

3. Teknik Analisis Frekuensi

Kelemahan sandi substitusi adalah hubungan frekuensi kemunculan huruf pada plainteks dan cipherteks yang tidak berubah. Jika huruf *a* dienkripsi menjadi huruf *X*, dan *a* muncul sebanyak *n* kali, maka *X* juga muncul sebanyak *n* kali. Hal ini dimanfaatkan penyerang untuk menemukan kuncinya. Penyerang memanfaatkan data statistik peluang kemunculan huruf berikut ini. Teknik seperti ini disebut dengan **teknik analisis frekuensi**.

Tabel 1. Peluang kemunculan huruf (Bahasa Inggris)

Huruf	Peluang	Huruf	Peluang
A	0.082	N	0.067
B	0.015	O	0.075
C	0.028	P	0.019
D	0.043	Q	0.001
E	0.127	R	0.060
F	0.022	S	0.063
G	0.020	T	0.091
H	0.061	U	0.028
I	0.070	V	0.010
J	0.002	W	0.023
K	0.008	X	0.001
L	0.040	Y	0.020
M	0.024	Z	0.001

Data di atas dapat dibagi menjadi lima kelompok, yaitu:

1. *E*, mempunyai peluang 0.120.
2. *T, A, O, I, N, S, H, R*, mempunyai peluang antara 0.06 dan 0.09.
3. *D, L*, mempunyai peluang sekitar 0.04.
4. *C, U, M, W, F, G, Y, P, B*, mempunyai peluang antara 0.015 dan 0.023.
5. *V, K, J, X, Q, Z*, peluangnya kurang dari 0.01.

Selain itu juga ditemukan data mengenai frekuensi kemunculan *digrams* dan *trigrams*. 30 digram paling sering muncul dalam Bahasa Inggris berturut-turut dari yang peluang terbesar adalah:

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, OF.

Sedangkan 12 trigram yang sering muncul dalam Bahasa Inggris berturut-turut dari peluang terbesar adalah:

THE, ING, AND, HER, ERE, ENT, THA, NTH, WAS, ETH, FOR, DTH.

Untuk lebih jelasnya mengenai cara kerja teknik analisis frekuensi dalam memecahkan cipher substitusi, perhatikan contoh di bawah ini.

Contoh 1.

Berikut ini diberikan sebuah cipherteks yang diperoleh dari sandi substitusi dengan plainteks dalam Bahasa Inggris. Coba Anda pecahkan cipherteks berikut ini menggunakan teknik analisis frekuensi.

Y I F Q F M Z R W Q F Y V E C F M D Z P C V M R Z W N M D Z V E J B T X
 C D D U M J N D I F E F M D Z C D M Q Z K C E Y F C J M Y R N C W J C S
 Z R E X C H Z U N M X Z N Z U C D R J X Y Y S M R T M E Y I F Z W D Y V
 Z V Y F Z U M R Z C R W N Z D Z J J X Z W G C H S M R N M D H N C M F Q
 C H Z J M X J Z W I E J Y U C F W D J N Z D I R

Hasil analisis frekuensi kemunculan huruf pada cipherteks di atas dapat dilihat pada Tabel 2 berikut ini.

Tabel 2. Frekuensi huruf cipherteks Contoh 1.

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

Karena *Z* paling sering muncul, dapat dicoba $d_K(Z) = e$. Huruf lainnya yang muncul lebih dari 10 kali adalah *C, D, F, J, M, R, Y*, maka kemungkinan huruf-huruf tersebut plainteksnya *t, a, o, i, n, s, h, r*. Akan tetapi hal ini tidak selalu menjamin korespondensi yang sama.

Digram yang sering muncul pada cipherteks adalah DZ dan ZW (4 kali), NZ dan ZU (3 kali), dan RZ , HZ , XZ , FZ , ZR , ZV , ZC , ZD , dan ZU (2 kali).

Perhatikan bentuk $-Z$ atau $Z-$, dan asumsi bahwa Z didekripsi menjadi e . Karena ZW muncul 4 kali dan WZ tidak muncul sama sekali, maka dapat dicoba bahwa $d_K(W) = d$. Karena DZ muncul 4 kali dan ZD muncul 2 kali, maka dapat dicoba bahwa $d_K(D) \in \{r, s, t\}$, tetapi tidak menutup kemungkinan bahwa D didekripsi selain dari r , s , atau t .

