URL Filtering
Feature Overview and Configuration Guide

Introduction

URL filtering blocks all HTTP access to a list of websites or portions of web sites. You can specify a short list of websites to block (up to 1000 blacklist and 1000 whitelist rules), and/or subscribe to the blacklist service offered by Kaspersky.

- A **white list** is a list of URLs that are known to comply with organisational policies.
- A **black list** is a list of URLs that are known to violate organisational policies.
- **Kaspersky** is a subscription-based service that classifies websites among dozens of pre-defined categories of content that will not comply with some organisations' policies.

  If you subscribe to the Kaspersky service, you can create additional blacklists to block extra URLs or whitelists to allow URLs that the Kaspersky service blocks. For more information on Kaspersky, visit their website at: [https://support.kaspersky.com](https://support.kaspersky.com)

URL filtering provides a fast efficient (stream-based) method of blocking web traffic from locations that are known to be undesirable. It acts on a global basis and should be used when traffic is to be blocked for everyone on the blacklist, or allowed for selective URLs as configured in a whitelist.

This contrasts with Web Control, which has finer grained control as URLs are proxied and categorised and access to websites are controllable on a per category basis. And since this service is proxy based there is increased latency compared to the (stream based) URL filtering service.

It is possible to use Web Control and URL Filtering at the same time. Connections must be permitted by both URL Filtering and Web Control in order to be allowed through the device. A block action in either feature will cause a failure to load the web page.
Products and software version that apply to this guide

This guide applies to the following AlliedWare Plus™ products, running version 5.4.6 or later.

- AR3050S
- AR4050S
- AR2050V (user configured lists only supported)

Feature support may change in later software versions. For the latest information, see the following documents:

- The product’s Datasheet
- The AlliedWare Plus Datasheet
- The product’s Command Reference

These documents are available from the above links on our website at alliedtelesis.com.

Feature support may change in later software versions. For the latest information, see the above documents.
How Does URL Filtering Work?

URL filtering works by sniffing all traffic as it traverses the AR-Series Firewall and detecting the HTTP transactions that are taking place. These transactions are then processed, and when an HTTP Request is detected, the URL in question is compared against the whitelists (if any) and blacklists configured.

- If a whitelist match is found, the traffic will not be blocked (and logged if configured to do so).
- If a blacklist match is found, the request will be dropped (and logged if configured to do so) - it will not be forwarded to the destination.
- If neither white nor blacklist matches are found, the traffic will not be blocked.
- Pattern checking stops as soon as a match is found. So if traffic matches any configured whitelist, then it will be allowed though the device. Or if traffic matches any configured blacklist then it will immediately be blocked. That same traffic will not be subsequently checked against additional whitelists or backlists.
How to Use URL Filtering

To use URL filtering, you can either use:

- a blacklist provided by Kaspersky
- custom lists (black/white)
- a combination of custom and Kaspersky lists.

Creating custom lists

A custom list is an ASCII formatted text file containing zero or more single-line pattern matches.

For example, the content of a text file named `blacklist-example.txt`, consisting of three patterns to match, (listed line-by-line) could look like this:

```plaintext
example.net/viruses/*
*/viruses/*
bad_url.com
```

URL pattern matches listed within the text file may take two forms:

- either a base domain, which will match all content of that domain, and all content of sub-domains:
  ```plaintext
  example.com
  ```

- or a wild-card match, where an asterisk will match zero or more characters in a URL:
  ```plaintext
  example.net/viruses/*
  */viruses/*
  ```

Once this list is available to the system (stored in Flash, USB, or on an SD card), the configuration to enable URL filtering is straight forward, as described below in the section "Configuring URL filtering" on page 8.
Details of the content of custom lists

A custom list is an ASCII formatted text file containing zero or more single-line pattern matches. So far, we have looked at the general syntax of the entries in these files. Here we look in more detail at the rules governing the content of these files:

- There is no ordering or precedence for patterns in the file.
- Spaces in the pattern are not allowed.
- The wildcard, asterisk ‘*’ can be used in the pattern to indicate a match on zero or more characters.
- If there are no ‘/’ or ‘*’ characters present, then all content of the domain is blocked.
- "Match everything" patterns are not allowed (e.g. ‘*’ or ‘*/’).
- Empty or comment lines (starting with ‘#’ or ‘;’) are ignored.
- The “www.” prefix should not be included in the pattern. However patterns and URLs are “normalized” before matching. More specifically;
  - The ‘www.’ prefix and authentication prefix ‘login:<password>@’, which may pre-pend a URL are automatically stripped from the URL before pattern matching.
  - Patterns are converted to lower case.

