CS 5602 Lecture 15 Historical Ciphers II

George Markowsky
Computer Science Department
Missouri University of Science and Technology

1

Symmetric Encryption

Encryption of most data is accomplished using fast block and stream ciphers. These are examples of symmetric encryption algorithms. In addition all historical, i.e. pre-1960, ciphers are symmetric in nature and share some design principles with modern ciphers.

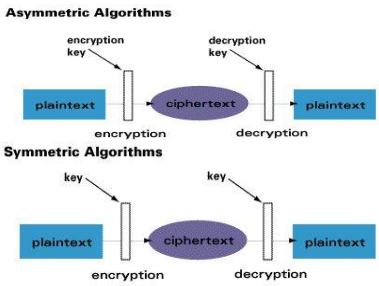
The main drawback with symmetric ciphers is that they give rise to a problem of how to distribute the secret keys between users, so we also address this issue.

We also discuss the properties and design of cryptographic hash functions and message authentication codes. Both of which will form basic building blocks of other schemes and protocols within this book.

In the following chapters we explain the theory and practice of modern symmetric ciphers, but first we consider historical ciphers.

3

Types of Algorithms



2

TABLE 1. English letter frequencies

Letter	Percentage	Letter	Percentage
A	8.2	N	6.7
B	1.5	O	7.5
C	2.8	P	1.9
D	4.2	Q	0.1
E	12.7	R	6.0
F	2.2	S	6.3
G	2.0	T	9.0
H	6.1	U	2.8
I	7.0	V	1.0
J	0.1	W	2.4
K	0.8	X	0.1
L	4.0	Y	2.0
M	2.4	Z	0.1

FIGURE 1. English letter frequencies

4

TABLE 2. English bigram frequencies

Bigram	Percentage	Bigram	Percentage
TH	3.15	HE	2.51
AN	1.72	IN	1.69
ER	1.54	RE	1.48
ES	1.45	ON	1.45
EA	1.31	TI	1.28
AT	1.24	ST	1.21
EN	1.20	ND	1.18

5

Trigrams

The most common trigrams are, in decreasing order,
THE, ING, AND, HER, ERE, ENT, THA, NTH, WAS, ETH, FOR.

6

3. Substitution Cipher

The main problem with the shift cipher is that the number of keys is too small, we only have 26 possible keys. To increase the number of keys a *substitution cipher* was invented. To write down a key for the substitution cipher we first write down the alphabet, and then a permutation of the alphabet directly below it. This mapping gives the substitution we make between the plaintext and the ciphertext

Plaintext alphabet ABCDEFGHIJKLMNOPQRSTUVWXYZ
Ciphertext alphabet GOYDSIPELUAVCRJWXZNHBQFTMK

Encryption involves replacing each letter in the top row by its value in the bottom row. Decryption involves first looking for the letter in the bottom row and then seeing which letter in the top row maps to it. Hence, the plaintext word HELLO would encrypt to the ciphertext ESVVJ if we used the substitution given above.

The number of possible keys is equal to the total number of permutations on 26 letters, namely the size of the group S_{26}, which is

$$26! \approx 4.03 \cdot 10^{26} \approx 2^{88}.$$

Since, as a rule of thumb, it is feasible to only run a computer on a problem which takes under 2^{80} steps we can deduce that this large key space is far too large to enable a brute force search even using a modern computer. Still we can break substitution ciphers using statistics of the underlying plaintext language, just as we did for the shift cipher.

7

XSO MJIWXVL JODIVA STW VAO VY OZJVCO'W LTJDOWX KVAKOAXJTXIVAW VY SIDS XOKSAVLVDQ IAGZWXJQ. KVUCZXOJW, KVUUZAIKTXIVAW TAG UIKJVOLOKXJ-VAIKW TJO HOLL JOCJOWOAXOG, TLVADWIGO GIDIXTL UOGIT, KVUCZXOJ DTUOW TAG OLOKXJVAIK KVUUOJKO. TW HOLL TW SVWXIAD UTAQ JOWOTJKS TAG CJVGZKX GONOLVCUOAX KOAXJOW VY UTPVJ DLVMTL KVUCTAIOW, XSO JO-DIVA STW T JTCIGLQ DJVHIAD AZUMOJ VY IAAVNTXINO AOH KVUCTAIOW. XSO KVUCZXOJ WKIOAKO GOCTJXUOAX STW KLVWO JOLTXIVAWSICW HIXS UTAQ VY XSOWO VJDTAIWTXIVAW NIT KVLLTMVJTXINO CJVPOKXW, WXTYY WOK-VAGUOAXW TAG NIWIXIAD IAGZWXJITL WXTYY. IX STW JOKOAXLQ IAXJVGZKOG WONOJTL UOKSTAIWUW YVJ GONOLVCIAD TAG WZCCVJXIAD OAXJOCJOAOZJITL WXZGOAXW TAG WXTYY, TAG TIUW XV CLTQ T WIDAIYIKTAX JVLO IA XSO GONOLVCUOAX VY SIDS-XOKSAVLVDQ IAGZWXJQ IA XSO JODIVA.

