How can I extract the hash inside an encrypted PDF file?

Asked 4 years, 8 months ago Active 1 year, 3 months ago Viewed 25k times

15

7

As long as I know, the **encrypted PDF** files don't store the decryption password within them, but a hash asociated to this password.

When auditing security, a good attemp to break PDF files passwords is extracting this hash and bruteforcing it, for example using programs like HashCat.

What is the proper method to **extract the hash inside a PDF** file in order to auditing it with, say, HashCat?

Answers for **John the Ripper** could be valid too, but I prefer HashCat format due to the easyness of making **GPU computing** work in Windows and bruteforce with **OCLHashCat** (the GPU version of HashCat). John the Ripper has a GPU version too, but JTR has no Windows version, at least with GPU enhancement.

hash pdf

edited Jul 4 '15 at 2:23

1 Answer

19

UPDATE 21 Dec 2017

The script pdf2john.py doesn't exist anymore. It has been substituted by a perl version, pdf2john.pl.

Extracted from HashCat Forums, this method works for me (**requires Perl**):

--Download **pdf2john.pl** from the suite John the Ripper (**OCLHashCat** works with the **same hash format** as John the Ripper):

```
wget https://github.com/magnumripper/JohnTheRipper/archive/bleeding-j
unzip bleeding-jumbo.zip
```

--Use it to **extract the hash** from your .pdf file:

```
perl JohnTheRipper-bleeding-jumbo/run/pdf2john.pl MyPDF.pdf > MyPDF-H
```

--Output file `MyPDF-Hash.txt` must be **edited**. Original would be something like:

```
MyPDF.pdf:$pdf$4*4*128*1028*1*16*652fc762fdb12c47a5f90ddd2b99b809*32*
```

so use your preferred editor:

```
nano MyPDF-Hash.txt
notepad MyPDF-Hash.txt
```

and leave only the **part inside double colons** `:` :

```
$pdf$4*4*128*1028*1*16*652fc762fdb12c47a5f90ddd2b99b809*32*dd86d858f9
```

--**Hint**: you can do the extraction and the edition in **one step** by using `sed` (UnxUtils version too, if you are doing it from Windows):

```
perl JohnTheRipper-bleeding-jumbo/run/pdf2john.pl MyPDF.pdf | sed "s/
sed "s/^.*://" > MyPDF-Hash.txt
```

--Your `MyPDF-Hash.txt` file is now **ready to use** with OCLHashCat (or John the Ripper).

NOTES:

- Tested working on CygWin (Windows).
- Tested working on Kali and Ubuntu Linux.

edited Jul 31 '18 at 18:51

vakus

3,513 3 17 29

answered Feb 21 '15 at 16:01

Sopalajo de Arrierez

561 1 5 14

pdf2john.py doesn't exist anymore. It has been substituted by a perl version – tpvasconcelos Dec 21 '17 at 2:59

1 Copying the perl file out of the directory does not work. It needs to be where it is, in the "run" directory, otherwise you'll get an error. – Eric Brandel Jun 10 '18 at 5:50

Hey, this answer doesn't work, first because you have the file extension py and the file is a perl script. – Philippe Delteil Jul 31 '18 at 16:56

1 To get rid of the irrelevante text on the hash, use this perl JohnTheRipper-bleeding-jumbo/run/pdf2john.pl MyPDF.pdf | awk -F":" '{ print $2}' > MyPDF-Hash.txt – Philippe Delteil Jul 31 '18 at 17:36

I suggest an edit to this answer: don't download pdf2john.pl from the repository, just download the whole repository and run pdf2john.pl from within it. – Baodad Jan 31 at 22:02