EECS 4980:805 Inside Cryptography

DES Analysis Project - CBC

David Carek

Analysis of DES CBC Mode

In this analysis we will compare the DES in electronic code book (ECB) mode with cipher block chaining (CBC) mode. We will compare statistics and graphs of both CBC and ECB modes to try to find any patterns in the CBC mode. Since CBC mode XORs the previous encrypted 8-byte block with the unencrypted data of the next 8-byte block we hope to see all potential patterns removed as well as frequencies of characters, di-grams, tri-grams, and octets kept beneath a normal noise ratio.

First, we will look at bit frequencies for each file used in the analysis. This information is displayed in the table below. The differences between the high and low bit frequencies for unencrypted frequency, encrypted ECB frequency, and encrypted CBC frequency are 4,769,536, 25,936, and 3,986, respectively. When CBC is compared to ECB or the unencrypted file, it has a much more random distribution. Ideally, after encryption, we would have the same number of high and low bits. When comparing ECB to CBC, CBC has a much lower difference in the high and low bits than ECB. This proves that CBC is more effective than ECB at the bit level.

|  |  |  |  |
| --- | --- | --- | --- |
| Bit Statistics | Unencrypted Frequency | Encrypted ECB Frequency | Encrypted CBC Frequency |
| 1 | 17,829,132 | 20,226,920 | 20,211,991 |
| 0 | 22,598,668 | 20,200,984 | 20,215,977 |

Next, we will look at the byte statistics in figure 1a. Since the focus of this analysis is on the comparison between ECB and CBC we will not discuss the unencrypted frequency much. Even though ECB mode raises the minimum and lowers the maximum byte frequencies by 18,924 and 765,031 respectively, CBC does a better job of evening out these values. CBC raises the minimum and lowers the maximum frequencies by 19,424 and 747,061 respectively. This outcome is also reflected in the range of the frequencies for both different modes of encryption. Since CBC helps eliminate byte frequency outliers in the encrypted data, we also get a mean and median that are closer together. The value of the median is typically a better representation of an average value since it is not affected by outliers. The difference in ECB and mean and median byte frequencies is 48.69. For CBC this difference is only 1.28. This data proves that CBC does a better job at reducing outliers in the byte frequencies. In other words, CBC better reduces any potential pattern in the encrypted file more than ECB does. Another important statistic to note is the standard deviation of byte frequencies. The standard deviation of CBC byte frequencies is 35% of the ECB byte frequency. This means that the distribution of the encrypted byte frequencies in CBC mode is more random than that of ECB mode. If we look at figures 2a-2c we can visually verify the statistics that were examined. If we look at the slope for each of the graphs we notice that the general trend levels off, going from the unencrypted file, to the ECB encrypted file, and finally the CBC encrypted file. The scale was set to the same range for both encrypted graphs for a more visible distinction between the two. When just looking at the byte data it seems that CBC mode is much better at distributing the byte frequencies evenly.

Since the byte data has been examined we will look at the di-gram data next. Looking at the di-gram statistics if figure 1b we see very similar results as in the byte statistics of figure 1a. The minimum for di-gram frequencies increases for both ECB and CBC mode by 26 and 41 respectively. The maximum for the di-gram frequencies decreases for both ECB and CBC by 183,432 and 183,669 respectively. The range is also drastically reduced for both ECB and CBC compared to the unencrypted file. CBC has a more favorable value for the range though since it is only 79, compared to ECB which has a range of 331. We also have the case where the mean and median are closer together for CBC mode compared to ECB mode. The standard deviation for CBC is also lower than the standard deviation for ECB when looking at di-gram frequencies. Again, a lower standard deviation means that the frequencies of the di-grams, like the bytes, are more evenly distributed. Looking at figures 3b and 3c we can easily tell that the CBC di-grams have many fewer outliers on either end of the graphs compared to ECB. Switching to CBC mode has helped remove the patterns in the data much more noticeably.

The tri-gram data is next on list to be examined. If we look at figure 1c we can begin to process the statistics. The interesting result of the minimum not changing between each of the different files is likely to be caused by the selected data. Keeping track of tri-grams that only occur once or no times causes this list of data to be too large for excel to process it. So the decision was made to leave those data out of the calculation which explains why all of the minimum tri-gram frequencies are equal to 2. The maximum, however, for each file follows the trend in the di-gram and byte statistics. The mean and median also follow the same pattern of converging more closely for CBC versus ECB. Another interesting statistic to note is the standard deviation for the CBC tri-grams. The patterns in the trigrams are so well distributed that the standard deviation is only 0.33. If we look at figures 4a-4c we can see the trend of each graph leveling out more than the previous. ECB reduces the initial spike quite a lot, but CBC reduces it to almost nothing.

