MID-TERM EXAM

Harshit Rai
2017152

Question 1

Total available characters = 10 (digits)+ 52 (letters) = 62
Total available space in password = 6
Total ways to form a password = 62x62x62x62x62x62 = (62)^6 = 56800235584

A) 1 password check = 10^{-1} seconds
So, 56800235584 passwords check = (10^{-1}/1)*56800235584 = **5680023558.4 seconds or 180 years. On an average = 90 years.**

B) 1 password check = 10^{-6} seconds
So, 56800235584 passwords check = (10^{-6}/1)*56800235584 = **56800.235584 seconds or 15.8 hours. On an average = 8 hours.**

C) Total ways to form a password = 62x62x62x62x62x62 = (62)^7 = 3.5216146x10^{12}

For one-tenth of a second.

1 password check = 10^{-1} seconds
So, 3.5216146x10^{12} passwords check = (10^{-1}/1)*3.5216146x10^{12} = **352161460621 seconds or 111.66 centuries. On an average = 56 centuries.**

For a microsecond.

1 password check = 10^{-6} seconds
So, 3.5216146x10^{12} passwords check = (10^{-6}/1)*3.5216146x10^{12} = **3521614.60621 seconds or 40.75 days. On an average = 21 days.**

For 10^{-1} speed=> len=7/len=6 : 111.66centuries/180years : 11166/180 = **62.03 ~ 62**
For 10^{-6} speed=> len=7/len=6 : 40.75days/15.8hours : 978/15.8 = **61.89 ~ 62**

D)
When length = 6

When 80% of the password is known to us. Therefore 80% of 6 = 4.8 ~ 4 (approx)
Remaining = 6 - 4 = 2
Total ways to form a password = 62x62 = (62)^2 = 3844

For one-tenth of a second.

1 password check = 10^{-1} seconds
So, 3844 passwords check = (10^{-1}/1)*3844 = **384.4 seconds or 6.4 hours. On an average = 3.2 hours.**

For a microsecond.

1 password check = 10^{-6} seconds
So, 3844 passwords check = (10^{-6}/1)*3844 = **0.003844 seconds or 230.64 milliseconds. On an average = 116 milliseconds.**

When length = 7

When 80% of the password is known to us. Therefore 80% of 7 = 5.6 ~ 5 (approx)
Remaining = 7 - 5 = 2
Total ways to form a password = 62x62 = (62)^2 = 3844

When we know 80% of the password, approximately 2 unknown characters remain when length=6 and length=7. Therefore, the total time is taken to guess the passwords when length=6 and length=7 for part A and part B are the same.

But yes obviously, when we know 80% of the password, our guessing time decreases enormously. As we can clearly see from the above-calculated figures. Therefore our task becomes easier.

Question 2

A) The cost of victimization can be estimated by considering various factors like the amount obtained, additional cost, prevention costs, response costs, intangible impacts, and lost outputs. These include almost all the losses incurred by the victim during the incident.

My method to estimate the same would be based on the amount lost and the victim's salary. How much time will it take to earn back the same lost amount.

