NCIIPC AIGC - Webinar (Second) on Problem Statement No. 7 at 11 am on 3rd September 2025 (Wednesday)-20250903_104409-Meeting Recording

September 3, 2025, 5:14AM

44m 23s

Raj Mehta started transcription

Raj Mehta 15:56

Good morning everyone. I'm Raj Mehta from FITT, IIT Delhi. Welcome to the 2nd and last webinar of on the problem statement #7, password extraction and decryption. I would like. I would also like to welcome the officer in charge, Dr.
Asti Lakshmi and Kandan Ajit the webinar. Sir, I would like to ask from you ma'am, whether we should start it or we should wait for some more time for members to join.

PS07 16:33

Sir, I think another 5 more minutes because I saw in last five minutes some more people have joined. So I think we can wait for another five more minutes. Is that OK with all?

Raj Mehta 16:46

You should be OK with that. OK, we should start it at.
Yeah, miss in 5 minutes. Thank you.
And now I think we can start the webinar. I would like to request Dr. Sri Lakshmi to kindly brief about the problem statement. There's document details in the documents mentioned and uploaded on the the on the portal.
So that the attendees in this webinar are aware of what we're going to talk about.
After that we can take doubts and on the portal on the on this in this webinar one by over to you Doctor Rastalakshmi.

Meenakshi 21:38

Mhm.

Raj Mehta 21:44
I request everyone to kindly mute themselves.

Meenakshi 21:45
It.

PS07 21:50
You're very good morning to all.

Raj Mehta 21:57
Voice is echoing.
Everyone kindly mute yourself and let the officer in charge speak.

PS07 22:24
Yeah, very good morning to all.
So this is problem statement seven. It's all about password recovery using AI and ML.
Since you all know despite various methods of authentication, password still remains the dominant method of authentication, especially if you go for securing files.
And you also know that conventional methods are just infeasible and it is increasingly inefficient and also it's resource intensive. So as the especially because of this policies of the passwords which has become more complex.
Which is not allowing you to keep the easiest password. So this is so we as this AI is changing the game, we want to exploit this AI approach for real password analysis.
So what? So we are posed to this problem.
So in the stage one you're looking into looking especially into Office 2013 and.
PDF files of version 1.7 and above. So these will be faster protected files.
And the stage one.
Just a minute, please.
Google.
OK, under stage one you have to develop your model for this recovery of this password protected files and this this we will be releasing you some mock test vectors mean some files which are password protected by 15th of September so that you can test with your model you have developed.
Developed and how far successful you're getting. So around 50 files we will be giving

you on 15th of September 10 AM and then we will be giving. There will be 3 levels of files released to you on 26th of October and 1st of November and 6th of November. And these files will be given to you and you can work on these files. But anyhow, this mock test files which we are releasing on 15th of September will not be evaluated at all. That is only to improve your model, but other files will be evaluated. So as and when you are recovering the password of.
Level 1, Level 2, Level 3 files. Please immediately post it back your result in the portal which has been provided for you to upload it. And so there will be two criteria on which your output will be evaluated. One is number of passwords you have recovered which will have a weightage of 75.
The other is the time. So as soon as you recover your password, please kindly ensure that you are posting it to the portal immediately your password so that your time will be counted from that only. So in level one we will be taking 20% of your this thing, level 2 is 30% and level 3 is 50%.
And whichever candidate has been and top 15 to 20 participants who has qualified on all this level one, level 2 and Level 3 will be taken for the final assessment and they will be called here in Delhi and they have to give a they have to.
Port their model on the reference reference platform we are providing. We will be giving a 800 GPU, 800 GPUs for you or to port your model and we will be giving you 24 hours to run your model and you to get back the result. There we will be giving you some test vectors again.

AS **Anil Sangwan** 25:47
Mode.

