# PAPI para acceso remoto a publicaciones electrónicas bajo contrato

## PAPI for Remote Access to Electronic Publications under Contract

Luis Meléndez


### Resumen

Este artículo presenta una solución basada en PAPI al problema de cómo puede una institución facilitar a su personal el acceso a recursos electrónicos remotos contratados cuando el control de acceso que realiza el proveedor es por dirección IP y se desea acceder desde fuera de la red interna.


**Palabras clave:** PAPI, control de acceso, proxy, biblioteca


### Summary

This paper presents a solution based on PAPI to solve the problem an institution has when access to purchase remote electronics resources is requested by its staff from outside the local network and the provider bases access control on IP adresss.


**Keywords:** PAPI, access control, proxy, library

PAPI proporciona una forma de solucionar el problema que se plantea para facilitar el acceso a determindos recursos cuando el investigador desea acceder desde fuera de la red interna

## 1.- Introducción

Hoy día la forma predominante en que los investigadores consultan bibliografía, bases de datos y publicaciones es mediante conexiones Web a las empresas proveedoras de dicha información. Para ello debe existir un contrato entre la institución y esas empresas, las cuales, como forma de control de acceso, suelen permitir las conexiones si la dirección IP del ordenador cliente pertenece al rango declarado por la institución contratante. Eso plantea el problema de cómo facilitar el acceso a esos recursos cuando el investigador desea acceder desde fuera de la red interna (por ejemplo desde su casa). PAPI proporciona una forma de solucionar ese problema.

El personal de la Universidad, por el simple hecho de serlo, tiene el derecho de acceder a los recursos contratados por ésta, independientemente del lugar desde el que desee hacerlo.

El control por dirección IP hace difícil garantizar ese derecho. Sea cual sea el mecanismo técnico que se implemente para solucionar el problema, debe cumplir dos premisas básicas:

- El acceso final al proveedor debe realizarse desde un ordenador de la Universidad
- Ésta debe controlar el acceso al mecanismo que se implemente de forma que sólo pueda ser usado por su propio personal.

El primero es un condicionante técnico; el segundo un requisito de la licencia contratada.

Existen varias tecnologías que permiten dar solución al problema cumpliendo esas premisas. Algunas de ellas son: VPN, VPN/SSL, Proxy Web, Productos específicos (Ezproxy, libproxy, etc.), PAPI.

Cada una de ellas tiene, como es habitual, sus ventajas e inconvenientes y no es propósito de este artículo realizar una comparativa. En nuestro caso, a pesar de tener previamente montada una infraestructura de VPN, decidimos en su momento probar PAPI por el conjunto de ventajas que ofrecía, y el resultado fue tan positivo que es la solución que actualmente utilizamos. Algunas de las características de PAPI son:

- No es necesario instalar ningún tipo de software ni configurar nada en el ordenador ni en el navegador.
- Funciona con cualquier navegador moderno.

- Es independiente del sistema operativo del usuario.
- Se puede usar desde cibercafés o puestos de uso público.
- El software es de código abierto y además está desarrollado por RedIRIS.

Además, si se configura adecuadamente, su uso es totalmente transparente al usuario. Si la universidad dispone de una página Web con los enlaces a los recursos contratados, el usuario simplemente tiene que acceder a ella tal como lo haría desde el PC de su despacho y se le pedirá que se autentifique en caso necesario.

Por contra, hay que tener en cuenta que PAPI es un software complejo, cuyo alcance va mucho más allá de la utilización para resolver el problema que nos ocupa. Por ello, aunque no es difícil de instalar y existe buena documentación, requiere un esfuerzo ponerlo en marcha. Sin embargo, la implementación que presentamos aquí se puede considerar casi como una solución "llave en mano". Cualquiera que desee montarlo de la misma forma, puede tener todo listo en muy poco tiempo partiendo de nuestros ficheros y configuraciones. De hecho, ya hay al menos un precedente que lo confirma.

## 2.- Conceptos de PAPI

Vamos a dar simplemente un breve repaso a los conceptos que conviene entender para su utilización:

- **Point of Access (PoA)**. Se trata de cualquier recurso Web cuyo acceso sólo debe permitirse a usuarios autorizados (previamente autentificados por un AS).
- **Authentication Server (AS).** Es la parte de PAPI que autentifica a un usuario y hace que su navegador cargue las cookies que le van a proporcionar acceso a los sitios Web protegidos (PoAs).
- **Proxy de reescritura (Rewriting Proxy).** Es una funcionalidad opcional de un PoA. Cuando se accede a una URL bajo éste, el proxy accede a la misma URL del servidor original (que tendrá configurado) y la sirve al navegador, cambiando enlaces si es necesario, para que apunten a el mismo en vez de al sitio original. La idea es que el usuario nunca se conecte directamente a éste.

En la versión actual (1.3.x) de PAPI, el Servidor de Autenticación está implementado como un CGI en Perl, y el PoA en mod_perl para Apache. Las futuras versiones de PAPI serán más independientes del servidor web.

En realidad PAPI permitiría que no existiera el problema que tenemos entre manos (aunque sirva igualmente para resolverlo). Los sitios Web de los proveedores serían Puntos de Acceso, configurados para confiar en los Servidores de Autenticación que montarían sus instituciones cliente. Un AS hace accesible a un Punto de Acceso, cuando el navegador de un usuario intenta acceder a la página Web que controla, la información que el PoA necesita para permitir o no el acceso (y que habitualmente se limita poco más que a 'este usuario se ha autentificado ante mí'). El Servidor de Autenticación puede usar cualquier medio para validar a los usuarios (LDAP, ficheros planos, etc.) y los PoAs deben confiar en que ha hecho bien su trabajo. Es decir, la idea sería separar la autentificación del control de acceso, elementos que estarían bajo dominios administrativos diferentes. Lógicamente, por tanto, AS y PoA son elementos de software independientes.

