IS2510/TEL2810 Information Security & Privacy

Homework 3

Total Points: 100

Due Date: by noon of April 23, 2013 (the solutions will be posted after that)

Q1: Assume that Alice and Bob are friends. Consider two conflict of class sets COI1 = {X, Y} and COI2 = {U, V}. Let CDX, CDY, CDU, and CDV be the company data sets of companies X, Y, U and V. **(Points 30)**

Show all the possible assignments that are allowed and prohibited by the Chinese wall policy - consider *read* only assignments as well as *read* and *write*.

Consider the following scenarios;

1. Only read accesses need to be provided to CDs
2. Both read and write accesses need to be provided together to any CD.
3. Both read/write required for CDX, and CDV, and read-only to remaining CDs.

Sample Answer

Note that Alice and Bob may not be able to cover all the CDs – for that you can assume there are others who can take on the assignment. You need to only consider Alice and Bob.

Assume that Alice and Bob are friends. Consider two conflict of class sets COI_1 = {X, Y} and COI_2 = {U, V}. Let CD_X, CD_Y, CD_U, and CD_V be the company data sets of companies X, Y, U and V. Illustrate using these, what kind of assignments are prohibited by the Chinese wall policy. (**Score 10**)

(An old solution below; You should be able to actually draw a diagram to illustrate this more easily)

Only read-accesses	
Alice can read only CDx	Bob can read only CDx
Or	Or
Alice can read only CDy	Bob can read only CDy
Or	Or
Alice can read only CDu	Bob can read only CDu
Or	Or
Alice can read only CDv	Bob can read only CDv
Or	Or
Alice can read CDx and CDu	Bob can read CDx and CDu
Or	Or
Alice can read CDx and CDv	Bob can read CDx and CDv
Or	Or
Alice can read CDy and CDu	Bob can read CDy and CDu
Or	Or
Alice can read CDy and CDv	Bob can read CDy and CDv

Read accesses can be any combination of 1 from left side (Alice permissions) and 1 from right side (Bob's permissions). Neither Bob nor Alice can read both CD's from the same COI.

Both read and write accesses need to be provided together to any CD.

Alice can r/w only CDx	Bob can r/w only CDx
Or	Or
Alice can r/w only CDy	Bob can r/w only CDy
Or	Or
Alice can r/w only CDu	Bob can r/w only CDu
Or	Or
Alice can r/w only CDv	Bob can r/w only CDv

Read/Write accesses can be any combination of 1 from left side (Alice Permissions) and 1 from right side (Bob's Permissions). Both Alice and Bob can't have r/w permissions for more than 1 company data set.

Both read/write required for *CDx*, and *CDv*, and read-only to remaining CDs.

Alice can r/w to CDx and Bob can r/w to CDv, but they are not permitted to read the other remaining CDs.
Or
Alice can r/w to CDv and Bob can r/w to CDx, but they are not permitted to read the other remaining CDs.
Or
Alice and Bob can both r/w to CDx, but neither of them could r/w to CDv or read other CDs.
Or
Alice and Bob can both r/w to CDv, but neither of them could r/w to CDx or read other CDs.
Or
Alice can r/w to CDx and Bob can read-only CDy or CDu or both (since CDy and CDu are in different COI's).
Or
Bob can r/w to CDx and Alice can read-only CDy or CDu or both (since CDy and CDu are in different COI's).
Or
Alice can r/w to CDv and Bob can read-only CDy or CDu or both (since CDy and CDu are in different COI's).
Or
Bob can r/w to CDv and Alice can read-only CDy or CDu or both (since CDy and CDu are in different COI's).

Q2: Let the following be the role hierarchy relationships, **(Points 10 + 20)**

RH = $\{(r_1, r_2), (r_1, r_3), (r_2, r_4), (r_2, r_5), (r_3, r_5), (r_3, r_8), (r_6, r_3), (r_6, r_7), (r_7, r_8)\}$.

Note that (r_i, r_j) means r_i is senior of r_j. Let each role be assigned to exactly one unique user. We can assume that u_i is assigned to role r_i. Further, note that if u is assigned to r that means u is authorized for r. Based on these, answer the following.

1. Find *authorized_users*(r_5) and *authorized_users*(r_4)?
2. For each of following SSD constraints, state if the above hierarchy and the user assignments would result in a conflict (Consider each of these individually only)
   - ($\{r_2, r_7\}$, 2)
   - ($\{r_2, r_3, r_4, r_7\}$, 3)

Sample Answer:

Authorized_users(r5) = {u1, u2, u3, u5, u6}

Authorized_users(r4) = {u1, u2, u4}

b. ({r2, r7}, 2) : first compute Authorized_user(r2) and Authorized_user(r7) - you will see that there are no common users that are authorized for both r2 and r7. Hence, we can add this constraint in the policy without conflicting with the hierarchy.

