

Characteristics of Thermally Sprayed NiCrBSi Coatings before and after Electromagnetic Induction Remelting Process

**Petru Cristian Vălean^{1,2}, Norbert Kazamer^{1,2},
Dragoș-Toader Pascal², Roxana Muntean², István Barányi³,
Gabriela Mărginean², Viorel-Aurel Șerban¹**

¹Politehnica University Timișoara, Department of Materials Science and Manufacturing Engineering, Piața Victoriei nr. 2, 300006 Timișoara, Romania
viorel.serban@upt.ro

²Westphalian University of Applied Sciences, Department of Materials Science and Testing, Neidenburgerstr. 43, 45897 Gelsenkirchen, Germany
petru-cristian.valean@studmail.w-hs.de, norbert.kazamer@w-hs.de,
dragos.pascal@w-hs.de, roxana.muntean@w-hs.de, gabriela.marginean@w-hs.de

³Óbuda University, Bánki Donát Faculty of Mechanical and Safety Engineering, Népszínház u. 8, H-1081 Budapest, Hungary, baranyi.istvan@bgk.uni-obuda.hu

Abstract: Active mechanical components that work in highly aggressive environments require protection against premature deterioration and, at the same time, it is necessary to ensure a long lifespan. In order to achieve these important features, a protective coating can be added. In this work, a NiCrBSi self-fluxing alloy powder was employed, as feedstock material. The coating was deposited by means of oxyacetylene flame spraying process, achieving a thickness of approximately 1000 μm and exhibiting a high degree of porosity and a weak mechanical adhesion to the substrate. To improve these characteristics, a remelting process using high frequency currents was applied. After the remelting process, the porosity decreased, from an initial value of approximately 15%, to a final value under 3%. Although the microhardness values did not change significantly, both wear-rate and corrosion behavior improvement, after the electromagnetic remelting, was observed.

Keywords: NiCrBSi; self-fluxing alloy; flame spraying; inductive remelting process; pin-on-disk; microhardness; corrosion behavior

1 Introduction

Parts that work in highly aggressive environments must be protected against premature deterioration, corrosion and wear and at the same time, it is strongly

necessary to ensure a long lifetime of the components. Thermally sprayed coatings offer a practical and economical solution to improve corrosion and wear resistance on these types of elements [1, 2]. Nickel based hard materials, such as NiCrBSi family; have been recently used in industries where corrosion and wear resistance is strongly required like: paper, petrol, hot working punches or heat exchangers [3-6]. Many studies showed that these coatings are an alternative for hard chromium coatings, which are harmful to the environment [7].

Thermal spraying is regarded as a suitable technique of coating components, in which a molten or semi-molten material is sprayed on a surface, which is commonly degreased and sand-blasted in advance. The technology can provide coatings thicknesses from a few micrometers to a few millimeters, depending on the used material and the applied coating technique. The majority of coating techniques which belong to thermal spraying use a powder as raw material. For example, high velocity oxygen fuel (HVOF) process can be used in order to obtain reduced porosity for the as sprayed coatings [8]. Another combustion process is flame deposition, using oxyacetylene. This process is also known as LVOF (low velocity oxygen fuel), which is generally used to repair damaged parts [9].

Currently, a great deal of conventional and expensive processes are studied and employed in order to deposit NiCrBSi powder on various substrates. An attempt of HVOF deposition of NiCrBSi coatings with a high content of Cr (14.5%) and Fe (4.5%) on a low alloyed steel substrate was successfully performed by L. Vieira *et al.* Though, the microscopic investigations revealed cracks at the coating-substrate interface, with possible peeling, it showed as well a high degree of micro porosity in the coating [10]. Another research in this field was performed by a team from the Netherlands, which tried to apply a coating using a similar type of powder, but with a higher Cr content (16.5%) by means of laser deposition process. The obtained NiCrBSi coating exhibited significant cracks and possessed high concentration of internal stress, although a pre-treatment of the substrate at 500°C was applied [11].

Therefore, this work aims to deposit the self-fluxing alloy NiCrBSi using a two-step deposition process. Firstly, an inexpensive deposition technique is considered (LVOF). The thermal spraying method consists of melting and propelling particles by a flame, towards a substrate, in order to form the NiCrBSi coating on a EN S355 J2 steel. Secondly, to solve the initial problems regarding the increased porosity, minimize the induced microcracks and to investigate the corrosion and wear behavior, flame remelting and high frequency current remelting was performed.

2 Experimental Procedure

2.1 Materials

The powder N330-FS6 delivered by the British company LSN Diffusion used in this study is a nickel based alloy produced through water atomizing process. The chemical composition of the powder can be seen in Table 1.

Table 1
The chemical composition of the NiCrBSi employed powder

Powder	Ni [%]	Cr [%]	B [%]	Si [%]	Fe [%]
N330-FS6	bal.	6	1	1.5	0.3

Employing this technique, spherical powder particles with small internal porosities and excellent flowability are produced [12]. In Figure 1, the morphology of the NiCrBSi powder used as a raw material is presented.

The elements Cr and B, which are present in the chemical composition of the powder increase the corrosion and respective the wear resistance, due to the fact that they are carbide-forming elements. Moreover, the presence of Si and B favors the wettability and deoxidation during the remelting process [13]. Additionally, these elements decrease the rate of un-remelted particles after the deposition process and after the heat treatment [14]. As substrate material, a non-alloyed EN S355 J2 quality steel was chosen.

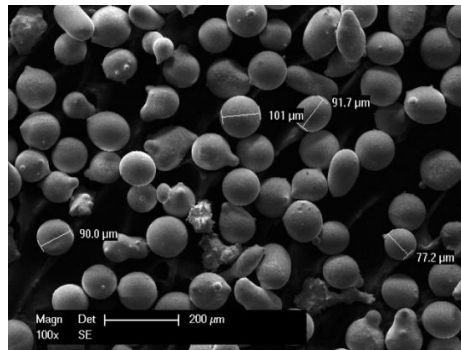


Figure 1

SEM micrograph of NiCrBSi gas atomized water collected powder

2.2 Sample Preparation

The NiCrBSi powder was sprayed using a LVOF technique, applying an oxidant flame, produced by a mix of gases, oxygen and acetylene. The coatings were deposited by the company Karl-Schumacher GmbH, Germany using a flame spraying gun produced by Metatherm, Germany. The geometry of the substrate was a cylindrical workpiece with a diameter of 70 mm and a length of 50 mm. Specimen degreasing and activation before deposition was done with alcohol followed by sandblasting creating the conditions for a good mechanical hooking of the splats. The sample was fixed in a lathe machine and the deposition process was performed. The spraying gun was advanced with a constant feed rate until a thickness of approximately 1 mm was achieved.

After the spraying process, the sample was subjected to two different remelting processes. The first sample was remelted by flame using an oxyacetylene gas process with a neutral stoichiometry. The second sample was remelted using an EKOHEAT 200/30 induction heater manufactured by Ambrell, The Netherlands. The induction heating unit requires from 15 kHz to 40 kHz frequency range. According to a similar research made by Hemmati and the team, the remelting process consisted of a preheating step [11] with a velocity of the inductor relatively four-time faster than the one used for the remelting procedure.