P **PS07** 26:06
Around some 10 files we will be giving and based on that we will be evaluating your methodology and what novel approach you have used, how your model is scalable, everything will be there and based on that six participants will be selected for stage 2.
And in stage two, we will be having.
Stage two we will be having the similar 6 participants will be working on our the same problem statement but with other different types of files also and the for them the test vectors will be given on 16th of March.
Sorry, mock test vectors will be given on 16th of March and test vectors will be given

on 1st of April 2026. OK, then three participants will be selected, but here the evaluation criteria is number of password cracked, the speed and methodology novelty used.
Everything will have a weightage of 20% OK and in stage two you will be working on two. You have to define your model model developed in stage two as well as the multilingual.
Password protected files will also be given here, so it may be either Indian language or even based language and that will have a weightage of 70%.
They will be and similarly the mock test for that will be given on 1st of September 2026 and the test vectors will be given on 15th of September 2026 and the evaluation will be based on the number of password cracks will have 60% and time of recovery will have a weightage of 20% and your novelty and the approach.
Or will have a weightage of 20%, so the final winner will be selected from the stage 3.
So any doubts you people have?

RM **Raj Mehta** 28:02
In case anybody anybody wants to ask a question, they can raise their hands.
Anupama.

AI **Anupama Itikirala** 28:05
Yeah, yeah. Madam, I want to understand what kind of a data set we can use. As for my understanding, the scenario for this is if I submit, if I submit something like IT returns, I get a PDF as the written statement, right? Where I use?

P **PS07** 28:16
Yeah, please go ahead.

AI **Anupama Itikirala** 28:25
Some kind of a combination of my date of birth plan and everything and I open it. So is the expectation is to see how can I recover that password kind of a scenario?

P **PS07** 28:37
Ma'am, can you repeat your question first? You are speaking something about letters.

Anupama Itikirala 28:41
My question is my question is this how I could not find any data set to take the samples of the to work in a data model. So can you give us some understanding?

PS07 28:52
So yeah, see like if you have you can just develop generate your own data set like you we are looking into Office 2033 and PDF version 1.7 and above files. So you can take one file, you can make it password protector.

Anupama Itikirala 29:06
Hey.

PS07 29:09
And you can take this as a your data set. You can generate as as any number of files you want and you can generate your own data set.

Anupama Itikirala 29:10
Oh.

Rutuja Ghagre 29:10
Hello.

Anupama Itikirala 29:18
That's exactly the question I was asking. If I find my IT returns from the portal that I get an IT returns file of a format which will have a combination of a bank card, date of birth, something, even bank statements. If you want to read a statement, it is the password protected with some combination at such files is what I need.

PS07 29:32
S.

Anupama Itikirala 29:38
To look at it.

PS07 29:39
Yeah, yeah, you can take any combination. You can generate your own password or you can take the password. Some leaked passwords will be available on the open source. Also you can take any password and you can protect yourself and you can check how your model is working.

Anupama Itikirala 29:53
A little more, two more minutes.

PS07 29:54
Right. Yeah, I understand like every banks will have their different format of keeping their credit card bill or your bill statement or anything they are protecting using different format, right. So you can look into those scenarios also like you can.

Anupama Itikirala 30:03
Thanks.
Oh.

PS07 30:12
Take that type of files. Uh, you can generate the password based on all the different bank types and everything.

Anupama Itikirala 30:18
Yeah, from the solution perspective, it is a NLP kind of an application where I need to look at the combination. You're able to apply some kind of a foreign sync algorithm, maybe brute force, all these things are old ones.

PS07 30:34
Yeah.

Anupama Itikirala 30:34
Some artificial intelligence way of fitting and crack how fast I can crack this. This is the solution expected right from the use case perspective.

PS07 30:40
Yes, yes, yes. Solution expected is how fast you can track your password and with.

Anupama Itikirala 30:46
I'll use some novel algorithm rather than the this boot what we say the normal algorithm root boost those kind of things and brute fort I'm I'm there. So I I I should use something else and crack this from the documents right recovery faster.

PS07 30:54
Google.
Yes.

Anupama Itikirala 31:06
Stuff and as many as I could. OK, I got it. So data says can you say somewhere because I did not find anything.

PS07 31:07
Yes.

Anupama Itikirala 31:18
Do you can you see you you some to where can I look for passwords? This kind of a data set sample data set?
Madam.
Hello Madam, can you please help me with the data set where I can look for open source data?

Raj Mehta 31:53
Dr. Ashtalakshmi, are you there?

Anupama Itikirala 32:00
Probably when she comes back, can one of you please ask her the question? Maybe there are more people in queue if I I can't take a lot of your time.

Rutuja Ghagre 32:07
Taarashtra Thakuri Sakuri Sakuri Saha Ki sinpi aapko pada hai.

