Indian Journal of History of Science, **24**(1) : 1–7 (1989)

SOME EARLY CODES AND CIPHERS

SUBHASH C. KAK

Department of Electrical and Computer Engineering
Louisiana State University
Baton Rouge, LA 70803, USA

(Received 8 September 1987 : after revision 7 September 1988)

Some early ciphers and codes used in India are reviewed. The significance to cryptology of the *Śiva sūtra* and the Āryabhaṭa and Kaṭapayādi ciphers is described.

I

The numerology of the *Ṛgveda*, and its profuse use of symbols that make no sense to the uninitated, can be viewed as examples of encryption of ideas. The later literature is also rich in references to cryptography. The *Arthaśāstra*, the *Lalitavistara*, and the *Kāmasūtra* refer to secret writing.

In a commentary on the *Kāmasūtra*, Yaśodhara describes two kinds of cryptographic schemes based on letter substitution.[1] In the method called *kauṭiliyam*, the letter substitutions are based on phonetic relations so that certain vowels become consonants and *vice versa*. In *mūladevīya* the cipher transformation is the reciprocal mapping

a	kh	gh	c	t	ñ	n	r	l	y
k	g	ṅ	ṭ	p	ṇ	m	ṣ	s	ś

with the remaining letters remaining unchanged. In the written form this cipher is called *gūḍhalekhya*. Of course, properly speaking, *gūḍhalekhana* means cryptography.

Amongst a variety of other cipher schemes known, one worthy of special mention is a finger language called *nirābhāṣa*, that can be used by deaf and dumb people. Here the phalanges stand for the consonants and the joints for the vowels. This method is sometimes used by traders in the marketplace with the hands under a piece of cloth.

The substitution ciphers of the *gūḍhalekhana* can be solved readily by means of statistical analysis. When the cipher words themselves are meaningful messages, which can be true only for short ciphertexts, the *gūḍhalekhana* method can be effective from the point of view of security.

Substitution ciphers have been used in other civilizations as well. But the ciphers described in the next paragraphs are unique in many respects.

II

In the *Daśagītika* chapter of *Āryabhaṭīya* a cipher is described that represents numbers by letter sequences or words. Āryabhaṭa later uses this cipher to represent numbers by such words that allow a versification of his mathematics and astronomy. Since letters may be readily mapped into numbers, his cipher could be used in more general applications.[2] It is conceivable that similar ciphers were used in military situations but we have no concrete evidence of that. As *Āryabhaṭīya* was an update of an earlier book of astronomy it cannot be established that Āryabhaṭa himself discovered the cipher.

The cipher is generally known to historians of science, but it does not appear to have been examined by cryptologists. Its appeal lies not only in its novelty but also in the fact that it is an ingenious technique that can yield cipher words that appear meaningful.

The reading of the cipher mapping in *Āryabhaṭīya* is as follows :

> Beginning with *ka* the *varga* letters (are to be used) in the *varga* places, and the *avarga* letters (are to be used) in the *avarga* places. *Ya* is equal to the sum of *na* and *ma*. The nine vowels each (are to be used) in two nines of *varga* and *avarga* places.

Since the cipher for the Sanskrit alphabet is well understood,[3] we present a version for the Roman alphabet. This detracts the flexibility of the cipher transformation somewhat, since the Roman alphabet is only half as large as the Sanskrit alphabet.

III

The Āryabhaṭa cipher (*AC*) divides up the Roman alphabet into 3 classes. The letters with corresponding numeral equivalence are shown below.

Class 1 Consonants

letter:	b	c	d	f	g	h	j	k	l	m
number:	1	2	3	4	5	6	7	8	9	10

Class 2 Consonants

letter:	n	p	q	r	s	t	v	w	x	y	z
number:	20	30	40	50	60	70	80	90	100	200	300

Class 3 Vowels

letter:	a	e	i	o	u
number: (exponents)	1	2	3	4	5

Letters from the first two classes ordinarily stand for the numbers written under each. However, when such a letter is followed by one from the Class Three it is multiplied by 10 to the power listed under it. In other words, the use of a letter from the Class Three shifts the digits generated by Class One or Class Two letters to positions of higher significance.

