Spring 2015
Chris Christensen
MAT/CSC 483

Cryptanalysis of the Vigenère Cipher: Kasiski Test

The keyword of a Vigenère cipher describes the rotation among the Caesar cipher alphabets that are used. That rotation leads to patterns that can be exploited by a cryptanalyst. If we know the length of the keyword, we can often determine the keyword and, hence, decrypt all messages encrypted with that keyword.

Here is a ciphertext message that has been encrypted with a Vigenère cipher.

```
nifon aicum niswt luvet vxshk nissx wsstb husle chsnv ytsro
cdsoy nisgx lnona chvch gnonw yndlh sfrnh npblr yowgf unoca
cossu ouoll iuvef issoe xgosa cpbew uormh lftaf cmwak bbbdv
cqvek muvil qbgnh ntiri ljgig atwnv yuvev iorim cpbsb hxviv
buvet vxshk uorim mjbdb pjrut fbueg ntgof yuwmx miodm ipdek
uuswx lfjek sewfy yssnm zscmm bpgeb huvez ysaag usaew mffvb
wfgim qpilw bbjeu yfbef vbfrt mtwnz uorig wpbvx hjsnm zpfag
uhsnm npglb jbqrh mttrh huwek mpfak ljjen hbbnh ooqew vzdak
udvum yucbx yoquf vffew vzonx hjumt lfgef vmwnz uxsiz bumag
xbbtb kvotx xumpx qswtx l
```

Assume that, somehow, we have discovered that the keyword has length five (which is conveniently the same as the size of the blocks). Then the first letter of each block is encrypted with the same row of the Vigenère square – they are encrypted with the same Caesar cipher. Similarly, the second letter of each block is encrypted with the same row – the same Caesar cipher. The third letters with the same Caesar cipher. The fourth letters with the same Caesar cipher. And, the fifth letters with the same Caesar cipher.

Because Caesar ciphers are easily broken by frequency analysis, we can discover the letters of the keyword. Here is how we can proceed.

Strip off the first letters of each block and do a frequency analysis on the result. They should have all been encrypted with the same Caesar cipher.

Alphabet number one – first letters of each block

```
A     11
B     11111
C     11111111
D
E
F     1
G     1
H     1111111
I     1111
J     1
K     1
L     11111111
M     1111111
N     11111111
O     11
P     1
Q     111
R
S     11
T
U     111111111
V     1111111
W     111
X     111
Y     1111111111
Z     11
```

It appears that ciphertext Y corresponds to plaintext e. (Not just because it is the most frequent letter but because all the high frequency letter patterns fit – U would correspond to a; C would correspond to i; H and I would correspond to n and o; and L, M, and N would correspond to r, s, and t.)

Now recall that when we are encrypting using a Vigenère square plaintext a corresponds to the first letter of the row being used – the letter of the keyword being used. So, it appears that (because U corresponds to a) the first letter of the keyword is u.

The keyword is u _ _ _ _.

Alphabet number two – second letters of each block

```
A
B    11111111
C
D    11
E    1
F    11111111
G    1
H    111
I    1111111
J    11111
K
L
M    11
N    1111
O    11111111
P    1111111111
Q    1
R
S    1111111
T    111111
U    11111111111111
V    1
W
X    1111
Y
Z    11
```

It appears that ciphertext F corresponds to plaintext e.

So, it appears that (because B corresponds to a) the second letter of the keyword is b.

The keyword is u b _ _ _.

Alphabet number three – third letters of each block

```
A    11
B    111111111
C    111
D    111
E
F    111111
G    1111111
H
I    11
J    111
K
L
M    11
N
O    11111111
P
Q    111
R    111111
S    1111111111111111
T    11
U    11
V    1111111111
W    111111111
X
Y
Z
```

It appears that ciphertext S corresponds to plaintext e.

So, it appears that the third letter of the keyword is o.

The keyword is u b o _ _. (Perhaps, you can already guess the keyword.)

