

## 15 Adjusting the Columns of a Disk Cipher

As a last application in this chapter we look at the problem: How to adjust the alphabets in the columns of a disk cipher? From Chapter 2 we know that this works only when the primary alphabet is known.

Imagine a ciphertext from a disk cipher whose period  $l$  we know already. Write the ciphertext in rows of length  $l$ . Then the columns are monoalphabetically encrypted, each with (in most cases) another alphabet. By Proposition 5 (iv) we expect a  $\chi$ -value of about  $\frac{1}{n}$  for each pair of columns. Since the alphabets for the columns are secondary alphabets of a disk cipher they differ only by a relative shift in the alphabet. There are 26 different possible shifts. These can be checked by exhaustion: We try all 26 possibilities (including the trivial one, bearing in mind that two columns can have the same alphabet). The perfect outcome would be 25 values near  $\frac{1}{n}$ , and one outcome around the coincidence index of the plaintext language, clearly indicating the true alphabet shift. The experimental results of Section 14 give hope that real outcome should approximate the ideal one in a great number of cases.

### Example 1

Let us try out this idea for the ciphertext from Section 9. We are pretty sure that the period is 7. (And we also adjusted the columns by visual inspection in Chapter 2.) The first two columns are

```
ARCPYMEAZKRWKHZLXRTRTMYRLMTVYCMRBZZKOLKKTOTCUKKOMVBLUYUYZALR
  OEKWMWZZRYZOOTUYURMTYYSOZEKLYVUYBYTZYKOVMYMZVYRORYTYMUWZ
  PZTZLSPLYLVVYYBYMQMWRXZYOKKMYZTZAKQZZT
OMZYDMPYQMHMFKAMMAACDNNZPIMYZHCJSCNCJQMMELEMPNPNPZYSNYHPNMOAM
  CAJMPZIVNMPADAHNKFNAHNVFJHFXYNPNYSYFMKNFMDNPZFGJMVMCMXYZZMQC
  MSYIMVAMKZOANZVSZFKMYEMQHZQMNDPMHDMKIYJF
```

Using the Perl script `adjust.pl` we get the results

Shift:	0	1	2	3	4	5	6
$\chi$ :	0.0499	0.0365	0.0348	0.0285	0.0320	0.0341	0.0298
7	8	9	10	11	12	13	14
0.0416	0.0307	0.0421	0.0402	0.0448	<b>0.0799</b>	0.0495	0.0373
15	16	17	18	19	20	21	22
0.0375	0.0293	0.0330	0.0276	0.0307	0.0306	0.0316	0.0352
23	24	25					
0.0338	0.0461	0.0529					

The result is clear without ambiguity: The correct shift is 12. Going through all  $7 \times 6/2 = 21$  pairs of columns we use the Perl script `coladj.pl` and get results in Table 37 that are consistent with each other and with the results of Chapter 2.

Table 37: *The optimal alphabet shifts for 7 columns*

Column:	0	1	2	3	4	5
1	12					
2	4	18				
3	15	3	11			
4	10	24	6	21		
5	24	12	20	9	14	
6	3	17	25	14	19	5

**Example 2**

The best guess for the period of the short ciphertext of Section 13 was  $l = 20$ . Therefore we consider 20 columns of lengths 8 or 7:

```

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

We have to assume the primary alphabet as known in order to know how to shift the columns, that is, how to identify the distance of the secondary alphabets of two columns relative to each other. The primary alphabet is QWERTZUABCFGHIJKLMNOPSVXY, the complete alphabet table is in Table 38.

The method from Example 1 gives  $20 \times 19/2 = 190$  proposals for optimal shifts between columns. However even for the first columns we already get inconsistent results. We face a complex optimization problem. Instead of continuing with the next columns we better would follow a proposal by SINKOV: Pick up the highest  $\chi$ -values and try to build clusters of fitting columns. But also this approach fails. After several hours off the track we try to understand why.

Let us imagine a plaintext of the same length, written in rows of length 20, columns of length 7 or 8. Take two columns that each have one letter twice and five or six single letters. Shifting the alphabets in such a way that the “twins” become identical letters, they contribute a summand of

$$\frac{4}{49} \approx 0.0818 \text{ for lengths } 7/7, \quad \frac{4}{56} \approx 0.0714 \text{ for } 8/7, \quad \frac{4}{64} \approx 0.0625 \text{ for } 8/8,$$

Table 38: *The alphabet table used in the example*

