Deniz Evrendilek (301340591)
CMPT 403 (Summer 2023)
4 Aug (Fri) 2023

Assignment 3

1(a). Under the assumption that 2020-02-01 is the 1st of Feb, 2020 (the same assumption is made for all the remaining answers), the total numbers of advertised bandwidth of relays with the flags "Guard" and "Exit" are 3109 and 1006, respectively. Therefore, the number of Guard relays is higher than that of the Exit relays. One possible reason for it is that more volunteer nodes are the Guard nodes rather than the Exit nodes. Suppose that an Exit node traffic is associated with a "bad", "blacklisted", or "illegal" website. In that case, the IPs of the Exit-node volunteer can refuse to provide internet services with the volunteer node.

1(b). Based on the download statistics for 50 KiB from metrics.torproject.org, the median value shows that the op-hk onion server takes 5.315 seconds (s) to download 51,200 bytes (50 KiB). Therefore, the download rate is 51,200 bytes / 5.315 s = approximately 9,633 bytes per s.

Based on the download statistics for 5 MiB from metrics.torproject.org, the median value shows that the op-hk onion server takes 28.580 s to download 5,242,880 bytes (5 MiB). Therefore, the download rate is 5,242,880 bytes / 28.580 s = approximately 183,445 bytes per s. To calculate the latency and transfer rate, we form and solve the following two equations:

5.315 s = latency + (51,200 bytes / transfer rate), where 51,200 bytes = 50 KiB

28.580 s = latency + (5,242,880 bytes / transfer rate), where 5,242,880 bytes = 5 MiB

This is equivalent to 23.265 s = (5,191,680 bytes / transfer rate), which gives that transfer rate = approximately 223,154 bytes per s. Using the newly estimated transfer rate yields that latency = approximately 5.080 s.

The percentage of the total load time due to the latency for 50 KiB data is estimated at 95.57% as a result of the following calculation: (5.080 s / 5.315 s × 100).

The percentage of the total load time due to the transfer rate for 50 KiB data is estimated at 4.42% as a result of the following calculation: ((5.315 s - 5.080 s) / 5.315 s × 100).

The percentage of the total load time due to the latency for 5 MiB data is estimated at 17.77% as a result of the following calculation: (5.080 s / 28.580 s × 100).

The percentage of the total load time due to the transfer rate for 5 MiB data is estimated at 82.22% as a result of the following calculation: ((28.580 s - 5.080 s) / 28.580 s × 100).

1(c). Reducing the node count from 3 to 2 would reduce the download time as well as the tor anonymity.

If 3 nodes are used:

$3T = (3L + 51{,}200 / R) \Rightarrow 3T = 3 \times (5.080\ s) + 51{,}200 /\ 223{,}154\ bytes\ per\ s \approx 15.460\ s$

If 2 nodes are used:

$2T = (2L + 51{,}200 / R) \Rightarrow 2T = 2 \times (5.080\ s) + 51{,}200 /\ 223{,}154\ bytes\ per\ s \approx 10.380\ s$

1(d). Increasing the nodes from 3 to 4 slows down the tor performance because the data would have not only to travel through one more additional node but also to wait for that additional node to decrypt/encrypt data, introducing additional latency to the traffic compared with the 3 nodes.

Regarding privacy, increasing the nodes from 3 to 4 does not provide any additional benefits, because an attacker only needs to take control over both the guard (first) and exit (last) nodes to de-anonymize the tor user. Therefore, increasing the number of nodes between the first and last nodes does not protect the privacy of the user if the first and last nodes are compromised.

2(a). k-anonymity should be used instead of Secure Multiparty Computation (SMPC). With k-anonymity, we can ensure that no data record in the database can be traced back and identify the user by removing or generalizing Personal Identifiable Information (PII) and generalizing the suppressing quasi-identifiers. The resulting data after the anonymization with k-anonymity still remain usable for improving its recommendation algorithm for the failing streaming entertainment company. The SMPC method is not compatible with or useful for what we are trying to achieve here. There are no multiple parties joining different data to compute a known function.

2(b). Private Information Retrieval (PIR) should be used instead of Differential Privacy (DP) and is a mechanism to hide queries from the data owner. We can leverage that by making the device fetch relevant data from the database without the database knowing what data we extract. Differential privacy is unfit for this case, because the method is used for adding noise to an already existing collection of data, responding with noisy aggregated data to its observers, and protecting the data providers while doing so.

2(c). Private Information Retrieval (PIR) should be used instead of k-anonymity. Utilizing PIR, we can hide the domain(s) we are interested in from the cooperating DNS server so that they cannot cybersquat. The k-anonymity method is not relevant for this case, because we do not need to anonymize a dataset.

2(d). Secure multiparty computation (SMPC) should be used instead of DP. SMPC would be used to protect the data of individuals and hospitals in order to determine and only respond with

whether or not an individual and someone who is infected is living in the same apartment as a definitive result. Differential privacy is unfit for this case, because it gives back noisy data as a result of a dataset query. While noisy data might be useful for aggregated data statistics, it goes against the goal of determining whether or not a person is living in the same apartment with someone who is infected with the CROW disease.