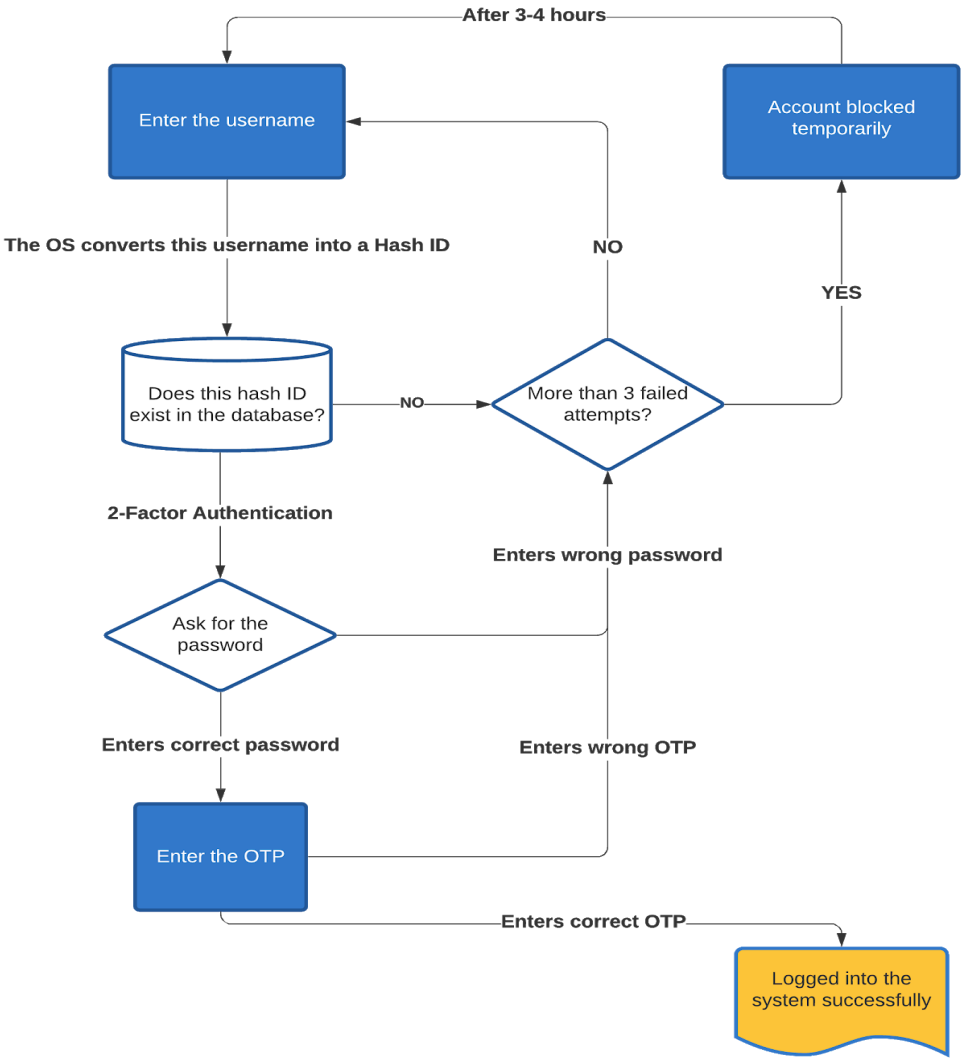
B) Identity Theft happens when the adversary steals personal information like name, password, credit card numbers, address, etc., to impersonate someone else's identity. It affects both the adversary and the person whose identity has been stolen. Identity Fraud happens when the adversary steals a fictitious person's private information to commit a crime like fraudulent withdrawals, fraudulent transactions, etc.

Source: Estimating the cost to Australian businesses of identity crime and misuse

---

Question 3

When the user enters their username, our **OS** converts into a unique hash ID. This will ensure that **no outsider** logs into the system. This hash ID is checked whether it exists in the database or not. If not then an error message should be displayed. Now as it is confirmed that the user exists in the database, their password is required for further authentication. The passwords are converted into cipher texts using appropriate encryption schemes like **RSA or AES** etc. A more secure system has **two-factor authentication.** It requires two pieces of private information to login to the system. The 1st step of authentication can be providing the **password/PIN** and the 2nd step can be providing the **OTP** which is sent on the mobile phone or some **biometric** authentication. It ensures secure authentication. Only after passing both the steps, the user is allowed to access their account. This whole process will ensure **confidentiality** of the system. Also, if the user's verification fails **more than 3 times**, then their account should be temporarily blocked for some time. If dealing with an online system, **HTTPS or SSH** etc internet protocols should be used for a safer and secure experience.



The **integrity** of the cryptosystem can be maintained by using a strong Hash function for encryption. The system should be robust enough to handle fast and crash free data retrieval. It should also maintain a backup on the cloud platform so that the data is not lost in case of a mishappening. In this manner the system's **availability** is also maintained.

Question 4

- A role is defined as a set of users, along with their permissions. Every individual user is given some role/permission according to their designation/post. On the other hand, a group is just a collection of users who share common roles/permissions.
- The role focuses on the role/power of the user. Whereas the group focuses more on the identity and not the role/activity of the user.
- A group always has at least two or more users. Users are put into the same group, which shares common control purposes. Whereas a role can have only one single user also.

