Questions

1. Cryptanalysis of General Substitution Cipher:

This question is about cipher text only attack on Substitution cipher. You are given below a cipher text encrypted using a mono alphabetic substitution cipher.

(a) Decrypt the above cipher text by following cipher text only attack and determine the encryption and decryption keys.

```
YNHPFAZNXPOHCQYAWYRNWRJSCFOFORYAWYRPFTXMCBXFCAWJRECW
CFCAWXWJYAWQFNTYFCAWAEYAVPTFCWZQHQFRVQFOXFXJJNRQQQRY
TNCFHYAWYRNWQFORJRQCZWAEYNHPFAZNXPOCYQHQFRVQVTQFURU
XQRJAWECNVEATWJXFCAWQXPPMCRJYNHPFAZNXPOHPNRQRWFQXNCZA
NATQXWJQHQFRVXFCYFNRXFVRWFAEEATWJXFCAWXMCQQTRQJRECWC
WZYNHPFAZNXPOCYFXQKQXWJQAMICWZWRSYNHPFAZNXPOCYPNAUMRV
QTQCWZRLCQFCWZFAAMQFORRVPOXQCQCQAWFORYMXQQCECYXFCAWA
EETWJXVRWFXMYAWYRPFQXWJJRVAWQFNXFCWZFORERXQCUCMCFHAEQ
AMICWZQRIRNXMYRWFNXMYNHPFAZNXPOCYPNAUMRVQ
```

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
o	**z**	**i**		**f**	**t**		**y**	**v**	**d**	**k**	**x**	**l**	**r**	**h**	**p**	**s**	**e**	**w**	**u**	**b**	**m**	**m**	**a**	**c**	**g**

(b) Give a detailed account on how you deducted the decryption mapping. This should include the sequence of guesses for the partial keys that you may have made using statistics such as relative frequency of letters and frequency of digrams and trigrams in the cipher text. And also you might have decided partial keys base on trial and error, and also by guessing output words. You should enumerate all such steps.

Since this is a cipher text only attack, the only data that is available for the cryptanalysis is the cipher text. Despite the number of possible keys in this type of encryption that is 26!, the mono alphabetic substitution cipher can be broken by analyzing the English language characteristics. Hence, the firsts steps followed for decrypting the message were the frequency analysis and generation of digrams and trigrams. After this procedure, some words were guessed and its letters substituted in the cipher text.
Is worth to point out that the capital letters represent the cipher text while the lower case letters represent the plain text.

Here is the cryptanalysis process to decrypt the message:

1) The analysis started with the frequency analysis of the alphabet in the cipher text. The command 'grep' was used for this purpose:

```
grep -o . monoalphabetic_ciphers | sort | uniq -c
```

monoalphabetic_ciphers is a file that contains the cipher text.

In the table below, the output of the command *grep* (second row) shows the frequency of each letter in the cipher text. The third row of the table represents the result expressed in percentages.

Frequency of each letter in the cipher alphabet:

F	Q	C	W	R	A	X	N	Y	P	J	Z	O	E	H	M	V	T	U	I	S	B	K	L
42	42	38	38	37	36	32	27	26	21	16	15	14	13	13	13	12	11	5	3	2	1	1	1
9.13	9.13	8.26	8.26	8.04	7.83	6.96	5.87	5.65	4.57	3.48	3.26	3.04	2.83	2.83	2.83	2.61	2.39	1.09	0.65	0.43	0.22	0.22	0.22

Here is the frequency of each letter in the English alphabet:

e	t	o	a	n	i	r	s	h	d	l	c	f	u	m	p	w	y	g	b	v	k	x	j	q	z
13	9	8.2	7.8	7.3	6.8	6.6	6.5	5.8	4.1	3.6	2.9	2.9	2.8	2.6	2.1	1.5	1.5	1.4	1.3	1	0.4	0.3	0.2	0.1	0.1

F and *Q* are the letters with highest frequency in the cipher text and the letter with highest frequency in the English alphabet is *e* so it is not correct to replace F or Q for e.

