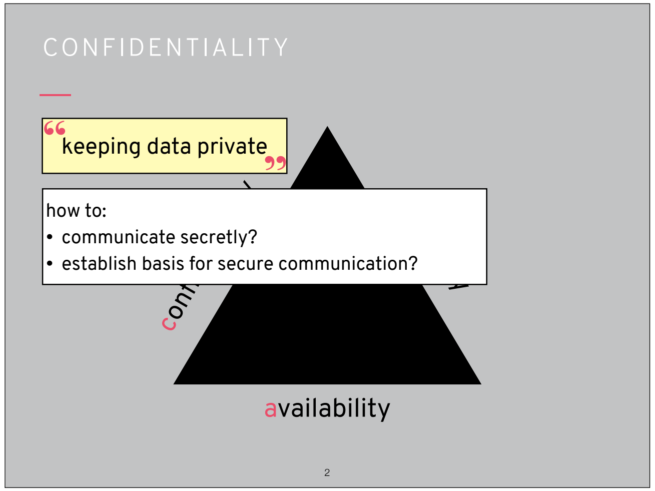
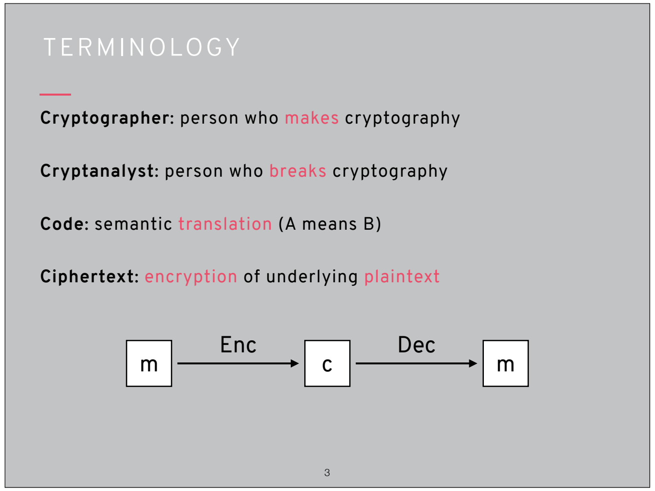
SECURITY (COMP0141): CONFIDENTIALITY



CONFIDENTIALITY

> “keeping data private”

how to:
- communicate secretly?
- establish basis for secure communication?



TERMINOLOGY

Cryptographer: person who makes cryptography

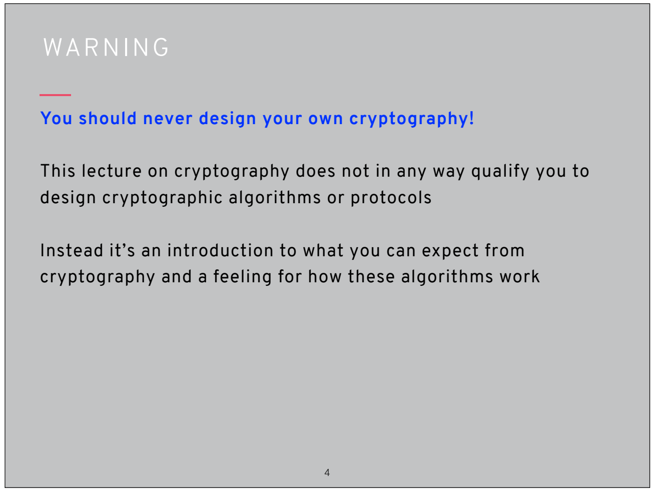
Cryptanalyst: person who breaks cryptography

Code: semantic translation (A means B)

Ciphertext: encryption of underlying plaintext

$$\begin{tikzcd} \boxed{m} \arrow[r, "Enc"] & \boxed{c} \arrow[r, "Dec"] & \boxed{m} \end{tikzcd}$$

Some terminology to get us on the same page. A code is not designed to hide any meaning, just to translate. In encryption, m represents a message (plaintext) that goes through an algorithm Enc to form a ciphertext c. This can be decrypted by putting c through an algorithm Dec to produce the message.



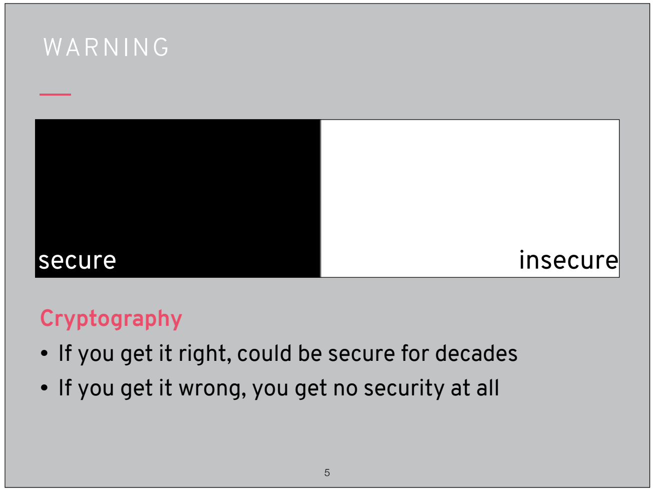
WARNING

You should never design your own cryptography!

This lecture on cryptography does not in any way qualify you to design cryptographic algorithms or protocols

Instead it’s an introduction to what you can expect from cryptography and a feeling for how these algorithms work

Don’t design your own crypto! Or at least don’t even deploy any crypto you designed yourself



WARNING

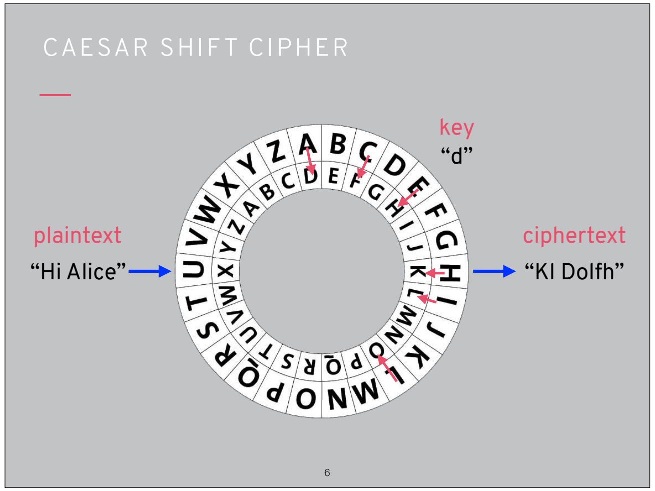
secure

insecure

Cryptography

- If you get it right, could be secure for decades
- If you get it wrong, you get no security at all

Why not? Goes back to idea of binary threat models, if you get the crypto wrong then you have no security at all



CAESAR SHIFT CIPHER

key

“d”