XSO GOCTJXUOAX STW T LTJDO CJVDJTUUO VY JOWOTJKS WZCCVJXOG MQ IAGZWXJQ, XSO OZJVCOTA ZAIVA, TAG ZE DVNOJAUOAX JOWOTJKS OWXTMLIW-SUOAXW TAG CZMLIK KVJCVJTXIVAW. T EOQ OLOUOAX VY XSIW IW XSO WXJ-VAD LIAEW XSTX XSO GOCTJXUOAX STW HIXS XSO KVUCZXOJ, KVUUZAIKTXIVAW, UIKJVOLOKXJVAIKW TAG UOGIT IAGZWXJIOW IA XSO MJIWXVL JODIVA. XSO TKT-GOUIK JOWOTJKS CJVDJTUUO IW VJDTAIWOG IAXV WONOA DJVZCW, LTADZTDOW TAG TJKSIXOKXZJO, GIDIXTL UOGIT, UVMILO TAG HOTJTMLO KVUCZXIAD, UTK-SIAO LOTJAIAD, RZTAXZU KVUCZXIAD, WQWXOU NOJIYIKTXIVA, TAG KJQCXVD-JTCSQ TAG IAYVJUTXIVA WOKZJIXQ.

8

We can compute the following frequencies for single letters in the above ciphertext:

Letter	Freq	Letter	Freq	Letter	Freq
A	8.6995	B	0.0000	C	3.0493
D	3.1390	E	0.2690	F	0.0000
G	3.6771	H	0.6278	I	7.8923
J	7.0852	K	4.6636	L	3.5874
M	0.8968	N	1.0762	O	11.479
P	0.1793	Q	1.3452	R	0.0896
S	3.5874	T	8.0717	U	4.1255
V	7.2645	W	6.6367	X	8.0717
Y	1.6143	Z	2.7802		

Letter	Percentage	Letter	Percentage
A	8.2	N	6.7
B	1.5	O	7.5
C	2.8	P	1.9
D	4.2	Q	0.1
E	12.7	R	6.0
F	2.2	S	6.3
G	2.0	T	9.0
H	6.1	U	2.8
I	7.0	V	1.0
J	0.1	W	2.4
K	0.8	X	0.1
L	4.0	Y	2.0
M	2.4	Z	0.1

In addition we determine that the most common bigrams in this piece of ciphertext are

TA, AX, IA, VA, WX, XS, AG, OA, JO, JV,

whilst the most common trigrams are

OAX, TAG, IVA, XSO, KVU, TXI, UOA, AXS.

Since the ciphertext letter O occurs with the greatest frequency, namely 11.479, we can guess that the ciphertext letter O corresponds to the plaintext letter E. We now look at what this means for two of the common trigrams found in the ciphertext

- The ciphertext trigram OAX corresponds to E * *.
- The ciphertext trigram XSO corresponds to * * E.

9

We examine similar common similar trigrams in English, which start or end with the letter E. We find that three common ones are given by ENT, ETH and THE. Since the two trigrams we wish to match have one starting with the same letter as the other finishes with, we can conclude that it is highly likely that we have the correspondence

- X = T,
- S = H,
- A = N.

Even after this small piece of analysis we find that it is much easier to understand what the underlying plaintext should be. If we focus on the first two sentences of the ciphertext we are trying to break, and we change the letters which we think we have found the correct mappings for, then we obtain:

THE MJIWTVL JEDIVN HTW VNE VY EZJVCE'W LTJDEWT
KVNKENTJTTIV NW VY HIDH TEKHNVLVDQ INGZWTJQ.
KVUCZTEJW, KVUUZNIKTTIVNW TNG UIKJVELEKTJVNIKW
TJE HELL JECJEWENTEG, TLVNDWIGE GIDITTL UEGIT,
KVUCZTEJ DTUEW TNG ELEKTJVNIK KVUUEJKE.

10

Recall, this was after the four substitutions

O = E, X = T, S = H, A = N.

We now cheat and use the fact that we have retained the word sizes in the ciphertext. We see that since the letter T occurs as a single ciphertext letter we must have

T = I or T = A.

The ciphertext letter T occurs with a probability of 8.0717, which is the highest probability left, hence we are far more likely to have

T = A.

We have already considered the most popular trigram in the ciphertext so turning our attention to the next most popular trigram we see that it is equal to TAG which we suspect corresponds to the plaintext AN*. Therefore it is highly likely that G = D, since AND is a popular trigram in English.

Our partially decrypted ciphertext is now equal to

THE MJIWTVL JEDIVN HAW VNE VY EZJVCE'W LAJDEWT
KVNKENTJATIV NW VY HIDH TEKHNVLVDQ INDZWTJQ.
KVUCZTEJW, KVUUZNIKATIVNW AND UIKJVELEKTJVNIKW
AJE HELL JECJEWENTED, ALVNDWIDE DIDITAL UEDIA,
KVUCZTEJ DAUEW AND ELEKTJVNIK KVUUEJKE.

This was after the six substitutions

O = E, X = T, S = H,
A = N, T = A, G = D.

11

We now look at two-letter words which occur in the ciphertext:

- IX
  This corresponds to the plaintext *T. Therefore the ciphertext letter I must be one of the plaintext letters A or I, since the only two-letter words in English ending in T are AT and IT. We already have worked out what the plaintext character A corresponds to, hence we must have I = I.
- XV
  This corresponds to the plaintext T*. Hence, we must have V = O.
- VY
  This corresponds to the plaintext O*. Hence, the ciphertext letter Y must correspond to one of F, N or R. We already know the ciphertext letter corresponding to N. In the ciphertext the probability of Y occurring is 1.6, but in English we expect F to occur with probability 2.2 and R to occur with probability 6.0. Hence, it is more likely that Y = F.
- IW
  This corresponds to the plaintext I*. Therefore, the plaintext character W must be one of F, N, S and T. We already have F, N, T, hence W = S.

