PMATH 329   Introduction to Cryptography

SOLUTIONS TO PROBLEMS 1-4 OF ASSIGNMENT 1

==

1. Cryptanalyze the following ciphertext. Show your thinking.

```
TWOZW VHFWK LXOXG MAWBO AHLMB EXYHK VXXLM BFTMX
WTLHG XKXZB FXGMB GYTGM KRTGW MPHIE TMHHG LVTOT
EKRFH OBGZL HNMAH GZXMM RLUNK ZKHTW AXTWH YVHEN
FGGXT KBGZK HTWCN GVMBH GYBOX XBZAM SXKHX TLMHY
IBMSX KLVAH HEYBK XWNIH GURHN KITMK HELAT OXWXL
MKHRX WUKBW ZXLHO XKFTK LAVKX XDYKH FZKXX GFHNG
MMHTI HBGMT LYTKG HKMAT LMAXU KBWZX WNXPX LMHYI
BMSXK LVAHH EPBEE WXYXG WABEE YBOXX BZAML BQHGX
FBEXG HKMAH YZKXX GFHNG MBYYH KVXWM HKXMB KXPBE
EWXLM KHRUK BWZXL HNMAH YZKXX GFHNG MTGWW XETRK
XWLTM FTKLA VKXXD EHHFB LVTIM
```

SOLUTION:

First, using the phi-statistic we determine the number of cipher alphabets. Here's the [sorted] frequencies for the letters in the ciphertext:

Symbol	Count
X	51
H	44
K	36
M	34
G	27
B	27
T	25
L	23
W	23
E	16
A	16
Y	15
Z	14
F	13
V	11
N	11
O	9
R	7
I	7
U	5
P	4
S	3
D	2
C	1
Q	1

J 0

Total number of symbols = 425
Phi statistic for this text = 11284
Expected phi for random text of the same length = 6937.7
Expected phi for English text of the same length = 11911.2

Hence, this is almost certainly a monoalphabetic substitution cipher.

First, we check whether it is a Caesar cipher using Kerckhoffs' shortcut to attempt to find the key K:

The most common plaintext letters are likely "e" (4), "t" (19), "a" (0), and "o" (14). The most common letters in the ciphertext are { X, H, K, M, G, B, T, L } = { 23, 7, 10, 12, 6, 1, 19, 11 }.

If an enciphered "e" is in the ciphertext list, then
4 + K (mod 26) is in { 23, 7, 10, 12, 6, 1, 19, 11 }, or
K is in { 19, 3, 6, 8, 2, 23, 15, 7 }.

If an enciphered "t" is in the ciphertext list, then
19 + K (mod 26) is in { 23, 7, 10, 12, 6, 1, 19, 11 }, or
K is in { 4, 14, 17, 19, 13, 8, 0, 18 }.

The intersection of the two sets for K is { 19, 8 }.

If an enciphered "a" is in the ciphertext list, then
0 + K (mod 26) is in { 23, 7, 10, 12, 6, 1, 19, 11 }, or
K is in { 23, 7, 10, 12, 6, 1, 19, 11 }.

Thus, K is likely 19 (T), and using this to decrypt assuming a Caesar cipher yields the plaintext.

Plaintext:

Adv gd comdr: Seventh div hostile force estimated as one regiment infantry and two platoons cavalry. Moving south on Gettysburg road, head of column nearing road junction five eight zero east of Pitzer school. Fired upon by our patrols, have destroyed bridges over marsh creek from Greenmount to a point as far north as the bridge due west of Pitzer school. Will defend hill five eight six one mile north of Greenmount. If forced to retire, will destroy bridge south of Greenmount and delay reds at Marsh Creek.

Loomis, capt.

