CSC221 Data Structures & Algorithms I Fall 2016

Programming Assignment #4 – Hash Tables

The authorization log on **gottlieb** shows a large number of failed login attempts; the vast majority of these attempts are from IP addresses in China. In a week, we receive approximately 40,000 login attempts to gain root access on gottlieb. In this lab, we will discover the IP addresses of the most persistent attackers.

Input: A text file named **hack_attempt.txt**. A sample from the file is shown below:

```
rhost=221.194.47.249
rhost=221.194.47.249
rhost=116.31.116.8
rhost=179.170.12.235
```

Each line in the input file represents a failed attempt to login from that IP address. You can download this file from: *http://menehune.opt.wfu.edu/csc221/hack_attempts.txt*.

Processing: Your program will read each line of input as a character string. Store each unique IP address in a **hash table** along with a count of how many times that IP address has been seen. Do not store any duplicate IP address as separate hash table records. Use a polynomial hash function. Resolve collisions by chaining.

Output: When input processing is complete, print the number of input lines processed. Skip a line and print the top 5 most frequently seen IP addresses, sorted in decreasing order along with the count of how many times that IP address has been seen in the file . There may be fewer than 5 distinct IP addresses; if so, print all of them.

Notes: You may "hard code" the file name as a string constant. E.g.:

```
const char * fname = "hack_attempts.txt" ;
```

When you open the file, be sure to check for success. If the file is missing or unreadable, your program should print a message and exit.

You must include a Makefile to compile your project.

Turn-In :

Keep all your work in a sub-directory named `Lab4`. Change to the parent directory of `Lab4` and create a tar archive of your work using the command:

```
% tar cf Lab4.tar Lab4
```

Upload the file `Lab4.tar` to your account on telesto.