INVESTIGATION OF INFLUENCE OF ENCODING FUNCTION COMPLEXITY ON DISTRIBUTION OF ERROR MASKING PROBABILITY

A.B. Levina[a], S.V. Taranov[a]

[a] ITMO University, Saint Petersburg, 197101, Russian Federation
Corresponding author: alla_levina@mail.ru

Article info
Received 19.01.16, accepted 26.02.16
doi:10.17586/2226-1494-2016-16-2-331-337
Article in English

For citation: Levina A.B., Taranov S.V. Investigation of influence of encoding function complexity on distribution of error masking probability. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2016, vol. 16, no. 2, pp. 331–337, doi:10.17586/2226-1494-2016-16-2-331-337

Abstract
Error detection codes are mechanisms that enable robust delivery of data in unreliable communication channels and devices. Unreliable channels and devices are error-prone objects. Respectively, error detection codes allow detecting such errors. There are two classes of error detecting codes - classical codes and security-oriented codes. The classical codes have high percentage of detected errors; however, they have a high probability to miss an error in algebraic manipulation. In order, security-oriented codes are codes with a small Hamming distance and high protection to algebraic manipulation. The probability of error masking is a fundamental parameter of security-oriented codes. A detailed study of this parameter allows analyzing the behavior of the error-correcting code in the case of error injection in the encoding device. In order, the complexity of the encoding function plays an important role in the security-oriented codes. Encoding functions with less computational complexity and a low probability of masking are the best protection of encoding device against malicious acts. This paper investigates the influence of encoding function complexity on the error masking probability distribution. It will be shown that the more complex encoding function reduces the maximum of error masking probability. It is also shown in the paper that increasing of the function complexity changes the error masking probability distribution. In particular, increasing of computational complexity decreases the difference between the maximum and average value of the error masking probability. Our results have shown that functions with greater complexity have smoothed maximums of error masking probability, which significantly complicates the analysis of error-correcting code by attacker. As a result, in case of complex encoding function the probability of the algebraic manipulation is reduced. The paper discusses an approach how to measure the error masking probability in the case of nonuniform distribution of the input code words. This approach can also be used to study the characteristics of security-oriented codes in case of strong and weak models of algebraic manipulation.

Keywords
coding theory, error masking probability, AMD codes, complexity of encoding function, side channel attacks

Acknowledgements
This work was presented on the Information Security and Protection of Information Technology Conference ISPIT 2015.

УДК 004.056.2

ИССЛЕДОВАНИЕ ВЛИЯНИЯ СЛОЖНОСТИ ФУНКЦИИ КОДИРОВАНИЯ НА РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТИ МАСКИРОВКИ ОШИБКИ

А.Б. Левина[a], С.В. Таранов[a]

[a] Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация
Адрес для переписки: alla_levina@mail.ru

Информация о статье
Поступила в редакцию 19.01.16, принята к печати 26.02.16
doi:10.17586/2226-1494-2016-16-2-331-337
Язык статьи – английский

Ссылка для цитирования: Левина А.Б., Таранов С.В. Исследование влияния сложности функции кодирования на распределение вероятности маскировки ошибки // Научно-технический вестник информационных технологий, механики и оптики. 2016. Т. 16. № 2. С.331–337. doi:10.17586/2226-1494-2016-16-2-331-337

Аннотация
Помехоустойчивые коды являются механизмом, который позволяет обеспечить надежную передачу данных в каналах с ошибками. Ненадежные каналы и устройства подвержены ошибкам внедрения. Помехоустойчивые коды позволяют обнаруживать такие ошибки. Существует два класса помехоустойчивых кодов – классические коды и

