Министерство образования Республики Беларусь
Учреждение образования
«Белорусский государственный университет
информатики и радиоэлектроники»

ТЕЛЕКОММУНИКАЦИИ: СЕТИ И ТЕХНОЛОГИИ, АЛГЕБРАИЧЕСКОЕ КОДИРОВАНИЕ И БЕЗОПАСНОСТЬ ДАННЫХ

МАТЕРИАЛЫ МЕЖДУНАРОДНОГО НАУЧНО-ТЕХНИЧЕСКОГО СЕМИНАРА
(Республика Беларусь, Минск, ноябрь – декабрь 2020 г.)

TELECOMMUNICATIONS: NETWORKS AND TECHNOLOGIES, ALGEBRAIC CODING AND DATA SECURITY

Минск БГУИР 2020

УДК 654:004.056
ББК 32.88+32.972.5
Т31

Т31 **Телекоммуникации**: сети и технологии, алгебраическое кодирование и безопасность данных = Telecommunications: Networks and Technologies, Algebraic Coding and Data Security : материалы междунар. науч.-техн. семинара (Республика Беларусь, Минск, ноябрь – декабрь 2020 г.) / редкол. : М. Н. Бобов [и др.]. – Минск : БГУИР, 2020. – 88 с.

ISBN 978-985-543-607-3.

Сборник содержит статьи, тематика которых посвящена научно-теоретическим разработкам в области сетей телекоммуникаций, информационной безопасности, алгебраического кодирования и обработки изображений.

Предназначен для научных сотрудников в области телекоммуникаций, преподавателей, аспирантов, магистрантов и студентов технических вузов.

Научное издание

Корректор *В.В. Чепикова*
Ответственный за выпуск *В.Ю. Цветков*
Компьютерный дизайн и верстка *Е.Г. Макейчик*

Подписано в печать 07.12.2020. Формат 60×84 ⅛. Бумага офсетная. Гарнитура «Таймс».
Отпечатано на ризографе. Усл. печ. л. 10,46. Уч.-изд. л. 8,7. Тираж 38 экз. Заказ 241.

Издатель и полиграфическое исполнение: учреждение образования
«Белорусский государственный университет информатики и радиоэлектроники».
Свидетельство о государственной регистрации издателя, изготовителя,
распространителя печатных изданий №1/238 от 24.03.2014,
№2/113 от 07.04.2014, №3/615 от 07.04.2014.
Ул. П. Бровки, 6, 220013, Минск

ISBN 978-985-543-607-3

СОДЕРЖАНИЕ

CONTENTS

УДК 621.391

MULTIPLE SEEDED REGION GROWING ALGORITHM FOR IMAGE SEGMENTATION USING LOCAL EXTREMA

A.T. NGUYEN, X.L. DAI, V.Yu. TSVIATKOU

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Submitted 28 October 2020

Abstract. In this paper, a multiple seeded region growing technique for image segmentation is presented. Conventional image segmentation techniques using region growing requires initial seeds selection, which increases computational cost and execution time. To overcome this problem, a seeded region growing technique for image segmentation is proposed, which starts from searching for local extrema of the image using morphology as the initial seeds, whose coordinates are saved in a pair of static FIFO queues, used for wave region growing. It grows regions according to the extreme values quasi-parallel. We use intensity based similarity index for the grow regions and adaptive threshold is used to calculate the criteria for the grow new waves. We apply the proposed algorithm to the Berkley segmentation dataset and discuss results based on F and $SSIM$ evaluation functions that show efficient segmentation.

Keywords: local extrema, image segmentation, region growing, seeded region growing, evaluation functions.

Introduction

Image segmentation is the basic requirement of any computer vision application because people are generally interested only in certain parts of the image. Image segmentation results in non-overlapping objects labeled with different region numbers. It should be noticed that no general technique has been developed yet to segment an image precisely, so different techniques are taking floor to perform segmentation [1].

Segmentation is used to detect the location of objects and boundaries in the tasks of visualizing medical images, searching, classifying, recognizing and tracking objects in images. This leads to the division of the image into areas that correspond to different objects or their parts [2, 3].

Segmentation accuracy determines the quality of the subsequent processing results. In some cases, the segmentation time may be limited, or it is necessary to control the number of image segments [4]. Segmentation errors are manifested in the accuracy and stability of the localization of regions when changing the conditions of video recording [5]. The main cause of errors in segmentation methods in real conditions is the uneven illumination of the scene, which arises due to the instability of the light source, uneven distribution of light over the surface of the object (especially large), and the inability to optically isolate the object from the shadow of other objects [6].

Threshold based image segmentation techniques discriminate regions on the basis of intensity value difference between pixels. Survey paper [7] shows analysis and comparison of various threshold based segmentation techniques. Thresholds for image segmentation have been calculated based on maximum entropy [8], interclass variation [9], histogram [10, 11]. The limitation of threshold based segmentation technique is that it performs well for images, which have only two components. For complex images, it is calculated to support further processes [12].

Texture describes the spatial distribution of gray intensity in the whole image. It provides a more accurate analysis of correlation, variance, and entropy at a lower level. Textures from an image have been calculated often with co-occurrence matrix and semi-variogram [13–15]. It is complicated to extract texture from low contrast or noisy images.

Clustering is an approach in which pixels are classified to a cluster, which is closest among all clusters. Pixels having homogeneous characteristics belong to the same cluster and different with respect to pixels of other clusters. The pixels must follow the homogeneity criteria in the same cluster. To perform clustering based segmentation, [16] present K-mean, [17] use LVQ efficiently. Fuzzy logic based Fuzzy C-Mean clustering method introduces fuzzy membership to pixels with respect to every cluster [18]. In cluster based image segmentation techniques, it is necessary to choose a certain number of clusters initially which eventually reduces the dynamicity of the technique.

Region splitting and merging techniques [19–21] starts with splitting an image into small regions and continued till regions with required degree of homogeneity are formed. Splitting phase impacts the overall segmentation of the image. This phase results in over segmented image which is followed by the merging phase. Thus, these techniques of region splitting and merging are complex and time consuming. The main objective of region growing is to map individual pixels called seeds in input image to a set of pixels called region. The original Seeded Region Growing (OSRG) [22] does not impose any constraint or restriction on the shape or boundary of the region, the new variant stabilized seeded region growing (SSRG) [23] is termed to prevent the leakage problem when the signal-to-noise ratio is low, the final segmented boundaries could be very rough even though if the true boundaries were smooth. Region growing method starts with initial seeds and grows with neighboring homogenous elements. Seed may be pixel or region. Due to its efficient results for realistic images, it is used widely in different manners. In [24] a region growing method based on the gradients and variances along and inside of the boundary curve is used. In [25] edge and smoothness factors as criterion to determine initial seed pixels are used and then seeded region growing method is used to segment images based on seed regions.

In the seed based region growing method, selection of initial seed is crucial because it decides the overall segmentation by region growing technique. To select initial seeds, the images can be first partitioned into a set of rectangular regions with fixed size and a simple automatic SRG algorithm can then be realized by selecting the centers of these rectangular regions as the seeds (RSRG) [26]. In Level Set based SRG (LSSRG) [27] base points are iteratively selected. A point has a higher likelihood of being a base point if it has smaller (with respect to a global maximum) gradient and variance values. Ideally, a base point should be at the center of the segment that it belongs to. To select initial seed watershed algorithm [28] used to first segment image to calculate no overlapped regions and then use centroid of region as initial seeds. Algorithm in [29] found out initial seed by applying edge based segmentation and then use centroid as initial seeds. Algorithm in [30] adopt the Harris corner detector to calculate initial seed. But seed selection affected by particular technique limitation and increases the computation overhead.

In the proposed algorithm, we start with local extreme pixels of the image as initial seed points. A pair of static FIFO queues with image size is used for saving seed points coordinates and region growing to decrease computational resources and increase algorithm speed. Then region growing is done according to grow adaptive threshold which follows the stopping criteria to start the new wave growing. We use Berkley segmentation database [31] which provide an empirical basis for research on image segmentation and boundary detection.

Research method

Seeded region growing method

Segmentation is a process of extracting required features or Region of Interest (ROI) from an image for future purpose like compression. The given or input image is sliced into multiple regions based on some properties like pixel intensity, texture, position or some local (or) global statistical parameter. Seeded Region Growing (SRG) method takes a set of seeds as input along with the image and it requires seeds as additional input. The seeds mark each of the objects to be segmented and compare with pixel value. The pixel with the smallest difference measured is allocated to the respective region the difference between a pixel's intensity value and the region's mean, is used as measures of similarity, this process continues until all pixels are allocated to a region [22, 23, 32]. The algorithm procedure is as follows:

Step 1. We start with a number of seed points which have been clustered into N clusters, called $R_1, R_2 ... R_N$. And the position of initial seed points is set as $P_1, P_2 ... P_N$.

Step 2. To compute the difference of pixel value of the initial seed point P_k and its neighboring points (y_N, x_N) ($|I(y_N, x_N) - I(P_k)|$), if the difference is smaller than the threshold (criterion σ_{SRG}) we define ($|I(y_N, x_N) - I(P_k)| \le \sigma_{SRG}$), the neighboring point (y_N, x_N) could be classified into R_k, where $k = \overline{1, N}$. For each set R_k, compute the value of the homogeneity criterion σ_{SRG} for all its immediate, unlabeled neighbors. The criterion σ_{SRG} can be any of σ_O [22], σ_S [23], σ_R [26], σ_{LS} [27].

Step 3. Recomputed the boundary of R_k and set those boundary points (y_N, x_N) as new seed points P_k. In addition, the mean pixel values of R_k have to be recomputed correspondingly.

Step 4. Repeat Step 2 and 3 until all pixels in image have been allocated to a suitable region.

Criterion selection

In [22] the criterion $\sigma(y, x)$ is defined to be a measure of how different neighbor unlabeled pixel (y, x) of the region H is from the region it adjoins. The simplest definition for $\sigma(y, x)$ is

$$\sigma(y, x, R_k) = \left| I(y, x) - \overline{I(R_k)} \right|, \tag{1}$$

where $I(y, x)$ is the gray value of the image point, $\overline{I(R_k)} = \frac{1}{N_k} \sum_{j=1}^{N_k} I(y_j, x_j)$ is a mean value of the region R_k, and $\sigma(y, x)$ is minimized

$$\sigma_O = \min_{(y, x) \in H} \left\{ \sigma(y, x, R_k) \mid k \in [1, N] \right\}. \tag{2}$$

In [23] the value $\sigma(y, x)$ is defined to be a measure of how different neighbor unlabeled pixels (y_u, x_v) in window size $(2L+1) \times (2L+1)$ of region H is from the region R_k it adjoins. The simplest definitions for $\sigma(y, x)$ and the criterion σ_S are

$$\sigma(y, x, R_k, L) = \frac{1}{(2L+1)^2} \left\{ \sum_{(u, v)=-L}^{L} \left| I(y_u, x_v) - \overline{I(R_k)} \right| \right\}. \tag{3}$$

$$\sigma_S = \min_{(y, x) \in H} \left\{ \sigma(y, x, R_k, L) \mid k \in [1, N] \right\}. \tag{4}$$

In [26] the value $D(y, x)$ is defined to be a measure of how different neighbor unlabeled pixels $(y \pm 1, x \pm 1)$ of the region H is from the region R_k it adjoins. The simplest definitions for $D(y, x)$ and the criterion σ_R are

$$D(y, x, R_k) = \sum_{(y \pm 1, x \pm 1) \in H} \left| I(y, x) - I(y \pm 1, x \pm 1) \right|, \tag{5}$$

$$\sigma_R = \min_{(y \pm 1, x \pm 1) \in H} \left\{ D(y, x, R_k) \mid k \in [1, N] \right\}. \tag{6}$$

In [27] the criterion σ_S depends on image bit depth m is

$$\sigma_{LS} = 3 \times 2^m / 64. \tag{7}$$

In this paper we propose a criterion $\sigma(y, x)$, which is defined to be a measure of how different neighbor unlabeled pixel (y, x) of the region H is from the current extreme pixel P_k of the region R_k it adjoins. The simplest definitions for $\sigma(y, x)$ and σ_{WG} are

$$\sigma\left(y, x, R_k\right) = \left|I\left(y, x\right) - I\left(P_k\right)\right|, \tag{8}$$

$$\sigma_{WG} = \min_{(y,\ x) \in H} \left\{\sigma\left(y, x, R_k\right) | \ k \in \left[1, N\right]\right\}, \tag{9}$$

where $I\left(P_k\right)$ is the gray value of the initial extreme point.

Proposed segmentation method

Seed selection is the first step of the region growing technique. Instead of selecting seeds initially we select extreme pixels (maxima and minima) of the image as initial seeds [32–35]. The proposed algorithm is executed as described in pseudo code. Pseudo code uses following variables:

N : number of all local extrema.

PG : static FIFO stack to store initial seed points and pixels to grow with same size of image.

NP : number of labeled pixels in FIFO stack PG.

NB : number of current labeled border pixels in FIFO stack PG.

PE : stack to store N intensity values of local extrema of the image.

REG : segmented matrix with same size of image I, storing the labels of grown region.

$CP_{8\text{-}nb}\left(j\right)$: 8-neighbours of current border pixel CP, where $j = \overline{1, 8}$.

σ_{WG} : region growing threshold (criterion).

PSEUDOCODE

Region_Growing(Gray image I)

$\sigma_{WG} = 0$, $NB = 1$, $NP = N$

Step 1: (region growing)

$NB = 0$

While $k \le NP$

$CP = PG\left(k\right)$, $NS = REG\left(CP\right)$, $EXT = PE\left(NS\right)$

For (8-nb of CP, $j = \overline{1, 8}$)

If ($REG\left(CP_{8-nb}\left(j\right)\right)$ not labeled)

If ($abs\left(EXT - I\left(CP_{8-nb}\left(j\right)\right)\right) \le \sigma_{WG}$)

$REG\left(CP_{8-nb}\left(j\right)\right) = NS$

$NP = NP + 1$

Else

$\sigma = \min\left\{abs\left(EXT - I\left(CP_{8-nb}\left(j\right)\right)\right)\right\}$

End if

End if

End for

If (one of 8-nb of CP not labeled)

$NB = NB + 1$, $PG\left(NB\right) = CP$

End if

$k = k + 1$

End while

Step 2: (starting a new wave)

While $NB > 0$

$k = 1,\ NP = NB,\ \sigma_{WR} = \sigma$

Go to step 1

End while.

Segmentation Evaluation Approach

We propose two unsupervised evaluation methods based on F [36, 37] and $SSIM$ [38] evaluation functions. F measures the average squared color error of the segments, penalizing over-segmentation. The structural similarity index $SSIM$ is used to determine the similarity of two images and is formed as a result of their comparison in terms of brightness, contrast and structure. Suppose a digital image I has been segmented into N regions, denoted as R_k, $k = \overline{1, N}$. For region R_k, denote its area (measured by the number of pixels) as $S_k = |R_k|$. The generalized F and $SSIM$ evaluation functions are defined as

$$F(I) = \sqrt{N} \times \sum_{k=1}^{N} e_k^2 \Big/ \sqrt{S_k}, \tag{10}$$

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \tag{11}$$

where $e_k^2 = \sum_{(y,x)\in R_k} \left(I(y,x) - \overline{I(R_k)}\right)^2$ is the squared color error, S_k is the number of pixels from region R_k, μ_x and μ_y are mean intensities, σ_x and σ_y are standard deviations, σ_{xy} is correlation coefficient of two grayscale images x and y, $C_1 = (k_1 L)^2$, $C_2 = (k_2 L)^2$, $L = 255$, $k_1 = 0.01$, $k_2 = 0.03$. Smaller F or higher $SSIM$ indicates better segmentation results.

Results and analysis

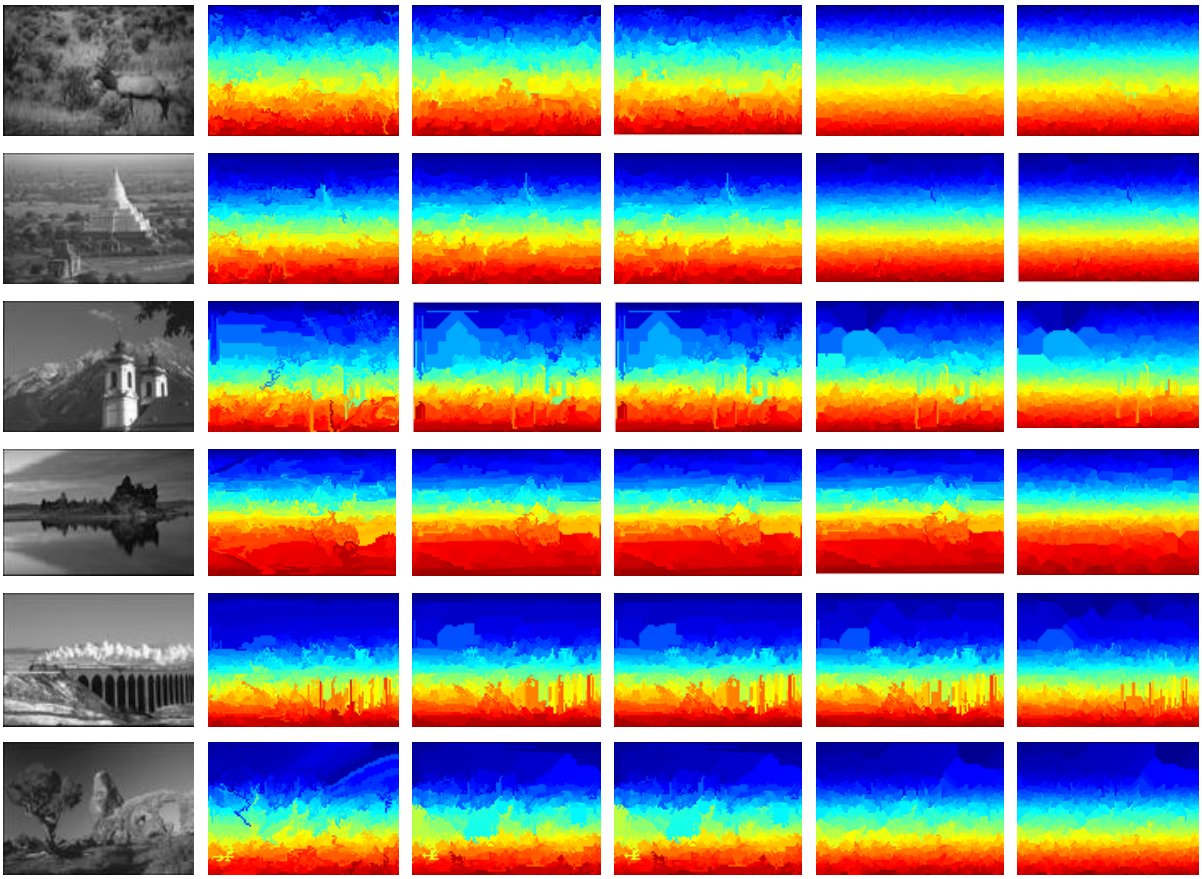
To examine the efficiency of our method we use six grayscale images from the Berkley segmentation database [31] shown in Fig. 1 *a*. The images are of size 120×80. Our algorithm takes approx 35 ms on system configured with Intel processor Core i3 2,3 GHz and 6 Gigabyte of RAM. We use Matlab 2015a tool to implement our method and others. The results after applying our proposed method on these images are shown in Fig. 1 *b*. These results are obtained by converting the region matrix containing labeled regions to an RGB image. We also compare our results with four algorithms [22, 23, 26, 27] for the initial seeds selection and the results are shown in Fig. 1 *c–f*.

The local maxima or local minima in this paper are selected as base points using mathematical morphology for automatic seeded region growing algorithm [32, 33]. Finding Local Extrema (LE) is often solved by mathematical morphology using dilation and erosion operations, respectively. It gives accurate results compared to block algorithms. However, the morphological algorithm has high computational complexity, which is associated with separate processing of maxima and minima, as well as iterative processing of the neighborhoods of all pixels. In this proposed system we developed two algorithms for extracting local extrema in grayscale images with low computational complexity, high accuracy and less memory [34, 35].

The average processing times of all algorithms for 100 grayscale images of size 120×80 using Berkley segmentation database [31], are shown as in Table 1 that our algorithm is fast compared to others [22, 23, 26]. The processing speed of the proposed algorithm is faster when implemented in C++ programming language. We evaluate the F (10) and $SSIM$ (11) for all images for all techniques. The results of comparison of the proposed method with the other techniques are given in Table 2. It is observed from Table 2 that the Liu's F is lower and $SSIM$ is higher for our method's results as compared to other segmentation algorithms.

Table 1. **Segmentation speed using Berkeley segmentation database**

Average speed	Using proposed	Using σ_O [22]	Using σ_S [23]	Using σ_R [26]	Using σ_{LS} [27]
$T(\sigma_{SRG})$, ms	35	60	220	38	29



a *b* *c* *d* *e* *f*

Fig. 1. Test images and results of segmentation: *a* – Original images; *b* – Segmented images using proposed algorithm; *c* – Segmented images using criterion σ_O; *d* – Segmented images using criterion σ_S; *e* – Segmented images using criterion σ_R; *f* – Segmented images using criterion σ_{LS}

Table 2. **F and SSIM evaluation functions**

No.	Image	Evaluation function	Using σ_{WG} (proposed)	Using σ_O [22]	Using σ_S, $L=1$ [23]	Using σ_R [26]	Using σ_{LS} [27]
1	Test1	*F*	$2,3223\times10^6$	$3,6731\times10^6$	$3,6734\times10^6$	$7,8243\times10^6$	$5,4289\times10^6$
		SSIM	0,9307	0,8809	0,8809	0,7902	0,8444
2	Test2	*F*	$1,6193\times10^6$	$2,5145\times10^6$	$2,5145\times10^6$	$4,5029\times10^6$	$4,5029\times10^6$
		SSIM	0,9401	0,8945	0,8945	0,8172	0,8172
3	Test3	*F*	$2,4755\times10^6$	$3,4630\times10^6$	$3,4630\times10^6$	$3,6331\times10^6$	$4,1901\times10^6$
		SSIM	0,8944	0,7808	0,7808	0,7676	0,7194
4	Test4	*F*	$1,1244\times10^6$	$2,1140\times10^6$	$2,1140\times10^6$	$2,1140\times10^6$	$3,2402\times10^6$
		SSIM	0,9025	0,7912	0,7912	0,7912	0,7040
5	Test5	*F*	$3,3733\times10^6$	$4,9943\times10^6$	$4,9943\times10^6$	$5,0785\times10^6$	$5,9831\times10^6$
		SSIM	0,9031	0,8110	0,8110	0,7892	0,7745
6	Test6	*F*	$1,4155\times10^6$	$2,4536\times10^6$	$2,4536\times10^6$	$2,8852\times10^6$	$3,2387\times10^6$
		SSIM	0,8964	0,7692	0,7692	0,7821	0,7779

Conclusion

A new approach to segment an image using a multiple seed based region growing algorithm has been proposed in this paper. In this method the extreme pixels of the image are selected as the initial seeds and the region is grown according to growing formula with the stopping criterion determined by adaptive threshold technique around the local extrema. The segmented result obtained by the proposed method is compared to other criterions of SRG algorithms [22, 23, 26, 27] and is observed to have lower F, higher $SSIM$ values and fast processing.

References

1. Verma O.P. [et al.] // International Conference on Communication Systems and Network Technologies. 2011. P. 500–503.
2. Chijindu V.C., Inyiama H.C., Uzedhe G. // African Journal of Computing & ICT. 2012. Vol. 5. P. 90–98.
3. Shivhare Gupta V. // International Journal of Engineering and Advanced Technology. 2015. Vol. 4. P. 153–157.
4. Gonzales R., Woods R., Eddins S. // Technosphere. 2006. P. 396–443.
5. Fabijańska A., Strzecha K., Sankowski D. // CADSM'2007, Polyana, Ukraine, 20–24 February. 2007. P. 439–441.
6. Chandrakala M., Devi P.D. // International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE). 2016. Vol. 5. P. 163–168.
7. Chang C.I. [et al.] // IEEE Proceedings Vision, Image and Signal Processing. 2007. Vol. 153. P. 837–850.
8. Kapur J.N., Sahoo P.K., Wong A.K.C. // Graphical Models and Image Processing. 1985. Vol. 29. P. 273–285.
9. Radhika K.R., Sekhar G.N., Venkatesha M.K. // International Conference on Multimedia Computing and Systems. 2009. P. 216–221.
10. Tian Junwei, Huang Yongxuan // IEEE International Symposium on Industrial Electronics. 2007. P. 1623–1627.
11. Zhenhua Zhang, Wenhui Li, Bo Li // International Conference on Information Assurance and Security. 2009. Vol. 2. P. 381–384.
12. Justice R.K., Stokely E.M. // IEEE International Conference on Bridging Disciplines for Biomedicine. 1996. Vol.3. P. 1083–1084.
13. Gambino O. [et al.] // IEEE Trans. Complex, Intelligent and Software Intensive Systems. 2010. P. 146–152.
14. Jie Wu [et al.] // International Conference on BioMedical Engineering and Informatics. 2008. Vol. 2. P. 263–267.
15. Zhy Chang-ming [et al.] // International Conference on Computer Science and Software Engineering. 2008. Vol. 1. P. 795–798.
16. Patill R.V., Jondhle K.C. // International Conference on Computer Science and Information Technology. 2010. Vol. 2. P. 117–121.
17. Hariadi M. [et al.] // Asia-Pacific Conference on Circuits and Systems. 2002. Vol. 2. P. 171–176.
18. Ock-Kyung Yoon [et al.] // IEEE International conference on Fuzzy Systems. 1999. Vol. 2. P. 853–857.
19. Jianping Fan [et al.] // IEEE Transaction on Image Processing. 2001. Vol. 10. P. 1454–1466.
20. Liu L., Sclaroff S. // IEEE International Conference on Computer Vision. 2001. Vol. 4. P. 98–104.
21. Kelkar D., Gupta S. // International Conference on Emerging Trends in Engineering and Technology. 2008. P. 44– 47.
22. Adams R., Bischof L // IEEE Trans. Pattern Anal. Mach. Intelligence. 1994. Vol. 16 (6). P. 641–647.
23. Fan M., Lee Thomas C.M. // Image Processing IET. 2015. Vol. 9(6). P. 478–485.
24. Deng Wankai [et al.] // International Conference on Biomedical Engineering and Informatics. 2010. Vol. 1. P. 393–396.
25. Chaobing Huang, Quan Liu, Xiaopeng L. // International Conference on Fuzzy Systems and Knowledge Discovery. 2010. Vol. 2. P. 533–536.
26. Fan J. [et al.] // Pattern recognition letters. 2005. Vol. 26(8). P. 139–1156.
27. Porikli F.M. // In Image Processing: Algorithms and Systems III. 2005. Vol. 5298. P. 536–543.
28. Jun Tang // International Conference on Computer Engineering and Technology. 2010. Vol 6. P. 634–637.
29. Fan Jianping [et al.] // IEEE Transaction on Image Processing. 2001. Vol. 10. P. 1454–1466.
30. Weihong Cui, Zequn Guan, Zhiyi Zhang // International conference on Computer Science and Software Engineering. 2008. Vol. 6. P. 93–96.
31. The Berkeley Segmentation Database and Benchmark: [Electronic resource]. URL: https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/
32. Sharma R., Sharma R. // International Journal of Innovative Research in Computer and Communication Engineering. 2014. Vol. 2(9). P. 5686–5693.
33. Soille P. Morphological Image Analysis: Principles and Applications. Springer, 2002.

34. Nguyen A.T., Tsviatkou V.Yu. // International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE). 2019. Vol. 8. P. 1–10.
35. Nguyen A.T., Tsviatkou V.Yu. // Informatics. 2019. Vol. 16(3). P. 23–36.
36. Zhang H., Fritts J.E., Goldman S.A. // Computer Vision and Image Understanding. 2008. Vol. 110(2). P. 260–280.
37. Liu J, Yang Y.H. // IEEE Trans. Pattern Anal. Mach. Intell. 1994. Vol. 16(7). P. 689–700.
38. Wang Zhou [et al.] // IEEE Transactions on Image Processing. 2004. Vol. 13(4). P. 600–612.

UDC 004.031.43:004.75

AUTOMATION TOOLS FOR THE DEVELOPMENT OF INTERNET OF THING NETWORKS

U.A. VISHNYAKOU, B.H. SHAYA, A.H. AL-MASRI, S.H. AL-HAJI

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Submitted 4 November 2020

Abstract. The analysis of the state and development of platforms for creating IoT networks is presented. The structure and purpose of the IoT platform components are considered. Features of the IoT platform with big data are given. Characteristics of the most common global platforms for building IoT networks and the features of the EAEU platform market are discussed. An algorithm for network modeling based on the AWS IoT platform is presented.

Keywords: Internet of things, IoT networks, IoT platform components, big data, IoT network modeling.

Introduction

With the continued development of large and small networks in Infocommunication (WAN, LAN) intended for the Internet of People (IoP), a variety of Internet of Things (IoT) networks are becoming more widespread [1]. IoT is a set of embedded systems, networks of wireless sensors, control systems and automation tools for processing information received from sensors. IoT networks allow a user to implement automation of production processes, management of transport, energy, agriculture, medicine, and create smart stores, smart homes, districts, and cities at a new level.

IoT is a concept of the network infrastructure development (physical basis) online, in which «smart» things without human intervention are able [2]:

− to connect to the network for remote interaction with other devices (Thing-Thing);

− to interact or interaction with autonomous or cloud data processing centers (Thing-Web Objects) for data transmission, storage, processing, analysis and management decisions aimed at changing the environment;

− to interact with user terminals (Thing-User) for the control and management of these devices.

This requires not only the development of scientific research and industrial development, but also the improvement of personnel training in this area [3].

IoT development automation

To automate the creation of IoT systems, leading global companies have developed design tools in the form of IoT platforms [4]. IoT platforms are becoming the central backbone of IoT deployments. The IoT platform market will reach $ 22,3 billion by 2023 year [5]. Due to the large volume of information received from IoT sensors, such platforms provide analytical tools for processing Big Data. Let's look at these questions in more detail.

In may 2020, the analytical company Counterpoint Research named the leading platforms for creating IoT networks and applications in terms of versatility (to meet the needs of users) and other parameters. The first position was taken by the Microsoft Azure platform, the second position − by Amazon Web Services (AWS), the third place − by Huawei OceanConnect, the fourth − by PTC ThingWorx, and the fifth place-by IBM Watson [4, 5]. This study took into account eight components: distribution, growth rate, integration and scaling capabilities, application support, cloud components, peripheral interaction, device data processing, and peripheral components [5].

An IoT platform is a hardware and software system for connecting end devices (sensors, sensors, etc.) to a cloud environment. The purpose of the IoT platform is to provide wireless communication of peripherals (sensors, devices), through additional interfaces (gateways) and communication protocols, as well as storage, processing and data mining [4]. The network structure using the platform includes end devices (sensors), a cloud platform, and applications for monitoring sensor information. The typical structure of the IoT platform includes the following components [6, 7]:

1. Communication: combines special protocols and data structures into a single software interface that provides accurate streaming of information and interaction with peripheral devices.

2. Device management: supports the work of connected sensors, the correction and updates for the software running on devices or boundary gateways.

3. Database: stores data in special formats transmitted from devices; supports requirements for hybrid cloud databases for a number of parameters (volume, speed, verification).

4. Event management: analyzes data using rule-based knowledge of events-actions that allow you to make decisions based on specific information received from sensors.

5. Analytics: performs complex analyses from basic data clustering and deep machine learning to predictive Analytics, extracting maximum value from the IoT data stream.

6. Visualization: allows users to identify patterns and observe trends using visualization dashboards, where data is depicted using linear or pie charts, 2D or 3D models.

7. Tools: allow IoT developers to prototype, test, and promote IoT network options, creating platform ecosystem applications for visualizing, managing, and controlling connected devices.

