secure Scripting

**Advanced scripting**

# Lab: Analyzing a spreadsheet

**Lab Description:** In this lab you will analyze a spreadsheet of IP connections. The spreadsheet is the same one as was used in an exercise for Module 8, but it has been converted into .csv (comma-separated value) format. Each field of the row is separated by a comma in this format.

**Lab Environment:** The students should have access to the contents of the “data” directories in lesson 3. They will also need access to a Linux system, preferably Linux Fedora 22. They will not need Internet or network connectivity once the directories and Laboratory Exercise files are downloaded.

For additional guidance, there are the specific programs that this lesson uses.

• Lesson 3 uses the commands *bash* (1) or *sh* (1), *date* (1), *expr* (1), *sed* (1), and *tr* (1).

**Lab Files that are Needed:**

For this lab you will need the following files:

* connect.csv

### **Lab Exercise 1**

The format of the "connect.csv" file is given on the slides. Please write a script that prints the date and time fields of each line of the file. The first line's fields should contain "DATE" and "TIME", respectively. Save the script as “ipconn1.sh”.

Hint: You can adapt the *while* loop on Slide 12 for this.

### **Lab Exercise 2**

Modify your script “ipconn1.sh” from Lab Exercise 1 to add a field called "TIMESTAMP" between the TIME and DURATION fields. Populate this field with the result of converting the contents of the DATE and TIME fields for that line to the number of seconds since the epoch (the \*nix timestamp). Save the script as “ipconn2.sh”; save the resulting output in a file named "connect-ts.csv".

Hint: The first line consists of the headings, so add the heading for this new field in the appropriate place.

### **Lab Exercise 3**

Write a script to print the first and last timestamp in the file "connect-ts.csv", followed by the difference in seconds between the two timestamps. Save the script as “ipconn3.sh”.

When you test the script, notice how hard it is to understand the large number of seconds as a meaningful duration of time. Modify the script so that it expresses the difference in the form “*hh*:*mm*:*ss*” where *hh*, *mm*, and *ss* are hours, minutes, and seconds respectively.

Hint: Use the program *expr* to divide by 60, then to divide the remainder by 60.

### **Lab Exercise 4**

Write a script that uses the file "connect.csv" as input and creates a new column called "DESTINATION NAME" in the output. Transform the DESTINATION IP based on the following information, and store the resulting names in the newly created column.

• 172.016.112.050 → pascal

• 172.016.112.194 → falcon

• 172.016.113.050 → zeno

• 172.016.114.050 → swan

• 172.016.114.168 → finch

• 172.016.115.005 → pc1

• 172.016.115.050 → mill

Save the script as “ipconn4.sh”, and save the output in a file called "connect-host.csv".

Hint: If you use *sed* to do this, it is easiest to put the substitution commands in a separate file, and use the -f option to *sed*.

### **Lab Exercise 5**

Write a script that takes input from the file "connect-host.csv” and adds a new field called "ANONYMIZED SOURCE IP" immediately after the SOURCE IP field in the output. Transform the SOURCE IP by replacing the last two octets of the IP address with 0, and store the transformation in the newly created field. For instance, the IP address 101.102.103.104 would become 101.102.0.0. Save the new script as “ipconn5.sh”.

### **Puzzler**

Modify the script you wrote for Lab Exercise 4 (ipconn4.sh) to also count the number of times each host name appears in the file. Save the modified version of the script as “ipconnP.sh”.

Hint: Keep a list of the names as they appear. Ensure the names are separated by newlines (*tr*). Then sort them (*sort*), eliminate duplicate lines, and put the number of times each line occurred before the output lines (*uniq –c*).

### **Big Puzzler**

Using the file “connect.csv” as input, write a shell script to create two new fields called “TIME (MINUTES)” and “DURATION (MINUTES)” right after the TIME and DURATION fields, respectively. Transform the original TIME and DURATION fields by rounding to the nearest minute and save the outputs in the newly created fields. Save the new script as “ipconnPB.sh”.

## what to submit

For the parts of the exercises that do not require you to write a script, put your answers in a PDF or text file numbered appropriately, and call that file “Unit3answers.pdf” or “Unit3answers.txt” respectively. For the parts of the exercises that do require scripts, create plain text files to hold your script, and name the file containing your script as indicated in the problem.