CS 6431 - Security and Privacy Technologies
Fall 2014

Homework #4

Due: 7:30pm EST, December 3, 2014

NO LATE SUBMISSIONS WILL BE ACCEPTED

YOUR NAME: _____

Collaboration policy

No collaboration is permitted on this assignment. Any cheating (*e.g.*, submitting another person's work as your own, or permitting your work to be copied) will automatically result in a failing grade.

Homework #4 (50 points)

Problem 1 (5 points)

Is Tor more vulnerable to traffic correlation attacks than the original onion routing or less? Hint: consider the differences in circuit construction.

Problem 2 (10 points)

Pick any Tor hidden service and use it. Record all IP addresses seen by your browser as you are using the service. Do all of them belong to Tor relays? Write down the name of the service, names of the relays, and any non-relay IP addresses you see.

Problem 3

Problem 3a (4 points)

What is the difference between "pseudonymity" and "anonymity"?

Problem 3b (4 points)

What is the most common way for websites to leak users' personal identifiers such as usernames and email addresses to third-party trackers, thus enabling the latter to identify pseudonymous profiles? Explain exactly how the leakage occurs.

Problem 4 (5 points)

If you are a third-party tracker whose iframe is included into multiple website, does canvas fingerprinting eliminate the need for third-party cookies? Explain.

Problem 5 (7 points)

For each of the three Web tracking mechanisms analyzed in "The Web never forgets" paper, explain whether tracking is possible if the user is browsing through Tor and, if no, why.

Problem 6

D is the dataset containing annual salaries of all Cornell employees. $bsdcount(D)$ returns the number of entries in D that are greater than $\$1,000,000$; $max(D)$ returns the maximum salary in the dataset.

Let San be the standard Laplacian mechanism for ϵ-differential privacy. Given any function f, San generates random ξ from the Laplacian distribution with variance that depends on the sensitivity of function f and the privacy parameter ϵ, and returns $f(D) + \xi$.

Problem 6a (5 points)

What is the sensitivity of *bsdcount* and max? State all assumptions you needed to calculate the answers.

Problem 6b (5 points)

For the "same level of privacy," which function requires "more noise" to be added? Given a function, how does the "noise distribution change" in order to achieve "higher level of privacy"? Your answers should make precise all terms in quotes.

Problem 6c (5 points)

Let $\epsilon = 0.001$, let $p = 0.01$ be your prior probability that Big Red Bear makes $\$10,000$ a year, and let p' be the probability after learning the differentially private values of *bsdcount* and max. What is the maximum value of p'?