PS-7: Extraction and Decryption of Passwords from Password Protected Documents, Disk and Compressed Files for Retrieval of Forensic Evidence

1. **Background**

   Passwords remain the most ubiquitous form of user authentication across digital systems. Despite the emergence of bio-metric and multi-factor authentication, passwords are still the first line of defense for the majority of users. Traditional password analysis and cracking methods, such as brute force, dictionary attacks, and rule-based approaches are time-consuming, computationally expensive, and often ineffective against complex or long passwords. AI/ML- driven solution will go beyond conventional techniques, proving a new paradigm for password analysis and cracking.

2. **Objective**

   The objective of this problem statement is to use artificial intelligence for the recovery of passwords from password-protected files. This involves developing a learning-based approach capable of identifying and replicating common patterns in user-created passwords. By training on real-world password datasets, the system aims to generate more accurate and efficient password guesses compared to traditional brute-force or dictionary methods. This also seeks to create an automated framework that can apply these AI-generated guesses to attempt decryption of protected file formats such as ZIP, PDF, or Office documents. Through systematic testing, the approach will be evaluated in terms of accuracy, speed, and computational efficiency.

3. Stages of the Problem Statement are as follows: -

Stage	Problem Statement
Stage-1	Password Extraction and Decryption of following documents: - i)Office 2013 & above ii) PDF version 1.7 & above Password length 12 Character and above (Character set consist of English alphabets, numbers & special characters)
Stage-2	Password Extraction and Decryption of following: - i) Archive ii) Documents iii) Disk encryption iv) Pretty Good Privacy (PGP) v) Operating System vi) Raw Hash Password length 14 Character and above (Character set consist of English alphabets, numbers & special characters)

Stage	Problem Statement
Stage-3	Password Extraction and Decryption of following: - i) Archive ii)Documents iii)Disk encryption iv)Pretty Good Privacy (PGP) v)Operating System vi)Raw Hash Multilingual Password length 08 Characters and above (Password may consist of English language alphabets, one UN language / Indian language alphabets, numbers & special characters) Improvements of model/solutions developed at Stage-2

RULES/ EVALUATION CRITERIA FOR SHARING IN OPEN DOMAIN

4. Stage-1

i. Training Data: There are no specific training data sets in this Problem Statement being provided by Challenge Team.

ii. Mock Test Facility for Self Assessment :The mock/testing dataset will be shared on 15 Sep 2025 and the same will not be part of evaluation. Leader board for the same will be displayed.

iii. Shortlisting of 15-20 Startups at the end of Stage I :Multiple level of evaluation will be adopted at stage- 1. For level-1 Shortlisting data set vectors will be provided on 26th Oct 2025. Level-2 Shortlisting data set vectors will be provided on 01st Nov 2025, and Level-3 Shortlisting data set vectors will be provided on 06th Nov 2025. The recovered password should be shared by the participants immediately after its recovery in the specific portal designed for uploading the recovered password for level 1, 2 & 3 at stage-1. Submission time as per timestamp will be noted as recovery time.

iv. The participants will be evaluated based on the following two criteria at level-1, 2, & 3. The weightage for the same mentioned is mentioned below.

   a) Number of Password recovered with a weightage of 75%.
   b) Time taken for recovery of Password with a weightage of 25%

v. The shortlisted participants will be published along with the cutoff score as per the evaluation criteria. Participants individual scores will be shared over the email.

