Cryptology I

(MTAT.07.002, 6 EAP)

Lectures:	Mon 12-14	hall 404	and	Tue 12-14	hall 403
Exercises:	Mon 14-16	hall 315	and	Wed 10-12	hall 405

homepage:

http://www.ut.ee/~peeter_l/teaching/kryptoi09s

(contains lecture materials)

For grade: exercises at home, during the midterm and the final 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.

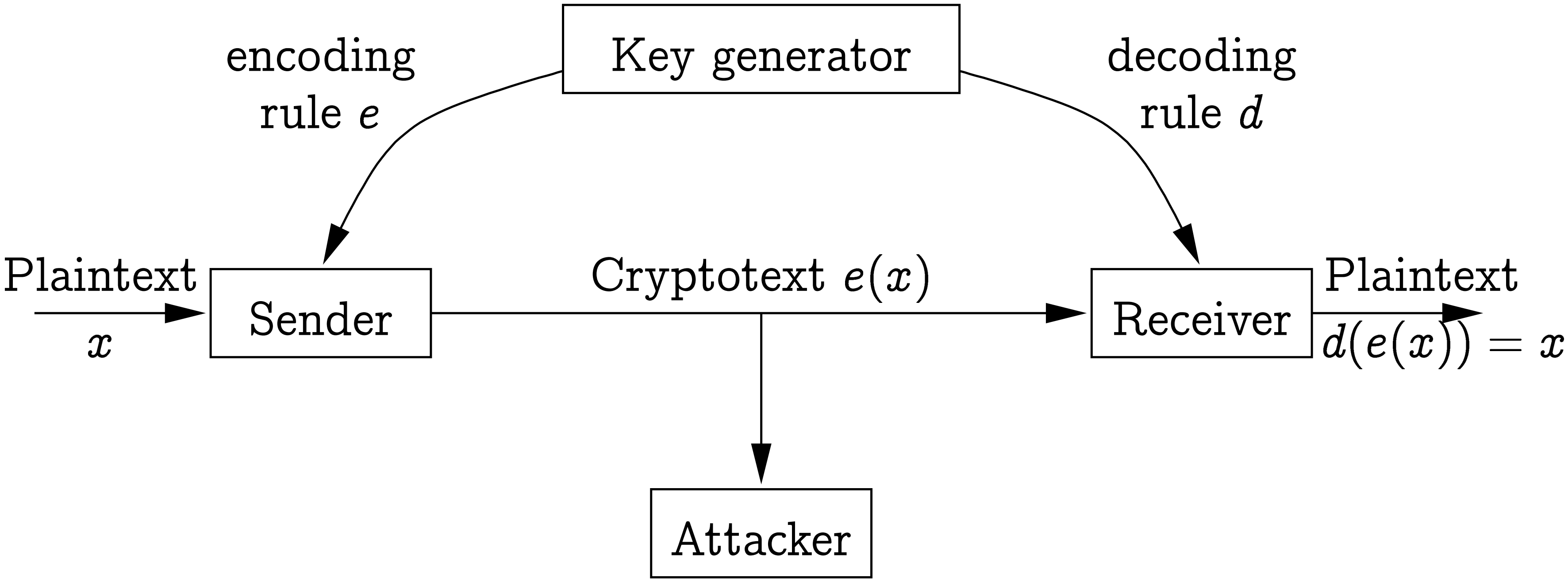
- (not speaking about availability)

Cryptography: Mathematical methods for ensuring system's security.

Cryptanalysis: Mathematical methods for breaking cryptography.

Cryptology: Cryptography and cryptanalysis.

Encryption and decryption:

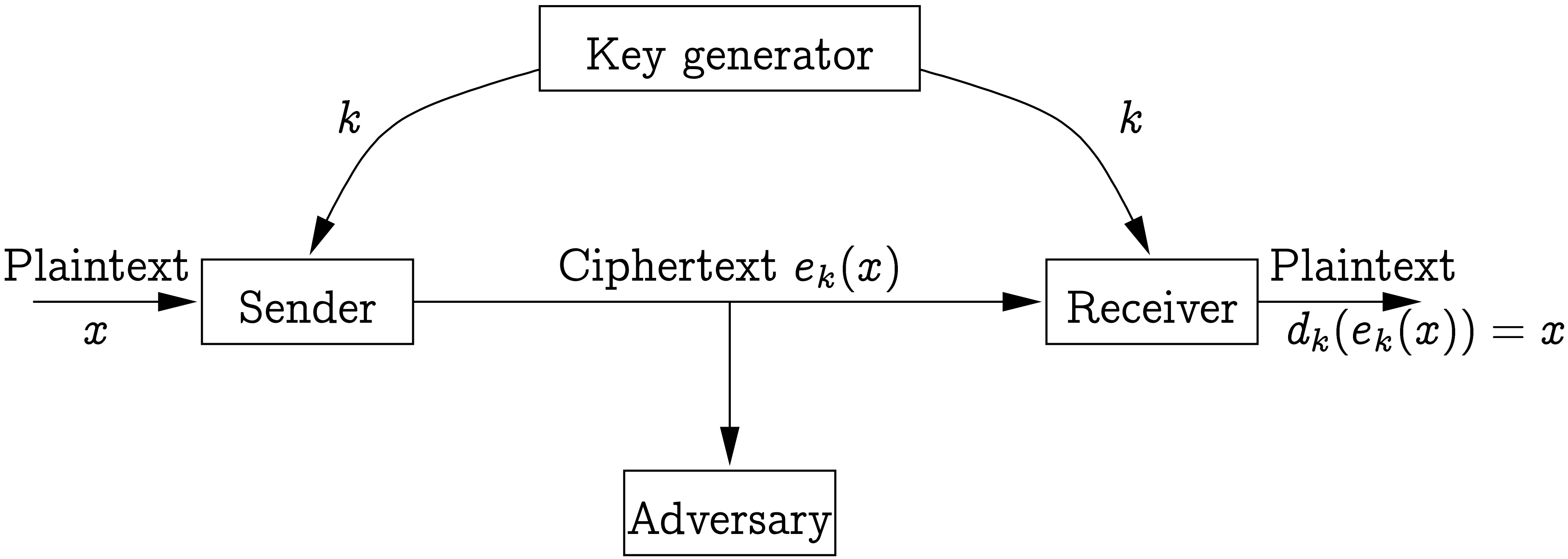


encoding- and decoding rules should have short descriptions.

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 $\mathcal{P}$.
  - $\forall x \in \mathcal{P} : d_k(e_k(x)) = x$

Encryption and decryption



k describes the rules for encryption and decryption.

Ancient greeks, esp. Spartans used as an encryption system a tool called *σκυτάλη* (*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 try out 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.