All these deductions leave the partial ciphertext as

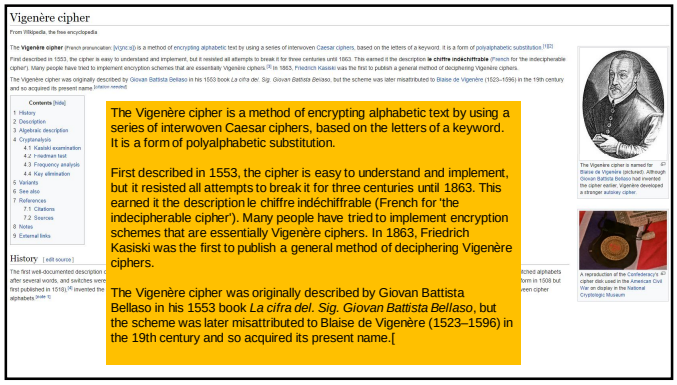
THE MJISTOL JEDION HAS ONE OF EZJOCE'S LAJDEST
KONKENTJATIONS OF HIDH TEKHNOLODQ INDZSTJQ.
KOUCZTEJS, KOUUZNIKATIONS AND UIKJOELEKTJONIKS AJE
HELL JECJESENTED, ALONDSIDE DIDITAL UEDIA,
KOUCZTEJ DAUES AND ELEKTJONIK KOUUEJKE.

This was after the ten substitutions

O = E, X = T, S = H, A = N, T = A,
G = D, I = I, V = O, Y = F, W = S.

Now you can finish this up on your own!

12

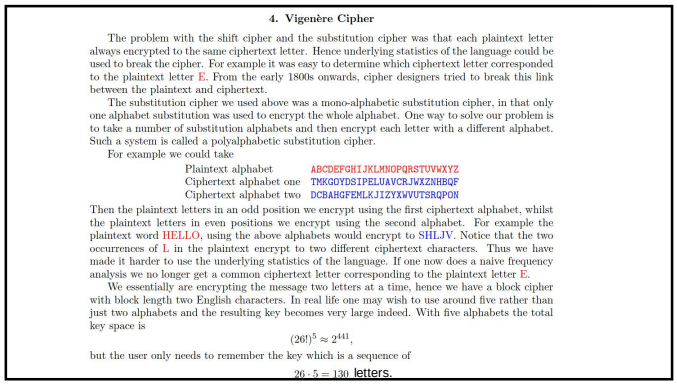


13

The Bellaso Cipher

- People know that Vigenere is an incorrect name, but they still insist on using it!
- Unbreakable for 300 years
- The key to breaking it was provided by Friedrich Kasiski

16



4. Vigenère Cipher

The problem with the shift cipher and the substitution cipher was that each plaintext letter always encrypted to the same ciphertext letter. Hence underlying statistics of the language could be used to break the cipher. For example it was easy to determine which ciphertext letter corresponded to the plaintext letter E. From the early 1800s onwards, cipher designers tried to break this link between the plaintext and ciphertext.

The substitution cipher we used above was a mono-alphabetic substitution cipher, in that only one alphabet substitution was used to encrypt the whole alphabet. One way to solve our problem is to take a number of substitution alphabets and then encrypt each letter with a different alphabet. Such a system is called a polyalphabetic substitution cipher.

For example we could take

Plaintext alphabet	ABCDEFGHIJKLMNOPQRSTUVWXYZ
Ciphertext alphabet one	TMKGOYDSIPELUAVCRJWXZNHBQF
Ciphertext alphabet two	DCBAHGFEMLKJIZYXWVUTSRQPON

Then the plaintext letters in an odd position we encrypt using the first ciphertext alphabet, whilst the plaintext letters in even positions we encrypt using the second alphabet. For example the plaintext word HELLO, using the above alphabets would encrypt to SHLJV. Notice that the two occurrences of L in the plaintext encrypt to two different ciphertext characters. Thus we have made it harder to use the underlying statistics of the language. If one now does a naive frequency analysis we no longer get a common ciphertext letter corresponding to the plaintext letter E.

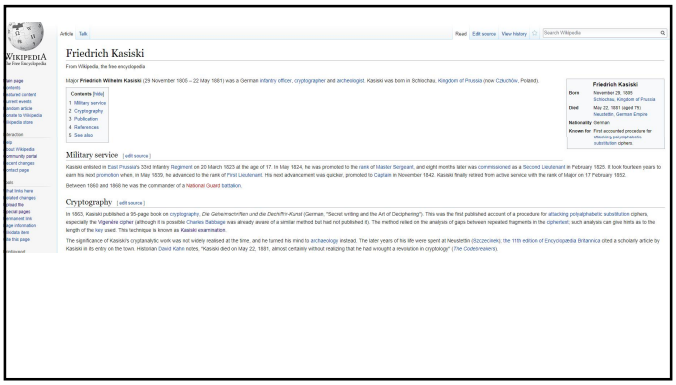
We essentially are encrypting the message two letters at a time, hence we have a block cipher with block length two English characters. In real life one may wish to use around five rather than just two alphabets and the resulting key becomes very large indeed. With five alphabets the total key space is

$$(26!)^5 \approx 2^{441},$$

but the user only needs to remember the key which is a sequence of

$$26 \cdot 5 = 130 \text{ letters.}$$

14



17

However, just to make life hard for the attacker, the number of alphabets in use should also be hidden from his view and form part of the key. But for the average user in the early 1800s this was far too unwieldy a system, since the key was too hard to remember.

