

Data Privacy Practicals

Madhav Agarwal -- 20221426

Question 1 : Write a program to perform encryption and decryption using Caesar cipher (substitutional cipher)

Solution :

The program shifts each letter in the input text by a given number of positions in the alphabet. For encryption, the letters are shifted forward, while for decryption, they are shifted backward. Non-alphabetic characters are not changed.

Time Complexity:

- $O(N)$, where N is the length of the input text.
- Each character is processed exactly once, and the operations (checking character type, shifting, etc.) are performed in constant time.

Space Complexity:

- $O(N)$, to store the result string, which is proportional to the length of the input text.

Output :

```
● madhav@machine:~/work/data_privacy_practicals$ /bin/python3 /home/madhav/work/data_privacy_pract
Caesar Cipher: Encryption and Decryption
Choose an option (encrypt/decrypt): encrypt
Enter the text: madhav
Enter the shift value: 3
Encrypted text: pdgkdy
○ madhav@machine:~/work/data_privacy_practicals$
○ madhav@machine:~/work/data_privacy_practicals$
● madhav@machine:~/work/data_privacy_practicals$ /bin/python3 /home/madhav/work/data_privacy_pract
Caesar Cipher: Encryption and Decryption
Choose an option (encrypt/decrypt): decrypt
Enter the text: pdgkdy
Enter the shift value: 3
Decrypted text: madhav
● madhav@machine:~/work/data_privacy_practicals$
```

Question 2 : Write a program to perform encryption and decryption using Rail Fence Cipher (transpositional cipher)

Solution :

For encryption, the text is written in a zigzag pattern across multiple rows (based on the key), and then read row by row to create the cipher. For decryption, the pattern is reconstructed to extract the original message

Time Complexity:

- **Encryption:** $O(N)$, where N is the length of the input text.
 - The text is traversed twice: first to place characters on the rails, and second to read the rails.
- **Decryption:** $O(N)$, where N is the length of the input cipher text.
 - The cipher is traversed three times: first to mark the rail positions, second to place characters, and third to read them.

Space Complexity:

- $O(N)$ for both encryption and decryption, as an auxiliary 2D list (the rail matrix) is used to hold the text, which is proportional to the input length.

Output :

```
madhav@machine:~/work/data_privacy_practicals$ /bin/python3 /home/madhav/work/data_privacy_practicals/rail_fence_cipher.py
Rail Fence Cipher: Encryption and Decryption
Choose an option (encrypt/decrypt): encrypt
Enter the text: madhav
Enter the key value: 3
Encrypted text: maahvd
madhav@machine:~/work/data_privacy_practicals$ /bin/python3 /home/madhav/work/data_privacy_practicals/rail_fence_cipher.py
Rail Fence Cipher: Encryption and Decryption
Choose an option (encrypt/decrypt): decrypt
Enter the text: maahvd
Enter the key value: 3
Decrypted text: madhav
madhav@machine:~/work/data_privacy_practicals$
```

Question 3 : Write a Python program that defines a function and takes a password string as input and returns its SHA-256 hashed representation as a hexadecimal string.

Solution :

The program takes a password from the user and uses the hashlib library to create a SHA-256 hash of the password. The hash is then converted to a hexadecimal representation, which is printed to the user for secure storage.

Time Complexity:

- Hashing with SHA-256: $O(1)$ with respect to the password length.
 - While SHA-256 performs a fixed number of operations regardless of the input size, the actual processing time may vary with the length of the input.

Space Complexity:

- $O(1)$ (constant space).
 - The space requirement is constant since the hash value produced by SHA-256 is always of a fixed length (256 bits).

Output :

```
madhav@machine:~/work/data_privacy_practicals$  
madhav@machine:~/work/data_privacy_practicals$  
madhav@machine:~/work/data_privacy_practicals$  
madhav@machine:~/work/data_privacy_practicals$ /bin/python3 /home/madhav/work/data_privacy_practicals/question3.py  
Enter the password: madhav123  
SHA-256 Hashed Password: 472bc037804023a43b186705e739d5e14ad0666ba34aa0d74a03d076c8653141d  
madhav@machine:~/work/data_privacy_practicals$  
madhav@machine:~/work/data_privacy_practicals$
```

Question 4 : Write a Python program that reads a file containing a list of usernames and passwords, one pair per line (separated by a comma). It checks each password to see if it has been leaked in a data breach. You can use the "Have I Been Pwned" API (<https://haveibeenpwned.com/API/v3>) to check if a password has been leaked.

