Theory Homework Assignment 1

Student Name:- Tolulope Olugbenga

Student Number:- 3643581

Questions / Answers

1a. Why data privacy matters to us? Please elaborate your view as detailed as possible in terms of the General Data Protection Regulation (GDPR).

Data privacy is important to us because

- We care - we are responsible for handling people's most personal information.
- This is an opportunity to make privacy central to what we do.
- By not handling personal data properly we could put individuals at risk and the entities reputation at stake.
- Getting it wrong could result in significant fines.
- We need robust systems and processes in place to make sure we use personal information properly and comply.

General data protection regulations (GDPR) which has come in to force on 25 of May 2018, is a European law that will replace the current Data Protection Act and the UK government will still implement the rules after Brexit. The aim is to strengthen and unify personal data protection for all individuals living in the European Union. In addition, the Information Commissioners Office (ICO) will lead on GDPR in the UK and will hand out penalties for organizations who are in breach of the new law.

1b. What are *k*-anonymity, *l*-diversity, and *t*-closeness in database privacy?

k-anonymity: Table T satisfies k-anonymity with regards to quasi-identifier QI if each tuple in (the multiset) T[QI] appears at least k times. Table T' is a k-anonymization of T if T' is a generalization/suppression of T, and T' satisfies k-anonymity. k-anonymity requires each tuple in (the multiset) T[QI] to appear k times, but does not say anything about the sensitive attribute values. If (almost) all sensitive attribute values in a QI group are equal, privacy is lost. This is homogeneity attack.

l-Diversity: a table is l-diverse if each of its QI groups contains at least l well-represented values for the sensitive attributes.

t-Closeness: a table has t-closeness if in each of its QI groups, the distance between the distribution of SA values in the group and in the whole table is no more than threshold t.

1c. What is differential privacy technique? Please describe the steps on how to add the proper Laplace noise to obtain the desirable privacy for the released dataset.

i.) The DP can guarantee that the privacy risk should not substantially increase as a result of participating in a statistical database. For example, there are two datasets X and X', and X is a neighbor of X' because they differ in one row. However, from the released statistics, it is hard to distinguish X and X'.

ii.) Let f(x) be the result of x and $\lambda = SF/\varepsilon$ be the noise.

$f(x)' = f(x) + Lap(\lambda)$

$= f(x) + Lap(SF/\varepsilon)$

1d. Describe the Big Data 4V's characteristics, including volume, velocity, variety, and veracity, as detailed as possible.

- Volume: Data volume is increasing exponentially, e.g., 44x increase from 2009 to 2020 and from 0.8 zettabytes to 35zb.

- Velocity: Data are generated fast and need to be processed fast.

- Variety: Different types of data are involved, including relational data, text data, graph data, etc., which become more complex. All these types of data need to be linked together.

- Veracity: The data inconsistency and incompleteness, ambiguities etc. will bring some uncertainty in big data. Veracity will consider some security issues in big data in order to protect data veracity.

1e. Describe the birthday attack in hash function

Birthday attack is to find two people with the same birthday, which is the same thing as finding a collision for a particular hash function. For example, in a group of 23 randomly chosen people, at least two will share a birthday with probability at least 50%. If there are 30, the probability is around 70%.

1f. Describe the homomorphic encryption technique as detailed as possible

Homomorphic encryption is an encryption that allows computation on cipher texts, which generates an encrypted result, which when decrypted will match the result of the operations as thou they have been executed on the plaintext. The main purpose of homomorphic encryption is to allow computation on encrypted data.

A major application of homomorphic technique can be seen in cloud computing. Due to Homomorphic techniques, the cloud can perform computation for the user and only the encrypted result will be returned.

2. Suppose that a message has been encrypted using DES in ciphertext block chaining mode. One bit of ciphertext in block Ci, and another bit of ciphertext in block Ci+1 are accidentally transformed from 0 to 1 during transmission. How much plaintext will be garbled as a result?

Changing the bit of Ci and Ci+1 during transmission, would cause three plaintext to be garbled Mi, Mi+1, Mi+2.

Mi = Dec(Ci) Xor Ci-1; Mi+1 = Dec(Ci+1) Xor Ci; Mi+2 = Dec(Ci+2) Xor Ci+1;

Mi+3 = Dec(Ci+3) Xor Ci+2

All plaintexts Mi+x such that x > 2 will not be garbled.

3. Using the extended Euclidean algorithm to find the multiplicative inverse of

(a) 12345 mod 54321

a = 12345, m = 54321; a has an inverse if the gcd(a,m) =1, but in this case the gcd is 3. Therefore a has no inverse.

K	0	1	2	3	4	5	6	7
R	12345	54321	12345	4941	2463	15	3	0
Q		0	4	2	2	164	5	
Xk	1	0	1	5	11	27	4439	
Yk	0	1	0	1	2	5	822	

(b) 350 mod 1769

a = 350, m = 1769.

gcd(a, m) = 1

Multiplicative inverse is ((-1)^n) * Xk

((-1)^7) * 652 = -652 mod m

= 1117 mod m

K	0	1	2	3	4	5	6	7	8
R	350	1769	350	19	8	3	2	1	0
Q		0	5	18	2	2	1	2	
Xk	1	0	1	5	91	187	465	652	
Yk	0	1	0	1	18	37	92	129	

4. In a public-key system using RSA, you intercept the ciphertext C = 9 sent to a user whose public key is e = 5, n = 35. What is the plaintext M?

