Computer Science Journal of Moldova, vol.**29**, no.2(86), 2021

Cybercrime Detection Using Semi-Supervised Neural Network

Abbas Karimi, Saber Abbasabadei, Javad Akbari Torkestani, Faraneh Zarafshan

Abstract

Nowadays, artificial intelligence is widely used in various fields and industries. Cybercrime is a concern of these days, and artificial intelligence is used to detect this type of crime. Crime detection systems generally detect the crime by training from the related data over a period of time, but sometimes some samples in a dataset may have no label. Therefore, in this paper, a method based on semi-supervised neural network is presented regarding crime types detection. As the neural network is a supervised classification system, therefore, this paper presents a pseudo-label method for neural network optimization and develops it to semi-supervised classification. In the proposed method, firstly the dataset is divided into two sections, labelled and unlabelled, and then the trained section is used to estimate the labelling of the unlabelled samples based on pseudo-labels. The results indicate that the proposed method improves the accuracy, Precision and Recall up to 99.83%, 99.83% and 99.83%, respectively.

Keywords: cybercrime, intrusion detection, neural network, semi-supervised classification.

1 Introduction

Today, with the development of information and communication technology (ICT), cybercrime has become a global concern [1]–[5]. Two factors, including time of using the computer and activity in the social

network, have been identified as the main factors and predictors of cybercrime. Cybercrime analysis is an important responsibility of the law enforcement system in every country [6]–[8]. As the crime exists in different and separable cases, the adaptability of the discovered patterns has concerns and challenges. Classification is often used to predict the process of crime, which reduces the time of offender's identification [3]–[9]. Failure to identify the crime causes and the criminal abuse make the society unsafe [2]–[10]. The best model for preventing crime is reducing the chance of crime in the society [11]. Criminal behavior is the result of an appropriate opportunity to commit a crime at a particular place and time, and eliminating or reducing those opportunities leads to reduction of crime in that situation, so it is the most important factor in preventing a crime [12]–[14]. Therefore, crime prevention has always been one of the most important issues of life that has been practiced through various ways. Intrusion was referred to is a [15]–[17]. Intrusion, that was referred to, is a series of unlawful acts that endanger the accuracy, privacy or access to a resource [16], [18], [19]. The intrusion can be divided into internal and external. External intrusion is carried out from outside a network into internal network by authorized or unauthorized persons, and internal intrusion is carried out within a network by authorized persons [20], [21]. Intruders generally use software defects, decryption and network sniffing to penetrate computers and networks. In order to deal with intrusion, several methods have been developed called intrusion detection methods that monitor the events occurred in a computer system or network [2], [22]–[24]. Due to the development of ICT and the launch of comprehensive information systems in police force and criminal intelligence registering, data mining and knowledge discovery techniques are used to analyse and detect the cybercrime, especially intrusion [25]–[27]. So, predicting, preventing and detecting the cybercrime using the data mining is a fascinating new idea performed by statistical techniques, machine learning, artificial intelligence and criminology [28]. For expanding the classification of cybercrime, we can use algorithms of supervised machine learning such as artificial neural networks. These methods are also used in data mining [29]–[32]. The basic idea of artificial neural networks is inspired

by the way the biological system uses for learning and knowledge creation [31]. A key element of this idea is creation of new structures for the information processing system. This consists of a large number of highly interconnected processing elements called neurons that work together to solve a problem and transmit information through synapses. The learning is adaptive, namely the weight of the synapses is changed by using samples to generate a correct response [34], [35]. Neural network is used in medical diagnosis [36], [37], reconstruction of digital elevation model [38], intrusion detection [39], [40], etc. This is a supervised classification, but the labelling is expensive and time consuming, so today there are various ways to provide semi-supervised methods [33]. Therefore, in this paper, an optimized semi-supervised neural network method for computer crime detection is presented. Here we propose a method called pseudo-label (PL) [41] in artificial neural network. Unlike supervised learning, our proposed method uses labelled and unlabelled samples during training. For unlabelled samples it produces predicted labels, which measure the overlap of detection probability based on probabilistic conditional entropy. Assuming the probability of detecting each type of independent class, the predicted label is calculated with the maximum probability in training process. Since the estimated label values may be erroneous, a variable coefficient called influence coefficient is used to control its impact on the classification process. Therefore, a reduction criterion is added for unlabelled samples. So, the main contribution of this paper is as follows:

- Modelling and extraction of cybercrime patterns using data mining.

- Increasing the accuracy in crime detection using semi-supervised neural networks.

- Evaluating the proposed method using evaluation parameters.

The rest of the paper is as follows: Section 2 briefly reviews related works. Section 3 provides an overview of the proposed method. In Section 4, the evaluation of the proposed method is presented and com-

pared with other methods. Conclusions and future work are presented in Section 5.

2 Literature review

In recent years, several researches on crime and intrusion detection problem have been conducted. Most of them used data mining and machine learning. Intrusion is a cybercrime, so this section is divided into two separate subsections: cybercrime detection and intrusion detection.

2.1 Cybercrime detection

Qayyum et al. (2018) investigated data mining techniques for crime detection. Crime detection is one of the hot topics in data mining, where different patterns of criminology have been identified. Its various steps include identifying the crime characteristics to identify the pattern of the crime. Data mining techniques have been widely used for crime detection. An analytical study by extracting the strengths and weaknesses of each technique is presented [6].

In 2018, Mingcheng Feng et al. analyzed big data and used data mining for crime analysis and prediction. The purpose of this study was to analyze exploratory data for criminal data analysis in the cities of San Francisco, Chicago, and Philadelphia. They examined data time series and predicted crime trends over the coming years. Experimental results indicate that the decision tree classification model fits better than k-NN and Bayesian approaches. These promising achievements will be useful for police and law forces to expedite the process of crime detection and provide insights that enable police and law forces to trace criminal activities, predict the likelihood of crime occurring, use resources effectively and make faster decisions [42]. Dutta et al. in 2017 investigated the detection of impersonation crime using data mining. This study mainly focuses on credit card related impersonation crime which is very common and costly nowadays. Current data mining techniques are not able to eliminate impersonation, and new data mining is

suggested to combat these crimes. This new method uses both public detection and Spike detection algorithms to detect fraudulent actions in programs [3]. Proffitt et al. (2018) studied the impact of cybercrime on economic crisis management. The purpose of this study was to find out how control of disaster recovery can help us to distinguish the cybercrime which disrupts the business. The results of the study are requirements of planning cyber security, aligning disaster recovery with cyber security, providing cyber security training for managers and employees, and applying the lessons learned from the experience. Implications for positive social change include the ability of organizations to return to acceptable levels of operations and continue to serve their employees, customers and other stakeholders [43].

Solak et al. (2015) studied the analysis of cybercrime perceptions among computer science students. Computer technology is growing rapidly and has become an inevitable part of modern life. While technology simplifies social life, at the same time it brings some security issues. So, it is easier to commit a crime and we are facing cybercrime. This study distinguishes the differences between broad perceptions of undergraduate students at Trakya University in terms of demographic factors. The research method was a questionnaire that was given to teachers and students in the Trakya University sample and was designed to measure and evaluate the level of interest in technology, the severity of cybercrime and people's perceptions of cybercrime in terms of ethics and law. The findings of this study can help us to identify the level of general understanding of cybercrime and the significant differences between groups [44].

Rosellini et al. (2017) investigated the use of data mining to identify US Army soldiers committing violent crimes. The purpose of this study was to use machine learning methods, stepwise regression and random forests, to develop models of predicting violent crime and crime among the US Army soldiers. Results indicated that using this method we can prevent the dangers which might happen to soldiers [45].

Li et al. (2017) researched the development of crime in England by using data mining with selforganized maps (SOM). The aim of this study was to study criminal phenomena in the United Kingdom and its

relationship with different crime factors. Data are collected by the SOM method. Clustering and properties are evaluated using the Scooter algorithm. Machine learning is applied to confirm the clustering result with SOM. As a result, 96.2% accuracy is achieved for crime prediction [46].

Chauhan and Sehgal (2017) studied crime analysis by using data mining techniques. With the rise of computer systems, crime intelligence analysts can help law enforcement officials speed up the process of solving crimes. Using the concept of data mining, we can analyze unknown and useful information of unstructured data. Using analytical and predictive techniques for criminal identification is very effective. Given the increasing crime rate over the years, we have to handle a huge amount of data and it will be extremely difficult to access. Criminals are progressing with the technology. Therefore, it is necessary to use advanced technologies to prevent crime. They focus on examining the algorithms and techniques used to identify criminals [25].

David and Suruliandi (2017) investigated crime analysis and use of data mining techniques at police stations and other similar criminal organizations. Databases contain a great deal of data that can be used to predict or analyze criminal movements and criminal interference in the society. Criminals can be predicted based on crime information. The main purpose of this study was conducting a survey on learning techniques of criminal identification. They investigated the method of crime analysis and its prediction using data mining techniques [47].

Hassani et al. (2016) investigate the use of data mining in crime. The main purpose of this paper is to present data mining applications in crime detection. It covers more than 100 applications of data mining in crime. Data mining techniques including data extraction, clustering, associated rules, decision tree, support vector machines, naive Bayesian, neural networks were applied and the desired results were significant for crime prediction [8].

2.2 Intrusion detection

Meera Gandhi et al. presented an intrusion detection model using decision trees. A 10-fold cross-validation metric was used to evaluate the proposed method. According to this metric, the proposed method detected known intrusion better than unknown intrusion [48]. Lakhina et al. [49] reduced the number of features taken from NSL-KDD dataset using PCA algorithm. In this study, they used principal component analysis and back propagation algorithm. Another research [38] used data mining to extract associated rules for attacks [50]. This framework produced a large number of rules, thereby increased the complexity of the system. Also, Dempster-Shafer and adaptive boosting (AdaBoost) [51] was used for intrusion detection. Meena et al. also presented a review paper on several classification algorithms with KDD CUP 99 and NSL-KDD datasets [52]. Elmasry et al. investigated a multiclass classification for intrusion detection. This is an empirical study, and it uses particle swarm optimization and deep learning to classify various datasets (KDD CUP 99, NSLKDD, CIDDS, and CICIDS2017) [16]. Verma et al. proposed a machine learning method for network intrusion detection. They used CIDDS-001, and the results show that accuracy in this method is 99.60% [53]. A system for HTTP DDoS Attacks detection was investigated in [54] based on information theory and Random Forest.

As it is shown above, most research has focused on cybercrime detection and intrusion detection using supervised methods. Therefore, the aim and main novelty of this paper is to improve the artificial neural network to use semi-supervised classification for intrusion detection.

3 Proposed method

In this section the proposed method is introduced. Initially, the methods used for data pre-processing are introduced, which include data normalization. The standard neural network is presented, and then the proposed method for using unlabelled data in the neural network is introduced.

3.1 Normalization

The goal of normalization is to normalize the elimination of data redundancy and maintain the dependency between related data. This process reduces the size of the database and guarantees improvement of data efficiency. Normalization by standard deviation works well in most cases by measuring the distance between intervals [55]. For sample i, the given value is converted using the following equation: If F is the feature, $\bar{F}$ is F mean, Std is standard deviation, and F' is the normalized value of the feature as follows:

$$\overline{F} = \frac{\sum F_i}{n}; \tag{1}$$

$$Std(F) = \sqrt{\frac{\sum (F_i - \overline{F})^2}{n-1}}; \tag{2}$$

$$F'_i = \frac{F_i - \overline{F}}{Std(F)}. \tag{3}$$

3.2 Artificial Neural Network

Artificial Neural Network (ANN) classification is one of the most effective methods in data classification, but this method has a critical problem which is getting stuck in local optimum [33], [34]. The purpose of network in training process is to minimize the total error of the network based on the weight of the network. We show ANN model as a function (4).

$$y' = M(F'). \tag{4}$$

Here y' is a predicted label, F' is an input feature vector extracted from equation (3), and M is a model of ANN. The back-propagation (BP) training algorithm is used. In layers (except input layer) we use a linear function like equation (5):

$$x = \sum F' \cdot W + b, \tag{5}$$

where W is a weight vector. We used tan as an activation function (equation (6)) to calculate outputs:

$$\text{Activation function} = f(\cdot) = \frac{1 - e^{-x}}{1 + e^{-x}} = y'; \tag{6}$$

$$\text{Error} = y - y'. \tag{7}$$

In this algorithm, after calculating the error in the output layer (equation 7), the values of the weights in the hidden layer are adjusted to reduce the error. Therefore, we need to have a differential of activation function according to the following equation:

$$\frac{d}{dx}\frac{1 - e^{-x}}{1 + e^{-x}} = 1 - \tanh^2 x = (1 - y')(1 + y'). \tag{8}$$

BP Error algorithm, which is an iterative gradient descent algorithm, is a simple way to train multilayer feed forward neural networks. The BP algorithm is based on the gradient descent rule:

$$W(n+1) = W(n) + \eta G(n) + \alpha[W(n) - W(n-1)], \tag{9}$$

where W is the weight vector, n is the iteration number, η is the learning rate, α is the momentum factor, and G is the gradient of error function that is given by 10:

$$G(n) = -\nabla E_p(n). \tag{10}$$

Here E_p is the sum squared error and calculated using equation (8).

This network can be defined as an information processing system consisting of a set of layers and mapping inputs (F') to a suitable set of outputs (y'). The neurons in each layer are fully connected with the neurons in the next layer. In ANN, each neuron has a nonlinear activation function except the input nodes. Updating weights continues to get the given level of error. Finally, ANN uses an appropriate gradient learning algorithm to train.

3.3 Proposed semi-supervised neural network

The main problem in ANN is getting stuck in local minima. Also, because this method is a supervised classification, therefore it is not usable for unlabelled data. We propose a method using pseudo-label (PL) [41] in the neural network. Based on the proposed method, the semi-supervised learning framework is used to train labelled and unlabelled samples. Unlabelled samples are labelled using equation (11).

$$y'_{Unlabelled} = \min d(M(F'), y), \tag{11}$$

where, d is a distance function. In fact, this relationship states that at each stage of training in the neural network the label of unlabelled data is estimated. The unrealistic label y' is calculated with the minimum distance. Since the estimated label values can be accompanied by an error, a coefficient is used to control its impact on the classification process. In this view, a reduction criterion is added for unlabelled samples. This criterion is shown in the equation (12).

$$E = \operatorname{argmin}\big[\text{norm2}_{lable}(y', y) + \alpha(t)\text{norm2}_{Unlabelled}(y'_{Unlabelled}, y)\big]. \tag{12}$$

In this equation, α is the influence coefficient for control the error between labelled and unlabelled samples in the training process. The equation (13) is used here to calculate the value of α.

$$\alpha(t) = \begin{cases} 0 & t \leq \varepsilon_1 \\ \left(\frac{t-T_1}{T_2-T_1}\right)^2 & \varepsilon_1 < t < \varepsilon_2 \\ 1 & \varepsilon_2 \leq t \end{cases} \tag{13}$$

where ε_1 and ε_2 are errors in the training process. Equation (13) represents the current iteration, T_1 is the first iteration with error equal to ε_1, and T_2 is the second iteration with error equal to ε_2 in the training process. We proposed three conditions as follows:

Condition 1: If $t \leq \varepsilon_1$, then in training process, equation (7) is used without unlabelled samples.

Condition 2: If $\varepsilon_1 < t < \varepsilon_2$, then in training process, equation (12) is used with labelled and unlabelled samples. In each epoch, updating $y'_{Unlabelled}$ is performed using equation (11).

Condition 3: If $\varepsilon_2 < t$, then in training process, equation (12) is used with labelled and unlabelled samples.

Pseudocode for the proposed method is illustrated in Table 3. Flow chart is shown in Figure 1.

Algorithm 1 Pseudocode for the proposed back propagation algorithm

Input: η:learning rate α:momentum value and designing multilayer network
Output: A trained neural network Method:

1: Create the initial amount of weights and bias in the network
2: Repeat loop until desired condition {
3: Repeat loop according to number of samples {
4: // feed forward
5: Repeat loop for each j of input layer {
6: $O_j = I_j$ // The output of an input unit is equal to its actual value.
7: Repeat loop for each j of the hidden layer or the output layer {
8: $I_j = \sum_i W_{ij} O_i + \theta_j$ // Calculating the unit network input j compared to i in the previous layer
9: $O_j = \frac{1-e^{-I_j}}{1+e^{-I_j}}$}// Calculate the output of each j
10: // back propagation {
11: If $t \leq \varepsilon_1${
12: Do line 22-31}
13: Elseif $\varepsilon_1 < t < \varepsilon_2${
14: Error = equation 12
15: updating $y'_{Unlabelled}$ using equation 11
16: Do line 22-31}
17: Else {
18: Combine labelled and unlabelled samples
19: Do line 22-31}
20: Repeat loop for each j in the output layer
21: Error = Target −Output // calculating error
22: $\Delta_O^{ij} = error \times (1 - y_{ij}) \times (1 + y_{ij})$ // calculating corrected error
23: Repeat loop for each unit j in the hidden layer from the last to first hidden layer
24: $\Delta_H = (1 + y_{ij}) \cdot (1 - y_{ij}) \cdot (\overline{\Delta_O} \times \overline{W})$
25: Repeat loop for W_{ij} weight and bias in the network {
26: $W_{ij}^{k+1} = W_{ij}^k + \eta G + \alpha[W_{ij}^k - W_{ij}^{k-1}]$}
27: $G = -\nabla E_p$}
28: If end of training = false go to line 10}

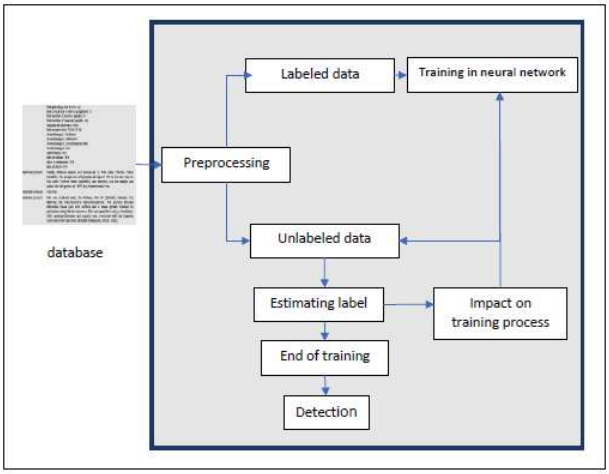


Figure 1. Flow chart for the proposed algorithm

4 Results and discussion

To prepare this paper, a computer with specific characteristics has been used, including:

Processor: Intel Pentium (R) CPU G620, 2.60 GHz 2.60GHz

Installed memory (RAM): 4.00 GB

System type: 32-bit Operating system

The operating system installed is Microsoft's Windows 10. MATLAB version R2017a 64-bit was used for modelling.

4.1 Database

In this paper we used Coburg intrusion detection dataset (CIDDS) to examine the proposed method. CIDDS is available in network-based intrusion detection system dataset. CIDDS has two versions called CIDDS001 and CIDDS-002, and we used CIDDS-001 version. This includes 12 features and 5 labelled classes. Class name and distribution

of dataset is Normal (32 000), DoS (32 000), brute force (32 000), port scan (32 000), and ping scan (32 000). Number of attacks and protocols are 92 and 5, respectively [16].

4.2 Evaluation Metrics

Confusion matrix [16],[56] is one of the criteria for evaluating each classifier. This matrix is a square matrix $N \times N$, where N is the number of classes in the classifier. The main diagonal in this matrix shows the number of correct diagnoses, and the other elements in this matrix show the wrong diagnoses. Table 1 illustrates an example of confusion matrix. Using this matrix, we obtain such metrics as Sensitivity, Specificity, Precision, Recall, F_1, and G-Means.

Table 1. Confusion matrix

True Positive (TP) crimes that are correctly identified as an intrusion	False positive (FP) Correct activity that has been wrongfully detected as an intrusion
False Negative (FN) intrusions that has been wrongfully detected as a correct activity	True Negative (TN) Correct activity that is correctly identified as a correct one

Mathematically speaking, Sensitivity through equation (14), consists of dividing the true positive into the sum of the true positive and false negative.

$$\text{Sensitivity} = \frac{TP}{TP + FN}. \tag{14}$$

Similarly, the Specificity through equation (15), is the result of dividing the true negatives into the sum of the true negatives and false positives.

$$\text{Specificity} = \frac{TN}{TN + FP}. \tag{15}$$

The Precision parameter, through the equation (16), is the result of dividing the true positives into the sum of the true positives and false

positives.

$$\text{Precision} = \frac{TP}{TP + FP}. \tag{16}$$

The Recall parameter through equation (17) is the result of dividing the true positives into the sum of the true positives and the false negatives.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{17}$$

And the F_1-measure and G-mean [56] are also obtained through equations (18) and (19):

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}; \tag{18}$$

$$\text{G-mean} = \sqrt{\text{Precision} \times \text{Recall}}. \tag{19}$$

It should be mentioned that since the dataset used here is multiclass, so each parameter is first separately calculated for each class. Then the average results are obtained.

Another tool for performance measure used here is the Receiver operating characteristic (ROC) [16]. To use the ROC curve in an evaluation, the area under the curve is suggested. This area is the probability that whenever the diagnostic classification variable is randomly measured for a negative response and a positive response, the value obtained is correct. Whenever the test is able to identify accurately, then the values will be low for positive responses and high for negative responses (or vice versa depending on status). The greater the test detection power, the more the ROC curve is above the diagonal, and the closer it is to the ideal (region 1) ROC curve. Inversely, if the ROC curve is under the diagonal or just at the bottom of the square, then the test is with low detection capabilities or useless.

4.3 Comparison of results

Figure 2 shows the mean square error for each epoch of the proposed algorithm in training process. Since one of the problems in neural network is getting stuck in local optimal, so here we use data for validation

as shown in Figure 2 with the green line to decide on the number of iterations of the training process. This figure shows that the lowest difference exists between training and validation data in epoch 20. From the curves, the training error is descending, and these three curves are increased in epoch 3. This is because the unlabelled samples are joined to training process.

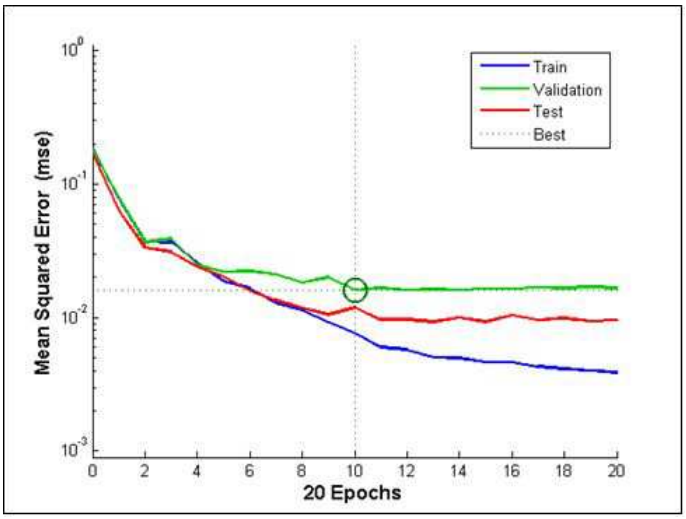


Figure 2. Mean square error in training process

Table 2 illustrates confusion matrix for Test set in the proposed method, called Semi-Supervised Neural Network (SSNN). Number of true predicted labels for Normal class (15976) and Ping scan class (15974) is more than for other classes. Table 3 shows evaluation metrics for SSNN method in each class separately. Precision metric for Normal (0.999) class and Ping scan (0.999) is more than for other classes in SSNN. Precisions for other classes such as DoS, Brute force and Port scan are 0.998, 0.998 and 0.997, respectively. Other evaluation metrics, such as Recall, F_1-measure, Sensitivity, Specificity and G-means are more than 0.99, and this means that SSNN can predict each class with high accuracy.

Tables 2 and 3 showed just the evaluation metrics for SSNN. There-

Table 2. Confusion matrix for each class

	Actual Classes				
	Normal	DoS	Brute force	Port scan	Ping scan
Normal	15976	8	2	10	2
DoS	6	15970	5	12	6
Brute force	5	9	15972	7	5
Port scan	9	10	13	15970	13
Ping scan	4	3	8	1	15974

Table 3. Performance evaluation of SSNN in terms of Precision, Recall, F_1-measure, Sensitivity, Specificity, and G-means for each class

Classes	Normal	DoS	Brute force	Port scan	Ping scan
Metrics					
Precision	0.999	0.998	0.998	0.997	0.999
Recall	0.999	0.998	0.998	0.998	0.998
F_1-measure	0.999	0.998	0.998	0.998	0.999
Sensitivity	0.999	0.998	0.998	0.998	0.999
Sensitivity	1.000	1.000	1.000	0.999	1.000
G-means	0.999	0.998	0.998	0.998	0.999

fore SSNN is compared with 7 other methods: ANN, SVM [53], NB [53], DT [53], VR [53], IAB [54], DBN [16]. ANN is a standard neural network explained in Section 3.1. SVM is support vector machine with Radial Basis Function (RBF) as a kernel function. NB and DT represent Naive Bayes and Decision Tree (J48), respectively. VR was investigated by Verma and Ranga using machine learning techniques to statistical analysis of dataset [53]. IAB was proposed by Idhammad et al. for attack detection in cloud environment [54]. DBN is deep belief networks and was explained by Elmasry et al. [16].

Accuracy is shown in Table 4. This metric for SSNN is equal to 99.83, and its value is greater than for ANN and other methods. Us-

ing of unlabelled samples in training process causes this improvement. After SSNN, VR and IAB have the greatest accuracy than in other methods, respectively. Another characteristic used here is AUC. The AUC for the SSNN (0.9987) shows better results.

Table 4. Accuracy and AUC

	SSNN	ANN	SVM	NB	DT	VR	IAB	DBN
Accuracy	99.83	95.71	98.19	98.70	98.90	99.60	99.54	94.66
AUC	0.9987	0.9571	0.9823	0.9871	0.9891	0.9961	0.9955	0.9625

Figure 3 shows the average of Precisions for all classes. As it is shown above, the SSNN method performs better than others. Precision is the ratio of classified samples by the classifier in a given class, to the total number of samples the classifier has classified in that class, either correctly or incorrectly. As it turns out from equation (16), the Precision shows in what proportion the detected positives are really positive. Precision in SSNN is equal to 99.83% and in ANN = 96.25%. DBN, VR and IAB have Precision 99.71%, 99.61% and 99.53%, respectively.

Figure 4 shows the value of average of Recall for classes. This parameter for SSNN is 99.83%, and in other methods, e.g., VR = 99.59% and IAB = 99.55%. So, it indicates that the proposed algorithm performs better than other methods. The Recall shows the ratio of true classification of samples in given classes by the classifier to the number of samples in that class. So, the Recall shows in what proportion true positives are correctly identified as positive. Therefore, Figure 4 shows that the SSNN predicts intrusion detection better.

F_1-Measure is proposed to compare Precision and Recall, in fact, F_1-Measure shows the harmonic mean between Precision and Recall (Figure 5). F_1-metric for SSNN is 99.83, and it is greater than for other methods like VR (99.60%). Increasing the number of training examples in the proposed method with the semi-supervised approach in it has improved the global search in NNSS.

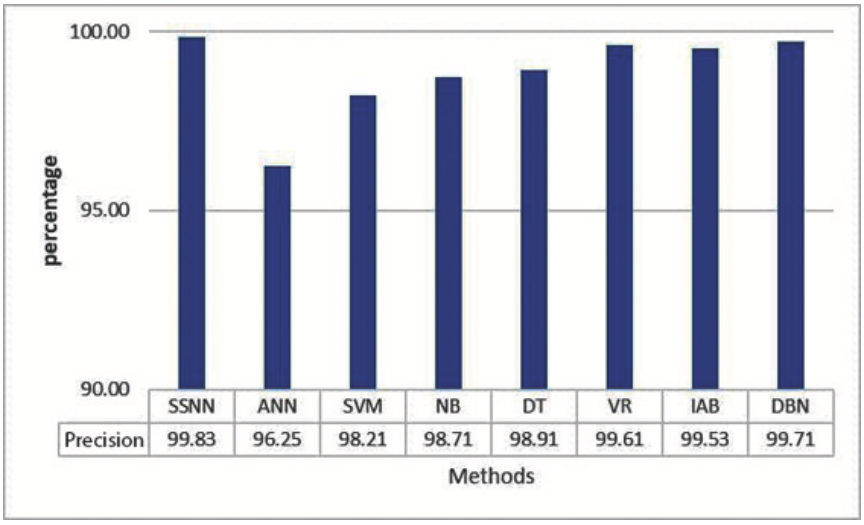


Figure 3. Comparison of Precision

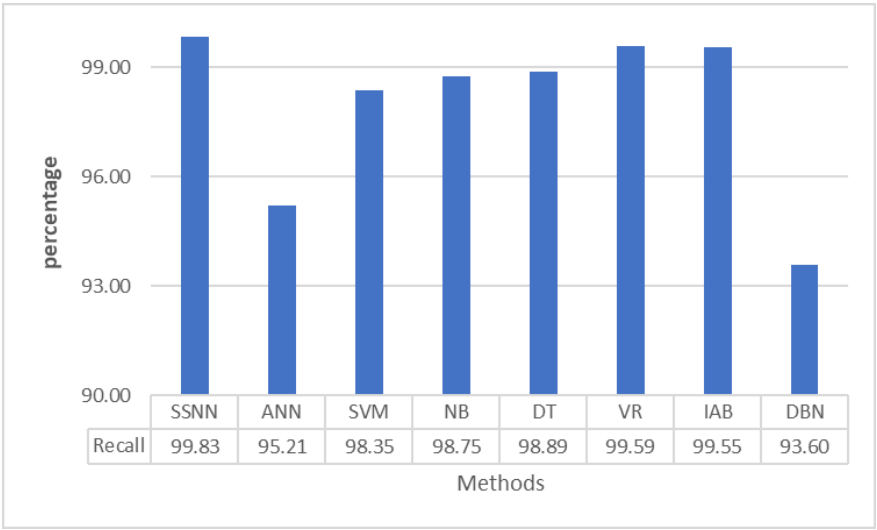


Figure 4. Comparison of Recall

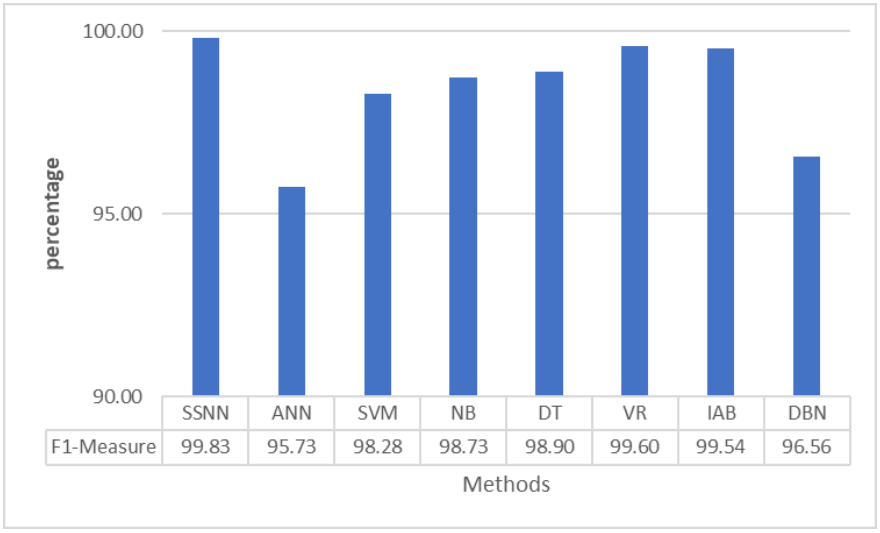


Figure 5. Comparison of F_1-measure

Figure 6 shows the values of the average Sensitivity parameter for all classes. The Sensitivity of SSNN is greater than 99.83%, which is better than for ANN (95.21%). The ability of a classifier to find the true positive of a class is called Sensitivity. According to equation (14), Sensitivity is ratio of true positives to the sum of true positives and false negative. So, the Sensitivity here shows what proportions of the intrusions are correctly identified. Therefore, this parameter also confirms that the SSNN predicts intrusions better than other methods.

