Guide to Web Filtering Deployments
Why Pass-By Filtering is Passé

Introduction

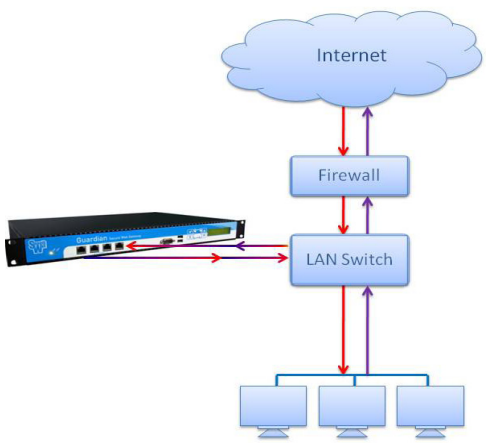
This white paper is designed to help organizations seeking to upgrade or replace their existing web filtering solution to make an informed choice on the deployment method that best suits their network. It has a particular focus on the intricacies of pass-by filtering solutions and their increasing obselescence in the context of today's ever-evolving web.

Common Filtering Deployment Methods

The aim of web content filtering is to be able to monitor and control all web traffic within an organization. There are many ways we might achieve that aim. Here we will explore the common modes. It is important to understand the differences between these filtering modes in order to reliably select the best option for your network.

Traditional Proxy

The traditional proxy is the original, and - some would say - the best mode of operation for a web content filter. Traditional filters tend to have come from a caching background. This mode uses the "web proxy settings" in your browser software.



Where they sit on the network:

Virtually anywhere. All clients must be able to route traffic to the proxy, and the proxy must be able to route to the wider Internet

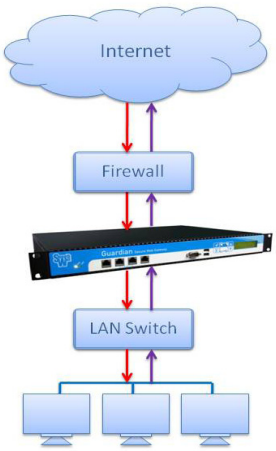
How do they intercept traffic:

Clients must be configured to use the proxy

How is direct access prevented:

Outside of the proxy, often using firewall rules

Transparent Proxy



Also called: Inline proxy, bridge proxy

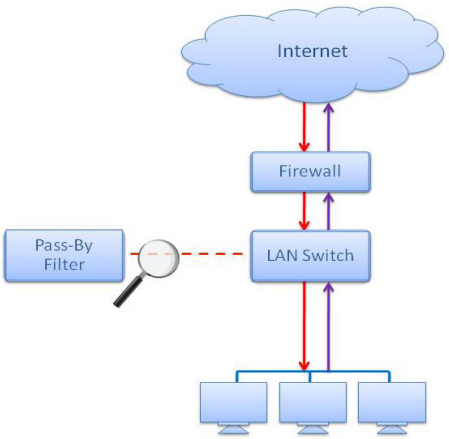
A transparent proxy transparently intercepts web requests in-line with the default route. It is commonly used to force all web traffic through a proxy, without users knowing. Transparent proxies operate without changing settings at client/browser-level.

Where they sit on the network:
Either at the gateway, or between the gateway and the client users

How do they intercept traffic:
All traffic must physically pass through the inline device, or gateway product

How is direct access prevented:
Usually as a setting within the filter product or on the host gateway device

Pass-by Filtering



Also called: Span-port filter, mirror port filter
Pass-by filtering solutions sit outside the flow of network traffic and 'watch' (as opposed to intercepting and scanning) all web requests.

Where they sit on the network:
Attached to a mirror port on a switch which carries all web-bound traffic

How do they intercept traffic:
By taking a copy from the switch

How is direct access prevented:
It isn't - clients in pass-by access the web "direct"

Pass-by Filtering - what's happening?

There are far fewer pass-by solutions available now than there were a year ago. Many filter vendors who previously offered pass-by no longer operate in that mode, or have relegated their pass-by products to the back catalogue. Even erstwhile pass-by kings 8e6 now regard the technology as past its best, preferring instead to concentrate on Finjan's "traditional proxy" platform (8e6, Marshall and Finjan have now merged to form M86). So what is causing this change, and why is it an issue?

Authentication
Because pass-by filters can't alter the data flow, it is very difficult to force a client to authenticate. Because of this, pass-by filters have had to rely on tricks to identify users for reporting purposes - directly querying the client PC, for example. These tricks can be unreliable, especially when admins would like to offer differing filter policies per group or per-user. As filters become

more advanced, and users demand more flexibility, correctly authenticating a user has become an important aspect of any web filter device.
Additionally, reporting and alerting functions have come to rely on correct user identification in order to produce rich, informative reports.

