Eighth International Workshop on Optimal Codes and Related Topics
July 10-14, 2017, Sofia, Bulgaria *pp. 5-10*

Cryptanalysis on short messages encrypted with M-138 cipher machine

Tsonka Baicheva, Miroslav Dimitrov

tsonka@math.bas.bg,miroslavdimitrov@fmi.uni-sofia.bg

Institute of Mathematics and Informatics,
Bulgarian Academy of Sciences, Bulgaria

Abstract. The M-138 (or CSP-845) encrypting device belongs to the family of mechanical cipher machines used by US Armed Forces [1]. The difficulty of applying ciphertext-only attacks on messages encrypted with M-138 significantly increase when the length of the encrypted message decrease. The best result publicly available successfully recovers encrypted texts with at least 100 characters [2] using 25 strips out of 100. In this paper we use different approach to recover encrypted texts with length of 75 or more, using 25 strips out of 100. Furthermore, we develop cryptanalysis strategies which can be used throughout similar mechanical cipher machines like Bazeries Cylinder, M-94, etc.

1 Introduction

Colonel Parker Hitt invented the M-138 strip system in 1916. This M-138 was an aluminum board with twenty-five channels in which sliding strips of mixed alphabets were used. In 1939 the device was changed to include 30 channels with paper strips and became M-138A in the Army, and then CSP-845 in the Navy [1]. By summer of 1940 most of the United States defense units were using this cipher system.

The M-138 cipher machine components are strip cipher device, an alphabet set, and a key list. It consists of an aluminum or plastic base on which is formed a series of channels or grooves into which alphabet strips may be inserted and slid horizontally. A vertical rule called the *reading guide*, which can be slid to the right or left, is used for marking a specific column of letters to be copied. Sets of alphabet strips, that is, strips of paper bearing random-mixed repeated alphabets of 26 letters, making a sequence of 52 letters.

The encryption procedure consists of dividing the plain-text into blocks of length 25. The strips defined in the key are successively arranged in the cipher device. Each block of the message is specially arranged, so it can be read following the reading guide, by using only the first half of the used strips, which leaves a possible offset from 1 to 25. Then, the reading guide slides with as necessary positions as defined in the key and this way the final (encrypted) message is constructed.

M-138 system was in use between 1939 and 1960 and it seemed that it served its purpose extremely well. It was a very cheap tool, easy to carry and operate, and it provided high security for the time of its usage. Up to now, not much has been published about cryptanalysis of the M-138. Therefore, it is not known how much effort it takes to break an M-138 message and if this effort was realistic for a code-breaking unit in World War 2. In addition, it would be interesting to know whether the M-138 could have been improved without major effort. **For instance, by using different offsets for different strips.**

2 The M-138A challenge

To stimulate research on this topic Klaus Schmeh created the M-138 challenge introduced on the following web pages: http://scienceblogs.de/klausis-krypto-kolumne/m-138-challenge/ and https://www.mysterytwisterc3.org/. The original challenge contains three ciphertexts generated with a fictive M-138 model. The task is to break the encryption and find the cleartext. It is divided in 3 parts. Part 1 contains 75 letters, which means that each strip in the frame is used three times. The complexity rises considerably with part 2 and 3 because the ciphertexts become shorter (50 and 25 letters).

Since the publication of this challenge at the end of 2013, 10 solutions on Part 1 have been found. There is no publicly available information how the participants revealed the message, which left us an impression that they were using non-automated (manual) techniques.

In the best available commercial product for cryptanalysis CT2 [3], the attack of M-138 is also applied. As the author of the attack Nils Rehwald described in his thesis, his attempts to recover an encrypted with M-138 machine message with length 75 using his attacks, failed. As CT2 depends on his thesis - the cryptanalysis software was not able to recover chunks of the original message (the best candidate for the offset of the key was not the correct one, too).

3 Brute-force attack against the M-138

In order to determine whether the brute-force attack is computationally feasible or not, we need to calculate the key space of the M-138. When an ordered subset from 100 strips with cardinality 25 is used, the key space is: $\Omega = \binom{100}{25} * 25! * 25 \approx 2^{166}$. For reference, the key space of $3DES$, using three different keys, is 2^{168}. As example, if we are able to process $300,000$ keys/s in attempt of revealing an encrypted with M-138 message, we will need approximately $2^{122.89}$ years to iterate through all possible elements of the key set, which makes a brute force strategy inapplicable.

4 Our approach to solve the Part 1 of the challenge

The aim of our work is not only to solve the particular challenge but to develop more general tool for cryptanalysis of M-138. We will illustrate the approach we have used with the following example.

Plain text: "Failures are trickling springs that feed into rivers of success, if only you keep creating".

We will call *parsed string* s over some alphabet A the string obtained from the plain text by removing all non-letter characters and then upper-casing the result.

In our example, the parsed string after dividing in blocks with length 25 will be the following:

F	A	I	L	U	R	E	S	A	R	E	T	R	I	C	K	L	I	N	G	S	P	R	I	N
G	S	T	H	A	T	F	E	E	D	I	N	T	O	R	I	V	E	R	S	O	F	S	U	C
C	E	S	S	I	F	O	N	L	Y	Y	O	U	K	E	E	P	C	R	E	A	T	I	N	G

Definition 1 *By* ***triplet*** *we define any three consecutive symbols positioned on the* ***reading guide*** *of a M-138 cipher machine. We define them as* $t(i, j)$*, where* i *defines the starting position of the triplet, while* j *defines the block index we got the triplet from. For consistency, we should include the limitation* $i > -1 \vee i < j - 1$*.*

Definition 2 *By* ***rotor*** *we define all triplets in a text, generated by M-138*

cipher machine, with a common starting position. We define them as ROT(i), where i defines the starting position of the rotor. For consistency, we should include the limitation $i > -1 \vee i < j - 1$.