Solution :

The program reads the passwords from a file, hashes them using SHA-1, and then checks if they have been breached by querying the "Have I Been Pwned" API. If a password is found to be compromised, the program notifies the user.

Time Complexity:

1. Hashing Password (SHA-1):

- $O(1)$, since SHA-1 hashing performs a fixed number of operations irrespective of the input length.

2. API Call to "Have I Been Pwned":

- $O(1)$ for making the API request, as it depends on network latency and the fixed response size for the prefix.
- Processing the response involves iterating over the returned hash suffixes, which takes $O(M)$, where M is the number of hash suffixes returned.

Space Complexity:

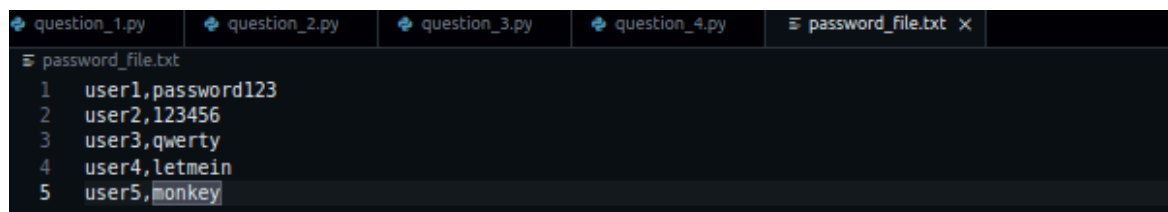
1. Hash Storage and Processing:

- $O(1)$, since only a constant amount of space is needed for the hash and suffix.

2. API Response Processing:

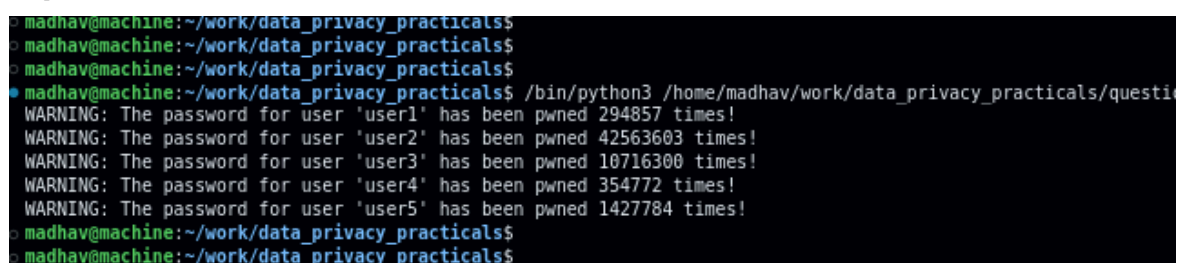
- $O(M)$, where M is the number of hash suffixes returned by the API (generally a small fixed size).

Dummy Test File :



```
question_1.py question_2.py question_3.py question_4.py password_file.txt x
password_file.txt
1 user1,password123
2 user2,123456
3 user3,qwerty
4 user4,letmein
5 user5,monkey
```

Output :



```
madhav@machine:~/work/data_privacy_practicals$
madhav@machine:~/work/data_privacy_practicals$
madhav@machine:~/work/data_privacy_practicals$
madhav@machine:~/work/data_privacy_practicals$ /bin/python3 /home/madhav/work/data_privacy_practicals/questi
WARNING: The password for user 'user1' has been pwned 294857 times!
WARNING: The password for user 'user2' has been pwned 42563603 times!
WARNING: The password for user 'user3' has been pwned 10716300 times!
WARNING: The password for user 'user4' has been pwned 354772 times!
WARNING: The password for user 'user5' has been pwned 1427784 times!
madhav@machine:~/work/data_privacy_practicals$
madhav@machine:~/work/data_privacy_practicals$
```

Question 5 : Write a Python program that generates a password using a random combination of words from a dictionary file..

Solution :

This program generates a password using a random combination of words from a dictionary file. It reads a list of words from the provided file, selects a specified number of words at random, and combines them to create a password.

