Optimised information system reliability techniques for knowledge society acceptance

Nikolaos G. Bardis*

Institute of Informatics & Telecommunications,
Net Media Lab, N.C.S.R. Demokritos
Terma Patriarchou Grigoriou & Neapoleos 27,
Agia Paraskevi, 153 10, Athens, Greece
and
Department of Mathematics and Engineering Science,
Hellenic Army Academy, University of Military Education,
BST 902, Vari, Greece
E-mail: bardis@ieee.org
*Corresponding author

Athanasios Drigas

Institute of Informatics & Telecommunications,
Net Media Lab, N.C.S.R. Demokritos
Terma Patriarchou Grigoriou & Neapoleos 27,
Agia Paraskevi, 153 10, Athens, Greece
E-mail: dr@imm.demokritos.gr

Nikolaos Doukas and Nikolaos V. Karadimas

Department of Mathematics and Engineering Science,
Hellenic Army Academy, University of Military Education,
BST 902, Vari, Greece
E-mail: doukasn@sse.gr E-mail: nkaradimas@sse.gr

Abstract: Two important obstacles averting the widespread adoption of knowledge society and e-government technologies in communities are distrust in their reliability and fear of the costs to be incurred. In order to overcome these obstacles, algorithmic advances are required that improve reliability of information systems without increasing their cost. In this paper a new approach to increase the reliability and effectiveness of transient error detection using checksum calculations is proposed, that is based on the optimisation of coding with special differential Boolean transformations. Two methods for obtaining such transformations are developed and examples of coding for checksum functions are given. The new methods are equivalent in terms of computational cost as classical checksum approaches.

Keywords: checksum; error detection; error control; differential Boolean transformations; data integrity; fault tolerance.

Reference to this paper should be made as follows: Bardis, N.G., Drigas, A., Doukas, N. and Karadimas, N.V. (2009) 'Optimised information system reliability techniques for knowledge society acceptance', *Int. J. Knowledge and Learning*, Vol. 5, Nos. 3/4, pp.207–221.

Biographical notes: Nikolaos Bardis is an Adjunct Researcher at National Centre for Scientific Research 'Demokritos' – Greece on Applied Technologies Department and Net Media Lab, Adjunct Assistant Professor in Automation Department at the Technological Education Institute (T.E.I) of Chalkida (Greece), and Research Scientist at University of Military Education (Hellenic Naval Academy, Hellenic Air force Academy, Hellenic Army Academy). His research interests include cryptography and data security, information theory, artificial intelligence, databases and software engineering. He is a Chartered Engineer and a member of the Greek Technical Chamber, member of the IEEE and at ComSoc as TPC member.

Athanasios Drigas is a Senior Researcher at N.C.S.R. Demokritos. He is the Coordinator of Telecoms and founder of Net Media Lab since 1996. From 1985 to 1999 he was Operational Manager of the Greek Academic network. He has been the Coordinator of Several International Projects, in the fields of ICTs, and e-services (e-learning, e-psychology, e-government, e-inclusion, e-culture, etc.). He has published more than 200 articles, seven books, 25 educational CD-ROMs and several patents. He has been a member of several international committees for the design and coordination of network and ICT activities and of international conferences and journals.

Nikolaos Doukas is a Lecturer (407/80) at the University of Military Education – Hellenic Army Academy and the Hellenic Air Force Academy. His current research focuses on security and reliability for information processing systems, the application of e-learning to teaching and further education within the context of military institutions of higher education, computer aided design and cryptography. He has also participated in research work in software engineering, broadband communications and computer graphics. He has presented publications in several aspects of his work and has presented his work in many international conferences.

Nikolaos V. Karadimas is a Postdoctoral Researcher at the Multimedia Technology Laboratory within National Technical University of Athens since 2007 and a Lecturer (407/80) at the University of Military Education – Hellenic Army Academy. He is teaching in Technical NCO Academy and in Technological Educational Institute of Piraeus, as well. He is a Chartered Engineer and a member of the Greek Technical Chamber, of the IEEE and of IET. His research interests are in the fields of databases, optimisation techniques, geographical information systems and decision support systems with emphasis in the military applications.

1 Introduction

The European institutions continuously emphasise that a knowledge and information society will be sustainable only if it ensures inclusion and broad electronic participation in society (EU4Journalists, 2008). Individual concerns about privacy, security and the use of information about their preferences and activities are an important barrier to the formation of an effective and broad-based knowledge society and a corresponding e-government system. If individuals distrust sending the identifying or financial information over the internet that is needed to complete transactions, the fraction of commercial and societal activities which can benefit from transition to the electronic medium will be significantly restricted. As a result, insufficient protection (or a

perception of insufficient protection) of personal privacy and security in these systems is a potentially serious impediment in the development of the information society, knowledge dissemination and very importantly, e-government and is therefore important from the perspective of policy making (SIBIS, 2002).

The present level of computer networks and telecommunication systems development is inseparably connected with the problem information integrity, which includes: the ensuring of high reliability during information transmission and the information storage in memories. The dynamic expansion of using the communication channels which are potentially exposed to interferences and the complex methods of packing the transferred information are combined with the increase of errors appearance during data transmission. A similar situation occurs also during the storage of data on magnetic means: a constant increase in the longitudinal and transverse density of information storage on such means is also combined with an increase of appearing errors (Klove and Korzhik, 1995; Klove, 2007). Data corruption may also occur due to hardware faults during processing.

