Des. Codes Cryptogr.
DOI 10.1007/s10623-015-0072-8

A general check digit system based on finite groups

Yanling Chen[1] · **Markku Niemenmaa**[2] · **A. J. Han Vinck**[3]

Received: 17 November 2014 / Revised: 5 March 2015 / Accepted: 19 March 2015
© Springer Science+Business Media New York 2015

Abstract In this paper, we review a new method for the universal design of a check digit system over an abelian group of arbitrary order. Furthermore, we challenge the current standards by comparing this system with several well-known and widely used systems such as ISBN, MEID, ISAN and the system over alphanumeric characters. We show that this novel design outperforms all of them in terms of the error detection capability with a comparable computational complexity. In particular, besides the well-known five types of errors to be detected (i.e., single error and four double errors which are adjacent/jump transposition and adjacent/jump twin errors), we address the t-jump transpositions and t-jump twin errors which generalize the four types of double errors, and aim to design the check digit system with a detection radius as long as possible that depends on t and reflects the capability of detecting these two special kinds of double errors. The results of this paper are based on the results of the article by Chen et al. (On some properties of a check digit system, 2012).

Keywords Abelian group · Field · Check digit · Error detection

Mathematics Subject Classification 20K01 · 11D88 · 94B05

Communicated by G. Mullen.

✉ Yanling Chen
yanling.chen-q5g@rub.de

Markku Niemenmaa
markku.niemenmaa@oulu.fi

A. J. Han Vinck
han.vinck@uni-due.de

[1] Institute of Digital Communication Systems, Ruhr University Bochum, 44780 Bochum, Germany

[2] Department of Mathematical Sciences, University of Oulu, 90014 Oulu, Finland

[3] Institute of Digital Signal Processing, University of Duisburg-Essen, 47057 Duisburg, Germany

Published online: 02 April 2015

1 Introduction

According to the statistical investigations by Beckley [1] and Verhoeff [2], when transmitting a sequence of digits, the most common transmission errors made by human operators are the following:

(1) single error: $\cdots a \cdots \to \cdots b \cdots$;
(2) adjacent transposition: $\cdots ab \cdots \to \cdots ba \cdots$;
(3) jump transposition: $\cdots abc \cdots \to \cdots cba \cdots$;
(4) twin error: $\cdots aa \cdots \to \cdots bb \cdots$;
(5) jump twin error: $\cdots aca \cdots \to \cdots bcb \cdots$.

For their relative frequency of occurrence, one can refer to Table 1. Note that insertion and deletion errors are not included in the above list. The reason is that they can be detected easily under the assumption that all codewords transmitted are of equal length. For practical purposes, it is desirable to design systems which are able to detect all of these errors; or, if infeasible, at least those with higher occurrence frequencies in order to achieve a satisfying error detection capability. In the literature and in practice, it is usually done by appending a check digit a_{n+1} to a given information sequence $a_1 \ldots a_n$.

According to Table 1, it is clear that the error types (1) single errors and (2) adjacent transpositions are the most prevalent ones. So research attention was first brought to design systems over groups with anti-symmetric mappings which ensure these two kinds of errors to be detected. One can refer to [3] and the reference papers therein for a survey on anti-symmetric mappings in different groups. Although most of the systems in use are defined over alphabets endowed with a group structure, nevertheless, possibilities of constructing error detecting codes based on quasigroups were also intensively discussed. A comprehensive investigation on the check digit systems over quasigroups was conducted in [4,5], where necessary and sufficient conditions were established in order to detect each of the 5 error types. So far, the approaches have been taken are in general mathematically analytical.

There are as well constructive approaches, attempting to detect at least errors of types (1)–(2), or if possible, all the 5 most frequent error types. The very first significant work relating to this was by Verhoeff [2], who proposed a decimal code over the Dihedral group $\mathbf{D}_5$ together with an appropriate permutation capable of detecting all the errors of types (1)–(2). Damm in [6], provided a more concise construction of decimal code over a totally anti-symmetric quasigroup of order 10. Its error detection capability is comparable with Verhoeff's method

Table 1 Error types and their frequencies [3]

Error type	Description in symbol	Frequency in %	
		Verhoeff	Beckley
(1) Single error	$\cdots a \cdots \to \cdots b \cdots$	79.0	86
(2) Transposition	$\cdots ab \cdots \to \cdots ba \cdots$	10.2	8
(3) Jump transposition	$\cdots acb \cdots \to \cdots bca \cdots$	0.8	
(4) Twin error	$\cdots aa \cdots \to \cdots bb \cdots$	0.6	6
phonetic error ($a \geq 2$)	$\cdots a0 \cdots \to \cdots 1a \cdots$	0.5	
(5) Jump twin error	$\cdots aca \cdots \to \cdots bcb \cdots$	0.3	
other errors		8.6	

[2], but without any dedicatedly constructed permutations involved. Moreover, a general construction for the check digit system over a group of order $r = 2s$, where s is an odd prime, was by Gumm [7], which uses the operation over dihedral groups $\mathbf{D}_s$ with an appropriate permutation. Similarly to the Verhoeff's method, the system detects all the errors of types (1)–(2). Another instance was in [8] which showed that the Sylow 2-subgroups of nearly all Chevalley groups in even characteristic allow the definition of a check digit system that is capable of detecting all errors of types (1)–(5).

Aside of these, in a recent work [9], Niemenmaa proposed a check digit system for hexadecimal numbers, based on a suitable automorphism of the elementary abelian group of order 16. Its design is concise and elegant, with the capability of detecting all the 5 types of errors. It is worth mentioning that this check digit system is currently a part of MISB Standard 1204.1 for the video community (for the details, see [10]). Inspired by this simple but effective design, the authors of [11] proposed check digit systems over an abelian group of order p^k, for any prime p and $k \geq 1$. Their systems could achieve the same error detection capability and beyond, over groups of prime power order which is a generalization of the check digit system over hexadecimal numbers. In [12], the authors further extended the results of [11] by proposing check digit system over a group of an arbitrary order. The idea is to employ several parallel subsystems as introduced in [11] which are based on the use of elementary abelian p-groups of order p^k.