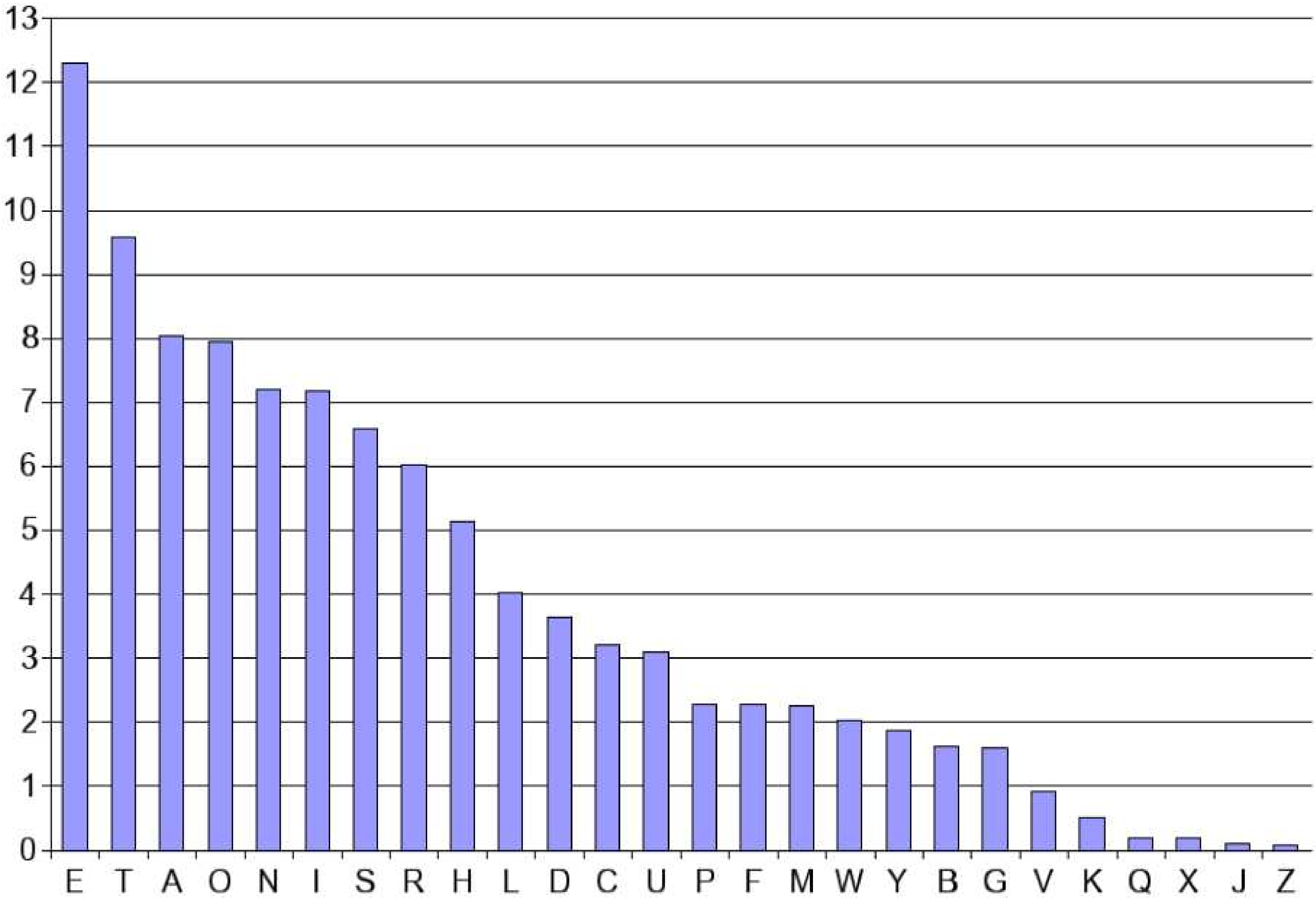
Exercise. How to encode a key of the substitution cipher in as few bits as possible?

We can break the substitution cipher 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		

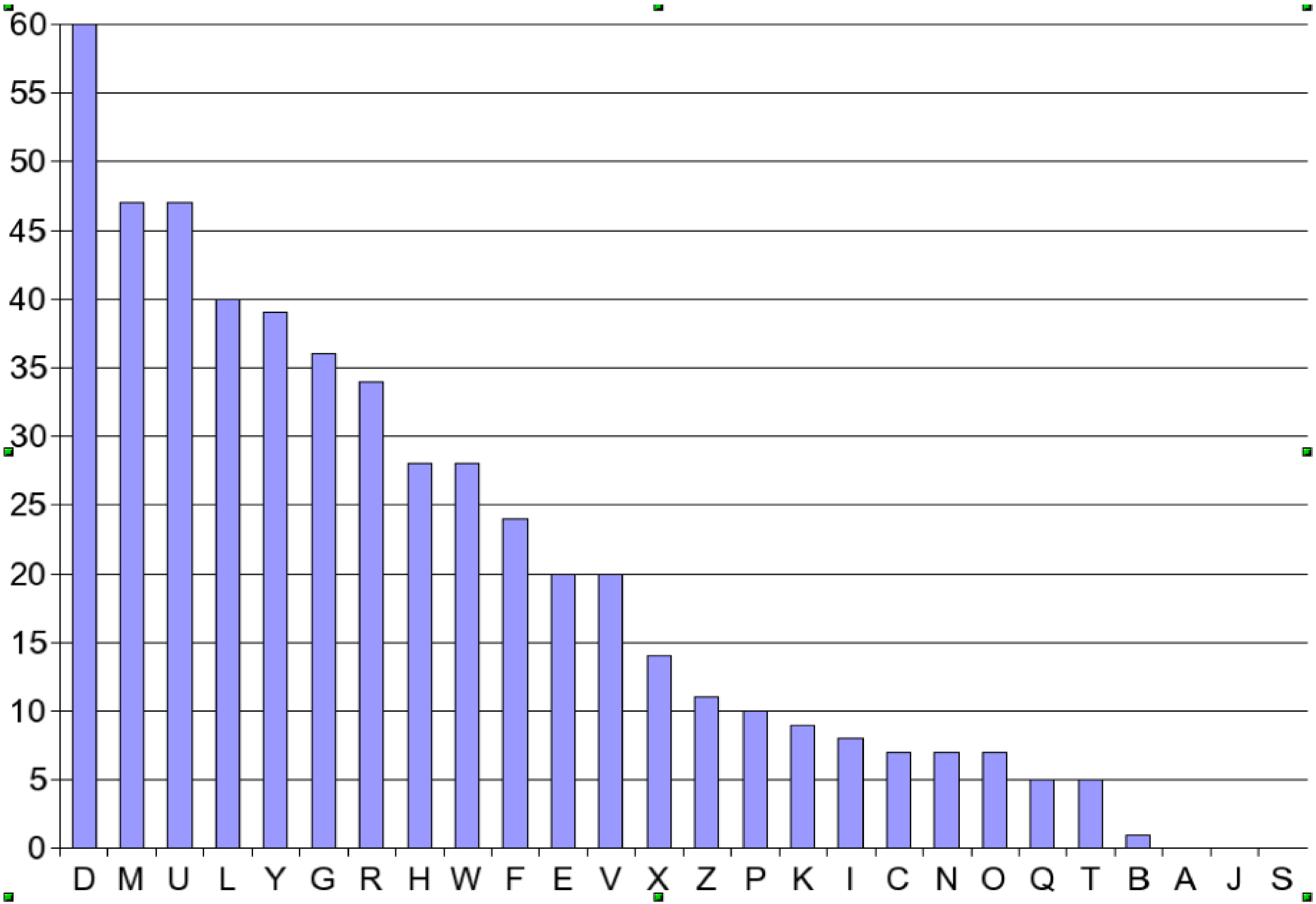
Most common trigraphs (descending): THE, ING, AND, HER, ERE, ENT, THA, NTH, WAS, ETH, FOR, DTH.

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.

