Queen Mary
University of London

# MTH6115 Cryptography

## Solutions 1

---

**1** a) In the ciphertext

`ORJNE RGURV QRFBS ZNEPU`

the most common letters are R (4) and N, E, U (2). So lets try the Caeser ciphers which take e to either R, N, E, or U as a first attempt. Taking e to R corresponds to a cyclic shift of 13 letters to the right and deciphers the ciphertext as

`beware the ides of march`

b) Both compositions are equivalent to a single substitution cipher, so they are equally difficult.

c) Yes, it does depend on the order. For example, consider the Caesar shift by $b = 1$ and the substitution cipher the reverses the English alphabet. If we apply the Caesar shift first, the resulting cipher encrypts `a` to `Y`. If we apply the Caesar shift last, the resulting cipher encrypts `a` to `A`.

**2** You should apply frequency analysis. The original message, of length 417, is:

```
KIRKY PIKTK IIPOX SZTPM UZIHK SSUNB SKITK RFKRP IFXSD RUEXR
FXYHG EYKOP RGRFX MXKYX IRIRK YRUXK YRFPI RFXIZ MQFPA FPIRF
XIUZY AXUNT UIRUN RFXXM XYEGU MXKYR FURFX YIRKY IKYXO PIPHS
XPMRF XMPEF RICGQ FXMRF XGKYX MURUZ RIFUM XHGRF XIZMF PIRUY
PAKSS GRFXT UIRBY UTPMX MRIRK YIUMR FXAXS XIRPK SIBFX YXQXY
XEYUZ BXDRU EXRFX YPMRU AUMIR XSSKR PUMIK MDRFX HYPEF RXIRI
RKYIE KPMXD BYUBX YMKTX IXVRX MIPOX AKRKS UEZXI UNIRK YIFKO
XHXXM KIIXT HSXDH GKIRY UMUTX YIQFP AFBYU OPDXI RKMDK YDPLX
DIRKY DXIPE MKRPU MI
```

The frequencies of the letters are as follows. The doubled letters that occur (`XX`, `II` and `SS` only) are also shown.

| X | R | I | K | Y | U | F | M | P | S | D | E |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 55 | 47 | 44 | 34 | 30 | 29 | 28 | 26 | 26 | 14 | 10 | 10 |
| T | G | H | Z | A | B | O | N | Q | C | L | V |
| 9 | 8 | 8 | 8 | 7 | 7 | 6 | 4 | 4 | 1 | 1 | 1 |
| J | W | XX | II | SS | | | | | | | |
| 0 | 0 | 2 | 2 | 3 | | | | | | | |

From this you suspect that X = e and either R or I decrypts to t. There is a small possibility that X does not decrypt to e; but really two of X, R, I should decrypt to e and t. Let us study some triples.

| triple | number of such |
|---|---|
| R*X | 14 (RFX 13 times and RUX once) |
| I*X | 3 (IFX, IRX and IIX) |
| X*R | 8 (XIR four times, XMR twice, XDR, XVR) |
| X*I | 4 (XYI twice, XMI and XDI) |
| R*I | 2 (RPI and RXI) |
| I*R | 0 |

This makes RFX = the pretty convincing, so let's substitute R = t, F = h and X = e. (Only 9 of the 13 occurences of the turn out to be the word *the*.) This gives the following.

```
KItKY PIKTK IIPOe SZTPM UZIHK SSUNB SKITK thKtP IheSD tUEet
heYHG EYKOP tGthe MeKYe ItItK YtUeK YthPI theIZ MQhPA hPIth
eIUZY AeUNT UItUN theeM eYEGU MeKYt hUthe YItKY IKYeO PIPHS
ePMth eMPEh tICGQ heMth eGKYe MUtUZ tIhUM eHGth eIZMh PItUY
PAKSS GtheT UItBY UTPMe MtItK YIUMt heAeS eItPK SIBhe YeQeY
eEYUZ BeDtU Eethe YPMtU AUMIt eSSKt PUMIK MDthe HYPEh teItI
tKYIE KPMeD BYUBe YMKTe IeVte MIPOe AKtKS UEZeI UNItK YIhKO
eHeeM KIIeT HSeDH GKItY UMUTe YIQhP AhBYU OPDeI tKMDK YDPLe
DItKY DeIPE MKtPU MI
```