In real life applications, there are many well-known examples for use of check digit systems, such as

- *EAN code* originally European Article Number (EAN), now renamed International Article Number with the abbreviation EAN being retained. Note that EAN was developed as a superset of the Universal Product Code (UPC) that is widely used for tracking trade items in stores;
- *ISBN code* the International Standard Book Number (ISBN) that uniquely identifies each specific edition of a book or book-like product. A 10-digit ISBN (i.e., ISBN-10) was used before 2007, whilst since Jan 1 in 2007 it changed to a 13-digit ISBN (i.e., ISBN-13), a format that is compatible with the 13-digit EAN code;
- *IBAN code* the International Bank Account Number (IBAN) which is an internationally agreed means of identifying bank accounts across national borders;
- *ISAN code* the International Standard Audiovisual Number (ISAN) that is a unique identifier for audiovisual works and related versions, similar to ISBN for books; and
- *MEID code* the International Mobile Equipment Identifier (MEID) that is a globally unique number identifying a physical piece of CDMA mobile station equipment. In practical terms, it can be seen as an International Mobile Station Equipment Identity (*IMEI*) but with hexadecimal digits.

In this paper, we first review the universal design proposed in [12]. Beyond the theoretical results which are addressed in more details here, we also provide a comprehensive comparison on error detection capability with several well-known and widely used systems. Interestingly, we demonstrate that this universal design outperforms the popular methods such as Luhn formula and Modulo method with a comparable computational complexity. In fact, the system is shown to be universal also in terms of the 'computational' cost, since the multiplications needed for the error detection can be bounded by a constant which is independent of the number of the information digits, unlike the other existing systems. Last but not least, we also give a brief report on the 'state of art' of the implementation of this universal system in standardization and real life applications.

The rest of the paper is organized as follows: First in Sect. 2, we introduce some preliminaries. In Sect. 3, we briefly review the recently introduced check digit system over a group of a prime power order. In Sect. 4, we present the general design of a check digit system over a group of an arbitrary order. In Sect. 5, we challenge current standards by comparing our system with several well-known and widely used systems such as ISBN, MEID, ISAN and the system over alphanumeric systems. Finally we conclude in Sect. 6.

2 Preliminaries

In [11], the authors looked into the following two categories of jump errors, which include and further extend the error types (2)–(5).

(6) t-jump transposition:

$$\cdots ab_1 \ldots b_t c \cdots \to \cdots cb_1 \ldots b_t a \cdots$$

(7) t-jump twin error:

$$\cdots ab_1 \ldots b_t a \cdots \to \cdots cb_1 \ldots b_t c \cdots$$

It is easy to see that error types (2) adjacent transposition and (3) jump transposition, can be regarded as t-jump transpositions for $t = 0$ and $t = 1$, respectively; (4) twin error and (5) jump twin error, can be regarded as t-jump twin errors for $t = 0$ and $t = 1$, respectively.

These two kinds of errors were first considered in [13,14] and treated as transposition and twin errors on places $(i, i + t + 1)$, where $1 \le i \le n$ and $i + t + 1 \le n$. They are of interest, not only because they simplify the list of the error types into a more compact form, but also because they may occur more frequently than expected, especially in nowadays when people input data while using a new keyboard with an unexpected layout, or when they forget to switch the language to the right one they intend to use.

For a given check digit system, we denote t^* to be the longest jump length such that for any $t \le t^*$, all the t-jump transpositions and t-jump twin errors will be detected. Intuitively we have the *detection radius*: $R = t^* + 1$, reflecting the capability of the system to detect these two kinds of generalized jump errors. By definition, a system capable of detecting error types (1)–(5) has an detection radius $R \ge 2$.

Let t_c be the maximum t^* that could be achieved, accordingly $R_c = t_c + 1$ be the *longest detection radius*. Then a check digit system capable of detecting all the single errors and double errors of types (6) and (7) within R_c, is of interest due to its desired error detection capability.

In [11], the authors proposed systems over an abelian group of order p^k, and demonstrated that

$$R_c = \begin{cases} 2^k - 2 & p = 2; \\ \frac{p^k - 1}{2} - 1 & p \text{ is an odd prime.} \end{cases} \tag{1}$$

They also provided easy construction of such systems that not only detect all the single errors and achieve the longest detection radius R_c. We will briefly review their constructive approach in Sect. 3.

Note that in case of $k = 1$, according to (1), we have $R_c = 0$ for $p = 3$, and $R_c = 1$ for $p = 5$, respectively. This implies that their proposed systems over a group of order 3 and 5, respectively, are unable to detect all the errors of types (1)–(5). This is consistent with the note as stated in [14, Remark 5].

3 Check digit system over a group of order p^k

In this section, we review the design of the check digit system proposed in [11] which is based on the use of elementary abelian p-groups of order p^k. This system has the ability to detect all the five error types (1)–(5) and beyond.

3.1 Abelian groups of order p^k

Consider a check digit system over a set of p^k numbers: 0, 1, 2, ..., $p^k - 1$, where p is a prime and k is a positive integer. One can represent these p^k numbers as elements of the abelian group $\mathbf{G}$ defined by

$$\mathbf{G} = \underbrace{\mathbf{Z}_p \oplus \mathbf{Z}_p \oplus \cdots \oplus \mathbf{Z}_p}_{k}, \tag{2}$$

in the manner that each number corresponds to a k-tuple in $\mathbf{G}$. One of the options is to take the k-tuple to be its base p representation.

