ENUNȚ PROBLEMĂ

Se dă un fișier text criptat **data.dat** (listat mai jos) și o arhivă ZIP care este parolată. Se știe că parola pentru arhiva ZIP este stocată în varianta decriptată a fișierului **data.dat**.

Se știu câteva exemple de criptare pentru câteva texte identificate anterior, și ele sunt redate mai jos (varianta necriptată în stânga, după săgeată urmează varianta criptată):

1. zer0 → 2hu3
2. Advanced → Dgydqfhg
3. dipped → glsshg

Misiunea voastră este să încercați să extrageți conținutul arhivei ZIP criptate.

DATA.DAT

zhofrphWrLWihvw5353

EXEMPLU DE REZOLVARE

Problema se poate aborda din unghiuri diferite:

1. Perspectiva cripto-aritmetică: analiza distanței Hamming (e.g. lexicografică) duce la următoarele observații:
    a. prima observație este că atât textul criptat, cât și cel decriptat, au aceeași lungime.
    b. la punctul 2, distanțaHamming(A,D) = 3, distanțaHamming(v, y) = 3 ➔ de aici observăm că există tratare specială a uppercase/lowercase (A și D vs v și y)
    c. la punctul 3, distanțaHamming(d,g) = 3 ➔ aici observăm că distanța rămâne egală între caracterele din textul criptat și cel decriptat

    De aici deja rezultă (datorită constantelor în distanța Hamming) că se pare că vorbim de un cifru Cezar, cu distanța 3 și mai rămâne să găsim lungimea.

    d. la punctul 1, distanțaHamming(z, 0) = 2, deci după z (ultimul caracter alfabetul lizibil de obicei) urmează un 2, așadar empiric deducem ordinea: …xyz0123… pentru alfabetul folosit.
    e. tot empiric putem deduce că, cel mai probabil, în alfabetul de intrare se folosesc literele până la cifra 9, după care se reia ordinea
    f. prin urmare, folosind alfabetul ab...xyz012...9 pentru cifrul Cezar, obținem DATA.DAT decriptat ca fiind: *zhofrphWrLWihvw5353 → welcomeToITfest2020*
2. Perspectiva entropică: se poate folosi entropia Huffman (frecvența simbolurilor) pentru a determina gradul de entropie al fiecăror din cele 3 texte, atât criptate cât și decriptate. Se va observa imediat că entropia este identică, ceea ce imediat conduce la ideea că, cel mai probabil, vorbim de un cifru Cezar (e.g. prin rotație). De aici, pașii sunt identici ca și mai sus, se obține distanța ca fiind 3 și se deduce în același mod lungimea alfabetului folosit.
3. Perspectiva brute force: se pot folosi tool-uri precum hashID (http://psypanda.github.io/hashID/) pentru a încerca dacă textul criptat reprezintă un hash sau nu. De asemenea, dacă se furnizează aplicația executabilă, se poate face reverse engineering

pentru a deduce care este exact algoritmul de criptare folosit. Dacă există o aplicație care furnizează datele criptate pe baza unor date necriptate, se pot deduce empiric legături relaționale între textele criptate și necriptate, pentru a încerca o decriptare pas-cu-pas.

(evident, toți pașii efectuați trebuie ilustrați cu screenshot-uri!!!)