Cryptanalysis of Arabic Poetry, ibn Tabataba Treatise
Bushra Mohamed Elamin Elnaim1, Hayder Abood S. Wsmi Al-Lami2

1Department of Computer Science & Information, Sattam bin Abdulaziz University, Al Sulail, Kingdom of Saudi Arabia
2Department of Computer & Information, Sattam bin Abdulaziz University, Al Sulail, Kingdom of Saudi Arabia

*Corresponding author
Bushra Mohamed Elamin Elnaim
Email: bushra_moh@hotmail.com

Abstract: Cryptography has been practiced to conceal messages science antiquity by different civilizations, including the ancient Egyptian, Chinese, Indian, Mesopotamian, Greek and Roman. But in none of them was there any cryptanalysis, only cryptography existed. Cryptology, the science of both making ciphers (cryptography) and breaking them (cryptanalysis), was born among the Arabs shortly after the rise of the Arab-Islamic empire. Many Arab scholars wrote on, and excelled in practicing, both branches of cryptology. The aim of this paper is to discuss not-widely-known ibn Tabataba contributions to cryptology in poet and his effort in this field.

Keywords: ibn Tabataba, Cryptanalysis of Poetry, Arab cryptology, ibn Tabataba’s Treatise on Cryptanalysis, Poetry Encipherment.

INTRODUCTION

It was stated earlier that the Arabs were the first to treat cryptography and cryptanalysis scientifically. They have written on the science of cryptology and developed it, thus becoming its founders in their own right. Kahn said in [1] that cryptology was born among the Arabs. This was confirmed by the 1980 discovery of some Arabic cryptologic treatises which were found in Istanbul’s Suleymanye library [2], in addition to the works of other scholars who wrote about cryptography and cryptanalysis from the Arab Middle East, such as Al Kindi (801–873), Ibn Dunainir (1187–1229), Ibn Adlan (1187–1268), Ibn ad-Duraihim (1312–1361), and Al-Qalqashandi (1355–1418). All these works were written by learned men from the Abbasid, Ayyubid, Mamluk, and Ottoman Empires [3, 4].

Cryptology, the science of both making ciphers (cryptography) and breaking them (cryptanalysis), was born among the Arabs shortly after the rise of the Arab-Islamic empire. Many Arab scholars wrote on, and excelled in practicing, both branches of cryptology [3].

In this paper discussion to ibn Tabataba contributions in cryptology and his effort in this field from his Treatise has been explained.

ABOUT IBN TABTABA'S LIVE

Muhammad ibn ‘Ahmad ibn Muhammad ibn ‘Ahmad ibn ‘Ibrahim Tabataba al-Alawi Abu al-Hasan was born in Isfahan and died there in AH 322/AD 934. Sources do not give detailed accounts of his life, but they are agreed that he was a leading literary authority, and a great poet and researcher. A renowned personality of widely-circulating poetry, he was well known for his intelligence, sagacity, serenity of mind, and seriousness of purpose [5].

ABOUT IBN TABATABA'S WORKS

Ibn Tabataba has left a number of works that tend primarily towards poetry, literature and the like. The following are the most important of his books:

- *Iyar as-si’r* (The Standard of Poetry): A book on literary criticism. Hailed by editors as significant, it has been printed and published several times, the most recent of which was in Riyadh 1985.
- *Tahdib at-tab* (Refinement of Taste): Includes an interesting poetic anthology, of his own selection, of the works of other poets.
TOOLS FOR CRYPTANALYSING PROSE AND POETRY

The author starts out by considering a few issues that are common to cryptanalyzing ciphers rendered in prose and poetry alike, before he expands on the cryptanalysis of poetry ciphers in particular. These general issues are:

a) The number of letters: Ibn Tabataba holds the Arabic alphabet to be made up of 28 letters, after the pattern of al-Mubarrid [an eminent Arabic-language figure, d. AH 286/ AD899], and counter to the view of the predominant majority of language scholars. (This view regards the letters as 29 in the aggregate, including the “hamza”.)

b) The space or word-spacer: The symbol used to denote a blank between every two words in a cipher, i.e. the end of a word and the beginning of the next. The difficulty, however, consists in the so called no-word-spacer encipherment, mentioned by ibn Adlan [7], and overlooked by ibn Tabataba.

c) Combination and noncombination of letters: This phenomenon has been tackled by most of those engaged in this art, with al-Kindi leading. An example of noncombinable letters can be referred to in [8].

d) Letter frequency of occurrence: According to ibn Tabataba, the high-frequency letters, in order of precedence, are: ل , ه ,ً ,ُ ,ٛ ,تو ,ع,ب.