Secure Traffic

An increasing amount of web traffic travels over secure (SSL/encrypted) connections. With the falling prices of hosting & SSL certificates, it is now much easier to set up a
secure site, and no longer the preserve of banks or online shops. Today, many anonymizers are running over SSL in order to avoid detection by filters. Pass-by filters are not well placed to intercept secure traffic, and as such, will fall prey to these avoidance techniques.

SaaS

One driver for the disappearance of pass-by is a vendor movement toward SaaS. Whether customers want it or not, it is attractive to vendors to be able to offer services which are externally hosted. It is not possible to host a pass-by filter as a service, and as such, vendors who have a service offering are tending to avoid this technology. This is an example of industry pressure driving technology, rather than customer requirements.

Content Filtering (for web 2.0)

This is perhaps the biggest reason to switch away from pass-by. In pass-by mode, only the URL is checked, as this is present in the outbound request from the client. Of course the pass-by filter is able to “copy” the response as well, however by this time, the client has already completed the transaction, and the user has the web content in their browser. Today’s rapidly changing web 2.0 world is no place for an old-fashioned URL filter. Sites such as anonymizers are impossible to accurately detect and block using URL blocklists. In order to combat this, many vendors have employed “web scraping” methods outside of the filter - but as always, the web wins. If you aren’t seeing the exact traffic your users are at a filter level, it is game over. (A good example of this is Facebook - simply seeing the URL www.facebook.com tells you nothing about user activity within the site itself - the content of which changes frequently and appears differently for different users).

Anti-Malware

Out with the old and in with the new - we now call our content filters “Secure Web Gateway” or something similar - why? because now it is common to perform security checks on web traffic. Everything from the validity of the connection itself to scanning for malware in downloaded files, or examining JavaScript for malicious intent. Once again, it isn’t possible to prevent web traffic based on a “body scan” of the received data in pass-by mode, so all anti-malware activity is restricted to simple URL blocklists.

But I chose a pass-by filter in 200x - are you saying I was wrong?

No. Quite the opposite. When you last looked at filtering, many of the issues described earlier were simply not as important as they are today. The web is a fast moving and complex environment, and decisions taken a few years ago may not be the one you'd make today. Additionally, it is very likely you based your purchasing decision on a few of the fantastic benefits pass-by filtering enjoys. For example, pass-by filtering is incredibly easy to deploy. You just connect the filter to a correctly configured switch, and walk away. Because of the low overheads involved, a single filter can process an incredible amount of data - so the hardware costs are generally low, even for a very large user base. In the rest of this document we will look at ways to ease the burden of changing your filter by examining some of the advantages of alternative methods, and solving some of the common problems encountered by those migrating from pass-by solutions.

So Which Deployment Type is Best?

Really, the question is "which is best for me?". The suitability of either transparent, or traditional proxying will depend to a large extent on your existing network. We will consider a few separate factors, and these should enable informed choice. Be aware that some vendors may be able to offer you a solution that combines traditional and transparent filtering proxies in one unit.

Network Size

Larger networks should consider traditional proxies as the favoured solution. A traditional proxy has a higher management overhead, but generally is an easier beast to scale. All solutions are likely to have greater overheads than your existing pass-by solution, and as such may need to operate in a load balanced environment. A pair of dedicated load-balancers can easily serve quite large proxy-farms for even the biggest deployments.
Smaller networks with a couple of hundred endpoints, on the other hand, may benefit from an "all in one" solution, such as a UTM. This is a good place to do transparent proxying if that mode of operation appeals to you.

Ease of Set-Up

If simplicity of deployment is a big deal for you, you might want to consider a transparent proxy. This won't need any configuration client-side, and will be as close to your existing pass-by filtering as you are likely to find.

Of course if you do go for a traditional proxy, it is possible to automate the client set-up. The resources section at the end of this whitepaper will give you a few pointers on this.

Authentication

If differentiating users is a key requirement, traditional proxies generally have the edge. This is because the browser client is aware of the proxy, and as such the proxy can use well-defined standards such as NTLM to ask for authentication. That's not to say authentication is impossible in transparent mode. Many vendors offer solutions in this area - including the tricks that

were used to identify pass-by users. Be sure to research each vendor and find out which identification and authentication methods are available in transparent mode.