8. External interfaces: enable integration with other systems and part of the broader it ecosystem through embedded application programming interfaces (APIs), software development kits (SDKs), and gateways.

IoT platform analysis

Let's look at the most well-known platforms from the world's leading IT-companies [8, 9]. The Microsoft Azure IoT Platform supports testing a user-developed network variant by modeling, as well as the ability to design new network solutions that meet the specifics of an original project. This platform has information security tools, the ability to expand and integrate with existing or planned systems. The platform allows a user to connect multiple devices from different manufacturers, to process data received from sensors, connect analytics tools and generate appropriate reports. So, it uses the information obtained from sensors for subsequent machine learning process. Let's consider the main features of other popular IoT platforms [9].

AWS IoT Core (Amazon) is a platform on which a user can create local networks or IoT applications. It supports special communication protocols, including custom ones, which allows a user to communicate between devices from different manufacturers. The AWS IoT Device Management platform component supports adding and managing external devices, monitoring and configuring them, and updating their operation. The WS IoT Analytics platform component implements automatic processing and analytical calculations of large amounts of data from various IoT devices and sensors. The AWS IoT Device Defender platform component supports configuring information security tools for IoT networks (authentication, encryption). It allows a user to create and adjust security policies, manage device authentication and authorization, provide a private transmission channel (encryption).

The Google Cloud IoT platform includes a number of components that a designer can use to create new IoT networks. This platform includes: Cloud IoT Core service for securely connecting, managing and retrieving data from peripherals; Cloud Pub/Sub service for processing event data and implementing analytical flow processing; Cloud Machine Learning Engine service for creating a machine learning model using data received from IoT sensors.

The Cisco IoT platform services additionally support voice and data transmission, enable the development of various IoT applications and perform many others functions to generate revenue from the project. By choosing the Cisco platform to host an IoT application, the user gets a set of convenient sensor management and monitoring functions, as well as advanced security measures, including identification of devices with the platform, use closed transmission channels, etc. All these tools are with the operation of mobile applications and remote interaction with the consumer. A number of additional services allow a user to implement other functions. For example, a user can take IoT services

for engineering networks development, when project systems intended for use by utilities to monitor and control the relevant sensors. The IoT Advisory service provides access to expert advice on the main business tasks in the field of the Internet of things.

The cloud platform of the German company SAP has all the necessary components for creating and managing IoT network applications. Remote devices can be connected either directly or via a cloud service. Advanced Analytics tools enable a client to receive, process, analyze, and explore Big Data received from sensors, meters, and other devices in IoT networks. Following the latest technological trends, the SAP IoT platform provides the ability to use IoT data to create custom applications with elements of artificial intelligence and machine learning.

More than half of all market projects in Belarus and Russia are developed by domestic companies from the EAEU countries − Belarus, Russia and Kazakhstan (51 %). The second supplier of IoT platforms to the EAEU market is the United States (23 %). The analysis of IoT platforms by industry reflects the overall market dynamics. Most competitive segments on this market are transport and logistics (42 %) and Smart city (32 %) projects [10].

Modeling an IoT network based on the Amazon platform

The typical network on base of IoT platforms includes:
– sensors, devices sending control information;
– IoT network gateway for converting the format of transmitted information;
– authentication and sensor management tools;
– cloud service components (infrastructure, platform blocks, etc.);
– user interfaces;
– user additional applications.

Let's consider an algorithm for modeling the Internet of things network using the IoT platform. Amazon IoT platform create services that meet multiple user requirements, and as a result, you can build a network quickly. This cloud platform has a significant advantage – the ability to independently model the network in a short time, without involving the corporate it service and additional information security tools. The generalized algorithm for creating a network on the AWS platform looks like this:

1. Sensors measure the parameters of processes (devices) that interact with the IoT platform using development tools (SDKs).

2. Devices send messages that are verified by the platform's authentication and authorization service. If the verification is not successful, the device IDs must be corrected.

3. Information from devices is sent to the gateway (Device Gateway), and various network protocols can be used. When information is converted in the gateway, it is sent to the rules processing unit (communication with Analytics) and in parallel to the Device storage unit (Device Shadows).

4. Device Shadows stores the current state of network peripherals for continuous access by software applications. If there is no connection to an individual device in the network, the Device Shadow block executes commands from applications, and when the connection is restored, it synchronizes the current state with the device.

5. Depending on the nature of the received data, the rule handler performs (programmed) actions: saves data in the database, sends SMS or e-mail information to the network Manager about their receipt, calls the HTTP API, sends data to the Analytics system, and so on.

6. Applications use this data to control devices by using the AWS API (application programming interface).

7. Information about all devices is stored on the AWS IoT platform.

Conclusion

The article considers the relevance of the development of Internet of things networks for many sectors of the economy, as evidenced by the analysis of the world, Russian and EAEU markets. IoT platforms are used to automate development. An analysis of their development is given, and the structure of such IoT platforms is discussed. The features of working with big data as part of IoT platforms due to the huge amount of information coming from sensors and devices are discussed. The features of the

most popular IoT platforms in the world are considered. An algorithm for network modeling based on the AWS IoT platform is presented.

References

1. Roslyakov A.V. Internet of things: textbook manual. Samara, Pgutii, 2015.
2. Tkachenko V. IoT – modern telecommunication technologies. [Electronic resource]. URL: http://www.lessons-tva.info/articles/net/013.html.
3. Vishnyakou U.A. // Bulletin of the connection. 2020. No. 3. P. 56–59.
4. IoT platform. [Electronic resource]. URL: https://iot.ru/wiki/iot-platforma.
5. IoT Platforms. [Electronic resource]. URL: http://www.tadviser.ru/index.php/.
6. Things to know about the IoT Platform ecosystem. [Electronic resource]. URL: https://iot-analytics.com/5-things-know-about-iot-platform.
7. Big Data-clouds of the Internet of things: what are IoT platforms and why are they. [Electronic resource]. URL: https://www.bigdataschool.ru/blog/iot-platform-big-data-cloud.html.
8. How the Internet of things uses Big Data. [Electronic resource]. URL: https://www.bigdataschool.ru/bigdata/iot-architecture-big-data.html.
9. Browse the best IoT platform in 2019. Tips for choosing a cloud solution. [Electronic resource]. URL: https://www.edsson.com/ru/blog/article?id=iot-platforms.
10. Technouklad company published a new study of Internet of things platforms in Russia and the EAEU. [Electronic resource]. URL: http://iotintelligence.ru/posts/2848858.

UDC 621.391.14

APPLY ERROR PATTERNS TO CORRECT AND ERASE TWO-DIMENSIONAL ITERATED CODES

REN XUN HUAN, V.K. KONOPELKO, V.Yu. TSVIATKOU

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Submitted 28 November 2020

Abstract. Two-dimensional iterated codes have large minimum Hamming distance and their complexity might be compared with Turbo codes. Familiar algorithms for iterating codes have low decoding capabilities and very high complexities. In order to ensure the applicability of iterated codes, in this paper, we propose a method for correcting and erasing errors of iterated codes for a two-dimensional (2D) interference channel with Hamming code which provides good performance and acceptable complexity.

Keywords: iterated codes, Hamming code, correct and erase errors.

Introduction

Hamming code or Hamming Distance Code is the best error correcting code we use in most of the communication network and digital systems. In Error control coding, parity check bits are calculated based on the input data. The input data and parity check bits are transmitted across a noisy channel. In the receiver, an ECC decoder is used to detect or correct the errors induced during the transmission. The number of parity bits depends upon the number of information bits. The hamming code uses the relation between redundancy bits and the data bits and this code can be applied to any number of data bits. A powerful ECC usually requires more redundant bits and more complex encoding and decoding processes, which increases the codec overhead [1].

At present, the most successful coding schemes are turbo codes and low-density parity-check codes, since their excellent capability, closely to the Shannon limit. Under some specific requirement (typically, code-rates near to the unity and low error rates required), iterated codes may turn into competitive. In addition, they are naturally advisable for high-speed parallel decoder implementation. Extended Hamming codes are representative constituents of iterated formulas since they are uncomplicated and can be iteratively decoded by fast suboptimal algorithms [2, 3]. More formidable constituent codes cause more powerful schemes but request more complex decoding algorithms, usually avoiding high data-rate performances.

The simplest two-dimensional iterated codes are single parity check (SPC) product codes, guaranteed to correct only one error by inverting the intersection bit in the erroneous row and column [1]. Multidimensional SPC iterated codes can be constructed to improve the error correction capability, but a more complex decoding process is required

In [8] proposed a method correct and erase errors by using the error pattern which is based on the process of calculating the pattern of the $t \times t$ (the error mode can be represented as a t by t matrix which can correct only t errors) type library and applying the identification parameters to recognise the errors. The method of correcting and erasing errors which are proposed in [8] cannot completely include all error patterns that can find through permutations of rows. Based on this defect, we propose an improved scheme.

Linear binary product codes are only considered In the thesis. However, the analytical results and the methods are also easily applied to other non-linear codes.

The rest of the paper is organized as follows. In Section 2, a brief introduction of the iterated codes will be presented. And in section 3, an improved algorithm which used to ameliorate pattern

library for correcting and erasing errors that exist through permutations of rows has proposed. The conclusion will be given in Section 4.

Syndrome (table-lookup) decoding

The use of syndrome for error detection and correction is discussed in [6–8], the standard array and its application to the decoding of linear block codes are presented.

Premeditate a (n,k) linear code with generator matrix *G* and parity-check matrix *H*. Let $v=(v_0,v_1,...,v_{n-1})$ be a codeword that transmitted over a noise channel. Let $r=(r_0,r_1,...,r_{n-1})$ be the received vector at the output of the channel. Because of the channel noise, the received words *r* may be different from *v*. The vector sum $e=r+v=(e_0,e_1,...,e_{n-1})$ is an *n* tuple, where $e_i=1$ is called the error pattern. When *r* is received, the decoder computes the following $(n-k)$ tuple

$$S=r\cdot H^T=(S_0,S_1,...,S_{n-k-1}).$$

Which is called the syndrome of *r*. Then, $S=0$ indicates that *r* is a codeword, and $S\neq 0$ means that produced errors. Therefore, we can employ the value of *S* to determine whether an error has arisen. Every (n,k) linear block code is capable of detecting 2^n-2^k pattern errors, however, it is capable of correcting only 2^{n-k} error patterns. For large *n*, 2^{n-k} is a small fraction of 2^n-2^k. Therefore, the probability of a decoding error is much higher than the eventuality of an undetected error.

A linear block code with $d_{\min}$ can assure to detect any errors less than or equal to $d_{\min}-1$. The theorem confirms the fact that a (n,k) linear code with minimum distance $d_{\min}$ is capable of correcting all the error patterns of $\lfloor (d_{\min}-1)/2 \rfloor$ or fewer errors, but it's not capable of correcting all the error patterns of weight $t+1$. A standard array has an important property that can be used to simpify the decoding process. There is a one-to-one correspondence between a correctable error patternand a syndrmes. Using this one-to-one correspondence relationship,we can form a decoding table,which is much simpler to use than a standard array. This table is either stored or wired in the receiver. The decoding of a received vector consists of three steps:

1. Compute the syndrome *S*.

2. Locate the coset leader e_l whose syndrome is equal to $r\cdot H^T$, Then e_l is assumed to the eroor pattern caused by the channel.

3. Decode the received vector *r* into the codeword $v^*=r+e_l$.

In theory, table-lookup decoding can be applied to any (n,k) linear code. It results in minimum decoding delay and minimum error possibility, however for large information redundancy, the implementation of this decoding scheme is not very reality, and either a major storage or a complicated logic circuitry is needed. Product (iterated) codes has the capability of constructing long, powerful quotes from short component codes. Therefore, our analysis is focused on the two-dimensional iterated codes.

Introduction to iterated code

Iterated codes (or product codes) are serially concatenated codes which were presented by Elias in 1954 [2]. The construction method of iterated codes allows us to construct long, powerful codes from short assembly codes. The concept of iterated codes is simple enough and comparatively efficient for constructing extremely long block codes by using at least two short block codes [3].

For a linear block code, the minimum distance is equal to the minimum codeword weight, which is defined as the number of nonzero symbols in a codeword. The minimum Hamming distance is also used to evaluate the error detection capability of a linear block code. The simplest two-dimensional product codes are single-parity check (SPC) product codes [1]. SPC product codes only guarantee correction of one error. The product codes, whose component codes are Hamming or extended Hamming product codes, are known as Hamming product codes.

The random-error-detecting and random-error-correcting capabilities of code are determined by its minimum distance. Hamming codes have a minimum distance of 3 and therefore are capable of correcting any single error over the span of the code block length. The weight enumerator of Hamming codes is known. Hamming codes are perfect codes and can be decoded easily using a table-lookup scheme. Good codes with a minimum distance of 4 for single-error correction and double-error detection can be acquired by properly shortening the Hamming codes. Hamming codes and their shortened editions have been proverbially used for error control in digital communication and data storage systems in these years owing to their high rates and decoding briefness.

Presume that two component codes $C_1(n_1, k_1, d_1)$ and $C_2(n_2, k_2, d_2)$ are used, where n_1, k_1 and d_1 are codeword width, input data width, and minimum Hamming distance for the code C_1, respectively n_2, k_2 and d_2 are codeword width, input data width, and minimum Hamming distance for the code C_2, separately. Here we use the simple Hamming codes construct the iterated codes, let $v = (1100, 0100, 1011, 1001)$ be a codeword, simultaneously $C_1(7,4,3)$ and $C_2(7,4,3)$ are used. Encoding process of iterated (product) codes shows as Fig. 1.

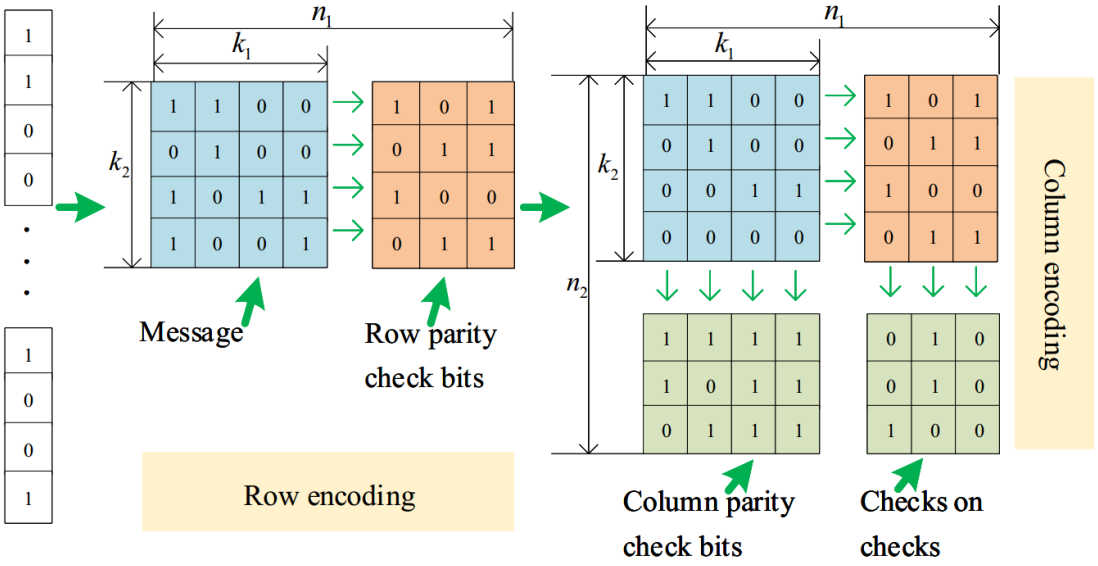


Fig. 1. Iterated codes encoding process

Fig. 1 shows the creation process of two-dimensional iterated codes, the iterated code $C_I((n_1 \times n_2 = 49), (k_1 \times k_2 = 16), (d_1 \times d_2 = 9))$ is constructed from C_1 and C_2 as follows:

The product code C is obtained from the codes C_1 and C_2 in the following manner:

1. Place $k_1 \times k_2$ information bits in an array $k_2 = 4$ rows and $k_1 = 4$ columns.

2. Coding the $k_2 = 4$ rows using the code C_1. Note that the result will be an array of $k_2 = 4$ rows and $n_1 = 7$ columns.

3. Coding the $n_1 = 7$ columns using the code $C_2(7,4,3)$.

Iterated (product) codes have a larger Hamming distance compared to that of the component codes. If the component codes C_1 and C_2 have minimum Hamming distance d_1 and d_2 respectively, then the minimum Hamming distance of the iterated code C_I is the product $d_1 \times d_2$, which greatly increases the error correction capability. Iterated codes can be constructed by a serial concatenation of simple component codes and a row-column block inter-leaver, in which the input sequence is written into the matrix row-wise and read out column-wise. Iterated codes can efficiently correct both random and burst errors. For example, if the received product codeword has errors located in a number of rows not exceeding $(d_2 - 1)/2$ and no errors in other rows, all the errors can be corrected during column decoding.

Algorithm decoding the product codes

Since in 1954, Elias introduced the product code, numerous decoding algorithms for decoding product codes were presented. The most obvious method of decoding is the one suggested by Elias himself in his original work [2]. In Elias's algorithm, the rows in the received message are decoded using a decoder for the code C_1 that decodes up to $\lfloor (d_{1\min})/2 \rfloor$. The columns of the resultant matrix are then decoded using a decoder for the code C_2 that decodes up to $\lfloor (d_{1\min})/2 \rfloor$. It can easily be shown that such a decoder can correct only up $\lfloor (d_1 \times d_2)/4 \rfloor$ [5]. In [8] proposed a new method erase the errors based on the two-step decoding with an error pattern library. Unfortunately, this two-step decoding method fails to correct certain error patterns. Besides, the method of correcting and erasing errors which are proposed cannot completely include all error patterns thast can find through permutations of rows, simultaneously, the algorithm didn't concrete presented how can we erase the error bytes, the disadvantage of the algorithm [6–8] showed in Fig. 2. Based on this defect, analysis of the structure of the error pattern [3], we proposed a more reasonable error correction algorithm. A three-stage pipelined Hamming product code decoding method is proposed, compared to the two-step row-column decoding method, the three-stage pipelined decoding method uses a row status vector and a column status vector to record the conducts of the row and column decoders. Instead of passing only the coded data between row and column decoder, these row and column status vectors are passed between stages to help make decoding decisions.

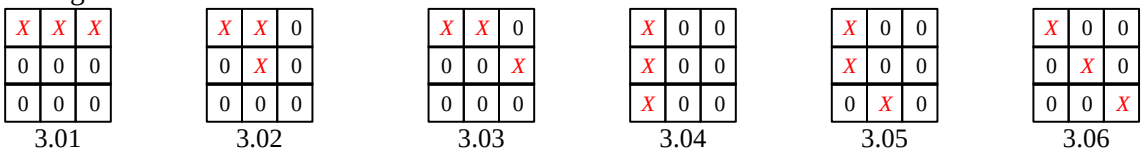


Fig. 2. All of the reduced combination error patterns for $t = 3$ [3]

If we apply Hamming code C_1 that distance is $d_1 = 3$ and Hamming code C_2, $d_2 = 3$ decoding the information that the error's patterns maintained the form as Fig. 2 or their transformer of the row and the column. Obviously, the property of the code itself can correct all of these pattern errors.

To illustrate the error correction process more evidently, Hamming codes are used as row and column component codes. An example of row and column status vectors after the first and second decoding stages shows as Fig. 3.

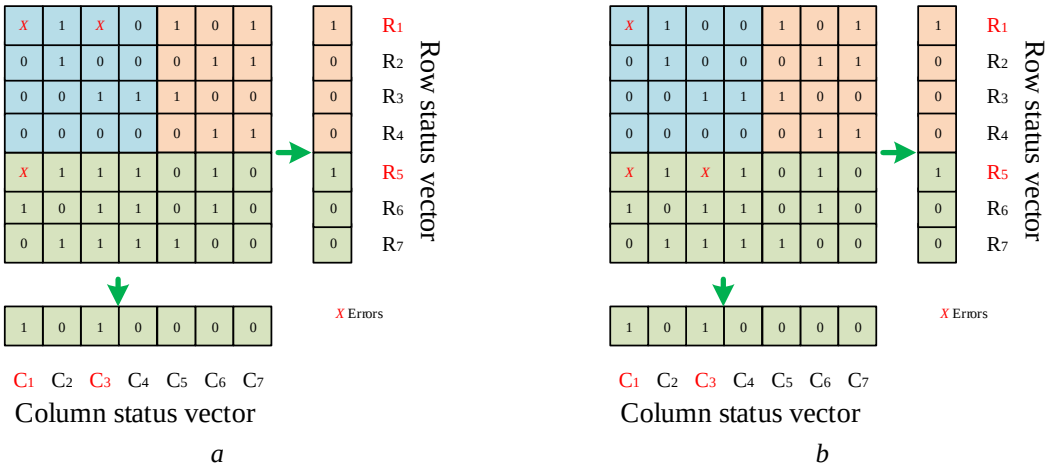


Fig. 3. An example of row and column status vectors after the first and second decoding stages

The simplified row and column status vector implementation can be described as follows.

The *i*-th position in the row status vector is set to «1» when there are detectable errors (regardless of whether the errors can be corrected or not). Then those locations where have only one error (determined by the value of the syndrome) are marked as R_{i_1}, otherwise mark the row as R_i However, if the value of the syndrome is 0, those positions are set as «0».

For the column status vectors, if in the *j*-th column the value of the syndrome is not equal 0, the *j*-th position in column status vector to be set to «1» when an error is detectable but not correctable, and also mark the location of the column as C_{i_1} when an error is correctable (means that only occurs one error), otherwise mark as C_i Fig. 4 shows an example of the row and column status vectors after the first and second stage decoding process.

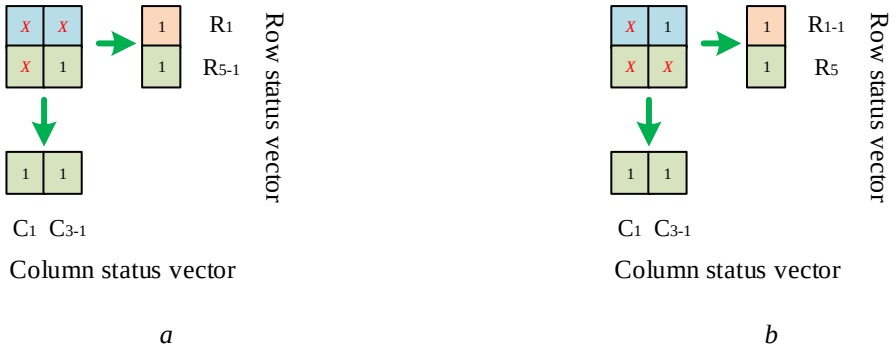


Fig. 4. Example of row and column status vectors after the first and second decoding stages

From the (*a*) and (*b*) of Fig. 4 we can perceive that the errors occurred in the same row (R_1, R_5) and column (C_1, C_3), yet the first graphic shows that there are two errors located in the row of R_1 and only one error located in the row of R_5, and the second graphic just the opposite, so when we erase the first errors of the graphic (*a*) we can apply three steps correct the errors, firstly, if the errors in a row are correctable, the error bit indicated by the syndrome is flipped. The row status vector is set to «0» if the syndrome is zero and «1» if the syndrome is non-zero. In this example we can correct the error which located in the row of R_5 and then correct the column of C_1 and C_3, however, for the graphic (*b*), at first we need to correct the R1and then correct the C_1 and C_3. Therefore we can know that the error pattern of (*a*) and (*b*) are different, the process of erasing is also different. The example is perfect because the Hamming code $d_{\min} = 3$ can correct 1 error and detect 2 errors.

And then we will describe the decoding process of the three-level pipeline Hamming product code. After initializing all state vectors to zero, the steps are as listed below.

The proposed iterative decoding method of two-dimensional Hamming products codes

Input: $r = v + e$
Output: v
Initialization: r_{sta}, $c_{sta} = 0$
While $S \neq 0$ program do.
Step 1. Row decoding of the received encoding matrix. If the error in the row is correctable, the error bit indicated by the syndrome will be reversed. If the syndrome is non-zero, the row state vector is set to «1».
Step 2. Update the column decoding of the matrix. The error correction process is similar to step 1, starting from step 1, use the column error vector and row state vector to calculate the column state vector.
Step 3. After the change from step 2, the matrix is decoded. The syndrome in each row has to be recalculated. If any remaining errors in each row can be corrected, use row syndrome to correct. If errors in each row are still detected but cannot be corrected, use the column status vector in step 2 to indicate which columns need to be corrected.
End while
Return v

Based on the theory of syndrome decoding and the capability of minimum Hamming distance which can correct $(d_{\min}-1)/2$ quantity errors and our proposed method which mark the state of error the row and column, obviously our method can correct permanent errors that are distributed in different rows.

From the analysis above we can derive that the method of three-step decoding can correct all of the random and burst errors that the combinations of error pattern less than 4, and also can correct some of the random and burst errors patterns which equal 5.

In the experiment, we apply our method to correct the error pattern library which proposed in [3], as a result, we can 100 % correct the random and bursts errors *t* less than 3, and for $t=4$ (the quantity of error pattern equal 16) we can correct and erase 93,75 % error patterns, and for $t=5$ (the sum of error pattern equal 34) we can correct and erase 91,17 % error patterns, but for $t=6$ (the sum of error pattern equal 90) we can only correct and erase 74,44 % error patterns, therefore for the error pattern which $t>4$ we need to use the minimum distance of Hamming code which $d_{\min}>3$.

From the process of decoding, we know that not only the encoded data is passed between the row and column decoders, but also the row and column state vectors are passed between stages to help make decoding decisions. We have proposed the method by dividing the encoded information into two transmissions, the reliability of the proposed method depends on both the error detection capability in the first transmission and error correction capability for the iterative decoding method. In the first transmission, the error patterns with single errors in different rows are corrected and the error patterns with two errors in a row are detected. The iterative decoding algorithm can correct up to six-bit errors, furthermore, our method can correct random errors that are distributed in different rows.

Conclusion

In this paper, we analyze methods for decoding the random errors on the basis of product coding, compared with two-step row-column decoding and this method solves the problem of rectangle four error patterns by recording the conduct of the row and column decoders using row and column status vectors. The iterative decoding algorithm can correct up to six-bit errors once the row and column parity check bits are received. Also, our method can correct permanent errors that are distributed in different rows.

References

1. Shu L., Daniel J. // ISBN 0-13-042672-5. 2004. P44–P63.
2. Elias P. // IRE Trans. Inform Theory. 1954. Vol. 4. P. 29–37.
3. Конопелько В.К., Смолякова О.Г. // Докл. БГУИР. 2008. № (7) 37, С. 19–28.
4. Pyndiah R. // IEEE Trans. 1998. Vol. 46. P. 1003–1010.
5. Ericson T. // Computer Science. 1986. № 307. P. 43–57.
6. Конопелько В.К., Смолякова О.Г. // Докл. БГУИР. 2008. № (9) 39, С. 142–153.
7. Конопелько В.К., Хоан Ф.Х. // Докл. БГУИР. 2007. № (8) 38, С. 55–60.
8. Смолякова О.Г., Хоан Ф.Х. // Докл. БГУИР. 2008. № (1) 31, С. 70–75.

УДК 621.391

КАСКАДНЫЙ АЛГОРИТМ LWE-E АЛГЕБРАИЧЕСКИХ РЕШЕТЧАТЫХ КОДОВ И ЭЛЛИПТИЧЕСКИХ КРИВЫХ

М.А. АЛИСЕЕНКО, С.Б. САЛОМАТИН

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 7 ноября 2020

Аннотация. Рассмотрены алгебраические решетки, криптосистема обучения с ошибками и свойства эллиптических кривых. Показан алгоритм кодирования на основе криптосистемы обучения с ошибками и эллиптических кривых. Приведено моделирование алгоритма LWE-E.

Ключевые слова: алгебраические решетки, криптосистема обучения с ошибками, эллиптические кривые.

Введение

Одной из задач разработки алгоритмов защиты данных является их потенциальная способность противостоять различного вида атакам, в том числе на основе пост-квантовых и параллельных вычислений.

Одним их методов решения задач такого рода является применение алгоритмов теории решеток, позволяющих создавать пространственно- временные многообразия кодовых структур и криптографию эллиптических кривых [1–5]. В настоящей работе рассматривается каскадная схема LWE-E, сочетающая алгоритмы с обучением на основе теории решеток и кодированием алфавита точками эллиптической кривой.

Криптосистема обучения с ошибками LWE

N-мерная целочисленная решетка $\mathbb{Z}^m$ – это решетка в евклидовом пространстве $\mathbb{R}^n$, точки которой являются *n*-кортежами целых чисел. Целочисленная решетка является нечетной унимодулярной решеткой [1–3].

Решетка может быть выражена через порождающую матрицу и целочисленный коэффициент аналогично линейным кодам. Под кратчайшим вектором решетки понимается наименьший радиус окружности, которая соединяет ближайшие точки от выбранной центральной точки. Решетчатая криптография относится к набору криптографических конструкций, которые относятся к дискретной аддитивной подгруппе. Среди особенностей решетчатой криптографии – квантовая безопасность, полностью гомоморфное шифрование.

Для того, чтобы решетки с высоким коэффициентом выигрыша от кодирования были применимы на практике, они должны удовлетворять ограничению по мощности [3]. В области решетки ограничение по мощности обеспечивается выбором набора кодирующих точек решетки, которые находятся в области формирования. Сложность области формирования возрастает с увеличением размерности решетки.

Криптосистема строится на основе решеток, поддерживается теоретическим доказательством безопасности. LWE (Learning with errors) параметризуется целыми числами n, m, l, t, r, q и распределением вероятностей χ над $\mathbb{Z}_q$. Функция χ обычно принимается как округленное нормальное распределение.

1. Алгоритм генерация ключа LWE.

Вход: $LWE = n,\ m,\ l,\ q$ – целые числа.

1.1. Выбрать $S \in \mathbb{Z}_q^{n\times l}$ случайным образом.

1.2. Выбрать $A \in \mathbb{Z}_q^{m\times n}$ случайным образом.

1.3. Выбрать $E \in \mathbb{Z}_q^{m\times l}$ согласно χ.

1.4. Вычислить $P = AS + E(\mathrm{mod}\, q)$, где $P \in \mathbb{Z}_q^{m\times l}$.

Выход: закрытый ключ S и открытый ключ $(A;P)$.

2. Алгоритм шифрования.

Вход: целые числа $n,\ m,\ l,\ t,\ r,\ q$, открытый ключ $(A;P)$, открытый текст $M \in \mathbb{Z}_t^{l\times 1}$.

2.1. Выбрать $a \in [-r,r]^{m\times 1}$ случайным образом.

2.2. Вычислить $A^T a(\mathrm{mod}\, q) \in \mathbb{Z}_q^{n\times 1}$.

2.3. Вычислить $c = P^T a + [Mq/t](\mathrm{mod}\, q) \in \mathbb{Z}_q^{l\times 1}$.

Выход: шифротекст (u,c).

3. Алгоритм расшифрования.

Вход: целые $n,\ m,\ l,\ t,\ r,\ q$, секретный ключ S, шифротекст (u,c).

3.1. Вычислить $v = c - S^T u$ и $M = [tv/q]$.

Выход: открытый текст M.

Криптосистемы на основе эллиптических кривых

Применение эллиптических кривых обеспечивает существенно более высокую стойкость при равной трудоемкости или существенно меньшую трудоемкость при равной стоимости. Для вычисления обратных функций на эллиптических кривых известны только алгоритмы с экспоненциальным ростом трудоемкости [2–5].

Эллиптические кривые, заданные в канонической форме, имеют вид $y^2 = x^3 + ax^2 + bx + c$, где a, b, и c – целые коэффициенты.

Полином $P(x) = x^3 + ax^2 + bx + c$ не имеет кратных корней. Многочлен третьей степени (без кратных корней), может иметь либо один, либо три вещественных корня. По предположению, будем считать, что все эти корни различны.

