CSCI 5271: Intro to Computer Security

Homework 3 **Due: Dec 23, 2020**

Ground Rules. You may choose to work with up to two other students if you wish. Only one submission is required per group, please ensure that all group members names are in a file named `students.txt`. Work must be submitted electronically via canvas.

Breaking a Fixed Capacity Chaum Mix

As discussed in class, Chaum mixes which do not force all users to participate each round are vulnerable to statistical attacks which leverage the fact that your friends are more likely to receive messages when you are online compared normal circumstances. This programming assignment will have you implement a simple statistical disclosure attack to de-anonymize conversing parties. Posted on the assignment is a file `target-rounds` which contains a number of rounds of message distribution via a Chaum mix. Each round takes up two lines, the first lists all senders for a given round, the second lists all receivers of message for that round. Your job is to leverage an attack spelled out in the paper posted on the class website to determine, based on these rounds of senders pushing messages into a Chaum mix and receivers getting messages out, who is talking with who.

In this assignment, there are 260 users, `a0` through `z9`. Each user has two "friends" chosen according to a power-law distribution (so that some friends are more popular than others). The friend relationship is *not* bi-directional (e.g. if `a0` has the friend `g5` that does *not* imply that `g5` has the friend `a0`). The mix itself has a batch size of 32. In each batch, a set of 32 users is picked uniformly at random, and each of these users picks one of its two friends to send a message through the mix. The output file lists the senders (on the lines beginning with "`S`") and recipients (on the lines beginning with "`R`") of each batch of 32 messages, but not the correspondence between them. For this problem, you should perform an "extended statistical disclosure attack" to discover the friends of twenty six "target" users. You can read more about this attack at from the paper on the course website.

Because this attack is statistical in nature, it will not correctly de-anonymize all users. You should be feeling good about your code if it is correct over 2/3rds of the time. I have also provided a second set of rounds (example-rounds) along with the ground truth for that set of rounds as to who is friends with whom. This is to help you test your code. Please note that users have different friends in the actual target and you will be responsible for attacking those users.

Your target users' "names" will be all those that end in the digit 0. Your submission should include the following items:

- Your source code for executing the extended statistical disclosure attack.
- A csv file named `results.csv` containing the target users and their friends as determined by the extended statistical disclosure attack. **Be sure this is for the target-rounds file and not the example-rounds file.**

Your CSV file should be machine readable in the form of:
`targetId,firstFriend,secondFriend`
For example if you find that `b0` has friends `t6` and `e4` your line would be:
`b0,t6,e4`