MATH 75: Cryptography	Spring 2022
PSET 1 — April 8, 2022	
Prof. Asher Auel	***Student: Amittai Siavava***

Credit Statement

I worked on these problems alone, with reference to class notes and the following books:

(a) ***The Code Book*** by **Simon Singh**.

(b) ***Cryptography*** by **Simon Rubinsen-Salzedo**

I also wrote some code to help in automating some of the problems.
Rather than upload all the source code, I tried to demonstrate the logic in my responses.
Please let me know if you would like to see the source-code.

Problems

1. What is the message embedded in the following?

> Dear George,
> Greetings to all at Oxford. Many thanks for your
> letter and for the Summer examination package.
> All Entry Forms and Fees Forms should be ready
> for final despatch to the syndicate by Friday
> 20th or at the very latest, I'm told, by the 21st.
> Admin has improved here, though there's room
> for improvement still; just give us all two or three
> more years and we'll really show you! Please
> don't let these wretched 16+ proposals destroy
> your basic O and A pattern. Certainly this
> sort of change, if implemented immediately,
> would bring chaos.

I thought it was unlikely that the text was actually encrypted (a strategy that did not work that well on problem 2). On this problem, I tried to look at the words in different positions in the text – every first word in a sentence, every last word, every nth word, in sentence n, etc.
Using this strategy, I deciphered this message (highlighted in red above):

`Your package ready Friday 21st, room three. Please destroy this immediately.`

2. In one of Dorothy Sayers' mysteries, Lord Peter is confronted with the following message:

```
I thought to see the fairies in the fields, but I saw only the evil
elephants with their black backs.  Woe! how that sight awed me!  The
elves danced all around and about while I heard voices calling clearly.
Ah! how I tried to see--throw off the ugly cloud--but no blind eye of a
mortal was permitted to spy them.  So then came minstrels, having gold
trumpets, harps and drums.  These played very loudly beside me,
breaking that spell.  So the dream vanished, whereat I thanked Heaven.  I
shed many tears before the thin moon rose up, frail and faint as a sickle of
straw.  Now though the Enchanter gnash his teeth vainly, yet shall he
return as the spring returns.  Oh, wretched man!  Hell gapes, Erebus now
lies open.  The mouths of Death wait on thy end.
```

He also discovers the key to the message, which is a sequence of integers:

787656543432112343456567878878765654343211234345656787887876565443321 1234

(a) Decrypt the message. *Hint. What is the largest integer value?*

The largest integer value is 8. All integer values in the key are non-zero, so it's unlikely that the mapping field includes 0. Using this strategy, I wrote a simple program that split the message into 8 lists of letters resulting in a matrix with 73 columns and 8 rows:

```
ITFNDWEHHAWAWECOBEOLRISOLUDOPDECTIRAREEYEGLRSEKNNETRAICRONGTNAUSTRAPSOOEO
TOATSOVATCOTELEUOIIILTEFYTERETMARNURUPRBBTLEHAEIYFHOINKAUCNELLRPUENENPUAN
HSIHBNINHKESDVDNUHCNYREFCNYTROSMEGMPMLYERHSAETDSTOISLTLWGHAEYLNRRTHSOETTT
OEREULLTEBHIMEADTEEGAITTLOEAMSOELGPSSALSEAOMDIHHERNEAAENHASTYHAINCEEWNHHH
UEIFTYESIAOGESLAWASCHEHHOBOLIPTMSOEATYOIATTVWTEEAEMUNSOOTNHHEESNSHLRLTSWY
GTEIITLWRCWHTDLNHRCLHDREULFWTYHIHLTNHEUDKSHAHHADRTOPDAFWHTHVTRTGOELEIHOAE
HHSESHEIBKTTHAADIDAEOTOUDIAATTENADSDEDDEIPENEAVMSHOFFSSTEEIASEHRHDGBEEFIN
TEILAEPTLSHAENRALVLAWOWGBNMSEHNSVTHDSVLMNEDIRNEABENRAITHERSIHTEEWMAUSMDTD
```

Using the key to pick out rows for each column — or word position in the sentence — I picked out the following message:

```
hesittethbetweenthecherubimstheisles
maybegladthereofastheriversinthesouth
```

When spaced out properly, the above reads:

```
He sitteth between the cherubims the isles
may be glad thereof as the rivers in the south
```

(b) If the algorithm is known, but not the key, how secure is this encryption scheme?

This scheme employs arbitrary substitutions to encrypt the message. This is more secure than other methods such as standard substitution ciphers.
Even if the algorithm is known, but not the key, then there are at least 8^n possibilities (assuming the interceptor knows enough to map the letters over 8-space), or 26^n if the interceptor does not know enough information to map the letters over the reduced space. These are both negligible possibilities — it would take immense computational power to decrypt such a message. However, that's perhaps *still* not as secure as modern methods including **RSA** and **AES**.

(c) If the key is known, but not the algorithm, how secure is this encryption scheme?

