Cryptanalysis of Caesar Cipher

Malgorzata Kupiecka

1.Introduction

Caesar Cipher is the most important one in cryptology. Take any book on the subject, you are sure to find its description right in the first chapter.

As there are no unbreakable ciphers[Bro98], we have tried to attack it.

For the specification of the cipher, the reader is referred to [Sue110].

2. Differential Cryptanalysis of Caesar Cipher

2.1. Differential Cryptanalysis of Caesar Cipher in Z_{26}

In differential cryptanalysis we try to recover the secret key, exploiting the relations between differences in an input and the resultant differences at the output of the cipher.

We have begun with creating a standard tool used in differential attacks, i.e. evaluating the difference distribution table of the cipher [Hey99] for all possible input differences in Z_{26}.

To see where the entries in our table come from, let us analyse what happens for an input difference equal to 3.
An input difference equal to 3 can come from 26 plaintext pairs listed below:
(a,d), (b,e), (c,f), (d,g), (e,h), (f,i), (g,j), (h,k), (i,l), (j,m), (k,n), (l,o), (m,p), (n,q), (o,r), (p,s), (q,t), (r,u), (s,v), (t,w), (u,x), (v,y), (w,z), (x,a), (y,b), (z,c).

For key k=0, this gives us following ciphertext pairs:
(a,d), (b,e), (c,f), (d,g), (e,h), (f,i), (g,j), (h,k), (i,l), (j,m), (k,n), (l,o), (m,p), (n,q), (o,r), (p,s), (q,t), (r,u), (s,v), (t,w), (u,x), (v,y), (w,z), (x,a), (y,b), (z,c).
The output difference is in all 26 cases equal to 3.

For key k=1, we get:
(b,e), (c,f), (d,g), (e,h), (f,i), (g,j), (h,k), (i,l), (j,m), (k,n), (l,o), (m,p), (n,q), (o,r), (p,s), (q,t), (r,u), (s,v), (t,w), (u,x), (v,y), (w,z), (x,a), (y,b), (z,c), (a,d).
Again, the output difference is 26 times equal to 3.

For key k=2:
(c,f), (d,g), (e,h), (f,i), (g,j), (h,k), (i,l), (j,m), (k,n), (l,o), (m,p), (n,q), (o,r), (p,s), (q,t), (r,u), (s,v), (t,w), (u,x), (v,y), (w,z), (x,a), (y,b), (z,c), (a,d), (b,e).
The output difference is 3 for all 26 pairs.

The results for keys k=3,4,5,...,25 also give us the output difference of 3, 26 times for each key.

We can see, that for an input difference equal to 3, there are together 26*26=676 output differences of 3, and zero output differences of other values.

We repeat this process. Carefully checking all possible keys with all remaining possible input differences, we get a surprising result:

ΔOUT ΔIN	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**
0	676	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	676	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	676	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	676	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	676	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	676	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	676	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	676	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	676	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	676	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	676	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	676	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	676	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	676	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	676	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	676	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	676	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	676	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	676	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	676	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	676	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	676	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	676	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	676	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	676	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	676

Now, if we try to attack the cipher, given the input difference of 3, and a ciphertext pair (f,i) we know that the following plaintext pairs are possible:
(f,i), (g,j), (h,k), (i,l), (j,m), (k,n), (l,o), (m,p), (n,q), (o,r), (p,s), (q,t), (r,u), (s,v), (t,w), (u,x), (v,y), (w,z), (x,a), (y,b), (z,c), (a,d), (b,e), (c,f), (d,g), (e,h); all with the same probability.

The possible keys are respectively:
k=0, k=1, k=2, k=3, k=4, k=5, k=6, k=7, k=8, k=9, k=10, k=11, k=12, k=13, k=14, k=15, k=16, k=17, k=18, k=19, k=20, k=21, k=22, k=23, k=24 and k=25.

This is however the entire key space, which means that we haven't obtained any new restrictions on key values. Therefore we cannot get this way any additional information about the secret key.

Notice, that Caesar Cipher's resistance against differential cryptanalysis is in this case remarkably better than in the case of many modern ciphers, like AES.

2.2. Differential Cryptanalysis of Caesar Cipher – case of Z_2

The situation changes, if we use ASCII codes and operations in Z_2.
The difference between 'a' = 01100001 and 'b' = 01100010 is now not 1, but $11_b = 3$
We get another very interesting difference distribution table, shown in Appendix A.

