VOLUME 8 NUMBER 2 BANJA LUKA, DECEMBER 2018 (45-96) ISSN 2232-9625 (Print) ISSN 2233-0194 (Online) UDC 004

Journal of Information Technology and Applications

(BANJA LUKA)

JITA

Exchange of Information and Knowledge in Research

APEIRON

THE AIM AND SCOPE

The aim and scope of the Journal of Information Technology and Applications (JITA) is:

- to provide international dissemination of contributions in field of Information Technology,
- to promote exchange of information and knowledge in research work and
- to explore the new developments and inventions related to the use of Information Technology towards the structuring of an Information Society.

JITA provides a medium for exchanging research results and achievements accomplished by the scientific community from academia and industry.

By the decision of the Ministry of Education and Culture of the Republic of Srpska, no.: 07.030-053-160-4/10 from 3/3/2010, the journal „Journal of Information Technology and Applications“ Banja Luka is registered in the Registry of public organs under the number 591.
Printed by Markos, Banja Luka in 300 copies two times a year.

Indexed in: LICENSE AGREEMENT, 3.22.12. **EBSCO** Publishing Inc., Current Abstracts

ebscohost.com
indexcopernicus.com
citefactor.org/contact
scholar.google.com
doisrpska.nub.rs
crossref.org
road.issn.org
citefactor.org
cosmosimpactfactor.com

Printed on acid-free paper

Annual subscription is 30 EUR
Full-text available free of charge at http://www.jita-au.com

UDC 004

ISSN 2232-9625 (Print)

ISSN 2233-0194 (Online)

JITA

VOLUME 8 NUMBER 2 BANJA LUKA, DECEMBER 2018 (45-96)

Contents

The content of this issue of JITA consists of five papers.

The first paper, entitled "Multidimensional Numbers and Semantic Numeration Systems. Theoretical Foundation and Application" by Alexander Ju. Chunikhin, presents a new class of numeration systems, namely Semantic Numeration Systems. The methodological background and theoretical foundations of such systems are considered. The concepts of abstract entity, entanglement and valence of abstract entities, and the topology of the numeration system are introduced. The proposed classification of semantic numeration systems allows to choose the numeration system depending on the problem being solved.

The next paper is "Implementation of The Neural Network Algorithm In Advanced Databases by Šikanjić Nedeljko, Zoran Ž. Avramović and Esad Jakupović. This paper presents the progress of humanity closely related to the progress of technology. The highlight of the development of technology will occur when the machines are able to do what they learn and know; think and make decisions on their own without human help. In this paper we.

The third article is "E-mail forensics: techniques and tools for forensic investigation of one court case" by Ljubomir Lazić. E-mail has emerged as the most important application on the Internet for communication of messages, delivery of documents and carrying out transactions and is used not only from computers, but many other electronic gadgets such as mobile phones. This paper is an attempt to illustrate e-mail architecture from forensics perspective. Also, this paper projects the need for e-mail forensic investigation and lists various methods and tools used for its realization.

The paper "Psychological connection between colors and certain characteristic terms", by Nedim, presents results of a research on psychological connection between 40 offered colors and 91 terms from everyday life. Similar researches have been conducted and published in a number of instances in domestic and foreign bibliography, but this research has certain particularities that are not often present in other articles. For one thing, colors whose associativity with certain terms is being analyzed are shown in a table, their name and code in the RGB system of color marking are provided.

The last article in this issue is "Safety aspects in shared medical IT environment" by Igor Dugonjić, Mihailo Travar, Gordan Bajić, Regional PACS and other shared medical systems are primary intended for sharing medical images. In these systems, the number of users is significantly increased in relation to local systems, and the fact is that the public network is very frequently used for data transfer. As medical data are very sensitive, such situation creates considerable risk regarding privacy, integrity and right to access to these data. This paper includes the most frequent risks and methods to solve these issues as well as recommendations for safe use of cloud computing systems in order to implement these systems.

On behalf of the Editorial Board we would like to thank the authors for their high quality contributions, and also the reviewers for the effort and time invested into the preparation of this issue of Journal of Information Technology and Applications.

Editors, Gordana Radic, Editor-in-Chief, Zoran Avramović, Dušan Starcevic

Multidimensional Numbers and Semantic Numeration Systems:Theoretical Foundation and Application

Alexander Ju. Chunikhin

Palladin Institute of Biochemistry, National Academy of Sciences of Ukraine, alchun@ukr.net

Contribution to the state of the art

DOI: **10.7251/JIT1802049C** UDC: **512.643:511.1**

Abstract: In this article, we present a new class of numeration systems, namely Semantic Numeration Systems. The methodological background and theoretical foundations of such systems are considered. The concepts of abstract entity, entanglement and valence of abstract entities, and the topology of the numeration system are introduced. The proposed classification of semantic numeration systems allows to choose the numeration system depending on the problem being solved. Examples of the use of a two-dimensional number system for image compression problems and computation of a two-dimensional convolution are given.

Keywords: Semantics, Abstract Entity, Entanglement, Numeration System.

Introduction

The modern world is characterized as an ever-increasing volume of stored and processed information, as the growing diversity and complexity of algorithms for its processing. A key role in such areas as the theory of formal languages and automata, control systems and artificial intelligence systems, cryptography and image processing belongs to digital data processing. Along with the search for new algorithms of the digital processing, the search for effective systems for representing numbers and operating with them (i.e. numeration systems) has begun.

From systems with non-natural bases, combinatorial and Fibonacci [5, 6] to numeration systems with double and multiple bases [4] - this is the range of theoretical searches and attempts of application. The generalized presentation of these efforts was found in the concept of abstract number systems [1, 8] as an infinite regular language over a totally ordered alphabet. However, all these systems have, one way or another, one basic object of speculation - the basis of the numeration system.

Is there a way to "look" at a numeration system in a different way? What else, besides the base, can become a generatrix of the numeration system? Are there other number systems, except systems of natural, integer, rational, real and complex numbers? And if so, what could be their representation systems? Where and how can this be applied?

Metaphysics of number

In modern mathematics the concept of a number is considered as basic, intuitively clear and not exactly defined. Only the main purpose of numbers is indicated: counting and ordering.

Perhaps, the first (and, as far as the author knows, the only) scientist who gave the definition of the concept of number (though not mathematical, but philosophical) is A.F. Losev. In [7] he points out that “a number is a definite form, or type, of pure semantic positing...”. One is a steady of the abstract act of positing. The homogeneity of acts of positing determines the uniformity of the ones that make up the number. Those the key moment in the set of

natural numbers is the equipoise - "it is the same always and everywhere" [7].

However, the number is not just a collection of ones. According to A.F. Losev "the number is nothing other than a definite totality of elements". Let us give the phrase "definitely totality" a modern vision. We present the following definition of a number:

The number is the steady of the entanglement of pure semantic positing acts.

The formalization of the collection of ones in a number is realized precisely by the abstract act of entangled them into one whole. We will call this "act-glueing" as semantic entanglement and denote it as $\diamondsuit$.

Semantic entanglement is a mental attitude in which the state of two or more objects (entities) must be described in a *semantic relationship* with each other (as one), even if the individual objects (entities) are not related to each other physically by the relations of generation, inclusion, interaction etc.

Only the semantically entangled ones form the cardinal characteristic of a set of ones, i.e. number as a pure quantity:

$$N = \text{card}(\{1, 1, \ldots, 1\}) = \diamondsuit(1, 1, \ldots, 1).$$

The limit of semantic entanglement of ones in the "whole" on the m-th act of positing leads to the stopping of the process of further generation of "next" numbers and fixation of *another one*: 1_m, i.e. m-ki. This possibility of the other ones positing also makes it possible to construct positional numeration systems. One is always a semantic one of some (abstract) entity. In positional numeration systems, entities form the positions.

Let us introduce the concept of the (cardinal) Abstract Entity (*Æ*), the basic properties of which will be the ability to take it essence in discrete portions (units), accumulate them, keep them, and in the case of excess, transfer the result of overflow to another Abstract Entities that entangled with it:

$$Æ_i = (i, n, P \mid \alpha \vee \omega, q, p),$$

where *i* is the name (identifier) of *Æ*i; *n* - cardinal capacity of *Æ* (the threshold of accumulation of entity ones 1i); *α* is the number of ones contained in *Æ*, within its capacity ($\alpha < n$); ω - an overflow - the number of ones equal to or greater than the *Æ* capacity ($\omega \geq n$); *q* is the number of ones arriving at the *Æ*i input from the connected *Æ*j; *p* is the overflow transfer value; *P* is the P-rule for determining the overflow transfer value (carry). Here $\forall$ n, α, ω, q, p $\in N_0$.

We call the "constant" part $Æ_i$ = (i, n, P) *the signature* of $Æ_i$, and the "variable" ($\alpha \vee \omega$, q, p) as its *state*.

The semantic postulate of positional numeration systems

Proposition 1. *The fact of overflowing of some entity $Æ_i$ makes sense for another entity $Æ_j$.*

On the assumption of the accepted methodological setting, it makes sense to talk about the directional semantic entanglement of abstract entities, in this case - the positions of the numeration system.

The directional semantic entanglement of abstract entities $Æ_i \triangleright Æ_j$ or $Æ_j$ = Ent ($Æ_i$) here means that the result of the overflow of the entity $Æ_i$ in the form of the number p_i (carry) becomes the one(-s) 1_j of the entity $Æ_j$ in a meaningful way. In this case, the fact of carry formation in the entity $Æ_i$ coincides with the fact of the assumption of the one 1_j in the entity $Æ_j$ that is semantically entangled with $Æ_i$.

Thus, any positional numeration system (PNS) is a collection of direct entangled abstract entities of a given signature:

$$PNS = \triangleright Æ_\bullet = \text{Ent}(\ldots\text{Ent}(Æ_i)\ldots).$$

Multinumbers and Polynumbers

The assumption of the heterogeneity of acts of positing with subsequent entanglement leads to the concept of a *multiset*, or a set with repeating elements. In other words, the multiset contains semantically entangled entities that have independent cardinals:

$$MM = \diamondsuit(a \mid \#a; b \mid \#b; \ldots; c \mid \#c).$$

If we introduce certain m-ary multirelations on the set of elements and assume the multiplicity of these elements, then we will speak of *polysets* [2].

Example. A = (one blue cube, eleven red pyramids, three black spheres, seven green cones, and five black cones).

A system of semantically entangled cardinals of multisets will be called *multinumbers*, and of polysets are called *polynumbers*. For polynumbers in [2, 3], the concept of a multidimensional natural number was introduced and the Peano system of such numbers was justified. What kind of number system should provide repre-

sentation and account of heterogeneous semantically entangled entities?

Let us move on to **semantic numeration systems**.

It is in the semantic numeration system the concept of directed entanglement of heterogeneous Abstract Entities arises due to the semantic entanglement of heterogeneous acts of positing.

Proposition 2. *The fact of overflow of some entity $Æ_i$ is meaningful for (several) other entities $Æ_j$, ..., $Æ_k$.*

$Æ_k$

$Æ_i \quad Æ_j \triangle\triangleright$ or $Æ_j = \mathrm{Ent}(Æ_i) \wedge Æ_k = \mathrm{Ent}(Æ_i)$.

Proposition 3. *For some entity Æi, the facts of overflowing (several) other entities Æj, ..., Æk are meaningful.*

$Æ_k \triangleright Æ_i \triangle$ ili $Æ_i = \mathrm{Ent}(Æ_j) \wedge Æ_i = \mathrm{Ent}(Æ_k)$.

$Æ_j$

Thus, in the general case, the abstract entity can both "perceive" the results of the overflow of *m* other entities, and "transfer" the result of its overflow to other abstract entities. The number of Æs, adhered to a given $Æ_i$ by its input, will be called the *passive valence* of $Æ_i$ and will be denoted as W_i. The number of Æs, adhered to a given $Æ_i$ by its output, will be called the *active valence* of $Æ_i$ and will be denoted as V_i.

For any Æ as the position of the semantic numeration system, now:

$$Æ_i = (\mathbf{i} \mid n_i, W_i, V_i, P_i \mid \alpha_i \vee w_i, q_i, p_i),$$

where *i* is the name (identifier) of Æ - in general, a tuple of partial names (characters, numbers) that constitute the full name of $Æ_i$; n_i - the capacity of $Æ_i$ (or the threshold); W_i is the passive valence; V_i is the active valence; α_i is the number of ones contained in $Æ_i$, within its capacity ($\alpha_i < n_i$); w_i - overflow ($w_i \geq n_i$); q_i - the number of units of the entity that have simultaneously entered the input $Æ_i$ from the others Æs connected to it; p_i is the overflow transfer value (carry). P_i is the P-rule for $Æ_i$.

The choice of Æs, their signatures and the method of the formation of a numeration system from them should be determined precisely by the given semantic interaction of entities within the framework of the problem being solved. The structure of the numeration system in accordance with the specified interaction semantics of Æs will be called the *topology of valence entanglement* (or valent entanglement matrix - VEM) of semantic numeration system (SNS). Thus, the SNS description will consist of a signature:

Signt (SNS) = <**I, i. | n., [VEM], P.**>,

and a state:

State (SNS) = < $\alpha. \vee \omega.$, q., p. >

Classification of semantic numeration systems

By the variability of the structure (topology of valence entanglement):

- constants;
- variables: - functionally defined / assigned;
  - uncertain (random, fuzzy).

By the regularity of the structure:

- regular, i.e. invariant to the structural shift (for example, a 2D lattice);
- irregular.

By the variability of the abstract entities signatures:

• By directions:
  - isotropic (identical in all directions);
  - anisotropic (different in different directions).
• Inside each direction:
  - homogeneous ($n_i(dir_i)$ = const);
  - heterogeneous ($n_i(dir_i)$ = var);
  - mixed (homogeneous in one direction and heterogeneous in others).

By type of valence:

- isovalent ($W_i = V_i$, "$Æ_i$);
- heterovalent ($\$Æ_i, W_i \neq V_i$);

By type of P-rule:

- standard numerical ($p = [\omega / n], \alpha` = \omega \bmod (n)$);
- special ($p = f(\omega, n, i)$).

On stability:

- stable (the procedure for representing a multinumber in SNS ends in a finite number of steps);
- unstable (the procedure is infinite).

By the controllability of the signature:

- autonomous (not controlled, hard-set);
- controlled (adaptive or under external control).

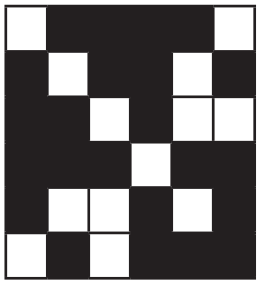
Along with the representation of numbers, any numeration system should provide the realization of elementary arithmetic operations with them. For SNS with a regular structure (2D lattice) in [2], operations of addition and multiplication of two-dimensional numbers are justified. A one-to-one

correspondence between the sum and products of multidimensional numbers and their representations in the 2D lattice SNS is proved.

Applications

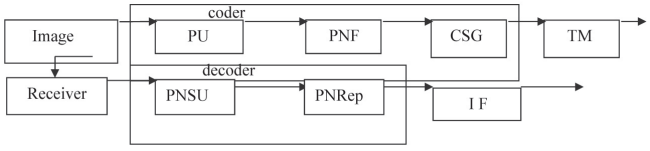
1. Compression of black and white images.

The main idea of the proposed method of image (or fragment) compression is to give the digital relief of the image the sense of the representation of the polynumber *A* in a regular isotropic SNS with $n_{ij} = 2$ and $W_{ij} = V_{ij} = 2$, "$Æ_{ij}$. By the inverse transformation from the given representation, we get the polynumber *A* as a more compact numerical set intended for storage or transmission. Restoration of the image (decompression) consists in making the procedure for representing the polynumber *A* in the same SNS.



$N_2 = 23 \rightarrow <23>_{00} = A$

Block diagram of a communication system based on a new data compression principle.



In the diagram: PU - partition unit, PNF - polynumber former, CSG - code sequence generator, TM- transmitter, PNSU – polynumber selection unit, PNRep – polynumber representation, IF – image former.

To compress halftone images, it is necessary to use a regular isotropic SNS with a base equal to the number of gray gradations.

Advantages of the proposed approach to data compression are:

1. The possibility of implementing progressive data(image) compression.
2. Potentially high compression ratio.
3. The possibility of lossless data compression.
4. The simplicity of the decoding (restoring) data algorithm on the receiving side.
5. The possibility of adaptive regulation of the transmitted (decoded) information volume depending on the permissible level of losses.

2. The method of calculating the two-dimensional convolution.

The basis for digital information processing is computational algorithms for convolution and discrete Fourier transform. The calculation of convolutions (one-dimensional and multidimensional) is currently carried out with the help of discrete Fourier transform algorithms, polynomial-theoretic and number-theoretic transformations. The main disadvantages of the Fourier transform are: the use of transcendental functions (sin and cos), the use of complex arithmetic even with real convolution. It leads to a doubling of the numerical fields dimension. Linear two-dimensional convolution in general form is expressed as

$$\sum\sum h(m_1, m_2)\, x(r_1-m_1, r_2-m_2) = y(r_1, r_2),$$

where $x(r_1-m_1, r_2-m_2)$ is the input two-dimensional sequence; h (m1, m2) is the two-dimensional impulse response of the system; y (r1, r2) is the output two-dimensional sequence.

We make use of the formal correspondence between the operations of two-dimensional convolution and the multiplication of two multinomials, on the one hand, and the one-to-one correspondence of products of multidimensional numbers (multinumbers and polynumbers) and their representations on the other. Then the calculation of the two-dimensional convolution can be performed in the 2D lattice SNS, carrying out instead of the discrete Fourier transform the transformation "2D representation of the number → polynumber", multiply the input polynomial and the impulse response, and then apply the transformation "resulting polynumber → 2D-representation of the result polynumber".

Advantages of the proposed method:

1. It does not require the use of complex quantities (spaces);
2. It does not use harmonic or special functions for the transformation;
3. It allows you to replace complex functional transformations with arithmetic ones;
4. Simplicity and clarity.

Conclusion

The Semantic Numeration Systems theory is at the initial stage of its development. Nevertheless, even now it can be assumed that SNS will be in demand in many areas related to the digital processing, among which are the following:

- cryptoprotection - the creation of fundamentally new cryptosystems to protect information of increased cryptographic strength;
- computer databases - compact representation, efficient storage and fast data transfer (exchange);
- geoinformation systems (GIS) - compact storage of digital terrain maps, their efficient transmission through communication channels;
- biometrics - effective identification of a person by fingerprints, the iris of the eye, photographs;
- medical technologies (tomography) - fundamentally new algorithms for 3D reconstruction;
- radars, sonars, and radio navigation - high-speed data processing;
- radio communication, including mobile communication - increasing the bandwidth of communication channels.

References

[1] Berthe V., Rigo M. (eds.) Combinatorics, Automata and Number Theory. CANT. – Cambridge University Press, 2010.
[2] Chunikhin A., Introduction to Multidimensional Number Systems. Theoretical Foundations and Applications, LAP LAMBERT Academic Publishing, 2012 (in Russian).
[3] Chunikhin A., Polymultisets, Multisuccessors, and Multidimensional Peano Arithmetics. – arXiv: 1201.1820v1.
[4] Dimitrov V., Jullien G., Muscedere R. Multiply-Base Number System: Theory and Application. – CRC Press, 2012.
[5] Fraenkel A., Systems of Numeration. – Amer. Math. Monthly, V.92, 1985, pp 105-114.
[6] Knuth D. E., The Art of Computer Programming, Vol.2. Seminumerical Algorithms. 3rd ed. – AW, 1989.
[7] Losev A.F., Dialectical Foundation of Mathematics. – M., Mysl, 1997 (in Russian).
[8] Rigo M. (ed.) Formal Languages, Automata and Numeration Systems 2. Applications to Recognizability and Decidability. – Wiley, 2014.