> A problem with this method is that the key, if known, gives away important information about how the message was encrypted. For instance, we were able to reduce the complexity from 26^n to 8^n just by exploiting a single fact about the encryption key.
> Knowing the full key made decrypting the entire message a trivial problem.

3. The message

`ofoxdryeqrsgkvudrbyeqrdrofkvvoiypdrocrknygypnokdrspokbxyofsv`

was encrypted using a shift cipher. Decrypt the message.

> Decrypting shift ciphers is trivial, since there is only a small set of possible actions that apply to every letter in the text.
> I wrote a program implementing the basic mechanics of shift ciphers.
> Running the text through the program, I got the following message with a shift of 10:

```
even though i walk through the valley of the
shadow of death i fear no evil
```

Full output

```
0: ofoxdryeqrsgkvudrbyeqrdrofkvvoiypdrocrknygypnokdrspokbxyofsv
1: nenwcqxdpqrfjutcqaxdpqcqnejuunhxocqnbqjmxfxomnjcqronjawxneru
2: mdmvbpwcopqeitsbpzwcopbpmdittmgwnbpmapilwewnlmibpqnmizvwmdqt
3: lcluaovbnopdhsraoyvbnoaolchsslfvmaolzohkvdvmklhaopmlhyuvlcps
4: kbktznuamnocgrqznxuamnznkbgrrkeulznkyngjuculjkgznolkgxtukbor
5: jajsymtzlmnbfqpymwtzlmymjafqqjdtkymjxmfitbtkijfymnkjfwstjanq
6: izirxlsyklmaepoxlvsyklxlizeppicsjxliwlehsasjhiexlmjievrsizmp
7: hyhqwkrxjklzdonwkurxjkwkhydoohbriwkhvkdgrzrighdwklihduqrhylo
8: gxgpvjqwijkycnmvjtqwijvjgxcnngaqhvjgujcfqyqhfgcvjkhgctpqgxkn
9: fwfouipvhijxbmluispvhiuifwbmmfzpguiftibepxpgefbuijgfbsopfwjm
10: eventhoughiwalkthroughthevalleyoftheshadowofdeathifearnoevil
11: dudmsgntfghvzkjsgqntfgsgduzkkdxnesgdrgzcnvnecdzsghedzqmnduhk
12: ctclrfmsefguyjirfpmsefrfctyjjcwmdrfcqfybmumdbcyrfgdcyplmctgj
13: bsbkqelrdeftxihqeolrdeqebsxiibvlcqebpexaltlcabxqefcbxoklbsfi
14: arajpdkqcdeswhgpdnkqcdpdarwhhaukbpdaodwzkskbzawpdebawnjkareh
15: zqziocjpbcdrvgfocmjpbcoczqvggztjaoczncvyjrjayzvocdazvmijzqdg
16: ypyhnbioabcqufenblioabnbypuffysiznbymbuxiqizxyunbczyulhiypcf
17: xoxgmahnzabptedmakhnzamaxoteexrhymaxlatwhphywxtmabyxtkghxobe
18: wnwflzgmyzaosdclzjgmyzlzwnsddwqgxlzwkzsvgogxvwslzaxwsjfgwnad
19: vmvekyflxyznrcbkyiflxykyvmrccvpfwkyvjyrufnfwuvrkyzwvriefvmzc
20: uludjxekwxymqbajxhekwxjxulqbbuoevjxuixqtemevtuqjxyvuqhdeulyb
21: tktciwdjvwxlpaziwgdjvwiwtkpaatnduiwthwpsdldustpiwxutpgcdtkxa
22: sjsbhvciuvwkozyhvfciuvhvsjozzsmcthvsgvorckctrsohvwtsofbcsjwz
23: riragubhtuvjnyxguebhtugurinyyrlbsgurfunqbjbsqrnguvsrneabrivy
24: qhqzftagstuimxwftdagstftqhmxxqkarftqetmpaiarpqmfturqmdzaqhux
25: pgpyeszfrsthlwvesczfrsespglwwpjzqespdslozhzqoplestqplcyzpgtw
```

4. The following message was encrypted using a simple substitution cipher:

```
53ddc305))6*;4826)4d.)4d);806*;48c8p60))85;;]8*;:d*8c83
(88)5*c;46(;88*96*?;8)*d(;485);5*c2:*d(;4956*2(5*-4)8p8*
;4069285);)6c8)4dd;1(d9;48081;8:8d1;48c85;4)485c528806*81
(d9;48;(88;4(d?34;48)4d;161;:188;d?;
```

Decrypt the message. *Hint. Use frequency analysis: consider e, ee, the, ...*

For fun, I tried to solve this problem using a brute-force algorithm — turns out, even when I limit branching to a factor of 2, for a search depth of 20 letters the search tree grows to an order of $2^{20} = 1048576$. I soon resolved to frequency analysis and more informed substitutions. To be able to use the tool shared in class, I substituted the symbols in the ciphertext for English letters, starting from 'A' and matching all the symbols and letters as they are encountered:

```
ABCCDBEAFFGHIJKSGFJCLFJCFIKEGHIJKDKMGEFFKAINCKHOCPIOCHKDKBPKKF
AHDIJGPIKKHRGHQIKFHCPIJKAFIAHDSOHCPIJRAGHSPAHTJFKMKHIJEGRSKAFI
FGDKFJCCINPCRIJKEKNIKOKCNIJKDKAIJFJKADASKKEGHKNPCRIJKIPKKIJPCQ
BJIJKFJCINGNIONKKICQI
```