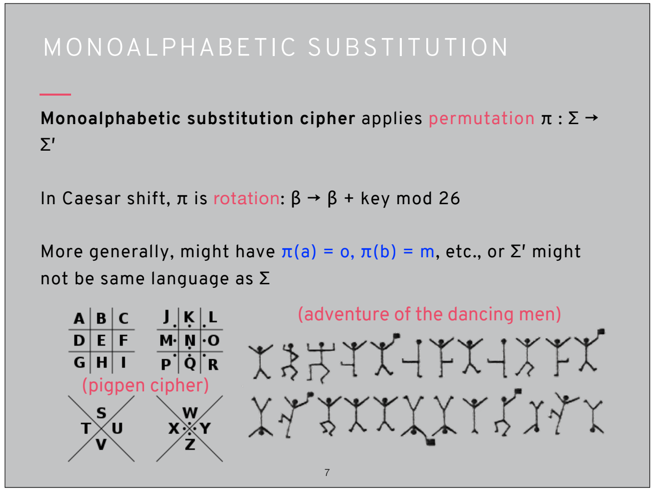
plaintext

“Hi Alice”

ciphertext

“Kl Dolfh”

If we use the key D then that means A maps to D. Using this rotation, H maps to K, so the first character of the ciphertext is K. The same is true for the rest of the plaintext: the character in the ciphertext is the one that the plaintext character is aligned with in the wheel.



MONOALPHABETIC SUBSTITUTION

Monoalphabetic substitution cipher applies permutation $\pi : \Sigma \rightarrow \Sigma'$

In Caesar shift, π is rotation: $\beta \rightarrow \beta$ + key mod 26

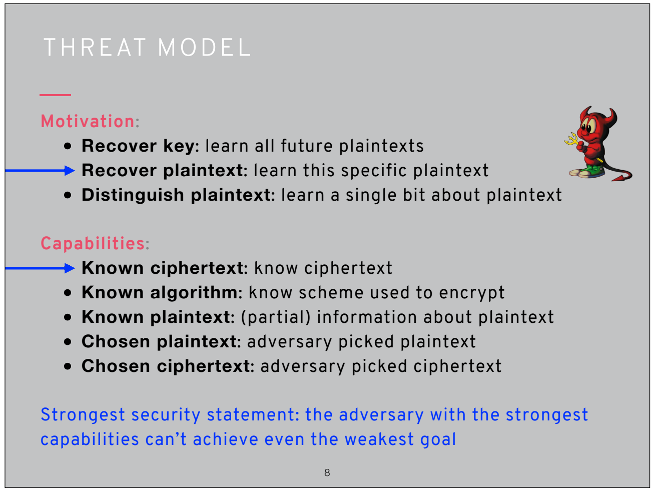
More generally, might have $\pi(a) = o$, $\pi(b) = m$, etc., or Σ' might not be same language as Σ

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

(pigpen cipher)

(adventure of the dancing men)

Caesar shift is thus a rotation, as you move a letter around the alphabet, and the number of times is dictated by the key (so T says rotate 20 times). More generally this is called a monoalphabetic substitution cipher: each character is replaced by a single other character (maybe in a different alphabet)



THREAT MODEL

Motivation:

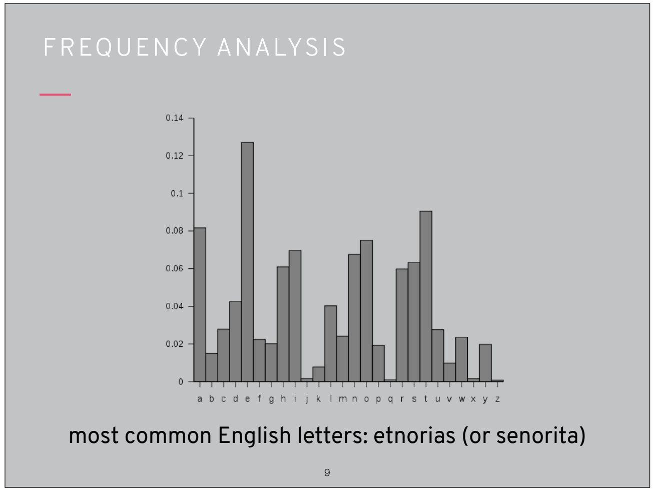
- **Recover key:** learn all future plaintexts
- **Recover plaintext:** learn this specific plaintext
- **Distinguish plaintext:** learn a single bit about plaintext

Capabilities:

- **Known ciphertext:** know ciphertext
- **Known algorithm:** know scheme used to encrypt
- **Known plaintext:** (partial) information about plaintext
- **Chosen plaintext:** adversary picked plaintext
- **Chosen ciphertext:** adversary picked ciphertext

Strongest security statement: the adversary with the strongest capabilities can't achieve even the weakest goal

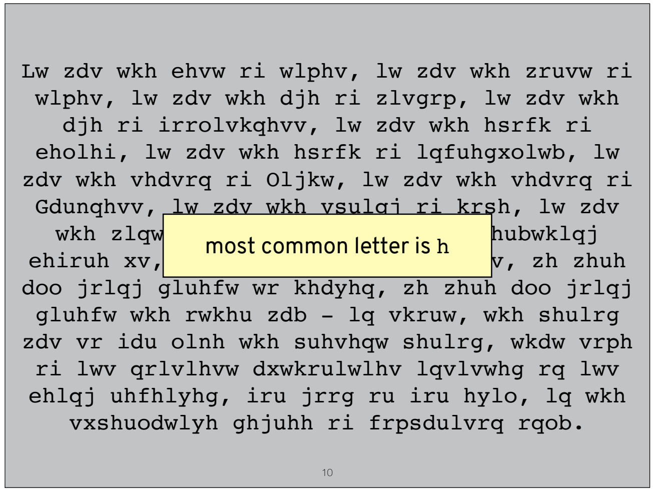
Go back to the idea of a threat model, and consider the different motivations and capabilities we should consider. For monoalphabetic substitution we'll see how even a weak attacker (one with few capabilities) can still break it



FREQUENCY ANALYSIS

most common English letters: etnorias (or senorita)

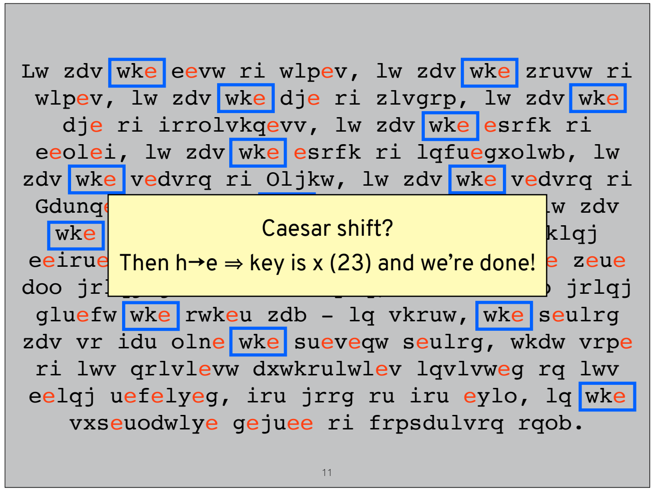
Monoalphabetic substitution ciphers can be pretty easily attacked because of frequency analysis