To consider the transformation between numbers and letters, consider the example of the number 89381. It may be mapped as *dekavilib*, because de is $3\times10^2=300$, ka is $8\times10^1=80$, vi is $80\times10^3=80{,}000$, b is 1 and li is $9\times10^3=9000$. This number may be alternatively mapped as any other permutation of de, ka, vi, b and li. One may break up 89381 in other ways to get different mappings. Thus breaking it up as 50000+30000+9000+300+81 could also be represented as *gizewezkab*, as well as other variants.

In Āryabhaṭa's cipher for the Sanskrit alphabet,[3] Class One has the 25 *varga* letters k to m numbered 1 to 25, Class Two has 8 *avarga* letters y to h numbered 30 to 100, and Class Three has 9 vowels, a, i, u, ṛ, ḷ, e, ai, o, au, representing multiplication by $10^0=1$ through 10^{16} in multiples of 10^2. Āryabhaṭa also requires the use of *varga* letters in odd places (*varga* places) counting from the right, and the use of *avarga* letters in even places (*avarga* places) counting from the right.

Āryabhaṭa also uses the rule that if two letters from Classes One and Two are strung together and a letter from Class Three follows, then both these letters are followed by the letter from Class Three. Symbolically, $l_1 l_2 l_3$ stands for $l_1 l_3 l_2 l_3$ therefore. This does not lead to confusion because Āryabhaṭa also insists that a letter from Classes One or Two should ordinarily be followed by a letter from Class Three. In other words, $l_1 l_2 l_3$ is just a shorthand notation for $l_1 l_3 l_2 l_3$. We shall not use this aspect of Āryabhaṭa's notation due to the constraints placed by the smaller size of the Roman alphabet.

The longest numbers represented by Āryabhaṭa in his treatise using his mapping run into 10 places. For example, cayagiyinusuchlr stands for 57,753,336. Āryabhaṭa's formulation is slightly more restrictive than our exposition of it for the Roman alphabet. On the other hand, the larger Sanskrit alphabet provides a much richer set of substitutions so that it is easier to obtain cipher words that in themselves are valid words of the language.

A variant :

The Āryabhaṭa cipher breaks up a number in any of the many additive components and constructs the cipher thereof. By defining other Classes of letters that imply multiplication and subtraction one can increase the flexibility even further. Of course, when using a letter code for multiplication one would have to keep the code letters for the multiplicands separate.

Example of AC on text :

By mapping A to 01, B to 02, as so on and space to 27, one may obtain a cipher for text. An example is shown below :

Message THE ARYABHATA CIPHER

T H E - A R Y A B H A T A - C I P H E R

Numerical equivalent 200805270118250102080120012703091608051 8

The encryption may be performed conveniently by grouping the numbers in sets of six.

200805	↔	*novag*
270118	↔	*yitixbak*
250102	↔	*cugoxc*
080120	↔	*vemca*
012703	↔	*bonejad*
091608	↔	*wixasak*
0518	↔	*rabak*

The cipherwords do often have different lengths. These cipherwords could be written in groups separated by spacings. Adding a Class of letters for subtraction makes it easy to devise cipherwords that have *fixed* length.

Number of cipher transformations :

Clearly there are 26 ! ways the 26 letters can be assigned to the numbers listed in the transformation. Therefore,

Number of keys = 26 !

Yet this figure does not give a fair idea of the effort of the cryptanalyst, since the large number of ways a given number can be broken up into additive factors. makes the effective cipher variation much greater.

Key size :

The key is the sequence of the letters in the transformation. It is 26 letters long.

IV

While we have mentioned the possibility of creating ciphertext that may be meaningful in itself, in practice this will work only for messages of very small length. For this reason this property can be exploited best when the objective is to hide numerical information that occurs in prose. For straight encryption of text one would not look for this characteristic in the ciphertext.

For a general setting to view the Āryabhaṭa cipher consider that the message and cipher alphabets are $(a_1, a_2, \ldots, a_n)$, and the code alphabet is $(b_1, \ldots, b_k)$ where $n > k$. The message block M is first converted into the code block R, which is then partitioned into the sum (with the choice of L left arbitrary) :

$$R = R_1' + R_2' + \ldots + R_L.$$

Each of R_i is encrypted by a one to many mapping $C_i = C(R_i)$ and a concatenation of the C_i's in any order defines the cipherblock.

