IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2018-2019

MEng Honours Degree in Electronic and Information Engineering Part IV
MEng Honours Degree in Mathematics and Computer Science Part IV
MEng Honours Degrees in Computing Part IV
MSc in Advanced Computing
MSc in Computing Science (Specialist)
for Internal Students of the Imperial College of Science, Technology and Medicine

This paper is also taken for the relevant examinations for the
Associateship of the City and Guilds of London Institute

PAPER C408

PRIVACY ENGINEERING

Wednesday 12th December 2018, 14:00
Duration: 120 minutes

Answer THREE questions

Paper contains 4 questions
Calculators not required

1a For the simple secure voting protocol, a voter (Alice) could compute the sum of the other votes. Isn't this a problem? What if Alice chooses her shares so that they add up to 2?

b Produce the garbled table for a NAND gate with input wires *w1* and *w2* and output wire *w3* where the *p*-bit is 1 for all three wires. Clearly show your solution on a diagram for the NAND gate, labelling the wires with their key/index-bit pairs as well as showing the encrypted table entries.

c Yao's garbled circuit construction assumes that the circuit generator (Alice) is semi-honest. Give an attack that Alice could make if she was malicious. How could the scheme be adapted to detect if the circuit evaluator (Bob) sends a fake result to Alice?

d Consider the following *1*-from-*2* oblivious transfer protocol that uses a trusted third-party Trent in an *honest-but-curious* setting. Alice's messages M_0 and M_1 are binary values of length *k*. Bob's message selection bit is *b*.

1.	Trent → Alice: R_0, R_1	Random binary values each of length *k*
2.	Trent → Bob: t, R_t	Random bit *t*
3.	Bob → Alice: *e*	$e = t \oplus b$
4.	Alice → Bob: $C_0,\ C_1$	$C_0 = M_0 \oplus\ R_e,\ C_1 = M_1 \oplus R_{1\text{-}e}$
5.	Bob: M_b	$M_b = C_b \oplus R_t$

i) For this protocol show the working for M_0=1101, M_1=0100, *b*=1, *t*=0, R_0=0101, R_1=0011.

ii) Explain how the protocol satisfies the properties of an oblivious transfer. Could Alice or Bob learn anything if they were malicious?

iii) Adapt the protocol to 1-from-*n* oblivious transfer. You can assume that *n* is a power of 2, i.e. 2, 4, 8, …

The four parts carry, respectively, 15%, 20%, 15%, and 50% of the marks.

2a Using a diagram describe how keys are constructed for proxy key encryption schemes and how they can be used to privately share data using an *honest-but-curious* server. Give three advantages of proxy-key schemes.

b Use the proxy key encryption scheme in part (a) to devise and describe a cryptographic scheme to perform encrypted keyword searches for encrypted documents held by a database running on an untrusted server.

In your description assume Alice inserts a new document and its associated keywords while Bob searches for documents with a particular keyword. You can assume that keys have already been computed as in part a.

Hint: Search on encrypted hashes of keywords and encrypt documents with a random symmetric key for better performance.

c In *functional encryption*, Alice can encrypt data with a master public key *mk* and also use it to compute private function keys for functions, for example *fk* for function *f*. Anyone given the private function key *fk* and the ciphertext c=$\mathrm{E}_{mk}(p)$ for some plaintext p can then compute $\mathrm{D}_{fk}(c)$=$f(p)$ without learning any other information about p. Explain how this could be used by an email service to privately filter spam encrypted emails sent to users.

The three parts carry, respectively, 40%, 50% and 10% of the marks.

3 Data anonymization

a i) What is an attribute? What is a quasi-identifier? What is a uniqueness attack?

ii) Define *k*-anonymity. How does it help prevent uniqueness attacks?

iii) What attack does *k*-anonymity not prevent but is prevented by ℓ-diversity? Give an example of a *k*-anonymous dataset that is vulnerable to such attack

b i) How does a unicity attack differ from a uniqueness attack? Can *k*-anonymity be applied to prevent unicity attacks?