For instance, consider the hexadecimal numbers: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E and F. One can represent them as elements of the abelian group $\mathbf{G} = \mathbf{Z}_2 \oplus \mathbf{Z}_2 \oplus \mathbf{Z}_2 \oplus \mathbf{Z}_2$ by denoting $0 = (0, 0, 0, 0)$, $1 = (0, 0, 0, 1)$, ..., $9 = (1, 0, 0, 1)$, $\mathrm{A} = (1, 0, 1, 0)$, $\mathrm{B} = (1, 0, 1, 1)$, $\mathrm{C} = (1, 1, 0, 0)$, $\mathrm{D} = (1, 1, 0, 1)$, $\mathrm{E} = (1, 1, 1, 0)$, and $\mathrm{F} = (1, 1, 1, 1)$.

3.2 Check equation

Suppose that the information digits $a_1, \ldots, a_n$ and the check digit a_{n+1} are all interpreted as elements of the abelian group $\mathbf{G}$. Then a_{n+1} can be determined by the check equation

$$a_1 \cdot \mathrm{P} + a_2 \cdot \mathrm{P}^2 + \cdots + a_n \cdot \mathrm{P}^n + a_{n+1} \cdot \mathrm{P}^{n+1} = 0, \tag{3}$$

where P is a $k \times k$ matrix over the prime field $\mathbb{F}_p$. Note that the operations (i.e., multiplication and addition) in the check equation are conducted in $\mathbb{F}_p$.

3.3 Check digit system

As one can see in (3), P is the only and the key parameter which determines the performance of the system of a fixed length in terms of error detection. In particular, we recall the following theorem:

Theorem 1 [11, Theorems 4.4 & 4.6] *Let* P *be a* $k \times k$ *matrix whose characteristic polynomial is a primitive polynomial over* $\mathbb{F}_p$. *Then a check digit system over an abelian group of order* p^k *built on* P *by using the check equation* (3) *is able to detect all the single errors and double errors of type (6) and (7) within detection radius* R_c *as defined in* (1).

It is known that there are $\phi(p^k - 1)/k$ primitive polynomials of degree k over $\mathbb{F}_p$, where $\phi(\cdot)$ is Euler's Totient function. As indicated in [11], given any of the primitive polynomials, an easy construction of matrix P suitable for Theorem 1, is to take its companion matrix. Recall that the companion matrix of a monic polynomial $g(z) = c_0 + c_1 z + \cdots + c_{k-1} z^{k-1} + z^k$ of a positive degree k over $\mathbb{F}_p$ is defined to be the following $k \times k$ matrix

$$\begin{pmatrix} 0 & 0 & 0 & \dots & 0 & -c_0 \\ 1 & 0 & 0 & \dots & 0 & -c_1 \\ 0 & 1 & 0 & \dots & 0 & -c_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -c_{k-1} \end{pmatrix}.$$

The character polynomial of such a matrix over $\mathbb{F}_p$ is exactly $g(z)$.

4 Check digit system over a group of order $p_1^{k_1} p_2^{k_2} \dots p_d^{k_d}$

In this section, we extend the design of the check digit system over an abelian group of order p^k to a more general case, i.e. over an abelian group of arbitrary order $N > 1$. Let $N = p_1^{k_1} p_2^{k_2} \cdots p_d^{k_d}$ be the unique prime factor decomposition of N. Assume that $p_1, p_2, \dots, p_d$ are distinct primes in the order such that $p_1^{k_1} < p_2^{k_2} < \cdots < p_d^{k_d}$.

4.1 Abelian groups of order $p_1^{k_1} p_2^{k_2} \dots p_d^{k_d}$

Recall the fundamental theorem of finite abelian groups that every finite abelian additive group of order $N = p_1^{k_1} p_2^{k_2} \dots p_d^{k_d}$ can be expressed as a direct sum of elementary abelian p_j-groups of order $p_j^{k_j}$, where $1 \le j \le d$. So we can express the abelian group $\mathbf{G}$ of order N to be the direct sum of $\mathbf{G}_j$, i.e.,

$$\mathbf{G} = \mathbf{G}_1 \oplus \mathbf{G}_2 \oplus \cdots \oplus \mathbf{G}_d, \tag{4}$$

where $\mathbf{G}_j$ is the elementary abelian p_j-group of order $p_j^{k_j}$ as described in Sect. 3.1: i.e., for $1 \le j \le d$,

$$\mathbf{G}_j = \underbrace{\mathbf{Z}_{p_j} \oplus \mathbf{Z}_{p_j} \oplus \cdots \oplus \mathbf{Z}_{p_j}}_{k_j}. \tag{5}$$

Consider a check digit system over a set of N numbers. Note that the N numbers, ranging from 0 to $p_1^{k_1} p_2^{k_2} \dots p_d^{k_d} - 1$, can be represented as elements of the abelian group $\mathbf{G}$. Each number x corresponds to a d-tuple in $\mathbf{G}$. For simplicity, we choose the representation $(x_1, x_2, \dots, x_d)$ such that

$$\begin{aligned} x &= q_1 * p_1^{k_1} + x_1; \\ q_1 &= q_2 * p_2^{k_2} + x_2; \\ q_2 &= q_3 * p_3^{k_3} + x_3; \\ &\vdots \quad \vdots \\ q_{d-2} &= q_{d-1} * p_{d-1}^{k_{d-1}} + x_{d-1}; \\ q_{d-1} &= x_d. \end{aligned} \tag{6}$$

Note that for $x \in \{0, 1, \dots \prod_{j=1}^{d} p_j^{k_j} - 1\}$, the quotient $q_j < \prod_{l=j+1}^{d} p_l^{k_l}$ for $1 \le j \le d-1$; the remainder $0 \le x_j < p_j^{k_j}$ for $1 \le j \le d$; and

$$x = x_1 + x_2 \cdot p_1^{k_1} + x_3 \cdot p_1^{k_1} p_2^{k_2} + \cdots + x_d \cdot \prod_{l=1}^{d-1} p_l^{k_l}. \tag{7}$$

Note that x_j can be further interpreted as an element in $\mathbf{G}_j$ as described in Sect. 3.1; and in the sequel we will use its base p_j representation in the calculation.