He places the letters ( ) and ( ) towards the end, in contrast with the mainstream trend of relatively arranging these letters as they appear in the word [9].

CRYPTANALYSIS OF POETRY CIPHERS:

Here is the longest and most important, discusses the algorithms of cryptanalyzing poetry ciphers. The author surveys the poetic features that aid in cryptanalysis, and for that matter mentions fifteen issues, mainly related to poetry. These are:

1. Extensive knowledge of prosody.
2. Refined appreciation of poetry.
3. Letter count of a verse for identifying metrical variations.
4. Utilization of the phenomenon of ”tasri” in a verse, i.e. the agreement between the last metrical units in both hemistiches of a line of poetry. A line of poetry so characterized is called ”musarra” [10]. Here ibn Tabataba observes four possibilities in such an agreement:
   a) agreement effecting tasri’, as in:
   الب اصبا نجب من نجد
   ٍَِ را اىزٛ جصف٘ ٍٔ اٗقاجُٔ       طشاً ٗٝثيغ مو ٍاٝخحاسُٓ
   in which the number of letters in one hemistich equals that in the other.

   b) agreement not effecting tasri’, as in:
   من ذا الذي تصفوه له أقائاته طرا وبيع كل مايحازه
   in which the count of letters is also the same in either hemistich. This kind of verse is called ”muqaffa” [10].

   c) agreement effecting tasri”, as in the following verse:
   أسٍِِ اً اٗفٚ دٍَِْة ىٌ جنيٌِ          تحٍ٘اّةِ اىذساج فاىَحثيٌِ
   in which the number of letters is not equal between the two hemistiches, and which contains geminated letters [11].

   d) disagreement effecting tasri”, as in:
   أسٍِِ اً اٗفٚ دٍَِْة ىٌ جنيٌِ          تحٍ٘اّةِ اىذساج فاىَحثيٌِ
   أقِيٜ عيٜ اىيًَ٘ ٝاتْثَ ٍزسِ         ّٗاٍٜ ٗاُ ىٌ جشحٖٜ اىًْ٘ فاسٖشٛ
   in which the same kind is prompted by lengthening the letter ( ) of ( ) in delivery to sound like ( ) , and thus correspond to [11].
(5) Looking out for letters that go together, such as (ا) and (ه). For example, a message ciphered using bird names, in which the letter (ا) is represented by "sparrow" and the (ه) by "crow", would have the twosome "sparrow crow" recurring more often than any else [11].

(6) Seeking three-letter and four-letter words (trigrams and tetra grams) with the letters (ا) and (ه), for the identification of these letters in a word would probably lead to the identification of the whole word [11].

(7) Pursuing two-letter words (bigrams), such as the particles: مثً او من, etc.; and imperative verbs, e.g. سلِّ، دُعِّ، خذ، etc. Note here that al-Kindi has already indicated the importance of the occurrence frequency of bigrams or the so-called contact count [11].

(8) Tracing long words along the lines of long poetic meters such as اسحفعاه, وفاعلات, and the like [11].

(9) Making use of the beginning of the second hemistich, so far made out, in determining the conjunctions (و) and (ف) if the context suggests words joined by means of conjunctions [11].

(10) Approximation of vocalized and neutral beats, based on the letters already identified, thus establishing a criterion for assessing the metrical measure of the cipher verse. The criterion is supposed to be equal to the letter count; otherwise the analogy should be changed over until the criterion fits the cipher exactly and evenly. By the criterion ibn Tabataba probably means scansion, i.e. the practice of marking off lines of poetry into metrical units known as feet, expressing the foot segments of a verse in terms of their metrically corresponding vocalized and neutral beats (e.g. فحوان // /o/o) [12].

(11) Applying the previous criterion in revealing intractable letters, through matching it up against the letters of the alphabet until it accords closely with the intended meter, and the word is written accordingly. ibn Tabataba warns the cryptanalyst here not to focus all his/her attention on certain letters at the expense of other letters, because “that is likely to protract your toil and disrupt your execution,” he says, maintaining that “cracking one letter often serves as a key by which to pierce through other letters that are still unresolved.” [13].

