RIT | B. Thomas Golisano College of Computing and Information Sciences

Department of Computing Security

Matthew Turi
(mxt9495)
Course: CSEC-380-01
Submission Date:
02-24-2020

HW3 - Activity 4 Writeup

Is there certain information about the web server that you can discern based on what files you can access?

Yes, although this was discovered through manual analysis and not automated scanning. The server doesn't respond with any server tokens in the response header, but the server does respond with a 403 when attempting to access /.htaccess. The 403 response isn't necessarily indicative that the file exists, but it leads me to believe that that file is intentionally being denied world read access. An htaccess file is a file with configuration directives for the Apache web server that apply settings on a per directory basis.

Are there any ways to improve the speed of your scanner?

I think that much of the speed degradation that my crawler experiences is due to its architecture. I essentially implemented a pseudo-BFS crawling scheme, and it uses up a large amount of system memory. Additionally, I think that scaling the application onto hardware with more CPU cores would allow it to run better, as I limit the number of threads my application spawns to the number of physical cores the system has available.

How can response codes be used in order to more efficiently search the site?

Basing the crawler's logic off of the response codes instead of say, the request body, is that I can choose to not process the entirety of the request body. This saves considerable amount of processing when handling HTTP responses. I only need to look at the response code and can make a decision as to whether the page exists and move on without ever having to run BS4 or something on the response content.

Are there any common naming patterns that you might expect would yield positive results?

I'd typically expect paths that store static assets on sites to yield many positive results (i.e. '/static', '/images', etc.). In the case of the Hooli website, the only path that I seemed to notice that was at least linked from the home page was to '/images'