Homework 1 Cybersecurity

L’analisi di frequenza è una tecnica utilizzata nel mondo del CyberSecurity per decifrare un testo criptato con un qualche tipo di algoritmo e con una chiave sconosciuta.

Questo metodo consiste nell’analizzare il messaggio criptato e fare una analisi delle frequenze di ogni carattere di quel messaggio.  
Ad esempio nel massaggio criptato XXVRRR abbiamo 2 volte il carattere X una volta il carattere V e 3 volte R

Una volta calcolate le frequenze di ogni carattere del messaggio si passa a fare un confronto statistico con le lettere più frequenti della lingua con cui è stato scritto il messaggio originale

Infatti, con una breve ricerca su Wikipedia sono risalito alle lettere più frequenti della lingua inglese



Ogni frequenza di ogni carattere del messaggio criptato viene confrontato con la frequenza di ogni carattere della lingua inglese e trovata la frequenza che più si avvicina a quella del messaggio cifrato basta sostituire la lettera

Ad esempio, se nel messaggio cifrato la lettera R compare il 13% di volte banalmente si sostituisce R, nel messaggio cifrato, con la lettera E. Ripetendo il procedimento per ogni carattere si riesce a creare un cifrario di sostituzione per il messaggio cifrato

In questo home work l’obiettivo era decifrare il seguente messaggio utilizzando l’analisi di frequenza : PIFFMKMQI'YRKJKPQMKDJ R KJKPXAKZQXVRKNGJMXZ'KXZKTHRKVTYRRTKAQZZJKPRKJKPXAKDJZKVQDRKFJMKMQIKAQTKDIFKQZKMQ'KSJORKMQIKPXAKFXVAYJORK XO XZ'KMQIYKOJZKJGGKQLRYKTHRKNGJORKVXZAXZ'

La metodologia che ho utilizzato è quella di scrivere un codice python non troppo elaborato che implementasse l’analisi di frequenza per il messaggio (codice allegato nella consegna)

Inizialmente con il codice non ho raggiunto grandissimi risultati infatti alla fine del processo ho avuto come risultato

BUUUYEYIU'REEOEBIYEYO E EOEBNGESINSEEVPOYNS'ENSEFBEESFREEFEGISSOEBEEOEBNGEYOSESIYEEUOYEYIUEGIFEYUUEISEYI'EKOGEEYIUEBNGEUNSGROGEE NG NS'EYIUREGOSEOPPEIXEREFBEEVPOGEESNSGNS'

Dopo ulteriore analisi ho notato come la lettera K compaia molte volte; quindi, ho pensato di sostituirla invece che con la E (la più frequente nella lingua inglese) direttamente con lo spazio ottenendo migliori risultati

DLCCS SOL'UT A DOS GAT A DIR NOIMT BYASIN' IN WPT MWUTTW RONNA DT A DIR GAN MOGT CAS SOL ROW GLC ON SO' KAFT SOL DIR CIMRUAFT IFIN' SOLU FAN AYY OVTU WPT BYAFT MINRIN'