Cryptology I

(MTAT.07.002, 4 AP / 6 ECTS)

Lectures and	T 12-16	aud. 404
Exercises:	N 8-12	aud. 404

homepage:

`http://www.ut.ee/~peeter_l/teaching/kryptoi07s`

(contains lecture materials)

For grade: exercises at home and during the exam.

Functionality: System's property to do things we want it to do.

Security: System's property to not do things we want it not to do.

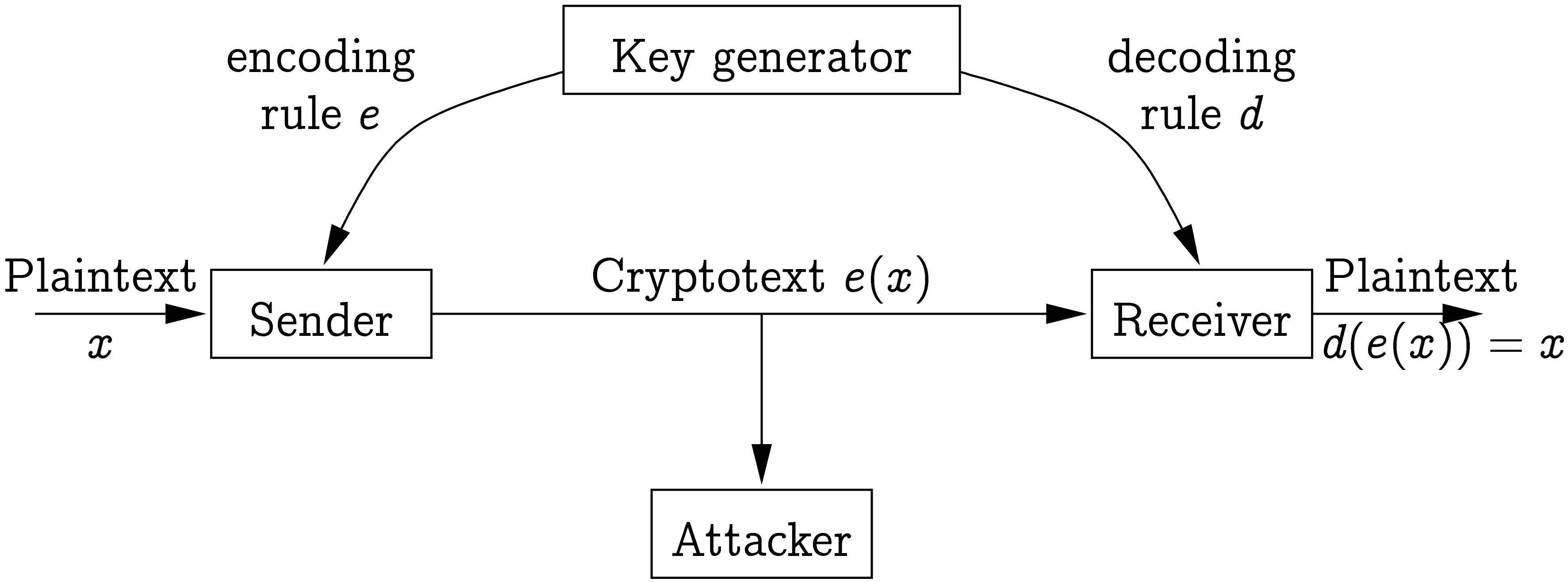
- (not speaking about availability)

Cryptography: Mathematical methods for ensuring system's security.

Cryptanalysis: Mathematical methods for breaking cryptography.

Cryptology: Cryptography and cryptanalysis.

Encryption and decryption:

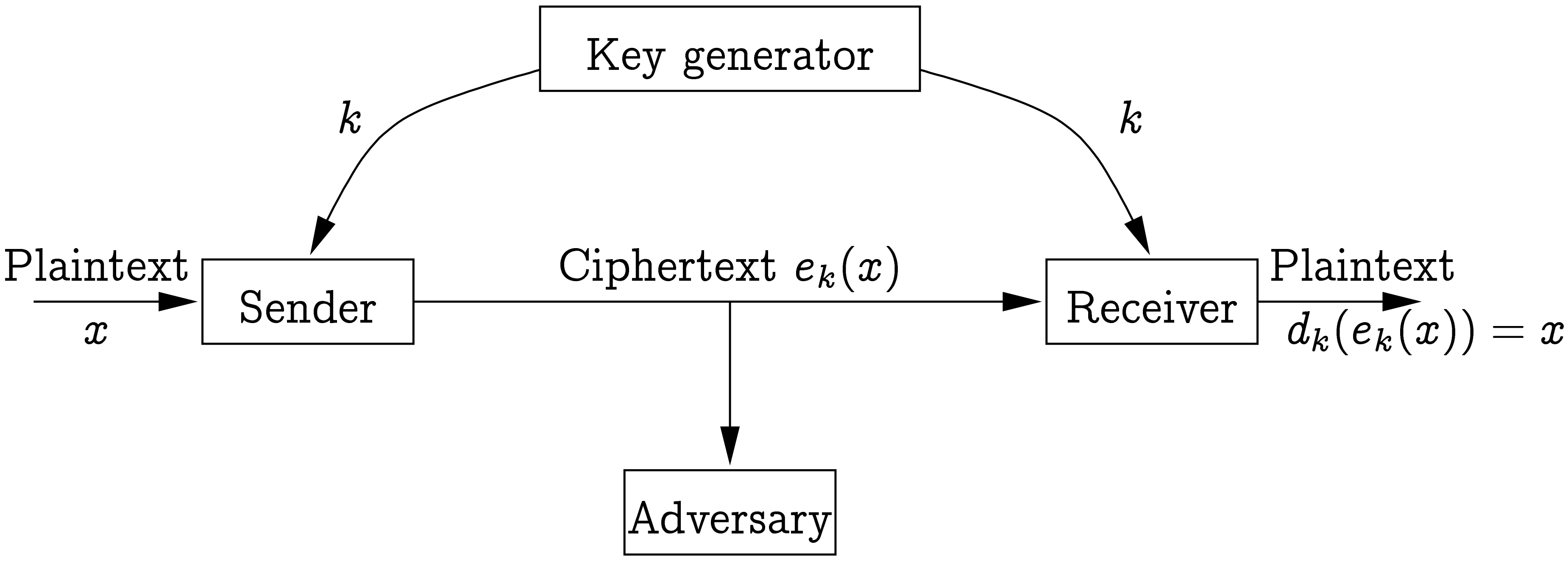


encoding- and decoding rules should have short descriptions.

Encryption system is a tuple $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$, where

- $\mathcal{P}$ is the set of possible plaintexts;
  - Often Σ^* for a suitable alphabet Σ.
- $\mathcal{C}$ is the set of possible ciphertexts;
- $\mathcal{K}$ is the set of possible keys;
- $\mathcal{E}$ and $\mathcal{D}$ are the sets of encoding and decoding rules.
  - If $e \in \mathcal{E}$, then $e : \mathcal{P} \longrightarrow \mathcal{C}$.
  - If $d \in \mathcal{D}$, then $d : \mathcal{C} \longrightarrow \mathcal{P}$.
- For all $k \in \mathcal{K}$ exist $e_k \in \mathcal{E}$ and $d_k \in \mathcal{D}$, such that $d_k \circ e_k$ is the identity function on $\mathcal{P}$.

Encryption and decryption



k describes the rules for encryption and decryption.

Ancient greeks, esp. Spartans used as an encryption system a tool called $\sigma\kappa\upsilon\tau\acute{\alpha}\lambda\eta$ (*skytale*; stick, *pulk*).

Decyption required a stick of the same girth. Key — diameter of the stick.

If the length of the text is not divisible with the number of letters on one round, then add nonsense letters to the end of the text.

Cryptanalysis: brute-force search of the key space.