Submitted: September 26, 2018
Accepted: October 12, 2018

About the authors

Alexander Ju. Chunikhin, Candidate in Engineering, received the diploma of engineer in radio electronics from Higher School of Military Aviation Engineering (Kiev, URSS) in 1983. He received the PhD (Candidate in Engineering) degree from Higher School of Military Aviation Engineering (Kiev, Ukraine) in 1991. Currently he works as a Senior Researcher of O.V. Palladin Institute of Biochemistry (The National Academy of Sciences of Ukraine). His research interests include complex systems, Petri nets, number systems and numeration systems. He published more than 80 scientific papers, two monographs.

For citation

Alexander Ju. Chunikhin, Multidimensional Numbers and Semantic Numeration Systems: Theoretical Foundation and Application *JITA – Journal of Information Technology and Applications*, PanEuropien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 8(2018) 2:49-53, (UDC: 512.643:511.1), (DOI: 10.7251/JIT1802049C), Volume 8, Number 2, Banja Luka, december 2018 (45-96), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004

IMPLEMENTATION OF THE NEURAL NETWORK ALGORITHM IN ADVANCED DATABASES

Nedeljko Šikanjić[1], Zoran Ž. Avramović[2], Esad F. Jakupović[2]

[1]PhD Student at Pan-European University APEIRON, Banja Luka

[2]Professor at Pan-European University APEIRON, Banja Luka

Contribution to the state of the art

DOI: **10.7251/JIT1802054S** UDC: **004.738.5:551.588:551.506**

Abstract: The progress of humanity is closely related to the progress of technology. The highlight of the development of technology will occur when the machines are able to do what they learn and know; think and make decisions on their own without human help. In this paper we will try to analyze how neural networks work and how they will further develop in terms of application in advanced databases. We will also explore how one of the major IT companies is developing and continuing to use neural networks.

Keywords: Neural network, Advanced database, SQL.

INTRODUCTION

People owe their development to their intelligence and by coping the world around them by their implementation of technology. One of the more advanced attempts is to imitate human behavior and its implementation in technologies is an example of a neural network. The neural network in technology is based on the imitation of the human brain and how it functions. Although it has made a lot of progress, it is not even close to the functioning of the human brain, which remains the biggest challenge, but its behavior is improving over time.

NEURON NETWORK

The neural network was created in 1943 when Waren McCulloh and Walter Pitts created a model for neural networks based on a mathematical and algorithmic model. In their model, Alan Turing later proposed a model which represents a functional construction according to the exchange of information flowing to the input, condition and output.

This model also works with the human brain. [1] The estimation is that the human brain has about 100 billion neurons. Neuron consists of the body (soma), dendrita (inlays) and axons (exits).

The artificial neuron is created on the same principle. The artificial neuron functions in the way that at the input it receives the value that it sums up and then forwards it to the output. Here we can find a layer system - an input layer, where we can have multiple input points, hidden layers (which represent the body of a neuron) and an output layer, where we can have more output points. The relationships between neurons are represented by numbers. These numbers are defined as the weight. The weight principle is set so that if the weight is greater then the effect of neurons on the other is also higher.

The way a neural network works is that it is "trained" for the first time, for example to learn how information should be processed, and the second time it is "tested" or performs certain tasks. This method is called propagation in advance.

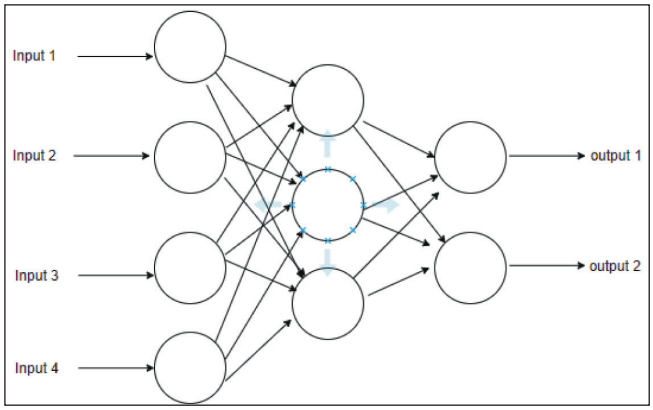


Image 1: *Neural network with hidden layer*

The way the neural network compares the output results through several iterations is called propagation backward. In this case, the information is returned to the neuron input. This is a system that people often use in attempts where, after a certain attempt, a person accurately or approximately accurately performs a certain operation.

Implementation of Algorithms of Neuron Networks in Basics - Sql Server

In this paper, we will work with an example of neural networks.What is amazing when it comes to machine learning is that it manages to transform us into adaptive thinkers [2]. We have created an example on a test dataset containing information about current trial subscription clients with their various attributes. Based on this set of data, we will find the probability of whether the possible client can become a paying customer through the algorithm. We will use a data set for displaying the neural network as it is implemented in Microsoft SQL Server Machine Learning Services along with the programming language R.

Microsoft made an important step towards statistical analysis as well as with data science [3]. Microsoft has made it possible to benefit from processing the data from where data are already located, in the database itself.

Some of the advantages of this approach are the following:

- The path to the data and its processing is reduced, thus improving security and facilitating access
- It is possible to obtain information in the short term
- The solution is scalable by being enabled as open source, so it can be applied in all versions that are offered
- New knowledge is being created that is already in place without having impact on the processing of existing data

Microsoft ML (Machine Learning) is a service responsible for transforming data or models that the R language and the Python language are processing [4]. The primary task is to enable data formatting. The functionalities provided by this service are the parallel execution, manipulation of data on disk and in memory as processing of huge amount of data.

The tasks or algorithms performed by the machine learning service are the following:

- The binary classification algorithm that teaches where one of the two classes of data instances belongs.

For example: It uses output of single two values such as 0 or 1 to detect true or false when search is given

$$R(\boldsymbol{x}) = P(+1|\boldsymbol{x}) - P(-1|\boldsymbol{x})$$

Formula 1: *Binary classifier [5]*

In the case of formula presented we use values of -1 and +1 where binary classifier for calculating probability is presented

- The multiclass classification algorithm learns to predict the categories to which the data belongs. The data are from 0 to k-1 where k represents the number of classes.

$$R_i(\boldsymbol{x}) = \frac{\sum_j a_{ij} P(j|\boldsymbol{x})}{\sum_j |a_{ij}| P(j|\boldsymbol{x})}$$

Formula 2: *Multi-class classification*

When we need more then single value as prediction we could use multi-class classification. Two sets are presented in this formula where *i* represents a row of matrix with partitioning of *j* row.

- A regression learning algorithm that teaches to predict the value of the dependent variable from a set of independent variables. The result that is given is a function used to display the value of a new data instance whose variables are not known.

- The anomaly detection algorithm that serves to identify which data or class of data does not belong to a given pattern. The sphere of application of these algorithms includes detection of fraud via credit cards, forecasting bankruptcy, estimating the value of mortgages or houses, checking the email whether it belongs to spam, etc.

Implementation use case

R language in combination with SQL Server objects such as Stored procedures is a very efficient and fast tool for work. By using this method, we shorten the way to the data, we do not have to think where to publish R scripts (for example, the web service) and we do not have to worry more about the performance since SQL Server will do it for us.

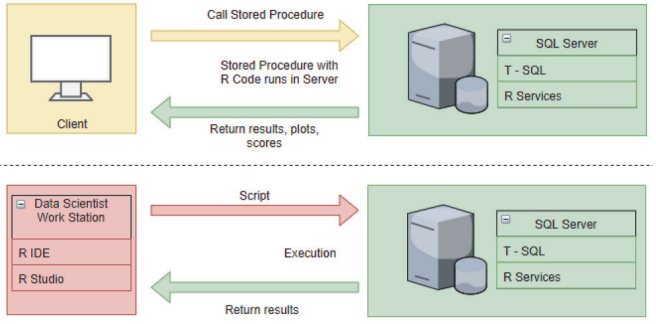


Image 2: *Diagram of the mode of operation and communication with SQL Server (R services)*

There are two options that we can use. One is to use the R Studio, which is very popular for the work of R language. Using libraries in R Studio, we can connect to SQL Server and execute the necessary commands and display the results in the R Studio. Another way is to use the SQL Server Management Studio environment, using functions and procedures, to process data using the R language in the form of enabled scripts, and to view data using various analysis and reporting tools that come with SQL Server.

In this example, we will use the Decision tree algorithm.

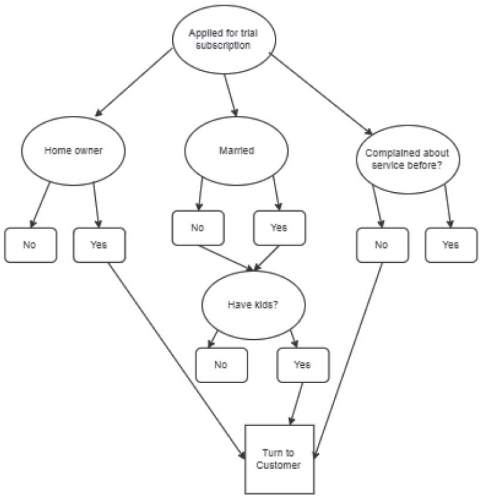


Image 3: *Decision tree diagram (partial) based on sample data*

As we can see on a decision tree diagram, we want to know, based on sample data, if this potential subscriber is a potential paying customer. With different sets of attributes we can determine if this is a valid model to predict an outcome with some percentage of certainty.

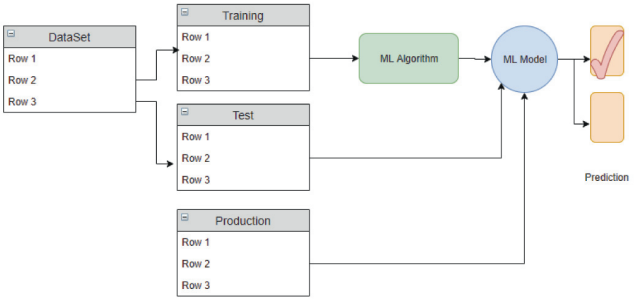


Image 4: *Training, testing, and prediction implementation diagram in SQL*

We will use a set of data to create a neural network. We will split this set of data into the training set and test set. The training set will use our algorithm to create a model and then use this model to test. Based on the testing, we can adjust our model so that we get the most accurate data. In the end, when we are satisfied with our model, then we can release it into production or production data to send it to us.

We will use a data model to predict whether a potential client will convert to a customer or not, based on a data set of some 25,000 data.

	Age	ClientId	CustomerSuspended	Education	Gender	HomeOwner	MaritalStatus	YearApplied	asMonthlySubScription
1	52	912	Yes	Bachelor or equivalent	Male	Yes	Married	2018	32
2	15	913	Yes	Bachelor or equivalent	Female	Yes	Married	2018	90
3	15	913	Yes	Bachelor or equivalent	Female	Yes	Married	2018	90
4	31	914	Yes	Master or equivalent	Female	Yes	Married	2018	97
5	31	914	Yes	Master or equivalent	Female	Yes	Married	2018	97
6	31	914	Yes	Master or equivalent	Female	Yes	Married	2018	97
7	61	915	Yes	High School or below	Male	No	Married	2018	35
8	61	915	Yes	High School or below	Male	No	Married	2018	35
9	77	916	Yes	High School or below	Male	No	Single	2018	43
10	77	916	Yes	High School or below	Male	No	Single	2018	43

In the data model, we keep data with all the necessary attributes on the basis of which we can come up with prediction of which attributes are crucial for converting the client into a permanent customer. Some of the attributes that exist in the set of data are the level of education, gender, whether the client is the owner of the house or real estate, marital status, etc.

```
USE JITA;
GO

DROP TABLE IF EXISTS ApeironNNModels;
GO
CREATE TABLE ApeironNNModels (
    [ID] VARCHAR(30) NOT NULL DEFAULT ('ModelType') PRIMARY KEY,
    [Value] VARBINARY(max) NOT NULL
);
GO
```

In order to be able to create a model in machine learning, we need to have a data set for training and a data set for testing.

The data were divided into two sets of data, a training data set and a test data set to a ratio of 70% in training and 30% in the test set of data [6].

```
DROP PROCEDURE IF EXISTS dbo.spCreateApeironNNModel;
GO
CREATE PROCEDURE dbo.spCreateApeironNNModel
AS
    EXECUTE sp_execute_external_script
    @language = N'R',
    @script = N'
    require("RevoScaleR")
    predictVar = "TurnToCustomer"
    idVar = "ClientId"
    removeVars = c(predictVar, idVar)
    trainVars <- rxGetVarNames(InputDataSet)
    trainVars <- trainVars[!trainVars %in% c(removeVars)]
    formula_string <- paste(c(predictVar, paste(trainVars, collapse = "+")), collapse = "~")
    formula <- as.formula(formula_string)

    decision_forest_model <- rxDForest(formula = formula,
                                       data = InputDataSet,
                                       nTree = 8,
                                       maxDepth = 32,
                                       mTry = 2,
                                       minBucket = 1,
                                       maxNumBins = 1001,
                                       replace = TRUE,
                                       importance = TRUE,
                                       seed = 8,
                                       parms = list(loss = c(0, 4, 1, 0)),
                    method="class" )
    OutputDataSet <- data.frame(payload = as.raw(serialize(decision_forest_model, connection=NULL)));',
    @input_data_1 = N'select * from trainData;'
    with result sets ((model varbinary(max)));
GO
```

In this stored procedure, we have an implemented part of the R language that creates the model over the variables we passed to them. This model is used by the decision tree, based on which we can get predictions for the data we send to them.

Regarding the decision tree, there is a function in the R language that we use as part of the SQL store procedure is rxDForest. We send the following parameters to it:

- Formula: the formula we pass to the columns that are key to the prediction model
- Data: data to be processed
- nTree: number of trees to which the model will grow

- maxDepth: maximum depth of trees used for calculating
- mTry: number of variables that are candidates for bending trees
- minBucket: number of observations in the node
- maxNumBins: control the maximum number of memory stores used for each variable
- replace: whether the observed variables will be replaced
- importance: a logical value that serves to evaluate the importance of the predictor
- seed: number used to initialize random numbers
- parms: additional parameters for bridging

```
USE JITA
GO

INSERT INTO ApeironNNModels ([value])
EXEC spCreateApeironNNModel;

UPDATE ApeironNNModels
SET [id] = 'ModelDecisionForest'
WHERE [id] = 'ModelType';

SELECT * FROM ApeironNNModels;
GO
```

The model is generated as binary data or output from stored procedure, which we will save in the table that will be used for sending parameters for test and production data.

```
CREATE PROCEDURE dbo.spPredictTurnToCustomerDecisionForest (@queryParam nvarchar(max))
AS
    declare @df_TurnToCustomer_model varbinary(max) =
        (select [value] from ApeironNNModels
        where [id] = 'ModelDecisionForest');

    EXECUTE sp_execute_external_script
    @language = N'R',
    @script = N'
    require("RevoScaleR")
    TurnToCustomer_model <- unserialize(df_TurnToCustomer_model)
    predictTurnToCustomer <- rxPredict(modelObject = TurnToCustomer_model,
                        data = InputDataSet,
                        type = "prob",
                        overwrite = TRUE)
    predictTurnToCustomer$X0_prob <- NULL
    predictTurnToCustomer$TurnToCustomer_Pred <- NULL
    names(predictTurnToCustomer) <- "TurnToCustomer_probability"
    threshold <- 0.6

    predictTurnToCustomer$TurnToCustomer_prediction <- ifelse(predictTurnToCustomer$TurnToCustomer_probability > threshold, 1, 0)
    predictTurnToCustomer$TurnToCustomer_prediction <- factor(predictTurnToCustomer$TurnToCustomer_prediction, levels = c(1, 0))

    OutputDataSet <- cbind(InputDataSet[, c("ClientId")], predictTurnToCustomer)
    ',
    @input_data_1 = @queryParam,
    @params = N'@df_TurnToCustomer_model varbinary(max)',
    @df_TurnToCustomer_model = @df_TurnToCustomer_model
    with result sets ((
                        ClientId int,
                        "TurnToCustomer_probability" float,
                        "TurnToCustomer_prediction" int));
GO
```

This stored procedure is used to predict based on the model we created.

The function we use in this stored procedure is rxPredict which serves to predict the results using the model we have already created.

The rxPredict functions are forwarded to the following parameters:

- Model: the model we previously created, in this case rxDForest

- Data: the data we use as a source for the prediction
- Type: prediction type, in our case we set the value "probe"
- Overwrite: a logical value that serves to overcome the output value

An example of passing test data is shown below in the SQL query.

```
USE JITA;
GO

EXEC dbo.spPredictTurnToCustomerDecisionForest
@queryParam = N'select * from testData;'
;
GO
```

Results | Messages

ClientId	TurnToCustomer_probability	TurnToCustomer_prediction
6799	0.25	0
6799	0.25	0
4642	0.625	1
4642	0.5	0
4946	0.125	0
4946	0.125	0
8831	0.111621660236196	0
8831	0.111621660236196	0

Based on the results we have received, we can further link this stored procedure to the reports to fine-tune the visual representation of the data, or we can further process the results both in the SQL environment and in the client application.

Generating visual representation of data

In order to create a visual representation of the results obtained, we can use the option within the SQL query.

It is important to note a hybrid approach using the R programming language within SQL Server where there are enough specific instructions to know.

For example, in order to generate plot diagrams in the R programming language, it is necessary to install certain libraries. For the plot diagram, we will use the ggplot2 library. The instruction for installing this library is as follows:

„install.packages("digest", dependencies = TRUE, lib = "PATH/Microsoft SQL Server/MSSQL14. MSSQLSERVER/R_SERVICES/library")“

In order to know which libraries we can use in the SQL environment, it is enough to execute a query to list all the libraries available in our SQL environment.

```
--install.packages("digest", dependencies = TRUE, lib = "C:/Program Files/Microsoft SQL Server/MSSQL14.MSSQLSERVER/R_SI

EXECUTE [sys].[sp_execute_external_script]
 @language = N'R'
 ,@script = N'AllPackages <- as.data.frame(installed.packages())
 OutputDataSet <- AllPackages[c("Package", "LibPath", "Version", "Priority", "Depends", "Imports", "LinkingTo", "Sugge
WITH RESULT SETS((PackageName NVARCHAR (100),
  LibPath NVARCHAR (255),
  Version NVARCHAR (20),
  Priority NVARCHAR (20),
  Depends NVARCHAR (255),
  Imports NVARCHAR (100),
  LinkingTo NVARCHAR (100),
  Suggests NVARCHAR (100),
  Enhances NVARCHAR (100),
  OS_type NVARCHAR (100),
  License NVARCHAR (100),
  Built NVARCHAR (100)))
```

Results Messages

	PackageName	LibPath	Version	Priority	Depends	Ir
16	ggplot2	C:/Program Files/Microsoft SQL Server/MSSQL14.MSSQLSERVER/R_SERVICES/library	2.2.1	NULL	R (>= 3.1)	c
17	ggplot2OLD	C:/Program Files/Microsoft SQL Server/MSSQL14.MSSQLSERVER/R_SERVICES/library	2.2.1	NULL	R (>= 3.1)	c
18	graphics	C:/Program Files/Microsoft SQL Server/MSSQL14.MSSQLSERVER/R_SERVICES/library	3.3.3	base	NULL	g
19	grDevices	C:/Program Files/Microsoft SQL Server/MSSQL14.MSSQLSERVER/R_SERVICES/library	3.3.3	base	NULL	n
20	grid	C:/Program Files/Microsoft SQL Server/MSSQL14.MSSQLSERVER/R_SERVICES/library	3.3.3	base	NULL	g
21	iterators	C:/Program Files/Microsoft SQL Server/MSSQL14.MSSQLSERVER/R_SERVICES/library	1.0.8	NULL	R (>= 2.5.0), utils	n
22	jsonlite	C:/Program Files/Microsoft SQL Server/MSSQL14.MSSQLSERVER/R_SERVICES/library	1.4	NULL	methods	n
23	KernSmooth	C:/Program Files/Microsoft SQL Server/MSSQL14.MSSQLSERVER/R_SERVICES/library	2.23-15	recommended	R (>= 2.5.0), stats	n

There are several options to generate a visual representation of the results obtained. One of them is through a SQL query.

```
execute sp_execute_external_script
@language = N'R'
, @script = N'

# Find the numeric columns and create a data frame
numeric_cols <- sapply(InputDataSet, is.numeric)

library("ggplot2")
library("reshape2")
cdrpivot <- melt(InputDataSet[, numeric_cols], id.vars = c("TurnToCustomer"))

iteo_image_file = "c:\\temp\\iteo_visual_representation.jpg";
jpeg(filename = iteo_image_file, width=1000, height = 1466);

print(ggplot(aes(x = value,
          group = TurnToCustomer,
          color = factor(TurnToCustomer)),
       data = cdrpivot) +
       geom_density() +
       facet_wrap(~variable, scales = "free"))
dev.off();
OutputDataSet <- data.frame(data=readBin(file(iteo_image_file, "rb"), what=raw(), n=1e6));
'
, @input_data_1 = N'
    SELECT [Age]
      ,[YearlyIncome]
      ,[MonthlySubscription]
      ,[TurnToCustomer]
      ,[Education]
      ,[Gender]
      ,[HomeOwner]
      ,[MaritalStatus]
```

In this SQL query, we use the R language programming commands to generate an image using the plot function to generate a diagram. Data is transmitted using the parameters used in the generic plot diagram. We can save this created diagram on a disk or in the database.