For instance, given the input difference of $5 = 101_b$, and a pair of ciphertexts 'g' and 'j', we have only a limited number of possibilities for plaintext pairs:

p1= a, p2 = d, key=6,	p1= I, p2 = l, key=24,	p1= q, p2 = t, key=16,
p1= c, p2 = f, key=4,	p1= k, p2 = n, key=22,	p1= s, p2 = v, key=14,
p1= d, p2 = a, key=6,	p1= l, p2 = i, key=24,	p1= t, p2 = q, key=16,
p1= f, p2 = c, key=4,	p1= n, p2 = k, key=22,	p1=v, p2=s, key=14,

That means, only 23% of keys are possible. With another pair we can further decrease the number of possible keys.
In some cases, this attack can have a slightly better complexity than a brute-force search.

Despite this small weakness we claim, that Caesar Cipher was designed to be pretty secure against differential attacks. That leads us to an astonishing conclusion, that differential cryptanalysis was known already in Roman Empire.

3. Linear Cryptanalysis of Caesar Cipher

In this section, we use binary variables; input variables are denoted by x_i. Each letter corresponds to its 8-bit ASCII code, with MSB denoted by x_0.[#] According to the cipher's specification [Sue110], only letters 'a' - 'z' are used.

Ex. 'a' corresponds to $x_0=0, x_1=1, x_2=1, x_3=0, x_4=0, x_5=0, x_6=0, x_7=1$.

Output variables are denoted by y_i ($y_0,..,y_7$), and key variables - by k_i ($k_0,..,k_4$)

We have found some linear equations [Hey99] of x_i, y_i and k_i, holding with probability not equal to ½.

With the notation given above, the linear approximations are as follows:

a) $k_4+k_2+y_5+y_6+x_7 = 0$ holding with probability 0,454
b) $k_4+k_3+y_7+y_6+x_4+x_5 = 0$ holding with probability 0,476

The first relation allows an attack with about 240 pairs ciphertext-plaintext, with probability of success 92 %. The attack would allow to recover 2 key bits. There remains however an open problem – how to find so many different pairs.

Linear attacks do not seem promising for us. As in the previous case - no attack with complexity better than brute-force attack has been discovered so far.
Again, we have to state, that linear cryptanalysis must be over 2 000 years old and was practiced already in ancient Rome.

[#] why not

4. Algebraic attacks on Caesar Cipher

This approach is quite a new idea, so we only present here some statements based on our intuition[*], without going into further details.

We suspect, that the cipher could be somehow described by a system of multivariate very low degree equations. We also think, that both the number of such equations and their degree would allow an efficient attack, probably with the help of some good Gröbner bases algorithm, like F4[**], for solving such systems.

We hope, the complexity of such an attack could be significantly better than the brute force search of the entire key space. We are currently working on this subject.

5. Acknowledgements

I wish to thank Alice and Bob.

6. References

[Bro98] Brown D., Digital Fortress, 1998
[Hey99] Heys H.M. A Tutorial on Linear and Differential Cryptanalysis, 1999
[Sue110] Suetonius, De Vita Caesarum Divus Julius,110 CE