Lw zdv wkh ehvw ri wlphv, lw zdv wkh zruvw ri
wlphv, lw zdv wkh djh ri zlvgrp, lw zdv wkh
djh ri irrolvkqhvv, lw zdv wkh hsrfk ri
eholhi, lw zdv wkh hsrfk ri lqfuhgxolwb, lw
zdv wkh vhdvrq ri Oljkw, lw zdv wkh vhdvrq ri
Gdunqhvv, lw zdv wkh vsulqj ri krsh, lw zdv
wkh zlqw hubwklqj
ehiruh xv, v, zh zhuh
doo jrlqj gluhfw wr khdyhq, zh zhuh doo jrlqj
gluhfw wkh rwkhu zdb – lq vkruw, wkh shulrg
zdv vr idu olnh wkh suhvhqw shulrg, wkdw vrph
ri lwv qrlvlhvw dxwkrulwlhv lqvlvwhg rq lwv
ehlqj uhfhlyhg, iru jrrg ru iru hylo, lq wkh
vxshuodwlyh ghjuhh ri frpsdulvrq rqob.

most common letter is h

The next set of slides represent an exercise in cryptanalysis. If the most common letter in the ciphertext is h then what do we think it represents?



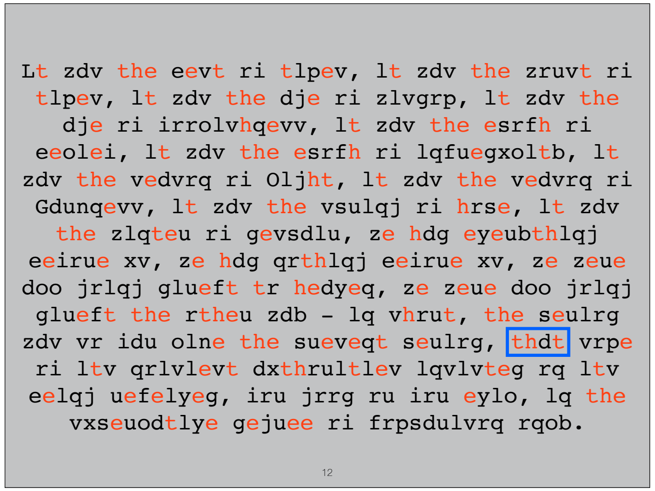
Lw zdv wke eevw ri wlpev, lw zdv wke zruvw ri
wlpev, lw zdv wke dje ri zlvgrp, lw zdv wke
dje ri irrolvkqevv, lw zdv wke esrfk ri
eeolei, lw zdv wke esrfk ri lqfuegxolwb, lw
zdv wke vedvrq ri Oljkw, lw zdv wke vedvrq ri
Gdunq lw zdv
wke klqj
eeirue e zeue
doo jr jrlqj
gluefw wke rwkeu zdb – lq vkruw, wke seulrg
zdv vr idu olne wke sueveqw seulrg, wkdw vrpe
ri lwv qrlvlevw dxwkrulwlev lqvlvweg rq lwv
eelqj uefelyeg, iru jrrg ru iru eylo, lq wke
vxseuodwlye gejuee ri frpsdulvrq rqob.

Caesar shift?

Then h→e ⇒ key is x (23) and we're done!

11

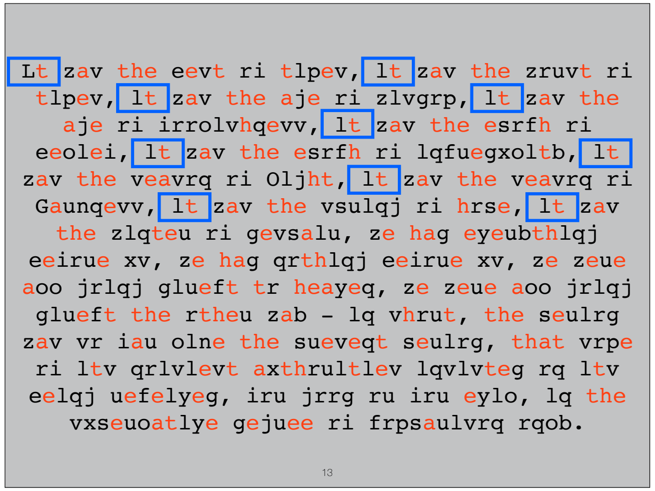
This isn't a Caesar shift so it's a little harder, we need to figure out the mapping bit by bit (whereas Caesar shift would give the whole thing at once)



Lt zdv the eevt ri tlpev, lt zdv the zruvt ri
tlpev, lt zdv the dje ri zlvgrp, lt zdv the
dje ri irrolvhqevv, lt zdv the esrfh ri
eeolei, lt zdv the esrfh ri lqfuegxoltb, lt
zdv the vedvrq ri Oljht, lt zdv the vedvrq ri
Gdunqevv, lt zdv the vsulqj ri hrse, lt zdv
the zlqteu ri gevsdlu, ze hdg eyeubthlqj
eeirue xv, ze hdg qrthlqj eeirue xv, ze zeue
doo jrlqj glueft tr hedyeq, ze zeue doo jrlqj
glueft the rtheu zdb – lq vhrut, the seulrg
zdv vr idu olne the sueveqt seulrg, thdt vrpe
ri ltv qrlvlevt dxthrultlev lqvlvteg rq ltv
eelqj uefelyeg, iru jrrg ru iru eylo, lq the
vxseuodtlye gejuee ri frpsdulvrq rqob.

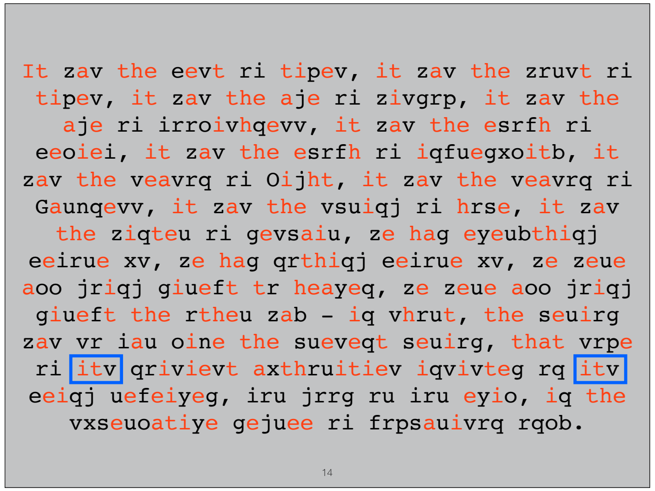
12

What word does this have to be?