2. Which of the following represent selections of English text enciphered monoalphabetically? Why?

(a) BQCKG WTNMC RZXUW EKACD SWAPO HLAIU

(b) KKPNV HHTZH TWEUH EYVAB YWKTQ MDALM

(c) QUXKG OYZNK RGTJU LZNKS UXTOT MIGRS

(d) AOZHO ZWSGW BHVSA SRWHS FFOBS OBGSO

SOLUTION:

Compute the phi-statistic for each ciphertext, and compare with the expected values for English text and random text.

(a) Cipher letter frequencies:

A 3
C 3
W 3
K 2
U 2
G 1
H 1
I 1
D 1
L 1
M 1
N 1
O 1
P 1
Q 1
R 1
S 1
T 1
E 1
B 1
X 1
Z 1

phi = 3(3)(2) + 2(2)(1) + 17(1)(0)
= 22

text length is 30

Expected phi for English text of length 30 = 0.0661 (30)(29)
= 57.5

Expected phi for random text of length 30 = 0.0385 (30)(29)
= 33.5

Since the phi-statistic is closer to the expected value for random text than for English text, this text is likely random.

(b) Cipher letter frequencies:

H 4
K 3
T 3
E 2
A 2
M 2

V 2
W 2
Y 2
B 1
L 1
N 1
P 1
Q 1
U 1
D 1
Z 1

phi = 1(4)(3) + 2(3)(2) + 6(2)(1) + 8(1)(0)
= 36

Since the phi-statistic is closer to the expected value for random text (33.5) than for English text (57.5), this text is likely random.

(c) Cipher letter frequencies:
G 3
K 3
T 3
U 3
N 2
O 2
R 2
S 2
X 2
Z 2
L 1
M 1
Q 1
I 1
Y 1
J 1

phi = 4(3)(2) + 6(2)(1) + 6(1)(0)
= 36

Since the phi-statistic is closer to the expected value for random text (33.5) than for English text (57.5), one would draw the conclusion that this text is likely random - especially given that phi is the same as in part (b). However, it turns out that this is the English text "Korea is the land of the morning calm", encrypted with a Caesar cipher with key K=6. This shows that performing the phi statistic on too small samples of ciphertext may lead to erroneous conclusions.

(d) Cipher letter frequencies:
S 6
O 5
H 3
B 3
W 3
F 2

G	2
A	2
Z	2
R	1
V	1

phi = 1(6)(5) + 1(5)(4) + 3(3)(2) + 4(2)(1) + 2(1)(0)
= 76

Since the phi-statistic is closer to the expected value for English text (57.5) than for random text (33.5), this text is likely a monoalphabetic substitution cipher. And this is indeed the English plaintext "Malta lies in the Mediterranean sea", encrypted with a Caesar cipher and key K=14.

3. Cryptanalyze the following ciphertext. Show your reasoning.
Hint: The title of this text is "The CADBURY Caramilk Cryptogram."

GHPCS PCEPMPEGAL YRREH ,Y TO,T,LAWR ACE TAXAIRL YX,I HIALL
O,DDRXH AF,MR ERD,HPW,XH HOW YL,K PH EPMPERE PXY, AH
IAC^ HRTWP,CH AH WOR CGIFRX ,Y XRNGPXRE ERD,HPWH AW WOR
R,WW,I ,Y WOR ERD,HPW,X PH WOR TRCWXR FL,TB KPWO DLAWRH
RATO DX,SXRHHPMRL^ HDLPWWPCS WOR YL,K ,Y RPWORX
PCSXREPRCW PCW, WK, HWXRAIH GCWPL WOR XRNGPXRE CGIFRX
PH XRATORE WOR PCEPKPEGAL ERD,HPWH AXR HOABRC
W,SRWORX W Y,XI WOR T,IDLRWR YPCPHORE FAX

SOLUTION:

The "," and "^" occuring in the ciphertext are nothing more than regular cipher characters.