The octet data is the most important outcome of CBC mode for DES. In figure 1d we can see that the unencrypted and ECB encrypted file statistics are the same. This is because ECB does not have intra-block confusion. However, in CBC mode DES uses XOR between the encrypted text of the previous block with the current block of plaintext. This allows any pattern of a given block to be thoroughly mixed with the encrypted block before it. This intra-block confusion worked very well for the test case of encrypting the Shakespeare text with the key ‘Pa$$w0rd’ because not a single octet was repeated. This is essentially best case scenario for what we wanted to accomplish with CBC. There existed a pattern between the unencrypted file and an ECB encrypted file. After using CBC instead of ECB to encrypt the file, the pattern was eliminated. Figures 5a and 5b show the correlation between the unencrypted file and ECB encrypted file. Figure 5c shows that this pattern is eliminated. For an even better visual representation we can look at Figures 6a-6c. Figure 6a is a bitmap of the Toledo Rockets logo. This image was sent through the DES encryption algorithm in ECB mode to produce figure 6b. Then figure 6a was run through DES in CBC mode to produce figure 6c. Figure 6b still has remaining patterns that allow you to see the original image, whereas figure 6c just looks like noise.

Another pattern found when analyzing ECB mode was that bytes, di-grams, and tri-grams would occur more frequently in octets that occurred more frequently. So the top 30 octets usually had one or more of the top 30 bytes, di-grams, and tri-grams. This pattern occurs in both the ECB encrypted file and unencrypted file. CBC eliminated this pattern because the top 30 octets were chosen arbitrarily since all frequencies for the octets were 1. Figure 7a and 7b display the pattern from the previous analysis. Figure 7a was for the ECB encrypted file and Figure 7b was for the unencrypted file. Since the first byte of Figure 7b occurred in almost all of the octets the color coating was stopped after the fifth byte so that overlap did not need to be considered.

To summarize, ECB was a good start to file encryption. It does a very good job at adding confusion and diffusion at the inter-block level. CBC is definitely a good step forward from ECB since it adds intra-block confusion. This allows for much better distributions at every level of analysis.

Figures

|  |  |  |  |
| --- | --- | --- | --- |
| **Byte Statistics** | **Unencrypted Frequency** | **Encrypted ECB Frequency** | **Encrypted CBC Frequency** |
| Min | 0 | 18924 | 19424 |
| Max | 767104 | 20723 | 20043 |
| Range | 767104 | 1799 | 619 |
| Median | 0.00 | 19691.50 | 19741.50 |
| Mean | 19740.14 | 19740.19 | 19740.22 |
| Standard Deviation | 70869.18 | 372.75 | 131.15 |

Figure 1a.

|  |  |  |  |
| --- | --- | --- | --- |
| **Di-gram Statistics** | **Unencrypted Frequency** | **Encrypted ECB Frequency** | **Encrypted CBC Frequency** |
| Min | 1 | 27 | 42 |
| Max | 183790 | 358 | 121 |
| Range | 183789 | 331 | 79 |
| Median | 267.00 | 74.00 | 77.00 |
| Mean | 2840.63 | 77.11 | 77.11 |
| Standard Deviation | 9405.39 | 21.22 | 8.73 |

Figure 1b.

|  |  |  |  |
| --- | --- | --- | --- |
| **Tri-gram Statistics** | **Unencrypted Frequency** | **Encrypted ECB Frequency** | **Encrypted CBC Frequency** |
| Min | 2 | 2 | 2 |
| Max | 67523 | 242 | 6 |
| Range | 67521 | 240 | 4 |
| Median | 368.06 | 3.49 | 2.11 |
| Mean | 41.00 | 2.00 | 2.00 |
| Standard Deviation | 1565.19 | 4.65 | 0.33 |

Figure 1c.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Octet Statistics** | **Unencrypted Frequency** | **Encrypted ECB Frequency** | | **Encrypted CBC Frequency** |
| Min | 1 | | 1 | 1 |
| Max | 241 | | 241 | 1 |
| Range | 240 | | 240 | 0 |
| Median | 1.62 | | 1.62 | 1.00 |
| Mean | 1.00 | | 1.00 | 1.00 |
| Standard Deviation | 2.80 | | 2.80 | 0.00 |

Figure 1d.

Figure 2a.

