Bot Monitor Plugin

—- plugin —- description: Keep an eye on bot traffic to your wiki. Now with Captcha option! author : Sascha Leib email : ad@hominem.info type : admin, action lastupdate : 2026-01-04 compatible : Librarian depends : conflicts : similar : tags : !experimental, admin, analytics, bots, data, log, logging, management, monitoring, security, statistics, tracking, captcha, blocking

downloadurl: https://github.com/saschaleib/dokuwiki-plugin-botmon/archive/refs/heads/main.zip bugtracker : https://github.com/saschaleib/dokuwiki-plugin-botmon/issues sourcerepo : https://github.com/saschaleib/dokuwiki-plugin-botmon donationurl:

screenshot_img : https://leib.be/sascha/_media/projects/dokuwiki/botmon/preview.jpg

---

About

This plugin measures visits and page loads to the wiki and makes an informed assessment how much of the traffic is caused by bots.

This is very much still «work in progress», and the assessment should be viewed critically, but it is a first attempt to even understand the problem. Once this is clearer, some ideas on how to solve (at least part of) it may follow.

Important: This plugin is still in *experimental* state, that mean some things may not work as expected and you may need to manually intervene at some points. Please make sure to read this page as well as the documentation thoroughly, in order to avoid problems.

Important: This plugin is designed to minimize the load on the server as much as possible. As a result, all processing is done in the admin's browser (using JavaScript). This works well for smaller sites (up to ca. 5000 to 10000 page views/day, depending on your machine), but can overload your browser on large sites. For these, you may want to use a different solution.

Installation

Search and install the plugin using the Extension Manager. Refer to Plugins on how to install plugins manually.

How does it work?

After installation, the plugin will start to collect data about wiki visits into its own «logs» directory. For every day (measured in UTC time) there is one file each with the following extensions:

- *<date>*`.src.txt` – each page load is logged here via PHP on the server side.
- *<date>*`.log.txt` – after page load is complete, the same event is logged again via JavaScript.
- *<date>*`.tck.txt` – as long as a page is open, every minute another entry is logged here (via JS).
- *<date>*`.captcha.txt` – log from the captcha solution to monitor its effectiveness (if enabled)

In combination, these log files allow to monitor a variety of bot activity – for example, bots that don't execute JS will be logged in the first file, but not the others. A bot that executes JS, but does not read the page like a human visitor would, will appear in the first two files, but not in the third, etc.

At the moment, the admin tool can only view the current date log files – this is the scenario that you check in the evening how your wiki is doing. A more comprehensive analysis (using a database and long-term data retention) is planned, but not implemented yet.

There is also an experimental Captcha solution implemented. However, it is a good idea to first monitor the site for a while to really understand the bot traffic and possibly try other strategies (explained in detail in the documentation!) first.

Configuration and Settings

The plugin has a number of configuration options which can be used to fine-tune the way it works:

- "Which data to show in the 'Latest tab" – select if the admin interface should show data from the ongoing day (best to monitor how changes affect the traffic), or the last full day (best for comparing traffic from day to day).
- "Combine visits from known IP-ranges into one entry" – if selected, separate visits that come from a known bot network, will be combined into a single «visit». Makes it easier to monitor these networks.
- "Add GeoIP Information" – If your server has the PHP GeoIP extension installed, you can enable country information in your log files. Note that this is not always very accurate.
- "Enable Captcha" – use a Captcha protection for your site. Options are:
  - "Disabled" – no Captcha is shown.
  - "Captcha with Lorem Ipsum text" – Use a captcha, show the traditional "Lorem ipsum" placeholder text to bots which can not solve the captcha.
  - "Captcha with Dada placeholder text" – Use a captcha, show an automatically generated Dadaist-style placeholder text to bots.
- "Captcha Seed" – Enter a random number here to use as "salt" for the encrypted cookie. IMPORTANT: Never leave this to the default value!
- "Automatically solve the Captcha, if …" – Select conditions where the captcha can automatically be solved.
  - "Client and page languages match" – if the client has a browser language that matches the language of the page requested (note: better don't enable this for English wikis!)

In addition to this, there are also several configuration files which can be modified. See the Developer information in the plugin documentation for more information

Development

 Feedback and contributions are always very much welcome.

The source code of the plugin is available at GitHub:
https://github.com/sascha-leib/dokuwiki-plugin-botmon.

Changelog

- 2025-09-06: Improved bot detection.
- 2025-09-11: General web-metrics for human visitors only.
- 2025-09-12: Cleanup and UI overhaul
- 2025-10-03: Now looks at the last *full* day (i.e. *yesterday*), improved web traffic analysis.
- 2025-10-17: General overhaul with more detailed information and automated cleanup
- 2025-10-19: Added icons for the various log types
- 2025-10-19: Reorganization of user-triggered actions: moved from clients to bots
- 2025-12-06: Added Captcha and more detailed request information

Known Bugs and Issues

See below.

ToDo/Wish List

The development plan includes:

- Fine-tuning the bot assessment – probably by adding more metrics and tests for bot activity.
- Processing the log files into a database and providing long-term information, not only for the latest data
- Adding a blacklist option for particularly annoying bots.

Other ideas:

- Use server-time instead of UTC.
- Show more web metrics in the dashboard.

FAQ

Q: How reliable is the assessment of the plugin?

A: Not very. There is still a lot of fine-tuning to be done; Many of the entries in the «probably humans» category are most likely bots, and all of them are categorized based on many assumptions that may or may not hold true.

Q: Why can't I see the countries of the requests?

Your server needs to have the «GeoIP» module installed for this feature. Then enable the country lookup in the configuration options. Note that this module is not very well maintained and often shows wrong information, or even more often, no country information at all. Still better than nothing, I reckon…

Q: How can I block bots?

As tempting as it is to think you can just install a plugin and that will block the bots for you, it is really not that easy. Please start with spending some time to analyze the traffic you have your web site and try to understand how the bots behave in your case, and which traffic (and in some cases: bots) you actually want to allow on your site.

Then see if you can use the methods described in the article “How to block bot traffic” in the documentation; first to tell bots they are not welcome (a lot of them will listen!) and then to actually block them on a system-level.

A last resort is the provided Captcha option. This will not only hide your content from bots, if they fail to solve the simple Captcha puzzle, but also feeds nonsense text to those bots that just won't listen. The idea is that botnet operators will voluntarily blacklist your site on their side, if they see that they are not getting good quality content from it.

From:
https://wwoss.ru/ - **worldwide open-source software**

Permanent link:
https://wwoss.ru/doku.php?id=wiki:plugin:botmon

Last update: **2026/01/18 15:39**