

How EZproxy is implemented at USQ
Corey Wallis, Electronic Services Officer – University of Southern Queensland
Library
wallis@usq.edu.au

The EZproxy learning curve

EZproxy is one of the core Library systems at USQ and compared to the others is very easy to configure and maintain. The learning curve for EZproxy can be very steep until you learn where the best information can be found.

It is also important to remember that hundreds of libraries around the world are using EZproxy to give access to resources for their students. If you are trying to configure a resource and are having troubles it is likely that another person who is also using EZproxy has already configured that resource.

Even in Australia over 45 libraries and institutions have purchased an EZproxy licence and many people are willing to provide assistance when configuring a resource that is specific to the Australian Market.

To help with the learning curve there are a number of places that have documentation and information available for those that are new to EZproxy.

Sources of information available to you to learn about EZproxy

The main source of information is the Useful Utilities website and more specifically the support section of the website. <http://www.usefulutilities.com/support/> It provides documentation on the initial evaluation, initial setup and maintenance of EZproxy. When you're trying to configure a resource for the first time the Database Specific Issues section of the website can be helpful <http://www.usefulutilities.com/support/db/>.

The best source of information regarding EZproxy is the EZproxy e-mail list that is provided by the University of New York <http://www.usefulutilities.com/support/list.html>. By subscribing to this list you gain access to a wealth of knowledge and experience from users of EZproxy around the world. The author of EZproxy, Chris Zagar, is also a member of the list and will usually reply to posts within a day. Chris is also very supportive and will work hard with you in resolving issues that you may come across.

Most, if not all, of the Australian users of EZproxy are on the e-mail list and it is therefore a good place to ask questions about those electronic resources that may be considered to be specific to the Australian market.

The basic EZproxy configuration commands

Configuring EZproxy can be a relatively simple process depending on the type and complexity of the resource that needs to be configured. EZproxy is configured by editing a text file and adding in configuration entries. The most common configuration entries are outlined below. Please note this is not meant to be a comprehensive list.

- Title
- URL
- Domain (DomainJavaScript)
- Host (HostJavaScript)
- AutoLoginIP
- ExcludeIP
- IncludeIP

The first four options outline the basic configuration options that all of the individual resource configurations in the EZproxy configuration file contain.

The `Title` option specifies the title of the resource that is to be proxied. For example if we were creating a basic configuration for the electronic resource EBSCOhost our title line would be as follows.

```
Title EBSCOhost
```

The second configuration option is the starting URL for the resource and in the case of EBSCOhost it would be as follows

```
URL http://search.global.epnet.com
```

This information is also used for the starting URL that makes the user access the resource via EZproxy.

The `Domain` option specifies what domains EZproxy should proxy when a user accesses this resource. In this instance the option would read

```
Domain epnet.com
```

The `Host` option specifies hosts of other starting URLs that may also be used for this configuration. For example in Science Direct it is possible to get URLs to the full text of articles that use the host linkinghub.elsevier.com. Therefore to complete our configuration we would add the following entry.

```
Host search.epnet.com
```

The complete configuration entry in the EZproxy configuration file would be as follows, regardless of the proxy method employed.

Title EBSCOhost
URL `http://search.global.epnet.com`
Domain `epnet.com`
Host `search.epnet.com`

This configuration will work for links such as

- `http://search.global.epnet.com/login.asp`
- `http://search.epnet.com/direct.asp?an=13956111&db=aph`

The first of these links is the standard login URL for the EBSCOhost suite of databases. When accessing an EBSCOhost database this is the URL that would typically be used. When accessing the resource via EZproxy the following URL would be provided to the user (assuming the proxy by host method is employed).

- `http://ezproxy.usq.edu.au/login?url=http://search.global.epnet.com/login.asp`

Access to the EBSCOhost suite of databases will be via the proxy because the URL specified in the url parameter to the login page matches the one that is specified in the URL command of the configuration. Subsequent pages that make up the EBSCOhost website will also be accessed via the proxy because the domain `epnet.com` is specified in the Domain command of the configuration.

The second URL is a persistent link to a specific journal article in one of the EBSCOhost databases. When accessing the article via EZproxy the following URL would be provided to the user (assuming the proxy by host method is employed).

- `http://ezproxy.usq.edu.au/login?url=http://search.epnet.com/direct.asp?an=13956111&db=aph`

Access to this URL will be via the proxy because the domain `search.epnet.com` is specified in the Host command of the configuration. The Host command takes effect in this instance because this will be the first EBSCOhost URL that the user will attempt to access in this session. In this instance a session is defined as starting when the login page is called with a url parameter and ends when the user accesses a different resource via EZproxy, the user logs out, or the session expires after 20 minutes of inactivity. Subsequent pages will be accessed via EZproxy because the domain `epnet.com` is specified in the Domain command of the configuration.

The above configuration will work well except that some links will not be proxied. This is because some of the links use JavaScript and the above configuration doesn't take this into account. Fortunately EZproxy extends the Domain and Host configuration options by providing `DomainJavaScript` and `HostJavaScript`.