As it has already been pointed out, the extensive usage of information technologies in a wide variety of human activities termed information society and the technogenic risks that are thus included, create an increasing demand for reliability in all aspects of operation of information systems, including reliability during the transmission and the storage and the processing of information (Saxena and McCluskey, 1987). A continuous increase in the speeds of data transmission in the telecommunications network systems dictates the stringent requirements for the effectiveness of the means of error control (Kounavis, and Berry, 2005). This effectiveness must be commensurate with the channel characteristics as the bandwidth and the rate of transmitted data. This condition determines the needs for a radical increase of the error control reliability of the means of error control. These means should have an increased speed and allow the parallel processing in hardware implementation (Lin and Costello, 1983).

The considerations outlined above imply that, from the point of view of promoting the information societies objectives on knowledge and learning, there is an urgent need for developing new or improving existing instruments that maintain reliability during data transmission, processing and storage within the context of information systems and communications networks. It should be pointed out that the term communication systems cover both the case of large scale networks (Ethernet, GSM etc.) but also micro networks (information exchange between individual components of a particular information processing device).

Increases in the processing and transmission rates of current information systems adversely affects the reliability of the transmission as the error rates due to interfering signals are also increased. Additionally, the widespread use of wireless networks as well as their broad penetration in recent years has lead to an increase in the number and intensity of electromagnetic interference sources. This interference is one of the principle causes of errors during information transmission.

As a consequence of these factors, a steady evolution of the technology dealing with error control during digital data transmission is required. Within this framework of evolution, a modification was recently proposed of one of the most common methods for error control, namely the checksum. A significant increase in the reliability of this error detection technique has been rendered feasible via the optimisation of the encoding of the components of the checksum, while maintaining its basic advantage, which is simplicity.

The principle scope for practical application of the proposed modified checksum method, are data transmission channels that interconnect computational systems. In such channels the error rates are relatively small and the application of complex control methods is considered inefficient. The reason is that the integration of such error control methods would in this case consume a disproportionately large part of computational resources, compared to the execution of the remaining administration functions.

The modified checksum guarantees a successful compromise between high reliability in error detection and simplicity of implementation of control and administration procedures. Consequently, robustness against the basic transmission error categories becomes feasible without any effect on the overall system computational load. Computerised management and command systems usually operate in real time conditions and possible disturbances in the flow of data could create significant technological risks. The modified checksum provides protection against such eventualities by enhancing robustness while maintaining the complexity at an acceptable level.

The risks arising from the interruption of data processing is particularly acute in computer systems that manage critical processes such as chemical plants, aircraft navigation, guidance systems, railway networks, electricity production plants (including nuclear ones), military applications, etc. Delays in the control and management processes of such applications, as well as transmission errors between their components may lead to catastrophic results. On the contrary, increases in the robustness of systems that support the Information Society contribute significantly to the overall social well being. In computer control and management systems the proposed modification of the checksum may be extensively used for the effective control of digital data transmission between system components, as well as for enhancing the reliability of communication between peer systems interconnected by local area networks.

This paper is organised as follows: in Section 2 presents the problems that appear in the context of detecting transient errors during the transmission, storage and processing of digital information. Section 3 presents two different optimisations of the checksum that result in different schemes for error detection. The new schemes are shown to have different merits, the first reducing dramatically the probability of undetected errors and the second one being suitable for the cases where symbols are likely to be repeated very often. Section 4 presents an analysis of the impact of the proposed technique in the aims and goals of knowledge society and e-government. Section 5 presents the following steps that are being pursued in this research and finally, Section presents the conclusions of the work presented in this paper.

2 Reliability problems in digital information systems

For guaranteeing the reliable data transmission in communication channels of computer networks a variety of means is used, which in most cases employ coding of the transmitted information. In the majority of systems, the transmission, processing and storage of information is carried out on a block by block basis and the respective integrity of the data of each block is controlled separately.

With the use of special coding it is possible to distinguish two approaches for the correction of the errors appearing:

- error detection by special codes and their correction by retransmission or reacquisition of the block upon request [automatic repeat request (ARQ)]
- correction of the appearing errors via the use of correcting codes without the need for repeated data exchanges [forward error correction (FEC)].

Elementary considerations show that the first approach cannot be used in the case of errors occurring during data storage. The reason is that in this case the corrupted data is the only available copy. The main advantage of FEC is that it is more efficient in exploiting channel bandwidth, since there is no need for traffic for data repetition. ARQ is better on the other hand, in the fact that the error detection requires simpler decoding hardware and smaller redundancy than the error correction method. The implementation of error detection has substantially smaller computational complexity and it is hence possible to calculate the error control functions considerably faster. In general however, the effectiveness of ARQ is smaller and depends on the multiplicity of errors.

The choice between the two approaches for the elimination of the appearing transient errors depends on intensity and the nature of the errors. The basic sources of errors in digital data channels are the inter-bit interferences, the externally produced noise and the thermal noise of transmission means (Klove and Korzhik, 1995). An additional source of errors may of course be sought in hardware faults.

The nature of the appearing errors depends not only on their source, but also on the type of transmission means and on the signal modulation method. Thus, in the ether communication channels the prevailing source of the transmitted errors is the externally produced noise and in this case the intensity of the appearing errors is sufficiently high so that the application of FEC technologies proves to be more preferable. In the wire line systems for digital data transmission, in which the intensity of errors is several orders lower in comparison with the wireless channels, the use of ARQ is considered to be more effective (Fletcher, 1983).

