ECE 498KL: eCrime and Internet Service Abuse (Fall 2018) **Due October 30, 10:00 P.M. CDT**

Assignment 1 Part A

100 pts

This assignment is the first part of a three-part assignment set that will teach you about search engine abuse. For this assignment, you will create your own Web site and collect data on visitors. **You must collect at least six consecutive hours of data.** This means that your Web site must be running and collecting data at least six hours before your submit your solutions.

1 Your Web Server

You will receive SSH login credentials to your own Web server virtual machine via email from your TA as well as the URL of your Web site. The file `index.php` contains the PHP script that generates the contents of the root page (the page displayed when "/" is requested by a Web browser.) The server will also be configured to run `404.php` when a non-existent document is requested (this would normally return a 404 error). You can find more information about PHP scripting at `http://php.net/docs.php`.

2 Problem 0: Page Content

Modify `index.php` to return some text content to the user. In addition to the login credentials for your Web server VM, you will also receiver three keywords. These keywords are made up words that currently do not occur anywhere on the Web. You content should include a paragraph of text that includes all three of these keywords on your page. See the page at `http://joshuar3.ece498kl-fa2018.org` for an example, but **do not use the same text**. Your page must be unique. *This problem is worth 10 points.*

3 Problem 1: Server-Side Logging

Modify `index.php` and `404.php` to log every visit to your Web site. Log at least the following information about each Web request:

- `Time`: Unix timestamp of request;
- `User-IP`: client IP address;
- `User-Agent`: the user agent as given in the `User-Agent` request header;
- `Referer`: the referring document as given in the `Referer` request header;
- `Request`: first line of the HTTP request, including request type (`GET`, `POST`, etc.) and path;
- `Accept-Language`: the set of languages given in the `Accept-Language` request header;
- `Accept-Encoding`: the set of encoding given in the `Accept-Encoding` request header.

Your may include additional fields. You must at least log requests for the root directory page (generated by `index.php`) and requests for all non-existent pages (using `404.php`).

Tip 1: You will find useful information in the `$_SERVER` PHP variable.

Tip 2: Remember that your Web server is asynchronous—several instances of your PHP script may run at the same time. You can use PHP's `flock()` function to guarantee an atomic write to your log file. *This problem is worth 45 points.*

4 Problem 2: Client-Side Logging

Use JavaScript in the document you return to the client to collect additional data about each Web visit. You must collect at least the following information for each client (identified by IP address and user agent):

- `User-IP`: client IP address (in ASCII);
- `User-Agent`: the user agent as given in the `User-Agent` request header;
- `Screen-Size`: client screen size as two integers separated by "`x`" (e.g. "`640x480`");
- `Time-Zone`: client time zone in minutes offset from UTC.

Your log may include additional fields. *This problem is worth 45 points.*

5 Solution Format

Your submission for this assignment must be two plain text files named `hw1pr1.txt` and `hw1pr2.txt`, containing your solutions for Problem 1 and 2, respectively. Each file must contain a record for each visit (Problem 1) or client (Problem 2). Records must be separated by a blank line. Each record must consist of named data fields, one per line. A data field consists of the field name, followed by a colon and a space, followed by the associated value. Figure 1 shows an example of this format for Problem 1; Figure 2 shows the format for Problem 2.

```
Time:  1539720731
User-IP: 130.126.255.136
Host:  www.google.com
User-Agent:  Mozilla/5.0 (Macintosh; Intel Mac OS X 10.9) Firefox/56.0
Referer:  https://www.google.com/
Request:  GET / HTTP/1.1
Accept-Encoding:  gzip, deflate, br
Accept-Language:  en-US,en;q=0.5
```

Figure 1: Solution formatting example for Problem 1 (`hw1pr1.txt`).

```
User-IP: 130.126.255.136
User-Agent:  Mozilla/5.0 (Macintosh; Intel Mac OS X 10.9) Firefox/56.0
Screen-Size:  1440x900
Time-Zone:  300
```

Figure 2: Solution formatting example for Problem 2 (`hw1pr2.txt`).

6 Submitting Solutions

You must submit your solutions on UIUC's local deployment of GitHub. 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.

6.1 Setting Up the Solution Submission Repository

You must sign in with your university credentials at:

https://edu.cs.illinois.edu/create-ghe-repo/ece-498-kl-fa2018/

You will be shown a menu with two buttons. Click the top one first to log into GitHub in a new tab. Your account will have no repositories unless you have used this system before. Return to the two-button tab and click "I've logged in." Once you do, there will be a Git repository automatically provisioned for you at:

https://github-dev.cs.illinois.edu/ece-498-kl-fa2018/*netid*

You can complete these steps regardless of your class registration status. We will grade a snapshot of the contents of your repository taken at the deadline for each assignment (or at the end of your extension). Only the latest commit to the master branch your repository will be considered. We will publish a solution skeleton at:

https://github-dev.cs.illinois.edu/ece-498-kl-fa2018/_release

Copy these exactly to your personal repo, and fill in your solutions. Do not introduce more directories or files than are found in `_release`. You should do something like this for this assignment:

```
$ mkdir ece498kl
$ cd ece498kl
$ git clone https://github-dev.cs.illinois.edu/ECE-498-KL-fa2018/_release.git
$ git clone https://github-dev.cs.illinois.edu/ECE-498-KL-fa2018/netid.git
$ cp -r _release/mp1 netid/.
```

6.2 Submitting Your First Assignment

```
$ git add mp1/hw1pr1.txt mp1/hw1pr2.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 parts B and C of this assignment set.

6.3 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.