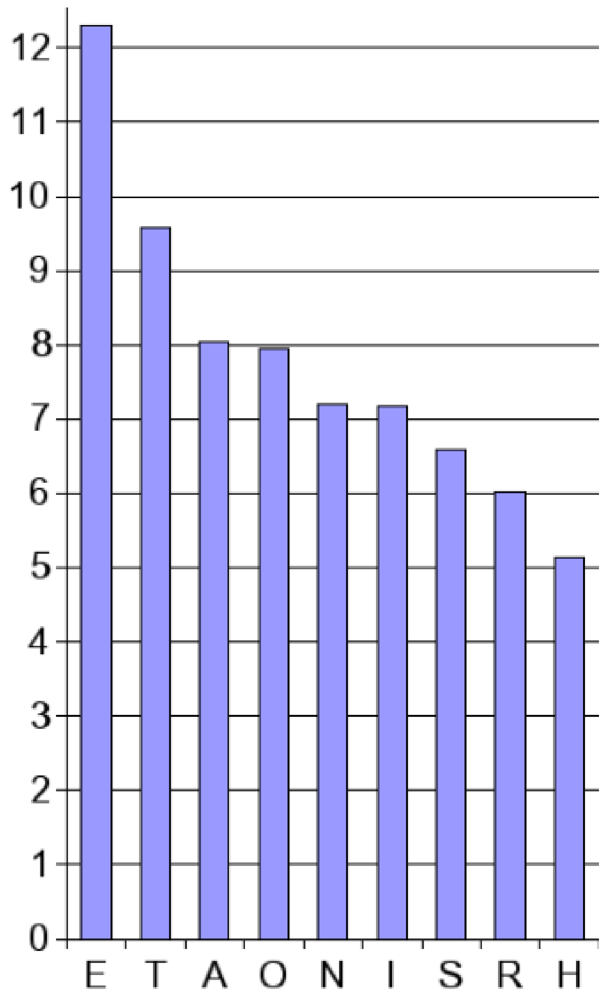
First step: count the number of occurrencies of each letter.

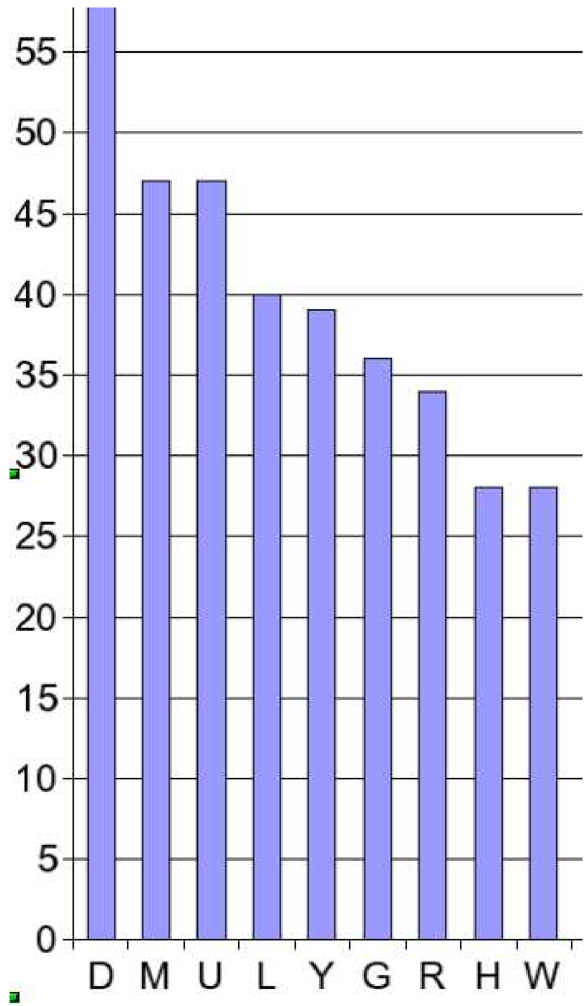
a	0	h	28	o	7	u	47
b	1	i	8	p	10	v	20
c	7	j	0	q	5	w	28
d	60	k	9	r	34	x	14
e	20	l	40	s	0	y	39
f	24	m	47	t	5	z	11
g	36	n	7				



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.


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


55
50
45
40
35
30
25
20
15
10
5
0
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...

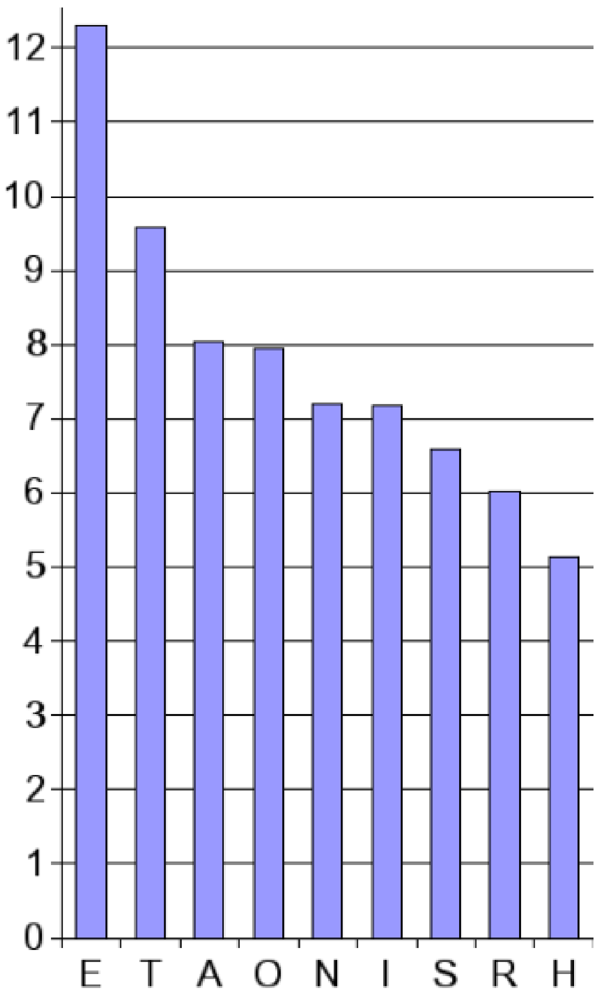
cryptotext plaintext

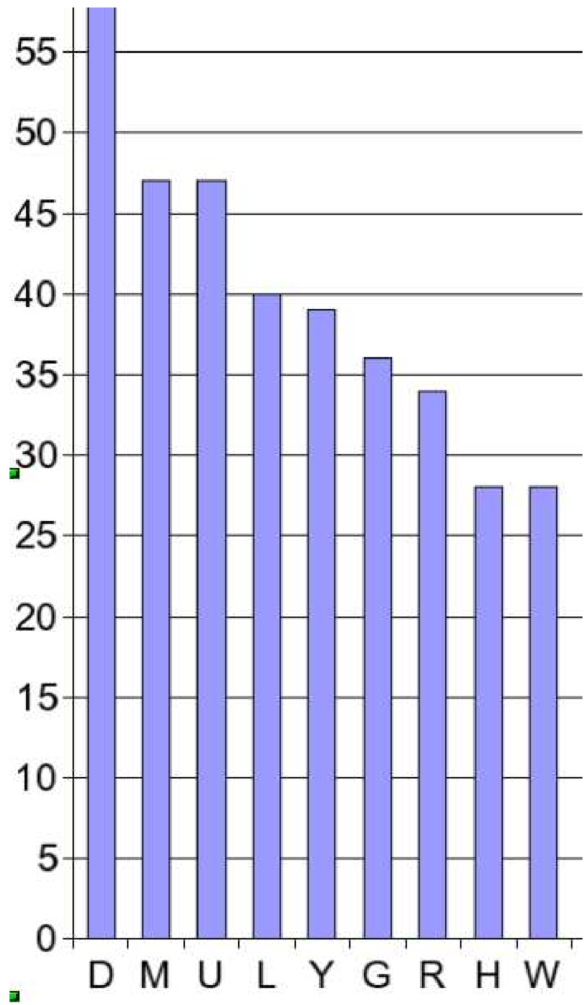
my	16	*yg*	9
yd	13	*yu*	9
df	11	*rd*	8
gw	11	*gm*	7
ur	11	*lh*	7
vu	11	*lx*	7
		xr	7

th	1.52	*at*	0.59
he	1.28	*on*	0.57
in	0.94	*nt*	0.56
er	0.94	*ha*	0.56
an	0.82	*es*	0.56
re	0.68	*st*	0.55
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.