Exercise: Break the following cryptogram of English text, encrypted with *skytale* (_ denotes space):

`Frh_a_rateolldre_f_ie_du_aedo_ntrs_ubfhm_ytaae_ae_`
`cs_o_likhwfslnwie_mo_aostawttnthl_heaeelserbr__u_s`
`lnGunu_e_anzno_slrycfepaexh_nixgwatdryav`

Cryptotext:

```
Frh_a_rateolldre_f_ie_du_aedo_ntrs_ubfhm_ytaae_ae_
cs_o_likhwfslnwie_mo_aostawttnthl_heaeelserbr__u_s
lnGunu_e_anzno_slrycfepaexh_nixgwatdryav
```

Decoding with *skytale* of girth 2:

```
Fhartolr__ed_eonr_bh_ta_ec__ihflwem_otwttlhaesrr__
lGn__nn_lyfpehnxwtrar__aeldefi_uad_tsufmyaea_solkw
sni_oasatnh_eeleb_usnuueazosrceax_igadyv
```

Cryptotext:

Frh_a_rateolldre_f_ie_du_aedo_ntrs_ubfhm_ytaae_ae_
cs_o_likhwfslnwie_mo_aostawttnthl_heaeelserbr__u_s
lnGunu_e_anzno_slrycfepaexh_nixgwatdryav

Decoding with *skytale* of girth 3:

F_rele___dnsbmtees_kfneooathhesb_sGu_z_rfahiwdaraa
od_idaot_f_a___lhsw__swnleeerulu_ansyee_xarvh_tlrf
eue_ruhyaacoiwlimattt_alr__nnenolcpxngty

Cryptotext:

```
Frh_a_rateolldre_f_ie_du_aedo_ntrs_ubfhm_ytaae_ae_
cs_o_likhwfslnwie_mo_aostawttnthl_heaeelserbr__u_s
lnGunu_e_anzno_slrycfepaexh_nixgwatdryav
```

Decoding with *skytale* of girth 4:

```
Fatl_e_orb_ae_ifwmowther_ln_nlfenwrr_edf_a_sfye_ok
siosthelbunuaorexiayhror_den_ht_c_hle_ttlasr_G_n_y
phxta_aleiudtumaaslwn_aan_ee_suezsca_gdv
```

Cryptotext:

```
Frh_a_rateolldre_f_ie_du_aedo_ntrs_ubfhm_ytaae_ae_
cs_o_likhwfslnwie_mo_aostawttnthl_heaeelserbr__u_s
lnGunu_e_anzno_slrycfepaexh_nixgwatdryav
```

Decoding with *skytale* of girth 5:

```
F_oeeanu_eclfi_ates_lunsfxxdrrl__etby_siseawhae_n_
zlehgrhalfddrfta_kl_otleruGenrp_wy_td_uoshaeohnmst
_eb_u_oyanaaaeri___ma__wwotnhlrsna_ceitv
```

Cryptotext:

`Frh_a_rateolldre_f_ie_du_aedo_ntrs_ubfhm_ytaae_ae_`
`cs_o_likhwfslnwie_mo_aostawttnthl_heaeelserbr__u_s`
`lnGunu_e_anzno_slrycfepaexh_nixgwatdryav`

Decoding with *skytale* of girth 6:

`Frl__nbte_feoths_G__fhwaradiatfa_ls_sneeuuase_avht`
`reerhacilmttar_nnlpnt_ee_dsmesknoahebsuzraidao_do_`
`___hw_wlerl_nyexr_lfu_uyaowiat_l_neocxgy`

Cryptotext:

`Frh_a_rateolldre_f_ie_du_aedo_ntrs_ubfhm_ytaae_ae_`
`cs_o_likhwfslnwie_mo_aostawttnthl_heaeelserbr__u_s`
`lnGunu_e_anzno_slrycfepaexh_nixgwatdryav`

Decoding with *skytale* of girth 7:

`Far_out_in_the_uncharted_backwaters_of_the_unfashi`
`onable_end_of_the_western_spiral_arm_of_the_Galaxy`
`_lies_a_small_unregarded_yellow_sunzyxwv`

Skytale is an example of transposition cipher.

We do not change the letters, but their order.

Next example is about a substitution cihper.

Letters are changed, but their order remains the same.

Ring of congruence classes $\mathbb{Z}_n$:

- elements $\{0, 1, \ldots, n-1\}$;
- addition and multiplication: as in $\mathbb{Z}$, but *modulo* n.

Let us identify Latin alphabet and $\mathbb{Z}_{26}$: A $\equiv 0$, B $\equiv 1$, $\ldots$, Z $\equiv 25$.

Shift cipher:

- $\mathcal{K} = \mathbb{Z}_{26}$.
- e_k: replace each letter x with $x + k$.
- d_k: replace each letter x with $x - k$.

Also known as Caesar's cipher.

ROT13 is shift cipher with the key 13.

Example:

- plaintext: "Quidquid latine dictum sit, altum viditur"
- key: 5

x	ABC DEF GHI JKL MNO PQR STU VWX YZ
$e_5(x)$	FGH IJK LMN OPQ RST UVW XYZ ABC DE

- ciphertext "Vznivzni qfynsj inhyzr xny, fqyzr aninyzw"

Cryptanalysis: brute-forcing the key space.

Exercise: break the following cryptogram of English text, encrypted with shift cipher:

Obr gc hvs dfcpzsa fsaowbsr; zchg ct hvs dscdzs ksfs asob, obr acgh ct hvsa ksfs awgsfopzs, sjsb hvs cbsg kwhv rwuwhoz kohqvsg.

Exercise: break the following cryptogram of English text, encrypted with shift cipher:

Obr gc hvs dfcpzsa fsaowbsr; zchg ct hvs dscdzs ksfs asob, obr acgh ct hvsa ksfs awgsfopzs, sjsb hvs cbsg kwhv rwuwhoz kohqvsg.

it is not hard to try out 26 keys, but…

Exercise: break the following cryptogram of English text, encrypted with shift cipher:

Obr gc hvs dfcpzsa fsaowbsr; zchg ct hvs dscdzs ksfs asob, obr acgh ct hvsa ksfs awgsfopzs, sjsb hvs cbsg kwhv rwuwhoz kohqvsg.

it is not hard to try out 26 keys, but. . .

Cryptogram contains several occurrences of “hvs”.

Exercise: break the following cryptogram of English text, encrypted with shift cipher:

Obr gc hvs dfcpzsa fsaowbsr; zchg ct hvs dscdzs ksfs asob, obr acgh ct hvsa ksfs awgsfopzs, sjsb hvs cbsg kwhv rwuwhoz kohqvsg.

it is not hard to text 26 keys, but. . .

Cryptogram contains several occurrences of “hvs”.

Could its corresponding plaintext be “the”?

$$\text{hvs} \equiv 7, 21, 18$$

$$\text{the} \equiv 19, 7, 4$$

$e_k(x) = x + k$, thus $k = e_k(x) - x$.

$7 - 19 = 21 - 7 = 18 - 4 = 14 \pmod{26}$

Cryptotext:

Obr gc hvs dfcpzsa fsaowbsr; zchg ct hvs dscdzs ksfs asob, obr acgh ct hvsa ksfs awgsfopzs, sjsb hvs cbsg kwhv rwuwhoz kohqvsg.

Decoded with the key 14:

And so the problem remained; lots of the people were mean, and most of them were miserable, even the ones with digital watches.

Exercise. Break the following cryptograms obtained from Latin texts using the Caesar's cipher:

LQ YLQR YHULWDV

RYWY RYWSXS VEZEC OCD

Shift cipher is a special case of substitution cipher.

- Key: A permutation σ of the alphabet Σ.
- e_σ: replace each letter x with $\sigma(x)$.
- d_σ: replace each letter x with $\sigma^{-1}(x)$.

Cryptanalysis: there are $\geqslant 4 \cdot 10^{26}$ keys, making brute-force search impossible.

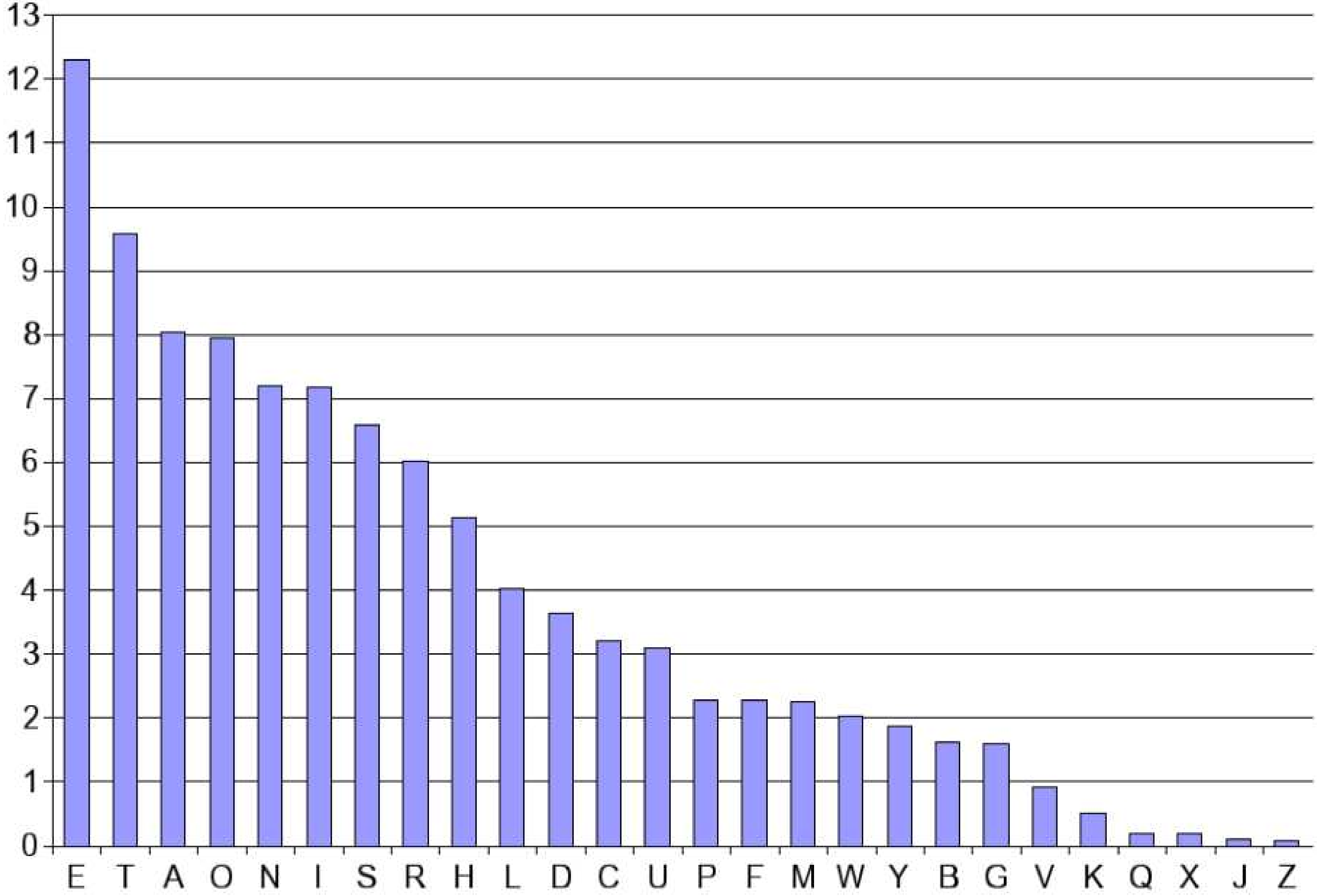
Exercise. How to encode a key of the substitution cipher in as few bits as possible?

