CS 5473 FINAL EXAMINATION

April 16th, 2014

Duration: Three hours

Name:

Student Number:

Note: This is a close notes and books exam. However, each student is allowed to bring an information sheet. There are total 6 problems to be solved. If the space is not enough for your answer, you can use the back of the pages. Documentation of the solutions are also important. If your solutions or written are difficult to read, then you may lose marks.

Problem 1. (15 marks)

In each of the following questions, cycle the correct answer.

1. An authorization table in a DBMS is used for
   (a) authenticating a client.
   (b) authenticating a server.
   (c) deciding if a user's query can be executed.
   (d) authenticating a table.

2. A Bigtable is a sparse map, because
   (a) it is very big.
   (b) it is persistent multi-dimensional.
   (c) the index can be arbitrary strings.
   (d) it is also indexed by a timestamp.

3. In a SQL injection attack,
   (a) the attacker inserts some codes into a SQL query.
   (b) the attacker tries to insert his own SQL commands in a data stream.
   (c) the attacker inserts his own data to a table.
   (d) the attacker tried to insert some harmful information to the database.

4. To avoid information leakage during trust negotiation,
   (a) a digital credential should include some unrelated information.
   (b) a party should not directly tell the other party the non-possession of a sensitive credential.
   (c) we should reduce the communication rounds between the two parties.
   (d) we should encrypt all the negotiations.

5. In RT statement: $A.r \leftarrow A.r_1.r_2 \cap B.r_3$, A asserts that
   (a) $A.r$ includes principal who is a member of both $A.r_2$ and $B.r_3$.
   (b) $A.r$ includes principal who is a member of both $A.r_1$ and $B.r_3$
   (c) $A.r$ includes principal who is a member of both $A.r_1$ and $B.r_3$ or of both $A.r_2$ and $B.r_3$
   (d) $A.r$ includes principal $C.r_2$ where C is a member of $A.r_1$ and $C.r_2$ is a member of $B.r_3$.

Problem 2. (20 marks)

Briefly answer the following questions:

1. In a Google Bigtable, why the column keys are grouped into column families?

2. Why there are concurrent access tables in DBMS architecture?

3. Explain why using data cubes in data warehouse may avoid many scans of the base tables.

4. Explain why in a policy composition framework, policies are treated as black boxes.

5. Explain why make query size restriction may reduce inference problems.

Problem 3. (25 marks)

Figure 1 shows medical records from a fictitious hospital.

ID	QI			SV
	Zip Code	Age	Nationality	Condition
1	13053	28	Russian	Heart Disease
2	13053	36	Japanese	Flu
3	13068	35	American	Cancer
4	13068	21	Japanese	Viral Infection
5	14850	46	Indian	Flu
6	13058	23	American	Viral Infection
7	14853	50	Indian	Cancer
8	14853	55	Russian	Heart Disease
9	14850	47	American	Viral Infection
10	13058	37	Indian	Cancer
11	13068	36	Japanese	Cancer
12	13068	38	Russian	Flu

Figure 1: Inpatient microdata

Answer the following questions.

1. To develop data anonymity, A and B generalized values of "Age". A generalized the values as $[21, 29], [30, 39], [40, 49], \geq 50$, while B generalized the values as $[21, 29], [28, 38], [35, 40], \geq 46$. Which method is better? Give your explanation. (Note that here we ignore how they generalized other QI values).

2. To generalize the value of “Zip Code”, three methods may be used:

   - Generalize the last digit: e.g., 1305*
   - Generalize the fourth digit: e.g., 130*3
   - Generalize the last two digits: e.g., 130**

   Which method is better? Explain why.

3. Generalize the QI values of the table for 3-diversity.

4. Using anatomy method to develop the 3-diversity for the table.

5. Compare the previous two methods for the 3-diversity and indicate the advantages and disadvantages for the different methods.

Problem 4. (20 marks)

Suppose we have a secure hash function $H()$ and a big database. Answer the following questions about data watermarking.

1. Briefly explain how to watermark numerical data which is acceptable to change ξ least significant bits.

2. For the security of the watermarking, what are the requirements for the hash function?

3. Suppose a database used a distortion-free watermarking. Now an attacker performs an attack by changing the positions of the rows with odd order to the rows with even order, i.e., exchange row 1 and row 2, exchange row 3 and row 4, . . ., etc. Do you think that kind of attack works? Why?

4. Suppose the database contains both numerical and non-numerical data. If we want to use both the numerical data watermarking and distortion-free watermarking on the database, what should we do? Do you think combination of these two methods have advantages? Why?

Problem 5. (10 marks)

In a trust management system, following SDSI name certificates are defined:
$(K_{Alice}$,bob, K_{Bob}, 1)
(K_{Bob}, friend, K_{Carol} friend, 1)

1. In above certificates, 3 public keys are involved. Do you think some or all of these keys require X.509 certificates (or PGP certificates)? Why?

2. Suppose Bob wants to issue a public key K_{Dave} to his friend Dave and Carol wants to issue a public key K_{Edward} to his friend Edward, what should they do? Explain why your method can certificate these two keys.

Problem 6. (10 marks)

Suppose an attacker wants to inference some information from a statistical database. He found that there are two databases provide the same data. But they use different access control method: one uses MAC and another uses RBAC. Which database will you recommend to the attacker? Why?