In the cable channels with the sequential data transmission without modulation, the transmitted error has the same nature, and the channels themselves correspond to the binary symmetrical channel model. This model assumes the appearance of erroneous transmission of zero or one with equal probabilities, since the probability p_j that j errors occur during the transmission of the n-bit code is determined for the binary symmetrical channel by the expression:

$$p_j = \binom{n}{j} \cdot p^j \cdot (1-p)^{n-j}, \tag{1}$$

where p is the probability of the erroneous transmission of one bit.

For error detection by the per block data transmission the most usual solutions employed are the cyclic redundancy check (CRC) codes and the checksum (CS) (Klove and Korzhik, 1995). The CS method in comparison with CRC is substantially simpler and ensures the maximum rate of the error control and it is an influential factor on a constant increase in the channel capacity of data transmission.

In contrast to CRC, the structure of the operations which are performed with the checksum calculation allow parallel processing which makes it possible to implement effectively this control by hardware. Therefore the time required for the calculation of error control will essentially not affect the performance of the data transmission.

Let us denote by D_1, D_2, …, D_k the k codes with n-bit size, which compose the transmission block, and by D_1', D_2', …, D_k' we denote the blocks on the receiver end. The CS's on the receiver and the sender are calculated with the same way: $S_S = D_1 \oplus D_2 \oplus \ldots \oplus D_k$ and $S_R = D_1' \oplus D_2' \oplus \ldots \oplus D_k'$.

The usual checksum calculation assumes that the data transmission is carried out as a block and organised as k codes D_1, D_2, …, D_k with a length of n bits. In this case the length n of the code is determined by the architecture of the control organisation and its value can coincide with the number of simultaneously transferred bits, and it can differ from it.

At the end of the transmission of the data block, the transmitter sends to the receiver the checksum S_S, which is XORed with the checksum that is calculated on the receiver S_R and obtains the differential code $\Delta = S_S \oplus S_R$ of size n. If $\Delta = 0$ then we consider that no errors have arisen. For the symmetrical binary channel and taking into account that in practice the relation $k >> n$ holds, then we can consider that during the transmission of one code only one error can arise.

The low reliability of the error detection of the even error multiplicity is the main disadvantage of the checksum. Actually, the most probable occurrence between them is the two-fold error (situation of appearance of single errors in two from k transmitted codes) and the code Δ can attain only n^2 different values of all the 2^n possible. It is means that for the studied model the usual checksum is ineffectively coding of two-fold error. Because of this, the probability P_2 of non-detection of the two-fold error is reasonably high and determined by the following error formula (Klove and Korzhik, 1995):

$$P_2 = \frac{1}{n} \tag{2}$$

Thus, the reliability level of the error detection when using the checksum can increase via the use of a coding optimisation, i.e., the calculation of checksums on the transmitter and the receiver in the form:

$$S_S = F(D_1) \oplus F(D_2) \oplus \ldots \oplus F(D_k) \text{ and } S_R = F(D_1') \oplus F(D_2') \oplus \ldots \oplus F(D_k'),$$

where F is the function of coding, defined by the system of Boolean functions.

In Bardis and Markovskyy (2004) an orthogonal system of Boolean functions that satisfy the SAC criterion has been proposed to be used as coding function. In this case, with the appearance of the two-fold error the code Δ has $n!/((n/2)!)^2$ different values and consequently, the probability of the non detection of the two-fold error substantially decrease in comparison with the usual checksum.

Usually these types of Boolean transformations are used in cryptographic algorithms and methods for their design have been developed in Bardis (2004) and Klove and Korzhik (1995).

A Boolean function $f(x_1, \ldots, x_n)$ defined on a set Z of all possible 2^n n-tuples of n variables, satisfies the SAC, if a complement of a single incoming n-tuple data bit changes the output of the Boolean function with probability 50%:

$$\forall j \in \{1, \ldots, n\}:$$
$$\sum_{x_1, \ldots, x_n \in Z} (f(x_1, \ldots, x_j, \ldots, x_n) \oplus f(x_1, \ldots, \overline{x_j}, \ldots, x_n)) = 2^{n-1} \tag{3}$$

If one of the n inputs of the avalanche transformation is changed then half of its outputs will be changed. This means, that there is an 'avalanche amplifier' which by changing one of the n-tuple incoming data bit transforms half of the outputs. Because every function of this system satisfies the avalanche criterion, these transformations are called 'avalanche'.

Let F(D) denote the Boolean orthogonal avalanche transformation on the n-bits code D. So the transformation $F(D)$ consists of the orthogonal Boolean functions $f_1(D)$, $f_2(D)$, …, $f_n(D)$, every one of which satisfies the avalanche criterion. The length of the transformed code $R = F(D)$ is n bits long, as well. The orthogonality of the $F(D)$ transformation indicates the one-to-one correspondence of codes D and R. The avalanche properties of the $F(D)$ transformation indicate that if one bit of the input code D is changed then, on average, $n/2$ bits of the output code $R = F(D)$ will be changed also.