Операция сложения точек на эллиптической кривой E определяется, отправляясь от графического изображения эллиптической кривой рис. 1.

На кривой E берутся две точки P и Q и проводится через них прямая. Эта прямая имеет третью точку пересечения с кривой E. Отражение этой точки от оси x образует новую точку, называемую суммой точек $(P+Q)$.

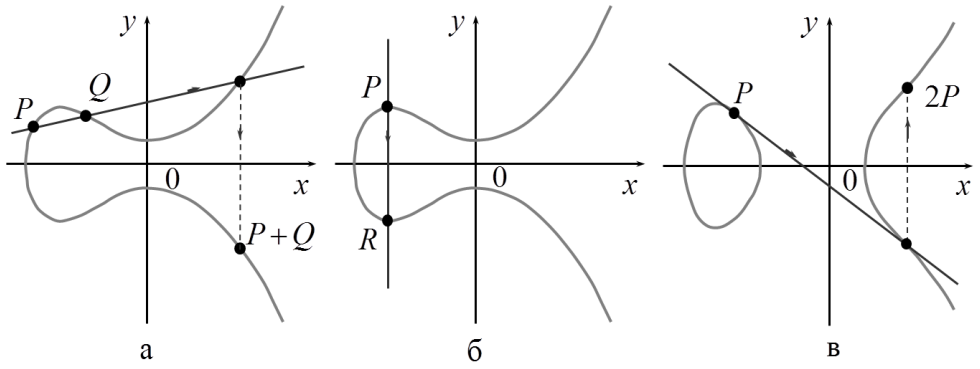


Рис. 1. Графическое изображение точек эллиптической кривой

Пусть точка P имеет координаты (x,y). Точку с координатами $(x,-y)$ будем обозначать как $-P$. Считаем, что вертикальная прямая, проходящая через P и $-P$, пересекает кривую в

бесконечно удаленной точке O, т.е. $[P+(-P)]=O$. По соглашению $P+O=O+P=P$. Точка O играет роль нуля в операциях на эллиптической кривой.

Представим, что точки P и Q сближаются друг с другом, и наконец сливаются в одну точку $P=Q$. Тогда композиция $R=P+Q=P+P$ будет получена путем проведения касательной в точке P и отражения ее второго пересечения с кривой R относительно оси абсцисс $R=P+P=2P$.

Для простого конечного поля Галуа, уравнение Вейерштрасса имеет вид $y^2=x^3+ax+b(\mathrm{mod}\,p)$, где a и b есть целые числа над конечным полем, но такие, что справедливо выражение $4a^3+27b^2\neq 0(\mathrm{mod}\,p)$.

Основной характеристикой эллиптической кривой есть ее порядок $\#E$. Под порядком эллиптической кривой понимается число различных точек на E, включая точку 0, который обозначается как $n=\#E(GF(p))$.

Ниже описаны свойства точек.

1. Сложение с нулем $P+O=O+P=P$, для всех точек $P=E(GF(p))$.

2. Для каждой точки P, существует точка $Q=E(GF(p))$, $P=(x,-y)$, такая что $P+O=O$. Точка Q называется обратным элементом и обозначается как $(-P)$.

3. Если $P=(x,-y)$, $Q=E(GF(p))$, то $(x,y)+(x,-y)=O$.

4. Абелевы группы точек, которые строятся по эллиптическим кривым, имеют одно значительное преимущество, которое объясняет их ценность для криптографии: для одного и того же большого основания p существует богатый выбор различных эллиптических кривых с разными значениями N. Эллиптические кривые составляют богатый источник «естественно возникающих» конечных абелевых групп, и это открывает большие возможности для применения в криптографии.

Операция сложения двух точек: пусть заданы две точки $Q=(x_2,y_2)\in E(GF(p))$ и $P=(x_1,y_1)\in E(GF(p))$. Сумма точек определяется как $R=P+Q=(x_3,y_3)$, где $y_3=(\lambda(x_1-y_3)-y_1)(\mathrm{mod}\,p)$, $x_3=(\lambda^2-x_1-x_2)(\mathrm{mod}\,p)$,

$$\lambda=\begin{cases}\dfrac{y_2-y_1}{x_2-x_1}(\mathrm{mod}\,p), & \text{если } P\neq Q \ \ (x_1\neq x_2)\\ \dfrac{3x_1^{\,2}+a}{2y_1}(\mathrm{mod}\,p), & \text{если } P=Q \ \ (x_1=x_2)\end{cases}.$$

Скалярное умножение определяется для каждой точки $P\in E(GF(p))$ эллиптической кривой $E(GF(p))$ как $kP=\underbrace{P\oplus P\oplus\ldots\oplus P}_{k\ раз}$, где $k\in N$, операция $\oplus$ есть операция сложения на эллиптической кривой.

Алгоритм кодирования алфавитов открытых текстов точками эллиптической кривой

Н. Коблиц в 1985 году предложил вероятностный алгоритм представления кодирования открытых текстов. Алгоритм переводит буквы алфавита A в набор точек на эллиптической кривой. Отображение является инъективным, однако оно будет обладать тем свойством, что, зная координаты точки $P=(x_i,y_i)\in E(GF(p))$ можно однозначно восстановить какому числу i они соответствуют. Таким образом, возможен обратный процесс декодирования.

Схема криптокодирования с использованием эллиптических кривых [6].

1. Задаемся модулем эллиптической кривой p. В соответствии с условием $4a^3+27b^2\neq 0(\mathrm{mod}\,p)$ выбираем коэффициенты a и b данной эллиптической кривой.

2. Согласно формуле $p+1-2\sqrt{p}\leq m\leq p+1+2\sqrt{p}$ производим оценку порядка точек m эллиптической кривой.

3. Согласно соотношениям $m = nq,\ n \in Z,\ n \geq 1,\ 2^{254} < q < 2^{256}$ выбираем q – порядок циклической подгруппы группы точек эллиптической кривой.

4. Образующую поля, точку $P(x_p, y_p)$, выбираем исходя из соотношения $qP = 0$.

5. Выбираем случайное число k, являющееся секретным ключом данной криптосистемы.

6. Производим вычисление точки $kP = P_k(x_k, y_k)$.

7. По формуле $\alpha = \sum_{i=0}^{255} \alpha_i 2^i$ производим преобразование входного двоичного вектора в целое число α, и вычисляем точку $\alpha P = P_\alpha(x_\alpha, y_\alpha)$.

8. Вычисляем $P_k(x_k, y_k) + P_\alpha(x_\alpha, y_\alpha) = Q(x_Q, y_Q)$. Полученная точка $Q(x_Q, y_Q)$ является зашифрованным представлением исходного числа α, а величина k – секретным ключом данной криптосистемы.

9. Для расшифрования необходимо, зная секретный ключ k, получить точку $P_k(x_k, y_k)$, после чего вычислить $Q(x_Q, y_Q) - P_k(x_k, y_k) = P_\alpha(x_\alpha, y_\alpha)$.

Схема криптокодирования приведена на рис. 2.

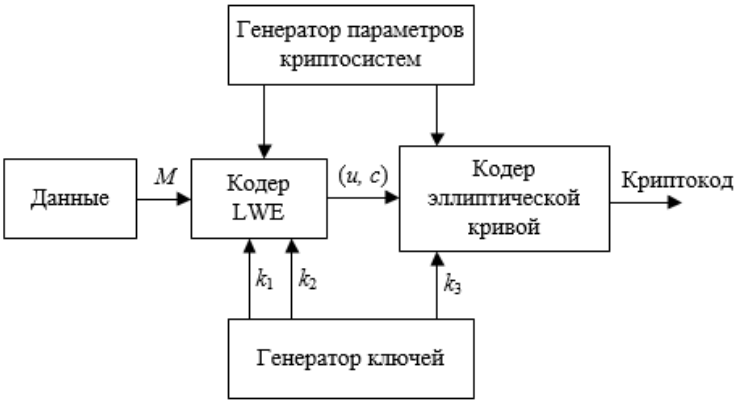


Рис. 2. Схема каскадного криптокодирования

Схема каскадного криптокодирования осуществляет последовательное преобразование информации сначала LWE системой, а затем кодером эллиптической кривой. Кодер LWE системы использует секретный ключ k_1 и открытый ключ k_2, используя для этого генератор нормального распределения. Кодер эллиптической кривой использует секретный ключ k_3, формируемый генератором случайных чисел.

Исходные данные для кодеров LWE системы и эллиптической кривой формирует генератор параметров криптосистем.

Моделирование алгоритма LWE-E

Результаты моделирования получены с помощью программного обеспечения Maple.

Открытый текст M сформируем случайным образом $M = [3\ \ 5\ \ 3\ \ 3\ \ 9\ \ 7\ \ 1]$.

1. Кодирование криптосистемой LWE.

Исходные данные $LWE = [3, 3, 6, 10, 9, 23]$. Секретный ключ формируется на основе нормального распределения randmatrix

$$S = \begin{bmatrix} 0 & 5 & 12 & 12 & 6 & 6 \\ 2 & 13 & 14 & 14 & 18 & 20 \\ 11 & 18 & 7 & 2 & 16 & 16 \end{bmatrix}.$$

Открытый ключ $(A;P)$: $A=\begin{bmatrix}2 & 4 & 4\\ 9 & 9 & 1\\ 7 & 19 & 14\end{bmatrix}$, $P=\begin{bmatrix}6 & 19 & 16 & 21 & 10 & 18\\ 6 & 19 & 11 & 15 & 2 & 20\\ 8 & 5 & 11 & 17 & 10 & 2\end{bmatrix}$

После криптокодирования LWE получим: первое слово криптокода $u=\begin{bmatrix}22 & 8 & 17\end{bmatrix}$ и второе слово криптокода $c=\begin{bmatrix}16 & 15 & 10 & 5 & 6 & 17 & 5\end{bmatrix}$.

2. Кодирование эллиптической кривой.

Найдем точки первого слова криптокода LWE $u=\begin{bmatrix}22 & 8 & 17\end{bmatrix}$. Точки имеют вид $Pu_1=\begin{bmatrix}8590 & 6540\end{bmatrix}$, $Pu_2=\begin{bmatrix}1864 & 1542\end{bmatrix}$, $Pu_3=\begin{bmatrix}9390 & 5333\end{bmatrix}$.

Для второго слова криптокода LWE $c=\begin{bmatrix}16 & 15 & 10 & 5 & 6 & 17 & 5\end{bmatrix}$ имеем множество точек: $Pc_1=\begin{bmatrix}4794 & 411\end{bmatrix}$, $Pc_2=\begin{bmatrix}5812 & 6876\end{bmatrix}$, $Pc_3=\begin{bmatrix}2395 & 7378\end{bmatrix}$, $Pc_4=\begin{bmatrix}5889 & 701\end{bmatrix}$, $Pc_5=\begin{bmatrix}6472 & 8143\end{bmatrix}$, $Pc_6=\begin{bmatrix}9390 & 5333\end{bmatrix}$, $Pc_7=\begin{bmatrix}5889 & 701\end{bmatrix}$.

В качестве примера проведем шифрование точки Pc_6: $EncrP_6=Pc_6+P_k=\begin{bmatrix}9728 & 490\end{bmatrix}$.

3. Декодирование точек криптокода осуществляется по правилу вычитания точки P_k. Проведем расшифрование точки $EncrP_6$: $DecrP_6=EncrP_6-P_k=\begin{bmatrix}9390 & 5333\end{bmatrix}$, что совпадает с точкой $Pc_6=\begin{bmatrix}9390 & 5333\end{bmatrix}$.

4. Расшифрование системой LWE для данного примера $l=\begin{bmatrix}3 & 5 & 3 & 3 & 9 & 7 & 1\end{bmatrix}$, что совпадает с исходным сообщением M.

Заключение

Исследование алгоритма LWE с кодированием алфавита точками эллиптической кривой позволяет построить трехключевые алгоритмы шифрования с мощность разнообразия, определяемой структурами многомерной решетки и эллиптических кривых, что повышает стойкость к пост-квантовым атакам. Криптосистема может быть рекомендована для защиты информации в системах связи и беспроводных сенсорных сетях.

CASCADE ALGORITHM LWE-E FOR ALGEBRAIC LATTICE CODES AND ELLIPTIC CURVES

M.A. ALISEYENKA, S.B. SALOMATIN

Abstract. Algebraic lattices, learning with errors cryptosystem and properties of elliptic curves are considered. An encoding algorithm based on learning with errors cryptosystem and elliptic curves are shown. The LWE-E algorithm is modeled.

Keywords: algebraic lattices, learning with errors cryptosystem, elliptic curves.

Список литературы

1. Ferdinand Nuwan Suresh. University of Oulu Graduate School, 2016. P. 178.
2. Olds C.D. Mathematical Association of USA, 2012. P. 192.
3. Johnson Norman W. Canadian Journal of Mathematics, 1999.
4. Stallings W. Cryptography and Network Security: Principles and Practic. Prentice-Hall, Upper Saddle River, New-Jersey, fifth edition, 2006.
5. Fadyn J.N. // Proceedings of the ICTCM 2014. P. 121–130.
6. Washington L.C. Elliptic Curves: Number Theory and Cryptography, Second Edition, CRC Press, 2008.

УДК 654.16

ВЛИЯНИЕ ПРОМЫШЛЕННОГО РАДИОШУМА НА РАДИУС СОТЫ С ТЕХНОЛОГИЕЙ УЗКОПОЛОСНОГО ИНТЕРНЕТА ВЕЩЕЙ

В.А. АКСЕНОВ, С.В. СМОЛЯК, М.Ю. ХОМЕНОК

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 24 октября 2020

Аннотация. Проведен анализ влияния промышленного радиошума на достижение максимального радиуса покрытия соты с технологией узкополосного интернета вещей NB-IoT в диапазоне 900 МГц. В среде Matlab выполнена аппроксимация оценки распределения плотности вероятности индустриального радиошума по результатам натурных измерений.

Ключевые слова: NB-IoT, промышленный радиошум, тепловой шум, чувствительность приемника, аппроксимация данных, радиус соты.

Введение

Опыт оператора сотовой связи А1 по планированию и эксплуатации сети узкополосного интернета вещей NB-IoT показывает, что достижению максимального радиуса покрытия в сотах этой сети зачастую мешает т.н. промышленный (индустриальный) шум, создаваемый множеством различных источников в промышленно развитых городах, вдоль электрифицированных железных дорог и т.п. На основе этого опыта ниже обсуждается вопрос об учете индустриального шума в расчетах радиуса покрытия соты с технологией NB-IoT.

Радиошум в диапазоне 900 МГц для сотовой сети с технологией NB-IoT

Радиошум определяется в Рекомендации МСЭ-R V.573 [1] следующим образом: радиочастотный шум это «…изменяющееся во времени электромагнитное явление, имеющее составляющие в радиочастотном диапазоне и явно не передающее информации, которое может налагаться на полезный сигнал или смешиваться с ним». Кроме того, «совокупность мешающих сигналов, если они отдельно неразличимы, может проявляться как радиочастотный шум.»

В некоторых случаях радиочастотный шум может передавать информацию о некоторых характеристиках своего источника, например о его природе или месте расположения. Например, промышленные импульсные шумы как правило имеют периодичность во временной области, пропорциональную периодам вращения соответствующих электродвигателей, электрических цепей импульсного регулирования и т.п.

В Рекомендации МСЭ-R P.372 [2] приводятся данные по радиошуму, внешнему по отношению к принимающей радиосистеме, который возникает вследствие следующих причин:

– излучение от грозовых разрядов (атмосферный шум, вызванный грозой);

– совокупное непреднамеренное излучение от электрических механизмов, электрического и электронного оборудования, линий электропередачи или систем зажигания двигателей внутреннего сгорания (промышленный шум);

– эмиссия от атмосферных газов и гидрометеоров;

– присутствие земной поверхности или других препятствий на пути луча антенны;

– излучение от небесных источников радиоволн.

Для описания шумов в Рекомендациях P.372 применяется величина F_a – коэффициент внешнего шума, определяемый как:

$$F_a = 10\lg\left(\frac{p_n}{kT_0B}\right), \qquad (1)$$

где p_n – допустимая мощность шума с выхода эквивалентной антенны без потерь; k – постоянная Больцмана, которая равна $1{,}38\times10^{-23}$ Дж/К; T_0 – эталонная температура (К), принятая равной 290 К; B – ширина полосы приемной системы на уровне мощности шума (Гц).

В соответствии с (1) коэффициент внешнего шума F_a показывает превышение (или наоборот) мощности некоторого внешнего шума над мощностью теплового шума.

Данные по медианным (средним) значениям указанных выше шумов в зависимости от частоты показаны на графиках рис. 1, заимствованных из указанных Рекомендаций Р.372 (стр. 6, рис. 3 в [2]). Можно видеть, что для частоты 900 МГц превышение промышленного шума над тепловым составляет 8 дБ.

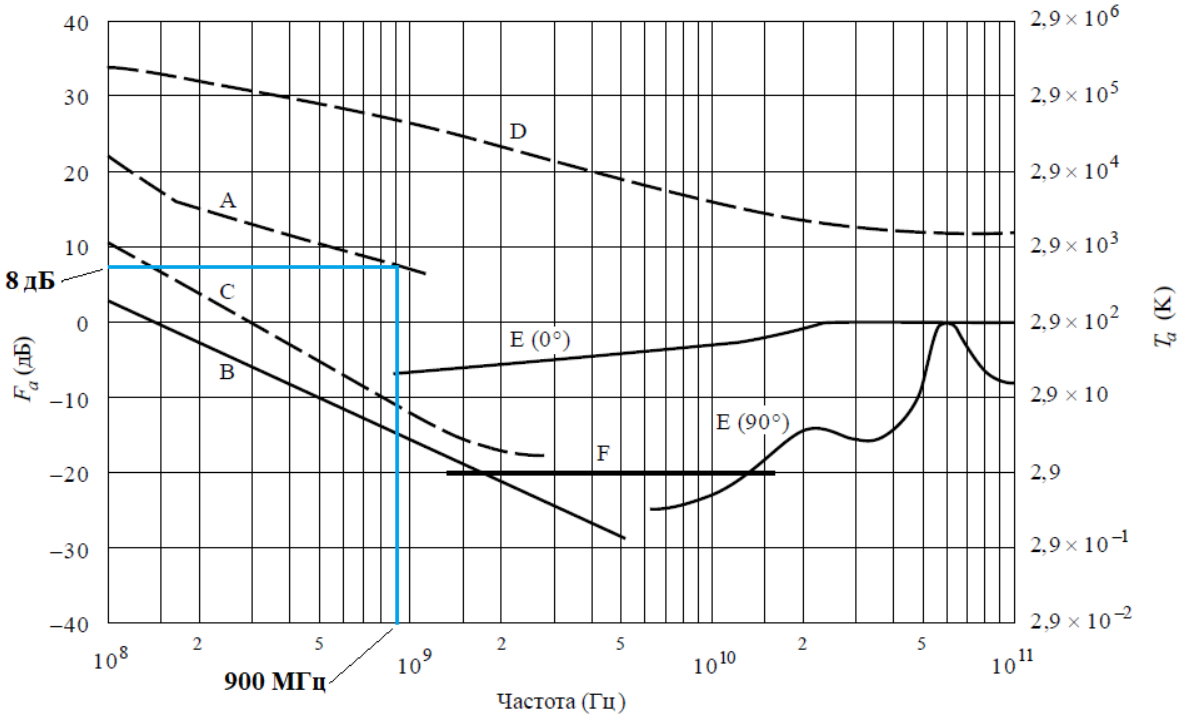


Рис. 1. F_a в зависимости от частоты (10^8–10^{11} Гц)

На рис. 1 приведены следующие графики: A – оценка медианного значения промышленного шума в деловой зоне; B – галактический шум; C – галактический шум (антенна, ориентированная к центру галактики, с бесконечно узким лучом); D – спокойное Солнце (ширина луча 0,5° в направлении Солнца); E – шум неба за счет кислорода и паров воды (антенна с очень узким лучом), верхняя кривая, угол места 0°, нижняя кривая, угол места 90°; F – черное тело (космический фон), 2,7 К.

Как указано в [3], в настоящее время коммерческая успешность эксплуатации сетей NB-IoT достигается при использовании радиусов сот с этой технологией, в 2–3 раза превышающих радиусы сот с иными технологиями в этом же диапазоне 900 МГц (GSM/GPRS/EDGE/U900). Соответственно, требуемая высокая чувствительность приемников в сетях NB-IoT по уровню до –138 дБм (режим передачи с одной несущей, шаг сетки 3,75 кГц), может оказаться недостижимой при столь сильном превышении промышленного шума над тепловым.

Аппроксимация результатов измерений в среде Matlab

Для оценки влияния промышленного шума на радиус сот представляет интерес не только среднее его значение, но и закон распределения плотности вероятности. К сожалению,

сложности с обеспечением метрологической чистоты измерений малых по уровню шумов в широких полосах наблюдения приводят к тому, что соответствующая информация почти отсутствует в доступных источниках и рекомендациях МСЭ. Наиболее интересные данные по закону распределения плотности вероятности промышленных шумов представлены в [4]. В этой статье имеется экспериментально полученная гистограмма распределения отсчетов мощности промышленного шума в осях «количество наблюдаемых значений – уровень (в дБм)» для частотного диапазона (450 – 500) МГц. Указано также, что наблюдаемое распределение мало зависит от времени суток, но более сильно связано с особенностями территории, где выполнялись измерения (промышленная городская среда, жилая территория, сельская местность).

Для получения математического описания (аппроксимации счетной во всех точках кривой) указанного распределения использовался пакет Curve Fitting Toolbox в среде Matlab версии R2020. Наилучший результат «подгонки кривой» (аппроксимации) дала функция из суммы двух гауссовых кривых (т.н. model Gauss2), представленная ниже

$$f(x) = a_1 \exp\left(-\left(\frac{(x-b_1)}{c_1}\right)^2\right) + a_2 \exp\left(-\left(\frac{(x-b_2)}{c_2}\right)^2\right), \qquad (2)$$

где коэффициенты с доверительной вероятностью 95 % принимают значения:

$a_1 = 0,06534\,(0,0288;0,1019)$,

$b_1 = 0,7857\,(0,7564;\ 0,8151)$,

$c_1 = 0,3887\,(0,313;\ 0,4643)$,

$a_2 = 0,04188\,(0,02176;\ 0,06201)$,

$b_2 = 0,324\,(-0,001546;\ 0,6496)$,

$c_2 = 0,6245\,(0,4504;\ 0,7987)$.

В скобках указаны значения доверительного интервала для каждого из коэффициентов.

Значение аргумента x в эксперименте менялось от –78 дБм до –114 дБм. Медианное значение аргумента x, при котором площадь распределения плотности вероятности делится пополам, составляет –90 дБм.

На рис. 2 показаны данные с гистограммы, полученной экспериментально, и аппроксимирующая кривая.

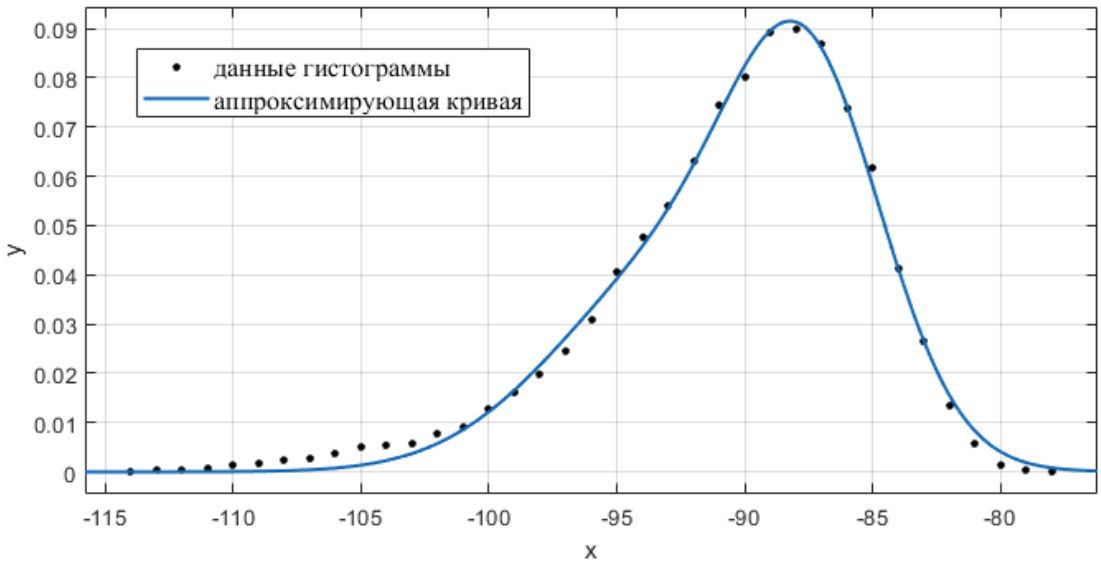


Рис. 2. Распределение плотности вероятности мощности промышленного шума x, дБм

С определенной долей приближения, полученное распределение можно распространить и на диапазон 900 МГц. Можно ожидать, что для этого диапазона и различных вариантов территории (жилая, деловая, сельская) будет меняться среднее значение мощности, однако сохранится форма распределения и его дисперсия.

Влияние параметров шума на радиус соты

Как известно [8], чувствительность приемника Sr (в дБм) определяется выражением

$$Sr = 10\lg\left(kT_0B\right) + NF + E_b/N, \quad (3)$$

или после преобразования

$$Sr = -174 + 10\lg\left(B\right) + NF + E_b/N, \quad (4)$$

где NF – Noise Figure – коэффициент шума приемника, E_b/N – запас на корректную демодуляцию, определяемый требуемым отношением энергии бита к энергии шума.

Вычисляемую таким образом чувствительность можно считать максимальной чувствительностью $Sr\max$. Если известна мощность передатчика Ptr и для расчета потерь при распространении используется, например, линейная от логарифма дистанции модель Хата $L\left(\lg d\right)$ [3], то для расчета радиуса соты можно использовать показанный на рис. 3 график в осях «мощность – логарифм дистанции». На нем точка пересечения линейной функции падения принимаемой мощности с уровнем максимальной чувствительности определит максимальный радиус соты $R\max$ (более точно, его логарифм). Однако, если уровень промышленного шума превышает уровень максимальной чувствительности, то будет происходить т.н. деградация чувствительности с ее уменьшением до некоторого значения $Sr\min$, определяемого параметрами распределения шума. Эта деградировавшая чувствительность определит новый, меньший, радиус соты $R\min$.

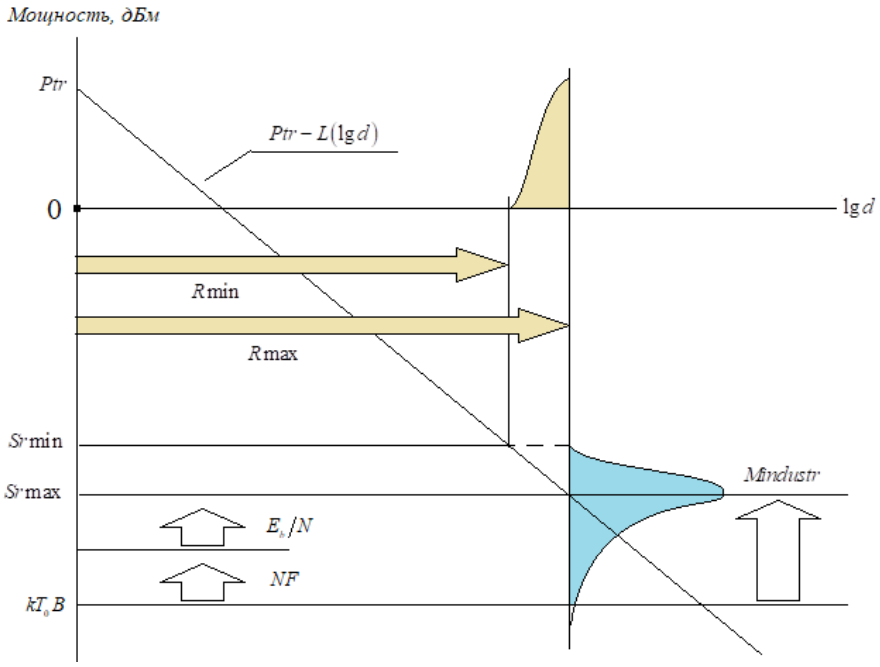


Рис. 3. Влияние параметров распределения мощности индустриального шума на изменение радиуса соты

Для примера, на графике рис. 3 показана ситуация, когда медианное значение $Mindustr$ распределения промышленного шума (полагаемое в Рекомендациях Р 372 для частоты 900 МГц равным 8 дБ выше теплового шума) оказалось в точности равно максимальной чувствительности

приемника $Sr\max$. Это совершенно реальный случай, при условии, что NF обычно лежит в приделах от 2 до 10 дБ [8], и отношение E_b/N, например, равно 3 дБ для модуляции несущей QPSK.

Из графика рис. 3 видно, что степень деградации чувствительности и, соответственно, уменьшение радиуса соты NB-IoT будет определятся взаимным соотношением уровней максимальной используемой чувствительности приемника и средним значением промышленного шума, а также формой кривой распределения уровня шума, лежащей выше уровня максимальной чувствительности. Если применить для графика на рис. 3 значения распределение плотности вероятности шума с рис. 2, поместив его медианное значение на уровень $Sr\max$, то деградация (уменьшение) чувствительности составит 12 дБ.

Заключение

Как следует из представленного выше анализа, учет промышленных радиошумов актуален при расчете параметров радиопокрытия сотовых систем, вынужденных использовать большие радиусы сот и, соответственно, требующих высокой чувствительности приемников, что как раз и присутствует в сетях NB-IoT. Постоянно возрастающая плотность источников промышленных радиошумов еще больше обостряет эту проблему.

INFLUENCE OF MAN-MADE RADIO NOISE ON A RANGE OF A CELL, WHICH USES THE NARROW-BAND INTERNET OF THINGS

V.A. AKSYONOV, S.V. SMOLYAK, M.Yu. KHOMENOK

Abstract. The analysis of the influence of industrial radio noise on the achievement of the maximum radius of the cell coverage with the technology of the narrowband Internet of Things NB-IoT in the 900 MHz range is carried out. In the Matlab environment, an approximation of the estimate of the distribution of the probability density of industrial radio noise based on the results of field measurements is performed.

Keywords: NB-IoT, industrial radio noise, thermal noise, receiver sensitivity, data approximation, cell radius.

Список литературы

1. Рекомендация МСЭ-R V.573-5. Словарь по радиосвязи. [Электронный ресурс]. URL: https:// /www.itu.int/dms_pubrec/itu-r/rec/v/R-REC-V.573-5-200709-S!!PDF-R.pdf.
2. Рекомендация МСЭ-R P.372-12. Радиошум. Серия P. Распространение радиоволн. Электронная публикация. Женева, ITU, 2016г. [Электронный ресурс]. URL: http://www.itu.int/publ/R-REC/en.
3. Аксенов В.А., Смоляк С.В., Хоменок М.Ю. // Телекоммуникации: сети и технологии, алгебраическое кодирование и безопасность данных. Минск. БГУИР, 2019. Стр. 43–48.
4. Frank Leferink [et al.] The Radio Science Bulletin No 334 (September 2010). [Электронный ресурс]. URL: https://www.researchgate.net/publication/47937063.
5. Man-Made Noise Measurement Programme (AY4119). Final Report. Issue 2. September 2003. [Электронный ресурс]. URL: https://www.researchgate.net/publication/3350485_Man-made_noise_measurement_programme.
6. Erik van Maanen. Practical radio Noise Measurements. 18th International Symposium and Exhibition on Electromagnetic Compatibility. June 28-30, 2006, Wroclaw, Poland. [Электронный ресурс]. URL: http://www.emc.wroc.plB.
7. Сообщество Экспонента. Учебное пособие по Curve Fitting Toolbox. [Электронный ресурс]. URL: https://hub.exponenta.ru/post/curve-fitting-toolbox796#1.
8. Kevin Faison. Understanding noise figures in radio receivers. [Электронный ресурс]. URL: https://www.eetimes.com/understanding-noise-figures-in-radio-receivers/#.