Looking at the frequency analysis - The most common 3-letter sequence is "IJK". In English, this is usually "the". Substitute in the three letters.

```
ABCCDBEAFFGHtheSGFhCLFhCFteEGHtheDeMGEFFeAtNCPtOCHeDeBPeeF
AHDthGPteeHRGHQteFHCPtheAFtAHDSOHCPthRAGHSPAHThFeMeHthEGRSeAFt
FGDeFhCCtNPCRtheEeNteOeCNtheDeAthFheADASeeEGHeNPCRthetPeethPCQ
BhtheFhCtNGNtONeetCQt
```

The most common letter in the ciphertext is "K". In English, this usually corresponds to the letter 'e'. This further validates the above exchange.
- The 4th, 5th, and 6th most common letters in the ciphertext are 'C', 'F', and 'H'. In English, the next 3 unused letters are 'a', 'i', and 'o'. Comparing these options: (a) 'C' appears in a pair, so it is more likely to be 'o' since English does not have many occurrences of "ii" or "aa". (b) 'F' appears before 't' on numerous occassions, so it might be either 's' or 'n'. Let's pick 's', since it is more common.

```
ABooDBEAssGHtheSGshoLshosteEGHtheDeMGEsseAtNoPtOoHeDeBPees
AHDthGPteeHRGHQtesHoPtheAstAHDSOHoPthRAGHSPAHThseMeHthEGRSeAst
sGDeshootNPoRtheEeNteOeoNtheDeAthsheADASeeEGHeNPoRthetPeethPoQ
BhtheshotNGNtONeetoQt
```

'E' completes the word "hosteE" — it is most likely 'l'. Another common two-letter pair is "GH". In English, this could be "to" or "in". However, it seems to appear frequently inside other words so it most likely is "in".

```
ABooDBlAssintheSishoLshostelintheDeMilsseAtNoPtOoneDeBPees
AnDthiPteenRinQtesnoPtheAstAnDSOnoPthRAinSPAnThseMenthliRSeAst
siDeshootNPoRtheleNteOeoNtheDeAthsheADASeelineNPoRthetPeethPoQ
BhtheshotNiNtONeetoQt
```

Next, 'D' and 'M' appear to complete the word "DeMil", which can be "devil".
Next, 'B' and 'A' complete "BlAss", and 'B' also completes 'Bood'. This combo is most likely 'g' and 'a'.

```
agoodglassintheSishoLshostelinthedevilsseatNoPtOonedegPees
andthiPteenRinQtesnoPtheastandSOnoPthRainSPanThseventhliRSeast
sideshootNPoRtheleNteOeoNthedeathsheadaSeelineNPoRthetPeethPoQ
ghtheshotNiNtONeetoQt
```

'P' completes "degPees" — it is most likely 'r'.
'S' and 'L' complete "SishoLs' — they're most-likely 'b' and 'p', respectively.
'Q' completes "throQgh" — it is most likely 'u'.
'R' completes "Rinutes" — it is most likely 'm'.
'O' and 'T' complete "bOnorthmainbranTh" — they're clearly 'y' and 'c'.
'N' completes "Nrom the" — it is most likely 'f'.

The decoded message becomes:

```
agoodglassinthebishopshostelinthedevilsseatfortyonedegrees
andthirteenminutesnortheastandbynorthmainbranchseventhlimbeast
sideshootfromthelefteyeofthedeathsheadabeelinefromthetree
throughtheshotfiftyfeetout
```

When spaced out:

```
A good glass in the bishop's hostel in the devil's seat
forty-onedegrees and thirteen minutes northeast and by
north main branchseventh limb east side, shoot from the
left eye of the death's heada bee-line from the tree
through the shot fifty feet out.
```

5. For fun, take a stab at this problem. In one of his cases, Sherlock Holmes was confronted with the following message:

534 C2 13 127 36 31 4 7 21 41
DOUGLAS 109 293 5 37 BIRLSTONE
26 BIRLSTONE 9 127 171

Although Watson was puzzled, Holmes was immediately able to deduce the type of cipher. Can you?

A possible explanation is a substitution scheme where entire words are substituted for specific numbers — for instance, "534" might be cryptic for "The queen", etc. This is perhaps less susceptible to the frequency analysis on letters, but individual words also have disparities on relative frequencies in most languages, so frequency analysis might still be able to decipher meaning behind the numbers.

It might also be referencing a specific location or book, with "DOUGLAS" and "BIRLSTONE" being some kind of identifier, such as the outhor, and the numbers telling where to look in the book.

This is a tougher system to scale — one would have to send the same book to everyone he wants to communicate with, in which case others might know about it, buy the same book, and learn how to decode the cryptic series of words and numbers.