

From EULER Project

PmWiki: Blocklist

The block list is one of a number of security measures that can be taken to protect your wiki from unwelcome postings.

Unfortunately, the open-editability of many wiki systems often makes them attractive targets for "link spam" or "wikispam", in which links are added to pages in an effort to increase search engine rankings or drive traffic to other sites. Also, many link spammers have developed automated systems to locate sites that accept visitor input and attempt to flood the site with unwanted links. Also, and harder to deal with, is just plain wiki vandalism where nonsense changes are made, often replacing entire pages.

By far the best countermeasure against wikispam is to restrict editing through the use of passwords (see Passwords and Passwords Admin). Experience has shown that passwords can be effective even if the password is widely known, and even if the password is publicly available on the site itself. However, there are many cases where passwording may be an impediment, so these will generally want to use some form of blocklist.

Blocklist basics

A *blocklist* is a list of IP addresses, phrases, and expressions which are prevented from being added into pages on the website. PmWiki is distributed with a built-in blocklisting capability; blocklists can be enabled by adding the following line to *local/config.php*:

```
$EnableBlocklist = 1;
```

This tells PmWiki to scan the SiteAdmin.Blocklist? page and the "SiteAdmin.Blocklist-Farm" page (and possibly other pages -- see below) looking for phrases and IP addresses to be excluded from posting to the site.

Blocking by word or phrase

The simplest form of block is simply a line containing "block:" followed by a word or phrase to be excluded from postings. For example, a line like

```
block:spam.com
```

in SiteAdmin.Blocklist will block any posts containing the string "spam.com" (case-insensitive) anywhere in the post.

Blocking by IP address

Sometimes we wish to restrict posts coming from particular addresses or address ranges that are known as sources of wikispam. If a blocklist page contains IP addresses of the form "a.b.c.d" or "a.b.c.*", then any posts coming from that address or range will be blocked.

To find an author's IP address, try hovering the mouse over the author name in the page history for a page.

Blocking by regular expression or pattern

Blocking on simple words can sometimes pose difficulties; for example, a simple `block:cial` entry will also block the word "specialist". For these cases it's often helpful to use a regular expression, as in:

```
block:/\bcial\b/
```

This says to block "cial" only if it doesn't occur in the middle of a larger word. The leading slash (/) after "block:" tells PmWiki to use a regular expression match instead of a simple string match. (Blocklist uses PCRE or "Perl Compatible Regular Expressions"; see <http://php.net/manual/en/ref.pcre.php> for more information.)

Regular expression to block 'href'

If you want to block 'href', you can use the following markup:

```
block:/[^\w\\]href\b/
```

which blocks 'href', but neither '\href' nor 'toughhref'.

The regular expression can be interpreted as follows: Match any character that is **neither** a word character **nor** a '\', followed by href which ends in a word boundary.

Letting authors know why they've been blocked

By default, blocklist only tells an author that a particular edit has been blocked, but doesn't give a specific reason for the blocking (e.g., the offending phrase). Setting the following in a local customization file will also provide the reasons for the block:

```
$EnableWhyBlocked = 1;
```

Managing multiple blocklists

PmWiki allows blocklist entries to come from multiple pages by setting the `$BlocklistPages` variable. By default `$BlocklistPages` is set to "SiteAdmin.Blocklist", as well as any automatically downloaded blocklists as described below. PmWiki will use all entries in all the blocklists for filtering wikispam. Setting a value of `$BlocklistPages` changes the default:

```
$BlocklistPages = array('Main.Blocklist', 'PmWiki.Blocklist');
```

The order of blocklists really doesn't matter -- all of the blocklist pages ultimately get used, and the `unblock:` entries are processed after all of the blocklist pages have been loaded.

Automatically downloaded blocklists

Maintaining blocklists is relatively easy to do, but can become tedious over time. Several groups have formed and maintain "shared blocklists", where a common blocklist is made available to all. PmWiki's blocklist capability has built-in features for automatically downloading and updating such shared blocklists.

If you're just in a hurry to make use of some standard blocklists, make the following setting in *local/config.php*:

```
$EnableBlocklist = 10;
```

This tells PmWiki to not only enable blocklists on the site, but to also configure itself to automatically retrieve and maintain local copies of well-known blocklists such as chongqed.org([approve sites](#)) and [MoinMaster](#)([approve sites](#)). These local copies will be saved in SiteAdmin.Blocklist-Chongqed and SiteAdmin.Blocklist-MoinMaster and refreshed once per day (as determined by the value of [\\$BlocklistDownloadRefresh](#)).

To automatically retrieve the [SiteAdmin.Blocklist?](#) page used at pmwiki.org, add the following setting in *local/config.php*:

```
$BlocklistDownload["$SiteAdminGroup.Blocklist-PmWiki"] = array('format'
=> 'pmwiki');
```

Ignoring specific entries in a blocklist (unblock)

When using a large master blocklist or blocklists automatically refreshed from external sites, it may be that some entries in the blocklists are inappropriate or overeager and block legitimate content. In this case a wikiadministrator can use "unblock" in a blocklist page to ignore an entry from the blocklist. For example, to allow "spam.com" even if another blocklist has a block entry for it:

```
unblock:spam.com
```

In order for unblocking to work the phrase or pattern following "unblock:" must be *exactly* the same as the original.