({r2, r3, r4, r7}, 3): here you need to show that any combination of the 3 roles from the set should not have a common authorized user - for this SSD constraint to be not conflicting with the hierarchy. So compute the authorized users for each of these. Then we can show that if you take r2, r3, r4, you will find u1 is authorized for all these three roles. Hence this SSD conflicts with the hierarchy.

Q3: **[40 Points]**

Suppose the following medical record dataset has been published.

SSN	Name	Race	DateOfBirth	Sex	ZIP	Marital Status	HealthProblem
		asian	09/27/64	female	94139	divorced	hypertension
		asian	09/30/64	female	94139	divorced	obesity
		asian	04/18/64	male	94139	married	chest pain
		asian	04/15/64	male	94139	married	obesity
		black	03/13/63	male	94138	married	hypertension
		black	03/18/63	male	94138	married	shortness of breath
		black	09/13/64	female	94141	married	shortness of breath
		black	09/07/64	female	94141	married	obesity
		white	05/14/61	male	94138	single	chest pain
		white	05/08/61	male	94138	single	obesity
		white	09/15/61	female	94142	widow	shortness of breath

Considering the publicly available voter registration list below, answer the following questions.

Name	Address	City	ZIP	DOB	Sex	Party	.....
.....	.....	.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....	.....	.....
Sue J. Carlson	900 Market St.	San Francisco	*94142*	*9/15/61*	*female*	democrat	.....
.....	.....	.....	.....	.....	.....	.....	.....

a. Answer the following questions:
   i. What kind of anonymization has been performed on the released dataset before being published, if any?
   ii. Explain if it is possible to infer any privacy-sensitive information from the released data using the voter list.

b. Consider the medical record dataset above. Given attributes *Race, DOB,* and *Sex* together as the quasi-identifier, and attribute *Health Problem* as the sensitive attribute, create and report a 2-anonymous version of the medical dataset (*k*-anonymity where *k=2*). How many equivalency classes are there in the result dataset?
   Note: There is no unique answer to this question. However, your result should have fairly low information loss.

c. Repeat the previous exercise for attributes *Sex*, *ZIP*, and *Marital Status* as quasi-identifier with $k = 3$. Explain what the l value of the result is in terms of the l-diversity principle (considering distinct l-diversity)?

Sample Answer (next page)

4. A

i. Naïve anonymization, i.e., only (explicit) identifier attributes have been removed from the table

ii. The only medical record that matches the highlighted record in the voter registration table based on quasi-identifiers DOB, Sex, and ZIP is the last one. Therefore, using such a quasi-identifier, one can know about Sue Carlson's health problem and marital status.

(Mentioning only about ZIP/DOB as quasi-identifier and health problem as sensitive field is a satisfactory answer)

4. B. 5 equivalency classes

PLEASE NOTICE THAT ALL THE COLUMNS NEED TO BE REPORTED, UNLESS THEY ARE EXPLICIT IDENTIFIERS!

SSN	Name	Race	DateOfBirth	Sex	ZIP	MaritalStatus	HealthProblem
		asian	09/64	female	94139	divorced	hypertension
		asian	09/64	female	94139	divorced	obesity
		asian	04/64	male	94139	married	chest pain
		asian	04/64	male	94139	married	obesity
		black	03/63	male	94138	married	hypertension
		black	03/63	male	94138	married	shortness of breath
		black	09/64	female	94141	married	shortness of breath
		black	09/64	female	94141	married	obesity
		white	1961	not_released	94138	single	chest pain
		white	1961	not_released	94138	single	obesity
		white	1961	not_released	94142	widow	shortness of breath

4. C. If we form a 3-anonymous table as below, the minimum distinct diversity of the sensitive attribute *Health Problem* is 3 in all the three equivalency classes. Therefore, we have *l*=3.

Org. RecNo	SSN	Name	Race	DateOfBirth	Sex	ZIP	MaritalStatus	HealthProblem
1			asian	09/27/64	female	941**	been_married	hypertension
2			asian	09/30/64	female	941**	been_married	obesity
7			black	09/13/64	female	941**	been_married	shortness of breath
8			black	09/07/64	female	941**	been_married	obesity
11			white	09/15/61	female	941**	been_married	shortness of breath
3			asian	04/18/64	male	9413*	married	chest pain
4			asian	04/15/64	male	9413*	married	obesity
5			black	03/13/63	male	9413*	married	hypertension
6			black	03/18/63	male	94138	not_released	shortness of breath
9			white	05/14/61	male	94138	not_released	chest pain
10			white	05/08/61	male	94138	not_released	obesity