Se configura todo de forma que al acceder a determinada/s URLs locales, 'salte' el PoA que tras hacer sus comprobaciones permitirá o no el acceso (el navegador debe presentar unos cookies que previamente habrá cargado mediante atributos SRC de elementos presentes en una página generada por el AS, que habrá previamente autentificado al usuario en el que confíe el PoA).

Usado de esta manera no sería necesaria la función de proxy de reescritura que está incorporada en el software de PoA.

Como PAPI (aún) no está implantado por los grandes proveedores, lo usamos de otra manera. Los PoAs que vamos a configurar no protegen páginas Web reales, sino que se limitan a hacer de intermediarios entre el navegador del usuario y los proveedores de recursos (RewritingProxy). Para poder acceder a esta funcionalidad (como PoAs que son) el AS nos habrá autentificado previamente como usuarios de la Universidad.

Nuestros PoAs van a tener cada uno un nombre DNS y un host virtual, y cada uno de ellos hará de proxy para un proveedor (más adelante se verá el porqué de estas decisiones).

Así, por ejemplo, papi13.uco.es es proxy para online.issn.org. De esta forma cuando el usuario accede a papi13.uco.es/assisted.html, ocurre lo siguiente:

- El Punto de Acceso que tenemos configurado en el servidor virtual papi13.uco.es comprueba que el navegador le ha presentado los cookies adecuados que demuestran que se ha autentificado al AS.
- Obtiene la página online.issn.org/assisted.html (la misma URL que se le ha pedido pero del servidor que tiene configurado para actuar de proxy).
- Cuando la obtiene, cambia en ella cualquier referencia que pueda haber a online.issn.org por papi13.uco.es.
- Devuelve la página resultado al navegador

No es conveniente decirle a los usuarios que deben conectarse al Servidor de Autenticación cuando quieran acceder desde fuera de la red interna, eso es transparente a ellos

El objetivo es que no se presente al navegador ninguna página que pueda contener un enlace directo al proveedor, porque en caso de acceder a él, éste denegará el acceso si el usuario no está en un ordenador de la universidad.

Según lo anterior, es necesario que el navegador se conecte a los distintos hosts virtuales (papi1, papi2, etc.) en vez de a los proveedores originales cuando accede desde fuera de nuestra red.

En vez de hacer copias de las páginas que contienen esos enlaces, usamos de nuevo la función de proxy de reescritura para que los cambie. Cuando el usuario se autentifica al AS, éste, como parte final de su tarea, nos redirige a un PoA especial que es proxy para nuestro propio servidor Web, y configurado para que cambie cada enlace original por otro que apunta al PAPI correspondiente, usando directivas como:

```
PAPI_Redirect http://online.issn.org/ http://papi13.uco.es/
```

No es conveniente decirle a los usuarios que deben conectarse al AS cuando quieran acceder desde fuera de la red interna. Eso es transparente a ellos. En su lugar se pueden usar varios mecanismos para que cuando un usuario acceda a una página con enlaces a recursos contratados desde fuera, se le redirija automáticamente al AS para autentificarse. Tras hacerlo, se le volverá a redirigir, esta vez al PAPI que hace de proxy del Web corporativo (el usuario está realmente viendo este Web, pero con los enlaces cambiados). En nuestro caso la primera redirección la hacemos simplemente con un fichero `.htaccess`

## 3.- Experiencia adquirida

En la Universidad de Cordoba llevamos usando PAPI de esta forma desde abril de 2004. El trabajo inicial fue importante, pero ha merecido la pena. El software es estable y no ocasiona problemas. El

único mantenimiento que necesita es cuando se contratan nuevos recursos o cuando las URLs de acceso a alguno de ellos cambia. Además de definir, en el primer caso, un nuevo host virtual (normalmente es cuestión de copiar y pegar) hay que configurar el proxy usando directivas `Remote_URL` (si el proveedor tiene un único servidor) o `Remote_Domain` (si tiene varios de los que carga distintos recursos o provoca redirecciones entre ellos), y a veces algunas `PAPI_Redirect`. El equipo de desarrollo de PAPI dispone de un repositorio con estas definiciones para una gran cantidad de proveedores y, en caso de necesitar una que no esté, no suele ser difícil definirla. En nuestro caso, casi siempre ha bastado con un simple `Remote_URL`.

En caso de encontrar problemas con algún proveedor, hay que recordar que PAPI puede ser una opción además de otras (tipo VPN). Los usuarios lo prefieren porque pueden usarlo desde cualquier sitio y no tienen que instalar nada en su PC (y los administradores porque no hay que dar soporte a problemas relacionados con el PC o navegador del usuario), pero si se tiene implantada una infraestructura de VPN o semejante, puede ofertarse ésta última como opción.

El equipo de desarrollo de PAPI dispone de un repositorio con estas definiciones para una gran cantidad de proveedores

**Luis Meléndez Aganzo,**
(luism@uco.es)
Servicio de Informática
(Equipo de Docencia e Investigación)
Universidad de Córdoba

## Referencias


- Instalación de PAPI en la Universidad de Córdoba para proporcionar acceso remoto a bases de datos y publicaciones electrónicas. *URL:* http://papi.rediris.es/doc/
- *PAPI para acceso remoto a publicaciones electrónicas bajo contrato*. Presentación en las JJTT RedIRIS 2004 URL: http://www.rediris.es/jt/jt2004/archivo/archivo-jt.es.html