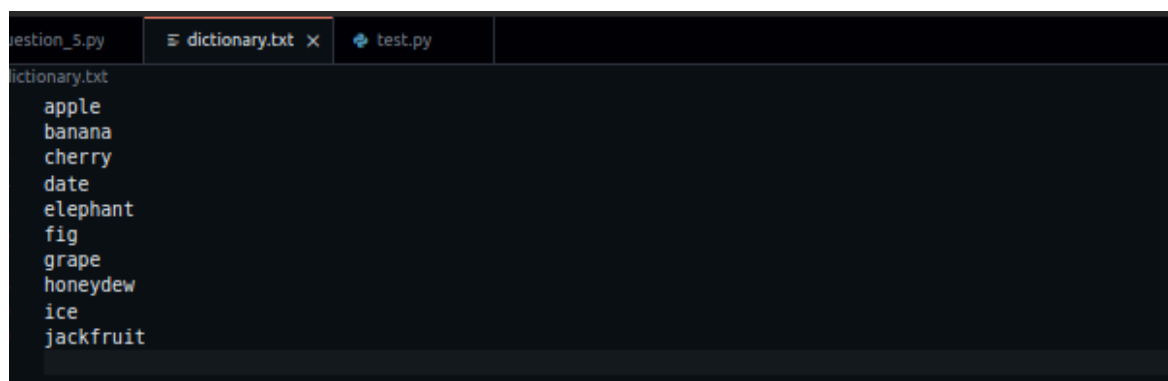
Time Complexity:

- $O(N)$, where N is the number of lines in the dictionary file.

Space Complexity:

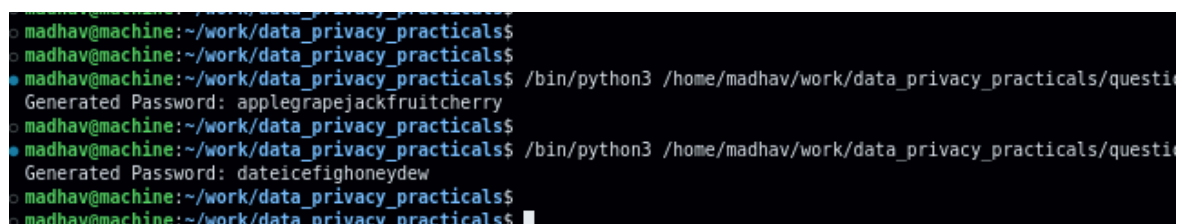
- $O(N)$, to store the words from the dictionary.

Dummy Test File :



```
question_5.py  dictionary.txt x  test.py
dictionary.txt
apple
banana
cherry
date
elephant
fig
grape
honeydew
ice
jackfruit
```

Output :



```
madhav@machine:~/work/data_privacy_practicals$
madhav@machine:~/work/data_privacy_practicals$
madhav@machine:~/work/data_privacy_practicals$ /bin/python3 /home/madhav/work/data_privacy_practicals/question_5.py
Generated Password: applegrapejackfruitcherry
madhav@machine:~/work/data_privacy_practicals$
madhav@machine:~/work/data_privacy_practicals$ /bin/python3 /home/madhav/work/data_privacy_practicals/question_5.py
Generated Password: dateicefighoneydew
madhav@machine:~/work/data_privacy_practicals$
madhav@machine:~/work/data_privacy_practicals$
```

Question 6 : Write a Python program that simulates a brute-force attack on a password by trying out all possible character combinations.

Solution :

This program simulates a brute-force attack on a given password. It generates all possible combinations of characters up to a specified length and attempts to match them with the target password.

Time Complexity:

- $O(|C|^L)$, where $|C|$ is the number of characters in the character set (62 in this case: lowercase, uppercase, digits) and L is the maximum length of the password (max_length). The time complexity grows exponentially with respect to the password length.

Space Complexity:

- $O(L)$, which is the space required to store the current combination being attempted.

Output :

```
madhav@machine:~/work/data_privacy_practicals$  
madhav@machine:~/work/data_privacy_practicals$  
madhav@machine:~/work/data_privacy_practicals$ /bin/python3 /home/madhav/work/data_privacy_practicals/question6.py  
Enter the password to simulate brute-force attack: abcde  
Enter the maximum length for brute-force attempts: 50  
Password found: abcde  
madhav@machine:~/work/data_privacy_practicals$ /bin/python3 /home/madhav/work/data_privacy_practicals/question6.py  
Enter the password to simulate brute-force attack: madhav  
Enter the maximum length for brute-force attempts: 50  
  
Password not found within the given length.  
madhav@machine:~/work/data_privacy_practicals$ /bin/python3 /home/madhav/work/data_privacy_practicals/question6.py
```

Question 7 : Demonstrate the usage/sending of a digitally signed document.

Solution :

This program demonstrates the process of digitally signing a document using RSA encryption. It generates a pair of RSA keys (public and private), signs a document using the private key, and then verifies the signature using the public key. This ensures that the document's integrity is maintained and confirms the authenticity of the sender.