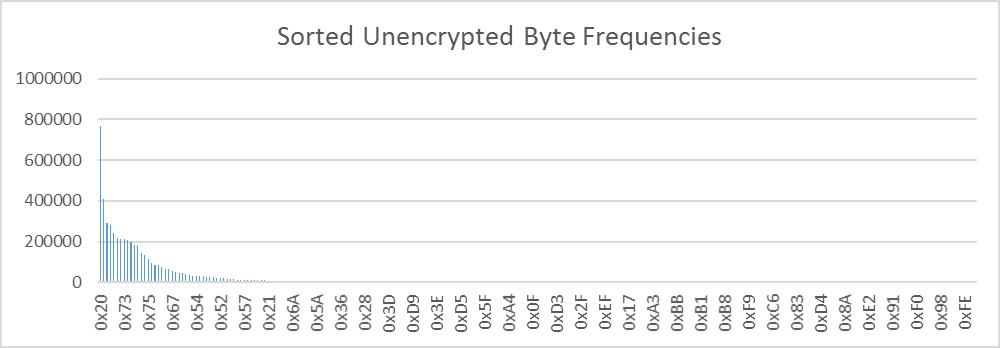


Figure 2b.

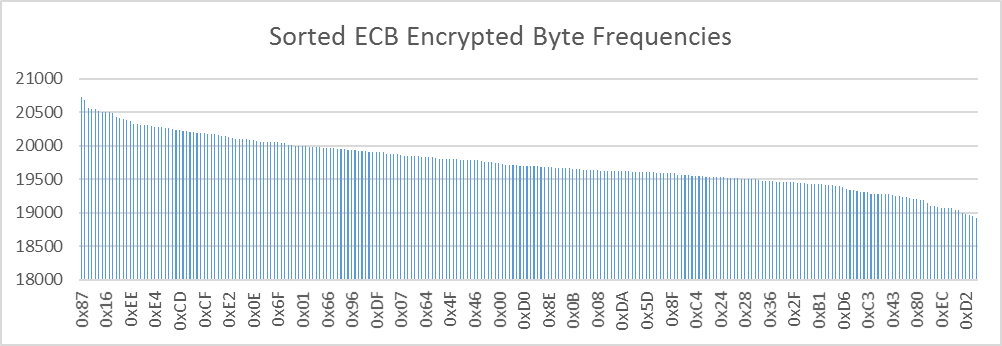


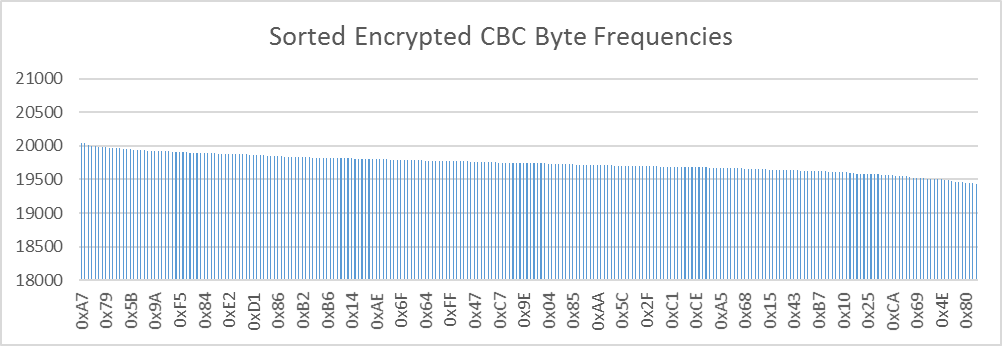
Figure 2c.

Figure 2d.

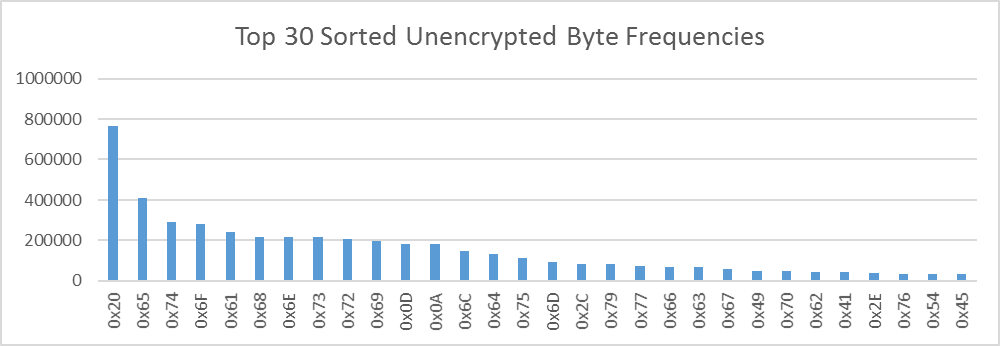


Figure 2e.