We can break it by analysing letter frequencies.

Letter frequencies in english (%):

A	$8,05$	H	$5,14$	O	$7,94$	U	$3,10$
B	$1,62$	I	$7,18$	P	$2,29$	V	$0,93$
C	$3,20$	J	$0,10$	Q	$0,20$	W	$2,03$
D	$3,65$	K	$0,52$	R	$6,03$	X	$0,20$
E	$12,31$	L	$4,03$	S	$6,59$	Y	$1,88$
F	$2,28$	M	$2,25$	T	$9,59$	Z	$0,09$
G	$1,61$	N	$7,19$				

Source: Jan Willemson, "Sissejuhatus krüptoloogiasse".



13
12
11
10
9
8
7
6
5
4
3
2
1
0
E T A O N I S R H L D C U P F M W Y B G V K Q X J Z

Most common digraphs:

th	1.52	*ha*	0.56	*is*	0.46	*se*	0.08
he	1.28	*es*	0.56	*or*	0.43	*le*	0.08
in	0.94	*st*	0.55	*ti*	0.34	*sa*	0.06
er	0.94	*en*	0.55	*as*	0.33	*si*	0.05
an	0.82	*ed*	0.53	*te*	0.27	*ar*	0.04
re	0.68	*to*	0.52	*et*	0.19	*ve*	0.04
nd	0.63	*it*	0.50	*ng*	0.18	*ra*	0.04
at	0.59	*ou*	0.50	*of*	0.16	*ld*	0.02
on	0.57	*ea*	0.47	*al*	0.09	*ur*	0.02
nt	0.56	*hi*	0.46	*de*	0.09		

Most common trigraphs (descending): THE, ING, AND, HER, ERE, ENT, THA, NTH, WAS, ETH, FOR, DTH.

Exercise: break the following cryptogram of an English text created with the substitution cipher:

Myd lwez odhrlw ilh vylp myd ylxrd vur gw uwz vuz rodngue vur Uhmyxh Fdwm, uwf myum vur lwez kdnuxrd gm yuoodwdf ml kd myd lwd yd egcdf gw. Yd yuf egcdf gw gm ilh uklxm myhdd zduhr, dcdh rgwnd yd yuf plcdf lxm li Elwflw kdnuxrd gm pufd ygp wdhclxr uwf ghhgmuked. Yd vur uklxm myghmz ur vdee, fuhq yughdf uwf wdcdh bxgmd um durd vgmy ygprdei. Myd mygwt myum xrdf ml vlhhz ygp plrm vur myd iunm myum odloed uevuzr xrdf ml urq ygp vyum yd vur ellqgwt rl vlhhgdf uklxm. Yd vlhqdf gw elnue hufgl vygny yd uevuzr xrdf ml mdee ygr ihgdwfr vur u elm plhd gwmdhdrmgwt myuw mydz ohlkukez mylxtym. Gm vur, mll - plrm li ygr ihgdwfr vlhqdf gw ufcdhmgrgwt.

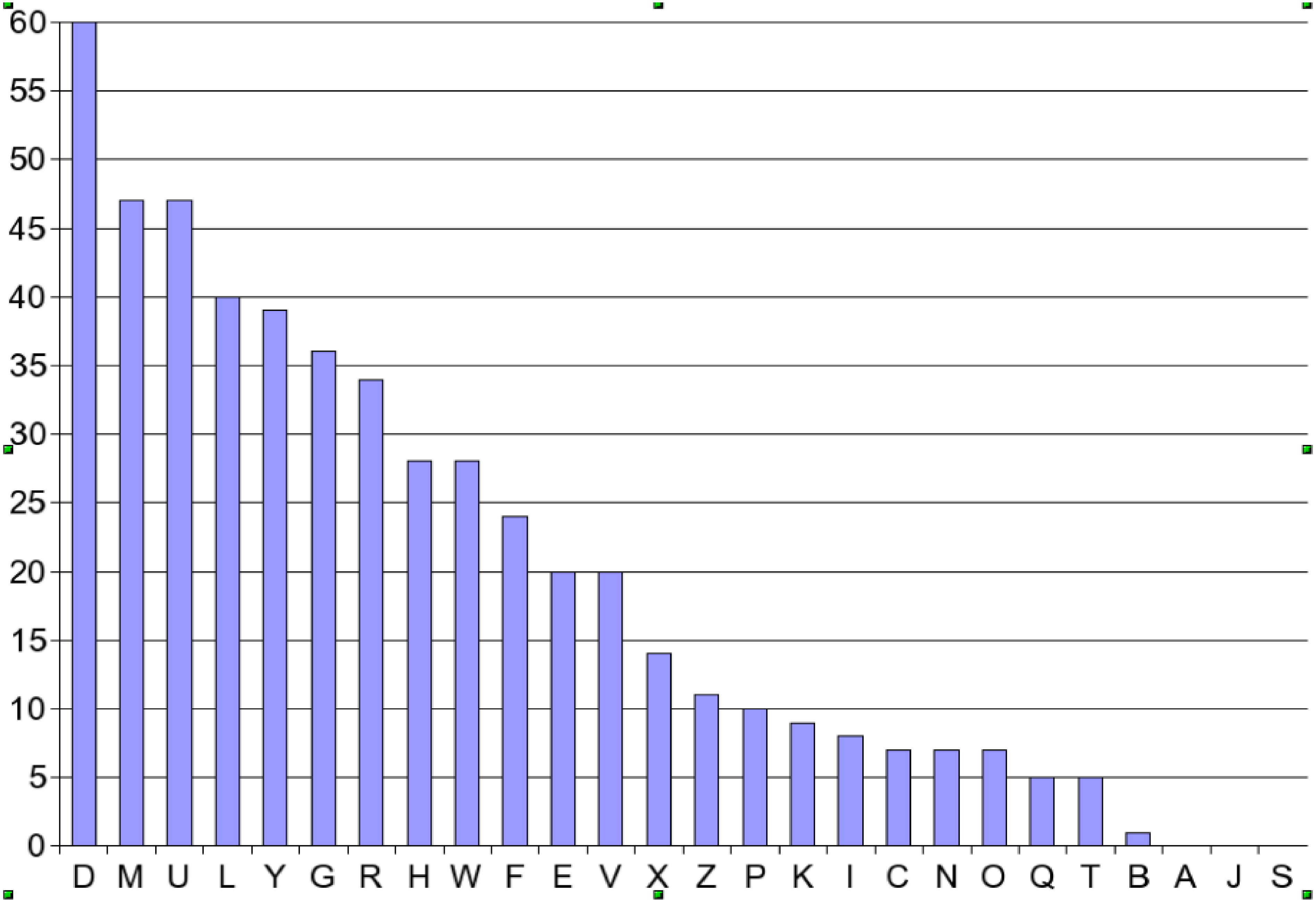
First step: count the number of occurrencies of each letter.