УДК 621.391

ЭФФЕКТИВНОСТЬ ФОРМАТОВ МОДУЛЯЦИИ ВЫСОКОСКОРОСТНЫХ ЦИФРОВЫХ ВОСП

Е.Ф. ЛЕОНОВИЧ, Н.В. ТАРЧЕНКО

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 8 ноября 2020

Аннотация. При использовании оптических транспортных сетей важен выбор способов формирования, передачи и приема оптических линейных сигналов, в связи с чем актуальной является задача определения критериев, позволяющих осуществлять их сравнительный анализ.

Ключевые слова: модуляция, эффективность, моделирование.

Введение

В настоящее время при постоянном росте объемов передаваемой информации увеличивается потребность в расширении пропускной способности волоконно-оптических систем передачи (ВОСП), что осуществляется путем совершенствования элементов оптического линейного тракта [1]. В частности, можно выделить следующие направления исследований в этой области:

– совершенствование и развитие когерентных систем связи: использование многоуровневых форматов модуляции и соответствующих устройств модуляции и демодуляции оптических сигналов; использование методов цифровой обработки сигналов для компенсации искажений и предварительной коррекции ошибок;

– совершенствование методов усиления и регенерации оптических сигналов;

– совершенствование структуры оптических направляющих систем.

Предметом исследования в данной работе являются многоуровневые форматы модуляции в высокоскоростных оптических системах передачи.

Критерии эффективности форматов модуляции

Статистическая теория связи для выбора формата модуляции в канале связи, наиболее целесообразного в заданных условиях, рекомендует проводить оптимизацию системы передачи по критериям эффективности.

Как известно [2], цифровой канал передачи характеризуется пропускной способностью – скоростью передачи информации $R_{\text{кан}}$, бит/с, полосой занимаемых частот $F_{\text{кан}}$, а качество передачи в канале – вероятностью ошибки.

Поэтому в качестве основных выбраны следующие критерии:

1 Информационная эффективность системы, определяющая степень использования пропускной способности канала:

$$\alpha = \frac{R_{\text{кан}}}{C_{\text{к}}}, \qquad (1)$$

где $C_{\text{к}}$ – пропускная способность канала, $R_{\text{кан}}$ – скорость передачи информации.

2 Спектральная эффективность системы, определяющая эффективность использования полосы частот канала:

$$\gamma = \frac{R_{\text{кан}}}{F_{\text{к}}}. \qquad (2)$$

3 Энергетическая эффективность системы E_b / N_0, определяющая отношение энергии, затрачиваемой на передачу одного бита информации, к спектральной плотности мощности шума при заданном качестве – вероятности ошибки.

Классификация форматов модуляции в ВОСП

Современные высокоскоростные ВОСП используют спектральное разделение каналов с определяемым Рекомендацией МСЭ-Т G.694.1 разносом частот оптических несущих [3]. При этом интервал между оптическими несущими для увеличения пропускной способности системы уменьшается от 100 ГГц, 50 ГГц, 25 ГГц до 12,5 ГГц. При этом с переходом к когерентным системам изменяются и форматы модуляции.

На рис. 1 представлена классификация форматов модуляции, используемых в высокоскоростных оптических системах передачи [4, 5], и приняты следующие обозначения: NRZ – код без возврата к нулю, RZ – код с возвратом к нулю, CS-RZ – код с возвратом к нулю и подавлением несущей, RZ-AMI – код с возвратом к нулю и альтернативной инверсией, ODB (PSBT) – бинарный код с формированием фазы, ADPSK – адаптивная дифференциальная фазовая манипуляция, DBPSK – дифференциальная двоичная фазовая манипуляция, DQPSK – дифференциальная квадратурная фазовая манипуляция

Рис. 1. Классификация форматов модуляции современных высокоскоростных ВОСП

Для исследования были выбраны следующие виды модуляции BPSK, QPSK, QAM-16, QAM-64.

Энергетическая эффективность форматов модуляции в ВОСП

Для оценки энергетической эффективности необходимо учесть специфику оптических когерентных систем передачи. Как известно [1, 5], основными источниками шумов в когерентных системах являются шумы спонтанной эмиссии оптических усилителей (ASE) и шумы, вызванные нелинейными эффектами в оптическом волокне, шумами когерентного оптического приемника можно пренебречь. При условии, что в ОВ обеспечивается ограничение по мощности группового оптического сигнала, шумами, вызванными нелинейными эффектами в оптическом волокне, также можно пренебречь или учитывать их в виде соответствующего штрафа оптической мощности.

Значение оптического отношения сигнал/шум (*OSNR*) необходимо для определения чувствительности оптического когерентного приемника при различных видах формирования канальных сигналов и поляризационного мультиплексирования.

Теоретическая оценка $OSNR$ оценивается отношением:

$$OSNR = \frac{P_i / \alpha}{N_{ASE} \cdot \Delta F_к}, \quad (3)$$

где P_i – мощность оптического канального сигнала, α – коэффициент, учитывающий поляризационное мультиплексирование ($\alpha = 2$ в случае поляризационного мультиплексирования, $\alpha = 1$ в его отсутствие), N_{ASE} – спектральная плотность мощности шума спонтанной эмиссии ($N_{ASE} = hf_0$, h – постоянна Планка, f_0 – центральная частота используемой длины волны оптического сигнала), $\Delta F_к$ – полоса частот оптического приемника, соответствующая полосе частот принимаемого оптического сигнала. При когерентном детектировании собственными шумами оптического приемника можно пренебречь, следовательно, ОСШ в точке принятия решения будет в основном определяться значением $OSNR$ на входе оптического приемника.

Определим связь $OSNR$ с отношением энергии, затрачиваемой на передачу одного символа, к спектральной плотности мощности шума SNR_S при приеме символа линейного сигнала, которое оценивается отношением:

$$SNR_S = \frac{E_S}{N_0}, \quad (4)$$

где E_S – энергия символа на тактовом интервале T_S ($E_S = P_i T_S$), T_S – длительность интервала для передачи символа сообщения, $B_S = 1/T_S$ – полоса частот сигнала, N_0 – спектральная плотность шума в полосе канала.

С учетом сказанного выше, получим:

$$OSNR = \frac{P_i / \alpha}{N_{ASE} \cdot \Delta F_к} = \frac{E_S / (\alpha T_S)}{N_{ASE} \cdot \Delta F_к} = SNR_S \frac{B_S}{\alpha \cdot \Delta F_к}. \quad (5)$$

Если число уровней (M) оптического сигнала больше двух, то соотношение между символьным (SNR_S) и битовым (SNR_B) отношениями сигнал/шум связаны выражением:

$$SNR_B = \frac{SNR_S}{\log_2 M}. \quad (6)$$

Тогда, при условии, что $m = \log_2 M$, получим:

$$OSNR = SNR_B \cdot m \frac{B_S}{\alpha \cdot \Delta F_к}, \quad (7)$$

где $SNR_B = E_b / N_0$ – отношение энергии бита к спектральной плотности мощности шума.

По формулам (1), (2), (7) рассчитаны значения информационной, спектральной и энергетической эффективностей для анализируемых видов модуляции в оптических системах передачи без и с использованием поляризационного мультиплексирования при условии, что полоса оптического канала (в том числе и оптического приемника) определяется по первым нулям спектра передаваемого сигнала. Результаты расчетов представлены в таблице. При этом энергетическая эффективность определяется минимальным $OSNR$ на входе оптического приемника. По известному значению $OSNR$ рассчитывается необходимая чувствительность оптического приемника.

Значения критериев эффективности различных форматов модуляции

Формат модуляции	Информационная эффективность (α)	Спектральная эффективность (γ)	OSNR при коэффициенте ошибки 10^{-12}, дБ
BPSK	0,107	0,5	5,5
QPSK	0,177	1	8,5
16-QAM	0,302	2	11,5
64-QAM	0,416	3	13,2
PM-BPSK	0,134	1	2,4
PM-QPSK	0,214	2	5,4
PM-16-QAM	0,354	4	8,5
PM-64-QAM	0,482	6	10,2

Заключение

Предложены критерии и получены выражения, позволяющие сравнивать эффективность различных форматов модуляции в современных цифровых ВОСП. Показано, как изменяются информационная и спектральная эффективности в оптических системах передачи при переходе к более сложным видам модуляции. В качестве критерия энергетической эффективности рекомендовано использовать минимальное значение *OSNR* на входе оптического приемника, при котором обеспечивается заданное качество восстановления сигнала при различных форматах модуляции. Оценив значение *OSNR*, можно определить чувствительность когерентного оптического приемника, что необходимо при проектировании оптических систем и сетей телекоммуникаций.

HIGH-SPEED DIGITAL FIBER-OPTIC COMMUNICATION SYSTEMS EFFICIENCY OF MODULATION FORMATS

E.F. LEONOVICH, N.V. TARCHENKO

Annotation. For optical transport networks the choice of methods for the formation, transmission and reception of optical linear signals is important. That is why determining the criteria that would allow their comparative analysis is very actual.

Keywords: modulation, efficiency, modeling.

Список литературы

1. Добавление 39 к Рекомендациям МСЭ-Т серии G. Рассмотрение вопросов расчета и проектирования оптических систем, МСЭ-Т, 2016.
2. Скляр Бернард. Цифровая связь. Теоретические основы и практическое применение М., 2010.
3. Рекомендация МСЭ-Т G.694.1 Спектральные сетки для применения технологий WDM: сетка длин волн для технологии DWDM. – МСЭ, 2012.
4. Наний О.Е., Трещиков В.Н. // Вестник связи. 2012. № 1. С.35–38.
5. Фокин В.Г. Оптические системы с терабитными и петабитными скоростями передачи. Новосибирск, 2015.

UDC 004.932.72'1; 004.93'14

RESEARCH ON PARALLEL ITERATIVE THINNING ALGORITHM

MA JUN, V.Yu. TSVIATKOU, V.K. KONOPELKO

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Submitted 29 October 2020

Abstract. Aiming at the problem of excessive erosion in thinning diagonal lines, complete deletion of patterns in the well-known thinning algorithm proposed by the Zhang and Suen, which is very good in respect to both connectivity and insensitively to boundary noise. Based on it, an improved parallel iterative thinning algorithm has proposed. The experiment shows that the proposed algorithm has better performance in visual quality by comparing with the ZS algorithm. The experiment shows that the proposed algorithm has better performance in visual quality by comparing with the ZS algorithm.

Keywords: Zhang's fast parallel thinning algorithm, single pixel.

Introduction

Image thinning algorithm is a method to extract the image skeleton, which is widely used in character recognition, bio-engineering, fingerprint identification and so on. The purpose of refinement is to eliminate a large number of unwanted points to extract the refined skeleton, preserve the geometric features of the pattern so that the computer can perform image-related tasks efficiently. Image thinning algorithms can be roughly divided into two categories: non-iterative thinning algorithms and iterative thinning algorithms. Noniterative thinning algorithm directly extracts the image skeleton after a round of calculation. The iterative thinning algorithm can be divided into serial thinning algorithm and parallel thinning algorithm. For the serial thinning algorithm, the deletion of pixels depends on all operations previously performed, and the pixels are deleted as soon as the deletion condition is satisfied. For the parallel thinning algorithm, the deletion of pixels depends on the result after the last iteration. All the pixels satisfying the deletion condition will be marked and deleted after completing an iteration [1–4].

Among the parallel thinning algorithms, Zhang's algorithm [1] has the good performance in continuity and it can relatively precisely describe the straight line, inflection point and cross point. However, at the same time, Zhang's algorithm also has some aspects can improved, which will have better performance, this paper will base on Zhang's algorithm to propose an improved one.

The rest of the paper is organized as follows. Section 2, the proposed improved method is illustrated. Section 3 describes the results of experiment. Conclusion is given in Section 4.

Improvement Algorithm

The proposed algorithm works on a 20 neighbors window, which can provide more information than the more common 3×3 neighbor window used in other iterative thinning algorithms. As shown in Fig. 1.

The *P* in the center of the 20 neighborhood is a candidate pixel selected for deletion when its neighborhoods meet some certain conditions.

The conditions consist of several logical criteria, 2 restoring templates, 1 compulsory deletion template and 10 Extra deletion templates.

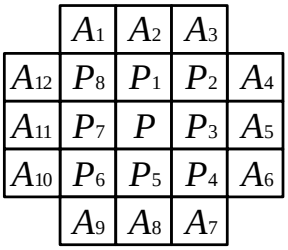


Fig. 1. 20-neighborhood used in proposed algorithm

In order to conduct the logical judgment, $AP(P)$ and $BP(P)$ should to computer at first. $BP(P)$ is the number of foreground pixels in the 8-neighbourhood field of *P*, and is defined as follows

$$BP(P) = \sum_{i=1}^{8} P_i. \tag{1}$$

$AP(P)$ is the total of the value of the corner state in the 8-neighborhood of *P*, $(P_9 = P_1)$ and is defined as follows

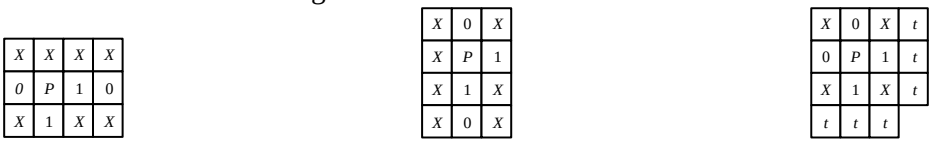
$$AP(P) = \sum_{i=1}^{4} (\overline{P_{2i-1}} P_{2i} + \overline{P_{2i}} P_{2i+1}), \tag{2}$$

$$\overline{P_{2i-1}} = 1 - P_{2i-1}. \tag{3}$$

All the deletable edge pixels are divided into two different groups to conduct the removing process according to their values $AP(P)$ and $BP(P)$.

For most pixels in one image, their values of the $AP(P)$ equal one and that of $BP(P)$ are within the range from 2 to 6. Under this circumstance, these pixels should be checked by the restoring template and compulsory deletion templates, as shown in Fig. 2. All the candidate pixels are removed and only those who match restoring templates but not match compulsory deletion templates are left.

Improvement algorithm including 4 main parts: Search Module, Connectivity Check Module, Single Pixel Correction Module and Contour Point Delete Module. The structure of these stages addressed in Flowchart as shown in Fig. 2.



a *b* *c*

Fig. 2. Restoring templates (*a*, *b*) and deletion template (*c*)

The symbols «0», «1», «*P*» and «*X*» in these templates denote a white pixel, which values are equal 0, a black pixel, which values are equal 1. The currently tested pixel and an ignorable condition, respectively, whereas «*t*» denotes that at least two of the pixels represented by the set of symbols should be a black pixel.

For those edge candidate pixels, whose $AP(P)$ equal two and value of $BP(P)$ are above 4 but less or equal than 5, should be examined by the extra deletion templates as shown in Fig. 3. Only those pixels, whose pattern of neighbors completely matched with one of the extra deletion templates are deletable.

In the extra deletion templates, the symbols «0», «1», «*P*» and «*X*» share the same meaning with the restoring templates and compulsory deletion template. But the symbols «*E*1», «*E*2», «*G*1» and «*G*2» are defined as special symbols that symbols should satisfy the following rule: the values of two arbitrary pixels in a given template marked as identical special symbols should be equal. The difference between «*E*1», «*E*2» and «*G*1», «*G*2» is mainly that the value of a pixel marked as «*E*1» and the value of a pixel

marked as «*E*2» are independent; however, the sum of the values of a pixel marked as «*G*1» and another pixel marked as «*G*2» should be greater than 1.

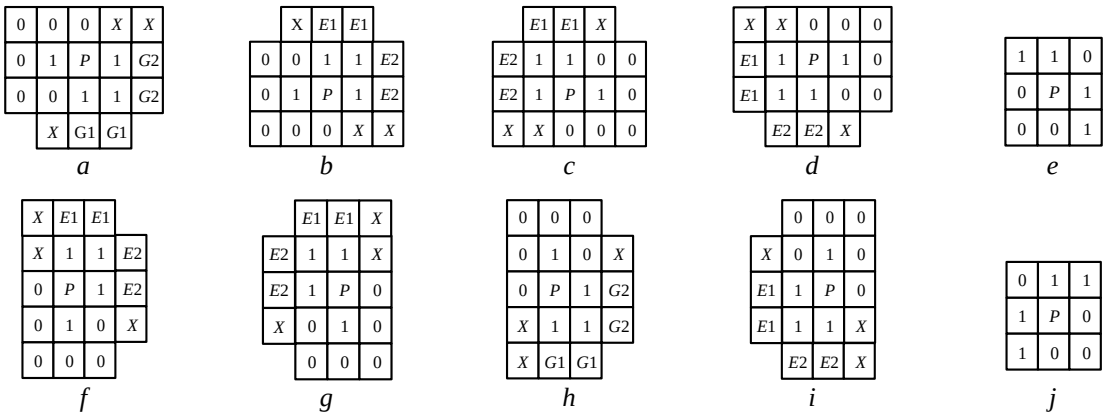


Fig. 3. Extra deletion templates

Tests and Results

To assess the performance, the proposed algorithm and the Zhang's algorithm were written in MATLAB R2018b. This data set compose characteristics and shapes. The results are shown in Fig. 4.

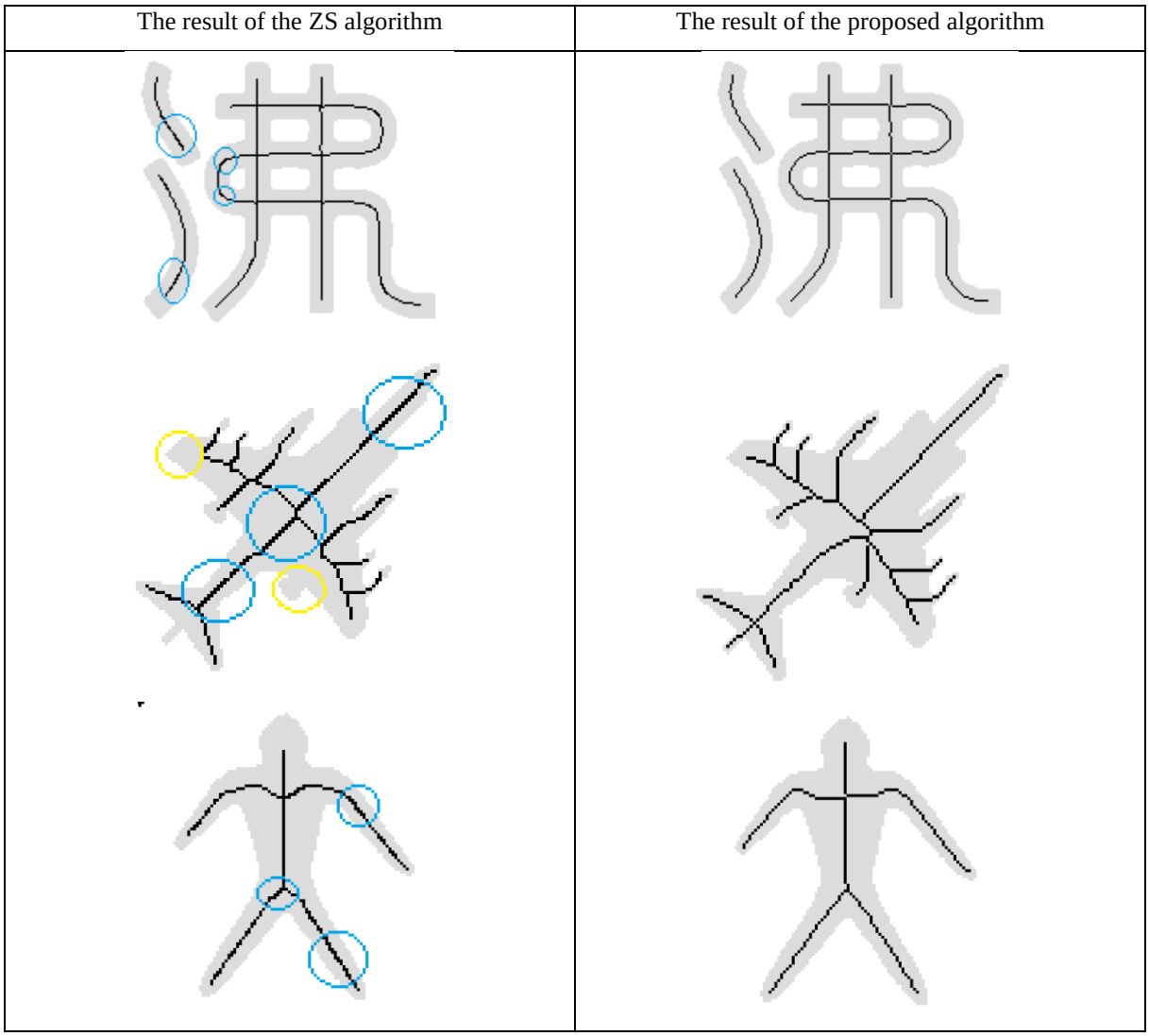


Fig. 4. Result comparison

Table has presented the number of iterations of both algorithms and real time consumed by the algorithms. From the perspective of these two parameters, it is convinced that the proposed algorithm has better performance in terms of the computational complexity, which can reduce about 30 % time consuming.

Comparison of the compute speed between algorithms

Original Binary image	Number of Iteration		CPU Time Consumed(s)	
	Zhang's Algorithm	Our Algorithm	Zhang's Algorithm	Our Algorithm
Character	18	16	0,2037	0,1446
Airplane	34	20	0,4983	0,3311
People	24	19	0,3693	0,2055

Conclusion

In this paper, we presented an improved algorithm based on the Zhang's algorithm, which has better performance in speed and single pixel. The experiments have proved the effectivity of the new algorithm.

References

1. Zhang T.Y., Suen C.Y. A fast thinning algorithm for thinning digital patterns. Communications of ACM.
2. Бушенко Д.А., Садыхов Р.Х. // Докл. БГУИР. 2009. № 7 (45). С. 81–86.
3. Mu S.M., Zhu X.H., Chen G.Y. // Microelectronics & Computer. 2013. Vol. 30, №1. P. 53–55.
4. Ye F.L. // Journal of Sichuan University. 2018. 32 (3). P. 91–93.

УДК 621.391.63

ОТНОШЕНИЕ ПЕРЕКРЕСТНЫХ ПОМЕХ К СИГНАЛУ В СЕТЯХ СЛЕДУЮЩЕГО ПОКОЛЕНИЯ NG-PON2

Н.Н. СЕРГЕЕВ, В.Н. УРЯДОВ, Д.Г. МИХНЮК

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 07 ноября 2020

Аннотация. Проведен анализ отношения перекрестных помех к сигналу в сетях следующего поколения NG-PON2. Показано, такой новый тип сети следующего поколения как NG-PON2 может спокойно функционировать с учетом возможных перекрестных помех сигналов.

Ключевые слова: перекрестные помехи, пассивная оптическая сеть, многоволнавая архитектура.

Введение

В системах PON характеристика отношения перекрестных помех к сигналу (*X*/*S*) применялись только относительно к ONT. Сеть NG-PON2 представляет собой многоволновую архитектуру, и поэтому сигналы с длинами волн NG-PON2, не предназначенные для конкретного приемника ONT либо OLT. Эти сигналы можно рассматривать как мешающие. В этой статье рассмотрены такие помехи и указываются характеристики *X*/*S* как для ONT, так и для OLT.

Отношение перекрестных помех к сигналу в NG-PON2

Ожидается, что ONT будет обеспечивать фильтрацию *X*/*S* непосредственно в приемнике ONT, тогда как *X*/*S* – фильтрация в OLT, вероятно, является внешней по отношению к приемнику OLT или в сочетании с приемником. Поскольку NG-PON2 является многоволновой системой, мешающие сигналы могут поступать из других каналов с длинами волн NG-PON2. В случае с ONT, мешающие сигналы могут также поступать из сосуществующих устаревших систем, таких как GPON, XG-PON1 [1], наложенное видео и длины волн от оптического рефлектометра при измерениях (OTDR). Чтобы свести к минимуму влияние мешающих сигналов, приемники NG-PON2 должны отклонять их, используя соответствующую фильтрацию по длине волны. Поэтому необходимо наличие допуска отношения *X*/*S* приемника NG-PON2. Пусть *S* – это оптическая мощность канала с индивидуальной длиной волны в совокупном сигнале NG-PON2, а *X* – это мощность совокупных мешающих сигналов, состоящих из устаревших сигналов и других каналов с длиной волны NG-PON2 [2]. Оба измерения измеряются в эталонной точке *R* приемника. Для измерения отношения допуска *X*/*S*, необходимо разобрать псевдослучайный код NRZ с той же линейной скоростью, что и основной сигнал NG-PON2 [3]. Фильтрация по длине волны разделена на две части: широкополосный *X*/*S*, основанный на соображениях устаревшей системы, и узкополосный *X*/*S*, основанный на каналах длины волны NG-PON2.

Рассмотрим общую маску допуска *X*/*S* в PON, которая позволяет приемнику ONT NG-PON2 удовлетворять свои требования к чувствительности (рис. 1).

Длины волн и общая оптическая мощность всех дополнительных услуг должны находиться под маской (рис. 1), чтобы обеспечить совместимость с ONT NG-PON2. Серые области обозначают диапазоны, в которых приемник NG-PON2 соответствует требованиям к чувствительности.

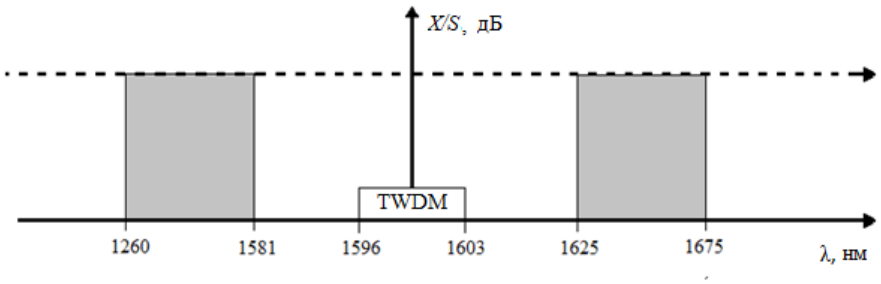


Рис. 1. Общая маска допуска отношения *X*/*S* в TWDM-PON ONT

Узкополосная маска допуска *X*/*S* в TWDM PON

Рассмотрим узкополосную маску *X*/*S* TWDM PON с использованием конкретных таблиц значений, приведенных ниже (рис. 2) [4]. Конфигурация самоинтерференции TWDM-PON является общей как для ONT, так и для OLT. Эта конфигурация с одной желаемой длиной волны и 14 потенциально мешающими длинами волн рассчитана для наихудшей ситуации перекрестных помех. Из 14 видов помех только семь смежных будут работать одновременно. То есть, если центральная частота находится на канале *C* (*C* имеет значение от 1 до 8), то источником помех с самой короткой длиной волны будет 1 канал, а источник помех с самой длинной волной будет на 8 канале. Приемник должен иметь указанную чувствительность для полезного сигнала в диапазоне длин волн λ_{S-} и λ_{S+}, несмотря на то, что присутствуют семь источников помех с уровнем мощности *X* и модулируются форматом сигнала TWDM PON с той же скоростью, что и основной сигнал.

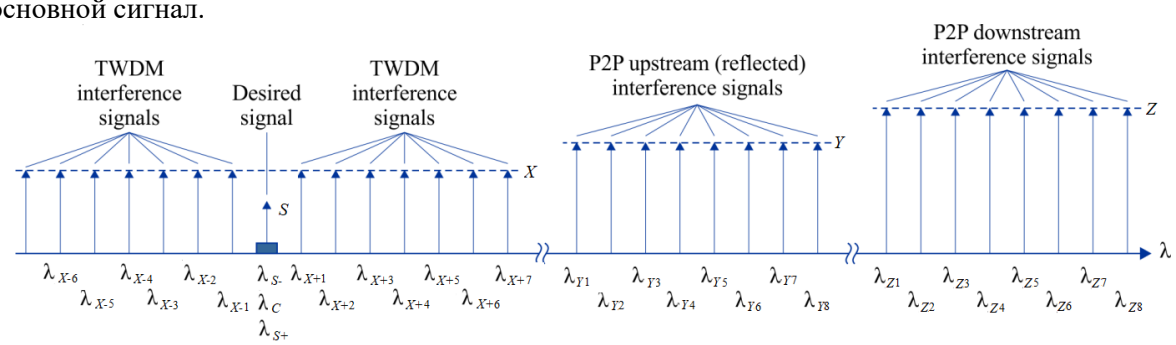


Рис. 2. Узкополосная маска допуска *X*/*S* в ONT

Каналы помех в восходящем направлении (которые возникают из-за возможного ORL 32 дБ в ODN) имеют уровень мощности *Y* и модулируются сигналом NRZ с той же скоростью передачи данных, что и канал сигнала. Нисходящие каналы помех имеют уровень мощности *Z* и модулируются сигналом NRZ с той же скоростью передачи данных, что и канал сигнала [4]. Ниже рассчитаны и приведены параметры маски нисходящего потока (табл. 1).

Таблица 1. **Параметры *X*/*S* нисходящего потока PON TWDM (прием ONT)**

Параметр	Значение
λ_{X-7}	$\lambda_C + MSE - 7 \times CS$
λ_{X-6}	$\lambda_C + MSE - 6 \times CS$
λ_{X-5}	$\lambda_C + MSE - 5 \times CS$
λ_{X-4}	$\lambda_C + MSE - 4 \times CS$
λ_{X-3}	$\lambda_C + MSE - 3 \times CS$
λ_{X-2}	$\lambda_C + MSE - 2 \times CS$
λ_{X-1}	$\lambda_C + MSE - CS$

Параметр	Значение
λ_{S-}	$\lambda_C - MSE$
λ_S	Центральная частота фильтра ONT
λ_{S+}	$\lambda_C + MSE$
λ_{X+1}	$\lambda_C - MSE + CS$
λ_{X+2}	$\lambda_C - MSE + 2 \times CS$
λ_{X+3}	$\lambda_C - MSE + 3 \times CS$
λ_{X+4}	$\lambda_C - MSE + 4 \times CS$
λ_{X+5}	$\lambda_C - MSE + 5 \times CS$
λ_{X+6}	$\lambda_C - MSE + 6 \times CS$
λ_{X+7}	$\lambda_C - MSE + 7 \times CS$
S	Чувствительность приемника + потеря оптического пути
X	$S + 4$ дБ
λ_{Y1}	
λ_{Y2}	
λ_{Y3}	
λ_{Y4}	
λ_{Y5}	
λ_{Y6}	
λ_{Y7}	
λ_{Y8}	
Y	$S + 7$ дБ
λ_{Z1}	
λ_{Z2}	
λ_{Z3}	
λ_{Z4}	
λ_{Z5}	
λ_{Z6}	
λ_{Z7}	
λ_{Z8}	
Z	2,48832 Гбит/с: $S + 7$ дБ 9,95328 Гбит/с: $S + 8$ дБ

Следовательно, 2,48832 Гбит/с (7 дБ) является максимальной разницей мощности передатчика для нисходящего канала (2,5 Гбит/с), а 9,95328 Гбит/с (8 дБ) – это максимальная разность мощности передатчика в OLT для нисходящего канала.

Ниже приведены параметры маски *X*/*S* для OLT (табл. 2). Для восходящего потока PON TWDM каналы расположены с разбросом по длине волны, что облегчает возможность их фильтрации приемниками [5].