Despite its shortcomings the most famous cipher during the 19th-century was based on precisely this principle. The Vigenere cipher, invented in 1533 by Giovan Batista Belaso, was a variant on the above theme, but the key was easy to remember. When looked at in one way the Vigenere cipher is a polyalphabetic block cipher, but when looked at in another, it is a stream cipher which is a natural generalization of the shift cipher.

The description of the Vigenere cipher as a block cipher takes the description of the polyal-phabetic cipher above but restricts the possible plaintext alphabets to one of the 26 possible cyclic shifts of the standard alphabet. Suppose five alphabets were used, this reduces the key space dow to

$$26^5\,(11{,}881{,}376) \approx 2^{23}\,(8{,}388{,}608) \approx 10^7$$

and the size of the key to be remembered as a sequence of five numbers between 0 and 25.

However, the description of the Vigenere cipher as a stream cipher is much more natural. Just like the shift cipher, the Vigenere cipher again identifies letters with the numbers 0, . . . , 25. The secret key is a short sequence of letters (e.g. a word) which is repeated again and again to form a keystream. Encryption involves adding the plaintext letter to a key letter. Thus if the key is SESAME, encryption works as follows,

```
THISISATESTMESSAGE
SESAMESESAMESESAME
LLASUWSXWSFQWWKASI
```

Again we notice that A will encrypt to a different letter depending on where it appears in the message.

0	A
1	B
2	C
3	D
4	E
5	F
6	G
7	H
8	I
9	J
10	K
11	L
12	M
13	N
14	O
15	P
16	Q
17	R
18	S
19	T
20	U
21	V
22	W
23	X
24	Y
25	Z

15

The Kasiski Examination (Wikipedia)

In cryptanalysis, Kasiski examination (also referred to as Kasiski's test or Kasiski's method) is a method of attacking polyalphabetic substitution ciphers, such as the Vigenère cipher. It was first published by Friedrich Kasiski in 1863, but seems to have been independently discovered by Charles Babbage as early as 1846.

How it works

In polyalphabetic substitution ciphers where the substitution alphabets are chosen by the use of a keyword, the Kasiski examination allows a cryptanalyst to deduce the length of the keyword. Once the length of the keyword is discovered, the cryptanalyst lines up the ciphertext in n columns, where n is the length of the keyword. Then each column can be treated as the ciphertext of a monoalphabetic substitution cipher. As such, each column can be attacked with frequency analysis. Similarly, where a rotor stream cipher machine has been used, this method may allow the deduction of the length of individual rotors.

18

The Kasiski Examination (Wikipedia)

The Kasiski examination involves looking for strings of characters that are repeated in the ciphertext. The strings should be three characters long or more for the examination to be successful. Then, the distances between consecutive occurrences of the strings are likely to be multiples of the length of the keyword. Thus finding more repeated strings narrows down the possible lengths of the keyword, since we can take the greatest common divisor of all the distances.

The reason this test works is that if a repeated string occurs in the plaintext, and the distance between corresponding characters is a multiple of the keyword length, the keyword letters will line up in the same way with both occurrences of the string.

19

The Kasiski Examination (Wikipedia)

- Suppose the plaintext is `crypto is short for cryptography`
- The string *crypto* is repeated
- Suppose the key chosen was abcdef, we do not benefit from the repetition of crypto because the repetition does not line up with the key

```
abcdefabcdefabcdefabcdefabcdefabc
crypto is short for cryptography.
```

0	A
1	B
2	C
3	D
4	E
5	F
6	G
7	H
8	I
9	J
10	K
11	L
12	M
13	N
14	O
15	P
16	Q
17	R
18	S
19	T
20	U
21	V
22	W
23	X
24	Y
25	Z

20

The Kasiski Examination (Wikipedia)

- If the chosen key were abcde, we would get better results

```
abcdeabcdeabcdeabcdeabcdeabcdeabc
crypto is short for cryptography.
```

0	A
1	B
2	C
3	D
4	E
5	F
6	G
7	H
8	I
9	J
10	K
11	L
12	M
13	N
14	O
15	P
16	Q
17	R
18	S
19	T
20	U
21	V
22	W
23	X
24	Y
25	Z

21

The Kasiski Examination (Wikipedia)

1. A cryptanalyst looks for repeated groups of letters and counts the number of letters between the beginning of each repeated group. For instance, if the ciphertext was FGXTHJAQWNFGXQ, the distance between FGX groups is 10. The analyst records the distances for all repeated groups in the text.
2. The analyst next factors each of these numbers. If any number is repeated in the majority of these factorings, it is likely to be the length of the keyword. This is because repeated groups are more likely to occur when the same letters are encrypted using the same key letters than by mere coincidence; this is especially true for long matching strings. The key letters are repeated at multiples of the key length, so most of the distances found in step 1 are likely to be multiples of the key length. A common factor is usually evident.
3. Once the keyword length is known, the following observation of Babbage and Kasiski comes into play. If the keyword is N letters long, then every Nth letter must have been enciphered using the same letter of the keytext. Grouping every Nth letter together, the analyst has N "messages", each encrypted using a one-alphabet substitution, and each piece can then be attacked using frequency analysis.

22

The Kasiski Examination (Wikipedia)

4. Using the solved message, the analyst can quickly determine what the keyword was. Or, in the process of solving the pieces, the analyst might use guesses about the keyword to assist in breaking the message.
5. Once the interceptor knows the keyword, that knowledge can be used to read other messages that use the same key.

