Bahan kuliah IF4020 Kriptografi

Kriptanalisis Sederhana
(Bagian 1)

Oleh: Dr. Rinaldi Munir

Prodi Informatika

Sekolah Teknik Elektro dan Informatika

Kriptanalisis pada *Cipher* Abjad-Tunggal

- Jumlah kemungkinan kunci = 26!
- Tidak dapat menyembunyikan hubungan antara plainteks dengan cipherteks.
- Huruf yang sama dienkripsi menjadi huruf cipherteks yang sama
- Huruf yang sering muncul di dalam palinteks, sering muncul pula di dalam cipherteksnya.

- Oleh karena itu, cipherteks dapat didekripsi tanpa mengetahui kunci (*ciphertext-only attack*)

- Metode yang digunakan:
  1. Terkaan
  2. Statistik (analisis frekuensi)

- Informasi yang dibutuhkan:
  1. Mengetahui bahasa yang digunakan untuk plainteks
  2. Konteks plainteks

Metode Terkaan

Asumsi: - bahasa plainteks adalah B. Inggris

Tujuan: mereduksi jumlah kunci

Contoh 1. Cipherteks: `G WR W RWL`

Plainteks: `i am a ma*`

`i am a man`

Jumlah kunci berkurang dari 26! menjadi 22!

Contoh 2.

Cipherteks: `HKC`

Plainteks:

- lebih sukar ditentukan,
- tetapi tidak mungkin

`z` diganti dengan `H`,

`q` dengan `K`,

`k` dengan `C`,

karena tidak ada kata “zqk” dalam Bahasa Inggris

Contoh 3.

Cipherteks: `HATTPT`

Plainteks: salah satu dari `T` atau `P` merepresentasikan huruf vokal, misal

```
cheese
misses
cannon
```

Contoh 4.

Cipherteks: `HATTPT`

Plainteks: diketahui informasi bahwa pesan tersebut adalah nama negara.

→ `greece`

- Proses menerka dapat menjadi lebih sulit jika cipherteks dikelompokkan ke dalam blok-blok huruf.
- Contoh:

```
CTBMN BYCTC BTJDS QXBNS GSTJC BTSWX CTQTZ CQVUJ QJSGS TJQZZ MNQJS
VLNSX VSZJU JDSTS JQUUS JUBXJ DSKSU JSNTK BGAQJ ZBGYQ TLCTZ BNYBN
QJSW
```

- Jika diberikan informasi bahwa cipherteks tersebut berasal dari perusahaan yang bergerak di bidang keuangan, maka proses menerka dapat lebih mudah

- Kata keuangan dalam Bahasa Inggris adalah `financial`

- Di dalam kata `financial` ada dua buah huruf `i` yang berulang, dengan empat buah huruf lain di antara keduanya (`nanc`) → `inanci`

- Cari huruf berulang dengan pola seperti itu di dalam cipherteks (tidak termasuk spasi). Ditemukan pada posisi 6, 15, 27, 31, 42, 48, 58, 66, 70, 71, 76, dan 82

```
      6          15                  27    31                42      58
CTBMN BYCTC BTJDS QXBNS GSTJC BTSWX CTQTZ CQVUJ QJSGS TJQZZ
MNQJS VLNSX VSZJU JDSTS JQUUS JUBXJ DSKSU JSNTK BGAQJ ZBGYQ
TLCTZ BNYBN  QJSW
```

- Hanya dua diantaranya, yaitu 31 dan 42 yang mempunyai huruf berikutnya yang berulang (berkoresponden dengan `n`) → `inanci`

- Dan dari keduanya hanya pada posisi 31 huruf `A` berada pada posisi yang tepat

- Jadi ditemukan `financial` pada posisi 30, yaitu untuk kriptogram `XCTQTZCQV`

```
CTBMN BYCTC BTJDS QXBNS GSTJC BTSWX CTQTZ CQVUJ QJSGS
TJQZZ MNQJS VLNSX VSZJU JDSTS JQUUS JUBXJ DSKSU JSNTK
BGAQJ ZBGYQ TLCTZ BNYBN  QJSW
```

- Diperoleh pemetaan:

X	→	f	C	→	i
T	→	n	Q	→	a
Z	→	c	V	→	l

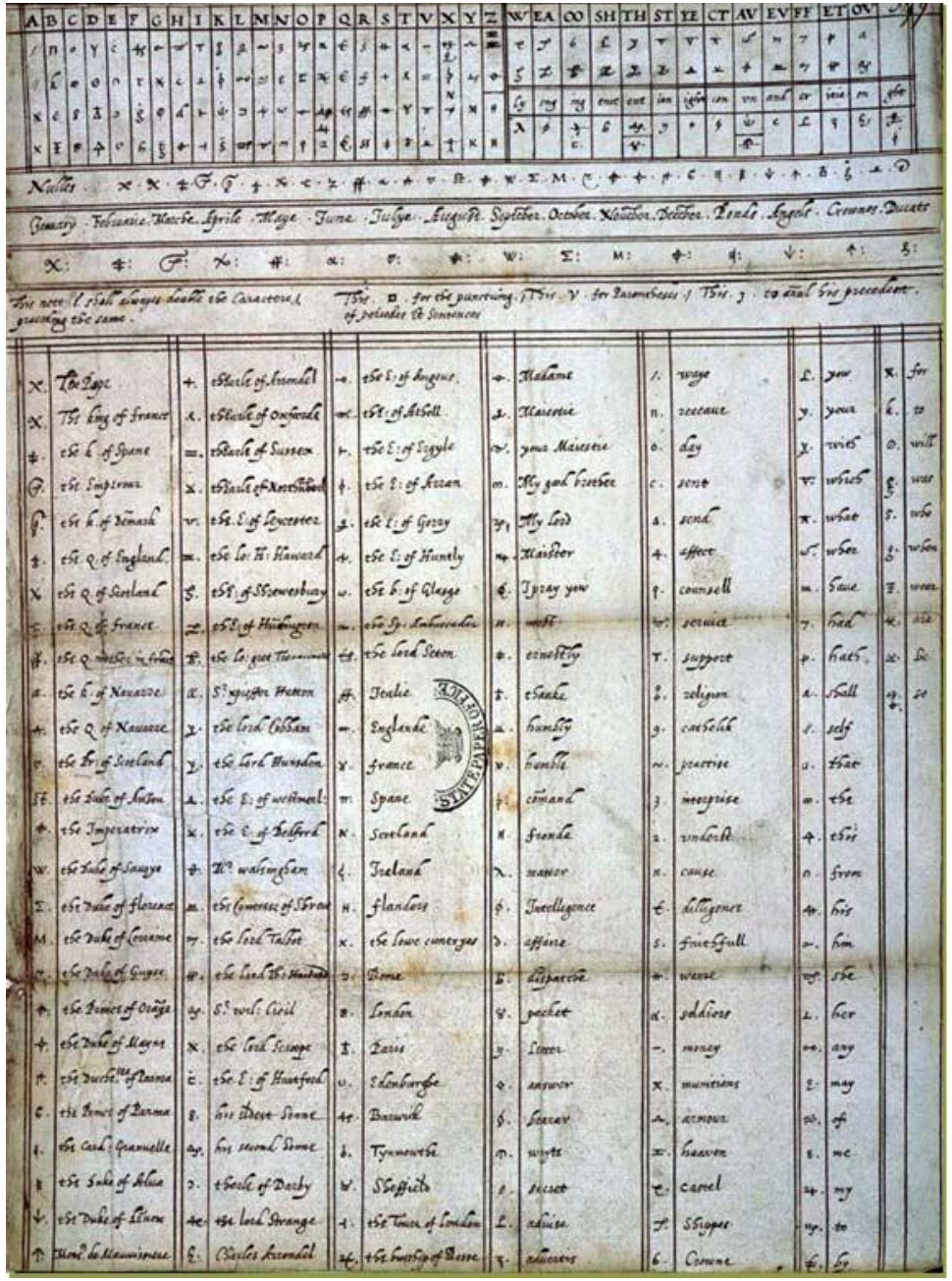
- Ganti semua huruf X, C, T, Q, Z, V dengan f, i, n, a, c, l:

```
CTBMN BYCTC BTJDS QXBNS GSTJC BTSWX CTQTZ CQVUJ
QJSGS TJQZZ MNQJS VLNSX VSZJU JDSTS JQUUS JUBXJ
DSKSU JSNTK BGAQJ ZBGYQ TLCTZ BNYBN QJSW

inBMN BYini BnJDS cfBNS GSnJi BnSWf inanc ialUJ
aJSGS nJacc MNaJS VLNSf VScJU JDSnS JaUUS JUBfJ
DSKSU JSNnK BGAaJ cBGYa nLinc BNYBN aJSW
```

- Jumlah kunci berkurang menjadi 20! Deduksi dapat diteruskan.