```

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

```

to the  $\chi$ -value. If accidentally there is another common letter, these values rise to

$$\frac{5}{49} \approx 0.1020 \text{ for lengths } 7/7, \quad \frac{5}{56} \approx 0.0893 \text{ for } 8/7, \quad \frac{5}{64} \approx 0.0781 \text{ for } 8/8.$$

And therefore we'll get many false alarms that will make the task of finding the correct solution very time-consuming. An experiment with plaintext confirms this. Here all shifts should be 0, however we found the maximal  $\chi$ -value for a shift of 0 in less than 20% of all cases.

To get better chances for success we need some known plaintext or more ciphertext or luck. We had luck and got more ciphertext. The following two messages  $b$  and  $c$ ,

```
AWYFN DHZPE PENES YGAVO YHGAD VTNLL TFKKH FHGYT DOGJI HJHHB
OOYFV EWDSJ MOIFY DRTLA BRRFE ZQGYQ AVYCH BQZPR RZTTH IONZE
SCEFH EFJBJ RNRWE TGVZR EYIIQ IZRWT OLGOC ICLFS EMYAH E
```

```
LIGJC KTNLF KBMZH XYWFB UWVPC RNYAJ WEVKV BRVPN PXYOT KVGLE
MBVHE WFZSM UOWFI EYXLB XRRKC XKGPT YONFY DKZLU CXRDC YJWZT
UWPDS VZWNU KORLK WUXUO VWHFL IEGXJ ZUKGC YJVDN EFYDK GJZON
BYXEV EWQSD MMHSS GJ
```

could be encrypted with the same key. Number 1 and 2 have a coincidence index  $\kappa(a, b) \approx 0.0411$  only. But  $\kappa(a, c) \approx 0.0811$ ,  $\kappa(b, c) \approx 0.1027$ . For both  $b$  and  $c$  the period 20 is confirmed by the SINKOV statistic and also by the autocoincidence spectrum. Therefore we align all three messages below each other with rows of length 20. From bad experience we know we should proceed very thoughtfully. Therefore we first look at the letter frequencies in the single columns (of lengths 22 to 25). The columns 2, 3, and 12 contain a letter in 7 exemplars. We try to adjust these columns in such a way that the most frequent letters match. For column 3 relative to column 2 we get a  $\chi$ -value of 0.1072 for a shift of 14, the next  $\chi$ -value being 0.0608. If we write the columns as rows, the result looks like this

```
Column 02: JRYDQYCBYGGIYEIYGVYWNPHYH
Column 03: JYFIQDOYFAJFCFIAJPOFFDFDS
shifted:   RHYEIXBHYPRYVVEPRCBYXXYXD
```

In both cases the most frequent letter with 7 occurrences is Y. For column 12 we get the optimal shift 22 relative to column 2 with a  $\chi$ -value of 0.1273, the next  $\chi$ -value being 0.0836. This also looks good and gives the result

```
Column 02: JRYDQYCBYGGIYEIYGVYWNPHYH
Column 12: MZRMYPYRANKYRTRGMVRRRKKX
shifted:   IWYIPYRJPYQYBINNYXXYGO
```

Also in the shifted column 12 the letter Y occurs 7 times. If we are right, comparing columns 3 and 12 should lead to the same result. Indeed the optimal shift is 8 with  $\chi \approx 0.1109$ , the next  $\chi$ -value being 0.0727.

This makes us confident that we are on the right track, and encourages us to set Y it to plaintext e. We continue our task under the hypothesis that columns 2, 3, and 12 match with the given shifts as

```

...
JRYDQYCBYGGIYEIYG VYWNPHYH
RHYEIXBHYPYVVEPRCBYXYXD
...
IWYIPYRJGPYQYBINNYYYGO
...

```

We take this text fragment as cluster “A” and try to match further columns. First we take columns where the most frequent letters occur 6 or 5 times.