Alphabet number four – fourth letters of each block

```
A     1111111
B     1
C     11
D     111
E     11111111111111111
F     1
G     11
H     11
I     11111111
J
K
L     111111
M     1111
N     1111111111111
O     1111
P     1
Q
R     11111
S     1111
T     1111
U     1111
V     11
W     11
X
Y
Z
```

It appears that ciphertext E corresponds to plaintext e.

So, it appears that the third letter of the keyword is a.

The keyword is u b o a _.

Alphabet number five – fifth letters of each block

```
A    11
B    1111111
C
D
E    11
F    1111111
G    111111
H    11111111
I    1
J
K    111111111
L    11
M    1111111111
N    11
O    1
P
Q
R    1
S
T    111111
U    11
V    11111
W    111111
X    1111111111
Y    11
Z    1111
```

It appears that ciphertext `X` corresponds to plaintext `e`.

So, it appears that the last letter of the keyword is `t`.

The keyword is `u b o a t`.

So, knowing just the length of the keyword, we were able to determine the keyword.

Two methods give us information about the length of the keyword of a Vigenère cipher – the Kasiski test and the Friedman test. We will discuss the first in this section and the second in the next section.

First, some history.

The Cryptanalysts

The Vigenère cipher might first have been broken by the English mathematician Charles Babbage (1792 – 1871); Kahn quotes Babbage as saying "an indistinct glimpse of defeating it presented itself vaguely to my imagination." But, if Babbage had a solution, he never published it. Babbage apparently had the tendency to never be satisfied with a work and to continue to refine things; so, he might never have been satisfied enough with his solution to publish it.

Friedrich Kasiski (1805 – 1881) is credited with breaking the Vigenère cipher in 1863. From the Sixteenth Century until the Nineteenth Century the cipher was generally considered to be secure. We will use Kasiski's technique to determine the length of the keyword.

In the Twentieth Century, William Frederick Friedman (1891 – 1969), the dean of American cryptologists, developed a statistical method to estimate the length of the keyword.

Friedrich Kasiski

"Friedrich Kasiski was born in November 1805 in a western Prussian town and enlisted in an East Prussian infantry regiment at the age of 17.

He moved up through the ranks to become a company commander and retired in 1852 as a major. Although he had become interested in cryptology during his military career, it was not until the 1860s that he put his ideas on paper. In 1863 his 95-page text *Die Geheimschriften und die Dechiffrirkunst* (*Secret Writing and the Art of Deciphering*) was published. A large part of its contents addressed the solution of polyalphabetic ciphers with repeating keywords, a problem that had tormented cryptanalysts for centuries.

Disappointed by the lack of interest in his findings, Kasiski turned his attention to other activities including anthropology. He took part in artifacts searches and excavations and wrote numerous archeological articles for scholarly journals. He died in May 1881 not realizing the significance of his cryptanalytic findings." Wrixon, Fred B., *Codes, Ciphers & other Cryptic & Clandestine Communication: Making and breaking secret messages from hieroglyphs to the internet*, Black Dog & Leventhal Publishers.

The Kasiski Test (or the Kasiski Attack)