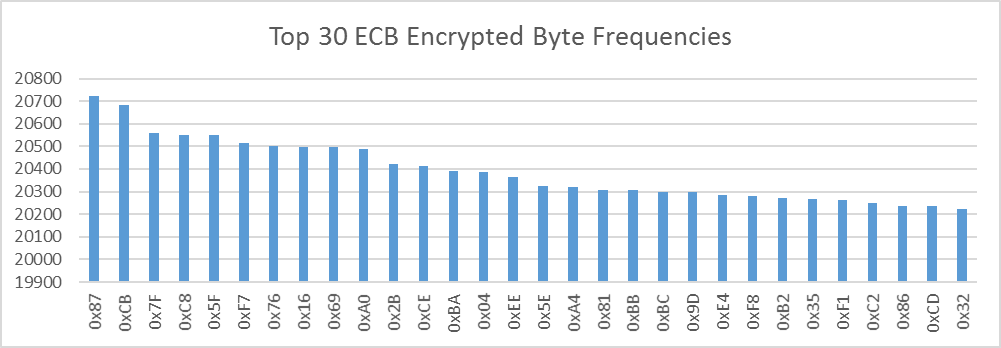


Figure 2f.

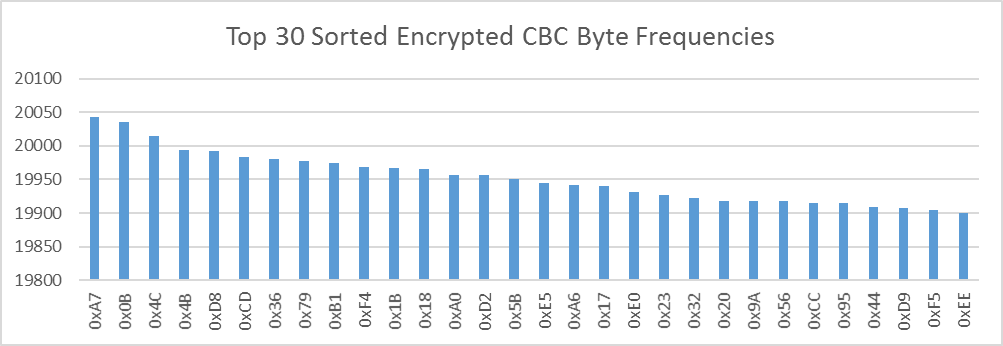


Figure 3a.

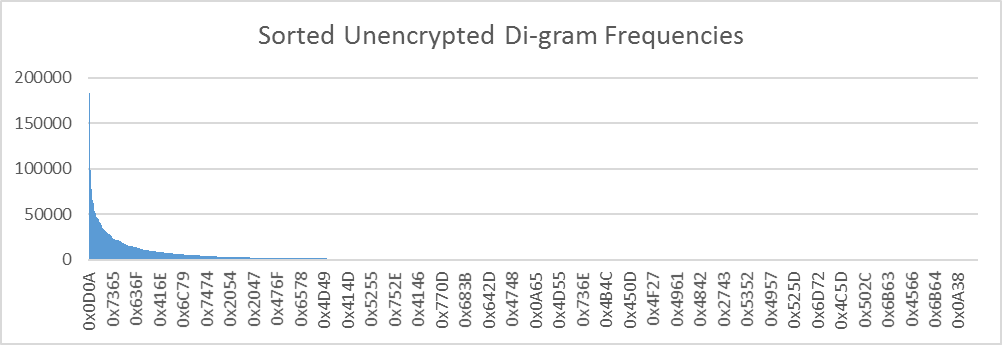


Figure 3b.

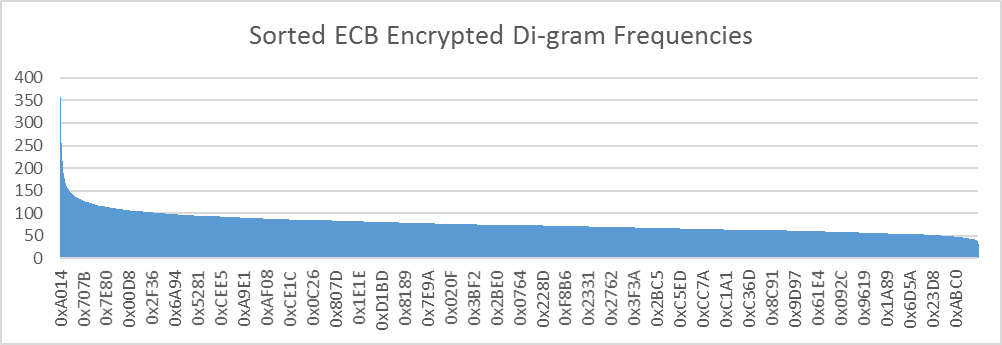


Figure 3c.

