For the second assignment (Assignment 1, Part 2), we were told to spend time trying to compromise the passwords that were given to us in a zip file containing 3 compressed folders each containing "text" files that held information within them.

My approach when first starting out was to check the contents of each to see if any of them were in plain text. I checked both the linkedin and formspring text files, but the passwords were hashed. I then changed the file extension of the yahoo file to a text file, and I was able to see all of the unhashed, plaintext emails and passwords.I quickly wrote a script to extract 100 of the passwords (that were unique) and proceeded to try cracking the other text files.

I then studied the linkedin file and saw that its file name was called "SHA1.txt". This gave me a clue as to what hashing algorithm it was in. I proceeded to look inside the file and found that a lot of the passwords seemed to have five leading 0s in the front. I figured that this may be significant and incorporated this idea as I wrote the next portion of my script. The next portion of my script relied on a "dictionary attack" method where I obtained the top 10,000 most used passwords by wikipedia[1] and hashed them and proceeded to place the five leading 0s and put them all in a list so that I could iterate through them comparing all of those masked hashes, with the list of hashes from the SHA1.txt file.

The formspring file was where it got a bit difficult. After receiving a hint from the TA's announcement about salted passwords, I proceeded to check out what the TA meant. I then used that information to create the next portion of my script where I first started doing the same method as above, but first salting the passwords before hashing them via SHA256. I didn't get any hits from the 10,000 most used passwords list, and it took me significantly longer as it wasn't only 10,000 iterations I had to do, I had 10,000 * 100 = 1,000,000 iterations. I then got a different dictionary I found online called rockyou[2]. With my weak processor on my laptop, I knew it wouldn't be able to run the script as fast as a better processor at school, but even then it wasn't working as fast as I would have liked. So I played around with the code, and after researching some python data structures, I realized I could just use a set instead of utilizing lists, and so I utilized a set for each of the sets of data (SHA256 hashed passwords from rockyou, and formspring.txt) and iterate through the comparisons.

Other techniques that I considered was brute forcing, but running polynomial time, even after running it overnight was just unbearable. So in the end, the Yahoo passwords were stored plaintext (in the clear), LinkedIn passwords were stored hashed via SHA1 with no salt or extra hashing, and FormSpring was hashed with a salt of a random value between 00-99.

In terms of difficulty cracking passwords, Yahoo < LinkedIn < FormSpring.

[1] https://en.wikipedia.org/wiki/Wikipedia:10,000_most_common_passwords
[2] https://wiki.skullsecurity.org/Passwords