Digunakan asumsi semula bahwa $d_K(Z) = e$ dan $d_K(W) = d$, kita perhatikan kembali cipherteks dan diketahui bahwa ZRW dan RZW keduanya terletak pada awal cipherteks, dan RW muncul kembali setelahnya. Karena R juga sering muncul pada cipherteks dan nd adalah digram yang juga sering muncul, maka dapat dicoba bahwa $d_K(R) = n$. Dari sini dapat diperoleh:

```

-----end-----e-----ned---e-----
YIFQFMZRWQFYVECFMDZPCVMRZWNMDZVEJBTX

-----e-----e-----n--d---
CDDUMJNDIFEFMZCDMQZKCEYFCJMYRNCWJCS

en---e---e---n-----n-----ed---
ZREXCHZUNMXZNZUCDRJXYYSMRTMEYIFZWDYV

e---e--ne-nd-e-e---ed-----n-----
ZVYFZUMRZCRWNZDZJJXZGWCHSMRNMDHNCMFQ

--e---ed-----d---e--n
CHZJMXJZWIEJYUCFWDJNZDIR
    
```

Langkah berikutnya adalah dengan mencoba $d_K(N) = h$, sebab NZ merupakan digram yang sering muncul dan ZN tidak. Jika hal ini benar, maka segmen dari plainteks $ne - ndhe$ menunjukkan bahwa $d_K(C) = a$. Sehingga dari sini diperoleh:

```

-----end-----a---e-a--nedh--e-----
YIFQFMZRWQFYVECFMDZPCVMRZWNMDZVEJBTX

a-----h-----ea---e-a---a---nhad-a-
CDDUMJNDIFEFMZCDMQZKCEYFCJMYRNCWJCS

en--a-e-h--eh--a-n-----n-----ed---
ZREXCHZUNMXZNZUCDRJXYYSMRTMEYIFZWDYV

e---e--neandhe-e---ed-a---nh---ha---
ZVYFZUMRZCRWNZDZJJXZGWCHSMRNMDHNCMFQ

a-e---ed-----a-d--he--n
CHZJMXJZWIEJYUCFWDJNZDIR
    
```

Selanjutnya, dicari huruf yang dienkrpsi menjadi o . Karena o umum digunakan, maka o dienkrpsi menjadi D , F , J , Y . Huruf Y kelihatan mempunyai peluang terbesar, sebab dapat ditemukan kata dalam plainteks dengan huruf vokal aoi dari CFM atau CJM . Oleh karena itu, dapat dicoba $d_K(Y) = o$.

Tiga huruf berikutnya yang sering muncul adalah D , F , J , yang mungkin didekripsi menjadi r , s , t dengan urutan yang sama. Dua kemunculan trigram NMD menunjukkan bahwa $d_K(D) = s$. Dengan mengambil trigram his pada plainteks (dengan tetap konsisten pada hipotesis percobaan sebelumnya bahwa $d_K(D) \in \{r, s, t\}$). Segmen $HNCMF$ kemungkinan merupakan hasil enkripsi dari $chair$, yang mengakibatkan $d_K(F) = r$ dan $d_K(H) = c$, sehingga dapat diperoleh $d_K(J) = t$. Dari sini diperoleh:

```

o-r-riend-ro--arise-a-inedhise--t---
YIFQFMZRWQFYVECFMDZPCVMRZWNMDZVEJBTX

ass-iths-r-riseasi-e-a-orationhad-a-
CDDUMJNDIFEFMZCDMQZKCEYFCJMYRNCWJCS

en--ace-hi-ehe-asnt-oo-in-i-o-redso-
ZREXCHZUNMXZNZUCDRJXYYSMRTMEYIFZWDYV

e-ore-ineandhesett-ed-ac-inhischair-
ZVYFZUMRZCRWNZDZJJXZGWCHSMRNMDHNCMFQ

aceti-ted--to-ardsthes-n
CHZJMXJZWIEJYUCFWDJNZDIR
    
```

Dari hasil terakhir di atas sudah cukup mudah bagi Anda untuk mencoba melanjutkannya sendiri.

4. Kesimpulan dan Saran

Walaupun mempunyai jumlah kemungkinan kunci yang sangat besar, yaitu $26! \approx 4 \times 10^{26}$, tetapi sandi substitusi mempunyai kelemahan yang cukup fatal, yaitu hubungan frekuensi kemunculan huruf antara plainteks dengan cipherteks yang tidak bisa dihilangkan.

Untuk mempersulit penyerang, sebaiknya digunakan bahasa yang tidak lazim digunakan, sehingga penyerang akan kesulitan dalam memecahkan cipherteks. Sedangkan untuk pihak penyerang harus mempunyai data-data frekuensi kemunculan huruf pada berbagai bahasa, termasuk juga bahasa-bahasa yang jarang digunakan, seperti bahasa daerah, bahasa kuno, dan sebagainya.

Daftar Pustaka

- Churchhouse, Robert, 2001, *Codes and Ciphers, Julius Caesar, the Enigma, and the Internet*, Cambridge University Press, UK.
- Stinson, D.R., 2006, *Cryptography: Theory and Practice Third Edition*, Capman and Hall/CRC Press, Boca Raton, Florida.