Example of the Boolean orthogonal avalanche transformation $F(D)$ on the 8-bits code $D(n = 8)$ is given below:

$$
\begin{aligned}
f_1(X) &= d_8d_7 + d_8d_6 + d_8d_5 + d_8d_4 + d_8d_2 + d_8d_1 + d_7d_5 + \\
&\quad + d_7d_4 + d_7d_2 + d_6d_3 + d_5d_3 + d_5d_1 + d_1 \\
f_2(D) &= d_8d_6 + d_8d_5 + d_8d_3 + d_8d_1 + d_7d_6 + d_7d_4 + d_7d_2 + d_7d_1 + \\
&\quad + d_6d_4 + d_6d_2 + d_5d_4 + d_5d_2 + d_5d_1 + d_4d_1 + d_3d_1 + d_2d_1 + d_2 \\
f_3(D) &= d_8d_7 + d_8d_5 + d_8d_3 + d_8d_2 + d_7d_6 + d_7d_5 + d_7d_3 + d_7d_2 + \\
&\quad + d_7d_1 + d_6d_5 + d_6d_3 + d_6d_2 + d_5d_4 + d_4d_2 + d_3d_2 + d_3 + d_2d_1 \\
f_4(D) &= d_8d_7 + d_8d_6 + d_8d_4 + d_8d_3 + d_8d_2 + d_8d_1 + d_7d_4 + d_7d_3 \\
&\quad + d_7d_2 + d_6d_5 + d_5d_4 + d_5d_3 + d_5d_2 + d_4d_3 + d_4 + d_1d_2 + d_3d_1 \\
f_5(D) &= d_8d_5 + d_8d_4 + d_8d_3 + d_8d_1 + d_7d_2 + d_7d_1 + d_6d_5 + d_6d_4 + \\
&\quad + d_6d_3 + d_6d_1 + d_5d_2 + d_5d_1 + d_5 + d_4d_3 + d_4d_2 + d_3d_1 + d_2d_1 \\
f_6(D) &= d_8d_3 + d_8d_2 + d_7d_5 + d_7d_4 + d_6d_3 + d_6d_2 + d_6 + d_5d_3 + d_5d_2 + \\
&\quad + d_5d_1 + d_4d_2 + d_3d_2 + d_2d_1 \\
f_7(D) &= d_8d_6 + d_8d_5 + d_7d_6 + d_7d_4 + d_7d_3 + d_7d_2 + d_7 + d_6d_4 + d_6d_3 \\
&\quad + d_6d_2 + d_6d_1 + d_5d_4 + d_5d_3 + d_5d_2 + d_4d_3 + d_3d_2 + d_3d_1 \\
f_8(D) &= d_8d_7 + d_8d_5 + d_8d_4 + d_8d_3 + d_8 + d_7d_6 + d_7d_5 + d_7d_4 + d_7d_3 \\
&\quad + d_7d_1 + d_6d_5 + d_6d_4 + d_6d_3 + d_5d_2 + d_4d_3 + d_4d_2 + d_4d_1
\end{aligned}
\tag{4}
$$

Thus, if a single error appears, then $n/2$ bits of the modified checksum will change. If a second error appears then another $n/2$ bits of the modified checksum will change. It is clear that the probability of the masking interaction of $n/2$ erroneous bit pairs is less than the probability of the masking interaction of a single bit pair.

It has been shown that the probability P_{2f} that the dual bit errors will not be detected in case orthogonal avalanche transformation $F(D)$ using is determined as follows:

$$
P_{2f} = \frac{1}{\binom{n}{n/2}} = \frac{((n/2)!)^2}{n!} = \prod_{j=0}^{n/2-1} \frac{j+1}{(n-j)} \tag{5}
$$

Thus, the probability of detecting dual errors during block transmission using the checksum control scheme orthogonal avalanche transformation $F(D)$, increases by t_2 times in comparison to the ordinary checksum scheme. The numerical value of the t_2 increase is determined by the formula:

$$t_2 = \prod_{j=1}^{n/2-1} \frac{n-j}{j+1} \tag{6}$$

For example, for $n = 8$, the probability that the dual errors will not be detected is decreased by 8.7 times in comparison to the traditional checksum.

As it has already been mentioned for the case of the orthogonal avalanche transformation $F(D)$, the code Δ has $n!/((n/2)!)^2$ different values and correspondingly the probability of the non detection of the two-fold error substantially decreases in comparison with the usual checksum. However in this case the $n!/((n/2)!)^2$ coding variants of the two-fold error are substantially less than the total number of all possible codes becomes $\Delta - 2^n$ and therefore in this case the optimisation of coding is not achieved. Consequently, an increase in the reliability of detection of prevailing type errors by checksum can be achieved via a further optimisation of its coding and the selection of a corresponding functional transformation.

The purpose of this research is to increase the reliability of checksum error control by developing functional transformations which optimise coding for the prevailing forms of errors in the binary symmetrical channel.

3 Checksum optimisation

For the practical implementation of error detection based on coding checksum optimisation for detecting transient errors, a calculation method of the modified checksum is proposed, similarly as in work (Bardis and Markovskyy, 2004). In this proposed method as terms are used codes which are obtained from Boolean transformations over the controlled codes. These Boolean transformations consist of a system of m Boolean functions with n variables:

$$F(D) = \{f_1(D), f_2(D), ..., f_m(D)\} \tag{7}$$

where D is an n-bit code: $D = \{d_1, d_2, \ldots, d_n\}$, $\forall j \in \{1, \ldots, n\}: d_j \in \{0, 1\}$.

