Fall 2005
Chris Christensen
Cryptology notes

Codes and Nomenclators

In common usage, there is often no distinction made between codes and ciphers, but in cryptology there is an important distinction.

Recall that a **cipher** is a method of concealment that replaces each letter (or string of letters) with another letter or number or symbol (or string of letters or numbers or symbols). Ciphers consist of a method of encryption and a key.

A **code**, on the other hand, is a method of concealment that replaces words or phrases with codewords or codenumbers (or codegroups). Codes require codebooks – dictionary-like books that list all possible words or phrases that might be used in communication and their corresponding codeword or codenumber.

One of the most famous coded messages was sent on January 16, 1917. The German Foreign Minister Arthur Zimmerman sent the following ciphertext message by telegram from Berlin (by two routes to detect and correct garbling that might occur during transmission) to the German ambassador in Mexico City:

```
130 13042 13401 8501 115 3528 416 17241 6491 11310
18147 18222 21560 10247 11518 23677 13605 3494 14936
98092 5905 11311 10392 10371 0302 21290 5161 39695
23517 17504 11269 18276 18101 0317 0228 17694 4473
22284 22200 19452 21589 67893 5569 13918 8958 12137
1333 4725 4458 5905 17166 15851 4458 17149 14471 6706
13850 12224 6929 14991 7382 15857 67893 14218 36477
5870 17553 67893 5870 5454 16102 15217 22801 17138
21001 17388 7446 23638 18222 6719 14331 15021 23845
3156 23552 22096 21604 4797 9497 22464 20855 4377
25610 18140 22260 5905 13347 20420 39689 13732 20667
6929 5275 18507 52262 1340 22049 13339 11265 22295
10439 14814 4178 6992 8784 7632 7357 6926 52262 11267
21100 21272 9346 9559 22464 15874 18502 18500 15857
2188 5376 7381 98092 16127 13486 9350 9220 76036 14219
5144 2831 17920 11347 17142 11264 7667 7762 15099 9110
10482 97556 3569 3670
```

The message was encoded with the German Code 13040 which had about 25,000 plaintext elements and 75,000 code numbers. Here are some examples from the codebook for 13040:

Februar	13605
fest	13732
finanzielle	13850
folgender	13918
Frieden	17142
Friedenschluss	17149
führung	17166
Ganz geheim	17214
Gebeit	17388
geheim	4377
Gemeinsame	4458

The message announced that on the first of February, 1917, the German government would begin unrestricted submarine warfare. And, more …

> We intend to begin on the first of February unrestricted submarine warfare. We shall endeavor in spite of this to keep the United States of America neutral. In the event of this not succeeding, we make Mexico a proposal of alliance on the following basis: make war together, make peace together, generous financial support and an understanding on our part that Mexico is to reconquer the lost territory of Texas, New Mexico, and Arizona. The settlement in detail is left to you. You will inform the President of the above most secretly as soon as the outbreak of war with the United States is certain and add the suggestion that he should, on his own initiative, invite Japan to immediate adherence and at the same time mediate between Japan and ourselves. Please call the President’s attention to the fact that the ruthless employment of our submarines now offers the prospect of compelling England in a few months to make peace.

In 1917, although we were materially supporting the British and the French, the United States was officially neutral in World War I (which had begun in 1914), and President Wilson had pledged to keep us out of the European war. Germany believed that, to win the war, they must cut the Atlantic supply lines to Britain. They knew that their action was likely to draw the United States into the war. Germany would try to keep the United States neutral; however, if they were unsuccessful in keeping the United States out of the war, they intended to try to occupy us in conflicts away from Europe. To distract the United States, the German government proposed an alliance between Germany and Mexico. The alliance would include the understanding that if Mexico made war on the United States then Mexico would reconquer its lost territories in Texas, New Mexico, and Arizona. There was also a suggestion that Mexico encourage Japan to attack the United States. The ciphertext message was intercepted and broken by the cryptanalysts of Britain's Room 40 who arranged for it to be made available to American cryptanalysts. The proposed alliance and the planned Mexican attack on the American Southwest, pushed the United States to declare war. President Wilson cited the Zimmerman Telegram in his address to Congress asking that a state of war be declared.

Codes vs Ciphers

> There is no sharp dividing line between codes and ciphers; the latter shade into the former as they grow larger. But in modern practice the differences are usually quite marked. Sometimes the two are distinguished by saying that ciphers operate on plaintext units of regular length (all single letters or all groups of, say, three letters), whereas codes operate on plaintext groups of variable length (words, phrases, … etc.). A more penetrating and useful distinction is that code operates on linguistic entities, dividing its raw material into meaningful elements like words and symbols, whereas cipher does not … . [Kahn, David, *The Codebreakers: The comprehensive history of secret communication from ancient times to the internet*, Scribner, 1996.]

A cipher can quickly be changed, if it is believed that it has been broken. The method of enciphering need not be changed – just the key. Of course, there is the classic problem of key distribution, but it is easier to change a cipher than a code.

Codes are not easy to change because codes require dictionary-sized books of codewords or numbers. To change a code requires construction of new codebooks and replacement of all the old codebooks. Construction of a codebook requires determining all the words of phrases that are likely to be encoded.