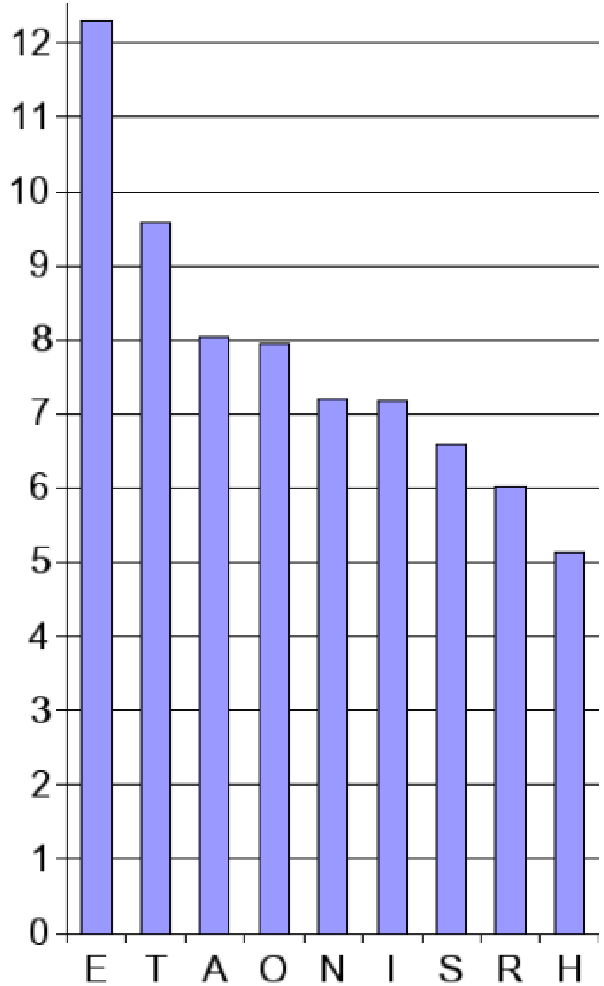
The lwez oerrlw ilr vhlp the hlxre var gw awz vaz roengae var Arthxr Fewt, awf that var lwez kenaxre gt haooewef tl ke the lwe he egcef gw. He haf egcef gw gt ilr aklxt three zearr, ecer rgwne he haf plcef lxt li Elwflw kenaxre gt pafe hgp werclxr awf grrgtakee. He var aklxt thgrtz ar veee, farq hagref 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)

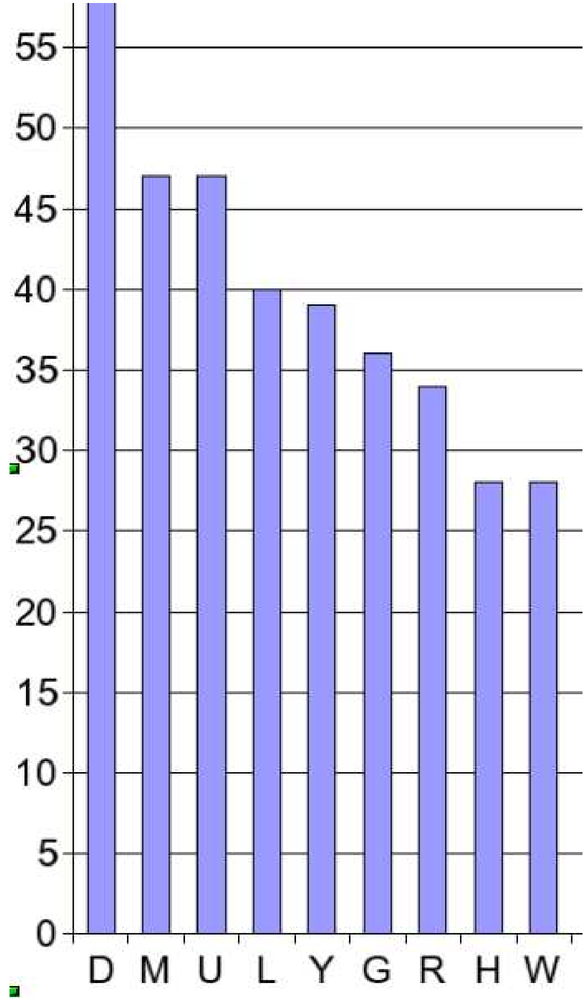
The lwez oerrlw ilr vhlp the hlure var gw awz vaz roengae var Arthur Fewt, awf that var lwez kenaure gt haooewef tl ke the lwe he egcef gw. He haf egcef gw gt ilr aklut three zearr, ecer rgwne he haf plcef lut li Elwflw kenaure gt pafe hgp werclur awf grrgtakee. He var aklut thgrtz ar veee, farq hagref 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. Gt var, tll - 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 iwterertiwt thaw thez orokakez thoutht. It var, too - port oi hir iriewfr vorqef iw afcertiriwt.

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 iwterertiwt thaw thez orokakez thoutht. It var, 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:

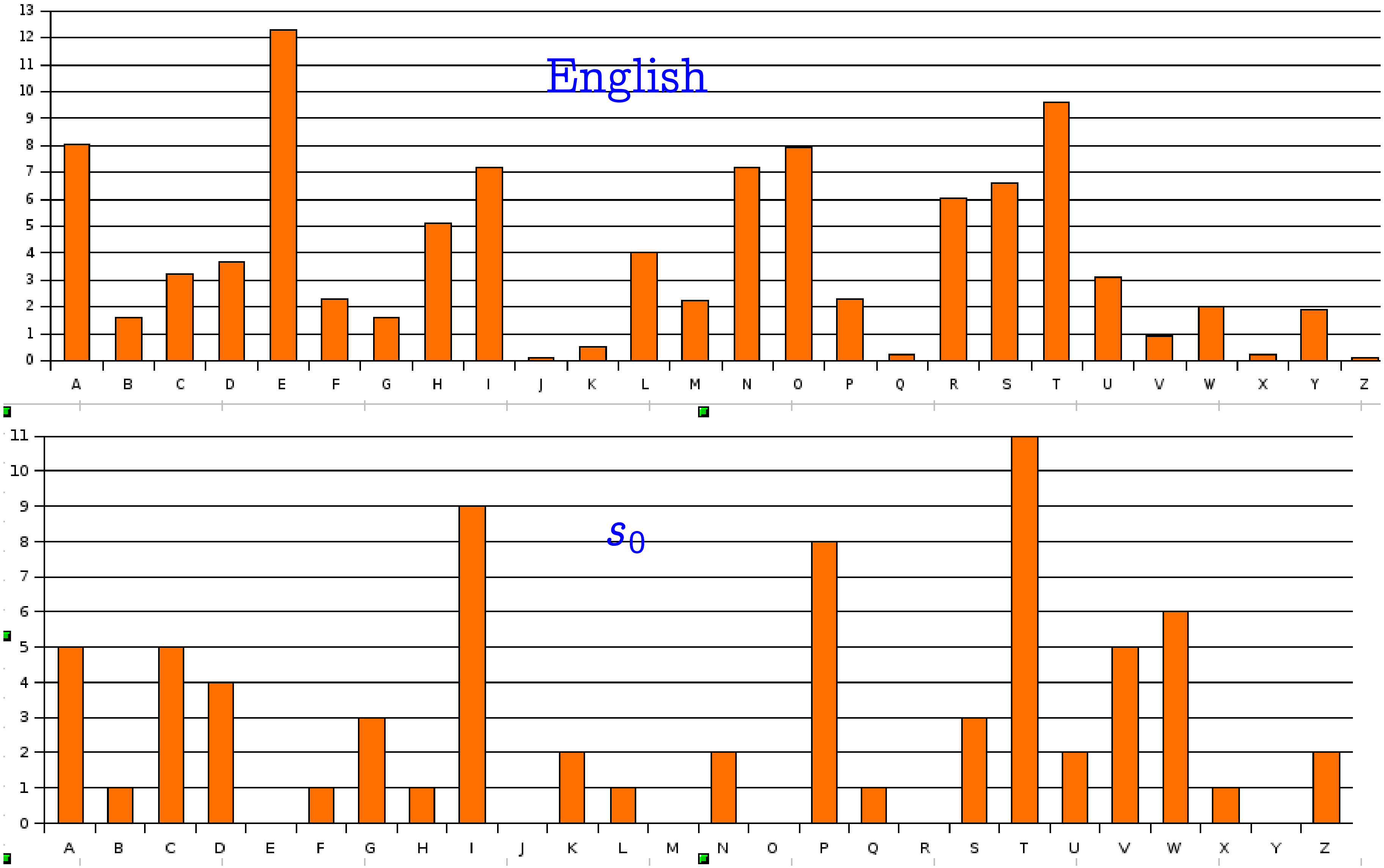
alytn, bmzuo, cnavp, dobwq, epcxr, fqdys, grezt, hsfau, itgbv, juhcw, kvidx, lwjey, mxkfz, nylga, ozmhb, panic, qbojd, rcpke, sdqlf, termg, ufsnh, vgtoi, whupj, xivqk, yjwrl