Figure 2 presents the inductive remelting process of the NiCrBSi coating. The as-sprayed specimen was inserted in a copper coil, fixed in a lathe machine, which was continuously rotated. The induction coil was moving from right to left with a constant speed.



Figure 2
Inductive remelting process of the NiCrBSi coating

For further analysis, metallographic samples were prepared according to the standard guide for metallographic preparation of thermal sprayed coatings [15].

3 Results and Discussion

3.1 Morphology and Porosity of the NiCrBSi Coating

In the case of as-sprayed NiCrBSi coating presented in Figure 3, it can be remarked a network of interconnected porosity and unbound particles due to the impact, rebound and contraction of the particles, with the substrate during the deposition process. Nonetheless, taking into account the technological advances, the flame-sprayed NiCrBSi coatings contain unwanted oxides, show moderate adhesion to the substrate and contain a large amount of porosity [16].

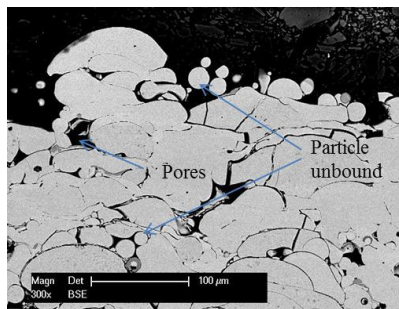


Figure 3

SEM micrograph of the as-sprayed NiCrBSi coating

The porosity degree was calculated with the aid of Leica QWin Image Processing and Analysis Software. The micrographs were acquired with an optical microscope and further processed. Figure 4 exhibits a considerably decrease of porosity from 15% in the case of as-sprayed coatings to a mean value of 1% after the flame remelting and respectively 0.5% after electromagnetic remelting treatment. The decrease of porosity in the case of the electromagnetic post-treatment can be caused by the capacity of the installation to control and set the temperature at a point close to the eutectic one where the wetting of the surface is done realizing a good gas extraction and a void closure.

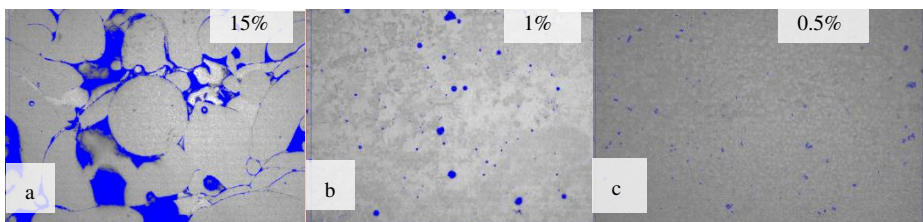


Figure 4

Optical micrographs of as-sprayed (a), flame remelted (b) and inductive remelted coatings (c)

The significant decrease of the porosity degree observed after the remelting process provides important information regarding the effectiveness of the post treatment, obtaining in this way a more compact NiCrBSi coating with a better adhesion to the substrate.

3.2 Microhardness and Tribological Behavior of the NiCrBSi Coating

Microhardness investigations were carried out on a Zwick Microhardness tester according to ISO 6507. The microhardness measurements were performed in 9 different spots along the cross-section of the NiCrBSi coating, the load used for determining the Vickers microhardness was 3 N applied for 15 seconds on each indentation and the distance between measurements was 0.15 mm. Figure 5 presents the values for the Vickers microhardness obtained for the NiCrBSi coating remelted by flame and respectively remelted using electromagnetic induction.

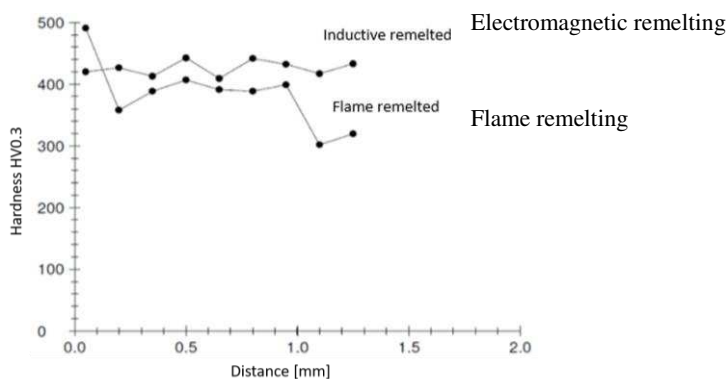


Figure 5

Microhardness values obtained for flame and electromagnetic remelted NiCrBSi coatings

A mean value of 390 HV0.3 was obtained for the flame remelted samples. For the electromagnetic fused samples 430 HV0.3 was measured with a smaller standard deviation, showing a homogenous coating along the surface. It is generally known that the hardness respectively the nanostructure of a material will directly influence its wear behavior [17]. Therefore, the sliding wear tests were performed using a pin-on-disc arrangement (POD) compliant to ASTM G99 and DIN 50324. This test design was chosen as the device is equipped with a friction coefficient transducer, providing on-line measurement of the friction coefficient. Testing parameters were maintained constant for all the samples: a counterbody of WC-Co, 6 mm diameter, 10 N load, linear speed of 15 cm s⁻¹ and 15000 laps (566 m).

Prior to the POD tests, all of the specimens were ground to a plane surface having the same roughness, cleaned with acetone and dried under warm air. The friction coefficient was continuously monitored during the tests. In this regard, the measured coefficients of the WC-Co ball against the coatings are presented in Figure 6.

For NiCrBSi coatings remelted by induction, the friction coefficient had a mean value of 0.603. Tribological tests concluded that the friction coefficient had a stable behavior, without fluctuations for the first 10000 laps and after that, it increased from 0.5 to 0.64 and stabilized at this value. The measurements showed that the NiCrBSi coating remelted by induction process exhibits a lower and steadier coefficient of friction in comparison with the flame remelted coating. The lower value of the friction coefficient obtained for the samples remelted by induction process, may be attributed to the lower porosity and a better adhesion between the particles of those samples, which leads to lower friction forces at the contact between sample and counterbody.

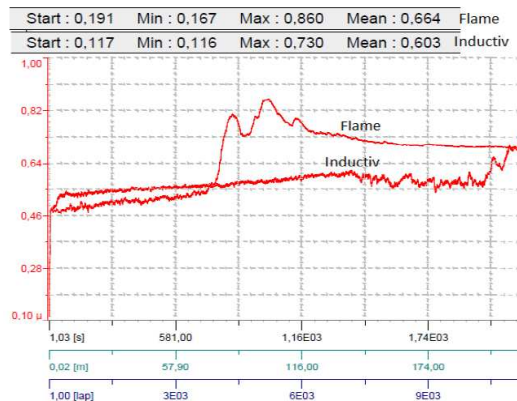


Figure 6

Friction coefficient of flame remelted and inductive remelted NiCrBSi coatings

After the test, the wear track investigations were performed on a confocal laser microscope, which facilitated the measurement of the depth of the tracks. The wear profiles of the NiCrBSi coating together with the mapping of the wear track are illustrated in Figure 7 and Figure 8.