These two configuration options make EZproxy check inside JavaScript code for the domains or hosts that need to be proxied. As this is a more process intensive task it isn't enabled by default. To update our EZproxy configuration to take into account the links in EBSCOhost that use JavaScript we can update the configuration so that it becomes as follows.

```
Title EBSCOhost
URL http://search.global.epnet.com
DomainJavaScript epnet.com
Host search.epnet.com
```

It is important to note that in previous versions the command options listed above were not available in the form listed. They were in an abbreviated form and this form is still valid. For example the following configuration means exactly the same as the one previously.

```
Title EBSCOhost
URL http://search.global.epnet.com
DJ epnet.com
H search.epnet.com
```

The last three commands refer to the way EZproxy authenticates a request for a particular page by determining the IP address of the computer that made the request. All three options can have one IP address or a range of IP addresses as parameters.

The `AutoLoginIP` command proxies requests from the specified IP address, or range of IP addresses, by the IP address parameter with the user being automatically logged in. For example the option detailed below would automatically login any request from the specified IP address range for any resource configurations that follow.

```
AutoLoginIP 192.168.0.0-192.168.0.255
```

The `ExcludeIP` command excludes a request from the specified IP range from being proxied. In the following example any request for any of the resources that are configured below this line in the EZproxy configuration file will be redirected to the resource directly when they come from the specified IP address.

```
ExcludeIP 192.168.0.100
```

The `IncludeIP` command proxies requests from the specified IP address parameter in a similar way to the `AutoLoginIP` command. However it does not automatically log the user in. A user attempting access to this resource from the following IP address would need to provide their username and password before EZproxy would proxy their request.

```
IncludeIP 192.168.0.55
```

In the following example requests for the EBSCOhost database from the IP range 192.168.0.0 to 192.168.0.255 would be excluded from being proxied by EZproxy. However requests from the IP address 192.168.0.111 would be proxied and the user who made the request from this IP would be prompted for their username and password. A request from any other IP address will also be proxied and the user prompted for their username and password before continuing.

```
ExcludeIP 192.168.0.0-192.168.0.255  
IncludeIP 192.168.0.111
```

```
Title EBSCOhost  
URL http://search.global.epnet.com  
DJ epnet.com  
H search.epnet.com
```

Any other database configuration following the EBSCOhost configuration will also have the same restrictions applied.

EZproxy is very versatile in the way it accommodates user authentication. EZproxy can be configured to use a variety of methods including, but not limited to:

- FTP
- IMAP
- POP3
- LDAP
- Referring URL

EZproxy also supports a mechanism for authentication by utilising an external script. In this way the exact authentication mechanism can be hidden from EZproxy. For example at the University of Southern Queensland EZproxy is configured to use LDAP for authentication by communicating directly with the central USQ LDAP server. The specific way in which EZproxy can be configured to authenticate users is outside the scope of this paper.

Basic configuration steps

These are the basic configuration steps that I follow when I need to configure a new resource.

1. Confirm that the subscription is indeed active
 - a. The first step is to confirm that the subscription is indeed active and operating as it should without going through EZproxy. This is determined by using the subscription details supplied by the publisher
 - b. Confirm that you can indeed get the full text of articles that you have a subscription for

2. Make a note of the domains used by the resource
 - a. Some resources store their full text on different servers and domains than the website
 - b. Some resources also redirect from one domain to another when you first visit the resource
 - c. One effective way is to watch closely the Address or Location bar of your Internet browser and to take note of the domains that are displayed when you access the resource
 - d. Another way to monitor exactly what is occurring between your browser and the web server is to use a product such as Paros from ProofSecure <http://www.proofsecure.com/> that allows you to see the entire HTTP, or HTTPS conversation between your browser and the publishers website
3. With this information start to build the configuration
 - a. Initially it is a good idea to not use the DomainJavaScript and HostJavaScript options to see if the website will proxy successfully without the need for checking for domains in JavaScript
4. Have a section of the configuration file that will allow you to test the configuration from within your own network
 - a. Most institutions will choose not to proxy accesses by users from on campus. This can make it difficult to check if the configuration is indeed working
 - b. By having a section of your configuration file that includes your local network you can check the configuration within your network and avoid the need for an external Internet connection
5. Restart the EZproxy service so that your changes to the configuration file take effect.
 - a. When restarting the service any users existing sessions will not be lost. However any transfers that were currently in progress will be broken and will be needed to be restarted
6. Check that all links are being re-written and that access is still correct
 - a. Confirm that all of the links in the website are being re-written and that you can still access the full text of resources
 - b. Don't forget to check that the URLs for such things as images and other multimedia components of the website are also be rewritten by your EZproxy server
 - c. If this is not the case go back to the configuration you have created and tweak as necessary
7. Once you have a successful configuration you can put it into production
 - a. You can move the configuration into production by either moving it outside of the test section of your configuration file or by copying it from your test server to the production server