In the following, you will compute ε_k, the expected unicity of a dataset given *k* points. Let *D* be a mobility dataset, containing points (position, time) for *U* different people. There are *M* different positions, and the dataset is collected over a period of *T* (discrete) timesteps. We assume that the users in the dataset follow some simple rules:

- At each timestep, each user is at one location taken uniformly at random in {antenna$_1$, …, antenna$_M$}, independently of her previous location and of other users.

- Each user is at one and only one place at each timestep.

ii) Let $P = (p_1, \ldots, p_k)$ a tuple of points for a random user, that is $p_i = (l_i, t_i)$ with l_i the antenna and t_i the timestamp. How many values can *P* possibly take? Does the length of the time period *T* matter?

iii) Let $P = (p_1, \ldots, p_k)$ a tuple of points for a random user. What is the probability that one other random user matches these *k* points?

iv) Using this result, what is the probability that one user is different from all other users in the population, according to *k* points randomly extracted from his or her trace? (This is the expected unicity of the population)

v) In order for this model to be more realistic, we modify the second rule: at each timestep, a user has a point with probability λ (independently of all other points and users). We call λ the sparsity of the dataset. Compute the unicity in this model. How does the expected unicity vary with the sparsity λ?

The two parts carry, respectively, 40%, and 60% of the marks.

4 Query-based systems and differential privacy

a Describe the technique of query set size restriction (QSR) for counts. What kind of attacks does QSR protect against? Give an example. What kind of attack does QSR not protect against? Give an example.

b **Sensitivity**. Let $f : \mathbb{D} \rightarrow \mathbb{R}$. Give the definition of the global sensitivity Δf of f. Explain briefly why the global sensitivity is useful for differential privacy.

c. **Differentially private proportions.** Let $\mathbb{D}$ be the family of all nonempty datasets with only one binary attribute (HIV). Suppose you have a dataset $D \in \mathbb{D}$, which is a set of records. Suppose that you want to release the proportion of users that have HIV=1.

More precisely, you are interested in the function $f : \mathbb{D} \rightarrow [0,1]$ defined as: $f(D) = |D_{\text{HIV}}{=}1| \,/\, |D|$ where D_{HIV}=1={records in D that have attribute HIV=1}.

i) Show that the global sensitivity Δf of f is greater or equal to 0.5. Hint: take two small neighboring datasets

ii) Consider the mechanism $M1 : \mathbb{D} \rightarrow \mathbb{R}$ defined as $M1(D) = f(D) + \text{Lap}(\Delta f \,/\, \varepsilon)$. Is this mechanism differentially private? Do you think it is a good mechanism for small epsilons (say, $\varepsilon = 0.1$)?

iii) Design a better ε-differentially private mechanism $M : \mathbb{D} \rightarrow \mathbb{R}$ for the function f. Explain why it is ε-differentially private.

iv) Consider the mechanism $M2 : \mathbb{D} \rightarrow \mathbb{R}$ defined as $M2(D) = (|D_{\text{HIV}}{=}1| + \text{Lap}(1/\varepsilon)) \,/\, |D|$. Is this mechanism differentially private?

Hint: To prove that a mechanism M is not differentially private, one can find two specific neighboring datasets D and D' and a specific subset $S \subseteq \mathbb{R}$ such that $\Pr[M(D) \in S] \nleq \exp(\varepsilon) \Pr[M(D') \in S]$. It is generally a good a idea to start looking at very small datasets. For example, you could take some D and D' of size 1 and 2 respectively. As for S, it is probably a good idea to take something simple, such as an interval. To compute the two probabilities, you might find this useful: $\Pr[\text{Lap}(b) > x] = ½ \exp(-x/b)$ for all $x \geq 0$.

The three parts carry, respectively, 30%, 10%, and 60% of the marks.