CS 6431: Security and Privacy Technologies
Homework #4

Due: Before class on April 25, 2018

Instructors: Tom Ristenpart and Vitaly Shmatikov

Problem 1: Consider a synchronous mixnet X that takes in ciphertexts on a set of messages $M = \{m_1, m_2, \ldots, m_n\}$ and outputs a set of messages $M' = \{m'_1, m'_2, \ldots, m'_n\}$, where messages are in $\{1, \ldots, q-1\}$ for some prime q (as for El Gamal plaintexts).

(a) A verifiable mixnet X is supposed to prove that $M = M'$. As discussed in class, a naïve way one might try to do this is to show that every $m_i \in M$ has a corresponding $m'_j \in M'$ such that $m_i = m'_j$. Why doesn't this approach work? (Give an example of an (M, M') such that $M \neq M'$ but a valid proof can be constructed.)

(b) Suppose that M consists of "yes" / "no" votes encoded respectively as messages $m = 1$ and $m = 2$. For how big a gap in "yes" / "no" votes can an adversarial mix X in principle change the outcome of an election if it uses the proof in (a)?

(c) Consider an alternative "proof" that shows that for any pair $(m_i, m_j) \in M \times M$, there exists a pair $(m'_u, m'_v) \in M' \times M'$ such that $m_i m_j = m'_u m'_v$. Does this fix the problem in (a)? If so, explain why. If not, show a counterexample using message values $m = 1$ and $m = 2$.

Problem 2: Consider the threat model for anonymous communications in which the adversary can monitor all network connections and control all but one server used in an anonymity system. Even with just two participants, the adversary should not be able to tell whether they are communicating, or not. For example, this is the threat model targeted by Vuvuzela, but not Tor.

- Explain attacks against Tor in this threat model.
- Does Tor's route selection algorithm make it more or less vulnerable vs. purely random selection of relays? Explain.

Problem 3: Recall the standard two-server IT-PIR scheme. Replicate a database of values $V_1, \ldots, V_n$ across two servers. A client wishing to retrieve V_i first chooses a random vector $R \leftarrow_{\$} \{0,1\}^n$. They let X_i be the one-hot vector that has $X_i[i] = 1$ and $X_i[j] = 0$ for $j \neq i$. That is, the ith component of X_i is set to one, and everywhere else is zero. Then the client sends one server $S_1 = R$ and the other server $S_2 = R \oplus X_i$. (The client chooses randomly which server each value is sent to.) The servers respond with $Y_b = \bigoplus_{j \,.\, S_b[j]=1} V_j$ for $b \in \{1,2\}$. The client finally computes $Y_1 \oplus Y_2$, which equals V_i.

Consider the following performance improvement. Choose R uniformly from the set of all one-hot vectors. Otherwise repeat the protocol as before.

- Describe the way in which the new protocol improves performance.
- Assume a client has queries for a single uniformly chosen index i a total of q times. With what probability does the adversary learn i when $q = 1$ (a single query)? What about after $q > 1$ queries?
- How might you adapt the performance improvement to achieve a mechanism that you can formally analyze to be ϵ-differentially private? Your solution should be faster (for servers) than the original IT-PIR scheme (which is trivially ϵ-differentially private for $\epsilon = 0$). Explain the performance achieved. Partial credit will be given for attempts; explain if you can why an analysis fails.

Problem 4: Many censorship circumvention tools tunnel communications over TLS. A censor might try to develop tests that seek to identify this tool by passively inspecting application layer headers or actively probing a destination IP/port. Discuss how you would go about determining both passive and active fingerprinting attacks against a given TLS-using tool (for which you have the ability to run the code). Discuss how you would measure its efficacy.

Problem 5: Dissent and Riposte aim to solve the problem of anonymous broadcast. Suppose you want instead to build a system for anonymous messaging between any two parties. How would you modify the design of Dissent or Riposte (your choice)? Which assumptions are no longer needed? How can you gain additional efficiency?