Computer Security Concepts
COMP 435, Fall 2017
Instructor: Cynthia Sturton

Release Date: 9/3/2017
Due Date: 9/10/2017, 11:59 PM

Introduction to Python

Overview

The purpose of this assignment is to give you time to set up the environment you will use in subsequent labs and to become familiar with Python, which you will also need in the next assignment.

Requirements

You will use a virtual machine (VM) for this assignment. It may take some time to set up the VM, but we will use it again in later assignments. There are two steps to installing the VM:

- Download and install VirtualBox, the virtual machine software.
- Download and install the VM from SEEDLabs.

Follow the instructions and download links for both steps from the SEED Labs site (`http://www.cis.syr.edu/~wedu/seed/lab_env.html`). Note that the VM is a 2GB file! This will take some time to download. You should do this on campus for faster download speeds.

Once you have the VM set up, you can install Python3. Login (the password for the VM is "dees"), open a terminal, and enter:

sudo apt-get update && sudo apt-get install python3

Be sure to **not** upgrade any other packages, since we will be highlighting vulnerable and outdated ones later in the class.

You will be writing a program in Python. A good starting point for learning Python is here: `https://docs.python.org/3/`. Some students may find an IDE for Python helpful. There are a number online, such as iPython and PyCharm, which you are encouraged to explore.

Detecting Duplicates

For this assignment you will be detecting duplicate hex values in a large text file. You will sort them by value and then display how many duplicates are present.

For example, lets say you have the following text file named "sample.txt":

```
9ec4c12949a4f31474f299058ce2b22a
9ec4c12949a4f31474f299058ce2b22a
43
aa
02
123
acd
aa
123
123
02
```

Your program should parse through the entire file and if there are more than 2 duplicates, print out how many instances occurred. In this case the correct output will be:

```
2 2
aa 2
123 3
9ec4c12949a4f31474f299058ce2b22a 2
```

Note that the output is sorted from the smallest value to the largest value in hex.
Your program must also take in the file as a command line argument. For example:

python3 `a1_csturton.py` sample.txt

There are many ways to complete the assignment. You should use the built-in data structures that Python provides (e.g., sets or dictionaries). You should not be looping through the hex values multiple times looking for duplicates.

Grading Rubric

70% will be graded on correctness

30% will be graded on style

Submission Instructions

To submit your assignment, go to sakai and upload your Python3 file titled:

`a1_<youronyen>.py`

So for example, my assignment would be titled:

`a1_csturton.py`