```

A vs 5: Optimal shift is 15 with chi = 0.0906 (next is 0.0683)
A vs 8: Optimal shift is 8 with chi = 0.1092 (next is 0.0758)
A vs 14: Optimal shift is 16 with chi = 0.1092 (next is 0.0859)

```

```

A vs 0: Optimal shift is 23 with chi = 0.0878 (next is 0.0817)
A vs 5: Optimal shift is 0 with chi = 0.0809 (next is 0.0619)
A vs 9: Optimal shift is 21 with chi = 0.0966 (next is 0.0663)

```

The most convincing match is with column 8, therefore we join it to our cluster, forming cluster “B”:

```

...
JRYDQYCBYGGIYEIYG VYWNPHYH
RHYEIXBHYPYVVEPRCBYXYXD
...
BHNRLWGRYPYRKCYJYYYWUE
...
IWYIPYRJGPYQYBINNYYYGO
...

```

Looking at the distribution of letters the Y stands out by far—that is no surprise because we picked columns with the most frequent letters and matched these. As a more meaningful check we transform our cluster to (presumed) plaintext; this means decrypting the fragments with the secondary alphabet that transforms e to Y, that is PSVXYQWERTZUABCDFGHIJKLMNO. This gives the supposed plaintext fragment (to be read top down):

```

...
uiepfeonerrtehtercegyases
isehtdnseaiiecehaioneededp
...
nsyiwgrieeaeivoeeeeeeglh
...
tgetaeiuraefentyyeeerz
...

```

This looks promising. Trying to extend this cluster by a formal procedure is dangerous because there could be columns with a most frequent (plaintext) letter other than *e*. Instead we look at neighboring columns, say at column 4 that should give a readable continuation of columns 2 and 3, in particular extending the digraph *th* in a meaningful way. The proposed shift should have a *Y* (for *e*) as 15th letter, or maybe a *P* (for *a*), or an *R* (for *i*).

Cluster B versus column 4 yields the optimal shift 3 with  $\chi \approx 0.0753$ , the 15th letter being *R* (for *i*). The next best values are  $\chi \approx 0.0664$  for a shift of 12, the 15th letter then being *G* (for *r*), and  $\chi \approx 0.0604$  for a shift of 25, the 15th letter being *Y* (for *e*). To decide between these possible solutions we decrypt the shifted columns and get the proposed cleartext columns

```

zoeiaetpbswhvivrmmwhezye
ixnrjncykbfqeereaavfqnihn
vkaewaplxosdrerrernnisdavua

```

Joining them to columns 3 and 4 the first one looks somewhat inauspicious but possible, the second one looks awkward, the third one looks best and is our first choice. This gives the three adjacent columns

```

uiepfeonerrtehtercegyases
isehtdnseaiiecehaioneededp
vkaewaplxosdrerrernnisdavua

```

and the new cluster “C” of (monoalphabetic) ciphertext, comprising columns 2, 3, 4, 8, 12:

```

...
JRYDQYCBYGGIYEIYGVYWNPHYH
RHYEIXBHYPYVVEPRCBYXYXD
KZPYLPDUMCHXGGYGBBRHXPKJP
...
BHNRLWGRYPYRKCYJYYYWUE
...
IWYIPYRJGPYQYBINNYYYGO
...

```

Note that for joining further columns we must not work with the (proposed) plaintext columns because the transformation between plaintext and ciphertext is not a simple shift.

Comparing the adjacent columns with cluster C we obtain

```
C vs 1: Optimal shift is 1 with chi = 0.0642 (next is 0.0632)
C vs 5: Optimal shift is 15 with chi = 0.0844 (next is 0.0686)
C vs 7: Optimal shift is 20 with chi = 0.0676 (next is 0.0621)
C vs 9: Optimal shift is 6 with chi = 0.0695 (next is 0.0653)
C vs 11: Optimal shift is 5 with chi = 0.0695 (next is 0.0638)
C vs 13: Optimal shift is 23 with chi = 0.0684 (next is 0.0588)
```

The best value seems that for column 13, so let's try this one first (skipping the dead end via column 5). The new cluster D is

```
...
JRYDQYCBYGGIYEIYGVYWNPHYH      uiepfeonerrtehtercegyases
RHYEIXBHYPYVVEPRCBYXXYXD        isehtdnseaiecehaioneededp
KZPYLPDUMCHXGGYGBBRHXPKJP        vkaewaplxosdrerrnnisdavua
...
BHNRLWGRYPYRKCYJYYYWUE           nsyiwgriaeivoeeeeeglh
...
IWYIPYRJGPYQYBINNYYYGO           tgetaeiuraefentyyeerz
EPJVIYDYHBBWXLEHDHAICY           hauctepesnngdwhspsmtoe
...
```

This looks good, and detecting the two th's between the cleartext columns 12 and 13 we try column 14 next.