With the appearance of a single error in the jth bit of the code D_i it is transformed into $D_i' = \{d_1, \ldots, d_j \oplus 1, \ldots, d_n\}$ and the differential Δ of the checksum can be represented in the form of the differentials values of the functions $f_1, f_2, \ldots, f_m$ with the variable d_j on the binary tuples $\{d_1, d_2, \ldots, d_j - 1, d_j + 1, \ldots, d_n\}$:

$$\begin{aligned} &\Delta = F(D_i) \oplus F(D_i') = \\ &\{f_1(D_i) \oplus f_1(D_i'), ..., f_m(D_i) \oplus f_m(D_i')\} = \\ &= \{\frac{\partial f_1}{\partial d_j}, \frac{\partial f_2}{\partial d_j}, ..., \frac{\partial f_m}{\partial d_j}\} \end{aligned} \tag{8}$$

The optimisation of single error coding in the modified m bits checksum can be achieved, if the number of possible values of the code of Δ equals 2^m. Since the number of versions

of the single error localisation in the code D is equal to n, so that the single error could be one way coded by checksum it is enough that $m = [\log_2 n]$ holds. In this case the binary code formed by a change in the functions with the appearance of error in the jth bit of the code D, i.e., with a change in the variable d_j, is equal to j–1:

$$\forall j \in \{1,...,n\}: \sum_{t=0}^{\lceil \log_2 n \rceil - 1} \frac{\partial f_t}{\partial d_j} \cdot 2^t = j - 1 \tag{9}$$

It is obvious that the condition (9) is satisfied if each of the functions $f_1, f_2, \ldots, f_m$ is linear, and the q function f_q includes the variable d_j (i.e., the value of the q bit of the binary number j–1 is equal to one). For example, if $n = 8$, then $m = 3$ and the system of functions which satisfy (9) can be as follows:

$$\begin{aligned} f_1 &= d_2 \oplus d_4 \oplus d_6 \oplus d_8 \\ f_2 &= d_3 \oplus d_4 \oplus d_7 \oplus d_8 \\ f_3 &= d_5 \oplus d_6 \oplus d_7 \oplus d_8 \end{aligned} \tag{10}$$

In this case, the length size of checksum of its coding is substantially lower than the code length size of Δ: $m < n$, however, the probability that the error of any multiplicity larger than one (single errors they are detected always) corresponds to expression (2). For example, the two-fold error is not detected only when both errors occurred in one and the same bit.

In order to decrease the probability of not detecting the multiplicity errors, it is necessary in addition of the Boolean system (9) which forms the set Ξ_1, to use a system of u functions which compose the set Ξ_2 so that $F = \{\Xi_1, \Xi_2\}$.

In order to detect the two-fold error with high reliability it is necessary that a number of conditions are fulfilled. Since two errors, localised in different bits of the transferred codes are always detected using the functions of the set Ξ_1, and so for detecting the errors which appear in the same bit on different codes of block it is necessary that the probability of the values agreeing of the differentials functions of the set Ξ_2 with the change of one variable, should be as small as possible or near to zero. Therefore the functions differentials of this set must not be constant, i.e., the functions must be non-linear. Moreover, the functions differentials $f_{m+1}, f_{m+2}, \ldots, f_{m+u}$, on any of the variables must constitute an orthogonal system of functions.

For the realisation of this condition two methods for the functions synthesis of the set of Ξ_2 are proposed. According to the first method the functions of $f_{m+1}, f_{m+2}, \ldots, f_{m+u}$ are defined on a set with n variables, which coincide with the values of the bits of the transferred codes and the set of their possible values forms the set Z. In this case the number of functions of the set Ξ_2 is equal to n–1, i.e., $u = n$–1, and the functions $f_{m+1}, f_{m+2}, \ldots, f_{m+u}$ must satisfy the condition:

$$\begin{aligned} &\forall j \in \{1,...,n\}, a_t \in \{0,1\}: \\ &\sum_{D \in Z} \bigoplus_{t=1}^{u} a_t \cdot \frac{\partial f_{m+t}}{\partial d_j} = 2^{u-1} \end{aligned} \tag{11}$$

Below is given an example of a system of seven Boolean functions with eight variables ($n = 8$), which compose the set Ξ_2 and satisfy the condition (11):

$$
\begin{aligned}
f_4 &= d_1 \cdot d_2 \oplus d_3 \cdot d_4 \oplus d_5 \cdot d_7 \oplus d_6 \cdot d_8 \\
f_5 &= d_1 \cdot d_3 \oplus d_2 \cdot d_4 \oplus d_5 \cdot d_8 \oplus d_6 \cdot d_7 \\
f_6 &= d_1 \cdot d_4 \oplus d_2 \cdot d_5 \oplus d_3 \cdot d_6 \oplus d_7 \cdot d_8 \\
f_7 &= d_1 \cdot d_5 \oplus d_2 \cdot d_6 \oplus d_3 \cdot d_7 \oplus d_4 \cdot d_8 \\
f_8 &= d_1 \cdot d_6 \oplus d_2 \cdot d_7 \oplus d_3 \cdot d_8 \oplus d_4 \cdot d_5 \\
f_9 &= d_1 \cdot d_7 \oplus d_2 \cdot d_8 \oplus d_3 \cdot d_5 \oplus d_4 \cdot d_6 \\
f_{10} &= d_1 \cdot d_8 \oplus d_2 \cdot d_3 \oplus d_4 \cdot d_7 \oplus d_5 \cdot d_6
\end{aligned}
\tag{12}
$$

