Criptanaliza Criptosistemului S(26)

Să presupunem că plaintextul este un text peste alfabetul $\{A,B,\ldots,Z\}$ (de exemplu, un text în limba engleză în care, am păstrat literele mari, am convertit literele mici în litere mari și am eliminat toate celelalte simboluri (cifrele, spaţiile, virgulele, punctele etc.)) și că vom folosi, de exemplu, cheia π specificată în tabelul următor:

A	В	С	D	Е	F	G	Н	Ι	J	K	L	М	N	О	Р	Q	R	S	Т	U	V	W	X	Y	Z
О	P	Q	R	S	Т	U	V	W	X	Y	Z	A	В	С	D	E	F	G	Н	I	J	K	L	M	N

Criptotextul va fi obținut înlocuind peste tot în plaintext 'A' cu 'O', 'B' cu 'P', ..., 'Y' cu 'M' și 'Z' cu 'N'.

Pentru partea de criptanaliză, esențial este să remarcăm că substituțiile (fiind funcții bijective) păstrează particularitățile textului inițial.

De exemplu, analizând criptotextul interceptat, vom găsi că cel mai frecvent simbol în criptotext este 'S' cu 12,5% și vom trage concluzia că acesta corespunde celei mai frecvente litere din plaintext. Deși noi nu cunoaștem plaintextul (tocmai de acest text suntem interesați...), știm că acesta este un text în limba engleză. Conform statisticilor, în textele în limba engleză, litera 'E' este cea mai frecventă (12,702%). Astfel, este clar că $\pi(E) = S$ (adică 'E'-urile din plaintext au fost înlocuite peste tot cu litera 'S').

Raţionamentul de mai sus poate fi continuat - dacă cel mai frecvent simbol în criptotext exceptând 'S' este 'H', cu 9,1%, putem trage concluzia că $\pi(T) = H$, deoarece litera 'T' este a doua literă (fiind devansată numai de litera 'E') ca frecvenţă în textele în limba engleză (cu 9,056%).

Marea problemă este că, în textele în limba engleză, există litere foarte apropiate ca frecvenţă, cum ar fi 'C' şi 'U' (2,782%, respectiv, 2,758%) sau chiar 'R' şi 'H' (5,987%, respectiv, 6,094%). În aceste cazuri, deciziile luate numai pe baza analizei frecvenţei literelor simple pot fi greşite.

Se pot folosi statisticile existente pentru digrame (grupe de câte două litere) sau trigrame (grupe de câte trei litere). De exemplu, să presupunem că cea mai frecventă trigramă în criptotext este 'HVS'. În textele în limba engleză, cea mai frecventă trigramă este 'THE' - astfel, rezultă că $\pi(H) = V$. Mergând mai departe, putem determina $\pi(O)$ în următorul mod - în limba engleză, cele mai frecvente digrame care conțin 'T' pe prima poziție sunt 'TH' și 'TO' (în această ordine). Căutând în criptotext cele mai frecvente digrame care conțin 'H' pe prima poziție (am determinat deja că $\pi(T) = H$) vom găsi 'HV' și 'HC' - rezultă că $\pi(O) = C$.

Se pot găsi astfel de reguli pentru a determina, pas cu pas, toate componentele cheii.