коды, ориентированные на безопасность. Классические коды имеют высокий процент обнаружения ошибок, но, в то же время, высокую вероятность пропустить ошибку в случае алгебраической манипуляции. В свою очередь, коды, ориентированные на безопасность, имеют маленькое кодовое расстояние и высокую защиту от алгебраических манипуляций. Вероятность маскировки ошибки является основным параметром кодов, ориентированных на безопасность. Детальное изучение данного параметра позволяет проанализировать поведение помехоустойчивых кодов в случае внедрения ошибки в устройство кодирования. Немаловажным параметром кодов, ориентированных на безопасность, является также сложность функций кодирования. Функции кодирования с низкой вычислительной сложностью и низкой вероятностью маскировки ошибки являются лучшей защитой устройств кодирования против действий злоумышленников. В работе исследуется влияние сложности функции кодирования на распределение вероятности маскировки ошибки. Показано, что вычислительно сложные функции имеют меньшую вероятность маскировки ошибки. Также изменение сложности функций кодирования влечет изменения в распределении вероятности маскировки ошибки. В частности, увеличение вычислительной сложности уменьшает разность между максимальным и средним значением вероятности маскировки ошибки. Показано, что функции кодирования с большой сложностью имеют менее различимые максимумы вероятности маскировки ошибки, что значительно упрощает анализ помехоустойчивого кода злоумышленником. Как результат, в случае более сложной функции кодирования вероятность алгебраической манипуляции уменьшается. Предлагается подход к измерению вероятности маскировки ошибки при неравномерном распределении входных кодовых слов. Данный подход может быть использован для изучения характеристик кодов, ориентированных на безопасность, для случаев сильной и слабой модели алгебраических манипуляций.

Ключевые слова

теория кодирования, вероятность маскировки ошибки, AMD-коды, сложность функции кодирования, атаки по сторонним каналам

Благодарности

Работа представлена на Конференции по информационной безопасности и защите информации ISPIT 2015.

Introduction

Being transmitted and stored, digital data suffer from noise that can introduce errors in the binary bits. Errors may occur either randomly or as a result of malicious acts. Error detection codes are mechanisms that enable robust delivery of data in unreliable communication channels and devices. Error-detecting codes add a redundancy data to a given digital data for detection of errors occurred during transmission, storage and injection by attackers. There are two classes of error detecting codes – classical codes and security-oriented codes. Classical error detecting codes are trying to maximize the Hamming distance between code words; at that, the larger Hamming distance is, the more errors can be detected and corrected. However, studies of Mark Karpovskiy [1–3] have shown that the classical error detecting codes are vulnerable to the fault injection attacks in the encoding device. Classical linear error-detecting codes are not optimum for error detection in communication channels and devices when the error distributions of a channel are non-stationary or unknown since they do not minimize the worst case error masking probability.

To protect against such attacks the class of security-oriented codes has been designed. Security-oriented codes provide detection of all possible errors in any distribution of input code words. Currently, there are two approaches to security-oriented code development: «robust codes» of Mark Karpovsky [4–6] and «algebraic manipulation detection (AMD) codes» of Ronald Cramer [7, 8]. AMD codes are considered as generalization of robust codes, so in this paper, the results will be shown in the framework of the conceptual apparatus of AMD codes. Notions of «algebraic manipulation» and «algebraic manipulation detection (AMD) codes» for the first time have been introduced in [7]. For constructing of the AMD codes various mathematical objects are used [9, 10].

The robust codes constructed by Karpovsky et al. in [4, 5, 11, 12] are the special class of weak AMD codes. Robust codes are AMD codes with deterministic encoding function. Deterministic encoding functions called one mapping of input values to code words of the some error-correcting code. This paper examines only deterministic encoding functions.

Fields of application of security-oriented codes are diverse and include communication channels with possible random errors, protect from age-related memory loss, secret sharing schemes and others. Some fault based side channel attacks are impossible or difficult to achieve, if the encoding device uses security-oriented codes [13–15]. These codes reduce significantly the number of undetected faults that can be exploited by an attacker.

The error masking probability is one of the main characteristic of the security-oriented codes; a number of undetected errors and complexity of encoding functions are also important characteristics. Since security-oriented codes are nonlinear, the coding rate is less than, for example, in linear codes. Accordingly, the complexity of the coding function is critical, since the more computational complexity is, the slower error detecting code will be.

This paper proposes an approach for analysis of coding functions in the case of non-uniform distribution of the input code words. The paper investigates the effect of the computational complexity of encoding functions on the probability of error masking for various distributions of input code words.

AMD codes and complexity of encoding function

AMD codes present the class of security-oriented codes. To define algebraic manipulation and AMD codes, it is necessary to consider the notion of an abstract storage device.

This device denotes as $\sum G$ and holds an element g from a finite abelian group G. An attacker is not able to obtain any information about the element g stored in the device $\sum G$. However, he can change the stored element g by adding another element $\delta \in G$. This tampering is called an algebraic manipulation. After algebraic manipulation, the abstract storage device $\sum G$ will store the value $g+\delta$. An adversary can choose the value $\delta \in G$ only based on what he had already known about g before it was stored in the device (a priory knowledge of g). AMD codes encode original information $s \in S$ as an element of $g \in G$ in such way that any algebraic manipulation is detected with high probability. Example of abstract storage device and algebraic manipulation is shown in Figure 1.