Therefore, given information digits $a_1, \ldots, a_n$, correspondingly we have their representations in $\mathbf{G}$, i.e., for $1 \le i \le n$,

$$a_i \leftrightarrow (a_{i,1}, a_{i,2}, \ldots, a_{i,d}) \in \mathbf{G}. \tag{8}$$

Furthermore, for $1 \le j \le d$, $a_{i,j}$ can be regarded as an element in $\mathbf{G}_j$, i.e.,

$$a_{i,j} \leftrightarrow (a_{i,j}(1), a_{i,j}(2), \ldots, a_{i,j}(k_j)) \in \mathbf{G}_j, \tag{9}$$

where $a_{i,j}(l) \in \mathbf{Z}_{p_j}$ for $1 \le l \le k_j$.

4.2 Check equation

Let the information digits $a_1, \ldots, a_n$ and the check digit a_{n+1} be interpreted as elements of the abelian group $\mathbf{G}$, i.e.,

$$\begin{array}{lll} a_1 & \leftrightarrow (a_{1,1}, & a_{1,2}, \ldots, a_{1,d}) \\ a_2 & \leftrightarrow (a_{2,1}, & a_{2,2}, \ldots, a_{2,d}) \\ \vdots & \vdots & \vdots \\ a_n & \leftrightarrow (a_{n,1}, & a_{n,2}, \ldots, a_{n,d}) \\ a_{n+1} & \leftrightarrow (a_{n+1,1}, & a_{n+1,2}, \ldots, a_{n+1,d}) \end{array} \tag{10}$$

Consider the j-th coordinate of $a_1, \ldots, a_n$ and a_{n+1}, for $1 \le j \le d$. We have $a_{1,j}, a_{2,j}, \ldots, a_{n,j}$ and $a_{n+1,j}$, all of which are elements of $\mathbf{G}_j$. Thus similar to the approach as described in Sect. 3, one can determine $a_{n+1,j}$ from $a_{1,j}, a_{2,j}, \ldots, a_{n,j}$ by applying the following check equation:

$$a_{1,j} \cdot \mathrm{P}_j + a_{2,j} \cdot \mathrm{P}_j^2 + \cdots + a_{n,j} \cdot \mathrm{P}_j^n + a_{n+1,j} \cdot \mathrm{P}_j^{n+1} = 0, \tag{11}$$

where $a_{i,j}$ for $1 \le i \le n+1$ are as defined in (9); P_j is a $k_j \times k_j$ matrix in the prime field $\mathbb{F}_{p_j}$; and the operations in (11) are conducted in $\mathbb{F}_{p_j}$. In a parallel manner, one can employ d such subsystems over $\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_d$, respectively, to obtain $(a_{n+1,1}, a_{n+1,2}, \ldots, a_{n+1,d})$ over $\mathbf{G}$ which corresponds to a_{n+1} according to (7).

4.3 Check digit system

We note that the representation defined in (6) is a bijection. So when a transmission error occurs, they will be reflected by the representations of the erroneous digits at at least one of the d coordinates. Recall that for each coordinate j, there is an underneath check digit system which is over $\mathbf{G}_j$ with a detection radius (as defined in Sect. 2) R_j, where $1 \le j \le d$. It is easy to see that the overall system over $\mathbf{G}$ could at least achieve a detection radius $R = \min\{R_1, R_2, \ldots, R_d\}$. Therefore, we easily have the following theorem as a generalization of Theorem 1.

Theorem 2 *For $1 \le j \le d$, let P_j be a $k_j \times k_j$ matrix whose characteristic polynomial is a primitive polynomial over $\mathbb{F}_{p_j}$. Then a check digit system over an abelian group of order*

$p_1^{k_1} p_2^{k_2} \cdots p_d^{k_d}$ built on $\{\mathrm{P}_j, 1 \le j \le d\}$ by using the check equation (11) is able to detect all the single errors and double errors of types (6) and (7) within the detection radius

$$R = \begin{cases} 2^{k_1} - 2 & p_1 = 2 \ \& \ 2p_1^{k_1} < p_2^{k_2}; \\ (p_2^{k_2} - 1)/2 - 1 & p_1 = 2 \ \& \ 2p_1^{k_1} > p_2^{k_2}; \\ (p_1^{k_1} - 1)/2 - 1 & p_1 \text{ is an odd prime}. \end{cases} \tag{12}$$

Here it is assumed that $p_1^{k_1} < p_2^{k_2} < \cdots < p_d^{k_d}$.

Proof Consider each subsystem over $\mathbf{G}_j$, for $1 \le j \le d$. Applying Theorem 1, we obtain its detection radius R_j defined by

$$R_j = \begin{cases} 2^{k_j} - 2 & \text{if } p_j = 2; \\ \dfrac{p_j^{k_j} - 1}{2} - 1 & \text{if } p_j \text{ is an odd prime.} \end{cases}$$

As discussed above, the overall system over $\mathbf{G}$ which is built on d parallel subsystem over $\mathbf{G}_j$, for $1 \le j \le d$, could achieve a detection radius $R = \min\{R_1, R_2, \ldots, R_d\}$. Assume that $p_1^{k_1} < p_2^{k_2} < \cdots < p_d^{k_d}$.

