

18739A: Foundations of Security and Privacy  
Final Exam  
(Due: Dec 8, 2011 11.59PM PST)

December 5, 2011

**Problem 1 (2.0 points)**

In this question, we will take a closer look at a mechanism for achieving  $\epsilon$ -*differential privacy* for query functions  $count_P$  (which returns the number of entries in a database with property  $P$ ) and  $max\text{-salary-at-Berkeley}$  (which returns the maximum salary of faculty and staff at UC Berkeley). This mechanism, denoted  $\kappa_f$  for query function  $f$ , computes  $f(X)$  and adds noise with a Laplace distribution with variance  $\sigma^2$  that depends on the *sensitivity* of the function and the differential privacy parameter  $\epsilon$ . State precisely any additional assumptions you make to answer the questions below.

1. Calculate the sensitivity  $\Delta count_P$  and  $\Delta max\text{-salary-at-Berkeley}$  for the query functions. Calculate the variances of the distributions from which noise is added in order to achieve  $\epsilon$ -*differential privacy* with  $\epsilon = 0.00001$  and  $\epsilon = 0.001$  for both query functions.

Now let's try to understand these results qualitatively. For the "same level of privacy", which function requires "more noise" to be added? For a fixed function, how does the "noise distribution change" in order to achieve "higher levels of privacy"? Your answers should make precise all the terms in quotes.

2. Let  $D_1 = \{A_P, B, C_P\}$  and  $D_2 = \{A_P, B, C_P, D_P\}$  be two neighboring databases where an element of the form  $X_P$  has the property  $P$  (e.g.,  $A_P$  has property  $P$ , but  $B$  does not). Calculate the values of  $count_P(D_1)$  and  $count_P(D_2)$ .

Recall that  $\kappa_f(D) = f(D) + Y$  where  $Y \sim Laplace(\Delta f/\epsilon)$ , i.e.,  $Y$  is a random variable that follows a Laplace distribution with mean 0 and variance  $(\Delta f/\epsilon)^2$ . Calculate  $Pr(\kappa_{count_P}(D_1) \in (2.3, 2.7])$  and  $Pr(\kappa_{count_P}(D_2) \in (2.3, 2.7])$  and verify that their ratio is less than  $e^\epsilon$ . Notice that you are verifying that the differential privacy definition is satisfied for  $S = (2.3, 2.7]$ . You can use  $\epsilon = 0.001$  in your calculations.

*Hint:* Look up the Wikipedia page for Laplace distribution.

**Problem 2 ( $0.5 \times 3 = 1.5$  points)**

This question requires you to understand the reading and lecture based on 'No Free Lunch in Data Privacy', discussed in class on Nov 29, 2011. List three scenarios where a differentially private

mechanism will not be effective in providing privacy guarantees and also justify why it doesn't work.

### Problem 3 (3.0 points)

This question requires you to understand the lecture based on 'Purpose Restrictions in Privacy Policies: Semantics and Audit', discussed in class on Nov 10, 2011.

1. Consider an MDP model modeling a scenario similar to the model in the empirical studies discussed in class (Slides 43-44):

Metropolis General Hospital and its employees follow the policy that they will share a patient's medical record with an outside specialist only for the purpose of providing that patient with treatment. A case worker employed by Metropolis General Hospital meets with a patient. The case worker develops a plan with the sole goal of treating the patient (we call this as purpose  $p$ ). The plan includes sharing the patient's medical record with an outside specialist. Consider two possible paths through the corresponding MDP- Path  $P_1$  describes the situation when the case worker sent the record to Specialist (we call this send action as  $a_1$ ) and the patient was treated successfully. Path  $P_2$  describes the situation when the case worker sent the record to Specialist (we call this send action as  $a_1$ ) and the patient was not treated successfully. Path  $P_3$  describes the situation when the case worker sent the record to the Insurance company (we call this send action as  $a_2$ ) and the patient was not treated successfully. Path  $P_4$  describes the situation when the case worker sent the record to the Insurance company (we call this send action as  $a_2$ ) and the patient was treated successfully. The case worker sent the record to the Insurance company in return for money. Based on this model, answer the following questions:

- (a) Is action  $a_1$  in path  $P_1$  for purpose  $p$ ? Why/ why not?
  - (b) Is action  $a_1$  in path  $P_2$  for purpose  $p$ ? Why/ why not?
  - (c) Is action  $a_2$  in path  $P_3$  for purpose  $p$ ? Why/ why not?
  - (d) Is action  $a_2$  in path  $P_4$  for purpose  $p$ ? Why/ why not?
2. The audit algorithm proposed in the work discussed in class, actually computes approximately optimal plans instead of exactly optimal plans. Does this approximation cause false positives (an action that did not violate the policy is flagged as a violation by the algorithm)/ false negatives (an action that violated the policy was not flagged as a violation by the algorithm) for "only for" and "not for" policies? Explain your answer.

### Problem 4 (3.5 points)

This problem shows the construction of an example estimator that can be used in RMA (as discussed in class on Nov 8, 2011). Remember that in RMA for each round of auditing we have to estimate the loss obtained if some other inspection action, other than the one actually used, had been played. Consider a repeated game between the Organization ( $Org$ ) and the Employee ( $Emp$ ). Suppose that in each round of the game  $Emp$  does 10 tasks.  $Org$  has 11 actions: do  $i$  inspections, where

$i \in \{0, 1, \dots, 10\}$ . *Emp* has 11 actions: Commit  $j$  violations, where  $j \in \{0, 1, \dots, 10\}$ . Assume the inspections are done by picking tasks to inspect at random without replacement. Based on these assumptions, answer the following questions:

1. Given that in round  $t$ , *Org* does  $i^t$  inspections and *Emp* does  $j^t$  violations, write down the probability mass function for the random variable  $V^t$ , where  $V^t$  represent the number of violations caught internally. Also, write down the expression for expected number of violations caught internally in terms of  $i^t$  and  $j^t$  (no need to derive the expected value of  $V^t$ ). (Hint: this is a hypergeometric distribution)
2. Violations can also be detected externally. The model for such detection is simple: any internally undetected violation has a probability  $p$  of being detected externally. Then, given  $j^t$  represents the total number of violations in round  $t$ , show the expected number of violations detected outside externally conditioned on  $j^t$  is  $p(j^t - E[V^t])$ .
3. Suppose we have a simple loss function  $L(v_{int}, v_{ext}) = 100 * v_{int} + 1000 * v_{ext}$ . Then the expected loss is given by  $100 * (E[V^t]) + 1000 * (p(j^t - E[V^t]))$ . Show that the expected loss is a function of  $i^t$  and  $j^t$ , and is linear in  $j^t$ . Observe that in this expression  $j^t$  is not known. Thus, by estimating  $j^t$  we can estimate the expected loss for different values of  $i^t$ .
4. Using the first part show that the estimate  $\hat{j}^t = 10 * V^t / i^t$  for  $j^t$  satisfies  $E[\hat{j}^t] = j^t$ . (Such estimators are called unbiased).
5. Suppose  $i^t = 5$  and  $V^t = 3$ ; now use the estimator  $\hat{j}^t$  derived in last part to write down the expected loss in terms of  $i^t$ . Now, plug in  $i^t = 6$  to get an estimate of the loss if  $i^t = 6$  would have been played. Assume  $p = 0.5$ .