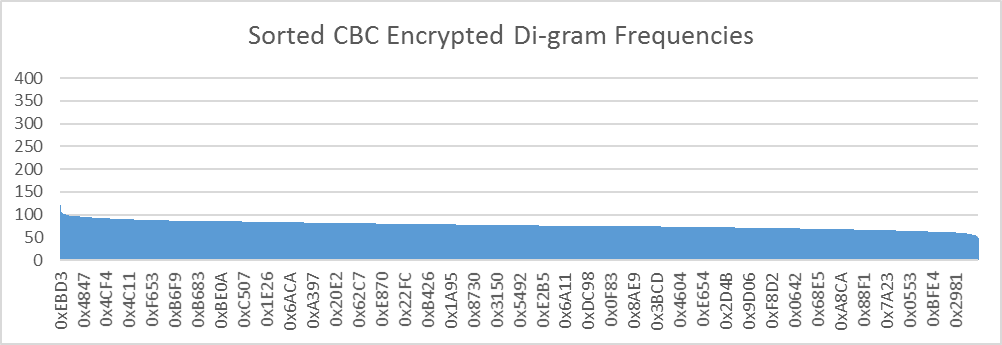


Figure 3d.

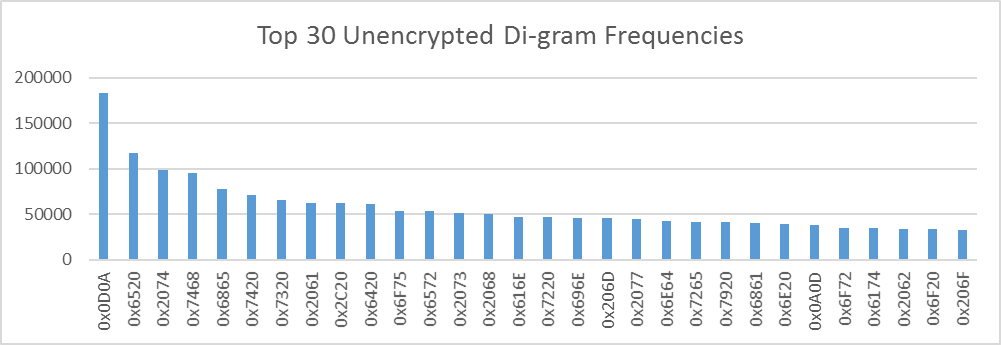


Figure 3e.

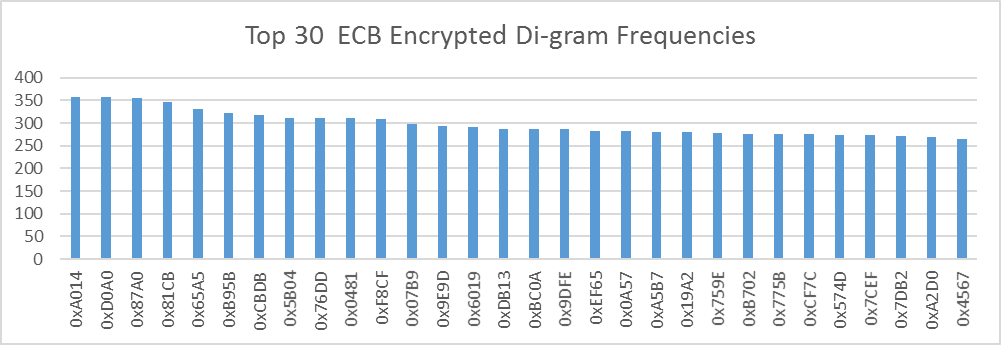


Figure 3f.

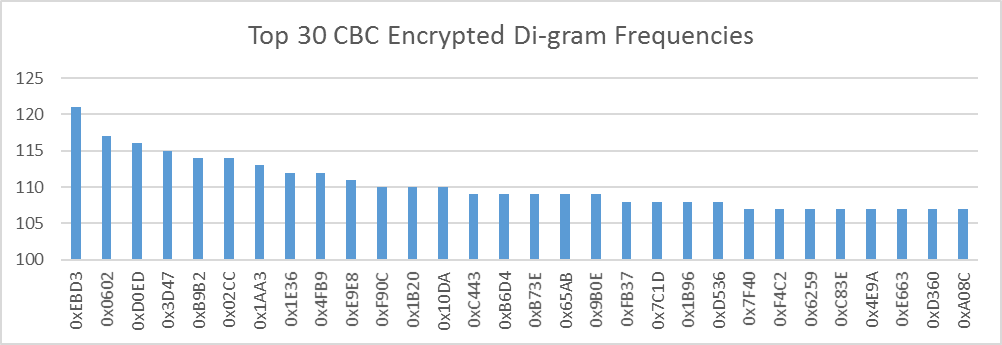


Figure 4a.