In case of $p_1 = 2$, we have that $p_2, \ldots, p_d$ are all odd prime and thus $R_2 = \min\{R_2, \ldots, R_d\}$. Furthermore,

- if $R_1 < R_2$, then we have $R = R_1 = 2^{k_1} - 2$.
  Note that $R_1 < R_2$ implies that

$$\begin{aligned} p_1^{k_1} - 2 &< \frac{p_2^{k_2} - 1}{2} - 1 \\ 2p_1^{k_1} &< p_2^{k_2} + 1 \\ \text{i.e.,} \quad 2p_1^{k_1} &< p_2^{k_2}, \end{aligned}$$

  where the last inequality holds since $2p_1^{k_1} = p_2^{k_2}$ is impossible due to the fact that the left-hand side of the equality results in even number while the right-hand side of the equality results in an odd number.
- if $R_1 \ge R_2$, then we have $R = R_2 = \frac{p_2^{k_2} - 1}{2} - 1$.
  Note that $R_1 \ge R_2$ implies that

$$\begin{aligned} p_1^{k_1} - 2 &\ge \frac{p_2^{k_2} - 1}{2} - 1 \\ 2p_1^{k_1} &\ge p_2^{k_2} + 1 \\ \text{i.e.,} \quad 2p_1^{k_1} &> p_2^{k_2}. \end{aligned}$$

In case of $p_1 \ne 2$, we have $R_1 = \min\{R_1, R_2, \ldots, R_d\}$ under the assumption that $p_1^{k_1} < p_2^{k_2} < \cdots < p_d^{k_d}$. Therefore, we have in this case $R = R_1 = \frac{p_1^{k_1} - 1}{2} - 1$.

We conclude the proof by summarizing the above discussions. □

Remark Although we use a bijection as defined in (6) to represent numbers in $\{0, \ldots, \prod_{j=1}^{d} p_j^{k_j} - 1\}$ as elements of the abelian group $\mathbf{G}$, however, it is worth mentioning that such a bijection is not unique. There are other alternatives. For instance, one can represent any

$x \in \{0, \ldots, \prod_{j=1}^{d} p_j^{k_j} - 1\}$ to be a d-tuple $(x_1, x_2, \ldots, x_d)$, where $x_j = x \mod p_j^{k_j}$ for $1 \le j \le d$; and in turn, given $(x_1, x_2, \ldots, x_d)$, one can retrieve x by applying the Chinese Remainder Theorem.

5 Challenging current standards

In this section, we compare our system with several well known and wildly used systems in real life applications. For the completeness, we first give a brief review on the comparisons of our system over groups of prime power orders [12] to those in current standards, such as ISBN-10 (which is over $\mathbf{Z}_{11}$), MEID and ISAN (which are over hexadecimal numbers, i.e., a group of order $16 = 2^4$). Furthermore, as an example of our general design, we compare our system over hexatridecimal numbers (i.e., over a group of order $36 = 2^2 3^2$) with the alphanumeric check digit system stated in current ISO/IEC 7064 standard [15].

5.1 Check digit system over $\mathbf{Z}_{11}$

Example 1 The 10-digit International Standard Book Number (ISBN-10) code is over $\mathbf{Z}_{11}$, and uses the check equation:

$$\sum_{i=1}^{10} i \cdot x_i \equiv 0 \pmod{11}. \tag{13}$$

This system detect all the errors of types (1)–(5) with the exception of the twin error at places $(5, 6)$ as $5 + 6 = 11$.

In [14, Example 18], a modification of the ISBN-10 code is proposed over $\mathbf{Z}_{11}$, which uses the following check equation:

$$\sum_{i=1}^{5} i \cdot x_i + \sum_{j=6}^{10} (5 - j) \cdot x_j \equiv 0 \pmod{11}. \tag{14}$$

This system improves the ISBN-10 code in the manner that it detects all the errors of types (1)–(5) without exceptions.

Our system is designed by applying Theorem 2 for $p = 11$ and $k = 1$. Note that over $\mathbf{Z}_{11}$, the primitive elements are 2, 6, 7, 8. We use $\alpha \in \{2, 6, 7, 8\}$ in the following check equation:

$$\sum_{i=1}^{10} \alpha^i \cdot x_i \equiv 0 \pmod{11}. \tag{15}$$

For instance, if we take $\alpha = 6$, then (15) has coefficients $\{\alpha^i \mod 11, 1 \le i \le 10\} = \{6, 3, 7, 9, 10, 5, 8, 4, 2, 1\}$. In particular, the MOD 11-2 system specified in ISO/IEC 7064 [15] can be considered as a special case of our system by taking $\alpha = 2$. Since the coefficients are fixed for a chosen α, compared to ISBN-10, our proposed system involves no computational complexity, for the calculation of the check digit and the verification.

By Theorem 2, our system could detect not only the errors of types (1)–(5), but also more generalized errors of types (6)–(7) within detection radius $R_c = 4$. It is easy to check that this holds also for the modification system defined by (14). In fact, both systems are able to detect all the single errors, all the t-jump transposition errors, and almost all the t-jump twin errors with the only exception of the 4-jump twin error.

Moreover, our system outperforms the ISBN-10 code and the modification system defined by (14) on detecting the phonetic errors ($\cdots a0 \cdots \to \cdots 1a \cdots$, where $a \geq 2$). It only fails to detect the phonetic errors as $x_i = 10$ for $1 \leq i \leq 9$. However, in the ISBN-10 code, number 10 is not used for x_i, $1 \leq i \leq 9$ at all. So to say, our system is able to detect all the possible phonetic errors; whilst both the ISBN-10 code and the modification proposed in [14, Example 18] fail to do so (one can refer [14, Example 18] for a detailed list of undetectable phonetic errors for both systems).

5.2 Check digit system over hexadecimal numbers

Example 2 A Mobile Equipment IDentifier (MEID) [16,17] is a globally unique 14-digit hexadecimal identification number for a physical piece of mobile station equipment. They are used as a means to facilitate mobile equipment identification and tracking and therefore should be resistant to modification.

An MEID is composed mainly of two basic components, the manufacturer code and the serial number, as shown in Table 2. Given 14 information digits $x_{14}x_{13}\ldots x_1$, the following different methods are employed to calculate the check digit, say x_{15}, as described in [17, Annex B]. Note that the check digit is not part of the MEID and is not transmitted when the MEID is transmitted.