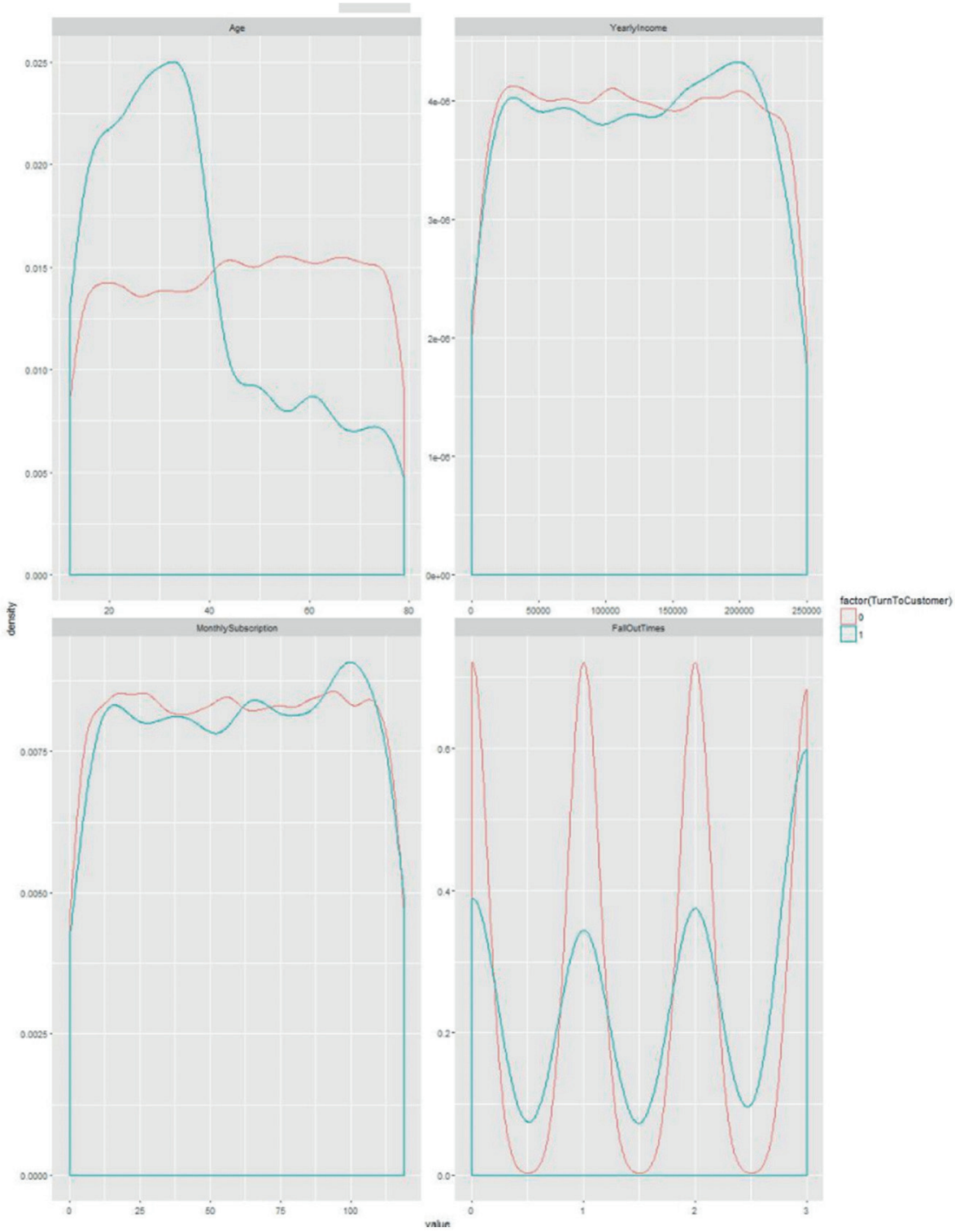


Image 3.2: *Generated plot diagram using SQL query*

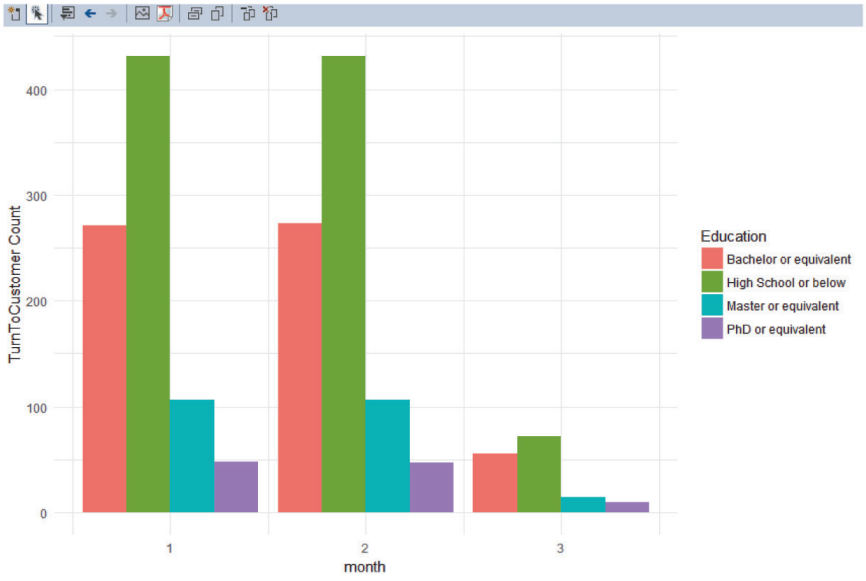
Another option is using Visual Studio. In addition to the existing project types, Visual Studio also has a project for the R programming language. Through the R tool, we can perform various commands, including data visualization.

JitaSqlWith.R

```
#Prepare SQL query to process the data
sqlQueryText <- "SELECT [Month], [Education], SUM([TurnToCustomer]) as TurnToCustomer
    FROM PotentialCustomerData
    GROUP BY [Month], [Education];"

# Set up a data source object based on the new query
pcdMonthTurnByEducationDataSet <- RxSqlServerData(
  sqlQuery = sqlQueryText,
  connectionString = sqlConnString
)

# Import the data from SQL Server to a local data frame
pcdMonthTurnByEducationDataFrame <- rxImport(pcdMonthTurnByEducationDataSet)

# Plot the data frame from SQL
ggplot(pcdMonthTurnByEducationDataFrame,
       aes(x = Month, y = TurnToCustomer,
           group = Education, fill = Education)) +
           geom_bar(stat = "identity",
           position = position_dodge()) +
           labs(x = "month", y = "TurnToCustomer Count") +
           theme_minimal()
```

Data is being prepared through the functions RxSqlServerData that enable loading the data into a data-set object, which then imports the local data frame.

After that we use the ggplot function by passing the data set object, defining what we are showing on the X and Y axes, along with other parameters such as labels and setting the theme of the diagram.



The diagram with the data that is generated can be further exported to other formats such as pdf, images or metadata [7].

Conclusion

In this paper we have shown how the neural network functions in advanced databases, which are the advantages of using neural networks in advanced databases and what is its further development. It is also important to know that the largest IT companies such as Microsoft have decided to implement their focus on the application of a neural network in a similar way, that large processing of data which the client stations find considerably difficult to process due to resources can move within the database and can thus apply security of access to data and use resources that their application algorithms or engine engine possess. For us as humans it is important to continue to strive to integrating different systems in order to unlock the knowledge that we could not imagine before.

References

[1] N. J. Nilsson, Introduction to Machine Learning, Robotics Laboratory, Department of Computer Science, Stanford University, 1998.
[2] Denis Rothman, Artificial Intelligence By Example: Develop machine intelligence from scratch using real artificial intelligence use cases, ISBN-13: 978-1788990547, May 2018
[3] Tomaz Kastrun, Julie Koesmarno, SQL Server 2017 Machine Learning Services with R: Data exploration, modeling, and advanced analytics, ISBN-13: 978-1787283572, February 2018
[4] https://blog.statsbot.co/bayesian-learning-for-statistical-classification-f2362d620428 accesed on 10.11.2018
[5] Raghav Bali, R Machine Learning By Example, ISBN-13: 978-1784390846, March 31, 2016
[6] https://docs.microsoft.com/en-us/sql/advanced-analytics/?view=sql-server-2017accessedon 15.11.2018

Submitted: November 21, 2018
Accepted: December 24, 2018

About the authors

Nedeljko Šikanjić holds a Magister degree in Informatics and Computer Science and has worked for more than 15 years as a Software and Database Architect/Engineer. His main fields of studies are in the area of advanced Databases and Software Architectures. He has been a holder of active Microsoft Certified Trainer Certificate since 2012 and has been teaching courses on various topics in Information Technologies.

Zoran Ž. Avramović was born in Serbia (Yugoslavia) on September 10th, 1953. He graduated from the Faculty of Electrical Engineering, University of Belgrade. At this Faculty he received a Master's degree, and then a PhD in technical sciences. He is:

- Academician of the Russian Academy of Transport (RTA, St. Petersburg, Russia, since 1995),
- Academician of the Russian Academy of Natural Sciences (RANS, Moscow, Russia, since 2001),
- Academician of the Yugoslav Academy of Engineering (YAE, Belgrade, Serbia, since 2004) (today: Engineering Academy of Serbia, EAS)
- Academician of the Academy of Electrotechnical Sciences of the Russian Federation (AES of the Russian Federation, Moscow, Russia, since 2007)
- Scientific Secretary of the Electrical Engineering Department of the Engineering Academy of Serbia.

Esad Jakupović, was born on 21 September 1950 in Ćeli, municipality of Prijedor, where he finished elementary and high school. He graduated from the Faculty of Industrial Pedagogy - the direction of physics in Rijeka in 1975. Postgraduate studies completed at the Faculty of Agricultural Sciences of the University of Zagreb in 1984 in the field of Mechanization of Agriculture. He defended his doctoral dissertation at the Faculty of Mechanical Engineering of the University of Ljubljana in 1991.

Since 2005 he has been employed as a regular professor at the Pan-European University "APEIRON" Banja Luka. At the same University, he served as Vice-Rector for Teaching, and the Vice President for Postgraduate and PhD Studies in the period 2005-2014.

He worked as the rector of the Pan-European University "APEIRON" from 2014 to 2018.

For citation

Nedeljko Šikanjić, Zoran Ž. Avramović, Esad F. Jakupović, Implementation of the Neural Network Algorithm in Advanced Databases, *JITA – Journal of Information Technology and Applications*, PanEuropien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 8(2018) 2:54-63, (UDC: 004.738.5:551.588:551.506), (DOI: 10.7251/JIT1802054S), Volume 8, Number 2, Banja Luka, december 2018 (45-96), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004

E-MAIL FORENSICS: TECHNIQUES AND TOOLS FOR FORENSICINVESTIGATION OF ONE COURT CASE

Ljubomir Lazić

Faculty Of Information Technology, Metropolitan University, Belgrade ljubomir.lazic@metropolitan.ac.rs

Case study

DOI: **10.7251/JIT1802064L** UDC: **004.42:004.738.5**

Abstract: E-mail has emerged as the most important application on the Internet for communication of messages, delivery of documents and carrying out transactions and is used not only from computers, but many other electronic gadgets such as mobile phones. This paper is an attempt to illustrate e-mail architecture from forensics perspective. Also, this paper projects the need for e-mail forensic investigation and lists various methods and tools used for its realization. A detailed header analysis of a multiple tactic spoofed e-mail message is carried out in this paper. It also discusses various possibilities for detection of spoofed headers and identification of its originator. Furthermore, difficulties that may be faced by investigators during forensic investigation of an e-mail message have been discussed along with their possible solutions. Our focus is on email header analysis phase offered by the tools. We examine the capability of a particular tools such as EmailTrackerPro and aid4mail in action. The paper describes the court case of cyber crime, the so-called identity theft in Internet communication via electronic mail by two business entities. Identity theft of e-mail addresses and false communications with a foreign company was carried out in order to indicate that a cash transaction of around EUR 100,000 was paid to the account of NN attackers and not to the account in the domestic Serbian bank.

Keywords: E-mail forensic, header analysis, E-mail message as evidence.

INTRODUCTION

Modern time communication is impossible without emails. In the field of business communication, emails are considered as its integral part. At the same time, emails are also used by criminals [1,2,4]. In digital forensics, emails are considered as evidence and Email Header Analysis has become important to collect evidence during forensics process [2,3]. Email clients are computer programs that allow users to send and receive emails. Over time, different types of email clients have been invented for the convenience of email users. We will discuss different types of email clients now. Broadly, email clients are divided into two types based on email saving location. These are web-based email clients and desktop-based email clients.

a) Web-based Email Clients: Web-based email clients save all their data to their web server. Some web-based clients are Gmail, Yahoo Mail, Hotmail, etc. The benefit of using web-based email clients is that they can be accessed from anywhere in the world, using Username and Password. One of their disadvantages is the users not knowing where their data is being stored.

b) Desktop-based Email Clients: Desktop-based email clients are the opposite of web-based clients. Outlook, Thunderbird, Mail Bird are some examples of desktop-based email clients. All data of desktop-based web browser is stored in the system of its users. Thus, users do not have to worry about data security. The same point can be considered as a disadvantage in some cases. This is especially the case when it is used in criminal activities, and the evidence cannot be collected from the server [3,5]. E-mail messages include transit handling envelope and trace information in the form of structured fields which are not stripped after messages are delivered, leaving a detailed record of e-mail transactions. A detailed header analysis can be used to map

the networks traversed by messages, including the information on the messaging software and patching policies of clients and gateways, etc. Over a period of year's e-mail protocols have been secured through several security extensions and producers, however, cybercriminals continue to misuse it for illegitimate purposes by sending spam, phishing e-mails, distributing child pornography, and hate e-mails besides propagating viruses, worms, hoaxes and Trojan horses. Further, Internet infrastructure misuse through denial of service, waste of storage space and computational resources are costing every Internet user directly or indirectly.

E-mail forensic analysis is used to study the source and content of e-mail message as evidence, identifying the actual sender, recipient and date and time it was sent, etc. to collect credible evidence to bring criminals to justice [1-5]. This paper is an attempt to illustrate e-mail architecture from forensics perspective. It describes roles and responsibilities of different e-mail actors and components, itemizes metadata contained in e-mail headers, and lists protocols and ports used in it. It further describes various tools and techniques currently employed to carry out forensic investigation of an e-mail message.

This paper projects the need for e-mail forensic investigation and lists various methods and tools used for its realization. A detailed header analysis of a multiple tactic spoofed e-mail message is carried out in this paper. It also discusses various possibilities for detection of spoofed headers and identification of its originator. Furthermore, difficulties that may be faced by investigators during forensic investigation of an e-mail message have been discussed along with their possible solutions [1,5].

This paper will also discuss tracing e-mail headers and issues associated with it. It will address both HTTP & SMTP initiated e-mails. It will discuss different ways used by e-mail senders to evade tracing and workarounds used by investigators to combat them. It will also discuss advanced measures and techniques used by investigators to track emails [4]. We will discuss particular tools in the paper, such as: *EmailTrackerPro* and *aid4mail* in action.

E-mail Service Architecture

E-mail system comprises of various hardware and software components that include sender's client and server computers and receiver's client and server computers with required software and services installed on each. Besides these, it uses various systems and services of the Internet. The sending and receiving servers are always connected to the Internet but the sender's and receiver's client connects to the Internet as and when required [2,3]. E-mail is a highly distributed service that involves several actors which play different roles to accomplish end-to-end e-mail exchange [2]. These actors fall under three groups, namely User Actors, Message Handling Service (MHS) Actors and ADministrative Management Domain (ADMD) Actors. User Actors are Authors, Recipients, Return Handlers and Mediators that represent people, organizations or processes that serve as sources or sinks of messages. They can generate, modify or look at the whole message. Message Handling Service (MHS) Actors are Originators, Relays, Gateways and Receivers which are responsible for end-to-end transfer of messages. These Actors can generate, modify or look at only transfer data in the message. ADministrative Management Domain (ADMD) Actors are Edges, Consumers and Transits which are associated with different organizations and have their own administrative authority, operating policies and trust-based decision making [2].

E-mail system is an integration of several hardware & software components, services and protocols, which provide interoperability between its users and among the components along the path of transfer. The system includes sender's client and server computers and receiver's client and server computers with required software and services installed on each of them. Besides, it uses various systems and services of the Internet [2].

The sending and receiving servers are always connected to the Internet but the sender's and receiver's client connects to the Internet as and when required. An e-mail communication, for example, between a sender 'Alice' having e-mail address 'alice@a.com' and recipient 'Bob' having e-mail address 'bob@b.com' is shown in Figure 1.

'Alice' composes an e-mail message on her computer called client for 'Bob' and sends it to her sending server 'smtp.a.org' using SMTP protocol. Sending server performs a lookup for the mail exchange record of receiving server 'b.org' through Domain Name System (DNS) protocol on DNS server [3] 'dns.b.org'. The

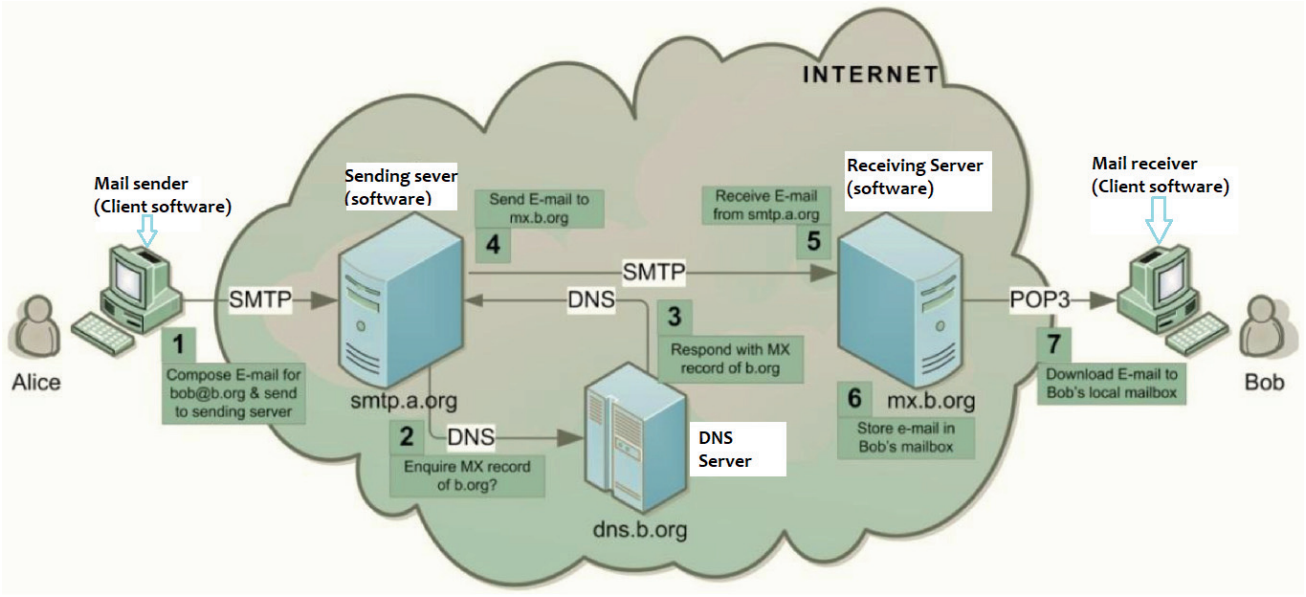


Figure 1. *E-mail communication between a sender 'Alice' and recipient 'Bob' [3]*

DNS server responds with the highest priority mail exchange server 'mx.b.org' for the domain 'b.org'. Sending server establishes SMTP connection with the receiving server and delivers the e-mail message to the mailbox of 'Bob' on the receiving server. 'Bob' downloads the message from his mailbox on receiving server to local mailbox on his client computer using POP3 [3] or IMAP [1] protocols. Optionally, 'Bob' can also read the message stored in his server mailbox without downloading it to the local mailbox by using a Webmail program.