Figure 7 shows the average Specificity value for all classes. The Specificity value for SSNN (99.96%) is greater than for the other methods. It indicates that the proposed method has performed well in Normal class. Figure 8 indicates average G-mean for all classes. G-mean in SSNN is 99.83% and proves that SSNN with higher accuracy predicts normal situation and attacks situation rather than ANN (97.21%), VR (99.60%), IAB (99.54%), DBN (98.90%), etc.

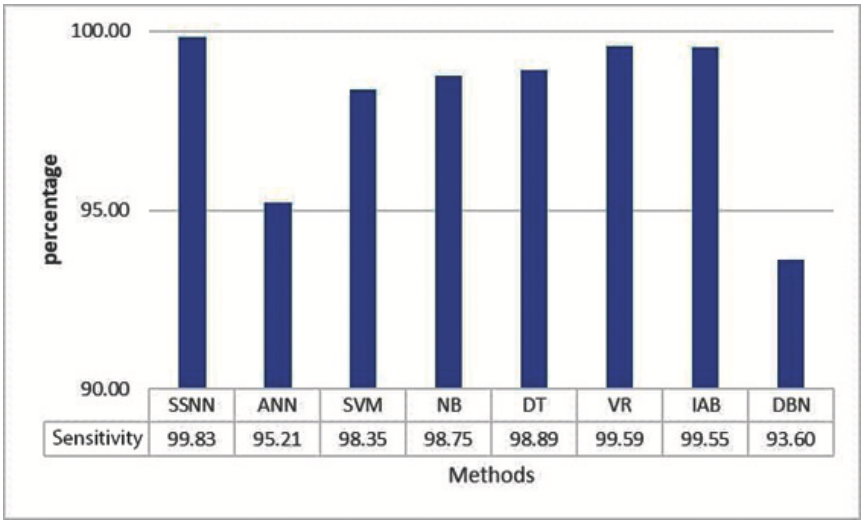


Figure 6. Comparison of Sensitivity

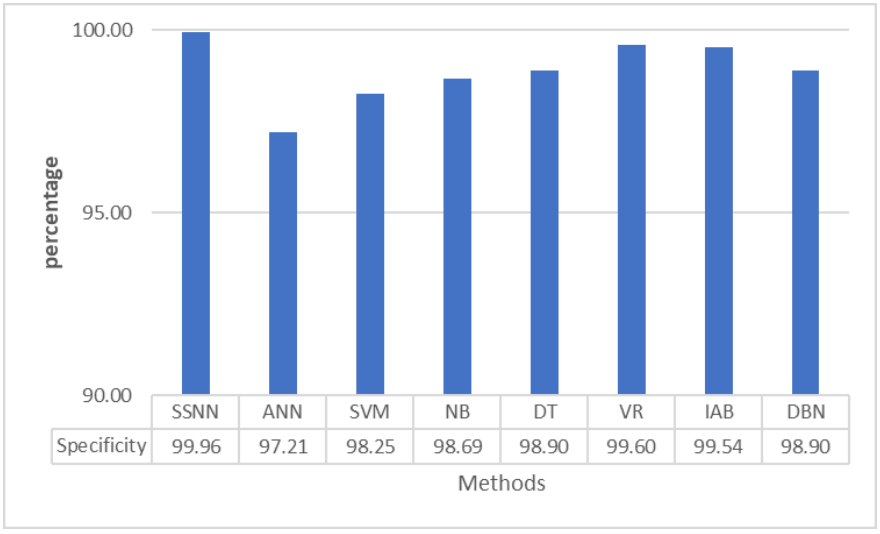


Figure 7. Comparison of Specificity

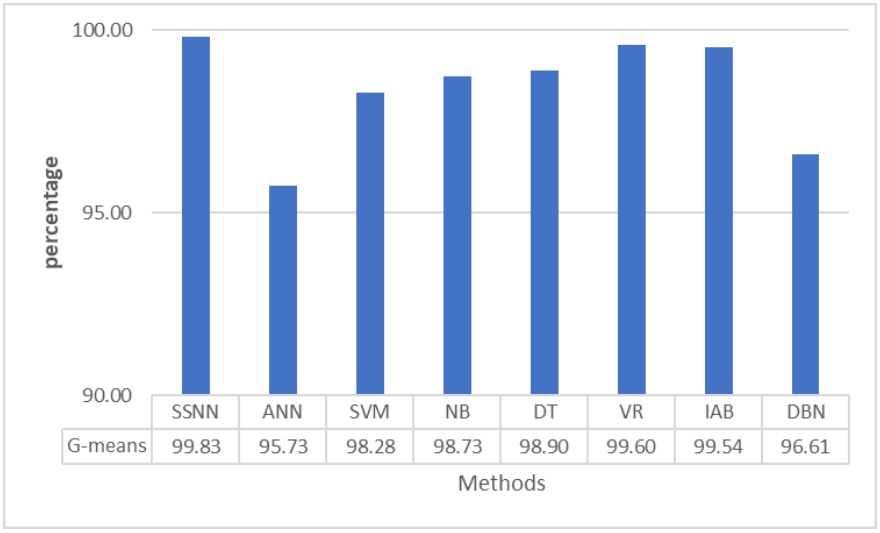


Figure 8. Comparison of G-means

5 Conclusion

Data mining techniques, including descriptive and predictive tools, have been introduced in various fields and a tremendous number of researches have been conducted on this issue. Data mining applications include business, management, medical, sports, econometrics, financial management, web business. One of the areas that has been the focus of data mining in recent years is the police enforcement, and one of the topics that has attracted a great deal of research is crime modelling. Therefore, this paper presents a semi-supervised method for detecting cybercrimes. The neural network is used here to classify the crimes. Since the neural network is a supervised classification technique, it is only usable for labelled data. On the other hand, it is a cost of fortune and time consuming to create labelled data. Thus, here the neural network is optimized so that it can be used in unlabelled data. Here the pseudo-labelling technique is used to estimate the labelled data during the neural network training process. The Precision, Recall, Sensitivity and Specificity values for the proposed network are obtained and

represent values of 99.83%, 99.83%, 99.83% and 99.96%, respectively. However, other researches have reported lower values.

We suggest for the future work, first, the detection of the probability of a crime before it happens, and this issue has not been addressed here. Second, a specific dataset is used here for crime detection, while many crimes occur today on social networks and data in the networks are a combination of text and images, so a hybrid method for crime detection is suggested.

Conflict of interest

We wish to inform that there is no known conflict of interest associated with this paper.

Acknowledgements

The authors are grateful for the support from Islamic Azad University and Atiye Andishan Group.

References

[1] A.-Z. Ala'M, J. f. Alqatawna, and H. Faris, "Spam profile detection in social networks based on public features," in *2017 8th International Conference on information and Communication Systems (ICICS)*, IEEE,2017, pp. 130–135.

[2] S. Aljawarneh, M. Aldwairi, and M. B. Yassein, "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model," *Computational Science*, vol. 25, pp. 152–160, 2018.

[3] S. Dutta, A. K. Gupta, and N. Narayan, "Identity Crime Detection Using Data Mining," in *2017 3rd International Conference on Computational Intelligence and Networks (CINE)*, IEEE, 2017, pp. 1–5.

[4] L. Y. Chang, L. Y. Zhong, and P. N. Grabosky, "Citizen co-production of cyber security: Self-help, vigilantes, and cyber-

crime," *Regulation & Governance*, vol. 12, no. 1, pp. 101–114, 2018.

[5] N. L. Piquero, "Causes and prevention of intellectual property crime," in *Combating Piracy*, Routledge, 2017, pp. 57–84.

[6] S. Qayyum and H. Dar, "A Survey of Data Mining Techniques for Crime Detection," *University of Sindh Journal of Information and Communication Technology*, vol. 2, no. 1, pp. 1–6, 2018.

[7] H. Benjamin Fredrick David and A. Suruliandi, "Survey on Crime Analysis and Prediction Using Data Mining Techniques," *ICTACT Journal on Soft Computing*, vol. 7, no. 3, pp. 1459–1466, 2017.

[8] H. Hassani, X. Huang, E. S. Silva, and M. Ghodsi, "A review of data mining applications in crime," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 9, no. 3, pp. 139–154, 2016.

[9] T. Tantawutho, "The computer crime incidents and the future of countermeasures in Thailand," in *2017 Third Asian Conference on Defence Technology (ACDT)*, IEEE, 2017, pp. 97–103.

[10] H. Fatima, G. Dash, and S. K. Pradhan, "Soft Computing applications in Cyber crimes," in *2017 2nd International Conference on Anti-Cyber Crimes (ICACC)*, IEEE,2017, pp. 66–69.

[11] J. D. Hawkins and J. G. Weis, "The social development model: An integrated approach to delinquency prevention," in *Developmental and Life-course Criminological Theories*, Tara Renae McGee and Paul Mazerolle, Eds. London: Routledge, 2017, pp. 3–27.

[12] R. F. Sparks, "Criminal opportunities and crime rates," *Indicators of crime and criminal justice: Quantitative studies*, vol. 1, pp. 18–28, 1980.

[13] K. Land, *Criminal circumstance. A dynamic multi-contextual criminal opportunity theory* (New Lines in Criminology Series), Taylor and Francis, 2018, 264 p. ISBN: 9781351524940.

[14] J. Li et al., "Signal-noise identification of magnetotelluric signals using fractal-entropy and clustering algorithm for targeted denoising," *Fractals*, vol. 26, no. 2, p. 1840011 (18 pages), 2018. DOI: https://doi.org/10.1142/S0218348X1840011X.

[15] E. E.-D. Hemdan and D. Manjaiah, "Cybercrimes Investigation and Intrusion Detection in Internet of Things Based on Data Science Methods," in *Cognitive Computing for Big Data Systems Over IoT* (Lecture Notes on Data Engineering and Communications Technologies), Arun Kumar Sangaiah, Arunkumar Thangavelu, and Venkatesan Meenakshi Sundaram, Eds. Springer International Publishing, 2018, pp. 39–62.

[16] W. Elmasry, A. Akbulut, and A. H. Zaim, "Empirical study on multiclass classification-based network intrusion detection," *Computational Intelligence*, vol. 35, no. 4, pp. 919–954, 2019.

[17] P. Mali, J. Sodhi, T. Singh, and S. Bansal, "Analysing the Awareness of Cyber Crime and Designing a Relevant Framework with Respect to Cyber Warfare: An Empirical Study," *International Journal of Mechanical Engineering and Technology*, vol. 9, no. 2, pp. 110–124, 2018, Article ID: IJMET_09_02_012.

[18] J. Zhao et al., "An "End-Network-Cloud" Architecture Key Technology with Threat Perception and Collaborative Analysis," in *IOP Conference Series: Materials Science and Engineering*, 2018, vol. 452, no. 3, IOP Publishing, p. 032091 (6 pages). DOI: 10.1088/1757-899x/452/3/032091.

[19] R. Gifty, R. Bharathi, and P. Krishnakumar, "Privacy and security of big data in cyber physical systems using Weibull distribution-based intrusion detection," *Neural Computing and Applications*, vol. 31, no. 1, pp. 23–34, 2019.

[20] D. A. Ziegler, J. R. Janowich, and A. Gazzaley, "Differential impact of interference on internally-and externally-directed attention," *Scientific reports*, vol. 8, no. 1, Article number 2498 (10 pages), 2018. DOI: 10.1038/s41598-018-20498-8.

[21] A. F. Sidiq, R. Umar, and A. Yudhana, "Research on Secure Virus Trojan in Cybersecurity Platform," *JSI (Jurnal sistem Informasi) Universitas Suryadarma*, vol. 5, no. 2, pp. 8–13, 2018. DOI: https://doi.org/10.35968/jsi.v5i2.247.

[22] P. A. A. Resende and A. C. Drummond, "A survey of random forest based methods for intrusion detection systems," *ACM Com-*

puting Surveys (CSUR), vol. 51, no. 3, Article No. 48 (36 pages), 2018. DOI: 10.1145/3178582.

[23] A. Tomlinson, J. Bryans, and S. A. Shaikh, "Towards viable intrusion detection methods for the automotive controller area network," in *2nd Computer Science in Cars Symposium - Future Challenges in Artificial Intelligence Security for Autonomous Vehicles (CSCS 2018)*, (Munich, Germany, 13-14 September 2018), 2018, United States: ACM. DOI: 10.1145/3273946.3273950.

[24] W. Wang, J. Liu, G. Pitsilis, and X. Zhang, "Abstracting massive data for lightweight intrusion detection in computer networks," *Information Sciences*, vol. 433, pp. 417–430, 2018.

[25] C. Chauhan and S. Sehgal, "A review: Crime analysis using data mining techniques and algorithms," in *2017 International Conference on Computing, Communication and Automation (ICCCA)*, (5-6 May 2017), 2017, pp. 21–25. DOI: 10.1109/CCAA.2017.8229823.

[26] O. E. Isafiade and A. B. Bagula, *Data mining trends and applications in criminal science and investigations*, IGI Global, 2016, 386 p. DOI: 10.4018/978-1-5225-0463-4, ISBN13: 9781522504634, ISBN10: 152250463X.

[27] J. Lin et al., "Automatic Knowledge Discovery in Lecturing Videos via Deep Representation," *IEEE Access*, vol. 7, pp. 33957–33963, 2019.

[28] S. Caneppele and M. F. Aebi, "Crime drop or police recording flop? On the relationship between the decrease of offline crime and the increase of online and hybrid crimes," *Policing: A Journal of Policy and Practice*, vol. 13, no. 1, pp. 66–79, 2017.

[29] A. V. Mbaziira and D. R. Murphy, "An empirical study on detecting deception and cybercrime using artificial neural networks," in *Proceedings of the 2nd International Conference on Compute and Data Analysis*, ACM,2018, pp. 42–46.

[30] K. Gajera, M. Jangid, P. Mehta, and J. Mittal, "A Novel Approach to Detect Phishing Attack Using Artificial Neural Networks Combined with Pharming Detection," in *2019 3rd International con-*

ference on Electronics, Communication and Aerospace Technology (ICECA), IEEE, 2019, pp. 196–200.

[31] D. Peraković, M. Periša, I. Cvitić, and S. Husnjak, "Artificial neuron network implementation in detection and classification of DDoS traffic," in *2016 24th Telecommunications Forum (TELFOR)*, (Belgrade, Serbia, 22-23 Nov. 2016), IEEE, 2016, pp. 1–4. DOI: 10.1109/TELFOR.2016.7818791.

[32] L. O. Batista et al., "Fuzzy neural networks to create an expert system for detecting attacks by sql injection," arXiv:1901.02868 [cs.AI], 2019.

[33] D. Hassabis, D. Kumaran, C. Summerfield, and M. Botvinick, "Neuroscience-inspired artificial intelligence," *Neuron*, vol. 95, no. 2, pp. 245–258, 2017.

[34] M. Van Gerven and S. Bohte, "Editorial: Artificial neural networks as models of neural information processing," *Frontiers in Computational Neuroscience*, vol. 11, Article number: 114 (2 pages), 2017. DOI: 10.3389/fncom.2017.00114.

[35] I. N. Da Silva, D. H. Spatti, R. A. Flauzino, L. H. B. Liboni, and S. F. dos Reis Alves, *Artificial neural networks. A Practical Course*, Switzerland: Springer International Publishing, 2017, XX+307 p. ISBN: 978-3-319-43161-1.

[36] S. Gong, W. Gao, and F. Abza, "Brain tumor diagnosis based on artificial neural network and a chaos whale optimization algorithm," *Computational Intelligence*, vol. 36, no. 1, pp. pp. 259–275, 2020. First Published: 20 November 2019, DOI: 10.1111/coin.12259.

[37] J. Mendoza and H. Pedrini, "Detection and classification of lung nodules in chest X-ray images using deep convolutional neural networks," *Computational Intelligence*, vol. 36, no. 2, pp. 370-401, 2020. First Published: 04 November 2019, doi: 10.1111/coin.12241.

[38] P. M. Buscema, G. Massini, M. Fabrizi, M. Breda, and F. Della Torre, "The ANNS approach to DEM reconstruction," *Computa-*

tional Intelligence, vol. 34, no. 1, pp. 310–344, 2018. First published: 28 November 2017, DOI: 10.1111/coin.12151.

[39] M.-J. Kang and J.-W. Kang, "Intrusion detection system using deep neural network for in-vehicle network security," *PloS one*, vol. 11, no. 6, ID: e0155781, 2016.

[40] E. Hodo et al., "Threat analysis of IoT networks using artificial neural network intrusion detection system," in *2016 International Symposium on Networks, Computers and Communications (ISNCC)*, (Yasmine Hammamet, Tunisia, 11-13 May 2016), IEEE, 2016, pp. 1–6. DOI: 10.1109/ISNCC.2016.7746067.

[41] A. Rasmus, M. Berglund, M. Honkala, H. Valpola, and T. Raiko, "Semi-supervised learning with ladder networks," in *Advances in neural information processing systems*, 2015, pp. 3546–3554.

[42] M. Feng, J. Zheng, Y. Han, J. Ren, and Q. Liu, "Big Data Analytics and Mining for Crime Data Analysis, Visualization and Prediction," in *International Conference on Brain Inspired Cognitive Systems*, Springer, 2018, pp. 605–614.

[43] T. G. Proffitt, "The Effects of Computer Crimes on the Management of Disaster Recovery," Ph.D. dissertation, Scholar Works Walden Dissertations and Doctoral Studies, Walden University, 2018. https://scholarworks.waldenu.edu/dissertations/5252.

[44] D. Solak and M. Topaloglu, "The Perception Analysis of Cyber Crimes in View of Computer Science Students," *Procedia - Social and Behavioral Sciences*, vol. 182, pp. 590–595, 2015. DOI: https://doi.org/10.1016/j.sbspro.2015.04.787.

[45] A. J. Rosellini et al., "Using administrative data to identify US Army soldiers at high-risk of perpetrating minor violent crimes," *Journal of psychiatric research*, vol. 84, pp. 128–136, 2017.

[46] X. Li, H. Joutsijoki, J. Laurikkala, and M. Juhola, "Development of crime in England and Wales 1898–2001: Data mining using self-organising map," in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, IEEE, 2017, pp. 1–8.

[47] H. B. F. David and A. Suruliandi, "Survey on Crime Analysis and Prediction Using Data Mining Techniques," *ICTACT Journal on Soft Computing*, vol. 7, no. 3, pp. 1459–1466, 2017.

[48] G. MeeraGandhi, K. Appavoo, and S. Srivasta, "Effective network intrusion detection using classifiers decision trees and decision rules," *Int. J. Advanced network and application*, vol. 2, no. 3, pp. 686–692, 2010.

[49] S. Lakhina, S. Joseph, and B. Verma, "Feature reduction using principal component analysis for effective anomaly–based intrusion detection on NSL-KDD," *International Journal of Engineering Science and Technology*, vol. 2, no. 6, pp. 1790–1799, 2010.

[50] J. Song, H. Xie, and Y. Feng, "Fast association rule mining algorithm for network attack data," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 20, no. 6–7, pp. 1465–1469, 2017. DOI: 10.1080/09720529.2017.1392464.

[51] P. Natesan and P. Balasubramanie, "Multi Stage Filter Using Enhanced Adaboost for Network Intrusion Detection," *International Journal of Network Security & Its Applications (IJNSA)*, vol. 4, no. 3, pp. 121–135, 2012.

[52] G. Meena and R. R. Choudhary, "A review paper on IDS classification using KDD 99 and NSL KDD dataset in WEKA," in *2017 International Conference on Computer, Communications and Electronics (Comptelix)*, IEEE, 2017, pp. 553–558.

[53] A. Verma and V. Ranga, "On evaluation of network intrusion detection systems: Statistical analysis of CIDDS-001 dataset using machine learning techniques," *Pertanika Journal of Science & Technology*, vol. 26, no. 3, pp. 1307–1332, 2018.

[54] M. Idhammad, K. Afdel, and M. Belouch, "Detection system of HTTP DDoS attacks in a cloud environment based on information theoretic entropy and random forest," *Security and Communication Networks*, vol. 2018, Article ID 1263123, 2018. DOI: https://doi.org/10.1155/2018/1263123.

[55] H. Schwarzenbach, A. M. Da Silva, G. Calin, and K. Pantel, "Data normalization strategies for microRNA quantification," *Clinical chemistry*, vol. 61, no. 11, pp. 1333–1342, 2015.

[56] R. Abdulhammed, M. Faezipour, A. Abuzneid, and A. AbuMallouh, "Deep and machine learning approaches for anomaly-based intrusion detection of imbalanced network traffic," *IEEE sensors letters*, vol. 3, no. 1, pp. 1–4, 2018.

Abbas Karimi, Saber Abbasabadi,
Javad Akbari Torkestani, Frane Zarafshan

Received April 04, 2020
Revised April 16, 2021
Accepted April 20, 2021

Abbas Karimi
Assistant Professor, Department of Computer Engineering,
Islamic Azad University, Arak Branch,
Arak, Markazi Provience, Iran.
E–mail: Akarimi@iau-arak.ac.ir

Saber Abbasabadei
PhD student, Department of Computer Engineering,
Islamic Azad University, Arak Branch,
Arak, Markazi Provience, Iran.
E–mail: saber.abbasabadey1@gmail.com

Javad Akbari Torkestani
Associate Professor, Department of Computer Engineering,
Islamic Azad University, Arak Branch,
Arak, Markazi Provience, Iran.
E–mail: j-akbari@iau-arak.ac.ir

Faraneh Zarafshan
Assistant Professor, Department of Computer Engineering,
Islamic Azad University, Arak Branch,
Arak, Markazi Provience, Iran.
E–mail: fzarafshan@aiau.ac.ir

Computer Science Journal of Moldova, vol.**29**, no.2(86), 2021

A Hybrid Method based on SA and VNS Algorithms for Solving DAP in DDS

Nasser Lotfi, Jamshid Tamouk

Abstract

Data allocation problem (DAP) is of great importance in distributed database systems (DDS). Minimizing the total cost of transactions and queries is the main objective of DAP which is mostly affected by the volume of transmitting data through the system. On the other hand, the volume of transmitting data depends on the fragment-to-site allocations method. DAP as a Np-hard problem has been widely solved by applying soft computing methods like evolutionary algorithms. In the continuation of our previously published research, this paper proposes a novel hybrid method based on Simulated Annealing Algorithm (SA) and Variable Neighborhood Search (VNS) mechanism for Solving DAP. To increase the performance, VNS mechanism is embedded into SA method in the proposed hybrid method. Technically speaking, in order to discover more promising parts of search space, the proposed method (VNSA) explores the search space via SA and fulfills more exploitation by applying neighborhood search mechanism. Moreover, due to the fact that both are a single solution-based method, they explore the search space faster than population-based methods. Performance of the proposed VNSA is experimentally evaluated using well-known benchmarks reported in state-of-the-art literature, and evaluation outcomes prove the robustness and fastness of the proposed hybrid method (VNSA). Furthermore, the results exhibit that VNSA outperforms its competitors and achieves better results in majority of test problems.

Keywords: Data Allocation Problem, Simulated Annealing, Variable Neighborhood Search.

1 Introduction

In DDS, minimizing the total cost of transactions and queries is of great importance which is mostly affected by amount of transmitting data through the system. Likewise, amount of transmitting data depends on fragment-to-site allocations method [1]. This NP-hard optimization problem in distributed systems is known as Data Allocation Problem (DAP). Data allocating methods in real world like sites, mail servers and search engines are significantly important because they deal with huge amount of data [2]. The detailed descriptions of DAP are given in Section 2. Due to it being an optimization problem, DAP has been widely solved by applying soft computing methods like evolutionary algorithms. The reason is that evolutionary algorithms are able to extract feasible and high quality solutions in acceptable computational time. Apart from a pretty wide literature can be found for solving DAP, two types of algorithms, namely static and dynamic algorithms, have been applied [3, 4, 32]. Taking the recent literature into account, researchers have been widely proposed hybrid methods for solving DAP [5]. The following part reviews some remarkable state-of-the-art methods for solving DAP.

Distributed database systems, fragmentation and allocation concepts were reviewed by Bhuyar et al. in [34]. Authors mentioned that fragmentation and allocation are two remarkable NP-hard challenges in DDS which can effectively affect performance of the system [34].

Wand et al. in [12] mapped DAP to well-known knapsack problem and solved it using Artificial Immune system method [13]. They also fulfilled comparison of the obtained results to some of state-of-the-art methods in order to measure the effectiveness of the applied method.

Biogeography-Based Optimization (BBO) method was applied by Singh et al. in [14] for minimizing total transmission and storage costs in database systems. Authors solved this fragment allocation problem and illustrated that their method is more effective than genetic algorithm [14].

Sen et al. [15] applied SA method for solving DAP and evaluated the method using standard benchmarks in CPLEX. Evaluation out-

comes proved that SA is faster and more effective.

Genetic Algorithm (GA), Ant Colony Optimization (ACO) [18] and Tabu-search method [19] alongside a new crossover method were applied by Tosun in [17] for solving DAP. Also, effectiveness of the proposed method was evaluated using QAPLIB benchmarks.

Abdalla [30] proposed an innovative data allocation method for allocating fragments to sites based on communication costs. Authors also aimed to enhance transmission cost to improve distribution performance. They improved data fragmentation and allocation methods; they also carried out site clustering to produce minimum possible number of clusters. Performance of the proposed method was evaluated using TC objective function and results proved the efficiency of suggested method.

A hybrid method combining ACO and local search was proposed by Adl et al. [11] for solving DAP. Evaluation results indicated that the proposed method is flexible and successful in extracting near-optimal solutions. Ahmed et al. [21] suggested an evolutionary method for solving DAP. They also introduced a dependency graph for modeling fragments dependencies. To measure the efficiency, the suggested method was evaluated using standard benchmarks.

To solve DAP, Mahi et al. [2] applied Particle Swarm Optimization (PSO) for minimizing total transmission cost. Authors evaluated their results over 20 benchmarks; and outcomes proved that the suggested method, PSO-DAP, outperforms state-of-the-art methods in terms of quality and time. A clustering-based fragmentation method was developed by Sewisty et al. [25] which emphasizes on generating clusters. In the proposed method, disjoint fragments are generated using clusters. Evaluation results indicated that the introduced method is valid and effective. In order to minimize total transmission cost, Apers et al. [33] proposed a fragment allocation mechanism. Authors described details of DAP complexity and compared the obtained results to existent methods.

Lotfi proposed a new hybrid method (DEVNS) in [35] based on differential evolution (DE) and VNS algorithms for solving data allocation problem. The author enhanced DE algorithm by proposing better se-

lection and crossover operators as well as embedding VNS into DE to increase its performance. The effectiveness of DEVNS was evaluated over well-known benchmarks and compared to nine existent methods reported in literature. The obtained results indicated that DEVNS outperforms all competitors in 13 out of 20 benchmarks. Even though DEVNS outperforms its competitors, it has some drawbacks. The main drawback and weakness of DEVNS is that it is slow and time consuming. DE is a population-based evolutionary algorithm, and working on a population of solutions is time consuming. Moreover, when it is combined with another algorithm like VNS, it becomes even slower. Hence, there is no comparison in terms of execution time in [35]. Another drawback is that according to the reported results, DEVNS does not work well on small size problems.

This paper proposes an innovative hybrid method, named as VNSA from this point on, for solving data allocation problem (DAP). The suggested method consists of SA algorithm [6, 7, 8] and VNS [9, 10].

All state-of-the art methods in [1] are population-based algorithms. Hence the aim of this study is to use combination of single-solution-based methods in order to achieve better results. On the other hand, since the single-solution-based methods are faster and time-efficient, the proposed hybrid method obtains better results in acceptable time. According to the proposed hybrid strategy, SA algorithm is mixed up with efficient neighborhood search mechanism to improve the exploration and exploitation performance. In the modified SA, instead of choosing the neighbors randomly, an efficient VNS is applied to fulfill more exploitation over the solution. Description of VNS mechanism and details of how to apply these operators are given in the following sections.

Performance of the proposed hybrid method (VNSA) is experimentally evaluated using the well-known benchmarks reported in state-of-the-art literature. Evaluation outcomes prove the robustness of the VNSA and exhibits that system outperforms its competitors and achieves better results. Moreover, the fact that they are single-solution-based methods, both SA and VNS explore the search space faster than the population-based methods. Hence, the proposed hybrid method

has been faster compared to state-of-the-art methods.

The rest of the paper is formed as following: The detailed description of Data Allocation problem and the proposed hybrid method are given in Section 2. Section 3 demonstrates the algorithm parameters, experimental and comparison results. Finally, the conclusions and some future research works are given in Section 4.

2 The proposed hybrid method for solving DAP in DDS

A detailed description of Data Allocation Problem is presented in this section. Likewise, it describes the proposed innovative hybrid method comprising modified SA algorithm [6, 7, 8] and VNS technique [9, 10] for solving DAP [33, 34]. Motivation and novelty of this study is to combine SA and VNS methods to increase the capability of discovering more promising parts of search space in acceptable time. These two methods are not able to achieve better results than state-of-the-art methods individually, but hybrid version of these algorithms is robust and works efficiently.

Every site in DDS deals with a part of database [3, 11]. During the running time, many transactions with different frequencies are given to the sites. For this reason, a huge amount of data is transmitted between the sites. Minimizing the total completion time of transactions is the main goal of DAP which is affected by data transmission speed [11]. Parameter notations in DAP are described in Table 1 [11].