Time Complexity:

- Key Generation: $O(k^3)$
- Hashing (SHA-256): $O(N)$
- Signing/Verification: $O(k^3)$

Space Complexity:

- $O(N + k)$, where N is the document length and k is the key size.

Output :

```
madhav@machine:~/work/data_privacy_practicals$  
madhav@machine:~/work/data_privacy_practicals$  
madhav@machine:~/work/data_privacy_practicals$ /bin/python3 /home/madhav/work/data_privacy_practicals/question7.py  
Document signed successfully.  
Sending document and signature...  
The document is verified successfully. The signature is valid.  
madhav@machine:~/work/data_privacy_practicals$  
madhav@machine:~/work/data_privacy_practicals$
```

Question 8: Students needs to conduct a data privacy audit of an organization to identify potential vulnerabilities and risks in their data privacy practices.

Output :

```
● madhav@machine:~/work/Assign_Privacy_MadhavAgarwal_20221426$ /bin/python3 /home/madhav/work/Assign_Privacy_MadhavAgarwal_20221426/audit.py
Starting Data Privacy Audit...

Category: DATA_COLLECTION
- Are users informed about data collection? (Yes/No): no
- Is data collection limited to what's necessary? (Yes/No): yes
- Are consent mechanisms in place? (Yes/No): yes

Category: DATA_STORAGE
- Is data encrypted at rest? (Yes/No): no
- Is sensitive data stored securely? (Yes/No): yes
- Are backup policies in place? (Yes/No): no

Category: DATA_ACCESS
- Is access to data restricted based on roles? (Yes/No): yes
- Are access logs maintained and monitored? (Yes/No): yes
- Are strong authentication mechanisms used? (Yes/No): no

Category: COMPLIANCE
- Is the organization GDPR compliant? (Yes/No): yes
- Are data retention policies clearly defined? (Yes/No): yes
- Is there a process for handling data subject requests? (Yes/No): no

Analyzing Responses...

Summary of Findings:
- DATA_COLLECTION: 1 issues identified.
  * Are users informed about data collection?
- DATA_STORAGE: 2 issues identified.
  * Is data encrypted at rest?
  * Are backup policies in place?
- DATA_ACCESS: 1 issues identified.
  * Are strong authentication mechanisms used?
- COMPLIANCE: 1 issues identified.
  * Is there a process for handling data subject requests?

Audit report generated: data_privacy_audit_report_2024-11-26_18-04-42.json
○ madhav@machine:~/work/Assign_Privacy_MadhavAgarwal_20221426$
○ madhav@machine:~/work/Assign_Privacy_MadhavAgarwal_20221426$
```

Generated file has been attached in the zip folder

Question 9 : Students needs to explore the requirements of the Data Protection Regulations and develop a plan for ensuring compliance with the regulation

Output :

```
madhav@machine:~/work/Assign_Privacy_MadhavAgarwal_20221426$  
madhav@machine:~/work/Assign_Privacy_MadhavAgarwal_20221426$ /bin/python3 /home/madhav/work/Assign_Privacy_Madhav  
Available Regulations:  
- GDPR  
- HIPAA  
  
Enter the regulation to comply with (e.g., GDPR, HIPAA): HIPAA  
  
Categories for HIPAA:  
- privacy_rule  
- security_rule  
- breach_notification_rule  
  
Enter the categories to include in the compliance plan (comma-separated): privacy_rule,breach_notification_rule  
  
Developing Compliance Plan for HIPAA...  
  
Compliance Plan:  
- Privacy_rule:  
  * Ensure protected health information (PHI) is safeguarded.  
  * Provide patients with rights over their PHI.  
  * Limit disclosures of PHI to the minimum necessary.  
- Breach notification rule:  
  * Notify affected individuals within 60 days of discovering a breach.  
  * Report breaches affecting more than 500 individuals to the Department of Health and Human Services.  
  
Compliance plan report generated: compliance_plan_HIPAA_2024-11-26_18-07-23.json  
madhav@machine:~/work/Assign_Privacy_MadhavAgarwal_20221426$
```

Generated file has been attached in the zip folder

Question 10 : Students needs to explore ethical considerations in data privacy, such as the balance between privacy and security, the impact of data collection and analysis on marginalized communities, and the role of data ethics in technology development.

