International Electronic Journal of Pure and Applied Mathematics

Volume 4 No. 3 2012, 159-165

ON SOME PROPERTIES AND LIMITATIONS IN THE ISBN-13 CODE

P. Waweru Kamaku[1 §], Bernard Kivunge[2], Cecilia Wangeci[3]

[1,3]Department of Pure and Applied Mathematics
Jomo Kenyatta University of Agriculture and Technology
P.O. Box 62000-00200, Nairobi, KENYA

[2]Department of Applied Mathematics
Kenya Polytechnic Univeristy College
P.O. Box 52428-00200, Nairobi, KENYA

Abstract: The International Standard Book Number system-13 (ISBN-13) is a thirteen digit code which uniquely identifies every book published internationally. The code was formed as an improvement to the ISBN-10 code which could detect and correct some errors but could generate fewer code words thus the great need for expansion. This paper highlights some properties, discusses major limitations and shows the error detection and correction capabilities in the code versus the effect the total number of code words that the code can generate.

AMS Subject Classification: 94B60

Key Words: code, dictionary, ISBN-13, error detection, error correction, parity

1. Introduction

The International Standard Book Number (ISBN-13) is a 13 digit code which uniquely identifies each book published internationally uniformly. It was generated to replace the ISBN-10 since the year 2007. In 1986, Raymond [8] wrote on ISBN-10 code and showed that it was designed to detect a single and a double-error which can be caused by the transposition of two digits and noted that ISBN-10 code cannot be used to correct an error unless one knows the digit in error. Doumen [1] described aims of cryptography in providing secure transmission of messages in the sense that two or more persons can communicate in a way that guarantees that confidentiality,

Received: December 9, 2011

§Correspondence author

data integrity and authentication are guaranteed. Leo and others [7] studied the Coding of the ISBN. They concluded that the minimum requirement for a useful code was that all single errors as well as all permutations of two symbols must be detectable and a minimum requirement for a useful code is that all single errors as well as all permutations of two symbols must be detectable.

The ISBN-13 code is generated such that the calculation of its check digit begins with the first 12 digits, thus excluding the check digit itself. Given a code word $X = x_1x_2x_3x_4x_5x_6x_7x_8x_9x_{10}x_{11}x_{12}x_{13}$ is a, then

$$x_{13} = 10 - (x_1 + 3x_2 + x_3 + 3x_4 + \cdots + x_{11} + 3x_{12}) \pmod{10}. \tag{1}$$

For example, the book "A first course in coding theory" by Raymond Hill has an ISBN-13: 9780198538035 [8]. Working out for the check digit

$$(9\times1)+(7\times3)+(8\times1)+(0\times3)+(1\times1)+(9\times3)+(8\times1)+(5\times3)+(3\times1)+(8\times3)$$
$$+(0\times1)+(3\times3) = 9+21+8+0+1+27+8+15+3+24+0+9 = 125 \equiv 5 \pmod{10}.$$

But $10 - 5 = 5 \equiv 5 \pmod{10}$ and hence the check digit is 5 as above.

Viklund [10] notes that this check system does not catch all errors in adjacent digit transposition. The ISBN-10 formula used the prime modulus 11 which avoided this blind spot.

If a, b and m are integers, then $a \equiv b \pmod{m}$ if and only if there is an integer k such that $a = b + km$. Let a and b are integers with $d = gcd(a, b)$. The diophantine equation $ax + by = c$ has no integer solutions if $d \nmid c$. If $d|c$, then there are infinitely many integral solutions [15].

2. Types of Errors in an ISBN Code

A Single error occurs as a result of incorrect typing of one digit in an ISBN codeword or as a result of a smudge. Double errors occur when two digits in a codeword are incorrect. A Transpose error occurs due to interchanging of digits in a codeword. Omission or insertion of a digit(s) in a code word result to an error. The parity check equation may not hold in all these types of errors. A Silent error occurs when the parity check equation holds despite the sent and received code words differing in some bit strings. Suppose C_1 and C_2 is are the sent and received code words respectively. To deect an error(s), the receiver computes if equation I is met; if not then error(s) exist in the received code word C_2. Contrary to ISBN-10, the ISBN-13 does not tell the exact bit string which is in error. It only identifies the existence of error in a code word. To identify the digit in error, the receiver may be forced to compare the sent and received code words and identify where they differ. The proposition below lays a foundation for this by simplifying the general condition that the code words must obey.

