1.2. THÁM MÃ CÁC HỆ MẬT MÃ CỔ ĐIỂN

PGS.TSKH. Vũ Đình Hòa

- Có nhiều kỹ thuật giải mã sử dụng các tính chất thống kê của ngôn ng tiếng Anh.
  - ***Giải mã hệ mã Affine***
  - ***Giải mã hệ mã thay thế***

- ***Xác suất xuất hiện của 26 ch cái:*** (theo Beker và Piper thống kê từ nhiều tiểu thuyết, tạp chí và báo)

Kí tự	A	B	C	D	E	F	G	H	I
Xác xuất	.082	.015	.028	.043	.127	.022	.020	.061	.070
kí tự	J	K	L	M	N	O	P	Q	R
Xác xuất	.002	.008	.040	.024	.067	.075	.019	.001	.060
Kí tự	S	T	U	V	W	X	Y	Z	
Xác xuất	.063	.091	.028	.010	.023	.001	.020	.001	

Beker và Piper phân 26 ch cái thành 5 nhóm:

- E: có xác suất khoảng 0.120
- T, A, O, I, N, S, H, R: có xac suất khoảng 0.06 đến 0.09
- D, L : có xác suất chừng 0.04
- C, U, M, W, F, G, Y, P, B: có xác suất khoảng 0.015 đến 0.023
- V, K, J, X, Q, Z mỗi ký tự có xác suất nhỏ hơn 0.01

- 30 bộ đôi thông dụng nhất (theo thứ tự giảm dần) là: TH, HE, IN, ER, AN, RE, ED, ON, ES, ST, EN, AT, TO, NT, HA, ND, OU, EA, NG, AS, OR, TI, IS, ET, IT, AR, TE, SE, HI và OF.
- 12 bộ ba thông dụng nhất (theo thứ tự giảm dần) là: THE, ING, AND, HER, ERE, ENT, THA, NTH, WAS, ETH, FOR và DTH.

1.2.1 Giải mã hệ mã Affine

- Mật mã Affine là một ví dụ đơn giản cho ta thấy cách giải mã nhờ dùng các số liệu thống kê.

- Bản mã nhận đợc từ mã Affine:

FMXVEDRAPHFERBNDKRXRSREFMORUDSDK
DVSHVUFEDKPKDLYEVLRHHRH

- Tần xuất xuất hiện của các ch cái trong bản mã.

kí tự	A	B	C	D	E	F	G	H	I	J	K	L	M
tần xuất	1	1	0	7	5	4	0	5	0	0	5	2	2
kí tự	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
tần xuất	1	1	2	0	8	3	0	2	4	0	2	1	0

- Các ký tự có tần suất cao nhất trong bản mã là: R (8), D (7), E, H, K (5) và F, S, V (4).
- Phỏng đoán ban đầu: Giả thiết R là ký tự mã của e và D là kí tự mã của t (e và t là 2 ch cái thông dụng nhất).

$\Rightarrow$ $e_K(4) = 17$ và $e_K(19) = 3$.

Giải hệ $4a + b = 17$

$19a + b = 3$

được $a = 6$, $b = 19$ (trong Z_{26}) không hợp lệ do $(6, 26) = 2$

- Phỏng đoán tiếp theo: R là ký tự mã của e và E là mã của t. $\Rightarrow e_K(4) = 17$ và $e_K(19) = 4$.

Giải hệ $\begin{matrix} 4a+b=17. \\ 19a+b=4 \end{matrix}$ được a = 13 . Loại

- Phỏng đoán: R là mã hoá của e và H là mã hoá của t. $\Rightarrow$ $e_K(4) = 17$ và $e_K(19) = 7$. được a = 8 (loại).
- Giả sử rằng R là mã hoá của e và K là mã hoá của t. Theo giả thiết này ta thu được a = 3 và b = 5 là khóa hợp lệ.

- Tính toán hàm giải mã ứng với K = (3,5) và giải mã bản mã
- Ta có d_K (y) = 9y - 19 và giải mã bản mã đã cho, ta đợc:

  algorithmsarequitegeneraldefinitionsof arithmeticprocesses

- Nh vậy khoá xác định trên là khoá đúng.