a	0
b	1
c	7
d	60
e	20
f	24
g	36

h	28
i	8
j	0
k	9
l	40
m	47
n	7

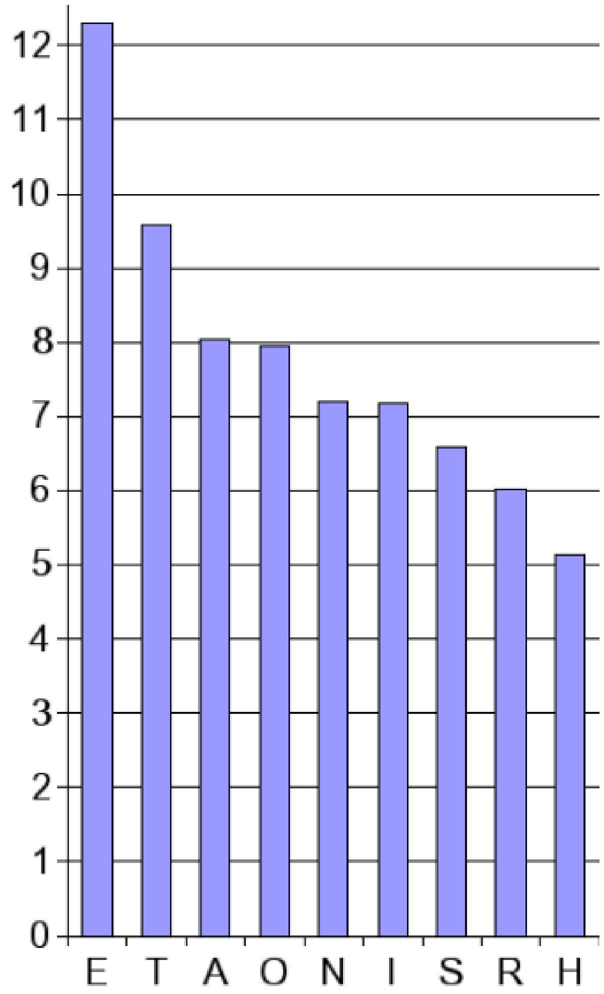
o	7
p	10
q	5
r	34
s	0
t	5

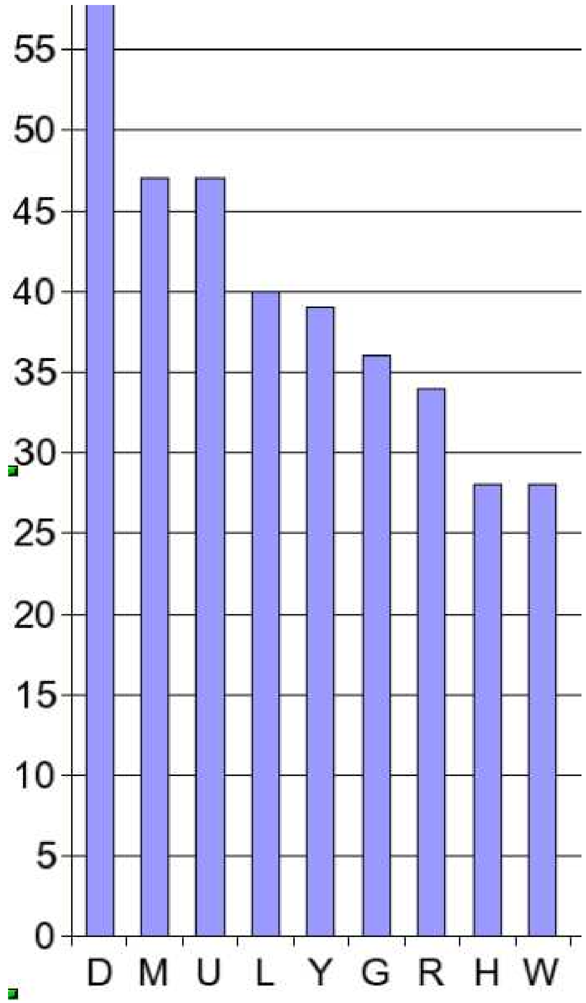
u	47
v	20
w	28
x	14
y	39
z	11



d in cryptotext is probably e in plaintext.

Mye lwez oehrlw ilh vylp mye ylxre vur gw uwz vuz roengue vur Uhmyxh Fewm, uwf myum vur lwez kenuxre gm yuooewef ml ke mye lwe ye egcef gw. Ye yuf egcef gw gm ilh uklxm myhee zeuhr, eceh rgwne ye yuf plcef lxm li Elwflw kenuxre gm pufe ygp wehclxr uwf ghhgmukee. Ye vur uklxm myghmz ur veee, fuhq yughef uwf weceh bxgme um eure vgmy ygpreei. Mye mygwt myum xref ml vlhhz ygp plrm vur mye iunm myum oeloee uevuzr xref ml urq ygp vyum ye vur ellqgwt rl vlhhgef uklxm. Ye vlhqef gw elnue hufgl vygny ye uevuzr xref ml meee ygr ihgewfr vur u elm plhe gwmehermgwt myuw myez ohlkukez mylxtym. Gm vur, mll - plrm li ygr ihgewfr vlhqef gw ufcehmgrgwt.


0
1
2
3
4
5
6
7
8
9
10
11
12
E T A O N I S R H


0
5
10
15
20
25
30
35
40
45
50
55
D M U L Y G R H W

Plaintext T — cryptotext M or U

Plaintext A and O — cryptotext U/M, L, Y

etc.

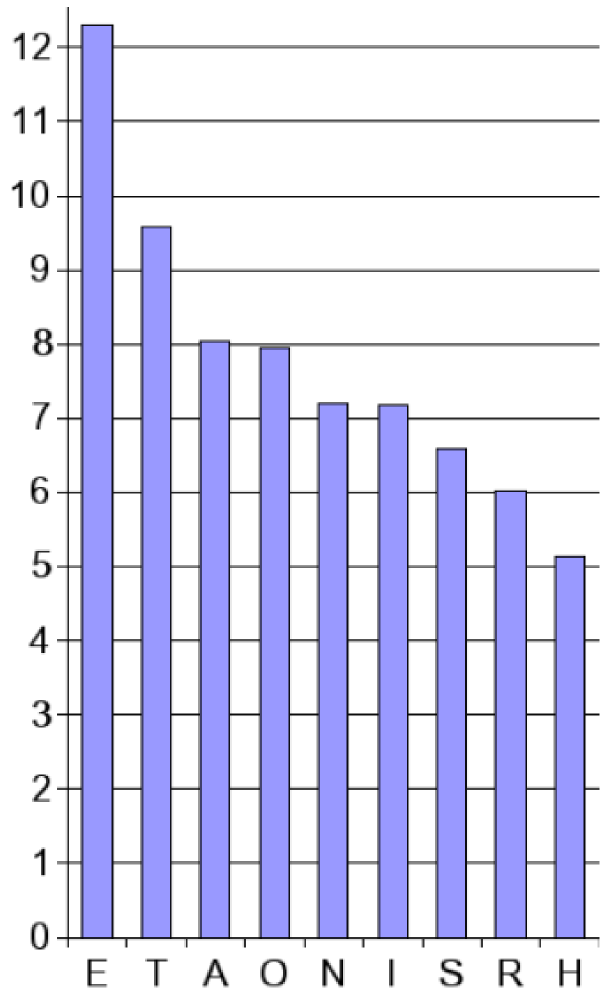
Count the most frequent digraphs...

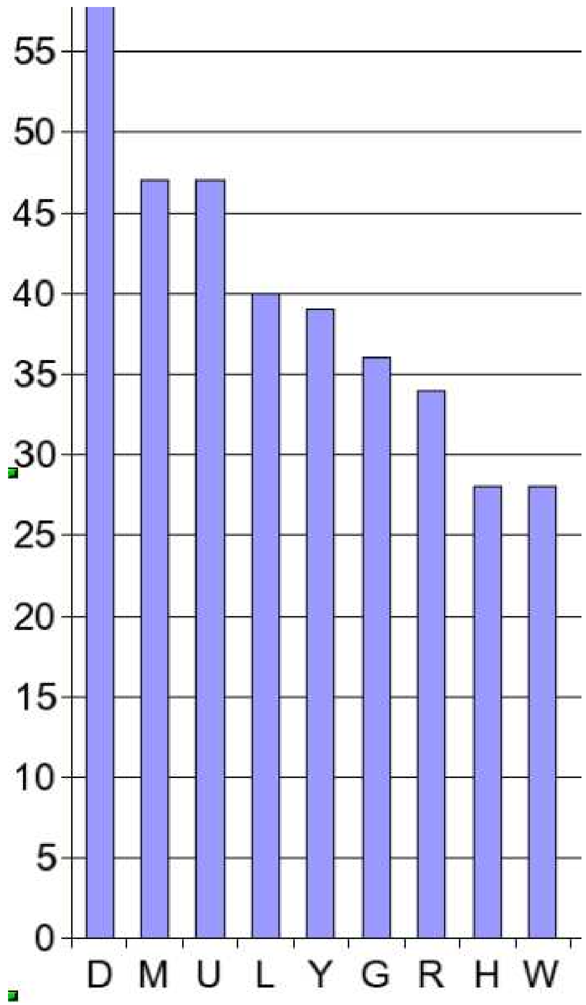
cryptotext plaintext

my	16	yg	9	th	1.52	at	0.59
yd	13	yu	9	he	1.28	on	0.57
df	11	rd	8	in	0.94	nt	0.56
gw	11	gm	7	er	0.94	ha	0.56
ur	11	lh	7	an	0.82	es	0.56
vu	11	lx	7	re	0.68	st	0.55
		xr	7	nd	0.63	en	0.55

m (crypto) — t(plain). y(crypto) — h(plain).