The table below describes how the pattern *mysite.com/ is matched (Blocked URLs) or not matched (Non-blocked URLs) for a blacklist.

Table 1: A pattern matching example with explanations.

<table>
<thead>
<tr>
<th>THIS PATTERN</th>
<th>BLOCKS THE URLs</th>
<th>NON-BLOCKED URLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>*mysite.com/</td>
<td>mmysub.mysite.com</td>
<td>mmysub.mysite.com/mypage</td>
</tr>
<tr>
<td></td>
<td><a href="http://www.mysite.com">www.mysite.com</a></td>
<td></td>
</tr>
</tbody>
</table>

Pattern matching explanations

- *mysub.mysite.com* is a match (and is therefore blocked ) because:
  - The wildcard, asterix ‘*’ matches the prepended text ‘mysub’ in the URL, and the remaining text in the URL matches the pattern.

- www.mysite.com is a match because:
  - The “www.” prefix is stripped off prior to matching, and the remaining text in the URL matches the pattern.

- *mysub.mysite.com/mypage* is not a match (and is therefore non-blocked) because:
  - The text ‘mypage’ in the URL is not part of the pattern.

The following table lists a series of blacklisted ‘domain and string pattern’ match criteria, and examples of URLs that would or would not be matched by these criteria.
<table>
<thead>
<tr>
<th>PATTERN</th>
<th>BLOCKED URLS</th>
<th>NON-BLOCKED URLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>com</td>
<td><a href="http://www.mydotcomurl.com">www.mydotcomurl.com</a></td>
<td>myausurl.com.au</td>
</tr>
<tr>
<td>ru</td>
<td>myrussian.pp.ru</td>
<td>myfakerussian.ru.org</td>
</tr>
<tr>
<td><em>z</em></td>
<td>faz.com</td>
<td></td>
</tr>
<tr>
<td></td>
<td>auzzi.id.au</td>
<td></td>
</tr>
<tr>
<td></td>
<td>zulu.com</td>
<td></td>
</tr>
<tr>
<td></td>
<td>me.kiwi.nz</td>
<td></td>
</tr>
<tr>
<td></td>
<td>fish.com/folder1/file.gz</td>
<td></td>
</tr>
<tr>
<td></td>
<td><a href="http://www.google.co.nz/search?client=ubuntu&amp;channel=fs&amp;q=ziare&amp;ie=utf-8&amp;oe=utf-8&amp;gfe_rd=cr&amp;ei=ZfKWVqgtk5PABN6YtqgD">www.google.co.nz/search?client=ubuntu&amp;channel=fs&amp;q=ziare&amp;ie=utf-8&amp;oe=utf-8&amp;gfe_rd=cr&amp;ei=ZfKWVqgtk5PABN6YtqgD</a></td>
<td></td>
</tr>
<tr>
<td>*mysite.com/</td>
<td>mysub.mysite.com</td>
<td>mysub.mysite.com/mypage</td>
</tr>
<tr>
<td></td>
<td><a href="http://www.mysite.com">www.mysite.com</a></td>
<td></td>
</tr>
<tr>
<td>mysite.com/*</td>
<td><a href="http://www.mysite.com/mypage.html">www.mysite.com/mypage.html</a></td>
<td>mysub.mysite.com/mypage</td>
</tr>
<tr>
<td></td>
<td><a href="http://www.mysite.com">www.mysite.com</a></td>
<td>mysub.mysite.com/mypage</td>
</tr>
<tr>
<td><em>mysite.com</em></td>
<td>mypage.mysite.com.au</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mysite.com.au</td>
<td></td>
</tr>
<tr>
<td></td>
<td><a href="http://www.mysite.com">www.mysite.com</a></td>
<td></td>
</tr>
<tr>
<td></td>
<td>mypage.mysite.com/folder/file.txt</td>
<td></td>
</tr>
<tr>
<td></td>
<td>somescrpt.sc?mysite.com.au</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mysite.com/folder/filename-bad.exe</td>
<td>search-engine.com/search?q=mysite.com/folder/filename.exe</td>
</tr>
<tr>
<td></td>
<td>mysite.com/folder/filename-bad/file.exe</td>
<td>mysite.com/subdomain/file.exe</td>
</tr>
<tr>
<td></td>
<td>mysite.com/subdomain/filename.exe1</td>
<td>mysite.com/subdomain/filename.html</td>
</tr>
</tbody>
</table>
Limits