Cryptanalysis of codes is usually a linguistic problem. Cryptanalysis of ciphers – looking for patterns – is usually a mathematical problem. We will deal (almost) exclusively with ciphers. In practice, the two methods of concealment can be combined. Enciphering an encoded message is called superenciphering.

Although when speaking precisely one should distinguish between *ciphers* and *codes*, the words are often used interchangeably. Similarly *encipher* and *encode* (or *encrypt*) are often used interchangeably, and *decipher* and *decode* are often used interchangeably. However, *decipher* (and *decode*) are usually used for authorized receivers in contrast with cryptanalysis.

The word *codemaking* is often used for *cryptography*, and *codebreaking* is often used for *cryptanalysis*.

Nomenclators

From the Fifteenth Century until the middle of the Nineteenth Century, nomenclators (*nomen*, name and *calator*, caller) were the primary form of cryptography. To protect communication, diplomats and popes had lists of codewords that could be substituted for names in communications. A typical nomenclator consisted of a simple substitution cipher to spell out words and codewords that could be used to substitute for names or common phrases.

To create a nomenclator, a cryptographer would try to determine the words that would be frequently used in communication and establish codewords to substitute for them. The cryptographer would also establish a simple substitution cipher to encrypt the remaining portions of the message.

To create a code, a cryptographer would try to determine the words that would be frequently used in communication and establish codewords to substitute for them. Codes typically require large, dictionary-like codebooks of substitutions.

A Code

We will construct a code for the 100 most frequent English words. The list is taken from *The Reading Teacher's Book of Lists, Third Edition*, by Fry, Kress, and Fountoukidis (the book also contains list of the second 100 most frequent words, the third 100 most frequent words, etc. The lists are available at www.duboislc.org/EducationWatch). The 100 most frequent English words listed by frequency are:

> the, of, and, a to, in, is, you, that, it, he, was, for, on, are, as, with, his, they, I, at, be, this, have, from, or, one, had, by, word, but, not, what, all, were, we, when, your, can, said, there, use, an, each, which, she, do, how, their, if, will, up, other, about, out, many, then, them, these, so, some, her, would, make, like, him, into, time, has, look, two, more, write, go, see, number, no, way, could, people, my, than, first, water, been, call, who, oil, its, now, find, long, down, day, did, get, come, made, may, part

We will use this list of words to construct a code. Like a nomenclator, we will also include a codeword for each letter of the alphabet to make it possible to spell words that are not in the list above.

We will establish codenumbers for the 100 words plus the 26 letters of the English alphabet. We will use a naive method to do that. A person encoding a message needs to be able to locate the word or letter for which the substitution will occur; so, will we construct an alphabetized list of the 100 words given above and their corresponding codenumbers.

```
a          001
about      002
all        003
an         004
and        005
are        006
as         007
at         008
b          009
be         010
been       011
but        012
by         013
c          014
call       015
can        016
come       017
```

could	018
d	019
day	020
did	021
do	022
down	023
e	024
each	025
f	026
find	027
first	028
for	029
from	030
g	031
get	032
go	033
h	034
had	035
has	036
have	037
he	038
her	039
him	040
his	041
how	042
i	043
I	044
if	045
in	046
into	047
is	048
it	049
its	050
j	051
k	052
l	053
like	054
long	055
look	056
m	057
made	058
make	059
many	060
may	061
more	062
my	063
n	064
no	065
not	066
now	067
number	068
o	069
of	070
oil	071
on	072
one	073
or	074

other	075
out	076
p	077
part	078
people	079
q	080
r	081
s	082
said	083
see	084
she	085
so	086
some	087
t	088
than	089
that	090
the	091
their	092
them	093
then	094
there	095
these	096
they	097
this	098
time	099
to	100
two	101
u	102
up	103
use	104
v	105
w	106
was	107
water	108
way	109
we	110
were	111
what	112
when	113
which	114
who	115
will	116
with	117
word	118
would	119
write	120
x	121
y	122
you	123
your	124
z	125

The message `the time has come to find the people who will go with you` would become `091 099 036 017 100 027 091 079 115 116 033 117 123`.

Decode the message `015 040 005 027 076 115 083 090 062 079 116 017`.

This is easy to do because the codenumbers arranged in increasing order correspond to the plaintext arranged in alphabetical order. Such a code is called a one-part code – one list may be used for both encoding and decoding.

Of course, this provides some help to a cryptanalyst. For example, if the cryptanalyst has determined that `the` is represented by `091` and `you` is represented by `123`, then `107` represents a word that is between `the` and `you` in alphabetical order.