2) According to (Stallings, 2006) a powerful way to do cryptanalysis is by the use of digrams and trigrams. Thus, by using a combination of Java code and *grep* tool in Unix, it was possible to determine the most frequent digrams and trigrams in the cipher text.

The *compare.sh* script, *digrams* file and *results_digrams* file can be found in the Appendix section of this document.

```
./compare.sh digrams > results_digrams
```

This script takes as input the *digrams* file found in the cipher text, and returns the number of occurrences of each digram in the file *results_digrams.* The file *digrams* was generated by a program written in Java.

Here are the 3 most frequent digrams:

DIGRAMS	
AW	14
FC	10
QF	10
NX	9

The most frequent digram “AW” in the cipher text could have been assumed as “th”: the most frequent digram in the English alphabet. However, before assigning it was necessary to check the letters, before and after, “AW”; in fact, some patterns were found. For instance: Y**AW**YR, Y**AW**YRPF, Y**AW**YRNW, FC**AW**X, FC**AW**ZJ. If ”AW” was “th” the next letter {Y,X or Z} must be the letter “e” in the English alphabet considering that “the” is the most common trigram and the letter “e” is the most frequent letter in the English alphabet.

Here it is the “AW” digram found in the cipher text:

```
YNHPFAZNXPOHCQYAWYRNWRJSCFOFORYAWYRPFTXMCBXFCAWJRECW
CFCAWXWJYAWQFNTYFCAWAEYAVPTFCWZQHQFRVQFOXFXJJNRQQQRY
TNCFHYAWYRNWQFORJRQCZWAEYNHPFAZNXPOCYQHQFRVQVTQFURU
XQRJAWECNVEATWJXFCAWQXPPMCRJYNHPFAZNXPOHPNRQRWFQXNCZA
NATQXWJQHQFRVXFCYFNRXFVRWFAEEATWJXFCAWXMCQQTRQJRECWC
WZYNHPFAZNXPOCYFXQKQXWJQAMICWZWRSYNHPFAZNXPOCYPNAUMRV
QTQCWZRLCQFCWZFAAMQFORRVPOXQCQCQAWFORYMXQQCECYXFCAWA
EETWJXVRWFXMYAWYRPFQXWJJRVAWQFNXFCWZFORERXQCUCMCFHAEQ
AMICWZQRIRNXMYRWFNXMYNHPFAZNXPOCYPNAUMRVQ
```

These patterns are composed of 5 to 6 letters; none of them has a pattern of 3 letters (which could correspond to “the”). Hence, this digram (AW) might not be replaced as “th”.

In order to find a trigram in the cipher text that is replaced by “the” the same code (Java and *grep* with some few changes) was used to generate the trigrams.

Here are the most frequent trigrams found in the cipher text:

TRIGRAMS	
YNH	6
NHP	6
HPF	6
PFA	6
FAZ	6
AZN	6
ZNX	6
NXP	6
XPO	6
XFC	6
FCA	6
CAW	6
CWZ	6
YAW	5
FOR	5
AWY	4

The results can be quite confusing. Some trigrams have the same frequency. Nevertheless, after analyzing them, it was possible to determine that they belong to the same pattern or word. For example YNH, NHP, HPF, PFA, FAZ and so on, can be found in the word: YNHPFAZNXPO. Thus, it means that none of those trigrams can be replaced by "the". The same analysis was done with the trigrams XFC, FCA, CAW, and CWZ. The trigrams YAW and AWY were not considered since they had the digram AW that was discarded previously. Consequently, the only possible option might be the trigram "**FOR**".

Thus, to make sure that this is the correct trigram, the letters **F**, **O** and **R** were analyzed individually.