- Peristiwa yang menimpa Queen Mary of Scotland pada abad 18 karena menggunakan *cipher* abjad-tunggal yang mudah diterka ➔ mudah dipecahkan.

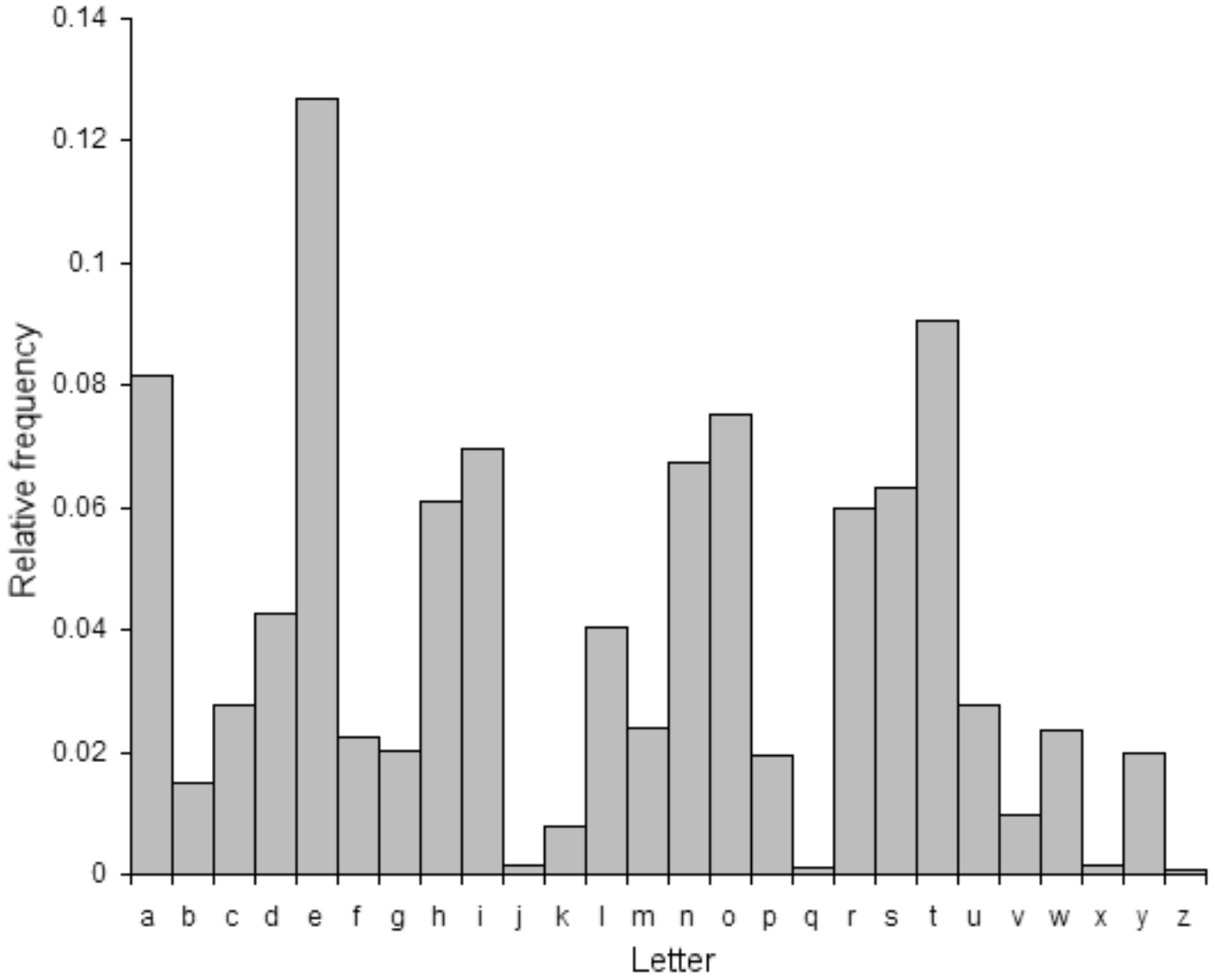


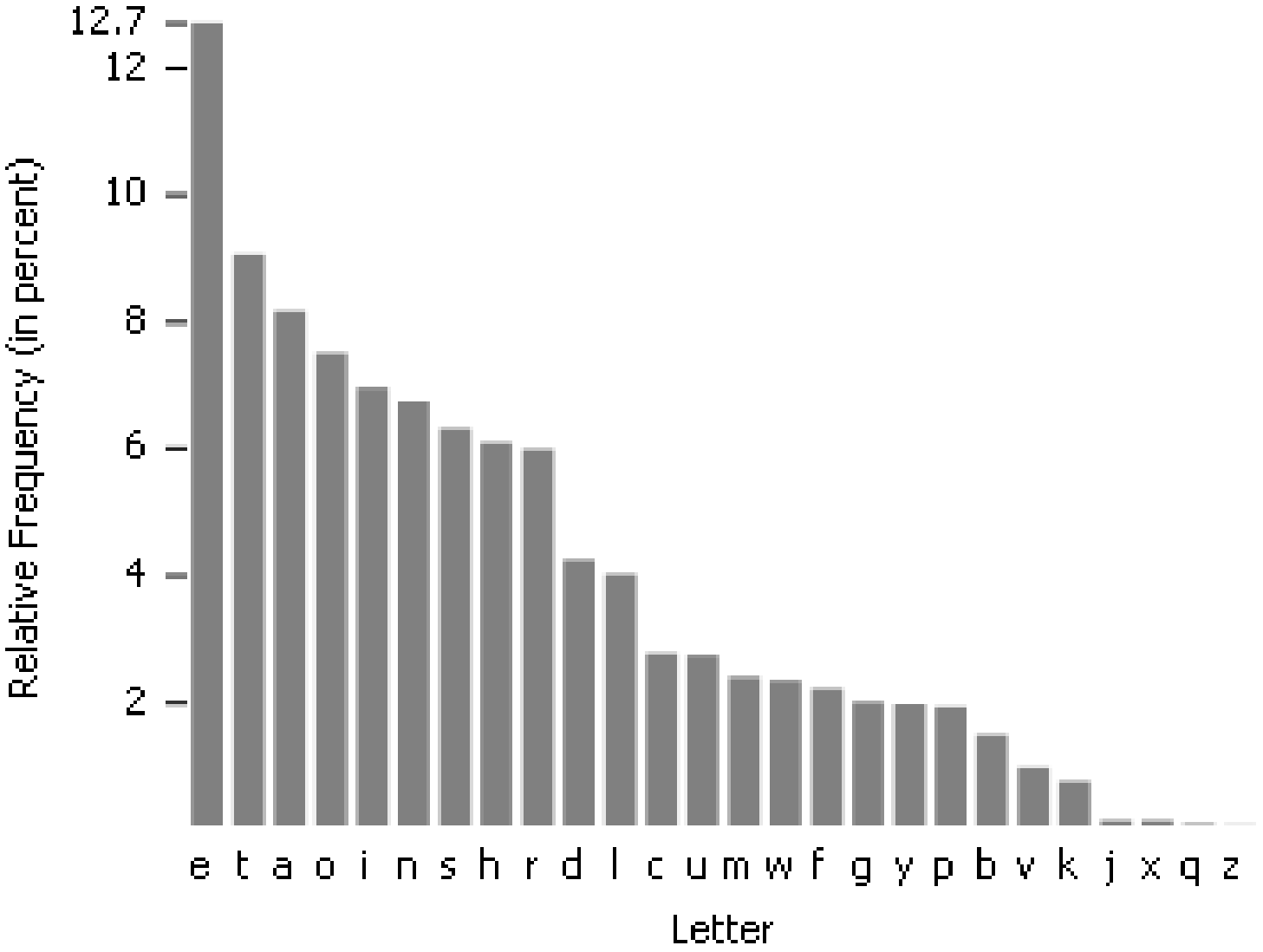
Cipher yang digunakan oleh *Mary Queen of Scott.*

