

WICC 2014 XVI Workshop de Investigadores en Ciencias de la Computación

Arquitectura para Acceso Remoto

Karina Cenci^{1,2} Leonardo de- Matteis^{1,2} Jorge R. Ardenghi¹ Pablo Fillottrani^{2,3}

¹Laboratorio de Investigación en Sistemas Distribuidos

²Laboratorio de Investigación en Ingeniería de Software y Sistemas de Información
Departamento de Ciencias e Ingeniería de la Computación

Universidad Nacional del Sur

³Comisión de Investigaciones Científicas, Provincia de Buenos Aires

Bahía Blanca - Buenos Aires - Argentina

e-mail: {ldm,kmc,jra,prf}@cs.uns.edu.ar

Resumen

En la actualidad, la utilización de fuentes de datos compartidas facilita el trabajo en equipo. Por este motivo las organizaciones implementan en sus locaciones sistemas con controles de acceso para compartir los datos de acuerdo a los privilegios de los usuarios. Nuevas formas de trabajo son la distribución de los miembros de un equipo en distintos lugares físicos, el trabajo desde las casas de los empleados, el traslado temporal a otra locación. Por todas estas razones, acceder a los datos en forma remota es una necesidad en crecimiento. Una respuesta a esta necesidad es la propuesta de una arquitectura referente en capas ICSAD (Interfaz, Control y Sistemas de Archivos Distribuidos). La misma permite construir una implementación que facilita la descarga de los documentos y el control de versionado para el caso en el que varios usuarios estén accediendo en modo modificación. En el laboratorio de sistemas distribuidos comenzamos a trabajar en la implementación de la arquitectura propuesta.

Palabras claves: sistemas de archivos distribuidos, Acceso remoto, implementación -sistemas distribuidos

Contexto

Esta línea de investigación y desarrollo se realiza en el ámbito del Departamento de Ciencias e Ingeniería de la Computación de la Universidad Nacional del Sur (UNS). En particular, como parte de las tareas que se realizan en el Laboratorio de Investigación en Sistemas Distribuidos (LISIDI) y

en el Laboratorio de Investigación y Desarrollo en Ingeniería de Software y Sistemas de Información (LISSI). El proyecto se financia parcialmente con fondos asignados por la UNS a proyectos de investigación.

Introducción

La utilización de recursos desde distintos espacios geográficos se ha incrementado y expandido a partir del auge de los dispositivos móviles, las redes de banda ancha y las conexiones inalámbricas. Las nuevas tecnologías motivan a las organizaciones a solicitar nuevos requerimientos para mejorar la calidad y eficiencia del trabajo de sus empleados.

La administración de datos compartidos dentro de una empresa u organización, comúnmente se realiza a través de un sistema de archivos distribuidos, con capacidades para el manejo de usuarios, permisos en una red de área local o nube propia. En algunos casos, estos sistemas no ofrecen seguridad, escalabilidad y operabilidad traspasando los límites organizacionales.

La alternativa de utilización de una red privada virtual (VPN) [6], en algunos casos, no es una vía de solución aceptable para las empresas, ya que no respeta las políticas de seguridad informática establecidas internamente. Cabe destacar, en este punto, que no es recomendable brindar a todos los usuarios acceso VPN a la red local de una organización por motivos de seguridad o en algunos casos también por motivos de velocidad.

Entre las alternativas existentes para gestionar el acceso a archivos compartidos y de uso habitual

en las organizaciones actuales pueden mencionarse *HFS*, *mod_dir* de Apache y *WebDAV*.

En primer lugar, *HFS* (HTTP File Server) [11] permite compartir archivos fácilmente entre un grupo de trabajo a través del protocolo *HTTP*. La compartición de los archivos se puede limitar a un grupo de usuarios o permitir que todos puedan acceder a los mismos. La diferencia con respecto a otros sistemas de archivos es que no se requiere de una red.

Una segunda alternativa, *Apache Module mod_dir* [3] se utiliza para redireccionar barra final (*trailing slash*) y servir como índice de directorio de archivos. Es una forma simple que se utiliza para compartir archivos, en especial es usada en las fuentes de datos de libre distribución para acceder a los servidores. Una de las ventajas de esta herramienta es que se puede instalar sobre diferentes sistemas operativos sobre los cuales se puede ejecutar Apache HTTP Server como: Microsoft Windows, GNU/Linux, Unix, OS X, etc.

En tercer lugar *WebDAV* (Web-based Distributed Authoring and Versioning) es un conjunto de extensiones de *HTTP*, que permite a los usuarios colaborar entre ellos para editar y manejar archivos en servidores web a través de la red. *WebDAV* está documentado en RFC 2518 y extendido en RFC 3253, RFC 2518 especifica el conjunto de métodos, encabezados y tipos de contenido secundario a *HTTP/1.1* para el manejo de propiedades, creación y administración de colecciones de recursos. Para usuarios comunes, *WebDAV* permite a equipos de desarrollo web y otros grupos de trabajo utilizar un servidor web remoto tan fácilmente como un servidor de archivos local.