E-mail system is an integration of several hardware and software components, services and protocols which provide interoperability between its users and among the components along the path of transfer. The e-mail architecture shown in Figure 2 below specifies the relationship between its logical components for creation, submission, transmission, delivery and reading processes of an e-mail message. Several communicating entities called e-mail nodes which are essentially software units working on application layer of TCP/IP model are involved in the process of e-mail delivery. Nodes working on lower layers such as routers and bridges which represent options to send e-mail without using SMTP

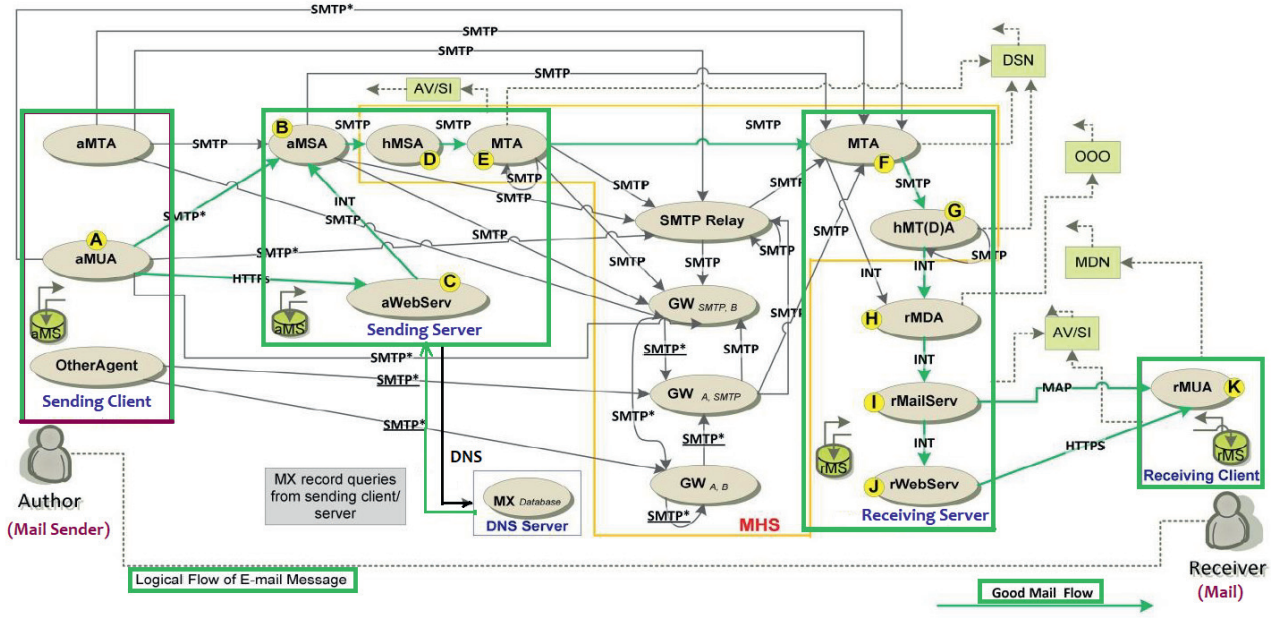


Figure 2. *E-mail Architecture [3]*

are not considered in this architecture because almost all e-mail communication uses SMTP directly or indirectly. Moreover, proprietary nodes used for internal deliveries at sending and receiving servers are also not considered in this architecture.

A mail message from Author to Receiver that traverses through aMUA, aMSA, hMSA, MTA (outbound), MTA (Inbound), hMDA, rMDA, rMailServ and rMUA is considered as good mail by the Sender Policy Forum (SPF). Mails following through other paths are either fully or partially non-SMTP based or uses non-standard transfer modes which are often suspected to contain viruses and spam. Delivery Status Notification (DSN) messages are generated by some components of MHS (MSA, MTA, or MDA) which provide information about transfer errors or successful deliveries and are sent to MailFrom addresses. Message Disposition Notification (MDN) messages are generated by rMUA which provide information about post-delivery processing are sent to Disposition-Notification-To address. Out Of Office (OOO) messages are sent by rMDA to return address [3].

E-mail forensic investigation techniques

E-mail forensics refers to the study of source and content of e-mail as evidence to identify the actual sender and recipient of a message, data/time of transmission, detailed record of e-mail transaction, intent of the sender, etc. This study involves investigation of metadata, keyword searching, port scanning, etc. for authorship attribution and identification of e-mail scams.

Various approaches that are used for e-mail forensic are described in [1] and are briefly defined below. E-mail forensic include header analysis, bait tactics, server investigations, and network device investigation. Besides mandatory headers, custom and MIME headers appearing in the body of the message are also analysed for sender mailer fingerprints and software embedded identifiers.

Email Forensics Analysis Steps

A forensic investigation of e-mail can examine both email header and body. This paper will look at header examination.

According to [3] an investigation should have the following:

- Examining sender's e-mail address
- Examining message initiation protocol (HTTP, SMTP)
- Examining Message ID
- Examining sender's IP address

Some other aspects that controls forensics step include the following properties (see Figure 3):

1) Storage format of email: Server side storage format may include maildir (each email is kept separate in a file, for each user), mbox format (all email files are in a single text file). Server-side stores email in SQL Server databases. Reading different types of formats can be done for forensics analysis by using notepad editor and applying regular expression-based searches [5]. At the client-side, an email is stored as mbox format (Thunderbird) [5]. Client side may also store emails as .PST (MSOutlook), and NSF (Lotus Notes) files.

2) Availability of backup copy of email: When checking from the serve side, all copies are transferred to the client. This requires seizing the client computer. For webmail, copies are always saved at the server side [4].

3) Protocol used to transport email: Email can be initiated and transported based on SMTP or HTTP [2] depending on the email server applications.

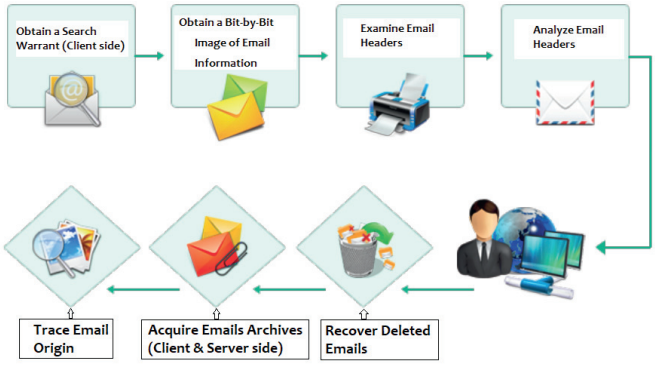


Figure 3. *Broad steps in email forensics for investigator*

Header Analysis

Meta data in the e-mail message in the form of control information i.e. envelope and headers including headers in the message body contain information about the sender and/or the path along which the message has traversed. Some of these may be spoofed to conceal the identity of the sender. A detailed analysis of these headers and their correlation is performed in header analysis. Besides header analysis, various other approaches that can be used for e-mail forensics include bait tactics, server investigations, and network device investigation. Cus-

tom and MIME headers appearing in the body of the message are also analysed for sender mailer fingerprints and software embedded identifiers [2].

Relevance of Headers & Components

Email header forensics basically denotes the examination done on the email message body and the source and path followed by it. This also includes the identification of genuine sender, time, or recipient of the emails. The email header forensic analysis can bring out the candid evidences from various components included in the header part. Let us see Figure 4 which components are helpful for header forensics:

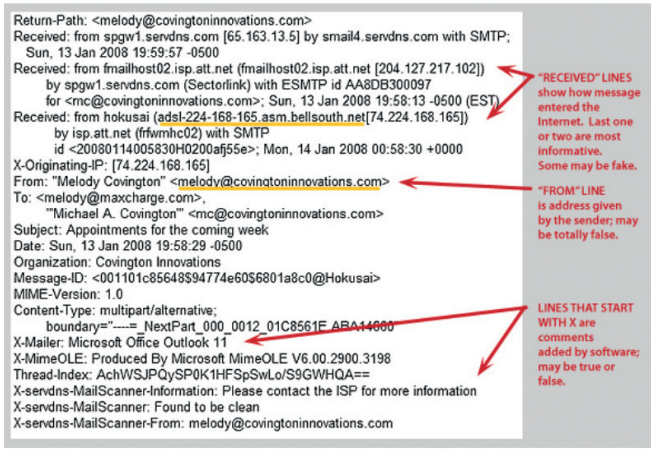


Figure 4. *A typical E-mail header*

`X-Apparently-To:` It will reveal recipient's email address while investigating. This can be the validation field for checking email service provider. Generally this field is referred to as "BCC, CC, or To" and is not restricted to "To".

`Delivery To:` This shows the address of the auto-mailer.

`Return-Path:` This field is used for the bounces of email messages. In case the mail server is sending the message and it cannot be delivered.

`Received-SPF:` During email header forensics, this field shows the information of email service used for the sending of mails. It is also having an ID number which is important for log examination for determining the validity of an email. In case of unavailability of the ID, the email must have been spoofed.

`Message ID:` This is a globally used unique identification ID which refers to the genuine time of the emails and version of message. It is highly important to know if investigators want to know whether spoofing is done to the email or not.

`MIME Version:` It stands for Multipurpose Internet Mail Extensions and is an Internet Standard which extends format of message.

`Content-type:` This shows the type of content or format used for the message like; XLML, Text, or HTML.

`X-Mailer:` It displays the email client which is used for sending the message.

`X-Originating-IP&Received:` This is an important field for tracing the IP address used for sending the email. This is the most important message when it comes to the email header forensic analysis as it has to be examined where the mail arrived from.

`DKIM-Signature:` This field stores the signature of an email and all key-fetching information in simple "tag=value" syntax. It is a crucial field to validate the domain name and identity allied to the message via cryptographic authentication.

SECURITY ISSUES IN INTERNET E-MAIL:

A. Secrecy: The content of email is in plain text format. While it is transmitting it never decrypted, so data can be easily revealed if one can get access of your mailbox and one can knows how to tap network and flow.

B. Integrity: Integrity means changes the original data. Email is mainly stored in plain text and also transmitted in plain text. Therefore, anyone can easily hack the way of email transmission and change the original data without being noticed by sender and receiver.

Security Issues In SMTP

Security in information technology is defined as to protect information against unauthorized revelation as well as unauthorized modification. The user needs to take care about possibility of malicious and fraudulent attacks by hackers as well as impact of viruses and denial-of-services attack. Some approaches that are useful for security of your system include:

A. Authentication

The technique can be used to identify and verify if anyone is seeking to access un authorized system.

B. *Access control*

Users can be restricted to ensure they only access data and services for which they have been authorized.

C. Encryption

Techniques that scramble data are used to protect information while data are transmitted over network.

D. Firewall

Firewall is mainly used to differentiate the internal and external information access. Firewall prevents the outsiders to access information within organization.

E. Intrusion detection

Techniques that monitor the system and network to check whether anyone is trying to access network without authentication.

F. Anti-virus software

It can detect viruses and prevent access to infected files.

The Threats to Email Security

A. Viruses

Email security contains multiple issues. Virus is the highest risk issue in network. Virus has capability to destroy complete data at a time. When virus is found in any email it can be bring down the entire mail system, often in a large amount in a single mail.

Many issues can affect the system but virus is stronger than any other. Virus stays long and destroys data immediately. It is not removed by any antivirus product. Virus leaves its impact for a long time and the recovery takes a large amount of money, resources and efforts as well as lost computer information.

B. SPAM

SPAM is another major issue in network security. Viruses and SPAM go hand in hand. Spam is also known as junk email. SPAM mail contains malicious code which affects mail system immediately. SPAM mail contains virus which can bring down the entire system. Users cannot request any mail but them getting number of mails of unintended user which can be a SPAM mail. Mail filtering cannot filter legitimate email from SPAM. Virus and SPAM have negligible difference.

Experiment: Man-in-the-Middle Attack

The main purpose of this experiment is to demonstrate the concept of the man-in-the-middle attack, the attacker being an NN person. This experiment is aimed at capturing data from a suspected user to connect to a WLAN and viewing unauthorized content that certainly happened in this court case. The experiment shows that the unauthorized content accessed by the suspicious user can be collected and can be used for a digital forensic investigation. The reader should take into account that all three actors in this experiment, i.e. router, attacker and legitimate user (see Figure 5), all at the same network address, i.e. 146.64 with the remaining two numbers indicating the address of each host in the network.

Execution of the experiment

In this experiment, a forensic researcher points out that the traffic for this experiment was not encrypted. The D-Link router is configured to be open, which means that no encryption keys such as WEP, WPA2, and WPS are configured.

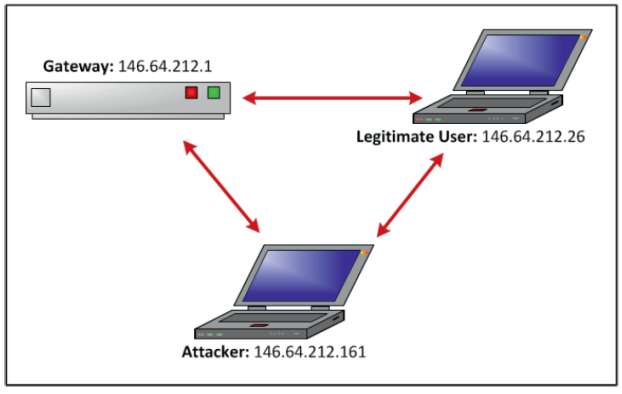


Figure 5. *Participants in an identity theft experiment*

In spite of this, the experiment would continue to be successful, even if encryption is established, although in this case more efforts should be made to crack the passwords first, but it should be emphasized that communication encryption continues to

be present as the greatest enemy of forensic scientists.

The main idea of this experiment is that the attacker uses an ARP spoofing mechanism to convince the legitimate user that they are a legitimate participant, device gateway [4]. After the response of a legitimate user, the attacker immediately confirms to the gateway that they are a legitimate user. Both the legitimate user and the gateway will think they have established a relationship with each other, and in fact they have both established a relationship with the attacker. This means that the gateway and legitimate user traffic is directed towards an attacker who can then intercept the communication between the two sides. For the purpose of this experiment, the attacker is only interested in the traffic of a legitimate user suspected of being searched for IMPORTANT online content.

EXAMINING E-MAIL FORENSIC TOOLS: CASE STUDIES

Email analysis, as we already mention, is the task performed in the network forensics. Email analysis is the process which involves analysis of emails sent and received at different ends. In current era, there are very less ways to analyse emails. Most widely accepted method is the Manual Method of Email Analysis [4,5]. Although there have been many attempts into securing e-mail systems, most are still inadequately secured. Installing antiviruses, filters, firewalls and scanners is simply not enough to secure e-mail communication. Some common examples of illegitimate uses of emails are spam, phishing, cyber bullying, botnets, disclosure of confidential information, child pornography and sexual harassment. The anonymity factor of e-mail has made it difficult for digital forensic investigators to identify the authorship of an email, and to aggravate this problem further; there is no standardised procedure to follow.

Therefore, a forensic investigator needs efficient tools and techniques to perform the analysis with a high degree of accuracy and in a timely fashion. It is evident that an email forensic tool may only assist the investigator during a specific stage of analysis [4,5].

While preforming manual method for email analysis, we try to spot spoofed messages which are sent through SMTP (Simple Mail Transfer Protocol). By analysing them we can decode the message being sent. After decoding, all IP addresses are analysed and their location is traced. A timeline of all event is made (in universal standard time) and is checked further for suspicious behaviour. Server logs are checked at the same time to ensure that all the activities are mentioned in the timeline so formed. If any suspicious activity is found, the mails are recovered and can be used as evidence against the sender. Email is extracted from the client server which keeps a copy of sent mails until a specific number.

First case study

First, we will describe a well-known case in court practice i.e. a case study involving the use of Manual Method for Email Analysis [4] using a whaling attack which is a spear-phishing attack directed specifically at high-profile targets like C-level executives, politicians and celebrities:

- An email attached to a $20 million dollar lawsuit purported to be from the CEO of "`tech.com`" to a venture capital broker. The message outlined guaranteed "warrants" on the next round of finding for the broker.
- "`tech.com`" filled counter claim and claimed the email was *forgery*. Their law firm engaged a team to determine the validity of the message.
- The team imaged all of the CEO's computers at his office and his home. Email server backup tapes were recalled from the client servers.
- All hard drivers and email servers were searched for "questioned" message. There were no traces of any such mail on any of the hard drive or mail spool.
- When the time stamps and message id's were compared with the server logs then it was found that the "questioned" message have not gone through either "`tech.com`'s" webmail or mail server at the time indicated by the date/time stamp on the message.
- Based on the analysis the defendants filed motion to image and examine broker's computers.
- Federal judge issued subpoena and the team arrived at the broker's business, he refused to allow his system to image.
- Broker's lawyer went into the state court, on a companion case, and got the judge to issue an order for a new court appointed examiner.
- The examination revealed direct proof of the

alteration of a valid message's header to create a "questioned" email.

The allegedly received email

The header of a problematic e-mail is presented as follows.

```
Return-Path: CEO Good_Guy@tech.com
Received: from mail.tech.com (mail.tech.com
[201.10.20.152])
by hedgefund.fund.com (8.11.0/8.11.0) ESMTP id
e73MfZ331592; Thu, 3 Aug 2000 15:45:31 -0400
Received: from webmail.tech.com (webmail.
tech.com
[10.27.30.190]) bymail.tech.com (Switch-2.0.1/
Switch-2.0.1) ESMTP id e73MfW903843; Thu, 3
Aug 2000 14:41:32 -0500
Received: from tech.com (ostrich.tech.com
[10.27.20.190])
by webmail.tech.com (8.8.8+Sun/8.8.8) with
ESMTP id RAA01318; Thu, 3 Aug 2000 14:41:31
-0500
content-class: urn:content-classes:message
Subject: Warrants on $25 Million Funding
Date: Thu, 3 Aug 2000 14:43:47 -0500
MIME-Version: 1.0
Content-Type: application/ms-tnef;
name="winmail.dat"
Content-Transfer-Encoding: binary
Message-ID: <3989e793.87BDEEE2@tech.com>
X-MS-Has-Attach:
X-MS-TNEF-Correlator:     <3989e793.87BDEEE2@
tech.com>
Thread-Topic: Warrants on $25 Million Funding
Thread-Index:     AcHatCZUSkaLe0ajEdaelQACpY-
cy8A==
From: "CEO Good_Guy@tech.com" <ceo_good_guy@
tech.com >
To: "Bad_Guy_Broker" <bad_guy@fund.com>
```

Information contained in the header can aid investigators in tracing the sender of the e-mail. A thorough investigation of e-mail headers should include examination of the sender's e-mail address and IP address, examination of the message ID as well as the messaging initiation protocol (HTTP or SMTP). To determine the source of the e-mail, *investigators must first examine the received section at the bottom of the header and work their way up in a bottom to top approach*.

It is also important that e-mail cases examine the logs of all servers in the received chain as soon as possible. Time is very important in e-mail cases as HTTP and SMTP logs are archived frequently; especially by large ISPs. If a log is archived, it could take time and effort to retrieve and decompress the log files needed to trace e-mails. Some e-mails have fake/forged headers in order to deceive investigators, so extreme caution and careful scrutiny should be practiced in investigating every part of the e-mail header.