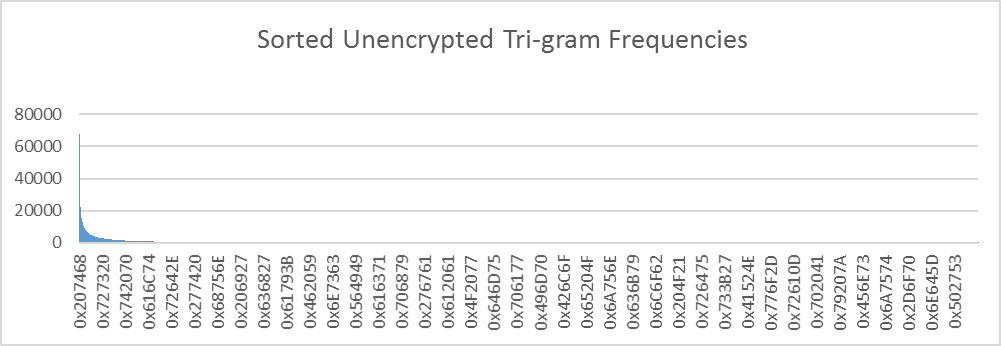


Figure 4b.

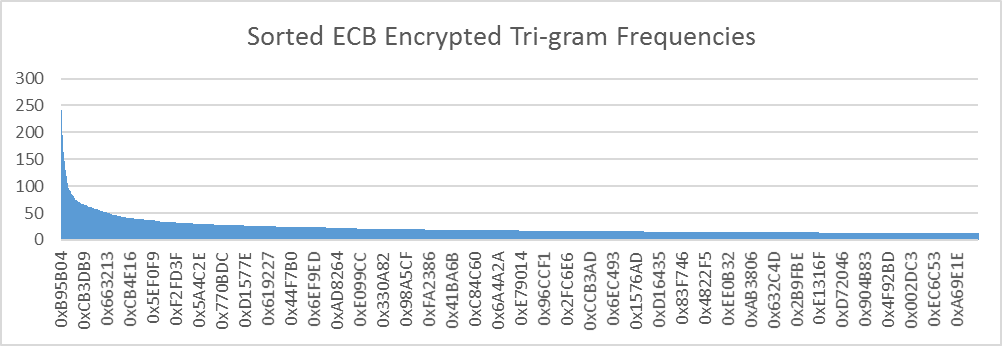


Figure 4c.

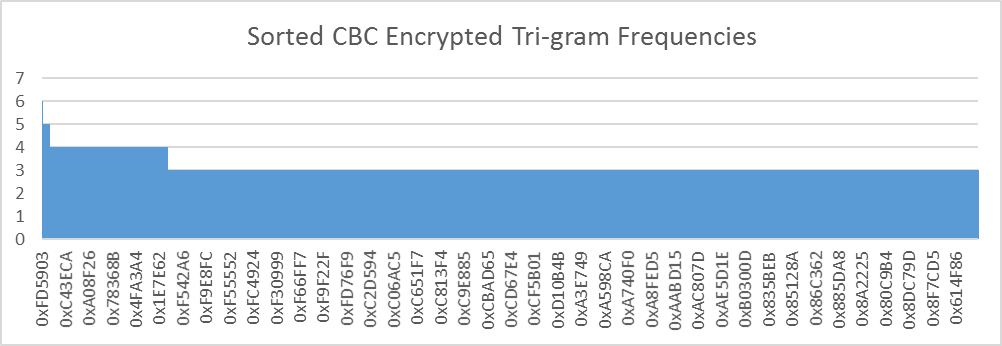


Figure 4d.

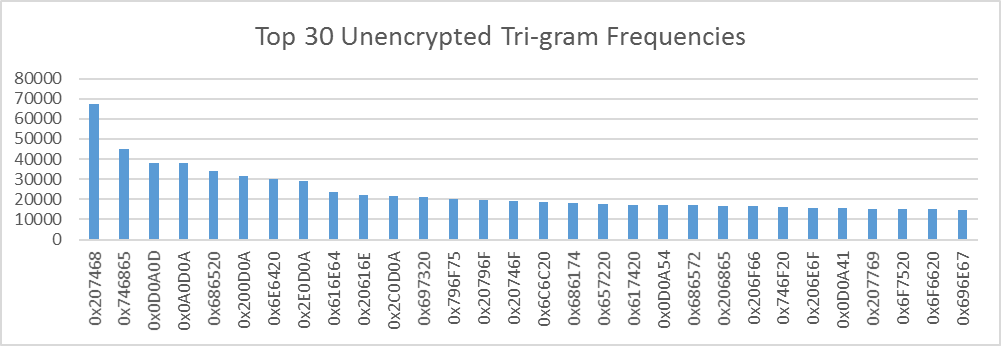


Figure 4e.