Metode Analisis Frekuensi

Tabel 2. Frekuensi kemunculan (relatif) huruf-huruf dalam teks Bahasa Inggris (sampel mencapai 300.000 karakter di dalam sejumlah novel dan suratkabar

Huruf	%	Huruf	%
A	8,2	N	6,7
B	1,5	O	7,5
C	2,8	P	1,9
D	4,2	Q	0,1
E	12,7	R	6,0
F	2,2	S	6,3
G	2,0	T	9,0
H	6,1	U	2,8
I	7,0	V	1,0
J	0,1	W	2,4
K	0,8	X	2,0
L	4,0	Y	0,1
M	2,4	Z	0,1


0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
Relative frequency
a b c d e f g h i j k l m n o p q r s t u v w x y z
Letter


12.7
12
10
8
6
4
2
Relative Frequency (in percent)
e t a o i n s h r d l c u m w f g y p b v k j x q z
Letter

- *Top* 10 huruf yang sering muncul dalam teks Bahasa Inggris: E, T, A, O, I, N, S, H, R, D, L, U

- Top 10 huruf *bigram* yang sering muncul dalam teks B. Inggris: TH, HE, IN, EN, NT, RE, ER, AN, TI, dan ES

- Top 10 huruf *trigram* yang sering muncul dalam teks B. Inggris: THE, AND, THA, ENT, ING, ION, TIO, FOR, NDE, dan HAS

- Top 10 huruf yang paling sering muncul dalam Bahasa Indonesia:

Huruf	Peluang (%)
A	17,50
N	10,30
I	8,70
E	7,50
K	5,65
T	5,10
R	4,60
D	4,50
S	4,50
M	4,50

- Kriptanalis menggunakan tabel frekuensi kemunculan huruf dalam B. Inggris sebagai kakas bantu melakukan dekripsi.

- Kemunculan huruf-huruf di dalam sembarang plainteks tercermin pada tabel tersebut.

- Misalnya, jika huruf “R” paling sering muncul di dalam cipherteks, maka kemungkinan besar itu adalah huruf “E” di dalam plainteksnya.

Langkah-langkah kriptanalisis dengan metode analisis frekuensi adalah sebagai berikut:

1. Hitung frekuensi kemunculan relatif huruf-huruf di dalam cipherteks.

2. Bandingkan hasil langkah 1 dengan Tabel 2. Catatlah bahwa huruf yang paling sering muncul dalam teks Bahasa Inggris adalah huruf E. Jadi, huruf yang paling sering muncul di dalam cipherteks kemungkinan besar adalah huruf E di dalam plainteksnya.

3. Langkah 2 diulangi untuk huruf dengan frekeuensi terbanyak berikutnya.

- Contoh: Diberikan cipherteks berikut ini (Stalling, 2011):

```
UZ QSO VUOHXMOPV GPOZPEVSG ZWSZ OPFPESX UDBMETSX AIZ
VUEPHZ HMDZSHZO WSFP APPD TSVP QUZW YMXUZUHSX
EPYEPOPDZSZUFPO MB ZWP FUPZ HMDJ UD TMOHMQ
```

Lakukakan kriptanalisis dengan metode analisis frekuensi untuk memperoleh plainteks. Asumsi: bahasa yang digunakan adalah Bahasa Inggris dan *cipher* yang digunakan adalah *cipher* abjad-tunggal.

Hitung frekuensi kemunculan huruf di dalam cipherteks tersebut:

Huruf	%	Huruf	%
P	13,33	Q	2,50
Z	11.67	T	2,50
S	8,33	A	1,67
U	8,33	B	1,67
O	7,50	G	1,67
M	6.67	Y	1,67
H	5,83	I	0,83
D	5,00	J	0,83
E	5,00	C	0,00
V	4,17	K	0,00
X	4,17	L	0,00
F	3,33	N	0,00
W	3,33	R	0,00

- Dua huruf yang paling sering muncul di dalam cipherteks: huruf P dan Z.
- Dua huruf yang paling sering muncul di dalam B. Inggris: huruf E dan T.
- Kemungkinan besar,

  P adalah pemetaan dari e

  Z adalah pemetaan dari t

- Tetapi kita belum dapat memastikannya sebab masih diperlukan cara *trial and error* dan pengetahuan tentang Bahasa Inggris.