Getting the next few letters after this tends to be the most difficult part of cracking a substitution cipher. Note that if we know where the spaces are the task is much easier. For example the first sentence is:

```
K ItKY PI K TKIIPOe, SZTPMUZI HKSS UN BSKITK thKt PI heSD
tUEetheY HG EYKOPRG.
```

The words K and thKt immediately yield K = a. However, the only two occurrences of th*t (SKITK thKtP IheSD on the 1st line and MeKYt hUthe YItKY on the 3rd line) have different letters for the *, so there does not seem to be a lot we can say without the word spaces to help us. Even after we identify that I = s, we cannot say much more than that K = a is now favoured by frequency analysis.

Now the block ItItK in the 2nd row forces I to be a vowel (other than e of course), y or s, noting that a word/sentence boundary can (and does) occur within

this block. Doubled letters appear quite infrequently [relative to expectation] in the text, and II occurs twice, so it is unlikely that I is a, i, u or y. Now eo is infrequent in English, even across word boundaries, and eI occurs 10 times in the text, which thus suggests that I = s (rather than o). (In fact es occurs more often than one would expect.) The block teItI in Row 6 backs up the hypothesis that I = s. (Note that s is less common than o, but not unreasonably so, so it is not absurd that it be the third most common letter in the plaintext.) Substituting I = s gives the following.

```
KstKY PsKTK ssPOe SZTPM UZsHK SSUNB SKsTK thKtP sheSD tUEet
heYHG EYKOP tGthe MeKYe ststK YtUeK YthPs thesZ MQhPA hPsth
esUZY AeUNT UstUN theeM eYEGU MeKYt hUthe YstKY sKYeO PsPHS
ePMth eMPEh tsCGQ heMth eGKYe MUtUZ tshUM eHGth esZMh PstUY
PAKSS GtheT UstBY UTPMe MtstK YsUMt heAeS estPK SsBhe YeQeY
eEYUZ BeDtU Eethe YPMtU AUMst eSSKt PUMsK MDthe HYPEh tests
tKYsE KPMeD BYUBe YMKTe seVte MsPOe AKtKS UEZes UNstK YshKO
eHeeM KsseT HSeDH GKstY UMUTe YsQhP AhBYU OPDes tKMDK YDPLe
DstKY DesPE MKtPU Ms
```

Now there is big cut off between the 9th and 10th most frequent letters in the ciphertext, which actually mirrors a similar discontinuity in the English letter frequencies. So frequency analysis would suggest that $\{\mathtt{K},\mathtt{Y},\mathtt{U},\mathtt{M},\mathtt{P}\} = \{\mathtt{a},\mathtt{o},\mathtt{i},\mathtt{n},\mathtt{r}\}$ (which does turn out to be the case). It does not seem we can positively identify any of these letters at this juncture. However, we can try to identify some vowels.

| text | where found | conclusion |
|---|---|---|
| KstKY | row 1 (start) | $\mathtt{K} \in \{\mathtt{a},\mathtt{i},\mathtt{o},\mathtt{u},\mathtt{y}\}$ |
| tGthe | row 2 | $\mathtt{G} \in \{\mathtt{a},\mathtt{i},\mathtt{o},\mathtt{u},\mathtt{y}\}$ |
| YthPs thesZ | row 2 | $\mathtt{P} \in \{\mathtt{a},\mathtt{i},\mathtt{o},\mathtt{u},\mathtt{y}\}$ |
| MeKYt hUthe | row 3 | $\mathtt{U} \in \{\mathtt{a},\mathtt{i},\mathtt{o},\mathtt{u},\mathtt{y}\}$ |