23

The Kasiski Examination (Wikipedia)

- Kasiski actually used "superimposition" to solve the Vigenère cipher. He started by finding the key length, as above. Then he took multiple copies of the message and laid them one-above-another, each one shifted left by the length of the key. Kasiski then observed that each column was made up of letters encrypted with a single alphabet. His method was equivalent to the one described above, but is perhaps easier to picture.
- Modern attacks on polyalphabetic ciphers are essentially identical to that described above, with the one improvement of coincidence counting. Instead of looking for repeating groups, a modern analyst would take two copies of the message and lay one above another.
- Modern analysts use computers, but this description illustrates the principle that the computer algorithms implement.
- The generalized method:
  - The analyst shifts the bottom message one letter to the left, then one more letters to the left, etc., each time going through the entire message and counting the number of times the same letter appears in the top and bottom message.
  - The number of "coincidences" goes up sharply when the bottom message is shifted by a multiple of the key length, because then the adjacent letters are in the same language using the same alphabet.
  - Having found the key length, cryptanalysis proceeds as described above using frequency analysis.

24

As an example, suppose the ciphertext is given by

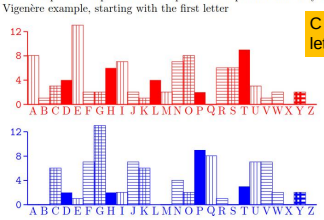
UTPDHUG NYH USVKCG MVCE FXL KQIB. WX RKU GI TZN, RLS BBHZLXMSNP
KDKS; CEB IH HKEW IBA, YYM SBR PFR SBS, JV UPL O UVADGR HRRWXF. JV ZTVOOV
YH ZCQU Y UKWGEB, PL UQFB P FOUKCG, TBF RQ VHCF R KPG, OU KFT ZCQU MAW
QKKW ZGSY, FP PGM QKFTK UQFB DER EZRN, MCYE, MG UCTFSVA, WP KFT ZCQU
MAW KQIJS. LCOV NTHDNV JPNUJVB IH GGV RWX ONKCGTHKFL XG VKD, ZJM VG
CCI MVGD JPNUJ, RLS EWVKJT ASGUCS MVGD; DDK VG NYH PWUV CCHIIY RD DBQN
RWTH PFRWBBI VTTK VCGNTGSF FL IAWU XJDUS, HFP VHCF, RR LAWEY QDFS
RVMEES FZB CHH JRTT MVGZP UBZN FD ATIIYRTK WP KFT HIVJCI; TBF BLDPWPX
RWTH ULAW TG VYCHX KQLJS US DCGCW OPPUPR, VG KFDNUJK GI JIKKC PL KGCJ
IAOV KFTR GJFSAW KTZLZES WG RWXWT VWTL WP XPXGG, CJ FPOS VYC BTZCUW
XG ZGJQ PMHTRAIBJG WMGFG. JZQ DPB JVYGM ZCLEWXR: CEB IAOV NYH JIKKC
TGCWXF UHF JZK.
WX VCU LD YITKFTK WPKCGVCWIQT PWVY QEBFKKQ, QNH NZTTW IRFL IAS
VFRPE ODJRXGSPTC EKWPTGEES, GMCG
TTVVPLTFFJ; YCW WV NYH TZYRWH LOKU MU AWO, KFPM VG BLTP VQN RD DSGG
AWKWUKKPL KGCJ, XY OPP KPG ONZTT ICUJCHLSF KFT DBQNJTWUG. DYN MVCK
ZT MFWCW HTWF FD JL, OPU YAE CH LQ! PGR UF, YH MWPP RXF CDJCGOSF, XMS
UZGJQ JL, SXVPN HBG!

25

Now we take every sixth letter and look at the statistics just as we did for a shift cipher to deduce the first letter of the keyword. We can now see the advantage of using the histograms to break the shift cipher earlier. If we used the naive method and tried each of the 26 keys in turn we could still not detect which key is correct, since every sixth letter of an English sentence does not produce an English sentence. Using our earlier histogram based method is more efficient in this case.

FIGURE 3. Comparison of plaintext and ciphertext frequencies for every sixth letter of the Vigenère example, starting with the first letter

C for first letter?



28

There is a way of finding the length of the keyword, which is repeated to form the keystream, called the *Kasiski test*. First we need to look for repeated sequences of characters. Recall that English has a large repetition of certain bigrams or trigrams and over a long enough string of text these are likely to match up to the same two or three letters in the key every so often. By examining the distance between two repeated sequences we can guess the length of the keyword. Each of these distances should be a multiple of the keyword, hence taking the greatest common divisor of all distances between the repeated sequences should give a good guess as to the keyword length.

Let us examine the above ciphertext and look for the bigram WX. The gaps between some of the occurrences of this bigram are 9, 21, 66 and 30, some of which may have occurred by chance, whilst some may reveal information about the length of the keyword. We now take the relevant greatest common divisors to find,

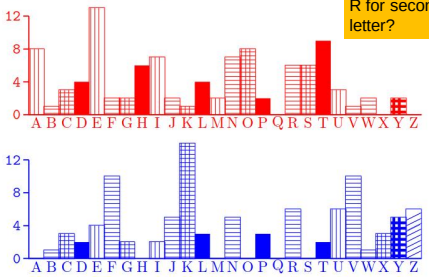
$$\gcd(30, 66) = 6,$$
$$\gcd(3, 9) = \gcd(9, 66) = \gcd(9, 30) = \gcd(21, 66) = 3.$$

We are unlikely to have a keyword of length three so we conclude that the gaps of 9 and 21 occurred purely by chance. Hence, our best guess for the keyword is that it is of length 6.

26

FIGURE 4. Comparison of plaintext and ciphertext frequencies for every sixth letter of the Vigenère example, starting with the second letter

R for second letter?



