List Management Issues When Filtering Using URL Blacklists

- Technology Whitepaper -

Introduction

Watchdog International Ltd has been involved in Internet filtering for many years, including the blocking of web sites on a countrywide basis most often for sites containing images depicting child sexual abuse (CSA). Through this experience, Watchdog has discovered that the importance of a well-managed URL list to ensure an effective service cannot be overstated. This document outlines considerations and recommendations in this area.

List Sources

There are two main sources of URLs for these lists, hotlines and law enforcement intelligence.

Hotlines

A number of countries have chosen to implement hotlines where members of the public can report web sites that they believe contain content that should be blocked. These sites are reviewed and then if the content matches the blocking criteria then the site is usually referred to enforcement authorities to take the site down. In many countries the hosting of child sexual abuse content is illegal so law enforcement can direct the hosting company to remove the content and the owners can be prosecuted. It is only when takedown is not possible within a reasonable time period that the URL will be added to the blacklist. This usually occurs when the site is hosted in a country where hosting CSA is not illegal or where law enforcement is ineffective. An example of one of these hotline organisations is the Internet Watch Foundation (IWF) in the UK.

Law Enforcement Intelligence

Law enforcement organisations that are involved in the policing of crimes involving CSA are able to collate lists of sites that have been accessed by offenders through forensic analysis of their computers. Again, once the URLs are identified, the first step is to investigate where these sites are hosted and to initiate takedown procedures, if they are hosted within their jurisdiction or to pass it on to the relevant authorities that do have jurisdiction. As with the hotline list if takedown is not possible within a reasonable time period then the site is added to the block list. An example of this sort of list is the one used by the Department of Internal Affairs (DIA) Censorship Compliance in New Zealand.

URL Analysis

It is very important that before a URL is added to the list that consideration is given to its effect on any blocking technology using the list. This is because certain URLs can have major effects on both the filters and the networks being filtered. Here are some examples of these:

Examples of Past URL Issues

ACMA Test List containing a "?" within the URL.
During the Australian Federal Government filtering trials a number of YouTube URLs were included in the list and these URLs contained a "?". The filtering technology being tested at the time had a limitation that the filter could not block any URL to the right of a "?" character. This created the problem where all YouTube URLs were blocked by the filter causing problems for a number of users on the network. This problem was temporarily solved by deleting the URLs from the library and adding the URL string after the "?" character to a custom URL keyword library on the filter. The filter vendor, in their next firmware upgrade, then solved this.

ACMA Test List containing a YouTube URL.
During the Australian Federal Government filtering trials a number of YouTube URLs were included in the list as mentioned above. Because YouTube is a site that receives a high amount of traffic from ISPs, adding a URL at this site can seriously affect the performance of hybrid BGP filtering systems such as NetClean White box or BT's Clean feed system. This is due to the fact that these systems route all of the traffic from the target ISP network that is destined for the IP addresses of these sites through the filter, which it is not designed to do. It normally handles the traffic from typical CSA web sites that are, by comparison, very low volume sites. The NetClean WhiteBox system has introduced a whitelist safeguard system that will not block any high traffic site that is added to the list to avoid this happening. As YouTube is a responsible site that has its own takedown policy, content will be removed eliminating the need for blocking.

IWF List containing a Wikipedia Image URL
In early 2009 the IWF list added a URL addressing an image on a Wikipedia page. This caused major problems with customers of some UK ISPs trying to access the site because of a technology limitation in the BT Cleanfeed system and possibly other filtering systems. This is because the Cleanfeed system passes all filtered traffic through a proxy server and that modifies the web site request by replacing the customer's source IP address with its own one. The Wikipedia site now suddenly had a large amount of traffic accessing it from one IP address, the one belonging to the proxy server at the ISP. This appeared to them as an attack so the IP address was blocked, thus blocking all access to all parts of Wikipedia for that ISP. The IWF very quickly removed the URL from the list to solve the problem and admitted that it was an error on their part. Fortunately, filtering systems using the NetClean WhiteBox technology do not use proxy servers and so do not have this Wikipedia URL problem.

Examples of Past URL Issues

ACMA Test List - very long URLs
During the Australian Federal Government filtering trials, a very long URL (257 characters) was included to the filtering list. The technology being trialled had a maximum URL length of 200 characters so any URL exceeding that length would bring up an error when the list was imported. Fortunately this URL redirected to another shorter one that was added to the list, instead of the original one, and this solved the problem.

List Management Considerations

Filtering Technologies being used to implement blocking of the list.
For example, DNS poisoning systems can only work at the whole domain level so parts of a web site cannot be blocked. For this reason lists for use with DNS poisoning filters should just contain the root domain name. Adding domains to the list that contain material that should not be blocked as well as content that should, is risky as all content on that domain will be blocked. It is for this reason that some regions such as the UK use URL filtering instead of DNS poisoning. URL filtering blocks the required parts of domains without blocking the whole site.

The examples above highlight some of the issues that can be introduced when the list is used for filtering. The issues that must be considered for every URL added to the list include:

- URL length
- URLs for high traffic sites such as YouTube
- Special characters within the URL

List Revision
Does the Blacklist still contain URLs with content that should be blocked? A regular revision of the list ensures that URLs still contain the content that matches the blocking criteria.