In fact, since the text commences with KstKY, this strongly suggests that $\mathtt{K} \in \{\mathtt{a},\mathtt{i}\}$, with an inclination towards K = a. (Thus with frequency analysis, we might posit that $\{\mathtt{K},\mathtt{U},\mathtt{P}\} = \{\mathtt{a},\mathtt{o},\mathtt{i}\}$ and $\{\mathtt{Y},\mathtt{M}\} = \{\mathtt{n},\mathtt{r}\}$, but we shall not rely much on this.)

There are two sequences ht in the text, namely eMPEh tsCGQ on the 4th row, and HYPEh t on the 6th row. The 4th row example is very interesting, for words in English seldom commence *ts-*, and few (if any) of these has a vowel or y as its 4th letter (which would be G here). So the ht sequence on the 4th row almost certainly occurs within a word, which makes it very very likely that Eht = ght, especially since both *ht sequences are Eht. The sequences ***ht are MPEht and YPEht, which are different. Since most **ught* sequences in English are *ought* (but what about *caught*, *slaughter* and so on?) and also that P is fairly common, we incline towards P = i, which would now suggest that K = a. Substituting E = g, P = i and K = a gives the text below.

```
astaY isaTa ssiOe SZTiM UZsHa SSUNB SasTa thati sheSD tUget
```

```
heYHG gYaOi tGthe MeaYe ststa YtUea Ythis thesZ MQhiA histh
esUZY AeUNT UstUN theeM eYgGU MeaYt hUthe YstaY saYeO isiHS
eiMth eMigh tsCGQ heMth eGaYe MUtUZ tshUM eHGth esZMh istUY
iAaSS GtheT UstBY UTiMe Mtsta YsUMt heAeS estia SsBhe YeQeY
egYUZ BeDtU gethe YiMtU AUMst eSSat iUMsa MDthe HYigh tests
taYsg aiMeD BYUBe YMaTe seVte MsiOe AataS UgZes UNsta YshaO
eHeeM asseT HSeDH GastY UMUTe YsQhi AhBYU OiDes taMDa YDiLe
DstaY Desig MatiU Ms
```

The end of the message is `Desig MatiU Ms`. Given that we now think that `U` = `o` anyway, and we know that `U` and `M` are both common, setting `U` = `o` and `M` = `n` looks right, especially since *-tion*, *-tions* are common word endings. This would make `DesigMatiUMs` into `designations` or `resignations`, probably the former from frequency analysis. The end of the 6th row is now `ionsa nDthe HYigh tests`, which really does force `D` = `d`. The sequence `tUgetheY` = `togetheY` occurs twice, which suggests that `Y` = `r`, as does the frequency analysis. The substitutions `UMDY` = `ondr` now give us.

```
astar isaTa ssiOe SZTin oZsHa SSoNB SasTa thati sheSd toget
herHG graOi tGthe neare ststa rtoea rthis thesZ nQhiA histh
esoZr AeoNT ostoN theen ergGo neart hothe rstar sareO isiHS
einth enigh tsCGQ henth eGare notoZ tshon eHGth esZnh istor
iAaSS GtheT ostBr oTine ntsta rsont heAeS estia SsBhe reQer
egroZ Bedto gethe rinto Aonst eSSat ionsa ndthe Hrigh tests
tarsg ained BroBe rnaTe seVte nsiOe AataS ogZes oNsta rshaO
eHeen asseT HSedH Gastr onoTe rsQhi AhBro Oides tanda rdiLe
dstar desig natio ns
```

So far we have found the 11 letters `RFXI EPK UMDY` = `thes gia ondr`. The remaining 13 letters in the ciphertext decrypt as:

`ZGHO BASTN QCLV` = `uybv pclmf wkzx`,

which is the order I found them in, though it is quite easy to proceed from here, and in lots of different ways too. In the key I had `JW` = `qj` (rather than `jq`), but you cannot know this, as `J` and `W` are not in the ciphertext. The deciphered text, with spaces removed, follows.

```
astarisamassiveluminousballofplasmathatisheldtoget
herbygravitythenearestartoearthisthesunwhichisth
esourceofmostoftheenergyonearthotherstarsarevisibl
einthenightskywhentheyarenotoutshonebythesunhistor
icallythemostprominentstarsonthecelestialspherewer
egroupedtogetherintoconstellationsandthebrightests
tarsgainedpropernamesextensivecataloguesofstarshav
ebeenassembledbyastronomerswhichprovidestandardize
dstardesignations
```

Now render it in good English by putting the spaces and punctuation back in, to get something like what is below.

A star is a massive, luminous ball of plasma that is held together by gravity. The nearest star to Earth is the Sun, which is the source of most of the energy on Earth. Other stars are visible in the night sky, when they are not outshone by the Sun. Historically, the most prominent stars on the celestial sphere were grouped together into constellations, and the brightest stars gained proper names. Extensive catalogues of stars have been assembled by astronomers, which provide standardized star designations.

Source of text: http://en.wikipedia.org/wiki/Star, first paragraph, as it was on 13th January 2010 at about 8pm.

This cipher is considerably easier if I leave the spaces in. Try it!

```
K IRKY PI K TKIIPOX, SZTPMUZI HKSS UN BSKITK RFKR PI FXSD RUEXRFXY
HG EYKOPRG. RFX MXKYXIR IRKY RU XKYRF PI RFX IZM, QFPAF PI RFX
IUZYAX UN TUIR UN RFX XMXYEG UM XKYRF. URFXY IRKYI KYX OPIPHSX PM
RFX MPEFR ICG, QFXM RFXG KYX MUR UZRIFUMX HG RFX IZM.
FPIRUYPAKSSG, RFX TUIR BYUTPMXMR IRKYI UM RFX AXSXIRPKS IBFXYX
QXYX EYUZBXD RUEXRFXY PMRU AUMIRXSSKRPUMI, KMD RFX HYPEFRXIR IRKYI
EKPMXD BYUBXY MKTXI. XVRXMIPOX AKRKSUEZXI UN IRKYI FKOX HXXM
KIIXTHSXD HG KIRYUMUTXYI, QFPAF BYUOPDX IRKMDKYDPLXD IRKY
DXIPEMKRPUMI.
```

**3** There is no point in applying Euclid's Algorithm to find inverses modulo $n$ for small $n$ like $13, 17, 26, 34$; trial and error is probably faster. A number $a$ is invertible modulo $n$ if and only if $\gcd(a, n) = 1$. In particular, the even numbers are not invertible modulo 26 or 34.

(a)

| Number | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Inverse | None | 1 | 7 | 9 | 10 | 8 | 11 | 2 | 5 | 3 | 4 | 6 | 12 |

(b)

| Number | 1 | 3 | 5 | 7 | 9 | 11 | 15 | 17 | 19 | 21 | 23 | 25 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Inverse | 1 | 9 | 21 | 15 | 3 | 19 | 7 | 23 | 11 | 5 | 17 | 25 |

(c)

| Number | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Inverse | None | 1 | 9 | 6 | 13 | 7 | 3 | 5 | 15 | 2 | 12 | 14 | 10 | 4 | 11 | 8 | 16 |

(d)

| Number | 1 | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 19 | 21 | 23 | 25 | 27 | 29 | 31 | 33 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Inverse | 1 | 23 | 7 | 5 | 19 | 31 | 21 | 25 | 9 | 13 | 3 | 15 | 29 | 27 | 11 | 33 |