As it was mentioned above, minimization of total cost is the main goal in DAP. During the process, storage capacity and data transmission cost are considered as problem limitation and problem cost respectively [31]. Figure 1 demonstrates the transaction-fragment and site-transaction dependencies. Total cost is mostly affected by data transmitting through the system. Variable X_{ij} is defined as below [11].

Table 1 Notation Description [11]

Notation	Description
n	Number of Sites
m	Number of fragments
s_i	Site number i
$SiteCap_i$	Capacity of s_i
$UC_{n\times n}$	Cost of data transmission between all sites
uc_{i1i2}	Data transmission cost transmitted from site s_i to site s_j
f_j	Fragment Number j
$fragSize_j$	Size of f_j
l	Number of transactions
t_k	Transaction Number k
$FREQ_{n\times l}$	Transactions Frequency
$freq_{ik}$	Frequency of t_k on s_i
$TRFR_{l\times m}$	Direct transaction-Fragment dependency
$trft_{kj}$	Volume of data transmission from site of f_j to the site which executes transaction t_k.
$Q_{l\times m\times m}$	Indirect transaction-fragment dependencies
q_{kj1j2}	Amount of data transmission from site of f_{j1} to the site of f_{j2}.
Ψ	Allocation scheme
Ψ_j	Site of f_j assigned in scheme Ψ
$COST(\Psi)$	Data transmission cost in Ψ
$COST1(\Psi)$	Data transmission cost obtained from $TRFR_{l\times m}$
$COST2(\Psi)$	Data transmission cost obtained from $Q_{l\times m\times m}$
$STFR_{n\times m}$	Site-fragment dependencies
$stfr_{ij}$	Amount of data from f_j accessed by s_t in time unit
$PARTIALCOST1_{n\times m}$	Fragment-site allocation $COST1(\Psi)$
$partialcost1_{ij}$	Cost of f_j allocated to s_i obtained from direct dependencies
$QFR_{l\times m\times m}$	Indirect transaction-fragment dependencies considering frequencies
qfr_{kj1j2}	Amount of data transmission from site of f_{j1} to site of f_{j2} considering frequency of t_k
$FRDEP_{m\times m}$	Inter fragment dependencies
$frdep_{j1j2}$	Amount of data transmission from site of f_{j1} to site of f_{j2} considering indirect dependencies

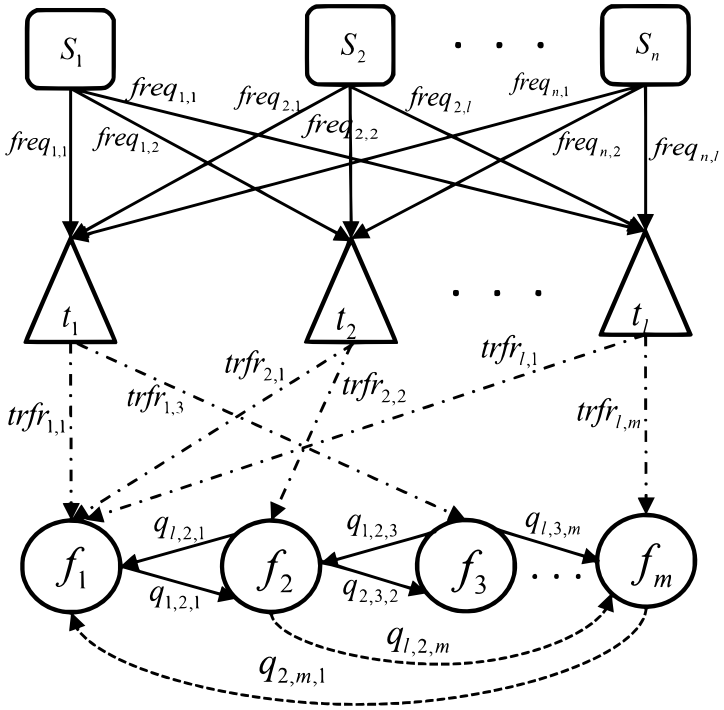


Figure 1. Transaction-fragment and site-transaction dependencies

$$X_{ij} = \begin{cases} 1, & \text{if } \Psi_j = s_i \ ; \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where ψ_j is the site, where f_i is assigned. Hence, the storage capacity limitation is declared as Equation (2) [11]:

$$\sum_{j=1}^{m} fragSize_j \times X_{ij} \leq siteCap, i = 1, ..., n. \tag{2}$$

Therefore, total data transmissions cost is computed as below [11]:

$$COST(\psi) = \sum_{j=1}^{m} partialcost1_{\psi,j} = \sum_{j1=1}^{m} \sum_{j2=1}^{m} frdep_{j1,j2} \times uc_{\psi_{j1}\psi_{j2}}, \quad (3)$$

where $partialcost_{\psi 1}$ is the cost of storing f_i on site $s_{\psi j}$ which is calculated as Equation (4) [11]:

$$partialcost1_{\psi_j j} = \sum_{q=1}^{n} uc_{\psi,q} \times stfr_{qj}, \quad (4)$$

where $stfr_{qj}$ is amount of data from f_j accessed by s_q which is computed by Equation (5) [11]:

$$stfr_{qj} = \sum_{k=1}^{l} freq_{qk} \times trfr_{kj}. \quad (5)$$

Also, in the second part of the COST(ψ), the value of $frdep_{j1j2}$ is amount of data transferred from site f_{j1} to the site f_{j2} by taking the indirect dependencies into account. The direct transaction-fragment dependency (TRFR) is a matrix in which for each execution of t_k, the value of $trfr_{kj}$ represents amount of data transferred from site holding f_j to the site holding t_k. The dependency is called as direct dependency if for each execution of t_k there is some data transferred from site f_j to the site t_k. Value is calculated as Equation (6) [11]:

$$frdep_{j1j2} = \sum_{k=1}^{l} qfr_{kj1j2}, \quad (6)$$

where $qfr_{kj_1j_2}$ is volume of data transmitted from site holding f_{j1} to the site holding f_{j2}, which is calculated as Equation (7) [11]:

$$qfr_{kj1j2} = q_{kj1j2} \times \sum_{r=1}^{n} freq_{kr}. \quad (7)$$

The proposed hybrid method uses a modified SA to generate the neighbors effectively. In the modified SA, a new solution is generated

by changing some sites randomly. Thereafter, more modification is carried out on new solution using VNS mechanism to discover more promising solutions around and replace them. Figure 2 demonstrates the flowchart for the proposed hybrid VNSA for solving DAP.

The flowchart in Figure 2 starts with parameter initialization in which the parameters are Neighborhood structure, Temperature, Cooling and Terminate. The neighborhood structure in VNS method is defined in a way that it makes somehow big modification over the solution. To do that, more sites are changed randomly to discover new solutions in far distance through the search space. This way, it would be possible to extract higher quality solutions through whole space. As well as temperature, cooling rate and terminate are initialized by 500, 0.2 and 0.1 respectively.

In the flowchart, there are some parameters – namely Y, K, Delta and P – in which Y is a new solution generated using neighborhood structure by changing site indexes, K is used as a counter for inner loop of VNS method, Delta is the difference between quality of current and new solution. The new solution is better if Delta holds positive value. Likewise, P is the acceptance probability of moving from current solution to a new generated solution.

In the next step, the current solution (allocation) is initialized by random. The solution (allocation) is represented as a one dimensional array with n columns, where n is the number of fragments in DDS. In this representation, the fragments and sites are shown by array indexes and array contents respectively. For instance, a sample allocation for 20 fragments and 4 sites is represented in Figure 3.

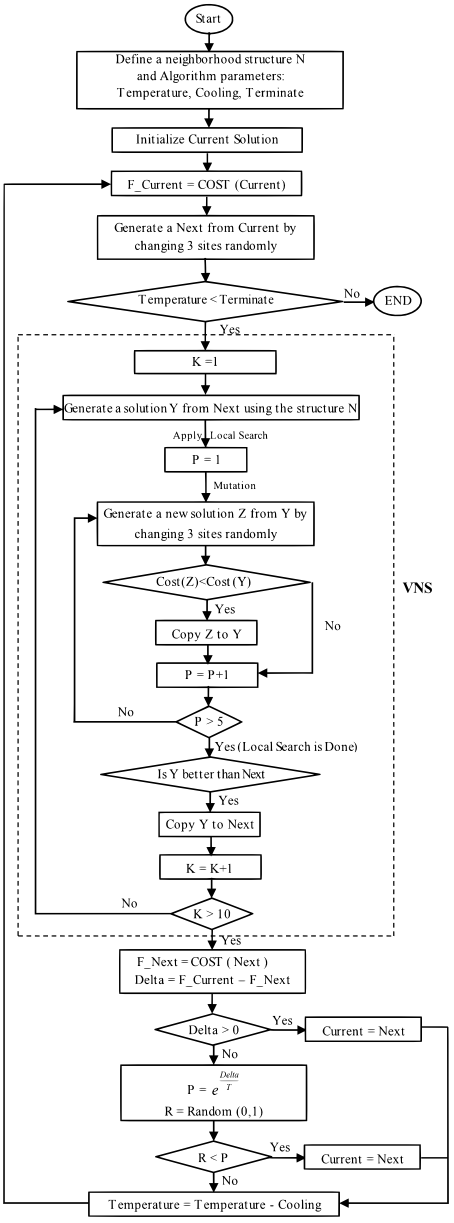


Figure 2. VNSA Flowchart

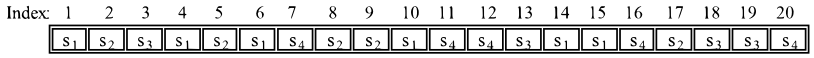


Figure 3. A sample solution representation for 20 fragments and 4 sites

The first solution is initialized randomly using the algorithm shown in Figure 4.

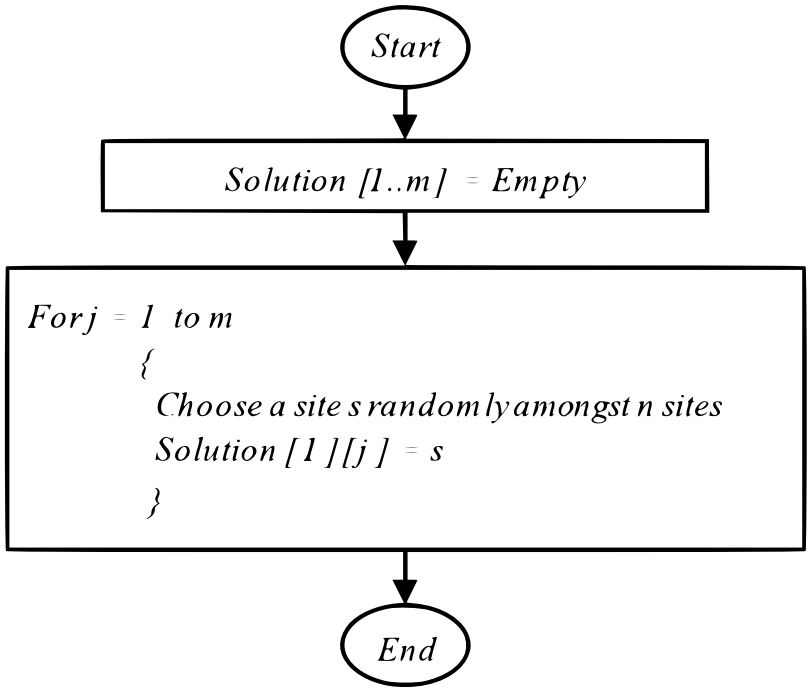


Figure 4. First solution initialization

Later on, system works in consecutive sessions until the temperature becomes less than termination value. The termination value is selected practically during the program running time. The termination value equal to 0.1 was adequate for the algorithm to execute fast and achieve good results. The total cost values are computed according

to the descriptions and equations given in the beginning of Section 2. Algorithm for computing the total cost value for a solution is given in Figure 5.

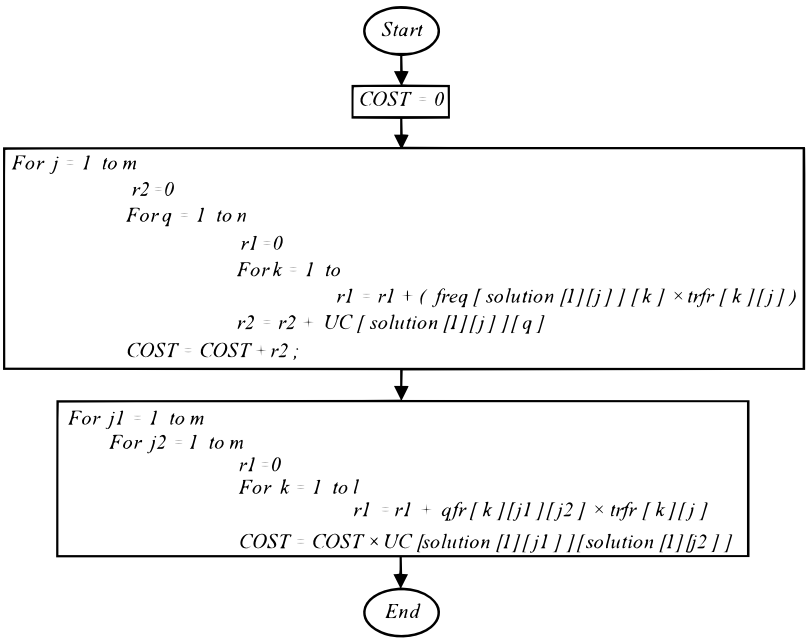


Figure 5. Cost value Calculation

Afterwards, a new solution named by 'Next' is generated by changing 3 sites on current solution randomly. Thereafter, VNS technique is applied over solution 'Next' to do more exploration and exploitation on search space. This way, solution 'Next' is adjusted and becomes more accurate. VNS algorithm is presented in Figure 2. In order to prevent time consuming by VNS, the inner loop is iterated 10 times. SA and VNS individually are single-solution-based algorithms and they are of fast kind. In case of combination, VNS adds more computational costs to the hybrid system. In order to have still fast hybrid system, a small size of VNS is applied. That's reason why the main loop in VNS is executed 10 times. VNS method starts with solution 'Next' and jumps

to somewhat far neighbor y (Exploration). Then it starts to search locally around solution y to find a solution better than solution 'Next'. If it finds such solution, 'Next' is replaced by it, otherwise it jumps from solution 'Next' to another solution y and continues until either better solution is found or loop is terminated.

Once the VNS part is terminated, VNSA continues with the rest of SA. In this part, first of all the cost of solution 'Next' and corresponding Delta value are calculated. Then, if Delta value is positive, algorithm decides to move directly from current to Next. Otherwise it will move to Next based on a probability value P. This way if a randomly generated value R is less than P, it moves to Next otherwise algorithm will continue with current solution. Afterwards, cooling value is subtracted from temperature value and VNSA continues with the next session.

3 Evaluation and Experimental results

Evaluation results and performance of the proposed VNSA are presented in this section. To evaluate VNSA, benchmark set reported in [1, 2] is taken into account. This benchmark set has been generated using the rules mentioned in Section 2 and has been used by all state-of-the-art methods in literature [11]. All parameter values commonly used by state-of-the-art methods are shown in Table 2. Moreover, the unit cost is considered between zero and one. Also, Number of fragments and sites are considered as equal [1].

Table 2 Parameter values used in proposed VNSA

Parameter Description	Notation	Value
Fragment Size	C	10
Transmission costs between two sites	UCN	[0-1]
Number of Transactions	L	20
Probability of transaction requested at a site	RPT	0.7
Probability of fragment accessed by transaction	APF	0.4
Probability of a transaction needing data transmission between two sites	APFS	0.025
Temperature	T	200
Cooling	C	0.1
Terminate	Tr	0.1

The proposed VNSA was developed in Matlab® programming language environment and performed over a system with 2.00 GHz CPU and 2 GB of memory.

The following state-of-the-art methods are compared to VNSA in terms of total cost and time: ACO (Ant Colony Optimization) [1, 27], RTS (Robust Tabu Search) [1, 26], GA (Genetic Algorithm) [1, 28], HG-MTS (Hybrid Genetic Multi-start Tabu Search) [1], PSO-DAP (Particle Swarm optimization) [2] and DEVNS (Differential Evolution Variable Neighborhood Search) [35]. The total cost achieved by all methods over DAP instances is presented in Table 3.

Table 3 Cost comparison of methods for different DAP instance s izes (cost value is column × 10^6)

Size	HG-MTS	GA3	GA2	GA1	RTS	ACO	PSO-DAP	DEVNS	VNSA
5	0.04	0.04	0.04	0.04	0.04	0.04	**0.02**	0.03	0.03
10	0.31	0.31	0.31	0.32	0.31	0.31	**0.05**	0.06	0.06
15	0.98	0.98	0.98	0.99	0.98	0.98	**0.41**	0.52	0.48
20	2.61	2.64	2.64	2.63	2.61	2.61	**0.77**	1.47	1.31
25	5.19	5.24	5.26	5.25	5.19	5.19	3.74	**3.35**	**3.35**
30	10.27	10.41	10.42	10.39	10.27	10.27	3.19	**2.98**	**2.98**
35	16.39	16.66	16.61	16.64	16.39	16.39	9.04	**8.51**	8.76
40	25.92	26.21	26.33	26.28	25.9	25.91	19.24	20.71	**19.05**
45	37.27	37.82	37.8	37.73	37.26	37.28	27.04	**25.56**	26.20
50	53.88	54.69	54.63	54.76	53.89	53.93	34.43	29.38	**28.76**
55	71.21	72.13	72.40	72.72	71.19	71.30	51.38	**46.19**	46.23
60	90.20	91.56	91.49	91.76	**90.16**	90.35	97.78	90.20	90.20
65	**112.08**	113.84	113.75	113.59	112.13	112.31	125.01	113.45	112.10
70	146.15	148.18	148.8	148.48	146.19	146.41	138.69	131.72	**130.55**
75	177.65	180.63	180.75	180.04	177.7	177.90	171.47	168.34	**165.81**
80	**219.18**	222.96	222.80	223.10	219.26	219.40	260.86	234.26	223.04
85	261.99	266.19	266.15	267.04	261.88	262.24	260.63	260.33	**260.12**
90	315.86	320.58	320.93	320.88	315.86	316.11	287.09	**279.49**	283.04
95	369.91	375.29	375.85	375.49	369.92	370.14	365.06	355.04	**352.93**
100	427.98	434.45	436.15	436.19	428.28	428.40	481.58	424.08	**421.32**

Even though it can be noticed from Table 3 that HG-MTS, RTS, PSO-DAP and DEVNS are better that VNSA in 2, 1, 4 and 6 problems respectively, differences between results are very small. Likewise, VNSA is better than all methods in Table 3 for 9 problem instances. In order to check similarity of VNSA to other eight methods as well as to indicate the rank of VNSA among 10 methods, Friedman Aligned Rank test is performed. The test is carried out based on procedure described in [22, 23, 24, 30]. All ranks assigned to all problem-method pairs by Friedman Aligned Rank test are illustrated in Table 4.

Table 4 Friedman aligned ranks for all problem-method pairs.

Size	HG-MTS	GA3	GA2	GA1	RTS	ACO	PSO-DAP	DEVNS	VNSA
5	118	118	118	118	118	118	110	112	112
10	85	85	85	86	85	85	76	78	78
15	50	50	50	51	50	50	43	45	44
20	25	28	28	26	25	25	19	21	20
25	6	7	9	8	6	6	3	2	2
30	15	17	18	16	15	15	12	11	11
35	133	138	135	137	133	133	68	64	65
40	130	134	139	136	128	129	70	72	69
45	141	145	144	143	140	142	59	56	57
50	156	165	163	166	157	159	38	34	32
55	152	164	168	170	151	154	37	31	33
60	102	121	119	123	99	103	173	102	102
65	91	109	108	107	93	94	178	106	92
70	125	149	161	153	126	127	62	42	40
75	120	160	162	148	122	124	67	58	52
80	60	74	73	79	61	63	179	177	75
85	98	158	155	167	95	105	88	87	80
90	147	174	176	175	147	150	30	22	29
95	96	169	172	171	97	104	66	39	35
100	53	71	89	90	54	55	180	41	36

Table 5 shows the rank averages, FAR value and p value. To calculate FAR, the equation in [23] with statistical significance of χ^2 distribution and k-1 degrees of freedom is used, where k denotes number of methods. Likewise to indicate the significant difference between all methods, p value is calculated. According to Table 6, rank average for VNSA is smaller than others which indicates that VNSA is best performing algorithm. Also, DEVNS and PSO-DAP take the second and third places respectively. It can be seen that p value is very close to zero which shows that there is significant statistical difference between all methods in Table 5. Likewise, very small p value implies that VNSA is statistically different than other 8 methods.

Table 5 Average Friedman aligned ranks, FAR and p value

Method	Average Friedman aligned ranks
HG-MTS	95.15
GA3	111.8
GA2	113.6
GA1	113.5
RTS	95.1
ACO	97.05
PSO-DAP	77.9
DEVNS	60
VNSA	**53.2**
FAR	**56.3264**
P values	**0.0012**

Running times of all methods for 20 problem instances are given in Table 6.

Table 6 Running time comparison of methods for increasing DAP instance sizes.

Size	HG-MTS	GA3	GA2	GA1	RTS	ACO	PSO-DAP	DEVNS	VNSA
5	1.44	88.11	56.11	76.27	0.83	9.26	**0.74**	5.41	0.81
10	2.45	94.91	50.37	87.80	2.73	14.52	**1.35**	4.92	1.55
15	2.65	104.13	66.22	90.76	5.66	13.74	**2.17**	5.39	2.72
20	4.17	167.22	84.13	123.79	8.89	17.91	3.65	11.45	**3.12**
25	5.21	125.30	81.96	131.98	14.52	25.86	4.25	13.07	**3.91**
30	7.38	137.02	104.64	132.46	20.89	31.17	6.45	16.50	**5.45**
35	10.73	151.02	111.87	150.06	29.06	43.31	6.81	14.27	**6.23**
40	15.60	173.21	128.75	166.80	37.05	56.59	8.85	22.43	**7.56**
45	20.80	202.10	159.10	191.93	48.67	80.92	8.42	20.41	**6.89**
50	26.80	359.57	207.56	471.98	62.74	105.33	9.60	23.45	**8.58**
55	27.22	261.71	201.43	258.31	76.07	126.00	13.74	27.78	**12.15**
60	39.56	290.46	208.37	315.31	91.79	166.55	**16.09**	35.43	16.72
65	48.92	336.01	284.08	421.93	109.20	204.35	16.09	38.20	**15.56**
70	63.13	358.03	34.20	536.15	131.54	320.62	17.34	27.12	**17.05**
75	73.41	380.81	379.07	609.77	155.31	309.51	17.01	26.69	**15.31**
80	87.84	416.18	331.17	464.17	193.63	396.18	16.29	33.27	**14.39**
85	102.79	586.21	364.71	532.05	195.80	807.43	**18.31**	39.61	18.68
90	123.19	531.13	400.37	563.15	215.58	621.55	20.98	44.53	**18.76**
95	143.16	569.92	974.24	629.55	250.72	725.93	18.15	41.93	**16.38**
100	179.07	808.82	568.73	1236.30	278.63	1203.99	20.97	48.22	**20.73**

Table 6 indicates that the suggested VNSA outperforms all competitors in terms of speed. The reason is that VNSA is mixed of two single-solution-based algorithms, SA and VNS, which are much faster

than population-based methods. Likewise, PSO-DAP is the second fastest method in which it is faster than VNSA in 5 out of 10 problem instances.

4 Conclusions and Future Works

This paper proposes an innovative hybrid method (VNSA) for solving data allocation problem in DDS. The method works based on a strategy to combine SA algorithm and VNS technique. To have an effective method, SA method is combined with effective neighborhood generation method which is called as VNS. The VNS technique is added to SA to provide more exploration and exploitation power and extract more accurate solutions. The obtained results demonstrate that the proposed hybrid method outperforms all state-of-the-art methods reported in literature. The proposed method was evaluated over different sizes of DAP in terms of cost and running time. According to the obtained results, VNSA took the first rank in 9 problems out of 20 problems in terms of cost value. Likewise, the results illustrate that VNSA is faster than all other methods under consideration. Also, the Friedman Aligned Rank test proved that significant difference between all methods exists, and VNSA is statistically different than all other methods. Future works are planned to add more details and parameters to DAP equations according to [25]. In new aspect of problem, Equation (3) will be enhanced by adding extra parameters like the number of sites involved in processing query and Communication costs between sites. Also variable X_{ij} will be inserted to Equation (3) to distinguish the existence of data fragment in the concerned site reached by the relevant query.

References

[1] U. Tosun, "Distributed database design using evolutionary algorithms," *Journal of Communications and Networks*, vol. 16, no. 4, pp. 430–435, 2014.

[2] M. Mahi, O. Baykan, and H. Kodaz, "A new approach based on particle swarm optimization algorithm for solving data allocation problem," *Applied Soft Computing*, Elsevier, vol. 62, pp. 571–578, 2018. DOI: https://doi.org/10.1016/j.asoc.2017.11.019.

[3] I. Ahmad, K. Karlapalem, and Y. K. Kwok, "Evolutionary Algorithms for Allocating Data in Distributed Database Systems," *Distributed and Parallel Databases*, vol. 11, pp. 5–32, 2012. DOI: https://doi.org/10.1023/A:1013324605452.

[4] A. Helal, "Dynamic Data Reallocation for Skew Management in Shared-Nothing Parallel Databases," *Distributed and Parallel Databases*, vol. 5, pp. 271–288, 1997.

[5] J. Liu, S. Zhang, C. Wu, J. Liang, X. Wang, and K. L. Teo, "A hybrid approach to constrained global optimization," *Applied Soft Computing*, Elsevier, vol. 47, no. C, pp. 281–294, 2016. DOI: 10.1016/j.asoc.2016.05.021.

[6] Y. Zhou, J. Wang, Z. Qiu, Z. Bi, and Y. Cai, "Differential evolution with guiding archive for global numerical optimization," *Applied Soft Computing*, Elsevier, vol. 43, no. C, pp. 424–440, 2016. DOI: https://doi.org/10.1016/j.asoc.2016.02.011.

[7] K.V. Price, "An Introduction to Differential Evolution," in *New Ideas in Optimization*, D. Corne, M. Dorigo, and F. Glover, Eds. London, UK: McGraw-Hill, 1999, pp. 79–108.

[8] R. Storn and K. Price, "Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, pp. 341-–359, 1997. DOI: https://doi.org/10.1023/A:1008202821328.

[9] P. Hansen and N. Mladenovic, "Variable neighborhood search: Principles and Applications," *European Journal of Operational Researchs*, Elsevier, vol. 130, no. 3, pp. 449–467, 2001. DOI: https://doi.org/10.1016/S0377-2217(00)00100-4.

[10] L. Liberti and M. Drazic, "Variable Neighbourhood Search for the Global Optimization of Constrained NLPs," in *Proceeding of GO*, 2005, pp. 1–5. Corpus ID: 15444192.

[11] R. K. Adl and S. M. T. R. Rankoohi, "A new ant colony optimization based algorithm for data allocation problem in distributed databases," *Knowledge and Information Systems*, vol. 20, no. 3, pp. 349–373, 2009. DOI: 10.1007/s10115-008-0182-y.

[12] M. Wang, S. Feng, C. Ouyang, and Z. Li, "RFID tag oriented data allocation method using artificial immune network," in *27th Chinese Control and Decision Conference (2015 CCDC)*, IEEE, pp. 5218–5223, 2015.

[13] D. Dasgupta, S. Yua, and F. Nino, "Recent Advances in Artificial Immune Systems: Models and Applications," *Applied soft computing*, Elsevier, vol. 11, pp. 1574–1587, 2011.

[14] A. Singh, K. S. Kahlon, and R. S. Virk, "Replicated static allocation of fragments in distributed database design using biogeography-based optimization," in *Int. Conf. on Advances in Communication, Network, and Computing, CNC*, 2014, pp. 462–472.

[15] G. Sen, M. Krishnamoorthy, N. Rangaraj, and V. Narayanan, "Mathematical models and empirical analysis of a simulated annealing approach for two variants of the static data segment allocation problem," *Networks*, pp. 4–22, 2016.

[16] V. F. Yu, A. A. N. P. Redi, Y. A. Hidayata, and O. J. Wibowo, "A simulated annealing heuristic for the hybrid vehicle routing problem," *Applied Soft Computing*, Elsevier, pp. 119–132, 2017.

[17] U. Tosun, "A new recombination operator for the genetic algorithm solution of the quadratic assignment problem," *Procedia Computer Science*, vol. 32, pp. 29–36, 2014. DOI: https://doi.org/10.1016/j.procs.2014.05.394.

[18] R. Jovanovic, M. Tuba, and S. Vob, "An ant colony optimization algorithm for partitioning graphs with supply and demand," *Applied Soft Computing*, Elsevier, pp. 317–330, 2017.

[19] P. Cortes, J. Munuzuri, L. Onieva, and J.Fernandez, "A Tabu Search algorithm for dynamic routing in ATM cell-switching networks," *Applied Soft Computing*, Elsevier, pp. 449–459, 2011.

[20] H. I. Abdalla, "A new data re-allocation model for distributed database systems," *International Journal of Database Theory and Application*, vol. 5, no. 2, pp. 45–60, 2012. Corpus ID: 741957.

[21] I. Ahmad, K. Karlapalem, Y. K. Kwok, and S. K. So, "Evolutionary algorithms for allocating data in distributed database systems," *Distributed And Parallel Databases*, vol. 11, no. 1, pp 5–32, 2002. DOI: http://dx.doi.org/10.1023/A:1013324605452.

[22] U. Tosun, T. Dokeroglu, and A. Cosar, "Heuristic algorithms for fragment allocation in a distributed database system," in *Computer and Information Sciences III*, London: Springer, 2013, ch. 41, pp. 401–408. DOI: https://doi.org/10.1007/978-1-4471-4594-3_41.

[23] J. Derrac, S. Garcia, D. Molina, and F. Herrera, "A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms," *Swarm Evolutionary Computing*, vol. 1, no. 1, pp. 3–18, 2011.

[24] A. Acan and N. Lotfi, "A multiagent, dynamic rank-driven multideme architecture for real-valued multiobjective optimization," *Artificial Intelligence Reviews*, Springer Netherlands, vol. 48, no. 1, pp. 1–29, 2017. DOI: https://doi.org/10.1007/s10462-016-9493-7.

[25] A. Sewisy, A. Amer, and H. Abdalla, "A Novel Query-Driven Clustering-Based Technique for Vertical Fragmentation and Allocation in Distributed Database Systems," *International Journal on Semantic Web and Information Systems*, vol. 13, no. 2, pp. 27–54, 2017. DOI: 10.4018/IJSWIS.2017040103.

