Lab 6 - Understanding the Host's Integrity

The objective of this lab is to write software that will detect changes across the filesystem by hashing all files across your virtual machine. This is accomplished by writing a Python script that will walk through all of the directories and files on your virtual server and hash each individual file, keeping track of the hash values so that subsequent runs can make note of the changes. In essence, you are writing a software package similar to the initial tripwire software.

Problems

1. Create a **hash.py** script that will recursively walk through all files on your file system. You should be able to configure your program to define certain files **and/or** directories as **unhashable** so that they will be ignored, **/dev** is a perfect example, as you **do not want to hash this directory**. **Note:** Below is a non-exhaustive list of directories you will probably want to ignore (if your program gets stuck on a directory you will want to consider ignore those as well):

    - /dev
    - /proc
    - /run
    - /sys
    - /tmp
    - /var/lib
    - /var/run

    A Part 1 solution will be able to walk through the entire file system, printing out the filenames (with their paths), and taking no action other than skipping the *unhashable* files or directories.

    Hint: In-built libraries and functions, such as **os.listdir**, **os.stat**, and **os.walk** can be quite helpful.

2. Integrate SHA2 (SHA256) so that each file is hashed as it moves through the file system. You are not expected to write SHA2 from scratch, use a python library such as *hashlib*.

3. Store the file and hash information so that it will be available for future runs. At a minimum store the following data:

    - filename with full path
    - hash
    - date/time file was observed

4. The final step, your program should run and update the hash information, upon completion it should print out summary information that includes all new files found, any missing files, and any file that was modified.

5. **Extra Credit** Detect that a file was moved, add to your summary section an output that documents where the file is now, and where it was, and the time of the last scan that saw it in the older location.

What to submit

When you have completed the lab **post all requested materials to your webpage**. Then, submit the link via email to your Professor.

The email subject line should be SY402 [Section Number] Lab [X]: Title of Lab (e.g., SY402 1111 Lab 6: Understanding the Host's Integrity). Email sent with a different subject line will reduce the overall grade by 5 points.

The web page link(s) should include:

1. (Attach) your hashing script (hash.py),
2. (Attach) the file in which you stored all of the hash data (**Submit the hashed data file compressed as a tar.gz file**)
3. (Attach) a README.TXT file that explains how you stored the data in the hash data file. If you completed the extra credit, make sure to note that in your README.TXT file.