1. For an MEID if all digits are decimal (i.e. the MEID is an IMEI for use with multi-mode phones), then the check digit x_{15} is calculated using the standard (decimal) Luhn formula as described in [18, ISO/IEC 7812-1:2006(E)]. Simply, the check equation is

$$\sum_{i=1}^{7}\left[\lfloor\frac{x_{2i-1}}{5}\rfloor + 2x_{2i-1} + x_{2i}\right] + x_{15} \equiv 0 \pmod{10}, \tag{16}$$

   where $\lfloor x \rfloor$ is the largest integer not greater than x. Note that the same method is also used to calculate the check digit for the 18-digit decimal representation of a 14-digit hexadecimal MEID.
2. For an MEID which contains at least one hexadecimal digit in the RR digits, the check digit x_{15} is calculated using a slight modification of the Luhn formula, in the manner that all arithmetic is performed in base 16. The check equation is therefore

$$\sum_{i=1}^{7}\left[\lfloor\frac{x_{2i-1}}{8}\rfloor + 2x_{2i-1} + x_{2i}\right] + x_{15} \equiv 0 \pmod{16}. \tag{17}$$

In this example, we focus only on the 2nd case: the hexadecimal MEID defined by (17). The reason is that the 1st case: the MEID defined by (16), in fact falls into the category of the check digit system over decimal numbers,whist in the 2nd case the MEID is practically a

Table 2 The format of MEID [16]

MEID														Check digit
Manufacturer code								Serial number						
R	R	X	X	X	X	X	X	Z	Z	Z	Z	Z	Z	C
x_{14}	x_{13}	x_{12}	x_{11}	x_{10}	x_9	x_8	x_7	x_6	x_5	x_4	x_3	x_2	x_1	x_{15}

Table 3 The format of ISAN [19]

ISAN																
Root												Episode				Check digit
R	R	R	R	R	R	R	R	R	R	R	R	E	E	E	E	C
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}

check digit system over hexadecimal numbers. It is easy to verify that the hexadecimal MEID defined by (17) fails to detect the following errors:

– generalized jump transposition errors:

1. all the t-jump transposition errors for all odd $t > 0$ (i.e., at places $(i, i + t + 1)$);
2. $(\mathrm{F}, 0) \leftrightarrow (0, \mathrm{F})$ at places $(i, i + t)$ for all odd t such that $1 \le i \le 14, i + t \le 14$;

– generalized jump twin errors:

1. $(x, x) \leftrightarrow (x + 4, x + 4)$ with $0 \le x \le 3$ or $8 \le x \le 11$ at places $(i, i + t)$ for all even t and odd i such that $1 \le i \le 14, i + t \le 14$;
2. $(x, x) \leftrightarrow (x + 8, x + 8)$ with $0 \le x \le 7$ at places $(i, i + t)$ for all even t and even i such that $1 \le i \le 14, i + t \le 14$;
3. $(x, x) \leftrightarrow (x + 5, x + 5)$ with $3 \le x \le 7$ at places $(i, i + t)$ for all odd t such that $1 \le i \le 14, i + t \le 14$.

Example 3 The International Standard Audiovisual Number (ISAN) [19] is a numbering system that enables the unique and persistent identification of any audiovisual works.

An ISAN consists of 16 hexadecimal digits, which can be divided into two segments: root segment and episode segment, as shown in Table 3. An appended check digit is calculated over the 16 ISAN digits according to a MOD 37, 36 system as specified in accordance with ISO/IEC 7064 [15]. That is, given $x_1 x_2 \ldots x_{16}$, the check digit x_{17} can be calculated by the check equation

$$(\cdots((((36 + x_1)||_{36} \cdot 2)|_{37} + x_2)||_{36} \cdot 2)|_{37} + \cdots + x_{17})\,||_{36} = 1, \tag{18}$$

where $x||_{36}$ is the remainder of x after dividing by 36 and if it is 0 then the value 36 will be substituted; and $x|_{37}$ is the reminder of x after dividing by 37. Note that although the 16 ISAN digits are hexadecimal, however, the check digit is hexatridecimal which belongs to the set of numbers '0' to '9' and letters 'A' to 'Z'.

According to [15] and our calculation given in Table 4, the MOD 37, 36 system is unable to detect all the errors of types (1)–(5). In fact, it fails to detect about 0.16 % of error type (2), 1.9 % of error (3), and 2.8 % of generalized jump twin errors (i.e., error type 7): t-jump twin errors for $t \ge 0$ which generalizes error types (4) as $t = 0$ and (5) as $t = 1$).

In summary, both MEID and ISAN, two widely used hexadecimal check digit systems, fail to detect all the errors of types (1)–(5).

Example 4 Following our approach as described in Theorem 1, one can construct alternative systems for MEID and ISAN as follows:

– Represent the hexadecimal numbers as elements of the abelian group $\mathbf{G} = \mathbf{Z}_2 \oplus \mathbf{Z}_2 \oplus \mathbf{Z}_2 \oplus \mathbf{Z}_2$ as described in Sect. 3.1.

Table 4 Undetected errors of error types (1)–(5) in percentage

Check digit system	Undetected errors in %				
	Type (1)	Type (2)	Type (3)	Type (4)	Type (5)
ISBN-10	0.0	0.0	0.0	11.11	0.0
(Hexadecimal) Luhn formula	0.0	0.833	100.0	4.167	6.667
MOD 37, 36	0.0	0.159	1.900	1.905	3.642
Our design as their alternatives	0.0	0.0	0.0	0.0	0.0

Calculations are based on the assumption that each character occurs in each position (of the information digits) with equal probability.
The five error types are as defined in Sect. 1. That is, (1) single error; (2) transposition error; (3) jump transposition error; (4) twin error and (5) jump twin error

– Find a 4×4 matrix P whose characteristic polynomial is either $z^4 + z + 1$ or $z^4 + z^3 + 1$ (both are primitive polynomial of degree 4). For instance, we take the companion matrix of $z^4 + z^3 + 1$ as an easy choice of such P. That is,

$$\mathrm{P} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

– Apply the following check equation to calculate the check digit a_{n+1} from the n information digits $a_1, \ldots, a_n$:

$$\sum_{i=1}^{n+1} a_i \cdot \mathrm{P}^i = 0.$$

It is easy to see that the above system serves as an alternative for the hexadecimal MEID by letting $n = 14$, and an alternative for ISAN by letting $n = 16$. In particular, our system, as an alternative for MEID, is capable of detecting all the possibles errors of types (1) and (6)–(7); whilst as an alternative for ISAN, is capable of detecting all the single errors, and almost all the possible jump errors of types (6)–(7) with the only exception of 14-jump transposition and twin errors at places (1, 16) and (2, 17).