0 A
1 B
2 C
3 D
4 E
5 F
6 G
7 H
8 I
9 J
10 K
11 L
12 M
13 N
14 O
15 P
16 Q
17 R
18 S
19 T
20 U
21 V
22 W
23 X
24 Y
25 Z

29

As an example, suppose the ciphertext is given by

UTPDHUG NYH USVKCG MVCE FXL KQIB. WX RKU GI TZN, RLS BBHZLXMSNP
KDKS; CEB IH HKEW IBA, YYM SBR PFR SBS, JV UPL O UVADGR HRRWXF. JV ZTVOOV
YH ZCQU Y UKWGEB, PL UQFB P FOUKCG, TBF RQ VHCF R KPG, OU KFT ZCQU MAW
QKKW ZGSY, FP PGM QKFTK UQFB DER EZRN, MCYE, MG UCTFSVA, WP KFT ZCQU
MAW KQIJS. LCOV NTHDNV JPNUJVB IH GGV RWX ONKCGTHKFL XG VKD, ZJM VG
CCI MVGD JPNUJ, RLS EWVKJT ASGUCS MVGD; DDK VG NYH PWUV CCHIIY RD DBQN
RWTH PFRWBBI VTTK VCGNTGSF FL IAWU XJDUS, HFP VHCF, RR LAWEY QDFS
RVMEES FZB CHH JRTT MVGZP UBZN FD ATIIYRTK WP KFT HIVJCI; TBF BLDPWPX
RWTH ULAW TG VYCHX KQLJS US DCGCW OPPUPR, VG KFDNUJK GI JIKKC PL KGCJ
IAOV KFTR GJFSAW KTZLZES WG RWXWT VWTL WP XPXGG, CJ FPOS VYC BTZCUW
XG ZGJQ PMHTRAIBJG WMGFG. JZQ DPB JVYGM ZCLEWXR: CEB IAOV NYH JIKKC
TGCWXF UHF JZK.
WX VCU LD YITKFTK WPKCGVCWIQT PWVY QEBFKKQ, QNH NZTTW IRFL IAS
VFRPE ODJRXGSPTC EKWPTGEES, GMCG
TTVVPLTFFJ; YCW WV NYH TZYRWH LOKU MU AWO, KFPM VG BLTP VQN RD DSGG
AWKWUKKPL KGCJ, XY OPP KPG ONZTT ICUJCHLSF KFT DBQNJTWUG. DYN MVCK
ZT MFWCW HTWF FD JL, OPU YAE CH LQ! PGR UF, YH MWPP RXF CDJCGOSF, XMS
UZGJQ JL, SXVPN HBG!

27

Continuing in a similar way for the remaining four letters of the keyword we find the keyword is

CRYPTO.

The underlying plaintext is then found to be:

Scrooge was better than his word. He did it all, and infinitely more; and to Tiny Tim, who did not die, he was a second father. He became as good a friend, as good a master, and as good a man, as the good old city knew, or any other good old city, town, or borough, in the good old world. Some people laughed to see the alteration in him, but he let them laugh, and little heeded them; for he was wise enough to know that nothing ever happened on this globe, for good, at which some people did not have their fill of laughter in the outset; and knowing that such as these would be blind anyway, he thought it quite as well that they should wrinkle up their eyes in grins, as have the malady in less attractive forms. His own heart laughed: and that was quite enough for him.

He had no further intercourse with Spirits, but lived upon the Total Abstinence Principle, ever afterwards; and it was always said of him, that he knew how to keep Christmas well, if any man alive possessed the knowledge. May that be truly said of us, and all of us! And so, as Tiny Tim observed, God bless Us, Every One!

The above text is taken from *A Christmas Carol* by Charles Dickens.

30

Dictionary Attack on the Bellaso Cipher

- According to https://en.oxforddictionaries.com/explore/how-many-words-are-there-in-the-english-language/
- *The Second Edition of the 20-volume Oxford English Dictionary, published in 1989, contains full entries for 171,476 words in current use, and 47,156 obsolete words. To this may be added around 9,500 derivative words included as subentries. Over half of these words are nouns, about a quarter adjectives, and about a seventh verbs; the rest is made up of exclamations, conjunctions, prepositions, suffixes, etc. And these figures don't take account of entries with senses for different word classes (such as noun and adjective).*

31

Dictionary Attack on the Bellaso Cipher

- Consequently, if the encoder used a regular word, there are "only" around 250,000 choices
- You can write a computer program that tries every word in a dictionary to see if it gets a decoding
- This is called a dictionary attack
- To foil that, the encoder could use a nonsense string (hard to remember) or technical terms not in a dictionary
- One can add strings such as abcde, abc, etc
- One can eliminate words of

32

5. A Permutation Cipher

The ideas behind substitution type ciphers forms part of the design of modern symmetric systems. For example later we shall see that both DES and Rijndael make use of a component called an S-Box, which is simply a substitution. The other component that is used in modern symmetric ciphers is based on permutations.

Permutation ciphers have been around for a number of centuries. Here we shall describe the simplest, which is particularly easy to break. We first fix a permutation group S_n and a permutation

$$\sigma \in S_n.$$

It is the value of σ which will be the secret key. As an example suppose we take

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 1 & 3 & 5 \end{pmatrix} = (1243) \in S_5.$$

Now take some plaintext, say

Once upon a time there was a little girl called snow white.

We break the text into chunks of 5 letters

onceu ponat imeth erewa salit tlegi rlcal ledsn owwhi te.

