

Homework 4
18-739: Foundations of Security and Privacy
Due: Thu Nov 17, 2011

Course: 18739, Carnegie Mellon University
Instructor: Professor Anupam Datta
Teaching assistant: Divya Sharma

Problem 1 [2.75pts]: De-anonymization Attacks

In the lecture on October 18, we discussed the definition of k-anonymity and a class of de-anonymization attacks that works on high-dimensional databases.

A central desideratum for database privacy definitions is that they be robust in the presence of adversaries with auxiliary information. For the first two questions below, please refer to the patient microdata table and its *k-anonymized* version in Figures on slide 9 from the October 18 lecture.

1. State clearly what privacy guarantee the table in Figure 2 provides in the presence of an adversary who has *no auxiliary information*. [0.75pt]
2. Explain how the privacy guarantee from the previous part can be compromised by an adversary with *auxiliary information* by demonstrating 2 different kinds of attacks. In presenting the attacks, be sure to identify the auxiliary information precisely. [1.5pt]
3. Examine why k-anonymity as a definition does not work for high dimensional databases with a short informal argument (2-3 sentences). [0.5pt]

Problem 2 [0.75 × 5 = 3.75 points]: Computational Soundness

This problem requires you to walk through the proof of the computational soundness theorem of the Abadi-Rogaway logic of encrypted expressions, which was discussed in class. Consider the following two expressions M and N . Assume that Π is a type-0 secure encryption scheme. We want to show that $\llbracket M \rrbracket_{\Pi} \approx \llbracket N \rrbracket_{\Pi}$.

$$M = \{K_3\}_{K_1} \{K_2\}_{K_1} \{11\}_{K_2} \{K_5\}_{K_4} K_4 \{00\}_{K_2} \{101\}_{K_5}$$
$$N = \{K_3\}_{K_1} \{K_2, 00\}_{K_3} \{K_5\}_{K_3} \{K_5\}_{K_4} K_4 \{K_6, K_6\}_{K_2} \{101\}_{K_5}$$

1. Write down the patterns for M and N and verify that they are equal.

2. Renumber the keys in M and N so that “deeper” keys have lower numbers to get the expressions M' and N' . Explain why $pattern(M') = pattern(N')$. State a syntactic restriction on expressions that is essential for this renumbering to be possible.
3. Write down the hybrids M_j and N_k where $0 \leq j \leq m$ and $0 \leq k \leq n$, such that $M_m = M'$, $N_n = N'$ and $M_0 = N_0 = pattern(M') = pattern(N')$.
4. To do the reduction proof, assume that $\llbracket M' \rrbracket_{\Pi} \approx \llbracket N' \rrbracket_{\Pi}$ does not hold. Thus, there exists a PPT adversary that can distinguish between $\llbracket M' \rrbracket_{\Pi}$ and $\llbracket N' \rrbracket_{\Pi}$ with non-negligible probability. Explain why this implies that there exists an adversary that has a non-negligible probability of distinguishing between ensembles of at least one pair of adjacent hybrids from the previous step.
5. Explain how an attacker that has a non-negligible probability of distinguishing between ensembles of adjacent hybrids can be used to construct another attacker that breaks the type-0 security of Π .

Problem 3 [3.5 points]: Privacy Logics and Auditing

This problem is intended to make you think about the nature of privacy policies and challenges in the expression and enforcement of such policies.

Part a (1 point)

1. As discussed in class during the lecture on Nov 3, explain the conceptual difference between application of runtime monitoring policies and post-hoc auditing.
2. List one scenario each where runtime monitoring or post-hoc auditing would be preferred.

Part b (2.5 points)

This question is about encoding clauses in a privacy policy as positive and negative norms. Consider the action where principal p_1 send a message m to principal p_2 , $\mathbf{Send}(p_1, p_2, m)$ refers to the \mathbf{Send} action with sender p_1 , recipient p_2 and message m . The message m is tagged with the tag t as it contains information about an attribute t about a subject q , and is sent for a purpose u . $\mathbf{tagged}(m, q, t, u)$ expresses this relation.

Consider the following clauses in a disclosure policy and express each of these as a positive norm or a negative norm (whichever one out of the two is applicable) using the privacy logic discussed in class on Nov 3. The policy states that:

Principal p can send a message m , that contains information about attribute t pertaining to subject q to principal p_2 , for the purpose u ,

1. if the disclosure is for treatment purposes
2. if p_1 has obtained consent from q

3. if the message contains a child's information, it can be sent to parents/guardians of the child *only if* the transmission is not prohibited by other laws
4. if the message contains psychotherapy notes about patient q , the disclosure is permitted *only if* q authorizes the disclosure.

You are required to do the following:

- Encode each of these clauses as positive or negative norms ϕ_1, ϕ_2, ϕ_3 and ϕ_4 as mentioned above
- Express the entire policy as:

$$\forall p_1, p_2, m, t, u, q. (\mathbf{Send}(p_1, p_2, m) \vee \mathbf{tagged}(m, q, t, u)) \Rightarrow f(\phi_1, \phi_2, \phi_3, \phi_4)$$

Express $f(\phi_1, \phi_2, \phi_3, \phi_4)$ using conjunctions/disjunctions of the norms specified by $\phi_1, \phi_2, \phi_3, \phi_4$.

Note: You can define any predicates you need while encoding the above clauses, for instance, in order to encode the statement that purpose u of a disclosure is for billing purposes, you can write $\mathbf{purpose}(u, \mathit{billing})$.