ECE 498KL: eCrime and Internet Service Abuse (Fall 2018) **Due November 13, 5:00 P.M. CST**

Assignment 1 Part B

100 pts

This assignment is the second part of a three-part assignment which will teach you about search engine abuse. For this assignment, you will analyze the Web visitor data you've collected while your server was running. **You must collect at least a week of data.** This means that your Web site should be running and collecting data at least seven days before your submit your solutions.

1 Your Web Server

You will use the same VM that you used for the first part of this assignment. Check your site regularly to make sure it is still running and collecting data.

2 Problem 3: Attacks

Miscreants on the Internet often try to exploit known vulnerabilities in Web services by issuing a request to a Web service that might exploit a vulnerability. Look through your logs for requests that may be attempting to exploit a vulnrability: such requests often requests an unusual path (the portion after GET or POST in the first line of the request).

In your solution, list four requests from different IP addresses that you think may be attempts to compromise your server. For each: (a) explain why the request looks suspicious; (b) identify the vulnerability being exploited; (c) provide the geographic location of the IP address; (d) provide the hostname corresponding to the client IP address, also called the "reverse DNS" name; (e) provide the organization to which the IP address is assigned. For questions (c) and (d), you may find free IP geolocation tools such as `https://tools.keycdn.com/geo` useful. For question (e), you may find the `whois` command-line tool useful, or the Whois search at `https://www.arin.net/public`. *This problem is worth 50 points.*

3 Problem 4: Web Crawlers

In order for your Web site to show up in search engine search results, the search engine must first visit your site and index its content. Look through your logs for requests that may be from search engine crawlers.

In your solution, list two requests from different IP addresses that you think may be a Web crawler attempting to index your site. For each: (a) explain why you think the request is from a search engine Web crawler; (b) identify the search engine the crawler represents; (c) provide the geographic location of the IP address; (d) provide the hostname corresponding to the client IP address, also called the "reverse DNS" name; (e) provide the organization to which the IP address is assigned. *This problem is worth 50 points.*

4 Solution Format

Your submission for this assignment are two plain text files named `hw1pr3-`*`netid`*`.txt` and `hw1pr4-`*`netid`*`.txt`, where *`netid`* is your NetID, containing your solutions for Problem 3 and 4, respectively. Both files must

contain free-form text (with examples of records as in Part A of the assignment, if necessary).

5 Submitting Solutions

You must submit your solutions on UIUC's local deployment of GitHub. Commit the two solution files to your student repository, also inside the mp1 directory, next to your answers to part A. We will grade a snapshot of the contents of your student repository taken at the deadline for each assignment. Only the latest commit to the master branch your repository that we create for you will be considered. If you would like to use one or more of your three 24-hour extensions, *you must email the instructor and the TA;* we will then pull your solution at the end of the extension.

```
$ git add mp1/hw1pr3.txt mp1/hw1pr4.txt
$ git commit -m "my solutions"
$ git push
```

Your server must continue running and logging visitor data after the deadline. You will need the data it collects for part C of this assignment set.

5.1 Academic Integrity

You may consult any Internet resources you wish, however, you must *not* discuss your solution with other students until three days after the assignment deadline.