Using the phi-statistic we determine the number of cipher alphabets. Here's the [sorted] frequencies for the symbols in the ciphertext:

Symbol	Count
R	47
W	33
P	31
,	28
H	27
X	23
A	19
E	18
O	18
C	16
L	14
Y	11
I	10
D	10
T	9

G	8
S	5
F	5
K	5
M	4
B	2
N	2
^	2

Total number of symbols = 347
Phi statistic for this text = 7984
Expected phi for random text of the same length = 4622.39
Expected phi for English text of the same length = 7936.1

Hence, this is almost certainly a monoalphabetic substitution cipher.

Given that symbols other than upper-case letters occur ("," and "^") in the ciphertext, we can rule out a simple Caesar cipher. Thus, we resort to cryptanalysis based on frequencies. The 20 most common digraphs and trigraphs, and the most common double letters are the following:

Digram	Count	Trigram	Count	Double	Count
------	-----	-------	-----	------	-----
OR	12	WOR	10	WW	3
WO	11	ERD	4	HH	2
XR	10	RD,	4	DD	1
RE	9	D,H	4	EE	1
PC	7	,HP	4	LL	1
PW	7	HPW	4	RR	1
HP	6	PCS	3		
RX	6	XRE	3		
EP	5	ORE	3		
ER	5	RCW	3		
,H	5	ORX	3		
W,	5	PCE	2		
CW	5	CEP	2		
,Y	4	EPM	2		
RA	4	PMP	2		
HA	4	MPE	2		
RD	4	PEG	2		
D,	4	EGA	2		
PH	4	GAL	2		
WP	4	LAW	2		

A good starting point is assuming that WOR = the, which results in the following:

```
           ee      h   te       e
GHPCS PCEPMPEGAL YRREH ,Y TO,T,LAWR ACE TAXAIRL YX,I HIALL

h  e     e e   t   ht          e
O,DDRXH AF,MR ERD,HPW,XH HOW YL,K PH EPMPERE PXY, AH

    e t    the    e    e   e  e   t t the
IAC^ HRTWP,CH AH WOR CGIFRX ,Y XRNGPXRE ERD,HPWH AW WOR
```

```
e tt     the  e    t       the  e t e          th     te
R,WW,I ,Y WOR ERD,HPW,X PH WOR TRCWXR FL,TB KPWO DLAWRH

e  h       e     e        tt     the           e the
RATO DX,SXRHHPMRL^ HDLPWWPCS WOR YL,K ,Y RPWORX

   e  e t   t  t    t e       t   the  e     e       e
PCSXREPRCW PCW, WK, HWXRAIH GCWPL WOR XRNGPXRE CGIFRX

   e  he  the               e    t     e  h   e
PH XRATORE WOR PCEPKPEGAL ERD,HPWH AXR HOABRC

t  ethe  t       the       ete       he
W,SRWORX W Y,XI WOR T,IDLRWR YPCPHORE FAX
```

Next, consider the ciphertext

```
      e the
   RPWORX
```

This probably corresponds to the plaintext "either." Under this assumption,

```
   t  ethe
   W,SRWORX
```

likely corresponds to "together." Filling in the new letters yields

```
 i g i  i i      ee    o    ho o  te        r   e    ro
GHPCS PCEPMPEGAL YRREH ,Y TO,T,LAWR ACE TAXAIRL YX,I HIALL

ho  er    o e  e o itor   ht    o  i    i i e  ir o
O,DDRXH AF,MR ERD,HPW,XH HOW YL,K PH EPMPERE PXY, AH

     e tio       the      er o  re  ire    e o it    t the
IAC^ HRTWP,CH AH WOR CGIFRX ,Y XRNGPXRE ERD,HPWH AW WOR

eotto  o  the  e o itor i  the  e tre   o     ith     te
R,WW,I ,Y WOR ERD,HPW,X PH WOR TRCWXR FL,TB KPWO DLAWRH

e  h  rogre  i e      itti g the   o  o  either
RATO DX,SXRHHPMRL^ HDLPWWPCS WOR YL,K ,Y RPWORX

i gre ie t i to t o  tre      ti  the re  ire       er
PCSXREPRCW PCW, WK, HWXRAIH GCWPL WOR XRNGPXRE CGIFRX

i  re  he  the i  i i       e o it    re  h  e
PH XRATORE WOR PCEPKPEGAL ERD,HPWH AXR HOABRC

together t  or  the  o    ete  i i he     r
W,SRWORX W Y,XI WOR T,IDLRWR YPCPHORE FAX
```

