Cryptography and murder – The Zodiac killer

Håvard Raddum
Selmersenteret
University of Bergen

Beginning of the story

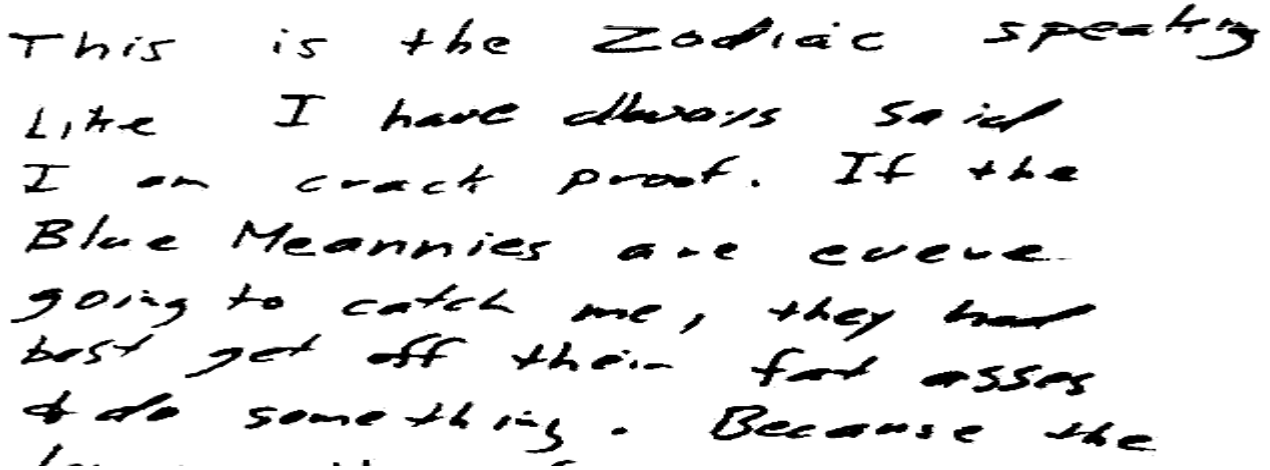
- Tor got an email
- Someone wanted help with a cipher
- Could the plaintext be in Norwegian?
- Decided to have a look at it

The Zodiac killer case

- Serial killer active in the San Fransisco area in the years 1967 – 1974 (at least)
- Linked to 7 murders, but claimed many more
- Never caught
- Police investigation closed in 2004
- Lots of theories and speculations around the case

This is the Zodiac speaking...

- Wrote many letters (at least 18) to the police and to newspapers in the SF area


This is the Zodiac speaking
Like I have always said
I am crack proof. If the
Blue Meannies are ever
going to catch me, they had
best get off their fat asses
& do something. Because the

www.zodiackiller.com

Ciphers

- Four letters contained ciphers
- Appear to be substitution ciphers, but with more ciphertext symbols than ordinary letters
- Two of them contain too little ciphertext to be solved

www.zodiackiller.com

This is the Zodiac speaking
By the way have you cracked
the last cipher I sent you ?
My name is ——

A E N ⊕ ⊗ K ⊗ M ⊗ ⊥ N A M

A l M o e F e n e r m a n ?

408 cipher

- First cipher the Zodiac sent
- Was divided in three parts and sent to three different newspapers
- All three parts published on the front pages
- Solved in a week by a high school teacher and and his wife

How did they do it?