Líneas de Investigación y Desarrollo

La tendencia del acceso remoto está en continuo incremento. El trabajo remoto y colaborativo es utilizado en diferente tipo de empresas. La necesidad de contar con herramientas de software que faciliten las operaciones con bajo costo en la comunicación y recursos físicos.

Los ejes centrales de esta línea de investigación dentro del proyecto son:

- La exploración, análisis y propuesta de una arquitectura referente para el desarrollo y despliegue de acceso remoto a fuentes de datos.

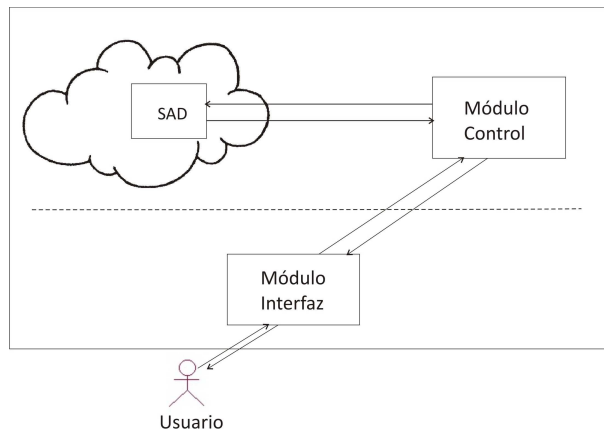


Figura 1: Arquitectura referente del Modelo IC-SAD

- Especificación de políticas para el acceso remoto.
- Diseño e implementación de una arquitectura que respete las políticas de derechos, restricciones y seguridad.

Resultados y Objetivos

Este proyecto comienza con el análisis de diferentes propuestas para acceder a fuentes de datos en forma remota y habilitar el trabajo desde distintas locaciones. Como primer resultado del trabajo y teniendo en cuenta las consideraciones de las propuestas analizadas, en los laboratorios (LISIDI - LISSI) se comenzó a trabajar en una nueva alternativa, la misma es una arquitectura modular denominada ICSAD, presentada en [1].

El desafío es brindar acceso remoto desde distintos tipo de dispositivos, garantizando que se respeten las políticas de seguridad utilizadas dentro de los límites de la organización. Así la arquitectura referente, que se denomina en forma abstracta INTERFAZ, CONTROL Y SISTEMA DE ARCHIVOS DISTRIBUIDOS (ICSAD), se muestra en la figura 1.

Como puede apreciarse, el modelo propuesto está organizado por capas. Cada capa es independiente, y se comunican a través de las interfaces definidas, de tal manera que si se modifica el comportamiento de las funciones no sea necesario modificar el resto de los componentes.

Los componentes principales son el sistema de archivos distribuidos (SAD), el módulo de control y el módulo de interfaz.

- Sistema de Archivos Distribuidos (SAD): este componente se encuentra dentro de los límites de la organización. Incluye todas las operaciones para el manejo de los archivos, las capacidades de acceso, la compartición de los directorios y archivos, la administración de los usuarios y permisos.
- Módulo de control: este componente es el encargado de conectar al módulo de interfaz con el sistema de archivos distribuidos, es la puerta de entrada a la organización desde el exterior. Incluye funciones para garantizar la seguridad en el acceso, permitiendo a los usuarios acceder a la información permitida desde el componente SAD. Además, se incluyen las funciones para leer, copiar, modificar, agregar documentos en el SAD.
- Módulo de interfaz: este componente se ejecuta en cada uno de los puntos de acceso remoto, como puede ser un teléfono celular, *tablet*, *notebook*, etc. Todas las operaciones requeridas sobre el SAD se realizan a través del módulo de control.

El componente principal de esta arquitectura es el módulo de control. Que incluye las siguientes funciones:

- Lectura: la política utilizada para la implementación de esta operación es la descarga del archivo (*downloading*).
- Escritura: esta operación es equivalente a la creación de un nuevo documento en el sistema de archivos, considerando que no puede tener el mismo nombre que un archivo existente.
- Modificación: en este caso se trata de modificar un archivo existente, para ello se puede adoptar una de las siguientes políticas en el módulo de control:
 - Semántica de sesión: cuando un archivo es modificado por varios usuarios en el mismo instante de tiempo se guarda en el sistema de archivos la última copia cargada (*uploading*).
 - Semántica de versionado: en este caso, se almacenan todas las versiones del archivo bajo el mismo nombre pero con algún atributo distintivo, como puede ser el

usuario, la fecha y hora de carga del archivo, o bien un identificador interno.