Since n = 35, we have p = 5 and q = 7.

Then,

phin = (p - 1)(q - 1) = 24

As we know ed =1 mod phin, so

d = (e^-1) mod phin = 5 mod 24

According to the RSA decryption algorithm, M = C^d mod n.

Therefore, M = 9^5 mod 35 = 4;

5. Use the Chinese Remainder Theorem (CRT) to solve x, where: x ≡ 1 mod 3, x ≡ 2 mod 5, x ≡ 3 mod 7

a.) M = m1*m2*m3 = 3*5*7 = 105

b.) Mi = M/mi; M1 = 105/3 = 35; M2 = 21; M3 = 105/7 = 15

c.) Mi*yi = 1 mod mi; 35y1 = 1 mod 3, 2y1 = 1 mod 3, -1y1 = 1 mod 3, y1 = -1 mod 3 = 2 mod 3

21y2 = 1 mod 5, 1y2 = 1 mod 5, y2 = 1

15y3 = 1 mod 7, 1y3 = 1 mod 7, y3 = 1

d.) $\sum_{i=1}^{n} ai * yi * Mi \ (mod\ M)$ = 1*2*35 + 2*1*21 + 3*1*15 = 70+42+45 = 157 mod 105 = 52.

6. Consider an Elgamal encryption scheme with a common prime q = 11 and a primitive root α = 2. If B has public key YB = 7 and A choose the random integer k = 3, what is the ciphertext of M = 9?

C1 = a^k = 2^3 = 8 mod 11. C2 = M * (YB ^ k) = 9 * (7^3) = 9*343 = 3087 mod 11 = 7 mod 11.

7. Let F(x) be the true answer on input x, and Geom(α) be the noise sampled from Geometric distribution with parameter α = e ^ (−ϵ/S(F)) . Please prove that the release of F(x) + Geom(α) = F(x) + Geom(e ^ (−ϵ/S(F))) can obtain ϵ-DP guarantee.

Proof

For $D1$, the probability density at any $T \in Ran(A)$ is proportional to (α−1/α+1) * (α^|A(D1)-T|) with parameter $\alpha = S(F)/\epsilon$. Similarly, for $D2$, the probability density at any $T \in Ran(A)$ is proportional to (α−1/α+1) * (α^|A(D2)-T|). Therefore,

$$\frac{\Pr[A(D1) \in T]}{\Pr[A(D2) \in T]} = \frac{\frac{a-1}{a+1} \cdot a^{|A(D1)-T|}}{\frac{a-1}{a+1} \cdot a^{|A(D2)-T|}} = a^{|A(D1)-T|-|A(D2)-T|}$$

$$\leq a^{|A(D1)-A(D2)|} = e^{-\frac{\epsilon}{S(F)}|A(D1)-A(D2)|} = e^{-\epsilon}$$

where the inequality follows form the triangle inequality. By the definition of sensitivity, $S(F) = Max_{D1,D2:|D1-D2|=1}|A(D1) - A(D2)|$. So the ratio is bounded by $e^{-\epsilon}$, yields ϵ-Differential Privacy.

8.

(B0, S1, Z2), k =2

DOB	Sex	Zip	Salary
1/21/76	*	537**	50,000
4/13/86	*	537**	55,000
2/28/76	*	537**	60,000
1/21/76	*	537**	65,000
4/13/86	*	537**	70,000
2/28/76	*	537**	75,000

(B1,S1,Z1), k=2

DOB	Sex	Zip	Salary
76-86	*	5371*	50,000
76-86	*	5371*	55,000
76-86	*	5370*	60,000
76-86	*	5370*	65,000
76-86	*	5370*	70,000
76-86	*	5370*	75,000

(B0,S0,Z1), k=1

DOB	Sex	Zip	Salary
1/21/76	M	5371*	50,000
4/13/86	F	5371*	55,000
2/28/76	M	5370*	60,000
1/21/76	M	5370*	65,000
4/13/86	F	5370*	70,000
2/28/76	F	5370*	75,000

9. Alice and Bob are good friends, they have shared a secret key sk in advance. Now, Alice wants to send 20 messages x1, x2,· · ·, x20 to Bob, because there may be errors occurring in communication channel and also possible injection false data attack from external attackers, Alice hopes to use the bloom filter to enhance the security of these messages in term of source authentication and data integrity. Can you help Alice and Bob to design an efficient bloom filter?

a) n = 160, m = 20

k = ln 2 * (n / m) = ln 2 * (160/20) = ln 2 * 8 = 5.5451 ≈ 6

FP = (0.5)^k = 0.5 ^ 6 = 0.015625

b) n = 160, m = 20

k = ln 2 * (n / m) = ln 2 * (160/20) = ln 2 * 8 = 5.5451 ≈ 6

$FP = (0.5)^{k/2} = 0.5^{6/2} = 0.5^3 = 0.0625$

Comparing D and (D1,D2), D is better because it has a smaller False-Positive

10. Let E() be a BGN homomorphic encryption scheme. Assume Bob has the public/private key pair (pk, sk) of E(), and Alice only has the public key pk of E(). Consider there is no collusion between Alice and Bob. When Alice has three ciphertexts $E(x_1)$, $E(x_2)$ and $E(x_3)$, please design a protocol run between Alice and Bob. With the protocol, Alice can finally obtain the ciphertext $E(x^4_1 + 2x^2_2 + x_3)$ while both Alice and Bob have no idea on the plaintexts x1, x2, and x3.