The differentials on any of the eight variables in the seven functions, which compose the set Ξ_2, form the system of orthogonal Boolean functions. For example, the differentials of the 4th variable d_4 form the system of linear functions, the orthogonality property of which is obvious:

$$
\begin{aligned}
&\frac{\partial f_4}{\partial d_4} = d_3; \frac{\partial f_5}{\partial d_4} = d_2; \frac{\partial f_6}{\partial d_4} = d_1; \frac{\partial f_7}{\partial d_4} = d_8; \\
&\frac{\partial f_8}{\partial d_4} = d_5; \frac{\partial f_9}{\partial d_4} = d_6; \frac{\partial f_{10}}{\partial d_4} = d_7
\end{aligned}
\tag{13}
$$

The differentials of functions f_1, f_2, f_3 in terms of the variable d_4 are equal and they correspond to the number of the variable x_4:j–$1 = 011_2 = 3$.

The validity of the conditions (11) ensures the detection of two errors, if they occur in the same bits of different codes. This means that the two-fold error will not be detected only when both errors will occur in one and the same bit j during the transmission of the two codes D_i and D_e, i.e., $i, e \in \{1, \ldots, k\}$, which either are equal or they are different only on the jth bit. This probability is determined by the formula:

$$
p_2 = \frac{1}{2^{n-1} \cdot n} \tag{14}
$$

It is obvious that formula (14) determines the probability of the non-detection not only of two-fold, but also of errors of any multiplicity larger than one. Comparison with the expression (2) shows that the reliability of the two-fold error detection is substantially increased in comparison with the usual checksum. Thus, the essence of the first of the proposed methods for the coding optimisation of the checksum lies in the fact that the transformation function of the total components of the set Ξ_2 is selected in such a way that their differentials on any of the variables will depend on the code D. The two-fold error will not be detected only when both errors will occur in one and the same bit of the pair of the same codes, or the pair of the codes is different only in this bit.

If we consider that the appearance of each of the n-bit codes in the block is equally probable, then the probability of the two-fold error which satisfied these conditions is determined by the formula (14). However, in practice very frequent is the situation, when some codes in the block are repeated sufficiently frequently. This situation is characteristic for the text documents and for images. Hence, in this case, the effectiveness of the proposed method of coding the components of checksum decreases.

The second coding method of the checksum components is deprived of this deficiency and consist of the fact that the functions $f_{m+1}, f_{m+2}, \ldots, f_{m+u}$ of the set Ξ_2 are obtained in such a way that their differentials on any of the variables will depend on the

number H of the code in the block. If the controlled block includes k codes then the number w of the bits of the H code is equal to $w = \lceil \log_2 k \rceil$ and consequently, $u = w$. In this case the functions $f_{m+1}, f_{m+2}, \ldots, f_{m+w}$ are defined on a set of $n + w$ variables, which includes n bits of the transmitted codes $d_1, d_2, \ldots, d_n$ and w is the number of the transmitted codes in the block $h_1, h_2, \ldots, h_w$. For the optimum coding of the components of checksum from the errors detection point of view, the Boolean functions $f_{m+1}, f_{m+2}, \ldots, f_{m+u}$ must be obtained in such a way, that their differentials on any of the variables $d_1, d_2, \ldots, d_n$ will form an orthogonal system of functions from the variables $h_1, h_2, \ldots, h_w$ which can attain 2^w values to compose the set Q.

In the particular case the differentials of the system of the Boolean functions $f_{m+1}, f_{m+2}, \ldots, f_{m+u}$ on any of the variables $d_1, d_2, \ldots, d_n$ can be orthogonal linear systems from the variables $h_1, h_2, \ldots, h_w$. In this case the following condition must hold:

$$\forall j \in \{1,\ldots, n\}, a_t \in \{0,1\}: \quad \sum_{H \in Q} \bigoplus_{t=1}^{w} a_t \cdot \frac{\partial f_{m+t}}{\partial d_j} = 2^{u-1} \tag{15}$$

As an example we can assume the number k of the codes in the controlled block to be equal to 64 ($k = 64$) and the code length to be equal to 8 ($n = 8$). Then $w = u = 6$ and correspondingly the code number H in the block consists of 6-bits $H = \{h_1, h_2, \ldots, h_6\}$. The system of the Boolean functions, which compose the set Ξ_2 including the six functions and satisfies the condition (15), has the following form:

$$\begin{aligned} f_4 &= h_1 \cdot (d_1 \oplus d_2 \oplus \ldots \oplus d_8) \\ f_5 &= h_2 \cdot (d_1 \oplus d_2 \oplus \ldots \oplus d_8) \\ f_6 &= h_3 \cdot (d_1 \oplus d_2 \oplus \ldots \oplus d_8\} \\ f_7 &= h_4 \cdot (d_1 \oplus d_2 \oplus \ldots \oplus d_8\} \\ f_8 &= h_6 \cdot (d_1 \oplus d_2 \oplus \ldots \oplus d_8) \\ f_9 &= h_6 \cdot (d_1 \oplus d_2 \oplus \ldots \oplus d_8) \end{aligned} \tag{16}$$