[26] E. Taillard, "Robust taboo search for the quadratic assignment problem," *Parallel Computing*, vol. 17, no. 4-5, pp. 443–455, 1991. DOI: https://doi.org/10.1016/S0167-8191(05)80147-4.

[27] M. Dorigo, V. Maniezzo, and A. Colorni, "Ant system: Optimization by a colony of cooperating Agents," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 26, no. 1, pp. 29–41, Feb. 1996. DOI: 10.1109/3477.484436.

[28] D. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Boston, MA, United States: Addison-Wesley Longman Publishing Co., Inc., 1989, 372 p. ISBN:978-0-201-15767-3.

[29] D. Nashat and A. Amer, "A Comprehensive Taxonomy of Fragmentation and Allocation Techniques in Distributed Database Design," *ACM Computing Surveys*, pp. 1–25, 2018. Article No.: 12, DOI: https://doi.org/10.1145/3150223.

[30] H. Abdalla and A. M. Artoli, "Towards an Efficient Data Fragmentation, Allocation, and Clustering Approach in a Distributed Environment," *Information*, vol. 10, no. 3, 2019. Article no. 112. DOI: 10.3390/info10030112.

[31] A. Sinha and P. Jana, "A hybrid MapReduce-based k-means clustering using genetic algorithm for distributed datasets," *Journal of Supercomputing*, vol. 74, pp. 1562–1579, 2018.

[32] A. A. Amer, M. H. Mohamed, and K. Al Asri, "ASGOP: An Aggregated Similarity-Based Greedy-Oriented Approach for Relational DDBSs Design," *Heliyon*, vol 6, no. 1, 2020. pp. e03172. DOI: https://doi.org/10.1016/j.heliyon.2020.e03172.

[33] P. M. J. Apers, "Data Allocation in Distributed Database Systems," *ACM Transactions on Database Systems*, vol. 13, no. 3, pp. 263–304, 1988. DOI: https://doi.org/10.1145/44498.45063.

[34] P. R. Bhuyar and A. D. Gawande, "Distributed Database Fragmentation and Allocation," *Journal of Data Mining and Knowledge Discovery*, vol. 3, no. 1, pp. 58–64, 2012. ISSN: 2229-6662 & ISSN: 2229-6670.

[35] N. Lotfi, "Data allocation in distributed database systems: a novel hybrid method based on differential evolution and variable neighborhood search," *SN Applied Sciences*, vol. 1, Article no. 1724, 2019. DOI: https://doi.org/10.1007/s42452-019-1787-3.

Nasser Lotfi, Jamshid Tamouk

Received December 14, 2020
Accepted December 25, 2020

Nasser Lotfi
Faculty of Engineering, Cyprus Science University, Girne, N.
Cyprus via Mersin 10, Turkey
E–mail: nasser.lotfi2020@gmail.com

Jamshid Tamouk
Faculty of Engineering, Eastern Mediterranean University,
Famagusta, N. Cyprus via Mersin 10, Turkey
E–mail: jamshid.tamouk20@gmail.com

Computer Science Journal of Moldova, vol.**29**, no.2(86), 2021

A practical digital signature scheme based on the hidden logarithm problem

D.N. Moldovyan

Abstract

A candidate for practical post-quantum digital signature algorithm based on computational difficulty of the hidden discrete logarithm problem is introduced. The used algebraic carrier represents a 4-dimensional finite non-commutative associative algebra defined over the field $GF(p)$, which is caracterized in using a sparse basis vector multiplication table for defining the vector multiplication operation. Structure of the algebra is studied. Three types of the commutative groups are contained in the algebra and formulas for number of groups of every type are obtained. One of the types represents groups of the order $(p-1)^2$, possessing 2-dimensional cyclicity, and one of them is used as a hidden group in the signature scheme developed using a new method for implementing a general criterion of post-quantum resistance proposed earlier.

Keywords: finite associative algebra, non-commutative algebra, commutative finite group, discrete logarithm problem, hidden logarithm problem, public key, digital signature, post-quantum cryptosystem.

MSC 2010: 68P25, 68Q12, 68R99, 94A60, 16Z05, 14G50

1 Introduction

Currently the development of the public-key digital signature algorithms and protocols that are resistant to quantum attacks (i. e., attacks including computations on a hypothetic quantum computer) attracts significant attention of the cryptographic community [1], [2].

Usually the research activity in the area of the post-quantum cryptography is focused on the development of the practical public-key cryptoschemes based on the computationally complex problems different from the factoring problem and the discrete logarithm problem (DLP). Actually, both the factoring problem and the DLP can be solved in polynomial time on a quantum computer [3]–[5].

Recently it was shown that the hidden discrete logarithm problem (HDLP) defined in finite non-commutative associative algebras (FNAAs) represents an attactive primitive for designing practical post-quantum cryptoschemes of the following types: commutative encryption algorithms [6], public key-agreement protocols [7], and digital signature schemes [8]–[10].

In the papers [11], [12] a general criterion for ensuring resistance of the HDLP-based signature schemes to hypothetic future quantum attacks based on quantum algorithms for computing the length of the periods of periodic functions was proposed. However, the signature schemes [11], [12] satisfying the said criterion had been developed using a method of doubling the signature verification equation.

In this paper a signature scheme implementing the said general design criteria without doubling the verification equation is developed. In addition a 4-dimensional FNAA is used as algebraic carrier of the signature scheme. Due to these features, the procedures of signature generation and verification are significantly faster and the size of the public key and size of signature are significantly smaller in comparision with the signature schemes [11], [12].

2 Preliminaries

2.1 Forms of the HDLP and design criteria

Usually the HDLP is defined in the m-dimensional ($m = 4, 6$, and 8) FNAAs as follows. One selects at random an integer $x < q$ and a generator G of a finite cyclic group of prime order q, which is contained in the used FNAA. To provide a required level of secutity the prime q should have sufficiently large size (≥ 256 bits). Then the vector G^x is

computed and two elements of the public key are formed: $Y = \psi_1(G^x)$ and $Z = \psi_2(G)$, where ψ_1 and ψ_2 are two different automorphism-map (or homomorphism-map) operations. The operations ψ_1 and ψ_2 are secret, therefore the potential attacker does not know the basic finite group in which the exponentiation operation had been performed. The masking operations ψ_1 and ψ_2 possess the property of mutual commutativity with the exponentiation operation that contributes mainly to the security, therefore, different known DLP-base signature algorithms can be used as prototypes of the HDLP-based algorithms.

In some of the HDLP-based signature schemes there is used the public key representing the triple of vectors (Y, Z, T), where the vector T is a fitting parameter in the verification equation. The hidden cyclic group is called the base group. The vectors Y, Z, and T are contained in other three different cyclic groups.

The rationale of the post-quantum resistance of the known HDLP-based signature schemes is quite straightforward: potential attacker knows no elements of the hidden cyclic group in which the exponentiation operation is performed, therefore, to compute the logarithm value x the Shor quantum algorithm [3] cannot be directly applied. Indeed, that algorithm is based on the ability of a quantum computer to perform a discrete Fourier transform (used to compute the period length of periodic functions) extremely efficiently for functions that take on values in a finite cyclic group [5]. In particular, to find the logarithm value x one constructs a periodic function whose values lie in a fixed cyclic group, which contains a period with the length depending on the value x.

For the case of the HDLP-based signature algorithms described in papers [8], [9] one can define the periodic function $F(i, j) = Y^i \circ T \circ Z^j$ in two integer variables i and j. This function contains a period with the length equal to $(-1, x)$:

$$F(i, j) = Y^i \circ T \circ Z^j = Y^{i-1} \circ T \circ Z^{j+x} = F(i - 1, j + x).$$

Thus, the design criterion related to the known HDLP-based signature algorithms can be formulated as follows.

Criterion 1. *The periodic functions constructed on the base of public parameters of the signature algorithm and containing a period with the length depending on the discrete logarithm value should take on values in different finite cyclic groups contained in the FNAA. Besides, no cyclic group can be pointed out as a preferable finite group for the values of the function* $F(i,j)$.

It is reasonable to assume that in the future, quantum algorithms will be developed that will effectively find the period length for functions that take on values within the framework of the whole FNAA used as algebraic support of the signature scheme. Taking into accoun such potential possibility, the following *advanced* criterion of the post-quantum resistance had been proposed in [11], [12].

Criterion 2. *Based on the public parameters of the signature scheme, the construction of a periodic function containing a period with the length depending on the discrete logarithm value should be a computationally intractable task.*

To implement a signature scheme satisfying the advanced criterion, one can use the idea of masking periodicity depending on the discrete logarithm value. To implement this idea, in the signature schemes proposed in [11], [12] a commutative group with two-dimensional cyclicity had been used as a hidden group. A finite commutative group is called group with the μ-dimensional cyclicity, if the group is generated by a generator system of μ independent elements possessing the same order value.

Suppose, in a hypothetic signature scheme the public key (Y, Z) includes elements computed as fallows: $Y = \psi_1(G^x)$ and $Z = \psi_2(GQ)$, where elements G and Q are generators of two different cyclic groups contained in the hidden commutative group with 2-dimensional cyclicity. Since each of the values G and Q has the same order, you cannot eliminate the Q multiplier effect by performing an exponentiation operation. Therefore, periodic functions, like the functions $F(i,j) = Y^i \circ Z^j$ or $F(i,j) = Y^i \circ T \circ Z^j$, will only show the periodicity associated with the value of the order of the elements G and Q.

This idea is quite trivial, but when it is implemented in specific signature algorithms, there are a number of complications that

must be overcome. The implementation of this idea in signature schemes [11], [12] required doubling the size of the public key and doubling the verification equation. At the same time, the size of the signature increased by three or more times compared to signature schemes that meet the Criterion 1.

In Section 3, a new 4-dimensional FNAA with a fast vector multiplication operation is proposed as an algebraic carrier of the HDLP-based signature algorithms. This algebra contains sufficiently large number of the commutative groups of the order $(p-1)^2$, which possess 2-dimensional cyclicity. Section 4 presents a novel method for designing HDLP-based signature schemes satisfying Criterion 2, which are free from the disadvantages of the implementations [11], [12].

2.2 Setting finite non-commutative algebras

Suppose a finite m-dimensional vector space is defined over the ground finite field $GF(p)$ and, additionally to the addition operation and scalar multiplication, a vector multiplication operation is defined so that it is distributive at the right and at the left relatively the addition operation. Then we have the algebraic structure called the m-dimensional finite algebra. Some algebra element A can be denoted in the following two forms: $A = (a_0, a_1, \ldots, a_{m-1})$ and $A = \sum_{i=0}^{m-1} a_i \mathbf{e}_i$, where $a_0, a_1, \ldots, a_{m-1} \in GF(p)$ are called coordinates; $\mathbf{e}_0$, $\mathbf{e}_1$, ... $\mathbf{e}_{m-1}$ are basis vectors.

The vector multiplication operation ($\circ$) of two m-dimensional vectors A and B is set as $A \circ B = \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} a_i b_j (\mathbf{e}_i \circ \mathbf{e}_j)$, where each of the products $\mathbf{e}_i \circ \mathbf{e}_j$ is to be substituted by a single-component vector $\lambda \mathbf{e_k}$, where $\lambda \in GF(p)$, which is indicated in the cell at the intersection of the ith row and jth column of the so-called basis vector multiplication table (BVMT). To define associative vector multiplication operation, the BVMT should define associative multiplication of all possible triples of the basis vectors $(\mathbf{e}_i, \mathbf{e}_j, \mathbf{e}_k)$: $(\mathbf{e}_i \circ \mathbf{e}_j) \circ \mathbf{e}_k = \mathbf{e}_i \circ (\mathbf{e}_j \circ \mathbf{e}_k)$.

The BVMT shown as Table 1 sets 2-dimensional finite commutative associative algebra that is a finite field $GF(p^2)$, if the structural constant $\lambda \neq 0$ is a quadratic non-residue in $GF(p)$ [14]. If λ is a

Table 1. The BVMT setting the 2-dimensional commutative algebra

$\circ$	$\mathbf{e}_0$	$\mathbf{e}_1$
$\mathbf{e}_0$	$\mathbf{e}_0$	$\mathbf{e}_1$
$\mathbf{e}_1$	$\mathbf{e}_1$	$\lambda\mathbf{e}_0$

quadratic residue, the set of invertible elements of the said algebra represents the multiplicative group Γ' possessing 2-dimensional cyclicity and having order equal to $(p-1)^2$. In general, a finite group is called group with μ-dimensional cyclicity if its minimum generator system includes μ independent elements of the same order [15].

When constructing the HDLP-based public-key cryptoschemes, one uses hidden cyclic groups whose order is equal to a prime of sufficiently large size. Therefore the used FNAAs are defined over the field $GF(p)$ whose characteristic is equal to a prime $p = 2q + 1$, where q is also a prime. In the last case the group Γ' includes the commutative subgroup Γ with the minimum generator system $< G_1, G_2 >$, in which the elements G_1 and G_2 have prime order q. Different pairs of integers i and j, such that $0 < i < q$ and $0 < j < q$, define different elements $G_{ij} = G_1^i \circ G_2^j$ having order q. Every element G_{ij} is a generator of some cyclic group of the prime order q. For a fixed pair of integers (i, j), where $i, j = 1, 2, \ldots, q-1$, each of the formulas $G_k = G_{ij} \circ G_1^k$ and $G_k = G_{ij} \circ G_2^k$, where $k = 0, 1, \ldots, q-1$, sets q generators of q different cyclic groups of the order q.

To set FNAAs of arbitrary even dimensions m, one can use two unified methods, [8], [16], each of which is represented by a simple mathematical formula parameterized by values $m = 2, 4, 6, \ldots 2i$, which describes the content of all cells of the BVMT as the basis vector $\mathbf{e}_{f_m(i,j)} = \mathbf{e}_i \circ \mathbf{e}_j$, where the function $f_m(i, j)$ takes on the values from the set $0, 1, 2, \ldots m-1$. For a fixed value m, each of that methods sets an m-dimensional FNAA in which the computational complexity of the vector multiplication operation is approximately equal to m^2 multiplications in $GF(p)$.

In order to provide a higher performance of the developed signature

scheme, in the present paper it is used a specially composed particular BVMT defining a 4-dimensional FNAA with the vector multiplication operation having complexity equal to ≈ 8 multiplications in $GF(p)$.

3 The used algebraic support and its properties

The 4-dimensional FNAA used as algebraic carrier is set by a sparse BVMT represented by Table 2, where the structural constant $\lambda \neq 0$. This algebra contains the global two-sided unit $E = (1, 1, 0, 0)$. The vectors $G = (g_0, g_1, g_2, g_3)$ satisfying the non-equality $g_0 g_1 \neq \lambda g_2 g_3$ are invertible. The vectors $N = (n_0, n_1, n_2, n_3)$ satisfying the condition $n_0 n_1 = \lambda n_2 n_3$ are non-invertible. It is easy to show that the number of non-invertible vectors is equal to $\eta_N = p^3 + p^2 - p$. Correspondingly, the number of invertible vectors, i. e. the order of the multiplicative group of the algebra, is equal to

$$\Omega = p^4 - \eta_N = p(p-1)\left(p^2 - 1\right). \tag{1}$$

To study structure of the algabra, consider different sets of the algebra elements X that are mutually permutable with a fixed vector A. The elements $X = (x_0, x_1, x_2, x_3)$ can be computed from the vector equation $A \circ X = X \circ A$ that can be reduced to solving the following system of three linear equations with the unknowns x_0, x_1, x_2, and x_3:

$$\begin{cases} \lambda\left(a_3 x_2 - x_3 a_2\right) = 0; \\ a_2\left(x_0 - x_1\right) + x_2\left(a_1 - a_0\right) = 0; \\ a_3\left(x_1 - x_0\right) + x_3\left(a_0 - a_1\right) = 0. \end{cases} \tag{2}$$

Consider the following cases: i) $a_0 = a_1 = s$ and $a_2 = a_3 = 0$; ii) $a_0 \neq a_1$ and $a_2 = a_3 = 0$; iii) $a_2 = 0$, and $a_3 \neq 0$; iv) $a_2 \neq 0$, and $a_3 = 0$; v) $a_2 \neq 0$, and $a_3 \neq 0$.

Case i) relates to the set of the scalar vectors $S = sE$. Evidently, every vector of the considered 4-dimensional FNAA is permutable with every scalar vector. The set of scalar vectors Ψ is contained in every set of pairwise permutable vectors.

Table 2. The sparse BVMT setting a 4-dimensional FNAA ($\lambda \neq 0$)

$\circ$	$\mathbf{e}_0$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$
$\mathbf{e}_0$	$\mathbf{e}_0$	0	0	$\mathbf{e}_3$
$\mathbf{e}_1$	0	$\mathbf{e}_1$	$\mathbf{e}_2$	0
$\mathbf{e}_2$	$\mathbf{e}_2$	0	0	$\lambda\mathbf{e}_1$
$\mathbf{e}_3$	0	$\mathbf{e}_3$	$\lambda\mathbf{e}_0$	0

Case ii). From system (2) one can easily see that the solution space includes p^2 vectors $X = (x_0, x_1, 0, 0)$, where $x_0, x_1 = 0, 1, \dots p-1$. This set of vectors X represents a commutative subalgebra containing $2p-1$ non-invertible vectors, multiplicative group of which has 2-dimensional cyclicity and order equal to $\Omega_1 = (p-1)^2$.

Case iii). System (2) is reduced to the following system of two equations:

$$\begin{cases} x_2 = 0; \\ x_1 = x_0 - x_3 \dfrac{a_0 - a_1}{a_3}, \end{cases} \tag{3}$$

which defines the solution space

$$X = (x_0, x_1, x_2, x_3) = \left(d, d - h\frac{a_0 - a_1}{a_3}, 0, h \right), \tag{4}$$

where $d, h = 0, 1, \dots p-1$. For the fixed value $h = 0$, formula (6) defines the set of scalar vectors. For the values $h \neq 0$, every vector V of set (4) satisfies the condition of the Case iii), therefore V defines the set of vectors Φ_V permutable with V, which coincides with set (4) that represents a commutative subalgebra Φ_A of the considered 4-dimensional FNAA.

If $a_0 \neq a_1$, then set (4) includes $2p-1$ non-invertible vectors and $p^2-(2p-1)$ invertible ones that compose multiplicative group of subalgebra Φ_A, which possesses 2-dimensional cyclicity and has order equal to $(p-1)^2$. Below, commutative groups of this type are denoted as Γ_1 and are called groups of the Γ_1 type.

If $a_0 = a_1$, then set (4) includes p non-invertible vectors and (p^2-p) invertible ones that compose a cyclic multiplicative group of the order $p(p-1)$. Commutative groups of this type are denoted as Γ_2 and are attributed to the Γ_2 type.

Case iv). System (2) is reduced to the following system of two equations:

$$\begin{cases} x_3 = 0; \\ x_1 = x_0 + x_2 \dfrac{a_1 - a_0}{a_2}, \end{cases} \tag{5}$$

which defines the solution space

$$X = (x_0, x_1, x_2, x_3) = \left(d, d + h\frac{a_1 - a_0}{a_2}, h, 0 \right), \tag{6}$$

where $d, h = 0, 1, \ldots p-1$. For the values $h \neq 0$, every vector V of set (6) satisfies the condition of the Case iv), therefore V defines the set of vectors Φ_V permutable with V, which coincides with set (6) that is a commutative subalgebra Φ_A. Note that for $h = 0$ formula (6) defines the scalar vectors.

If $a_0 \neq a_1$, then set (6) includes $2p-1$ non-invertible vectors and $(p-1)^2$ invertible ones that compose a multiplicative group of the Γ_1 type. If $a_0 = a_1$, then set (6) includes p non-invertible vectors and $p(p-1)$ invertible ones that compose a multiplicative group of the Γ_2 type.

Case v). System (2) is reduced to the following system of two equations:

$$\begin{cases} x_3 = x_2 \dfrac{a_3}{a_2}; \\ x_1 = x_0 + x_2 \dfrac{a_1 - a_0}{a_2}, \end{cases} \tag{7}$$

which defines the solution space

$$X = (x_0, x_1, x_2, x_3) = \left(d,\ d + h\frac{a_1 - a_0}{a_2},\ h,\ h\frac{a_3}{a_2} \right), \tag{8}$$

where $d, h = 0, 1, \ldots p-1$. For arbitrarily fixed integers d and $h \neq 0$, every vector V of set (8) satisfies the condition of the Case v),

therefore V defines the set of vectors Φ_V permutable with V, which is described by formula (8) written for coordinates of the vector $V = (v_0, v_1, v_2, v_3) = \left(d,\ d + h\frac{a_1 - a_0}{a_2},\ h,\ h\frac{a_3}{a_2}\right)$:

$$X' = \left(x'_0, x'_1, x'_2, x'_3\right) = \left(d',\ d' + h'\frac{v_1 - v_0}{v_2},\ h',\ h'\frac{v_3}{v_2}\right), \qquad (9)$$

where $d', h' = 0, 1, \ldots p-1$. Substitution of the coordinates $v_0 = d$, $v_1 = d + h\frac{a_1 - a_0}{a_2}$, $v_2 = h$, and $v_3 = h\frac{a_3}{a_2}$ in formula (9) gives $\Phi_V = \Phi_A$. Due to the latter result one can conclude that Φ_A is the set of pairwise permutable vectors. Actually, Φ_A is a commutative subalgebra of the order p^2. Like in the cases ii), iii), and iv), for the fixed value $h = 0$ formula (8) describes the set of scalar vectors: $X = (d,\ d,\ 0,\ 0)$, where $d = 0, 1, \ldots p-1$. Thus we have come to the following conclusion:

Proposition 1. Every 4-dimensional vector, except scalar vectors, is included in a single commutative subalgebra Φ.

Consider possible types of the Φ_A relating to the Case v). Using the non-invertibility condition, one can write the following equation for coordinates of non-invertible vectors contained in Φ_A:

$$d\left(d + h\frac{a_1 - a_0}{a_2}\right) = \lambda h^2 \frac{a_3}{a_2}, \qquad (10)$$

where different pairs of integers (d, h) that satisfy equation (10) set different non-invertible vectors contained in Φ_A, that is a commutative subalgebra. Consider solution of equation (10) relatively the unknown value d for a fixed value of h:

$$d = \left(\frac{a_0 - a_1}{a_2} \pm \frac{\sqrt{(a_0 - a_1)^2 + 4\lambda a_2 a_3}}{2a_2}\right) h. \qquad (11)$$

The number of non-invertible vectors that are contained in Φ_A depends on the value $\Delta = (a_0 - a_1)^2 + 4\lambda a_2 a_3$.

If $\Delta = \delta \neq 0$ is a quadratic residue in $FG(p)$, then for every value $h = 1, 2, \ldots p-1$, formula (11) gives two different values of d, i. e.

$2(p-1)$ non-invertible vectors different from $(0,0,0,0)$. Taking the zero vector into account we have $\eta_N = 2p-1$. The multiplicative group of the subalgebra has order $\Omega_1 = (p-1)^2$ and possesses 2-dimensional cyclicity, i. e. it is a group of Γ_1 type.

If $\Delta = 0$, then for every value $h = 0, 1, \ldots p-1$ formula (11) gives the single value of d, i. e. p non-invertible vectors including the zero vector. Thus, the number of non-invertible vectors is equal to $\eta_N = p$. The multiplicative group is cyclic and has order $\Omega_2 = p^2 - p = p(p-1)$, i. e. it is a group of Γ_2 type.

If $a_0 \neq a_1$, then set (5) includes $2p-1$ non-invertible vectors and $(p-1)^2$ invertible ones that compose a multiplicative group of the Γ_1 type. If $a_0 = a_1$, then set (5) includes p non-invertible vectors and $p(p-1)$ invertible ones that compose a multiplicative group of the Γ_2 type.

If $\Delta = \delta \neq 0$ is a quadratic non-residue in $FG(p)$, then for every value $h = 1, 2, \ldots p-1$ formula (11) gives no solution of equation (10), with exception $(d,h) = (0,0)$ corresponding to the zero vector $(0,0,0,0)$. Subalgebra Φ_A represents a finite ground field $GF(p^2)$, the multiplicative group of which is cyclic and has the order $\Omega_3 = p^2 - 1$. A group of this type is called a group of the Γ_3 type.

Thus, the considered 4-dimensional FNAA contains commutative groups of the types Γ_1, Γ_2, and Γ_3.

4 The number of commutative groups of every type

Proposition 2. Number η_Φ of the commutative Φ subalgebras equals to $p^2 + p + 1$.

Proof. The set of the scalar vectors Ψ is contained in every Φ subalgebra. Due to the Proposition 1, every vector that is different from a scalar vector is contained in a single subalgebra of Φ type. Therefore, for the number η_Φ of the Φ subalgebras one can write:

$$\eta_\Phi = \frac{p^4 - p}{p^2 - p} = p^2 + p + 1. \qquad (12)$$

The Proposition 2 is proven.

Suppose k, t, and u denote number of different commutative groups of the types Γ_1, Γ_2, and Γ_3 correspondingly. Then we have $\eta_\Phi = k+t+u$ and

$$k + t + u = p^2 + p + 1. \tag{13}$$

Due to the Proposition 1 and formula (1) one can write:

$$\begin{aligned}
&(\Omega_1 - (\#\Psi - 1))\, k + (\Omega_2 - (\#\Psi - 1))\, t + (\Omega_3 - (\#\Psi - 1))\, u = \\
&\qquad = \Omega - (\#\Psi - 1)\,; \\
&\Big((p-1)^2 - (p-1)\Big)\, k + (p(p-1) - (p-1))\, t + \\
&+ \big(p^2 - 1 - (p-1)\big)\, u = p(p-1)\left(p^2 - 1\right) - (p-1);
\end{aligned}$$

$$(p-2)k + (p-1)t + pu = p^3 - p - 1. \tag{14}$$

Using formulas (13) and (14) one can easily derive the following two equations:

$$2k + t = (p+1)^2; \tag{15}$$

$$2u + t = p^2 + 1. \tag{16}$$

To find the value t, consider the number of non-invertible vectors A relating to the Case v), which define the Φ_A subalgebras containing multiplicative groups of the Γ_2 type. Such vectors satisfy the conditions $a_0 a_1 = \lambda a_2 a_3 \neq 0$ and $\Delta = (a_0 - a_1)^2 + 4\lambda a_2 a_3 = 0$ from which it is easy to get the conditions

$$a_0 \neq 0; \quad a_1 \neq 0; \quad a_1 = -a_0; \quad a_3 = \frac{a_0 a_1}{\lambda a_2},$$

where $a_0, a_2 = 1, 2, \ldots p-1$. Therefore, $(p-1)^2$ different non-invertible vectors A relating to the Case v) define the Φ subalgebras containing the multiplicative group of the Γ_2 type. Every one of the said Φ subalgebras includes $p-1$ non-invertible vectors that are different from the zero vector, therefore, the Case v) gives $\eta_{v)} = \frac{(p-1)^2}{p-1} = p-1$ different groups of the Γ_2 type. In addition, each of the Cases iii) and iv) gives one unique Γ_2-type group. Thus, we have got

$$t = p + 1.$$

Substituting the last value in (15) and (16) one gets:

$$k = \frac{p(p+1)}{2}; \tag{17}$$

$$u = \frac{p(p-1)}{2}. \tag{18}$$

5 Signature algorithm satisfying the advanced criterion of post-quantum resistance

The introduced 4-dimensional FNAA is well suited for development a HDLP-based digital signature scheme satisfying the advanced criterion of post-quantum resistance, since it contains a large number of the Γ_1-type commutative groups having 2-dimensional cyclicity. The proposed signature scheme is described as follows.

5.1 Computation of the signature scheme parameters

The FNAA used as algebraic support is defined over the finite ground field $GF(p)$ with prime $p = 2q + 1$, where q is a 256-bit prime. The required value p is set as generation of different 256-bit primes q until the value $2q + 1$ will satisfy a test for primality (for example, trying several arbitrary different integers $b < p$, one gets a value b such that $b^2 \bmod p \neq 1$ and $b^q \bmod p \neq 1$). Generation of the requiered prime p introduces the main contribution in the computational complexity of generating the parameters of the proposed signature scheme. Taking into account that on the average about 10^2 different primes q are to be tried, one can estimate that the generation of the value p takes about 10^7 multiplications modulo 257-bit prime.

The secret hidden group $\Gamma_{<G,U>}$ with 2-dimensional cyclicity is set as computation of its minimum generator system $< G, U >$ including invertible vectors G and U having order q.

Algorithm for generating a hidden group with 2-dimensional cyclicity:

1. Using the invertibility condition $a_0a_1 \neq a_2a_3$, generate a random invertible vector $A = (a_0, a_1, a_2, a_3)$ such that $\{a_2 \neq 0;\ a_3 \neq 0\}$ and compute the value

$$\Delta = (a_0 - a_1)^2 + 4\lambda a_2 a_3.$$

2. If $\Delta \neq 0$ is a quadratic non-residue, then go to step 1.

3. Calculate the vector $G = A^{\frac{p-1}{q}} = A^2$.

4. If $G = E$, then go to step 1. Otherwise generate a primitive element $s \in GF(p)$ and compute the scalar vector $S = sE = (s, s, 0, 0)$.

5. Generate a random integer $k < q$ and compute the vector

$$U = S^{\frac{p-1}{q}} \circ G^k.$$

6. Output the permutable vectors G and U each of which has order equal to 256-bit prime q.

Note the step 2 outputs a vector A that is an element of the commutative group of the Γ_1-type. This algorithm works quickly due to the fact that the number of the Γ_1-type groups is equal to $k = 2^{-1}p(p+1)$, and the latter contain about $k(p-1)^2 \approx 2^{-1}p^4$ elements of the used 4-dimensional FNAA, i. e. about half of all invertible 4-dimensional vectors.

Generation of the parameters of masking operations:

1. Select at random an invertible vector A possessing order equal to the value $p^2 - 1$, which satisfies the condition $G \circ A \neq A \circ G$.

2. Select at random an invertible vector B possessing order equal to the value $p^2 - 1$, which satisfies the conditions $B \circ A \neq A \circ B$ and $G \circ B \neq B \circ G$

The private values A and B are used as parameters of masking operations.

Computation of the public key (W, Y, Z):

1. Select a random integer $1 < x < q$ and compute the vector $W = A \circ G^x \circ B^{-1}$.

2. Compute the vectors $Y = B \circ G \circ B^{-1}$ and $Z = B \circ U \circ A^{-1}$.

The value x is an element of private key. The size of the public key (W, Y, Z) is equal to $\approx$ 384 bytes. Computational difficulty of the

public-key generation procedure is roughly equal to one exponentiation in the 4-dimensional FNAA (computational complexity of one exponention in the FNAA equals on the average to ≈ 3072 multiplications in $GF(p)$).

