SPCS Cryptography Class Lecture 2

June 23, 2015

Vigenere cipher

Recall how Vigenere cipher works.

Example

Encrypt `cryptography` using the key `lemon`.

Solution

Plaintext	*cryptography*
Key	*lemonlemonle*
Ciphertext	*NVKDGZKDOCSC*

What can we do to crack this cipher?

Vigenere cipher

- The important thing here is the length of ciphertext vs length of the key.
- If the key has about the same length (or even longer) as the ciphertext, then the encryption is essentially random (why?), and it's impossible to decipher the text.
- However, if the length of ciphertext is much longer than the key, then we can do something.
- This was first figured out by Kasiski, a German military officer 150 years ago.

Vigenere cipher

Here is the basic idea:

- Some words/letter combinations (for example, "the") in the plaintext may repeat itself.
- Moreover, when the key is much shorter than the plaintext, the words/letter combinations may get encrypted by the SAME shift.
- This would give us information about the KEY LENGTH.
- In particular, the distance between two occurences of same words would be a multiple of keylength.

Vigenere cipher

Example

Suppose the key is `MATH`, and the plaintext `him` appears at place 25, 53, and 164.

Plaintext	`thisissom...`
Key	`mathmathm...`
Position	`123456789...`

- At 25, `him` would be shifted by MAT, so the ciphertext becomes `TIF`.
- At 53, `him` would be shifted by MAT, so the ciphertext becomes `TIF`.
- At 164, `him` would be shifted by HMA, so the ciphertext becomes `OUM`.

In particular, `TIF` would appear at place 25 and 53 in the ciphertext. This is entirely because of having the same shift, which in turn is because the keylength 4 divides $53 - 25 = 28$.

Vigenere cipher

How does this help us?

A Vigenere cipher

```
ZPGDL RJLAJ KPYLX ZPYYG LRJGD LRZHZ QYJZQ
REPVM SWRZY RIGZH ZVREG KWIVS SAOLT NLIUW
OLDIE AQEWF IIYKH BJOWR HDOGC QHKWA JYAGG
EMISR ZQOQH OAVLK BJOFR YLVPS RTGIU AVMSW
LZGMS EVWPC DMJSV JQBRN KLPCF IOWHV KXJBJ
PMFKR QTHTK OZRGQ IHBMQ SBIVD ARDYM QMPBU
NIVXM TZWQV GEFJH UCBOR VWPCD XUWFT QMOOW
JIPDS FLUQM OEAVL JGQEA LRKTI WVEXT VKRRG
                  XANI
```

http://www.cryptoclub.org/tools/crack_vigenerecipher.php

Vigenere cipher

- We look for repeated trigrams,

Trigram	Appears at places	Difference
AVL	117 and 258	$141 = 3 \cdot 47$
BJO	86 and 121	$35 = 5 \cdot 7$
DLR	4 and 25	$21 = 3 \cdot 7$
GDL	3 and 24	$16 = 2^4$
LRJ	5 and 21	$98 = 2 \cdot 7^2$
MSW	40 and 138	$84 = 2^2 \cdot 3 \cdot 7$
PCD	149 and 233	$13 = 13$
QMO	241 and 254	$98 = 2 \cdot 7^2$
VMS	39 and 137	$84 = 2^2 \cdot 3 \cdot 7$
VWP	147 and 231	$84 = 2^2 \cdot 3 \cdot 7$
WPC	148 and 232	$21 = 3 \cdot 7$
ZHZ	28 and 49	$21 = 3 \cdot 7$

- This suggests that the keylength is probably 7.

Vigenere cipher

- What next?
- Since the alphabet at position 1, 8, 15, ... are all shifted by the same key, this is one single Caesar cipher!
- Can we just test all the 26 shifts as before?

Vigenere cipher

- Fortunately, frequency analysis still works.
- Key observation: The most frequent letter in English is E, T and A, and the four consecutive letters that appear most infrequently is $VXYZ$.
- So if we have a consecutive batch of five letters with low frequency (corresponding to *vwxyz*), surrounded by two alphabets with high frequency (corresponding to t and a), it's likely that we found the shift.