SOURCE: (7) Roles Versus Groups

Question 5

A) The basic definition of PII in both GDPR and CCPA is almost identical. Although, the critical difference between PII perception in GDPR and CCAP is that the former considers PII, which is provided by a consumer only and excludes data that was purchased from a third party. On the other hand, GDPR considers all PII the same and does not create any separation regardless of the source of information. So, GDPR is much more secure in terms of confidentiality and keeping the data safe.

B) Customer loyalty history is private information because if the shopping history of an individual is leaked publicly, the adversary will come to know basic details like age, physique, likes & dislikes, wealth, total members in the family, spending habits, the date, time, location of the purchase and other small details. Later these pieces of information can be used to make something useful out of it. It can result in a potential threat to the customer.

C) Giving the phone number to the cashier at Big Bazaar is technically providing our PII to them. Adding our PII in another database can result in a potential threat or data breach in the future. As it is also written in section 72 of the Information Technology Act, 2000, if someone forces an individual or PII is taken without the individual's consent, shall be liable for a punishment of imprisonment or a fine up to 2 lakhs or both.

D) A proper entry-log is maintained for every student who eats in the mess. There is publicly identifiable information too in those records. The entry-log contains our name,

phone number, address, date, time, and menu of the food. As our PII is available to the people of the mess, data breaching can take place. Now there arises a scenario where Food Monk provides this PII of students to Swiggy. Swiggy can use this data to intelligently recommend students' food according to their preferences to make greater profits for the company. Overall, the data has been breached and used in the company's profit without the student's consent.

SOURCE: California Consumer Privacy Act, THE INFORMATION TECHNOLOGY ACT, 2000

Question 6

- A digital virus can spread either via direct contact with an external connecting device like pen-drive, hard-disk, cd's or via wireless contact over the internet or network.
- The reach and speed of contamination of a digital virus is very huge and rapid if it spreads over the wireless internet. Digital packets travel at the speed of light.
- If your device has the virus then all the devices connected to the common router catches the virus in just a couple of seconds because the affected packets will travel from one router to another and transmit the virus in between all the nodes.
- The virus can also be spread when someone sends an email to someone.
- All the systems in our ISP will become the virus' hotspot.
- Just like there are antibodies in humans, computers have antiviruses which help them to fight with the digital virus.
- As a humans' have an immune system, on which the severity of the virus depends. A computer's immune system can be defined in terms of its firewalls, antivirus, anti-spyware, anti-malware capability.
- Both human viruses and the digital virus have the capability of mutation.

Question 7

Apple collects users' personal information by adding some noise to it and removing their respective name tags. They collect a large amount of noisy and biased data. Randomize it and process it to make something meaningful out of it. This whole dataset can not link back to a particular user rather than the entire community. The company uses this information about the community for a limited period to train their algorithms for a better user experience without any threat of data breach or loss of confidentiality. So the user's identity remains anonymous in the whole process.

Data is collected every three months or so and deleted after that. Limited retention of data helps to make their algorithms more robust and powerful. As the data gets deleted after the retention period, one can not breach someone's privacy if their data is not present in the database.

SOURCE: Differential Privacy Overview

Question 8

A) The advantages of logical deletion over cryptographic erasure are:

- If data is deleted by mistake, it can be recovered within the recovery period in logical deletion. Which is not the case in crypto erase.
- If data is deleted through cryptographic erasure, the encryption keys are destroyed. The data can not be recovered because it is computationally infeasible to crack the encryption key (at least 128 bits) with the available computational power and time. So it can not be recovered even if it still has time before its backup expires.
- Cryptographic erasure relies heavily on the manufacturer. The adversary could crack the keys if the cryptographic algorithms have some flaw or loophole due to human error.
- Logical deletion allows us to keep the file's history, which is very useful in file auditing.

