EECS 4980:805 Inside Cryptography

DES Analysis Project

David Carek

The purpose of this analysis is to understand the result of applying the Data Encryption Standard (DES) algorithm. The data collected will help to paint a picture of what the algorithm does as a whole which is difficult to observe by just looking at an encrypted file.

The first data collected is on byte frequencies. The graphs below illustrate the byte frequencies before and after encryption. Prior to encryption the byte frequencies have widely varying values. The most important take away from these graphs is the variation in the range. In the plaintext the space character occurs 767104 times. No character in the encrypted file comes anywhere near to having this many occurrences. The most frequent character that occurs in the encrypted file occurs 20724 times and is not even an ASCII character. At a glance it is fairly easy to tell that the points are much more evenly distributed. This validates the requirements for confusion that DES is supposed to produce.

Unsorted Byte Frequencies:

The data in the charts above is a helpful tool to understand

|  |  |  |
| --- | --- | --- |
| Statistics of sorted frequencies | Unencrypted | Encrypted |
| Min | 0.00 | 18924.00 |
| Max | 767104.00 | 20724.00 |
| Range | 767104.00 | 1800.00 |
| Mean | 19740.14 | 19740.19 |
| Median | 0.00 | 19692.00 |
| Standard Deviation | 70869.18 | 372.69 |

One of the more interesting parts of this project was comparing a bitmap image with its encrypted pair. One of the problems with the DES in ECB mode is that any given block always encrypts to the same encrypted block. If there is a reoccurring pattern of 8 bytes in the plaintext, then the ciphertext will similar patterns. This is most easily seen by looking at the images below. The white background and the yellow text of the Toledo Rocket image always encrypts to the same pattern. This makes the original image more visible in the ciphertext even though it has been encrypted. When patterns are present in the plaintext, we start to lose some of the confusion. If we made an analysis of frequent octets in the English language and had a large enough ciphertext, then we might be able to back into the plaintext more easily.

Unencrypted and Encrypted Bitmap File:



This problem of reoccurring octets is also visible in the octet analysis of the Shakespeare text.