The lwez oehrlw ilh vhlp the hlxre vur gw uwz vuz roengue vur Uhthxh Fewt, uwf thut vur lwez kenuxre gt huooewef tl ke the lwe he egcef gw. He huf egcef gw gt ilh uklxt thhee zeuhr, eceh rgwne he huf plcef lxt li Elwflw kenuxre gt pufe hgp wehclxr uwf ghhgtukee. He vur uklxt thghtz ur veee, fuhq hughef uwf weceh bxgte ut eure vgth hgpreei. The thgwt thut xref tl vlhhz hgp plrt vur the iunt thut oeloee uevuzr xref tl urq hgp vhut he vur ellqgwt rl vlhhgef uklxt. He vlhqef gw elnue hufgl vhgnh he uevuzr xref tl teee hgr ihgewfr vur u elt plhe gwtehertgwt thuw thez ohlkukez thlxtht. Gt vur, tll - plrt li hgr ihgewfr vlhqef gw ufcehtgrgwt.





eht dym

u(crypto) is either a or o(plain).

The lwez oehrlw ilh vhlp the hlxre vur gw uwz vuz roengue vur Uhthxh Fewt, uwf thut vur lwez kenuxre gt huooewef tl ke the lwe he egcef gw. He huf egcef gw gt ilh uklxt thhee zeuhr, eceh rgwne he huf plcef lxt li Elwflw kenuxre gt pufe hgp wehclxr uwf ghhgtukee. He vur uklxt thghtz ur veee, fuhq hughef uwf weceh bxgte ut eure vgth hgpreei. The thgwt thut xref tl vlhhz hgp plrt vur the iunt thut oeloee uevuzr xref tl urq hgp vhut he vur ellqgwt rl vlhhgef uklxt. He vlhqef gw elnue hufgl vhgnh he uevuzr xref tl teee hgr ihgewfr vur u elt plhe gwtehertgwt thuw thez ohlkukez thlxtht. Gt vur, tll - plrt li hgr ihgewfr vlhqef gw ufcehtgrgwt.

u(crypto) is a(plain).

The lwez oehrlw ilh vhlp the hlxre var gw awz vaz roengae var Ahthxh Fewt, awf that var lwez kenaxre gt haooewef tl ke the lwe he egcef gw. He haf egcef gw gt ilh aklxt thhee zeahr, eceh rgwne he haf plcef lxt li Elwflw kenaxre gt pafe hgp wehclxr awf ghhgtakee. He var aklxt thghtz ar veee, fahq haghef awf weceh bxgte at eare vgth hgpreei. The thgwt that xref tl vlhhz hgp plrt var the iant that oeloee aevazr xref tl arq hgp vhat he var ellqgwt rl vlhhgef aklxt. He vlhqef gw elnae hafgl vhgnh he aevazr xref tl teee hgr ihgewfr var a elt plhe gwtehertgwt thaw thez ohlkakez thlxtht. Gt var, tll - plrt li hgr ihgewfr vlhqef gw afcehtgrgwt.

The lwez oehrlw ilh vhlp the hlxre var gw awz vaz roengae var Ahthxh Fewt, awf that var lwez kenaxre gt haooewef tl ke the lwe he egcef gw. He haf egcef gw gt ilh aklxt thhee zeahr, eceh rgwne he haf plcef lxt li Elwflw kenaxre gt pafe hgp wehclxr awf ghhgtakee. He var aklxt thghtz ar veee, fahq haghef awf weceh bxgte at eare vgth hgpreei. The thgwt that xref tl vlhhz hgp plrt var the iant that oeloee aevazr xref tl arq hgp vhat he var ellqgwt rl vlhhgef aklxt. He vlhqef gw elnae hafgl vhgnh he aevazr xref tl teee hgr ihgewfr var a elt plhe gwtehertgwt thaw thez ohlkakez thlxtht. Gt var, tll - plrt li hgr ihgewfr vlhqef gw afcehtgrgwt.

h(crypto) is r(plain)

The lwez oerrlw ilr vhlp the hlxre var gw awz vaz roengae var Arthxr Fewt, awf that var lwez kenaxre gt haooewef tl ke the lwe he egcef gw. He haf egcef gw gt ilr aklxt three zearr, ecer rgwne he haf plcef lxt li Elwflw kenaxre gt pafe hgp werclxr awf grrgtakee. He var aklxt thgrtz ar veee, farq hagref awf wecer bxgte at eare vgth hgpreei. The thgwt that xref tl vlrrz hgp plrt var the iant that oeloee aevazr xref tl arq hgp vhat he var ellqgwt rl vlrrgef aklxt. He vlrqef gw elnae rafgl vhgnh he aevazr xref tl teee hgr irgewfr var a elt plre gwterertgwt thaw thez orlkakez thlxtht. Gt var, tll - plrt li hgr irgewfr vlrqef gw afcertgrgwt.

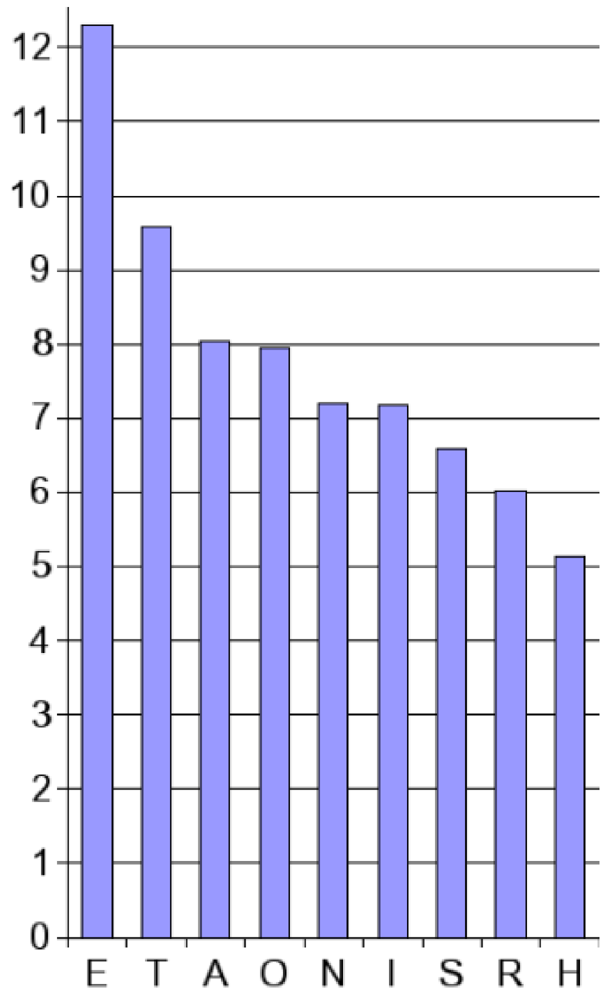
The lwez oerrlw ilr vhlp the hlxre var gw awz vaz
roengae var Arthxr Fewt, awf that var lwez kenaxre gt
haooewef tl ke the lwe he egcef gw. He haf egcef gw gt ilr
aklxt three zearr, ecer rgwne he haf plcef lxt li Elwflw
kenaxre gt pafe hgp werclxr awf grrgtakee. He var aklxt
thgrtz ar veee, farq hagref awf wecer bxgte at eare vgth
hgpreei. The thgwt that xref tl vlrrz hgp plrt var the iant
that oeloee aevazr xref tl arq hgp vhat he var ellqgwt rl
vlrrgef aklxt. He vlrqef gw elnae rafgl vhgnh he aevazr
xref tl teee hgr irgewfr var a elt plre gwterertgwt thaw
thez orlkakez thlxtht. Gt var, tll - plrt li hgr irgewfr
vlrqef gw afcertgrgwt.