vi. The following weightage will be assigned to each levels:

   a) level – 1: 20%

b) level – 2: 30%
c) level – 3: 50%

vii. The 15-20 selected participant, after level 3, will have to do reference implementation of their solution to particular H/W for further evaluation at IIT Delhi on the dates communicated to them. The number may vary based on the overall performance at the discretion of the Jury for this Problem Statement. The selected participants have to provide detailed documents which include the following:

a) Methodology used
b) Dockerized Model
c) Model architecture
d) Justification for model selected
e) Novelty of the approach
f) Scalability.
g) Any other relevant detail

viii. **Selection of Participants for Stage II**

Evaluation criteria for 15-20 shortlisted participants during thier physical evaluation at IIT Delhi (with a holdout dataset) will be the following

a) Number of Password recovered with a weightage of 60%.
b) Time taken for recovery of Password with a weightage of 20%
c) Methodology/Novelty of the approach with a weightage of 20%

a. Shortlisted participants will be asked to demonstrate their solution at IIT Delhi on completion of stage-1 deadline.
b. Participants will be allotted slots in which they need to run their solution on holdout data provided by the organizers on given resources with following specifications: -
    i. OS – Ubuntu 24.04 LTS
    ii. CPU – 48+ core
    iii. RAM – 256+ GB
    iv. GPU - A-100, 40/80 GB*

1. Solution Demo Duration: Upto 24 Hours for each selected participant
2. After evaluation at Stage 1, a maximum of six participants will be selected for stage-2 of competition. Selected participants must submit containerized solution along with all dependencies.

*Thismay change based on feedback.

5. Stage-2

i. The mock test vector will be shared with participants on 16th Mar 2026.

ii. The test vectors will be shared 01st Apr 2026

iii. Detailed documents to be prepared by the participants, which should include the following:

   a) Methodology used
   b) Dockerized Model
   c) Justification for model selected
   d) Novelty of the approach
   e) Scalability
   f) Any other relevant detail

iv. Their containerized solution along with all dependencies must be submitted.

v. Recovered password to be immediately shared.

vi. Further criteria if any, will be intimated later.

vii. **Evaluation criteria**

   a) Efficiency: Number of passwords cracked with a weightage of 60%.
   b) Speed: Time taken to recover passwords with a weightage of 20%
   c) Methodology/Novelty of the approach with a weightage of 20%

6. Stage -3

i. The mock test vector will be shared with participants on 01th Sept 2026, including contextual information in respect of few.

ii. Multilingual test vector (70%) and test vector for the improved solution of stage-2 (30%) will be shared for evaluation on 15th Sep 2026 including contextual information in respect of few.

iii. Detailed documents to be prepared by the participants, which should include the following:

   a) Methodology used
   b) Dockerized Model
   c) Justification for model selected
   d) Novelty of the approach
   e) Scalability
   f) Any other relevant detail

iv. Access to the model developed to be shared with evaluators as and when it is required.

v. Recovered password to be immediately shared.

vi. The solution should be compatible/implementable as per custom requirement environment.

vii. **Evaluation Criteria**

a) Efficiency: Number of passwords cracked with a weightage of 60%.

b) Speed: Time taken to recover passwords with a weightage of 20%

c) Methodology/Novelty of the approach with a weightage of 20%

7. Sessions with Mentors\Experts

a. For Stage-1, the organisers plan to meet participants via online meet or email to resolve their doubts, if any. This provision will be made active from 15th Aug 2025 and details regarding interaction will be shared on this website. Kindly keep viewing this website regularly for updates on this.

b. There will be sessions with Mentors\Experts in Stage-2 and Stage-3 for the willing selected participants to help them in achieving the best solutions.

Annexure B

Guide to Output Submission Format and File/Folder Naming Structure for PS-07

1. Introduction

a. Purpose: This document provides a standard format and naming structure for PS-07 developers to ensure consistency and ease of access and review of results for the organiser.

b. Scope: This guide applies to the output report, evidences and documentation formats.

2. Mock Test Vector Submission

a. At each instance of cracking of password,team must immediately report <Problem_Statement_Number>_<Application ID>_<File Name>_<Password> as per arrangement provided on the portal.

3. Shortlisting Data Set Vector Submission

a. At each instance of cracking of password,team must immediately report <Problem_Statement_Number>_<Application_ID>_<Test_Vector_Level_Number><File Name>_<Password> as per arrangement provided on the portal.