
POLYALPHABETIC CIPHERS AND THEIR MATHEMATICS

CIS 400/628 — Spring 2005
Introduction to Cryptography

This is based on Chapter 2 of Lewand.

A SIDE DISCUSSION ON FUNCTIONS: INJECTIVE, SURJECTIVE, AND BIJECTIVE

Consider the following:

$$L = \{ a, b, c, \dots, x, y, z \}$$

$$U = \{ A, B, C, \dots, X, Y, Z \}$$

$$f : L \rightarrow U$$

Some examples that match the type of f :

$$f_1(l) = \text{toupper}(l)$$

$$f_2(l) = \begin{cases} Z & \text{if } l = a \\ A & \text{otherwise} \end{cases}$$

$$f_3(l) = \begin{cases} B & \text{if } l = a \\ C & \text{if } l = b \\ \dots & \\ Z & \text{if } l = y \\ A & \text{if } l = z \end{cases}$$

INJECTIVE FUNCTIONS

▶ An **injective** function is **1-to-1**, i.e., each element in the output is mapped to by only one element in the input.

▶ For example, consider

$$f : \mathbb{Z}^+ \rightarrow \mathbb{Z}^+, f(i) = i + 1$$

▶ This is **1-to-1**.

▶ **But**, does $f^{-1}(0)$ exist?

▶ From the earlier definition of L, U, f_1 , and f_2 :

• $f_1(l) = \text{toupper}(l)$ is **injective**.

• $f_2(l) = \begin{cases} Z & \text{if } l = a \\ A & \text{otherwise} \end{cases}$ is **not!**

SURJECTIVE FUNCTIONS

- ▶ A **surjective** function maps at least one value onto every member of the output set.
- ▶ This is sometimes called an **onto** mapping.
- ▶ For example, consider

$$f : \mathbb{Z}^+ \rightarrow \mathbb{Z}^+, f(i) = i \operatorname{div} 2 + 1$$

$$f(1) = 1, f(2) = 2, f(3) = 2, f(4) = 3, f(5) = 3, \dots$$

- ▶ **But**, is it **injective**?

- ▶ Again, from our earlier definitions,

- $f_1(l) = \text{toupper}(l)$ is **surjective**.

- $f_2(l) = \begin{cases} Z & \text{if } l = a \\ A & \text{otherwise} \end{cases}$ is **not!**

BIJECTIVE FUNCTIONS

- ▶ A bijective function is both **1-to-1** and **onto**.
- ▶ I.e., it is both **injective** and **surjective**.
- ▶ Again recalling our definitions:

$$\bullet f_3(l) = \begin{cases} B & \text{if } l = a \\ C & \text{if } l = b \\ \dots & \\ Z & \text{if } l = y \\ A & \text{if } l = z \end{cases} \text{ is bijective.}$$
$$\bullet f_2(l) = \begin{cases} Z & \text{if } l = a \\ A & \text{otherwise} \end{cases} \text{ is not!}$$

- ▶ Recall that monoalphabetic ciphers are **bijective**.

SCAMBLING FREQUENCIES

- ▶ To get around the weakness of monoalphabetic ciphers, we need to scramble letter frequencies somehow.
- ▶ A **polyalphabetic substitution cipher** is a cipher in which there is **not** a 1–1 map between plaintext and ciphertext letters
- ▶ An example from Lewand:
 - Let $S = \{ 00, 01, 02, \dots, 99 \} =$ two digit strings
 - Define a map $a_i \mapsto S_i$, a subset of $S \quad \ni$
 - ★ S_0, \dots, S_{25} are a partition of S .
 - ★ (freq. of a_i) $\approx \|S_i\| / \|S\|$.
 - When encoding a_i pick a random element of S_i .
 - In the ciphertext, the freq. of all two digit seqs. is about the same.
- ▶ To analyze these schemes we need counting & probability.

EXAMPLE

Letter	Subset of S
a	15, 33, 37, 55, 57, 72, 91, 96
b	24
c	03, 39, 67
e	08, 12, 20, 46, 47, 59, 64, 79, 81, 85, 90, 94, 97
h	05, 16, 30, 42, 69, 99
k	77
r	21, 25, 65, 68, 92, 95
s	00, 28, 52, 63, 74, 78
t	07, 19, 23, 35, 38, 54, 70, 84, 89
u	09, 32

Starbucks at three \Rightarrow 52 38 33 65 24 32 39 77 00
15 70 07 69 68 59 46

COUNTING: I

The Multiplication Principle

If task 1 can be done p_1 ways and

task 2 can be done p_2 ways and

⋮

task k can be done p_k ways,

then the total number of ways of doing all k tasks is

$$p_1 \times p_2 \times \dots \times p_k$$

Examples How many are there of:

- ▶ License Plates with three letters followed by four digits
- ▶ License Plates as before, but no repeated chars.
- ▶ Monoalphabetic ciphers

PERMUTATIONS

A **permutation** is an ordering of a set of objects.