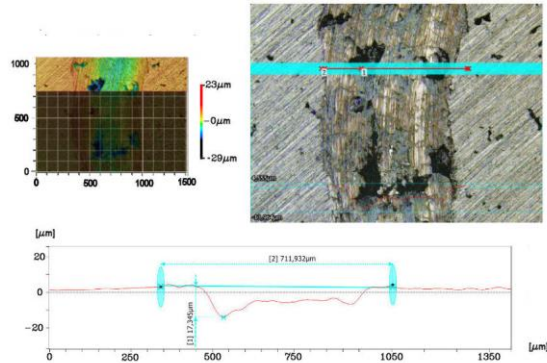


Figure 7

Wear track for sample remelted by flame

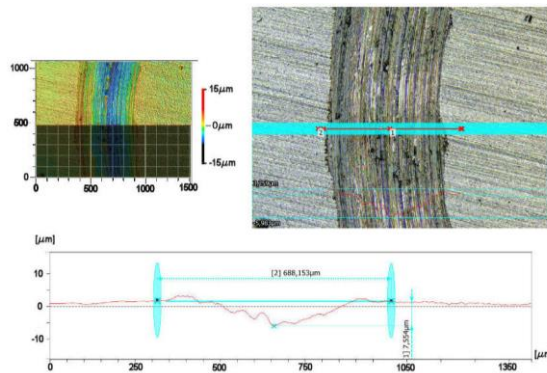


Figure 8

Wear track for sample remelted by inductive process

The wear rate calculated for induction remelted samples was $1.178 \cdot 10^{-4} \text{ mm}^3 \text{ N}^{-1} \text{ m}^{-1}$ respectively $2.338 \cdot 10^{-4} \text{ mm}^3 \text{ N}^{-1} \text{ m}^{-1}$ for the flame remelted ones. The obtained results demonstrate that a higher hardness of the tested surfaces leads in this situation to a lower material removal from the coating, giving a better resistance to wear in the case of induction remelted samples. In the previous figures one may observe that penetration depth of the ball was much lower for the inductively remelted coating ($7.5 \text{ } \mu\text{m}$) for the inductive remelted coating respectively $17.3 \text{ } \mu\text{m}$ for the flame fused coating).

3.3 Adhesive Properties

The adhesion to the substrate is one of the most important properties of the coated components. In order to determine if the coating has a good adhesion to the

substrate, the samples were investigated by realizing indentations on the substrate-coating interface.

The indentations were performed in different spots of the cross-section along the interface between the NiCrBSi coating and the substrate. The load used for examining the adhesion was 1200 N applied for 15 seconds on each test. In Fig. 9, before the remelting process, delamination, low mechanical bonding and cracks can be observed at the interface and in the coating. Figure 10 illustrates the indentations on both samples.

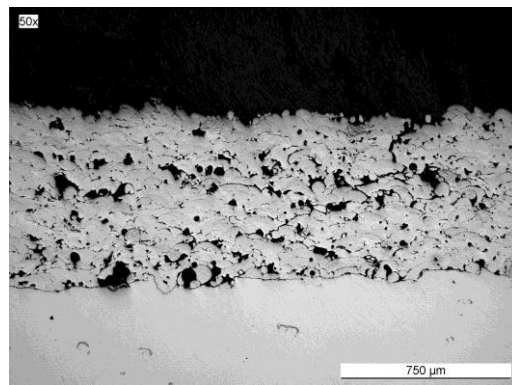


Figure 9

As-sprayed coating and interface overview

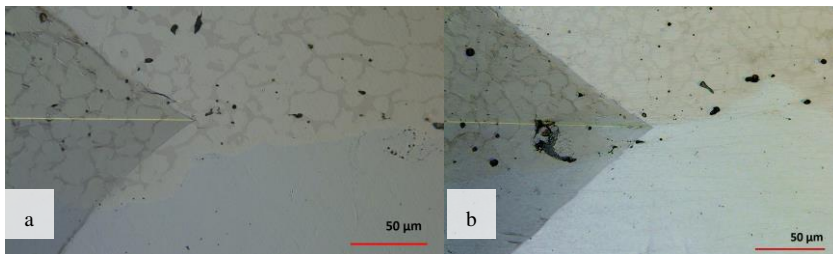


Figure 10

HV 120 indentation for flame (a) and inductive remelted coating (b)

The results show that both remelting processes offered a good adhesion to the substrate and delamination did not occur.

3.4 Corrosion Behavior

The corrosion behavior of the NiCrBSi coatings was investigated by electrochemical polarization in a three electrode cell in 3.5% NaCl aqueous solution. A saturated calomel electrode (SCE) was used as reference electrode, a

platinum electrode as auxiliary electrode and the NiCrBSi samples represented the working electrode. The samples were polarized in a potential interval from - 500 mV to +500 mV, with a scan rate of 0.16 mVs⁻¹.

Table 2
Corrosion potential and corrosion current densities of NiCrBSi coatings

Sample	E _{Corr} vs SCE (i=0) [mV]	i _{Corr} [mA cm ⁻²]
Flame	-268	91 10 ⁻⁴
Inductive	-217	27 10 ⁻⁴

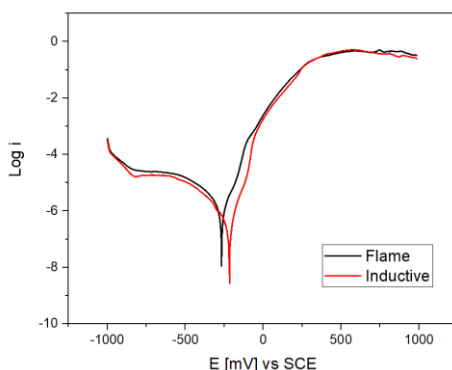


Figure 11

Polarization curves of the flame respectively inductively fused NiCrBSi coatings tested in 3.5% NaCl aqueous solution

It is important to mention that Bergant *et al.* have demonstrated in previous studies that the Ni-based treated coatings exhibit a better corrosion resistance with a factor of 10 compared to the substrate and the as-sprayed one [18]. Considering this data, the analyzed samples were the post-treated ones. The results presented in Table 2 and the polarization curves illustrated in Figure 11 reveal a higher corrosion rate for the NiCrBSi coating remelted by flame process compared to the inductive remelted one. As can be noticed, the corrosion current density i_{corr} is three times lower in the case of the inductive remelted sample ($\sim 27 \cdot 10^{-4} \text{ mA cm}^{-2}$) compare to the flame remelted one ($\sim 91 \cdot 10^{-4} \text{ mA cm}^{-2}$). The corrosion potential E_{corr} of the inductive post-treated sample which is slightly shifted to a positive value than the flame is another sign of superiority of the red marked coating. Regarding the anodic region of the graph, one can see that the curves present a similar behavior, presenting a slight repassivation plateau with an overlapping point at about 50 mV.

4 Conclusions

This work presents a new approach for obtaining NiCrBSi coatings with low porosity and good adhesion to substrate, using a two-step deposition process. As-sprayed, NiCrBSi coatings present high porosity, low cohesion between particles and poor adhesion to the substrate, as well as irregular surface geometry. The coatings which were remelted by induction are more compact and the porosity decreased with a factor of 30, from the as-sprayed one to the induction treated ones.

