajas que encierra el servicio de Disco Virtual, en comparación a otras alternativas para compartir documentos con otras personas, como por ejemplo, adjuntado archivos a los mensajes.

8 Aquí hablamos de tres tipos de permisos: lectura, escritura y ejecución. Se debe tener en cuenta que para ejecutar un archivo, también se necesita tener permiso de leerlo. En el caso de los directorios, el permiso de lectura permite listar los contenidos del directorio, mientras el de ejecución permite entrar en él.

9 El Alojamiento Virtual es un método empleado por los Servidores Web para alojar más de un nombre de dominio en una misma máquina y con la misma dirección IP.