

# **The Ultimate Proxy Guide**

For people who want to use proxies, host proxies and learn how they work

**Righteous Ninja**

**2010**

## Table of Contents

1	Introduction .....	2
2	Proxy Basics .....	3
2.1	Proxy Use .....	3
2.1.1	Anonymity .....	3
2.1.2	Security .....	3
2.1.3	Access .....	3
2.2	How They work .....	4
2.2.1	Access .....	4
2.2.2	Anonymity and Security .....	5
2.3	How a Proxy Engine Works .....	6
3	Finding Proxies .....	8
4	Running Your Own Proxy .....	9
4.1	Starting My Own Proxy .....	9
4.1.1	Scripts .....	9
4.1.2	Hosting .....	9
4.1.3	Monetizing your proxy .....	9
4.2	Running Your Proxy .....	10
5	Blocking Proxies .....	11
5.1	Domain or IP Blocking .....	11
5.2	URL Blocking .....	11
5.3	Deep Packet Inspection .....	11
6	References .....	12

# 1 Introduction

Web proxies come in many guises and perform many functions. In this document we will focus on suffix based proxies which are colloquially referred to as web based proxies or web proxies. These proxies do not require any changes to browser settings and are commonly used to bypass firewalls.

This document is aimed at proxy users, would be proxy users, proxy owners/webmasters, proxy developers and system administrators who are trying to block the use of these proxies. There is a continuous power struggle between proxy users who search for new proxies and system administrators who attempt to block these proxies. The aim is to give all these disparate users a thorough understanding of this industry and the challenges they face.

Buying, selling and running proxies is big business these days. You have proxy owners with over 100 proxies to their names. The document will also cover this aspect of the proxy industry by showing you how you can start your own proxy, monetize it, as well as some of the pitfalls of being a proxy owner.

## **2 Proxy Basics**

In this chapter we will cover the basics of web proxies. Who uses them, how they work and an example of how a proxy engine works.

### **2.1 Proxy Use**

Proxies are generally used by people in three scenarios. The first is anonymity, the second is security and the third is access.

#### **2.1.1 Anonymity**

In this scenario a user want's to remain anonymous to the website he or she is visiting. When a user visits a site, details such as your IP, what browser you are using etc are available to the site. Using this information the site is usually able to keep track of you. Using a proxy prevents this as the site you are visiting sees the proxy as the end user, effectively hiding your IP.

#### **2.1.2 Security**

If users are afraid that somebody is keeping a track of the sites they are visiting via routers and firewalls, then a proxy will go some way to stopping that. These web proxies usually encode the URLs so firewalls and routers have no idea what site the user is visiting.

#### **2.1.3 Access**

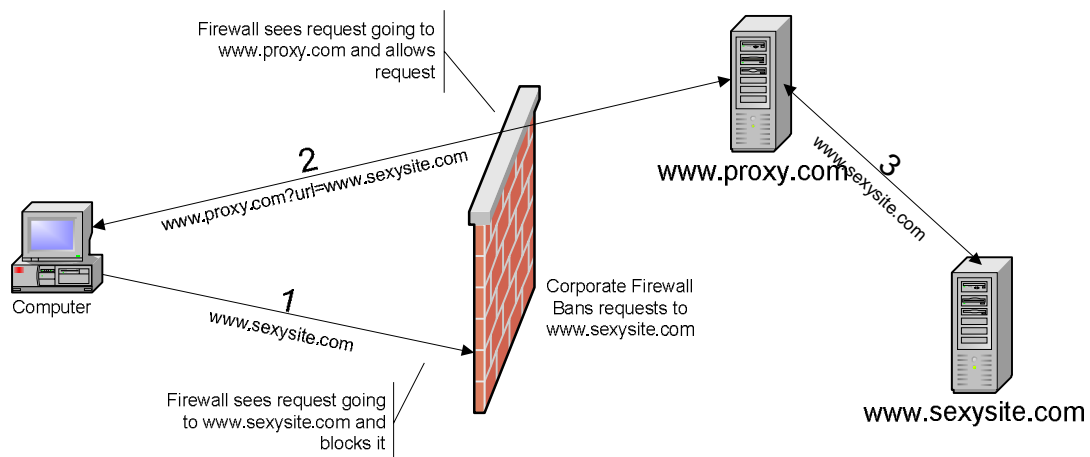
The most common use of web proxies is to get around firewalls and blocks. Social networking sites such as Facebook and Myspace etc are usually considered detrimental to productivity and are banned from work and school networks. These firewall rules are usually based on IPs and hostnames and are easily overcome by web proxies that rewrite these URLs.