We first pad the message, with some random letters, so that we have a multiple of five letters in each chunk.

onceu ponat imeth erewa salit tlegi rlcal ledsn owwhi teahb.

Then we take each five-letter chunk in turn and swap the letters around according to our secret permutation σ. With our example we obtain

coenu npaot eitmh eewra lsiat etgli crall dlsdn wohwi atheb.

We then remove the spaces, so as to hide the value of n, producing the ciphertext

coenunpaoteitmheewralsiatetglicralldlsdnwohwiatheb.

33

Known-Plaintext Attack (Wikipedia)

- *The known-plaintext attack (KPA) is an attack model for cryptanalysis where the attacker has access to both the plaintext (called a crib), and its encrypted version (ciphertext). These can be used to reveal further secret information such as secret keys and code books. The term "crib" originated at Bletchley Park, the British World War II decryption operation.*
- *The usage "crib" was adapted from a slang term referring to cheating (e.g., "I cribbed my answer from your test paper"). A "crib" originally was a literal or interlinear translation of a foreign-language text—usually a Latin or Greek text—that students might be assigned to translate from the original language.*

34

Known-Plaintext Attack (Wikipedia)

- *The idea behind a crib is that cryptologists were looking at incomprehensible ciphertext, but if they had a clue about some word or phrase that might be expected to be in the ciphertext, they would have a "wedge," a test to break into it.*
- *If their otherwise random attacks on the cipher managed to sometimes produce those words or (preferably) phrases, they would know they might be on the right track.*
- *When those words or phrases appeared, they would feed the settings they had used to reveal them back into the whole encrypted message to good effect. The usage "crib" was adapted from a slang term referring to cheating (e.g., "I cribbed my answer from your test paper").*
- *A "crib" originally was a literal or interlinear translation of a foreign-language text—usually a Latin or Greek text—that students might be assigned to translate from the original language.*

35

Known-Plaintext Attack (Wikipedia)

- *In the case of Enigma, the German High Command was very meticulous about the overall security of the Enigma system and understood the possible problem of cribs.*
- *The day-to-day operators, on the other hand, were less careful. The Bletchley Park team would guess some of the plaintext based upon when the message was sent, and by recognizing routine operational messages.*
- *For instance, a daily weather report was transmitted by the Germans at the same time every day. Due to the regimented style of military reports, it would contain the word Wetter (German for "weather") at the same location in every message.*
- *(Knowing the local weather conditions helped Bletchley Park guess other parts of the plaintext as well.) Other operators, too, would send standard salutations or introductions. An officer stationed in the Qattara Depression consistently reported that he had nothing to report.*
- *"Heil Hitler," occurring at the end of a message, is another well-known example.*

36

Known-Plaintext Attack (Wikipedia)

- *At Bletchley Park in World War II, strenuous efforts were made to use (and even force the Germans to produce) messages with known plaintext. For example, when cribs were lacking, Bletchley Park would sometimes ask the Royal Air Force to "seed" a particular area in the North Sea with mines (a process that came to be known as gardening, by obvious reference). The Enigma messages that were soon sent out would most likely contain the name of the area or the harbour threatened by the mines.*
- *The Germans themselves could be very accommodating in this regard. Whenever any of the turned German Double cross agents sent a message (written by the British) to their respective handlers, they frequently obligingly re-encrypted the message word for word on Enigma for onward transmission to Berlin.*

37

Known-Plaintext Attack (Wikipedia)

- *When a captured German revealed under interrogation that Enigma operators had been instructed to encode numbers by spelling them out, Alan Turing reviewed decrypted messages and determined that the number "eins" ("one") was the most common string in the plaintext. He automated the crib process, creating the Eins Catalogue, which assumed that "eins" was encoded at all positions in the plaintext. The catalogue included every possible position of the various rotors, starting positions, and key settings of the Enigma.*
- *The Polish Cipher Bureau had likewise exploited "cribs" in the "ANX method" before World War II (the Germans' use of "AN", German for "to", followed by "X" as a spacer to form the text "ANX").*

38

Known-Plaintext Attack (Wikipedia)

- *The United States and Britain used one-time tape systems, such as the 5-UCO, for their most sensitive traffic.*
- *These devices were immune to known-plaintext attack; however, they were point-to-point links and required massive supplies of one time tapes.*
- *Networked cipher machines were considered vulnerable to cribs, and various techniques were used to disguise the beginning and ends of a message, including cutting messages in half and sending the second part first and adding nonsense padding at both ends.*
- *The latter practice resulted in the world wonders incident. The KL-7, introduced in the mid-1950s, was the first U.S. cipher machine that was considered safe against known-plaintext attack.*

39

Known-Plaintext Attack (Wikipedia)

- *Classical ciphers are typically vulnerable to known-plaintext attack.*
- *For example, a Caesar cipher can be solved using a single letter of corresponding plaintext and ciphertext to decrypt entirely.*
- *A general monoalphabetic substitution cipher needs several character pairs and some guessing if there are fewer than 26 distinct pairs.*

40

The World Wonders Incident (Wilipedia)

- *"The world wonders" was a phrase used as security padding in an encrypted message sent from Admiral Chester Nimitz to Admiral William Halsey, Jr. on October 25, 1944, during the Battle of Leyte Gulf.*
- *The words, intended to be without meaning, were added to hinder Japanese attempts at cryptanalysis, but were mistakenly included in the decoded message given to Halsey and interpreted by him as a harsh and sarcastic rebuke.*
- *As a consequence, Halsey dropped his pursuit of a Japanese carrier task force in a futile attempt to aid United States forces in the Battle off Samar.*

41

A Permutation Cipher (Smart)

- *However, breaking a permutation cipher is easy with a chosen plaintext attack, assuming the group of permutations used (i.e. the value of n) is reasonably small.*
- *To attack this cipher we mount a chosen plaintext attack, and ask one of the parties to encrypt the message*

abcdefghijklmnopqrstuvwxyz,

to obtain the ciphertext

cadbehfigjmknlorpsqtwuxvyz.

