CSCI/CSIS 631

Fall 2019

Homework 3

Solution

First, we test to be sure this is a Vigenère cipher. We compute the frequency counts of each letter, and from that get the index of coincidence (IC), which is 0.040. Using Figure 9–4, this indicates a polyalphabetic cipher, so we are justified in assuming a Vigenère cipher. We next look for repetitions. The following table summarizes them:

Repetition	Begins at	Ends at	Interval Length	Factorization of Interval Length
ts	0	40	40	2 2 2 5
ts	0	59	59	59
sm	1	105	104	2 2 2 13
pp	6	30	24	2 2 2 3
pc	7	38	31	31
hp	15	120	105	3 5 7
fa	21	79	58	2 29
fa	21	118	97	97
ue	23	83	60	2 2 3 5
pi	31	109	78	2 3 13
ms	33	104	71	71
sf	34	74	40	2 2 2 5
ip	37	92	55	5 11
ts	40	59	19	19
pk	43	57	14	2 7
lvf	61	116	55	5 11
krg	68	88	20	2 2 5
fa	79	118	39	3 13
iyo	110	135	25	5 5

Of these, the interval lengths have the factor 2 appear 9 times, 3 appears 5 times, 4 appeears 6 times, 5 appears 7 times, 6 appears 3 times, 7 appears 2 times, and 8 and 10 appear 4 times each. Given the IC indicates a key length of greater than 2, we will first try a key of length 5, as that is the most frequent factor greater than 2.

Splitting the ciphertext into 5 parts, we compute the IC for each alphabet. They are:

alphabet #1: 0.042	alphabet #3: 0.050	alphabet #5: 0.48
alphabet #2: 0.066	alphabet #4: 0.066	

Of these, alphabets 2 and 4 have an IC indicating a key length of 1, alphabets 3 and 5 have an IC indicating a key length of between 2 and 3, and alphabet 1 has an IC indicating a key length of between

5 and 10. Given the short ciphertext, we will accept that our hypothesis of a key length of 5 is worth pursuing, and proceed accordingly.

We next lay out the frequencies for each alphabet:

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

alphabet #1: 1 0 1 1 2 2 2 3 3 0 0 0 1 1 1 0 2 3 3 0 0 1 0 0 0

alphabet #2: 1 0 0 3 0 1 0 1 0 0 0 4 1 1 2 5 0 1 3 0 0 0 0 1 2 2

alphabet #3: 2 2 0 1 1 0 1 0 5 0 0 0 3 1 2 2 2 0 1 1 1 2 1 0 0 0

alphabet #4: 0 0 4 1 0 2 2 0 1 2 4 0 1 0 0 3 1 0 0 0 4 2 0 0 0 1

alphabet #5: 2 0 1 0 2 1 0 1 1 0 1 2 2 0 0 2 0 3 3 1 0 0 4 2 0 0

First consider the second alphabet. The frequencies in the middle (4 1 1 2 5 0 1 3), representing counts for the letters l through s, are similar to the frequency counts expected at the beginning of the unshifted alphabet. So, let us assume that the second alphabet maps A into L. In the fourth alphabet, the frequencies from w to z, and a and b, are 0 0 0 1 0 0, and are similar to the frequency counts expected at the end of the unshifted alphabet. So, let us assume the fourth alphabet maps A into C. For the first alphabet, the frequencies from the letter on match the frequency counts expected at the end of the unshifted alphabet, so let us assume the first alphabet maps A into A.

We now substitute back into the ciphertext. As in the book, the bold letters are plaintext and the unbolded letters are the ciphertext:

THM**T**M	**ME**P**A**W	**CO**U**E**X	**HE**E**A**P	**RU**A**S**E	**ID**B**O**W	**PE**I**K**S	**FM**I**N**C	**TH**Q**N**K	**SO**N**S**L
OEA**A**R	**DS**P**I**T	**SA**V**D**W	**EA**T**I**R	**GW**I**X**S	**FC**I**B**F	**AG**M**S**E	**ND**S**I**R	**GS**I**N**H	**WH**G**T**L
ESM**A**M	**SB**W**I**P	**IN**O**H**S	**TA**V**D**A	**HE**B**H**I	**RP**Q**G**W	**HA**D**E**A	**IN**O**S**X		

The first word is either “THE” or “THAT”. But if it’s “THAT”, then the third alphabet is unshifted, and the frequency counts do not match those of the unshifted alphabet. If it’s “THE”, on the other hand, then E maps into M, meaning A maps into I, and the frequency counts are closer to those of the unshifted alphabet than if the first word were “THAT”. Adopting as a working hypothesis that the third alphabet maps A into I, and updating the text as before, we have:

THETM	**MEHA**W	**COME**X	**HEWA**P	**RUSS**E	**IDTO**W	**PEAK**S	**FMAN**C	**THIN**K	**SOFS**L
OESAR	**DSHI**T	**SAND**W	**EALI**R	**GWAX**S	**FCAB**F	**AGES**E	**NDKI**R	**GSAN**H	**WHYT**L
ESEAM	**SBOI**P	**INGH**S	**TAND**A	**HETH**I	**RPIG**W	**HAVE**A	**INGS**X		

At this point, the cipher falls apart. The “W” at the end of the second block is obviously “S”, so the fifth alphabet maps A into E. The plaintext is:

THETI	**MEHAS**	**COMET**	**HEWAL**	**RUSSA**	**IDTOS**	**PEAKO**	**FMANY**	**THING**	**SOFSH**

OESAN	**DSHIP**	**SANDS**	**EALIN**	**GWAXO**	**FCABB**	**AGESA**	**NDKIG**	**GSANN**	**WHYTH**
ESEAI	**SBOIL**	**INGHO**	**TANDW**	**HETHE**	**RPIGS**	**HAVEW**	**INGST**		

So, the key is ALICE, and the plaintext is:

‘The time has come,’ the Walrus said,

‘To speak of many things:

Of shoes—and ships—and sealing wax—

Of cabbages—and kings—

And why the sea is boiling hot—

And whether pigs have wings.’

This is a part of a poem from Through the Looking Glass (and yes, I know it should be “talk”, not “speak”, in the second line!)