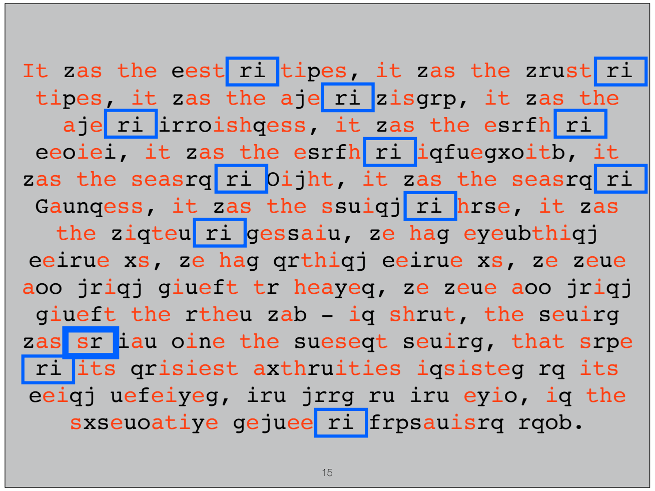
Lt zav the eevt ri tlpev, lt zav the zruvt ri
tlpev, lt zav the aje ri zlvgrp, lt zav the
aje ri irrolvhqevv, lt zav the esrfh ri
eeolei, lt zav the esrfh ri lqfuegxoltb, lt
zav the veavrq ri Oljht, lt zav the veavrq ri
Gaunqevv, lt zav the vsulqj ri hrse, lt zav
the zlqteu ri gevsalu, ze hag eyeubthlqj
eeirue xv, ze hag qrthlqj eeirue xv, ze zeue
aoo jrlqj glueft tr heayeq, ze zeue aoo jrlqj
glueft the rtheu zab – lq vhrut, the seulrg
zav vr iau olne the sueveqt seulrg, that vrpe
ri ltv qrlvlevt axthrultlev lqvlvteg rq ltv
eelqj uefelyeg, iru jrrg ru iru eylo, lq the
vxseuoatlye gejuee ri frpsaulvrq rqob.

And what about this? Need to consider the letters that we've already used



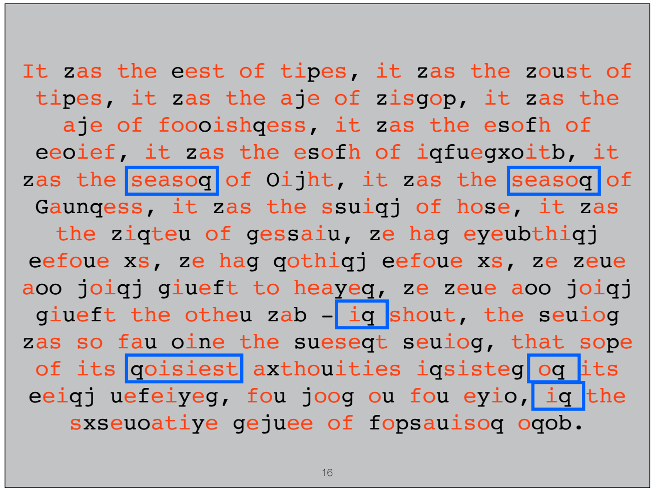
It zav the eevt ri tipev, it zav the zruvt ri
tipev, it zav the aje ri zivgrp, it zav the
aje ri irroivhqevv, it zav the esrfh ri
eeoiei, it zav the esrfh ri iqfuegxoitb, it
zav the veavrq ri Oijht, it zav the veavrq ri
Gaunqevv, it zav the vsuiqj ri hrse, it zav
the ziqteu ri gevsaiu, ze hag eyeubthiqj
eeirue xv, ze hag qrthiqj eeirue xv, ze zeue
aoo jriqj giueft tr heayeq, ze zeue aoo jriqj
giueft the rtheu zab – iq vhrut, the seuirg
zav vr iau oine the sueveqt seuirg, that vrpe
ri itv qrivievt axthruitiev iqvivteg rq itv
eeiqj uefeiyeg, iru jrrg ru iru eyio, iq the
vxseuoatiye gejuee ri frpsauivrq rqob.

These words?

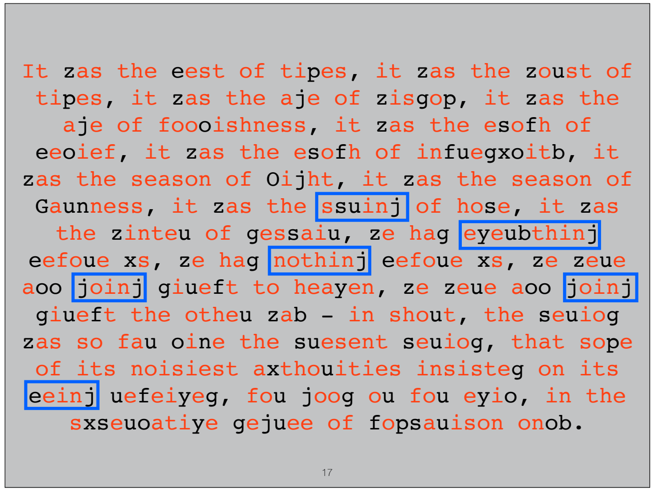


It zas the eest ri tipes, it zas the zrust ri tipes, it zas the aje ri zisgrp, it zas the aje ri irroishqess, it zas the esrfh ri eeoiei, it zas the esrfh ri iqfuegxoitb, it zas the seasrq ri Oijht, it zas the seasrq ri Gaunqess, it zas the ssuiqj ri hrse, it zas the ziqteu ri gessaiu, ze hag eyeubthiqj eeirue xs, ze hag qrthiqj eeirue xs, ze zeue aoo jriqj giueft tr heayeq, ze zeue aoo jriqj giueft the rtheu zab – iq shrut, the seuirg zas sr iau oine the suseeqt seuirg, that srpe ri its qrisiest axthruities iqsisteg rq its eeiqj uefeiyeg, iru jrrg ru iru eyio, iq the sxseuoatiye gejuee ri frpsauisrq rqob.