The letter F has the highest occurrence in the cipher text (9.13%) and the letter t in the English alphabet as well (9%). The letter O in the cipher text has a relative low occurrence (3.04%) in the cipher text as well as letter h (5.8%) in the English alphabet. The letter R in the cipher text has one of the highest occurrences (8.04%) while the letter e in the English alphabet has the highest (13%). Consequently, the "FOR" trigram in the cipher text is a good guessing for the "the" trigram in the English alphabet.

After this analysis, the "FOR" trigram was replaced in the cipher text by the trigram "the"; each letter "F","O", and "R" by its correspondent "t", "h" and "e".

The three letters **QQQ** in the cipher text cannot belong to the same word (according to the English language word characteristics), so they were split and considered as two ending letters in one word and one beginning letter on the next word. According to the table below, the most common ending and double letters in the alphabet are: e, o, t, and s.

	a	e	i	o	t	r	n	h	s
common	x	**X**	x	x	x				
start	x		x	x	**X**				x
end		**X**		x	x	x	x		x
doubles		x		x	x				x

Source: (Sauerberg)

The letters "t" and "e" were already assigned to "F" and "R" in the cipher text respectively, so the letter "Q" in the cipher text might correspond either to "o" or "s" in the English alphabet. However, according to the table below, the letter "s" is more likely to be found at the end of the words in the English alphabet than the letter "o". Then, "Q" in the cipher text was replaced by "s" in the English alphabet.

2.9	0.2	0.6	10.0	20.3	4.5	2.8	2.5	0.4	0	0.1	3.7	1.3
a	b	c	d	e	f	g	h	i	j	k	l	m
9.7	4.5	0.5	0	5.5	12.7	9.7	0.2	0.1	1.0	0.2	5.5	0
n	o	p	q	r	s	t	u	v	w	x	y	z

Letter Frequencies – Final Letters.

Source: (Sauerberg)

The table below shows the process to decipher the text based on the considerations already explained. The second column shows further considerations for decipher other letters based on assumptions and guessing of words.
Capital letters represent the cipher text and lower case letters represent plain text.