Таблица 2. **Параметры *X*/*S* восходящего потока PON TWDM (прием OLT)**

Параметр	Значение
λ_{X-7}	$\lambda_C + MSE - 7\times CS$
λ_{X-6}	$\lambda_C + MSE - 6\times CS$
λ_{X-5}	$\lambda_C + MSE - 5\times CS$
λ_{X-4}	$\lambda_C + MSE - 4\times CS$
λ_{X-3}	$\lambda_C + MSE - 3\times CS$
λ_{X-2}	$\lambda_C + MSE - 2\times CS$
λ_{X-1}	$\lambda_C + MSE - CS$
λ_{S-}	$\lambda_C - MSE$
λ_S	Центральная частота фильтра ONT
λ_{S+}	$\lambda_C + MSE$
λ_{X+1}	$\lambda_C - MSE + CS$
λ_{X+2}	$\lambda_C - MSE + 2\times CS$
λ_{X+3}	$\lambda_C - MSE + 3\times CS$
λ_{X+4}	$\lambda_C - MSE + 4\times CS$
λ_{X+5}	$\lambda_C - MSE + 5\times CS$
λ_{X+6}	$\lambda_C - MSE + 6\times CS$
λ_{X+7}	$\lambda_C - MSE + 7\times CS$
S	Чувствительность приемника + потеря оптического пути
X	S + 20 дБ
20 дБ – это худший случай разницы мощности передатчика TWDM ONT (5 дБ) и дифференциальных потерь на оптическом тракте (15 дБ).	

Заключение

Ожидается, что в NG-PON2 ONT будет обеспечивать фильтрацию *X*/*S* непосредственно в приемнике ONT, тогда как *X*/*S* – фильтрация в OLT, вероятно, является внешней по отношению к приемнику OLT или в сочетании с приемником. Поскольку NG-PON2 является многоволновой системой, мешающие сигналы могут поступать из других каналов с длинами волн NG-PON2. Чтобы свести к минимуму влияние мешающих сигналов, приемники NG-PON2 должны отклонять их, используя соответствующую фильтрацию по длине волны. Поэтому необходимо наличие допуска отношения *X*/*S* приемника NG-PON2. Анализ расчетов допуска показывает, что для восходящего потока каналы расположены с разбросом по длине волны, что дает возможность их хорошей фильтрации. Приемник должен иметь указанную чувствительность для полезного сигнала в диапазоне длин волн λ_{S-} и λ_{S+}. Используя приведенные значения NG-PON2, сеть может спокойно функционировать с учетом возможных перекрестных помех сигналов.

CROSSTALK TO SIGNAL RATIO IN NEXT GENERATION NETWORKS NG-PON2

N.N. SERGEEV, V.N. URYADOV, D.G. MIHNUK

Abstract. The ratio of crosstalk to signal in next-generation NG-PON2 networks is analyzed. It is shown that a new type of next-generation network such as NG-PON2 can safely function with possible signal crosstalk.

Keywords: crosstalk, passive optical network, multi-wave architecture

Список литературы

1. ITU-T Recommendation G.983.1. Broadband optical access networks based on passive optical networks. 2005. P. 11–20.
2. Nesset D. // Journal of Lightwave Technology. London. 2015. Vol. 33. P. 1136–1143.
3. Giorgi L. [et al.] // Opt. Fiber Commun. Conf. Expo. 2013. P. 91–93.
4. ITU-T Recommendation G.989.2. NG-PON2: Physical media dependent layer specification. 2019. P. 9–11.
5. ITU-T Recommendation G.989.3. 40-Gigabit-capable passive optical networks (NG-PON2): Transmission convergence layer specification**.** 2020. P. 21–28.

УДК 004.056.57:032.26

АЛГОРИТМЫ И РЕАЛИЗАЦИЯ ИНТЕЛЛЕКТУАЛЬНЫХ АГЕНТОВ ИНТЕГРАЦИИ ПРЕДПРИЯТИЙ И УЧРЕЖДЕНИЙ ОБРАЗОВАНИЯ

Д.А. КАЧАН, В.А. ВИШНЯКОВ

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 1 ноября 2020

Аннотация. Разработаны алгоритмы запросов требуемых специалистов от предприятий и оптимизации их выпуска учреждениями образования на основе мультиагентной технологии. Для информационной поддержки представлен алгоритм работы интеллектуального агента. Показано применение блокчейн технологии при формировании запросов предприятий на ИТ-специалистов и их удовлетворение. Приведены элементы реализации смарт контракта.

Ключевые слова: интеллектуальный агент, блокчейн технология, смарт контракт.

Введение

Построение интеллектуальной маркетинговой системы с подсистемой интеллектуального агента формирования запросов предприятий на ИТ-специалистов относится к процессам создания единого информационного пространства виртуального предприятия с применением информационных технологий и многоагентной модели управления инновационным кластером. Под виртуальным предприятием стоит понимать межпроизводственную кооперацию ряда юридически независимых организаций [1]. В контексте данной работы стоит рассматривать кооперацию производственных предприятия и учреждений образования в части подготовки специалистов для различных отраслей.

Основой подобной кооперации является единое информационное пространство, в пределах которого осуществляется информационный обмен, обеспечиваемый интеллектуальными агентами, формирующими запросы и агрегирующими необходимые данные. В таком случае интеллектуальные агенты представляют собой необходимые составляющие процессов интернет и интранет-маркетинга.

Мультиагентные системы

В работе [2] используется понятие мультиагентной системы – системы, образованной несколькими взаимодействующими интеллектуальными агентами, в которых агенты могут использоваться для решения одной или нескольких задач. Организация интернет-маркетинга, предложенная авторами, представлена в работе [3], где рассмотрены вопросы работы интеллектуальных агентов (ИА) в среде облачных вычислений. Через портал осуществляется взаимодействие интеллектуальных агентов для анализа потребностей предприятий в специалистах и их выпуск учреждениями образования.

Агент представляет собой вычислительную систему, помещенная во внешнюю среду, способную взаимодействовать с ней, совершая автономные рациональные действия для достижения определенных целей [2]. Характер взаимодействия агентов достаточно разнообразен и определяется реализуемыми функциями, характером информационной маркетинговой системы, взаимодействием организаций в рамках единой информационной среды:

– прямое взаимодействие;

– косвенное взаимодействие (опосредованное через окружающую среду и других агентов);

– сотрудничество агентов;

– конкуренция агентов.

Для описания алгоритма ИА агента формирования запросов предприятий на ИТ-специалистов необходимо принять во внимание решения по архитектуре информационной среды для многоагентной системы [3]:

– состав МАС обнаружения инноваций включает агентов: товарных, ценовых, коммуникационных, исследователей рынка, позволяющих делать вывод о состоянии и тенденциях его развития;

– метод принятия агентами совместного решения, позволяющий сформировать его на основании анализа сведений, полученных из различных источников, оценить состояние рынка в целом и свести для контрактов заказчиков и потребителей;

− методика, позволяющая обучить МАС обнаружению инноваций и использовать ее для поддержки новых разработок (товаров, услуг).

Информационная среда интеграции производства и образования

Сформированная единая информационная среда в интеграции производства и образования представляет собой платформу для дальнейшего развития информационных технологий и применения современных методов. Применяемые информационные системы зачастую не являются механизмом для самостоятельного принятия решений в тех или иных ситуациях, так как при создании ИС практически невозможно предусмотреть все логические цепочки событий и описать их программными алгоритмами. Формирование кластера мультиплицирует сложность системы, и учет негативных последствий становится невозможным.

Выходом из данной ситуации является применение мультиагентных систем, для которых информационная система является внешней средой, с которой агентам приходится взаимодействовать. Интеллектуальный агент не обладает возможностью контролировать внешнюю среду, работа агента сводится к принципу «не навреди» и может не принести ожидаемого результата. На основании этого вытекает ряд требований к алгоритму работу агента: самообучаемость; целеполагание; прогнозирование; планирование; автономность.

Рассмотрим 2 варианта, определяющих алгоритм агентов: простой, представляющий собой, по сути, автоматизацию деятельности человека и комплексный, определяющий интеллектуальную составляющую агента, которая обеспечивает его автономные и самостоятельность действий.

Первый вариант представляет собой программную составляющую информационной системы, существующую в рамках единого информационного пространства и обеспечивающей прямую связь предприятия и учреждения образования. В таком случае, архитектура интеллектуального агента формирования запросов предприятий на ИТ-специалистов сводится к совместно используемой базе данных с учетом обеспечения необходимых требований информационной безопасности.

При росте количества участников единой информационной среды, возможно, осуществить репликацию БД – каждый участник хранит свою копию БД, которые объединяются в синхронизируемый кластер. При развитии единого информационного пространства может возникнуть ситуация, когда рост вычислительных мощностей нецелесообразен и хранение реплик БД на всех узлах нецелесообразно. Дальнейшее развитие среды возможно за счет использования метода шардинга, когда имея в единой среде n серверов, репликация данных осуществляется на m серверов, причем $m < n$. При этом возникает парадокс, описанный в теореме Брюера: доступно лишь 2 состояния из трех: доступность данных, согласованность данных, устойчивость к разделению.

В случае наличия единой БД, алгоритм агента сводится к линейной логике обработки данных запросов, сформированных кадровой службой предприятия, внесения их в некий информационный ресурс учреждения образования, в котором оператор оценивает запрос и осуществляет распределение выпускников, оценку рынка труда и т.п. Фактически, данный алгоритм отражает традиционный подход формирования планов подготовки специалистов, принятый, в том числе и за рубежом: сбор данных о потребностях, обработка, формирование планов подготовки.

Другим вариантом является использование интеллектуального агента, осуществляющего не только обработку первичных данных, поступающих от оператора, но и осуществляющего различные типы анализа данных [4] (семантический анализ, синтаксический анализ и проч.).

Преимущество данного подхода заключается в отсутствии формализованности работы каждого участника среды и, как итог, меньшее количество ошибок при попытках подогнать реальные процессы под некое среднее принятое значение либо формат.

Интеллектуальные агенты для формирования запросов предприятий на ИТ-специалистов

Инновации, связанные с использованием ИТ-индустрии, являются важнейшей отличительной чертой современной экономики. Эта особенность обусловила рост потребности в профильных специалистах, включая наукоемкие производства, образование, технологии, культуру. По этой причине рассматривается именно алгоритм формирования запросов предприятий на примере ИТ-специалистов, хотя данная специфика не ограничивает применение алгоритмов для других видов деятельности. Работа интеллектуальных агентов основана на использовании следующих технологий: синтаксический и семантический анализы, прогнозирование, генетические алгоритмы, классическая и нечеткая логики и др.

Формирование запросов предприятий на ИТ-специалистов осуществляется двумя группами агентов: интеллектуальные агенты оценки кадрового потенциала предприятия (определение спроса рынка труда) и интеллектуальные агенты мониторинга образовательного процесса (подготовка квалифицированных специалистов системой образования).

Алгоритм работы агента первого типа, осуществляющих оценку кадрового потенциала предприятия, на основании которого может происходить формирование заявки на выпускников учреждений образования, сформированный на основании результатов исследования в работе [5] включает шаги:

1. Сбор и предварительная обработка данных;
2. Выбор системы показателей с учетом имеющейся базы данных;
3. Оценка состава, структуры, движения кадрового потенциала организации;
4. Анализ факторов, влияющих на кадровый потенциал и эффективности его использования;
5. Сравнительная оценка кадрового потенциала по совокупности объектов исследования;
6. Оценка эффективности управления кадровым потенциалом в организации;
7. Принятие решения о потребности предприятия в молодых специалистах.

На этапе обмена данными интеллектуальными агентами предприятия и учреждения образования происходит сопоставление приобретаемых квалификаций обучающихся с потребностью предприятия.

Применение технологии блокчейн совместно с использованием интеллектуальных агентов для формирования запросов предприятий на ИТ-специалистов

Применение блокчейн технологии при формировании запросов предприятий на ИТ-специалистов и их удовлетворении, определяется следующими факторами:

– повышение точности планирования кадровой политики предприятий за счет возможности установления предварительных договорных отношений с будущим сотрудником;

– повышение доверия обучающегося в вопросах последующего трудоустройства, что особенно актуально для ИТ-индустрии;

– обеспечение учреждения образования достоверной информацией относительно трудоустройства выпускника при оформлении смарт-контракта с участием всех сторон.

В случае заключения договоренностей между работодателем и выпускником учреждению образования отводится роль арбитра, который по условию смарт-контракта может потребовать от предприятия внесения залога в криптовалютных токенов до момента исполнения контракта.

Реализация смарт-контракта в сети блокчейн

Для реализации алгоритмов ИА в сети блокчейн и целей последующей проверки используем возможности смарт-контрактов [6]. Для разработки и тестирования смарт-контакта используем тестовую блокчейн сеть на локальной вычислительной машине, создаваемую на основе приложения Ganache [7]. Ganache предоставляет возможность создания виртуальной сети блокчейн, состоящий максимально из 10 участников (в работе используется 3 участника) – адресов Ethereum. При активации создаются адреса с закрытыми ключами, а также каждому адресу присваивается сумма криптовалюты в размере 100 единиц Ethereum для осуществления транзакций и отладки взаимодействий.

Другим инструментом, используемым для разработки смарт-контактов, является Remix [8]. Remix – это среда разработки, применяемая не только для создания смарт-контактов, но и для их отладки, публикации в сеть блокчейн (в том числе, созданную для проведения тестов на локальной машине), проверки работоспособности и т.д., используя язык разработки solidity. Общее взаимодействие осуществляется посредством программных алгоритмов на основе языка Python.

Программный алгоритм представлен в виде трех функций, позволяющих выполнить следующие задачи:

– сохранение значения в сети блокчейн (в контексте работы – значения хэш-функции файла электронного документа об образовании);

– задание значения, сохраняемого в сети блокчейн;

– возврат значения по запросу.

После публикации контракта в среде Remix контракт становится доступным для его использования с заданными параметрами. По итогам выполнения получает следующие результаты:

– сформированный файл электронного документа об образовании установленного образца, содержащий необходимые данные – 1014598.pdf;

– значение хэш-функции SHA1 электронного документа 1014598.pdf – 6d80deff635c5e3b4c6d1cbd491cd8c558027688;

– номер транзакции публикации хэш-суммы в сети блокчейн – 0x22a3057d8a4aba41720c54b246f50bc32af5d89c81eb63a7bc0e371eaab5c4a5;

– адрес отправителя – 0x3994Dacf68A445Af612F8fDd2ff34c9a5c4a8D23;

– значение хэш-функции в виде tx data, опубликованное в сети блокчейн – 0x9f2ad4c00000000000000000000000000000000000000000000000000000000000000020000000000
0000000000000000000000000000000000000000000000000000028366438306465666636333563356
53362346336643163626434393163643863353538303237363838000000000000000000000000000000
000000000000000000.

Алгоритм проверки достоверности выполняется следующим образом:

− пользователь, получивший ранее обозначенные данные и осуществляющий проверку, вычисляет хэш-значение электронного документа, используя либо отдельное программное обеспечение, либо интернет-сервисы (например, https://emn178.github.io/online-tools/sha1_checksum.html);

− полученное значение хэш-функции совпадает с предоставленным, и далее необходимо осуществить проверку значения, опубликованного в сети блокчейн. Для этого необходимо найти транзакцию 0x22a3057d8a4aba41720c54b246f50bc32af5d89c81eb63a7bc0e371eaab5c4a5 в сети Ethereum по адресу https://etherscan.io, и по номеру транзакции получить данные tx data (input data);

− необходимо осуществить конвертацию данных из шестнадцатеричной системы счисления в ASCII текст, используя программные продукты для преобразования или онлайн-сервисы (например, https://www.rapidtables.com/convert/number/hex-to-ascii.html).

Заключение

Разработан алгоритм интеллектуального агента формирования запросов предприятий на ИТ-специалистов с использованием мультиагентных информационных систем. Рассмотрены

алгоритмы работы интеллектуальных агентов предприятия и учреждения образования, объединенными единой информационной средой. Для ИА предприятия рассмотрен принцип и алгоритм анализа кадрового потенциала, на основании которого принимается управленческое решение и происходит взаимодействие с ИА УО. Разработан алгоритм взаимодействия ИА с обучающимся на примере заключения смарт-контракта на последующее трудоустройство. Разработан и реализован смарт-контракт и проверена работа программных алгоритмов.

ALGORITHMS AND IMPLEMENTATION OF INTELLIGENT AGENTS FOR INTEGRATION OF ENTERPRISES AND EDUCATIONAL

D.A. KACHAN, U.A. VISHNYAKOU

Abstract. Algorithms for requesting the required specialists from enterprises and optimizing their output by educational institutions based on multi-agent technology are developed.an algorithm for the operation of an intelligent agent is presented for information support. The article shows the use of blockchain technology in the formation of enterprise requests for its specialists and their satisfaction. Elements of smart contract implementation are given.

Keywords: intelligent agent, blockchain technology, smart contract.

Список литературы

1. Афанасьев М.Я., Грибовский А.А. // Научно-технический вестник информационных технологий, механики и оптики. 2011. № 6. С. 113–117.
2. Бугайченко Д.Ю., Соловьев И.П. // Системное программирование. 2005. № 1. С. 36–67.
3. Вишняков В.А., Качан Д.А. // Докл. БГУИР. 2020. № 2. С. 30–36.
4. Вишняков В.А. // Системный анализ и прикладная информатика. 2018. № 1. С. 45–50.
5. Выборова Е.Н., Шатохин А.А. // Аудитор. 2019. № 2. С. 33–40.
6. Chen J. [et al.] // IEEE Transactions on Software Engineering [Electronic resource]. URL: https://ieeexplore.ieee.org/document/9072659.
7. Oliva G.A., Hassan A.E., Jiang Z.M. // Empirical Software Engineering [Electronic resource]. URL: https://link.springer.com/article/10.1007/s10664-019-09796-5.
8. Solaiman E, Wike T., Sfyrakis I. // Wiley special issue paper [Electronic resource]. URL: https://onlinelibrary.wiley.com/doi/epdf/10.1002/cpe.5811.

УДК 621.373

УМЕНЬШЕНИЕ ВИБРАЦИОННОЙ ЧУВСТВИТЕЛЬНОСТИ КВАРЦЕВЫХ ГЕНЕРАТОРОВ

В.В. МУРАВЬЕВ, С.А. КОРЕНЕВСКИЙ, Н.М. НАУМОВИЧ, В.Н. КИЙКО

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 8 ноября 2020

Аннотация. Показано, что применение виброгасителей позволяет обеспечить эффективное уменьшение фазовых шумов кварцевых генераторов на частотах более 300 Гц. Для уменьшения фазовых шумов на частотах менее 300 Гц проведена разработка и экспериментальные исследования электронной схемы компенсации, которая позволяет уменьшить мощность фазовых шумов генератора на 20–25 дБ.

Ключевые слова: фазовые шумы генератора, виброгасители, электронные схемы компенсации.

Введение

До настоящего времени наиболее распространенным видом опорных генераторов для малогабаритных синтезаторов частот являются кварцевые генераторы. Стабильность частоты кварцевых генераторов зависит от многих параметров, в частности, от воздействия ускорения или вибрации, что особенно критично для некоторых вариантов применения кварцевых генераторов, например, при их работе на подвижной технике. Данный параметр принято называть G-чувствительностью генератора. Он определяется как относительное изменение выходной частоты генератора при воздействии ускорения 1g. Наибольший сдвиг частоты генератора наблюдается в случае, если приложенное ускорение направлено параллельно вектору G-чувствительности. Величину и ориентацию вектора G-чувствительности (Г) определяют путем измерения отдельных взаимно ортогональных компонент по x, y, z [1].

Описание эксперимента

Проведенные экспериментальные исследования показали, что основным источником фазовых шумов синтезаторов при воздействии вибраций является кварцевый генератор.

На рис. 1 приведены результаты экспериментальных исследований шумовых характеристик кварцевого генератора при отсутствии (нижняя кривая) и наличии (верхняя кривая) вибраций с спектральной плотностью мощности 0,04 g2/Гц. Видно, что наличие вибраций приводит к увеличению спектральной плотности фазовых шумов кварцевого генератора в области нижних частот до 50 дБ. Для уменьшения спектральной плотности фазовых шумов разработана электронная схема компенсации. На кварцевый генератор устанавливается акселерометр, измеряющий временную зависимость изменения вибраций на кварцевом резонаторе. Сигнал акселерометра усиливается и поступает на вход коррекции частоты кварцевого генератора. Фаза выходного сигнала акселерометра выбирается таким образом, чтобы девиация частоты, обусловленная сигналом акселерометра, была равна по величине и противоположна по знаку девиации частоты, обусловленной воздействием вибрации на кварцевую пластину.

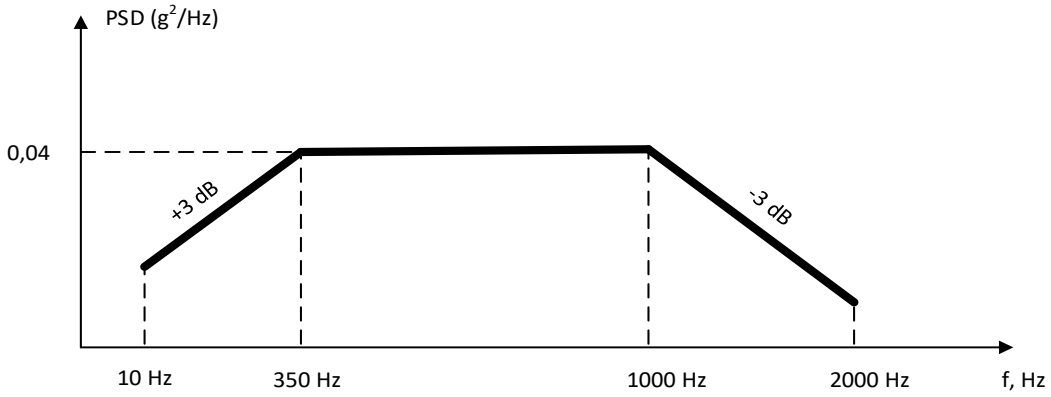


Рис. 1. Спектральная плотность мощности вибраций, воздействующих на кварцевый генератор

На рис. 2 приведены результаты экспериментальных исследований фазовых шумов кварцевого генератора. Нижняя кривая – шумы синтезатора на частоте 100 МГц при отсутствии вибраций; средняя кривая – фазовые шумы кварцевого генератора при наличии электронной схемы компенсации фазовых шумов; верхняя кривая – фазовые шумы кварцевого генератора при спектральной плотности вибраций 0,04 g2/Гц; средняя кривая – фазовые шумы кварцевого генератора при наличии электронной схемы компенсации.

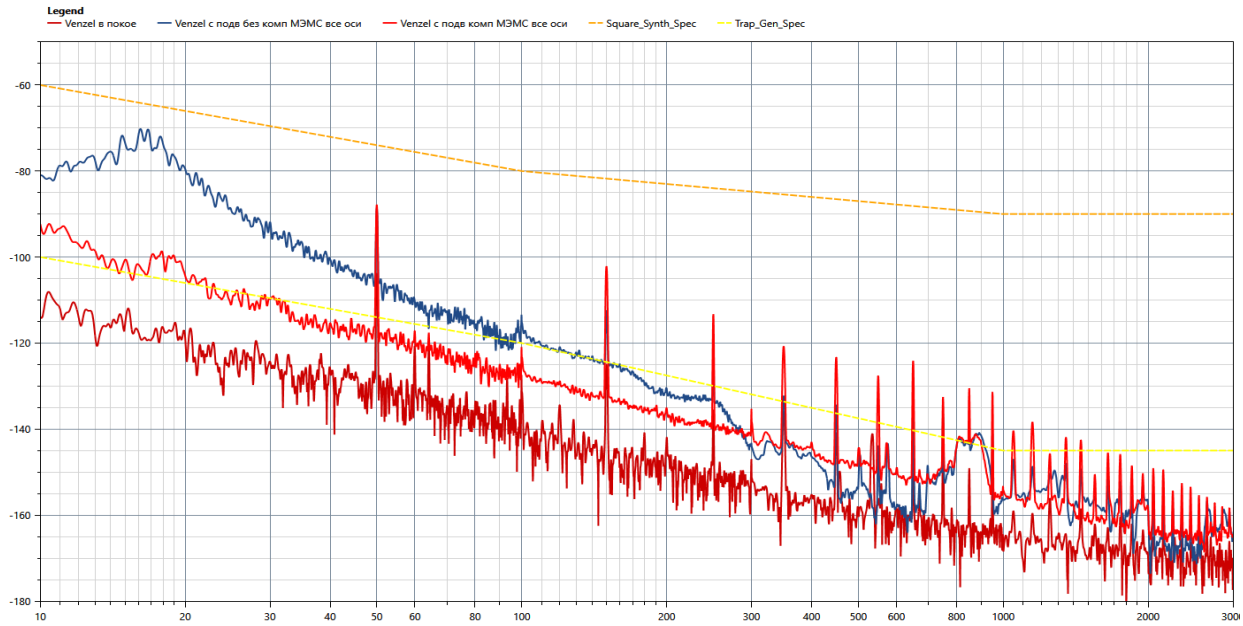


Рис. 2. Фазовые шумы кварцевого генератора при использовании в схеме электронной компенсации МЭМС датчика и мощности вибраций 0,04 g2/Гц

Из рис. 2 видно:

– схема электронной компенсации позволяет обеспечить эффективную компенсацию фазовых шумов генератора в диапазоне частот 20–300 Гц (до 25 дБ);

– уменьшение эффективности схема электронной компенсации на частотах менее 20 Гц, обусловлена неравномерностью АЧХ акселерометра;

– уменьшение эффективности схема электронной компенсации на частотах более 300 Гц, обусловлена временем задержки сигнала в акселерометре, выполненном по МЭМС технологии.

Эффективное уменьшение фазовых шумов генератора на частотах более 300 Гц, может быть обеспечено:

– применением серийных виброгасителей;

– увеличением массы кварцевого генератора;

– применением акселерометров с малым временем задержки.

Полученные результаты показывают, что применение электронных схем компенсации позволит обеспечить значительное уменьшение мощности фазовых шумов кварцевого генератора, и использовать серийные, малогабаритные виброгасители.

Заключение

Для обеспечения малой вибрационной чувствительности кварцевых генераторов необходимо одновременное использование виброгасителей и электронной схемы компенсации. Это позволяет уменьшить спектральную плотность мощности фазовых шумов кварцевых генераторов на 20–25 дБ.

REDUCING THE VIBRATION SENSITIVITY OF QUARTZ GENERATORS

V.V. MURAVIEV, S.A. KARANEUSKI, N.M. NAUMOVICH, V.N. KIYKO

Abstract. It was shown that the use of vibration dampers makes it possible to effectively reduce the phase noise of quartz oscillators at frequencies above 300 Hz. To reduce phase noise at frequencies less than 300 Hz, the development and experimental research of an electronic compensation circuit has been carried out, which makes it possible to reduce the power of the phase noise of the generator by 20–25 dB.

Keywords: generator phase noise, vibration dampers, electronic compensation circuits.

Список литературы

1. Alabaster C. // SciTech Publishing, Edison, NJ. 2012. P. 139–145.

УДК 621.391.1

ПЕРСПЕКТИВЫ ИСПОЛЬЗОВАНИЯ СИГНАЛОВ С OFDM В СИСТЕМАХ СВЯЗИ ДИАПАЗОНА ОВЧ

А.Л. ХОМИНИЧ

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 3 ноября 2020

Аннотация. Рассматривается возможность использования радиосигналов с множественными несущими с ортогональным уплотнением (OFDM) в тактических системах связи диапазонов очень высоких (ОВЧ) и частично ультравысоких (УВЧ), т.е. от 30 МГц до 512 МГц. Анализируются перспективы развития систем связи с точки зрения увеличения пропускной способности и сопутствующие этому проблемы.

Ключевые слова: пропускная способность, символьная скорость, несущая, межсимвольная интерференция, многолучевое распространение.

Введение

Практически свершившийся полный переход на цифровые методы обработки и передачи речи, изображений и данных оказывает влияние и на использование тактических систем связи, большинство из которых работают в диапазоне частот от 30 до 512 МГц. Долгое время эти системы использовались только для передачи речи, с преимущественным использованием частотной модуляции (ЧМ) и узкополосных, с шириной полосы частот 25, 12,5 или 8,33 кГц, каналов связи. В конце XX века эти системы получили возможность передачи данных со скоростями в пределах от 1,2 до 19,2 кбит/с и в практически неизменном с тех пор виде выпускаются как отечественными, так и зарубежными предприятиями [1–6].

В настоящее время требования к тактическим системам связи, в том числе и системам боевого управления существенно меняются [7]. Они должны обеспечивать не только передачу речи в цифровом виде, но и данных, а желательно еще и видеоинформацию. При этом для передачи речи, в зависимости от требований к качеству и типа кодека, достаточной будет скорость передачи в пределах от 1 до 10 кбит/с, средняя скорость передачи данных может быть невысокой (единицы, максимум десятки кбит/с), но эти данные следуют, как правило, в пакетном режиме и для достижения минимальной задержки их передачи пиковая скорость должна быть достаточно высокой (сотни кбит/с, а в ряде случаев и выше). Если же ставить задачу передачи видео, то здесь потребуется средняя пропускная способность в сотни кбит/с на один канал.

Использование узкополосных каналов с полосами 5/8,3/12,5/25 кГц не позволяет добиться перечисленных скоростей передачи даже при использовании многопозиционных видов модуляции, к тому же существенно снижающих помехоустойчивость. Поэтому очевиден переход к широкополосным системам или за счет объединения нескольких каналов, или за счет использования иной сетки частот, при этом и методы формирования радиосигналов также становятся иными [8, 9]. Подобные решения уже существуют и количество их растет [10–15], прежде всего осваивается диапазон частот от 225 до 450 МГц, при этом может использоваться сетка частот с шагом 125 или 625 кГц и полосой частот радиосигнала 660 кГц или 4 МГц соответственно. В немалой степени этому способствует использование концепции программно определяемого радио (SDR), позволяющей в пределах заданной полосы частот гибко варьировать параметры модуляции и помехоустойчивого кодирования, адаптируя систему к конкретным условиям использования.

В тоже время переход к широкополосным системам порождает проблемы, которые были мало актуальны ранее, одной из наиболее важных которых является влияние на качество связи многолучевого распространения радиоволн (РРВ). Один из способов ее решения и рассматривается далее.

Анализ использования сигналов с OFDM в системах связи диапазона ОВЧ

Известно, что реальная пиковая скорость передачи данных в радиоканале определяется выражением

$$B = \frac{\log_2(M) R \Delta F}{1+\alpha} = \log_2(M) B_{\text{СИМВ}},$$

где M – размерность сигнального созвездия; R – относительная скорость помехоустойчивого кодирования; ΔF – ширина полосы частот радиоканала; α – коэффициент расширения полосы частот радиосигнала по сравнению с теоретическим пределом (пределом Найквиста), обычно выбирается в пределах 0,1…0,35; $B_{\text{СИМВ}}$ – символьная скорость.

Средняя скорость при этом будет несколько ниже, поскольку при ее вычислении учитываются интервалы времени на передачу синхронизирующих и служебных сигналов, пилот-сигналов и защитных интервалов (при наличии), а также процент времени активной работы станции.

В дальнейшем, для упрощения рассуждений будем руководствоваться параметрами символьной скорости, поскольку именно она напрямую определяет длительность символа, важную при анализе межсимвольной интерференции из-за многолучевого РРВ.

Рассмотрим потенциальные проблемы в каналах связи с полосой от 125 кГц до 4 МГц, что соответствует увеличению в 5…160 раз по сравнению с используемой в настоящее время полосой 25 кГц. В дальнейшем результаты могут быть экстраполированы и на другие полосы частот. Принимая $\alpha = 0,25$, получим соответствующие значения символьной скорости от 0,1 до 3,2 Мсимв/с.

Межсимвольная интерференция начинает оказывать значительное влияние на качество связи при разности задержек распространения лучей $\Delta t_3 \geq 0,25 T_{\text{СИМВ}}$, что при принятых выше значениях символьной скорости равно 2…0,0625 мкс и соответствует разности хода лучей $r = \Delta t_3 \times c = 800...25\text{м}$, здесь c – скорость распространения электромагнитных волн в вакууме. Расстояние между абонентами типовой тактической системы связи может быть от нескольких десятков метров до нескольких десятков километров, поэтому в условиях многолучевого РРВ, характерного для крупного города либо холмистой местности, такие разности путей лучей легко достижимы. В условиях отсутствия прямой радиовидимости, когда вероятен приход нескольких лучей с примерно равной мощностью, их задержка, приближающаяся к длительности символа, приведет к полной потере связи практически при любом отношении сигнал/шум на входе приемника.