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



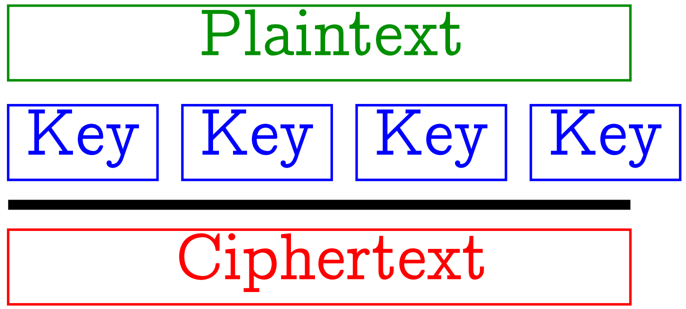
To find k_0, shift the lower chart to match the upper chart as well as possible.

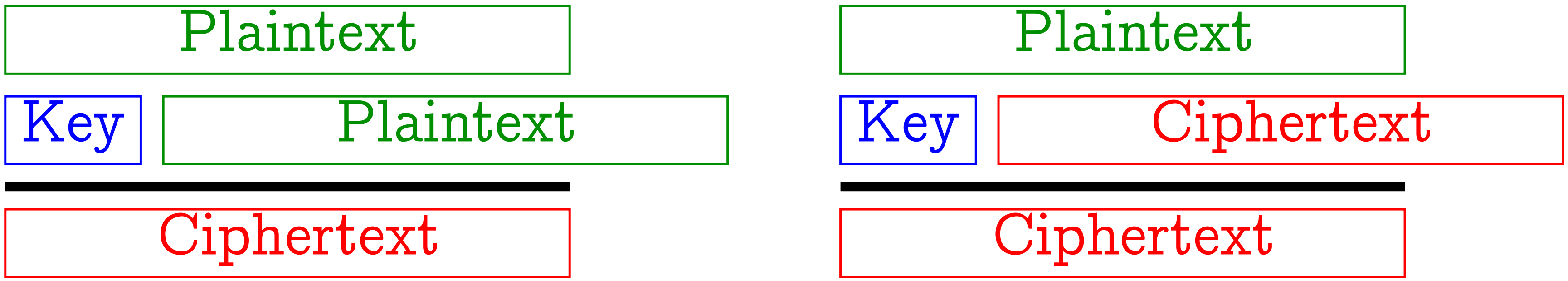
The quality of match can again be expressed by the index of mutual coincidence.

- 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?

Hill's cipher

- Key: a number m and an invertible sqare matrix $M \in \mathbb{Z}_{26}^{m \times m}$.
- Encoding: split the text to sequences of length m. The ciphertext corresponding to $x \in \mathbb{Z}_{26}^{m}$ is $x \cdot M$.
- Decoding: the plaintext corresponding to the ciphertext $y \in \mathbb{Z}_{26}^{m}$ is $y \cdot M^{-1}$.

Example: let $m = 3$ and

$$M = \begin{pmatrix} 15 & 2 & 13 \\ 8 & 21 & 1 \\ 14 & 16 & 7 \end{pmatrix} .$$

Then $\det M \equiv 9 \pmod{26}$, i.e. M is invertible in $\mathbb{Z}_{26}^{3\times 3}$ (because 9 is invertible in $\mathbb{Z}_{26}$).

Let the plaintext be CRYPTOGRAPHY or
$(2, 17, 24), (15, 19, 14), (6, 17, 0), (15, 7, 24)$.

Multiplying all these four vectors with M (from the right) gives us the ciphertext $(8, 17, 3), (1, 3, 0), (18, 5, 17), (19, 15, 6)$
or
IRDBDASFRTPG.

To decode, let us find M^{-1}...

$$\left(\begin{array}{ccc|ccc} 15 & 2 & 13 & 1 & 0 & 0 \\ 8 & 21 & 1 & 0 & 1 & 0 \\ 14 & 16 & 7 & 0 & 0 & 1 \end{array}\right) \to \left(\begin{array}{ccc|ccc} 1 & 14 & 13 & 7 & 0 & 0 \\ 8 & 21 & 1 & 0 & 1 & 0 \\ 14 & 16 & 7 & 0 & 0 & 1 \end{array}\right) \to$$

Multiplied the first row with $7 = 15^{-1}$.

$$\left(\begin{array}{ccc|ccc} 1 & 14 & 13 & 7 & 0 & 0 \\ 0 & 13 & 1 & 22 & 1 & 0 \\ 0 & 2 & 7 & 6 & 0 & 1 \end{array}\right) \to \left(\begin{array}{ccc|ccc} 1 & 14 & 13 & 7 & 0 & 0 \\ 0 & 1 & 11 & 12 & 1 & 20 \\ 0 & 2 & 7 & 6 & 0 & 1 \end{array}\right) \to$$

Added the right multiples of the first row to the second and third rows. Then subtracted the sixfold third row from the second.

$$\left(\begin{array}{ccc|ccc} 1 & 14 & 13 & 7 & 0 & 0 \\ 0 & 1 & 11 & 12 & 1 & 20 \\ 0 & 0 & 11 & 8 & 24 & 13 \end{array}\right) \to \left(\begin{array}{ccc|ccc} 1 & 14 & 13 & 7 & 0 & 0 \\ 0 & 1 & 11 & 12 & 1 & 20 \\ 0 & 0 & 1 & 22 & 14 & 13 \end{array}\right) \to$$

$$\left(\begin{array}{ccc|ccc} 1 & 14 & 0 & 7 & 0 & 13 \\ 0 & 1 & 0 & 4 & 3 & 7 \\ 0 & 0 & 1 & 22 & 14 & 13 \end{array}\right) \rightarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 3 & 10 & 19 \\ 0 & 1 & 0 & 4 & 3 & 7 \\ 0 & 0 & 1 & 22 & 14 & 13 \end{array}\right)$$

Added the multiples of the third row to the first and second row. Then added the multiple of the second row to the first row. Hence

$$M^{-1} = \begin{pmatrix} 3 & 10 & 19 \\ 4 & 3 & 7 \\ 22 & 14 & 13 \end{pmatrix}$$

To decode, the vectors making up the ciphertext must be multiplied with M^{-1} from the right.

$(8, 17, 3) \cdot M^{-1} = (2, 17, 24)$, etc.