[*] Yes, I'm a woman.
[**] In other notation $GF(2^2)$

Appendix A

DIFFERENCE (XOR) DISTRIBUTION TABLE FOR CAESAR CIPHER

	XOR OUT																														
XOR IN	00000	00001	00010	00011	00100	00101	00110	00111	01000	01001	01010	01011	01100	01101	01110	01111	10000	10001	10010	10011	10100	10101	10110	10111	11000	11001	11010	11011	11100	11101	11110	11111
00000	676	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
00001	0	288	0	168	0	0	0	72	0	0	0	0	0	0	0	48	0	0	0	0	0	0	0	0	0	0	0	24	0	0	0	24
00010	0	0	288	0	0	0	144	0	0	0	0	0	0	0	96	0	0	0	0	0	0	0	0	0	48	0	0	0	0	0	48	0
00011	0	168	0	148	0	60	0	72	0	0	0	0	0	40	0	48	0	0	0	0	0	0	0	0	0	20	0	24	0	20	0	24
00100	0	0	0	0	242	0	0	0	0	0	0	0	154	0	0	0	0	0	0	0	0	0	22	0	0	0	44	0	88	0	22	0
00101	0	0	0	60	0	122	0	66	0	0	0	30	0	78	0	44	0	0	0	0	0	10	0	10	0	24	0	42	0	54	0	32
00110	0	0	144	0	0	0	122	0	0	0	60	0	0	0	78	0	0	0	0	0	30	0	0	0	24	0	40	0	30	0	44	0
00111	0	72	0	72	0	66	0	62	0	24	0	30	0	42	0	40	0	0	0	4	0	14	0	10	0	28	0	38	0	38	0	32
01000	0	0	0	0	0	0	0	0	200	0	0	0	0	0	0	0	0	0	80	0	0	0	40	0	160	0	0	0	0	0	40	0
01001	0	0	0	0	0	0	0	24	0	104	0	60	0	0	0	20	0	32	0	36	0	24	0	36	0	80	0	40	0	24	0	40
01010	0	0	0	0	0	0	60	0	0	0	104	0	0	0	52	0	80	0	0	0	36	0	32	0	0	0	80	0	36	0	40	0
01011	0	0	0	0	0	30	0	30	0	60	0	56	0	22	0	26	0	36	0	30	0	26	0	36	0	48	0	46	0	34	0	40
01100	0	0	0	0	154	0	0	0	0	0	0	0	116	0	6	0	0	0	36	0	42	0	32	0	0	0	28	0	80	0	26	0
01101	0	0	0	40	0	78	0	42	0	0	0	22	0	62	0	36	0	24	0	24	0	34	0	26	0	16	0	34	0	50	0	32
01110	0	0	96	0	0	0	78	0	0	0	52	0	6	0	60	0	40	0	12	0	32	0	22	0	16	0	40	0	26	0	40	0
01111	0	48	0	48	0	44	0	40	0	20	0	26	0	36	0	34	0	20	0	26	0	28	0	26	0	24	0	32	0	36	0	32
10000	0	0	0	0	0	0	0	0	0	0	80	0	0	0	40	0	200	0	0	0	0	0	80	0	0	0	80	0	0	0	40	0
10001	0	0	0	0	0	0	0	0	0	32	0	36	0	24	0	20	0	104	0	60	0	48	0	60	0	32	0	40	0	24	0	40
10010	0	0	0	0	0	0	0	0	80	0	0	0	36	0	12	0	0	0	0	0	84	0	52	0	64	0	0	0	48	0	40	0
10011	0	0	0	0	0	0	0	4	0	36	0	30	0	24	0	26	0	60	0	74	0	64	0	62	0	32	0	28	0	40	0	40
10100	0	0	0	0	0	0	30	0	0	0	36	0	42	0	32	0	0	0	0	0	116	0	42	0	0	0	24	0	74	0	40	0
10101	0	0	0	0	0	10	0	14	0	24	0	26	0	34	0	28	0	48	0	64	0	78	0	54	0	16	0	30	0	54	0	40
10110	0	0	0	0	22	0	0	0	40	0	32	0	32	0	22	0	80	0	0	0	42	0	60	0	32	0	36	0	32	0	38	0
10111	0	0	0	0	0	10	0	10	0	36	0	36	0	26	0	26	0	60	0	62	0	54	0	56	0	32	0	34	0	38	0	40
11000	0	0	48	0	0	0	24	0	160	0	0	0	0	0	16	0	0	0	0	0	0	0	32	0	136	0	0	0	0	0	40	0
11001	0	0	0	20	0	24	0	28	0	80	0	48	0	16	0	24	0	32	0	32	0	16	0	32	0	72	0	36	0	24	0	36
11010	0	0	0	0	44	0	40	0	0	0	80	0	28	0	40	0	80	0	0	0	24	0	36	0	0	0	72	0	40	0	36	0
11011	0	24	0	24	0	42	0	38	0	40	0	46	0	34	0	32	0	40	0	28	0	30	0	34	0	36	0	46	0	38	0	40
11100	0	0	0	0	88	0	30	0	0	0	36	0	80	0	26	0	0	0	0	0	74	0	32	0	0	0	40	0	82	0	36	0
11101	0	0	0	20	0	54	0	38	0	24	0	34	0	50	0	36	0	24	0	40	0	54	0	38	0	24	0	38	0	58	0	40
11110	0	0	48	0	22	0	44	0	40	0	40	0	26	0	40	0	40	0	0	0	40	0	38	0	40	0	36	0	36	0	42	0
11111	0	24	0	24	0	32	0	32	0	40	0	40	0	32	0	32	0	40	0	40	0	40	0	40	0	36	0	40	0	40	0	40