Существует два основных пути решения проблемы многолучевого РРВ – расширение спектра передаваемого сигнала и увеличение длительности символа за счет использования передачи на нескольких несущих.

Первый способ заключается в замене передаваемого символа последовательностью, и широко используется в системах связи с кодовым разделением каналов (CDMA). Однако он имеет два недостатка, существенных в рассматриваемом случае. Во-первых, полоса частот радиосигнала и так предполагается достаточно большой (до 4 МГц), реализовать существенное ее расширение, особенно в диапазоне частот до 225 МГц, не представляется возможным. Во-вторых, для реализации кодового разделения требуется, чтобы уровни сигналов, поступающих на вход приемника от разных абонентов, были близки. В условиях, когда расстояния между абонентами сети могут различаться на несколько порядков, такое реализовать крайне сложно. Тем не менее, системы с прямым расширением спектра являются наиболее вероятным кандидатом для работы в условиях, требующих высокой помехоустойчивости, в том числе и радиоэлектронной борьбы. Естественно, это возможно только при наличии необходимого частотного ресурса.

В настоящей работе анализируется второй путь – использование множественных несущих. При этом скорость передачи на каждой несущей будет снижена в N_A раз, а длительность символа во столько же раз увеличится. Здесь и далее N_A – количество активных (используемых) несущих.

Если принять за допустимую задержку распространения $\Delta t_З = 10$ мкс, что соответствует разности хода лучей $r = 3000$ м, то длительность символа должна быть порядка 40...50 мкс. Приняв полезную длительность символа $T_A = 40$ мкс и относительную длительность защитного интервала (ЗИ) $T_{GI} = 1/4$, получим полную длительность OFDM символа $T_{OFDM} = T_A / (1 + T_{GI}) = 50$ мкс. Расстояние между несущими при соблюдении условия ортогональности при этом составит $\Delta F = 1/T_A = 25$ кГц, что хорошо согласуется с шагом сетки частот, принятой в нижней части ОВЧ-диапазона.

В этом случае при одинаковых с одночастотным методом условиях использования полосы частот, к примеру, в условном 2 МГц канале можно разместить 64 активных несущих, из которых порядка 50…60 будут использованы для передачи данных, остальные – как пилот-сигналы. Достижимая символьная скорость при этом составит

$$B_{\text{СИМВ}} = \frac{1}{T_A(1+T_{GI})N_A} = 1...1,2 \text{ Мсимв/с.}$$

Это несколько меньше, чем в одночастотном режиме, что обусловлено наличием пилот-сигналов и защитного интервала, однако помехоустойчивость в условиях многолучевого РРВ обещает быть существенно выше. Кроме того, скорость спада спектральной плотности мощности OFDM-сигнала за пределами активной полосы частот существенно выше, чем у QAM-сигнала, поэтому коэффициент запаса α может быть меньше, соответственно символьная скорость приблизится к одночастотному случаю.

Безусловно, проблемы при использовании OFDM-сигналов тоже есть. Это более высокие требования к стабильности частот гетеродинов передатчика и приемника, необходимость обеспечения частотной и временной синхронизации при приеме, соответственно усложненная конструкция. Также недостатком является значительное отношение пиковой мощности радиосигнала к средней (пик-фактор), из-за чего требуется повышенная линейность выходных каскадов передатчика, что увеличивает энергопотребление. Но в целом эти проблемы решаемы, что подтверждается широким распространением OFDM-систем в цифровом теле- и радиовещании, системах широкополосного беспроводного доступа, сотовых системах связи 4-го (LTE) и последующих поколений [16–18].

Заключение

Проведенный анализ свидетельствует о том, что применение сигналов с OFDM вполне актуально и в системах связи ОВЧ и близких к нему диапазонов с учетом перспектив их развития. Изучение зарубежного опыта показало, что работы в этом направлении ведутся, и достаточно интенсивно [19–21]. Также анализ показывает, что современные тактические системы связи должны быть универсальными, т.е. поддерживать как различные перспективные стандарты связи, так и существующие, вплоть до аналоговых. Концепция SDR вполне позволяет такую реализацию.

Также следует отметить, что в данной статье не ставилась задача привязки к конкретной системе, принятые исходные данные являются типовыми (усредненными), однако при необходимости могут быть уточнены либо экстраполированы на нужный класс систем, либо устройств.

PROSPECTS FOR USE OF OFDM SIGNALS IN VHF RANGE COMMUNICATION SYSTEMS

A.L. KHAMINICH

Abstract. The possibility of using multi-carrier radio signals with orthogonal multiplexing (OFDM) in tactical communication systems of the very high (VHF) and partially ultrahigh (UHF) frequency band, i.e., from 30 MHz to 512 MHz, was considered. The analysis of the development of communication systems in terms of increasing throughput and the accompanying problems was given.

Keywords: throughput, symbol rate, carrier, intersymbol interference, multipath propagation.

Список литературы

1. Портативные радиостанции сети боевого управления [Электронный ресурс] URL: https://topwar.ru/64725-portativnye-radiostancii-seti-boevogo-upravleniya.html.
2. SINCGARS: Семейство радиостанций армии США [Электронный ресурс] URL: https://trcvr.ru/2016/04/07/sincgars-semejstvo-radiostancij-armii-ssha.
3. Средства связи и боевая экипировка. Каталог ОАО «Техника связи» [Электронный ресурс]. URL: https://t-c.by/wp-content/uploads/2019/10/Katalog-TVN.pdf.
4. Носимая радиостанция УКВ-диапазона Р-188 [Электронный ресурс URL: http://www.agat-system.by/upload/ iblock/574/574213990e8f7b3c7bf5246d1e60e629.pdf.
5. Комплекс возимых радиостанций КВ и УКВ диапазонов Р-181 [Электронный ресурс]. URL: https://bte.by/katalog/sredstva-svyazi/kompleks-vozimykh-radiostantsiy-kv-i-ukv-diapazonov-r-181.html.
6. PR4G F@stnet Product Family [Электронный ресурс]. URL: https://www.thalesgroup.com/en/markets/defence-and-security/radio-communications/land-communications/ tactical-radios/pr4g-fstnet.
7. Joe L. Future army bandwidth needs and capabilities [Электронный ресурс] URL: https://www.rand.org/content/dam/rand/pubs/monographs/2004/RAND_MG156.pdf.
8. Business Models for new entrants in SDR Tactical Radio Market [Электронный ресурс]. URL: https://www.wirelessinnovation.org/assets/work_products/Reports/winnf-15-p-0064-v1.0.0%20%20sdr% 20business%20models.pdf.
9. Blyskun A. [et. al.] //MILCOM 2013 – 2013 IEEE Military Communications Conference, San Diego, CA, 2013, P.396–399.
10. HF XL Concept: the right answer to current crowded HF spectrum [Электронный ресурс]. URL: https://www.thalesgroup.com/en/markets/defence-and-security/radio-communications/land-communications/ tactical-radios/hf-wideband.
11. Wideband vehicular Software Defined Radio (SDR) [Электронный ресурс]. URL: https://www.thalesgroup.com/en/markets/defence-and-security/radio-communications/land-communications/ tactical-radios/flexnet.
12. L3HARRIS RF-7880NR. Enhanced High-Capacity Data Radio (EnHCDR) [Электронный ресурс]. – URL: https://www.harris.com/sites/ default/files/downloads/solutions/rf-7880nr-enhcdr-enhanced-high-capacity-data-radio-datasheet.pdf.
13. L3HARRIS AN/PRC-163. Multi-channel Handheld Radio [Электронный ресурс]. URL: https://www.harris.com/sites/default/files/an-prc-163-multi-channel-handheld-radio-datasheet.pdf.
14. TC5 Wideband HF & VHF Modem and ALE Controller Module [Электронный ресурс]. URL: https://www.rapidm.com/product/tc5-wideband-hf-vhf-modem-and-ale-controller-module.
15. The Case for Software-Defined Capabilitie [Электронный ресурс]. URL: https://www.harris.com/perspectives/modernizing-the-battlefield/the-case-for-software-defined-capabilities.
16. Zhang X., Zhou X. LTE – Advanced Air Interface Technology. CRC Press, 2013.
17. Rumney M. [et al.] LTE and the evolution to 4G wireless: design and measurement challenges. John Wiley & Sons, Ltd., 2013.
18. Hanzo L. [et al.] MIMO-OFDM for LTE, WiFi and WiMAX: coherent versus non-coherent and cooperative turbo-transceivers. John Wiley & Sons, Ltd., 2011.
19. Soliman M. [et al.] // International Journal of Electrical, Electronics and Data Communication. 2018. Vol. 6, Is. 9, P. 70–73.
20. Chang-Ying L. // 2018 4th Annual International Conference on Network and Information Systems for Computers (ICNISC), Wuhan, China, 2018, P. 101–110.
21. Puspitaningayu P., Hendrantoro G.// 2014 6-th International Conference on Information Technology and Electrical Engineering (ICITEE), Yogyakarta, 2014, P. 1–5.

УДК 621.391

ОЦЕНКА РАБОТЫ АЛГОРИТМА ВОЛНОВОГО ВЫРАЩИВАНИЯ ОБЛАСТЕЙ ЛОКАЛЬНЫХ МАКСИМУМОВ С ВЫБОРОМ ПИКСЕЛЕЙ В ПОРЯДКЕ УБЫВАНИЯ ЗНАЧЕНИЙ ДЛЯ РАЗЛИЧНЫХ ТИПОВ АСМ-ИЗОБРАЖЕНИЙ

В.В. РАБЦЕВИЧ, В.Ю. ЦВЕТКОВ

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 8 ноября 2020

Аннотация. Произведена оценка работы алгоритма волнового выращивания областей локальных максимумов с выбором пикселей в порядке убывания значений для различных типов изображений атомно-силового микроскопа. Показано, что предложенный алгоритм показывает значительно лучший результат для изображений, состоящих из множества слитно стоящих областей.

Ключевые слова: атомно-силовая микроскопия, волновое выращивание областей.

Введение

Различные узкоспециализированные изображения, применяемые для анализа различных явлений в медицине, металлургии, нанотехнологиях и т.д. требуют определенных алгоритмов для их обработки и последующей оценки результата. При анализе неорганических наноструктур, полученных с помощью атомно-силовой (далее АСМ) или сканирующей-зондовой микроскопии (далее СЗМ) необходимо учитывать, что исходными данными являются матрицы зондирования, содержащие информацию о высоте нахождения объектов на подложке, которую можно интерпретировать, как яркость [1]. Рассматриваемые изображения с позиции оценки работы алгоритмов сегментации можно разделить на три в зависимости от расположения наноструктур на подложке относительно друг друга: группа объектов, располагающихся на поверхности слитно; отдельно стоящие объекты; объекты со сложной топологией.

Для автоматической сегментации АСМ-изображений предлагается алгоритм волнового выращивания областей локальных максимумов с их выбором в порядке убывания значений (ВОЛМА) [2]. Сущность алгоритма состоит в использовании изменяющегося от максимума к минимуму порога яркости для выбора пикселей роста областей (локальных максимумов) или пикселей, присоединяемых к пикселям смежных существующих областей, которые имеют такую же или большую яркость. В отличие от обычного выращивания областей [3], использующего последовательную обработку сегментов, в предложенном алгоритме границы всех областей расширяются волнообразно за счет присоединения необработанных соседних значимых пикселей, яркости которых удовлетворяют порогу, понижаемому после обработки всех значимых пикселов минимуму.

Целью работы является оценка работы автоматической сегментации АСМ-изображений без предварительного выделения начальных точек роста областей на двух типах изображений.

Оценка эффективности алгоритма поиска локальных экстремумов на основе центрально-симметричного сканирования

Для проведения оценки была создана база тестовых изображений в среде Gwyddion [4], которые были разбиты на две большие категории: изображения первого типа (рис. 1 *а, б*), имеющие отдельно стоящие элементы на подложке и второго типа (рис. 1 *в, г*), имеющие слипшиеся группы элементов на подложке. Для реализации 205 тестовых АСМ-изображений с

объектами разных свойств и размеров использовался Gwyddion и инструмент «синтезировать». Каждое из них содержит различные по размеру, особенностям и взаимному расположению объекты. Трехмерные модели некоторых изображений представлены на рис. 1.

Для оценки работы алгоритма использовались такие параметры как мера однородности признака внутри сегмента, контраст на границе сегментов, комплексный критерий и количество сегментов [5].

Критерий однородности основан на значении дисперсии признака внутри сегментов и вычисляется по формуле

$$U_{Mz} = 1 - \sum_{S_i \in Mz} \frac{\omega_i \sigma_i^2}{N}, \quad (1)$$

где $\omega_i = \frac{A_i \sigma_i^2}{A_{Mz} \sigma_{\max}^2}$ – вес, определяющий вклад сегмента в АСМ-изображение, A_i – площадь рассматриваемого сегмента, A_{Mz} – площадь всего региона интереса, σ_i^2 – дисперсия признака (здесь и далее под признаком понимается значение яркости или высоты для канала topography, АСМ-изображений) в рассматриваемом сегменте, которая определяется по формуле (2), $\sigma_{\max}^2$ – максимальное значение дисперсии признака, определяется по формуле (3), N – нормирующий множитель

$$\sigma_i^2 = \sum_{i \in S} \frac{(f_i - \overline{f_i})}{A_i}, \quad (2)$$

$$\sigma_{\max}^2 = \frac{1}{2}(f_{\max} - f_{\min})^2, \quad (3)$$

где $f_{\max}$, $f_{\min}$ – максимальные и минимальные значения признака на всем регионе интереса (для тестовых изображений – вся матрица зондирования Mz), $\overline{f_i}$ – среднее значение яркости в рассматриваемом сегменте. Данная мера используется для расчета комплексного критерия.

Контраст на границе соседних сегментов определяется как

$$C_{Mz} = \frac{\sum_{S_i \in Mz} \omega_i c_i}{\sum_{S_i \in Mz} \omega_i}, \quad (4)$$

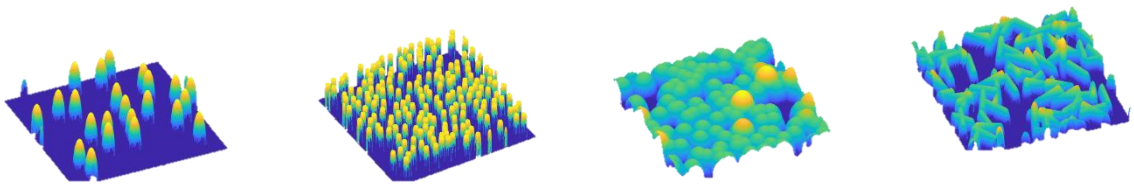
где $c_i = \sum_{S_i} p_{ij} c_{ij}$, – контраст рассматриваемого сегмента, p_{ij} – коэффициент смежности рассматриваемого сегмента; $c_{ij} = \frac{\left|\overline{f_i} - \overline{f_j}\right|}{\overline{f_i} + \overline{f_j}}$ – контраст двух соседних сегментов, $\omega_i = \frac{A_i \sigma_i^2}{A_{Mz} \sigma_{\max}^2}$ – вес, определяющий вклад сегмента в АСМ-изображение, A_i – площадь рассматриваемого сегмента, A_{Mz} – площадь всего региона интереса, σ_i^2 – дисперсия признака (здесь и далее под признаком понимается значение яркости или высоты для канала topography, АСМ-изображений) в рассматриваемом сегменте, которая определяется по формуле (2) $\overline{f_i}$ – среднее значение яркости в рассматриваемом сегменте.

Для конечной оценки работы алгоритмов сегментации будет применен комплексный критерий, учитывающий однородность признака внутри сегментации и их количество

$$Q = \frac{1}{10000N} \sqrt{R} \sum_{i=1}^{R} \left[\frac{e_i^2}{1 + \log A_i} + \left(\frac{R(A_i)}{A_i} \right)^2 \right], \quad (5)$$

где N – количество пикселей на изображении, R – количество сегментов, A_i – площадь i-го сегмента, e_i – величина, характеризующая степень однородности i-го сегмента, $R(A_i)$ – количество сегментов, имеющими площадь A.

Полученные статистические данные для двух режимов работы ВОЛМА (с использованием остановки на уровне подложки и без) представлен в табл. 1 (для тестовых АСМ- изображений первого типа) и табл. 2 (для тестовых АСМ-изображений второго типа).



а *б* *в* *г*

Рис. 1. Трехмерная поверхность некоторых тестовых изображений, синтезированных в Gwyddion: *а, б* – изображения, относящиеся к первому типу; *в, г* – изображения, относящиеся ко второму типу

Таблица 1. **Статистическая оценка результатов сегментации тестовых изображений первого типа алгоритмом ВОЛМА**

Сегмен-тация	Статистические показатели	Количество сегментов	Однородность сегмента	Контраст на границе сегментов	Комплексный критерий
ВОЛМА без остановки	Дисперсия	157188,5050	0,0078	0,0037	7,03E-07
	Медиана	116,0000	0,7432	0,1315	1,58E-05
	Среднее значение	299,3714	0,7549	0,1446	1,05E-04
	Стандартное отклонение	396,4700	0,0886	0,0605	8,38E-04
ВОЛМА с остановкой	Дисперсия	154805,7842	0,0011	0,0171	1,05E-02
	Медиана	116,0000	0,9176	0,9945	2,05E-06
	Среднее значение	298,1524	0,9307	0,9466	1,96E-02
	Стандартное отклонение	393,4537	0,0326	0,1308	1,03E-01

Таблица 2. **Статистическая оценка результатов сегментации тестовых изображений второго типа алгоритмом ВОЛМА**

Сегмен-тация	Статистические показатели	Количество сегментов	Однородность сегмента	Контраст на границе сегментов	Комплексный критерий
ВОЛМА без остановки	Дисперсия	17126,0464	0,0112	0,0041	1,21E-09
	Медиана	85	0,8150	0,1506	8,99E-06
	Среднее значение	127,0714	0,8087	0,1615	2,47E-05
	Стандартное отклонение	130,8665	0,1059	0,0642	3,48E-05
ВОЛМА с остановкой	Дисперсия	17135,3269	0,0021	0,0464	3,19E-01
	Медиана	85,0000	0,9452	0,5791	4,37E-08
	Среднее значение	127,1531	0,9401	0,5404	5,76E-02
	Стандартное отклонение	130,9020	0,0457	0,2155	5,64E-01

Как видно из полученных данных алгоритм волнового выращивания областей показывает лучший результат по однородности признака внутри сегментов (0 – минимальное значение, 1 – максимальное) при применении алгоритма ко второму типу изображений. Контраст на границе сегментов (0 – минимальное значение критерия, 1 – максимальное), показывает лучший результат для первого типа изображений, однако глобальный контраст АСМ-изображений на порядок выше для первого типа изображений за счет большей площади контрастного фона, на котором находятся элементы. Комплексный критерий показывает лучший результат (чем

меньше, тем лучше) для изображений второго типа, вне зависимости от режима использования алгоритма.

Заключение

Показана работа волнового алгоритма выращивания областей применительно для тестовых изображений двух типов. Установлено, что алгоритм ВОЛМА показывает лучшие результаты по комплексному критерию для второго типа изображений более чем в 10 раз (по среднему значению на выборке). Так же возрастает показатели критерия однородности внутри сегмента. Установлено, что для повышения точности работы алгоритма применительно для первого типа изображений необходимо дополнительно использовать критерий остановки, основанный на отделении фона от основных значимых элементов. При увеличении площади объектов и при уменьшении общей площади контрастной подложки – качество работы алгоритма возрастает.

EVALUATION OF THE WORK OF THE ALGORITHM OF WAVE GROWTH OF LOCAL MAXIMUM REGIONS WITH THE CHOICE OF PIXELS IN ORDER OF DECREASING VALUES FOR DIFFERENT TYPES OF AFM IMAGES

V.V. RABTSEVICH, V.Yu. TSVIATKOU

Abstract. The work of the algorithm of wave growth of the regions of local maxima with the choice of pixels in decreasing order of values for various types of images of an atomic force microscope was evaluated. It is shown that the proposed algorithm shows a significantly better result for images consisting of many contiguous areas.

Keywords: atomic force microscopy, wave growing areas.

Список литературы

1. Рабцевич В.В. [и др.] // Телекоммуникации: сети и технологии, алгебраическое кодирование и безопасность данных. 2019. С. 58–63.
2. Рабцевич В.В. [и др.] // Вес. Нац. акад. навук Беларусі. Сер. фіз.-тех. навук. 2019. № 2. С. 215–231.
3. Adams R., Bischof L. // IEEE Transactions on Pattern Analysis and Machine Intelligence 1994. Vol. 16, No. 6. P. 641–647.
4. Gwiddion [Electronic resource]. URL: http://gwyddion.net.
5. Захаров А.В. [и др.] // Труды НИИСИ РАН. 2012. Т. 2, № 2. С. 87–99.

UDC 621.396

WEB-APPLICATIONS VULNERABILITIES TESTING TECHNIQUE

M. ABOUKRA, N.V. NASONOVA

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Submitted 9 November 2020

Abstract. The technique for detecting vulnerabilities in web applications is developed. The technique combines exploration, GrayBox and testing principles to detect the riskiest vulnerabilities rated by OWASP.

Keywords: Web-applications, vulnerabilities, BlackBox, WhiteBox testing.

Introduction

An unsafe program is a potential target for an attacker who can use existing vulnerabilities to unauthorized access to available information, influence the operation of programs and services (start or stop), and inject malicious code into the system [1]. Software vulnerabilities are mainly caused by logical errors and software errors.

Adversary model

A web application testing methodology is needed to protect a company's web resources from the most common, dangerous, malicious threats associated with programming and configuration errors in web applications. First an adversary model should be developed. Assume, that the attacker has no access to the application source code. It can interact with it indefinitely (it is on the same local network or the application is available via the Internet and there is no firewall). This option is common. Also, assume, that the attacker understands the architecture of web applications and is qualified enough to perform the attack. The task of the vulnerabilities testing technique is to find the most critical vulnerabilities, spending a small number of resources.

Application vulnerabilities

There are several popular classifiers of vulnerabilities [2]:

– CWE (Common Weakness Enumeration) – a database of vulnerability types. The main task of the project is to provide descriptions of common types of vulnerabilities, ways to prevent, detect and fix them;

– CVE (Common Vulnerabilities and Exposures) – a vocabulary of specific vulnerabilities of software;

– SecurityFocus BID;

– IBM ISS X-Force.

The CWE classifier contains generalized classes of vulnerabilities that are often found in software products [3]. The main threats to web application security are shown in Fig. 1.

The most common at the moment are the following web application vulnerabilities: XSS (Cross Site Scripting), various injections (PHP, SQL), RCE (Remote Code Execution), errors in the implementation of authentication and authorization, and unsafe deserialization of objects [3].

Vulnerability detection methods

Successful testing of web applications requires a systematic approach or methodology. The most popular are OWASP and WASC. They are the most complete and formalized methodologies to date.

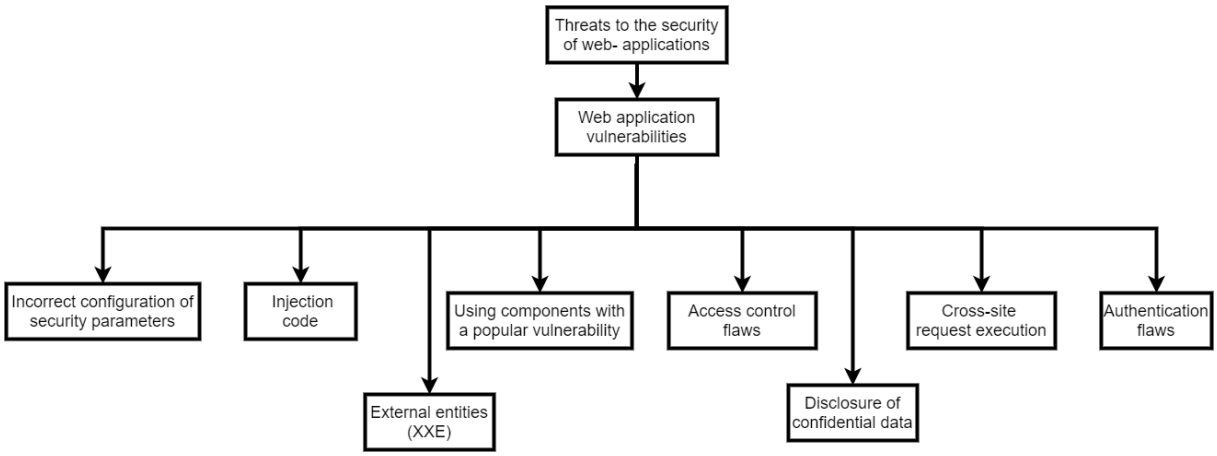


Fig. 1. The main threats to the security of web applications

There are several testing principles that we can apply to detect vulnerabilities [4]:

– DAST – dynamic (i.e. requiring execution) analysis of the application without access to the source code and the server side, the so called BlackBox testing;

– SAST – static (i.e., not requiring execution) analysis of an application with access to the source code of the web application and to the web server; in fact, it is an analysis of the source code for formal signs of vulnerabilities and a server security audit;

– IAST – dynamic analysis of the security of a web application, with full access to the source code, the web server – WhiteBox testing;

– Source code analysis – static or dynamic analysis with access to the source code without access to the server environment.

The following techniques are used for BlackBox testing:

– equivalent partition;

– boundary value analysis;

– analysis of cause and effect relationships;

– assumption of error.

This technique detects the following categories of errors:

– incorrectly implemented or missing features;

– interface errors;

– errors in data structures or organization of access to external databases;

– behavioral errors or insufficient system performance.

BlackBox testing is limited by its disadvantages: only a very limited number of program execution paths are tested and some tests may be redundant if they have already been conducted by the developer at the unit testing level. It is also quite difficult to create effective test cases without a clear specification. Moreover, this testing method has a number of advantages, as testing is done from the perspective of the end user and can help to detect inaccuracies and inconsistencies in the specification. The tester does not need to know programming languages and delve into the specifics of the program implementation. Testing can be performed by experts independent of the development department, which helps to avoid bias.

The opposite of the BlackBox technique is WhiteBox testing. WhiteBox testing is divided into the following levels:

– unit testing;

– integration testing;

– regression testing;

– hacking testing.

WhiteBox testing can be done at an early stage: there is no need to wait for the creation of the user interface. More thorough testing can be done, covering a large number of program execution paths. But at the same time, this method also has disadvantages, as it requires a lot of specialized knowledge to perform WhiteBox testing. When using test automation at this level, maintaining test scripts can be difficult if the program changes frequently.

The main stages of web application penetration testing include intelligence, access control, selection of parameters, checking the logic of the web application, checking the server environment. Having a test plan for an application, you can step by step examine all its components for the presence of certain vulnerabilities. Based on the specifics of each web application, certain points can be supplemented with checks specific to this application.

Vulnerability scanners are used to facilitate security testing. They allow you to avoid many routine actions and significantly reduce the time required for testing as shown in Fig. 2.

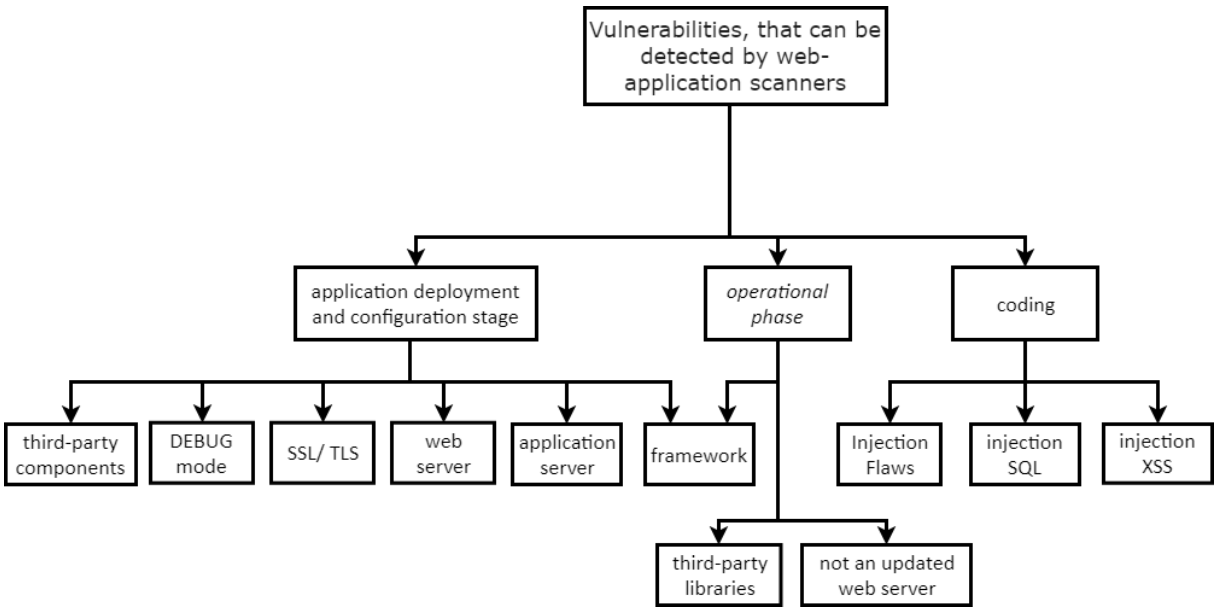


Fig. 2. Vulnerabilities detection using web application scanners

The task of searching for vulnerabilities in the development phase is much more successful when using the scanner in manual mode, rather than automatically. The task of searching for vulnerabilities of the second and third types is automated quite well (especially if the scanner is configured for application technologies and its environment). In manual mode, the operator works with the analyzed application through a web browser, which is connected to the application through the scanner as through an intermediate proxy server. In this case, the operator observes all HTTP requests and HTTP responses in the scanner and, at his choice, can compose and send the necessary test requests based on them. When working with a scanner in this mode, there are the following features: firstly, increased requirements are imposed on the qualifications of the operator, and secondly, it is important for the operator to get the maximum of convenient tools from the scanner to automate the creation and sending of test requests and analyze application responses to them. Burp Suite and ZAP Proxy are typical representatives of this class of scanners. In the automatic mode of operation, the operator is only required to configure the parameters for launching the scanner: specify the application URL and username/password (when testing the closed part). Typical representatives of this class are HP WebInspect, Acunetix, etc.

Development of a technique for detecting vulnerabilities in web applications

The developed technique stages are given in Fig. 3. Vulnerability search starts with exploration. For these purposes, we use the nmap network scanner. The next stage includes the so-called GrayBox testing. It includes both BlackBox and WhiteBox testing elements. First, we conduct an automatic scan for vulnerabilities using ZAP. It also helps you to discover entry points to the application and facilitate further exploration. To bypass authorization, we use dictionary search. As a dictionary we take the top 10,000 most frequently encountered passwords, which allows us to sort out most weak passwords in a relatively short time. Then, after the automated search, we manually test the detected entry points to the application (web pages that accept user input and API endpoints). For manual testing, the application source code should be accessed. These steps will help you to find most of the basic vulnerabilities that an attacker can detect. Then it is necessary to conduct WhiteBox testing, which will make it possible to detect more complex vulnerabilities. The algorithm of the proposed technique is shown in Fig. 3.

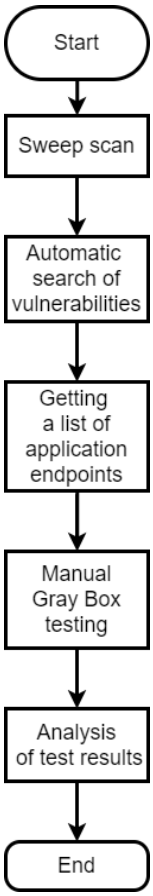


Fig. 3. Algorithm of the developed methodology for searching for vulnerabilities

Thus, this technique allows you to combine the existing methods of searching for vulnerabilities, which makes it possible to detect the most critical vulnerabilities inherent in web applications, the criticality of which is rated by OWASP [5].