YNHPFAZNXPOHCQYAWYRNWRJSCFOFORYAWYRPFTXMCBXFCAWJRECW CFCAWXWJYAWQFNTYFCAWAEYAVPTFCWZQHQFRVQFOXFXJJNR**QQQ**RY TNCFHYAWYRNWQFORJRQCZWAEYNHPFAZNXPOCYQHQFRVQVTQFURU XQRJAWECNVEATWJXFCAWQXPPMCRJYNHPFAZNXPOHPNRQRWFQXNCZA NATQXWJQHQFRVXFCYFNRXFVRWFAEEATWJXFCAWXMCQQTRQJRECWC WZYNHPFAZNXPOCYFXQKQXWJQAMICWZWRSYNHPFAZNXPOCYPNAUMRV QTQCWZRLCQFCWZFAAMQFORRVPOXQCQCQAWFORYMXQQCECYXFCAWA EETWJXVRWFXMYAWYRPFQXWJJRVAWQFNXFCWZFORERXQCUCMCFHAEQ AMICWZQRIRNXMYRWFNXMYNHPFAZNXPOCYPNAUMRVQ	CIPHER TEXT The letters QQQ represent two different words since no word in English can have three same letters.
YNHPtAZNXPh HCsYAWYeNWeJSCth theYAWYePtTXMCBXtCAWJeECW CtCAWXWJYAWstNTYtCAWAEYAVPTtCWZ sHsteV sthXtXJJNess seY TNCtHYAWYeNWstheJesCZWAEYNHPtAZNXPhCY**sHsteV**sVTstUeU XseJAWECNVEATWJXtCAWsXPPMCeJYNHPtAZNXPhHPNeseWtsXNCZA NATsXWJsHsteVXtCYtNeXtVeWtAEEATWJXtCAWXMCssTesJeECWC WZYNHPtAZNXPhCYtXsKsXWJsAMICWZWeSYNHPtAZNXPhCYPNAUMeV sTsCWZeLCstCWZtAAMstheeVPhXsCsCsAWtheYMXssCECYXtCAWA EETWJXVeWtXMYAWYePtsXWJJeVAWstNXtCWZtheEeXsCUCMCtHAEs AMICWZseIeNXMYeWtNXM YNHPtAZNXPh CYPNAUMeVs	Replacement of “FOR” by “the” and “Q” by “s” Some words were split based on the patterns previously found: YNHPtAZNXPh. The word sHsteV was assumed as “system”
YNyPtAZNXPh yCsYAWYeNWeJSCth theYAWYePtTXMCBXtCAWJeECW CtCAWXWJYAWstNTYtCAWAEYAmPTtCWZ system sthXtXJJNess seY TNCtyYAWYeNWstheJesCZWAEYNyPtAZNXPhCY systems **mTst** UeU XseJAWECNmEATWJXtCAWsXPPMCeJYNyPtAZNXPhyPNeseWtsXNCZA NATsXWJ system XtCYtNeXtmeWtAEEATWJXtCAWXMCssTesJeECWC WZYNyPtAZNXPhCYtXsKsXWJsAMICWZWeSYNyPtAZNXPhCYPNAUMem sTsCWZeLCstCWZtAAMstheemPhXsCsCsAWtheYMXssCECYXtCAWA EETWJXmeWtXMYAWYePtsXWJJemAWstNXtCWZtheEeXsCUCMCtyAEs AMICWZseIeNXMYeWtNXM YNyPtAZNXPh CYPNAUMems	Replacement of “H” by “y” and “V” by “m”. The word mTSt was assumed as “must”.
YNyPtAZNXPh yCsYAWYeNWeJSCth theYAWYePtuXMCBXtCAWJeECW CtCAWXWJYAWstNuYtCAWAEYAmPutCWZ system sthXtXJJNess seY uNCtyYAWYeNWstheJesCZWAEYNyPtAZNXPhCY systems **must** UeU XseJAWECNmEAuWJXtCAWsXPPMCeJYNyPtAZNXPhyPNeseWtsXNCZA NAusXWJ system XtCYtNeXtmeWtAEEAuWJXtCAWXMCssuesJeECWC WZYNyPtAZNXPhCYtXsKsXWJsAMICWZWeSYNyPtAZNXPhCYPNAUMem susCWZeLCstCWZtAAMstheemPhXsCsCsAWtheYMXssCECYXtCAWA EEuWJXmeWtXMYAWYePtsXWJJemAWstNXtCWZtheEeXsCUCMCtyAEs AMICWZseIeNXMYeWtNXM YNyPtAZNXPh CYPNAUMems	Replacement of “T” as “u”
YNyPtAZNXPh yCsYAWYeNWeJSCth theYAWYePtuXMCBXtCAWJeECW CtCAWXWJYAWstNuYtCAWAEYAmPutCWZ system sthXtXJJNess seY uNCtyYAWYeNWstheJesCZWAEYNyPtAZNXPhCY systems must UeU XseJAWECNmEAuWJXtCAWsXPPMCeJYNyPtAZNXPhyPNeseWtsXNCZA NAusXWJ system XtCYtNeXtmeWtAEEAuWJXtCAWXM **Cssues** JeECWC WZYNyPtAZNXPhCYtXsKsXWJsAMICWZWeSYNyPtAZNXPhCYPNAUMem susCWZeLCstCWZtAAMstheemPhXsCsCsAWtheYMXssCECYXtCAWA EEuWJXmeWtXMYAWYePtsXWJJemAWstNXtCWZtheEeXsCUCMCtyAEs	The word Cssues was guess as “issues”.