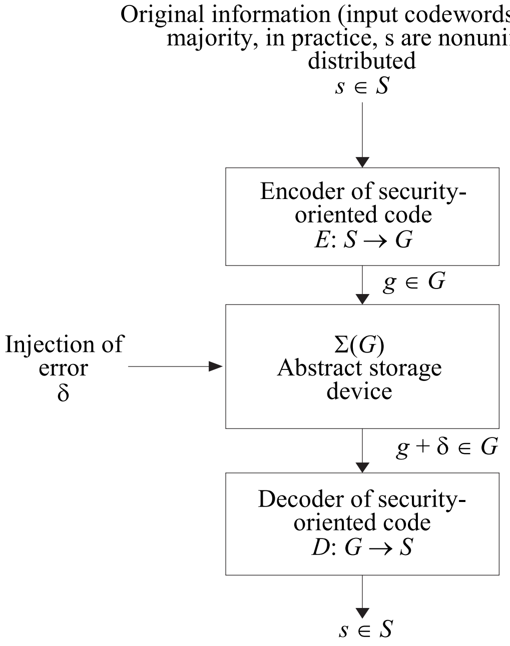
Original information (input codewords). In vast majority, in practice, s are nonuniform distributed
$s \in S$
Encoder of security-oriented code
$E: S \to G$
$g \in G$
Injection of error
δ
$\Sigma(G)$
Abstract storage device
$g+\delta \in G$
Decoder of security-oriented code
$D: G \to S$
$s \in S$

Fig. 1. Scheme of abstract storage device $\sum G$. An adversary can inject an error $\delta \in G$ and control distribution of $s \in S$, so it is necessary to arrange a protection based on security-oriented codes

In [7] Cramer et al. allocate weak and strong type of fault injection attack. In weak attack an adversary cannot choose the input and thereby control the output of system. So, from an adversary's point of view the source s is uniformly distributed and the attacker only can inject any specific error pattern $\delta \in G$ in the storage device $\sum G$, but he/she cannot change value s at their own discretion.

In strong attack the adversary can determine the outputs by choosing the inputs. In this case the adversary knows the value $s \in S$ and, moreover, he can choose it himself. In both types of fault injection attacks the value g stored in $\sum G$ is hidden from the attacker.

One of the main criteria for evaluating the effectiveness of AMD code is the error masking probability $Q(e)$. The error masking probability $Q(e)$ can be defined as

$$Q(e)=\frac{\{g \in C, g+e \in C\}}{M}, \qquad (1)$$

where C is the robust code, g is a codeword that belongs to the code C, e is an error injected by an adversary, M is the number of codewords in the code C. Error vector e can take all possible values of the group G, which also includes all the codewords g. In the case of weak attack model, calculation of error masking probability is performed simply. The distribution of occurrence probability of the input codewords is uniform, so $p(g_1) = p(g_2) = \cdots = p(g_M) = \frac{1}{M}$, where $p(g)$ denotes the occurrence probability of corresponding codewords g. The error masking probability for certain error e will be equal to the number of vectors g which satisfy the formula (1).

In a strong attack model input codewords can be controlled by an attacker and have a non-uniform distribution. Thus, for calculation of the error masking probability, it is necessary to add corresponding probability $p(g_i)$ to the current error masking probability $Q(e_j)$ whenever the condition $g_i + e_j \in C$, where i and j are the sequence numbers of vectors in group G. This approach of error masking probability allows not only measuring the maximum value of the error masking probability, but also getting the full probability distribution of error masking for each error e from group G.

Influence of encoding function complexity on distribution of error masking probability in the example of Maiorana-McFarland functions

The characteristics of the AMD codes depend on the properties of their encoding function This paper explores the influence of encoding function complexity on the error masking probability distribution at the example of comparison of functions $F(x, y) = xy$ and $F(x, y) = xy^{-1}$. In these functions parts x and y represent the two halves of the information part of the codeword. That is, the input vector s may be represented by concatenating of these two parts x and y so $s = (x, y)$. Accordingly, entire codeword includes the redundant part and has a form $g = (x, y, F(x, y))$, where $F(x, y)$ is a result of encoding function.

Let’s consider the behavior of the error masking probability of two encoding functions $F(x, y) = xy$ and $F(x, y) = xy^{-1}$. The second function is computationally more difficult than the first one, because it further comprises taking the multiplicative inverse in the field.