- Tetapi ini adalah langkah awal yang sudah bagus.

Iterasi 1:

```
UZ QSO VUOHXMOPV GPOZPEVSG ZWSZ OPFPESX UDBMETSX AIZ
 t           e   e te    t  t  e e             t

VUEPHZ HMDZSHZO WSFP APPD TSVP QUZW YMXUZUHSX
   e t    t  t     e  ee    e   t     t

EPYEPOPDZSZUFPO MB ZWP FUPZ HMDJ UD TMOHMQ
 e  e e t t  e    t e   et
```

- ZWP dan ZWSZ dipetakan menjadi t*e dan t**t
- Kemungkinan besar W adalah pemetataan dari H sehingga kata yang mungkin untuk ZWP dan ZWSZ adalah the dan that

- Diperoleh pemetaan:

  P → e
  Z → t
  W → h
  S → a

- Iterasi 2:

```
UZ QSO VUOHXMOPV GPOZPEVSG ZWSZ OPFPESX UDBMETSX AIZ
 t  a         e   e te  a  that  e e a        a    t

VUEPHZ HMDZSHZO WSFP APPD TSVP QUZW YMXUZUHSX
   e t    ta t  ha e  ee   a e   th     t  a

EPYEPOPDZSZUFPO MB ZWP FUPZ HMDJ UD TMOHMQ
 e  e e tat  e     the   et
```

- `WSFP` dipetakan menjadi `ha*e`.

- Dalam Bahasa Inggris, kata yang mungkin untuk `ha*e` hanyalah `have`, `hate`, `hale`, dan `haze`

- Dengan mencoba mengganti semua `F` di dalam cipherteks dengan `v`, `t`, `l`, dan `z`, maka huruf yang cocok adalah `v` sehingga `WSFP` dipetakan menjadi `have`

- Dengan mengganti `F` menjadi `v` pada kriptogram `EPYEPOPDZSZUFPO` sehingga menjadi `*e*e*e*tat*ve*`, maka kata yang cocok untuk ini adalah `representatives`

- Diperoleh pemetaan:

  E → r
  Y → p
  U → i
  O → s
  D → n

- Hasil akhir bila diselesaikan):

  It was disclosed yesterday that several informal but direct contacts have been made with political representatives of the viet cong in moscow

- Analisis frekuensi tetap bisa dilakukan meskipun spasi dihilangkan:

LIVITCSWPIYVEWHEVSRIQMXLEYVEOIEWHRXEXIPFEMVEWHKV
STYLXZIXLIKIIXPIJVSZEYPERRGERIMWQLMGLMXQERIWGPSR
IHMXQEREKIETXMJTPRGEVEKEITREWHEXXLEXXMZITWAWSQWX
SWEXTVEPMRXRSJGSTVRIEYVIEXCVMUIMWERGMIWXMJMGCSMW
XSJOMIQXLIVIQIVIXQSVSTWHKPEGARCSXRWIEVSWIIBXVIZM
XFSJXLIKEGAEWHEPSWYSWIWIEVXLISXLIVXLIRGEPIRQIVII
BGIIHMWYPFLEVHEWHYPSRRFQMXLEPPXLIECCIEVEWGISJKTV
WMRLIHYSPHXLIQIMYLXSJXLIMWRIGXQEROIVFVIZEVAEKPIE
WHXEAMWYEPPXLMWYRMWXSGSWRMHIVEXMSWMGSTPHLEVHPFKP
EZINTCMXIVJSVLMRSCMWMSWVIRCIGXMWYMX

- Hasil perhitungan frekuensi kemunculan huruf, bigram, dan trigram:
  - huruf `I` paling sering muncul,
  - `XL` adalah bigram yang paling sering muncul,
  - `XLI` adalah trigram yang paling sering muncul.