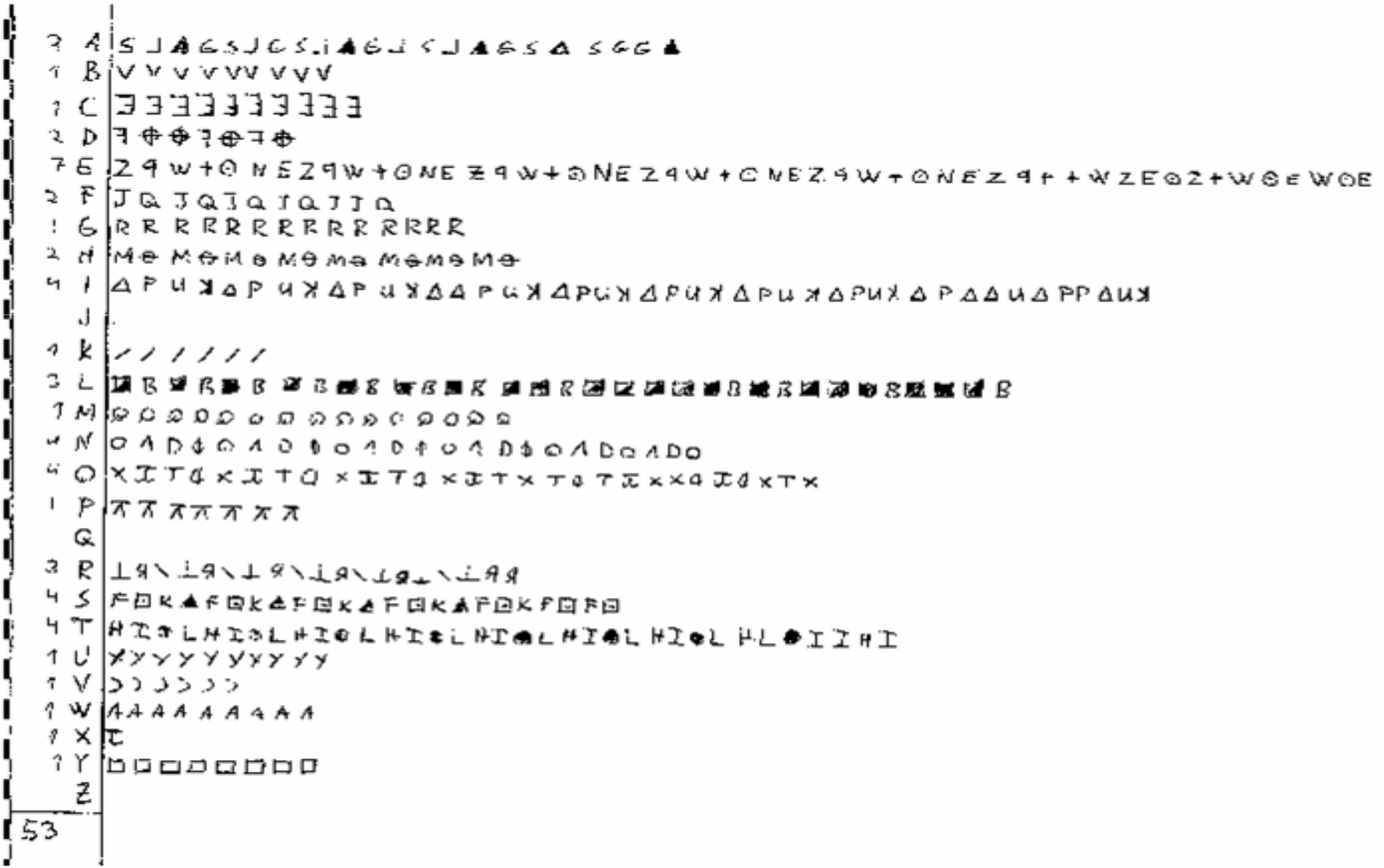
- Knew the killer had a big ego, assumed he would start with ’I’
- Assumed the word ’kill’ would appear several times
- This was 1969 – only pencil and paper!

www.zodiackiller.com

Observations

- One symbol = one letter
- More symbols than letters => one letter may be represented by several symbols
- Simple frequency analysis will not work
- 53 symbols used, each one appearing 408/53 = 7.7 times on the average

Order of symbols

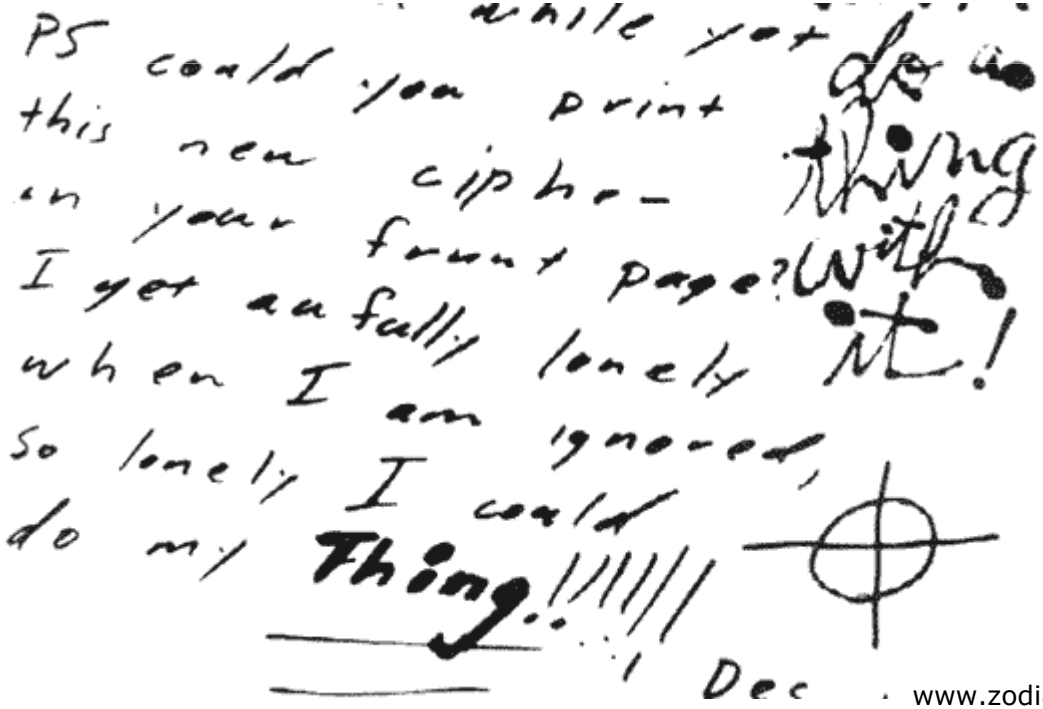


Cycle structure

- Symbols representing the same letter used in order
- Makes sure symbols are used equally often
- However, system messed up a bit