A safer scheme would be to randomize the codenumbers. For example,

```
a           141
about       592
all         653
an          589
and         793
are         238
as          462
at          643
b           383
be          279
been        502
but         884
by          197
c           169
call        399
can         375
come        105
could       820
d           974
day         944
did         307
do          816
down        406
e           286
each        208
f           998
find        628
first       034
for         825
from        342
g           117
get         067
go          982
h           148
had         086
```

has	513
have	282
he	306
her	647
him	093
his	844
how	609
i	550
I	582
if	231
in	725
into	359
is	408
it	128
its	481
j	450
k	284
l	102
like	701
long	938
look	521
m	559
made	644
make	622
many	948
may	954
more	930
my	381
n	964
no	428
not	810
now	975
number	665
o	334
of	461
oil	756
on	482
one	337
or	867
other	831
out	652
p	712
part	019
people	091
q	456
r	856
s	692
said	346
see	861
she	045
so	432
some	664
t	821
than	339
that	607
the	260
their	249

them	273
then	724
there	870
these	066
they	063
this	155
time	881
to	488
two	152
u	092
up	096
use	254
v	715
w	364
was	367
water	892
way	590
we	360
were	011
what	330
when	305
which	204
who	213
will	841
with	469
word	519
would	415
write	116
x	094
y	572
you	703
your	657
z	595

It would be useful to anticipate the words that would be used in communication so that encryption of individual letters would not be necessary and, therefore, individual letter frequencies could not be attacked.

An attack on a code is a linguistic attack based upon frequencies of words. It would, therefore, be useful to anticipate common phrases and have codenumbers assigned to them. Like with ciphers frequencies of blocks of words are harder to attack.

The list above works well for encoding, but it does not work well for decoding because it is necessary to search for the codenumbers. To aid decoding, usually a list of the words arranged by increasing order of their codewords is constructed. Such a code is called a two-part code. There are two lists – one for encoding and one for decoding.

011	were

019	part
034	first
045	she
063	they
066	these
067	get
086	had
091	people
092	u
093	him
094	x
096	up
102	l
105	come
116	write
117	g
128	it
141	a
148	h
152	two
155	this
169	c
197	by
204	which
208	each
213	who
231	if
238	are
249	their
254	use
260	the
273	them
279	be
282	have
284	k
286	e
305	when
306	he
307	did
330	what
334	o
337	one
339	than
342	from
346	said
359	into
360	we
364	w
367	was
375	can
381	my
383	b
399	call
406	down
408	is
415	would
428	no

432	so
450	j
456	q
461	of
462	as
469	with
481	its
482	on
488	to
502	been
513	has
519	word
521	look
550	i
559	m
572	y
582	I
589	an
590	way
592	about
595	z
607	that
609	how
622	make
628	find
643	at
644	made
647	her
652	out
653	all
657	your
664	some
665	number
692	s
701	like
703	you
712	p
715	v
724	then
725	in
756	oil
793	and
810	not
816	do
820	could
821	t
825	for
831	other
841	will
844	his
856	r
861	see
867	or
870	there
881	time
884	but
892	water

```
930   more
938   long
944   day
948   many
954   may
964   n
974   d
975   now
982   go
998   f
```

This message has been encoded with the two-part code given above.

```
091 346 607 306 086 502 381 998 856 550 286 964 974
```

We will superencipher it by adding to it a random additive key:

```
Code        091 346 607 306 086 502 381 998 856 550 286 964 974
Key         275 778 960 917 363 717 872 146 844 090 122 495 343
Ciphertext  266 014 567 213 349 219 153 034 690 540 308 359 217
```

Coding Theory is Something Else

Another area of mathematics is coding theory. Coding theory is not concerned with the concealment of messages but rather with the correct transmission of messages. **Coding theory** develops techniques to detect and even correct errors occurring during transmission. The techniques involve building redundancy into the information being transmitted so that errors can be detected and corrected. Repetition is a standard technique. For example, if we wanted to transmit the digit 1 and we suspected that our communication channel might be noisy, we might send five ones 11111 because we are concerned that the single digit 1. If the message were received as 11010, it would be decoded by a majority of digits as 1 under the assumption that the majority of the digits were transmitted correctly. Universal Product Codes (UPC) which are found on items in the grocery store and International Standard Book Numbers (ISBN) are more complicated examples of such codes.

JN-25

JN-25 is the U.S. designation of a World War II **J**apanese **n**aval code. The code exhibits characteristics of both codes and coding theory. It was a two-part code that was superenciphered with additives. The basic code consisted of 33,333 five-digit code groups which replaced words (characteristic of codes), but the sum of the five digits was divisible by 3 (characteristic of coding theory) to detect garbling during transmission -- for example, 58743 and 78225 are JN-25 code groups. The codenumbers were then superenciphered with a string of additives.

Exercises

1. Encode the following message with the two-part code given above:

`Many more people now call him friend.`

2. Decode the following message that was encoded with the two-part code given above:

`820 703 282 273 399 381 665`

3. Decode the following message that was encoded with the one-part code given above:

`123 016 067 015 040 008 041 075 068`

4. Construct a new, two-part code that incorporates coding theory by taking the two-part code given above and inserting a new digit prior to the three digits that are given. Place an odd in the first spot if the sum of the given three digits is odd, and place an even digit in the first spot if the sum of the given digits is even. For example, for the word `first` that has code `034`, place an odd digit before this string; e.g., `5034`.

5. The following message was encoded with the one-part code given above and the superenciphered with the following additive string

`164 062 862 089 986 280 348 253`

The ciphertext message is

`277 078 985 036 900 204 326 293`

Decrypt the message.