B) Various data erasure methods are:

1. **U.S. Air Force System Security Instruction 5020:** 3 overwriting rounds with all 0's, all 1's, any character as pattern and verification at the end.
2. **Gutmann Algorithm:** 35 overwriting rounds with various patterns. A time-consuming method but effective.
3. **Bruce Schneier's Algorithm:** 7 overwriting rounds with all 1's, all 0's, pseudo-random sequence five times as pattern.
4. **German Federal Office for Information Security:** 2-3 overwriting rounds with non-uniform pattern and its complement.
5. **British HMG Infosec Standard 5, Baseline Standard:** 1 overwriting round with random pattern and verification at the end.
6. **NIST SP-800-88 Rev. 1:** 1 overwriting round with all 0's as pattern and outlines solutions based on the media type.

SOURCE: Data erasure, What is Cryptographic Erasure (Crypto Erase)?

Question 9

Original data contains PII of millions of individuals, causing a catastrophe if released publicly. So it becomes very crucial to ensure the confidentiality of the data. Various techniques, like differential privacy, k-anonymization (generalization and suppression), de-identification, data-masking, and pseudonymization etc can be practiced to establish secure data collaboration. K-anonymization will ensure that a particular individual's information can not be used to identify the individual uniquely. De-identification will prevent an individual's PII from being revealed explicitly.

Terms and conditions for fair use of data:

- Data must only be available to download after verifying its identity and purpose for using the data.

- Auditing the person from time to time to ensure the safe use of data.
- There must be a limit to the collection of personal information.
- Data must be masked or encrypted before making it public.
- We must not be held accountable if, somehow, data breaching takes place.

This is how we can preserve the confidentiality of our data and also use it for good deeds.

Question 10

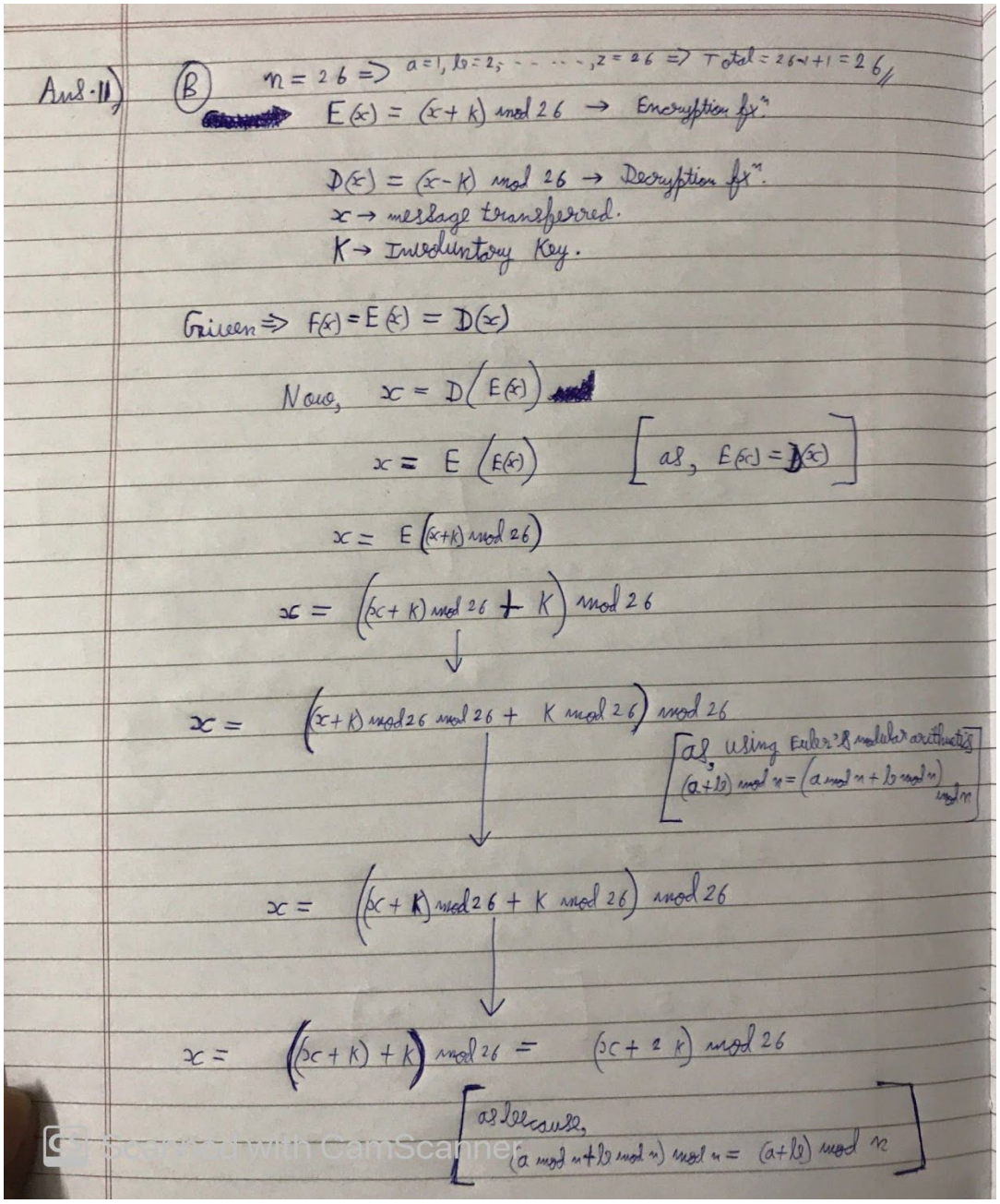
The timing attack exploits the vulnerabilities of a security system by studying the response time for various inputs. The timing attack effectively discovers the relationship between a client and a server on the TOR network. Currently, no particular policies or strategies are followed in TOR networks to counter a timing attack. Although introducing an **'entry guard'** reduces the chances of timing attack.