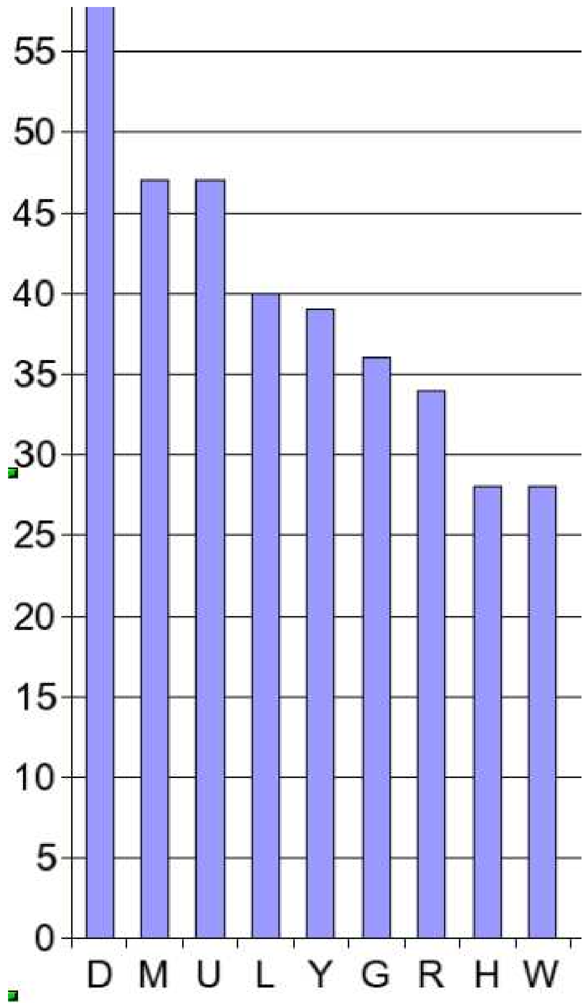
x(crypto) on u(plain)

The lwez oerrlw ilr vhlp the hlure var gw awz vaz roengae var Arthur Fewt, awf that var lwez kenaure gt haooewef tl ke the lwe he egcef gw. He haf egcef gw gt ilr aklut three zearr, ecer rgwne he haf plcef lut li Elwflw kenaure gt pafe hgp werclur awf grrgtakee. He var aklut thgrtz ar veee, farq hagref awf wecer bugte at eare vgth hgpreei. The thgwt that uref tl vlrrz hgp plrt var the iant that oeloee aevazr uref tl arq hgp vhat he var ellqgwt rl vlrrgef aklut. He vlrqef gw elnae rafgl vhgnh he aevazr uref tl teee hgr irgewfr var a elt plre gwterertgwt thaw thez orlkakez thlutht. Gt var, tll - plrt li hgr irgewfr vlrqef gw afcertgrgwt.

The lwez oerrlw ilr vhlp the hlure var gw awz vaz roengae var Arthur Fewt, awf that var lwez kenaure gt haooewef tl ke the lwe he egcef gw. He haf egcef gw gt ilr aklut three zearr, ecer rgwne he haf plcef lut li Elwflw kenaure gt pafe hgp werclur awf grrgtakee. He var aklut thgrtz ar veee, farq hagref awf wecer bugte at eare vgth hgpreei. The thgwt that uref tl vlrrz hgp plrt var the iant that oeloee aevazr uref tl arq hgp vhat he var ellqgwt rl vlrrgef aklut. He vlrqef gw elnae rafgl vhgnh he aevazr uref tl teee hgr irgewfr var a elt plre gwterertgwt thaw thez orlkakez thlutht. <u>Gt</u> var, <u>tll</u> - plrt li hgr irgewfr vlrqef gw afcertgrgwt.

g(crypto) and l(crypto) are vowels.





aehrtu

udyhmx

g(crypto) and l(crypto) are frequent, y(plain) is infrequent.

g(crypto) is i(plain) and l(crypto) is o(plain).

The owez oerrow ior vhop the houre var iw awz vaz roeniae var Arthur Fewt, awf that var owez kenaure it haooewef to ke the owe he eicef iw. He haf eicef iw it ior akout three zearr, ecer riwne he haf pocef out oi Eowfow kenaure it pafe hip wercour awf irritakee. He var akout thirtz ar veee, farq hairef awf wecer buite at eare vith hipreei. The thiwt that uref to vorrz hip port var the iant that oeooee aevazr uref to arq hip vhat he var eooqiwt ro vorrief akout. He vorqef iw eonae rafio vhinh he aevazr uref to teee hir iriewfr var a eot pore iwtererti wt thaw thez orokakez thoutht. It var, too - port oi hir iriewfr vorqef iw afcertiriwt.

The owez oerrow ior vhop the <u>houre</u> var iw awz vaz roeniae var Arthur Fewt, awf that var owez kenaure it haooewef to ke the owe he eicef iw. <u>He haf</u> eicef iw it ior akout three zearr, ecer riwne he haf pocef out oi Eowfow kenaure it pafe hip wercour awf irritakee. He var akout thirtz ar veee, farq hairef <u>awf</u> wecer <u>buite</u> at eare <u>vith</u> hipreei. The thiwt that uref to vorrz hip port var the iant that oeooee aevazr uref to arq hip vhat he var eooqiwt ro vorrief akout. He vorqef iw eonae rafio vhinh he aevazr uref to teee hir iriewfr var a eot pore iwterertiwt thaw thez orokakez thoutht. <u>It var</u>, too - port oi hir iriewfr vorqef iw afcertiriwt.

r(crypto) is s(plain). f(crypto) is d(plain). b(crypto) is q(plain). v(crypto) is w(plain). w(crypto) is n(plain).

The onez oerson ior whop the house was in anz waz soeniae was Arthur Dent, and that was onez kenause it haooened to ke the one he eiced in. He had eiced in it ior akout three zears, ecer sinne he had poced out oi Eondon kenause it pade hip nercous and irritakee. He was akout thirtz as weee, darq haired and necer quite at ease with hipseei. The thint that used to worrz hip post was the iant that oeooee aewazs used to asq hip what he was eooqint so worried akout. He worqed in eonae radio whinh he aewazs used to teee his iriends was a eot pore interestint than thez orokakez thoutht. It was, too - post oi his iriends worqed in adcertisint.

and now it's easy. . .

The only person for whom the house was in any way special was Arthur Dent, and that was only because it happened to be the one he lived in. He had lived in it for about three years, ever since he had moved out of London because it made him nervous and irritable. He was about thirty as well, dark haired and never quite at ease with himself. The thing that used to worry him most was the fact that people always used to ask him what he was looking so worried about. He worked in local radio which he always used to tell his friends was a lot more interesting than they probably thought. It was, too - most of his friends worked in advertising.

x	abcdefghijklmnopqrstuvwxyz
$\sigma(x)$	uknfditygsqepwlobhrmxcvjza

All encoding rules of a substitution cipher constitute a group.

The same holds for the shift cipher.

- For all $k, k' \in \mathcal{K}$ exists $k'' \in \mathcal{K}$ such that $e_{k'} \circ e_k = e_{k''}$.
  - Shift c.: $k'' = k + k'$, Substitution c.: $k'' = k' \circ k$.
- Exists a key $k \in \mathcal{K}$, such that e_k is the identity transformation.
- For all $k \in \mathcal{K}$ exists $k' \in \mathcal{K}$ such that $e_k = d_{k'}$.

Substitution cipher is monoalphabetic — each letter is always encoded to the same letter.

Example of a polyalphabetic cipher — Vigenère cipher.

Basically, it applies shift ciphers with different keys to different text positions.

Example: let the key be “secret” and plaintext “this has been hidden well”. The key is $(18, 4, 2, 17, 4, 19)$.

t	h	i	s	h	a	s	b	e	e	n
19	7	8	18	7	0	18	1	4	4	13
18	4	2	17	4	19	18	4	2	17	4
11	11	10	9	11	19	10	5	6	21	17
l	l	k	j	l	t	k	f	g	v	r

h	i	d	d	e	n	w	e	l	l
7	8	3	3	4	13	22	4	11	11
19	18	4	2	17	4	19	18	4	2
0	0	7	5	21	17	15	22	15	13
a	a	h	f	v	r	p	w	p	n

Ciphertext: “llkj ltk fgvr aahfvr pwpn”.

Exercise: break the following cryptogram of English text created with Vigenère cipher:

We ywqzeq iddug bjt cnjkc bhb eduyl ute imtn lvbvae; fbtpntm odnfbtduf ajpdbeu aobugs aal ntacmf ligp vwe gqpn fyqezeeqpv fyiot, bhb cal jiu fuvmv. We ozgptumf p svtgct gpcck lww io gpg Seabtpsfqu. Ihr Lgcteiuhif itt aa cpguyg vgiom qu gbctbaalu, p wvtf qug xntafipi bhvew wuwo ihr Dqvoaa jpd emetngta iaxmp io ruraolqpv af kcieeqpv sgihu oa bjtie tqcg uiwa fymgis, bv vwe fbtxcg cpseeavpnqqpv tuiv ihrg mtec bjtmfmnkef dggy zcew tb bjtmfmnkef.