## 2.2 How They work

Let us now take a look at how these proxies work. How they anonymize requests, and how they get around firewalls.

### 2.2.1 Access

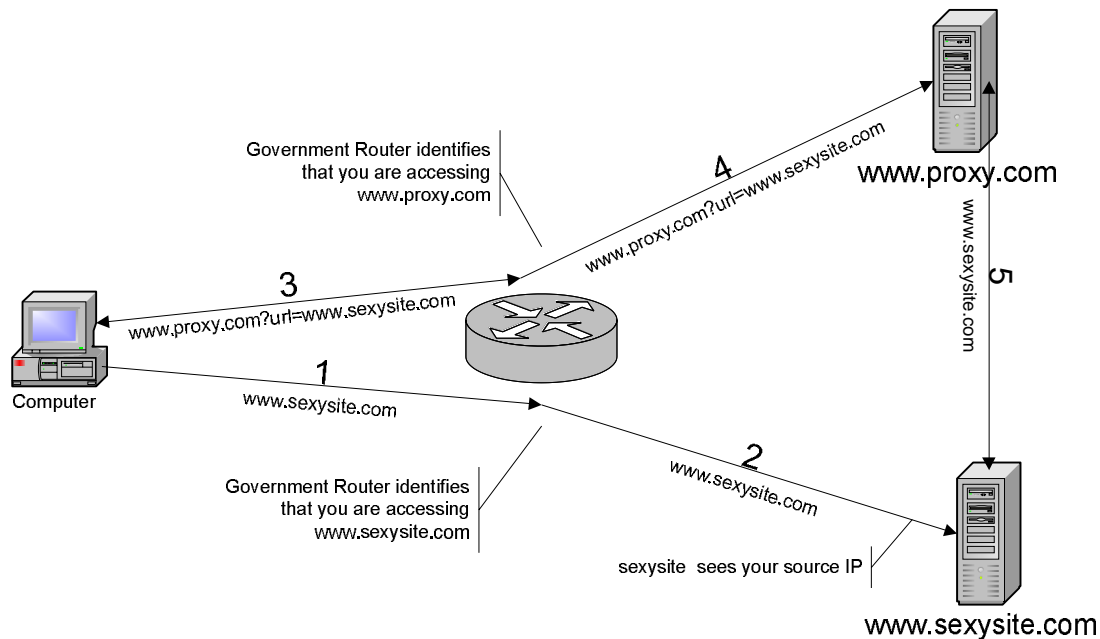
First let us see how proxies allow us to go through firewalls and bypass bans. Let us assume that you want to access a site called *www.sexysite.com*. But your network administrator has blocked this site.



1. When you try to access *www.sexysite.com* from your browser, the request first goes to the firewall. The firewall sees that you are trying to access *www.sexysite.com* and blocks it because it is a banned site.
2. When you go through the proxy, the firewall sees *www.proxy.com*. As this is not a banned site, the firewall allows it.
3. Once the request reaches *www.proxy.com*, the proxy will fetch the data from *www.sexysite.com* and return it. All this time the firewall thinks you have made a request to *http://www.proxy.com*

## 2.2.2 Anonymity and Security

Now let us look at how proxies protect your anonymity. Let us assume again that you want to visit *www.sexysite.com* but you don't want anyone to know about it.



1. When you try to access *www.sexysite.com* from your browser, the request goes through an ISP router. This router will detect what site you are visiting and store the information.
2. The site *www.sexysite.com* will then be able to detect your location through your source IP. Therefore both the site and the government know what you are doing
3. When you access *www.sexysite.com* through your proxy, the router sees a request to *www.proxy.com*. The router does not know that you are accessing *www.sexysite.com*.
4. The request goes to the proxy
5. Then the proxy makes a request to *www.sexysite.com*. The site then sees the proxies source IP. Your original source IP is not available. This way the URL you are visiting and your source IP is kept hidden.

## 2.3 How a Proxy Engine Works

Now let us take a look at how a proxy engine works. This extremely simple example is in PHP. However, you do not need any PHP experience or any programming experience to understand the process.

First let us look at a simple example.

1. Accept a url from a webpage (get request)
2. Open a cURL (or any http ) connection to the url you received.
3. Read the html from your connection.
4. Replace all the links in the html file to point to your PHP application
5. Return the output to the client.