Here is a message enciphered with a Vigenère cipher. (It is taken from: Beutelspacher, Albrecht, *Cryptology: An introduction to the Art and Science of Enciphering, Encrypting, Concealing, Hiding and Safeguarding Described Without any Arcane Skullduggery but not Without Cunning Waggery for the Delectation and Instruction of the General Public*, Mathematical Association of America, 1996.)

```
DBZMG   AOIYS   OPVFH   OWKBW   XZPJL   VVRFG   NBKIX
DVUIM   OPFQL   VVPUD   KPRVW   OARLW   DVLMW   AWINZ
DAKBW   MMRLW   QIICG   PAKYU   CVZKM   ZARPS   DTRVD
ZWEYG   ABYYE   YMGYF   YAFHL   CMWLW   LCVHL   MMGYL
DBZIF   JNCYL   OMIAJ   JCGMA   IBVRL   OPVFW   OBVLK
OPVUJ   ZDVLQ   XWDGG   IQEYF   BTZMZ   DVRMM   ANZWA
ZVKFQ   GWEAL   ZFKNZ   ZZVCK   VDVLQ   BWFXU   CIEWW
OPRMU   JZIYK   KWEXA   IOIYH   ZIKYV   GMKNW   MOIIM
KADUQ   WMWIM   ILZHL   CMTCH   CMINW   SBRHV   OPVSO
DTCMG   HMKCE   ZASYD   JKRNW   YIKCF   OMIPS   GAFZK
JUVGM   GBZJD   ZWWNZ   ZVLGT   ZZFZS   GXYUT   ZBJCF
PAVNZ   ZAVWS   IJVZG   PVUVQ   NKRHF   DVXNZ   ZKZJZ
ZZKYP   OIEXX   MWDNZ   ZQIMH   VKZHY   DVKYD   GQXYF
OOLYK   NMJGS   YMRML   JBYYF   PUSYJ   JNRFH   CISYL
N
```

We begin the attack by frequency analysis.

A	11111111111111111		19
B	11111111111111		14
C	11111111111111111		16
D	11111111111111111111	20	
E	11111111		8
F	11111111111111111		19
G	11111111111111111111	20	
H	111111111111		12
I	1111111111111111111111111		26
J	11111111111111111		16
K	1111111111111111111111111	25	
L	111111111111111111111		22
M	11111111111111111111111111111111		32
N	11111111111111111		16
O	111111111111111111		18
P	11111111111111111		16
Q	1111111111	10	
R	11111111111111		14
S	11111111111		11
T	111111		6
U	11111111111		11
V	1111111111111111111111111111111111		34
W	111111111111111111111111111111		28
X	1111111111	10	
Y	1111111111111111111111111111		27
Z	11111111111111111111111111111111111111111		41
			491

The "relatively equal" frequencies suggest multiple alphabets – a polyalphabetic cipher, which, for us, would suggest a Vigenère cipher.

Here's the idea behind the Kasiski test. Consider a Vigenère cipher with keyword *Galois*. (My favorite mathematician.)

```
abcdefghijklmnopqrstuvwxyz
GHIJKLMNOPQRSTUVWXYZABCDEF
ABCDEFGHIJKLMNOPQRSTUVWXYZ
LMNOPQRSTUVWXYZABCDEFGHIJK
OPQRSTUVWXYZABCDEFGHIJKLMN
IJKLMNOPQRSTUVWXYZABCDEFGH
STUVWXYZABCDEFGHIJKLMNOPQR
```

Think of a common trigraph – say, `the` – and assume that it appears twice in the plaintext message.

Depending on the way the keyword and `the` align, there are six different ways that `the` could be enciphered.

```
GALOISGALOISGALOISGALOISGALOISGALOIS
the    the    the    the    the    the
ZHP    TSS    EVM    HPW    BZK    LNE
```

If we are lucky and `the` is encrypted by the same three alphabets, we would see a duplicate trigraph.

```
GALOISGALOISGALOISGALOIS...GALOISGALOISGALOIS
the                           the
ZHP                           ZHP
```

What is important to notice is that the distance between the beginnings of the `ZHP` trigraphs is a multiple of the length of the keyword. This provides information about the length of the keyword.

So, we search through the ciphertext for trigraphs (or strings of other lengths), and we look for repetitions. Sometimes, of course, the repetitions are just accidental – two different strings of three letters are encrypted into the same three-letter string by different alphabets, but sometimes the repetitions correspond to the same three-letter string being encrypted by the same three alphabets. These are the occurrences that we would like to discover.