Now condider

i gre ie t
PCSXREPRCW

likely "ingredient", and assuming C=n,

o
,Y

is probably "of"

ing indi id feed of ho o te nd r e fro
GHPCS PCEPMPEGAL YRREH ,Y TO,T,LAWR ACE TAXAIRL YX,I HIALL

ho er o e de o itor ht f o i di ided irfo
O,DDRXH AF,MR ERD,HPW,XH HOW YL,K PH EPMPERE PXY, AH

n e tion the n er of re ired de o it t the
IAC^ HRTWP,CH AH WOR CGIFRX ,Y XRNGPXRE ERD,HPWH AW WOR

eotto of the de o itor i the entre o ith te
R,WW,I ,Y WOR ERD,HPW,X PH WOR TRCWXR FL,TB KPWO DLAWRH

e h rogre i e itting the f o of either
RATO DX,SXRHHPMRL^ HDLPWWPCS WOR YL,K ,Y RPWORX

ingredient into t o tre nti the re ired n er
PCSXREPRCW PCW, WK, HWXRAIH GCWPL WOR XRNGPXRE CGIFRX

i re hed the indi id de o it re h en
PH XRATORE WOR PCEPKPEGAL ERD,HPWH AXR HOABRC

together t for the o ete fini hed r
W,SRWORX W Y,XI WOR T,IDLRWR YPCPHORE FAX

Obtaining the rest of the plaintext from this point is relatively straightforward. The complete plaintext is given below, with the errors corrected.

Plaintext:

Using individual feeds of chocolate and caramel from small
hoppers above depositors the flow is divided into as
many sections as the number of required deposits at the
bottom of the depositor is the centre block with plates
each progressively splitting the flow of either
ingredient into two streams until the required number
is reached the individual deposits are shaken
together to form the complete finished bar.

(with punctuation)

Using individual feeds of chocolate and caramel from small hoppers above depositors, the flow is divided into as many sections as the number of required deposits. At the bottom of the depositor is the centre block with plates, each progressively splitting the flow of either ingredient into two streams until the required number is reached. The individual deposits are shaken together to form the complete finished bar.

4. Cryptanalyze the following Vignere cipher. Show your reasoning.
   Hint: 7 alphabets were used to encipher the plaintext.