For example let us assume that the site *http://mysite.com* contains the html below, and your proxy is hosted at *http://myproxy.com*

```
1. <html>
2. <head>
3.   <link rel="stylesheet" type="text/css"
4.     href="mystyle.css"/>
5.   <script type="text/javascript"
6.     src="http://codediaries.com/jslib/ads.js"></script>
7. </head>
8. <body>
9.   <h1>PHP Proxy</h1>
10.  This is a test page to show
11.  what a php web-page proxy does.
12.  <br/>
13.  You can find more information at
14.  <a href="http://codediaries.blogspot.com">Code diaries</a>
15.  <br/>
16.  <a href="help.html">Help</a> |
17.  <a href="contact.html">Contact</a>
18. </body>
19. </html>
```

Overleaf you will find what this html get converted to. The idea is to point all the css and javascript links, the href links the image sources etc back to your proxy via the *url* parameter. So let's assume you go to this site via the URL below.

*http://myproxy.com?url=http://mysite.com*

```

1. <html>
2. <head>
3.   <link rel="stylesheet" type="text/css"
4.
href="http://myproxy.com?url=aHR0cDovLzEyNy4wLjAuMS9waHAvYi9teXN0eWxlLmNzcw_
_"/>
5.   <script type="text/javascript"
6.
src="http://myproxy.com?url=aHR0cDovL2NvZGVkaWFyaWVzLmNvbS9qc2xpYi9hZHMuanM_
"></script>
7. </head>
8. <body>
9.   <h1>PHP Proxy</h1>
10.   This is a test page to show
11.   what a php web-page proxy does.
12.   <br/>
13.   You can find more information at
14.   <a
href="http://myproxy.com?url=aHR0cDovL2NvZGVkaWFyaWVzLmJsb2dzcG90LmNvbQ__">C
ode diaries</a>
15.   <br/>
16.   <a
href="http://myproxy.com?url=aHR0cDovLzEyNy4wLjAuMS9waHAvYi9oZWxwLmh0bWw_">H
elp</a> |
17.   <a
href="http://myproxy.com?url=aHR0cDovLzEyNy4wLjAuMS9waHAvYi9jb250YWN0Lmh0bWw_
_">Contact</a>
18. </body>
19. </html>

```

Above is the modified html. As you will have noticed, the only difference in the transformed html above and the original html in the previous section are the links and references to other sites and pages.

As you will notice all the URLs have been replaced by something like *http://myproxy.com?url=aHR0cDovLzEyNy4wLjAuMS9waHAvYi9teXN0eWxlLmNzcw\_\_*. The string after the 'url' actually contains the fully qualified url. If we look at line 4, the "aHR0cDo..". string contains the full link <http://mysite.com/mystyle.css> encoded in base 64. What it really looks in decoded format is <http://myproxy.com?url=http://codediaries.com/jslib/ads.js>

All the links should be rewritten in this format. Now when your script receives a request such as the one below,

```
http://myproxy.com?url=aHR0cDovLzEyNy4wLjAuMS9waHAvYi9teXN0eWxlLmNzcw__
```

You will have to base64 decode the string and make a cURL request to the site.

### 3 Finding Proxies

How do you find a proxy? The easy answer would be to search for free web proxies in Google and you will get millions of sites. Most of these site will be sites with lists of proxies that are rated according to uptime or speed. Proxy owners usually have to pay to get onto these lists. The downside of this is that these lists are rated according to who paid the most and not necessarily according to the best and fastest proxies. These sites are notorious for inaccurate results and you will find that most proxies listed on these sites are interminably slow or don't quite work with the sites you are interested in.

The best way is trial and error. Try out a couple of sites and stick with the one that gives you the best surfing experience. Also, most proxies will look and behave in an identical manner because they are all using the same scripts with minor visual customisations. However, some proxies will have code customisations that target a specific site such as Facebook or Myspace etc. These proxies will give you a better experience for these specific sites.

You can buy lists of proxies from sites. These sites claim to have over 100,000 proxies on their lists. The price of these lists vary, but on average, are about \$25 each. These lists are targeted more towards the webmasters who own proxy lists and not the end consumer. A consumer would have no use for a list of 100,000 proxies.

## 4 Running Your Own Proxy

What do people get out of hosting these proxies? Surely they are not doing it as a service to the public. Most proxy owners wish to make money from their proxies. There are a couple of ways of doing this. But the two generally accepted ways. The first is to sell advertising space on your proxy index page and proxified pages. The second is to sell your proxy. To make money from a proxy you first need a proxy.