Secure or SSL Traffic

In some environments, passing secure traffic unfiltered may be acceptable, or conversely, it may be ok to simply block all SSL access. If this isn't the case, traditional proxies will prove to be the easier choice. Like authentication, SSL control has long been the preserve of the traditional proxy, although more transparent proxy solutions are now offering SSL features. Be aware that there are a number of types of SSL filtering. Some products will offer simple domain level blocking, others will add certificate checking to that, and the most advanced systems will include full inspection of SSL content. What's best for you depends largely upon the security issues you face, although without full inspection you will not be able to scan for malware travelling over secure/encrypted connections.

Control of User Environment

If you have good control of the filtered devices - for example, if they're all Windows PCs which are owned and maintained by your organization, it is easy to run a traditional proxy. You can push the settings in Group Policy, and you're done. On the other hand, if you have a mixed environment with a lot of guests or untrusted computers, a transparent proxy offers an easier answer.

I'd like to move away from Pass-By - How do I sell the benefits to my organization?

Now you've looked at the opportunities ahead, you have hopefully come to a decision on how to proceed. Moving from pass-by filtering may be a bone of contention for some stakeholders in your organization. Few people really enjoy change, so you may need to help them reach the same conclusion you did. Here are some pointers:

True Content Filtering

With real content filtering, you can be sure the quality of your filtering will rise. This will be especially true in risk areas such as anonymizers and sites which change frequently.

Keep up with "Web 2.0"

Scan pages that your users see - not what the filter sees. Understand how your users interact with Facebook, Google and YouTube, the biggest sites on the web, all of which offer rapidly changing content which URL filtering can't touch.

Malware protection

Web downloads can now be scanned for viruses, malware and Trojans, augmenting the desktop anti-virus scanners your users should already have.

More reliable authentication

Different filter levels will be more accurately applied than before, which means managers will get better reports and web traffic can be interpreted in a more meaningful way.

Be part of the future

It's likely that your existing product is coming to the end of its commercial existence - whether the vendor admits it or not. This is an open secret within the content filtering industry.

Resources for Pointing Clients at a Traditional Proxy

Getting previously unconfigured user PCs to access your new filter proxy can be hard. Here's how you might do it

1. Group Policy
In a new or existing Group policy object, applied to your domain, the IP address for your proxy server can be set Under "User Configuration\Windows Settings\Internet Explorer Maintenance\Connection". Double click the "proxy settings" icon and enter the IP address and port number of your proxy server.

2. WPAD
WPAD or "web proxy auto detect" uses DNS to tell your browser where to find a proxy.pac file. Set a DNS record to point to wpad.yourdomain, which should be a web server hosting the proxy.pac file. In this case the proxy.pac file should be named "wpad.dat"

3. DHCP
DHCP option 252 is used to inform a browser of the location of a proxy.pac file. This option can be set in all major DHCP servers. This will work with Internet Explorer if it is set to "automatically detect proxy settings".

4. Get your users to do it!
This works surprisingly often - if you have an area where guest users normally access your network, a poster explaining how to set up the proxy is a great last resort, in case none of the automatic methods worked

5. Offer transparent as a backup
Finally, some vendors will allow you to configure a transparent proxy simultaneously with a traditional proxy. If your network architecture allows this, it might be a good option

Writing Proxy.pac files
Proxy.pac files are useful for "cheap" fault tolerance and load balancing in traditional proxy setups. They also figure in some auto-config modes. A Proxy.pac file is essentially a chunk of JavaScript that tells your browser which proxy to use and when. You can use any of the features of JavaScript to help the broswer "decide" which proxy is appropriate.

For lots of good information on writing your pac files visit:
www.returnproxy.com

A Note on Guardian Web Filters

Guardian web filters can be configured to operate in either traditional proxy mode or transparent proxy mode. Content is filtered dynamically, in real time and users are authenticated seamlessly (single sign-on) without the need to enter additional credentials. Guardian also offers an exceptional level of protection against web-based threats via Sunbelt Software's market-leading VIPRE anti-malware solution. Features such as URL-specific and search string filtering also help Guardian to understand user activity on popular web 2.0 based sites such as YouTube, Facebook and Wikipedia. New features due Q4, 2010 will also facilitate HTTPS/SSL content filtering in transparent mode.

UK + INTERNATIONAL

Smoothwall Ltd
1 John Charles Way
Leeds LS12 6QA
United Kingdom

+44 (0)800 5 999 040 UK
+44 (0)870 1 999 500 International
sales@smoothwall.net
www.smoothwall.net

USA + CANADA

Smoothwall Inc.
6201 Fairview Road, Suite 320
Charlotte, NC 28210-4274
United States of America

1-800-959-3760 US + Canada
1-888-899-9164 Fax
sales@smoothwall.com
www.smoothwall.com