Assignment 3

1. a)

- Advertised bandwidth of relays with the “Guard” flag (but not the “Exit” flag) = 249.6358 Gbits/s

- Advertised bandwidth of relays with the “Exit” flag (but not the “Guard” flag) = 19.30847 Gbits/s

- “Guard” flag has higher bandwidth than “Exit” flag.

- Exit nodes connect and pass data to the servers and receive reply from servers then send back to cached adjacent node. They do not do much work, but they are more vulnerable due to legal issues so that not many people want to become exit nodes.

b) Median download rate for 50KiB file = 209.4 Kbits per second

Median download rate for 5MiB file = 1848.8 Kbits per second

The file size is 100 times larger, but the download rate is only 10 times faster than smaller file. Therefore, the larger file will take more time to complete its download. The reason behind it is fragmentation of large file where server divides into small chunks and send them over the wires. It takes some time to gather all chunks in order so that it might take longer in time. There’s another factor is speed of transmission which could be a bottleneck at the “Exit” node in Tor.

c)

Disadvantages:

- Using three nodes requires more communications for key exchange for each node while a single node only need one key exchange.
- Faster throughput since data only passes single node.

Advantages:

- Only “Guard” node knows who you are, and “Exit” node knows destination. Good anonymity level.
- 3 nodes provide sufficient latency to complete round trip data transfer at the optimal level.

2.

(a)

- **k-anonymity**: the technique satisfies to preserve patients’ privacy. The distorted data table will have 2 columns: location and CROW infection. Location column can be distorted to be a radius

of 40m around user. The app can query and return result with 2-diversity (Yes/No) in CROW infection column. User can calculate the percentage of infected patient over all people in the database and see whether he/she has high chance to be exposed.

- **Differential privacy**: the technique is suitable for this problem. Collecting data of infected patients and it is good to count numbers of CROW patients in the area to notify user.

- **Secure multiparty computation**: this technique works fine because 2 companies could jointly compute query without sharing their users data in order to preserve privacy.

- **Private information retrieval**: this technique is not suitable because it requires user to download database or involve in complicated encryption scheme.

(b)

- **k-anonymity**: the technique is not suitable to use because it only involves single person, and no sensitivity attribute exists here.

- **Differential privacy**: this technique is not suitable because it reveals user's queries on personal eating habits.

- **Secure multiparty computation**: this technique is not suitable because it requires 2 parties to compute together.

- **Private information retrieval**: the technique is suitable for data privacy where owner does not know what user is asking for but just sending the result to user.

(c)

- **k-anonymity**: the technique does not work here because your main problem is to send private DNS query.

- **Differential privacy**: the technique does not work because it requires non private query.

- **Secure multiparty computation**: this technique is not suitable because it requires 2 parties to compute together.

- **Private information retrieval**: this technique is suitable for data privacy which you DNS does not know you send query but return you answer for such query.

(d)

- **k-anonymity**: this technique works for the problem that does not reveal raw data to public.

- **Differential privacy**: this technique can be used to determine real-life usage habits and online usage pattern in order match whether they should recommend talking to each other.

- **Secure multiparty computation**: this technique is suitable for this problem for ussers' perspective. User must input their genuine info and the algorithm can compute to see if you are matched with whom. None of users know each other information.

- **Private information retrieval**: the technique does not work for this case because the data owner needs your information to match with other people. By hiding your information aka query and only return solution. The database would never grow and can hardly find any matches.