DES Analysis

# Character Frequency Analysis

While analyzing the performance of DES (from a security standpoint, not necessarily a performance standpoint), four collections of data were obtained through a slightly modified DES program. In some instances, certain functions of DES were disabled to analyze how each “piece of the puzzle” helps secure the data (either through confusion or diffusion) from certain patterns in the output. These four collections of character frequencies are:

1. Character frequencies in the collective works of Shakespeare (plaintext)
2. Character frequencies in the collective works of Shakespeare after being encrypted with DES (using the password 0x0123456789ABCDEF)
3. Digraph / Trigraph analysis of the plaintext
4. Digraph / Trigraph analysis of the encrypted output

## 1. Character Frequencies in the Collective Works of Shakespeare (Plaintext)

Figures 1 and 2 show the most popular character frequencies recorded (only characters with at least one occurrence were recorded in the entire data table and chart). Unsurprisingly, vowels are the most popular letters, followed by T, S, and N. The most used non-A-to-Z character is the space character, along with the carriage return and line feed characters. Fitting a chart and data table of the necessary size is quite large, so it is recommended to view the included Excel document (DESAnalysis.xlsx) to view the full data table and chart at a more reasonable size.



Figure 1 - Character Frequency for A-Z and Some Special Characters

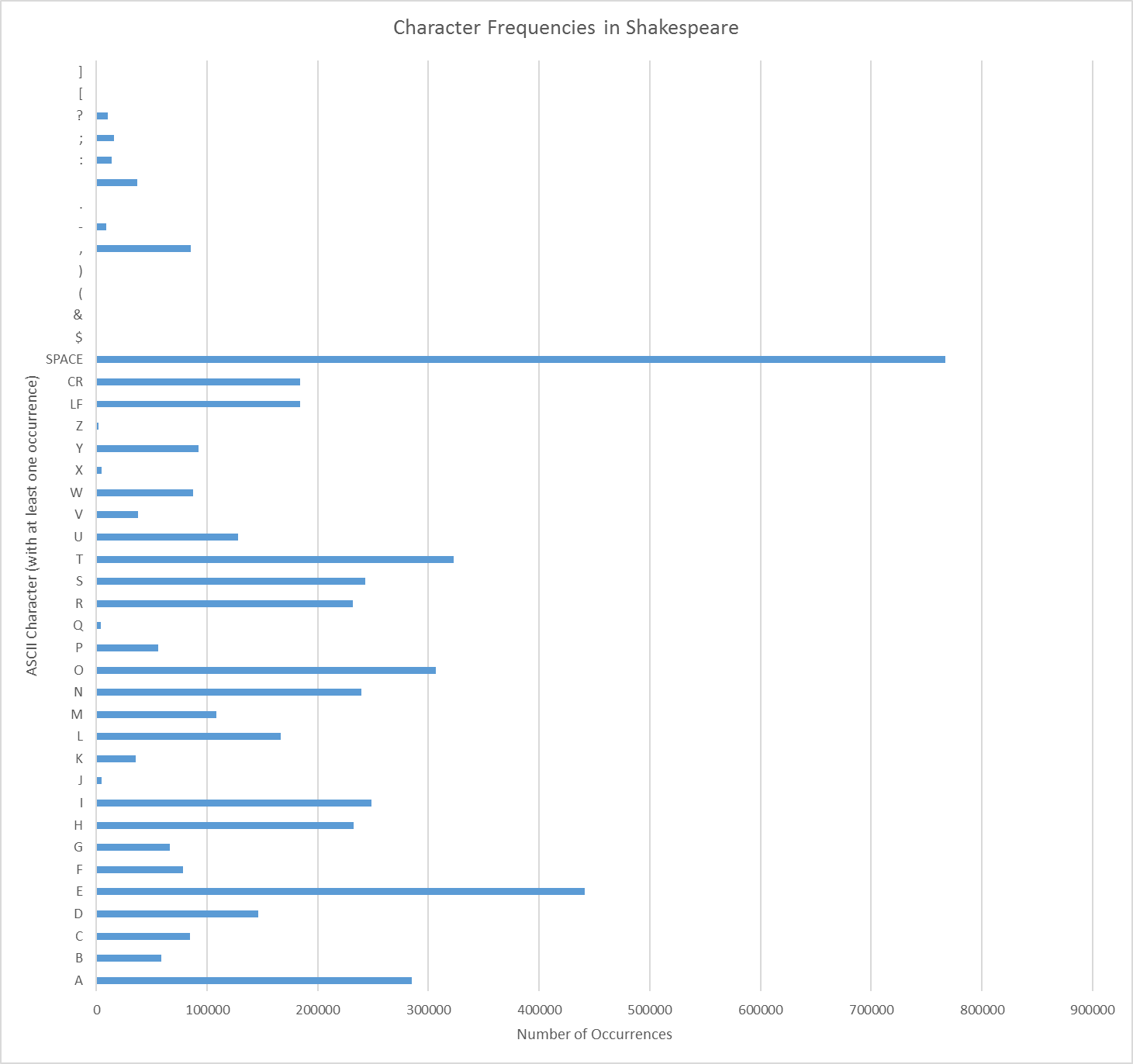


Figure 2 - A Really Tiny Graph Displaying Figure 1's Values

## 2. Character frequencies in the collective works of Shakespeare after being encrypted with DES

Figures 3 and 4 below show a data table and graph of character frequencies after encrypting the document using DES with a given password (in this case, 0x0123456789ABCDEF). As with any good algorithm, DES should be able to adequately obscure any byte-level patterns (such as character frequencies), and the data below demonstrates this. Please note that in the graph below, both the uppercase and lowercase variants of A-Z are *combined*, as they are combined in Figures 1 and 2 (to maintain consistency). Due to the combining of these characters, it appears that A-Z occurs twice as often as the other characters; however, that is not the case. Another note worth mentioning is that ASCII characters 128 to 255 (decimal) all have zero occurrences.