In a practical implementation, one may represent the elements of code digits themselves to be the partition. Now each b is mapped into one of the several candidate a's, the choice being greater if $k \ll n$. The cipher block letters could be chosen so as to follow the language statistics as closely as possible.

The Āryabhaṭa encryption paradigm is likely to have uses in encryption of computer data and communications signals. Its main disadvantage is that the ciphertext is larger than the plaintext.

V

Even more versatile than the Āryabhaṭa cipher is the *Kaṭapayādi* system that is believed to have been devised by Vararuci (4th Century A.D.).[4] It is also described in one of the extant manuscripts of the *Laghu Bhāskarīya* of Bhāskara I (629 A.D.), and in *Grahacāranibandhana* of Haridatta (683 A.D.). It was known in several variants[5] and used to represent numbers as words. Again, as text could be converted into a number sequence one may thus talk about a *Kaṭapayādi* or *Vararuci* cipher.

In the *Kaṭapayādi* cipher the numerals are mapped into different letters. The conjoint vowels have no numerical significance, and in a conjoint consonant only the last one denotes a number. This provides a much greater flexibility in converting numbers into meaningful words than does the Āryabhaṭa cipher. The mapping table is shown below :

Numeral	*Letters*
1	k, ṭ, p, y
2	kh, ṭh, ph, r
3	g, ḍ, b, l
4	gh, ḍh, bh, v
5	ṅ, ṇ, m, ś
6	c, t, ṣ
7	ch, th, s
8	j, d,
9	jh, dh
0	ñ, n, and vowels by themselves

The words are written starting from the right. Thus the number 644 may be coded in many ways to give words such as *bhavati* (bha : 4, va : 4, ti : 6), *vibhata*, and so on. By choosing other numeral/letters mappings other *Kaṭapayādi* type ciphers are obtained.

VI

Amongst early codes the *Śiva sūtras*[6] of *Aṣṭādhyāyī* are most remarkable. It has been argued that these *sūtras* were anterior to Pāṇini, but in any case their structure is closely linked to Pāṇini's grammar. The *sūtras* are listed in Figure 1.

Number	Speech-sound	Anubandha(end-marker)
1	a i u	ṇ
2	ṛ ḷ	k
3	e o	ṅ
4	ai au	c
5	h y v r	ṭ
6	l	ṇ
7	ñ m ṅ ṇ n	m
8	jh bh	ñ
9	gh ḍh dh	ṣ
10	j b g ḍ d	ś
11	kh ph ch ṭh th c ṭ t	v
12	k p	y
13	ś ṣ s	r
14	h	l

Figure 1. The Śiva sūtra

The letters a,i,u are, on the use of the *Śiva sūtra* code, listed as a^{ṇ}. The entire set of letters is represented as a^{l}, and so on. The *Śiva sūtra* code is certainly a mnemonic device. The manner in which the letters are broken into the fourteen sets has important grammatical reasons. The *sūtrās* are 'contraction/summation' (*pratyahāra*) rules.

A *Śiva sūtra* type rule lends itself to use in conjunction with a cryptographic transformation for added security. The *anubandha* letters may be repeated to distinguish them from the list letters. A subset of a given list can thus be represented very succinctly and further coded by a substitution, Āryabhaṭa, or Vararuci cipher.

VII

The preceding sections show the great interest of Indians in codes and ciphers. It is not known whether they developed a systematic science of cryptography, however. It is hoped that this article would spur interest in the writing of a systematic account of the contributions of Indians to the subject of cryptology.

REFERENCES AND NOTES

[1] Kahn, D., *The Codebreakers*, New York, 1967 presents a brief outline of the history of cryptography in India. An authoritative account of this history is yet to be written, however.

[2] Kak, S.C., The Āryabhaṭa Cipher, *Cryptologia*, **12**, 113-117, 1988.

[3] Datta, B. and Singh, A.N., *History of Hindu Mathematics*, Bombay, 1962 ; Shukla, K.S. and Sarma, K.V., *Āryabhaṭīya of Āryabhaṭa*, New Delhi, 1976.

[4] Datta and Singh, "History", p.71.

[5] Sarma, K.V., *A History of the Kerala School of Hindu Astronomy*, Hoshiarpur, 1972.

[6] Vasu, S.C., *The Aṣṭādhyāyī of Pāṇini*, Delhi, 1981.