5.3 Check digit system over hexatridecimal numbers

Alphanumeric is a combination of alphabetic and numeric characters, containing in total 36 symbols. Naturally it is very popular in human interfaces. So a good design of check digit system over hexatridecimal numbers is of great interest. For convenience, we usually keep the numbers 0 to 9 while map the characters 'A' to 'Z' to numbers 10 to 35 in calculation.

Example 5 A MOD 37, 36 system as specified in accordance with ISO/IEC 7064 [15], belongs to the category of check digit system over hexatridecimal numbers and used in the real life application for livestock identification [20]. In general, for a given information sequence $x_1 x_2 \ldots x_n$, the check digit x_{n+1} can be calculated by the following check equation:

$$(\cdots((((36 + x_1)||_{36} \cdot 2)|_{37} + x_2)||_{36} \cdot 2)|_{37} + \cdots + x_{n+1})||_{36} = 1. \tag{19}$$

We note that the ISAN code introduced in Sect. 5.2 is a MOD 37, 36 system by taking $n = 16$ and additionally limiting the information digits $x_1, x_2, \ldots, x_{16}$ to be hexadecimal numbers.

Recall the fact that the MOD 37, 36 system fails to detect all the most frequent 5 types errors (see [15] or Table 4.) In the following, we provide an alternative design that detects all the most frequent 5 types errors.

Example 6 Following our approach as described in Theorem 2, one can construct alternative alphanumeric systems as follows:

- Since $36 = 2^2 \cdot 3^2$, we can represent the hexatridecimal numbers 0 to 35 as elements of the abelian group $\mathbf{G} = \mathbf{G}_1 \oplus \mathbf{G}_2$ where $\mathbf{G}_1 = \mathbf{Z}_2 \oplus \mathbf{Z}_2$ and $\mathbf{G}_2 = \mathbf{Z}_3 \oplus \mathbf{Z}_3$.
- More specifically, given n characters, $a_1, \ldots, a_n$, we easily have $(q_1, \ldots, q_n)$ and $(r_1, \ldots, r_n)$, where q_i and r_i are the quotient and remainder, respectively, when dividing a_i by 9 for $1 \leq i \leq n$. Note that in the calculation, q_i is also used for its base 2 representation as an element in $\mathbf{G}_1 = \mathbf{Z}_2 \oplus \mathbf{Z}_2$; and r_i for its base 3 representation as an element in $\mathbf{G}_2 = \mathbf{Z}_3 \oplus \mathbf{Z}_3$.
- To set up the check equation, we need two matrices: one binary matrix P_1 and one ternary matrix P_2, such that their characteristic polynomials are primitive polynomials over the prime fields $\mathbb{F}_2$ and $\mathbb{F}_3$, respectively. That is, P_1 should be a matrix which has $x^2 + x + 1$ (which is the only primitive polynomial of degree 2 in $\mathbb{F}_2$) as its characteristic polynomial; and P_2 should be a matrix which has $x^2 + x + 2$ as its characteristic matrix (there are only 3 monic irreducible polynomials of degree 2 in $\mathbb{F}_3$: $x^2 + 2x + 2$, $x^2 + 1$, and $x^2 + x + 2$; and $x^2 + x + 2$ is the only primitive one). Simply taking the companion matrices of these primitive polynomials, we have, respectively,

$$\mathrm{P}_1 = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}; \quad \mathrm{P}_2 = \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}.$$

- From the n information digits $a_1, \ldots, a_n$, correspondingly we have the representations $(q_1, \ldots, q_n)$ and $(r_1, \ldots, r_n)$. Apply the following check equations to obtain (q_{n+1}, r_{n+1}) :

$$\begin{aligned} \sum_{i=1}^{n+1} q_i \cdot \mathrm{P}_1^i &= 0 \text{ over } \mathbb{F}_2; \\ \sum_{i=1}^{n+1} r_i \cdot \mathrm{P}_2^i &= 0 \text{ over } \mathbb{F}_3. \end{aligned} \tag{20}$$

- The check digit is $a_{n+1} = q_{n+1} * 9 + r_{n+1}$.

By Theorem 2, the above system has a detection radius $R = 2$ and thus can detect all the most frequent error types (1)–(5). It performs better than the existing alphanumeric system in ISO/IEC 7064 such as MOD 37, 36 system (with one check digit). In fact, there are another two alphanumeric systems stated in ISO/IEC 7064: MOD 37-2 and MOD 1271-36. However, we note that MOD 37-2 is over 37 characters ('0' to '9', and 'A' to 'Z', plus '*') and MOD 1271-36 is with 2 check digits.

One can refer to [21] for a web interface of our proposed alphanumeric check digit system as given in Example 6. And, more supportive documentations such as implementation code and detailed explanations can be found at [22]. Therein the alphanumeric system is implemented to check the correctness of the check digits of the Electric Vehicle Contract IDs. Note

that the Contract Identifier (short: CID; also known as eMA-ID or EVCO-ID) is described by the eMI^3 Group and standardized in ISO/IEC-15118 [23, Annex H].

5.4 Computational complexity

Now let us consider the check digit systems based on different methods from the perspective of the computational complexity. We mainly focus on the 'expensive' operations such as addition and multiplication, and do not include in our comparison the table look-ups caused by character representation since they are often faster in terms of processing time.