5.2 Signature generation procedure

Suppose one is to compute a signature to the electronic document M, using some specified 256-bit hash-function f_H.

The signature to the electronic document M is computed using the private key (x, A, B, G, U) as folows:

1. Using random integers $k < q$ and $t < q$, compute the vector

$$V = A \circ G^k \circ U^t \circ A^{-1}.$$

2. Compute the value e (the first signature element) from the document M to which the vector V is concatenated: $e = f_H(M, V)$, where f_H is a specified hash-function.

3. Compute the second signature element s as solution of the following quadratic equation

$$es^2 - s + xt + t = k \bmod q.$$

If the last equation has no solution, then go to step 1.

4. Compute the third signature element

$$d = s^{-1}(t - s) \bmod q.$$

On the average, computation of a signature (e, s, d) requires performing the first, second, and third steps of the signature generation procedure two times. Therefore, computational difficulty of the signature generation is roughly equal to four exponentiations in the 4-dimensional FNAA (≈ 12300 multiplications in $GF(p)$). The signature size is equal to ≈ 768 bits (96 bytes).

5.3 Signature verification procedure

The verification of the signature (e, s, d) to the document M is performed on the public key (W, Y, Z) as folows:

1. Using the public key, compute the vector V':

$$V' = \left(W \circ Y^{es} \circ Z \circ (W \circ Y \circ Z)^d\right)^s.$$

2. Compute the hash-function value e' from the document M to which the vector V' is concatenated: $e' = f_H(M, V')$.

3. If $e' = e$, then the signature is genuine. Otherwise reject the signature.

The computational difficulty of the signature verification procedure is roughly equal to three exponentiation operations in the FNAA ($\approx$ 9200 multiplications modulo p).

Correctness proof of the signature algorithm consists in proving that the correctly computed signature (e, s, d) will pass the verification procedure as a genuine signature. Due to the mutual commutativity of the automorphism-map operation with the exponentiation operation we have the following:

$$\begin{aligned}
V'_1 &= \left(W \circ Y^{es} \circ Z \circ (W \circ Y \circ Z)^d\right)^s = \\
&= \left(A \circ G^x \circ B^{-1} \circ \left(B \circ G \circ B^{-1}\right)^{es} \circ B \circ U \circ A^{-1} \circ \right. \\
&\quad \left. \circ \left(A \circ G^x \circ B^{-1} \circ B \circ G \circ B^{-1} \circ B \circ U \circ A^{-1}\right)^{s^{-1}(t-s)}\right)^s = \\
&= A \circ G^{xs} \circ G^{es^2} \circ U^s \circ G^{x(t-s)} \circ G^{t-s} \circ U^{t-s} \circ A^{-1} = \\
&= A \circ G^{es^2-s+xt+t} \circ U^t \circ A^{-1} = \\
&= A \circ G^k \circ U^t \circ A^{-1} = V \quad \Rightarrow \quad V' = V \;\Rightarrow\; e' = e.
\end{aligned}$$

6 Discussion

In the proposed signature scheme, the idea [11] of using a hidden commutative group possessing 2-dimensional cyclicity is exploited. However, the used method for implementing the advanced criterion of post-quantum security is different. Due to a novel design the signature size is reduced significantly and performance of the signature verification

procedure is improved. The used 4-dimensional FNAA set by a sparse BVMT provides about a twofold (and fourfold) increase in performance of the signature scheme in comparison with [11] (and [12]).

Like in the signature scheme [17], in the introduced one generation of the public key is performed using two different types of operations masking the hidden commutative group. The first type relates to the automorphism map operation possessing the property of mutual commutativity with the exponentiation operation (see the formula for computing the element Y of the public key). The second type relates to a map operation that is free of the property of mutual commutativity with the exponentiation operation (see formulas for computing the elements W and Z of the public key). However, masking operations of the second type are not arbitrary. Their parameters are chosen taking into account the fact that their compositions form a composite operation that has the said property of commutativity. An advantage of the introduced signature scheme against [17] is elimenation of doubling of both the verification equation and the public key.

An advantage of the introduced signature scheme is the use of the algebraic support for which one can evidently demonstrate the existence of the sufficiently large number of different commutative groups having 2-dimensional cyclicity (see formula (17)).

Table 3, where a procedure execution time* is estimated in multiplications in $GF(p)$, presents a rough comparison of the proposed signature scheme with the introduced earlier in [11], [12], [17]) ones (for the case of using a 257-bit prime).

To illustrate fulfillment of Criterion 1, consider the construction of some periodic functions based on public parameters of the proposed signature algorithm.

1. Suppose the function

$$F_1(i,j) = (W \circ Y \circ Z)^i \circ (W \circ Z)^j = A \circ G^{xi+i+j} \circ U^{i+j} \circ A^{-1}$$

includes a period with the length (δ_i, δ_j). Then, taking into account that G and U are generators of different cyclic groups of the same order q, we have $x\delta_i + \delta_i + \delta_j \equiv 0 \bmod q$ and $\delta_i + \delta_j \equiv 0 \bmod q$. From these two congruencies one gets: $x\delta_i \equiv 0 \bmod q \Rightarrow \delta_i \equiv \delta_j \equiv 0 \bmod q$. The last

Table 3. Comparison of the proposed and known signature schemes

Signature scheme	signature size, bytes	public-key size, bytes	sign. gener. time*	sign. verific. time*
[11]	192	768	18,432	24,576
[12]	256	1158	41,472	55,296
[17]	192	768	30,720	24,576
Proposed	96	384	12,300	9,200

means the function $F_1(i,j)$ possesses only the periodicity connected with the value q that is order of cyclic groups contained in the hidden commutative group with 2-dimensional cyclicity.

2. Suppose the function

$$F_2(i,j) = (Z \circ W)^i \circ Y^j = B \circ U^i \circ G^{xi} \circ G^j \circ B^{-1}$$

includes a period with the length (δ_i, δ_j). Then, we have $\delta_i \equiv 0 \bmod q$ and $x\delta_i + \delta_j \equiv 0 \bmod q \Rightarrow \delta_i \equiv \delta_j \equiv 0 \bmod q$, i. e., the function $F_2(i,j)$ also possesses only the periodicity connected with the value q.

7 Conclusion

An alternative design of the HDLP-based signature schemes satisfying the advanced criterion of post-quantum security is implemented in the introduced signature scheme. A new 4-dimensional FNAA set by a sparse BVMT is used as the algebraic carrier. For the first time, the types of the commutative groups contained in the algebraic carrier have been studied, and formulas for computing the number of the groups of every type have been obtained. To study the properties of the used FNAA, a method of computing sets of pairwise permutable vectors has been applied. This method is attractive to study properties of FNAAs of other types and dimensions.

The proposed signature scheme seems to be quite competitive with known candidates for post-quantum signature schemes. At the same time, one can suppose that other efficient designs for implementing Criterion 2 are possible.

Acknowledgement. *This work was partially supported by RSF and by the budget theme No. 0060-2019-010.*

References

[1] *Post-Quantum Cryptography. 9th International Conference, PQCrypto 2018, Fort Lauderdale, FL, USA, April 9-11, 2018, Proceedings* (Lecture Notes in Computer Science, vol. 10786, Security and Cryptology), Tanja Lange, Rainer Steinwandt, Eds. Springer International Publishing, 2018. ISBN: 978-3-319-79062-6.

[2] *Post-Quantum Cryptography. 10th International Conference, PQCrypto 2019, Chongqing, China, May 8–10, 2019 Revised Selected Papers*, (Lecture Notes in Computer Science, vol. 11505, Security and Cryptology), Jintai Ding, Rainer Steinwandt, Eds. Springer International Publishing, 2019. ISBN: 978-3-030-25509-1.

[3] P. W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on quantum computer," *SIAM Journal of Computing*, vol. 26, pp. 1484–1509, 1997.

[4] A. Ekert and R. Jozsa, "Quantum computation and Shorś factoring algorithm," *Rev. Mod. Phys.*, vol. 68, p. 733, 1996.

[5] R. Jozsa, "Quantum algorithms and the fourier transform," *Proc. Roy. Soc. London Ser A*, vol. 454, pp. 323 – 337, 1998.

[6] D.N. Moldovyan, N. A. Moldovyan, and A. A. Moldovyan, "Commutative Encryption Method Based on Hidden Logarithm Problem," *Bulletin of the South Ural State University. Ser. Mathematical Modelling, Programming & Computer Software*, vol. 13, no. 2, pp. 54–68, 2020. DOI: 10.14529/mmp200205.

[7] D. N. Moldovyan, "Post-quantum public key-agreement scheme based on a new form of the hidden logarithm problem," *Computer Science Journal of Moldova*, vol. 27, no. 1(79), pp. 56–72, 2019.

[8] N. A. Moldovyan and A. A. Moldovyan, "Finite Non-commutative Associative Algebras as carriers of Hidden Discrete Logarithm Problem," *Bulletin of the South Ural State University. Ser. Mathematical Modelling, Programming & Computer Software*, vol. 12, no. 1, pp. 66–81, 2019. DOI: 10.14529/mmp190106.

[9] N. A. Moldovyan and I. K. Abrosimov, "Post-quantum electronic digital signature scheme based on the enhanced form of the hidden discrete logarithm problem," *Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes*, vol. 15, no. 2, pp. 212–220, 2019. https://doi.org/10.21638/11702/spbu10.2019.205 (in Russian).

[10] N. A. Moldovyan, "Finite Non-commutative Associative Algebras for Setting the Hidden Discrete Logarithm Problem and Post-quantum Cryptoschemes on Its Base," *Buletinul Academiei de Stiinte a Republicii Moldova. Matematica*, no. 2(93), pp. 62–67, 2020.

[11] D. N. Moldovyan, A. A. Moldovyan, and N. A. Moldovyan, "Digital signature scheme with doubled verification equation," *Computer Science Journal of Moldova*, vol. 28, no. 1(82), pp. 80–103, 2020.

[12] D. N. Moldovyan, A. A. Moldovyan, and N. A. Moldovyan, "An enhanced version of the hidden discrete logarithm problem and its algebraic support," *Quasigroups and Related Systems*, vol. 28, no. 2, pp. 269–284, 2020.

[13] A.A. Moldovyan and N.A. Moldovyan, "Post-quantum signature algorithms based on the hidden discrete logarithm problem," *Computer Science Journal of Moldova*, vol. 26, no. 3(78), pp. 301–313, 2018.

[14] N.A. Moldovyan and P.A. Moldovyanu, "New primitives for digital signature algorithms," *Quasigroups and Related Systems*, vol. 17, no. 2, pp. 271–282, 2009.

[15] N.A. Moldovyan, "Fast signatures based on non-cyclic finite groups," *Quasigroups and Related Systems*, vol. 18, no. 1, pp. 83–94, 2010.

[16] N. A. Moldovyan, "Unified Method for Defining Finite Associative Algebras of Arbitrary Even Dimensions," *Quasigroups and Related Systems*, vol. 26, no. 2, pp. 263–270, 2020.

[17] N. A. Moldovyan and A. A. Moldovyan, "Candidate for practical post-quantum signature scheme," *Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes*, vol. 16, no. 4, pp. 455–461, 2020. https://doi.org/10.21638/11701/spbu10.2020.410 (in English).

D. N. Moldovyan

Received December 28, 2020
Accepted June 14, 2021

St. Petersburg Federal Research Center of
the Russian Academy of Sciences (SPC RAS),
St. Petersburg Institute for Informatics and
Automation of the Russian Academy of Sciences
14 Liniya, 39, St.Petersburg, 199178
Russia
E–mail: mdn.spectr@mail.ru

Computer Science Journal of Moldova, vol.**29**, no.2(86), 2021

Cube-root-subgroups of SL_2 over imaginary quadratic integers*

Miroslav Kureš

Abstract

All cube roots of the identity in the special linear group of 2×2-matrices with entries in the ring of integers in $\mathbb{Q}[\sqrt{d}]$ are described. These matrices generate subgroups of the third order; it is shown that such subgroups may contain non-elementary matrices in the sense of P. M. Cohn. All this is viewed with respect to possible applications in lattice cryptography.

Keywords: Imaginary quadratic field, ring of integers, non-elementary matrices, special linear group, public-key cryptography, lattice-based cryptosystems.

MSC 2010: 13F07, 15A16, 11R11, 11T71.

1 Introduction

Some public-key cryptographic algorithms are believed to be secure against attacks by quantum computers. Lattice-based cryptography is believed to be one of them ([1]). Another system is e. g. multivariate public key cryptography (we discuss some of its features and possible improvements in the papers [3] and [4], where the multivariate system is called TTM, which means the Tame Transformation Method).

On the other hand, Shor's algorithm for quantum computers is designed to solve prime factorization of large primes and the discrete logarithm problem in polynomial time.

In the paper, we will have in mind – in the background – lattices represented by their generator matrices. We are concerned with design

*The research has been supported by Brno University of Technology, the specific research plan No. FSI-S-20-6187.

ideas based on linear algebra over certain domains. Regarding the submission of these ideas into a lattice-based cryptosystem, we can think, for instance, the GGH cryptosystem, which is perhaps the most intuitive encryption scheme based on lattices. We remark that the classical GGH cryptosystem has been subject to cryptanalytic attacks and should be considered insecure, but we are not particularly limited to this system, which we mention as an example only, and, moreover, we present a completely different algebraic situation in which no attacks have yet been proposed.

In some cryptographic protocols working over finite groups, we have to be careful that we do not fall victim to what is called a *small subgroup attack*. In order to counter this attack, the prime order groups are used, in which all elements are primitive and small subgroups do not exist ([8]).

So, our design is to consider lattice cryptography over imaginary quadratic integers, involving not only large-size matrices but also 2×2 matrices, then selecting non-elementary matrices that are difficult to generate. In this paper we note that there exist also inappropriate, small subgroups in this case. Our result can also be seen as a contribution to matrix theory.

2 Some facts about orders of imaginary quadratic fields and problem formulation

Let d be a negative square-free integer and C a positive integer. We will distinguish two cases:

(I) $d \equiv 1 \pmod 4$

(II) $d \equiv 2$ or $d \equiv 3 \pmod 4$

We set

$$\varepsilon = \begin{cases} 1 & \text{for the case (I)} \\ 0 & \text{for the case (II);} \end{cases}$$

and

$$\theta = \sqrt{d} + \frac{\varepsilon}{2}\left(1 - \sqrt{d}\right)$$

and

$$D = -d + \frac{\varepsilon}{4}(1 + 3d).$$

Further, we denote by $\mathbb{Z}[C\theta]$ an *order of the imaginary quadratic field* $\mathbb{Q}[\sqrt{d}]$, so

$$\mathbb{Z}[C\theta] = \{x + yC\theta; x, y \in \mathbb{Z}\}.$$

The order $\mathbb{Z}[C\theta]$ is a normed ring with the norm $|\ |\colon \mathbb{Z}[C\theta] \to \mathbb{R}^+$ equal to the complex numbers absolute value. Then for $z = x + yC\theta \in \mathbb{Z}[C\theta]$ we have

$$|z|^2 = x^2 + \varepsilon xyC + y^2C^2D.$$

All orders are domains. For $C = 1$, the order $\mathbb{Z}[C\theta]$ is called the *maximal order* or the *ring of integers* in $\mathbb{Q}[\sqrt{d}]$ (often denoted also by $\mathcal{O}_{Q[\sqrt{d}}$).

The maximal orders are principal ideal domains (PID's) if and only if d is one of the numbers

$$-1, -2, -3, -7, -11, -19, -43, -67, -163.$$

For the (more usual) norm defined as the square root of our $|\ |$, there does not exist any $x \in \mathbb{Z}[\theta]$ with this norm between 1 and 2 (or in our $|\ |$ between 1 and 4) which holds for every negative square-free d, excerpt when

$$d = -1, -2, -3, -7, -11$$

which are just the domains in which is a Euclidean algorithm for computing the greatest common divisor. All other $\mathbb{Z}[\theta]$ are non-Euclidean.

2.1 The fact that not all matrices are elementary

Let us consider $\mathrm{SL}(2, \mathbb{Z})$. By an *elementary transvection* $V_{ij}(r)$ $1 \leq i \neq j \leq m$, $r \in R$, we mean a matrix $[a_{\nu\mu}]$ with

$$a_{\nu\mu} = \begin{cases} 1_R & \text{for } \nu = \mu \\ r & \text{for } \nu = i, \mu = j \\ 0_R & \text{otherwise.} \end{cases}$$

Finite products of elementary transvections form a subgroup $\mathrm{SE}(2,\mathbb{Z})$ of *elementary matrices* in $\mathrm{SL}(2,\mathbb{Z})$. In other words, one can obtain elementary matrices by a finite sequence of multiplyings a row (or column) by a non-zero number r and adding the result to another row (or column), starting from the identity matrix. However, $\mathrm{SE}(2,\mathbb{Z}) = \mathrm{SL}(2,\mathbb{Z})$.

When we replace $\mathbb{Z}$ by another ring R, then it can happen that $\mathrm{SE}(2,R)$ is a proper subgroup of $\mathrm{SL}(2,R)$. This problem was studied by P. M. Cohn in 1966. He was the first to give an example of a matrix with determinant 1 which is not elementary. In the paper [5] (published in 2011), we introduced an algorithm how to detect such matrices for any order of imaginary quadratic field.

2.2 Some classical examples of non-elementary matrices

Here are some known examples of non-elementary matrices that have been studied before. By way of illustration, we also calculate their twelfth power: these matrices are of infinite order, so the norms of their entries are not bounded.

P. M. Cohn, [2], 1966: $d = -19$,

$$M_{\mathsf{Cohn}} = \begin{pmatrix} 3-\theta & 2+\theta \\ -3-2\,\theta & 5-2\,\theta \end{pmatrix}.$$

Then

$$M_{\mathsf{Cohn}}^{12} = \begin{pmatrix} -138533292392+7105318818\,\theta & 1003585011-60934202901\,\theta \\ -29430829974+110698224819\,\theta & -249231517211+23358797787\,\theta \end{pmatrix}.$$

R. Tuler, [9], 1983: $d = -37$,

$$M_{\mathsf{Tuler}} = \begin{pmatrix} 29 & 7-\theta \\ 7+\theta & 3 \end{pmatrix}.$$

We compute

$$M_{\mathsf{Tuler}}^{12} = \begin{pmatrix} 1033551428421627457 & 249747318595287744-35678188370755392\,\theta \\ 249747318595287744+35678188370755392\,\theta & 105918530781987265 \end{pmatrix}.$$

This paper shows that there are also small-order non-elementary matrices for which even the third power is the identity matrix. For example, for $d = -163$ all matrices

$$M_1 = \begin{pmatrix} 1 - 5\,\theta & 2 + 2\,\theta \\ 19 - 12\,\theta & -2 + 5\,\theta \end{pmatrix},$$

$$M_2 = \begin{pmatrix} 4 - \theta & 5 + 2\,\theta \\ 4 & -5 + \theta \end{pmatrix},$$

$$M_3 = \begin{pmatrix} 8 - \theta & 4 - 2\,\theta \\ -8 & -9 + \theta \end{pmatrix},$$

and

$$M_4 = \begin{pmatrix} -10 - 3\,\theta & 25 + 4\,\theta \\ -2 - 2\,\theta & 9 + 3\,\theta \end{pmatrix}$$

are examples of non-elementary matrices, the third power of which is I.

Theorem 2.7 of [6] allows us to calculate easily the possible orders of elements of $\mathrm{GL}\,(2, \mathbb{Z})$. The result is that this group can have subgroups of orders 2, 3, 4, and 6. We note that all these numbers are divisors of 12, whose power we have calculated above. [1]

2.3 The problem

We search for matrices M whose entries are integers in $\mathbb{Q}[\sqrt{d}]$ satisfying

$$M^3 = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

[1] When we have already mentioned divisors of the number 12, we will also remember the efforts of the duodecimalists, by recalling Limerick from the Duodecimal Bulletin 2015, No.1:
A base boasting reason and rhyme,
Sported factors full four at a time.
"Why the fourth and the third?"
"Cause just three is absurd;
And only two, sir, is a crime."

Our next requirement is that the determinant of the matrix M is equal to 1. So, we do not consider matrices like $\begin{pmatrix} 1 & 0 \\ 0 & \frac{-1+\sqrt{-3}}{2} \end{pmatrix}$; in other words, we are only in the group $\mathrm{SL}\,(2, \mathbb{Z}\,[\theta])$.

Of course, trivial solutions of the problem $M^3 = I$ are diagonal matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} \frac{-1+\sqrt{-3}}{2} & 0 \\ 0 & \frac{-1+\sqrt{-3}}{2} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \frac{-1-\sqrt{-3}}{2} & 0 \\ 0 & \frac{-1-\sqrt{-3}}{2} \end{pmatrix},$$

the last two only in the case of Eisenstein integers. We want to classify all other nontrivial matrices.

3 The result

Theorem 1. *Let d be a negative square-free integer. Then the complete classification of matrices of $\mathrm{SL}\,(2, \mathbb{Z}\,[\theta])$ representing nontrivial cube roots of identity is as follows.*

(i) For every $d \neq -3$, a matrix of $\mathrm{SL}\,(2, \mathbb{Z}\,[\theta])$ represents nontrivial cube roots of identity if and only if it is of a form

$$\begin{pmatrix} a + A\,\theta & b + B\,\theta \\ c + C\,\theta & -a - 1 - A\,\theta \end{pmatrix}, \tag{1}$$

where

$$bc + BC\left(d - \varepsilon\frac{3d+1}{4}\right) = -1 - a - a^2 - A^2\left(d - \varepsilon\frac{3d+1}{4}\right) \tag{2}$$

$$bC + Bc + \varepsilon BC = -A - 2aA - A^2, \tag{3}$$

$a, A, b, B, c, C \in \mathbb{Z}$.

(ii) For $d = -3$, a matrix of $\mathrm{SL}\,(2, \mathbb{Z}\,[\theta])$ is of the same form as in the case (i).

(iii) The cardinality of the intersection of a 3-element subgroup generated by a nontrivial cube root of identity with $\mathrm{SE}\,(2, \mathbb{Z}\,[\theta])$ can be either 1 or 3.

Proof. (i) Let us start with a matrix

$$M = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix},$$

where $\alpha, \beta, \gamma, \delta \in \mathbb{C}$. One can easily deduce that requirement $M^3 = I$ implies the following system of equations:

$$\begin{aligned} \alpha^3 + 2\alpha\beta\gamma + \beta\gamma\delta &= 1; && (4)\\ \beta\left(\alpha^2 + \beta\gamma + \alpha\delta + \delta^2\right) &= 0; && (5)\\ \gamma\left(\alpha^2 + \beta\gamma + \alpha\delta + \delta^2\right) &= 0; && (6)\\ \alpha\beta\gamma + 2\beta\gamma\delta + \delta^3 &= 1. && (7) \end{aligned}$$

Equations (5) and (6) are directed to three variants.

Variant I. Let $\beta = 0$. Then $\alpha^3 = 1$ and $\delta^3 = 1$. Let us denote by $\sigma = \frac{-1+\sqrt{3}}{2}$ and by $\bar{\sigma}$ its complex conjugate. Then the set of complex cube roots of 1 is $S = \{1, \sigma, \bar{\sigma}\}$. Then $\alpha = \sigma_1$, $\delta = \sigma_2$, where σ_1 and σ_2 are some elements from S. Now, it follows from the equation (6) that if $\sigma_1 = \sigma_2$, then $\gamma = 0$, and if $\sigma_1 \neq \sigma_2$, then γ can be taken completely arbitrarily. The conclusion of this variant is therefore

$$\begin{pmatrix} \sigma_1 & 0 \\ \gamma & \sigma_2 \end{pmatrix} \qquad \text{with } \sigma_1,\ \sigma_2,\ \gamma \text{ as mentioned.} \qquad (8)$$

Variant II. Let $\gamma = 0$. Analogous to the previous Variant I. The conclusion is

$$\begin{pmatrix} \sigma_1 & \beta \\ 0 & \sigma_2 \end{pmatrix} \qquad \text{with } \sigma_1,\ \sigma_2,\ \beta \text{ similarly as above.} \qquad (9)$$

Variant III. Let $\beta \neq 0 \wedge \gamma \neq 0$. Then

$$\beta\gamma = -\alpha^2 - \alpha\delta - \delta^2.$$

If we substitute this into (4) or (7), we obtain equally

$$(\alpha + \delta)^3 = -1. \qquad (10)$$

Let $\widehat{S} = \{-1, -\sigma, -\bar{\sigma}\}$ and let $\widehat{\sigma}_0$ be an element from $\widehat{S}$ and $\bar{\widehat{\sigma}}_0$ its complex conjugate. Then

$$\beta\gamma = -\alpha^2 - \widehat{\sigma}_0 + \bar{\widehat{\sigma}}_0. \tag{11}$$

In conclusion, we have

$$\begin{pmatrix} \alpha & \beta \\ \gamma & -\alpha + \widehat{\sigma}_0 \end{pmatrix}, \qquad \text{where } \widehat{\sigma}_0, \beta \text{ and } \gamma \text{ satisfy what has been just described.}$$

Non-Eisenstein integers ($d \neq -3$) do not contain elements σ, $-\sigma$, $\bar{\sigma}$ and $-\bar{\sigma}$. This brings a much simpler situation. In the *Variant I* as in the *Variant II*, we obtain only the identity matrix I. The *Variant III* reads as

$$\begin{pmatrix} \alpha & \beta \\ \gamma & -\alpha - 1 \end{pmatrix}, \qquad \text{where } \beta\gamma = -\alpha^2 - \alpha - 1.$$

For a precise description of the condition $\beta\gamma = -\alpha^2 - \alpha - 1$, we write down complex numbers into components:

$$\alpha = a + A\theta, \;\; \beta = b + B\theta, \;\; \gamma = c + C\theta.$$

We derive from $\theta = \sqrt{d} + \frac{\varepsilon}{2}\left(1 - \sqrt{d}\right)$ that $\theta^2 = d + \varepsilon\left(\theta - \frac{3d+1}{4}\right)$ and compute directly (2) and (3).

We remark that for given $a, A \in \mathbb{Z}$, we can denote $A_1 = -1 - a - a^2 - A^2\left(d - \varepsilon\frac{3d+1}{4}\right)$, $A_2 = -A - 2aA - A^2$ (the right hand sides of (2) and (3)) and observe that there are many integer solutions for b, B, c and C, for example, $b = A_1$, $B = A_2$, $c = 1$, $C = 0$.

(ii) For $d = -3$, $\theta = \sqrt{-3} + \frac{1}{2}\left(1 - \sqrt{-3}\right) = \frac{1+\sqrt{-3}}{2}$. Then

$$S = \left\{1, -\theta, -\bar{\theta}\right\} \quad \text{and} \quad \widehat{S} = \left\{-1, \theta, \bar{\theta}\right\}.$$

The *Variant I* and *Variant II*, namely (8) and (9), lead to three trivial solutions, I, $\begin{pmatrix} -\theta & 0 \\ 0 & -\theta \end{pmatrix}$ and $\begin{pmatrix} -\bar{\theta} & 0 \\ 0 & -\bar{\theta} \end{pmatrix}$ and for solutions described in the special case.

In the *Variant III*, we have to add to equations (10) and (11) the additional equation, the requirement that the determinant of M is equal to 1. However,

$$\det M = \alpha\delta - \beta\gamma = \alpha\left(\widehat{\sigma}_0 - \alpha\right) - \left(-\alpha^2 + \alpha\widehat{\sigma}_0 + \bar{\widehat{\sigma}}_0\right) = -\bar{\widehat{\sigma}}_0.$$

Nevertheless, $-\bar{\widehat{\sigma}}_0 = 1$ means that $\bar{\widehat{\sigma}}_0 = -1$ and $\widehat{\sigma}_0 = -1$. Therefore, we have no equations other than those already derived in (iA).

(iii) We have a group with 3 elements

$$M = \begin{pmatrix} \alpha & \beta \\ \gamma & -\alpha - 1 \end{pmatrix}, \quad M^2 = M^{-1} = \begin{pmatrix} -\alpha - 1 & -\beta \\ -\gamma & \alpha \end{pmatrix},$$
$$M^3 = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

with $\beta\gamma = -\alpha^2 - \alpha - 1$. Of course, M^3 is elementary and M is elementary if and only if M^2 is elementary as elementary matrices form the group $\mathrm{SE}\left(2, \mathbb{Z}\left[\theta\right]\right)$. It remains to show that M can be both elementary and non-elementary. This can be demonstrated already for $d = -5$. Suitable matrices are, for example,

$$M_{\mathcal{E}} = \begin{pmatrix} 1 + \sqrt{-5} & -2 + 3\sqrt{-5} \\ -1 & -2 - \sqrt{-5} \end{pmatrix}$$

which is elementary and

$$M_{\mathcal{N}} = \begin{pmatrix} 2 - 7\sqrt{-5} & 1 + 2\sqrt{-5} \\ 28 - 21\sqrt{-5} & -3 + 7\sqrt{-5} \end{pmatrix}$$

which is non-elementary.

We present elementary transvections as elementary row opera-

tions. We observe

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \xrightarrow{\mathrm{r}_1=\mathrm{r}_1+\left(3+\sqrt{-5}\right)\mathrm{r}_2} \begin{pmatrix} 1 & 3+\sqrt{-5} \\ 0 & 1 \end{pmatrix}$$
$$\xrightarrow{\mathrm{r}_2=\mathrm{r}_2-\mathrm{r}_1} \begin{pmatrix} 1 & 3+\sqrt{-5} \\ -1 & -2-\sqrt{-5} \end{pmatrix}$$
$$\xrightarrow{\mathrm{r}_1=\mathrm{r}_1-\sqrt{-5}\mathrm{r}_2} \begin{pmatrix} 1+\sqrt{-5} & -2+3\sqrt{-5} \\ -1 & -2-\sqrt{-5} \end{pmatrix} = M_{\mathcal{E}}.$$

and that is why the matrix $M_{\mathcal{E}}$ is elementary.

For the proof of the non-elementarity of $M_{\mathcal{N}}$, we use approach presented in [5]. The $(1,2)$-submatrix of $M_{\mathcal{N}}$ is

$$A = \begin{pmatrix} 2-7\sqrt{-5} & 1+2\sqrt{-5} \end{pmatrix},$$

and the reduction ellipse $\mathcal{E}_{\mathrm{red}}$ has the equation

$$x^2 + 5y^2 + \frac{136}{21}x + \frac{110}{21}y + \frac{76}{7} = 0$$

(for the procedure of the computation see [5]). We can verify (or observe on the Figure 1) that there are no interior lattice points of the reduction ellipse with integer coordinates. Hence there are no reduction elements for A, and therefore $M_{\mathcal{N}}$ is non-elementary.

