Queen Mary
University of London

MTH6115 Cryptography

Solutions 2

---

1 The last two lettes are padding, so we remove them (you could also remove them at the end to be sure they are indeed paddings). We then decrypt using the Vigenère key to obtain

```
NGHBP TEBME BECNQ JBGNE HBPLN UUGNQ KECDE NENTZ PTEBF DIH
```

Next we decrypt this using the inverse of the affine map $\Theta_{11,3}$, namely $\Theta_{19,21}$ (check that this is indeed the inverse), to obtain the plaintext:

```
ifyou stopt othin kofit youwi llfin dthat itisc ustom ary
```

2 (a) The answer is

```
SBOC NK XBP JUZ YSWH
```

(b) The answer is

```
SUQW ND ZVP CWT YLYB
```

(c) The answer is

```
SUQW ND ZVP CWT YLYB
```

This is identical to what we obtained in part (b).

(d) The answer is

```
JAYDHPLBWCJQ
```

This is the key that we used in part (c).

(e) The conclusion is that composing two Vigenère ciphers with keys `JHWJHW` (lenght 6) and `ATCU` (length 4) is the same is applying a Vigenère with the single key `JAYDHPLBWCJQ` (length 12). The way we obtained the single key from the other two keys was by just adding them, but note that the initial keys had lengths 6 and 4, so before adding we should make them equal in length by repeating them, that is, we add `JHWJHWJHWJHW` (length 12) to `ATCUATCUATCU` (length 12). Where did 12 come from? Think about it!

Quastion: what if we had used `JHW` instead of `JHWJHW`? Or `ATCUATCU` instead of `ATCU`?

3 (a) The m of the question is 12. Below, we consider composing first two Vigenère ciphers of the same length r, and then Vigenère ciphers of arbitrary lengths r and s.

Let the keyword length be r. If a letter is at position i modulo r, then the first Vigenère cipher will Caesar shift it (say) a_i places to the right, and the second will shift it by b_i places. Their composition (either way round) thus shifts such a letter by $a_i + b_i$ places, and so this composition is a Vigenère cipher with a keyword of length r. If the original keywords were words in a particular language, it is unlikely that the keyword of their composition is also a word in the same language. For example, the composition of Vigenère ciphers with keywords CRIB and CODE has keyword EFLF.

Now a Vigenère cipher with keyword of length r is also a Vigenère cipher with keyword of length kr for any $k \in \mathbb{N}^+$ (just concatenate k copies of the original keyword). Thus Vigenère ciphers with keywords of lengths r and s are also Vigenère ciphers with keywords of length $\mathrm{lcm}(r,s)$, and so their composition is also a Vigenère cipher with keyword of length $\mathrm{lcm}(r,s)$ (or possibly a proper divisor thereof).

Composing Vigenère ciphers of lengths r and s with keywords `aa...aab`, with just one `b` in each case, which occur at positions $\equiv 0$ modulo r and s respectively, yields a Vigenère cipher with a keyword of length $\mathrm{lcm}(r,s)$ with just one `c` in it, as `c` can only occur when `b` occurs in both keywords, that is only at positions whose number is divisible by r and s, and thus is divisible by $\mathrm{lcm}(r,s)$. Thus $\mathrm{lcm}(r,s)$ is minimal.

(b) In general, the answer is n^m.

(c) Arguably, the number of keys is the number of pairs (K_1, K_2), where K_1 is a key of a V_r, and K_2 is the key of a V_s, thus giving the answer $n^r \times n^s = n^{r+s}$ (so n^{10} in question). What I'm looking for is the number of encryptions of any (suffiently long) piece of text, which is *not* the same as the above. The issue is that we can have $v_1 v_2 = v_3 v_4$, where $v_1 \neq v_3$ are V_r's and $v_2 \neq v_4$ are V_s's. So let $v_1 v_2 = v_3 v_4$ where v_1 and v_3 are V_r's and v_2 and v_4 are V_s's. Then $v := v_3^{-1} v_1 = v_4 v_2^{-1}$ is both a V_r and a V_s, and so is a $V_{\gcd(r,s)}$. Thus $v_3 = v_1 v^{-1}$ and $v_4 = v v_2$, and so there

are at most $n^{\gcd(r,s)}$ such pairs (v_3, v_4) for any pair (v_1, v_2). On the other hand if v_1 is a V_r, v_2 is a V_s and v, v' are $V_{\gcd(r,s)}$'s then $v_1 v^{-1}$ is a V_r and vv_2 is a V_s and $v_1 v^{-1}.vv_2 = v_1 v_2$. Moreover, $vv_2 = v'v_2$ implies that $v = v'$ (just right-cancel the v_2's), and so there are at least $n^{\gcd(r,s)}$ relevant pairs (v_3, v_4). So each composition of a V_r and a V_s decomposes as such in precisely $n^{\gcd(r,s)}$ ways, thus giving the required answer as $n^{r+s}/n^{\gcd(r,s)} = n^{r+s-\gcd(r,s)}$ (so n^8 in question).