First step: find the length of the key.

One way of doing it is the Kasiski's test:

Let us find identical sequences of length $\geqslant 3$ from the ciphertext. It is likely that they correspond to identical plaintexts and their distance is divisible by the length of the key.

We ywqzeq iddug bjt cnjkc bhb eduyl ute imtn lvbvae; fbtpntm odnfbtduf ajpdbeu aobugs aal ntacmf ligp vwe gqpn fyqezeeqpv fyiot, bhb cal jiu fuvmv. We ozgptumf p svtgct gpcck lww io gpg Seabtpsfqu. Ihr Lgcteiuhif itt aa cpguyg vgiom qu gbctbaalu, p wvtf qug xntafipi bhvew wuwo ihr Dqvoaa jpd emetngta iaxmp io ruraolqpv af kcieeqpv sgihu oa bjtie tqcg uiwa fymgis, bv vwe fbtxcg cpseeavpnqqpv tuiv ihrg mtec bjtmfmnkef dggy zcew tb bjtmfmnkef.

Distance of “bjtmfmnkef”-s is 20. Distance of “ajpd”-s is 175. Distance of “bjt”-s is 265 and 55. The key length is probably 5.

Other way: index of coincidence.

The index of coincidence $I_c(s)$ of a string s is the probability that two randomly chosen positions of s contain the same letter.

Let $p_{s,x} = \dfrac{\text{num. of occurrences of } x \text{ in } s}{|s|}$. Then $I_c(s) = \sum_{x \in \Sigma} p_{s,x}^2$.

For a random string s: $I_c(s) \approx 0.038$ ($|\Sigma| = 26$).

For an English text s: $I_c(s) \approx 0.066$ (the probabilities are from the table above).

For an English text encrypted with a monoalphabetic cipher s: also $I_c(s) \approx 0.066$.

If, from the ciphertext, we choose the positions where the same shift has been applied, then the I_c of the corresponding subsequence should be ≈ 0.066.

If we choose positions where several different shifts are used then the result looks more random and its I_c should be lower.

Assume that the length of key is 1. The I_c of the entire cryptotext is ≈ 0.049. Hence there are several shifts in use and our assumption is wrong.

Assume $|k| = 2$. Then s_{even} is

```
wyqeidgjcjcheultitlbaftnmdftuapbuousanamlgveqnyeeqvyobbajuumwogtmpv
gtpclwopsatsqirgtihftacgyvimubtalpvfuxtfpbvwuohdvajdmtgaamirroqvfce
qvghobteqgiaygsvwftccsevnqvuvhgtcjmmkfgycwbjmmkf
```

and s_{odd} is

```
ewzqdubtnkbbdyuemnvvebptonbdfjdeabgaltcfipwgpfqzepfithclifvvezpufst
cgckwiggebpfuhlceuiitapuggoqgcbauwtqgnaiihewwirqoapeentixpoualpakie
psiuajitcuwfmibvebxgpeapqptiirmebtfnedgzetbtfne
```

Indices of coincidence are respectively 0.049 and 0.056. Probably too small.

Assume $|k| = 3$. Then s_0 is

`wwedgtjbeytmlvfpmntfpeogatmivgnqepytbluvwztfvcpkwgsbsurciitag`
`giqbblwfgtibeuidojettapraqaceviojecifgbwbcpepqtvrtbmnfgctjfk`

s_1 is

`eqqdbckhdletvabnofdadubslafgwqfeevibcjfmegupttclipetfiltuftcu`
`voucauvqxaphwwhqapmnaxiuopfiqshattgwyivetgsanpuigejfkdyebtme`

s_2 is

`yziujncbuuinbettdbujbauanclpepyzqfohaiuvopmsggcwogapqhgehiapy`
`gmgtaptunfivworvadegimorlvkepgubiquamsvfxcevqvihmctmegzwbmnf`

and the indices of coincidence are respectively 0.056, 0.052 ja 0.049.

For $|k| = 4$ indices of coincidence are
0.054, 0.064, 0.053, 0.059.

For $|k| = 5$ indices of coincidence are
0.081, 0.083, 0.082, 0.090, 0.076.

For $|k| = 6$ indices of coincidence are
0.055, 0.069, 0.057, 0.065, 0.054, 0.059.

So probably $|k| = 5$. The size of indices of coincidence is caused by the shortness of the text.

Five ciphertexts, each of them obtained by some shift:

```
wzdtcdtnapddpastlwnzvtafwppccispichtggubpqtiwivptiiavivutcaiwxspvittkgwtk
eeucbuelennudoaaiefefbluetstkoeshtiauigawuabwhodnaooaesoigfsecenthemeytme
yqgnhyivftffbbacggyeyhjvouvglgafrefayobavgfhuraegxrlfegaeuybfgequrcffzbff
wibjblmbbmbaeulmpqqqibimzmtpwpbqliicgmcltxivwdamtmuqkqibtimvbcaqigbmdcbm
qdjkeutvtotjugnfvpepocuvgfgcwgtugutpvqtufnpeoqjeaprpcphjqwgvtpvpvmjngejn
```

Denote these texts by $s_0, \ldots, s_4$. Let the letters of the key be $k_0, \ldots, k_4$.

Subtracting k_i from the letters of s_i gives something where the letters are distributed as in English.

Next step: find $k_i - k_j$ for different i and j.

Mutual index of coincidence $MI_c(s, s')$ of the strings s and s' is the probability that a randomly chosen letter of s and a randomly chosen letter of s' are equal.

$$MI_c(s, s') = \sum_{x \in \Sigma} p_{s,x} p_{s',x}$$

If s and s' are English texts then $MI_c(s, s') \approx 0.066$.

$MI_c(s, s')$ does not change when we apply the same monoalphabetic cipher (with the same key) to both s and s'.

Let p_x be the frequence of the letter x in English. Let s be English text. Let s' be obtained from English text by applying to it shift cipher with the key ℓ.

$$MI_c(s, s') = \sum_{i=0}^{25} p_i p_{i+\ell},$$

I.e. $MI_c(s, s')$ depends only on ℓ.

If s [resp. s' had been obtained from English text with the key i [resp. $i + \ell$] then MI_c would have been the same.

The respective values of MI_c are (depending on ℓ):

0	0.066	7	0.038	14	0.039	20	0.036
1	0.040	8	0.033	15	0.045	21	0.033
2	0.032	9	0.035	16	0.038	22	0.044
3	0.033	10	0.038	17	0.035	23	0.033
4	0.044	11	0.045	18	0.033	24	0.032
5	0.033	12	0.039	19	0.038	25	0.040
6	0.036	13	0.043				

We can probably recognize if s and s' have been obtained using the same key of the shift cipher.

We had $s_0, \ldots, s_4$. Let s_i^ℓ be obtained from s_i by shift cipher using the key ℓ.

Then s_i^ℓ has been obtained from a text with the frequency of letters as in English, by applying the shift cipher with the key $k_i + \ell$.

For all i, j, ℓ check whether the keys of the shift cipher for obtaining s_i and s_j^ℓ have been equal.

If yes, then $k_i = k_j + \ell$.

$MI_c(s_0, s_1^\ell)$:

0	0.039	7	0.038	14	0.050	20	0.031
1	0.042	8	0.046	15	0.069	21	0.046
2	0.044	9	0.031	16	0.033	22	0.044
3	0.032	10	0.027	17	0.035	23	0.030
4	0.042	11	0.044	18	0.043	24	0.031
5	0.030	12	0.032	19	0.037	25	0.042
6	0.036	13	0.027				

Probably $k_0 = k_1 + 15$.

$MI_c(s_0, s_2^\ell)$:

0	0.027	7	0.027	14	0.055	20	0.042
1	0.039	8	0.040	15	0.051	21	0.054
2	0.055	9	0.033	16	0.034	22	0.037
3	0.038	10	0.042	17	0.049	23	0.036
4	0.033	11	0.029	18	0.036	24	0.044
5	0.033	12	0.029	19	0.029	25	0.035
6	0.026	13	0.046				

We can't be sure of the value $k_0 - k_2$. Maybe its 2 or 14 or 21... or maybe 15.

$MI_c(s_0, s_3^\ell)$:

0	0.049		7	0.074		14	0.049		20	0.045
1	0.027		8	0.027		15	0.029		21	0.035
2	0.032		9	0.034		16	0.030		22	0.041
3	0.048		10	0.041		17	0.040		23	0.027
4	0.030		11	0.039		18	0.058		24	0.029
5	0.029		12	0.030		19	0.034		25	0.044
6	0.037		13	0.041						

Probably $k_0 = k_3 + 7$.