□

Remark 1. It is easy to search for *square roots* from the identity. But there are no non-trivial solutions in $\mathrm{SL}\,(2,\mathbb{Z})$, which the reader can verify. Another problem would be to look for the *fourth roots*, where the same procedure as in our paper could be used.

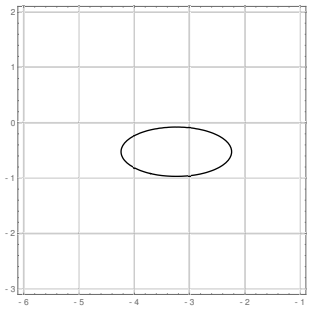


Figure 1. The reduction ellipse $\mathcal{E}_{\mathrm{red}}$ for the matrix A possessing no interior lattice point.

4 The Goldreich–Goldwasser–Halevi cryptosystem (GGH)

The private key is a generator matrix Z of a complete lattice $\mathcal{L}$ with good properties such as short and nearly orthogonal generators together with a unimodular matrix M.

The public key is another ("bad") generator matrix Y of $\mathcal{L}$ obtained as $Y = MZ$.

Given an open message expressed as a vector $\mathbf{m}$ and a random (so-called error) vector $\mathbf{e}$ which has to be small, we cipher by $\mathbf{c} = \mathbf{m} \cdot Y + \mathbf{e}$.

To decrypt $\mathbf{c}$, first compute $\mathbf{d} = \mathbf{c} \cdot Z^{-1}$ and apply the rounding-off which provides $\lfloor \mathbf{d} \rceil$. Finally, we compute $\mathbf{m}$ as $\lfloor \mathbf{d} \rceil \cdot M^{-1}$.

4.1 GGH cryptosystem – introductory example over Gaussian integers

Let $d = -1$, then $\theta = \sqrt{-1}$. We will present only a very basic example to remind the reader of the principles of the GGH system.

Let the generator matrix of the complete 3-lattice $\mathcal{L}$ be

$$Z = \begin{pmatrix} 12+\theta & 0 & 1+2\,\theta \\ 1 & 15+\theta & 1-\theta \\ 4-\theta & 3-3\,\theta & 11-19\,\theta \end{pmatrix},$$

and the unimodular matrix

$$M = \begin{pmatrix} 8+3\,\theta & -2+20\,\theta & 3+2\,\theta \\ 33-3\,\theta & 39+8\,\theta & 4-6\,\theta \\ 2+4\,\theta & 2+5\,\theta & 1 \end{pmatrix}$$

($\det M = 1$). Then the "bad" generator matrix of $\mathcal{L}$ is

$$Y = M \cdot Z = \begin{pmatrix} 105+69\,\theta & -35+295\,\theta & 91+6\,\theta \\ 448-23\,\theta & 571+129\,\theta & 16-110\,\theta \\ 26+54\,\theta & 28+74\,\theta & 12-8\,\theta \end{pmatrix}.$$

Now, for the open message $\mathbf{m} = (50+\theta, 11-\theta, 34+15\,\theta)$ and the error vector $\mathbf{e} = (1+\theta, 2-\theta, 3-2\,\theta)$, we compute the ciphered message as

$$\mathbf{c} = \mathbf{m} \cdot Y + \mathbf{e} = (10161+5081\,\theta, 4209+18498\,\theta, 5141-929\,\theta)\,.$$

The decryption: first, we compute $\mathbf{d} = \mathbf{c} \cdot Z^{-1}$ as

$$\left(\frac{2989188391}{3907301} + \frac{1008368124}{3907301}\theta, \frac{2423186121}{7814602} + \frac{4872178230}{3907301}\theta, \frac{860187474}{3907301} + \frac{375699873}{7814602}\theta\right)$$

which we round-off to

$$\lfloor \mathbf{d} \rceil = (765+258\,\theta, 310+1247\,\theta, 220+48\,\theta)\,.$$

Finally, $\mathbf{m}$ is recovered with

$$\mathbf{m} = \lfloor \mathbf{d} \rceil \cdot M^{-1} = (50+\theta, 11-\theta, 34+15\,\theta)\,.$$

4.2 GGH cryptosystem – a use of non-elementary matrices, where $d = -163$

For example, we can use a non-elementary unimodular matrix $M_{\mathcal{U}} = (M_1 \cdot M_2 \cdot M_3 \cdot M_4)^3$ as

$$\begin{pmatrix} 52573221941851385 + 7734170153866877\,\theta & -108339682589377105 - 8450168415298437\,\theta \\ 97004133663118053 + 22955204128735434\,\theta & -224828457848433395 - 29213244459035477\,\theta \end{pmatrix}.$$

We can use it for construction of a "bad" generator matrix. Then, a finding of the inverse matrix is complicated not only by that we have

used an "innovation" ring of integers with completely different multiplication compared to $\mathbb{Z}$ but also by the fact that MZ can also be non-elementary and attempting to break such an element of a cryptosystem precludes, for example, a use of Gauss-Jordan elimination for the inverse.

5 Conclusion

Computer science develops cryptographic protocols and judges how secure certain protocols are. The purpose of cryptography is to prevent other parties from accessing information they should not access. In our digitized world, these questions have considerable importance as communicating sides do not meet directly and use asymmetric cryptographic protocols. Modern public-key cryptography uses advanced algebraic methods. The paper deals with questions coinciding with the so-called small subgroup attack. There are classified all cube roots of the identity in the special linear group of second-order matrices with entries in the ring of integers in imaginary quadratic fields. The result may be essential for designing new protocols, e.g. in lattice-based cryptosystems.

Acknowledgements

The research has been supported by Brno University of Technology, the specific research plan being No. FSI-S-20-6187.

References

[1] *Post-Quantum Cryptography*, D. J. Bernstein, J. Buchmann, and E. Dahmen, Eds. Springer-Verlag Berlin Heidelberg, 2009, 256 p. DOI: 10.1007/978-3-540-88702-7. Hardcover ISBN: 978-3-540-88701-0.

[2] P. M. Cohn, "On the structure of GL_2 of a ring," *Publications Mathématiques de l'I.H.É.S.*, vol. 30, pp. 5–53, 1966. DOI: https://doi.org/10.1007/BF02684355.

[3] J. Hrdina, M. Kureš, and P. Vašík, "A note on tame polynomial automorphisms and the security of TTM cryptosystem," *Applied and Computational Mathematics*, vol. 9, no.2, pp. 226–233, 2010.

[4] M. Kureš, T. Decome, and G. Drecourt, "LoRi-TTM cryptosystem," *Annals of the University of Craiova, Mathematics and Computer Science Series*, vol. 45, no. 1, pp. 137–150, 2018.

[5] M. Kureš and L. Skula, "Reduction of matrices over orders of imaginary quadratic fields," *Linear Algebra and Its Applications*, vol. 435, no. 8, pp. 1903–1919, 2011.

[6] J. Kuzmanovich and A. Pavlichenkov, "Finite groups of matrices whose entries are integers," *The American Mathematical Monthly*, vol. 109, no. 2, pp. 173–186, 2002.

[7] B. Nica, "The unreasonable slightness of E_2 over imaginary quadratic rings," *The American Mathematical Monthly*, vol. 118, no. 5, pp. 455–462, 2011.

[8] C. Paar and J. Pelzl, *Understanding cryptography: a textbook for students and practitioners*, Springer, 2009, 390 p. ISBN: 978-3-642-04100-6. DOI: 10.1007/978-3-642-04101-3.

[9] R. Tuler, "Detecting products of elementary matrices in $GL_2(\mathbb{Z}[\sqrt{d}])$," *Proceedings of the American Mathematical Society*, vol. 89, no. 1, pp. 45–48, 1983.

Miroslav Kureš

Received May 3, 2021
Accepted July 18, 2021

Miroslav Kureš
Brno University of Technology
Technická 2, 61669 Brno, Czechia
Phone: +420 541142714
E–mail: kures@fme.vutbr.cz

Computer Science Journal of Moldova, vol.**29**, no.2(86), 2021

BeyondBenford R Package to compare Benford's and BDS's Distributions

Stéphane Blondeau Da Silva

Abstract

The package **BeyondBenford** compares the goodness of fit of Benford's and Blondeau Da Silva's (BDS's) digit distributions in a given dataset. It first enables to check whether the data distribution is consistent with theoretical distributions highlighted by Blondeau Da Silva or not; indeed, this ideal theoretical distribution must be at least approximately followed by the data for the use of BDS's model to be well-founded. It also allows to draw histograms of digit frequencies or probabilities (and their confidence or prediction intervals), both observed in the dataset and given by the two theoretical approaches. Finally, it proposes to quantify the goodness of fit of these laws via Pearson's chi-squared tests.

Keywords: Benford's law, digits, experimental data, R-package.

MSC 2010: 60E05.

1 Introduction

Benford's Law, also called Newcomb-Benford's Law, is somewhat surprising; indeed, the first digit d, $d \in [\![1, 9]\!]$, of numbers in many naturally occurring collections of data does not follow a discrete uniform distribution, as might be thought, but a logarithmic distribution (see the recent books of [1] and [2]). Discovered by the astronomer Newcomb in 1881 [3], it was definitively brought to light by the physicist Benford in 1938 [4]. The probability that d is the first digit of a number

is approximately:

$$\log(1+\frac{1}{d}) \ .$$

It can also be extended to digits beyond the first one [5], the probability for d, $d \in [\![0,9]\!]$, to be the p^{th} digit of a number being:

$$\sum_{j=10^{p-2}}^{10^{p-1}-1} \log(1+\frac{1}{10j+d}) \ .$$

It was quickly admitted that numerous empirical data sets follow Benford's law: economic data [6], social data [7], demographic data [8], [9], physical data [10], [11] or biological data [12], [13] for instance; to such an extent that this law was used to detect possible frauds in lists of socio-economic data [14], [15] or in scientific publications [16].

Nevertheless many discordant voices brought a significantly different message. By putting aside the distributions known to fully disobey Benford's law [17]–[19], this law often appeared to be a good approximation of the reality, but no more than an approximation [20]–[22].

Similar to first digit case, the distributions of digits beyond the first have been observed in various application areas [11], [16], [23] and have also been used to detect frauds [24]–[26]. Once more, limits of such methods were emphasized [24], [25], [27].

Blondeau Da Silva, considering data as realizations of a homogeneous and expanded range of random variables following discrete uniform distributions, showed that, the proportion of each d, $d \in [\![0,9]\!]$, as leading digit [28] or as other digit [29] structurally fluctuates. He demonstrated that, in his models, the predominance of 1 as digit (followed by 2 and so on) is all but surprising, and that the observed fluctuations around the values of probability determined by Benford's Law are also predictible: there is not a single Benford's Law but numerous distinct laws each of them determined by a parameter, the upper-bound of the considered data.

The huge and growing literature on Benford's law is available on the online database `www.benfordonline.net` with well over 1000 papers.

Two R packages [30] already enable to check whether datasets conform to Benford's law or not: **BenfordTests** [31] and **benford.analysis** [32]. The package **BeyondBenford** [33] compares the goodness of fit of Benford's law, on the one hand, and BDS's laws, on the other hand. Indeed, these latter, under certain conditions that we will recall, allow a better reliability of adjustment. The package **BeyondBenford** calculates the digit distribution in the considered dataset and determines whether it is consistent with BDS's or Benford's one. It also provides plotting tools for the visual evaluation of these distributions. We will walk through a detailed example to give an overview of the **BeyondBenford** package.

2 An example to get familiar with the main functions

Street addresses of Pierre-Buffiere, a small town of approximately 1200 inhabitants in Haute-Vienne (France), are available on:

`www.data.gouv.fr/fr/datasets/base-d-adresses-nationale-ouverte-bano/` ,

which is an open platform for French public data.

After loading the package (`library(BeyondBenford)`), the code `data(address_PierreBuffiere)` gives us access to the sample data. This factor contains 346 rows, with each row representing an address number.

3 Is the data consistent with BDS's model?

This is the first essential question that must be answered in the affirmative. If this is not the case, comparisons are not relevant: the package should not be used. Indeed the use of the package is appropriate when the studied data can be considered as realizations of a homogeneous and expanded range of random variables approximately following discrete uniform distributions. In this model, the data is strictly positive and is lower and upper-bounded, constraint which is often valid in datasets,

the physical, biological, demographic, social and economical quantities being limited [28].

Among the different domains studied by Benford [4], some could be well adapted to our model: sizes of populations or street addresses for example (see [28] for a detailed explanation). [34] advised precisely to use their own similar model in the case of street addresses or when considering the first-page numbers of articles in a bibliography.

[29] showed that the model induces a specific distribution of positive integers determined by a lower and an upper-bound. Hence, in order to conform as closely as possible to the model, the studied database must have a distribution similar to that described in [29]. In p^{th} digit case, the probability p_k to obtain the number $k \in [\![l_b; u_b]\!]$ (where l_b is the lower-bound and u_b is the upper-bound) verifies:

$$p_k = \frac{1}{u_b - l_b + 1} \sum_{i=k-l_b+1}^{u_b-l_b+1} \frac{1}{i} \ .$$

In the studied example, the minimum value of the street number is 1 and the maximum value is 74. The associated theoretical distribution for the second digit is plotted in Figure 1.

The **BeyondBenford** package provides plotting tools to determine whether the data is consistent with BDS's model: the function `dat.distr`. The function's main arguments are as follows:

- **dat**: the considered dataset, a data frame containing non-zero real numbers.
- **theor**: if `theor=TRUE` BDS's theoretical distribution is plotted, otherwise only the histogram is represented. Default value: `theor=TRUE`.
- **nclass**: a strictly positive integer: the number of classes in the histogram. Default value: `nclass=50`.
- **conv**: if `conv=1`, all values of the dataset are multiplied by 10^k where k is the smallest positive integer such that all non-zero

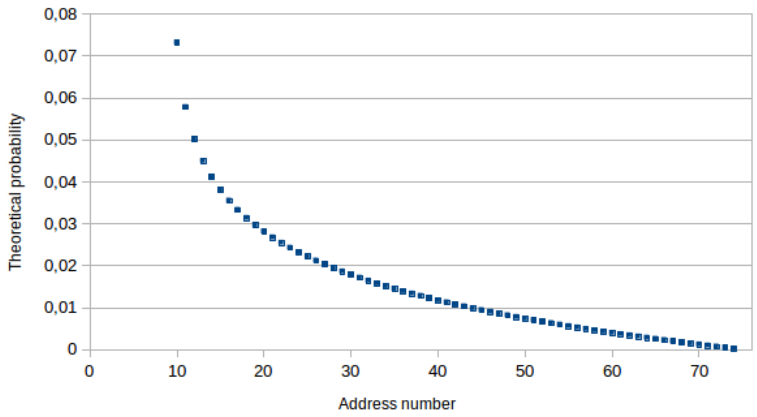


Figure 1. Theoretical distribution of street numbers in the second digit case, the upper-bound being 74.

numerical values in the newly multiplied data frame have an absolute value greater than or equal to 1. Default value: `conv=0`.

- **lwbound**: a positive integer, which characterizes the data. All (or most) of the data are greater than this "lower-bound". Default value: `lwbound=max(floor(min(abs(dat)))+1,(10**(dig -1)))`.

- **upbound**: a positive integer, which characterizes the data. All (or most) of the data are lower than this "upper-bound". Default value: `upbound=ceiling(max(dat))`.

- **dig**: the chosen position of the digit (from the left). Default value: `dig=1`.

- **nchi**: the number of classes for values from 10^{p-1} to max(max(dat a),upbound). If `nchi>0`, the function returns the chi-squared statistic (with $nchi - 1$ degrees of freedom) of goodness of fit determined by the different classes. The null hypothesis states

that the studied distribution is consistent with the considered theoretical distribution. Default value: `nchi=0`.

Let us apply the `dat.distr` function to the `address_PierreBuffiere` dataset. The output from `dat.distr` is the data histogram along with optional BDS's theoretical distributions (Figure 2).

It can be noted that there are also other arguments for the formatting of histograms.

Example 1. *## Both the histogram and theoretical distribution are represented*

```
dat.distr(address_PierreBuffiere, dig=2, nclass=65)
```

Note that, out of the 346 values, only 217 are taken into consideration here because the values need to have at least 2 digits.

The data distribution looks similar to the one described by BDS's model (Figure 2). Let us provide a second example in which we numerically determine whether the studied distribution is conform to the theoretical distribution or not:

Example 2. *## The function returns the chi-squared statistic of goodness of fit determined by nchi classes.*

```
dat.distr(address_PierreBuffiere, dig=2, nchi=4)
[1] "Class freq.:" "130" "51" "25"
[5] "11"
[1] "Theor.  freq.:" "127.0509775044 73" "54.6446927640259"
"26.7009825219254"
[5] "8.60334720957567"
            chi2             pval
1   Chi2 value is:   The p-value is:
2 1.08754607580421 0.780081329402347
```

The `dat.distr` function returns, if requested, the frequencies of each equal-sized class of the dataset and the associated theoretical frequencies. It also returns a data frame containing the value of the chi squared test-statistic and its p-value. Note that the number of classes is limited by the theoretical frequencies that cannot exceed 5 in Pearson's chi-squared test [35]. In our example, the null hypothesis cannot

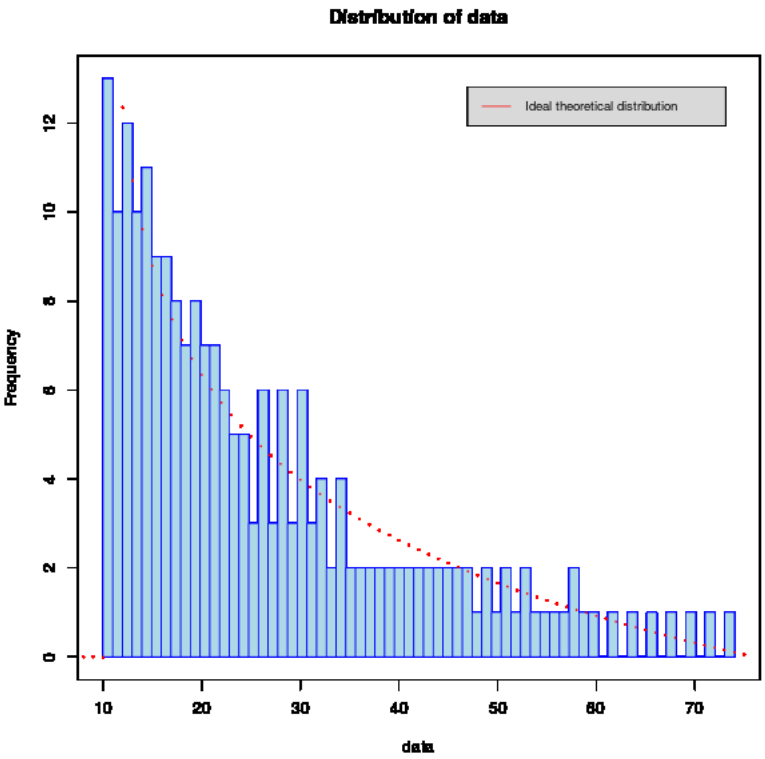


Figure 2. The data histogram of street numbers in the second digit case and its BDS's theoretical distribution.

be rejected: the studied distribution is consistent with the theoretical distribution.

4 Comparisons of the goodness of fit of Benford's and BDS's digit distributions in a dataset

4.1 Raw data

The package **BeyondBenford** provides a function returning the frequencies of each figure at a given position in the considered dataset: `obs.numb.dig`. Its two arguments **dat** and **dig** have already been defined in `dat.distr` function. The function output is a vector containing the frequencies of each figure in ascending order.

Let us give an example:

Example 3.
```
obs.numb.dig(address_PierreBuffiere, dig=2)
[1] 31 24 27 21 25 17 21 16 19 16
```

For instance, there are 25 values with the second digit being a 4.

The package **BeyondBenford** also provides a function returning Benford's probability that a figure is at a given position: `Benf.val`. Its argument **dig** has already been defined in `dat.distr` function. Its second argument **fig** stands for the considered figure.

Let us give an example:

Example 4.
```
Benf.val(4, dig=2)
[1] 0.1003082
```

The package **BeyondBenford** at last provides a function returning BDS's probability that a figure is at a given position (once the associated upper-bound has been specified): `Blon.val`. Its four arguments – **dig**, **lwbound**, **upbound**, and **fig** – have already been defined above.

Let us give an example:

Example 5.
```
Blon.val(fig=4, dig=2, upbound=74)
[1] 0.09836642
```

In Table 1 below, the frequencies of each figure in the considered dataset, regarding the second digit of address numbers, are listed.

Table 1. Values of frequency of each figure as second digit in the database, Benford's and BDS's theoretical values (the chosen upper-bound being 74). These values are rounded to the nearest ten-thousandth.

Figure	Freq. in the database	BDS's values	Benford's values
0	0.1429	0.1436	0.1197
1	0.1106	0.1247	0.1139
2	0.1244	0.1136	0.1088
3	0.0968	0.1053	0.1043
4	0.1152	0.0984	0.1003
5	0.0783	0.0924	0.0967
6	0.0968	0.0872	0.0934
7	0.0737	0.0825	0.0904
8	0.0876	0.0782	0.0876
9	0.0737	0.0742	0.0850

The BDS's theoretical values seem slightly better; in particular the frequency range is higher, both for the observed data and for BDS's theoretical values.

4.2 Plotting tools

The **BeyondBenford** package provides plotting tools to perform the comparison between the two models with the function `digit.distr`. In addition to arguments that are shared with `dat.distr` (**dat**, **dig**, **lwbound**, **upbound**), the `digit.distr` function has mainly the following additional arguments:

- **mod**: if `mod="ben"`, the data histogram and that of Benford are displayed, if `mod="ben&blo"`, the data histogram, that of Benford and that of BDS are plotted, and otherwise the data histogram and that of BDS are given. Default value: `mod="ben"`.

- **No.sd**: the positive decimal number of standard deviation that defines the confidence and prediction intervals i.e. the error bars. If No.sd=0, no error bars are drawn. Default value: `No.sd=0`.

- **Sd.pr**: If Sd.pr=1, error bars for proportions are plotted (with No.sd standard deviation confidence intervals). If Sd.pr=0, they are not plotted. Default value: `Sd.pr=0`.

Let us apply the `digit.distr` function to the `address_PierreBuffiere` dataset. The output from `dat.distr` is a histogram of theoretical and experimental digit distribution (Figure 3).

There are also other arguments for the formatting of histograms.

Example 6. *`digit.distr(address_PierreBuffiere, dig=2, mod="ben&blo", No.sd=1, Sd.pr=1)`*

Naturally, Figure 3 is consistent with Table 1.

4.3 Pearson's chi-squared test

To quantify the quality of theorical models, we use Pearson's chi-squared test of goodness of fit [35]: the null hypothesis states that the studied distribution is consistent with the considered theoretical distribution, *i.e.* Benford's or BDS's ones. The function `chi2` determines the test statistic and its associated p-value. In addition to arguments that are shared with `dat.distr` (**dat**, **dig**, **lwbound**, **upbound**), the `chi2` function has the following specific arguments:

- **mod**: if mod="ben", the theorical distribution considered is that of Benford, else it is BDS's ones which is chosen. Default value: `mod="ben"`.

- **pval**: if pval=0, the p-value is not returned, else it is available. Default value: `pval=0`.

Let us apply this function to the `address_PierreBuffiere` dataset. The output from `chi2` is a data frame containing the Pearson's chi-squared statistic (and the associated p-value if requested).

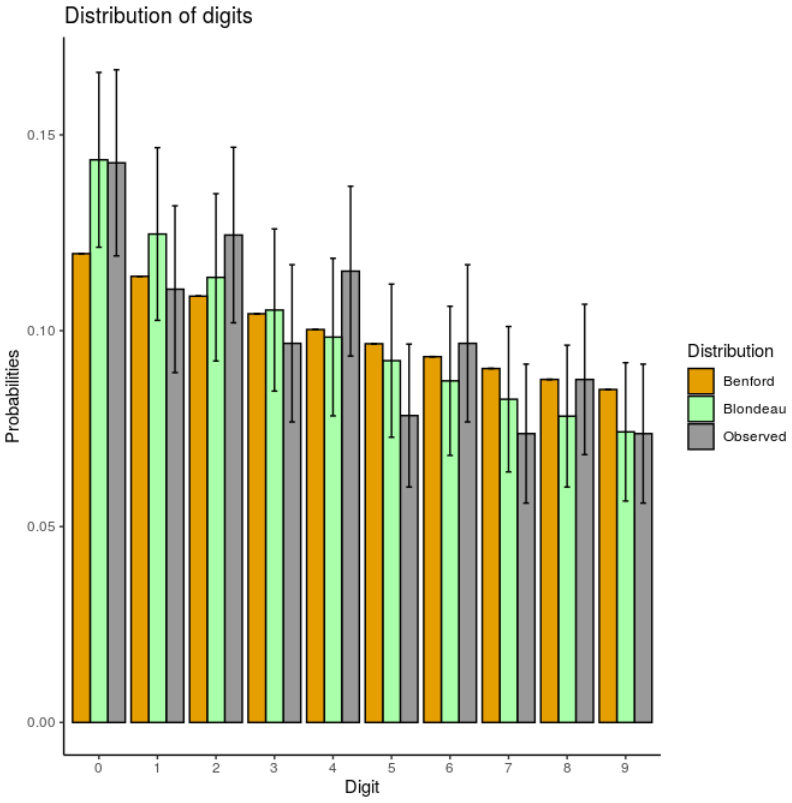


Figure 3. The histogram of Pierre Buffière's street numbers in the second digit case along with Benford's and BDS's distributions of digit.

Example 7. *## Measure of Benford's Law goodness of fit*

```
chi2(address_PierreBuffiere, dig=2, pval=1)
            chi2             pval
1   Chi2 value is:   The p-value is:
2 3.84793221030181 0.921132993758269
```

```
   ## Measure of BDS's Law goodness of fit
chi2(address_PierreBuffiere, dig=2, pval=1, mod="BDS")
            chi2             pval
1   Chi2 value is:   The p-value is:
2 2.47996278328848 0.981417506807756
```

In both cases the null hypothesis cannot be rejected: the studied distribution is consistent with the theoretical distributions. It can be noted that the quality of the adjustment seems slightly better with BDS's model.

5 Conclusion

The use of Benford's Law has increased rapidly in the last few years in extremely diverse fields such as mathematics, physics, biology, economics and demography, to name but a few. But the adjustment proposed by Benford is often only approximate. In some precisely described cases, BDS's probability distributions should be preferred. Indeed probabilities of occurrence of digits in these distributions fluctuate around Benford's values [28], [29].

The **BeyondBenford** package is thus a relevant tool to compare the goodness of fit of Benford's and BDS's distributions in a given collection of data. It offers new laws to find a better approximation of digits distribution in a considered dataset.

References

[1] S. J. Miller, Ed., *Benford's Law: Theory and Applications.* Princeton, NJ: Princeton University Press, 2015.

[2] A. Berger and T. Hill, *An Introduction to Benford's Law*. Princeton, NJ: Princeton University Press, 2015.

[3] R. Newcomb, "Note on the frequency of use of the different digits in natural numbers," *American Journal of Mathematics*, vol. 4, pp. 39–40, 1881.

[4] F. Benford, "The law of anomalous numbers," *Proceedings of the American Philosophical Society*, vol. 78, pp. 127–131, 1938.

[5] T. Hill, "The significant-digit phenomenon," *The American Mathematical Monthly*, vol. 102, no. 4, pp. 322–327, 1995.

[6] T. Sehity, E. Hoelz, and E. Kirchler, "Price developments after a nominal shock: Benford's Law and psychological pricing after the euro introduction," *International Journal of Research in Marketing*, vol. 22, no. 4, pp. 471–480, 2005.

[7] J. Golbeck, "Benford's law applies to online social networks," *Plos One*, vol. 10, no. 8, 2015.

[8] M. Nigrini and W. Wood, "Assessing the integrity of tabulated demographic data," 1995, preprint.

[9] L. Leemis, B. Schmeiser, and D. Evans, "Survival distributions satisfying Benford's Law," *The American Statistician*, vol. 54, no. 4, pp. 236–241, 2000.

[10] M. Nigrini and S. Miller, "Benford's Law applied to hydrology data-results and relevance to other geophysical data," *Mathematical Geology*, vol. 39, no. 5, pp. 469–490, 2007.

[11] T. Alexopoulos and S. Leontsinis, "Benford's Law in astronomy," *Journal of Astrophysics and Astronomy*, vol. 35, no. 4, pp. 639–648, 2014.

[12] E. Costasa, V. Lopez-Rodasa, F. Torob, and A. Flores-Moya, "The number of cells in colonies of the cyanobacterium microcystis aeruginosa satisfies benford's law," *Aquatic Botany*, vol. 89, no. 3, pp. 341–343, 2008.

[13] J. L. Friar, T. Goldman, and J. Pérez-Mercader, "Genome sizes and the Benford distribution," *Plos One*, vol. 7, no. 5, 2012.

[14] C. Durtschi, W. Hillison, and C. Pacini, "The effective use of benford's law to assist in detecting fraud in accounting data," *Journal of Forensic Accounting*, vol. V, pp. 17–34, 2004.

[15] K. Tödter, "Benford's Law as an indicator of fraud in economics," *German Economic Review*, vol. 10, pp. 339–351, 2009.

[16] A. D. Alves, H. H. Yanasse, and N. Y. Soma, "Benford's Law and articles of scientific journals: comparison of JCR and Scopus data," *Scientometrics*, vol. 98, pp. 173–184, 2014.

[17] R. A. Raimi, "The first digit problem," *American Mathematical Monthly*, vol. 83, no. 7, pp. 521–538, 1976.

[18] T. Hill, "Random-number guessing and the first digit phenomenon," *Psychological Reports*, vol. 62, no. 3, pp. 967–971, 1988.

[19] T. W. Beer, "Terminal digit preference: beware of benford's law," *Journal of Clinical Pathology*, vol. 62, no. 2, p. 192, 2009.

[20] P. D. Scott and M. Fasli, "Benford's Law: an empirical investigation and a novel explanation," CSM Technical Report 349, University of Essex, 2001, https://cswww.essex.ac.uk/technical-reports/2001/CSM-349.pdf.

[21] A. Saville, "Using Benford's Law to detect data error and fraud: An examination of companies listed on the Johannesburg Stock Exchange," *South African Journal of Economic and Management Sciences*, vol. 9, no. 3, pp. 341–354, 2006.

[22] N. Gauvrit and J.-P. Delahaye, *Scatter and Regularity Imply Benford's Law... and More*. World Scientific, 2011, pp. 53–69.