However, this is quite a bit long and tiring procedure which would involve too many mails to be analysed, which would be excessively time-consuming. Time being the most expensive entity, we need to save the time as much as we can. To save this time certain tools are present which helps to reduce the work burden. So, we need a software tools, such as eMailTrackerPro (http://www.emailtrackerpro.com/) and Aid4Mail Forensic (http://www.aid4mail.com/) that are discussed in the next section.

In this case investigator should look at ESMTP id which is a unique identification assigned by each intermediate relay or gateway server. This id is usually in a hexadecimal string that is reset each day. Resulting in an id that can be resolved to a time window on a particular server. The investigator should also compare the header information against server logs: `webmail@tech.com`. Analysis of the webmail server logs revealed several issues regarding the validity of the suspect message:

- Matching trace header timestamps and ESMTP ids revealed that RAA01318 was issued at 17:41:31 to the authentic message
- Comparing the 14:41:31 timestamp of the suspect message with the log revealed the server was assigning ESMTP ids beginning with "OAA" not "RRA" as represented in the header.

Analysis of the mail server logs confirmed that the suspect message was not authentic:

- Matching trace header timestamps and ESMTP ids revealed that the authentic Message-ID was logged at 17:41:32 and assigned ESMTP id e73MfW903843 then it was sent to the `hedgefund@fund.com` server and it was assigned a new ESMTP id e73MfZ331592
- Comparing the 14:41:32 timestamp of the suspect message with the log revealed there were no messages for over an hour during that time frame.

Second case study

This section describes the court case of cybercrime so called "identity theft in Internet communication by electronic mail by two business entities". Based on the analysis of the method of communica-

tion (e-mails, SMS messages and voice), languages in business correspondence, frequency of transactions, problems in business, ways of solving them in over 100 collected e-mails in communication between two companies during three years of successful cooperation, the author of the work came to indisputable indicators of cybercrime [4]. Identity theft of e-mail addresses and false communication with a foreign company was carried out in order to indicate that a cash transaction of around EUR 100,000 was paid to the account of NN attackers in the London bank, and not to the account in the domestic Serbian bank to which the money was paid up to then in the process of electronic payment of goods and services between the parties to the dispute. The process of examining e-mails is described using the *eMailTrackerPro* tool in the event of identity theft by an NN person (attacker, hacker), an e-mail forensic investigation plan, restrictions, an attacker detection process as the third NN person in an email communication, **Man-in-the-Middle Attack experiment** that served as the basis for forensic analysis of e-mail in the case study. As for this case, it is necessary to see from which address the hacker sent a message, and through which hopes (jumps through the Internet) a message was sent to reach its destination, as can be seen in the following Figure 6 [4].

Address of Hop	Name of Hop	Location
192.168.0.1		(Private)
10.41.0.1		(Private)
185.89.137.165		*Australia*
185.89.136.22		*Australia*
212.73.241.201		*Italy*
4.69.142.225	ae-2-13.bear1.Italy2.Level3.net	*USA*
212.133.7.34	MC-LINK-SPA.bear1.Italy2.Level3.net	*Slovakia*
213.21.130.38		*Italy*
77.43.83.155	net77-43-83-155.mclink.it	*Italy*
213.203.157.195	mail.cinellipiumini.com	**Italy**

Figure 6. *Hopes through which the hacker's mail passed*

As far as the hopes through which the message goes, we can see that it is a little unusual that everything is going from Italy, going to the server in Slovakia, to the US (forged email address of `xxxxx@yahoo.com`), then back to Italy and then to Australia. The following Figure 7 will show the path on the map as the message was traveling.

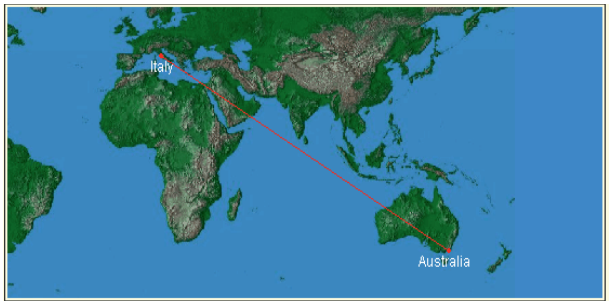


Figure 7. *Path on the map the message travelled*

After this knowledge, it was necessary to analyse other suspicious e-mails, as well as the email server on the victim's side, as we have described earlier. It was found that during the time of the hacker attack, the actual sender did not send any messages. There are many tools which may assist in the study of source and content of e-mail message so that an attack or malicious intent of the intrusions may be investigated. This section introduces some of these tools: eMailTrackerPro and Aid4Mail Forensic.

Software eMailTrackerPro [5] is a proprietary email forensic solution that analyses email files stored in local disk and supports automatic email analysis for the identification of spamming incidents. eMailTrackerPro is capable of recovering the IP address that sends the message along with its associated geographical location (city) to determine the threat level or validity of an e-mail message. It can find the network service provider (ISP) of the sender. A routing table is provided to identify the path between the sender and receiver of an email. It also can check a suspected email against Domain Name Server blacklists to safeguard against spam.

The *disadvantage* associated with this software is that it would be unable to find a spammer which is not blacklisted into its database.

Add4Mail forensic software tool

This is another tool developed for helping in the mail sorting purpose only. This software can find emails which can be searched by any particular keyword. As with *EmailTrackerPro* and on this tool, we need to configure our mail. Let us choose which mail we will use for analysis. In this case, we will use *gmail*. Once we have completed the mail configuration, we are going to the next step that allows us to select the time frame in which we want to search for mail by keywords, and in the window where Vaccky,

VacckY, etc., are located. It is actually a keyword search box as in Figure 8.

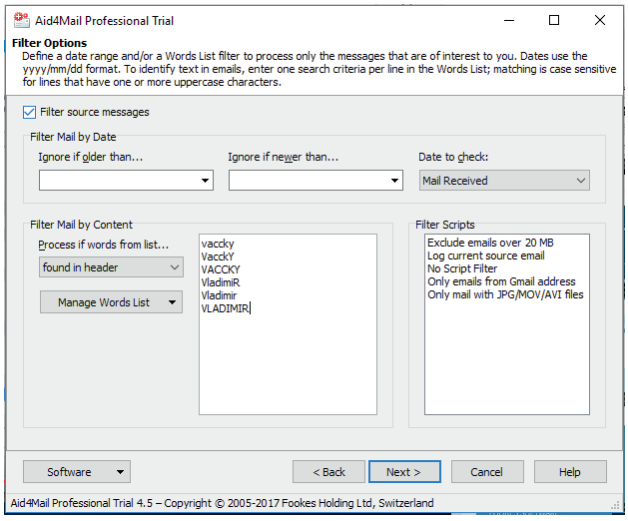


Figure 8. *Example: Search keywords for a mailbox by Add4Mail*

The output provided by this software program is the message written in the email along with the date, time and other information specific to the mail as in Figure 9. This software program can also be used to fetch some deleted mails from their trash folder. Unlike email tracker pro, this tool does not only serve to track the message, but also for detailed forensic mail analysis. This tool can be found at http://www.aid-4mail.com/, but unlike *EmailTrackerPro* it is not an open source, you must actually purchase a license.

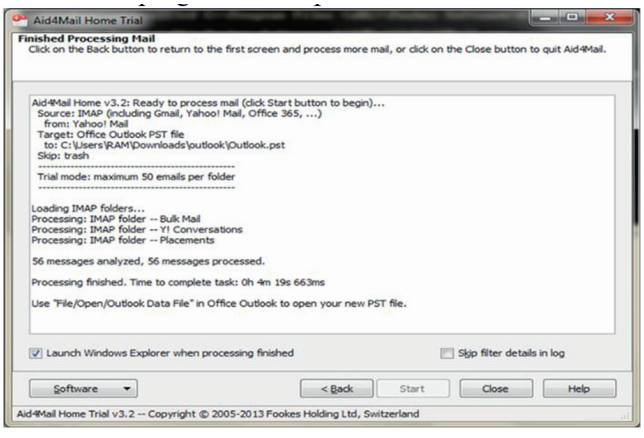


Figure 9. *Example of processing mail by Add4Mail*

The major *disadvantage* of this software is that it can only find keywords that the user searches. It has no artificial intelligence and therefore is a completely manual software program developed to sort and find mails.

Conclusion

Digital forensic analysis is a complex and time-consuming process which involves the analysis of digital evidence. Emails might contain valuable information that could lead investigators to the identity and/or location of the offender. Additionally, email forensic tools through email header analysis may even reveal information related to the host machine used during the composition of the message. In this paper, we have discussed key information related to email forensic analysis as well as important aspects of header tracing. Finally, we have demonstrated two forensic tools that can be utilised for email analysis emphasising on their key features in an effort to assist investigators in the selection of the appropriate tools.

High-tech crime, also known as e-crime or cyber-crime, includes a set of offenses that involve the use of the Internet, a computer, or some other electronic device. This paper describes the court case of cyber-crime, the so-called identity theft in Internet communication via electronic mail by two business entities. Based on the analysis of the method of communication (e-mails, SMS messages and voice), languages in business correspondence, frequency of transactions, problems in business, ways of solving them in over 100 collected e-mails in communication between two companies during three years of successful cooperation, the author of the research came to indisputable indicators of cyber-crime. Identity theft of e-mail addresses and false communications with an Italian firm was carried out in order to indicate that a cash transaction of around EUR 100,000 was paid to the account of NN attackers in the London Bank, and not to the account in the domestic Serbian bank to which the money had been paid by then in the process of electronic payment of goods and services between the parties to the dispute.

Acknowledgements

This research was financially supported by the Ministry of Science and Technological Development of the Republic of Serbia, as part of the TR35026 "Software Environment for the optimal management of the quality software development process" project.

REFERENCES

[1] Al-Zarouni M (2004) Tracing E-mail Headers, Australian Computer, Network & Information Forensics Conference, pp. 16–30.
[2] Banday MT (2011) Analysing E-Mail Headers for Forensic Investigation, Journal of Digital Forensics, Security and Law, Vol. 6(2).
[3] Banday MT (2011) Techniques and Tools for Forensic Investigation of E-mail, International Journal of Network Security & Its Applications, Vol. 3, No. 6.
[4] Lazic Lj (2018) E-Mail Forensics: The Case From The Court Practice Of Theft Of Identity, Conference: ITeO2018, 28. September, Banjaluka, pp. 368- 383.
[5] Mrityunjay UC et al. (2017) Novel Approach for Email Forensics, International Journal of Engineering Research & Technology (IJERT), Special Issue.

Submitted: November 10, 2018
Accepted: November 23, 2018

ABOUT THE AUTHORS

Ljubomir Lazić was born on December 18, 1955. He is software engineering and computer science professor at METROPOLITAN University, Belgrade, Serbia. He received the bechelor degree in electrical engineering from School of Electrical Engineering, Belgrade University in 1979. He was a Post-Doctoral Researcher at The WSEAS (The World Scientific and Engineering Academy and Society) of computer science from 2009 to 2010. He successfully defended PhD thesis: "Integrated and Optimized Software Testing Process" in January, 2007 at University of Belgrade, Faculty of Electrical Engineering.

So far, he have authored over 100 research papers. Courses teach: Software Engineering, Software Project Management, Software Testing, Human Computer Interaction, Component Based Engineering. Current research interests are: Optimal software project management, Software Metrics, Effort Estimation Modeling etc. He continue to serve industry in a variety of roles, including consulting, executive education, and expert testimony.

FOR CITATION

Ljubomir Lazić, E-Mail Forensics: Techniques And Tools For ForensicInvestigation Of One Court Case, *JITA – Journal of Information Technology and Applications*, PanEuropien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 8(2018) 2:64-74, (UDC: 004.42:004.738.5), (DOI: 10.7251/JIT1802064L), Volume 8, Number 2, Banja Luka, december 2018 (45-96), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004

Psychological Connection between Colors and Certain Characteristic Terms

Nedim Smailović
Pan-European University APEIRON, Banja Luka

Case study
DOI: **10.7251/JIT1802075S** UDC: **159.937.51.072**

Abstract: This paper presents results of a research on psychological connection between 40 offered colors and 91 terms from everyday life. Similar researches have been conducted and published in a number of instances in domestic and foreign bibliography, but this research has certain particularities that are not often present in other articles. For one thing, colors whose associativity with certain terms is being analyzed are shown in a table, their name and code in the RGB system of color marking are provided. Colors were not merely described, e.g. blue, green, yellow, etc., as it may be confusing or lead to misinterpretation of answers since there are many degrees of blue color, for example. In the second part of the poll, the subjects answered the Ishihara test, in order to check the ability to correctly interpret the colors. The third specific is triple visual interpretation of received results using colored graphs.

Keywords: color, Ishihara test, psychology, visualization of data.

Introduction

The electromagnetic waves spectrum spreads in a large range of frequencies i.e. wavelengths. At one end of the spectrum is gamma radiation (gamma particles) with wavelength < 0,02 nm and frequency > 15·1018 Hz, and at other end are radio waves with wavelength of 1 mm – 100 km and frequency 300 GHz – 3 kHz. The electromagnetic spectrum can be divided into seven bands, which are the following (from highest to lowest frequencies i.e. shortest to longest wavelengths):

1. Gamma rays, 2. X – rays, 3. Ultraviolent radiation, 4. Visible light, 5. Infrared radiation, 6. Microwaves, and 7. Radio waves.

Visible light is only the segment of electromagnetic spectrum containing waves of wavelength 390 nm – 750 nm and frequency 770 THz – 400 THz. Humans have receptors (eyes) for these waves and at their detection after passing through optical nerve, the brain creates the sense of receiving the light.

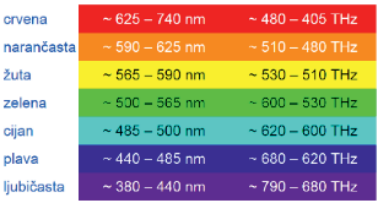
The eye retina contains two types of cells sensitive to light, rod and cone cells. Rod cells react even to the small quantity of light, but do not differentiate colors. That is why at night we only see the shapes of objects. Poets say: "Dawn is the birthtime of colors". That is when the light gets stronger and is received by cone cells. There are three kinds of cone cells and each recognizes one of three primary colors – red, green or blue. These signals are combined in the brain, enabling recognition of a whole spectrum of colors. Retina has about 130 million rod cells and 6,5 million cone cells. They are mainly all situated in the area called the macula lutea, where an up-side-down image of the observed scene is created.

Wavelengths of the visible part of electromagnetic spectrum are mixed in many combinations, and human eye can differentiate over 10 million colors. When a Sun light passes through a transparent glass prism, certain wavelengths diffract differently, creating a spectrum of colors.

Larousse encyclopedia interprets colors in this way:

"A color is an impression created by electromagnetic waves as they pass through our eye; a property

ascribed to light or objects to create a special visual impression." [9]



Picture 1. *Names of colors, their wavelengths and frequency ranges*

The word "color" (in the sense of a matter or a visual feeling) was not separately defined in the *Leksikon stranih reči i izraza*, by Milan Vujaklija, Prosveta – Belgrade, 1980, (Lexicon of Foreign Words and Expressions), or in *Rječnik stranih riječi: tuđice, posuđenice, izrazi, kratice i fraze* (Dictionary of Foreign Words) by Sima Anić, Nikola Klaić, Želimir Domović - Zagreb: Sani-plus, 1998. As if it were not of foreign origin, while it is. In the dictionary "Turkish Words in Serbo-Croat Language" by Abdulah Skaljic, it states that the word color (boja) has its origin in Turkish (boya) in the sense of farba (paint), though that word is Germanism. Such a frequent word in everyday speech, and we do not have our own term for it. Only in Russian and Bulgarian language we can still find Slavic words 'cvet' and 'краска'. The notion of colors may be connected with old words 'mast' ('masnica' – bruise) and 'шаръ' (colorfulness, to draw). However, the word color has stayed unchanged from Turkish in everyday speech, just as there are no adequate substitute for some other words of Turkish origin: bakar, čarapa, česma, dugme, jastuk, jogurt, kaiš, limun, makaze, testera, tavan, kreč, šećer, sapun, majmun, rakija, čelik, kičma, pare, puž, komšija, kesa, kašika, đubre… (eng. copper, socks, tap, button, pillow, yogurt, belt, lemon, scissors, (wood)saw, attic, chalk, sugar, soap, monkey, spirit drink, steel, spine, money, snail, neighbor, bag, spoon, rubbish…).

From a wider perspective, the notion of a color has several different meanings:

1. color as a psychological phenomenon, i.e. as a stimulus in an eye and a subjective perception;
2. color as an optical phenomenon;
3. color as means of expression;
4. color as a matter.

In everyday speech the word color is often accompanied with attributes such as: (of) light color, dark color, calm color, dead color, vivid color, joyful color, etc.

Isaac Newton (1642 -1727) experimentally proved in 1676 that white light separates into seven different colors, if passed through a glass prism. White and black are not among those colors. White color is obtained by mixing all the colors of that spectrum.

Newton also experimented with associations of colors and tones of a musical scale, while a century later Johann Wolfgang Goethe (1749-1832) studied psychological effects of colors. Colors are inseparable from everyday life situations. In business world, colors are often an important segment of presenting and doing business. Well selected colors or their combinations may be an important factor of market success, while a bad choice may produce negative effect, fully opposite from the desired one. It is particularly important when planning an appearance on markets of different cultural, religious or geographic communities. Observers of different age, ethnicity, gender or local communities have different perceptions of particular colors. Often the symbolic meaning ('the language') of colors is changing.

SYSTEMATIZATION OF COLORS

Thomas Young and Hermann von Helmoltz established in early 19th century that any color of light may be obtained by mixing the three primary light sources. That meant that colors first needed to be systematized. Many great scholars wrote about systematization of colors, and among them were: Leonardo da Vinci (1452.-1519), Sir Isaac Newton (1642-1727), Johann Wolfgang Goethe (1749-1832), James Clerk Maxwell (1831-1879), Albert Henry Munsell (1858-1918), Wilhelm Oswald (1853-1932).

There are several different interpretations of primary colors (primaries), which is a consequence of the fact that colors may be observed from different standpoints: psychological, physics, chemical, etc. The science of colors is a true example of interdisciplinary area.

Physics standpoint interprets behavior of colored rays of light and their mixing. In that sense the primaries are additive: red (R), green (G) and blue (B), or subtractive: cyan (C), magenta (M) and yellow (Y). Subtractive primaries are complementary colors to corresponding additive primary colors.

Psychological standpoint interprets colors based on their noticing and perception of color.

Hence, the psychological primaries colors were defined: red, green, blue, yellow, white and black.

The painters will emphasize, among their primaries, those colors they use as a matter (paint in a tube) to mix and get all other colors. **Painters' primaries** are red, blue, yellow, white (noncolor) and black (noncolor). White and black noncolors are used to obtain the desired degree of saturation of mixed colors.

The Art primaries are the same as Painters' primaries, excluding white and black noncolor.

The theory is even more complicated by the fact that painters' primary "red" is not necessarily the same as the additive primary "red".

Mixing the primaries gives secondary colors, while further mixing of secondary colors produces tertiaries, etc.

Colors rarely appear isolated in nature and almost always combine mutually next to each other. There we have harmony of colors. Just as some paired sounds may be in harmony or disharmony, two or more colors next to each other may look nice but they can also produce a completely different feeling. All this is, of course, subjective, as something that looks nice to some may not look like that to others, since the perception of colors is very complex.