Output :

```
madhav@machine:~/work/Assign_Privacy_MadhavAgarwal_20221426$ /bin/python3 /home/madhav/work/Assign_Privacy_MadhavAgarwal_20221426/question_10.py
```

Exploring Ethical Considerations in Data Privacy...

Topic: Privacy vs Security

- How should organizations balance individual privacy with the need for security?

Your response: Organizations should implement transparent policies, minimize data collection, use encryption, and ensure consent to balance privacy with security needs effectively.

- What measures can be implemented to protect privacy without compromising security?

Your response: Use encryption, minimize data collection, apply access controls, anonymize data, ensure transparency, and implement consent-based data handling to protect privacy securely.

- What are examples of when privacy and security have conflicted?

Your response: Government surveillance programs (e.g., PRISM), contact tracing during pandemics, and corporate data collection often create privacy versus security conflicts.

Topic: Impact on Marginalized Communities

- How can data collection practices disproportionately affect marginalized communities?

Your response: Data collection can reinforce biases, enable targeted surveillance, and limit opportunities, disproportionately impacting marginalized communities' privacy, rights, and freedoms.

- What steps can be taken to ensure inclusivity and fairness in data analysis?

Your response: Use diverse datasets, eliminate biases, involve marginalized groups, conduct fairness audits, and implement ethical guidelines to ensure inclusive data analysis.

- Are there cases where biased data has caused harm? If so, how could it have been prevented?

Your response: Yes, biased data in predictive policing, hiring algorithms, and healthcare has caused discrimination. Prevention involves diverse datasets, fairness testing, and bias mitigation strategies.

Topic: Role of Data Ethics in Technology Development

- What ethical principles should guide the development of data-driven technologies?

Your response: Key ethical principles include transparency, accountability, fairness, privacy, informed consent, minimizing harm, inclusivity, and promoting social good in data-driven technologies.

- How can organizations ensure accountability in the use of AI and machine learning?

Your response: Organizations can ensure accountability by adopting transparent policies, conducting regular audits, using explainable AI, documenting processes, and involving diverse stakeholders in oversight.

- What are the consequences of neglecting data ethics in technology development?

Your response: Neglecting data ethics can lead to discrimination, privacy violations, biased outcomes, public distrust, legal repercussions, and harmful societal impacts, undermining technology's intended benefits.

Summary of Ethical Considerations:

Topic: Privacy vs Security

- How should organizations balance individual privacy with the need for security?
 - * Organizations should implement transparent policies, minimize data collection, use encryption, and ensure consent to balance privacy with security needs effectively.
- What measures can be implemented to protect privacy without compromising security?
 - * Use encryption, minimize data collection, apply access controls, anonymize data, ensure transparency, and implement consent-based data handling to protect privacy securely.
- What are examples of when privacy and security have conflicted?
 - * Government surveillance programs (e.g., PRISM), contact tracing during pandemics, and corporate data collection often create privacy versus security conflicts.

Topic: Impact on Marginalized Communities

- How can data collection practices disproportionately affect marginalized communities?
 - * Data collection can reinforce biases, enable targeted surveillance, and limit opportunities, disproportionately impacting marginalized communities' privacy, rights, and freedoms.
- What steps can be taken to ensure inclusivity and fairness in data analysis?
 - * Use diverse datasets, eliminate biases, involve marginalized groups, conduct fairness audits, and implement ethical guidelines to ensure inclusive data analysis.
- Are there cases where biased data has caused harm? If so, how could it have been prevented?
 - * Yes, biased data in predictive policing, hiring algorithms, and healthcare has caused discrimination. Prevention involves diverse datasets, fairness testing, and bias mitigation strategies.

Topic: Role of Data Ethics in Technology Development

- What ethical principles should guide the development of data-driven technologies?
 - * Key ethical principles include transparency, accountability, fairness, privacy, informed consent, minimizing harm, inclusivity, and promoting social good in data-driven technologies.
- How can organizations ensure accountability in the use of AI and machine learning?
 - * Organizations can ensure accountability by adopting transparent policies, conducting regular audits, using explainable AI, documenting processes, and involving diverse stakeholders in oversight.
- What are the consequences of neglecting data ethics in technology development?
 - * Neglecting data ethics can lead to discrimination, privacy violations, biased outcomes, public distrust, legal repercussions, and harmful societal impacts, undermining technology's intended benefits.

Summary report generated: ethical_considerations_summary_2024-11-26_18-11-58.json
madhav@machine:~/work/Assign_Privacy_MadhavAgarwal_20221426\$

Generated file has been attached in the zip folder