Report

Assignment 1
Binary Entropy Analysis Tool

Submitted by
Vaibhav Gajanan Lokhande
2017-18

Contents

Chapter 1

Assignment

1. Generate a file containing 100 bytes, the values of 0's.
2. Calculate entropy of this file.
3. Generate another binary file having 50 0's and 50 1's, and calculate entropy.
4. Apply RAR or ZIP and create test3.rar or test.zip and calculate its entropy.
5. Apply encryption of test2.bin and calculate its entropy.
6. Calculate entropy of any windows app like notepad.exe or paint.exe.
7. Try to download a malware file and calculate entropy.

Chapter 2

Theory

Entropy is the measure of randomness of data in a message or file. By calculating entropy, we can determine the randomness of data or occurrence of different characters and thus determine the complexity of any message or file. For example a simple text file will have less entropy compared to its compressed form.
The entropy is given by formula

$$H = -\sum p_i * log_2(p_i)$$

where, p_i is the probability of occurance of character i.
Using entropy, we can identify the encrypted or compressed executable, which can be used to detect the malware.

Chapter 3

Code

The program is developed in matlab.

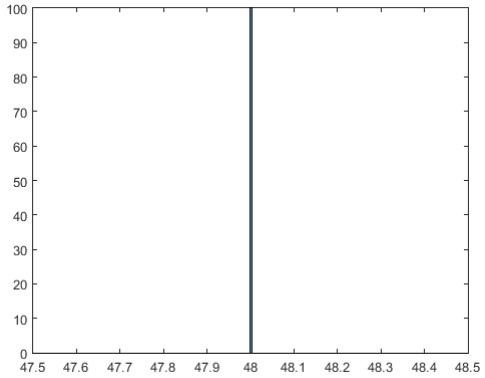
Entropy.m

```
fileName=input('Enter name of file:','s');
file=fopen(fileName,'rb');
fileData=fread(file);
fileLength=length(fileData);
h=histogram(fileData,256);
frequency=h.Values;
fclose(file);
probability=frequency./fileLength;
i=probability>0;
entropy=-sum(probability(i).*log2(probability(i)))
```

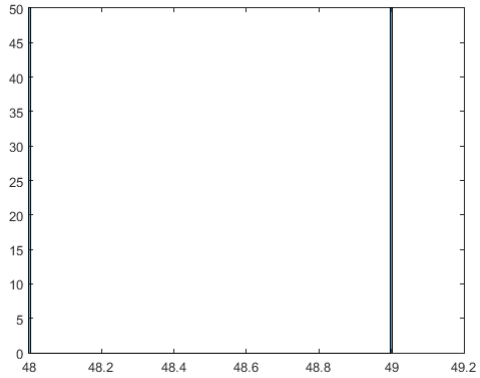
Chapter 4

Observations

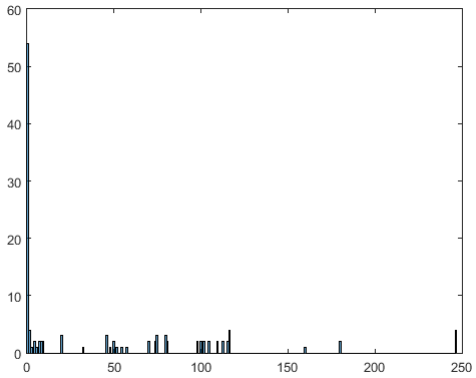
Sr.No.	File Name	Description	Entropy
1	test1.bin	File with all 0's and size 100 bytes.	0
2	test2.bin	File with 50 0's & 50 1's and size 100 bytes.	1
3	test3.zip	test2.bin compressed to zip.	3.7898
4	test2.bin.gpg	Encrypted test2.bin	7.2973
5	notepad.exe	Windows notepad application.	6.9442
6	test5.bin	File with half 0's & half 1's and size is 1MB.	1



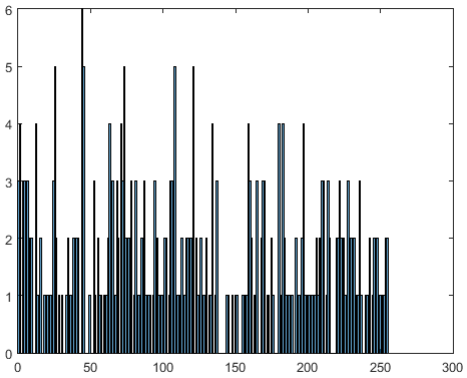
(a) Result for test1.bin



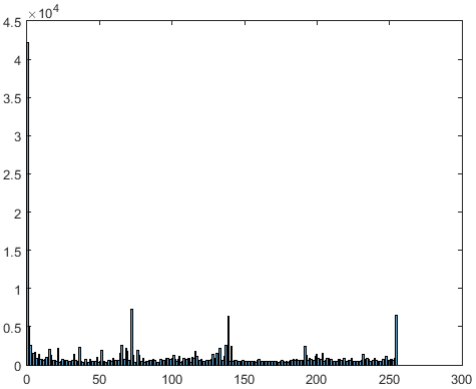
(b) Result for test2.bin



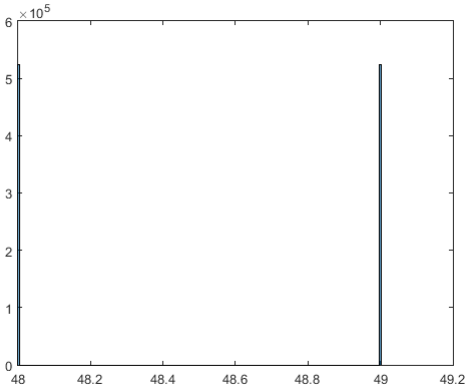
(a) Result for test3.zip



(b) Result for test2.bin.gpg



(a) Result for notepad.exe



(b) Result for test5.bin

Chapter 5

Conclusion

From the observations, we can conclude following points.

1. Entropy of file increases with increase in randomness of data, 0 when all characters are same and 8 when all characters are different.

2. Entropy file can be anything irrespective of file size. This can be concluded from observation no 2 and 6, where size of file is different but the type of data is same. So their entropy is same.

3. Entropy of encrypted files is very high, more than 7. Using this observation we can determine whether a binary file is encrypted or not.