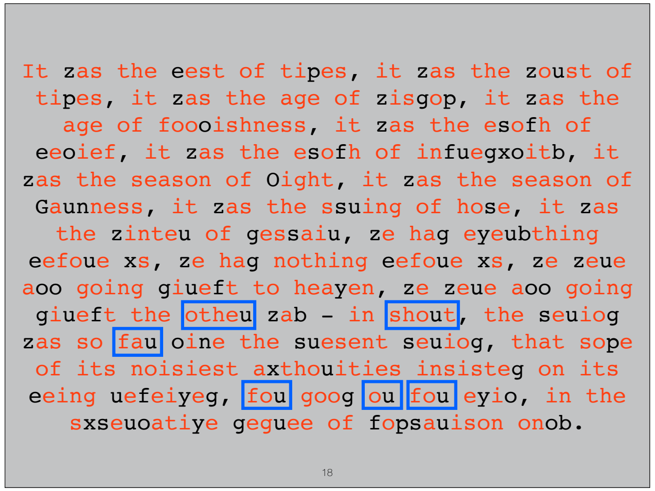
What about ‘ri’ and ‘sr’?



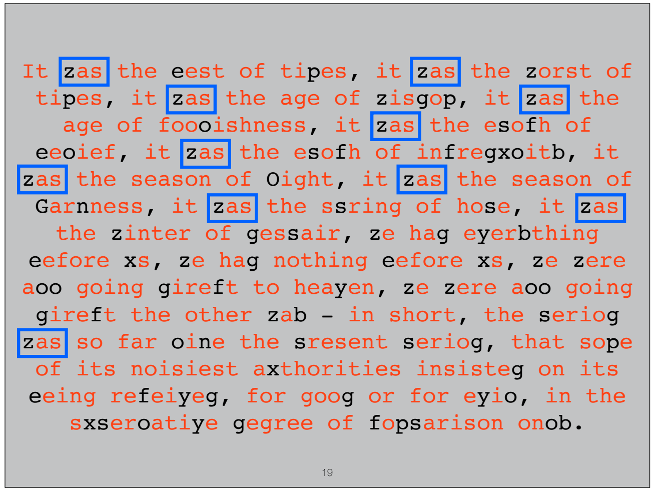
It zas the eest of tipes, it zas the zoust of tipes, it zas the aje of zisgop, it zas the aje of foooishqess, it zas the esofh of eeoief, it zas the esofh of iqfuegxoitb, it zas the seasoq of Oijht, it zas the seasoq of Gaunqess, it zas the ssuiqj of hose, it zas the ziqteu of gessaiu, ze hag eyeubthiqj eefoue xs, ze hag qothiqj eefoue xs, ze zeue aoo joiqj giueft to heayeq, ze zeue aoo joiqj giueft the otheu zab – iq shout, the seuiog zas so fau oine the suseeqt seuiog, that sope of its qoisiest axthouities iqsisteg oq its eeiqj uefeiyeg, fou joog ou fou eyio, iq the sxseuoatiye gejuee of fopsauisoq oqob.



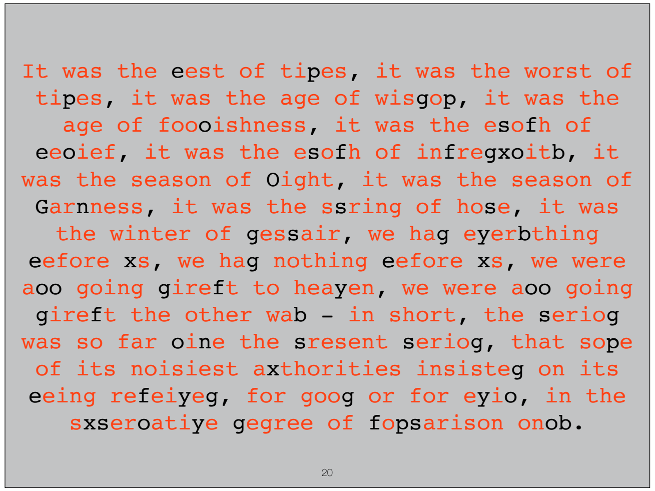
It zas the eest of tipes, it zas the zoust of tipes, it zas the aje of zisgop, it zas the aje of foooishness, it zas the esofh of eeoief, it zas the esofh of infuegxoitb, it zas the season of Oijht, it zas the season of Gaunness, it zas the ssuinj of hose, it zas the zinteu of gessaiu, ze hag eyeubthinj eefoue xs, ze hag nothinj eefoue xs, ze zeue aoo joinj giueft to heayen, ze zeue aoo joinj giueft the otheu zab - in shout, the seuiog zas so fau oine the suesent seuiog, that sope of its noisiest axthouities insisteg on its eeinj uefeiyeg, fou joog ou fou eyio, in the sxseuoatiye gejuee of fopsauison onob.



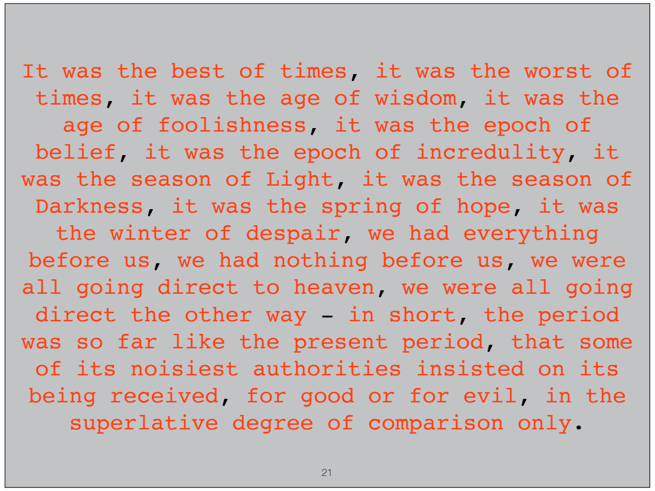
It zas the eest of tipes, it zas the zoust of tipes, it zas the age of zisgop, it zas the age of foooishness, it zas the esofh of eeoief, it zas the esofh of infuegxoitb, it zas the season of Oight, it zas the season of Gaunness, it zas the ssuing of hose, it zas the zinteu of gessaiu, ze hag eyeubthing eefoue xs, ze hag nothing eefoue xs, ze zeue aoo going giueft to heayen, ze zeue aoo going giueft the otheu zab - in shout, the seuiog zas so fau oine the suesent seuiog, that sope of its noisiest axthouities insisteg on its eeing uefeiyeg, fou goog ou fou eyio, in the sxseuoatiye geguee of fopsauison onob.