Distribution	$F(x, y) = xy$	$F(x, y) = xy^{-1}$
Uniform distribution	$\max Q(e) = 0{,}0625$	$\max Q(e) = 0{,}0625$
Binomial distribution	$\max Q(e) = 0{,}4987$	$\max Q(e) = 0{,}2227$
$p_1(g) = \begin{cases} \frac{0{,}8}{56}, & 51 \le g < 106 \\ \frac{0{,}2}{200}, & \textit{otherwise} \end{cases}$	$\max Q(e) = 0{,}2285$	$\max Q(e) = 0{,}1222$
$p_2(g) = \begin{cases} \frac{0{,}7}{100}, & 101 \le g < 200 \\ \frac{0{,}3}{156}, & \textit{otherwise} \end{cases}$	$\max Q(e) = 0{,}1222$	$\max Q(e) = 0{,}0917$
$p_3(g) = \begin{cases} \frac{0{,}1}{150}, & 1 \le g < 150 \\ \frac{0{,}9}{106}, & \textit{otherwise} \end{cases}$	$\max Q(e) = 0{,}1358$	$\max Q(e) = 0{,}1045$
$p_4(g) = \begin{cases} \frac{0{,}9}{30}, & 101 \le g < 130 \\ \frac{0{,}1}{226}, & \textit{otherwise} \end{cases}$	$\max Q(e) = 0{,}48$	$\max Q(e) = 0{,}1844$

Table. Comparison of the functions $F(x, y) = xy$ and $F(x, y) = xy^{-1}$ for various distribution and value $r = 4$. g denotes the codeword of code and $p(g)$ is a probability of codeword occurrence

For comparison it is necessary to generate codewords of error correcting codes with encoding function $F(x, y) = xy$ and $F(x, y) = xy^{-1}$, where x and y are, respectively, the first and second half of the information part. Both codes have a codeword length equal to $3r$, the number of codewords is equal to 2^{2r}, where r is the length of redundancy part. The results of measurement of error masking probability produced by the scripts in MatLab software for these functions are shown in the Table.

The Table results show that the code based on function $F(x,y)=xy^{-1}$ provides less maximum of error masking probability under nonuniform codeword distribution. The difference between the codes is clearer if we compare the codes under distribution and piecewise function number 4.

However, for the security-oriented code, not only the maximum value of the error masking probability is important, but also the distribution of the error masking probability. Let us consider the distribution of the error masking probability for code with encoding function $F(x,y)=xy$ for distribution of codeword.

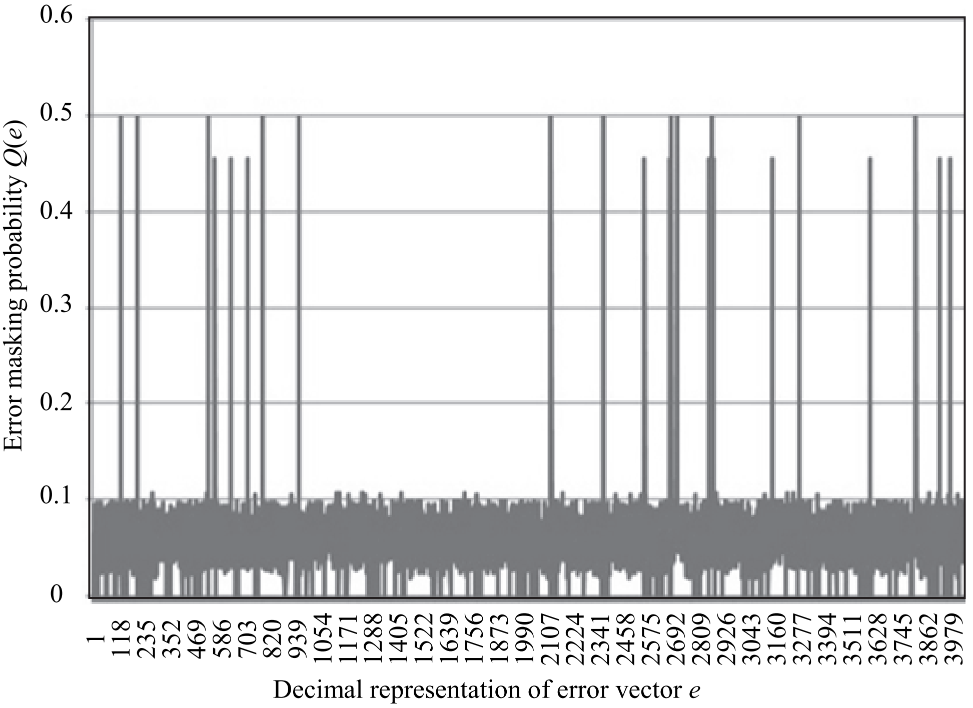

Fig. 2. The error masking probability of each error for code with encoding function $F(x,y)=xy$ ($r=4$) under binomial distribution. The ordinate shows the value of the masking probability $Q(e)$. Abscissa is a decimal representation of error vectors e