International Associations for Light and Colors Research

For the purpose of interpreting all aspects of color research, the AIC – The International Color Association (https://www.aic-color.org/) was established. The AIC aims to have close cooperation with existing international organizations, such as Associations with topics relating to light: CIE, Commission internationale de l'éclairage; ICO, International Commission for Optics; ISO, International Organization for Standardization; SPIE, The International Society for Optical Engineering; AISV, Association internationale de semiotique visuelle; Applied Vision Association (Great Britain); Royal Photographic Society (Great Britain); IS&T, Society for Imaging Science & Technology (USA); ICC, International Color Consortium; ICVS, International Colour Vision Society; IACM, International Association of Color Manufacturers, Centre d'information de la couleur (France); Fédération française de la couleur (France), Ad chroma (France); SDC, Society of Dyers and Colourists (Great Britain); Color Marketing Group (USA); China Fashion Colour Association (China).[1]

Picture 2. *Logo of the AIC*

Picture 3. *Logo of the International Color Day*

In 2009 the AIC accepted founding the International Color Day, which is commemorated in many countries around the world. The establishing of the international color day (ICD) is considered justified as the color, thanks to the visual perception, is one of the most influential phenomenon in life of people and one of ways that contribute most to the perception of reality. It could be said that the contemporary world communicates through colors.

The ICD was proposed by the Portuguese organization for colors in 2008. The proposal was agreed in 2009 among members of national associations in more than 30 countries.

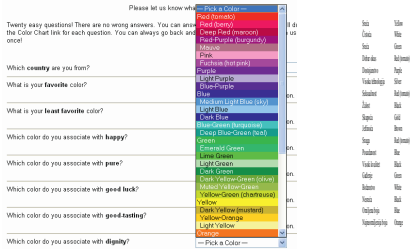
Some of the activities and events that take place on ICD day are: art exhibitions, architectonic projects, design, decoration, fashion, meetings, discussions, scientific events, workshops on the use of color and light for children and adults, competitions, etc.

Subjective Perception of Colors

The Internet address http://express. colorcom. com until recently held a comprehensive interpretation about colors, dealing with topics such as: colors and science, colors and computers, colors and the world, etc. There also was, inter alia, a survey where site visitors could say something about associativity of a term – notion and a color. There were 18 questions asked. Colors for answers are shown visually (as displayed in the picture) in order to avoid misunderstanding regarding which color it is in the words

description. The same color may be associated to different terms.

The authors stated that over 30.000 persons took part in the survey, and they published the following results on associating colors with terms:



Picture 4. *An Internet research on associating colors with certain terms*

SURVEY ON SUBJECTIVE PERCEPTION OF COLORS

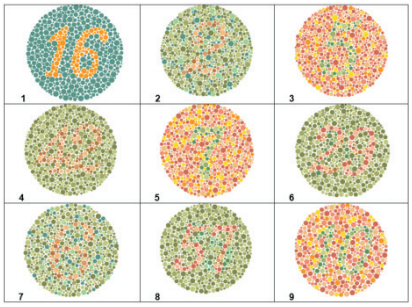
The polled subjects had 40 concrete colors in front of themselves, which belong to Microsoft standard palette of colors with a unique code to differentiate the colors among themselves. For example, one green-blue color was titled teal, it was number 14 and a code in the RGB system was 000 255 204. The subject saw exactly that color, and not some abstract green-blue. According to the opinion of the polled subjects, that color best corresponds to some of the offered 92 terms.

Picture 5. *A survey form with colors and associating terms*

The second part of the survey was Ishihara test.

Dr. Shinobu Ishihara (1879 - 1963) was a Japanise military surgeon and an ophthalmologist who created a test for color-blindness, called after him afterwards. The test comprises of recognizing a number or a pattern in a set of differently colored circles.

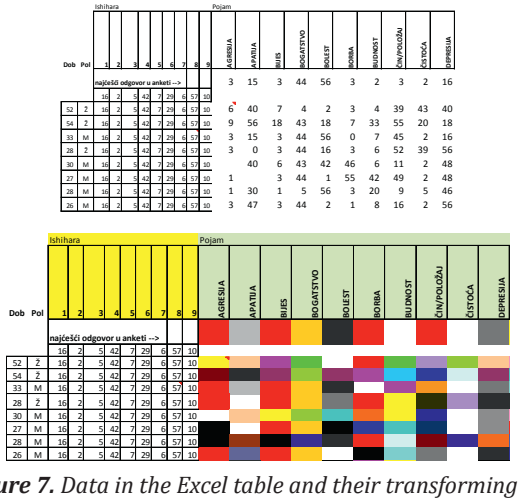
The subjects were tasked to recognize the number in the given circles and write down the result in the provided table.



1	2	3
4	5	6
7	8	9

Picture 6. *Ishihara test*

The third specific of this work is the processing of the obtained results. The survey results were imported in the Excel table. The subjects wrote down the color under its given number, and in Excel that number in the table was again presented in the same color of the color from the answer. In that way the research results were visualized. We used the opportunity of visualization by chart type Treemap, which was presented only in the latest version of the program MS Excel 2016.



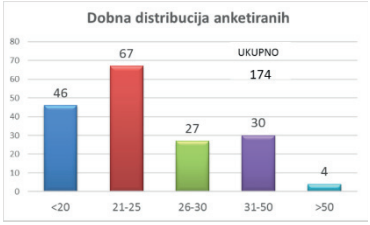
Picture 7. *Data in the Excel table and their transforming into corresponding colors*

Survey Results

Average age of the polled subjects	25,8 years
Total number of polled subjects:	174
No. of men	147
No. of women	27
Incorrect answers in Ishihara test (dichromatism)	12 (7%)

Dischromatism (Greek dys-, chroma, color, opsis – vision, seeing), medically – color blindness, inability of the sense of vision to perceive all specter colors equally. [10].

Normal color vision is technically called normal trichromacy.

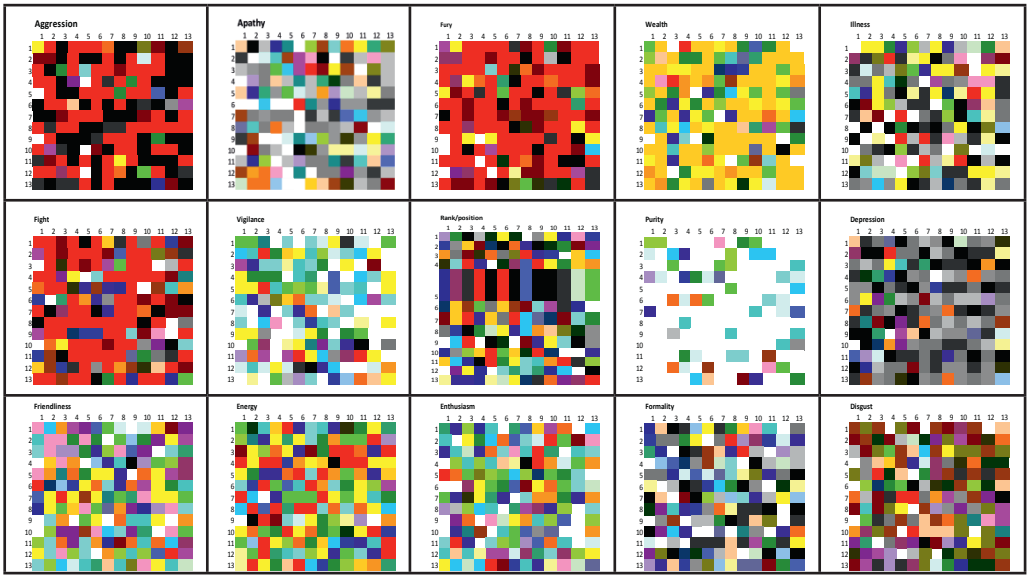


Picture 8. *Chart and a table of age distribution of subjects*

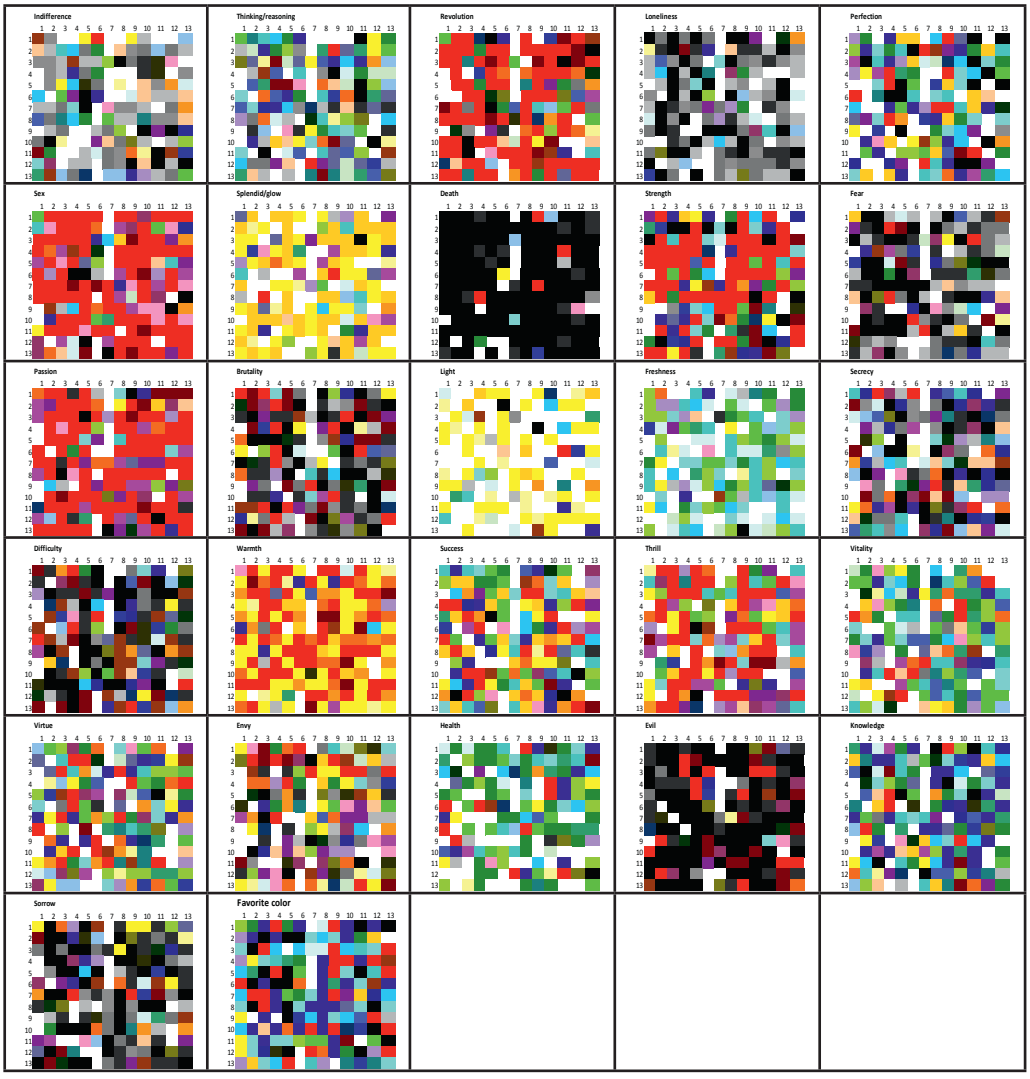
The following picture shows answers received in the survey, each in the 13x13 square. Every square presents 169 color associations with a certain term. Even in such a relatively small sample, we could notice interesting results. Domination of certain colors are very present with certain terms. The most prominent examples are with terms: anger (red), wealth (yellow), hygiene (white), nature (green), revolution (red), death (black) …

There also are surprising examples, where terms such as aggression, anger or death are associated with yellow color, or terms isolation, death and fear are associated with green, or cleanliness is associated with brown or dark-blue color. A detailed analysis showed that one person that did not answer correctly to Ishihara test (does not recognize colors properly) unexpectedly associated terms of negative connotation (apathy, illness, disguise, stupidity, fear, deceit, poison, infertility…), as a rule, with yellow color.

Survey results were obtained used the features of MS Excel (cell format and conditional formatting). Graphic designers might find an inspiration in these examples, e.g. combination of colors with the term energy, concentration, luxury, gentleness…



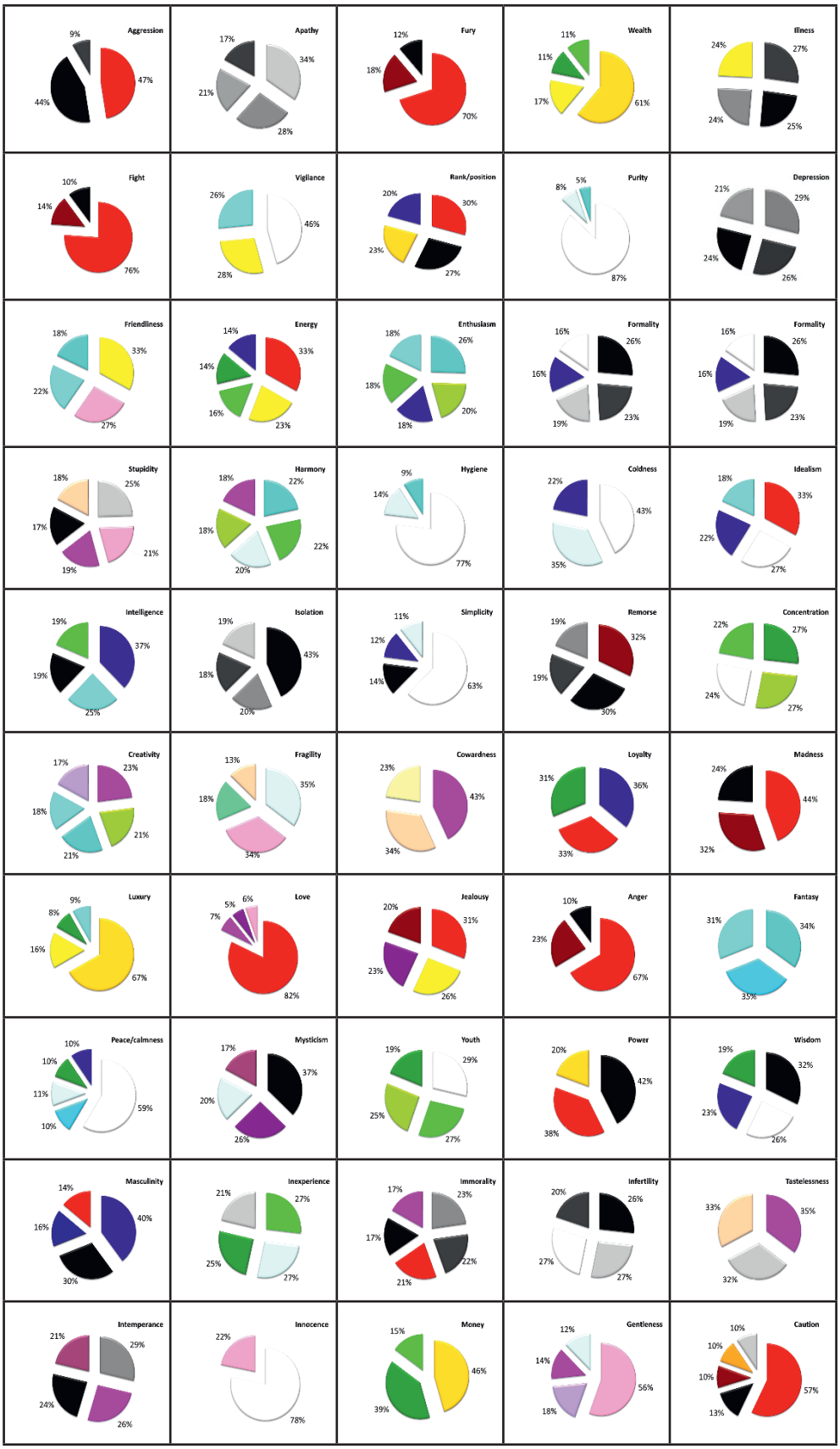

Stupidity
Harmony
Hygiene
Coldness
Idealism
Intelligence
Isolation
Simplicity
Remorse
Concentration
Creativity
Fragility
Cowardness
Loyalty
Madness
Luxury
Love
Jealousy
Anger
Fantasy
Peace/calmness
Mysticism
Youth
Power
Wisdom
Masculinity
Inexperience
Immorality
Infertility
Tastelessness
Intemperance
Innocence
Money
Gentleness
Caution
Optimism
Poison/venom
Seriousness
Fertility
Incentive
Specialty
Fairness
Justice
Emptiness
Refinement
Exaggeration
Deceit
Nature
Joy
Birth

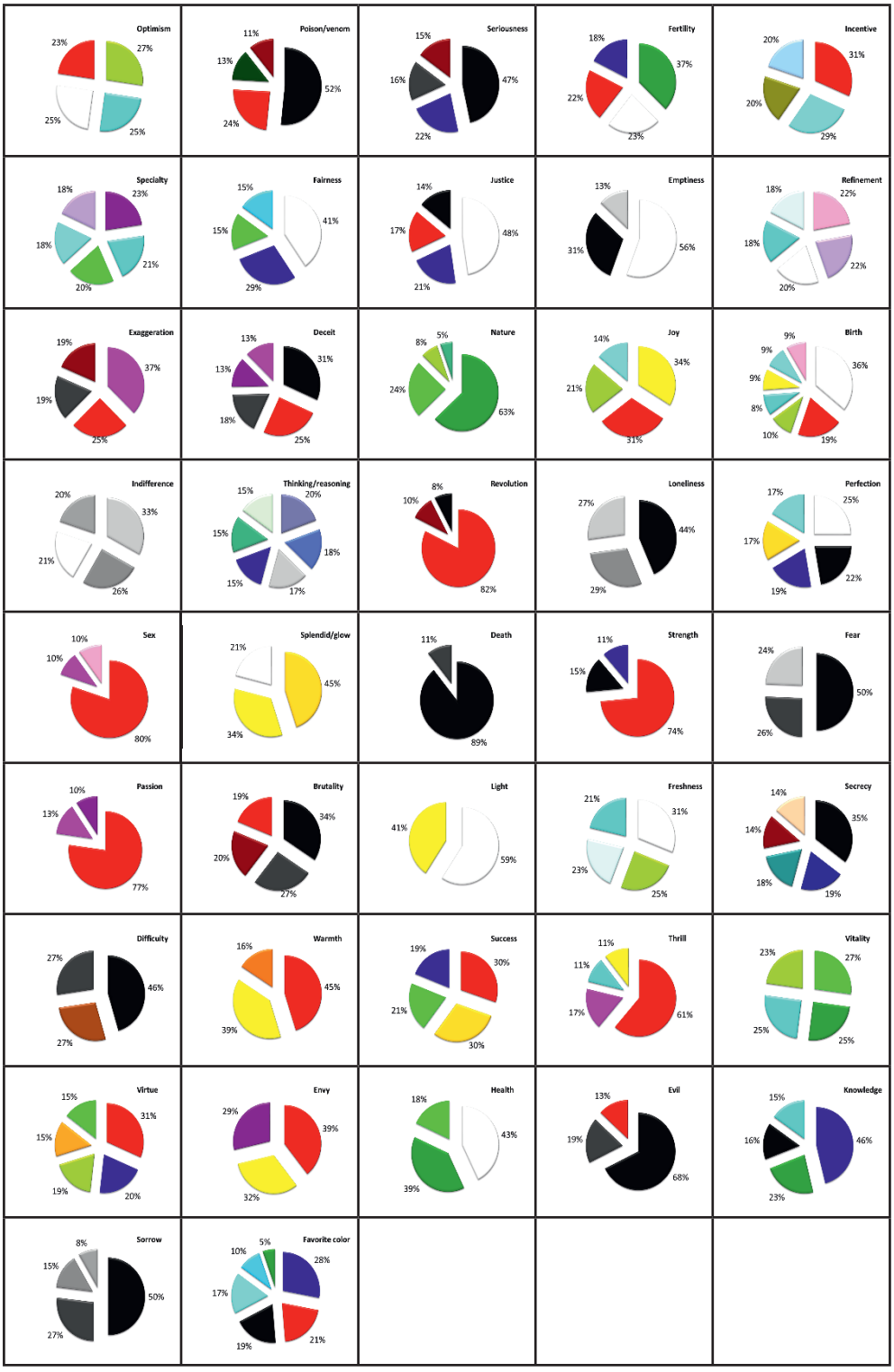


Picture 9. *Overview of all answers to the associating colors with certain terms*

The following picture shows the results of the same survey in the Pie Chart type of chart. For each term the most associated colors were presented, from two to six dominating ones. The given percentages show only the mutual relation of dominant colors. It is noticeable that terms with negative psychological prefix: apathy, depression, isolation, infertility, emptiness, loneliness, death, fear, remorse, had only shades of grey among dominating colors. Some terms with clearly understood meaning have only two or three dominating color associations, such as: aggression, anger, fight, purity, hygiene, innocence, emptiness, revolution, death, light, while other colors were very scarce. More (four to six) dominant color associations have terms of less specific meaning, such as: enthusiasm, immorality, thinking, birth, etc.