We can then deduce that the permutation looks something like

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & \dots \\ 2 & 4 & 1 & 3 & 5 & 7 & 9 & 6 & 8 & 10 & 12 & 14 & 11 & 13 & 15 & \dots \end{pmatrix}.$$

We see that the sequence repeats (modulo 5) after every five steps and so the value of n is probably equal to five. We can recover the key by simply taking the first five columns of the above permutation.

42

Venona Project – Wikipedia

The Venona project (1943–80) was a counter-intelligence program initiated by the U.S. Army's Signal Intelligence Service (later the National Security Agency). The purpose of the Venona project was the decryption of messages transmitted by the intelligence agencies of the Soviet Union, e.g. the NKVD, the KGB (foreign intelligence) and the GRU (military intelligence). During the 37-year duration of the Venona project, the Signal Intelligence Service decrypted and translated approximately 3,000 messages from Russian to English; among the signals-intelligence yielded was discovery of the Cambridge Five espionage ring in Britain and Soviet espionage of the Manhattan Project in the U.S. The Venona project remained secret for more than fifteen years after it concluded, and some of the decoded Soviet messages were not declassified and published until 1995.

43

Venona Project – Wikipedia

This message traffic, which was encrypted with a one-time pad system, was stored and analyzed in relative secrecy by hundreds of cryptanalysts over a 40-year period starting in the early 1940s. Due to a serious blunder on the part of the Soviets, some of this traffic was vulnerable to cryptanalysis. The Soviet company that manufactured the one-time pads produced around 35,000 pages of duplicate key numbers, as a result of pressures brought about by the German advance on Moscow during World War II. The duplication—which undermines the security of a one-time system—was discovered and attempts to lessen its impact were made by sending the duplicates to widely separated users. Despite this, the reuse was detected by cryptologists in the US.

44

Venona Project – Wikipedia

The decrypted messages gave important insights into Soviet behavior in the period during which duplicate one-time pads were used. With the first break into the code, Venona revealed the existence of Soviet espionage at Los Alamos National Laboratories. Identities soon emerged of American, Canadian, Australian, and British spies in service to the Soviet government, including Klaus Fuchs, Alan Nunn May, and Donald Maclean. Others worked in Washington in the State Department, the Treasury, Office of Strategic Services, and even the White House.

The decrypts show the U.S. and other nations were targeted in major espionage campaigns by the Soviet Union as early as 1942. Among those identified are Julius and Ethel Rosenberg; Alger Hiss; Harry Dexter White, the second-highest official in the Treasury Department; Lauchlin Currie, a personal aide to Franklin Roosevelt; and Maurice Halperin, a section head in the Office of Strategic Services. CHECK ALSO NSA.GOV

45

Block Cipher – Wikipedia

- *In cryptography, a block cipher is a deterministic algorithm operating on fixed-length groups of bits, called blocks, with an unvarying transformation that is specified by a symmetric key. Block ciphers operate as important elementary components in the design of many cryptographic protocols, and are widely used to implement encryption of bulk data.*
- *The modern design of block ciphers is based on the concept of an iterated product cipher. In his seminal 1949 publication, Communication Theory of Secrecy Systems, Claude Shannon analyzed product ciphers and suggested them as a means of effectively improving security by combining simple operations such as substitutions and permutations. Iterated product ciphers carry out encryption in multiple rounds, each of which uses a different subkey derived from the original key.*
- *One widespread implementation of such ciphers, named a Feistel network after Horst Feistel, is notably implemented in the DES cipher. Many other realizations of block ciphers, such as the AES, are classified as substitution-permutation networks.*

46

DES

- *Data Encryption Standard* -- came from IBM
- Certified by NBS (NIST) in 1976 for 10-15 years for cryptographic protection of sensitive, but unclassified computer data
- 56 bit key -- can be broken by NSA, but not by most private concerns
- Encoding and decoding is rapid

47

Contemporary Cryptography

- In 1973 NBS (now NIST) issued a call for a new standard with the following requirements
  - high level of security
  - completely specified and "easy" to understand
  - secrecy should depend on key
  - available to all
  - adaptable
  - efficient and economical
  - exportable
  - verifiable

48

DES

- Two rounds of request were put forward
- In 1975, IBM's algorithm based on Horst Feistel's *Lucifer was* selected to become *DES* (data encryption standard)
- Was certified only for 15 years
- Key length was limited to 56 bits by NBS (originally proposed to be 112 bits)

49

DES

- Has been *decertified*
- In the public domain
- Very widely used
- A secret key system
- Replaced by AES (Advanced Encryption Standard)

50

Public vs. Private Concerns

- The Federal Government does not want criminals to have strong encryption, so it wants to withhold it from the general public
- Forbids export of cryptographic software and systems
  - Believe that only people in the US can write programs!
  - OK to publish research!

51

Public vs. Private Concerns

- The *Clipper Chip*
- Various sorts of key escrow systems
- *Salt II* treaties prohibited encryption to deny telemetry data
- PGP -- Pretty Good Privacy
- Public key cryptosystems have laid the foundations for modern e-commerce

52