340 cipher

- Sent to the San Francisco Chronicle in Nov. 1969



www.zodiackiller.com

UNSOLVED

www.zodiackiller.com

Initial observations

- Appears to be same kind of cipher as the 408
- Each line contains 17 symbols, as in the 408
- Contains 63 symbols, each appearing
  340/63 = 5.4 times each on the average

Frequency analysis

```
 1: **** - 4
 2: *** - 3
 3: ******** - 8
 4: **** - 4
 5: ********** - 11
 6: ******* - 7
 7: ****** - 6
 8: ****** - 6
 9: *** - 3
10: ***** - 5
11: ********** - 10
12: ** - 2
13: ****** - 6
14: ***** - 5
15: ****** - 6
16: ********** - 10
17: ***** - 5
18: ***** - 5
19: ************************ - 24
20: ********** - 12
21: ******* - 7
```

7.7% - T,A,O,I,N,S?

```
22: ***** - 5
23: ********** - 10
24: ** - 2
25: **** - 4
26: ****** - 6
27: **** - 4
28: ****** - 6
29: ****** - 6
30: ****** - 6
31: ******* - 7
32: **** - 4
33: ***** - 5
34: ***** - 5
35: ** - 2
36: ********* - 9
37: ******* - 7
38: ***** - 5
39: **** - 4
40: ********* - 9
41: **** - 4
42: **** - 4
```

```
43: *** - 3
44: **** - 4
45: ** - 2
46: *** - 3
47: **** - 4
48: ** - 2
49: **** - 4
50: ** - 2
51: ********** - 11
52: ********** - 10
53: *** - 3
54: *** - 3
55: ***** - 5
56: ****** - 6
57: ** - 2
58: ** - 2
59: *** - 3
60: ** - 2
61: * - 1
62: *** - 3
63: ** - 2
```

Same method as the 408?

- Maybe symbols representing the same letter appear in cyclic order?

- Knowing which symbols represent the same letter makes cipher vulnerable to frequency analysis

- Assume cycle system used, identify different symbols representing the same letter

2-cycles

- Try all pairs of symbols and see which ones appear in alternating pattern (2-cycle)
- Number of 2-cycles in the 340 cipher is 90
- Most of them consist of symbols only appearing a few times
- Exceptions:
  - ( , ) – 6 times each
  - (,) – 7 times each

^ and ▪ same letter?

- 9 symbols appear 6 times each, what is the probability that two of them form a 2-cycle?
- Given two of the symbols, probability they appear in alternating pattern:

$$\frac{\#\text{ of alt. patterns}}{\#\text{ of pos. patterns}} = \frac{2}{\frac{12!}{(6!)(6!)}} = \frac{1}{66}$$

- Pr(at least one 2-cycle)= $1 - (65/66)^{\binom{9}{2}} \approx 0.423$
- Can not conclude anything

 and  same letter?

- 4 symbols appear 7 times each
- Given two of the symbols, probability they appear in alternating pattern:

$$\frac{2}{\frac{14!}{(7!)(7!)}} = \frac{1}{1716}$$

- Pr(at least one 2-cycle)= $1 - (1715/1716)^{\binom{4}{2}} \approx 0.0035$
- Good basis for guessing = in the 340 cipher

n-cycles

- May find all n-cycles ($n > 2$) by trying all $\binom{63}{n}$ possibilities
- Better way, use following result :

$(s_2,s_3,\ldots,s_n)$
$(s_1,s_3,\ldots,s_n)$
$\vdots \qquad \vdots$
$(s_1,s_2,\ldots,s_{n-1})$

all form $(n-1)$-cycles $\Leftrightarrow$ $(s_1,s_2,\ldots,s_n)$ forms an n-cycle

- Example: (1,5), (1,13) and (5,13) are 2-cycles

$\Updownarrow$

(1,5,13) is a 3-cycle

Overview of n-cycles

- Knowing all 2-cycles, may use result to recursively produce all n-cycles, $n>2$

n	# of n-cycles
2	90
3	62
4	14
5	2

- Using this to identify symbols representing same letter => too many letters (45)

Are we on the right track?

- Questions:
  - Maybe cipher shold be read columnwise?
  - Has the cycle system been used at all?

- If cycle system used, we should see more cycles than what to expect in a random symbol sequence

- Generated 10.000 random sequences from the set of symbols found in 340 cipher, counted the number of n-cycles in them

Result

Actual 340 cipher

n	# of n-cycles
2	90
3	62
4	14
5	2

10.000 random seq.

n	min	avg	max	# of times above actual 340
2	8	34.9	79	0
3	0	16.0	102	46
4	0	4.2	113	713
5	0	0.6	79	626

Transposed cipher:
(reading columnwise)

n	# of n-cycles
2	35
3	10
4	1
5	0

Findings (?)

- Strong bias in number of n-cycles, $n = 2,3$

  Evidence that cycle system has been used

- No bias when reading columnwise

  Cipher to be read linewise

- Analysis that allows for small deviations in cycle system needed