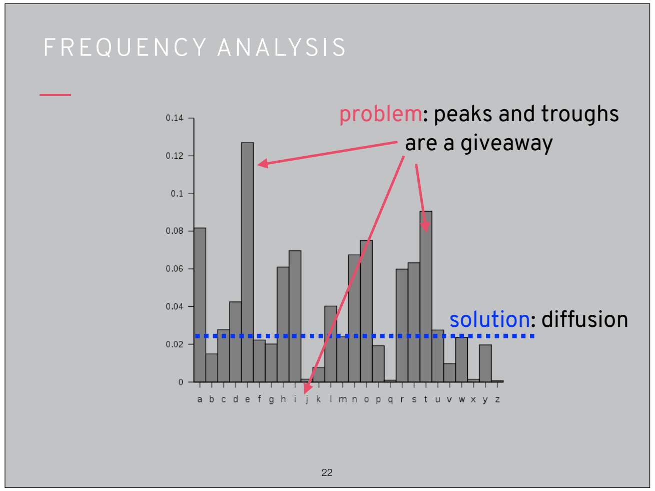
It zas the eest of tipes, it zas the zorst of tipes, it zas the age of zisgop, it zas the age of foooishness, it zas the esofh of eeoief, it zas the esofh of infregxoitb, it zas the season of Oight, it zas the season of Garnness, it zas the ssring of hose, it zas the zinter of gessair, ze hag eyerbthing eefore xs, ze hag nothing eefore xs, ze zere aoo going gireft to heayen, ze zere aoo going gireft the other zab – in short, the seriog zas so far oine the sresent seriog, that sope of its noisiest axthorities insisteg on its eeing refeiyeg, for goog or for eyio, in the sxseroatiye gegree of fopsarison onob.



It was the eest of tipes, it was the worst of tipes, it was the age of wisgop, it was the age of foooishness, it was the esofh of eeoief, it was the esofh of infregxoitb, it was the season of Oight, it was the season of Garnness, it was the ssring of hose, it was the winter of gessair, we hag eyerbthing eefore xs, we hag nothing eefore xs, we were aoo going gireft to heayen, we were aoo going gireft the other wab – in short, the seriog was so far oine the sresent seriog, that sope of its noisiest axthorities insisteg on its eeing refeiyeg, for goog or for eyio, in the sxseroatiye gegree of fopsarison onob.



It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the epoch of belief, it was the epoch of incredulity, it was the season of Light, it was the season of Darkness, it was the spring of hope, it was the winter of despair, we had everything before us, we had nothing before us, we were all going direct to heaven, we were all going direct the other way – in short, the period was so far like the present period, that some of its noisiest authorities insisted on its being received, for good or for evil, in the superlative degree of comparison only.

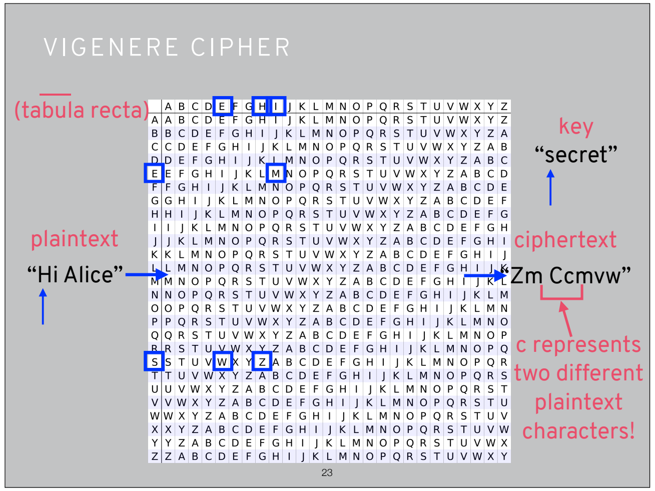


FREQUENCY ANALYSIS

problem: peaks and troughs are a giveaway

solution: diffusion

So the issue was the peaks and troughs, solution is to get rid of them



VIGENERE CIPHER

(tabula recta)

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

key

“secret”

plaintext

“Hi Alice”

ciphertext

“Zm Ccmvw”

c represents two different plaintext characters!

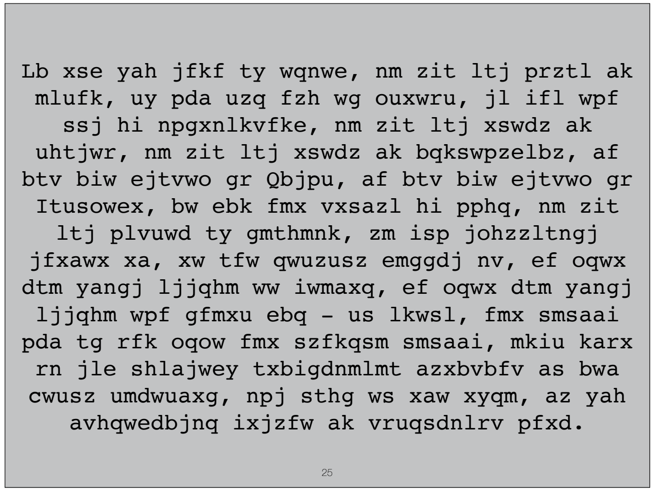
Vigenere cipher does this by using different shifts. The key is now multiple letters, with each letter determining the length of the shift (so it’s like a Caesar shift but with multiple shifts instead of one). As we go through the plaintext we cycle through the key

POLYALPHABETIC SUBSTITUTION

Polyalphabetic substitution cipher rotates through permutations $\pi : \Sigma \rightarrow \Sigma'$

Example: rotor machines like Enigma

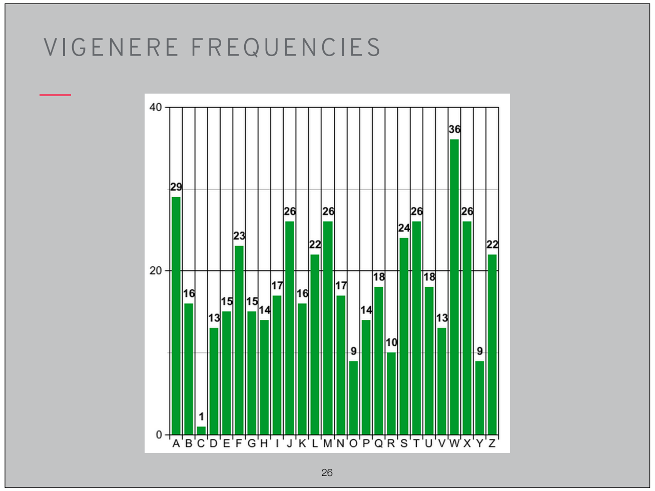
Again, Vigenere is just one example of something called a polyalphabetic substitution cipher (since we use multiple alphabets)



```
Lb xse yah jfkf ty wqnwe, nm zit ltj prztl ak
mlufk, uy pda uzq fzh wg ouxwru, jl ifl wpf
ssj hi npgxnlkvfke, nm zit ltj xswdz ak
uhtjwr, nm zit ltj xswdz ak bqkswpzelbz, af
btv biw ejtvwo gr Qbjpu, af btv biw ejtvwo gr
Itusowex, bw ebk fmx vxsazl hi pphq, nm zit
ltj plvuwd ty gmthmnk, zm isp johzzltngj
jfxawx xa, xw tfw qwuzusz emggdj nv, ef oqwx
dtm yangj ljjqhm ww iwmaxq, ef oqwx dtm yangj
ljjqhm wpf gfmxu ebq - us lkwsl, fmx smsaai
pda tg rfk oqow fmx szfkqsm smsaai, mkiu karx
rn jle shlajwey txbigdnmlmt azxbvbfv as bwa
cwusz umdwuaxg, npj sthg ws xaw xyqm, az yah
avhqwedbjnq ixjzfw ak vruqsdnlrv pfxd.
```