Vigenere cipher

- So let us break the message into seven chunks, $s_1, \cdots, s_7$, with s_1 stands for the string for 1st, 8th, 15th alphabet, s_2 stands for 2nd, 9th, 16th alphabet and so on.
  - $s_1 =$ `zlxrhrrhwloehdweoklilwvlhphqbynwhwfjulrxx`
- The frequency table for s_1 is as follows

	A	B	C	D	E	F	G	H	I	J	K	L	M
Freq	0	1	0	1	2	1	0	6	1	1	1	6	0
	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Freq	1	2	1	1	4	0	0	1	1	5	3	1	1

- Rarest chunks: *ZABCD*, *YZABC*.

Vigenere cipher

- If $ZABCD = vwxyz$, then $E = a$, $X = t$, and $I = e$. Is this likely?
- If $YZABC = vwxyz$, then $D = a$, $W = t$, $H = e$. Is this likely? What should wvxyz correspond to?
- Shift should be 3 (D), and decrypted s_1 is

  wiuoeooetilbeatblhifitsiemenyvktetcgriouu

Vigenere cipher

What about the other $s_2, \cdots, s_7$?

$$s_2 = \texttt{pazjzezzitlwboamqbvuzpjpvmtiimiquptiqjkta}$$
$$s_3 = \texttt{gjpgqpyvvndfjgjihjpagcqckfkhvqvvccqpmgtvn}$$
$$s_4 = \texttt{dkydyvrrsliiocysoosvmdbfxkobdmxgbdmdoqiki}$$
$$s_5 = \texttt{lpyljmiesieiwqarafrmsmrijrzmapmeoxoseewr}$$
$$s_6 = \texttt{rygrzsggauayrhgzvrtsejnobqrqrbtfruofaavr}$$
$$s_7 = \texttt{jllzqwzkowqkhkgqlygwvskwjtgsduzjvwwlvleg}$$

Let's try it out!

Vigenere cipher

A Vigenere cipher

`Whether I shall turn out to be the hero of my own life, or whether that station will be held by anybody else, these pages must show. To begin my life with the beginning of my life, I record that I was born (as I have been informed and believe) on a Friday, at twelve o'clock at night. It was remarked that the clock began to strike, and I began to cry, simultaneously.`
(Excerpt from *David Copperfield*, by Charles Dickens.)

Another method of finding key length: Friedman's index of coincidence.

Vigenere cipher

Another one?

```
TOGMG GBYMK KCQIV DMLXK KBYIF VCUEK CUUIS
VVXQS PWWEJ KOQGG PHUMT WHLSF YOVWW KNHHM
RCQFQ VVHKW PSUED UGRSF CTWIJ KHVFA THKEF
FWPTJ GGVIV CGDRA PGWVM OSQXG HKDVT WHUEV
KCWYJ PSGSN GFWSL JSFSE OOQHW TOFSH ACIIN
GFBIF GABGJ ADWSY TOPML ECQZW ASGVS FWRQS
    FSFVQ RHDRS NMVMK CBHRV KBLXK GZI
```

Variants of Vigenere cipher

- What if key length is comparable to length of message?
- Does Kaisiski's method still work?
- In any case, this idea is called *one time pad*, when the key length is at least the length of message.

One time pad

Example

Plaintext	`ihaveanarmywehaveahulk`
Key	`laegaaftqofawgdjnewuuv`
Ciphertext	`THEBEASTHADWANDEREDOFF`

One time pad

Example

Plaintext	`ihaveanarmywehaveahulk`
Key	`oxrjuohpjmkrcttyanmodd`
Ciphertext	`WEREYOUPAYINGATTENTION`

One time pad

Example

Plaintext	`ihaveanarmywehaveahulk`
Key	`yhdxjerzngylwmdtuonypo`
Ciphertext	`GODSNEEZESWHATDOYOUSAY`

One time pad

- Historically, one-time pad was first used in 1920s by German foreign office to protect their diplomatic correspondences.
- One-time pad and its variants were used extensively during WWII by Germany, UK (Special Operations Executive), US (NSA) etc.
- Attacks? If used properly, impossible. But problem becomes one of carrying the codebook.
- Colonel Oleg Penkovsky, spying for UK and USA during Cuban Missile Crisis, was executed in 1960s.
- Reuse of one time pad is bad!
  VERONA Project: US effort from 1943 to 1980 to decrypt messages sent by Soviet Union Intelligence agencies.