Proposition 1. *Any ISBN-13 codeword must satisfy the condition below*

$$\sum_{i=1}^{13} \begin{cases} n^{th} \ digit, & n \ odd, \\ 3(n^{th}) \ digit, & n \ even, \end{cases} \equiv 0 \pmod{10}, \tag{2}$$

or $\sum_{n=0}^{6}((2n+1)^{th} digit) + \sum_{n=0}^{6}(3(2n)^{th} digit) \equiv 0 \pmod{10}$.

Proof. From equation (1) above,

$$x_{13} + (x_1 + 3x_2 + x_3 + 3x_4 + \cdots + x_{11} + 3x_{12}) \pmod{10} = 10 \equiv 0 \pmod{10}.$$

Since x_{13} must be between 0 - 9 then clearly, $x_{13} \equiv x_{13} \pmod{10}$, thus

$$a(x_1 + 3x_2 + x_3 + 3x_4 + \cdots + x_{11} + 3x_{12} + x_{13}) \pmod{10} \equiv 0 \pmod{10},$$

$$(x_1+x_3+x_5+x_7+x_9+x_{11}+x_{13})+(3x_2+3x_4+3x_6+3x_8+3x_{10}+3x_{12}) \pmod{10} \equiv 0 \pmod{10}.$$

Therefore

$$\sum_{i=1}^{13} \begin{cases} n^{th} \text{ digit}, & n \text{ odd}, \\ 3(n^{th}) \text{ digit}, & n \text{ even}, \end{cases} \equiv 0 \pmod{10}.$$

Proposition 2. *In an ISBN-13 codeword, the check bit string does not permute.*

Proof. By contradiction; Suppose the check digit permutes. Let C_1 and C_2 be two ISBN-13 code words given by

$$C_1 = a_1 a_2 \dots \ a_{11} a_{12} a_{13}$$

and

$$C_2 = a_1 a_2 \dots a_{11} a_{12} a'_{13}.$$

From equation (1)

$$a_{13} = 10 - (a_1 + 3a_2 + a_3 + \dots \ + 3a_{10} + a_{11} + 3a_{12}) \pmod{10},$$

$$a'_{13} = 10 - (a_1 + 3a_2 + a_3 + \dots \ + 3a_{10} + a_{11} + 3a_{12}) \pmod{10}.$$

But since C_1 and C_2 are similar in the first 12 bit strings, then the right hand side of both equations are the same hence $a_{13} = a'_{13}$. Hence the check bit string does not permute.

Proposition 3. *In an ISBN-13 code, any transposition of adjacent bit strings on a sent code word that yields same sum modulo 10 cannot be detected nor corrected.*

Proof. For an error to be corrected, its existence must first be detected. We therefore show that the error cannot be detected and thus cannot to talk about being corrected. Let a_i and a_{i+1}, be adjacent bit strings in a code word. Suppose during transmission, the bit strings are transposed. From the operation defined in equation 1 for the two bit strings:

$$3a_i + a_{i+1} \equiv b \pmod{10} \text{ and } a_i + 3a_{i+1} \equiv b \pmod{10}$$

for $1 \leq i \leq 12$.

From basic modulo operations, adding up the two equations yields: $4a_i + 4a_{i+1} \equiv 2b \pmod{10}$. Therefore $10 \mid 4a_i + 4a_{i+1} - 2b$ and hence there exists an integer n such that $4a_i + 4a_{i+1} - 2b = 10n$, $n \in \mathbb{Z}$. Simplifying yields $2a_i + 2a_{i+1} - b = 5n$, $n \in \mathbb{Z}$. Any transposition of adjacent bit strings a_i and a_{i+1},where $1 \leq i \leq 12$ such that equation $2a_i + 2a_{i+1} - b = 5n$, $n \in \mathbb{Z}$ is obeyed will go undetected.