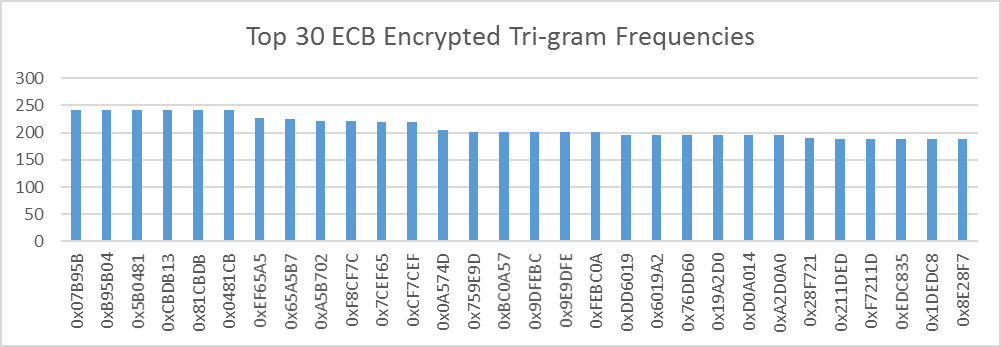


Figure 4f.

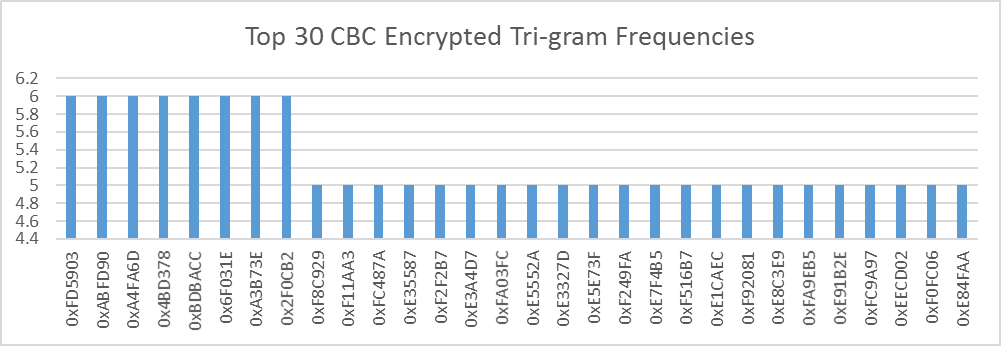


Figure 5a.

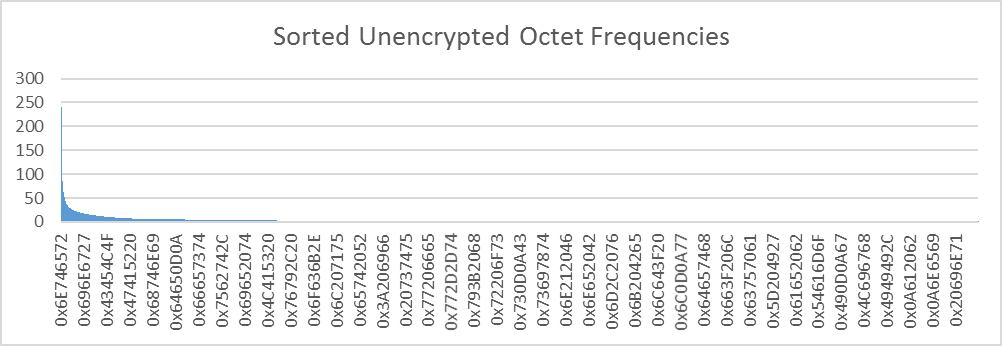


Figure 5b.

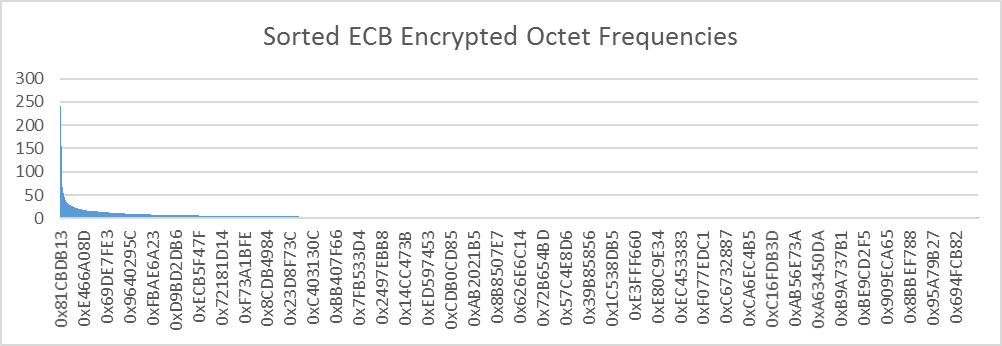


Figure 5c.