Figure 3 - Letter Frequencies after DES Encryption (A-Z are 2x)

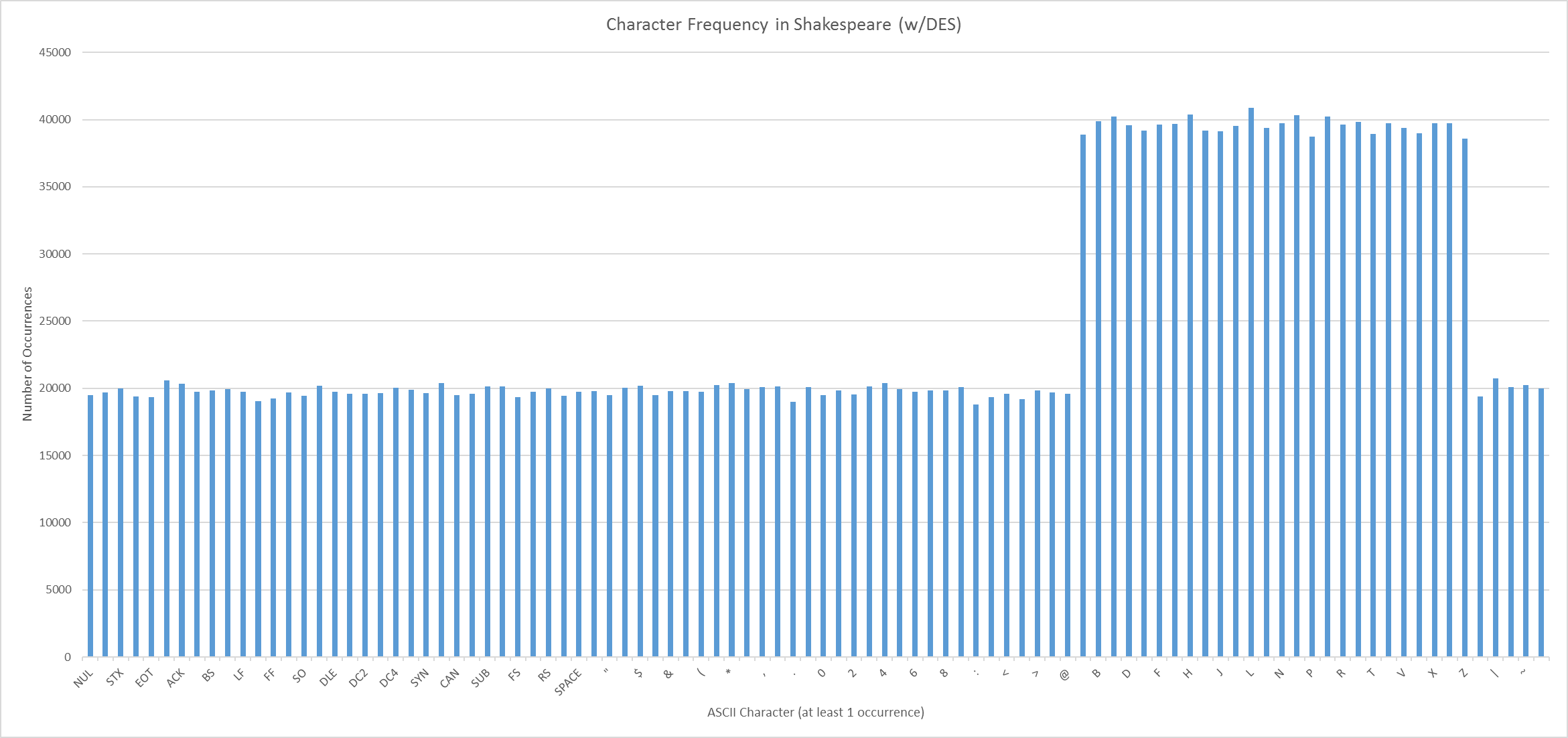


Figure 4 - A Really Tiny Graph Displaying Figure 3's Values (A-Z are 2x)

## 3. Digraph / Trigraph Analysis of the Plaintext

To see how well DES performs against obscuring digraph and trigraph patterns, a C# application was ran against the plaintext to find the most common digraph / trigraph patterns (only non-whitespace count as a digraph/trigraph pair) as a benchmark. Below are the results of the digraph / trigraph analysis (Figure 5) of the top ten digraphs and trigraphs in the plaintext.

Figure 5 - Top 10 Digraphs and Trigraphs – Plaintext

## 4. Digraph / Trigraph Analysis of the Encrypted Output

The same C# application mentioned in (3) was ran a second time, this time against the encrypted output that was analyzed in (2). As shown in Figure 6, the distribution of digraphs and trigraphs is larger, resulting in a fewer number of common digraphs and trigraphs. Due to the nature of encryption, several of the characters displayed in the digraph / trigraph table are not valid “display” ASCII characters, and therefore either have a question mark or box-like symbol. Looking at the output, we can conclude that DES does a sufficient job obscuring digraph and trigraph patterns.

Figure 6 - Top 10 Digraphs and Trigraphs - Encrypted Text