EXAMPLES

- a. How many perms. of $\{a, b, c\}$ are there?
- b. Four spies: A, B, C, D. We choose one a pilot and another as co-pilot. **Q:** How many ways are there of doing this?
- c. There are five spies. Choose one to go to Miami and another to go to Watertown. **Q:** How many ways can we do this?
- d. Same as above, but choose a 3rd to go to Jersey City.
- e. **Q:** How many perms. are there of r objects selected from a set of size n . (Notation: $P(r, n)$.)

COMBINATIONS

- ▶ A **combination** is a selection of r objects from a set of size n . (We don't worry about order.)
- ▶ The number of ways of selecting (choosing) r objects from a set of size n is:

$$C(n, r) = \frac{n!}{r!(n-r)!} = \frac{1}{r!} \cdot P(n, r).$$

We also write this $\binom{n}{r}$.

▶ EXAMPLE

Suppose a lottery ticket contains 6 numbers from $\{0, \dots, 39\}$.

Q: How many tickets are possible when order matters?

Q: How many when order doesn't matter?

PROBABILITY

TERMINOLOGY

sample space: the possible outcomes of an experiment

event: a subset of a sample space

In this course, sample spaces are usually finite.

To Determine the prob. of an event in a finite sample space

1. Determine the elements of S , the sample space
2. Assign a weight to each element of $S \ni$
each weight is ≥ 0
the weights sum to 1.
3. Probability of $E = \sum_{a \in E} \text{weight}(a)$.

BASIC PROPERTIES OF PROBABILITY

▶ $\neg E = \{ x \in S \mid x \notin E \}$.

▶ $p(\neg E) = 1 - p(E)$.

▶ For all E , $0 \leq p(E) \leq 1$.

▶ If E and F are events of S , then

$$p(E \cup F) = p(E) + p(F) - p(E \cap F).$$

▶ Computing $p(E \cap F)$ can be tricky.

EXAMPLE: Roll a 6 sided die.

▶ $p(\text{rolling an odd number}) = 1/2$.

▶ $p(\text{rolling a prime}) = 1/2$.

▶ $p(\text{rolling an odd prime}) = 1/3 \neq \frac{1}{2} \cdot \frac{1}{2}$.

INDEPENDENCE

DEFINITION Suppose $E, F \subseteq S$.

E and F are **independent** iff $p(E \cap F) = p(E) \cdot p(F)$.

E and F are **dependent** iff $p(E \cap F) \neq p(E) \cdot p(F)$.

DEFINITION

If an experiment is repeated in n independent trials

& if the probability of an event E is p ,

then the expected number of events ($\text{Exp}(E)$) is $p \cdot n$.

EXAMPLE:

Flipping a coin 5 times, $S = \{ \text{HHHHH}, \dots, \text{TTTTT} \}$

$$E(\text{no heads}) = 1/32 \quad E(1 \text{ head}) = 5/32$$

$$E(2 \text{ heads}) = 10/32 \quad E(3 \text{ head}) = 10/32$$

$$E(4 \text{ heads}) = 5/32 \quad E(5 \text{ heads}) = 1/32$$

$$\text{Exp. num. of heads} = \frac{1}{2} \cdot 5 = 2.5. \quad \leftarrow \text{how to interp.?$$

BACK TO CIPHERS

The problem with the monoalphabetic ciphers is that the frequency of characters is unchanged.

Vigenère Cipher

Plaintext = mollywillneverbreakthis. Key = chaos.

	m	o	l	l	y	w	i	l	l	n	e	v	e	r	b	r	e	a	k	t	h	i	s
+	c	h	a	o	s	c	h	a	o	s	c	h	a	o	s	c	h	a	o	s	c	h	a
<hr/>																							
	O	V	L	Z	Q	Y	P	L	Z	F	G	C	E	F	T	T	L	A	Y	L	J	P	S
-	c	h	a	o	s	c	h	a	o	s	c	h	a	o	s	c	h	a	o	s	c	h	a
<hr/>																							
	m	o	l	l	y	w	i	l	l	n	e	v	e	r	b	r	e	a	k	t	h	i	s

THE VIGENÈRE TABLE

	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
a	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
b	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	A
c	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
d	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
e	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
f	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
g	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
h	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
i	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
j	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
k	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
l	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
m	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
n	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
o	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
p	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
q	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
...																										
y	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
z	Z	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

CRYPTANALYSIS OF THE VIGENÈRE CIPHER

- ▶ Look at repeated strings & the distance between them (why?)

t h e l i g h t b r o w n f o x j u m p s
h o u n d h o u n d h o u n d h o u n d h

A V Y Y L N V N O U V K H S R E X O Z S Z

o v e r t h e l a z y d o g
o u n d h o u n d h o u n d

C P R U A V Y Y D G M X B J

- ▶ The Kasiski test: the g.c.d. of these distances is likely to be a multiple of the keyword length. (why?)
- ▶ In real text, there would be several repeated sequences.