URL filtering is limited to 1000 custom whitelist and 1000 custom blacklist rules, spread over any number of list files.

Configuring URL filtering

URL filtering is turned on by configuring a whitelist that uses a custom file, a blacklist that uses a custom file, or blacklisting that uses the Kaspersky service.

1. To add a whitelist that uses a custom file (that is stored on USB, for example) and then enable URL filtering, use the commands:

   `awplus#configure terminal`  
   `awplus(config)#url-filter`  
   `awplus(config-url-filter)#whitelist usb:/my_whitelist.txt`  
   `awplus(config-url-filter)#protect`

2. To add a blacklist that uses a custom file (that is stored on Flash, for example) and then enable URL filtering, use the commands:

   `awplus#configure terminal`  
   `awplus(config)#url-filter`  
   `awplus(config-url-filter)#blacklist flash:/blacklist-example.txt`  
   `awplus(config-url-filter)#protect`

3. To add a blacklist provided by Kaspersky and then enable URL filtering, use the commands:

   `awplus#configure terminal`  
   `awplus(config)#url-filter`  
   `awplus(config-url-filter)#provider kaspersky`  
   `awplus(config-url-filter)#protect`

   - To check that Kaspersky is active, enter the command `show url-filter`:

     ```
     awplus#show url-filter
     Status:     Enabled (Loading)
     Provider:   Kaspersky
     Status:     Enabled
     Resource version:  not set
     Update interval:  1 hour
     Blacklist entries:  -
     Custom blacklists  Entries
     blacklist-example.txt  3
     Custom whitelists  Entries
     ```

   - Invalid entries in URL filter lists are ignored (not loaded).
   - Expiry of the Kaspersky URL Filtering Subscription License will cause URL filtering to reload without a Kaspersky blacklist.
Using multiple white and black lists

The AR-Series Firewall supports pattern checking against multiple whitelists and multiple blacklists.

Multiple custom whitelists or blacklists can be configured and checked as follows:

awplus(config)#url-filter
awplus(config-url-filter)#blacklist blacklist1.txt
awplus(config-url-filter)#blacklist blacklist2.txt
awplus(config-url-filter)#blacklist blacklist3.txt
awplus(config-url-filter)#whitelist whitelist1.txt
awplus(config-url-filter)#whitelist whitelist2.txt
awplus(config-url-filter)#whitelist whitelist3.txt
awplus(config-url-filter)#protect

You can check the configuration using the `show url-filter`, `show running-config url-filter` and `dir` commands:

```
Awplus#show url-filter
Status:    Enabled (Active)
Provider:  not set
Custom blacklists   Entries
  blacklist1.txt       18
  blacklist2.txt       23
  blacklist3.txt       39
Custom whitelists   Entries
  whitelist1.txt       11
  whitelist2.txt       26
  whitelist3.txt       33
```

```
Awplus#show running-config url-filter
url-filter
  blacklist blacklist1.txt
  blacklist blacklist2.txt
  blacklist blacklist3.txt
  whitelist whitelist1.txt
  whitelist whitelist2.txt
  whitelist whitelist3.txt
  protect
!
```

```
Awplus#dir
  107 -rw- May 11 2016 04:52:44  whitelist1.txt
  229 -rw- May 11 2016 04:52:39  whitelist2.txt
  372 -rw- May 11 2016 04:51:50  blacklist3.txt
  202 -rw- May 11 2016 04:51:38  blacklist2.txt
  170 -rw- May 11 2016 04:51:31  blacklist1.txt
```
Rules for the processing of lists