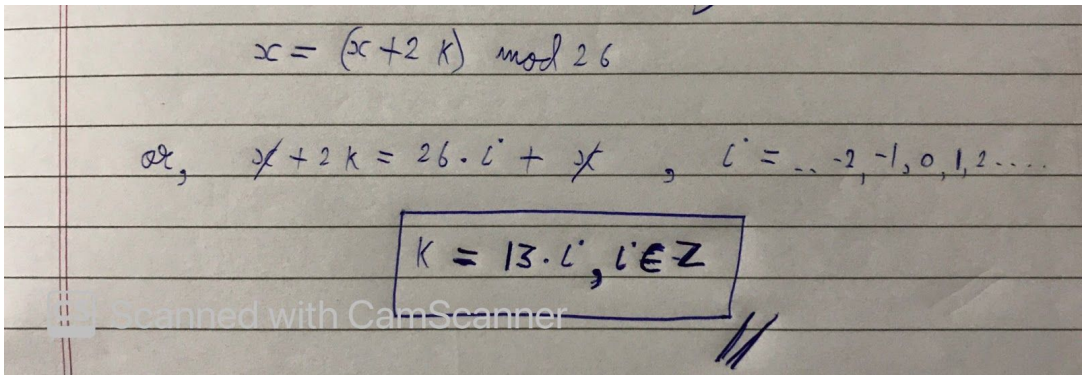
A timing attack is majorly practiced on low computing devices, such as smart cards. They can crack the private keys from low computing devices like OpenSSL based web servers and smart cards. However, other web services are relatively safe as they mask differences in timing.

SOURCE: What is timing attack? - Definition from WhatIs.com , Remote Timing Attacks are Practical