8. Initially closely monitor access to the resource
 - a. Once the electronic resource has been made available to your users it is advisable to monitor closely the performance of the resource to ensure the your configuration is correct
 - b. Occasionally it is necessary to check access to the electronic resource from outside of your IP address to track down issues that were not immediately apparent

For some of the bigger publishers it is also advisable to e-mail the EZproxy e-mailing list because someone at another Library may already have a working configuration. It is also important to note that the simplest type configuration is outlined here. Some publishers' sites require a more advanced configuration. In this instance it is highly likely that Chris Zagar has dealt with the publisher directly and will have a configuration on the Useful Utilities website. For example below is the configuration for the Thompson ISI product web of Science.

```
Option DomainCookieOnly
Title ISI Databases
URL http://isiknowledge.com/
DJ isiknowledge.com
DJ isihighlycited.com
DJ newisiknowledge.com
DJ newisiknowledge.com
DJ webofscience.com
DJ jcrweb.com
DJ isicc.com
Find value="http://"
Replace value="http://^A
Find VALUE="http://"
Replace VALUE="http://^A
Find rurl=http://"
Replace rurl=http://^A
Find product_st_thomas=http://"
Replace product_st_thomas=http://^A
Find return_url=http://"
Replace return_url=http://^A
Find ST_URL=http://"
Replace ST_URL=http://^A
Option Cookie
```

The exact meaning of this configuration is outside the scope of this paper.

Proxy by Port vs. Proxy by Hostname

There are two ways that EZproxy can provide access to electronic resources and these are referred to as Proxy by Port and Proxy by Hostname.

When EZproxy is configured to use Proxy by Port, EZproxy uses a unique port number to distinguish between one resource and another, or more literally one hostname and port combination from another.

Each starting URL needs to connect to port 2048 on the EZproxy server and the user, once authenticated, would be directed to a port above this number to access their desired resource. This proxy technique does have some drawbacks. The main drawback is that port 2048 and most port numbers above this number are non-standard ports; and as such they would be closed by firewall administrators of off campus users, particularly those in large corporate environments.

This means that users behind a securely configured firewall can not access resources via an EZproxy server that is working in this manner. At the time USQ moved from Proxy by Port to Proxy by Host we knew that at any time one of our users could be directed to port 2048 to port 2848 a total of 800 ports.

It can prove to be very problematic to get a firewall administrator to open up so many ports because the more ports that are open the less secure a firewall becomes. With the increase in home users using a firewall product the problem compounded 100 fold. With a large corporation it was typically possible to ask the user to contact their network department to request that these ports be open. A home user in general doesn't know how to re-configure their firewall and so made it impossible for themselves to gain access to our resources.

Proxy by Hostname changes this by operating on the standard World Wide Web port, port 80. What this means is that if a user can access the Internet they can, once properly authenticated, gain access to our resources. EZproxy manages access by creating a unique hostname based on the host name of the resource that it is accessing on the users behalf.

For example in a proxy by port scenario once the user has gained access to the electronic resource Blackwell Synergy they would see this URL in their browser:

`http://ezproxy.usq.edu.au:2148/servlet/useragent?func=showHome`

In contrast using the Proxy by Hostname method the user would see this URL in their browser:

`http://www.blackwell-synergy.com.ezproxy.usq.edu.au/servlet/useragent?func=showHome`

(URL is wrapped)

Both provide a unique hostname for this resource while the user is accessing it.

The benefit is that the Proxy by Hostname alternative allows users to access the resource via the standard World Wide Web port, port 80, which means that no firewall reconfiguration needs to occur.

The cost of this solution is that it is more complex to set-up due to the nature of DNS and sharing one server and one Network interface card with multiple IP addresses. While Proxy by Hostname is more complicated to set-up the benefit is that it is easier for users to gain access to resources and once implemented there are less support calls from users because there are no firewall issues to contend with.

Once the changeover from Proxy by Port to Proxy by Hostname was completed at USQ our support calls in relation to access difficulties related to our Electronic Resource fell from approximately 15 per week to 1-2 per week.

How EZproxy is implemented at USQ

EZproxy is implemented at USQ in the Proxy by Hostname mode, this change was made at the beginning of 2004. We have a production server that is running Windows 2000 and provides a number of other services not just EZproxy.

We also have a test server that we use for testing new configurations, troubleshooting existing configurations and implementing new projects. The test server is currently running Linux and is providing other services to our systems team. In both cases the EZproxy service is very easy to maintain and easy on process and memory usage.

EZproxy is available for a number of different platforms and each platform specific version is functionally the same which means we can take the configuration file from one and use it on the other. The only thing that would need to be changed is some of the IP exclusions and inclusions.

In August 2004 the USQ EZproxy server serviced 2,022,033 successful requests with an average request rate of 65,228 per day. Approximately 24 gigabytes of data was transferred with 26.57% of the total traffic served being attributed to Adobe PDF documents.