(e) **Solution 1.** Cnnsider the affine map $\Theta_{a,0} : \mathbb{Z}_n \to \mathbb{Z}_n$, namely $\Theta_{a,0}(x) = ax$. By a theorem from the lectures, this map is bijective if and only if $\gcd(a, n) = 1$. We now show that $\Theta_{a,0}$ is bijective if and only if $a$ has an inverse modulo $n$. If $\Theta_{a,0}$ is bijective, then there is some $a'$ such that $\Theta_{a,0}(a') = 1$, so $aa' \equiv 1$

modulo $n$. Conversely, if $a$ has an inverse $a'$ modulo $n$, then $\Theta_{a',0}$ is inverse to $\Theta_{a,0}$, so $\Theta_{a,0}$ is invertible.

**Solution 2.** If $a$ has an inverse modulo $n$, then there exists $x$ such that $ax-1$ is divisible by $n$, that is $ax-1=ny$ for some integer $y$. So $ax+ny=1$. Now, if $\gcd(a,n)=d>1$, then $d$ divides both $a$ and $n$, so it divides $ax+ny=1$, a contradiction. If $\gcd(a,n)=1$, then we know (e.g., by Euclid's Algorithm) that there are $x$ and $y$ such that $ax+ny=1$. This implies that $ax\equiv 1$ modulo $n$. Therefore, $a$ has an inverse modulo $n$.

(f) Here is the working for Euclid's Algorithm.

$$\begin{aligned} 241 &= 17\cdot 14+3 \\ 17 &= 3\cdot 5+2 \\ 3 &= 2\cdot 1+1 \end{aligned}$$

Now work backwards from the last equation and write 1 as a linear combination of 241 and 17.

$$\begin{aligned} 1 &= 1\cdot 3-1\cdot 2 \\ 1 &= 1\cdot 3-(17-3\cdot 5)\cdot 1=(-1)\cdot 17+6\cdot 3 \\ 1 &= (-1)\cdot 17+6(241-17\cdot 14)=6\cdot 241+(-85)\cdot 17 \end{aligned}$$

Thus, $-85\cdot 17\equiv 1 \pmod{241}$, so the inverse of 17 is $-85$, i.e. 156.

**4** Recall that to find the inverse of $\Theta_{a,b}$ in the form $\Theta_{a',b'}$, we have to solve the congruences

$$\begin{aligned} aa' &\equiv 1 \pmod{n}, \\ ba'+b' &\equiv 0 \pmod{n}. \end{aligned}$$

(a) Using the previous exercise, the inverse of 5 (mod 13) is $a'=8$. Substitute this in the second equation to obtain $b'\equiv -(4)(8)=7 \pmod{13}$. So, modulo 13, the inverse of $\Theta_{5,4}$ is $\Theta_{8,7}$.

(b) Using the previous exercise, the inverse of 5 (mod 26) is $a'=21$. Substitute this in the second equation to obtain $b'\equiv -(4)(-5)=20 \pmod{26}$. So, modulo 26, the inverse of $\Theta_{5,4}$ is $\Theta_{21,20}$.

(c) Using the previous exercise, the inverse of 5 (mod 17) is $a'=7$. Substitute this in the second equation to obtain $b'\equiv -(4)(7)=6 \pmod{17}$. So, modulo 17, the inverse of $\Theta_{5,4}$ is $\Theta_{7,6}$.

Do you notice a pattern here? Can you quickly (in your head!) find the inverse of $\Theta_{6,5}$ (mod 35)? What about $\Theta_{6,4}$ (mod 35)?

**5** All we need to do is to find two elements of $G=\mathrm{AGL}_1(\mathbb{Z}_n)$ which do not commute. Going for an easy example, I pick $g=\Theta_{1,1}$ and $h=\Theta_{-1,0}$. For all $n$,

the elements 1 and $-1$ have multiplicative inverses modulo $n$ (namely themselves), so these maps are permutations. We have $hg = \Theta_{-1,-1}$ and $gh = \Theta_{-1,1}$, and since $n \geqslant 3$, these maps are distinct. (Note that there are pairs of elements $g, h \in G$ such that $gh = hg$, for example we can take $g = h = \Theta_{1,-1}$.)