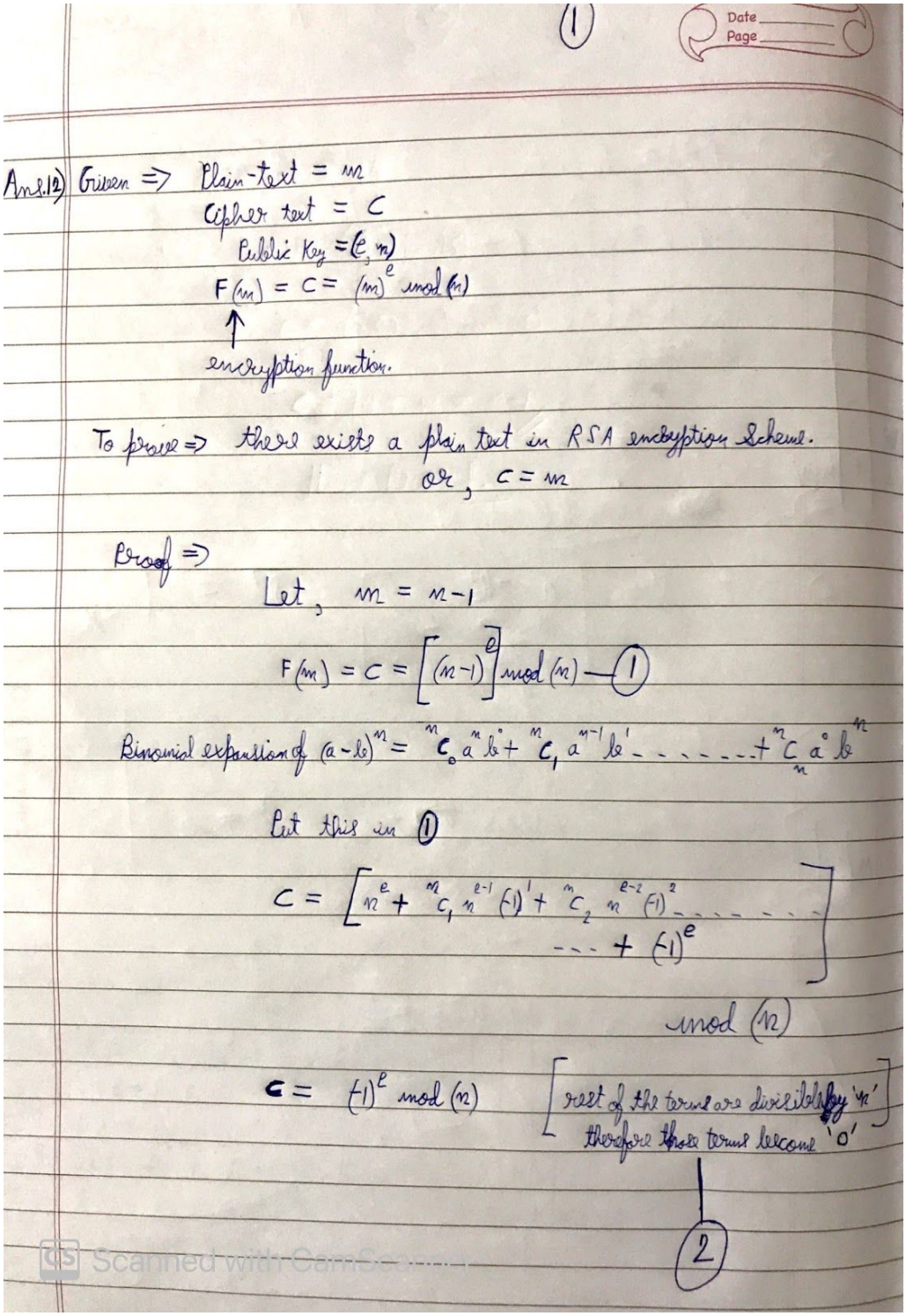
---

Question 11.B





Question 12



(1)

Date
Page

Ans.12) Given ⇒ Plain-text = m
Cipher text = C
Public Key = (e, n)
$F(m) = C = (m)^e \bmod (n)$
↑
encryption function.

To prove ⇒ there exists a plain text in RSA encryption Scheme.
or, $C = m$

Proof ⇒

Let, $m = n - 1$

$$F(m) = C = \left[(n-1)^e\right] \bmod (n) \quad \text{— (1)}$$

Binomial expansion of $(a-b)^n = {}^nC_0\, a^n b^0 + {}^nC_1\, a^{n-1} b^1 - \dots + {}^nC_n\, a^0 b^n$

Put this in (1)

$$C = \left[n^e + {}^nC_1\, n^{e-1}(-1)^1 + {}^nC_2\, n^{e-2}(-1)^2 - \dots - \dots + (-1)^e\right] \bmod (n)$$

$C = (-1)^e \bmod (n)$ [rest of the terms are divisible by 'n' therefore those terms become '0']

(2)

Scanned with CamScanner

Now, we know that $GCD(e, \phi(n)) = 1$

Also, the range of totient function $\phi(n)$ is always even for all $n > 2$.