(1) Soal Latihan Pemecahan Sandi Substitusi:

Seorang sandiman berhasil memperoleh pesan tersandi dari seseorang musuh asing yang isinya:

EMGLOSUDCGDNCUSWYSFHNSFCYKDPUMLWGYICO
 XYSIPJCKQPKUGKMGOLICGINCGACKSNISACYKZ
 SCKXECJCKSHYSXCGOIDPKZCNKSHICGIWYGKK
 KGOLDSILKGOIUSIGLEDSPWZUGFZCCNDGYYSFU
 SZCNXEOJNCGYEOWEUPXEZGACGNFGLKNSACIGO
 IYCKXCJUCIUZCFZCCNDGYYSFEUEKUZCSOCFZC
 CNCIACZEJNCSHFZEJZEGMXCYHCJUMGKUCY

Dia berhasil mengetahui bahwa bahasa yang digunakan adalah Bahasa Inggris, algoritma yang digunakan adalah sandi substitusi, dan w dienkrip menjadi F . Berikut ini diberikan hasil analisis frekuensinya. Coba Anda temukan pesan yang sebenarnya.

Tabel 3. Hasil analisis frekuensi cipherteks (1)

C	37	CG	7	CCN	3
G	24	ZC	7	FZC	3
S	20	AC	5	GOI	3
K	18	CK	5	YSF	3
I	15	CN	5	ZCC	3
Y	15	GO	5	CFZ	2
U	14	YS	5	CGI	2
N	13	CY	4	CJU	2
Z	13	FZ	4	CKS	2
E	12	GK	4	CKX	2
O	10	GY	4	CND	2
F	9	NC	4	CYK	2
D	8	SF	4	DGY	2
J	7	CC	3	GAC	2
L	7	CI	3	GOL	2
X	7	CJ	3	GYI	2
P	6	GL	3	ICG	2
A	5	IC	3	JCK	2
H	5	KS	3	JNC	2
M	5	KU	3	KSH	2
W	5	MG	3	NDG	2
Q	1	OI	3	UZC	2
B	0	SH	3	YYS	2
R	0	SI	3	ZCN	2
T	0	US	3	ZEJ	2
V	0	XC	3	ACG	1
		XE	3	ACI	1
		CF	2	ACK	1
		CS	2	ACY	1
		DG	2	ACZ	1

LEMBAR KERJA 1
 Pemecahan Sandi Rahasia

EMGLOSUDCGDNCUSWYSFHNS
 FCYKDPUMLWGYICOXYSIPJC
 KQPKUGKMGOLICGINCGACKS
 NISACYKZSCKXECJCKSHYSX
 CGOIDPKZCNKSHICGIWYGKK
 KGOLDSILKGOIUSIGLEDSP
 WZUGFZCCNDGYYSFUSZCNXE
 OJNCGYEOWEUPXEZGACGNFG
 LKNSACIGOIYCKXCJUCIUZC
 FZCCNDGYYSFEUEKUZCSOCF
 ZCCNCIACZEJNCSHFZEJZEG
 MXCYHCJUMGKUCY

(2) Soal Latihan Pemecahan Sandi Substitusi:

Diberikan cipherteks dengan plainteks dalam Bahasa Inggris yang dienkrif menggunakan cipher substitusi.

MJZYBLGESECNCMQYGXYSPLYZDZPMYGIIRLL
 CPAYCKYKGWZMCWZKYFRMZVVCXXZLZPMYX
 LGWYTJSMYGPZYWCAJMYCWSACPZYXGLYZHS
 WBNZYXZTYTGRNVYMJCPOYMJSMYCXYMJZLZ
 YSLZYMTZPMQYMJLZZYBZGBNZYCPYSYLGGW
 YMJZPYMJZLZYCKYSPYZDZPKYIJSPIZYMJS
 MYMJZLZYSLSZYMTGYGXYMJZWYTCMJYMJZYK
 SWZYECLMJVSQYERMYMJKYCKYKG

Tabel 4. Hasil analisis frekuensi cipherteks (2)

Y	49	MJ	14	YMJ	10
Z	33	YM	12	MJZ	5
M	27	ZY	10	CKY	3
C	18	MY	6	JZL	3
J	17	JZ	5	GXY	2
G	14	YC	5	MJC	2
L	14	CK	4	MYC	2
S	14	KY	4	MYG	2
P	13	LZ	4	PYZ	2
K	9	YS	4	SLZ	2
W	9				
X	8				
T	6				
B	4				
E	4				
I	4				
N	4				
R	4				
A	3				
Q	3				
V	3				
D	2				
F	1				
H	1				
O	1				
U	0				

LEMBAR KERJA 2
 Pemecahan Sandi Rahasia

MJZYBLGESECNCMQYGXYSPLYZ
 DZPMYGIIRLLCPAYCKYKGWZM
 CWZKYFRMZVVCXXZLZPMYXL
 GWYTJSMYGPZYWCAJMYCWSAC
 PZYXGLYZHSWBNZYXZTYTGRN
 VYMJCPOYMJSMYCXYMJZLZYS
 LZYMTZPMQYMJLZZYBZGBNZY
 CPYSYLGGWYMJZPYMJZLZYCK
 YSPYZDZPKYIJSPIZYMJSMYM
 JZLZYSLSZYMTGYGXYMJZWYTC
 MJYMJZYKSWZYECLMJVSQYER
 MYMJCKYCKYKG