Introduction to Ciphers

Last Revised –May 3, 2006

Cunsheng Ding
cding@cs.ust.hk
www.cs.ust.hk/˜cding

Mordecai Golin
golin@cs.ust.hk
www.cs.ust.hk/˜golin

WELCOME TO HKUST.
THANK YOU FOR COMING TO THIS CLASS.
I WILL TRY TO MAKE THIS
AS MUCH FUN I CAN.

YGNEQOG VQ JMWUV
VJCPM AQW HQT EQOKPI VQ VJKU ENCUU
K YKNN VTA VQ OCMG VJKU
CU OWEJ HWP CU K ECP

DVOXLNV GL SPFHG
GSZMP BLF ULI XLNRMT GL GSRH XOZHH
R DROO GIB GL NZPV GSRH
ZH NFXS UFM ZH R XZM

Monoalphabetic Substitution Ciphers

`plaintext`

WELCOME TO HKUST.
THANK YOU FOR COMING TO THIS CLASS.
I WILL TRY TO MAKE THIS
AS MUCH FUN I CAN.

`shift every letter forward by two`

YGNEQOG VQ JMWUV
VJCPM AQW HQT EQOKPI VQ VJKU ENCUU
K YKNN VTA VQ OCMG VJKU
CU OWEJ HWP CU K ECP

`reverse the alphabet.  A→Z, B→Y, etc.`

DVOXLNV GL SPFHG
GSZMP BLF ULI XLNRMT GL GSRH XOZHH
R DROO GIB GL NZPV GSRH
ZH NFXS UFM ZH R XZM

Cryptography: The science of analyzing and deciphering codes and ciphers.

Cipher: Letters are replaced by other letters

Code: Words/phrases/concepts replaced by other words.

In a Monoalphabetic Substitution Cipher every character of a message is replaced with a uniquel alternate character. The mapping (replacement) is the key.

Monoalphabetic Substitution Ciphers (more)

Atbash Cipher: A MSC in which the alphabet is reversed, i.e., A$\longrightarrow$Z, B$\longrightarrow$Y, etc.
Very easy to remember.

Caesar Cipher: A MSC in which each letter is *shifted* (forwards or backwards) by the same amount.
The key is the shift amount.
Number of possible *keys* is 26.
Still easy to remember.

For an arbitrary MSC, the *key* is a *permutation* of
$\{a, b, c, \ldots, z\}$.
Number of possible *keys* is $26! \sim 4 * 10^{26}$.
Much harder to remember (must write down the key and it can be lost/stolen).

Tentative Plan

- Review of Modulo Arithmetic
- Additive, Multiplicative and Affine Ciphers
- Cryptanalysis: Breaking some Simple Ciphers

Review of Modulo Arithmetic

1) $x \pmod{n}$ is the *remainder* when x is divided by n.

$5 \pmod{3} = 2; \quad 32 \pmod{26} = 6.$

2) We use $-x$ to represent $n - x$.

$-7 = 3 \pmod{10}$

3) $[[x \pmod{n}] + [y \pmod{n}]] \pmod{n}$
$= [x + y] \pmod{n}$

$$
\begin{aligned}
[[28 \pmod{10}] + [19 \pmod{10}]] &= 17 \pmod{10} \\
&= 7 \pmod{10} \\
&= [28 + 19] \pmod{10}
\end{aligned}
$$

4) $[[x \pmod{n}] * [y \pmod{n}]] \pmod{n}$
$= [x * y] \pmod{n}$

$$
\begin{aligned}
[[28 \pmod{10}] * [19 \pmod{10}]] &= 72 \pmod{10} \\
&= 2 \pmod{10} \\
&= 572 \pmod{10} \\
&= [28 * 19] \pmod{10}
\end{aligned}
$$

Map the letters in the alphabet to the integers modulo 26.

A	B	C	D	...	W	X	Y	Z
1	2	3	4	...	23	24	25	0

We can define at least three natural types of ciphers:

- Additive (Caesar) Ciphers with shift k.
  $f(x) = x + k \pmod{26}$

- Multiplicative Ciphers
  $f(x) = x * k \pmod{26}$

- Affine Ciphers
  $f(x) = s(x + k) \pmod{26}$

Additive (Caesar) Ciphers with shift k

To *encipher*, map letter x to $f(x) = x + k \pmod{26}$

To *decipher* map letter x to $f(x) = x + (-k) \pmod{26}$

Example with $k = 3$.

```
meet me at the usual place at eight oclock
phhw ph dw wkh xvxdo sodfh dw hljkw rforfn
```

Note that after the enciphering step we will often ignore the word boundaries and write the text in blocks of fixed size, e.g.,

```
phh wph dww khx vxd oso
dfh dwh ljk wrf orf n
```

Multiplicative Ciphers

To *encipher*, map letter x to $f(x) = x * k \pmod{26}$

To *decipher* map letter x to $f(x) = x*(k^{-1}) \pmod{26}$
k^{-1} is number t such that $k * t = 1 \pmod{26}$.