Here are the trigraphs of the ciphertext with the number of repetitions shown in ( ):

AOI ARL AWI AKB AKY ARP ABY AFH AJJ AIB ANZ AZV ALZ AIO ADU ASY AFZ AVN AVW

BZM BWX BKI BWM BYY(2) BZI BVR BVL BTZ BWF BRH BZJ BJC

CGP CVZ CMW CVH CYL CGM CKV CIE CMT CHC CMI CMG CEZ CFO CFP CIS

DBZ(2) DVU DKP DVL(3) DAK DTR DZW DGG DVR DUG DTC DJK DZW DVX DNZ DVK DGQ

EYG EYM EYF EAL EWW EXA EZA EXX

GAO GNB GPA GAB GYF GYL GMA GGI GIQ GWE GMK GHM GAF GMG GBZ GTZ GXY GPV GQX GSY

HOW HLC(2) HLM HZI HCM HVO HMK HFD HVK HYD HCI

IYS IXD IMO INZ IIC ICG IFJ IAJ IBV IQE IEW IYK IOI IYH IKY IIM IMK IMI ILZ INW IKC IPS IJV IEX IMH ISY

JLV JNC JJC JCG JZD JZI JKR JUV JDZ JCF JVZ JZZ JGS JBY JJN JNR

KBW(2) KIX KPR KYU KMZ KOP KFQ KNZ KVD KKW KWE KYV KNW KAD KCE KRN KCF KJU KRH KZJ KYP KZH KYD KNM

LVV(2) LWD LMW LWQ LCM(2) LWL LCV LMM LDB LOM LOP LKO LQX LZF LQB LZH LGT LYK LJB

MGA MOP MWA MMR MRL MZA MGY(2) MWL MMG MIA MAI MZD MMA MAN MUJ MKN MOI MKA MWI MIL MTC MIN MGH MKC MIF MGB MWD MHV MJG MRM MLJ

NBK NZD NCY NZW NWM NWS NWY NZZ(5) NKR NMJ NRF

OIY OPV(4) OWK OPF OAR OMI(2) OBV OPR OIY OII ODT OIE OCL OLY

PVF(2) PJL PFQ PUD PRV PAK PSD PVU(2) PRM PVS PSG PAV POI PUS

QLV QII QXW QEY QGW QBW QWM QNK QIM QXY

RFG RVW RLW(2)RPS RVD RLO RMM RMU RHV RNW RHF RML RFH

SOP SDT SBR SOD SYD SGA SGX SIJ SYM SYJ SYL

TRV TZM TCH TCM TZZ TZB

UIM UDK UCV UJZ(2) UCI UQW UVG UTZ UVQ USY

VFH VVR VRF VUI VVP VPU VWO VLM VZK VDZ VHL VRL VFW VLK VUJ VLQ(2) VRM VKF VCK VDV VGM VOF VSO VGM VLG VNZ VWS VZG VUV VQN VXN VKZ VKY

WKB WXZ WOA WDV WAW WIN WMM WQI WEY WLW WLC WOB WDG WAZ WEA WFX WWO WOP WEX WMO WMW WIM WSB WYI WWN WNZ WSI WDN

XZP XDV XWD XUC XAI XYU XNZ XXM XMW XYF

YSO YUC YGA YYE YEY YMG YFY YAF YLD YLO YFB YKK YMZ YVG YDJ YIK YUT YPO YDV YDG YFO YKN YMR YYF YFP YJJ YLN

ZMG ZPJ ZDA ZKM ZAR ZWE ZIF ZDV(2) ZMZ ZWA ZVK ZFK ZZZ(2) ZZV(2) ZVC ZIY ZIK ZHL ZAS ZKJ ZJD ZWW ZVL ZZF ZFZ ZSG ZBJ ZZA ZAV ZGP ZZK(2) ZKZ ZJZ ZKY ZZQ ZQI ZHY

Whew!

After identifying the repeated trigraphs, we look at their spacing.