Proposition 4. *In an ISBN-13 code, any single transposition (two bit strings transposed) of bit strings on a sent code word of the same parity (even or odd) digit position cannot be detected nor corrected.*

Proof. As above, we show that the error cannot be detected. For this proof to be complete, it must be shown that it holds for both even and odd position. Without loss of generality, let a_i and a_j , where $i = 2n, 1 \leq n \leq 6$ and $j = 2m, 1 \leq m \leq 6$, be the two bit strings in the sent codeword which are transposed to yield to the received codeword. Upon computation of the check digit, each bit string at an even position is multiplied by 3. Clearly $3a_i + 3a_j = 3a_j + 3a_i, \forall\ i, j$ since addition of integers is commutative. Thus upon computation of athe check digit, both codewords yield the same result.

Similarly, without loss of generality, suppose a_l and a_h , where $i = 2n + 1, 1 \leq n \leq\ 6$ and $j = 2m + 1, 1 \leq\ m \leq\ 6$, and are the two bit strings in the sent codeword which are transposed to yield to the received codeword. Upon computation of the check digit, each bit string at an odd digit position is multiplied by 1. Clearly $a_l + a_h = a_h + a_l, \forall\ l, h$. Thus upon computation of the check digit, both codewords yield the same result. In both cases, the error is not noticed hence as discussed earlier, cannot be corrected.

Example. Consider the sent codeword 9780198538035 and 9087198538035 as the received codeword. As shown earlier, 9780198538035 is a valid codeword. Consequently, 9087198538035 will yield the same results since it is just a transposition of the second and fourth bit strings. Any error yielding the second codeword from the first sent code word will go undetected.

Proposition 5. *In an ISBN-13 code, any double or multiple silent errors on odd positions of a sent code word that yields same sum modulo 10 with the ones on the received code word cannot be detected nor corrected.*

Proof. Suppose two silent errors occur on two odd positions of the code word $C_1 = a_1a_2 \dots a_{12}a_{13}$ namely a_i and a_j where $1 \le i \le 11, 1 \le j \le 11, a_i \ne a_j$ to yield to a code word $C_2 = a_1a_2 \dots a_{12}a_{13}$ which differ with C_1 in only the two odd positions such that a_i is replaced by a'_i and a_j is replaced by a'_j. Without loss of generality suppose $i = 1$ and $j = 3$. This choice can generalize the other odd positions since upon calculation of the check digit; the bit stings at the odd positions are each multiplied by 1 thus the digit position does not matter provided it is in an odd position.

$$a_{13} = 10 - (a_1 + 3a_2 + a_3 + \cdots + 3a_{10} + a_{11} + 3a_{12}) \pmod{10},$$

$$a'_{13} = 10 - (a'_1 + 3a_2 + a'_3 + \cdots + 3a_{10} + a_{11} + 3a_{12}) \pmod{10}.$$

Suppose a_1 and a_3 yields same sum modulo 10. That is $a_1 + a_3 \equiv b \pmod{10}$ and $a'_1 + a'_3 \equiv b \pmod{10}, 1 \le b \le 9$.

Then we have that

$$(a_1 + a_3) - (a'_1 + a'_3) \equiv (b - b) \pmod{10} (a_1 + a_3) - (a'_1 + a'_3) \equiv 0 \pmod{10}$$

$$(a_1 + a_3) \equiv (a'_1 + a'_3) \pmod{10}.$$

This means that any such error on bit strings satisfying this equation will go unnoticed hence cannot be corrected.