PS07 32:08
Yeah.

Anupama Itikirala 32:21
Yeah.

Raj Mehta 32:21
Uh, is there any other person who wants to ask any question?
Because I don't see any hand raised.
So I don't see any hand raised who wants to ask any question and since this is second and last webinar on this particular problem statement #7.
Yeah, Imam, go on.

IMAM 32:50
Yeah, so my question is we have to create a complete solution in which your team will provide us the PDF or document in which we have to crack or open that file or we have to just?
Uh, provide the algorithm solution that using this and this combination we can crack this uh password.

PS07 33:13
So you at initial stage you forward your password whatever you have recovered.
And while submitting your final reports, you can give a brief what all is required. You can what a brief of your what you have used. But when you're selected for the final testing, they may ask what novelty you have used and how you have approached the problem, what is the scalability of your.

IMAM 33:37
OK, that's that's I got that. But my main question is like do I have to create a complete GUI based app or just for?

PS07 33:42
Muchilpati.

IMAM 33:56
Providing a solution only.

PS07 33:58
No, no, you can just provide the solution only during at final stage we can think of those things.

IMAM 34:03
OK, OK, so for the initial stage I have to provide the algorithms and the strategy.

PS07 34:09
Your model should work and your model should extract the passwords. You can concentrate on those things and improving your model to get more passwords. Train your model to get more passwords cracked. Finally this outlook and all like GUI and all we can look into the.

IMAM 34:26
OK. And is there any context will be given that which type of document is going to be input so that based on the given input document the AI and like the AI model will be predict?

PS07 34:29
The final speech.

IMAM 34:45
That for this kind of document, which type of algorithm should I have to design? Is there any that type of scenario or it can be anything or any random random document can become?

PS07 34:59
Random document you can take and initial stage may you can look into the

passwords of this whatever leak passwords.
In stage two, we make use of context.
Taste models.

Danies Alex K 35:25
Even in the even the.

PS07 35:25
Whatever password, keep the password sets are there.

Raj Mehta 35:36
I request others. I request others to call and mute themselves.
Yeah.

PS07 35:46
So what I'm trying to say is this thing like initial stage may you you can look into your password sets that will give you the idea how to train your models and when you're fine fine tuning your model at stage 3.
That time we will give you some contextual information based on that you can train your models.
Right now you can concentrate on human thinkable passwords and you can look into the various sites where the leaked passwords are available, how the passwords people use to keep the passwords. Based on that you can train your model.

IMAM 36:25
That I got your point. But again main thing is like suppose if you say we are giving the input as a banking document.

PS07 36:33
Oh.
OK.

IMAM 36:42
So based on banking I can decide and I can train the model that in the banking sector this kind of password combination are can be presentable.

PS07 36:56
Hmm.

IMAM 36:57
Send then if you say it is like a one card database so that format is will be will be different so that in that case I can train model in that way and merge everything. So is that.
That kind of context will be given or like you will give us random document like anything suppose.

PS07 37:16
No, like.
Anything, anything like we cannot specific go to some specific this thing only only banking. It will be only banking documents or something like that so.

IMAM 37:28
OK, so basically under the algorithm I have to design the algorithm in that way that algorithm will identify first, then I will crack the code and that everything all the scenario.

PS07 37:39
Yes.
Yes.

Ajeet 37:42
Pasha, Pasha, if I may just, if I may just intervene here, Imam, the point is the files which will be given to you, we are not categorizing them under any of the heads that is banking or anything. It could be any head. So you can train.

IMAM 37:55
OK, OK, got it.

Ajeet 38:02
And your model to address maximum kind of things. Now you could you could you

could go in for identification of the system first or you could apply all of them in a random form. So that is the approach which you will have to take and we judge what kind of approach you have.

IMAM 38:05
Got it.
OK, that depends on me.

Ajeet 38:19
As far as the document which we are going to offer you are going to be from all kinds of things which are imaginable, which you can imagine. So your model has to be a holistic kind. It has to holistically answer the things your model, your approach, your approach could be a better way of doing a brute forcing also intelligent.

PS07 38:23
OK.

IMAM 38:27
OK.

Ajeet 38:38
You're being unfortunate. It could be anything for that matter, so OK.

IMAM 38:41
I got it, Sir. OK.