Given n information digits and one check digit, computations involved in the check equation to detect whether error occurs are listed as follows:

- in MEID, the underlying Luhn formula employed in its check equation (17) for the hexadecimal case (or (16) for decimal case), requires about
  - $n/2$ multiplications modulo 16 (or modulo 10),
  - $2n$ additions modulo 16 (or modulo 10);
- in the MOD 37, 36 system, its check equation (19) requires about
  - n multiplications modulo 37, and
  - $(n+1)$ additions modulo 36;
- in the system over a group of order p^k as described in Sect. 3,
  - as $n < p^k - 2$, the check equation (3) requires about
    - $k^2(n+1)$ multiplications modulo p, and
    - $k(kn+k-1)$ additions modulo p;
  - as $n \geq p^k - 2$, one can first rearrange of the check equation (3) (using the fact $\mathrm{P}^i = \mathrm{I}$ if $(p^k - 1)|i$) to reduce the calculations. After the rearrangement, it requires about
    - $k^2(p^k - 2)$ multiplications modulo p, and
    - $k^2(p^k - 2) + kn$ additions modulo p;

  So to say, in general, it needs at most $k^2(p^k - 2)$ multiplications modulo p, and at most $k^2(p^k - 2) + kn$ additions modulo p. We notice that unlike the Luhn formula and MOD 37, 36 , the number of multiplications needed here is bounded by a constant rather than a linear increase with respect to n therein.
- in our proposed system over a group of order $p_1^{k_1} p_2^{k_2} \ldots p_d^{k_d}$ as described in Sect. 4, its check equation (11) requires for $1 \leq i \leq d$ at most
  - $k_i^2(p_i^{k_i} - 2)$ multiplications modulo p_i, and
  - $k_i^2(p_i^{k_i} - 2) + k_i(n+1)$ additions modulo p_i.

  Similarly, the number of multiplications needed in the system is bounded by some constant which is independent of the length of the information digits.

6 Conclusion

In this paper, we study a check digit system over an abelian group of an arbitrary order, which simply employs several parallel subsystems which are based on the use of elementary abelian p-groups. Furthermore, we challenge the current standards by comparing this system with several well-known and widely used check digit systems such as ISBN, MEID, ISAN and

the system over alphanumeric characters. We show that this universal design outperforms all of them in terms of the error detection capability. Moreover, from the perspective of complexity, the number of multiplications needed in this design can be bounded by some constant (which is independent of the length of the information digits), rather than a linear increase as by Luhn formula and methods such as MOD 37, 36. The design concept is in general simple, constructive and easy to be adopted in practice and it therefore serves as an attractive alternative for ISBN code, MEID, ISAN, alphanumeric identification system and for other applications. As one can see in [10,21], promising steps have been taken in the implementation of our system over hexadecimal or alphanumeric numbers in either standardization or real life applications.

References

1. Beckley D.F.: An optimum system with modulo 11. Comput. Bull. **11**, 213–215 (1967).
2. Verhoeff J.: Error Detecting Decimal Codes. Mathematical Centre Tracts, vol. 29. Mathematica Centrum, Amsterdam (1969).
3. Schul R.H.: On check digit systems using anti-symmetric mappings. Numbers, Information and Complexity. Kluwer Academic Publishers, Boston (2000).
4. Belyavskaya G.B., Izbash V.I., Mullen G.L.: Check character systems using quasigroups: I. Des. Codes Cryptogr. **37**, 215–227 (2005).
5. Belyavskaya G.B., Izbash V.I., Mullen G.L.: Check character systems using quasigroups: II. Des. Codes Cryptogr. **37**, 405–419 (2005).
6. Damm H.M.: Total anti-symmetrische Quasigruppen (Dr. rer. nat.). Philipps-Universität Marburg (2004).
7. Gumm H.P.: A new class of check-digit methods for arbitrary number systems. IEEE Trans. Inf. Theory **31**, 102–105 (1985).
8. Broecker C., Schulz R.-H., Stroth G.: Check character systems using Chevalley groups. Des. Codes Cryptogr. **10**(2), 137–143 (1997).
9. Niemenmaa M.: A check digit system for hexadecimal numbers. Appl. Algebra Eng. Commun. Comput. **22**, 109–112 (2011).
10. MISB ST 1204.1: Motion Imagery Identification System (MIIS) - Core Identifier (2013).
11. Chen Y., Niemenmaa M., Han Vinck A.J., Gligoroski D.: On some properties of a check digit system. IEEE International Symposium on Information Theory (ISIT 2012), Cambridge, MA, 1–6 July (2012).
12. Chen Y., Niemenmaa M., Han Vinck A.J.: A check digit system over a group of arbitrary order. In: 8th International ICST Conference on Communications and Networking in China, pp. 897–902 (2013).
13. Mullen G.L., Shcherbacov V.A.: Properties of codes with one check symbol from a quasigroup point of view. Izv. AN RM Math. **3**, 71–86 (2002).
14. Mullen G.L., Shcherbacov V.: n-T-quasigroup codes with one check symbol and their error detection capabilities. Comment. Math. Univ. Carolinae **45**(2), 321–340 (2004).
15. ISO/IEC 7064: 2003(E): Information technology—security techniques—check character systems.
16. 3GPP2 report S. R0048: 3G Mobile Equipment Identifier (MEID)—Stage 1 (2005).
17. 3GPP2 X. S0008–0 v3.0: MAP Support for the Mobile Equipment Identity (MEID) (2009).
18. ISO/IEC 7812–1:2006(E): Identification cards—identification of issuers—part 1: numbering system.
19. ISO 15706 2002(E): Information and documentation—International Standard Audiovisual Number (ISAN).
20. NAIS Program Standards and Technical Reference, Version 2.2, APHIS, USDA (2008).
21. http://www.ochp.eu/id-validator/.
22. http://www.ochp.eu/downloads/.
23. ISO 15118–1:2013: Road vehicles—vehicle to grid communication interface—part 1: general information and use-case definition.