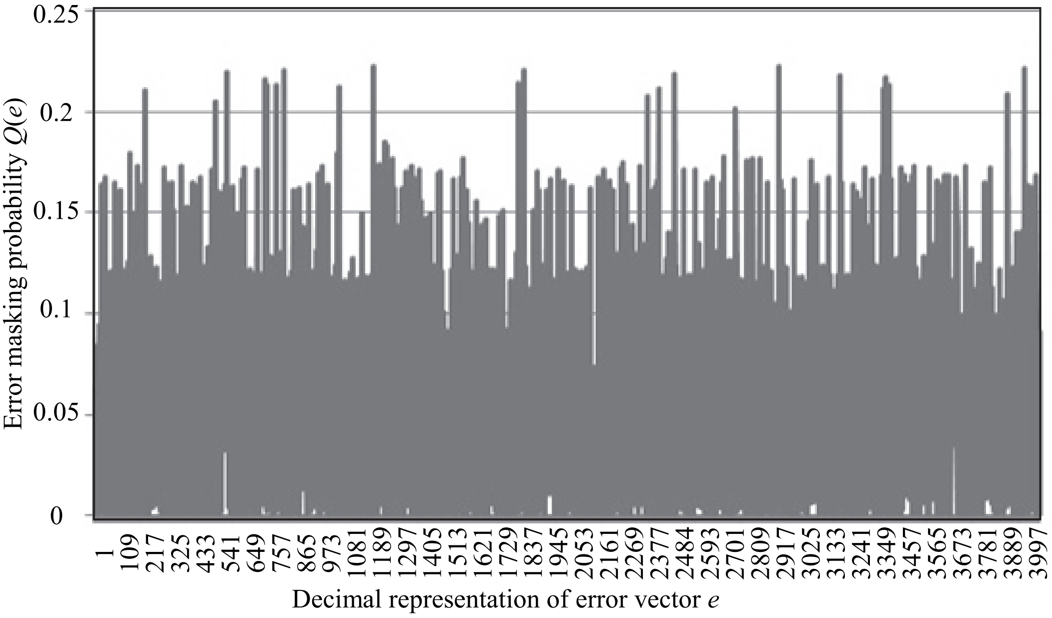

Fig. 3. The error masking probability of each error for code with encoding function $F(x,y)=xy^{-1}$ ($r=4$) under binomial distribution. The ordinate shows the value of the masking probability $Q(e)$. Abscissa is a decimal representation of error vectors e

It is clear from the Figure 2 that there is a huge difference between the maximum values of $Q(e) \approx 0.5$ and the average value over the whole distribution $Q(e) \approx 0.12$. As a result, errors corresponding to the maximum values are potentially dangerous to implement, since they have a higher chance of being missing by error-correcting codes. Let's analyze what happens with the distribution of the error masking probability in case of minor complication of encoding function the condition that coding function remains perfectly nonlinear.

In addition to reduction of the maximum of error masking probability, the difference between the maximum and average value of $Q(e)$ decreases. Maximum values of $Q(e)$ in the Figure 3 are not as distinct as in the Figure 2 that hampers their localization in error injection attack. The probability distribution in Figure 3 is more difficult to analyze by the attacker, because error masking probability values are close to each other.

Conclusion

This paper investigates the influence of encoding function complexity on the error masking probability distribution in the example of comparison of functions $F(x, y) = xy$ and $F(x, y) = xy^{-1}$. The paper assesses the maximum of error masking probabilities, the difference between the maximum and average values of $Q(e)$, and also the probability of the potential error injection. Assessment shows that even insignificant increase of the function complexity can improve the characteristics of AMD codes. The paper proposes an approach to measure the error masking probability for the case of nonuniform distribution of input codewords. This approach allows a detailed analysis of the error masking probability for each possible error.