Is the URL a repeatable one and not just a snapshot of an individual session on a browser? URLs that are reported to hotlines come directly from the address line in a browser. It is not always possible to access the material when this URL is put into a new browser session. The accessibility of all URLs need to be checked in new browser session for their validity before adding them to a blacklist.

List Management Considerations

List Revision continued
Takedown possibility - The first priority of the list organization is to issue the 'take down' of the site in question. A site should not be added to a list if it can taken down within a reasonable timeframe.

Because URL lists contain the addresses of sites that are illegal to access, protection of the list is an important issue. Some URL lists have already been leaked on the Internet so this is a real issue to be considered. There are a number of options available to address this as follows:

List Security
Automatic List Distribution
This is where a server is set up in a secure facility and it distributes the list out to the ISPs automatically when required. Encryption and authentication can be used to ensure that the list is secure during transmission and also that the target server is the correct one. This system requires the target servers to communicate via a standard protocol and also to accept the list in a standard format. These requirements are simple to meet if every ISP is using the same filtering system and much more complex if they are using a range of different systems.

Managed Filtering Service
Using a managed filtering service where the list is kept secure on the servers providing the service is the most secure method, as the list never needs to leave the secure server. The NetClean WhiteBox hosted system is currently the only managed service for the filtering of illegal content available at present. It is currently being used by ISPs in New Zealand managed by the Department of Internal Affairs as well as in the UK, managed by NetClean who host the service and use the URL list from the IWF.

Password-Secured File
This system is used to ensure that the list can only be accessed by those who are given the password and these people would usually be those that need to install it on the filtering servers. However, there has to be a trust relationship between the list managers and the recipients to ensure that those who have access to the password do not reveal the password to others or distribute the list in an insecure manner. This method would have to be used in conjunction with a non-disclosure agreement to ensure that if the list was leaked that the list managers have legal redress.

List Organisation	Industry Organisation	List Source	Blocking Technologies and Region
Internet Watch Foundation (IWF)	ISPA	Internet Watch Foundation Hotline	BT's Cleanfeed, other ISP Filters including Optenet WOLF, NetClean WhiteBox in UK and Europe
Swedish Police		ECPAT Sweden Hotline	Majority of Swedish ISPs using DNS Poisoning
Norwegian Police (KRIPOS)		Save the Children Norway	Majority of Norwegian ISPs using DNS Poisoning
Danish Police	Danish Internet Providers Association	Red Barnet Danish Police Hotline	Majority of Danish ISPs using DNS Poisoning
Finnish Police (National Bureau of Investigation)		Save the Children Finland	Majority of Finnish ISPs using DNS Poisoning
The Swiss Coordination Unit for Cybercrime Control (CYCOS)		Originally Danish Police and hotline	Majority of Swiss ISPs using DNS Poisoning
Cybertip.ca as part of the Canadian Coalition Against Internet Child Exploitation (CCAICE)		RCMP's National Child Exploitation Coordination Centre (NCECC) and Department of Justice	Project Cleanfeed Canada where 6 major ISPs use a modified version of BT's Cleanfeed internal BGP filtering system
"Stop-it" managed by Save the Children Italy		"Centro nazionale per il contrasto della pedopornografia" (The National Centre against Child Pornography)	Some Italian ISPs using DNS poisoning
Voluntary Self Control Multimedia Service Providers (FSM e.V.)		Federal Criminal Police (BKA)	Some German ISPs using DNS poisoning
Meldpunt Kinderporno, the Dutch Hotline on child pornography	NLIP (the Dutch Association of Internet Providers)	Dutch National Police Forces (KLPD)	One Dutch ISP (UPC) using DNS Poisoning
Department of Internal Affairs (DIA) and ECPAT		DIA	Most New Zealand ISPs shortly to be filtering using NetClean WhiteBox system hosted and managed by DIA

Recommendations

List Sharing
Due to the global presence of child sexual abuse content on the Internet, it can be accessed from anywhere. For this reason a site that is added to the blocking list in one country will be accessible in all other countries. If list providers shared their lists then all of them would have more effective filtering. List authorities are reluctant to share lists partly due to differing legislation and definitions around child sexual abuse.

Reviewing of Filtering Technologies
It would be prudent for authorities that manage lists to periodically review filtering technologies used to implement the blocking of web sites. Many countries are using DNS poisoning because it was easy to implement and cost effective but it has list limitations and is easy to bypass. Newer technologies are available that are more secure and effective but not much more expensive to implement.

Conclusions

The creation and management of a URL list to be used in an effective regional blocking service involves addressing a number of important issues including the filtering technologies being used, the security of the list and the inclusion of 'high traffic' web sites. There are a number of organisations around the world that have experience in this area such as the Internet Watch Foundation in the UK and the Department of Internal Affairs in New Zealand.

Contact Watchdog

For More Information visit www.watchdoginternational.net
Or email us at filtering@watchdoginternational.net
Phone: +64 9 424 9060