The microhardness investigations revealed that the inductive remelted NiCrBSi coatings do not present fluctuation of hardness and the mean value is 10% higher than the flame remelted ones. This fact can be attributed to a more compact microstructure with lower porosity, and a finer distribution of the phases.

Regarding the wear behavior, the wear rate of the electromagnetic samples was more than two times lower than the flame remelted one. The electrochemical corrosion tests showed that inductive remelted specimen, which is treated with a method that allows a precise control of the temperature and energy and exhibits a 3X better corrosion resistance than the flame remelted samples.

References

- [1] J. Davies, Handbook of Thermal Spray Technology, Materials Park, OH, USA: ASM International, 2004
- [2] P. Mrva and D. Kottfer, "Influence of Thermal Spray Coatings on the Thermal Endurance of Magnesium Alloy ML-5", *Acta Polytechnica Hungarica*, Vol. 6, No. 2, pp. 71-75, 2009
- [3] I. Hemmmati, R. Huizenga, V. Ocelik and J. D. Hosson, "Microstructural design of hardfacing Ni-Cr-B-Si-C alloys", *Acta Materialia*, Vol. 61, No. 16, pp. 6061-6070, 2013
- [4] S. Houdkova, M. Vostrak, M. Hruska, et al. "Comparison of NiCrBSi coatings, HVOF sprayed, re-melted by flame and by high-power laser", Proceedings of the 22nd International Conference on Metallurgy and Materials, May 15th -17th, Brno, Czech Republic, 2013
- [5] F. Battez, J. L. Viesca, R. Gonzales, et al. "Friction reduction properties of a CuO nanolubricant used as a lubricant for a NiCrBSi coating", *Wear*, Vol. 268, pp. 325-328, 2010
- [6] M. C. Lin, L. S. Chang, H. C. Lin, et al. "A study of high-speed slurry erosion of NiCrBSi thermal-sprayed coating", *Surface & Coatings Technology*, Vol. 201, pp. 3193-3198, 2006
- [7] S. Houdkova, F. Zahalka, M. Kasparova and L. Berger, "Comparative study

- of thermally sprayed coatings under different types of wear conditions for hard chromium replacement”, *Tribology Letters*, Vol. 43, pp. 139-154, 2011
- [8] N. Kazamer, D. Pascal, V. Serban and G. Marginean, “A comparison between hardness, corrosion and wear performance of APS sprayed WC-CoMo and WC-Co coatings”, *Solid State Phenomena*, Vol. 254, pp. 71-76, 2016
- [9] B. Zoran, T. Uros and G. Janez, “Effect of high-temperature furnace treatment on the microstructure and corrosion behavior of NiCrBSi flame-sprayed coatings”, *Corrosion Science*, Vol. 88, pp. 372-386, 2014
- [10] L. Vieira, H. Voorwald and M. Cioffi, “Fatigue Performance Of AISI 4340 Steel Ni-Cr-B-Si-Fe HVOF Thermal Spray Coated”, *Procedia Engineering*, Vol. 114, pp. 606-612, 2015
- [11] H. I. V. Ocelik and J. D. Hosson, “Effects of the alloy composition on phase constitution and properties of laser deposition Ni-Cr-B-Si coatings”, *Physics Procedia*, Vol. 41, pp. 302-311, 2013
- [12] L. Pawlowski, “The Science and Engineering of Thermal Spray Coatings”, 2nd edition, Wiley, 2008
- [13] ASM International, Alloy Phase Diagram, Vol. 3, ASM Handbook, 2005
- [14] Z. Bergant, J. Grum, “Quality improvement of flamesprayed, heat treated and remelted NiCrBSi coatings”, *Journal of Thermal Spray Technology*, Vol. 18, pp. 380-391, 2009
- [15] Standard Guide for Metallographic Preparation of Thermal Sprayed Coatings, ASTM Standard E1920-03, West Conshohocken, PA: ASM International, 2003
- [16] Bergant Z. and Grum J., “Porosity evaluation of flame-sprayed and heat-treated nickel-based coatings using image analysis”, *Image Analysis & Stereologz*, Vol. 30(1), pp. 53-62. 2011
- [17] I. Konyashin, B. Ries, D. Hlawatschek, Y. Zhuk, A. Mazilkin, B. Straumal, F. Dorn and D. Park, “Wear-resistance and hardness: Are they directly related for nanostructured hard materials?”, *International Journal of Refractory Metals and Hard Materials*, Vol. 49, pp. 203-211, 2015
- [18] Z. Bergant, U. Trdan and J. Grum., “Effect of high-temperature furnace treatment on the microstructure and corrosion behaviour of NiCrBSi flame-sprayed coating”, *Corrosion Science*, Vol. 88, pp. 372-386, 2014

Potential Therapeutic Modalities of Reawakening Fetal Hemoglobin Simulated by Reaction Systems

Mani Mehraei¹, Benedek Nagy¹, Nimet Ilke Akcay²,
Şükrü Tüzmen³

¹ Department of Applied Mathematics and Computer Science,
Eastern Mediterranean University, Famagusta, North Cyprus, Mersin-10, Turkey,
mani.mehraei@emu.edu.tr, benedek.nagy@emu.edu.tr

² Faculty of Medicine, Eastern Mediterranean University,
Famagusta, North Cyprus, Mersin-10, Turkey, ilke.cetin@emu.edu.tr

³ Department of Biological Sciences, Eastern Mediterranean University,
Famagusta, North Cyprus, Mersin-10, Turkey, sukru.tuzmen@emu.edu.tr

Abstract: Thalassemia syndromes are a diverse group of inherited genetic disorders. There are different types of thalassemia disorders, such as, β -thalassemia, which is also called Mediterranean anemia, that is an inherited disease that played a major role in the American thriller movie, "Dying of the light" starring Nicolas Cage (Dec. 2014). In this study, we focus on the beta-globin (β -globin) gene family related disorders. We seek potential amelioration strategies for β -thalassemia and sickle cell anemia via γ -globin gene induction. In this work, a simulation model is developed, utilizing a reaction systems methodology. These systems are finite and based on a discrete time scale and can be used to describe and analyze complex biological systems and biological phenomenon. In our model, simulations of normal and abnormal cases of fetal, to adult hemoglobin switching developmental stage are illustrated. Various types of known and potential treatment strategies for β -thalassemia and sickle cell anemia cases from the literature have been utilized to validate our model, used for identifying new potential treatments to be tested by molecular biologists, in the future studies. Moreover, we propose a novel potential simulation, as a therapeutic means, for β -thalassemia and sickle cell anemia, by identifying FOG1 as a potential target. Finally, our proposed model, based on a reaction systems methodology, shows that inhibition of FOG1 expression by using methods, such as, RNAi induces γ -globin gene expression and can compensate for the lack of beta-globin in patients suffering from β -globin gene related diseases, such as, β -thalassemia and sickle cell anemia.