The Differentials on any of the eight variables of the given six functions, which compose the set Ξ_2, forms the system of orthogonal Boolean functions. For example, the differentials of the 4th variable d_4 form the system of linear functions, for which the property of orthogonality is obvious:

$$\begin{aligned} &\frac{\partial f_4}{\partial d_4} = h_1; \frac{\partial f_5}{\partial d_4} = h_2; \frac{\partial f_6}{\partial d_4} = h_3; \\ &\frac{\partial f_7}{\partial d_4} = h_4; \frac{\partial f_8}{\partial d_4} = h_5; \frac{\partial f_9}{\partial d_4} = h_6 \end{aligned} \tag{17}$$

This ensures the optimality of coding error with each fixed bits in which it occurred, i.e., the unique dependence of the error code on the number of the transmitted code in the block.

With the appearance of two-fold error in different codes the differentials of the functions $f_{m+1}, f_{m+2}, \ldots, f_{m+u}$ will be different, and consequently $\Delta \neq 0$. With the appearance of the two-fold error in one code, the differentials of functions $f_1, f_2, \ldots, f_m$

will be different and so $\Delta \neq 0$. Thus, the second proposed method for the optimisation of coding checksum ensures the detection of all one and two-fold errors which are the more often appearing errors for the symmetrical binary channels. The probability P_3 that errors with larger multiplicity will not be detected with the proposed method is determined by the following expression:

$$P_3 = \frac{1}{k \cdot n} \tag{18}$$

The comparison of expressions (14) and (18) shows that the use of the second checksum coding method is more preferable in the case of large number of codes in the blocks. In comparison with the usual checksum, an increase in the reliability of the error detection of even multiplicity with the use of the proposed coding method composes several orders.

In comparison with the use of SAC transformations (Klove and Korzhik, 1995), the reliability by using the first proposed method for multiple errors detection increases t_1 times and t_2 times by using the second method. In these cases the numerical values of t_1 and t_2 are determined by the expressions:

$$\begin{aligned} t_1 &= \frac{2^{n-1} \cdot ((n/2)!)^2}{(n-1)!} \\ t_2 &= \frac{k \cdot ((n/2)!)^2}{(n-1)!} \end{aligned} \tag{19}$$

The analysis of the values of formulas (19) for the most frequently in practice values k and n shows that the reliability of the multiply error detection increases by several orders of magnitude in comparison with the use of the SAC transformations (Bardis and Markovskyy, 2004).

The calculations which realise the functional transformations (10), (12), and (16) are considerably simpler than the operations of polynomials divisions which are performed in CRC and allow the multilevel parallel processing. Thus, with the hardware realisation the logical operations of functions (12), the functions of systems (10) (12) and (16) can be simultaneously calculated in parallel.

The above observation makes it possible to implement a fast realisation of the control operations which are unattainable by using the CRC method and to effectively introduce the error detection functions without delaying the processing of the high-speed transmission channels of digital information.

4 Impact of IS reliability on knowledge society and e-government and e-banking

The development of information system integration via computer networks is one of the fundamental factors of progress of the Information society and of the social structure in general. The information transmission technology is the most important constituent part in this information integration revolution. Progress in the field of information

technologies is impossible without an increase in the bandwidth and the reliability available for the transmission of digital data.

One of the difficulties in promoting the adoption of new processes and technologies aiming at more efficient knowledge management in society is trust towards the innovation (Angehrn, 2005). In other words people need to be reassured that the new technology is going to work in their own interest. Any guarantee that a new communication network is less susceptible to errors than the one that they are already using, contributes significantly in this direction. Such a guarantee may be offered by the proposed modified checksum technique.

Society is beginning to appreciate that organisational knowledge is a source of a competitive advantage in the long run (Ordonez de Pablos, 2005). Thus, the creation, deployment, transfer, transformation, renewal, and accumulation of knowledge turned out to be key strategic activities for all kinds of firms and organisations, as well as for society in general (Ordonez de Pablos, 2005). The proposed modified checksum technique safeguards a significant proportion of this strategic activity which includes both the deployment and the transfer of knowledge.

One of the factors that promote learning in society is 'information and knowledge literacy', i.e. the ability to recognise an information or knowledge need, and then to locate, evaluate, and use effectively the needed information (King, 2005). In this context, the proposed technique can contribute in locating, obtaining and sharing the knowledge needed.

In electronic learning platforms, a recurring and persistent problem is keeping learning material up-to-date in the presence of progress. One of the solutions to this problem is automated reengineering techniques (Pankratius and Vossen, 2005). Such techniques can significantly benefit from the proposed modified checksum, since it reduces the need for post update, manual verification of the automatically uploaded data. In Gupta and Bostrom (2005) the fact that knowledge flow between teams facilitates decision-making is pointed out. This flow greatly benefits from the new technique, since the new technique guarantees the integrity of this flow.

The advantage of implementing semantic portals that allow students to look for courses distributed across many locations is pointed out in Vargas-Vera and Moreale (2005). One of the basic enabling technologies for such portals is error detection and the proposed scheme contributes significantly in guaranteeing course data integrity.