Example with $k = 3$.

```
meet me at the usual place at eight oclock
mooh mo ch hxo kekcj vjcio ch oauxh sijsig
```

$$\begin{array}{rcl} m &=& 13 \\ e &=& 5 \\ t &=& 20 \end{array} \qquad \begin{array}{rcl} 13*3 &=& 13 \pmod{26} \\ 5*3 &=& 15 \pmod{26} \\ 20*3 &=& 8 \pmod{26} \end{array} \qquad \begin{array}{rcl} && \\ 15 &=& o \\ 8 &=& h \end{array}$$

To decipher multiply every letter in coded message by $9 = 3^{-1} \pmod{26}$.

$$\begin{array}{rcl} m &=& 13 \\ o &=& 15 \\ h &=& 8 \end{array} \qquad \begin{array}{rcl} 13*9 &=& 13 \pmod{26} \\ 15*9 &=& 5 \pmod{26} \\ 8*9 &=& 20 \pmod{26} \end{array} \qquad \begin{array}{rcl} && \\ 5 &=& e \\ 20 &=& t \end{array}$$

Multiplicative Ciphers (more) $f(x) = x * k \pmod{26}$

- If $gcd(k, n) = 1$ we say that k, n are *relatively prime*.

- If k, n are relatively prime then
  $\{k, 2k, 3k, \ldots (n-1)k\}$
  is a permutation of
  $\{1, 2, 3, \ldots, n-1\}$

- If k, n are *not* relatively prime then there exists some $r < n$ such that $rk = 0 \pmod{n}$
  Then $\{k, 2k, 3k, \ldots (n-1)k\}$
  is not a permutation of
  $\{1, 2, 3, \ldots, n-1\}$

- If k, n are relatively prime, then
  there exists some r such that $r = k^{-1} \pmod{n}$.

Multiplicative Ciphers (more) $f(x) = x * k \pmod{26}$

- No matter what k is, z always gets mapped to z.
  $f(0) = k * 0 = 0 \pmod{26}$

- For cipher to work correctly, must have that
  k and 26 are *relatively prime*,
  i.e., $k \in \{1, 3, 5, 7, 11, 15, 17, 19, 21, 23, 25\}$
  In this case, $f(x)$ defines a cipher,
  since it induces a permutation of the letters.

- If k and 26 are *not* relatively prime,
  $\exists r$ such that $k * r = 0 \pmod{26}$.
  Then $f(r) = 0$ so two letters get mapped to z
  and $f(x)$ is *not* a cipher.
  Example: $k = 4$.
  $f(13) = 72 \pmod 26 = 0$
  so "m" (13) maps to "z".
  Also, both "a" (1) and "n" (14) would
  map into "d" (4) since
  $14 * 4 = 56 = 4 \pmod{26}$

Affine Ciphers $f(x) = s(x+k) \pmod{26}$

In order to ensure that $f(x)$ defines a cipher we must, as in multiplicative ciphers, require that s and 26 be relatively prime.

To decipher use function
$g(y) = s^{-1} * y - k \pmod{26}$
Note that $g(f(x)) = x$.

Example with $k = 11$ and $s = 7$.
`meet me at the usual place at eight oclock`
An additive shift of $k = 11$ gives
`xppe xp le esp fdflw awlnp le ptrse znwznv`
A further multiplicative "shift" of $s = 7$ gives
`lhhi lh fi ich pbpfe gefth fi hjvci zteztx`

Cryptanalysis: Breaking some Simple Ciphers

Use *Statistical Analysis*. In a given language, each character has a characteristic *frequency*. By trying to match these frequencies to frequencies of characters appearing in coded message, one could try and guess the key.

More sophisticated analyses would use *digraph (pairs of letters)* frequencies or even *trigraph (triples of letters)* frequencies.

Example: The coded message is

```
YQQ FYQ MFF TQQ EGM XBX
MOQ MFQ UST FAO XAO W
```

Frequency table is

Letter	A	B	E	F	G	M	O	Q	S	T	U	W	X	Y
Frequency	2	1	1	5	2	4	3	6	1	2	1	1	3	2

Since Q is the most frequent letter in the message we guess that e in the plaintext maps to Q in the code. If this is an additive cipher then $k = 12$ (since $e = 5$ and $Q = 17$).

Decoding under this assumption gives

```
mee tme att heu sua lpl
ace ate igh toc loc k
```

Life is usually not this easy!

Next time we will see more sophisticated statistical attacks.