

NETCLEAN WHITEBOX

Child Sexual Abuse Web Site Filtering System

- Technical Description

Introduction

Whitebox from NetClean is a system designed to block access to Web sites containing child sexual abuse images (CSAI) through Internet Service Providers. This document describes the implementation and operation of the system.

The Server

The server is designed as a filtering appliance and uses the BSD Unix variant as its operating system. It is configured with a GUI via a Web browser and has a database consisting of the list of URLs that are to be blocked. It is available in various configurations depending on form factor, performance and redundancy requirements.

Installation

1. The server is installed in a data centre that has direct access to an international Internet gateway. To avoid routing loops it is normally installed within a network that is not part of the target network to be filtered. If this is a problem, then it can be installed within that network but the outgoing web traffic from the filter must be tunnelled through to a device that is connected to a network outside of the target network – it needs to be part of another Autonomous System. Or alternatively, you could make a router in the same Autonomous System not aware of the /32 hostroutes for the traffic to reach the correct recipient and create a tunnel from the filter to this router.
2. A tunnel is set up between a router on the target network that carries the external BGP routing and the Whitebox server. This tunnel can be either GRE or IPIP.
3. The Whitebox server is configured as an external BGP neighbour to the router within the filtered ISP.

Operation

1. The URL list to be blocked is loaded into the Whitebox server. Note that the filter can work down to directories and pages on web sites, should just parts of sites be required to be blocked. For example, if the URL <http://www.badstuff.com> was added to the block list then access to the page <http://www.badstuff.com/goodstuff/good.html> would be blocked. However, if the URL <http://www.badstuff.com/badstuff> was added instead then access to <http://www.badstuff.com/goodstuff/good.html> would not be blocked but <http://www.badstuff.com/badstuff/bad.html> would be.
2. The server checks the IP address of each URL on the list by doing a DNS lookup on it. This is done often, for example at least daily would be recommended as CSAI sites can change their IP addresses often to try to avoid detection.
3. The server then distributes /32 routes via BGP to the target ISP router for the IP addresses obtained in step 2. These routes advertise the Whitebox tunnel IP address as the next hop for these hostroutes.
4. If a user on the target network tries to access a blocked web page, say for example, <http://www.badstuff.com/badstuff/bad.html> and the URL <http://www.badstuff.com/badstuff> was on the URL list then their web request would be routed through the tunnel to the filtering server when that request passed through

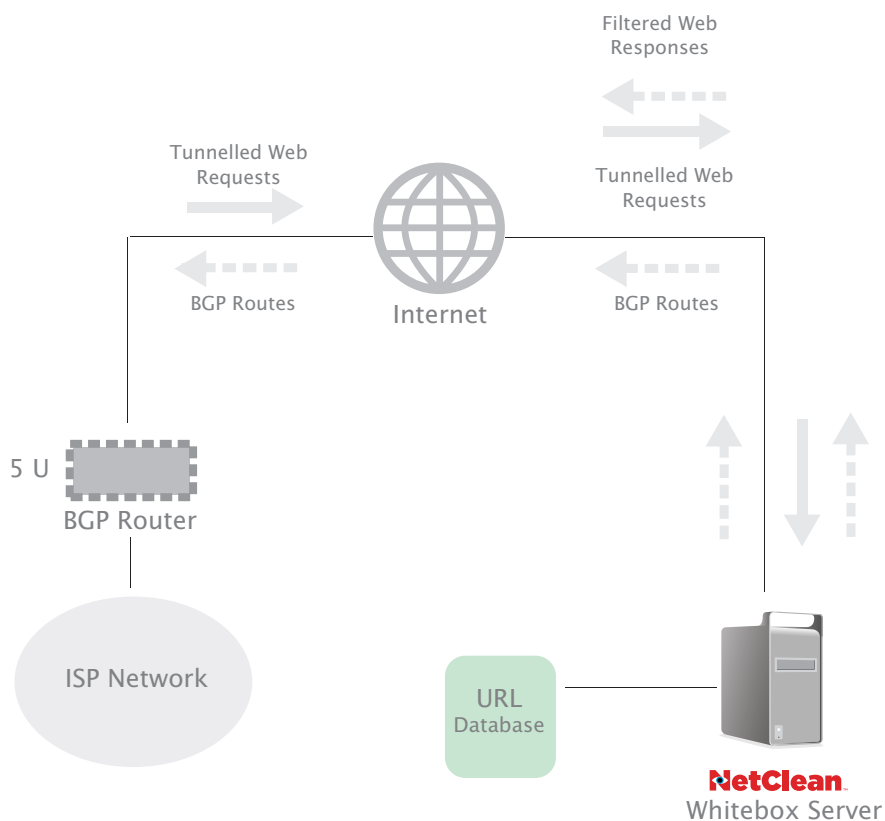
the border router. This is because the route to the IP address for that site will have been sent via BGP through to that router from the server.

5. The Whitebox server then does a URL lookup on the requested URL, `http://www.badstuff.com/badstuff/bad.html`. It gets a match to the URL on the list, `http://www.badstuff.com/badstuff`, and sends a framed redirect to a block page of your choice back to the source IP address requesting the web site, spoofing the tcp-session, thus fulfilling the request.

6. If a user on the target network tries to access a web page that is not blocked, but resides on the same IP address as a web site that is blocked, for example, `http://www.badstuff.com/goodstuff/good.html` and the URL `http://www.badstuff.com/badstuff` was on the URL list then their web request would be routed through the tunnel to the filtering server when that request passed through the border router. This is because the route to the IP address for that site will have been sent via BGP through to that router from the server.

7. The Whitebox server then does a URL lookup on the requested URL, `http://www.badstuff.com/goodstuff/good.html`. It gets NO match to the URL on the list, `http://www.badstuff.com/badstuff`, so the web request gets passed directly to the target web site, thus fulfilling the request. This request must be sent out to an Autonomous System separate from the target network being filtered, to avoid the request looping.

8. Any web site requests from the target network that are going to Web sites hosted on IP ranges not associated with the blocking URL list will route through the network as before so this system has no effect on this traffic whatsoever.



Filtering Capacity

We estimate that one filtering server has the capacity to support up to around 500,000 customers, based on real world results. This is dependent on a number of factors, but mainly on the amount of “clean traffic” that the filter has to handle.

Redundancy and Load Balancing

A high-availability installation on a large network (>500,000 users) could be delivered by using a number of servers.

Multiple Service Provider Support

A significant benefit of the external BGP implementation is that one Whitebox server can support multiple service providers. The number of providers that can be supported by one server is normally only limited by the throughput as we have tested the system with many tunnels. For example, one server could support up to 10 ISPs with 50,000 customers each or 100 ISPs with 5,000 customers each.

More Information

For more information or questions in regard to this document, please contact peter@watchdog.net.nz.

For general information please email info@netclean.com or visit www.netclean.com