AMICWZseIeNXMYeWtNXM YNyPtAZNXPh CYPNAUMems	
YNyPtAZNXPh yisYAWYeNWeJSith theYAWYePtuXMiBXtiAWJeEiW itiAWXWJYAWstNuYtiAWAEYAmPutiWZ system sthXtXJJNess seY uNityYAWYeNWstheJesiZWAEYNyPtAZNXPhiY systems must UeU XseJAWEiNmEAuWJXtiAWsXPPMieJYNyPtAZNXPhyPNeseWtsXNiZA NAusXWJ system XtiYtNeXtmeWtAEEAuWJXtiAWXM **issues** JeEiWi WZYNyPtAZNXPhiYtXsKsXWJsAMIiWZWeSYNyPtAZNXPhiYPNAUMem susiWZeListiWZtAAMstheemPhXsisisAWtheYMXssiEiYXtiAWA EEuWJXmeWtXMYAWYePtsXWJJemAWstNXtiWZtheEeXsiUiMityAEs AMIiWZseIeNXMYeWtNXM YNyPtAZNXPh iYPNAUMems	Replacement of “C” by “i”
YNyPtAZNXPh yisYAWYeNWeJSith theYAWYePtuXMiBXtiAWJeEiW itiAWXWJYAWstNuYtiAWAEYAmPutiWZ system sthXtXJJNess **seY uNity** YAWYeNWstheJesiZWAEYNyPtAZNXPhiY systems must UeU XseJAWEiNmEAuWJXtiAWsXPPMieJYNyPtAZNXPhyPNeseWtsXNiZA NAusXWJ system XtiYtNeXtmeWtAEEAuWJXtiAWXM issues JeEiWi WZYNyPtAZNXPhiYtXsKsXWJsAMIiWZWeSYNyPtAZNXPhiYPNAUMem susiWZeListiWZtAAMstheemPhXsisisAWtheYMXssiEiYXtiAWA EEuWJXmeWtXMYAWYePtsXWJJemAWstNXtiWZtheEeXsiUiMityAEs AMIiWZseIeNXMYeWtNXM YNyPtAZNXPh iYPNAUMems	The word seYuNity was guessed as “security”
cryPtAZrXPh yiscAWcerWeJSith thecAWcePtuXMiBXtiAWJeEiW itiAWXWJcAWstructiAWAEcAmPutiWZ system sthXtXJJress **sec urity** cAWcerWstheJesiZWAEcryPtAZrXPhic systems must UeU XseJAWEirmEAuWJXtiAWsXPPMieJcryPtAZrXPhyPreseWtsXriZA rAusXWJ system XtictreXtmeWtAEEAuWJXtiAWXM issues JeEiWi WZcryPtAZrXPhictXsKsXWJsAMIiWZWeScryPtAZrXPhicPrAUMem susiWZeListiWZtAAMstheemPhXsisisAWthecMXssiEicXtiAWA EEuWJXmeWtXMcAWcePtsXWJJemAWstrXtiWZtheEeXsiUiMityAEs AMIiWZseIerXMceWtrXM cryPtAZrXPh icPrAUMems	Replacement of “Y” by “c” and “N” by “r”. The word cryPtAZrXPh was assumed as “cryptography”.
cryptograph yiscoWcerWeJSith thecoWceptuaMiBatioWJeEiW itioWaWJcoWstructioWoEcomputiWg system sthataJJress **sec urity** coWcerWstheJesigWoEcryptographic systems must UeU aseJoWEirmEouWJatioWsappMieJcryptographypreseWtsarigo rousaWJ system atictreatmeWtoEEouWJatioWaM issues JeEiWi WgcryptographictasKsaWJsoMIiWgWeScryptographicproUMem susiWgeListiWgtooMstheemphasisisoWthecMassiEicatioWo EEuWJameWtaMcoWceptsaWJJemoWstratiWgtheEeasiUiMityoEs oMIiWgseIeraMceWtraM cryptograph icproUMems	Replacement of “P” by “p” and “A” by “o”.
cryptography is coWcerWeJ Sith the coWceptuaMiBatioW JeEiWitioWaWJ coWstructioW oE computiWg systems that aJJress **security** coWcerWs the JesigWoE cryptographic systems must UeUaseJoWEirmEouWJatioWsappMieJ cryptography preseWtsarigorousaWJ systematic treatmeWt oEEouWJatioWaM issues JeEiWiWg cryptographic tasKsaWJsoMIiWgWeS cryptographic proUMem susiWgeListiWgtooMstheemphasisisoWthecMassiEicatioWo EEuWJameWtaMcoWceptsaWJJemoWstratiWgtheEeasiUiMityoEs oMIiWgseIeraMceWtraM cryptograph icproUMems	Split words that look familiar. i.e. coWceptuaMiBatioW, coWcerWeJ