[23] D. Geyer, "Detecting fraud in financial data sets," *Journal of Business and Economics Research*, vol. 8, no. 7, pp. 75–83, 2010.

[24] W. R. Mebane Jr, "Election forensics: The second-digit benford's law test and recent american presidential elections," in *Proceedings of the Election Fraud Conference*, 2006.

[25] W. K. T. Cho and B. J. Gaines, "Breaking the (Benford) Law: Statistical fraud detection in campaign finance," *The American Statistician*, vol. 61, no. 3, pp. 218–223, 2007.

[26] D. W. Joenssen, "Two digit testing for benford's law," in *Proceedings of the ISI World Statistics Congress, 59th Session in Hong Kong*, 2013.

[27] A. Diekmann, "Not the first digit! Using benford's law to detect fraudulent scientific data," *Journal of Applied Statistics*, vol. 34, no. 3, pp. 321–329, 2007.

[28] S. Blondeau Da Silva, "Benford or not Benford: a systematic but not always well-founded use of an elegant law in experimental fields," *Communications in Mathematics and Statistics*, 2019.

[29] ——, "Benford or not Benford: new results on digits beyond the first," *arXiv:1805.01291 [stat.OT]*, 2018.

[30] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2013. [Online]. Available: http://www.R-project.org/

[31] D. W. Joenssen, *BenfordTests: Statistical Tests for Evaluating Conformity to Benford's Law*, 2015, r package version 1.2.0. [Online]. Available: https://CRAN.R-project.org/package=BenfordTest

[32] C. Cinelli, *benford.analysis: Benford Analysis for Data Validation and Forensic Analytics*, 2018, r package version 0.1.5. [Online]. Available: https://CRAN.R-project.org/package=benford.analysis

[33] S. Blondeau Da Silva, *BeyondBenford: Compare the Goodness of Fit of Benford's and Blondeau Da Silva's Digit Distributions to a*

Given Dataset, 2020, r package version 1.3. [Online]. Available: https://CRAN.R-project.org/package=BeyondBenford

[34] E. Janvresse and T. De La Rue, "From uniform distributions to Benford's Law," *Journal of Applied Probability*, vol. 41, no. 4, pp. 1203–1210, 2004.

[35] K. Pearson, "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," *Philosophical Magazine*, vol. 50, no. 302, pp. 157–175, 1900.

Stéphane Blondeau Da Silva,

Received August 30, 2020
Accepted July 25, 2021

Stéphane Blondeau Da Silva
Institution Académie de Limoges
Address 13 Rue François Chénieux, 87000 Limoges France
Phone: 07 81 49 13 18
E–mail: Stephane.Blondeau-Da-Silva@ac-limoges.fr

Computer Science Journal of Moldova, vol.**29**, no.2(86), 2021

Magic Sigma Coloring of a Graph

Narahari Narasimha Swamy, Badekara Sooryanarayana,
Akshara Prasad S. P.

Abstract

A sigma coloring of a non-trivial connected graph G is a coloring $c : V(G) \to \mathbb{N}$ such that $\sigma(u) \neq \sigma(v)$ for every two adjacent vertices $u, v \in V(G)$, where $\sigma(v)$ is the sum of the colors of the vertices in the open neighborhood $N(v)$ of $v \in V(G)$. The minimum number of colors required in a sigma coloring of a graph G is called the sigma chromatic number of G, denoted $\sigma(G)$. A coloring $c : V(G) \to \{1, 2, \cdots, k\}$ is said to be a magic sigma coloring of G if the sum of colors of all the vertices in the open neighborhood of each vertex of G is the same. In this paper, we study some of the properties of magic sigma coloring of a graph. Further, we define the magic sigma chromatic number of a graph and determine it for some known families of graphs.

Keywords: Sigma Coloring, open neighborhood sum, magic sigma coloring, sigma chromatic number.

MSC 2010: 05C15.

1 Introduction and preliminaries

Graph coloring is a very important branch of graph theory which has been studied by various authors. While the most commonly studied type of vertex coloring is proper coloring, several of its variations have been introduced and extensively studied. For related work on graph coloring, we refer [11]–[13]. Many of these colorings have been introduced so as to ensure vertex-distinguishing, edge-distinguishing or neighbor-distinguishing properties in a graph. One such coloring is a neighbor-distinguishing coloring, named the sigma coloring of a graph.

As introduced by Chartrand et al. in the year 2010, a k-vertex coloring $c : V(G) \to \mathbb{N}$ of a non-trivial graph G is said to be *sigma coloring* of G if $\sigma_c(u) \neq \sigma_c(v)$ for every two adjacent vertices $u, v \in V(G)$, where $\sigma_c(v) = \sum_{w \in N(v)} c(w)$, called the *open neighborhood sum* of v, is the sum of the colors of all the vertices in the open neighborhood $N(v)$ of $v \in V(G)$. The *sigma chromatic number* of a graph G, denoted $\sigma(G)$, is the minimum number of colors required in a sigma coloring of G. In the paper, it has been proved that the sigma chromatic number of a graph G is bounded by its chromatic number $\chi(G)$. Also, many characterizations of the sigma chromatic number have been established. It is worth mentioning here that the sigma coloring of a graph has been independently studied as lucky labeling by Czerwinski et al. [3] and additive labeling of a graph by Bartnicki et al. [1].

Since its introduction, several studies on the sigma chromatic number have been carried out. In particular, the complexity of sigma partitioning and sigma chromatic number has been discussed by Dehghan et al. [4], [5]. Further, the sigma chromatic number of some particular families of graphs has been obtained by various authors [6], [7].

One interesting question pertaining to graphs G that are not sigma colorable is whether there exists a coloring c of G such that all its vertices receive the same open neighborhood sum, i.e., $\sigma_c(u) = \sigma_c(v)$ for all $u, v \in V(G)$. We introduce the notion of magic sigma coloring to answer this question. Further, we study some of its properties and identify certain families of graphs that admit magic sigma coloring. Also, we obtain the magic sigma chromatic number of some such families of graphs. For standard graph related terminologies, we refer [2], [8], [9].

Definition 1. *Given a simple connected graph $G = (V, E)$, a coloring $c : V(G) \to \{1, 2, \cdots, k\}$ is said to be a magic sigma coloring of G if $\sigma_c(u) = \sigma_c(v)$ for all $u, v \in (G)$. Further, a graph G which admits a magic sigma coloring is said to be magic sigma colorable and $\sigma_c(G) = \sigma_c(v), v \in V(G)$, is called the open neighborhood sum of G w. r. t. the coloring c. Further, a disconnected graph G is said to be magic sigma colorable if each of its components is magic sigma colorable.*

Here, it has to be noted that the magic sigma coloring of a graph

need not be surjective. That is, all the elements in the co-domain, called colors, need not be used in a magic sigma coloring.

Definition 2. *Let G be a graph with c being a magic sigma coloring of G. The c-color sum of G, denoted $S_c(G)$, is the sum of the colors of all the vertices in G, i.e., $S_c(G) = \sum_{v \in V(G)} c(v)$.*

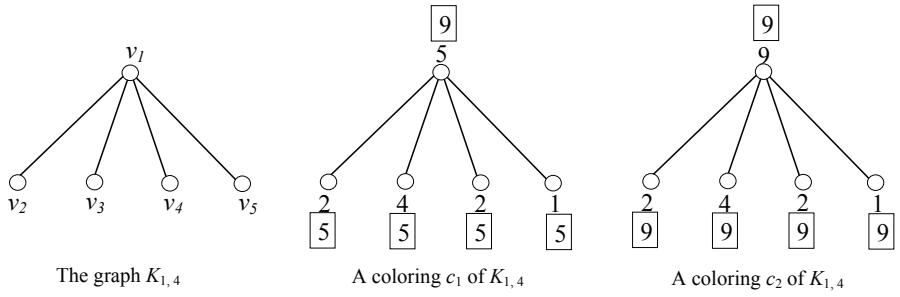


Figure 1. The graph $K_{1,4}$ and its two colorings

To illustrate, consider the graph $K_{1,4}$ and its two colorings c_1 and c_2 in Fig. 1. Since $\sigma_{c_1}(v_1) = 9 \neq 5 = \sigma_{c_1}(v_4), c_1$ is not a magic sigma coloring of $K_{1,4}$. However, c_2 is a magic sigma coloring of $K_{1,4}$ as the open neighborhood sum of each vertex w. r. t. c_2 is the same.

It is easy to observe that some graphs are magic sigma colorable and some others are not. For instance, consider the path P_4 with $V(P_4) = \{v_1, v_2, v_3, v_4\}$ and $E(P_4) = \{v_1v_2, v_2v_3, v_3v_4\}$. Further, for a coloring c to be a magic sigma coloring of P_4, we must have $N_c(v_1) = N_c(v_2) = N_c(v_3) = N_c(v_4)$. However, $N_c(v_1) = c(v_2)$ and $N_c(v_3) = c(v_2) + c(v_4)$. Consequently, we need to have $c(v_2) + c(v_4) = c(v_2)$, i.e., $c(v_4) = 0$. This is not feasible as each color in a magic sigma coloring has to be positive. Thus, P_4 is not magic sigma colorable.

2 Magic colorable graphs

We begin this section with some fundamental results pertaining to the magic sigma colorability of a graph. Further, we discuss about some particular families of graphs which are/are not magic sigma colorable.

Lemma 1. *If a graph G has two vertices such that the open neighborhood of one vertex is a proper subset of the other, then G is not magic sigma colorable.*

Proof. Let G be a graph having two vertices, say u and v, such that $N(u) \subset N(v)$. Then, $\sigma_c(u) < \sigma_c(v)$ in any coloring c of G. Hence, G is not magic sigma colorable. □

Remark 1. *The converse of the above lemma is not true. That is, it is not necessary that, if the graph is not magic sigma colorable, then there exist two of its vertices such that the open neighborhood of one vertex is a proper subset of the other.*

Proof. Consider the graph G_1 in Figure 2. Suppose G_1 is magic sigma

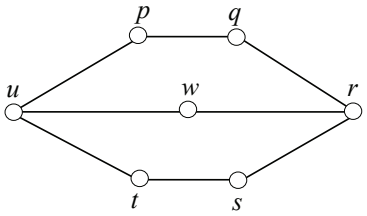


Figure 2. The graph G_1

colorable. Then, there exists a magic sigma coloring, say c of G_1 so that $\sigma_c(v)$ to be the same for all $v \in V(G_1)$. In particular, $\sigma_c(w) = \sigma_c(p) = \sigma_c(q) = \sigma_c(s) = \sigma_c(t)$ implies $c(q) = c(r) = c(s)$ and $c(p) = c(t) = c(u)$.

Similarly, $\sigma_c(u) = \sigma_c(r)$ gives $c(p) + c(t) = c(q) + c(s)$ so that $c(q) = c(r) = c(s) = c(p) = c(t) = c(u)$.

Now, $\sigma_c(u) = \sigma_c(w)$ gives $c(t) + c(w) = c(r)$ which implies that $c(w) = 0$, a contradiction since each color in a magic sigma coloring is positive. Hence, G_1 is not magic sigma colorable.

However, it is easy to observe that no two vertices of G_1 are such that the open neighborhood of one vertex is a proper subset of the other. □

Theorem 2. *The star graph $K_{1,n}, n \geq 1$, is magic sigma colorable.*

Proof. Consider a star graph $K_{1,n}$ with $V(K_{1,n}) = \{v_0, v_1, \cdots, v_n\}$ such that v_0 is the central vertex. Define a coloring $c : V(K_{1,n}) \to \{1, n\}$ as

$$c(v_i) = \begin{cases} n & \text{if } i = 0 \\ 1 & \text{otherwise} \end{cases}.$$

It is easy to observe that $\sigma_c(v_0) = \sigma_c(v_1) = \cdots = \sigma_c(v_n) = n$. Hence, $K_{1,n}$ is magic sigma colorable. □

Lemma 2. *A graph with a pendant vertex has a path of length three if and only if it is not magic sigma colorable.*

Proof. Consider a graph G with a pendant vertex v. Suppose G has a path of length three. Since G is connected, there is a path of length three starting from v. Let the path be $v - w - x - u$. Then, we have $N(v) \subset N(x)$. Therefore, by Lemma 1, G is not magic sigma colorable.

In order to prove the converse, we use the method of contraposition. Suppose G has no path of length three. Then, $diam(G) \leq 2$. Further, G has no cycle. This implies that G is isomorphic to the star graph $K_{1,n}, n \geq 1$ so that it is magic sigma colorable from Theorem 2. □

Corollary 3. *A graph G with $diam(G) \geq 3$ and having a pendant vertex is not magic sigma colorable.*

As a direct consequence of Lemma 2, we have the following result.

Theorem 4. *A non-trivial tree T is magic sigma colorable if and only if $T \cong K_{1,n}, n \geq 1$.*

Theorem 5. *Every k-regular graph with $k \geq 1$ is magic sigma colorable.*

Theorem 6. *The wheel graph W_n is magic sigma colorable.*

Proof. Let W_n be wheel on n vertices $v_1, v_2, \cdots, v_n$, with v_1 as the central vertex and the vertices $v_2 - v_3 - \cdots - v_n - v_2$ forming the cycle. Define a coloring $c : V(W_n) \to \{1, n-3\}$ as

$$c(v_i) = \begin{cases} n-3, & \text{if } i = 1 \\ 1, & \text{otherwise} \end{cases}.$$

It is easy to verify that c is a magic sigma coloring of W_n so that it is magic sigma colorable. □

Theorem 7. *The complete k-partite graph $K_{n_1,n_2,\cdots,n_k}, k \geq 2$, is magic sigma colorable, where each $n_i \geq 1$.*

Proof. Consider the complete k-partite graph $K_{n_1,n_2,\cdots,n_k}$, where each $n_i \geq 1$. Without loss in generality, let $n_1 \geq n_2 \geq \cdots \geq n_{k-1} \geq n_k$. Let $V_1 = \{v_{11}, v_{12}, \cdots, v_{1n_1}\}, V_2 = \{v_{21}, v_{22}, \cdots, v_{2n_2}\}, \cdots, V_k = \{v_{k1}, v_{k2}, \cdots, v_{kn_k}\}$ be the k-partite sets of $V(K_{n_1,n_2,\ldots,n_k})$.

Define a coloring $c : V(K_{n_1,n_2,\cdots,n_k}) \to \{1, 2, \cdots, n_1 + n_2 + \cdots + n_k\}$ as

$$c(v_{ij}) = \begin{cases} n_1 - j + 1 & \text{if } j = n_i \\ 1 & \text{otherwise} \end{cases}$$

for each $i = 1, 2, \cdots, k$ and $j = 1, 2, \cdots, n_i$.

By the definition of c, it follows that

$$\sigma_c(v_{ij}) = \sum_{l(\neq i)=1}^{k} \sum_{m=1}^{n_l} c(v_{lm})$$

for each $i = 1, 2, \cdots, k$ and $j = 1, 2, \cdots, n_i$.

In particular, we have the following:

i) For each $j = 1, 2, \cdots, n_1$,

$$\begin{aligned}\sigma_c(v_{1j}) &= \sum_{l=2}^{k}\sum_{m=1}^{n_l} c(v_{lm}) \\ &= \sum_{l=2}^{k}\sum_{m=1}^{n_l-1} c(v_{lm}) + \sum_{l=2}^{k} c(v_{ln_l}) \\ &= \sum_{l=2}^{k}(n_l - 1) + \sum_{l=2}^{k}(n_1 - n_l + 1) \\ &= \sum_{l=2}^{k} n_1 = (k-1)n_1.\end{aligned}$$

ii) For each $i = 2, 3, \cdots, k$ and $j = 1, 2, \cdots, n_i$,

$$\begin{aligned}\sigma_c(v_{ij}) &= \sum_{l(\neq i)=1}^{k}\sum_{m=1}^{n_l} c(v_{lm}) \\ &= n_1 + \sum_{l(\neq i)=2}^{k}\sum_{m=1}^{n_l-1} c(v_{lm}) + \sum_{l(\neq i)=2}^{k} c(v_{ln_l}) \\ &= n_1 + \sum_{l(\neq i)=2}^{k}(n_l - 1) + \sum_{l(\neq i)=2}^{k}(n_1 - n_l + 1) \\ &= n_1 + \sum_{l(\neq i)=2}^{k} n_1 \\ &= n_1 + (k-2)n_1 = (k-1)n_1.\end{aligned}$$

Thus, we see that the open neighborhood sum of all vertices is the same and is equal to $(k-1)n_1$. As a result, the complete k-partite graph $K_{n_1,n_2,\cdots,n_k}$ is magic sigma colorable. □

Corollary 8. *[10] If a graph G is such that a set consisting of any two adjacent vertices in G forms a minimal dominating set of G, then G is magic sigma colorable.*

Theorem 9. *The k^{th} power graph P_n^k of a path on $n \geq 3$ vertices is magic sigma colorable if and only if $k = n-2$ or $n-1$.*

Proof. Consider the graph P_n^k with $V(P_n^k) = \{v_1, v_2, \cdots, v_n\}$. Since $diam(P_n) = n-1$, we have $k \leq n-1$.

We prove the necessary part by the method of contraposition. Suppose $k \leq n-3$. Then, we see that in the graph P_n^k, $N(v_1) =$

$\{v_2, v_3, \cdots, v_{k+1}\}$ and $N(v_{k+2}) = \{v_2, v_3, \cdots, v_{k+1}, v_{k+3}, v_{k+4}, \cdots, v_{2k+2}\}$ so that $N(v_1) \subset N(v_{k+2})$. Hence, by Lemma 1, P_n^k is not magic sigma colorable.

We prove the converse considering two cases as follows.

Case (1): $k = n - 1$. In this case, $P_n^k \cong K_n$ so that by Theorem 5, P_n^k is magic sigma colorable.

Case (2): $k = n - 2$. Define a coloring $c : V(P_n^k) \to \{1, 2\}$ as

$$c(v_i) = \begin{cases} 1 & \text{if } i = 1, n \\ 2 & \text{otherwise} \end{cases}.$$

Since $k = n - 2 = diam(P_n) - 1$, v_1 is adjacent to all the vertices except v_n and vice-versa. Further, each of the other vertices is adjacent to all the vertices in P_n^k. Thus, we have $N(v_1) = N(v_n) = \{v_2, v_3, \cdots, v_{n-1}\}$ so that $\sigma_c(v_1) = \sigma_c(v_n) = 2(n-2)$ and $N(v_i) = V(P_n^k)$ so that $\sigma_c(v_i) = 1 + 2(n-3) + 1 = 2(n-2)$ for each $i = 2, 3, \cdots, n-1$.

Consequently, P_n^k is magic sigma colorable. □

Theorem 10. *Two graphs G and H are magic sigma colorable if and only if $G + H$ is magic sigma colorable.*

Proof. Let G and H be magic sigma colorable with c_1 and c_2 being the magic sigma colorings of G and H respectively. Let k and m be any non-negative integers such that $k\,(S_{c_1}(G) - \sigma_{c_1}(G)) = m\,(S_{c_2}(H) - \sigma_{c_2}(H))$.

Let c be a coloring of $G + H$ defined by

$$c(v) = \begin{cases} kc_1(v) & \text{if } v \in V(G) \\ mc_2(v) & \text{if } v \in V(H) \end{cases}.$$

Then, we have, for any vertex $v \in V(G + H)$, the following:

Case (1): If $v \in V(G)$, then $\sigma_c(v) = k\sigma_{c_1}(G) + mS_{c_2}(H)$.

Case (2): If $v \in V(H)$, then $\sigma_c(v) = m\sigma_{c_2}(G) + kS_{c_1}(G)$.

It is easy to verify that the open neighborhood sum of every vertex in $G+H$, w. r. t. c, is the same so that $G+H$ is magic sigma colorable.

Conversely, suppose $G+H$ is magic sigma colorable, with c being a magic sigma coloring of $G+H$. Then, for two vertices $u, v \in V(G)$, we have $\sigma_c(u) = \sigma_c(v)$.

This implies that

$$\sum_{\substack{x\in V(G)\\(x,u)\in E(G)}} c(x) + S_c(H) = \sum_{\substack{y\in V(G)\\(y,v)\in E(G)}} c(y) + S_c(H)$$

$$\text{so that} \quad \sum_{\substack{x\in V(G)\\(x,u)\in E(G)}} c(x) = \sum_{\substack{y\in V(G)\\(y,v)\in E(G)}} c(y)$$

$$\text{which implies that } \sigma_c(u)\big|_G = \sigma_c(v)\big|_G.$$

Thus, G is magic sigma colorable with the coloring c restricted to its vertices. The magic sigma colorability of H follows similarly. □

Theorem 11. *A graph G is magic sigma colorable if and only if its complement $\bar{G}$ is magic sigma colorable.*

Proof. Suppose G is magic sigma colorable with c being its magic sigma coloring.
Based on the fact that any graph G or its complement $\bar{G}$ is connected, we consider the following cases:

Case (1): Suppose G and $\bar{G}$ are both connected.
Since G is magic sigma colorable, $\sigma_c(u) = \sigma_c(v) \quad \forall\ u, v \in V(G)$ so that
$S_c(G) - \sigma_c(u) = S_c(G) - \sigma_c(v) \ \ \forall\ u, v \in V(G)$.
This implies that c is a magic sigma coloring of $\bar{G}$ with $S_c(G) - \sigma_c(G)$ as its open neighborhood sum. Hence, $\bar{G}$ is magic sigma colorable.

Case (2): Suppose G is connected, but $\bar{G}$ is disconnected.
Let H be an arbitrary component in $\bar{G}$. Then, there exists an edge between every vertex in $V(H)$ and every vertex in $V(G) - V(H)$

in G since H is a disconnected component of $\bar{G}$.
Let $u, v \in V(H)$ be arbitrary.
Then $\sigma_c(u)\big|_G = \sigma_c(v)\big|_G$ so that $S_c(G) - \sigma_c(u)\big|_G = S_c(G) - \sigma_c(v)\big|_G$.
Choosing a coloring c_1 of H such that $\sigma_{c_1}(v) = S_c(G) - \sigma_c(v)\big|_G$, for each $v \in H$, ensures that H is magic sigma colorable, which in turn, implies that $\bar{G}$ is also magic sigma colorable.

Case (3): Suppose G is disconnected, but $\bar{G}$ is connected, and $H_1, H_2, \cdots, H_k$ are the components of G. Then, each H_i is trivial or connected and is magic sigma colorable so that, from cases (1) and (2), its complement $\bar{H}_i$ is magic sigma colorable too. Further, $\bar{G} = H_1 + H_2 + \cdots + H_k$, so that, by Theorem 10, $\bar{G}$ is magic sigma colorable.

□

3 Magic sigma chromatic number of some graphs

In this section, we define the magic sigma chromatic number of a graph which is magic sigma colorable. Further, we determine this parameter for some classes of graphs.

Definition 3. *Suppose a graph G is magic sigma colorable. Then, the least k for which G admits a magic sigma coloring is called the magic sigma chromatic number of G, denoted by $\sigma_m(G)$.*

Observation 12. *If a graph G is magic sigma colorable, then $\sigma_m(G) \geq 1$.*

Theorem 13. *A graph G is regular if and only if $\sigma_m(G) = 1$.*

Proof. Suppose G is a regular graph. Then, by Theorem 5, G is magic sigma colorable. Further, the coloring $c : V(G) \to \{1\}$ is a magic sigma coloring of G so that $\sigma_m(G) = 1$.

Conversely, let G be a graph with $\sigma_m(G) = 1$. Then, there exists a magic sigma coloring, say $c : V(G) \to \{1\}$ i. e., every vertex is given the same color 1 in G. Suppose G is not regular. Then, there exist at least two vertices, say u and v, such that $deg(u) \neq deg(v)$. Then, $\sigma_c(u) \neq \sigma_c(v)$, a contradiction to the fact that c is a magic sigma coloring of G. Thus, G is a regular graph. □

Lemma 3. *For a star graph $K_{1,n}, n \geq 1$, $\sigma_m(K_{1,n}) = n$.*

Proof. Let $V(K_{1,n}) = \{v_0, v_1, \cdots, v_n\}$ such that v_0 is the central vertex. By Theorem 2, $K_{1,n}$ is magic sigma colorable and uses n colors. Thus, $\sigma_m(K_{1,n}) \leq n$.

For $n = 1$, we have $K_{1,1} \cong P_2$ so that $\sigma_m(K_{1,1}) = 1$ by Theorem 7. Consider the case $n \geq 2$. For a coloring c of $K_{1,n}$ to be a magic sigma coloring, we must have $\sigma_c(v_i) = \sigma_c(v_j)$ for all $i, j = 0, 1, \cdots, n$. Also, we have $\sigma_c(v_0) = \sum_{i=1}^{n} c(v_i)$ and $\sigma_c(v_i) = c(v_0)$ for each $i = 1, 2, \cdots, n$. Thus, $c(v_0) = \sum_{i=1}^{n} c(v_i) \geq n$ since each color is positive. Consequently, $\sigma_m(K_{1,n}) \geq n$. Therefore, $\sigma_m(K_{1,n}) = n$. □

Theorem 14. *The magic sigma chromatic number of the complete k-partite graph $K_{n_1,n_2,\cdots,n_k}$ with $k \geq 2$ and $n_i \geq n_{i+1} \geq 1, i = 1, 2, \cdots k-1$, is $\lceil \frac{n_1}{n_k} \rceil$.*

Proof. Consider the complete k-partite graph $K_{n_1,n_2,\cdots,n_k}$ with $1 \geq n_i \geq n_{i+1}, i = 1, 2, \cdots k-1$. Let V be the vertex set of $K_{n_1,n_2,\cdots,n_k}$ with $V_1 = \{v_{11}, v_{12}, \cdots, v_{1n_1}\}$, $V_2 = \{v_{21}, v_{22}, \cdots, v_{2n_2}\}$, ..., $V_k = \{v_{k1}, v_{k2}, \cdots, v_{kn_k}\}$ being its k-partite sets.

By Theorem 7, $K_{n_1,n_2,\cdots,n_k}$ is magic sigma colorable. For a coloring c to be a magic sigma coloring of $K_{n_1,n_2,\cdots,n_k}$, we should have $\sigma_c(u) = \sigma_c(v)$ for all $u, v \in V$. In particular, $\sigma_c(u) = \sigma_c(v)$ for all $u \in V_1$ and $v \in V_k$. This implies that

$$\sum_{v_{ij} \in V(i \neq 1)} c(v_{ij}) = \sum_{v_{ij} \in V(i \neq k)} c(v_{ij}).$$

Simplifying, we get $\sum_{l=1}^{n_k} c(v_{kl}) = \sum_{m=1}^{n_1} c(v_{1m}) \geq n_1$ since each color is positive. Thus, by the generalized pigeon hole principle, we see that there exists at least one vertex $v \in V_k$ with $c(v) \geq \lceil \frac{n_1}{n_k} \rceil$. We therefore conclude that $\sigma_m(K_{n_1,n_2,\cdots,n_k}) \geq \lceil \frac{n_1}{n_k} \rceil$.

To prove the reverse inequality, define $c : V \to \{1, 2, \cdots, \lceil \frac{n_1}{n_k} \rceil\}$ as

$$c(v_{ij}) = \begin{cases} \lceil \frac{n_1}{n_i} \rceil, & \text{if } j = 1 \\ \left\lceil \frac{n_1 - \sum_{m=1}^{j-1} c(v_{im})}{n_i - j + 1} \right\rceil, & \text{otherwise} \end{cases}$$

for each $i = 1, 2, \cdots, k$.
By the definition of c, we see that $c(v_{k1}) = \lceil \frac{n_1}{n_k} \rceil \geq c(v_{ij})$ for all i, j so that the greatest color used in c is $\lceil \frac{n_1}{n_k} \rceil$. Further, it is easy to observe that c is a magic sigma coloring of $K_{n_1,n_2,\cdots,n_k}$. Consequently, $\sigma_m(K_{n_1,n_2,\cdots,n_k}) \leq \lceil \frac{n_1}{n_k} \rceil$.

Therefore, $\sigma_m(K_{n_1,n_2,\cdots,n_k}) = \lceil \frac{n_1}{n_k} \rceil$. □

Conclusion The concept of magic sigma coloring of graphs has been introduced in this paper. Some families of graphs which are/are not magic sigma colorable have been identified. Further, the magic sigma chromatic number of some such graphs have been established. As a continuation of the work carried out in the paper, one can attempt to characterize magic sigma colorable graphs. Also, graphs with a specific magic sigma chromatic number can be constructed. Further, forbidden graphs pertaining to magic sigma coloring can be identified.

Acknowledgment The authors are thankful to the managements of Tumkur University, Tumakuru and Dr. Ambedkar Institute of Technology, Bengaluru for their constant support and encouragement during the preparation of this paper. The authors are indebted to the learned referees for their thoughtful comments and suggestions.

References

[1] T. Bartnicki, B. Bosek, S. Czerwiński, J. Grytczuk, G. Matecki, and W. Żelazny, "Additive coloring of planar graphs," *Graphs Combin.*, vol. 30, no. 5, pp. 1087–1098, 2014.

[2] F. Buckley and F. Harary, *Distance in Graphs*, Redwood City, Calif.: Addison-Wesley Pub. Co., 1990, 335 p. ISBN-10: 0201095912, ISBN-13: 978-0201095913.

[3] S. Czerwiński, J. Grytczuk, and W. Zelazny, "Lucky labelings of graphs," *Inform. Process. Lett.*, vol. 109, no. 18, pp. 1078–1081, 2009.

[4] A. Dehghan, M. R. Sadeghi, and A. Ahadi, "The Complexity of the Sigma Chromatic Number of Cubic Graphs," 2014, ArXiv abs/1403.6288.

[5] A. Dehghan, M. R. Sadeghi, and A. Ahadi, "Sigma Partitioning: Complexity and Random Graphs," *Discrete Mathematics and Theoretical Computer Science*, vol. 20, no. 2, 2018, #19.

[6] A. D. Garciano, M. C. T. Lagura, and R. M. Marcelo, "On the sigma chromatic number of the join of a finite number of paths and cycles," *Asian-European Journal of Mathematics*, 2021, Article no. 2150019.

[7] L. G. S. Gonzaga and S. M. Almeida, "Sigma Coloring on Powers of Paths and Some Families of Snarks," *Electronic Notes in Theoritical Computer Science*, vol. 346, pp. 485–496, 2019. DOI: 10.1016/j.entcs.2019.08.043.

[8] F. Harary, *Graph theory*, Avalon Publishing, 1969, 274 p. ISBN-10: 0201410338, ISBN-13: 9780201410334.

[9] N. Hartsfield and G. Ringel, *Pearls in Graph Theory: A Comprehensive Introduction*, Revised, Subsequent Edition, Academic Press, 1994, 249 p. ISBN-10: 0123285534, ISBN-13: 978-0123285539.