Keywords: bioinformatics; hemoglobin switching; beta-globin; beta-thalassemia; reaction systems; simulation; modeling biological systems

1 Introduction

Diseases related to the hemoglobin (Hb) protein can result from either structural aberrations, that give rise to “hemoglobinopathies”, or defects in the synthesis of one or more of the polypeptide chains of Hb, which lead to thalassemias [17]. The thalassemia syndromes are a diverse group of inherited disorders that can be characterized according to their insufficient synthesis or absent production of one or more of the globin chains. Here we will emphasize the beta-globin (β -globin) gene family related disorders, especially the potential amelioration strategies for β -thalassemias via γ -globin gene upregulation. To date more than 200 mutations are reported causing various levels of β -globin gene defects [17], which are known to produce β -thalassemia. β -thalassemia and sickle-cell anemia diseases are caused by mutation(s) in β -globin gene, which result in defective adult Hemoglobin (HbA) and lead to various abnormal phenotypes. β -thalassemia itself is a common blood disorder worldwide, especially endemic in regions such as Mediterranean basin, Middle Eastern countries, Central and South Eastern Asia, Northern part of Africa, and India. Every year, thousands of infants worldwide are born with this disease.

β -thalassemia is an inherited blood disorder and the global annual incident rate has been reported to be 1 in 100,000 live births [8]. Thus, preventing it by educating and informing people, and introducing novel treatment strategies is essential. Hemoglobin Switching refers to a developmental stage of globin gene regulation (Fig. 1). In the case of the fetal to adult globin gene expression switch, γ -globin gene expression is up-regulated during the first six months of gestational age before birth, and then starts going down, continuing after birth. This gradual down-regulation of β -globin gene is compensated by the gradual up-regulation of γ -globin gene expression starting from just before the third month of gestational age before birth, and continuing to be up-regulated after birth replacing γ -globin gene production in a healthy human being. Overall, this phenomenon results in down regulation of fetal hemoglobin (HbF) and up regulation of adult hemoglobin (HbA) in a healthy adult developmental stage (Fig. 1). In case of any defect in β -globin gene, this picture is presented with a lack of or no production of HbA. These defects should be supported/corrected by means of treating the individuals with these syndromes, utilizing blood transfusions and certain drug treatments. Therefore, one of the strategies in recent decades was related to compensating the lack or loss of HbA by inducing γ -globin gene production, which would result in up-regulation of Fetal Hemoglobin (HbF) molecule in turn replacing the lost function of the HbA [2].

Facilitation and inhibition are two fundamental mechanisms for functioning biochemical reactions [6]. Reaction Systems are formal models that work on a discrete timescale in a parallel manner to investigate and analyze biochemical reactions and the interaction among them [7]. A formal reaction has been considered as a triplet $a = (R, I, P)$, where R is the set of reactants, I is the set of

inhibitors, and P is the set of products [6]. A reaction is enabled if all elements of set R , are present and all elements of set I are absent in the actual “state” of the system. If a reaction goes, then the result will be set P , as the product that is set R produced the elements of set P . Otherwise a reaction is not enabled and the product of such reaction is the empty set. For example, if we define reactions $a_1 = (\{A,B\},\{C\},\{A,D\})$, $a_2 = (\{E\},\{A\},\{A,B\})$, and initial state to be $\{A,B,E\}$, then reaction a_1 is enabled since both reactants A and B are available and there is no C present to inhibits the reaction. Thus, the products of reaction a_2 are produced which are A and D . On the other hand, a_2 is not enabled since A as an inhibitor is available and does not let this reaction to go. Therefore, the result will be union of $\{A, D\}$ and empty set which is the set of products $\{A, D\}$.

In a biological system, there are phenomena that can easily be represented in a binary way, e.g., a gene can be down regulated or upregulated, a complex could be formed or being in parts. Therefore, we believe that relatively simple models as Reaction Systems are opt to model some biological phenomena. In this paper, we have exploited a Reaction System protocol to simulate hemoglobin switching process in case, where a genetic defect occurs in β -globin gene. This enabled us to validate qPCR data of known and experimental drugs to show the efficacy of these strategies. Ultimately, by considering a novel initial state, we came up with new strategy, which can be used to increase γ -globin gene induction.

2 Biological Context (Targets for Various Potential Therapeutic Models)

In a healthy human adult, c-Myb as a member of MYB family activates transcription of KLF1 gene, which is a transcription factor needed for γ -globin gene expression. KLF1 binds to promoter of BCL11A gene and activates the transcription of BCL11A gene [13]. BCL11A protein binds to NuRD complex, which contains HDAC1/2, CHD3/4, and MBD2 [10, 17]. BCL11A together with NuRD complex physically interact with SOX6 and has molecular interaction with FOG1 and GATA1 transcription factors [18]. These transcription factor complexes are also essential for γ -globin gene expression. Finally, this multi-protein complex which includes NuRD complex, BCL11A, and Erythroid Transcription Factors, SOX6, FOG1, and GATA1 [3] inhibits γ -globin gene expression [17]. This process is illustrated in Fig. 2.

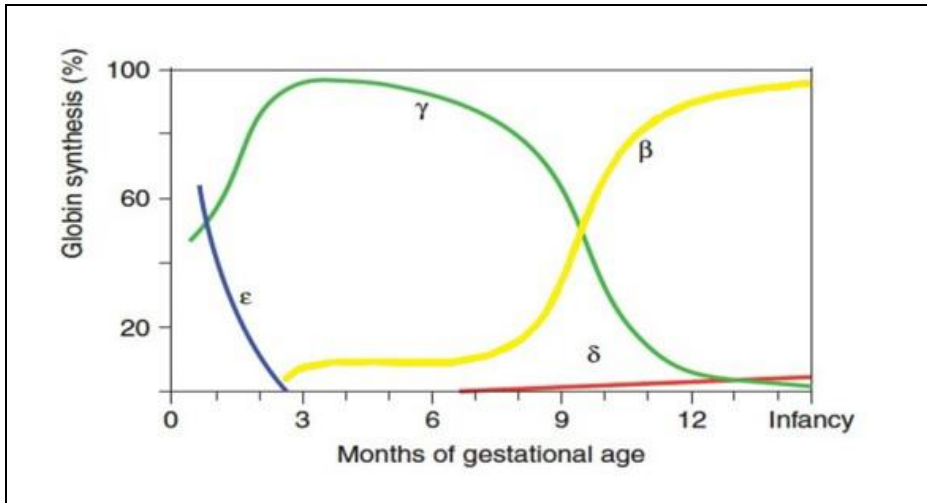


Figure 1
Fetal to adult hemoglobin switching (Adapted from [17])