BIPIZ VYPVK JLXAD VUBPP QDVEY XTMIM TCLRV SIVKN SEBCP SNELX
WPCES CTMRD SIAGW ROCEL XWPCC YFCQP QDQGD BJVJF QBIMS YKACN
QBXEL XWPWZ VWPKE ODWAQ TGDZQ GGMRO ZTDTE KDSMF IPGYX KSCQB
KEMGC JWDLW AIJGM ZQWHU QBPEU RMUCT FDTQP PZVEP BKYYV WFCKB
PWEDZ GZCSL TKVSZ RLWIP IZYEP GPYHL SKMAY FGSCV QGAVG IMEDT
RXDZO KEMGC AVYCI VDVAY FVHGM OSDIK QGRRD CARIN WPEKJ ZGCEL
SITKW TXSRK GCDXG PCVRZ VAOMF ZPSHA MUSXM DPZNI TFEWI TNHEJ
TIPND SXIEC PBTJD LWMEW ZPDGP PELJZ GCELS CKCXM IMHMF DZMVT
VVSQC SCLER PGCIP GKFXZ DZKJL XADVQ PAIGE TGDCC ACOVY REACI
EMPWK IWCCJ WLTUC XOMLH QPPZV EPBKY YRGLB JOCIA HIYKJ XGEWT
DPGLX VHYCQ SIQQX PZWCN WBELW GBJOT FERZA ZESYG IRRTG KJJUI
DXWBK CXPBL TVFNL XSRWP DCSDP VFCZS LTKVS ZRLDB JOOEL PKQWX
YFXKC DTSFH BGBXM FPTUK YHDXV MCELS IAROP HACNQ BXELX WPPCS
EDVGV ZGSIQ QXESS CWVRP VAICU ODEKD XJSDX ARIVO OEDVW TSELE
PAVBT GLHMV YQVMA MUDZI FRZAZ ESJHK TKXFD TLCDL FWUWT OTXAH
AVYCI VDZVB LRKBQ VDPHL DIPYE LWGTQ MLXAD VCXOH WRZAZ EMLLP
GXYIW SMFPZ VHGWE ODWAC OKDPQ HAWAC PRUGG RDTSF IMERY MIJMU
DSEFR IPBPH MRMKX QSJBI VSZRW MXQCF VWHEK DSMFN WWBNS EBCPS
NELXW PCYIL LWTUL WOTTN KDTJD DKNPE KNAVO XFSHM HYCXZ TLGFP
PGEUG XESXT VEBJT LXWPZ CSYGI OCELW XJOMC CHIWI BLTZX KUEMW
QHBGW TWSKM TCLXA AMVYZ PXDZE YYXJD TNSYK SCLRB ZXWRB KXRMF
UWTWL XADVV RCSMV PGXNV QEBKY YFQPK QWMMF PBKYY SXEZQ QCEEB
QPQLR VHVCD PVEXV CVSEJ SECBP JWPBW BPWAI KCXPR UGGRD LRVSM
EBJTL XVHYC QSIQQ WLYLD UCDTG SATAK YHOXB JYFXA CBGBG IFIQQ
XMCLW MVOCQ ACINE DIJDZ CZAPA RIVSZ RMHQP QLRSA OQBTX ZBIPN
LOWNE JSNLA CLKFT HMPTK JPWLW MCVRS JXBJW ELWHC DCJWL TUGXN
VQEBU KATDX KCDTS FXVHY CQSIQ QXMIX DZGSE MKHMP DQVGB IVOCQ
ACINY CGGBX WDPVD DKCDT SFPVF OYXWG AAYFV VPBCM ZQEJV KMLXA
DVUXP XODZM KEXZT ZGMPM NXVID PVEXV CVZVU DUREE IJAWE KEMGC
BJODE ETSGI TWMHM FDZHW RZAZE XZTQP PZVEP BKYYE XIMTS EPWPD
GCELW CMVGZ VCXVC NOMLX WPDZX ZTINQ ZVAIP ODSIA QUUEM WQHBG
WAVGK QFODO WNOGX PVSIQ QXVIQ BIPKR IETVV FPVAU QEKEM GCIPN
ZTWGI VSZRS ANGKE YJTAV RLXWC PCXNI LWMDK DMURZ AZESY GIRRT
GKTKW BTXQD NVRPW MQAAC EIE

SOLUTION:

By the analysis in class, we know that the keyword likely has length 7.
We need the frequency counts for each of the seven subtexts (taken from the notes).