References

1. Karpovsky M.G., Taubin A. New class of nonlinear systematic error detecting codes. *IEEE Transactions on Information Theory*, 2004, vol. 50, no. 8, pp. 1818–1820. doi: 10.1109/TIT.2004.831844
2. Kulikowski K.J., Karpovsky M.G., Taubin A. Fault attack resistant cryptographic hardware with uniform error detection. *Lecture Notes in Computer Science*, 2006, vol. 4236, pp. 185–195.
3. Kulikowski K.J., Karpovsky M.G., Taubin A. Robust codes and robust, fault tolerant architectures of the advanced encryption standard. *Journal of System Architecture*, 2007, vol. 53, pp. 138–149. doi: 10.1016/j.sysarc.2006.09.007
4. Ge S., Wang Z., Luo P., Karpovsky M. Reliable and secure memories based on algebraic manipulation detection codes and robust error correction. *Proc. 6th Int. Conf. on Dependability*. Barcelona, Spain, 2013.
5. Luo P., Lin A.Y.-L., Wang Z., Karpovsky M.G. Hardware implementation of secure shamir's secret sharing scheme. *Proc. IEEE 15th Int. Symposium on High-Assurance Systems Engineering*. Miami, USA, 2014, pp. 193–200. doi: 10.1109/HASE.2014.34
6. Wang Z., Karpovsky M.G., Kulikowski K.J. Design of memories with concurrent error detection and correction by non-linear SEC-DED codes. *Journal of Electronic Testing: Theory and Applications*, 2010, vol. 26, no. 5, pp. 559–580. doi: 10.1007/s10836-010-5168-5
7. Cramer R., Fehr S., Padro C. Algebraic manipulation detection codes. *Science China Mathematics*, 2013, vol. 56, no. 7, pp. 1349–1358. doi: 10.1007/s11425-013-4654-5
8. Cramer R., Dodis Y., Fehr S., Padró C., Wichs D. Detection of algebraic manipulation with applications to robust secret sharing and fuzzy extractors. *Lecture Notes in Computer Science*, 2008, vol. 4965, pp. 471–488. doi: 10.1007/978-3-540-78967-3_27
9. Wang Z., Karpovsky M.G. Reliable and secure memories based on algebraic manipulation correction codes. *Proc. 2012 IEEE 18th Int. On-line Testing Symposium*. Sitges, Spain, 2012, art. 6313861, pp. 146–149. doi: 10.1109/IOLTS.2012.6313861
10. Wang Z., Karpovsky M.G., Joshi A. Nonlinear multi-error correcting codes for reliable MLC nand flash memories. *IEEE Transactions on VLSI Systems*, 2012, vol. 20, no. 7, pp. 1221–1234. doi: 10.1109/TVLSI.2011.2157183
11. Jongsma E. *Algebraic Manipulation Detection Codes*. Bachelorscriptie, Universiteit Leiden, 2008.
12. Ge S., Wang Z., Luo P., Karpovsky M. Secure memories resistant to both random errors and fault injection attacks using nonlinear error correction codes. *Proc. 2nd Int. Workshop on Hardware and Architectural Support for Security and Privacy, HASP 2013*. Tel-Aviv, Israel, 2013, art. 5.
13. Shumsky I., Keren O., Karpovsky M. Robustness of security-oriented binary codes under non-uniform distribution of codewords. *Proc. 6th Int. Conf. on Dependability*. Barcelona, Spain, 2013.
14. Wang Z., Karpovsky M. New error detecting codes for design of hardware resistant to strong fault injection attacks. *Proc. Int. Conference on Security and Management*, *SAM*. Las-Vegas, USA, 2012.

15. Akdemir K.D., Wang Z., Karpovsky M.G., Sunar B. Design of cryptographic devices resilient to fault injection attacks using nonlinear robust codes. In: *Fault Analysis in Cryptography*. Eds. M. Joye, M. Tunstall. Springer, 2011, pp. 1036–1048. doi: 10.1007/978-3-642-29656-7

Alla B. Levina – PhD, Associate professor, ITMO University, Saint Petersburg, 197101, Russian Federation, alla_levina@mail.ru

Sergey V. Taranov – assistant, ITMO University, Saint Petersburg, 197101, Russian Federation, serg.tvc@mail.ru

Левина Алла Борисовна – кандидат технических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, alla_levina@mail.ru

Таранов Сергей Владимирович – ассистент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, serg.tvc@mail.ru