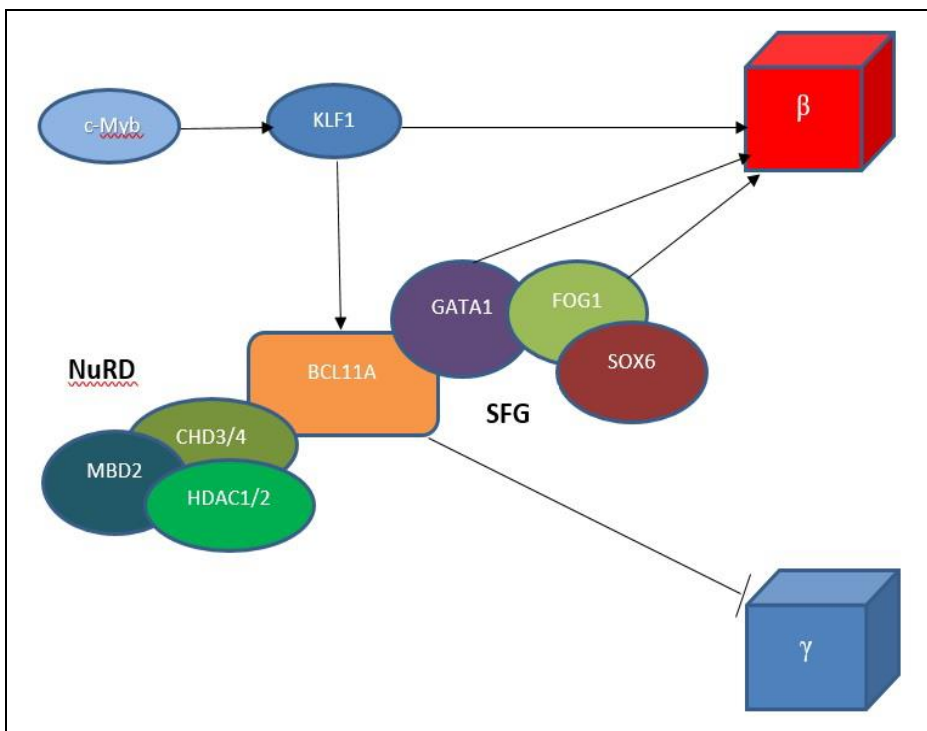


Figure 2
Hemoglobin switching pathway (Adapted from [3])

3 Exploiting Reaction Systems to Model Hemoglobin Switching Process

As mentioned earlier, a reaction can be shown as a triplet $a = (R, I, P)$, where R is the set of reactants, I set of inhibitors, and P is the set of products [6]. A reaction system is a pair (O, A) of sets of objects and reactions. To model our system in term of reaction systems we are considering these objects as possible elements of the set $O = \{C, K, B, H, N, F, G, S, SFG, \beta, \gamma\}$. Where objects corresponding to these symbols are illustrated in Table 1.

If one, or some of these objects appear to be in a state or in reactions, it means those certain gene expressions are upregulated, otherwise disappearing or nonexistence means down regulation. We have not considered other elements of NuRD complex to emphasize mostly on the role of HDAC1/2 in γ -globin gene induction. Furthermore, SOX6 along with GATA1 and FOG1 are called erythroid transcription factors (SFG) which are interacting with BCL11A [3, 19].

The reactions, i.e., the elements of A in our system:

$$\begin{array}{ll}
 a_1 = (\{C\}, \{\}, \{C, K\}) & a_2 = (\{C, K\}, \{\}, \{C, K, B\}) \\
 a_3 = (\{S, F, G\}, \{\}, \{S, F, G, SFG\}) & a_4 = (\{H\}, \{\}, \{H, N\}) \\
 a_5 = (\{B, N, SFG, \gamma\}, \{\}, \{B, N, SFG\}) & a_6 = (\{K, G, F\}, \{\}, \{K, G, F, \beta\}) \\
 a_7 = (\{B\}, \{K\}, \{B\}) & a_8 = (\{\}, \{B\}, \{\gamma\}) \\
 a_9 = (\{N\}, \{H\}, \{N\}) & a_{10} = (\{\}, \{N\}, \{\gamma\}) \\
 a_{11} = (\{SFG\}, \{S, F, G\}, \{SFG\}) & a_{12} = (\{\}, \{SFG\}, \{\gamma\}) \\
 a_{13} = (\{G\}, \{\}, \{G\}) & a_{14} = (\{F\}, \{\}, \{F\}) \\
 a_{15} = (\{C\}, \{\}, \{C\}) & a_{16} = (\{S\}, \{\}, \{S\})
 \end{array}$$

Reaction a_1 illustrates that up regulation of C-Myb results into up regulation of KLF1. Reaction a_2 shows that up regulation of C-Myb and KLF1 leads to up regulation of BCL11A. Then, a_3 illustrates that up regulation of SOX6, FOG1, and GATA1 indicates up regulation of these erythroid transcription factors as a complex; and a_4 shows that up regulation of HDAC1/2 leads to up regulation of NuRD complex. We have not considered other components of NuRD complex in this system. Reaction a_5 indicates that binding of BCL11A with NuRD complex and erythroid transcription factors GATA1, FOG1, and SOX6 leads to inhibition of γ -globin gene expression. Further, a_6 shows that KLF1, GATA1, and FOG1 are transcription factors of β -globin gene. Then a_7 mentions that down regulation of KLF1 leads to down regulation of BCL11A. The reactions a_8 , a_{10} , and a_{12} indicate that down regulation of either BCL11A, NuRD complex, or erythroid transcription factors GATA1, FOG1, and SOX6 leads to γ -globin gene induction; while a_9 shows that down regulation of HDAC1/2 leads to down regulation of NuRD complex. Further, a_{11} indicates that down regulation of either GATA1, FOG1, or SOX6 leads to down regulation of these erythroid transcription factors as a complex. Finally, a_{13} , a_{14} , a_{15} , and a_{16} indicate that if GATA1, FOG1, C-MYB, or SOX6 are upregulating, they will remain to be upregulated.

Table 1
Descriptions related to objects in set O

Object's symbol	Object
C	C-Myb
K	KLF1
B	BCL11A
H	HDAC1/2
N	NuRD complex
F	FOG1
G	GATA1
S	SOX6
SFG	Erythroid Transcription Factors SOX6, FOG1, and GATA1
β	Beta globin
γ	Gamma globin

The reaction system $H_0 = (O, A)$ is able to describe (simulate) the hemoglobin switching process of a healthy person. We considered initial state for the normal case of fetal hemoglobin switching to be $\{C, H, G, F, S, \gamma\}$. Then reactions indexed by 1, 3, 4, 8, 10, 12, 13, 14, 15, and 16 are allowed. After the first round, new state $\{C, K, H, N, G, F, S, SFG, \gamma\}$ is obtained. Then reactions with indices 1, 2, 3, 4, 6, 8, 13, 14, 15, 16 are allowed. After the second round, new state is $\{C, K, B, H, N, G, F, S, SFG, \beta, \gamma\}$. Then reactions with numbers 1, 2, 3, 4, 5, 6, 13, 14, 15, and 16 are enabled. After third round, the new state of the system is $\{C, K, B, H, N, G, F, S, SFG, \beta\}$. This is the last state which is a fix point of the system and it does not change anymore. That is β has started to be up regulated while γ is down regulating. The result of the simulation is the expected outcome for the healthy case.

4 Extended Reaction System to Deal with Disease and Treatment Cases

To extend Reaction System to cover cases related to β -thalassemia and treatment options, new objects are added and reactions are edited to the previous ones: $O' = \{I, D, M\}$, where objects corresponding to these symbols are illustrated on Table 2.