The order of processing of lists is:

- First - Whitelists
- Second - Custom blacklists
- Third - Kaspersky-provided blacklists

The matching logic is that as soon as a URL matches an entry in a list that it is being compared against, then comparing stops and the relevant action (allow, if the match occurs in a whitelist, or deny if the match occurs in a blacklist) is taken.

Because Whitelist matching precedes blacklist matching, you can use custom whitelists to override any corresponding blacklist entries. An HTTP request that includes a URL that matches an entry in a whitelist will be permitted immediately, and the URL will not be matched against an blacklists.

So, if websites you actually want to access are being blocked by the Kaspersky blacklist, or some subsection of an otherwise dangerous site is desirable, a whitelist may be created.

For this example, the example.net/viruses/research folder contains information that is needed within the otherwise completely blocked site.

This can be allowed by creating a whitelist file named whitelist-example.txt in Flash memory, with the contents:

example.net/viruses/research/*

And configuring it as follows:

```
awplus#configure terminal
awplus(config)#url-filter
awplus(config-url-filter)#whitelist whitelist-example.txt
awplus(config-url-filter)#protect
```

This whitelist will be processed prior to the blacklist, and will allow matching traffic through.

Updating lists

Updating the Kaspersky blacklist

When subscribed to the Kaspersky URL Filter service, updates to the Kaspersky blacklist will be made available. By default URL filtering checks for updates to the Kaspersky blacklist every hour.

You can configure the update interval via the `update-interval` command in `url-filter` configuration mode. The update process is managed by the Update Manager utility.
You can see the update status in two show command outputs: `show url-filter` and `show resource`.

```
awplus#show url-filter
Status: Enabled (Loading)
Provider: Kaspersky
Status: Enabled
Resource version: urlfilter_kaspersky_stream_v48
Update interval: 1 hour
Blacklist entries: 63457
...
```

```
awplus#show resource
+-------------+----------+----------+--------------------------+--------------------------+--------------------------+
<table>
<thead>
<tr>
<th>Resource Name</th>
<th>Status</th>
<th>Version</th>
<th>Interval</th>
<th>Last Download</th>
<th>Next Download</th>
</tr>
</thead>
</table>
+-------------+----------+----------+--------------------------+--------------------------+--------------------------+
```

When the Update Manager finds a new version is available, it downloads and instructs URL Filter to start using the new blacklist. An update check can be manually initiated with the `update urlfilter_kaspersky_stream now` or `update all now` commands.

**Updating a user-defined blacklist or whitelist**

You can modify blacklist and whitelist files that you have created. Once you have completed all the desired changes, use the `url-filter reload custom-lists` command to reload the modified files.

When a new blacklist or whitelist is configured and URL filter is already enabled, it automatically starts using the new file.
Monitoring URL Filtering

The **show url-filter** command displays a summary of the state of URL filtering, including the provider state, and counts of entries in each provided list. Any lists that contain too many entries to load will be noted here.

```plaintext
awplus#show url-filter
Status: Enabled (Active)
Provider: Kaspersky
  Status: Enabled
Resource version: not set
Update interval: 1 hour
Blacklist entries: -
Custom blacklists Entries
  blacklist-example.txt 3
Custom whitelists Entries
  whitelist-example.txt 1
```

Logging

By default logging is performed when there are:

- Blacklist and whitelist hits—logged at informational level.
- Invalid match criteria, detected while loading third party and custom blacklist and whitelist files—logged at error level.
- Missing configured custom blacklist and/or whitelist files, while starting/restarting the feature—logged at warning level.

For more information about logging, please refer to the Logging Feature Overview and Configuration Guide.