Introduction to Image Analysis

EN | ZH 图像文件能够很好地包含黑客文化,因此 CTF 竞赛中经常会出现各种图像文件。

图像文件有多种复杂的格式,可以用于各种涉及到元数据、信息丢失和无损压缩、校验、隐写或可视化数据编码的分析解密,都是 Misc 中的一个很重要的出题方向。涉及到的知识点很多(包括基本的文件格式,常见的隐写手法及隐写用的软件),有的地方也需要去进行深入的理解。

元数据 (Metadata)

元数据(Metadata),又称中介数据、中继数据,为描述数据的数据(Data about data),主要是描述数据属性(property)的信息,用来支持如指示存储位置、历史数据、资源查找、文件记录等功能。

元数据中隐藏信息在比赛中是最基本的一种手法,通常用来隐藏一些关键的 Hint 信息或者是一些重要的如 password 等信息。

这类元数据你可以 右键 --> 属性 去查看, 也可以通过 strings 命令去查看, 一般来说, 一些 隐藏的信息(奇怪的字符串)常常出现在头部或者尾部。

接下来介绍一个 identify 命令, 这个命令是用来获取一个或多个图像文件的格式和特性。

-format 用来指定显示的信息,灵活使用它的 -format 参数可以给解题带来不少方便。 format 各个参数具体意义

例题

Break In 2017 - Mysterious GIF

这题的一个难点是发现并提取 GIF 中的元数据,首先 strings 是可以观察到异常点的。

```
GIF89a
!!!""###$$$%%&&&'''((()))***+++,,,--
-...///000111222333444555666777888999:::;;;<<<===>>>???
@@@AAABBBCCCDDDEEEFFFGGGHHHIIIJJJKKKLLLMMMNNNOOOPPPQQQRRRSSSTTTUUUVVVWWWXXXYYY

4d494945767749424144414e42676b71686b6947397730424151454641415343424b6b77676753

NETSCAPE2.0
ImageMagick
...
```

这里的一串 16 进制其实是藏在 GIF 的元数据区

接下来就是提取,你可以选择 Python,但是利用 identify 显得更加便捷

```
root in ~/Desktop/tmp λ identify -format "%s %c \n" Question.gif
0
4d494945767749424144414e42676b71686b6947397730424151454641415343424b6b77676753

1
5832773639712f377933536849507565707478664177525162524f72653330633655772f6f4b38
...
24
484b7735432b667741586c4649746d30396145565458772b787a4c4a623253723667415450574c
25 724b3052485a6b745062457335797444737142486435504646773d3d
```

其他过程这里不在叙述,可参考链接中的 Writeup

像素值转化

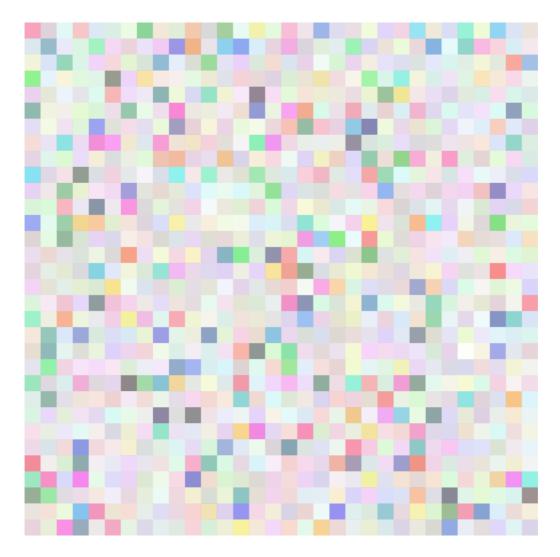
看看这个文件里的数据, 你能想到什么?

```
255,255,255,255,255.....
```

是一串 RGB 值,尝试着将他转化为图片

而如果反过来的话,从一张图片提取 RGB 值,再对 RGB 值去进行一些对比,从而得到最终的 flag。

这类题目大部分都是一些像素块组成的图片, 如下图



相关题目:

- CSAW-2016-quals:Forensic/Barinfun
- breakin-ctf-2017:A-dance-partner

评论