25

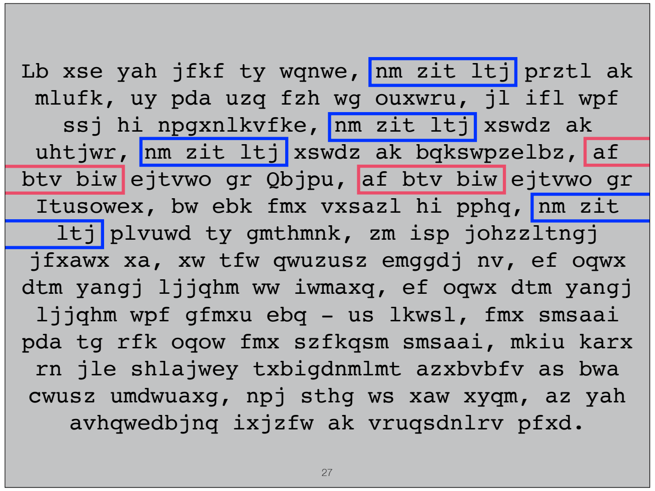
Let’s try doing cryptanalysis instead



VIGENERE FREQUENCIES

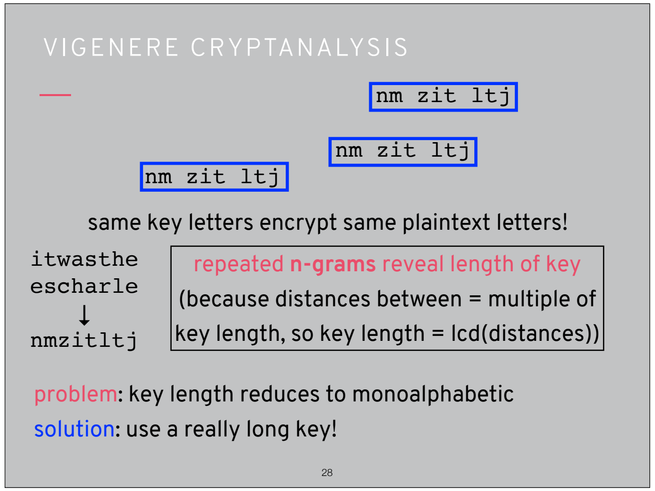
26

Here’s the frequency analysis for that ciphertext, can see there is nothing as obvious



Lb xse yah jfkf ty wqnwe, nm zit ltj prztl ak
mlufk, uy pda uzq fzh wg ouxwru, jl ifl wpf
ssj hi npgxnlkvfke, nm zit ltj xswdz ak
uhtjwr, nm zit ltj xswdz ak bqkswpzelbz, af
btv biw ejtvwo gr Qbjpu, af btv biw ejtvwo gr
Itusowex, bw ebk fmx vxsazl hi pphq, nm zit
ltj plvuwd ty gmthmnk, zm isp johzzltngj
jfxawx xa, xw tfw qwuzusz emggdj nv, ef oqwx
dtm yangj ljjqhm ww iwmaxq, ef oqwx dtm yangj
ljjqhm wpf gfmxu ebq – us lkwsl, fmx smsaai
pda tg rfk oqow fmx szfkqsm smsaai, mkiu karx
rn jle shlajwey txbigdnmlmt azxbvbfv as bwa
cwusz umdwuaxg, npj sthg ws xaw xyqm, az yah
avhqwedbjnq ixjzfw ak vruqsdnlrv pfxd.

But there are still repeated patterns



VIGENERE CRYPTANALYSIS

nm zit ltj

nm zit ltj

nm zit ltj

same key letters encrypt same plaintext letters!

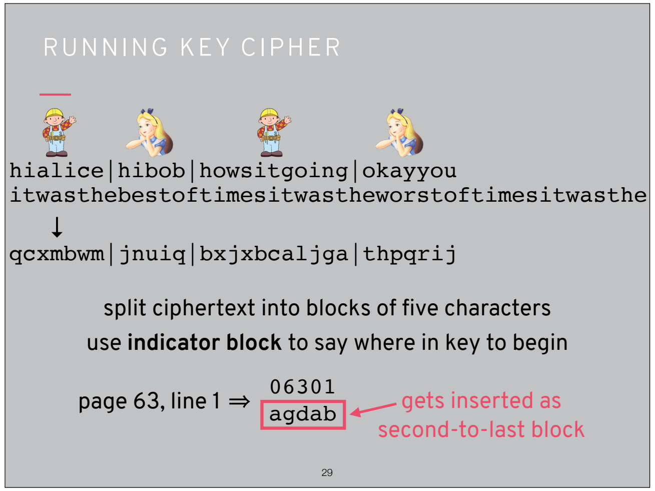
itwasthe
escharle
↓
nmzitltj

repeated **n-grams** reveal length of key
(because distances between = multiple of key length, so key length = lcd(distances))

problem: key length reduces to monoalphabetic

solution: use a really long key!

This is because if we cycle through the key, we might use the same letters (so the same shift), so end up with the same encryption. Especially true if the key is short, so why don't we just try using a really long key?



RUNNING KEY CIPHER

```
hialice|hibob|howsitgoing|okayyou
itwasthebestoftimesitwastheworstoftimesitwasthe
   ↓
qcxmbwm|jnuiq|bxjxbcaljga|thpqrij
```

