Introdução a Investigação Forense, e Análise de Metadados PDF

Segurança da Informação

Charles Tim Batista Garrocho

Instituto Federal de São Paulo – IFSP
Campus Campos do Jordão

garrocho.ifspcjo.edu.br/SEGA6

charles.garrocho@ifsp.edu.br

Curso Superior de TADS

Ciência Forense e Computação Forense

A **Ciência Forense** é compreendida como o conjunto de todos os conhecimentos científicos e técnicas que são utilizados para **desvendar** não só crimes, como também variados assuntos legais.

Ela é considerada uma área **interdisciplinar** pois envolve física, Química, biologia, entre outras. Tem como objetivo principal o suporte a investigações referentes a justiça civil e criminal.

A **Computação Forense** é o uso de métodos científicos para preservação, coleta, validação, identificação, análise, interpretação, documentação e apresentação de evidência digital com validade probatória em juízo.

Com os discos rígidos atingindo a capacidade de TeraBytes, milhões de arquivos podem ser armazenados. Logo, é necessário a utilização de **métodos e técnicas** de Computação Forense para encontrar a prova desejada que irá solucionar um crime, por exemplo.

Anonymous Metadata Fail

Em 10 de dezembro de 2010, o grupo de hackers Anonymous publicou um **comunicado** de imprensa descrevendo as motivações por trás de um recente ataque chamado Operation Payback.

Irritado com as empresas que abandonaram o suporte para o site WikiLeaks, a Anonymous pediu retaliação ao realizar um ataque distribuído de negação de serviço (DDoS) contra algumas das partes envolvidas.

A Anonymous publicou o comunicado de imprensa sem assinatura e sem atribuição. Distribuído como um arquivo de formato de documento portátil (**PDF**), o boletim de imprensa continha metadados.

Além do programa usado para criar o documento, os metadados PDF continham o **nome do autor**, o Sr. Alex Tapanaris. Em poucos dias, a polícia grega prendeu o Sr. Tapanaris.

Investigação Forense com Python

Vamos **recriar** rapidamente a investigação forense de um documento que se mostrou útil na prisão de um membro do grupo hacker Anonymous.

O site Wired.com ainda **disponibiliza** o comunicado. Podemos começar baixando o documento usando o utilitário wget:

```
tim@charles:~$ wget http://www.wired.com/images_blogs/threatlevel/2010/12/ANONOPS_The_Press_Release.pdf
```

Agora será necessário **instalar** a biblioteca PyPDF:

```
tim@charles:~$ sudo apt-get install python-pypdf
```

A biblioteca **PyPDF** vai permitir manipular arquivos PDF. Sendo assim, útil para descobrir metadados nesse comunicado.

Investigação Forense com Python e PyPDF

O script aceita como entrada o arquivo de PDF a ser analisado. A biblioteca **optparse** é utilizada para analisar a entrada.

```
16 def main():
17     analisador = optparse.OptionParser("use %prog "+\
18       "-F <arquivo PDF>")
19     analisador.add_option('-F', dest='arquivo',\
20         type='string', help='especifique o arquivo PDF')
21
22     (opcoes, args) = analisador.parse_args()
23     arquivo = opcoes.arquivo
24     if arquivo == None:
25         print analisador.usage
26         exit(0)
27     else:
28         imprimir_meta(arquivo)
```

Para extrair metadados, é utilizado o método **getDocumentInfo()**. Este método retorna uma matriz de tuplas.

Cada **tupla** contém uma descrição do elemento de metadados e seu valor.

```
 8 def imprimir_meta(arquivo):
 9     arq_pdf = PdfFileReader(file(arquivo, 'rb'))
10     info_doc = arq_pdf.getDocumentInfo()
11     print '[*] PDF MetaData para: ' + str(arquivo)
12     for item_meta in info_doc:
13         print '[+] ' + item_meta + ':' + info_doc[item_meta]
```

Iterando através desta matriz imprime todos os metadados do documento PDF.

Investigação Forense com Python e PyPDF

Abaixo é apresentado um exemplo de execução deste script.

```
tim@charles:~$ python metadados_pdf.py -F ANONOPS_The_Press_Release.pdf
[*] PDF MetaData para: ANONOPS_The_Press_Release.pdf
[+] /Author:Alex Tapanaris
[+] /Producer:OpenOffice.org 3.2
[+] /Creator:Writer
[+] /CreationDate:D:20101210031827+02'00'
tim@charles:~$
```

Neste caso, com a execução do script contra o comunicado do grupo Anonymous, pode-se visualizar os metadados que levaram as autoridades gregas a prender o Sr. Tapanaris.

Atividades

Agora que demonstramos que podemos descobrir um autor de um arquivo PDF, vamos expandi-lo para fazer uma investigação um pouco mais complexa.

Você deverá procurar na Internet, pelo menos 10 arquivos PDF de locais distintos.

Após isso desenvolva um script que irá receber o nome de um autor e irá verificar se o mesmo consta nos 10 PDFs. O script irá receber como entrada uma pasta que irá conter os arquivos PDF.

Coloque o resultado em um arquivo, contendo na primeira linha o autor, e nas próximas linhas os nomes dos arquivos que o pertence.