$MI_c(s_0, s_4^\ell)$:

0	0.052	7	0.039	14	0.038	20	0.038
1	0.035	8	0.028	15	0.045	21	0.030
2	0.044	9	0.039	16	0.032	22	0.034
3	0.035	10	0.037	17	0.036	23	0.028
4	0.045	11	0.030	18	0.026	24	0.038
5	0.033	12	0.036	19	0.041	25	0.047
6	0.053	13	0.062				

It is reasonable to guess that $k_0 = k_4 + 13$.

$MI_c(s_2, s_4^\ell)$:

0	0.044	7	0.032	14	0.031	20	0.028
1	0.042	8	0.030	15	0.045	21	0.033
2	0.038	9	0.033	16	0.045	22	0.042
3	0.028	10	0.045	17	0.048	23	0.030
4	0.031	11	0.068	18	0.044	24	0.034
5	0.040	12	0.056	19	0.029	25	0.039
6	0.030	13	0.036				

Probably $k_2 = k_4 + 11$.

$$k_0 = k_1 + 15$$

$$k_0 = k_3 + 7$$

$$k_0 = k_4 + 13$$

$$k_2 = k_4 + 11$$

Possible keys are “zkxsm” and all words that can be obtained by shifting its letters. These are:

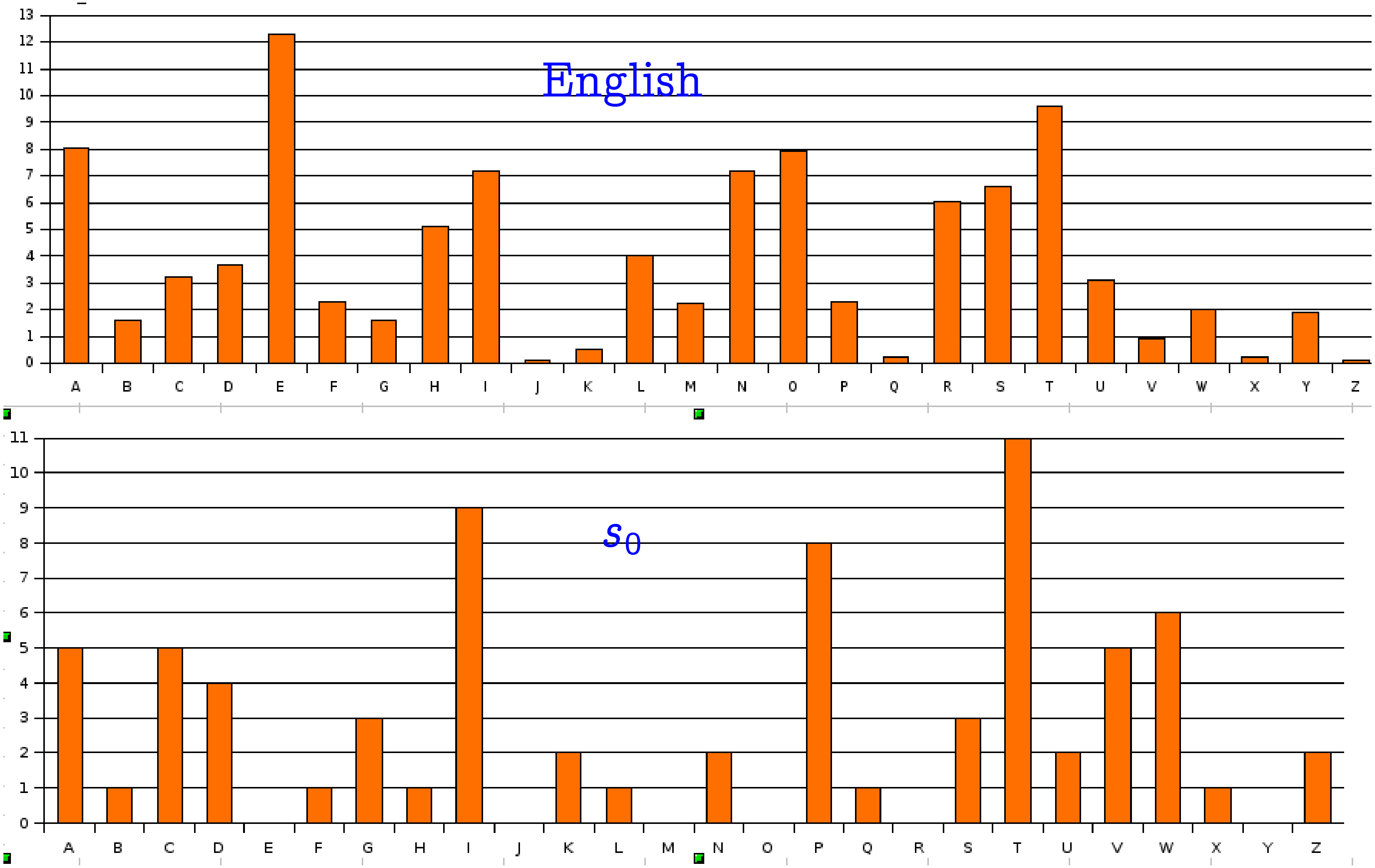
alytn, bmzuo, cnavp, dobwq, epcxr, fqdys, grezt, hsfau, itgbv, juhcw, kvidx, lwjey, mxkfz, nylga, ozmhb, panic, qbojd, rcpke, sdqlf, termg, ufsnh, vgtoi, whupj, xivqk, yjwrl

Let us try them all.

The key “panic” gives

He looked about the cabin but could see very little; strange monstrous shadows loomed and leaped with the tiny flickering flame, but all was quiet. He breathed a silent thank you to the Dentrassis. The Dentrassis are an unruly tribe of gourmands, a wild but pleasant bunch whom the Vogons had recently taken to employing as catering staff on their long haul fleets, on the strict understanding that they keep themselves very much to themselves.

Another way: consider



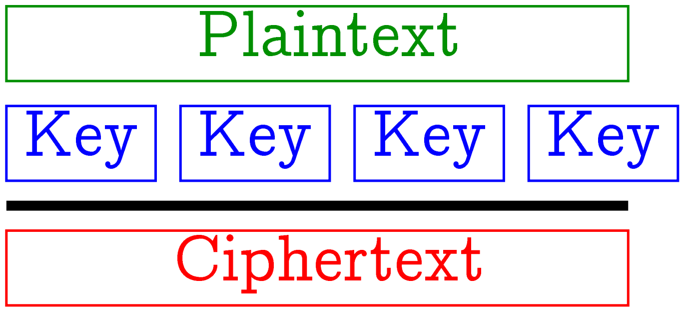
To find k_0, shift the lower chart to match the upper chart as well as possible.

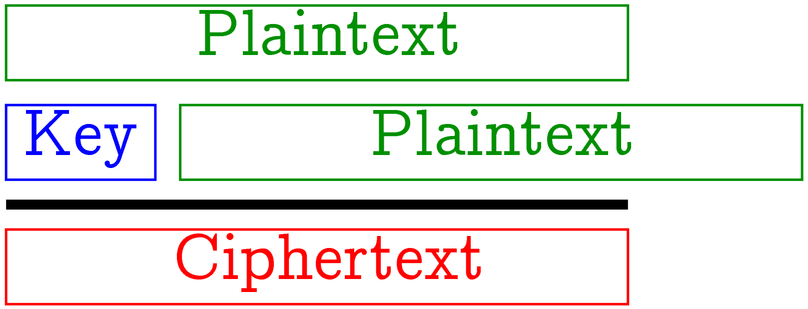
The quality of match can again be expressed by the index of mutual coincidence.

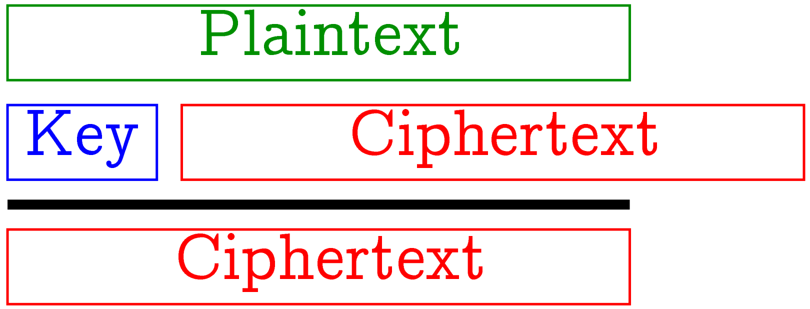
- Let p_i be the frequency of letter i in English.
- Let p'_i be the frequency of letter i in s_0.

Find ℓ that maximises

$$\sum_{i=0}^{25} p_i p'_{(i+\ell) \bmod 26}$$







- Vigenère or key autokey cipher (above).
- Text autokey cipher (two variants) (below).

Exercise. One of those two variants has serious problems. Which one? Break DCOWRWBZKDFJOBQNBHJU

Exercise. How to break the “good” variant? Assume we know the key length. How to derive “subsequences” where the letter frequency is similar to English?