Permissions on blocklist pages

In general, an [administrator](#) will want to edit-protect the SiteAdmin.Blocklist and any other blocklist pages to prevent arbitrary changes to the blocklist (see [Passwords](#)). Since most pages in the SiteAdmin.* group are edit-protected by default anyway, this usually isn't a problem.

Administrators may also wish to read-protect the various blocklist pages so that others do not know the exact phrases and/or IP addresses that are being blocked. (By their nature blocklists tend to contain phrases or terms that may be offensive or inappropriate to some.)

Any pages created via automatic download (see above) are automatically locked against viewing except by administrators.

```
administrators (intermediate)
```

Detailed configuration of automatically downloaded blocklists

Automatic downloading of blocklist information is controlled by the [\\$BlocklistDownload](#) array. An entry for MoinMaster might look like:

```
$BlocklistDownload["$SiteAdminGroup.Blocklist-MoinMaster"] = array(
'url' => 'http://moinmaster.wikiwikiweb.de/BadContent?action=raw(approve sites)',
'format' => 'regex',
'refresh' => 86400);
```

This says to download the blocklist data from the given url into the SiteAdmin.Blocklist-MoinMaster page, that the entries in the blocklist are regular expressions, and to refresh the information every 86,400 seconds (one day).

If 'refresh' is omitted, then the page will be refreshed at the time interval given by `$BlocklistDownloadRefresh` (default one day). If 'format' is omitted, the page is assumed to have PmWiki-formatted entries as described above. If 'url' is omitted, then the blocklist information is downloaded from a standard location on the pmwiki.org site.

To force a refresh of an automatically downloaded blocklist, simply delete the existing page -- a new version will be installed upon the next blocklist scan. Blocklist pages are checked for download in response to any `?action=edit` request.

If you are specifying your Blocklist-Pages in the config.php you have to specify the automatically updated pages too, else they won't be updated or created even if you use `$EnableBlocklist = 10`; .

Farm-wide blocklist

A blocklist can be applied farm-wide (see [SharedPages](#)). After these pages are created they can be moved into the farm *shared.d/* directory:

Blocklist Variables

The following variables help control the configuration and operation of blocklists:

\$EnableBlocklist

If set to a non-zero value, then blocklists are enabled on the site. If set to a value of ten or higher, then add entries for automatic downloads of standard blocklists.

`$EnableBlocklist = 1`; # enable blocklists

`$EnableBlocklist = 10`; # auto-configure standard blocklists

\$EnableWhyBlocked

By default, authors are not told which particular phrases or IP addresses are causing a particular post to be blocked; setting `$EnableWhyBlocked` to 1 provides this information.

`$EnableWhyBlocked = 1`; # give reasons for blocking

\$BlocklistPages

An array of pages to be checked for blocklist entries. The elements of the array may contain [page variables](#). Defaults to "Site.Blocklist", plus any other automatically downloaded blocklist pages.

\$BlocklistMessageFmt

The message to provide the author whenever a post has been blocked.

\$BlockedMessagesFmt

If `$EnableWhyBlocked` is set, defines the text to use for each type of block being performed.

Currently only 'ip' and 'text' are recognized.

`BlockedMessagesFmt['ip'] = "IP address blocked from posting: ";`

`$BlockedMessagesFmt['text'] = "Text blocked from posting: ";`

\$BlocklistDownload

An array of automatically-downloaded blocklists. The keys of the array are the pages in which the blocklists should be stored, the values contain the url, format, and refresh interval for the downloaded blocklist.

```
# Download the MoinMaster blocklist every twelve hours
```

```
$BlocklistDownload["$SiteAdminGroup.Blocklist-MoinMaster"] = array(  
  'url' => 'http://moinmaster.wikiwikiweb.de/BadContent?action=raw',  
  'format' => 'regex',  
  'refresh' => 43200);  
# Download a shared blocklist from pmwiki.org every day  
$BlocklistDownload["$SiteAdminGroup.Blocklist-Shared"] = array(  
  'format' => 'pmwiki');
```

\$BlocklistDownloadRefresh

The default refresh interval for any \$BlocklistDownload entries that don't explicitly specify a 'refresh' value.

perform automatic downloads once per week by default

\$BlocklistDownloadRefresh = 86400 * 7;

\$BlocklistDownloadFmt

The format to use when saving automatically downloaded blocklists.

\$EnableBlocklistImmediate

Some cookbook recipes update pages with author input but don't use the built-in data posting routines.

If \$EnableBlocklistImmediate is set (default) and the current action is listed in

\$BlocklistActions (below), then an immediate blocklist scan is performed on the incoming text.

\$BlocklistActions

A list of actions for which immediate blocklist checks should be performed (see

\$EnableBlocklistImmediate above).

perform immediate checks for ?action=comment

\$BlocklistActions['comment'] = 1;

perform immediate checks for ?action=postdata

\$BlocklistActions['postdata'] = 1;

Retrieved from <https://www-sop.inria.fr/mascotte/EULER/wiki/pmwiki.php/PmWiki/Blocklist>

Page last modified on February 14, 2009, at 03:03 PM