Types of attacks against encryption systems

- ciphertext-only (*tuntud krüptotekstiga*)
  - Given a ciphertext, find the plaintext and/or the key.
- known-plaintext (*tuntud avatekstiga*)
  - The attacker knows a number of plaintext-ciphertext pairs. With their help, find the key or the plaintext corresponding to some other ciphertext.
- chosen-plaintext (*valitud avatekstiga*)
  - The attacker can invoke the encoding function. Find the key or the plaintext.
- chosen-ciphertext (*valitud krüptotekstiga*)
  - The attacker can invoke the decoding function. Find the key or the plaintext. The decoding function may not be invoked on the ciphertext to decode.

Known-plaintext attack on Hill's cipher

Let m be known (if not, guess). let (x_i, y_i) be the pairs of known plaintext-ciphertext pairs corresponding to an unknown key. I.e. $y_i = x_i \cdot M$.

- Let $x_{i_1}, \ldots, x_{i_m}$ be linearly independent plaintexts.
- Let X be a matrix with the rows $x_{i_1}, \ldots, x_{i_m}$.
- Let Y be the matrix with the rows $y_{i_1}, \ldots, y_{i_m}$.
- $Y = X \cdot M$, hence $M = X^{-1} \cdot Y$.
- If m was unknown then we can use the other plaintext-ciphertext pairs to verify the correctness of M.

Exercises

- What is the number of $m \times m$-keys of Hill's cipher?
- A square matrix M is involutory if $M = M^{-1}$. Mr. Hill himself suggested using an involutory matrix as a key. How many $m \times m$ involutory matrices exist?
  - Why would Hill have suggested so? Hint: he proposed this cipher in 1929.

Affine Hill's cipher

Hill's cipher is just a linear transformation of $\mathbb{Z}_{26}^m$.

A more general form of it is:

- Key: $m \in \mathbb{N}$, $M \in \mathbb{Z}_{26}^{m \times m}$, $v \in \mathbb{Z}_{26}^m$, such that M is invertible.

- Encryption of $x \in \mathbb{Z}_{26}^m$ is $x \cdot M + v$.

- Decryption of $y \in \mathbb{Z}_{26}^m$ is $y \cdot M^{-1} - v$.

Exercises

- How to do a known-plaintext attack on affine Hill's cipher (assuming that m is known)?
  - How many plaintext-ciphertext pairs we need if everything necessary turns out to be linearly independent?
- If M in the key of the affine Hill's cipher is the unit matrix, what sort of cryptosystem results?

More exercises

- How resistant are Caesar cipher (a.k.a. shift cipher, *nihkešiffer*), substitution cipher (*asendusšiffer*) and Vigenère cipher against known-plaintext and chosen-plaintext attacks?

- How much corresponding plaintext and ciphertext is needed for a known-plaintext attack on a multiply applied Vigenère cipher, if the number of keys and their lengths are known?

Affine cipher

If $m = 1$ in affine Hill's cipher, then the result is called just the affine cipher.

In an affine cipher

- $\mathcal{K} = \mathbb{Z}_{26}^* \times \mathbb{Z}_{26}$;
- $e_{(k,a)}(x) = k \cdot x + a \bmod 26$ for a character x;
- $d_{(k,a)}(y) = (y - a) \cdot k^{-1} \bmod 26$ for a character y.

(to encrypt a text: encrypt each character separately)

known-plaintext cryptanalysis

It is usually sufficient to have two pairs (x_1, y_1), (x_2, y_2) of corresponding characters in plaintext and ciphertext.

Then

$$\begin{cases} y_1 = x_1 \cdot k + a \\ y_2 = x_2 \cdot k + a \end{cases} \Longrightarrow (y_1 - y_2) = (x_1 - x_2) \cdot k \Longrightarrow$$

$$k = (y_1 - y_2) \cdot (x_1 - x_2)^{-1} \text{ and } a = y_1 - x_1 \cdot k \pmod{26}$$

If $(x_1 - x_2)$ is not invertible in $\mathbb{Z}_{26}$ then we get several solutions for k.

Then we need more plaintext-ciphertext pairs.

Transposition cipher

- Key: $m \in \mathbb{N}$ and a permutation σ of $\{1, \ldots, m\}$.
- To encrypt a plaintext:
  - Write it down on rows, with m symbols per row.
    * Pad or do not pad the text, to make its length divisible by m.
  - Permute the resulting m columns according to σ.
  - Read out the ciphertext, row by row.
- To decrypt, do everything in reverse.
  - If the plaintext was unpadded, figure out which columns were taller.

Exercise: what is the relation between transposition cipher and Hill's cipher?

Example: let $m = 8$ and $\sigma = \frac{1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8}{3 \quad 5 \quad 2 \quad 7 \quad 4 \quad 1 \quad 6 \quad 8}$.

Let the plaintext be

THEFIRSTHOMEASSIGNMENTISDUEATTHETHURSDAYNEXTWEEK

T	H	E	F	I	R	S	T		R	E	T	I	H	S	F	T
H	O	M	E	A	S	S	I		S	M	H	A	O	S	E	I
G	N	M	E	N	T	I	S		T	M	G	N	N	I	E	S
								permuted:								
D	U	E	A	T	T	H	E		T	E	D	T	U	H	A	E
T	H	U	R	S	D	A	Y		D	U	T	S	H	A	R	Y
N	E	X	T	W	E	E	K		E	X	N	W	E	E	T	K

The ciphertext is

RETIHSFTSMHAOSEITMGNNIESTEDTUHAEDUTSHARYEXNWEETK

Cryptanalysis

- Recognizing transposition cipher: the letters in the ciphertext have the same frequency as in the plaintext.
- First, somehow guess the number of columns m.
- Write text in m columns (as by decryption) and look for anagrams.
  - Look for anagrams in rows, but also consider two rows (following each other) together.
- For example, the last row in the previous example was EXNWEETK.
  - Probably an anagram of NEXTWEEK.
  - This already fixes 5 of 8 rows.

Frequencies of di-, tri-, . . . -graphs

- Pick a column.
  - . . . with largest number of common characters.
- Put another column beside it; consider the sum of frequencies (in plaintext) of resulting bigrams.
  - Also consider row breaks; you may want to shift the other column a position up or down.
- The column with the largest such sum is the most probable neighbour.

- Using a substitution cipher and a transposition cipher together usually gives good results:
- Determining the plaintext characters for some (frequent) characters in the ciphertext does not reveal parts of words.
- Anagramming, or looking for frequent digraphs is hard if we do not know the alphabet.

Confusion and diffusion

A cipher provides good

- **diffusion** if the statistical structure of the plaintext leading to its redundancy is "dissipated" into long range statistics — into statistical structure involving long combinations of letters in the cryptotext.
- **confusion** if it makes the relation between the simple statistics of the cryptotext and simple description of the key a very complex and involved one.

(paraphrased from: Claude Shannon. *Communication Theory of Secrecy Systems*. Bell System Technical Journal **28**(4):656–715, 1949.)

