log-file-decrufter

Console application designed to reduce the data within web server logs (tested against apache access logs).

When dealing with a DFIR breach there is normally so much data, that we need to get it to a manageable level. With this is mind we generally don't care about requests to *.png files or* .css files, so the application removes those lines and creates a new file.

As a example, it can reduce a 3 GB log file down to less than 1 GB in 10 mins.

log-file-decrufter uses pre-defined regexes to remove the lines. The regexes are stored in the config file. Feel free to add/remove/modify them. If you add/modify then make sure the regexes are escaped e.g. backslashes

If you have useful additions/modifications then get in touch and I will see about adding them.

Currently the regexes remove the following:

- GET *.png HTTP/1.1
- GET *.jpg HTTP/1.1
- GET *.js HTTP/1.1
- GET *.css HTTP/1.1
- GET *.gif HTTP/1.1
- GET *.pdf HTTP/1.1
- GET *.woff HTTP/1.1
- GET *.mp3 HTTP/1.1
- GET /favicon.ico HTTP/1.1

- GET / HTTP/1.1"

The configuration file also allows the modification of the number of workers that are used at once. By default the value is set to 0 which sets the number of workers == to the number of CPU cores.