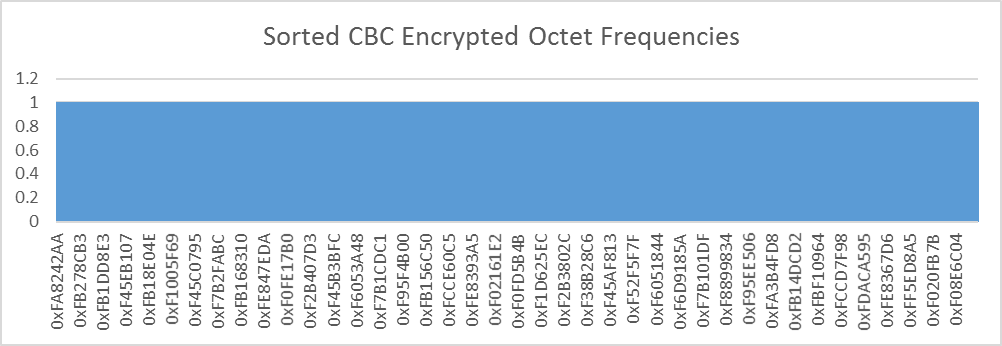


Figure 5d.

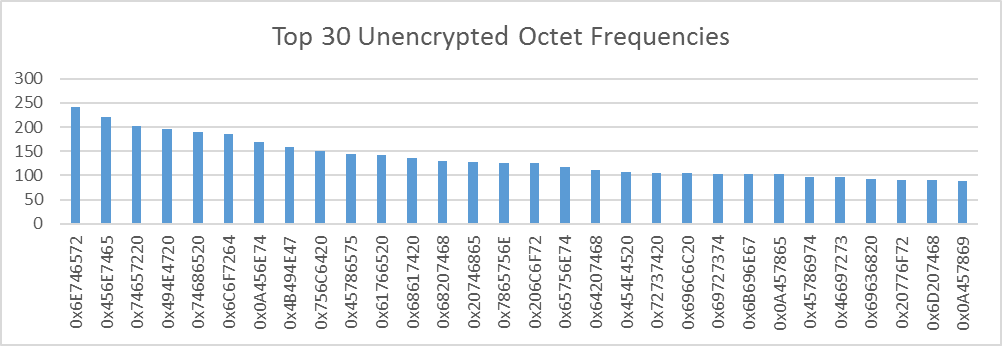


Figure 5e.

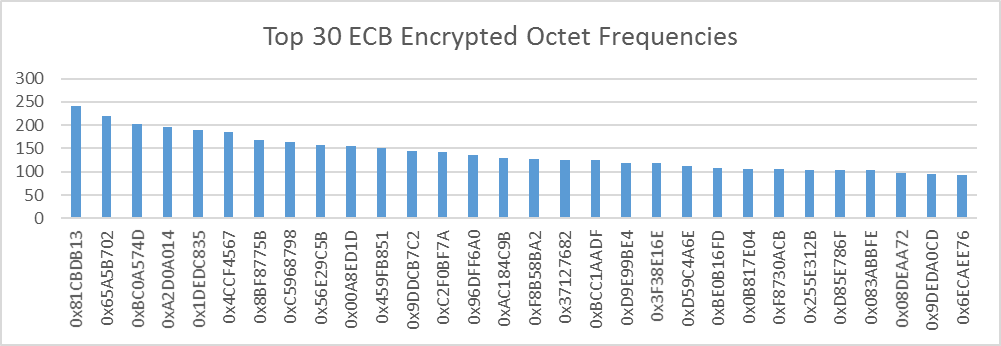


Figure 5f.

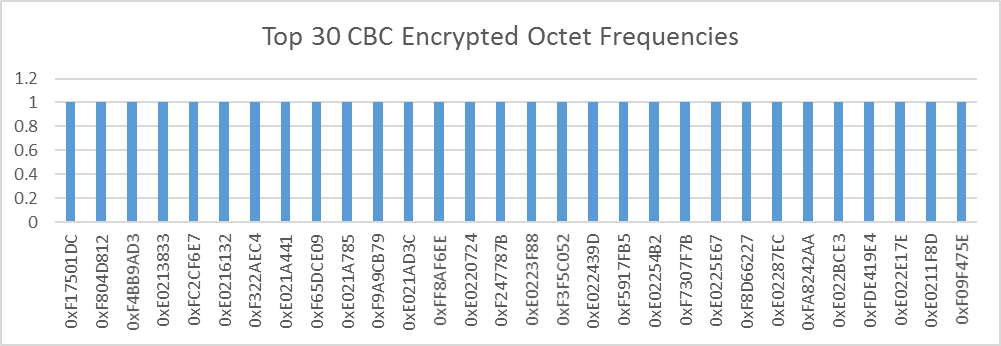


Figure 6a.



Figure 6b.



Figure 6c.



Figure 7a.



Figure 7b.