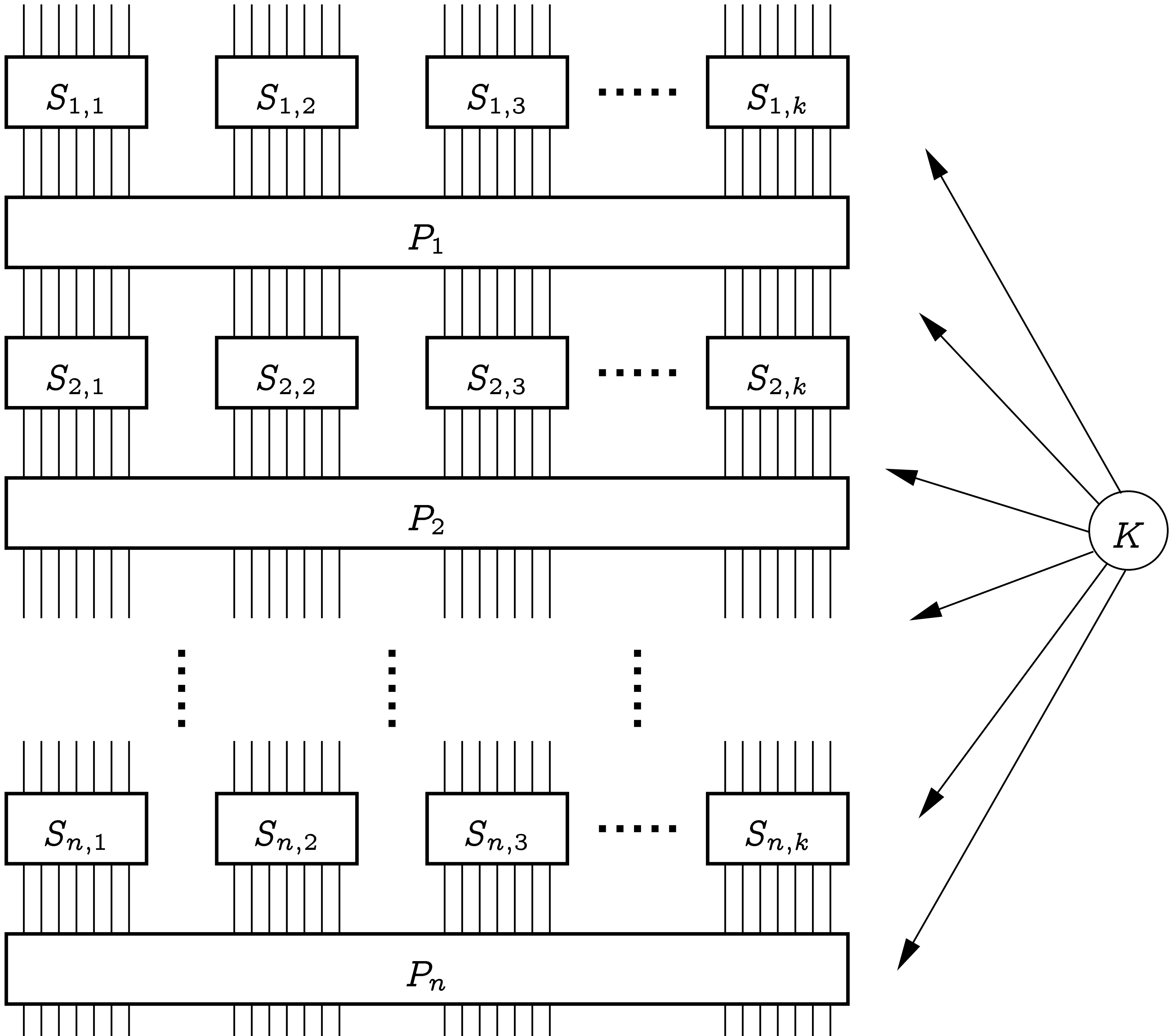
Achieving confusion and diffusion

- Diffusion is usually obtained by permuting the characters.
  - Or applying a more complex linear operation on long vectors of characters.
- Confusion is achived by substituting characters (or short sequences of them).

Iterating substitution and permutation may produce good ciphers.

Somewhere the key has to be mixed in, too.

Substitution-permutation network



Substitution gives good confusion

- When substitution cipher has been used, it is usually easy to find the cryptotext character corresponding to "E".
  - This maps a simple statistic of the cryptotext (counts of characters) to a simple property of the key.
- Maybe the cryptotext characters corresponding to some other frequent plaintext characters can be found this way, too.
- But for finding the rest of the substitution key, longer stretches of ciphertext have to be considered.
  - A simple property of the key can only be derived from a complex statistic of the ciphertext.
- This is confusion.

Fractionation

A character from the Latin alphabet does not have to be the “smallest unit” operated on by a cipher.

If we sacrifice a letter then we can encode each character in the plaintext as two elements of $\mathbb{Z}_5$.

This gives us a “plaintext” with $\mathbb{Z}_5$ as the alphabet.

We must have designed our cipher to work on $\mathbb{Z}_5^*$. We get the ciphertext as a string from $\mathbb{Z}_5^*$.

Optionally we may encode it back into Latin alphabet.

Instead of $\mathbb{Z}_5^2$ we may use $\mathbb{Z}_6^2$ (allowing us to encode Latin alphabet and numbers 0–9) or $\mathbb{Z}_3^3$ (allowing one extra symbol).

Fractionation helps to destroy frequency statistics.

Limits of pre-modern ciphers

- A combination of ciphers and techniques seen here can give us a quite strong cipher. But…
- Before the invention of computing machines, encryption and decryption had to be done by hand.
- The construction of a cipher had to be simple enough, such that this hand-operation produced reliable results even if performed in a stressful situation.
- For more complex ciphers, mechanical machines (like ENIGMA) were used.

A primer on algebra / number theory

Let S be a set. Let $\star : S \times S \to S$ be a function. Then $(S, \star)$ is a groupoid.

If $(a \star b) \star c = a \star (b \star c)$ for all $a, b, c \in S$ then $(S, \star)$ is a semigroup.

Let $\mathbf{1}$ be an element of the semigroup S.

If $\forall a \in S : a \star \mathbf{1} = a = \mathbf{1} \star a$ then $\mathbf{1}$ is a unit element.

A semigroup S that has a unit element is a monoid.

Let $\cdot^{-1} : S \to S$ be a function (S is a monoid).

If $a \star a^{-1} = a^{-1} \star a = \mathbf{1}$ for all $a \in S$, then S is a group.

If $a \star b = b \star a$ for all $a, b \in S$ (S is a group) then S is an Abelian group.

Theorem. The unit $\mathbf{1}$ and the inverse $\cdot^{-1}$ are unique.

Let $(S, \star, \ldots)$ be an algebraic structure. A substructure is a set $T \subseteq S$, such that all operations of S are closed on T.

- Applying the operations to elements of T gives elements of T.

Denote $T \leqslant S$.

Let $(G, \cdot, \mathbf{1}, \cdot^{-1})$ be a group and $H \subseteq G$. Then $H \leqslant G$ if

- $ab \in H$ for all $a, b \in H$;
- $\mathbf{1} \in H$;
- $a^{-1} \in H$ for all $a \in H$.

Theorem (Lagrange). Let G be a finite group and $H \leqslant G$. Then $|H|$ divides $|G|$.

$(R, +, \cdot)$ is a semiring if

- $(R, +)$ is a commutative monoid;
- $(R, \cdot)$ is a monoid;
- $\cdot$ distributes over $+$:
  $a \cdot (b + c) = ab + ac$ and $(a + b) \cdot c = ac + bc$.

A semiring R is a ring if $(R, +)$ is an Abelian group.

The multiplicative group R^* of a ring R is the set

$$\{a \in R \,|\, \exists b \in R : ab = ba = \mathbf{1}\}$$

together with the operation $\cdot$.

A ring R is a field if $R^* = R \backslash \{\mathbf{0}\}$.

Let G be a group and $g_1, \ldots, g_n \in G$. Let $\langle g_1, \ldots, g_n \rangle$ be the smallest set, such that

- $\mathbf{1} \in \langle g_1, \ldots, g_n \rangle$;
- $g_1, \ldots, g_n \in \langle g_1, \ldots, g_n \rangle$;
- If $a \in \langle g_1, \ldots, g_n \rangle$ then also $a^{-1} \in \langle g_1, \ldots, g_n \rangle$;
- If $a, b \in \langle g_1, \ldots, g_n \rangle$ then also $ab \in \langle g_1, \ldots, g_n \rangle$;

Theorem $\langle g_1, \ldots, g_n \rangle$ is a subgroup of G.