1.2.2 Giải mã hệ mật mã thay thế

- Bản mã nhận đợc từ hệ mật mã thay thế là:

YIFQFMZRWQFYVECFMDZPCVMRZWNMDZVEJBTXCDDUMJ
NDIFEFMDZCDMQZKCEYFCJMYRNCWJCSZREXCHZUNMXZ
NZUCDRJXYYSMTMEYIFZWDYVZVYFZUMRZCRWNZDZJT
XZWGCHSMRNMDHNCMFQCHZJMXJZWIEJYUCFWDINZDIR

kí tự	A	B	C	D	E	F	G	H	I	J	K	L	M
tần xuất	0	1	15	13	7	11	1	4	5	11	1	0	16
kí tự	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
tần xuất	9	0	1	4	10	3	2	5	5	8	6	10	20

- Z xuất hiện nhiều hơn so với bất kỳ một ký tự nào khác trong bản mã nên có thể phỏng đoán $d_K(Z) = e$.
- Các ký tự còn lại xuất hiện ít nhất 10 lần (mỗi ký tự) là C, D, F, J, R, M, Y. Ta hy vọng rằng, các ký tự này là mã khoá của t, a, c, o, i, n, s, h, r, tuy nhiên sự khác biệt về tần suất không đủ cho ta có đợc sự phỏng đoán thích hợp.

- Xem xét các bộ đôi. Các bộ đôi thờng gặp nhất:
  - DZ và ZW (4 lần mỗi bộ);
  - NZ và ZU (3 lần mỗi bộ);
  - RZ, HZ, XZ, FZ, ZR, ZV, ZC, ZD và ZJ (2 lần mỗi bộ)
- Vi ZW xuất hiện 4 lần còn WZ không xuất hiện lần nào và W xuất hiện ít hơn so với nhiều ký tự khác, nên có thể phỏng đoán là $d_K(W) = d$.
- Vi DZ xuất hiện 4 lần và ZD xuất hiện 2 lần nên ta có thể $d_K(D) \in \{r,s,t\}$, tuy nhiên vẫn còn cha rõ là ký tự nào trong 3 ký tự này là ký tự đúng.

- Từ giả thiết $d_K(Z) = e$ và $d_K(W) = d$ mà RW xuất hiện 1 lần, vì R thờng xuất hiện trong bản mã và nd là một bộ đôi thờng gặp nên ta nên thử $d_K(R) = n$.
- Tiếp theo thử $d_K(N) = h$ vì NZ (he) là một bộ đôi thờng gặp còn ZN (eh) không xuất hiện.
- Từ đó $d_K(C) = a$

- Xét M là ký tự thờng gặp nhất sau Z.
- Đoạn bản mã RNM sẽ giải mã thành nh- gợi ý h- sẽ bắt đầu một từ, bởi vậy M sẽ biểu thị môt nguyên âm. Phỏng đoán $d_K(M)$ = i hoặc o (vì đã có $d_K(Z)$=e, $d_K(C)$=a). Vì ai là bộ đôi thờng gặp hơn ao nên từ bộ đôi CM trong bản mã thử $d_K(M)$ = i.

- Vì o là một ch thờng gặp nên giả định ch cái tơng ứng trong bản mã là một trong các ký tự D,F,J,Y. Y thích hợp nhất, nếu không ta sẽ có các xâu dài các nguyên âm, chủ yếu là aoi (từ CFM hoặc CJM). Bởi vậy giả thiết $d_K(Y) = o$.
- Ba ký tự thờng gặp nhất còn lại trong bản mã là D,F,J, ta phán đoán sẽ giải mã thành r,s,t theo thứ tự nào đó. Hai lần xuất hiện của bộ ba NMD gợi ý rằng $d_K(D) = s$ ứng với bộ ba his trong bản rõ (phù hợp với giả định trớc là $d_K(D) \in \{r,s,t\}$).
- Đoạn HNCMF có thể là bản mã của chair, điều này sẽ cho $d_K(F) = r$ (và $d_K(H) = c$) và bởi vậy (bằng cách loại trừ) sẽ có $d_K(J) = t$.

- Bản rõ:

  Our friend from Pais examined his empty glass with surprise, as if evaporation had taken place while he wasn't looking. I poured some more wine and he settled back in his chair, face tilted up towards the sun.