CS 365: Digital Forensics
Spring 2020
Assignment #4

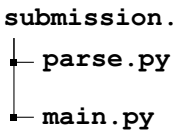
Arun Dunna
adunna@cs.umass.edu

Revision 1.0 - February 27, 2020

Due: March 6, 11:55pm

Submission Instructions

1. Your programming solutions for this assignment are to be written in **Python 3.6**.
2. Your solutions to written response portions of this assignment are to be typed and submitted to the Gradescope written assignment in PDF format with the following specifications: **12pt font, 1-inch margins, clearly labeled and appropriately sectioned solutions, full name included, student ID number included, submission date included.**
3. Your programming solutions are to be submitted to the corresponding Gradescope programming assignment. These should be submitted in a **ZIP file** containing your Python code files, in the following format:

   This will match the format of the assignment code distributable.

   Alternatively, you can upload each Python file individually in Gradescope instead of using a ZIP file.

Prerequisites

1. Ensure that you are enrolled in the course on Gradescope. The assignment submission will open a few days before the deadline. If you have not been automatically enrolled, you can use the following enrollment code: **9DZ53K**
2. Setup and test your own environment for executing **Python 3.6** code. For example, this can be in an IDE such as PyCharm, or in a terminal with the **`python3.6`** command.
3. Obtain the assignment distributable from the course website: **`asgn04distrib.zip`**

1 Programming

The assignment distributable contains skeleton code in one file: **`parse.py`**. It also contains **`main.py`** to test the program. Edit the **`parse.py`** Python file to complete following functions, further described in the docstrings of each function in the skeleton code. You may assume that the files you are reading (passed into your functions as strings) exist.

1.1 Password Extraction [50 points]

(50 points) `extract_passwords(inputFile)`**:** extract ASCII-printable passwords of length $\geq N$ from TCP data

2 Written Response

The following questions should be answered in complete, readable sentences, and adhere to the relevant specifications in the submission instructions. They should be labeled according to their numbers. These questions will be focused around network forensics, but any content from previous sections in the course can be brought in to aid in answering the questions. These are open-ended questions, and the more specifics you give, the better.

2.1 Inside Job [15 points]

Aerotyne International is a cutting-edge tech firm based out of the Midwestern United States. They have around 4,000 employees working out of over 73 countries around the globe, and gross upwards of USD $2 billion in net profits per year. You are their Head of Security, and just discovered that company trade secrets were sold on the Dark Web. The CEO of Aerotyne comes barging into your fancy 11th-floor office, and demands that you find out how the secrets were stolen within the next 24 hours. You are also told that this happened on company time, and was an inside job using the company's systems and networks.

Assume that your company has a zero-log policy, and you do not monitor network traffic at all.

1. **What, if anything, can you and your team do to narrow down who may have leaked company data?**
2. **How can you respond to this to better handle similar incidents in the future?**

2.2 Malicious Activity [15 points]

Unfortunately, it turned out that Aerotyne International had been in cahoots with Stratton Oakmont and their worth was incredibly overvalued. Instead of being worth $30 billion, they were now worth $30, period. You therefore had to find a new job, and decided to become your own boss by starting an Incident Response Team called "The Analysis Team" (often referred to as "The A-Team"). On your first day as Head of Investigations, a call comes by your office. It's your old pal, Gerardo! But unfortunately, he's in trouble - his company, Jack's Magic Jacks (JMJ) was recently the victim of a ransomware attack. He has made sure to disconnect all of his computers from the Internet and from the network, and luckily, his primary access server was untouched by the virus. He tells you that he logs most of the network traffic, but that logs are deleted after just one week of being recorded to save space. He also tells you that no workers have access to this particular network or set of systems, but that multiple server applications (such as web servers and email servers) are hosted on them, and he admits to not updating them regularly. He needs your help to isolate the entry point that the hackers used to infect the systems. Remembering that you are working with a dangerous virus, you make sure to isolate all systems involved and get to work.

1. **What steps would you take before starting to investigate the systems, and why?**
2. **What steps would you take to determine possible entry points, and what are additional ways, if any, that you can narrow that pool down to a single entry point?**
3. **Lastly, what steps would you take to prevent incidents like this from happening in the future, based off of the possible entry points?**

2.3 Criminal Images [20 points]

It's been several years, and you've now ended up working often with law enforcement on criminal investigations as a digital forensics expert. On one summer day, you get a message on your pager (yes, you have a pager) to meet with your primary contact at the FBI, Lauren, immediately. Lauren tells you that a popular image hosting company, Imgyour, has become the focal point of two separate criminal investigations regarding their images. Two separate users, in two separate cases, uploaded images to the system, and both users are suspects in two separate homicide cases. These images were uploaded around the same time that the crimes were found to be committed. Lauren knows you've worked with image metadata before, and asks you for your help with trying to extract any information that might narrow down whether the suspects could have committed the crimes in their respective cases.

Unfortunately, Lauren has some extra information that makes this a bit more difficult. While the first suspect's (Suspect A) images remain in a secure drive in Imgyour's data facility, the second suspect (Suspect B) had their images deleted from Imgyour's drives by an automatic content removal program. Additionally, Lauren does not know if Imgyour strips metadata when users upload images, or if they retain original copies. She does know that Imgyour keeps meticulous server logs of who uploaded which files, but those logs don't include the actual content - only the file names and hashes of the file content.

Given all this information, Lauren asks you to map out a plan for each suspect that would aid in narrowing down whether they could have committed their respective crimes.

Describe your plan for each suspect, and note that while you have access to all of Imgyour's servers, you would need warrants issued to search the computers of the suspects.