If $\langle g_1, \ldots, g_n \rangle = G$ then $g_1, \ldots, g_n$ generate G.

If $\exists g \in G$, such that $\langle g \rangle = G$, then G is cyclic.

The order of $g \in G$ is $|\langle g \rangle|$. It divides $|G|$ (if it is finite).

Let $a, b \in \mathbb{Z}$. We say that a divides b if $\exists k \in \mathbb{Z}$: $ak = b$.

- Write $a \mid b$ or $b \vdots a$.

$a, b \in \mathbb{Z}$ are congruent *modulo* $n \in \mathbb{Z} \backslash \{0\}$ if $(a - b) \mid n$.

- Write $a \equiv b \pmod{n}$.

For any $n \in \mathbb{Z} \backslash \{0\}$, the congruence *modulo* n is a congruence relation on $\mathbb{Z}$:

- reflexive, symmetric, transitive (i.e. equivalence)
- If $a \equiv b$ and $c \equiv d \pmod{n}$, then also $a + c \equiv b + d$, $ac \equiv bd$ and $-a \equiv -b \pmod{n}$.

Let $\mathbb{Z}_n$ be the set of equivalence classes of $\cdot \equiv \cdot \pmod{n}$.

$|\mathbb{Z}_n| = n$. Denote the class containing $k \in \mathbb{Z}$ with $\overline{k}$.

One can define operations on $\mathbb{Z}_n$ through the operations on $\mathbb{Z}$.

- works, because $\equiv$ is a congruence

$\mathbb{Z}_n$ together with the defined operations is a ring.

$\overline{a}$ is invertible in $\mathbb{Z}_n$ iff a and n are coprime (denote $a \perp n$).

$\mathbb{Z}_n$ is a field iff n is a prime.

- $(\mathbb{Z}_n, +)$ is a cyclic group.
  - Generated by 1.
- If n is a prime then $(\mathbb{Z}_n^*, \cdot)$ is a cyclic group.
  - (the multiplicative group of any finite field is cyclic)

Exercise. If G is a finite cyclic group and $|G| = n$ then how many $g \in G$ are there, such that g generates G?

A common divisor of some $a, b \in \mathbb{Z}$ is a $d \in \mathbb{Z}$, such that $d \mid a$ and $d \mid b$.

A common divisor d of a and b is the greatest common divisor if for any common divisor d' of a and b we have $d' \mid d$.

Euclidean algorithm for finding $\gcd(a, b)$:

1. Let $a_0 = \max(|a|, |b|)$, $a_1 = \min(|a|, |b|)$.
2. Let $a_{i+1} = a_{i-1} \bmod a_i$ for $i = 1, 2, \ldots$
   - Stop when $a_{n+1} = 0$ for some n.
3. Return a_n.

Theorem. For all $a, b \in \mathbb{Z}$ and $d = \gcd(a, b)$ there exist $u, v \in \mathbb{Z}$, such that $au + bv = d$.

Proof. (**Extended Euclidean Algorithm** (EEA)).

1. Assume $a \geqslant b > 0$. Let $a_0 = a$, $b_0 = b$, $u_1 = v_0 = 0$, $u_0 = v_1 = 1$.

2. For $i = 1, 2, \ldots$ do:

$$a_{i+1} = a_{i-1} \bmod a_i$$
$$u_{i+1} = u_{i-1} - u_i \cdot \lfloor a_{i-1}/a_i \rfloor$$
$$v_{i+1} = v_{i-1} - v_i \cdot \lfloor a_{i-1}/a_i \rfloor \quad .$$

   until $a_{n+1} = 0$.

3. Then $a_n = \gcd(a, b) = au_n + bv_n$.

Let $\overline{a} \in \mathbb{Z}_n$ and $a \perp n$. To find a^{-1} in $\mathbb{Z}_n$:

- Using EEA, find u, v, such that $au + nv = 1$.
- The answer is $\overline{u}$.
  - Because in $\mathbb{Z}_n$, $1 = au + nv = au + 0 \cdot v = au$.

Chinese remainder theorem

Theorem. Let $m_1, m_2, \ldots, m_r$ be pairwise coprime natural numbers and $a_1, a_2, \ldots, a_r$ some integers. The system of congruences

$$\begin{aligned} x &= a_1 \bmod m_1 \\ x &= a_2 \bmod m_2 \\ &\cdots \\ x &= a_r \bmod m_r \end{aligned}$$

has exactly one solution *modulo* $m_1 \cdot m_2 \cdot \ldots \cdot m_r$.

Proof. We'll find x as follows. Define

- $M = m_1 \cdot m_2 \cdot \ldots \cdot m_r$.
- $M_i = M/m_i$, $1 \leqslant i \leqslant r$.
- $M_i' = M_i^{-1} \pmod{m_i}$.
- $x = (M_1 M_1' a_1 + M_2 M_2' a_2 + \ldots + M_r M_r' a_r) \bmod M$.

Then $x \equiv M_i M_i' a_i \equiv a_i \pmod{m_i}$, because $M_j \equiv 0 \pmod{m_i}$, when $i \neq j$.

We showed that there exists at least one solution. There cannot be more than one, because a (different) solution exists for each of the possible tuples $(a_1, \ldots, a_r)$.

Euler's totient function φ

…is defined as

$$\varphi(n) := |\mathbb{Z}_n^*| = |\{x \in \mathbb{Z}_n : \gcd(x, n) = 1\}|.$$

Theorem. If $p \in \mathbb{P}$ and $e \in \mathbb{N}$, then

$$\varphi(p^e) = p^e - p^{e-1}.$$

What is $\varphi(n)$ for any $n \in \mathbb{N}$? Any n can be uniquely represented as the product of powers of its prime factors:

$$n = p_1^{e_1} \cdot p_2^{e_2} \cdot \ldots \cdot p_r^{e_r}.$$

Theorem. $\varphi(n) = (p_1^{e_1} - p_1^{e_1-1}) \cdot \ldots \cdot (p_r^{e_r} - p_r^{e_r-1})$.

This follows from

Lemma. If $\gcd(m, n) = 1$, then $\varphi(m \cdot n) = \varphi(m) \cdot \varphi(n)$.

$\varphi(m \cdot n) = \varphi(m) \cdot \varphi(n)$: example

Consider the case $n = 72$.

$$\begin{aligned}
\varphi(72) &= \varphi(8 \cdot 9) = \varphi(8) \cdot \varphi(9) = \\
&= \varphi(2^3) \cdot \varphi(3^2) = (2^3 - 2^2) \cdot (3^2 - 3^1) = \\
&= (8 - 4) \cdot (9 - 3) = 4 \cdot 6 = 24.
\end{aligned}$$

	0	1	2	3	4	5	6	7	8
0	0	64	56	48	40	32	24	16	8
1	9	1	65	57	49	41	33	25	17
2	18	10	2	66	58	50	42	34	26
3	27	19	11	3	67	59	51	43	35
4	36	28	20	12	4	68	60	52	44
5	45	37	29	21	13	5	69	61	53
6	54	46	38	30	22	14	6	70	62
7	63	55	47	39	31	23	15	7	71

Let $m \perp n$. There is is a "natural" isomorphism between $\mathbb{Z}_{mn}^*$ and $\mathbb{Z}_m^* \times \mathbb{Z}_n^*$:

$$x \in \mathbb{Z}_{mn}^* \mapsto (x \bmod m, x \bmod n)$$

$$(u, v) \in \mathbb{Z}_m^* \times \mathbb{Z}_n^* \mapsto \big(n \cdot (n^{-1} (\bmod m)) \cdot u + m \cdot (m^{-1} (\bmod n)) \cdot v\big) \bmod mn$$

Second row is just an application of CRT.