Proposition 6. *In an ISBN-13 code, any double silent errors on even positions of a sent code word that yields same sum modulo 10 cannot be detected nor corrected.*

Proof. Suppose two silent errors occur on two even positions of the code word $C_1 = a_1a_2 \cdots a_{12}a_{13}$ namely a_i and a_j where $2 \le i \le 12, 2 \le j \le 12, a_i \ne a_j$ to yield to a code word $C_2 = a_1a_2 \cdots a_{12}a_{13}$ which differ with C_1 in only the two even positions such that a_i is replaced by a'_i and a_j is replaced by a'_j. Without loss of generality suppose $i = 2$ and $j = 4$. This choice can generalize for the other even positions since upon calculation of the check digit; the bit stings at the even positions are each multiplied by 3 thus the digit position does not matter provided it is in an even position $a_{13} = 10 - (a_1 + 3a_2 + a_3 + 3a_4 \cdots + a_{11} + 3a_{12}) \pmod{10}$ then

$$a'_{13} = 10 - (a_1 + 3a'_2 + a_3 + 3a'_4 \cdots + a_{11} + 3a_{12}) \pmod{10}.$$

Suppose $3a_2 + 3a_4 \equiv b \pmod{1}$ and $3a'_2 + 3a'_4 \equiv b \pmod{1}$, $1 \le b \le 9$. Then $3a_2 + 3a_4 - b = 10k, k \in \mathbb{Z}$ and $3a'_2 + 3a'_4 - b = 10k'$, $k' \in \mathbb{Z}$,

$$(3a_2 + 3a_4) - (3a'_2 + 3a'_4) = 10(k - k'), (k - k') \in \mathbb{Z}$$

$(3a_2 + 3a_4) \equiv (3a'_2 + 3a'_4) \pmod{10}$ and thus $(a_2 + a_4) \equiv (a'_2 + a'_4) \pmod{10}$.

Any interchange of a_i with a_j satisfying this equation will be an error which would go unnoticed hence cannot be corrected.

Proposition 7. *The ability to detect an error(s) in the ISBN-13 code does not affect (increase or reduce) its dictionary size.*

Proof. An arbitrary code word x which does not satisfy equation I is an invalid ISBN-13 codeword thus its existence does not increase or reduce the dictionary. If a transposition or silent error occurs, it yields a different codeword which is also a member of the code and similarly does not increase or reduce the dictionary.

3. Conclusion

The cataloguing in libraries may be faced a big problem if error occur due to smudge on their books which may lead to users borrowing the wrong books. A buyer interested in purchasing a book coded a, he/she may decide place the order by sending the ISBN code word to the seller. Suppose in the process due to errors the seller receives the codeword a1 which is also satisfies the equation I hence a valid code word but differ from a as shown above. If no reconfirmation is made, the reader will end up purchasing the wrong book. Due to this, reconfirmation is very vital which makes the process long and expensive.

A code should thus not only guarantee the uniqueness of a book for cataloguing but also ensure it avoids this blind spot. ISBN-13 is an improvement of the ISBN-10 but as seen, it faces some major limitations.

Acknowledgment

Special acknowledgement to the late Prof. Cecilia Wangeci for the great work she did in the development of this paper before resting in peace. Being the first woman to become a Professor in Pure Mathematics in Kenya has left us with a reason to learn more and a legacy to pride in.

References

[1] J.M. Doumen, *Some Applications of Coding Theory in Cryptography*, Eindhoven University Press, Germany (2003).

[2] Eric Weisstein, *ISBN Code*, http://mathworld.wolfram.com/ISBN.html (2010).

[3] V.D. Goppa, A new class of linear error correcting codes, *Problems of Information Transmission*, **6** (1970), 207-212.

[4] R.W. Hamming, Error detecting and error correcting codes, *Bellsyst Technology Journal*, **29** (1950), 147-160.

[5] Kenneth H. Rosen, *Elementary Number Theory and its Applications*, Addison-Wesley Publishing Company (1993).

[6] E. Leo, Detection and correction of multiple errors in general block codes, *Mathematical and Computer Modelling*, **30** (2005), 113-121.

[7] E. Leo, *The Coding of the ISBN*, http://en.scientificcommons.org/leo-egghe (2010).

[8] H. Raymond, *A First Course in Coding Theory*, Oxford University Press (1986).

[9] R. Ronald, E. Leo, On the detection of double errors in ISBN and ISSN-like codes, http://en.scientificcommons.org/leo-egghe (2010).

[10] A. Viklund, *ISBN Information Home*, http://isbn-information.com/index.html (2007).