(12) Making full use of the Arabic syntax, which calls for certain succession that is sine qua non for correct wording in Arabic. For example:

- A relative pronoun (e.g. الذي) requires an antecedent.
- Letters specific to verbs are not followed by nouns.
- Letters specific to nouns are not followed by verbs.
- Adverbs of time and place require the nouns by which they are governed, or to which they are annexed.
- Positions of nouns, verbs, and particles should be observed in context. ibn Tabataba notes that confusion of meaning and structure in poetic ciphers, as well as non-observance of what is easy and familiar, would in all probability lead to difficult cryptanalysis [12].

(13) Utilization of the special order of letters dictated by unusual poetic necessities. This is what is termed by Arabic language scholars "poetic license" — the acceptable deviation, on the part of the poet, from rule or conventional form of poetry to produce the desired effect, such as the use of trip tote where diptote is typically appropriate; e.g. أدعتُ ىيحشب اىحٜ اُعْٜ تٖا ق٘افٞاً ىٌ اعَٜ تاجحلاتٖا

where the word ق٘افٞاً is used instead of the syntactically correct form قفاقيا.

(14) Trying all over again in case one letter only remains obscure of the cipher verse.

(15) ibn Tabataba sums up three requisites for efficacy in poetry cryptanalysis, i.e.:

a. acquaintance with the general texture and constitution of the Arabic word (letter combinability), the frequently used words and those less or never used.

b. familiarity with the Arabic syntax and grammar.

c. adequate knowledge of versification and metrical patterns [14].
METHODS OF ENCIPHERMENT BY SIMPLE SUBSTITUTION

Although ibn Tabataba’s treatment of encipherment does not generally cut across a single type, i.e. simple substitution, it can be classified into three categories:

a) Substitution of certain generic names for letters: Of these genera he mentions birds, beasts, people, aromatic plants, tools and instruments, etc [14].

b) Encipherment by bead threading, mentioned later by ibn Dunaynir [15].

ibn Tabataba’s reference to this method here is an indication to its popularity in the fourth century of the Hegira (Ninth AD), i.e. two centuries prior to ibn Dunaynir [16].

c) Picturing different marks, through devising special signs and symbols substituted for letters, such as those used by ibn ad-Durayhim in his two practical examples [17]. To be noted is that ibn Tabataba never touches upon any of the many encipherment methods treated by his predecessor al-Kindî; he rather restricts himself to simple substitution in its simplest varieties, with the employment of word-spacers [18].

AN EXAMPLE OF POETRY ENCIPHERMENT

The example given by ibn Tabataba concerns itself with the encipherment of a well-known line of poetry by 'Imru' al-Qays:

قِفا ّثِلِ رمشٙ حثٞةٍ ٍْٗض

by simple substitution (category (a) above). Encipherment has been performed as follows:

<table>
<thead>
<tr>
<th>Letter</th>
<th>Bird Name</th>
<th>Letter</th>
<th>Bird Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>ق</td>
<td>peacock</td>
<td>ن</td>
<td>Buzzard</td>
</tr>
<tr>
<td>ف</td>
<td>pheasant</td>
<td>م</td>
<td>Crow</td>
</tr>
<tr>
<td>ا</td>
<td>Hawk</td>
<td>ل</td>
<td>Raven</td>
</tr>
<tr>
<td>ن</td>
<td>falcon</td>
<td>ك</td>
<td>Francolin</td>
</tr>
<tr>
<td>ب</td>
<td>harrier</td>
<td>ب</td>
<td>Grouse</td>
</tr>
<tr>
<td>ك</td>
<td>merlin</td>
<td>م</td>
<td>Bobwhite</td>
</tr>
<tr>
<td>م</td>
<td>Eagle</td>
<td>د</td>
<td>Ringdove</td>
</tr>
<tr>
<td>د</td>
<td>Saker</td>
<td>ر</td>
<td>Pigeon</td>
</tr>
<tr>
<td>ر</td>
<td>vulture</td>
<td>خ</td>
<td>Duck</td>
</tr>
</tbody>
</table>

It is worthy to remark that the arrangement of bird names has suffered confusion in the cipher text of the original manuscript, probably owing to a scribe’s omission. However, the arrangement has been set right here in light of the plaintext as follows: peacock pheasant hawk falcon harrier Merlin eagle falcon saker merlin vulture buzzard crow harrier buzzard harrier raven eagle falcon francolin grouse harrier bobwhite peacock ringdove hawk grouse raven buzzard harrier buzzard falcon grouse pigeon duck raven grouse pheasant crow raven eagle grouse, the number of names is identical to the number of letters of the verse (i.e. 44) [19].