One of the most interesting results in this reserach is the analysis of the answer to the question about the favorite color. It could be expected it would be blue. That color was among the most popular ones in other works as well. The second place went to red color, which could also be expected. The surprise was that the third color per affinity was black (noncolor). From a total of 174 surveyed subjects, blue was the favorite color for 31 of them (18%), red for 23 (13%) and black for unexpected 21 (12%). Looking at it in a different way (see the last round chart), the order among the six most favorite colors is the following (Microsoft names of colors): Blue 28% (0, 0, 255), Red 21% (255, 0, 0), Black 19% (0, 0, 0), Aqua 17% (0, 255, 255), Turquoise 10% (0, 204, 255), Dark Green 5% (0, 128, 0).

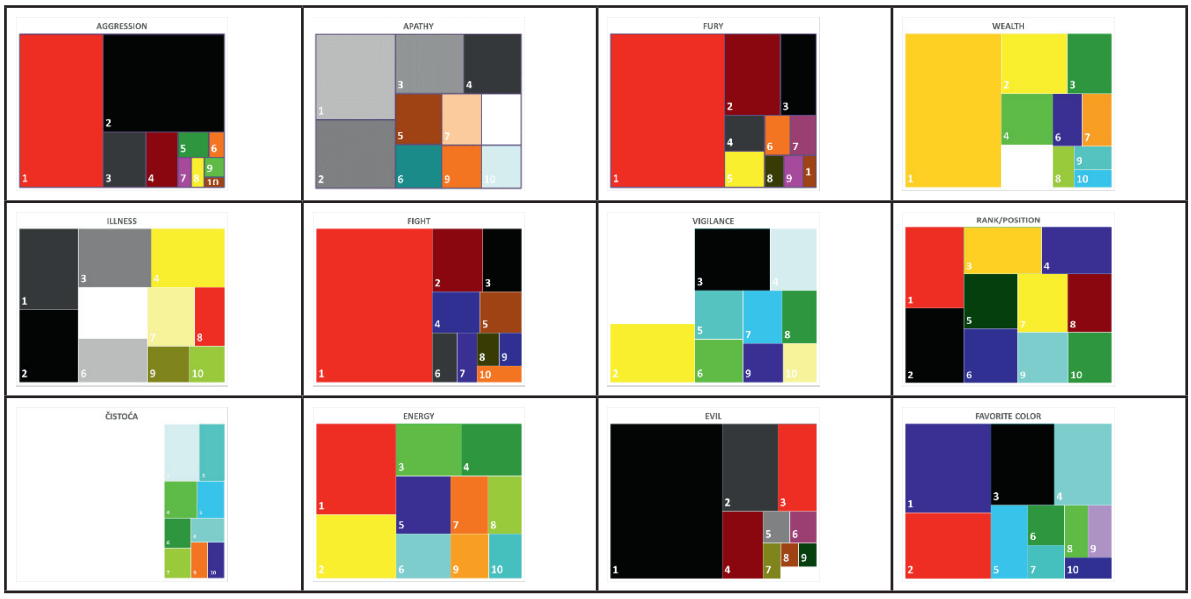

Aggression
9%
47%
44%
Apathy
17%
34%
21%
28%
Fury
12%
18%
70%
Wealth
11%
11%
17%
61%
Illness
24%
27%
24%
25%
Fight
10%
14%
76%
Vigilance
26%
46%
28%
Rank/position
20%
30%
23%
27%
Purity
8%
5%
87%
Depression
21%
29%
24%
26%
Friendliness
18%
33%
22%
27%
Energy
14%
33%
14%
16%
23%
Enthusiasm
18%
26%
18%
20%
18%
Formality
16%
26%
16%
19%
23%
Formality
16%
26%
16%
19%
23%
Stupidity
18%
25%
17%
21%
19%
Harmony
18%
22%
18%
22%
20%
Hygiene
9%
14%
77%
Coldness
22%
43%
35%
Idealism
18%
33%
22%
27%
Intelligence
19%
37%
19%
25%
Isolation
19%
43%
18%
20%
Simplicity
11%
12%
14%
63%
Remorse
19%
32%
19%
30%
Concentration
22%
27%
24%
27%
Creativity
17%
23%
18%
21%
21%
Fragility
13%
35%
18%
34%
Cowardness
23%
43%
34%
Loyalty
31%
36%
33%
Madness
24%
44%
32%
Luxury
9%
8%
16%
67%
Love
5%
6%
7%
82%
Jealousy
20%
31%
23%
26%
Anger
10%
23%
67%
Fantasy
31%
34%
35%
Peace/calmness
10%
10%
11%
59%
10%
Mysticism
17%
37%
20%
26%
Youth
19%
29%
25%
27%
Power
20%
42%
38%
Wisdom
19%
32%
23%
26%
Masculinity
14%
16%
40%
30%
Inexperience
21%
27%
25%
27%
Immorality
17%
23%
17%
22%
21%
Infertility
20%
26%
27%
27%
Tastelessness
33%
35%
32%
Intemperance
21%
29%
24%
26%
Innocence
22%
78%
Money
15%
46%
39%
Gentleness
12%
14%
56%
18%
Caution
10%
10%
10%
57%
13%



Picture 10. *Overview of percentage participation of most frequent colors associated with certain terms using the Pie type of chart*

A chart of Treemap type shows the data in a specific way, using rectangles that fully fill a certain surface.

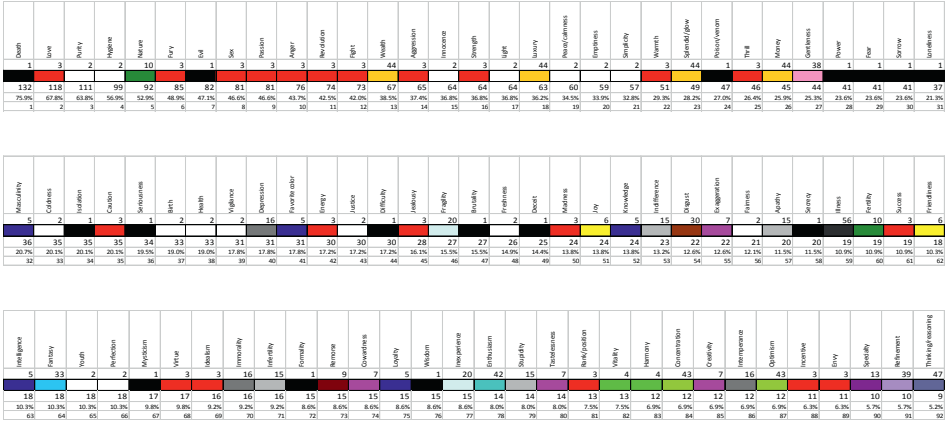
The largest rectangle presents the largest set of data, and to the right of it there are smaller rectangles with lower values in the descending order. This chart type does not have sub-types and may be used in Office 2016. It did not exist in earlier versions. The picture shows some characteristic results of processing answers using the Treemap type chart.



Picture 11. *Overview of ten most frequent colors associated with certain terms using the Treemap type of chart*

Table 1. *Overview of colors with most associations with certain terms*

Color	RGB code	Most frequently associated terms
	255 000 000	love, fury, sex, passion, anger, revolution, fight, aggression, strength, warmth, thrill, caution, energy, jealousy, madness, success, virtue, idealism, rank/position, incentive, envy
	000 000 000	death, evil, poison, power, fear, sorrow, loneliness, isolation, seriousness, difficulty, brutality, deceit, secrecy, mysticism, formality, wisdom
	000 000 255	masculinity, favorite color, knowledge, intelligence, loyalty
	128 128 128	depression, immorality, intemperance
	000 128 000	nature, fertility



Picture 12. *The order of the most prominent connection between a color and a term*

Conclusion

In this paper we analyzed psychological connection between 40 offered colors and 91 terms from everyday life. The answers were provided by 147 men and 27 women who were asked, besides the color-term associations, also to fill out the Ishihara test that checks the ability of correct perception of colors and to state their favorite color. A part of received results was expected, so even the result of Ishihara test, showing that 7% of subjects (12 polled, 10 men and 2 women) do not recognize colors properly (dichromatism) – do not differ significantly from the results of other research. Color blindness (color vision deficiency, or CVD) affects approximately 1 in 12 men (8%) and 1 in 200 women in the world. In Britain this means that there are approximately 2.7 million colorblind people (about 4.5% of the entire population), most of whom are male. [6] The biggest surprise for the author was high position (third place) of black on the list of favorite colors, given that 80% of the polled subjects were 30 or under 30 years of age. The answers suggest that light colors caused mainly positive emotional associations, while dark colors caused mainly negative emotional associations. The analysis was not made separately for men and women. We will leave it for the future survey that should be posted onto a website, where a far larger number of participants will be able to take part in the survey.

Bibliography

[1] AIC - International Colour Association: https://aic-color.org/ [Accessed 16. 7 2018]

[2] Arnhajm, R. (1998). Umetnost i vizuelno opažanje. Beograd: Univerzitet umetnosti u Beogradu i Studentski kulturni centar Beograd.

[3] Barni, D. (2006). Svetlost. Beograd: Knjiga-komerc i Politika NM.

[4] Color Matters. (26. 7 2018). Preuzeto od Color Matters: https://www.colormatters.com/

[5] Colorcom - The Color Consultant Experts. (12. 7 018). Preuzeto od Colorcom - The Color Consultant Experts: https://www.colorcom.com/index.php

[6] colourblindawareness. http://www.colourblindawareness.org/colour-blindness/ [Accessed 1. 8. 2018.]

[7] Illustrated Oxford Dictionary. (1998). London, England: Dorling Kinderslay Limited and Oxford University Press.

[8] Nauka, velika ilustrirana enciklopedija (DK Science the definitive visual guide). (2011). Zagreb: Mozaik knjiga.

[9] Nova enciklopedija u boji Vuk Karadžić Larousse. (1977). Beograd: Izdavačko preduzeće Vuk Karadžić.

[10] Rječnik stranih riječi: tuđice, posuđenice, izrazi, kratice i fraze (Dictionary of foreign words) by Sima Anić, Nikola Klaić, Želimir Domović - Zagreb: Sani-plus, 1998. pg. 308)

[11] Tanhofer, N. (2008). O boji na filmu i srodnim medijima. Zagreb: Akademija dramskih umjetnosti Sveučilišta u Zagrebu i Novi Lliber d.o.o.

[12] Trstenjak, A. (1987). Čovek i boje. Beograd: Nolit.

[13] What's your favorite color? [Infographic]: https://www.hotdesign.com/marketing/whats-your-favorite-color [Accessed 10. 7. 2018.]

Submitted: December 12, 2018
Accepted: December 22, 2018

About the authors

Nedim Smailovic was born in Tuzla. He has been living in Banja Luka since 1973. He graduated from the Faculty of Electrical Engineering, department of Telecommunications. Since 1982 he has worked in RO PTT traffic of Bosnia and Herzegovina, and a series of organizational transformation it is now called Mtel doo Banja Luka. His first work experience was in designing and maintaining the PTT capacities. He obtained his Master's degree from Pan European University 'Aperion' Banja Luka, in 2005. There he also defended his doctoral thesis titled: Computer information graphics in presenting Bosnia and Herzegovina on the road to accessing the European Union. He was elected Associate Professor in 2013 and he has been teaching since in three universities in Bosnia and Herzegovina subjects relating to computer technology. He is an author and co-author of several books from the field of information technology and mathematics. He is married, father of two daughters.

For citation

Nedim Smailović, Psychological Connection between Colors and Certain Characteristic Terms, *JITA – Journal of Information Technology and Applications*, PanEuropien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 8(2018) 2:75-85, (UDC: 159.937.51.072), (DOI: 10.7251/JIT1802075S), Volume 8, Number 2, Banja Luka, june 2018 (45-96), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004

SAFETY ASPECTS IN SHARED MEDICAL IT ENVIRONMENT

Igor Dugonjić[1], Mihajlo Travar[2], Gordan Bajić[1]

[1]Pan-European University "Apeiron", Banja Luka, BiH

[2]University of Business Studies, Banja Luka, BiH

Critical Review

DOI: **10.7251/JIT1802086D** UDC: **001.3:61]:615.849**

Abstract: Regional PACS and other shared medical systems are primary intended for sharing medical images. In these systems, the number of users is significantly increased in relation to local systems, and the fact is that the public network is very frequently used for data transfer. As medical data are very sensitive, such situation creates considerable risk regarding privacy, integrity and right to access to these data. This paper includes the most frequent risks and methods to solve these issues as well as recommendations for safe use of cloud computing systems in order to implement these systems.

Keywords: PACS, DICOM, IHE.

INTRODUCTION

In the era of fast changes and increasingly less time necessary to put an innovation in use, there is a great urge to accomplish as much as possible with the minimum of effort and tools. It is common for radiology wards in hospitals to possess and use a central information system such as Picture Archiving and Communication System (PACS), Radiology Information System (RIS) and the like.

While Electronic Medical Record (EMR) is able to coordinate handling with medical content (medical image with meta-data) which come from different viewers, EMR users must face the work with different viewers from different wards [17]. Due to this fact, it is easy to predict users' dissatisfaction because of the need to work in several different applications that are basically intended for the same job [13][5]. This why there is also a problem on the level of a health institution due to allocation of funds for maintenance of several small or medium archives.

The logical solution to this problem is the centralisation of the medical data archiving system on the regional or national level rather than local level only. Such organisation can improve the level of health services in human resources, education, and R&D, develop better and more comprehensive patient monitoring systems and possibility of improving new healthcare services.

The use of PACS creates a data sharing mechanism. These data may be delivered to a regional PACS client or any XDS client (document user). Apart from advantages such system can offer, it also poses certain risks. The growth of a system to the regional level significantly increases the risk of its users, which also leads to significant increase in safety risks. Therefore, it is necessary to analyse the risks and develop mechanisms to enhance medical image security.

SECURITY CONSIDERATION

The Royal College of Radiologist (RCR) prescribes standards for patient confidentiality, and RIS and PACS [11]. Confidentiality is one of the basic patient-physician relationship principles. However, attention should be paid in introducing modern technologies since this relationship may include other participants as well. Pursuant to Data Protec-

tion Act (1998) [3], the Human Rights Act (1998) [8] and National Health Service Act (NHS) (2006) [10], RCR defines the following key identification data on patients: patient's name, address, full post code, date of birth, pictures, photographs, videos, audio tapes, NHS number, local patient identifiable codes and anything else that maybe used to identify a patient directly or indirectly, for example rare diseases, drug treatments etc.

Although PACS brings many possibilities to improve radiological practice on one hand, there is a significant risk to patient confidentiality on the other. This risk may occur in case of improper use of PACS. Some of these examples are searching images and results by medical staff not involved in a patient's care as well as searching images and results by individuals unauthorized for the access to such medical archives [11].

The 1996 Health Insurance Portability and Accountability Act (HIPAA) requires, among other things, the protection from unauthorized access to patient data [7]. Medical images obtained from CR, CT, MRI, US etc. cannot be used without other information such as description of diagnosis, possible reference to the history of disease, previous treatments and other information on the patient. This complex set of information on the patient is highly sensitive. This is why a high level of security is required for the data handled by regional PACS.

Training has always been one of the most important parts in radiology. Images adequate for training or research need to be created as anonymous i.e. personal data on patient and other information that may indicate the patient's identity during sending to the regional PACS that is used for training and research need to be replaced with fictitious data.

Besides economic advantages, the service quality such as greater availability, high reliability and scalability is the main incitement to use cloud computing. However, in case of outsourcing i.e. clinical data transfer to the cloud, health institutions face many challenges that should be taken into consideration in the preparation phase of integration. Namely, privacy and confidentiality of data in the cloud is the main obstacle for cloud to be widely accepted due to the risk of compromised user's data confidentiality when the data are transferred to cloud.

Privacy Protection and Regulation of Right to Access

Data need to be protected at three different levels, as follows: a) data transferred in the public network, b) data stored on the regional PACS server and c) user's access to these data.

Health institutions not covered by own optical network are oriented to public networks, which is much more economical but requires the use of data protection cryptography.

The safety of data sent through public network may be ensured by using dedicated optical cables whenever possible, and by using encrypted tunnels with a high level of protection in all shared lines used in another network traffic. For this purpose, IPSEC tunneling (Internet Protocol Security tunelling) may be used, most frequently with Advanced Encryption Standard AES-256 encrypting algorithm. IPSEC is the most frequently used safety control on network layer for private data protection through public IP networks. Depending on how it is implemented and configured, IPSEC can provide any combination of the following protection types: Confidentiality, Integrity, Authentication, Replay Attack Protection, Traffic Analysis Protection, and Access Control [15]. The most frequent method of using IPSEC implementations is Virtual Private Network (VPN) service. VPN is a private network built on the already existing physical network, which can ensure a safe communication mechanism for data and information sent through the network.

The safety of data stored in a shared medical archive can be attained by using a hardware intended for this purpose as well as strict restriction of physical and network access to the equipment. One of the ways to implement the access control is to use two firewalls, one of which is controlled by the local health institution and its staff, and the other (external) firewall by the regional PACS engineering staff. This allows the regional PACS administrators to control the access to central resources, while the health institution network administrators can control the access to their network. This allows anybody to have the access to resources they are responsible for and which they control. This refers to both network connection variants i.e. those implemented with optical cables and those using IPSEC tunnel between the regional PACS server and local institution.

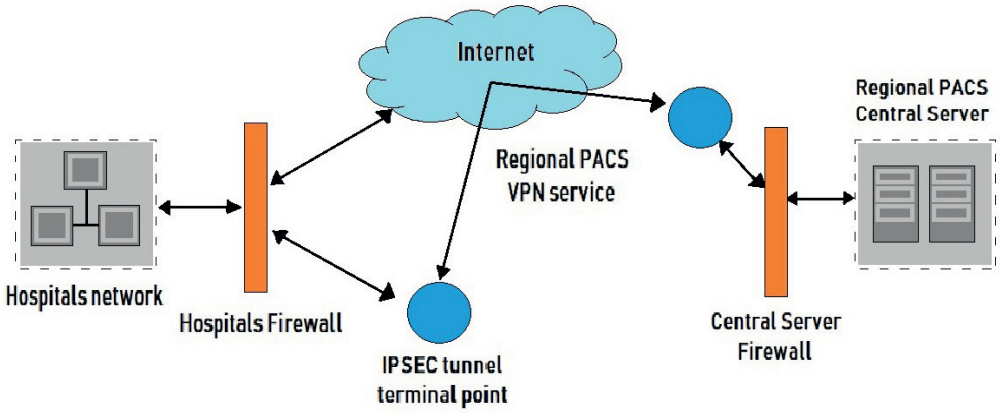


Figure 1. *Health institutions interconnection model using IPSEC tunnel*

As with a local medical data exchange server, the regional system must have the access to data; however, generally in this case the data are in a remote location and need to be accessed through the public network. The aspect of data privacy protection from unauthorized reading and changing in the regional systems expands the requirements the local PACS needs to fulfil. Data security must primarily be protected from those who are not system users.

When the local PACS is integrated into the regional system, there is the risk of unauthorized reading of medical data. This risk is present primarily due to the fact that local centres need to be connected into a joint network. The right to access to the regional system must be reduced in relation to the access to the local PACS. This comes from the fact that system users, generally, should not have the same rights in a remote centre in comparison to a local centre.

The access to medical images and related data must be defined in line with users' rights. The situation in the regional PACS is more complicated when it comes to the local centre. The ownership structure of local centres is generally heterogeneous, and the right of access policy and users' privileges in the system must be in line with the specific needs and requirements. Table 1. shows an example of establishing the regional PACS users' rights.