```
D vs 14: Optimal shift is 16 with chi = 0.0945 (next is 0.0793)
```

If we rely on this result, we get the next cluster E:

```
...
JRYDQYCBYGGIYEIYGVYWNPHYH      uiepfeonerrtehtercegyases
RHYEIXBHYPYVVEPRCBYXXYXD        isehtdnseaiecehaioneededp
KZPYLPDUMCHXGGYGBBRHXPKJP        vkaewaplxosdrerrnnisdavua
...
BHNRLWGRYPYRKCYJYYYWUE           nsyiwgriaeivoeeeeeglh
...
IWYIPYRJGPYQYBINNYYYGO           tgetaeiuraefentyyeerz
EPJVIYDYHBBWXLEHDHAICY           hauctepesnngdwhspsmtoe
PBDCAHPBYCIYIPYPCIPPEPC         anpomasneotetaeotaahao
...
```

Good! Let's continue with column 15:

```
E vs 15: Optimal shift is 0 with chi = 0.0719 (next is 0.0574)
```

Joining the resulting “cleartext” to columns 12, 13, 14 gives the disturbing result

```
tgetaeiuraefentyeeerz
hauctepesnngdwhspsmtoe
anpomasneotetaeotaahao
evkpceqeqhktjtdngdegeh
```

Therefore we dismiss this proposal. Unfortunately also the next best  $\chi$ -value gives no sensible result. On the other hand the shifts giving a possible complement to the **th** have a quite small  $\chi$ -value. Therefore we leave column 15 and retry column 1:

E vs 1: Optimal shift is 1 with chi = 0.0631 (next is 0.0577)

This would give us cluster F:

```
...
FXGRCKGYEIPDXDQNJJEYPPEXGN      qdriovrehtaadpfyuheaahdry
JRYDQYCBYGGIYEIYGWVWNPYH      uiepfeonerrtehtercegyases
RHYEIXBHYPYVVEPRCBYXXYXD      isehtdnseaiecehaioneededp
KZPYLPDUMCHXGGYGBBRHXPKJP      vkaewaplxosdrerrnisdavua
...
BHNRLWGRYPYRKCYJYYWUE          nsyiwgrieaeivoeeeglh
...
IWYIPYRJGPYQYBINNYYYGO          tgetaeiuraefentyeeerz
EPJVIYDYHBBWXLEHDHAICY          hauctepesnngdwhspsmtoe
PBDCAPHBYCIYIPYCIPPEPC          anpomasneotetaeotaahao
...
```

The plaintext now begins with `.quiv...`. A dictionary search finds hits such as “equivalent”, “equivocal”, and “a quiver”. We compare cluster F with column 1 and look for shifts that make the first letter **a** (P in our secondary alphabet) or **e** (Y). We have luck! The optimal shift gives **e**, so we take this as our favourite solution:

F vs 0: Optimal shift is 7 with chi = 0.0717 (next is 0.0696)

and form the next cluster G:



YGCVHCYXIULYIRCCXHEHUHBCY	erocsoedtlwetioodshslnoe
FXGRCKGYEIPXDQNJEYPPXGN	qdriovrehtaadpfyuheaahdry
JRYDQYCBYGGIYEIYGWVWNPYH	uiepfonerrtehtercegyases
RHYEIXBHYPYVVEPRCBYXXYXD	isehtdnseaiecehaioneedep
KZPYLPDUMCHXGGYGBBRHXPKJP	vkaewaplxosdrerrnisdavua
...	
BHNRLWGRYPYRKCYJYYYWUE	nsyiwgriaeivoeeeglh
...	
IWYIPYRJGPYQYBINNYYYGO	tgetaeiuraefentyyeerz
EPJVIYDYHBBWXLEHDHAICY	hautepesngdwhspsmtoe
PBDCAPHBYCIYIPYCIPPEPC	anpomasneotetaeotaahao
...	

Noting the fragments `ciphe` in “line” 4 (fourth column in the schema above) and `ipher` in “line” 14, we cannot resist completing them as `cipher`.

G vs 5: Optimal shift is 11 with  $\chi = 0.0708$  (next is 0.0697)

G vs 19: Optimal shift is 21 with  $\chi = 0.0775$  (next is 0.0585)

Note that we now see how misleading our former results for column 5 were. This is caused by the six a’s in this column that the  $\chi$ -method tried to associate with the e’s of other columns.

Adding both of these results in one step gives cluster H:

YGCVHCYXIULYIRCCXHEHUHBCY	erocsoedtlwetioodshslnoe
FXGRCKGYEIPXDQNJEYPPXGN	qdriovrehtaadpfyuheaahdry
JRYDQYCBYGGIYEIYGWVWNPYH	uiepfonerrtehtercegyases
RHYEIXBHYPYVVEPRCBYXXYXD	isehtdnseaiecehaioneedep
KZPYLPDUMCHXGGYGBBRHXPKJP	vkaewaplxosdrerrnisdavua
YDLKHDYYYGEYVLRRLZMZLYGRWW	alsrolaaandaysesgtgsanecc
...	
BHNRLWGRYPYRKCYJYYYWUE	nsyiwgriaeivoeeeglh
...	
IWYIPYRJGPYQYBINNYYYGO	tgetaeiuraefentyyeerz
EPJVIYDYHBBWXLEHDHAICY	hautepesngdwhspsmtoe
PBDCAPHBYCIYIPYCIPPEPC	anpomasneotetaeotaahao
...	
YAYIAECJYDPVXLRIHYJIZ	emetmhouepacdwitseeutk

We see that column 6 should start with 1 (U). And this is also the “ $\chi$ -optimal” solution:

H vs 6: Optimal shift is 10 with  $\chi = 0.0734$  (next is 0.0554)

And column 7 should start with e (Y):

H vs 7: Optimal shift is 20 with  $\chi = 0.0647$  (next is 0.0639)

We are not amused, also the next best  $\chi$  is unwanted. However the shift that gives **e** has a  $\chi$ -value of 0.0639 that is acceptable. We fill in columns 6 and 7:

YGCVHCYXIULYIRCCXHEHUHBCY	erocsoedtlwetioodshslnoe
FXGRCKGYEIPPXDQNJEYPPXGN	qdriovrehtaadpfyuheaahdry
JRYDQYCBYGGIYEIYGVYWNPHYH	uiepfionerrtehtercegyases
RHYEIXBHYPYVVEPRCBYXYXD	isehtdnseaiecehaioneededp
KZPYLPDUMCHXGGYGBRHXPKJP	vkaewaplxosdrerrnisdavua
YDLKHDYYYGEYVLRZMZLYGRWW	alsrolaaandaysesgtgsanecc
UYRHGCDIVIIYHDPJIRACQJGY	leisroptctespautimofuree
YHUUCBYHIESHIXGEBPAIDPI	esllonesthbstdrhnampat
BHNRLWGRYPYRKCYJYYWUE	nsyiwgriaeivoeeeeeeglh
...	
IWYIPYRJGPYQYBINNYYYGO	tgetaeiuraefentyyeeerz
EPJVIYDYHBBWXLEHDHAICY	hautepesnngdwhspsmtoe
PBDCAPHBYCIYIPYCIPPEPC	anpomasneotetaeotaahao
...	
YAYIAECJYDPVXLRIHYJIZ	emetmhouepacdwitseeutk

It's time, for easier reading, to arrange our findings in the right order where "columns" are columns:

equivalen...tha....e	rdiskless...gan....m
oreeasily...eup....e	ciphersli...tco....t
softworow...atm....m	ovedalong...eea....h
eronpaper...ips....o	denslats
theexacti...uen....u	ltraonthe...rse....e
warisdeba...ano....p	eatedasse...ent....a
tdecrypti...fge....c	iphersadv...edt....d
oftheeuro...nwa....w	oyears
duringthe...the....i	shcontinuu...yso....t
heenigmae...ypt....s	sagessome...esa....e
layedafte...ema....e	shadanupg...eth....u
ndseveral...roa....t	oreduceth...zeo....k
eyspace	

Now its easy to complete the text: In the first row read **equivalent** and complete column 9. In the fourth row read **cipher slide** and complete column 10. Then read **with** in the first row and complete column 11. Then in the last two rows we recognize **the size of ... keyspace**, this allows us to complete column 15. Now in the first two rows we read **cipher disk** and complete the remaining columns 16, 17, 18.

This is the final solution:

equivalentwithaciphe	rdisklesselegantbutm
oreasilymadeupisthe	cipherslideitconsist
softworowsthatmaybem	ovedalongsideeachoth
eronpaperstripsorwoo	denslats
theexactinfluenceofu	ltraonthecourseofthe
warisdebatedanoftrep	eatedassessmentistha
tdecryptionofgermanc	iphersadvancedtheend
oftheeuropeanwarbytw	oyears
duringthewarthebriti	shcontinuallysolvedt
heenigmaencryptedmes	sagessometimesabitde
layedafterthemachine	shadanupgradetheyfou
ndseveralapproachest	oreducethesizeofthek
eyspace	