### 4.1 Starting My Own Proxy

To start your own proxy you need a proxy script, a place to host it and a way to monetize your proxy.

#### 4.1.1 Scripts

There are freely available proxy scripts that can be readily downloaded. These scripts are mostly written in PHP. A few of the more popular ones are PHProxy, Surrogaffier, Glype Proxy. They are all roughly equivalent when it comes to performance.

#### 4.1.2 Hosting

Once you have downloaded a proxy you need a place to host it. You can usually find cheap hosting that costs less than \$10 a month with around 500 GB bandwidth. Some hosting providers do not allow proxies because they use a lot of CPU. Some hosting providers such as xproxyhost.com have special proxy plans. What you essentially need is PHP support and 99.9% of the hosting providers support PHP.

#### 4.1.3 Monetizing your proxy

The fastest way to monetize a proxy is to use an affiliate program to display ads on your index page and on your proxified pages. The 800 pound gorilla in this area is Google's AdSense. You can insert AdSense on your home page or start up page, but it is usually not a good idea to insert AdSense on proxified pages. Proxified pages are the pages you are proxying such as Facebook etc. This is because you cannot control what sites users go to and some of those sites may violate Google's terms of service. This will lead to Google banning ads to your domain.

There many other AdSense alternatives such as AdBrite, Bidvertiser, Clicksor and AdultFriendFinder. It is up to the individual user to try them out and see what works best. It is generally accepted that Google's AdSense is the most favoured. The

consensus amongst proxy owners is that to use AdSense on your main page or start up page and to use another less stringent affiliate such as AdBrite on the proxified pages.

The final way to make money from your proxy is of course to sell it. As a rule of thumb the cost of a site is usually how much you can earn from it in 3 months. So if your site earns about \$10 a day then the cost of that site would be about  $3 \times 30 \times 10$  about \$900, plus the cost of your domain. But the truth is that a PR2 (page rank) proxy site will retail for close to \$100. There are many forums on the web where you can find buyers and sellers for proxies.

## **4.2 Running Your Proxy**

What are the things you should look out for when you run your proxy?

### **4.2.1 Finite Proxy Life**

A lot of people believe that it is best to sell off a proxy after a certain amount of time. They believe that once your proxy is well established it will be blocked by network administrators and so will start to lose traffic. By the time you reach the

### **4.2.2 Ads on Proxified Pages**

Some affiliates such as AdSense read the text on the pages they are displayed. In this an affiliate such as Google will read the HTML on the proxified pages. If the pages have adult or violent content, Google's AdSense will register it as adult or violent content and stop serving ads to that IP. The IP being the proxies IP.

### **4.2.3 Phishing**

This is a very interesting aspect of proxies. What happens is a user navigates to a famous site such as eBay or PayPal. The user then copies the URL in the browser's address bar, pastes it in an email and sends it to the eBay or PayPal complaints mail address stating that the site is a phishing site. The problem is that the URL points to your proxy but it looks like the eBay or PayPal site. This constitutes a phishing attack.

## 5 Blocking Proxies

As a system or network administrator you might want to block access to these proxies. You have to bear in mind that new proxies keep springing up at the rate of thousands a day. There are essentially two ways you can block proxies. The first is to block the domain name of the proxy and the second is to block using URL pattern matching.

### 5.1 Domain or IP Blocking

Considering there must be over a million proxies in operation, blocking proxies individually by domain is almost next to impossible. You can trawl through proxy lists and block all the top proxies. But when will you stop? When your list reaches 10,000 entries? 100,000 entries?

### 5.2 URL Blocking

If you have the option of blocking URLs by pattern matching, then a majority of proxies use something like `www.proxy.com?u=ww.facebook.com` or `www.proxy.com?u=wey6YHL92JAbsgw`. You can use the `?u=` to block. However, some of the more advanced the customized proxies will use very individual looking URL patterns.

### 5.3 Deep Packet Inspection

Unfortunately it is next to impossible to use deep packet inspection to block a proxy. Because the content of the files do not give any information that the file is coming through a proxy. You could theoretically block a certain website by pattern matching blocks of static html. But this would be an extremely expensive exercise in terms of time and CPU.

## 6 References

<http://codediaries.blogspot.com>

For the snippets of proxy code. And help in building your own proxy.

<http://codediariesproxy.blogspot.com>

For all the proxy information.