Digital identity can be defined as a set of claims made by one digital subject about itself or another digital subject [2], while determining the digital identity is reduced to what the subject: is (fingerprint), knows (password), has (token), does (motor skills), where he/she is etc.

Organized pairs of username and password may be used to log into the local system. Although such authentication system carries certain risks, it may

Table 1. *Example of establishing a regional PACS system user's rights*

		Local center		Remote center	
		Radiographer	**Radiologist**	**Radiologist**	**System Administrator**
Local center	Image saving	Yes	Yes	Yes	Yes
	Image describing	Yes	Yes	Yes	Yes
	Raport writing	-	Yes	Yes	-
	Raport reading	Yes	Yes	Yes	Yes
	Administration	-	-	-	Yes
Remote center	Image saving	-	-	-	Yes
	Image describing	-	-	Yes	Yes
	Raport writing	-	-	Yes	-
	Raport reading	-	-	Yes	Yes
	Administration	-	-	-	Yes

yield satisfactory results in some specific cases. PACS users identity in more complex systems may be based on Public Key Infrastructure (PKI). In that case, it is expected that every user that uses more than one workstation or who shares a workstation with someone else must possess Dongle that generates a private key. The competent body prescribes an appropriate public key. The issue of medical specialists' electronic identity should be solved globally i.e. for the whole system [16].

When regional PACS is used for educational or scientific purposes, it is important to establish the coordination in marking medical images with fictitious data so that the patient's identity always remain hidden. On the other hand, students or researchers have the access not only to individual medical images but also to series of images for a more complex insight into the history of disease and applied treatments. It is very important to deny access to sensitive and confidential information on the patient. A frame structure for training consists of the case study object. Every such object is a hypertext object describing a specific medical case and directing to relevant, also anonymous medical images. The data for a real patient originating from different clinical centres must use the same fictitious identity. Such principle allows students a more complex insight into the history of disease. Using Web Access to DICOM Objects (WADO) service, web client may require the access to DICOM objects such as medical images or medical reports in a remote repository. One of important parameters supported by WADO service is data anonymisation. It enables removal of all identification patient's data from DICOM objects if those have not been previously removed [4].

The current state of cloud computing does not guarantee privacy and confidentiality of stored data due to possible publishing and unauthorized use of those data. World Privacy Forum (WPF) suggests several pieces of advice for cloud computer users: a) the terms of service provision need to be carefully considered prior to storing any information into cloud and, if the terms are inappropriate or unclear, another cloud provider should be taken into consideration. b) information stored in cloud are more available than those in a private computer, due to which sensitive information should not be transferred into the cloud. c) in case of data withdrawal by users, it is important that the cloud provider does not retain the right to those data. d) During a change in requirements for service provision, make sure the user is informed on it [6]. This clearly shows that there are great risks regarding data confidentiality in using of cloud service.

Even in case of data anonymisation it is possible that they can be illegally used or sold to a third party for statistical purpose. Therefore, cloud computing without additional systems of data privacy and confidentiality is inadequate for storing any kind of confidential information. It is often impossible to provide a system so that it can avoid information 'leak' or data mining performed to separate certain patterns from medical data. Apart from that, there are many unclear situations such us 'if provider goes bankrupt, what will happen to the data in cloud?' Furthermore, there is the issue of provider's headquarters, users' headquarters and server location not being in the same countries, each of them having different legislation: In case of a dispute, which law is applicable? Therefore, these situations should also be taken into account when creating a system working under cloud.

In order to combine two concepts - Cross Enterprise Document Sharing for Imaging **(**XDS-I) and public cloud - it is necessary to ensure the privacy and confidentiality of protected medical information (PHI) without removing the interoperability between XDS-I profile. The way the privacy and confidentiality, and at the same time interoperability in migration of two XDS-I profiles to cloud, are ensured is by applying some of protection procedures such as coding during storing and on-the-fly decoding, or middleware coding/decoding and XDS-I with privacy protection. XDS-I with privacy protection allows protection and interoperability on the architecture level because trends and actors would be the same as in XDS profile. The shortcoming of this is the lack of interoperability on the document level because such profile has not been planned by IHE yet [12].

Digital Watermarking

DICOM is the standard for transmission and storage of medical images. All installed equipment in a shared system should be fully DICOM compatible. The main content of DICOM image is medical visu-

al information, while DICOM header also contains metadata. These very sensitive data should be adequately protected.

Digital watermarking offers a suitable technique to ensure the authenticity and copyrights of medical images. For watermarking to be useful, the process of adding watermark on a medical image should be done immediately after obtaining the image on modality.

In respect of the visibility, watermarks are classified into visible and hidden, while in terms of resistance, watermarks are classified into robust and fragile watermarks.

A visible watermark is easily visible with the naked eye. This watermark is inserted into an image to be almost impossible to separate it and get the original image. This method can be used to prevent illegal distribution of medical images.

There are much more methods of marking with invisible watermark than method of marking with visible watermark. Invisible watermarks are hidden within the content. They can be detected only with authorized programs. They are used for protection and copyright authentication, ownership confirmation and detection of unauthorized copying. Inserted watermarks are resistant to image processing.

Robust watermarks are resistant to attacks and can be used for copyright protection.

Fragile digital watermarks can be easily destroyed during each attempt of data manipulation and hence they are used to detect changes in digital content. They allow data authentication.

Depending on domain insertion they are classified into spatial, transform and parametric watermarks.

In spatial domain technique, a digital watermark is inserted into positions of lower bits that are not very significant for image display so that the display quality is not impaired. Digital watermark insertion techniques are quite simple and are relatively efficient way to insert an invisible watermark inside an image. Unfortunately, spatial domain techniques are not very resistant to common forms of data manipulation. This method is mainly used to provide authenticity.

The parametric method is based on transforming the original image on the parametric level, where the original image is manipulated by changing its parameters such as illumination, contrast or even colour in case of colour medical images.

Transformational (frequency) domain is the way to insert a digital watermark inside an image that begins with transformation of the original image in a frequency domain. The most used mathematical transform in digital watermark protection methods are Discrete Fourier transform (DFT), Discrete cosine transform (DCT) and Discrete wavelet transform (DWT).When using these techniques, watermarks are not added to intensities of individual parts of an image but to transform ratios, and the image with a digital watermark is obtained by inverse transformation. Such techniques are far more efficient since they allow for the human visual system properties in determining the watermark position inside the image.

This method (Transformational)is the most suitable for medical imaging as it neither impairs the quality of the original image nor changes the image parameters such as i.e. contrast, illumination or bit depth [14]. This method belongs to highly resistant methods in terms of the use of filters and compression.

A valuation of the existing protection methods using a digital watermark in terms of their resistance to imprint and subsequent digitalisation process has shown that the highest potential is in methods that disperse (distribute) the energy of a digital watermark across the entire digital signal. Such approach is inherently resistant to degradation caused by the process of imprinting and subsequent digitalisation, which arises from the very properties of mathematical transform, which is the basis of the approach.

The use of such method is unsuitable since even the tiniest bit changes may render a proper diagnosis problematic. However, RONI (Region of Non Interest), a method that inserts a watermark into a part of medical image that is not used in diagnostics, may be a compromise solution. The solution is a compromise because it uses the positive aspect of watermarking. However, ROI (Region of Interest) remains unprotected i.e. without watermark.

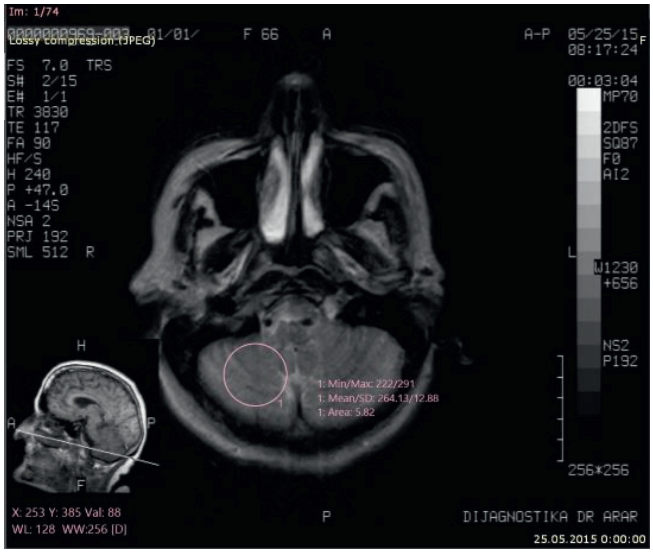


Figure 2. *Manual circular ROI in MR image of head*

Watermarking can be very important in standard radiological practice and training of students and medical personnel staff since it can protect authenticity, proof of ownership and data immutability.

Conclusion

For a shared regional medical system to be acceptable, it is necessary to pay attention to the rights and privileges of users and administrators on one hand, and confidentiality of patient medical data as well as information on health institutions themselves on the other. The issue of authentication is especially pronounced with mobile users who access the system from different locations or in case a single access point is used by several users. The need to protect medical data from unauthorized access is also pointed out, as well requirements for the use of regional PACS for educational and scientific purposes. Cloud computing is a technology that offers a simple way to allow PACS functionality and possibility of regional integration of local diagnostic centres. However, this technology carries certain risks related to data privacy. This paper deals with these risks and provides certain recommendations for their elimination. Besides, the paper examines the possibility of implementation of XDS-I infrastructure in cloud as well as medical image watermarking.

References

[1] Avramović Ž. Zoran, Radojičić Z. Radomir, Mirković D. Saša,(2015) A new Approach to Computer Analysis of Queuing Systems Without Programming, JITA- Journal of Information Technology and Applications, JITA 5(2015) 1:25-32

[2] Cameron K., The laws of identity, Microsoft Corp.,https://msdn.microsoft.com/en-us/library/ms996456.aspx, (last accessed 27/10/17)

[3] Data Protection Act (1998), http://www.legislation.gov.uk/ukpga/1998/29/contents, (last accessed 27/11/17)

[4] Digital Imaging and Communications in Medicine (DICOM) Part 18: Web Access to DICOM Persistent Objects (WADO), National Electrical Manufacturers Association, USA 2011

[5] Fuller SS, Kethcall DS, Tarzy-Hornach P: Integrating knowledge resources at the point of care: opportunities for librarians, Bull Med LibrAssoc 87(4):393–403, 1999, Oct

[6] Gellman Robert ,WPF, Privacy in the Clouds: Risks to Privacy and Confidentiality from Cloud Computing, February 23, 2009

[7] HIPAA, Federal Register / Vol. 68, No. 34 / Thursday, February 20, 2003 / Rules and Regulations

[8] Human Rights Act (1998), http://www.legislation.gov.uk/ukpga/1998/42/contents, (last accessed 27/11/17)

[9] MaksimovićMirjana, Implementation of Fog computing in IoT-based healthcare system, (2017) JITA- Journal of Information Technology and Applications, JITA 7(2017) 2:100-107

[10] National Health Service Act(2006), http://www.legislation.gov.uk/ukpga/2006/41/contents (last accessed 27/11/17)

[11] RCR, https://www.rcr.ac.uk/publication/standards-patient-confidentiality-and-ris-and-pacs, (last accessed 27/11/17)

[12] Ribeiro S. Luís, Costa Carlos and Oliveira José Luís: Current Trends in Archiving and Transmission of Medical Images, Medical Imaging, InTech, 2011

[13] Richardson M, MIND scape and PubMed: Web sites that can change the way we work, AcadRadiol 5(7):519–520, 1998, Jul

[14] Rocek Ales, Medical image data security based on principles of digital watermarking methods, Advances in Data Networks, Communications, Computers and Materials ISBN:978-1-61804-118-0

[15] SheilaFrankel, Kent Karen , LewkowskiRyan , Orebaugh D. Angela , Ritchey W. Ronald:Guide to IPSEC VPNs: Recommendations of the National Institute of Standards and Technology Paperback – December 31, 2005

[16] SlavicekKarel, Javornik Michal, Dostal Otto: MeDiMed – Regional Center for Medicine Multimedia Data Exchange, WSEAS TRANSACTIONS on INFORMATION SCIENCE & APPLICATIONS ISSN: 1790-0832 Issue 4, Volume 5, 2008

[17] Stewart BK and Langer SG: Integration of DICOM images into an electronic medical record using thin viewing clients, Proc AMIA Symp 902-6,1998

Submitted: December 17, 2018
Accepted: December 22, 2018

ABOUT THE AUTHORS

Igor Dugonjić earned his Master's degree in computer science at the Faculty of Electrical Engineering, University of Banja Luka. He is currently doing his PhD at Pan-European University 'APEIRON' in Banja Luka. Mr Dugonjić works as a medical equipment programming and maintenance engineer at the University Clinical Centre of Republika Srpska. He has written several scientific papers on medical ICT research.

Mihajlo Travar earned his PhD at the Faculty of Mechanical Engineering, University of Belgrade. He is a member of Regulatory Commission for Energy of Republika Srpska. Mr Travar is Associate Professor at the 'University of Business Studies' in Banja Luka, where he gives lectures on the following subjects: Databases, Software Engineering, CASE Tools, Design Engineering and ERP Systems. He has written more than forty scientific papers in ICT, mechanical engineering and business organisation.

Gordan Bajić earned his Doctor's degree in health sciences in the field of physiotherapy and work therapy at the Pan-European University "Apeiron", Banja Luka. Gordan Bajic is Assistant Professor of Health Sciences at the Pan-European University, Faculty of Health Sciences "Apeiron" and is vice-dean for teaching at the Faculty of Health Sciences, Pan-European University "Apeiron", Banja Luka. He is a member of many symposiums and has written many scientific papers in the field of medicine, physiotherapy, etc.

FOR CITATION

Igor Dugonjić, Mihajlo Travar, Gordan Bajić, Safety Aspects In Shared Medical It Environment, *JITA – Journal of Information Technology and Applications*, PanEuropien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 8(2018) 2:86-92, (UDC: 001.3:61]:615.849), (DOI: 10.7251/JIT1802086D), Volume 8, Number 2, Banja Luka, june 2018 (45-96), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004

Instructions for Authors

The *Journal of Information Technology and Application (JITA)* publishes quality, original papers that contribute to the methodology of IT research as well as good examples of practical applications.

Authors are advised that adherence to the Instructions to Authors will help speed up the refereeing and production stages for most papers.

- Language and presentation
- Length of submissions
- Submission
- Contact details/biographies
- Title of the paper
- Abstract and keywords
- Figures and tables
- Sections
- Footnotes
- Special characters
- Spelling
- References
- Proofs
- PDF offprint
- Copyright and permissions
- Final material
- Correspondence
- Publication ethics

Language and presentation

Manuscripts should be written in English. All authors should obtain assistance in the editing of their papers for correct spelling and use of English grammar. Manuscripts should have double spacing, with ample margins and pages should be numbered consecutively. The Editors reserve the right to make changes that may clarify or condense papers where this is considered desirable.

Length of submissions

Papers should not normally exceed 12 Journal pages (about 8000 words). However, in certain circumstances (e.g., review papers) longer papers will be published.

Submission

Manuscripts must be submitted through the JITA online submission system.

Please read the instructions carefully before submitting your manuscript and ensure the main article files do not contain any author identifiable information.

Although PDF is acceptable for initial submission original source (i.e. MS Word) files will be required for typesetting etc.

Contact details/biographies

A separate file containing the names and addresses of the authors, and the name and full contact details (full postal address, telephone, fax and e-mail) of the author to whom correspondence is to be directed should be uploaded at the time of submission (you should select Contact details/Biographies as the file type). This file is not shown to reviewers. This file should also contain short biographies for each author (50 words maximum each) which will appear at the end of their paper.

The authors' names and addresses must not appear in the body of the manuscript, to preserve anonymity. Manuscripts containing author details of any kind will be returned for correction.

Title of the paper

The title of the paper should not be longer than 16 words.

Abstract and keywords

The first page of the manuscript should contain a summary of not more than 200 words. This should be self-contained and understandable by the general reader outside the context of the full paper. You should also add 3 to 6 keywords.

Figures and tables

Figures which contain only textual rather than diagrammatic information should be designated Tables. Figures and tables should be numbered consecutively as they appear in the text. All figures and tables should have a caption.

Sections

Sections and subsections should be clearly differentiated but should not be numbered.

Footnotes

Papers must be written without the use of footnotes.

Special characters

Mathematical expressions and Greek or other symbols should be written clearly with ample spacing. Any unusual characters should be indicated on a separate sheet.

Spelling

Spelling must be consistent with the Concise Oxford Dictionary.

References

References in the text are indicated by the number in square brackets. If a referenced paper has three or more authors the reference should always appear as

the first author followed by et al. References are listed alphabetically. All document types, both printed and electronic, are in the same list. References to the same author are listed chronologically, with the oldest on top. Journal titles should not be abbreviated.

Journal

Avramović ZŽ (1995) Method for evaluating the strength of retarding steps on a marshalling yard hump. *European Journal of Operational Research*, 85(1), 504–514.

Book

Walsham G (1993) Interpreting Information Systems in Organizations. Wiley, Chichester.

Contributed volume

Huberman AM and Miles MB (1994) Data Management and analysis methods. In Handbook of Qualitative Research (Denzin NK and Linkoln YS, Eds), pp 428-444, Sage, Thousand Oaks, California.

Conference Paper

Huberman AM and Miles MB (1994) Data Management and analysis methods. In Handbook of Qualitative Research (Denzin NK and Linkoln YS, Eds), pp 428-444, Sage, Thousand Oaks, California.

Unpublished reports/theses

Nandhakumar JJ (1993) The practice of executive information systems development: and in-depth case study. PhD Thesis, Department of Engineering, University of Cambridge.

Proofs

Proofs of papers will be sent to authors for checking. Alterations to diagrams should be avoided where possible. It will not be possible to accept major textual changes at this stage. Proofs must be returned to the publishers within 48 hours of receipt by fax, first-class post, airmail or courier. Failure to return the proof will result in the paper being delayed.

PDF offprint

Corresponding authors will receive a PDF of their article. This PDF offprint is provided for personal use. It is the responsibility of the corresponding author to pass the PDF offprint onto co-authors (if relevant) and ensure that they are aware of the conditions pertaining to its use.

The PDF must not be placed on a publicly-available website for general viewing, or otherwise distributed without seeking our permission, as this would contravene our copyright policy and potentially damage the journal's circulation. Please visit **http://www.apeiron-journals.com/JITA/authors/rights_and_permissions.html** to see our latest copyright policy.

Copyright and permissions

The copyright of all material published in the Journal is held by Paneuropean University APEIRON. The author must complete and return the copyright form enclosed with the proofs.

Authors may submit papers which have been published elsewhere in a foreign language, provided permission has been obtained from the original publisher before submission.

Authors wishing to use material previously published in JITA should consult the publisher.

For citation

Bestemyanov Petr Filimonovich, Evaluation of the period of sensors motion parameters of the train , JITA – Journal of Information Technology and Applications, PanEuropien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 8(2018) 1:14-17, (UDC: 656.08:352.07), (DOI: 10.7251/ JIT1801014B), Volume 8, Number 1, Banja Luka, june 2018 (1-44), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004

About the authors - Biographies

Authors should submit their short biographies up to 100 words at the end of their work and, possibly, an picture.

Final material

All final material must be submitted electronically in its original application format (MS Word is preferred). The file must correspond exactly to the final version of the manuscript.

Correspondence

Business correspondence and enquiries relating to advertising, subscriptions, back numbers or reprints should be addressed to the relevant person at:

Paneuropean University APEIRON
Journal JITA
Pere Krece 13, P.O.Box 51
78102 Banja Luka
Bosnia and Hercegovina / RS

Publication ethics

We take an active interest in issues and developments relating to publication ethics, such as plagiarism, falsification of data, fabrication of results and other areas of ethical misconduct. Please note that submitted manuscripts may be subject to checks using the corresponding service, in order to detect instances of overlapping and similar text.

JITA

PUBLISHER
Paneuropean University APEIRON,
College of Information Technology
Banja Luka, Republic of Srpska, B&H
www.apeiron-uni.eu
Darko Uremović, Person Responsible for the Publisher
Aleksandra Vidović, Editor of University Publications

ISSN 2232-9625
9 772232 962005