Table 2
Descriptions related to objects in set O'

Object's symbol	Object
I	BCL11A down regulator
D	KLF1 deactivator
M	Beta globin gene mutation

The reactions of our extended reaction system are as follows:

$$\begin{array}{ll}
 a'_1 = (\{C\}, \{D\}, \{C, K\}) & a'_2 = (\{C, K\}, \{I\}, \{C, K, B\}) \\
 a'_3 = (\{S, F, G\}, \{\}, \{S, F, G, SFG\}) & a'_4 = (\{H\}, \{\}, \{H, N\}) \\
 a'_5 = (\{B, N, SFG, \gamma\}, \{\}, \{B, N, SFG\}) & a'_6 = (\{K, G, F\}, \{M\}, \{K, G, F, \beta\}) \\
 a'_7 = (\{B\}, \{K\}, \{B\}) & a'_8 = (\{\}, \{B\}, \{\gamma\}) \\
 a'_9 = (\{N\}, \{H\}, \{N\}) & a'_{10} = (\{\}, \{N\}, \{\gamma\}) \\
 a'_{11} = (\{SFG\}, \{S, F, G\}, \{SFG\}) & a'_{12} = (\{\}, \{SFG\}, \{\gamma\}) \\
 a'_{13} = (\{G\}, \{\}, \{G\}) & a'_{14} = (\{F\}, \{\}, \{F\}) \\
 a'_{15} = (\{M\}, \{\}, \{M\}) & a'_{16} = (\{C\}, \{\}, \{C\}) \\
 a'_{17} = (\{D\}, \{\}, \{D\}) & a'_{18} = (\{I\}, \{\}, \{I\}) \\
 a'_{19} = (\{S\}, \{\}, \{S\}) &
 \end{array}$$

In reaction a'_1 , D stands for KLF1 deactivator. When D is presents, it doesn't let KLF1 to be transcribed so that its level decreases. In a'_2 , I plays the role of BCL11A down regulator. Thus, whenever I is present, KLF1 cannot activate transcription of BCL11A. In a_6 , when mutation has happened in β -globin gene, transcription factors of β are not able to transcribe this DNA to mRNA anymore, so that β -globin gene expression decreases. a'_{15} , a'_{17} , and a'_{18} show that if mutation has happened, KFL1 deactivator exist, or BCL11A down regulator is present, they remain present.

The reaction system $H = (O \cup O', \{a'_1, a'_2, \dots, a'_{19}\})$ is able to describe (simulate) the hemoglobin switching process also in a patient who has β -globin gene mutation.

5 Simulation Results

In this section, simulation results in a severe β -thalassemia patient with possible treatments found in the literature, are illustrated. Moreover, potential therapeutic modalities are proposed by referring to Reaction System simulation results. For obtaining the simulation results, we have written a C++ program on a personal computer.

5.1 Simulation of β -globin Gene Mutations

To simulate the case when mutations occur in β -globin gene, we have considered initial state to be $\{C, H, G, F, S, \gamma, M\}$. Then reactions with indices 1, 3, 4, 8, 10, 12, 13, 14, 15, 16, and 19 are allowed. Thus, after these reactions take place, after the first round, the new state $\{C, K, H, N, G, F, S, SFG, \gamma, M\}$ is reached. In this state reactions indexed by 1, 2, 3, 4, 8, 13, 14, 15, 16 and 19 are allowed. After second round, the new state is $\{C, K, B, H, N, G, F, S, SFG, \gamma, M\}$. Then reactions indexed by 1, 2, 3, 4, 5, 7, 13, 14, 15, 16 and 19 are enabled. Then the last state which is the fix point of the system and does not change anymore is $\{C, K, B, H, N, G, F, S, SFG, M\}$. Thus, in case of mutation in β -globin gene, expression of β -globin gene is down regulated. Moreover, γ -globin gene expression is down regulated after fetal to adult hemoglobin switching. The steps are illustrated in Fig. 3. As one can observe the simulation gives exactly the expected result also for the ill case.

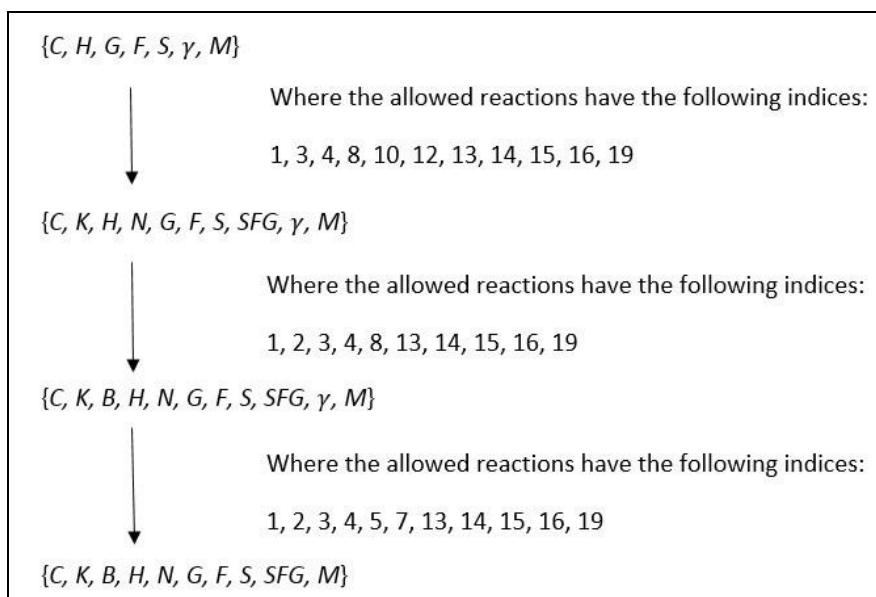


Figure 3

Simulation of steps in Reaction Systems in case of mutation in β -globin gene

5.2 Validation of Current Strategies to Induce γ -globin Gene Expression

5.2.1 HDAC1/2 Inhibition

Drugs such as Lovastatin [11], Romidepsin, and Vorinostat [5] can be used as inhibitors of Histone deacetylase enzyme. By eliminating “H” from initial state, we can simulate this case. Initial state is $\{C, G, F, S, \gamma, M\}$. Then reactions indexed by 1, 3, 8, 10, 12, 13, 14, 15, 16 and 19 are enabled. After first round, new state is $\{C, K, G, F, S, SFG, \gamma, M\}$. Then reactions 1, 2, 3, 8, 10, 13, 14, 15, 16, and 19 are allowed. After second round, the new state $\{C, K, B, G, F, S, SFG, \gamma, M\}$ is obtained. Then reactions with numbers 1, 2, 3, 10, 13, 14, 15, 16 and 19 are allowed, which leads to the same state that is the fix point of the system and does not change anymore. Thus HDAC1/2 inhibition has resulted in γ -globin gene expression induction. The simulation steps are shown in Fig. 4. The simulation works well in this case too.