MANAGEMENT OF POETRY ENCIPHERMENT

By this management is meant the placement of cipher words on an endless circle, from which all poetical meters relevant to that circle disengage themselves, provided that those words are capable of such disengagement. Also the first word of the cipher verse is not identified each word of it is a potential start.

ibn Tabataba illustrates his point with a verse whose composition and metrical units fit in quite harmoniously with the meters peculiar to the circle of the similar. This is the third of the metrical circles, so named because all its parts are similar seven-bit feet adding up to 21 bits in all. Three meters develop or emanate from this circle, i.e. hazag, razaz and ramal, as shown in the following figure [19];
Accordingly, the circle has been divided into 21 equal parts serially numbered. Against the numbers are placed the corresponding strokes (denoting vocalized letters -harakat-) or bubbles (denoting neutral letters -sukuns-). If we start from the number (1), al-hazag meter emanates with its feet ٍفاعٞيِ ٍفاعٞيِ ٍفاعٞيِ. If we start from the number (4), i.e. leaving out the first syllable ٍفا of al-hazag foot, ar-ragaz meter emerges with its feet ٍسحفعيِ ٍسحفعيِ ٍسحفعيِ. If, however, we start from the number (6), i.e. skipping the second syllable عٞـ of the foregoing hazag foot, ar-ramal meter disengages itself: فاعلاجِ فاعلاجِ فاعلاجِ.

Given that ibn Tabataha's illustrative example is:

(а Brachycatalectic ragaz ), it is interesting to distinguish the possible alternatives, represented by the following model [20].

Let us start with the second word ( كريم ), from which emerges the hazag meter. Notice how the letters of the hazag verse are written (inside the circle) against the corresponding scansion represented by strokes and bubbles (outside the circle). The verse is:

كريم ماجد بحَر جَواد سَابق بَدر

But the circle has room for no more than 3 feet, interrupting the verse at ( سا ). With the first syllable left out, the second alternative (ragaz) emanates, the letters of which are written next to the first starting from the second syllable. The verse is:

بَدر كِرم ماجَد بحَر جَواد سَابق
Likewise, skipping the second syllable prompts the emergence of the third alternative (ramal) whose letters are written next to the second starting from the third syllable. This verse is:

ٍَاجذُ تَحشُ جََ٘اد   ساتِق َتذس مشٌٝ

ibn Tabataba stresses the point that in this type of cipher, letter cryptanalysis should be performed before looking out for the meter. He maintains that meter may straighten by whatever word it starts something likely to cause ambiguity to the cryptologue through driving him/her to mistakenly believe that their cryptanalysis is correct as long as the meter is sound, while in fact it is not. The previous example is a case in point [21].

PHOTOCOPIES OF IBN TABATABA’S TREATISE

Fig-3: A photocopy of the title page of ibn Tabataba's Treatise
(Document No. 5300, as-Sulaym niyya Ottoman Archives, Istanbul, Turkey)[22]

Fig-4: A photocopy of the first page of ibn Tabataba's Treatise
(Document No. 5300, as-Sulaym niyya Ottoman Archives, Istanbul, Turkey)[23]
CONCLUSIONS

There were some discoveries about Arab Contributions in Cryptography, especially in poet. Here we have described ibn Tabataba's Treatise, but there remain a lot more to discover in our academic libraries, personal libraries, and also in all places where our Arabic cultural treasures are still imprisoned.

Ibn Tabataba's Treatise on Cryptanalysis can be divided into five chapters as follows:
1. Tools for cryptanalyzing prose and poetry.
2. Cryptanalysis of poetry ciphers.
3. Methods of encipherment by simple substitution.
4. An example of poetry encipherment.
5. Management of poetry encipherment.

Ibn Tabataba, the author of this treatise, is a poet of great note who practiced the art of encipherment in his poetry, and was reported to have left special poetic selections in cipher. The reader will immediately realize that this treatise is so straightforward and brief that it only just needs any further elaboration. It is sufficient therefore to state the most remarkable views covered by ibn Tabataba in each chapter, highlighting the relevant technical terms in the edited text.

REFERENCES

Available online at http://saspublisher.com/sjet/