**6** You should apply frequency analysis. The original message is:

```
JZQE TTCR QKCX ZQDC GEJY XQUT OURU ELXZ EHQJ CKGM HGJC JMJC
URKC XZQD SZQD QCRQ EKZL QJJQ DCRE RELX ZEHQ JCGO EXXQ NJUC
JGRM OQDC KQAM CPEL QRJQ RKDY XJQN MGCR WEGC OXLQ OEJZ QOEJ
CKEL TMRK JCUR ERNK URPQ DJQN HEKI JUEL QJJQ DXXX
```

The letters with highest frequencies are: `Q`:24, `J`:20, `C`:17 and `E`:17. So you suspect that `Q` = `e` and `J` = `t`. If we consider the bijection between the English alphabet and $\mathbb{Z}_{26} = \{0, 1, 2, \ldots, 25\}$, we have `E` $= 4$, `J` $= 9$, `Q` $= 16$, and `T` $= 19$. To decrypt the message we use the affine susbtitution $\Theta_{\alpha,\beta}$ sending `Q` to `e` and `J` to `t`. This means

$$\begin{cases} 16\alpha + \beta \equiv 4 & (\text{mod } 26) \\ 9\alpha + \beta \equiv 19 & (\text{mod } 26) \end{cases}$$

Solving this system of equations, we find $\alpha \equiv 9 \pmod{26}$ and $\beta \equiv 16 \pmod{26}$. Applying $\Theta_{9,16}$ to the ciphertext gives us the plaintext:

```
thea ffin ecip heri saty peof mono alph abet icsu bsti tuti
onci pher wher eine achl ette rina nalp habe tism appe dtoi
tsnu meri cequ ival ente ncry pted usin gasi mple math emat
ical func tion andc onve rted back toal ette rppp
```

which is part of the Wikipedia entry for Affine cipher:

`http://en.wikipedia.org/wiki/Affine_cipher`

**7** The trouble here is that the encrypt map is not invertible, and each possible ciphertext letter has 2 letters that can encrypt to it. The decryption key is:

| Ciphertext | B | D | F | H | J | L | N | P | R | T | V | X | Z |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Plaintext | an | hu | bo | iv | cp | jw | dq | kx | er | ly | fs | mz | gt |

Let us write out the ciphertext with the possible decrypts underneath.

```
HZRRB TTTNF RVDRT JFFFH VTFDB THJRJ BBXBP RDRRV DFVZH ZDZHF
igeea llldb efhel cbbbi flbha licec aamak eheef hbfgi ghgib
vtrrn yyyqo rsury pooov syoun yvprp nnznx rurrs uostv tutvo

BJHJD RRHBL RJZHH RFZDR RLHVR DRLHT TDBHR HXXRB VRNHV VHJDT
acich eeiaj ecgii ebghe ejife hejil lhaie immea fedif fichl
npvpu rrvnw rptvv rotur rwvsr urwvy yunvr vzzrn srqvs svpuy

ZTNRJ HJDRR HBZXR VVBZR VRBJR TJZRN DVHBZ HZ
gldec ichee iagme ffage feace lcged hfiag ig
tyqrp vpurr vntzr ssntr srnpr yptrq usvnt vt
```

I expect everyone to have got this far. Bob now has to decide which alternative decryption holds at each position. This is English text, and there are enough footholds to allow Bob to extract the original text, possibly with the help of some inspired guesswork, but if Alice had transmitted a random letter string then Bob would have no chance. The decrypted text reads:

> It really does help Bob if (you) Alice can make her substitution cipher injective. Otherwise he will have immense difficulty deciphering messages encrypted using it!

(The spurious 'you' is a mistake borne of trying to get a good wording for the above passage.)