In the given example, we have: t(0,0) = FAI; t(1,0) = AIL; t(0,1) = GST;
t(4,2) = IFO and ROT(3)=

L	U	R
H	A	T
S	S	I

A random generated key is used for the encryption:

42	19	26	28	02	17	49	38	87	08	94	64	92	88	37	63	39	35	30	31	05	27	34	78	60

The result when applying the offset 6 is:

S	I	K	J	I	I	C	A	O	O	F	D	B	M	Q	X	C	Z	S	W	Q	O	L	N	O
M	G	Z	E	P	W	P	V	J	H	Q	L	D	N	P	Z	E	H	C	H	B	Y	A	G	U
L	P	B	W	B	A	E	C	R	V	N	M	Z	G	V	J	D	W	C	F	P	H	J	H	N

The final encrypted text is:

SIKJIICAOOFDBMQXCZSWQOLNOMGZEPWPVJHQLD
NPZEHCHBYAGULPBWBAECRVNMZGVJDWCFPHJHN

Analyzing various sources of parsed English texts, we estimated the expected logarithmic probability of chosen rotors on positions $0, 3, 6, 9, 12, 15, 18, 21$ over specially chosen unencrypted texts (sentences or paragraphs), but divided by blocks of length 25. Few things should be considered:

- The logarithmic probability of a rotor is the sum of all the logarithmic probability of triplets it posses.
- The unencrypted texts were chosen in such a way, to guarantee at least 1 rotor with exactly 3 triplets.
- When organizing the texts by rotors, all the rotors with few or more than 3 triplets were discarded.
- Following the previous consideration, we discarded sentences (or paragraphs) with parsed length less than 53 or more than 72.

4.1 Triplets Cut Attack

When we split an encrypted with M-138 cipher machine English text by rotors, and using the generated logarithmic probabilities of the rotors, we expect a logarithmic probabilities greater than -13.0. Moreover, the rotor consist of 9 letters, but its construction depends only on 3 strips. For each rotor and a fixed offset value, we can try all possible strip configurations, cut the undesired values and sort the results. Our experiments revealed, that iterating through all possible combinations of strips 970200, for a given fixed offset value, roughly 97% of the rotors have a logarithmic probability less than -13.0.

4.2 Deterministic Wave Search

By the Triplets Cut Attack we have considerably shrank the possible space of meaningful English texts. Then, we have implemented a search strategy called *Deterministic Wave Search - DWS*. By deterministic we mean that given two different instances of the same algorithm having equal starting states, will assure yielding of equal final results. Furthermore, an instance of the algorithm will always lead to an ending state. The strategy is the following:

1. The starting state is generated by a fixed position of $ROT(0)$, and random positions of remaining rotors. For simplicity, we will declare the final strip as the last rotor. $ROT(0)$ is the current rotor.

2. We iterate through all possible indexes of the right-adjacent rotor to the current rotor. The optimal logarithmic probability defines our next current rotor. If the index of the current rotor is i, the index of the next current rotor is $(i+3) mod 25$ - in the general case. If the current rotor is the eight rotor, the next current rotor will be the last rotor (the strip). If the current rotor is the last rotor (the strip), the next current rotor will be $ROT(0)$.

3. We observe the searching results. If we make 8 consecutive hops (a cycle), without improving the optimal logarithmic probability, we announce the final result as a local optimum.

4. We iterate through all possible offsets, but not all possible indexes of $ROT(0)$. To further optimize the speed of the algorithm, we limit the fixed starting position of $ROT(0)$ to the top 150 optimal indexes of it. In case the desired configuration of plaintext' $ROT(0)$ is not among the top 150 results, almost always exists some rotor J in top 150 indexes, for

which the common characters between the real $ROT(0)$ and itself is no less than 6, which doesn't affect the attack.

5 Applying the attack on the M-138 challenge, Part 1

We used the given in [4] strips and an encrypted message with length 75:

PTIJJHDJPKYTMTKUVEPDHYKLHDEYMGLIJLNWKX
VGZILQNCJRHW JNBJFUAQHNBJGXWZBESXNXPZH

We initiated a total of $150 * 25 = 7500$ instances of the *DWS* on 8 rotors + 1 strip. On iteration 471 (with current offset 4) we successfully recovered the offset value, the indexes of 4 rotors (including the strip) and last column of the adjacent to the left rotor. This yielded total of 33 characters of the plaintext. Fixing those rotor's indexes and repeating the *DWS* with non-fixed index of the first rotor, revealed the entire plaintext.

After the successful attack, the key can be recovered by *Known-Plaintext Attack* [2].

Finally, we will note that the M-138 is almost unbreakable when the message' length is about 25 characters long (Part 3 of the challenge), because there are hundreds of meaningful candidates for the decrypted text.

References

[1] War Department (1945) Special Instructions For Using The Strip Cipher Device (short title: FEDB-1), *NSA Historical Collections 190/37/7/1*, NR 2288 CBLL35 12804A 19450205 , 1945 (Box 798, F: 2288, pp 12).

[2] N. Rehwald, Implementation and Cryptanalysis of the M-138, *CrypTool 2.0*, Universitet Kassel, 2015.

[3] https://www.cryptool.org/trac/CrypTool2/browser/trunk/CrypPluginsExperimental/M138Analyzer/M138Analyzer.cs?rev=6462

[4] http://scienceblogs.de/klausis-krypto-kolumne/m-138-challenge/.