OPTIMIZED HASH FUNCTIONS FOR INTEGRITY CONTROL AND DATA RECOVERY IN EMBEDDED SYSTEMS

Rozlomii I. O.[0000-0001-5065-9004]**, PhD, Yarmilko A.**[0000-0003-2062-2694]**, PhD, Naumenko S.**[0000-0002-6337-1605]

Bohdan Khmelnytsky National University of Cherkasy, Ukraine
inna-roz@ukr.net

Abstract: The method of constructing a generalized cryptographic hashing method for integrity control and data recovery with the introduction of minimal redundancy is discusses. It is suggested to use a hash code system to control data integrity. They are built according to rules similar to the rules for building linear redundant codes. The method has the potential to be implemented in embedded systems, in particular, in IoT systems.

Keywords: Data integrity control, data recovery, hash function, matrix crypto transformations, redundant codes.

The rapid development of embedded systems technologies demands reliable methods for data integrity assurance that take into account the limited resources of embedded systems. Among various methods, hash function schemes are of significant interest. They are successfully applied in tasks related to localizing faulty blocks; however, they are not without drawbacks. The main one among them is the high redundancy when verifying the integrity of small-sized message block sequences.

Considering the mentioned drawback of existing solutions, the search for ways to reduce redundancy, introduced for a given level of data protection in the presence of random errors and destructive influences from an attacker, is relevant. Moreover, despite the widespread use of hash functions, they are relatively under-researched. Practical proposals for their application mainly focus on finding ways to enhance their cryptographic strength [1]. However, proposals for the application of hash functions that would allow reducing the introduced redundancy for a given level of data protection are very scarce.

The results obtained up to the current period allow for the substantiated formulation of requirements for information protection means, including its authenticity and integrity control. A proven means of managing the authenticity of information transmission is error-resistant encoding [2].

Today, Hamming codes are considered a fundamental approach to building error-resistant systems for encoding and decoding digital data. Existing research and publications provide a sufficient foundation for generalizing the concepts of error-resistant coding and hashing methods [3]. They confirm the fact that the theory of error-resistant coding, particularly Hamming correction codes, has been adapted to address data integrity issues in various application domains.

The introduction of redundant information into transmitted data, which can be altered during network transmission, enables the detection and correction of errors at the receiver's end of the information message. The mathematical theory of constructing redundant (error-resistant) codes has made significant advancements. However, there exists a substantial gap between the level of theoretical achievements in error-resistant coding theory and the level of practical results derived from applying this theory.

We have proposed a method of cryptographic hashing based on Hamming codes for information protection and recovery. Using the mathematical framework of vector system theory, we have developed an algorithm for constructing linear hash codes to ensure data integrity in information systems. The construction rules of hash codes are influenced by the specified (or necessary) level of information resource protection. The redundancy of the control information depends on the need for error-correcting properties. The construction rules of linear hash code systems are analogous to the rules for constructing Hamming codes. Thus, a well-developed theory of linear redundant codes can be applied in the new area of constructing linear hash code systems.

The main advantage of the proposed method is the ability to control information integrity and correct defects for a given level of protection with minimal redundancy and the capability to localize integrity violations and correct a specified number of errors.

The developed linear hash codes, constructed by analogy to Hamming codes, allow for error correction in message blocks. However, the number of errors that can be corrected (the code's error-correcting capability) depends on the size of the redundant code (control blocks of information). Therefore, it's essential to rationally select the necessary redundant code to ensure the required information reliability while not burdening communication channels with excessive redundant data. In other words, the goal is to achieve data integrity with minimal redundant code.

The proposed solutions can be applied both in traditional information systems and in the implementation of intelligent systems with hybrid human-machine intelligence. An example of this is the contemporary concept of creating production systems (such as Industry 4.0, Industry 5.0), which focus on utilizing artificial intelligence as a secure, reliable, and responsible component within a unified human-machine functional and communication space. The intellectualization of human-machine interaction, the cooperative nature of intelligent agent activities, and their interaction in an open functional space undoubtedly emphasize the issues of efficiency, security, and predictability of such systems. The urgency of these problems, directly related to the state of communication, is likely to intensify as these technologies transition from being unique and experimental projects to becoming widespread utilitarian applications.

Since a common solution for creating modules of cooperative systems of this kind involves their replication on an embedded systems platform, the presence of constrained devices and battery life limitations will directly affect the operation of security subsystems. Therefore, the security problem in such systems needs to be addressed comprehensively. Methods and algorithms that are not only aligned with the main functional capabilities but also resource-efficient for the technical platform are of high priority.

In general, the practical implementation of relevant software tools should contribute to enhancing trust among communication process participants by enabling the identification and recovery of damaged message fragments in the information stream. Considering these aspects, the proposed method of ensuring message integrity aligns with the general requirements for such means.

The obtained results provide a scientific and engineering toolkit for controlling and ensuring data integrity, with the ability to verify their authenticity after recovery in case of integrity violations. They also provide the necessary conditions for creating advanced and improving existing data storage systems.

[1] J. Andress, The basics of information security: understanding the fundamentals of InfoSec in theory and practice, 2nd. ed., Syngress, 2014.

[2] J. Sima, J. Bruck, On optimal k-deletion correcting codes, IEEE Transactions on Information Theory 67.6 (2020) 3360-3375.

[3] A. Yarmilko, I. Rozlomii, H. Kosenyuk, Hash method for information stream’s safety in dynamic cooperative production system, in: S. Shkarlet et al. (Eds): Mathematical Modeling and Simulation of Systems, volume 344 of Lecture Notes in Networks and Systems, Springer, Cham, 2022, pp. 173-183. doi.org/10.1007/978-3-030-89902-8_14.