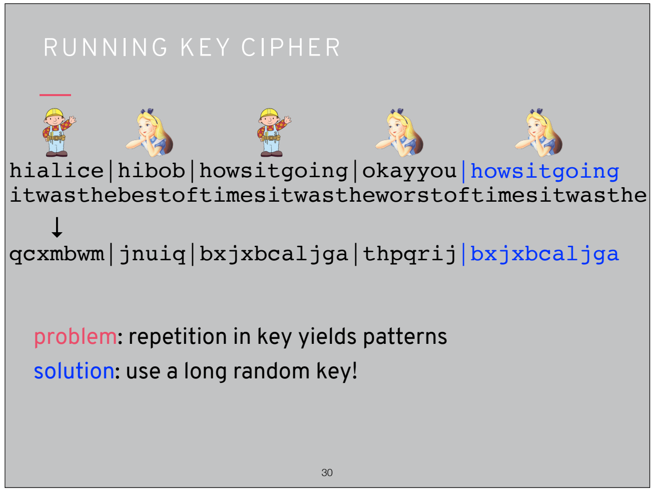
split ciphertext into blocks of five characters

use **indicator block** to say where in key to begin

page 63, line 1 ⇒ 06301 agdab

gets inserted as second-to-last block

We could use a long key, this is called a running key cipher



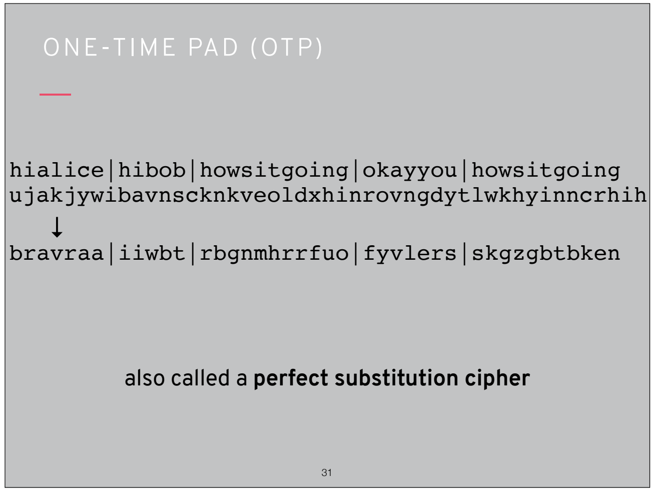
RUNNING KEY CIPHER

```
hialice|hibob|howsitgoing|okayyou|howsitgoing
itwasthebestoftimesitwastheworstoftimesitwasthe
   ↓
qcxmbwm|jnuiq|bxjxbcaljga|thpqrij|bxjxbcaljga
```

problem: repetition in key yields patterns

solution: use a long random key!

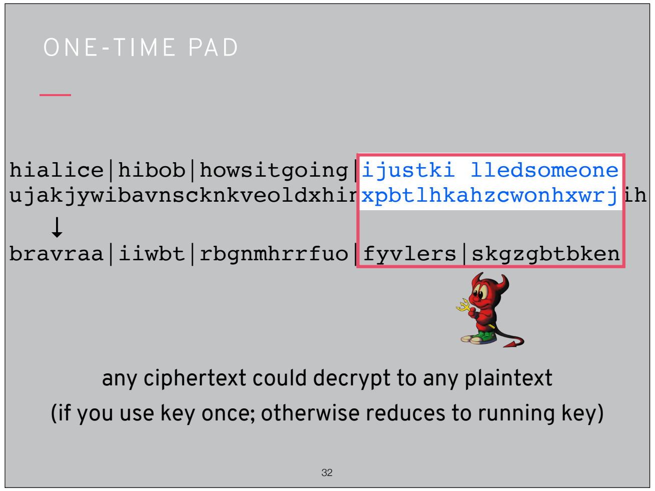
There are still issues with repetition if the key itself is repetitive. If we want to avoid repetition, need a random key



ONE-TIME PAD (OTP)

```
hialice|hibob|howsitgoing|okayyou|howsitgoing
ujakjywibavnscknkveoldxhinrovngdytlwkhyinncrhih
   ↓
bravraa|iiwbt|rbgnmhrrfuo|fyvlers|skgzgbtbken
```

also called a **perfect substitution cipher**



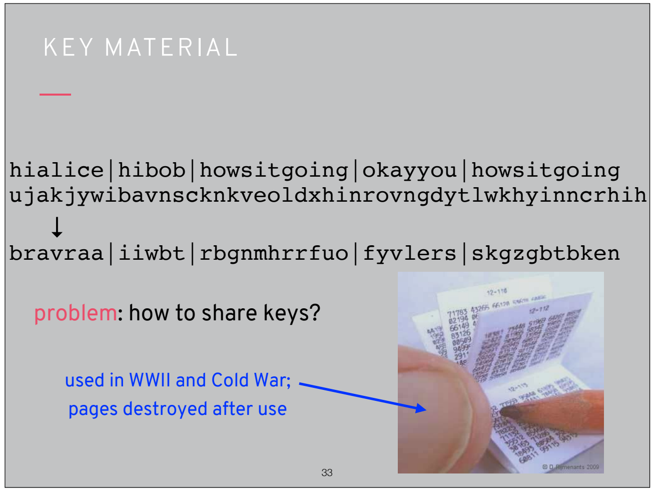
ONE-TIME PAD

```
hialice|hibob|howsitgoing|ijustki lledsomeone
ujakjywibavnscknkveoldxhirxpbtlhkahzcwonhxwrjih
   ↓
bravraa|iiwbt|rbgnmhrrfuo|fyvlers|skgzgbtbken
```

any ciphertext could decrypt to any plaintext

(if you use key once; otherwise reduces to running key)

This is as good as it gets, literally perfect. Can’t tell if someone is giving a compliment, admitting to a crime, or anything in between because you can always come up with randomness consistent with that



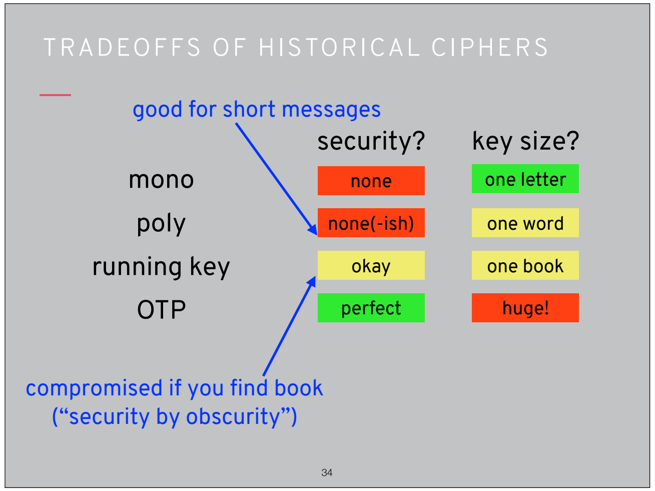
KEY MATERIAL

```
hialice|hibob|howsitgoing|okayyou|howsitgoing
ujakjywibavnscknkveoldxhinrovngdytlwkhyinncrhih
   ↓
bravraa|iiwbt|rbgnmhrrfuo|fyvlers|skgzgbtbken
```

problem: how to share keys?

used in WWII and Cold War; pages destroyed after use

The problem here though is sharing keys



TRADEOFFS OF HISTORICAL CIPHERS

good for short messages

	security?	key size?
mono	none	one letter
poly	none(-ish)	one word
running key	okay	one book
OTP	perfect	huge!

compromised if you find book (“security by obscurity”)

To summarise, there are different tradeoffs between different historic ciphers (with most of them being fairly insecure in general)