BYY	**360**		**BYY**				
DBZ	**140**		**DBZ**				
DVL	121		**DVL**	**50**		**DVL**	
HLC	**170**		**HLC**				
KBW	**55**		**KBW**				
LVV	**20**		**LVV**				
LCM	**170**		**LCM**				
MGY	**20**		**MGY**				
NZZ	**140**		**NZZ**	**25**		**NZZ**	**25**
	NZZ	**20**		**NZZ**			
OPV	**155**		**OPV**	**10**		**OPV**	**135**
	OPV						
OMI	**190**		**OMI**				
PVF	**155**		**PVF**				
PVU	224		PVU				
RLW	**20**		**RLW**				
UJZ	71		UJZ				
VLQ	**50**		**VLQ**				
ZZV	139		ZZV				
ZZK	6		ZZK				
ZDV	19		ZDV				
ZZZ	**195**		**ZZZ**				

Remember that we are looking for a length that is a common divisor of "all" of these lengths – well, not “all” because some repetitions are accidental – but most. The bolded portions of the table seem to indicate that the length of the keyword might be five.

Now we return to the ciphertext and separate it into its five alphabets. We begin with the first letter and take every fifth letter after it. Then take the second letter and every fifth letter after it. Then take the third letter and every fifth after it. Etc.

If we determined the length of the keyword correctly, we should have partitioned the ciphertext into five sets of ciphertext letters each of which was encrypted with a Caesar cipher. Then we proceed as we did for the first example of this section.

Let us look at each alphabet separately.

Alphabet number one

```
A     1111
B     11
C     111111
D     1111111111
E
F
G     111111
H     1
I     11111
J     1111111
K     111
L     1
M     1111
N     1111
O     1111111111111
P     1111
Q     1
R
S     1
T
U
V     1111
W     1
X     11
Y     1111
Z     1111111111111111
```

It appears that Z might correspond to e; that would make V correspond to a. The first letter of the keyword would be v. v _ _ _ _

Alphabet number two

```
A    1111111111
B    1111111111
C    11
D    11
E
F    1
G
H
I    111111
J    1
K    1111
L    1
M    1111111111111
N    111
O    1111
P    1111111
Q    111
R
S
T    111
U    11
V    11111111111
W    111111111
X    1
Y
Z    11111
```

It appears that `M` might correspond to `e`; that would make `I` correspond to `a`. The second letter of the keyword would be `i`. `v i _ _ _` (Guess?)

Alphabet number three

```
A
B
C    11
D    111
E    111111
F    11111
G    111
H
I    1111111111
J    11
K    111111111111
L    111
M
N
O
P    11
Q
R    1111111111111
S    111
T    1
U    11
V    11111111111111
W    111
X    11
Y    111
Z    111111111
```

It appears that V might correspond to e; that would make R correspond to a. The third letter of the keyword would be r. v i r _ _ (Guess?)

Alphabet number four

```
A     11
B     11
C     111111
D
E
F     11111
G     1111
H     111111
I     11111
J     111
K     1
L     111111
M     111111111
N     111111111
O
P     11
Q     1
R     1
S     1
T
U     1111
V     111
W     111
X     111
Y     1111111111111111111
Z     111
```

It appears that `Y` might correspond to `e`; that would make `U` correspond to `a`. The fourth letter of the keyword would be `u`. `v i r u _` (Not many possibilities.)

Alphabet number five

```
A    111
B
C
D    11111
E    11
F    11111111
G    1111111
H    11111
I
J    111
K    11111
L    11111111111
M    111111
N
O    1
P    1
Q    11111
R
S    111111
T    11
U    111
V    11
W    111111111111
X    11
Y    1
Z    11111111
```

It appears that `W` might correspond to `e`; that would make `S` correspond to `a`. The fifth letter of the keyword would be `s`. The keyword would be `v i r u s`. Notice that we know the keyword, but we have not yet deciphered the message.