THE FRIEDMAN TEST

The index of coincidence = the probability of selecting two random letters from a text and getting the same letter

n = the number of letters in a text

n_i = the number of a_i in a text

Prob. of selecting two random letters & getting two 'a's

$$= \binom{n_0}{2} / \binom{n}{2} = \frac{n_0(n_0 - 1)/2}{n(n - 1)/2} = \frac{n_0(n_0 - 1)}{n(n - 1)}$$

∴ Prob. that two randomly chosen letters are the same

$$= \sum_{i=0}^{25} \frac{n_i(n_i - 1)}{n(n - 1)} = \text{IC}(T) \approx \sum_{i=1}^{25} \left(\frac{n_i}{n} \right)^2$$

For English: $\text{IC} \approx 0.065$ For random text: $\text{IC} \approx 0.038$

So?

USING THE FRIEDMAN TEST

- ▶ For a monoalphabetic cipher, the $IC \approx 0.065$.
- ▶ So, if the IC for the ciphertext is closer to 0.038 than to 0.065 , it's probably not a monoalphabetic cipher.
- ▶ **Key observation:** for a keyword of length r , a Vigenère cipher is really r monoalphabetic additive ciphers used in rotation!
- ▶ Therefore, every r^{th} letter is from the same cipher \Rightarrow
 - the IC for pairs of letters from the same cipher will be ≈ 0.065 .
 - the IC for pairs not from the same cipher will be ≈ 0.038 .
- ▶ So each cipher is used on $\approx n/r$ letters.

CALCULATING THE LENGTH OF THE KEY

Number of ways to choose two chars. from a **particular** cipher:

$$\binom{\frac{n}{r}}{2} = \frac{\frac{n}{r}!}{2! \left(\frac{n}{r} - 2\right)!} = \frac{\frac{n}{r} \left(\frac{n}{r} - 1\right)}{2}$$

But there are r ciphers, so the overall number of ways to choose 2 letters from the same cipher:

$$r \cdot \frac{\frac{n}{r} \left(\frac{n}{r} - 1\right)}{2} = \frac{n \cdot \left(\frac{n}{r} - 1\right)}{2}$$

And out of that, we'd expect to get

$$0.065 \cdot \frac{n \cdot \left(\frac{n}{r} - 1\right)}{2}$$

Pairs of letters.

(Why?)

CALCULATING THE LENGTH OF THE KEY, II

How about the number of ways to choose letters **not** from a **particular** cipher?

(# of ways to choose 1 letter from the cipher) \times
(# of ways to choose 1 letter not from the cipher) $\div 2 =$

$$\frac{\frac{n}{r}(n - \frac{n}{r})}{2}$$

Why divide by 2?

As before, there are r ciphers, so the overall number of ways to choose 2 letters not from the same cipher is:

$$r \cdot \frac{\frac{n}{r}(n - \frac{n}{r})}{2} = \frac{n \cdot (n - \frac{n}{r})}{2} \quad \text{and} \quad 0.038 \cdot \frac{n \cdot (n - \frac{n}{r})}{2}$$

is the expected number of pairs from this choosing.

CALCULATING THE LENGTH OF THE KEY, III

Given all that, the total number of expected pairs, EN , is thus:

$$EN = 0.065 \cdot \frac{n \cdot \left(\frac{n}{r} - 1\right)}{2} + 0.038 \cdot \frac{n \cdot \left(n - \frac{n}{r}\right)}{2} =$$
$$0.065 \cdot \frac{n \cdot (n - r)}{2r} + 0.038 \cdot \frac{n^2 \cdot (r - 1)}{2r}.$$

So the probability of both pairs being the same is

$$IC \approx \frac{EN}{\binom{n}{2}} = \frac{2EN}{n \cdot (n - 1)} =$$
$$\frac{1}{r \cdot (n - 1)} \cdot (0.027n + r \cdot (0.038n - 0.065)).$$

PUTTING IT ALL TOGETHER

Solving for r yields

$$r \approx \frac{0.027n}{IC \cdot (n - 1) - 0.038n + 0.065}$$

But r **is the key length!** So, for a given ciphertext:

1. Calculate the IC for the ciphertext.
2. If it's nearer to 0.038 than 0.065, try to solve for Vigenère
3. Use the Friedman test to calculate r .
4. Use the Kasiski test to produce the g.c.d. of the expected key length.
5. Try keys whose length is the multiple of the Kasiski test result nearest the result of the Friedman test.

THE ONE TIME PAD

- ▶ This is the only provably secure cipher.
- ▶ Requires a key as long as the cleartext.
- ▶ Keys may **never** be reused.
- ▶ Keys must be **randomly** generated (but finding true randomness is **hard**).
- ▶ That means that **every** cleartext is as likely as any other.
- ▶ Use the mod 26 addition on a letter by letter basis:

	n	o	w	i	s	t	h	e	t	i	m	e	f	o	r	a	l	l	...
+	F	F	I	W	M	L	J	Z	A	J	N	O	R	V	E	K	P	B	...
<hr/>																			
	I	T	E	E	E	E	Q	D	T	R	Z	S	W	J	V	K	A	M	...