Si se profundiza sobre los aspectos de implementación del módulo de control se puede distinguir una serie de aspectos principales.

El módulo de control puede ubicarse físicamente:

- en un servidor que hace de intermediario;
- en el mismo servidor donde se encuentra una parte del sistema de archivos distribuido;
- presente en cada uno de los servidores del sistema de archivos distribuido.

El conjunto de actividades a realizar para el diseño, desarrollo y despliegue de la propuesta son:

1. Implementación completa del sistema, es decir de todos los componentes del módulo de control y de interfaz del usuario.
2. Poder definir la autenticación sobre diferentes implementaciones de archivos distribuidos que hagan uso de capacidades de acceso mediante: Samba, FTP o SFTP.
3. Se instalará y configurará un ambiente de trabajo virtual para poder llevar a cabo experimentos y pruebas de uso de los programas/módulos a implementar:
 - Configuración de un sistema de archivos distribuidos (utilizando Microsoft Windows Server 2012)
 - Configuración de acceso remoto mediante: Samba, FTP y SFTP.
 - Implementación del módulo de control.
 - Implementación del módulo de interfaz con el usuario.
 - Pruebas con navegadores en Workstations y dispositivos móviles (celulares y tablets).

Formación de recursos humanos

En relación con la formación de los recursos humanos, el trabajo a desarrollar permite una capacitación práctica adecuada de los profesionales

involucrados, permitiendo plasmar con logros concretos las investigaciones realizadas y las especificaciones de la arquitectura definida, todo dentro del ámbito del Departamento de Ciencias e Ingeniería de la Computación. Además el uso por parte de usuarios finales de nuestra propuesta en empresas del medio es posible, debido a que ya se han realizado consultas y se ha confirmado el interés de contar en ámbitos productivos con la misma, redundando en beneficios en diversas actividades económicas de la sociedad.

Referencias

- [1] K. Cenci, L. de Matteis, and J. Ardenghi. Arquitectura en capas para acceso remoto. In *XIX Congreso Argentino de Ciencias de la Computación (CACIC 2013)*, pages 21–25, 2013.
- [2] The Apache Software Foundation. *Module Apache*. <https://www.modules.apache.org>.
- [3] The Apache Software Foundation. *Apache Module mod_dir*, 2013. http://httpd.apache.org/docs/current/mod/mod_dir.html.
- [4] G. Harlan, A. Baranovski, K. Bloom, and et al. Hadoop distributed file system for the grid. In *IEEE Nuclear Science Symposium Conference Record*, pages 1056–1061, 2009.
- [5] L. Hernández and M. Pegah. Webdav: What it is, what it does, why you need it. In *SIGUCCS '03, ACM*, pages 249–254, 2003.
- [6] R. Hills. *Common VPN Security Flaws*, 2005. <http://www.nta-monitor.com/>.
- [7] N. Islam, M. Rahman, J. Jose, R. Rajachandrasekar, H. Wang, H. Subramoni, C. Murthy, and D. Panda. High performance rdma-based design of hdfs over infiniband. In *IEEE*, pages 1–12, 2012.
- [8] T-J Liu, C-Y Chung, and C-L Lee. A high performance and low cost distributed file system. In *Software Engineering and Service Science (ICSESS), 2011 IEEE 2nd International Conference on*, pages 47–50, 2011.
- [9] S. Malgaonkar, S. Surve, and T. Hirave. Distributed files sharing management: A file sharing application using distributed computing concepts. In *IEEE International Conference on Computational Intelligence and Computing Research*, 2012.
- [10] A. Mashtizadeh, A. Bittau, Y. Huang, and D. Mazières. Replication, history, and grafting in the ori file system. In *SOSP'13, ACM*, pages 151–166, 2013.
- [11] M. Melina. *HFS Http File Server*, 2002. <http://www.rejetto.com/hfs/>.
- [12] S. Miltchev, J. Smith, V. Prevelakis, A. Keromytis, and S. Ioannidis. Decentralized access control in distributed file systems. *ACM Comput. Surv.*, 40(3):10:1–10:30, August 2008.
- [13] J. Siegel and P. Lu. User-level remote data access in overlay metacomputers. In *Proceedings of the IEEE International Conference on Cluster Computing (CLUSTER'02)*, pages 1–4, 2002.
- [14] J. B. Weissman, M. Marina, and M. Gingras. Optimizing remote file access for parallel and distributed network applications.
- [15] F. Zaffar and G. Kedem. Paranoid: A global secure file access control system. In *Proceedings of the 21st Annual Computer Security Applications Conference (ACSAC 2005)*, IEEE, 2005.