Exercises

1. Here is a message that was encrypted with a Vigenère cipher with a keyword of length 6.

```
wgixf irtnx amwpz gfcln bztef roozn maour tlrno dsxjw
xxdan zhdix nqtta hogcm rwrvj numyb gxavt mgzdt ewlqs
wvwtm lgblk nrins ozgif bgnlm fpsqn xhvja ufgmj xyxum
hqsxv vztea bzrpt lrijy ivnto fywew uyfse beiaw vbimm
igwhq ceytk ppien udmiq nkmtw bnidy tgitm lcfqy fhegp
ghewv viqbi pwsql itmtp avlzk mdmao gxmsf bgxls sokdm
eagyz azntg zdhvx rameg qifre snood yeqts tlreg dirpt
apirt bnqfe zwaez mzukd meavz qlitz gytln bxqvi ntkpg
ieugz ywczu bowrl gxlmn viqwm gpsqx mpwgs amaaz fhihv
ofehf bgfew nvjih sfmju sgywy grium rbehc zubep nukdi
gnqtf oxumc mr
```

1a. Find the first letter of the keyword.
1b. Find the second letter of the keyword.
1c. … the third letter … .
1d. …the fourth letter … .
1e. … the fifth letter … .
1f. … the sixth letter … .
1g. Find the keyword.
1h. Decrypt the message.

2a. Use a Kasiski test to determine the length of the keyword.

```
tfrvg akceo ekiii brjgy obqgr nfbuk zimme tfyeb puwyr vqibj
ymeyv bfwyc actpu gwvvm akout cnzxx zvnaz ojbgu tpzkg ukcrv
punhk zlsth jqbtu fvbpn eypxn xdvcm giafv gvzly cnjgl ujunr
enfea uyiil vttlp lhreh vafgu lzkxh nccbh xuuel vymlz kgogn
ymoxb wbnvk exjws hnwbq amaud auoky scgom frgnz mbvor ufykz
vivye brxhz zvgke fvkga tfvvt rthue ukkyk prstc eznoe qrgsx
hgcwa jhhkw gnxhf zgnuc jmpzx xkprh kueku rbhvn eufio nbxwn
fknsu lzltk cojbg umbvv bxmbn mfzhz kprxt ccene coekg ohhfv
nyecx pgnbf coegv yujlg guekv kgntp hxvbr vquoy itbud ceogn
xwcil vbnjw szaym iyrxs jbbuw vcmgi afvgc gke
```

Here is a list of trigraphs that occur more than once:

Trigraph	**Frequency**		
kpr	3	kgo	2
afv	2	lzk	2
bgu	2	mbv	2
ceo	2	mgi	2
cmg	2	oek	2
coe	2	ogn	2
fvg	2	ojb	2
gia	2	rvq	2
gke	2	uek	2
gnx	2	ulz	2
iaf	2	vcm	2
ilv	2	vkg	2
jbg	2	yeb	2
jws	2	zkg	2
		zxx	2

2b. Find the keyword.
2c. Decrypt the message.

3a. The word `the` is encrypted exactly the same way twice in a ciphertext, and the distance between the two appearances is 24. What are the possible lengths of the keyword?

3b. The word `the` is encrypted exactly the same way twice in a ciphertext, and the distance between the two appearances is 15. What are the possible lengths of the keyword?

3c. The word `the` is encrypted exactly the same way twice in a ciphertext, and the distance between the two appearances is 7. What are the possible lengths of the keyword?

3d. The word `the` is encrypted exactly the same way twice in a ciphertext, and the distance between the two appearances is 6. What are the possible lengths of the keyword?

4. The following are the intervals between repeated trigraphs in a ciphertext. For each, what do you suspect the length of the keyword to be?

4a. 6, 12, 48, 7.
4b. 11, 32, 56, 3, 8, 16.
4c. 7, 35, 105, 2, 7, 42.

5. If the keyword of a Vigenère cipher has repeated letters, what problems might this cause for the Kasiski test? (It causes the same problems for the Friedman test.)

6. In the section "Cryptography of the Vigenère Cipher," exercises 3, 6, 7, 8, and 9 describe variants of the Vigenère cipher. For each of these variants, describe how successful the Kasiski attack would be and how you might modify the attack.

7. Vigenère suggested using long keyphrases. Why is this a good idea?

8. If a short message were encrypted with a Vigenère cipher, how likely is it that we could successfully attack it?