5 Future research

From a technical point of view this research is currently being expanded to exploit the ideas of the modified checksum in developing a corresponding FEC scheme. This will be done by a suitable modification of the Hamming code calculation in order to reduce the amount of memory used and the complexity of indexing calculations required. Additionally, the benefits of the introduction of the modified checksum scheme for transient error detection in high reliability storage systems will need to be quantified within the context of particular code designs. Finally, the application of the modified checksum towards improving the tolerance of information processing systems to hardware faults and transient algorithmic errors is also being investigated.

6 Conclusions

The proposed method is based on the coding optimisation to increase the effectiveness of the error detection during data transmission and data storage using the checksums. This makes it possible to significantly decrease the probability of the mutual masking of even multiplicity errors both in comparison with the usual checksum and the use of SAC transformations for coding of its components (Bardis and Markovskyy, 2004).

Two methods for obtaining special differential Boolean functional transformations which optimise the checksum coding of the codes in block are developed. These two methods are developed from the point of view of the criterion of error detection which appears in the binary symmetrical channel. Examples of the transformations which optimise coding checksum for both the proposed methods are given.

The estimations of the probability of error detecting of different multiplicity are theoretically substantiated. It is proven that during coding of the checksum components using functions which depend both on the transmitted codes and on the number of the code in the block, all errors of multiplicity less than three will be detected. The analysis carried out showed that using the proposed methods make it possible to decrease the probability of non detection of multiple errors by several orders of magnitude, in comparison with the known schemes of checksum calculation.

The structure of the functional transformation is considerably simpler in comparison with the transformations of the CRC method and allows multilevel parallel process on hardware implementation which makes it possible to ensure the high performance of the error control without delays in the process of data transmission (Albertengo and Sisto, 1990; Bardis et al., 2005; Partrige et al., 1995). The developed methods have been shown to be a significant contribution towards the goal of enhancing the community acceptance of knowledge society, e-government and e-banking technologies.

References

Albertengo, G.A. and Sisto, R.J. (1990) 'Parallel CRC generation', *IEEE Micro*, Vol. 11, No. 10, pp.84–91

Angehrn, A.A. (2005) 'Designing innovation games for community-based earning and knowledge exchange', *Int. J. Knowledge and Learning*, Vol. 1, No. 3.

Bardis N.G (2004) 'Echoplex error control system using avalanche transformations', *WSEAS Transactions on Communications*, ISSN 1109-2742, Vol. 3, No. 2, pp.741–745.

Bardis, E.G. and Markovskyy, A.P. (2004) 'Utilization of avalanche transformation for increasing of echoplex and checksum data transmission control reliability, *2004 International Symposium on Information Theory and its Applications (ISITA-2004)*, 10–13 October, Parma, Italy, pp.656–660.

Bardis, N.G, Bardis E.G., Markovskyy A.P. and Economou, C. (2005) 'Hardware implementation of data transmission control based on Boolean transformation, WSEAS Transactions on Communications, ISSN 1109-2742, Vol. 4, No. 7, pp.363–371.

EU4Journalists (2008) *Information Society and Media*, available at http://www.eu4journalists.eu/index.php/dossiers/english/C78/255/, (accessed on October).

Fletcher, J. (1983) 'An arithmetic checksum for serial transmissions', *IEEE Transactions on Communications*, Vol. 30, No. 1, pp.76–85.

Gupta, S. and Bostrom, R.P. (2005) 'Theoretical model for investigating the impact of knowledge portals on different levels of knowledge processing', *Int. J. Knowledge and Learning*, Vol. 1, No. 4, pp.287–304.

King, W.R. (2005) 'An androgogy model for IS and management education', *Int. J. Knowledge and Learning*, Vol. 1, No. 3.

Klove, T. (2007) 'Codes for error detection, serial on coding theory and cryptography', *World Scientific*, Vol. 2, p.201.

Klove, T. and Korzhik, V. (1995) *Error Detecting Codes: General Theory and Their Application in Feedback Communication Systems*, Kluwer, Norwell, MA, p.433.

Kounavis, M.E. and Berry, F.L. (2005) 'A systematic approach to building high performance software based CRC generators', *10th IEEE Symposium on Computers and Communications (ISCC'05)*, pp.855–862.

Lin, S. and Costello, D.J. Jr, (1983) *Error Control Coding*, Prentice-Hall, Inc., Englewood Cliffs, N.J.07632, ISBN 0-13-283796-X, p.603.

Ordonez de Pablos, P. (2005) 'Intellectual capital statements: what pioneering firms from Asia and Europe are doing now', *Int. J. Knowledge and Learning*, Vol. 1, No. 3, pp.249–268.

Pankratius, V. and Vossen, G. (2005) 'Reengineering of educational material: a systematic approach', *Int. J. Knowledge and Learning*, Vol. 1, No. 3, pp.229–248.

Partrige, C., Hughes, J. and Stone, J. (1995) 'Performance of checksums and CRCs over real data', *Proc. SIGCOMM'95 Conf.*, ASM, pp.68–76.

Saxena, N.R. and McCluskey, E.J. (1987) 'Extended precision checksums', *Proc.17-th Intern. Symp. Fault Tolerant Comput.: FCTS-17*, Pittsburgrh, USA, pp.142–147.

SIBIS Statistical Indicators Benchmarking the Information Society (2002) *Workpackage 1: eEurope Benchmarking Framework*, IST publication, October.

Vargas-Vera, M. and Moreale, E. (2005) 'Automatic extraction of knowledge from student essays', *Int. J. Knowledge and Learning*, Vol. 1, No. 4, pp.318–331.