Ketiga data terbanyak ini menghasilkan dugaan bahwa

`I` berkoresponden dengan huruf plainteks `e`,

`XLI` berkoresponden dengan `the`,

`XL` berkoresponden dengan `th`

Pemetaan:

`I` → `e`

`X` → `t`

`L` → `h`

- XLEX dipetakan menjadi th*t.
- Kata yang cocok untuk th*t. adalah that.
- Jadi kita memperoleh: E → a
- Hasil iterasi pertama:

heVeTCSWPeYVaWHaVSReQMthaYVaOeaWHRtatePFaMVaWHKVSTY
htZetheKeetPeJVSZaYPaRRGaReMWQhMGhMtQaReWGPSReHMtQa
RaKeaTtMJTPRGaVaKaeTRaWHatthattMZeTWAWSQWtSWatTVaPM
RtRSJGSTVReaYVeatCVMUeMWaRGMeWtMJMGCSMWtSJOMeQtheVe
QeVetQSVSTWHKPaGARCStRWeaVSWeeBtVeZMtFSJtheKaGAaWHa
PSWYSWeWeaVtheStheVtheRGaPeRQeVeeBGeeHMWYPFhaVHaWHY
PSRRFQMthaPPtheaCCeaVaWGeSJKTVWMRheHYSPHtheQeMYhtSJ
theMWReGtQaROeVFVeZaVAaKPeaWHtaAMWYaPPthMWYRMWtSGSW
RMHeVatMSWMGSTPHhaVHPFKPaZeNTCMteVJSVhMRSCMWMSWVeRC
eGtMWYMt

- Selanjutnya,

  `Rtate` mungkin adalah `state`,

  `atthattMZE` mungkin adalah `atthattime`,

  `heVe` mungkin adalah `here`.

- Jadi, kita memperoleh pemetaan baru:

  `R` → `s`

  `M` → `i`

  `Z` → `m`

  `V` → `r`

- Hasil iterasi ke-2:

```
hereTCSWPeYraWHarSseQithaYraOeaWHstatePFairaWHKrSTYhtm
etheKeetPeJrSmaYPassGaseiWQhiGhitQaseWGPSseHitQasaKeaT
tiJTPsGaraKaeTsaWHatthattimeTWAWSQWtSWatTraPistsSJGSTr
seaYreatCriUeiWasGieWtiJiGCSiWtSJOieQthereQeretQSrSTWH
KPaGAsCStsWearSWeeBtremitFSJtheKaGAaWHaPSWYSWeWeartheS
therthesGaPesQereeBGeeHiWYPFharHaWHYPSssFQithaPPtheaCC
earaWGeSJKTrWisheHYSPHtheQeiYhtSJtheiWseGtQasOerFremar
AaKPeaWHtaAiWYaPPthiWYsiWtSGSWsiHeratiSWiGSTPHharHPFKP
ameNTCiterJSrhisSCiWiSWresCeGtiWYit
```

- Teruskan, dengan menerka kata-kata yang sudah dikenal, misalnya `remarA` mungkin `remark`, dsb

- Hasil iterasi 3:

```
hereuponlegrandarosewithagraveandstatelyairandbr
oughtmethebeetlefromaglasscaseinwhichitwasenclos
editwasabeautifulscarabaeusandatthattimeunknownt
onaturalistsofcourseagreatprizeinascientificpoin
tofviewthereweretworoundblackspotsnearoneextremi
tyofthebackandalongoneneartheotherthescalesweree
xceedinglyhardandglossywithalltheappearanceofbur
nishedgoldtheweightoftheinsectwasveryremarkablea
ndtakingallthingsintoconsiderationicouldhardlybl
amejupiterfor*hisopinionrespectingit
```

- Tambahkan spasi, tanda baca, dll

```
Here upon Legrand arose, with a grave and stately
air, and brought me the beetle from a glass case in
which it was enclosed. It was a beautiful
scarabaeus, and, at that time, unknown to
naturalists—of course a great prize in a scientific
point of view. There were two round black spots
near one extremity of the back, and a long one near
the other. The scales were exceedingly hard and
glossy, with all the appearance of burnished gold.
The weight of the insect was very remarkable, and,
taking all things into consideration, I could
hardly blame Jupiter for his opinion respecting it.
```