Conclusion

A technique for detecting vulnerabilities in web applications is developed. The technique combines exploration, GrayBox and testing principles to detect the riskiest vulnerabilities rated by OWASP. It is suggested to be applied by the companies for their inside testing of the vulnerabilities of the proprietary applications.

References

1. Introduction to Secure Coding Guide [Electronic source]. URL: https://developer.apple.com/library/archive/documentation/Security/Conceptual/SecureCodingGuide.
2. Top 25 Software Errors [Electronic source]. URL: https://www.sans.org/top25-software-errors.
3. BlackBox and WhiteBox testing [Electronic source]. URL: https://quality-lab.ru/blog/key-principles-of-black-box-testing/.
4. Damn Vulnerable Web Application (DVWA) [Electronic source]. URL: http://www.dvwa.co.uk/.
5. OWASP Broken Web Applications [Electronic source]. URL: https://owasp.org/www-project-broken-web-applications/.

УДК 621.391

ИССЛЕДОВАНИЕ ХАРАКТЕРИСТИК РЕЧЕВОГО КОДЕРА ДЛЯ БЫСТРОГО ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ НА ОСНОВЕ АЛГОРИТМА МАЛЛА

М.С. АНТОНЕНКО, Т.М. ПЕЧЕНЬ

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 7 ноября 2020

Аннотация. Рассмотрен метод компрессии, основанный на применении дискретного вейвлет-преобразования для обработки речевого сигнала. Предложена модель речевого кодера для быстрого вейвлет-преобразования, основанного на кратномаштабном анализе с использованием алгоритма Малла и исследованы его характеристики.

Ключевые слова: вейвлет-преобразование, алгоритм Малла, речевой кодер, компрессия.

Введение

Основной вид коммуникации между людьми – это передача речи. По этой причине актуальным остается совершенствование методов цифровой обработки и передачи речевых сообщений [1]. Новые средства обработки сигналов и соответствующие специализированные инструментальные программные среды создают возможности появления эффективных систем передачи и обработки речевых сообщений [2]. Компрессия и кодирования речи позволяют решать сложные проблемы по минимизации числа бит необходимых для передачи сигнала.

В настоящее время одной из важных задач является создание систем низкоскоростной передачи с высоким качеством восприятия сигнала, способных корректно функционировать в реальных условиях окружающей среды [3]. Решение проблемы компрессии речевого сигнала особенную роль играет в таких сферах как Интернет-телефония, сотовая связь, запись и хранение речевых сообщений в специальных и портативных устройствах [4].

При низких скоростях передачи информации параметрическое кодирование более эффективное. В данном виде кодирования в качестве параметров используют различные характеристики представления сигнала в частотной области – спектральные кодеры, использующие вейвлет-преобразование [5].

Одномерное вейвлет-преобразование

Малла предложил эффективный алгоритм реализации дискретного вейвлет-преобразования, который представляет собой классическую схему обработки: входной одномерный сигнал раскладывается на высокочастотные и низкочастотные компоненты с использованием пары ортогональных фильтров.

В основе преобразования заложено, что любую функцию $s(t)$ можно рассматривать на любом m' -уровне разрешения. На этом уровне между ее усредненными значениями и флуктуациями вокруг средних значений для разделения функции используем формулу

$$s(t) = \sum_{k=-\infty}^{\infty} C_{m'}, k(t) + \sum_{m=m'}^{\infty} \sum_{k=-\infty}^{\infty} D_m, k\psi_m, k(t). \quad (1)$$

Как видно из (1) при бесконечных пределах первая сумма стремится к нулю и поэтому может быть опущена. В результате получаем «чистое» вейвлет-преобразование. При решении практических задач цифровые данные представлены в виде конечного набора отсчетов. Таким

образом, наилучший уровень разрешения определяется интервалом, в котором содержится один отсчет. В результате суммирование будет выполняться в конечных пределах. Значение $m = 0$ как правило принимается для наилучшего уровня разрешения. Принятая форма вейвлетов $\varphi_{m,k} = 2m / 2\varphi(mt - k)$ усредняет отсчеты. Это происходит при уменьшении значений $m = 0, -1, -2, \ldots$. Для удобства расчетов обычно исключается использование отрицательных индексов масштабирования. В этом случае знак «–» обычно вводится непосредственно в функции вейвлетов $\varphi_{m,k} = 2 - m/2\varphi(2 - mt - k)$. После выполнения данной операции вычисляются вейвлет-коэффициенты для значений $m > 0$.

Такой кратномасштабный анализ для случая с последовательным увеличением значений m способствует получению алгоритма, состоящего с быстрых итерационных вычислений:

$$C_{m+1}, k = \sum_n h_n C_m, 2k + n, \tag{2}$$

$$D_{m+1}, k = \sum_n g_n C_m, 2k + n, \tag{3}$$

$$C_0, k = \int_{k\Delta t}^{(k+1)\Delta t} s(t)\,\varphi\,(t - k)\,dt. \tag{4}$$

Алгоритм Малла представляет собой пирамидальный алгоритм вычисления вейвлет-коэффициентов. В этом случае вейвлеты Добеши обеспечивают наиболее приемлемые условия для обработки речевого сигнала. Уравнение (4) используются для случаев если известна аналитическая форма функции $s(t)$.

Поясним суть операций, выполняемых формулами (2) и (3). На первой итерации преобразования цифровой фильтр h_n из сигнала $s_k = C_0$, k выделяет низкие частоты, а октавный фильтр g_n выделяет верхние частоты. На выходе фильтра h_n отсутствует верхняя половина частот. Выполняется децимация выходного сигнала: осуществляется по формуле (2) [4]. На выходе фильтра g_n освобождается место в области низких частот, и аналогичное прореживание выходного сигнала приводит к транспонированию верхних частот на освободившееся место [5].

Таким образом, каждый из выходных сигналов несет информацию о своей половине частот, при этом выходная информация представлена таким же количеством отсчетов, что и входная [3].

Восстановление сигнала

Поскольку в формулах (2), (3) вместо базисных функций используются фильтры, то обратные преобразования, т.е. последовательную сборку сигнала от больших m к малым и реконструкцию сигналов по значениям его вейвлет-коэффициентов с любого уровня разрешения, имеет смысл также выразить через фильтры реконструкции

$$C_{m-1} = \sum_{n \in I} C_{m,n} hr(k) - 2n + \sum_{n \in I} D_{m,n} gr(k) - 2n. \tag{5}$$

Алгоритм вычислений по формуле (5) является обратным алгоритму декомпозиции. Таким образом, аппроксимируются коэффициенты C_m и D_m в 2 раза меньший шаг дискретизации с двукратным увеличением частоты Найквиста и восстановлением спектра коэффициентов C_m в низкочастотную часть нового главного диапазона спектра $C_m - 1$, а спектра коэффициентов D_m в высокочастотную часть спектра $D_m - 1$. Далее выполняется расстановка нулевых значений между коэффициентами C_m и D_m, фильтрация полученных массивов низкочастотным $hr(k)$ и высокочастотным $gr(k)$ фильтрами реконструкции, и сложением результатов фильтрации. Модули частотных характеристик фильтров $hr(k)$ и $gr(k)$ должны повторять модули

частотных характеристик фильтров $hr(k)$ и $gr(k)$ [3]. Однако, фильтры декомпозиции $hr(k)$ и $gr(k)$ являются односторонними и фазосдвигающими, и при реконструкции коэффициентов $C_m - 1$ этот сдвиг фазы должен быть исключен. Этого можно достичь реверсом значений коэффициентов фильтров декомпозиции по следующей формуле [4]

$$hr(k) = \text{reverse}(h(k)),\ gr(k) = \text{reverse}(g(k)). \quad (6)$$

Точность преобразования сигналов зависит от потерь информации при выполнении прореживания спектров. Главным образом, эти потери возникают на срезах полос пропускания фильтров низких и высоких частот, крутизна которых зависит от порядка фильтров, их согласованности, и типа вейвлетных функций [2].

Обязательным условием преобразования сигнала является его задание количеством отсчетов, равном $N = 2m$, где значение $m \geq 1$ определяет максимально возможное число уровней декомпозиции сигнала при целочисленных значениях кратности сдвигов операторов фильтров количеству отсчетов вейвлетных коэффициентов на каждом уровне декомпозиции. Требованием является выполнение условия количество отсчетов сигнала равного $N = 2m$ [4].

Структурная схема декомпозиции и реконструкции речевого сигнала

Структурные схемы декомпозиции и реконструкции сигнала алгоритма Малла представлены на рис. 1.

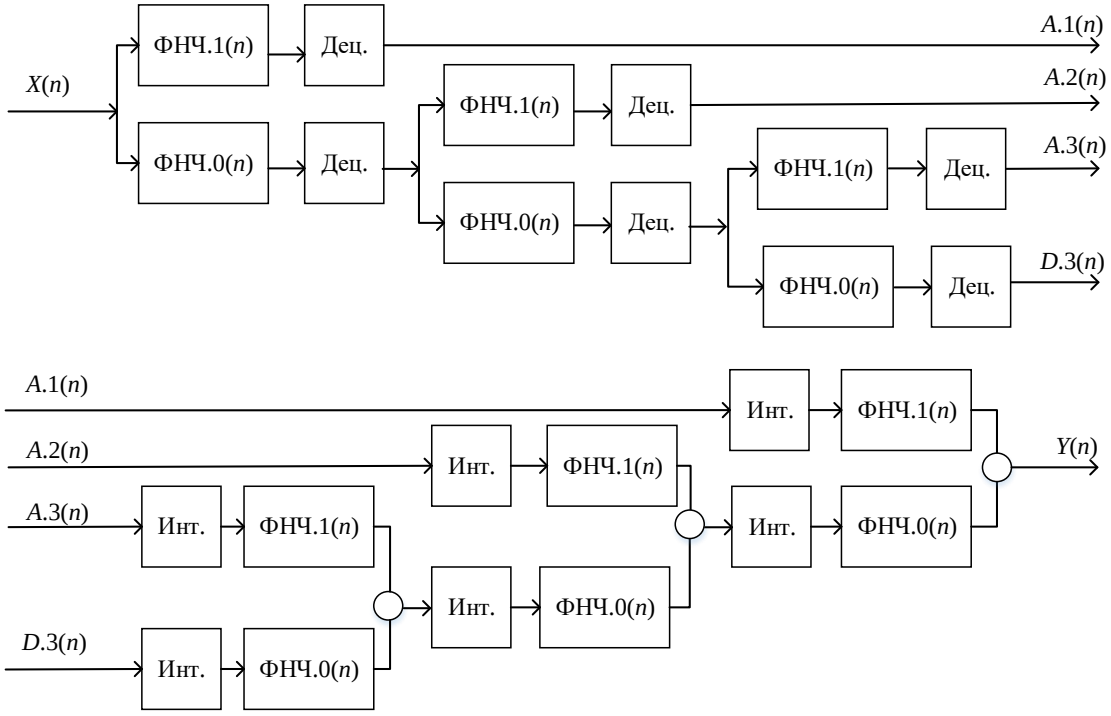


Рис. 1. Структурные схемы декомпозиции и реконструкции сигнала

Алгоритм основан на четырех операциях: фильтрации, децимации, интерполяции и суммирования. Функционально фильтры представляют собой низкочастотные и высокочастотные фильтры с конечной импульсной характеристикой (КИХ-фильтры). Дециматор (Дец.) – это устройство, которое позволяет уменьшать частоту дискретизации в 2 раза. Интерполятор (Инт.) повышает частоту дискретизации в 2 раза.

Моделирование процесса кодирования декодирования речевого сигнала

Моделирование работы системы кодер/декодер производилось в специализированной программе MATLAB R2016b с использованием реальных речевых сигналов, ранее записанных в файл. Схема моделирования работы системы кодер/декодер представлена на рис. 2.

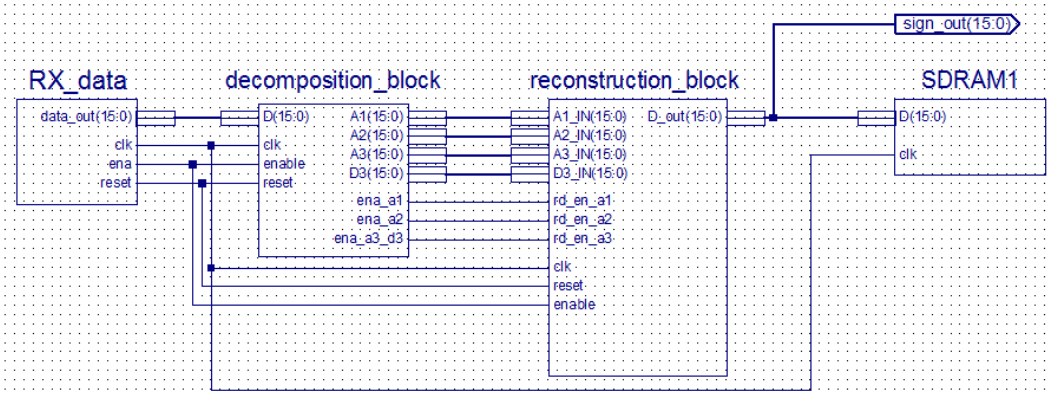


Рис. 2. Схема моделирования работы системы кодер/декодер

Как видно из рис. 2 в состав схемы моделирования работы системы кодер/декодер входят следующие блоки:

– RX_data – блок для генерации управляющих сигналов;

– decomposition_block – блок декомпозиции;

– reconstruction_block – блок реконструкции сигнала;

– SDRAM1 – блок сохранения результата работы сигнала.

Входной блок моделирования генерирует сигналы «clk», «enable», «reset», а также на вход основной схемы подает входные отсчеты. Отсчеты сигнала читаются из файла, который был создан в MATLAB R2016b. Для моделирования был взят wav-аудио файл с записью голоса. С помощью функции «wavread» были считаны значения сигнала. На рис. 3 представлена осциллограмма исходного речевого сигнала.

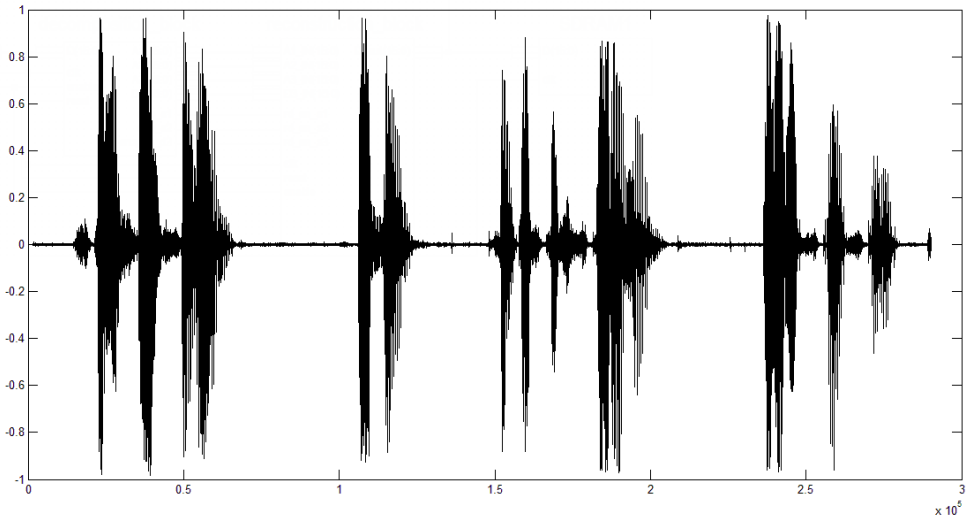


Рис. 3. Осциллограмма исходного речевого сигнала

Файл, содержащий отсчеты сигнала, называется «speech_data.txt». Выходной блок моделирования SDRAM записывает отсчеты восстановленного сигнала в файл. Отсчеты сигнала записываются в файл «SIGNAL_OUT.txt».

Поскольку амплитуда сигнала находится в пределах от –1 до 1, поэтому необходимо умножить каждый отсчет на 32767, для перевода в диапазон от –32767 до 32767. Затем числа преобразовываем в двоичный вид и записываем в файл.

Рассмотрим листинг *m* -файла (wav_ADC) данной процедуры:

```
fSpData = fopen('speech_data.txt','w');
Y = wavread('D:/test.wav');
Z = 0;
for ii = 1:290304
  tmp = round(Y(ii) * 32767);
if tmp < 0
    Z = dec2bin(abs(tmp),15);
    Z = strcat('1',Z);
    fprintf(fSpData,'%s\n',Z);
else
    Z = dec2bin(tmp,16);
    fprintf(fSpData,'%s\n',Z);
end;
  Z =0;
end;
fclose(fSpData);
```

Затем пропустив сигнал через разработанную схему и, сохранив в файл результаты в блоке SDRAM1, обратно преобразуем в wav-файл. Листинг *m*-файла (wav_DAC) для преобразования обратно в wav-файл.

```
fSpData = fopen('SIGNAL_OUT.txt','r')
dout = 0;
for ii = 1:290304
  strTmp = fscanf(fSpData,'%s',[1]);
  decTmp = bin2dec(strTmp(2:16));
  decTmp1 = decTmp/32767;
if strTmp(1) == '1'
    dout(ii) = decTmp1 *(-1);
else
    dout(ii) = decTmp1;
end;
end;
fclose(fSpData);
```

На рис. 4 представлена осциллограмма восстановленного сигнала.

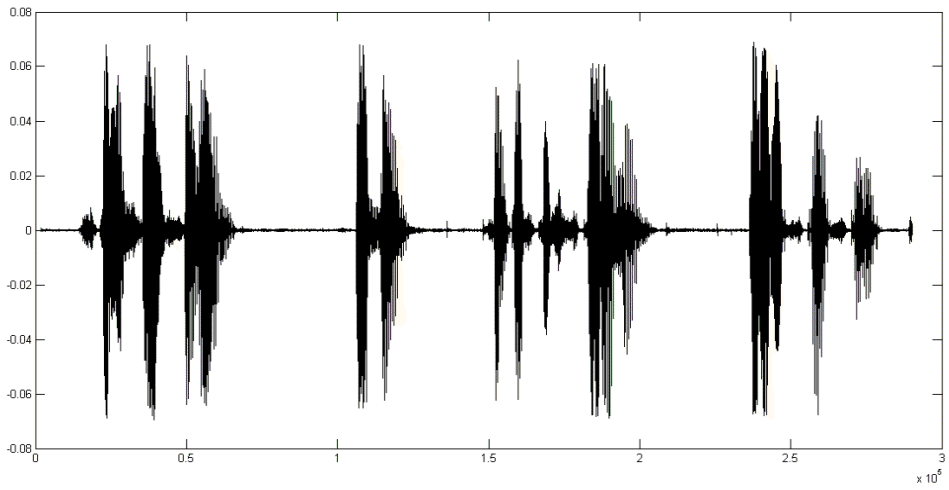


Рис. 4. Осциллограмма восстановленного сигнала

Для оценки качества восстановления сигнала оценим разность вход/выход сигналов и корреляция вход/выход.

Разность исходного сигнала и полученного представлена на рис. 5.

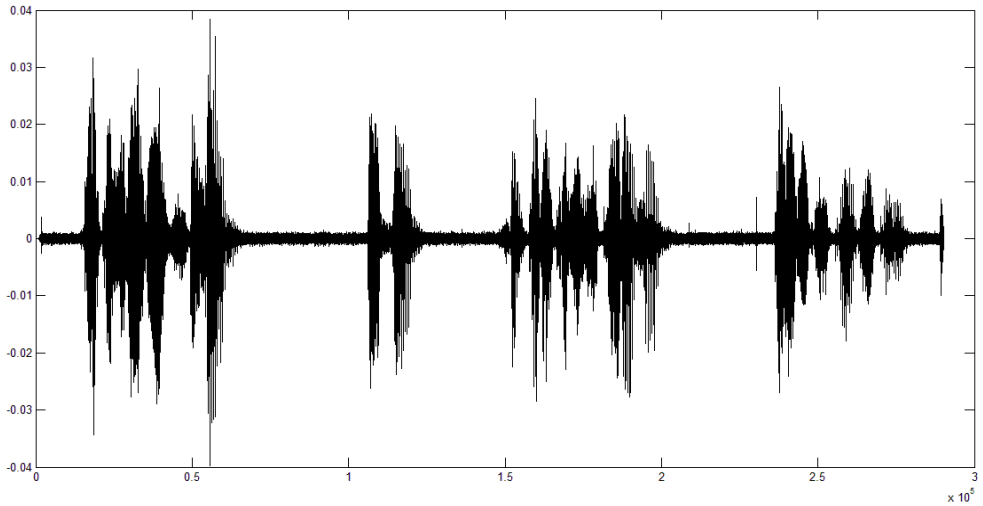


Рис. 5. Разность исходного и полученного сигнала

Из графика видно, что ошибки достигают 0,04 по амплитуде, а относительно исходных сигналов $0,04/0,9 = 0,044 = 4\%$. Эти ошибки нормальны, так как происходило преобразование коэффициентов фильтра к целым двоичным числам до 2^3, а также при реконструкции происходило деление, отбрасывались младшие разряды.

На рис. 6 показан ненормированный уровень корреляции двух исследуемых сигналов.

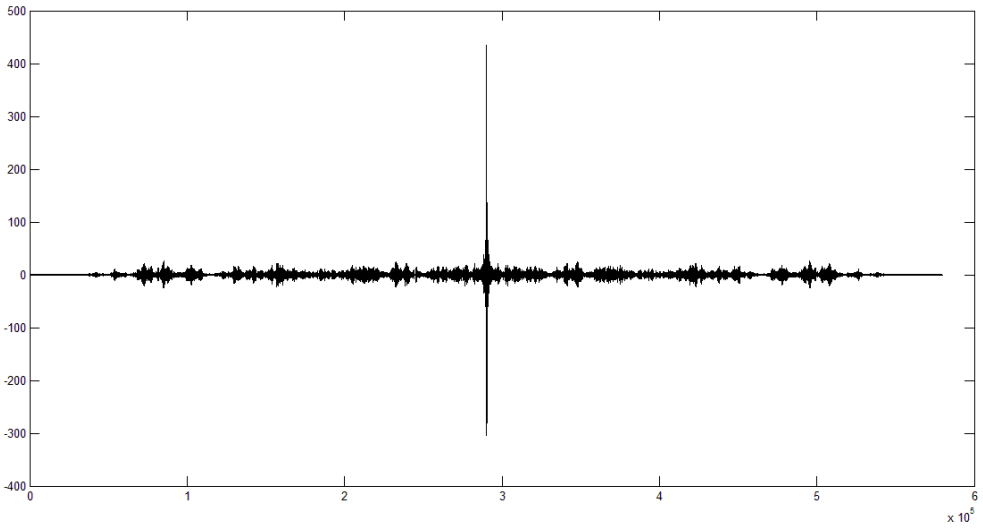


Рис. 6. Уровень корреляции исходного и полученного сигнала

Как видно из представленного графика на рис. 6, имеется пик в точке со значением по оси абсцисс равным $2,9 \cdot 10^5$. Это подтверждает явное сходство сигналов.

Заключение

Проведено моделирование системы кодер/декодер в специализированной программе MATLAB R2016b с использованием реального речевого сигнала. В результате установлена практически полная идентичность преобразованного и восстановленного сигналов, что позволяет судить о высоком качестве восстановления. Точность восстановления составила величину примерно равную 4 %. Алгоритм при этом обеспечивает трехуровневую компрессию речевого сигнала. Разработанный модуль кодирования/декодирования может использоваться при создании высокоэффективных кодеков речевого сигнала для низкоскоростных линий коммуникаций.

INVESTIGATION OF THE CHARACTERISTICS OF SPEECH ENCODER FOR FAST WAVELET TRANSFORMATION BASED ON THE MALL ALGORITHM

M.S. ANTONENKO, T.M. PECHEN

Abstract. A compression method based on the use of a discrete wavelet transform for processing a speech signal is considered. A model of a speech encoder for fast wavelet transform based on multiple-scale analysis using the Mall algorithm is proposed and its characteristics are investigated.

Keywords: wavelet transform, the Mall algorithm, speech encoder, compression.

Список литературы

1. Graf M., Truong H.L. // Computer networks. 1999. Vol. 31, Issue 3. P. 273.
2. Куньсянь Л., Франц Дж.А., Саймар мл. Р. // Цифровые процессоры обработки сигналов серии TMS320. 1987. Т. 75, № 9. С. 8–27.
3. Kondoz A.M. Digital speech: coding for low bit rate communication systems. NY.: John Wiley & Sons, 1996.
4. Makhoul J., Roucos S., Gish H. // Proc. IEEE 1985 Nov. Vol. 73, 1551–1588.
5. Das A., Gersho A. // Int. J. of Speech Technology. 1999. Vol. 2., P. 317–327.

УДК 336.203

РАБОТА С БОЛЬШИМИ ДАННЫМИ И БАЗОЙ ДАННЫХ В СЕТИ ИНТЕРНЕТА ВЕЩЕЙ

В.А. ВИШНЯКОВ, С.К. ЭЛЬ ХАДЖИ

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 9 ноября 2020

Аннотация. Представлены десять характеристик сущности больших данных (БД) для работы в сетях интернета вещей (ИВ). Рассмотрены особенности не реляционных баз данных и приведены характеристики трех наиболее распространенных из них: Apache Cassandra, MongoDB, HBase. На примере БД Cassandra даны концепции для хранения больших данных. Рассмотрены ключевые различия в моделировании данных для БД Cassandra по сравнению с реляционной базой данных. Приведен пример построения БД Cassandra для обработки в сети ИВ информации от десяти датчиков качества продукции.

Ключевые слова: большие данные, не реляционные базы данных, сеть ИВ.

Введение

Большие данные – это область приложений, которая рассматривает способы анализа, систематического извлечения информации из наборов данных, которые слишком велики или сложны для обработки традиционными (реляционными) системами управления базами данных.

Большие данные обладают специфическими характеристиками и свойствами, которые могут помочь понять как проблемы, так и преимущества инициатив в области больших данных. В работе [1] авторы представляют свою модель больших данных − 10V, в которой обсуждают пять дополнительных характеристик, добавляя их к пяти, ранее известных, для описания больших данных. Рассмотрим все десять характеристик.

1. Объем. Является наиболее известной характеристикой больших данных.

2. Скорость. Скорость – это быстрота, с которой данные генерируются, производятся, создаются или обновляются.

3. Разнообразие. Когда речь заходит о больших данных, необходимо обрабатывать не только структурированные данные, но и полу структурированные данные.

4. Изменчивость. Первое – это количество несоответствий в данных. Большие данные также изменчивы из-за множества измерений данных, возникающих из множества различных типов и источников данных. Изменчивость может относиться к несогласованной скорости, с которой большие данные загружаются в базу данных.

5. Правдивости. По мере увеличения любого или всех вышеперечисленных свойств достоверность (уверенность или доверие к данным) падает.

6. Действия. Как и достоверность, валидность относится к тому, насколько точны и корректны данные для их предполагаемого использования. Поэтому необходимо принять эффективные методы управления данными, чтобы обеспечить согласованное качество данных, общие определения и метаданные.

7. Уязвимость. Большие данные порождают новые проблемы безопасности. В конце концов, нарушение данных с большими данными – это большое нарушение. Сбор и хранение больших данных часто приводит к утечкам данных.

8. Волатильность. Волатильность относится к принятию решения о том, сколько лет должны быть данные, прежде чем они будут считаться неактуальными, историческими или бесполезными больше, и как долго данные должны храниться.

9. Визуализация. Еще одна особенность больших данных заключается в том, насколько сложно их визуализировать. Современные средства визуализации больших данных сталкиваются с техническими проблемами из-за ограничений технологии in-memory и плохой масштабируемости, функциональности и времени отклика.

10. Значение. Другие характеристики больших данных бессмысленны, если не извлекается из них бизнес-ценность. В больших данных можно найти существенную ценность, оптимизировать процессы и улучшать производительность машин или бизнеса.

Базы данных NoSQL

Базы данных NoSQL (не SQL или не только SQL) стали стандартной платформой данных и основной промышленной технологией для работы с огромным ростом объема данных. Концепция баз данных NoSQL [2] была предложена для эффективного хранения и обеспечения быстрого доступа к большим наборам данных, объем, скорость и изменчивость которых трудно решить с помощью традиционных систем управления реляционными базами данных. Базы данных NoSQL используют различные структуры данных (например, ключ-значение, широкий столбец, документ, график и т.д.) По сравнению с табличными реляционными базами данных эти базы данных не содержат схем, поддерживают легкую репликацию, имеют простой API, согласованы и могут обрабатывать огромные объемы данных.

На рынке существует множество баз данных NoSQL, различные отраслевые тенденции свидетельствуют о том, что Apache Cassandra входит в тройку лучших используемых сегодня вместе с MongoDB и HBase [3].

Apache Cassandra – это распределенная и децентрализованная система хранения данных с открытым исходным кодом для управления очень большими объемами структурированных данных, распределенных по нескольким центрам обработки данных. Она обеспечивает высоко доступное обслуживание без единой точки отказа.

Apache HBase – это не реляционная распределенная база данных с открытым исходным кодом, созданная по образцу BigTable от Google и написанная на Java. Она разработан как часть проекта Apache Hadoop и работает поверх HDFS, предоставляя BigTable-подобные возможности для Hadoop.

MongoDB – это кроссплатформенная документо-ориентированная система баз данных, которая не используем традиционную табличную структуру реляционных баз данных в пользу JSON-подобных документов с динамическими схемами, что упрощает и ускоряет интеграцию данных в определенные типы приложений.

Гипотеза CAP [4] определяет компромисс между доступностью системы, согласованностью и допуском разбиения, утверждая, что только два из трех свойств могут быть сохранены в распределенных реплицированных системах одновременно.

База данных Cassandra

Cassandra предлагает надежную поддержку кластеров, охватывающих несколько центров обработки данных, с асинхронной репликацией без мастера, позволяющей выполнять операции с низкой задержкой (особенно для записи/обновления) [5].

Согласованность в Cassandra может быть настроена таким образом, чтобы обеспечить компромисс между доступностью и задержкой в сравнении с точностью данных. Уровень согласованности чтения определяет, сколько узлов реплики должны отвечать на запрос. Уровень согласованности записи определяет количество реплик, на которых запись должна завершиться успешно, прежде чем клиент получит подтверждение. Cassandra поддерживает два типа операций записи с небольшой разницей между ними: insert и update. Эта БД использует следующие концепции для хранения данных.

1. Кластер – это совокупность узлов или центров обработки данных, расположенных в кольцевой архитектуре. База данных Cassandra распределена по нескольким машинам, которые работают вместе. Самый внешний контейнер называется кластером. Для обработки сбоев каждый узел содержит реплику, и в случае сбоя реплика берет на себя ответственность. Кассандра упорядочивает узлы в кластере в кольцевом формате и назначает им данные.

2. Перебора всех вариантов. Пространство ключей – это самый внешний контейнер для данных в Cassandra.

3. Семейство столбцов – это контейнер для упорядоченной коллекции строк (строка – единица репликации). Каждая строка, в свою очередь, представляет собой упорядоченную коллекцию столбцов. Семейства столбцов в Cassandra подобны таблицам в реляционных базах данных. Каждое семейство столбцов содержит коллекцию строк, которые представлены картой <RowKey, SortedMap<ColumnKey, ColumnValue>>. Ключ дает возможность получить доступ к связанным данным вместе; в отличие от реляционных таблиц схема семейства столбцов не является фиксированной, и Cassandra не заставляет отдельные строки иметь все столбцы. Пользователь может свободно добавлять любой столбец в любое семейство столбцов в любое время.