cryptography is **coWcerWeJ Sith** the **coWceptuaMiBatioW** JeEiWitioWaWJ coWstructioW oE computiWg systems that aJJress **security** coWcerWs the JesigWoE cryptographic systems must UeUaseJoWEirmEouWJatioWsappMieJ cryptography preseWtsarigorousaWJ systematic treatmeWt oEEouWJatioWaM issues JeEiWiWg cryptographic tasKsaWJsoMIiWgWeS cryptographic proUMem susiWgeListiWgtooMstheemphasisisoWthecMassiEicatioWo EEuWJameWtaMcoWceptsaWJJemoWstratiWgtheEeasiUiMityoEs oMIiWgseIeraMceWtraM cryptograph icproUMems	**coWcerWeJ** assumed as “concerned” **Sith** assumed as “with”
cryptography is concerned with the conceptualization **deEinitionand** construction oE computing systems that address **security** concerns the designoE cryptographic systems must UeU**asedonEirmEoundationsapplied** cryptography **presentsarigo**rousand systematic treatment oEEoundational issues deEining cryptographic tasKsandsolIingnew cryptographic proUlem susingeListingtoolstheemphasisisontheclassiEicationo EEundamentalconceptsanddemonstratingtheEeasiUilityoEs olIingseIeralcentral cryptograph icproUlems	Replacement of “W” by “n”, “T” by “d” and “S” by “w”.
cryptography is concerned with the conceptualization **definition** and construction of computing systems that address **security** concerns the design of cryptographic systems must UeUasedonfirm foundations applied cryptography presents a rigorous and systematic treatment of foundational issues defining cryptographic tasKsandsolIingnew cryptographic **proUlem**susingeListing tools the emphasis is on the classification of fundamental concepts and demonstrating the feasiUility of solIingseIeral central cryptographic proUlems	**proUlem** guessed as “problem”
cryptography is concerned with the conceptualization **definition** and construction of computing systems that address **security** concerns the design of cryptographic systems must be based on firm foundations applied cryptography presents a rigorous and systematic treatment of foundational issues defining cryptographic tasKsandsolIingnew cryptographic **problem** susingeListing tools the emphasis is on the classification of fundamental concepts and demonstrating the feasibility of **solIing** seIeral central cryptographic problems	Replacement of “U” by “b”
cryptography is concerned with the conceptualization **definition** and construction of computing systems that address **security** concerns the design of cryptographic systems must be based on firm foundations applied cryptography presents a rigorous and systematic treatment of foundational issues defining cryptographic **tasKs** and solving new cryptographic **problem**s using **eListing** tools the emphasis is on the classification of fundamental concepts and demonstrating the feasibility of **solving** several central cryptographic problems	assume **SolIng as** “solving”**, tasKs** as “tasks” and **eListing** as “existing”

cryptography is concerned with the conceptualization **definition** and construction of computing systems that address **security** concerns the design of cryptographic systems must be based on firm foundations applied cryptography presents a rigorous and systematic treatment of foundational issues defining cryptographic **tasks** and solving new cryptographic **problem**s using **existing** tools the emphasis is on the classification of fundamental concepts and demonstrating the feasibility of **solving** several central cryptographic problems	Replacement of “l” by “v”, “K” by “k” and “L” by “x” **PLAIN TEXT**