4 Since we have been given the key length, the first step is to tabulate the number of times each letter appears at positions $\equiv i \pmod 4$ where $i \in \{1, 2, 3, 0\}$. This information is given below.

letter	posns $\equiv 1 \pmod 4$	posns $\equiv 2 \pmod 4$	posns $\equiv 3 \pmod 4$	posns $\equiv 0 \pmod 4$	total
A	3	0	11	0	14
B	0	2	20	17	39
C	17	9	7	2	35
D	3	5	2	4	14
E	6	10	2	7	25
F	8	15	0	24	47
G	28	5	3	4	40
H	8	0	0	6	14
I	10	13	19	12	54
J	16	16	4	14	50
K	16	16	3	0	35
L	0	5	11	2	18
M	2	2	24	11	39
N	8	1	3	2	14
O	3	0	4	10	17
P	14	8	13	14	49
Q	7	0	18	1	26
R	3	20	0	0	23
S	0	2	1	12	15
T	9	1	10	13	33
U	12	9	3	22	46
V	12	26	12	3	53
W	3	1	12	2	18
X	0	4	3	9	16
Y	6	16	0	0	22
Z	0	8	8	2	18

(The last column is irrelevant to the question, but does show that the letter frequencies do not have a uniform distribution.) We try to match the columns up 'by eye' to shifts of the expected distribution of English letters, using the distinctive letter distributions whereby R, S, T (and U) are relatively common, V, W, X, Y, Z are rare, and A and E are common. Looking at the first column suggests that text in positions $\equiv 1 \pmod 4$ has

been shifted by 2 places (in spite of the rarity of plaintext u = ciphertext W). Similarly, by looking at other columns, we get shifts of 17 (a ↦ R), 8 (a ↦ I) and 1 (a ↦ B) for the other cases. So the keyword is probably CRIB. [If you should perform the chi-squared tests (which I did using possibly different letter frequencies to you), you should find they clearly indicate the keyword to be CRIB.]

The plaintext is

```
youo ught tobe asha medo fyou rsel fsai dali ceag reat
girl like yous hemi ghtw ells ayth isto goon cryi ngin
this ways topt hism omen tite llyo ubut shew ento nall
thes ames hedd ingg allo nsof tear sunt ilth erew asal
arge pool allr ound hera bout four inch esde epan drea
chin ghal fdow nthe hall afte rati mesh ehea rdal ittl
epat teri ngof feet inth edis tanc eand sheh asti lydr
iedh erey esto seew hatw asco ming itwa sthe whit erab
bitr etur ning sple ndid lydr esse dwit hapa irof whit
ekid glov esin oneh anda ndal arge fani nthe othe rhec
amet rott inga long inag reat hurr ymut teri ngto hims
elfa shec ameo hthe duch esst hedu ches sohw onts hebe
sava geif ivek epth erwa itin gali cefe ltso desp erat
etha tshe wasr eady toas khel pofa nyon esow hent hera
bbit came near hers hebe gani nalo wtim idvo icei fyou
plea sesi rthe rabb itst arte dvio lent lydr oppe dthe
whit ekid glov esan dthe fana ndsk urri edaw ayin toth
           edar knes sash arda shec ould go
```

Which after adding punctuation and cases you get the following text taken from *Alice's Adventures in Wonderland*:

> 'You ought to be ashamed of yourself,' said Alice, 'a great girl like you,' (she might well say this), 'to go on crying in this way! Stop this moment, I tell you!' But she went on all the same, shedding gallons of tears, until there was a large pool all round her, about four inches deep and reaching half down the hall.
>
> After a time she heard a little pattering of feet in the distance, and she hastily dried her eyes to see what was coming. It was the White Rabbit returning, splendidly dressed, with a pair of white kid gloves in one hand and a large fan in the other: he came trotting along in a great hurry, muttering to himself as he came, 'Oh! the Duchess, the Duchess! Oh! won't she be savage if I've kept her waiting!' Alice felt so desperate that she was ready to ask help of any one; so, when the Rabbit came near her, she began, in a low, timid voice, 'If you please, sir–' The Rabbit started violently, dropped the white kid gloves and the fan, and skurried away into the darkness as hard as he could go.