Семейство столбцов Cassandra имеет следующие атрибуты:

– keys_cached − он представляет собой количество местоположений, которые будут храниться в кэше для каждого SSTable;

– rows_cached − представляет собой количество строк, все содержимое которых будет кэшироваться в памяти;

– preload_row_cache − указывает, хотите ли вы предварительно заполнить кэш строк.

4. Колонки. Столбец – это базовая структура данных Cassandra с тремя значениями, ключ или имя столбца, значение и отметка времени; это единица хранения в Cassandra. Столбцы и количество столбцов в каждой строке могут отличаться в отличие от реляционной базы данных, где данные хорошо структурированы.

5. Суперколонки. Суперстолбец – это специальный столбец, поэтому он также является парой ключ-значение. Но суперколонка хранит карту подколонок, как правило, семейства столбцов хранятся на диске в отдельных файлах.

6. Коллекции. В отличие от концепций внешних ключей и соединений, используемых реляционными базами данных, отношения в Cassandra представлены с помощью коллекций. Cassandra избегает объединения между двумя таблицами, сохраняя адреса электронной почты пользователя в столбце коллекции в таблице user. Каждая коллекция определяет тип хранящихся данных и может быть одной из следующих трех: набор, список, карта.

БД использует язык Cassandra Query Language (CQL) как простой интерфейс для доступа, который является альтернативой традиционному языку структурированных запросов (SQL), используемому СУБД. CQL добавляет слой абстракции, который скрывает детали реализации этой структуры и предоставляет собственные синтаксисы для коллекций и других распространенных кодировок. Языковые драйверы доступны для Java (JDBC), Python (DBAPI2), Node.JS (Helenus), C++ и т.д.

Основные правила моделирования данных в БД Cassandra

Ключевые различия в моделировании данных для Cassandra по сравнению с реляционной базой данных включают следующее [6].

1. Не присоединяется. Нельзя выполнять соединения в Cassandra. Если при разработке модели данных, нужно что-то вроде соединения, придется либо выполнить работу на стороне клиента, либо создать денормализованную вторую таблицу, представляющую результаты соединения.

2. Отсутствие ссылочной целостности. Хотя Cassandra поддерживает такие функции, как облегченные транзакции и пакеты, в ней нет понятия ссылочной целостности между таблицами. В реляционной базе данных можно указать внешние ключи в таблице, чтобы ссылаться на первичный ключ записи в другой таблице, такие операции, как каскадное удаление, недоступны.

3. Денормализация. При проектировании реляционных баз данных важно понятие нормализации. Это не является преимуществом при работе с Cassandra, она работает лучше всего, когда модель данных денормализована.

4. Запрос. Реляционное моделирование, означает, что работа начинается с концептуальной области, а затем представляются сущности в этой области в таблицах. Затем назначаются первичные ключи и внешние ключи для отношений в модели. Когда есть отношение «многие ко многим», создаются таблицы соединений, которые представляют только эти ключи. Таблицы

соединений не существуют в реальном мире и являются необходимым побочным эффектом работы реляционных моделей. В Cassandra начинается работа не с модели данных, а с модели запросов. С помощью Cassandra моделируются запросы и данные должны быть организованными вокруг них.

5. Проектирование для оптимального хранения. Поскольку таблицы Cassandra хранятся в отдельных файлах на диске, важно, чтобы связанные столбцы определялись вместе в одной таблице.

6. Сортировка дизайнерское решение. В Cassandra сортировка трактуется иначе; это дизайнерское решение. Порядок сортировки, доступный для запросов, является фиксированным и полностью определяется выбором столбцов кластеризации, предоставленных в команде CREATE TABLE. Оператор CQL SELECT поддерживает порядок по семантике, но только в порядке, указанном столбцами кластеризации.

Разработка базы данных Cassandra для сети Интернета вещей

Система IoT включает в себя 10 датчиков качества продукции (температура). Каждый из датчиков посылает данные 1 раз в секунду. Таким образом, основные требования можно сформулировать следующим образом:

– 1 система должна продолжать запись, если один узел перестает работать;

– 2 система должна записывать 10 новых записей в секунду, несмотря на возможные сбои узла/сети;

– 3 система должна сообщать обо всех измерениях, собранных любым датчиком за указанный день в течение нескольких миллисекунд;

– 4 система должна как можно быстрее сообщать обо всех измерениях, собранных любым датчиком в течение указанного периода времени.

Проектирование базы данных. Cassandra выполняет запись гораздо быстрее, чем другие базы данных SQL и NoSQL, поэтому она может быть идеальным выбором для работы с десятками запросов на запись в секунду. Кроме того, для удовлетворения требований надежности разработчики могут выбрать мультиреплицированный режим.

Архитектура кластера Cassandra (например, коэффициент репликации = 3), в которой один из узлов развернут в облаках. Для выполнения третьего требования можно принять решение о хранении всех значений, собранных в течение одного дня, в одном необработанном виде. Для этого нам нужно создать составной первичный ключ, который состоит из времени события используемого в качестве ключа кластеризации, а также составного ключа раздела, включая текущую дату и идентификатор датчика.

```
CREATE TABLE temperature_events_by_day (
day text,
sensor_id uuid,
event_time timestamp,
temperature double,
PRIMARY KEY ((day, sensor_id), event_time)
)
WITH CLUSTERING ORDER BY event_time DESC;
```

В этом фрагменте day – это текст в формате: «ГГГГ-ММ-ДД»; (day, sensor_id) – представляет собой комбинированный ключ (формируются по дате показания всех датчиков); event_time – это ключ кластеризации (в кластере – значения датчиков в конкретный момент времени в секундах). Благодаря обратной сортировке внутри строки (с порядком кластеризации по event_time DESC) можно получать самые важные (последние) данные всех датчиков.

Ключ раздела используется для идентификации раздела или узла в кластере, в котором хранится эта строка. Когда данные считываются или записываются из кластера, для вычисления хэш-значения ключа раздела используется функция Partitioner. Это хэш-значение используется для определения узла/раздела, содержащего эту строку.

Ключ кластеризации предназначен для хранения данных строк в отсортированном порядке. Сортировка данных осуществляется по столбцам, которые входят в ключ

кластеризации. Такое расположение позволяет эффективно извлекать данные с помощью ключа кластеризации.

Поиск ключа раздела, который является day/sensor_id, является очень быстрой операцией в Cassandra. Таким образом, предлагаемая модель данных удовлетворяет третьему требованию.

Для реализации четвертого требования потребуется поиск определенного дня (дней) на первом этапе и сравнения временных меток внутри каждого дня на втором этапе. Однако это может занять некоторое время, если база данных включает собранные данные в течение нескольких дней. Учитывая тот факт, что в системе всего 10 датчиков можно рассмотреть возможность создания второго ключевого пространства, где каждая строка соответствует определенному датчику, игнорируя дни.

```
CREATE TABLE temperature_events_by_day (
sensor_id uuid,
event_time timestamp, temperature double,
PRIMARY KEY (sensor_id, event_time)
)
WITH CLUSTERING ORDER BY event_time DESC;
```

В этом фрагменте sensor_id – это ключ раздела (формируются показание конкретного датчика); event_time – это ключ кластеризации (в строке – значение датчика в конкретный момент времени).

Благодаря обратной сортировке внутри строки (с порядком кластеризации по event_time DESC) можно получать самые важные (последние) данные конкретного датчика.

Заключение

1. Представлены десять характеристик сущности больших данных (БД) для работы в сетях интернета вещей (ИВ). Рассмотрены особенности не реляционных баз данных и приведены характеристики трех наиболее распространенных из них: Apache Cassandra, MongoDB, HBase. На примере БД Cassandra даны концепции для хранения данных, такие как кластер, набор ключей, семейство столбцов, колонки, суперколонки, коллекции.

2. Рассмотрены ключевые различия в моделировании данных для Cassandra по сравнению с реляционной базой данных, включающие соединения, ссылки, денормализацию, запрос, хранение и сортировку.

3. Рассмотрен пример построения БД Cassandra для обработки в сети ИВ информации от датчиков качества продукции. Приведены элементы работы с БД, используя язык Cassandra Query Language.

WORKING WITH BIG DATA AND DATABASE IN THE INTERNET OF THINGS

U.A. VISHNYAKOU, S.K. EL-HAJJ

Abstract. Ten characteristics of the big data entity (DB) for working in the Internet of things (IoT) networks are presented. The features of non-relational data bases are considered and the characteristics of the three most widely distributed such databases are given: Apache Cassandra, MongoDB, and HBase. Using the example of Cassandra DB, concepts for storing big data are given. Key differences in data modeling for the Cassandra DB compared to a relational database are considered. An example of building the Cassandra data base for processing information from ten product quality sensors in the IoT network is given.

Keywords: big data, non-relational data bases, IoT network.

Список литературы

1. Firican G. The 10 Vs of Big Data. 8 February 2017. [Electronic resource]. URL: https://tdwi.org/articles/2017/02/08/10-vs-of-big-data.aspx.
2. Evans E. NoSQL 2009. 12 May 2009. [Electronic resource]. URL: http://blog.sym-link.com/2009/05/12/nosql_2009.html.
3. Github. Benchmarking Cassandra and other NoSQL databases with YCSB [Electronic resource]. URL: https://github.com/cloudius- systems/osv/wiki/Benchmarking-Cassandra-and-other-NoSQL-databases-with-YCSB.
4. Brewer E. Towards Robust Distributed Systems in 19-th Annual ACM Symposium on Principles of Distributed Computing, Portland, USA, 2000.
5. DataStax. Apache Cassandra 2.1 for DSE. About data consistency. 14 February 2018. [Electronic resource]. URL: https://docs.datastax.com/en/cassandra/2.1/cassandra/dml/dmlAboutDataConsi stency.html.
6. Carpenter J., Hewitt J. Cassandra: The Definitive Guide. O'Reilly Media, 2016.

UDC 004.852

VULNERABILITIES OF DEEP LEARNING BIOMETRIC VOICE APPLICATIONS IN BANKING AND FINANCIAL SERVICES

S.N. PETROV, O.S. ELSAYED, T.A. PULKO

The Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Submitted 8 November 2020

Abstract. The analysis of the market of virtual digital voice assistants is carried out. It is established that the popularity of such applications in the fintech industry is growing, especially starting in 2019. Made a review of the most popular voice assistants used in banks. The main vulnerabilities and threats are analyzed.

Keywords: speech recognition, neural network, voice technologies, biometric technologies, voice spoofing, virtual voice assistants.

Introduction

The usage of biological measurements or Biometrics' traits as key values for smartly identifying someone or verifying their claim of being a certain person, increasingly through means of Artificial Intelligence and Deep Learning methods, is enthusiastically being propagated as other remote technologies in the post-coronavirus pandemic are, abruptly being introduced into the lives of the ordinary people within their regular interactions and daily activities, and gradually obtaining popularity as the «New Normals» among both consumers and services providers across multiple different areas and applications, such as Automotive, Healthcare, Education, Governmental & Public Services, Legal & Forensic, Military & Defense, Consumer Electronics & IoT Gadgets, and Retail, E-Commerce and the Banking and Financial Services and Insurance .

Speech and Voice Recognition market is expected to grow between the years 2019–2025 almost 17,2 % CAGR (Compound Annual Growth Rate) reaching an estimated 26,8 billions dollars by 2025. And the selected Biometric measures when ranking by Consumer Preference was: voice recognition (32 %), fingerprints (27 %), facial scan (20 %), hand geometry (12 %), and iris scan (10 %), where high preference to most convenience and familiarity when choosing the biometric technology.

This growing is contributed equally by the consumers themselves and by driver and orchestrated by the Authorities i.e. both the Service Regulators & Service Providers in the Fintech (Financial Technology) sector to comply with such mandatory, caused by number of interconnected reasons pushing both parties to participate in this cycle of pushing forward towards further broader adoption of advanced smart voice-based technologies and solutions; Governments and Regulators are both strictly obliged to comply with the anti-money-laundry and criminal funding acts like the American KYC regulation or the European PSD2 regulation both with punitive compensations of multi million dollars for the entities that fall short of fulfilling them.

On the other hand, voice-enabled devices are driving the interest of several well established banking and ICT market players and start-ups equally towards developing consumer-centric solutions for this market highlighted by the convince and none-intrusive features of the technology. Voice recognition technologies are increasingly being considered as relatively cost-effective and convenient mechanisms to gain access or exercise control over different types of connected devices that are part of smart homes, connected cars, and other smart technology segments in the Internet of Things Ecosystem.

However, it is very important to keep in mind no system is perfect and the same applies to Machine or Deep-Learning-based systems, which in themselves are based on statistical and mathematical methods with scientifically forecasted margins of errors.

Specially with the COVID-19 era, when due to the various degrees mobility restrictions of individuals that ranged from simple social distancing measures and guideline to be followed strictly by people and businesses in their daily work-habits, to partial or full local and regional lockdowns.

The person themselves being the password or key of access is not new and as technology is spreading it is being used more and more in the daily lives of ordinary including their daily interactions and activities. Banking and financial services being a daily activity for people it was not very far behind in the following of the trend, this was specially capitalized and with the outbreak of the Coronavirus pandemic and the consequent lockdowns that followed. To meet the minimum requirements of banks and monitory and in the need to authenticate and identify their customers/clients.

Speech technologies are widely used in the Fintech sphere. There are several main areas of use for biometrics systems.

1. Speaker Identification. Automatically verify and authenticate speakers in seconds by using their voice as a highly accurate biometric identifier and provide every client with a truly immersive call experience.

2. Fraud Detection. Prevent yourself from fraudsters hiding behind someone else's identity with a speaker verification system running in the background marking suspicious speakers whose voices don't match.

3. Defense and Security. Identify and quickly search for speakers in the large quantities of audio recordings to stay ahead of crime and perform detailed forensic voice analyses faster with the help of AI-powered technology.

A large and fast-growing market is the use of virtual voice assistants, which greatly simplifies the management of a Bank account, making this process similar to normal communication with a human operator. Company «Tractica» defines a virtual digital assistant as an automated software application or platform that assists humans through understanding natural language in written or spoken form and leverages some form of artificial intelligence in doing so [1]. According to their research, the virtual digital assistant market will grow rapidly further (Fig. 1).

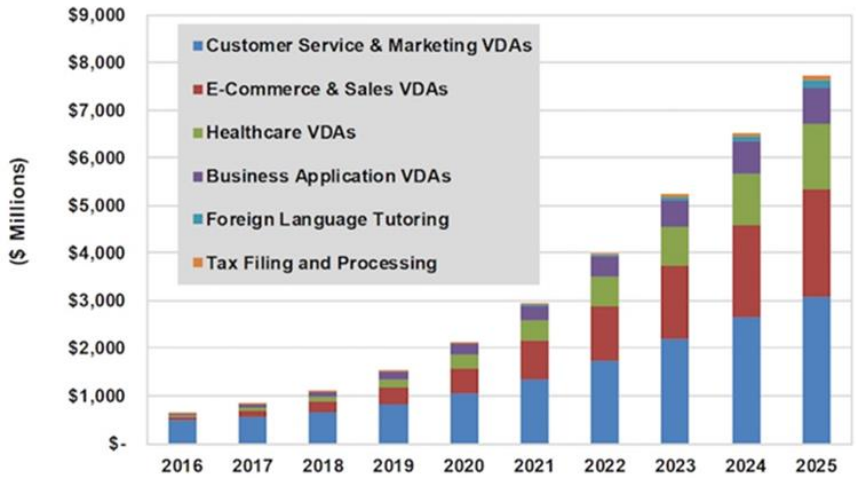


Fig. 1. Enterprise virtual digital assistant software revenue by use case, 2016–2025

Many financial organizations, mainly in the US, are already using voice-activated virtual assistants to access accounts and make payments. Under the Alexa skills category of «banking and finance» (such as YNAB, Business Voice Apps, Create My Voice, Marketplace, Commercial Trends and others) more than 3000 results come up, including American Express, Capital One, PayPal and US Bank.

Some banks have also integrated Apple's Siri, including Mashreq Bank in the UAE. In 2016, Mashreq started allowing customers to transfer payments of up to Dh500 using Siri. Mashreq Bank customers can tell Siri how much they want to transfer and to whom, but have to authenticate by using either their pin or touch ID fingerprint recognition.

In 2017, UK bank Barclays rolled out a similar concept, allowing customers to ask Siri to make a payment and then authenticate it with Apple's touch ID.

OCBC Bank in Singapore has offered voice banking through Google Assistant on a smartphone or Google Home device since April 2018. Customers can mainly inquire about the bank's services and plan their financial future; for example, they can calculate the mortgage loan amount they can afford.

Bank of America, which serves more than 65 million consumer and clients, created its own virtual assistant Erica in June 2018.

Biometrics' voice recognition systems vulnerabilities

However, such systems have certain disadvantages. Many challenges connected with accuracy level. Voice-based systems have very high error rates, especially false rejections.

Another problem is speaker verification spoofing. According to [2] most biometric systems are vulnerable to imposture. Spoofing attacks are performed on a biometric system at the sensor or acquisition level to bias score distributions toward those of genuine clients, increasing the False Acceptance Rate (FAR) [3].

The most common options for implementing spoofing today are:

1. Impersonation. Impersonation refers to spoofing attacks with human-altered voices and is one of the most obvious forms of spoofing.

2. Replay attacks, using speech recordings of a genuine client, or concatenation of shorter segments. The equal error rate (EER) of 1 % can increase to 70 % using replayed spoof attacks.

3. Voice conversion, which is a technique that electronically converts one speaker's voice towards that of another.

4. Speech synthesis. In this approach a speech synthesizer is used which is adapted to the voice of genuine clients. Using an HMM-based speech synthesizer, the FAR can rise up to 91 %.

Conclusion

Virtual voice assistants are widely used in the banking sector. The integration of such systems allows you to reduce customer service time, increase security, but at the same time, these solutions currently have the disadvantages inherent in all probabilistic technologies (FAR and FRR), and are also subject to targeted attacks such as voice spoofing.

References

1. Virtual Digital Assistant Software to Reach $7.7 Billion and 1 Billion Users in 2025 [Electronic source]. URL: https://voicebot.ai/2018/02/01/virtual-digital-assistant-software-reach-7-7-billion-1-billion-users-2025/.

2. Spoofing countermeasures for the protection of automatic speaker recognition from attacks with artificial signals, F. Alegre, R. Vipperla and N. Evans, INTERSPEECH 2012, 13th Annual Conference of the International Speech Communication Association [Electronic source]. URL: http://www.eurecom.fr/en/publication/3731/download/mm-publi-3731.pdf.

3. Spoofing and countermeasures for automatic speaker verification, N. Evans, T. Kinnunen, J. Yamagishi, INTERSPEECH 2013, 14th Annual Conference of the International Speech Communication Association [Electronic source]. URL: https://www.researchgate.net.

UDC 621.396

INFORMATION SECURITY REQUIREMENTS FOR A SMALL BUSINESS COMPANY

LIANG JINHUI, N.V. NASONOVA

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Submitted 8 November 2020

Abstract. The most common types of threats, as well as attacks and intrusion methods were analyzed. Adaptive Security Architecture model was used to effectively mitigate the considered threats. Using the information classes and its characteristics, information security requirements have been developed for storing and processing critical information in the network of a small business.

Keywords: information security attacks and intrusions, information security requirements, small business companies.

Introduction

The number of external attacks on the infrastructure of organizations is growing significantly. This problem is especially relevant now, when many companies are in a hurry to transfer employees to remote work. Hackers look for any open breach in systems at the perimeter of the network, for example, a forgotten unsecured web application, not updated software, or an incorrectly configured server with a weak administrator password. The larger the compromised company and the higher the privileges obtained, the more profitable the transaction can be made by the criminal.

The legal definition of «small business» is not strictly fixed and varies by country and by industry. Commonly it depends on a number of employees, annual sales, value of assets and net profit, alone or as a combination of factors. In most of the countries the number of employees in a small business company is assumed up to 50–250 persons.

Relevance of information security for a small business company

It is widely believed that the problem of cyberattacks by low-skilled hackers (the so-called script kiddies) is more relevant for small companies that are not ready to invest heavily in protecting their resources. Large organizations invest much more in information security and, it would seem, should be better protected. But penetration testing demonstrates the vulnerability of even large companies. Refer to the statistics collected by Positive technologies for 2019 [1]. Information about the most common types of threats, as well as attacks and intrusion methods based on the analysis of statistics, is presented in the table.

Suspicious network activity was detected in the infrastructure of 97 % of companies. In 28 % of companies, the activity of a number of utilities and tools was revealed, which may indicate a compromise. There is a trend towards «living off the land» attacks. Also, attempts to discredit or destroy a competitor's business led to an increase in DDoS attacks. In 81 % of organizations, in-depth analysis of network traffic revealed malware activity. Multiple attempts to connect to external servers on port 445/TCP (SMB) indicate a malware infection. But miners and adware were more common in the infrastructure. Thirty-three percent of companies use dictionary passwords, which gives rise to a large number of brute-force attacks or dictionary attacks. At the network perimeter of 11 % of companies, the main security problem is inadequate protection of web applications. Thus, after analyzing the results of the above statistics, we can conclude that the security of the corporate network, and therefore the continuity of the technological process, directly depends on the efficiency of administration of networks

and network equipment, as well as the timely installation of relevant security updates for the software used. The statistics analysis results are given in Table 1.

Table 1. **Types of attacks and intrusions in corporate networks**

Categories of identified threats	Percentage of companies (%)	Intrusion methods	Types of attacks
Suspicious network activity	97	– hiding traffic; – network scanning; – attempts to remotely start the process; – collection of information about active network sessions on nodes, users, groups, password policy, etc.; – perimeter scan.	– DDOS; – IP spoofing; – living off the land.
Violation of IS parameters	94	– use of unprotected data transfer protocols; – use of software for remote access; – using BitTorrent; – open network ports on the perimeter.	– phishing; – mailbombing; – SPAM; – telephone phreaking; – pre-texting.
Malware activity	81	– miners; – adware; – spyware; – reading arbitrary files.	– WannaCry; – worms; – viruses; – Trojan horse; – spyware; – ransomware.
Attempts to exploit software vulnerabilities	28	availability of exploits in the public access	Rootkit
Password guessing attempts	19	– using dictionary passwords; – saved authentication parameters; – obtaining and increasing privileges.	– brute force; – by dictionary.
Attempts to exploit web vulnerabilities	11	– using vulnerabilities in XML and WEB services; – forging cross-site requests.	– cookies infection; – SQL injection; – XSS Scripting.

When developing corporate information security systems for companies for which the cost of information leaks and other incidents may be the highest, it is necessary to rely on existing standards that have already proven themselves in this area of working with data and combating cybersecurity threats. In a narrower sense, the Adaptive Security Architecture (ASA) model, first introduced to the market in 2014 [2], will be a successful solution. Within the framework of this model, protection against targeted attacks is prioritized, but at the same time it provides the maximum possible defense against all internal and external threats provided for by the security policies of a particular organization. The ASA model assumes building a security architecture at four levels [3]:

– prediction (forecasting);
– warning (prevention);
– identification (detection) of threats;
– response.

First of all, when developing methods for protecting a corporate network, one should rely on the classification of data and processes circulating in the network [4, 5]. This is necessary in order to understand what exactly is to be protected and what are the priority areas of protection, since without prioritization the proper level of protection will not be achieved. It is necessary to find out which systems need to be protected in the first place, that is, those without which the organization's action will simply stop. Also, determine the systems on the protection of which you can save money or not at all. For small businesses, the following types of information circulating in the company's network can be distinguished, which are subject to protection:

– client database with important confidential data (numbers of documents, cards, details of visa services);
– information about quotas, insurance payments and policies;
– accounting software, accounting, taxes;

– individual schemes and developments for attracting clients, details of promotions and other information constituting scientific, technical and technological information related to the company's activities;

– personal data of employees of the enterprise and partners, stored in the database and transmitted over the network;

– e-mail messages and database information containing service information, information about the activities of the enterprise, etc.

The assignment of security categories to an information network is based on an assessment of the damage that can be caused by security breaches. There are three main aspects of information security: availability, confidentiality, and integrity. IS violations can affect only a part of these aspects, just as safety regulators can be specific for certain aspects. Therefore, it is advisable to assess the possible damage separately for violations of accessibility, confidentiality and integrity, and if necessary, an integral assessment can be obtained. When categorizing an information system, the categories of information stored, processed and transmitted by means of IS are taken into account, as well as the value of the assets of the IS itself in accordance with the scale (Table 2) [5].

Table 2. **Types of attacks and intrusions in corporate networks**

Damage level	An impact, produced by loss of availability, confidentiality and / or integrity on the organization's operations, assets and people	Impact to the company’s business
High	a severe or catastrophic impact	the company loses the ability to perform all or some of its main functions
Moderate	a serious detrimental impact	the company remains able to fulfill the mission assigned to it, but the effectiveness of the main functions is significantly reduced
Low	a limited detrimental impact	the company remains capable of fulfilling the mission entrusted to it, but the effectiveness of the main functions is significantly reduced

Thus, in most commercial private small businesses, the following categories of confidential information can be distinguished:

– personal data;

– commercial information.

The most important aspect of the information security policy of any company is ensuring the security of personal data. The security of personal data is achieved by eliminating unauthorized, including accidental, access to personal data, which may result in the destruction, modification, blocking, copying, distribution of personal data, as well as other unauthorized actions. Statistics have shown that the largest number of threats observed in 2019 is associated with suspicious network activity detected in 97 % of organizations, which leads to the collection of information about active network sessions on nodes, about users, groups, password policies, etc. This includes an increase in the number of social engineering methods, respectively, an increase in such attacks as phishing, telephone phreaking, pre-texting, etc. Accordingly, these types of threats can lead to a violation of the confidentiality and availability of stored confidential information, including personal data [4].

Therefore, the following information security requirements have been developed for storing and processing confidential information in the network of a small business:

– information and related resources must be available to authorized users;

– differentiation of access of registered users to hardware, software and information resources of the network (the ability to access only those resources and perform only those operations with them that are necessary for specific users to perform their official duties);

– registration of user actions when using protected resources in system logs and periodic control of the correctness of system users by analyzing the contents of these logs;

– protection of personal data from leakage through technical channels;

– protection of personal data from unauthorized disclosure;

– compliance with the procedure for storing personal data on paper and electronic media;

– development of regulations for responding to incidents and related procedures;

– protection against unauthorized modification and control of the integrity of the software used in the network, as well as protection of the system from the introduction of unauthorized programs;

– implementation of mechanisms for automatic blocking of detected malware by removing them from program modules or destroying them;

– checking the integrity of anti-malware protection modules required for its correct functioning;

– availability of means for restoring the personal data protection system;

– allocation of a communication channel that ensures the protection of personal data;

– exchange of personal data, during their processing in the information system, through communication channels, the protection of which must be ensured by the implementation of appropriate organizational measures and (or) the use of technical means;

– restriction of unauthorized outgoing traffic from applications used to process, store or transmit confidential information and personal data;

– the use of strong encryption to protect confidential information and personal data, transmission or remote access to which is carried out using mobile and portable devices that support network authentication.

Conclusion

97 % of organizations reveal suspicious network and malware activity. Adaptive security architecture helps businesses stay ahead of cybercriminals, suggests flexible security measures to protect data and systems in as agile a way as possible, rather than relying on outdated perimeter defense strategies. Using the information classes and its characteristics, information security requirements have been developed for storing and processing critical information in the network of a small business.

References

1. Current Cyber Threats: Results of 2019 [Electronic source]. URL: https://www.ptsecurity.com/ru-ru/research/analytics/cybersecurity-threatscape-2019/.

2. What is Adaptive Security Architecture? [Electronic source]. URL: https://adaptivesecurityarchitecture247.wordpress.com/2016/04/16/what-is-adaptive-security-architecture-2/

3. Weise J. Designing an Adaptive Security Architecture. Sun BluePrints Online, 2008.

4. Top Cybersecurity Threats in 2020 [Electronic source]. URL: https://onlinedegrees.sandiego.edu/top-cyber-security-threats/

5. Vacca J.R. Computer and Information Security Handbook. Morgan Kaufmann, 2017.

СВЕДЕНИЯ ОБ АВТОРАХ

1.	Абукара Мохамед Абдусалям	– магистрант кафедры защиты информации БГУИР
2.	Аксенов Вячеслав Анатольевич	– старший преподаватель кафедры инфокоммуникационных технологий БГУИР
3.	Алисеенко Маргарита Александровна	– аспирант кафедры инфокоммуникационных технологий БГУИР
4.	Антоненко Мария Сергеевна	– студент кафедры инфокоммуникационных технологий БГУИР
5.	Вишняков Владимир Анатольевич	– д.т.н., профессор кафедры инфокоммуникационных технологий БГУИР
6.	Дай Суан Лой	– исследователь Вьетнамского государственного технического университета им. Ле Куй Дона
7.	Жэнь Сюньхуань	– аспирант кафедры инфокоммуникационных технологий БГУИР
8.	Качан Дмитрий Александрович	– соискатель кафедры инфокоммуникационных технологий БГУИР
9.	Кийко Вадим Николаевич	– ассистент кафедры инфокоммуникационных технологий БГУИР
10.	Конопелько Валерий Константинович	– д.т.н., профессор кафедры инфокоммуникационных технологий БГУИР
11.	Кореневский Святослав Александрович	– к.т.н., доцент кафедры инфокоммуникационных технологий БГУИР
12.	Леонович Екатерина Федоровна	– выпускник кафедры инфокоммуникационных технологий БГУИР
13.	Лян Цзиньхуэй	– магистрант кафедры защиты информации БГУИР

14.	Ма Цзюнь	– аспирант кафедры инфокоммуникационных технологий БГУИР
15.	Михнюк Дмитрий Геннадьевич	– магистрант кафедры инфокоммуникационных технологий БГУИР
16.	Муравьев Валентин Владимирович	– д.т.н., профессор кафедры инфокоммуникационных технологий БГУИР
17.	Насонова Наталья Викторовна	– д.т.н., доцент кафедры защиты информации БГУИР
18.	Наумович Николай Михайлович	– к.т.н., начальник Центра 1.6 НИЧ БГУИР
19.	Нгуен Ань Туан	– аспирант кафедры инфокоммуникационных технологий БГУИР
20.	Петров Сергей Николаевич	– к.т.н., доцент кафедры защиты информации БГУИР
21.	Печень Татьяна Михайловна	– старший преподаватель кафедры инфокоммуникационных технологий БГУИР
22.	Пулко Татьяна Александровна	– к.т.н., доцент кафедры защиты информации БГУИР
23.	Рабцевич Виолетта Викторовна	– ассистент кафедры инфокоммуникационных технологий БГУИР
24.	Саломатин Сергей Борисович	– к.т.н., доцент кафедры инфокоммуникационных технологий БГУИР
25.	Сергеев Николай Николаевич	– аспирант кафедры инфокоммуникационных технологий БГУИР
26.	Смоляк Сергей Владимирович	– магистрант кафедры инфокоммуникационных технологий БГУИР
27.	Тарченко Надежда Владимировна	– к.т.н., доцент кафедры инфокоммуникационных технологий БГУИР

28.	Урядов Владимир Николаевич	– к.т.н., доцент кафедры инфокоммуникационных технологий БГУИР
29.	Хоменок Михаил Юлианович	– к.т.н., доцент кафедры инфокоммуникационных технологий БГУИР
30.	Хоминич Александр Леонидович	– старший преподаватель кафедры инфокоммуникационных технологий БГУИР
31.	Цветков Виктор Юрьевич	– д.т.н., заведующий кафедрой инфокоммуникационных технологий БГУИР
32.	Шайа Бахаа Хикмат	– аспирант кафедры инфокоммуникационных технологий БГУИР
33.	Эль Масри Абдель Хуссейн Али	– аспирант кафедры инфокоммуникационных технологий БГУИР
34.	Эль Хаджи Слейман Кхалед Кхалиль	– аспирант кафедры инфокоммуникационных технологий БГУИР
35.	Эльсайед Омния Сулиман Сид	– магистрант кафедры защиты информации БГУИР