T0	T1	T2	T3	T4	T5	T6
D 27	M 27	G 26	D 28	P 31	V 26	A 29
C 23	V 26	P 26	Y 24	Z 29	X 26	L 27
I 21	I 25	V 26	B 20	E 28	S 23	W 24
T 21	W 21	C 21	X 19	C 20	E 19	S 21
X 21	B 20	K 18	C 18	L 19	I 19	E 18
P 20	Z 18	Q 17	O 17	T 19	M 18	G 15
H 17	Q 16	E 12	S 17	D 15	W 16	F 14
R 14	K 15	U 12	K 16	Y 14	R 14	J 11
G 13	T 9	A 11	R 10	X 9	L 11	K 9
B 12	A 8	R 11	W 10	N 8	Q 10	M 9
A 8	U 7	T 10	Z 9	M 6	T 9	Q 9
E 6	L 6	J 9	M 8	J 6	G 9	D 9
J 5	P 6	F 8	P 8	S 6	F 8	Z 8
W 5	C 6	N 7	N 7	F 6	H 8	V 7
Q 4	N 5	W 7	Q 5	O 5	P 5	X 7
S 4	E 5	O 5	I 5	A 5	C 5	U 5
N 4	G 4	H 4	V 5	Q 4	Y 5	H 5
K 3	O 4	M 3	E 4	R 4	K 3	Y 4
U 3	J 4	I 2	F 3	H 2	O 2	C 3
L 3	D 3	D 2	J 3	G 1	B 1	O 3
V 2	X 3	B 1	G 2	V 1	J 1	N 1
F 1	S 1	Y 1	U 1	B 1	Z 1	I 1
O 1	R 0	L 0	T 0	W 0	D 0	B 0
M 1	H 0	X 0	A 0	I 0	N 0	T 0
Y 0	Y 0	S 0	L 0	K 0	U 0	R 0
Z 0	F 0	Z 0	H 0	U 0	A 0	P 0

Example: Finding the key K5 for the Caesar cipher used to encode subtext T5:

The most common plaintext letters are likely "e" (4), "t" (19), "a" (0), and "o" (14). The most common letters in the ciphertext are { V, X, S, E, I, M, W, R } = { 21, 23, 18, 4, 8, 12, 22, 17 }.

If an enciphered "e" is in the ciphertext list, then
4 + K5 (mod 26) is in { 21, 23, 18, 4, 8, 12, 22, 17 }, or
K5 is in { 17, 19, 14, 0, 4, 8, 18, 13 }.

If an enciphered "t" is in the ciphertext list, then
19 + K5 (mod 26) is in { 21, 23, 18, 4, 8, 12, 22, 17 }, or
K5 is in { 2, 4, 24, 11, 15, 19, 3, 24 }.

The intersection of the two sets for K5 is { 19, 4 }.

If an enciphered "a" is in the ciphertext list, then
0 + K5 (mod 26) is in { 21, 23, 18, 4, 8, 12, 22, 17 }, or
K5 is in { 21, 23, 18, 4, 8, 12, 22, 17 }.

Thus, K5 is likely 4 (E).

Repeating this for each of K0, K1, ..., K6 yields the cipher key PICKLES.
The plaintext (with the errors corrected) is given below.

Cipher key: PICKLES

Plaintext:

Many organizations rely on computers and data communications to keep their operations running smoothly by making information more accessible to more people within the organization. But as it becomes more accessible, information requires more protection than you may now have. You can now protect information stored on your premises by physical measures that limit access to authorized people. Similarly, IBM hardware and software products have features that can be used to identify and check the authorization of people trying to gain access to a system and its info mation. Now there is a way to protect information even further. The IBM cryptographic subsystem can extend data control and protection to the data communications terminals and links that speed information from one location to another. It uses a sophisticated algorithm, a strict set of rules, to encrypt or scramble data before it is stored or transmitted to another location and decrypt it when needed for processing. It employs encryption techniques that can reduce information exposures within your communications network as well as provide a system base for the development of encryption programs. The IBM cryptographic subsystem is a versatile tool for controlling and protecting information through encryption by a combination of programming and SNA terminal hardware features. It can encrypt and decrypt information automatically and without intervention by the terminal, user, or application using an algorithm and a key which individualizes the algorithm. The subsystem encrypts application information before it is sent from a terminal or computer location and enters your data communications network. At the receiving terminal or computer location, the same key is used to decrypt the information after it leaves the network. In addition to the algorithm, the IBM subsystem provides key generation, key management, verification, and operational features that enhance the basic cryptographic security of the subsystem.