Ajeet 38:48
Asha, I suppose I've not shared anything which is away from what we are looking at.

PS07 38:50
It's.
And soon.

Swaroop M N 39:02
Hello, I have a question. May I? Yeah, there are two types of opening the password.

One is opening the PDF itself and another is printing the PDF. Are you looking at only opening the PDF or opening the document?

Raj Mehta 39:03
It's a loop here, yeah.

PS07 39:18
So it is like owner password and user password. Like owner password will look into that opening the opening the PDF and user password. User password will look into the right.
The the rights which you're going to give for rights like whether it should be writable or non editable like that so.
So we are looking for both. If it has been, if it has been protected by both, we are looking for both. If it is not, it has been used to only this thing, then this.

Swaroop M N 40:00
OK. Thank you. And I have another question that is during level one, you said you will be giving us some data sets on 26th of October and we have to submit as early as possible and what would be the latest date that we need to submit?

PS07 40:20
So for the no we yeah that is level one. We as soon as you recover the password you can immediately upload it on the portal. That is what I'm trying to say because your time will be calculated from there how much time you have taken.
Because we will be publishing this on 26th of October at 10:00 AM in the morning. So the time you submit will be your recovery time. So that will also be taken as one of the parameter to evaluate. So it is as soon as you recover the part, how many part we will be sharing around 20 files per.
For level one and 20 files for level 2 and 20 files for Level 3. So as soon as you recover the password you keep posting it immediately. But for we will be closing it on 10th of November 2025. Till that you can post your.

Swaroop M N 41:12
Thank you very much. That's it.

Raj Mehta 41:17
Anupama, would you like to repeat your question when?

Anupama Itikirala 41:23
Hey, Sir, you said something to me.

Raj Mehta 41:26
Yeah, because he had asked some question and Madam was not there at that time. So you can repeat your question now.

Anupama Itikirala 41:30
I was asking Madam Master Lakshmi if she can refer to us some data sets to look into it.
Can see if she can say this because I did not find any such thing. There are a lot of things you talk about the brute force and something. She can suggest something to get started.

PS07 41:54
Oh, you are asking about the leaked data set sites.

Anupama Itikirala 41:58
Yeah, if there's some leak is that that says where I can find because it's very difficult to collect the data, right? So.

PS07 42:06
So you so you just go into the sites of like rock you rock you.com.

Anupama Itikirala 42:16
Rockyou.com OK.

PS07 42:20
GitHub.

Anupama Itikirala 42:21
OK.

PS07 42:23
So you can find on these sites. If you just go to Internet and you type leaked password, you will get so many different sites where the leaked passwords are there. So you can take the you can use that as for training your models.

Anupama Itikirala 42:36
OK. Thank you.
Yeah, this is a little difficult to pick a different problem from all the ones that I saw in this thing. So little different, more algorithmic approach rather than the data science here. It's not a very data science, it's more of an algorithm can show the efficacies and the.

PS07 42:46
Oh, yeah.
Yeah.

Anupama Itikirala 42:58
In cracking that, that's that's where I ask you because I did not find the right way out. Thank you.

PS07 43:03
OK.

Raj Mehta 43:07
Is there any other question?
Yeah, me to settle, Yeah.

Swaroop M N 43:12
Sorry, may I? Yeah, when we will be having a next meeting.

Raj Mehta 43:18

This is the second and last webinar on this particular problem statement. But in the meantime I would like to update all of you that we have started communities in problem statement with communities on the portal and you can post your queries, your doubts question there and the we should also soon be adding all the officers in charge.
In that problem statement with communities and they shall be responding to your queries. But as far as webinar is concerned, this is the last webinar of this particular problem statement.

SN **Swaroop M N** 43:52
Thank you.

RM **Raj Mehta** 43:56
I don't see any any other hand raised. With the permission of the officer Dr. Asthalashmi and Kanna Ajay, shall we close the webinar now?

AS **Anil Sangwan** 44:07
Yes, Sir.

RM **Raj Mehta** 44:09
Thank you. Thank you everyone for being here and asking your questions and wish you all the best. Now I'm ending the webinar for all. Thank you once again.

S **Swati** 44:19
Thank you, Sir.

AI **Anupama Itikirala** 44:19
Thank you.

Raj Mehta stopped transcription