Therefore to have $GCD(e, \phi(n)) = 1$, 'e' must be 'odd'
(Even)

Put this in (2)

$C = (-1) \bmod (n)$

$C = n - 1$ [both $(n-1)$ and (-1) belong to the same congruent class.]

or, $C = m = n - 1$

Hence Proved that there exists a Plain text in the RSA encryption scheme.

Example.

Let, $p = 3$ and $q = 5$

$n = (p \times q) = 15$

$\phi(n) = (2) \times (4) = 8$

Let, $e = 5$ $[GCD(e, \phi(n)) = 1]$

CS Scanned with CamScanner

$m = n - 1 = 14$

Now, $C = (14)^5 \bmod (15)$

$C = (537824) \bmod (15)$

$C = 14 = m = n - 1$

Hence Proved.

CS Scanned with CamScanner

Question 13

A) l-diversity is an extension of k-anonymization. l-diversity is more secure as compared to k-anonymization due to more diversity of sensitive information. We can achieve l-diversity using k-anonymization if we use k-anonymity in a specific manner. By having at least 'k-1' unique quasi-identifier and having at least 'l' different values of sensitive information in each group of 'k-1' unique quasi-identifier. Also, having l=k.

In the example given below, we have achieved 3-diversity using 3-anonymization. We have three groups of unique quasi-identifier. Each of these groups has three unique (diversity) values of sensitive information.

Customer ID	Name	Place	City	Country	No items purchased
C000**	*	New ****	New ****	U**	1
C000**	*	New ****	New ****	U**	2
C000**	*	New ****	New ****	U**	3
C000**	*	*	*	*	1
C000**	*	*	*	*	2
C000**	*	*	*	*	3
C000**	*	*	*	*n**a	1
C000**	*	*	*	*n**a	2
C000**	*	*	*	*n**a	3

B) We use generalization (global recoding) and local suppression to hide some of the values or characters from our data. This does not alter the original data points but only hides/masks them to maintain confidentiality. So the overall distribution (mean, variance) of the dataset is preserved. In a table, the primary key (attributes which can identify a value uniquely) is always made hidden. The sensitive information is shown as it is. The remaining columns are called the quasi-identifiers. These columns are used for k-anonymization to ensure that the data does not lose the distribution.

SOURCE: Anonymization Methods — SDC Practice Guide documentation

Question 14

Keyloggers are software installed on a system to monitor the productivity of their employees. This software secretly monitors keyboard strokes through pattern recognition when someone is typing something. However, this can also be used to spy on someone. One can also exploit someone's PII just by using these programs.

Shoulder surfing is more renounced to happen in a crowded place due to the adversary's very low distance range. Modern technologies like secret cameras and hidden microphones make it easier for the adversary to exploit the vulnerabilities. It is done either by eavesdropping on someone's conversation or by literally peeking from the victim's shoulder, his keystrokes, and typing patterns. Individual's private information is at risk.
Therefore shoulder surfing and keylogger act as catalysts to social engineering attacks.

SOURCE: What is shoulder surfing? - Definition from WhatIs.com

---

Question 15

Customer ID	Name	Place	City	Country	No items purchased	Price
C000**	*	*	*	*	2	6000
C000**	*	New ****	New ****	U**	2	3000
C000**	*	New ****	New ****	U**	3	5000
C000**	*	*	*	*n**a	2	5000
C000**	*	*	*	*	1	10000
C000**	*	New ****	New ****	U**	3	5000
C000**	*	*	*	*	2	7000
C000**	*	*	*	*	1	7000
C000**	*	*	*	*n**a	1	8000
C000**	*	*	*	*n**a	1	7000
C000**	*	*	*	*n**a	1	7000
C000**	*	*	*	*n**a	2	7000

- Customer ID and Name are the primary keys, so they have to be made hidden.
- Place, city, country are quasi-identifier, so these have to be anonymized.
- Items purchased and Price is the sensitive information (reason mentioned in Q5), so written as it is.