5 We should look at the differences between these numbers and see what common factors they have. For instance, within the group $\{1, 89, 201, 289\}$ the difference between every two of them is divisible by 8. The difference $15 - 1 = 14$ can only have a common factor of 2 with the previous pairs, so it is unlikely to be relevant, so probably 15 is an outlier. Similarly, you can convince yourself that 320 is probably an outlier too. So the best guess for the key length is 8.

6 The frequent trigrams occur as follows.

trigram	frequency	positions	gcd(posn diffs)
BWL	3	26, 92, 104	6
ILV	3	141, 153, 207	6
IOI	3	171, 201, 213	6
LVJ	3	142, 208, 418	6

The calculations for the first two rows are: $\gcd(92-26, 104-92) = \gcd(66, 12) = 6$ and $\gcd(153-141, 207-153) = \gcd(12, 54) = 6$. Thus the Babbage–Kasiski method suggests a keyword of length 6. (The Friedman method does too.) Note that the Babbage–Kasiski method is usually *not* as decisive as this.

Now tabulate the number of times each letter appears at positions $\equiv i \pmod 6$ where $i \in \{1, 2, 3, 4, 5, 0\}$. This information is given below.

letter	posns $\equiv 1\ (6)$	posns $\equiv 2\ (6)$	posns $\equiv 3\ (6)$	posns $\equiv 4\ (6)$	posns $\equiv 5\ (6)$	posns $\equiv 0\ (6)$	total
A	1	5	1	6	0	0	13
B	1	15	2	1	0	1	20
C	5	3	6	0	2	3	19
D	1	0	7	1	0	3	12
E	5	2	5	1	8	3	24
F	2	0	0	1	2	5	10
G	8	4	2	0	2	0	16
H	2	0	4	9	2	0	17
I	4	8	10	2	9	6	39
J	1	1	5	2	1	8	18
K	6	2	1	2	3	3	17
L	0	0	1	16	4	2	23
M	0	6	0	1	5	2	14
N	3	3	3	1	1	3	14
O	3	3	0	9	0	0	15
P	9	5	6	7	10	2	39
Q	4	4	3	0	0	0	11
R	3	0	2	0	10	8	23
S	0	1	4	4	3	2	14
T	7	2	6	1	1	3	20
U	8	1	0	2	0	2	13
V	6	5	2	6	7	12	38
W	1	4	8	1	5	0	19
X	1	6	2	1	4	3	17
Y	1	0	0	4	1	3	9
Z	0	2	1	3	1	7	14

(The last column is irrelevant to the question, but does show that the letter frequencies do not have a uniform distribution.) We try to match the columns up 'by eye' to shifts of the expected distribution of English letters, using the distinctive letter distributions whereby R, S, T (and U) are relatively common, V, W, X, Y, Z are rare, and A, E and I are common. Looking at the first (real) column suggests that text in positions $\equiv 1 \pmod 6$ has been shifted by 2 places ($\mathtt{a} \mapsto \mathtt{C}$), in spite of the rarity of plaintext $\mathtt{u}$ = ciphertext $\mathtt{W}$.

Similarly, by looking at other columns, we get shifts of 8 ($\mathtt{a} \mapsto \mathtt{I}$), 15 ($\mathtt{a} \mapsto \mathtt{P}$), 7 ($\mathtt{a} \mapsto \mathtt{H}$), 4 ($\mathtt{a} \mapsto \mathtt{E}$) and 17 ($\mathtt{a} \mapsto \mathtt{R}$) for the other cases. So the keyword is probably CIPHER. If you should perform the various statistical tests, you should find they clearly indicate the keyword to be CIPHER. (This is even the case, though not quite as clearly, with Gadsby or French statistics, with frequency assignments of 0.02% to the zero-frequency letters. The values of the statistics differ somewhat from what you get when more usual frequencies are employed.)

The decryption with keyword CIPHER makes sense, and the decrypted text is below. (It is also usually quite easy to detect if one or two letters are wrong in the keyword.)

“The quick brown fox jumps over the lazy dog” is an English-language pangram, that is, a phrase that contains all of the letters of the alphabet. It has been used to test typewriters and computer keyboards, and in other applications involving all of the letters in the English alphabet. Owing to its shortness and coherence, it has become widely known and is often used in visual arts.

Short pangrams are probably slightly annoying when it comes to decrypting ciphers, but I’m sure you will cope anyway. There are ciphers that are much more badly affected by such language games than the Vigènere cipher.