[10] S. R. Jayaram, "Minimal dominating sets of cardinality two in a graph," *Ind. J. of Pure and App. Math.*, vol. 28, no. 1, pp. 43–46, 1997.

[11] N. Narahari, B. Sooryanarayana, and K. N. Geetha, "Open Neighborhood Chromatic Number of an Antiprism graph," *Appl. Math. E-Notes*, vol. 15, pp. 54–62, 2015.

[12] N. Narahari and B. Sooryanarayana, "Open Neighbourhood Coloring of some Path related graphs," *Eurasian Math. Journal*, vol. 6, no. 4, pp. 77–91, 2015.

[13] B. Sooryanarayana and N. Narahari, "The Neighborhood Pseudochromatic Number of a Graph," *Int. J. Math. Comb.*, vol. 4, pp. 92–99, 2014.

Narahari Narasimha Swamy, Badekara Sooryanarayana, Akshara Prasad S. P.

Received May 31, 2020
Accepted April 27, 2021

Narahari Narasimha Swamy
Department of Mathematics
University College of Science
Tumkur University, Tumakuru - 572103
Karnataka, India
Phone:+919739482878
E–mail: narahari_nittur@yahoo.com

Badekara Sooryanarayana
Department of Mathematical and Computational Studies
Dr. Ambedkar Institute of Technology, Bengaluru - 560056
Karnataka, India
Phone:+919844236450
E–mail: dr_bsnrao@yahoo.co.in

Akshara Prasad S. P.
Department of Mathematics
University College of Science
Tumkur University, Tumakuru - 572103
Karnataka, India
Phone:+9190355 83206
E–mail: akshara.prasad.sp@gmail.com

Computer Science Journal of Moldova, vol.**29**, no.2(86), 2021

Bounds for Degree-Sum adjacency eigenvalues of a graph in terms of Zagreb indices

Sumedha S. Shinde, Narayan Swamy, Shaila B Gudimani, H. S. Ramane

Abstract

For a graph G the degree sum adjacency matrix $DS_A(G)$ is defined as a matrix, in which every element is sum of the degrees of the vertices if and only if the corresponding vertices are adjacent, otherwise it is zero. In this paper we obtain the bounds for the spectral radius and partial sum of the eigenvalues of the DS_A matrix. We also find the bounds for the DS_A energy of a graph in terms of its Zagreb indices.

Keywords: Adjacency eigenvalues, degree sum adjacency matrix, Zagreb index, Eigenvalues, Energy.

MSC 2010: 05C05.

1 Introduction

Association of Graph theory with Chemistry has resulted in introducing more molecular structure descriptors, in particular Topostructural descriptors (Wiener index, Hosoya Z index, Zagreb indices, Mohar indices, and many more). In [6] Gutman and Trinajstić observed that the total π-electron energy depended on the molecular structure. So some expressions were deduced for the π-electron energy containing these two terms:

$$M_1 = \sum_{vertices} (d_u)^2$$

and

$$M_2 = \sum_{edges} d_u \cdot d_v.$$

It was observed that both the terms reflect the extent of branching of a molecular structure and hence were responsible for decreasing the total π-electron energy with increasing branches. Later M_1 and M_2 were renamed as first Zagreb index and second Zagreb index respectively [12]. Numerous results are obtained by mathematicians on M_1 and M_2 [1], [7], [9], [15], [16].

The degree sum matrix for a graph was defined by [13] and the characteristic polynomial of the degree sum matrix for a graph in terms of its adjacency polynomial was also obtained. In the first part of this paper, we discuss the bounds for the DS_A-spectral radius and bounds for the partial sum of the DS_A-eigenvalues in terms of the Zagreb indices. In the later part, we obtain the bounds for the energy of a degree sum adjacency matrix in terms of the Zagreb indices.

The degree sum adjacency matrix $DS_A(G)$ of a graph G is defined as

$$DS_A(G) = [ds_{ij}] = \begin{cases} d_i + d_j, & \text{if there is an edge between } v_i \text{ and } v_j; \\ 0, & \text{otherwise.} \end{cases}$$

The characteristic polynomial of $DS_A(G)$ is defined as

$$P_{DS_A(G)} = \beta^n + a_1\beta^{n-1} + a_2\beta^{n-2} + \cdots + a_n.$$

As $DS_A(G)$ matrix is a real and symmetric, its eigenvalues are real and can be arranged as $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_n$. The largest DS_A eigenvalue is known as the DS_A spectral radius of a graph G.

The first and the second Zagreb indices introduced in the year 1972 by I. Gutman [6] are

$$Zg_1 = Zg_1(G) = M_1(G) = \sum_{i=1}^{n} d_i^2 = \sum_{edge\ e=ij} (d_i + d_j).$$

$$Zg_2 = Zg_2(G) = M_2(G) = \sum_{edge\ e=ij} d_i d_j.$$

Lemma 1. *Let G be a simple n-ordered graph, with every vertex v_i having the degree d_i, $i = 1, 2, \ldots, n$. Let $DS_A(G)$ be the degree sum adjacency matrix of G, then*

$$\sum_{i=1}^{n} \beta_i = 0; \tag{1}$$

$$\sum_{i=1}^{n} \beta_i^2 = 2 \sum_{1 \le i < j \le n} (d_i + d_j)^2. \tag{2}$$

Lemma 2. *Let G be a simple n-ordered graph, with every vertex v_i having the degree d_i, $i = 1, 2, \ldots, n$. Let $DS_A(G)$ be the degree sum adjacency matrix of G with $\beta_1, \beta_2, \ldots, \beta_n$ as its eigenvalues. Let $Zg_1(G)$ and $Zg_2(G)$ be the Zagreb indices. Then,*

$$\begin{aligned}
\sum_{i=1}^{n} \beta_i^2 &= 2 \sum_{1 \le i < j \le n} (d_i + d_j)^2 = 2 \sum_{1 \le i < j \le n} (d_i^2 + d_j^2 + 2 d_i d_j) \\
&= 2 \left[\sum_{i=1}^{n} d_i (d_i)^2 + 2 \sum_{edge\ e=ij} d_i d_j \right] \\
&= 2 \left[\sum_{edge\ e=ij} (d_i + d_j) + \sum_{i=1}^{n} d_i^2 (d_i - 1) + 2 \sum_{edge\ e=ij} d_i d_j \right] \\
&= 2 \left[Zg_1(G) + \sum_{i=1}^{n} d_i^2 (d_i - 1) + 2 Zg_2(G) \right].
\end{aligned} \tag{3}$$

Lemma 3. *If $(c_1, c_2, \ldots, c_n)$ and $(d_1, d_2, \ldots, d_n)$ be n vectors, then by Cauchy-Schwartz inequality [14]:*

$$\left(\sum_{i=1}^{n} c_i d_i \right)^2 \le \left(\sum_{i=1}^{n} c_i^2 \right) \left(\sum_{i=1}^{n} d_i^2 \right). \tag{4}$$

Lemma 4. *Let G be a graph having n vertices and m edges, with adjacency eigenvalues as $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$. Let H be another graph*

with n vertices having $d_1, d_2, \ldots, d_n$ as its vertex degrees and let the degree sum adjacency eigenvalues of H be $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_n$. Then

$$\sum_{i=1}^{n} (\lambda_i \beta_i) \leq \sqrt{4m[Zg_1(G) + 2Zg_2(G) + \sum_{i=1}^{n} d_i^2(d_i - 1)]}. \qquad (5)$$

Proof. By using Lemma 2 and Lemma 3 we have,

$$\begin{aligned} \sum_{i=1}^{n} (\lambda_i \beta_i)^2 &\leq \left(\sum_{i=1}^{n} \lambda^2\right)\left(\sum_{i=1}^{n} \beta^2\right) \\ &= 2m(2[Zg_1(G) + 2Zg_2(G) + \sum_{i=1}^{n} d_i^2(d_i - 1)]) \\ &= 4m[Zg_1(G) + 2Zg_2(G) + \sum_{i=1}^{n} d_i^2(d_i - 1)] \\ \sum_{i=1}^{n} \lambda_i \beta_i &\leq \sqrt{4m[Zg_1(G) + 2Zg_2(G) + \sum_{i=1}^{n} d_i^2(d_i - 1)]}. \end{aligned}$$

□

2 Bounds for spectra of $DS_A(G)$

There are various bounds obtained for largest eigenvalue of an adjacency matrix in literature. In [5], [11] various bounds on the other eigenvalues of signless Laplacian and adjacency matrices are given.
If G is a simple graph of order n having e edges with adjacency eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$, then for $1 \leq p \leq n$,

$$\sqrt{\frac{(n-p)2e}{np}} \geq \lambda_p \geq -\sqrt{\frac{(p-1)2e}{n(n-p+1)}}. \qquad (6)$$

For a adjacency matrix we have $\sum_{i=1}^{n} \lambda_i^2 = 2e$. Here, in degree sum adjacency matrix, the term $2[Zg_1(G)+2Zg_2(G)+\sum_{i=1}^{n} d_i^2(d_i-1)]$ plays the same role. So the direct consequence of Eq.(6) will be Eq.(7).

Theorem 1. *For a graph G, with degree sum adjacency eigenvalues $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_n$ and for $1 \leq p \leq n$*

$$\sqrt{\frac{(n-p)2M}{np}} \geq \beta_p \geq -\sqrt{\frac{(p-1)2M}{n(n-p+1)}}, \tag{7}$$

where $M = [Zg_1(G) + 2Zg_2(G) + \sum_{i=1}^{n} d_i^2(d_i - 1)]$.

Theorem 2. *Let G be a simple n-ordered graph. Let β_1 be the spectral radius of $DS_A(G)$ and $Zg_1(G)$ be the first Zagreb index. Then*

$$\beta_1 \geq \frac{2}{n} Zg_1(G). \tag{8}$$

Proof. Let G be a simple connected graph with n vertices with every vertex v_i having the degree d_i respectively. By the definition of $DS_A(G)$ we observe that the sum of all the entries of $DS_A(G)$ is $\sum_{i \neq j} ds_{ij} = \sum_{i \neq j}(d_i + d_j)$. Let $x = [1, 1, \ldots, 1]$ be the all one vector. Then by Rayleigh principle we have:

$$\begin{aligned} \beta_1 &\geq \frac{xDS_Ax^T}{xx^T} = \frac{1}{n}\sum_{i \neq j}(d_i + d_j) \\ &= \frac{1}{n} 2 \sum_{i<j}(d_i + d_j) \\ &\geq \frac{2}{n} Zg_1(G). \end{aligned}$$

If G is a r-regular graph, then $Zg_1(G) = nr^2$.

$$\beta_1 = \frac{2}{n} nr^2 = 2r^2.$$

Hence the equality holds for regular graph. □

Theorem 3. *Let G be a graph with n vertices and m edges, with $d_1, d_2, \ldots, d_n$ as its vertex degrees and the degree sum adjacency eigenvalues be $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_n$. Then*

$$\beta_1 \leq \sqrt{\frac{2p}{p-1}[Zg_1(G) + 2Zg_2(G) + \sum_{i=1}^{n} d_i^2(d_i - 1)]} + \frac{1}{p-1}\sum_{i=2}^{p} \beta_{n-p+i}. \tag{9}$$

Proof. Let $\beta_1, \beta_2, \ldots, \beta_{n-p+1}, \beta_{n-p+2}, \ldots, \beta_n$ be the degree sum adjacency eigenvalues of G. Let $H = K_p \cup \overline{K_{n-p}}$. The adjacency eigenvalues of H are

$$\lambda_i = \begin{cases} p-1, & 1 \text{ time;} \\ 0, & (n-p) \text{ times;} \\ -1, & (p-1) \text{ times;} \end{cases}$$

and the number of edges of H is $m = \dfrac{p(p-1)}{2}$. Using Lemma 4 we get

$$\sum_{i=1}^{n} (\lambda_i \beta_i) \leq \sqrt{4m[Zg_1(G) + 2Zg_2(G) + \sum_{i=1}^{n} d_i^2(d_i - 1)]} \ ;$$

$$(p-1)\beta_1 + (0) \sum_{i=2}^{n-p-1} \beta_i - \sum_{i=n-p+2}^{n} \beta_i$$

$$\leq \sqrt{4\frac{p(p-1)}{2}[Zg_1(G) + 2Zg_2(G) + \sum_{i=1}^{n} d_i^2(d_i - 1)]} \ ;$$

$$(p-1)\beta_1 \leq \sqrt{2p(p-1)[Zg_1(G) + 2Zg_2(G) + \sum_{i=1}^{n} d_i^2(d_i - 1)]} + \sum_{i=n-p+2}^{n} \beta_i \ ;$$

$$\beta_1 \leq \sqrt{\frac{2p}{(p-1)}[Zg_1(G) + 2Zg_2(G) + \sum_{i=1}^{n} d_i^2(d_i - 1)]} + \frac{1}{p-1} \sum_{i=2}^{p} \beta_{n-p+i} \ .$$

□

Corollary 1. *Let G be a graph on n vertices and m edges, with $d_1, d_2, \ldots, d_n$ as its vertex degrees. Then,*

$$\beta_1 \leq \sqrt{\frac{2(n-1)}{n}[Zg_1(G) + 2Zg_2(G) + \sum_{i=1}^{n} d_i^2(d_i - 1)]}. \qquad (10)$$

Proof. Putting $p = n$ in (9) and using Eq.(1) we get,

$$\begin{aligned}
\beta_1 &\leq \sqrt{\frac{2n}{(n-1)}[Zg_1(G) + 2Zg_2(G) + \sum_{i=1}^{n} d_i^2(d_i-1)] + \frac{1}{n-1}\sum_{i=2}^{n}\beta_i}\,; \\
\beta_1 &\leq \sqrt{\frac{2n}{(n-1)}[Zg_1(G) + 2Zg_2(G) + \sum_{i=1}^{n} d_i^2(d_i-1)] + \frac{1}{n-1}(-\beta_1)}\,; \\
\beta_1 &\leq \sqrt{\frac{2(n-1)}{n}[Zg_1(G) + 2Zg_2(G) + \sum_{i=1}^{n} d_i^2(d_i-1)]}.
\end{aligned}$$

□

Remark 1. *The equality in (10) is satisfied for complete graphs. As* $Zg_1(G) = nr^2 = n(n-1)^2$, $\sum_{i=1}^{n} d_i^2(d_i-1) = n(n-2)(n-1)^2$, $Zg_2(G) = mr^2 = \dfrac{n(n-1)^3}{2}$, *substituting this in (10) we get*

$$\beta_1 = 2(n-1)^2.$$

Corollary 2. *The spectral radius of* $DS_A(G)$ *is bounded by*

$$\frac{2}{n}Zg_1(G) \leq \beta_1 \leq \sqrt{\frac{2(n-1)}{n}[Zg_1(G) + 2Zg_2(G) + \sum_{i=1}^{n} d_i^2(d_i-1)]}. \tag{11}$$

Proof. Combining Eq.(8) and Eq.(10), we get the bounds for the DS_A spectral radius of graph G. □

Remark 2. *The equality in (11) holds for complete graphs.*

Theorem 4. *Let* G *be an* n*-ordered graph with vertex degrees* $d_1, d_2, \ldots, d_n$ *and its degree sum adjacency eigenvalues as* $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_n$. *Then,*

$$\sum_{i=1}^{k} \beta_i \leq \sqrt{\frac{2k(p-1)}{p}[Zg_1(G) + 2Zg_2(G) + \sum_{i=1}^{n} d_i^2(d_i - 1)]}\ ,$$
$$1 \leq k \leq n. \tag{12}$$

Proof. Let $\beta_1, \beta_2, \ldots, \beta_k, \beta_{k+1}, \ldots, \beta_n$ be the degree sum adjacency eigenvalues of G. Let H be the union of k copies of complete graph K_p, that is $H = \cup_k K_p$, where $kp = n$. The adjacency eigenvalues of H are

$$\lambda_i = \begin{cases} p-1, & k \text{ times;} \\ -1, & (n-k) \text{ times.} \end{cases}$$

Then the number of vertices of H is $n = pk$, and therefore its edges are $\frac{kp(p-1)}{2}$. Using Lemma 4,

$$\begin{aligned}
(p-1)\sum_{i=1}^{k} \beta_i - \sum_{i=k+1}^{n} \beta_i &\leq \sqrt{\frac{4kp(p-1)}{p}[Zg_1(G) + 2Zg_2(G) + \sum_{i=1}^{n} d_i^2(d_i-1)]}; \\
p\sum_{i=1}^{k} \beta_i - \sum_{i=1}^{n} \beta_i &\leq \sqrt{2kp(p-1)[Zg_1(G) + 2Zg_2(G) + \sum_{i=1}^{n} d_i^2(d_i-1)]}; \\
p\sum_{i=1}^{k} \beta_i &\leq \sqrt{2kp(p-1)[Zg_1(G) + 2Zg_2(G) + \sum_{i=1}^{n} d_i^2(d_i-1)]}; \\
\sum_{i=1}^{k} \beta_i &\leq \sqrt{\frac{2k(p-1)}{p}[Zg_1(G) + 2Zg_2(G) + \sum_{i=1}^{n} d_i^2(d_i-1)]}.
\end{aligned}$$

Thus we have obtained the bound for the sum of k, DS_A eigenvalues of a graph G. If $k = 1$, we observe that the Eq.(12) get reduced to Eq.(10). □

Theorem 5. *Let G be a graph on n vertices and m edges, with $d_1, d_2, \ldots, d_n$ as its vertex degrees and degree sum adjacency eigenvalues $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_n$. Then,*

$$\sum_{i=1}^{k}(\beta_i - \beta_{n-k+i}) \leq \sqrt{4k[Zg_1(G) + 2Zg_2(G) + \sum_{i=1}^{n} d_i^2(d_i - 1)]}. \quad (13)$$

Proof. Let $\beta_1, \beta_2, \ldots, \beta_k, \beta_{k+1}, \ldots, \beta_{n-k}, \beta_{n-k+1}, \ldots, \beta_n$ be the degree sum adjacency eigenvalues of G. Let H be the union of k copies of $K_{p,q}$ a complete bipartite graph, that is $H = \cup_k K_{p,q}$, where $kp = n$. The adjacency eigenvalues of H are

$$\lambda_i = \begin{cases} \sqrt{pq}, & k \text{ times;} \\ 0, & (n-2k) \text{ times;} \\ -\sqrt{pq}, & k \text{ times.} \end{cases}$$

The number of edges of H is kpq. Using Lemma 4, we get:

$$\sqrt{pq}\sum_{i=1}^{k}\beta_i + 0\sum_{i=k+1}^{n-k}\beta_i - \sqrt{pq}\sum_{i=k+1}^{n}\beta_i \leq \sqrt{4kpq[Zg_1(G) + 2Zg_2(G) + \sum_{i=1}^{n} d_i^2(d_i - 1)]};$$

$$\sqrt{pq}\sum_{i=1}^{k}\beta_i - \sum_{i=1}^{k}\beta_{n-k+i} \leq \sqrt{4kpq[Zg_1(G) + 2Zg_2(G) + \sum_{i=1}^{n} d_i^2(d_i - 1)]};$$

$$\sum_{i=1}^{k}(\beta_i - \beta_{n-k+i}) \leq \sqrt{4k[Zg_1(G) + 2Zg_2(G) + \sum_{i=1}^{n} d_i^2(d_i - 1)]}.$$

□

3 Bounds for Energy of a DS_A matrix

The Energy of a Degree Sum adjacency matrix $DS_AE(G)$ can be defined as the sum of the absolute DS_A eigenvalues of a graph G, analogous to the various energy concepts like energy of an adjacency matrix [8] and distance matrix [3]. This energy is also referred to as Zagreb energy in [10]:

$$DS_AE(G) = \sum_{i=1}^{n} |\beta_i|. \tag{14}$$

Hyper-Zagreb index was recently introduced in [4], which is defined as $HM(G) = \sum_{edge\ e=ij}(d_i+d_j)^2$. So using Lemma (2), we can express hyper-Zagreb index in terms of first two Zagreb indices.

$$HM = [Zg_1(G) + 2Zg_2(G) + \sum_{i=1}^{n} d_i^2(d_i - 1)]. \tag{15}$$

In [10], authors have expressed bounds for Zagreb energy in terms of hyper-zagreb index. So using Eq.(15) we state that bounds for Zargeb energy can also be expressed in terms of first and second Zagreb indices.

$$\sqrt{2[Zg_1(G) + 2Zg_2(G) + \sum_{i=1}^{n} d_i^2(d_i - 1)]} \leq DS_AE(G)$$
$$\leq \sqrt{2n[Zg_1(G) + 2Zg_2(G) + \sum_{i=1}^{n} d_i^2(d_i - 1)]};$$

$$DS_AE(G)$$
$$\geq \sqrt{2[Zg_1(G) + 2Zg_2(G) + \sum_{i=1}^{n} d_i^2(d_i - 1)] + n(n-1)|det(DS_A(G))|^{2/n}}.$$

Theorem 6. *Let G be an r-regular graph with n vertices. Then*

$$DS_AE(G) \geq 4r^2. \tag{16}$$

Proof. Let G be an r regular graph with n vertices and $2r^2, 2r\lambda_2, 2r\lambda_3, \ldots, 2r\lambda_n$ be its DS_A eigenvalues in terms of its adjacency eigenvalues. Then

$$\begin{aligned} DS_AE(G) &= |2r^2| + \sum_{i=2}^{n} |2r\lambda_i| \\ &\geq 2r^2 + \left| \sum_{i=2}^{n} 2r(-r) \right| \\ &\geq 2r^2 + \left| -2r^2 \right| \\ &\geq 4r^2. \end{aligned}$$

□

References

[1] B. Zhou and I. Gutman, "Further properties of Zagreb indices," *MATCH Commun. Math. Comput. Chem*, vol. 54, no. 1, pp. 233–239, 2005.

[2] D. M. Cvetković, M. Doob, and H. Sachs, *Spectra of Graphs: Theory and Applications*, New York: Academic Press, 1980.

[3] G. Indulal, I. Gutman, and A. Vijayakumar, "On distance energy of graphs," *MATCH Commun. Math. Comput. Chem*, vol. 60, no. 2, pp. 461–472, 2008.

[4] G. H. Shirdel, H. Rezapour, and A. M. Sayadi, "The hyper-Zagreb index of graph operations," *Iran. J. Math. Chem.*, vol. 4, no. 2, pp. 213–220, 2013.

[5] H. S. Ramane, S. S. Shinde, S. B. Gudimani, and J. B. Patil, "Bounds for the signless Laplacian eigenvalues of graphs," *International J. of Math. Sci. and Engg. Appls. (IJMSEA)*, vol. 6, no. 6, pp. 129–135, 2012.

[6] I. Gutman and N. Trinajstić, "Graph theory and molecular orbitals. Total φ-electron energy of alternant hydrocarbons," *Chemical Physics Letters*, vol. 17, no. 4, pp. 535–538, 1972. DOI: https://doi.org/10.1016/0009-2614(72)85099-1.

[7] I. Gutman and K. Das, "The first Zagreb index 30 years after," *MATCH Commun. Math. Comput. Chem*, vol. 50, pp. 83–92, 2004.

[8] I. Gutman, "The energy of a graph," *Ber. Math. Stat. Sekt. Forschungszentrum Graz*, vol. 103, 1978, pp. 1–22.

[9] K. Das and I. Gutman, "Some properties of the second Zagreb index," *MATCH Commun. Math. Comput. Chem*, vol. 52, pp. 103–112, 2004.

[10] N. Jafari Rad, A. Jahanbani, and I. Gutman, "Zagreb Energy and Zagreb Estrada Index of Graphs," *MATCH. Commun. Math. Comput. Chem*, vol. 79, no. 2, pp. 371–386, 2018.

[11] R. C. Brigham and Dutton, "Bounds on the graph spectra," *J. Combin. Theory*, vol. 37, no. 3, pp. 228–234, 1984. DOI: https://doi.org/10.1016/0095-8956(84)90055-8.

[12] R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley-vch, 2000. Print ISBN: 9783527299133, DOI: 10.1002/9783527613106.

[13] R. K. Zaferani, "A study of some topics in the theory of graphs," Ph.D. Dissertation, University of Mysore, Mysore, 2009.

[14] S. Bernard and J. M. Child, *Higher Algebra*, New Delhi: Macmillan India Ltd., 1994, xiv+585p.

[15] S. Nikolić, G. Kovačević, A. Miličević, and N. Trinajstić, "The Zagreb indices 30 years after," *Croatica chemica acta*, vol. 76, no. 2, pp. 113–124, 2003.

[16] T. Došlic, B. Furtula, A. Graovac, I. Gutman, S. Moradi, and Z. Yarahmadi, "On vertex–degree–based molecular structure descriptors," *MATCH Commun. Math. Comput. Chem*, vol. 66, no. 2, pp. 613–626, 2011.

Sumedha S. Shinde, Narayan Swami,
Shaila Gudimani, H. S. Ramane

Received June 25, 2020
Accepted March 16, 2021

Sumedha S. Shinde
Depathment of Mathematics,
KLE Technological University,
Hubballi-580031, India
Phone: 9448631697
E–mail: sumedhanarayan@gmail.com

Narayan Swamy
Depathment of Mathematics,
KLE Technological University,
Hubballi-580031, India
Phone: 9448631697
E–mail: nswamy@kletech.ac.in

Shaila Gudimani
Depathment of Mathematics,
KLE Technological University,
Hubballi-580031, India
Phone: 9448631697
E–mail: sbgudimani@kletech.ac.in

Harishchandra S. Ramane
Department of Mathematics,
Karnatak University,
Dharwad - 580003, India
Phone: +919945031752
E—mail: hsramane@yahoo.com

Constantin Gaindric – 80th anniversary

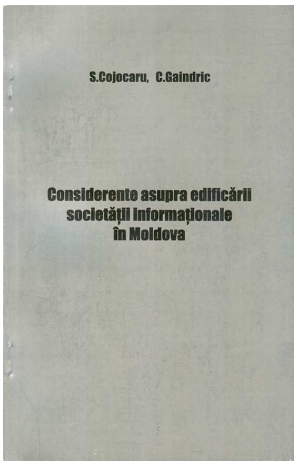
On September 11, 2021 Prof. Constantin Gaindric turns 80! This age does not track in any way with this person when you look at him.

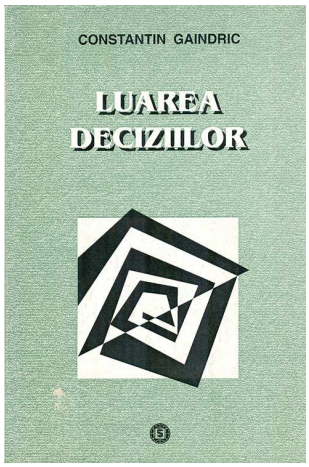
Always fit, neat, impetuous, smart, vigorous, decisive, attentive, curious, with a smile on his face, ready to explain, suggest, discuss, tell an interesting story from his busy life, a lover of good music and literature – it's all about one person.

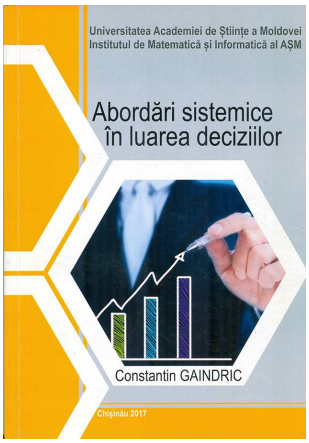
Doctor habilitatus, Full Professor, Corresponding Member of the Academy of Sciences of Moldova, National Prize Laureate of Moldova, Editor-in-Chief of the "Computer Science Journal of Moldova", Member of the Mathematical Society of the Republic of Moldova, Member of the Romanian Society of Mathematical Modeling, Member of the Balkan Union for Fuzzy Systems – these are several, but by no means all items from the large list that may characterize the same one person.

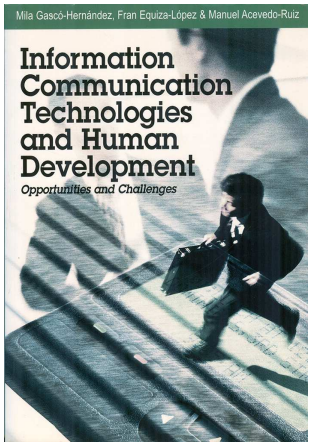
Constantin Gaindric published his first scientific paper being still a student at the Pedagogical Institute in Bălți. As a co-author, there was the name of Professor Israel Gohberg, a name that will soon become notorious in the world of functional analysis and operator theory. Then

none of them assumed that they will become colleagues at the Institute of Mathematics of the Academy of Sciences of Moldova. Later the words "and Computer Science" were added to the Institute title thanks to the insistence of Professor Gaindric, who was already the director of this institute. It was here that he fully manifested his talent as a mathematician, computer scientist, but also as a leader and manager. He had a rather difficult mandate, in a period of disintegration of the USSR and the formation of the independent state of the Republic of Moldova. The transition was not easy at all: half-year overdue salaries, unheated offices in winter, disconnected telephones. The great merit of director Gaindric was the preservation of the staff, which survived even in those harsh conditions. Moreover, at that time he initiated publication of the Computer Science Journal of Moldova, concluded the first collaboration agreements with research institutions abroad. And when a representative of an organization, one of those that was well funded at any time, proposed to "buy" the institute, offering funding in exchange for hiring the entire staff that worked on the issue related to its profile, he replied firmly: "The institute is not for sale".









He hesitated a lot when he was proposed to head the Supreme Attestation Commission (SAC), which was in charge of organizing the doctoral theses defenses and conferring the respective titles. He accepted the challenge and carried out a radical reform of this field, abandoning the Soviet model and establishing the one close to that of European countries. Specifically, there were abandoned the specialized scientific councils of 20 people and more, of which there were two or three members who understood what the thesis was about. Instead of this, a flexible mechanism was adopted, with a commission of 5-7

people, specialists in that subject. Moreover, the theses began to be posted on the SAC website. This was the first step in the direction of Open Science, which is now talked about a lot; in the Republic of Moldova it was done 20 years ago. These things seem natural now, but 20 years ago Prof. Gaindric had to overcome the harsh resistance of people rooted in the old system, who could not imagine how a doctoral commission could work without having a permanent chair for all theses' defenses in the field, or how the text of a thesis can be displayed for public access.

Always young even at this anniversary time, energetic, full of new ideas, he remains one of the most active members of the Academy of Sciences of Moldova and the most devoted researchers of the Institute of Mathematics and Informatics. Happy birthday, dear colleague!

We wish you health, energy, creative success, and a good mood!

The Editorial Board of Computer Science Journal of Moldova and the staff of the Vladimir Andrunachievici Institute of Mathematics and Computer Science