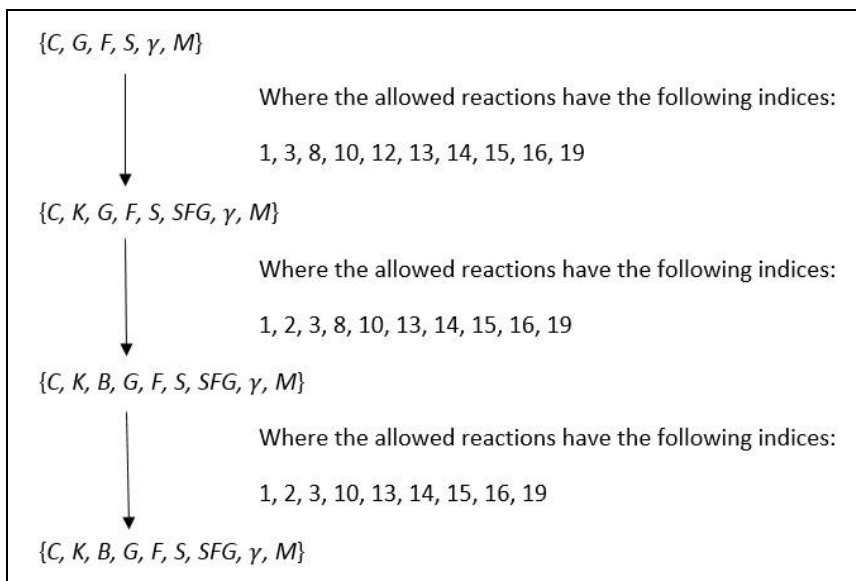


Figure 4

Simulation of steps in Reaction Systems in case of HDAC1/2 inhibition

5.2.2 KLF1 Down-Regulation

A recent example of KLF1 down regulator is herbal drug Ninjin'yoeito [10]. To simulate it, we have considered initial state to be: $\{C, D, H, G, F, S, \gamma, M\}$. Then reaction numbers 3, 4, 8, 10, 12, 13, 14, 15, 16, 17 and 19 are allowed. After the

first round, the state $\{C, D, H, N, G, F, S, SFG, \gamma, M\}$ is reached. Then reactions 3, 4, 8, 13, 14, 15, 16, 17 and 19 are allowed. The last state, which is the fix point of the system and does not change anymore is the same as previous one $\{C, D, H, N, G, F, S, SFG, \gamma, M\}$, which means down regulation of KLF1 has resulted in γ -globin gene expression induction. Down-regulation of KLF1 leads to down regulation of β -globin gene, which results in the induction of γ -globin gene expression. Thus, this scenario might be useful only in the case of severe β -thalassemia cases, where the production of β -globin gene is already defective or no β -globin is expressed at all. The stages are explained in Fig. 5.

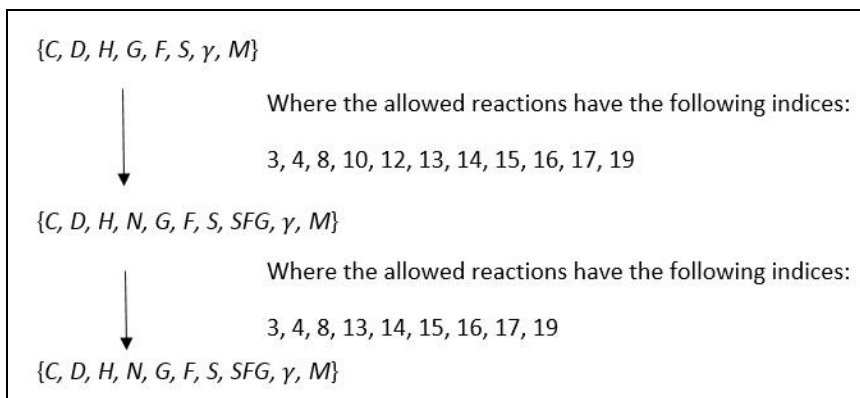


Figure 5

Simulation of steps in Reaction Systems in case of KLF1 down-regulation

5.2.3 BCL11A Down-Regulation

Hydroxyurea (HU) [9] and simvastatin together with t-BHQ decrease expression of BCL11A gene [13]. To simulate it, we have defined object I to represent inhibition of BCL11A. To simulate such case, we considered initial state to be $\{C, H, G, F, S, \gamma, M, I\}$. Then reactions 1, 3, 4, 8, 10, 12, 13, 14, 15, 16, 18 and 19 are allowed. New state after first round is $\{C, K, H, N, G, F, S, SFG, \gamma, M, I\}$. Then reactions 1, 3, 4, 8, 13, 14, 15, 16, 18 and 19 are enabled. The last state which is the fix point of the system which does not change anymore is $\{C, K, H, N, G, F, S, SFG, \gamma, M, I\}$. Thus, the simulation shows that down regulation of BCL11A leads to γ -globin gene expression induction. The simulation stages are illustrated in Figure 6. Our model also captures well that type of treatment.

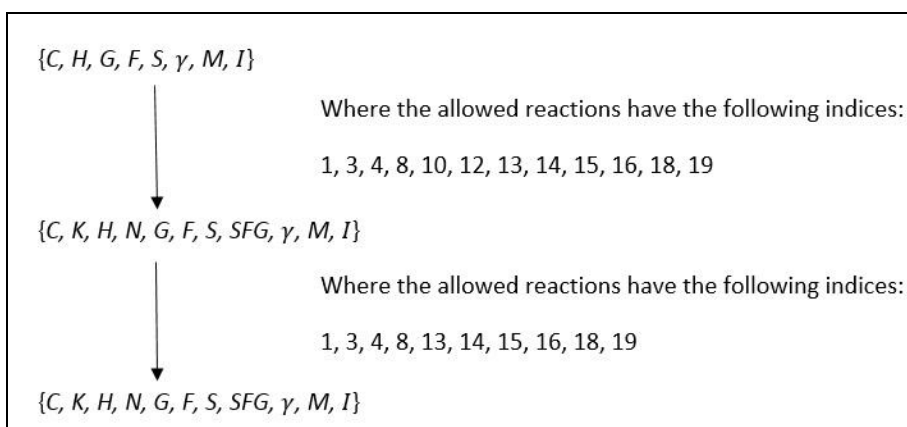


Figure 6

Simulation of steps in Reaction Systems in case of BCL11A down-regulation

5.2.4 SOX6 Down-Regulation

There are drugs such as HU which their utilization leads to SOX6 down regulation [9]. To simulate, we may consider initial state as: $\{C, H, G, F, \gamma, M\}$. Then reactions indexed by numbers 1, 4, 8, 10, 12, 13, 14, 15, and 16 are allowed. After the first round, new state is $\{C, K, H, N, G, F, \gamma, M\}$. Then reactions 1, 2, 4, 8, 12, 13, 14, 15, and 16 are allowed. After second round, new state is $\{C, K, B, H, N, G, F, \gamma, M\}$. Then reactions indexed by 1, 2, 4, 12, 13, 14, 15, and 16 are allowed. After third round, new state is the same as previous one, which is the fix point of the system and does not change anymore. Thus, down regulation of SOX6 led to γ -globin gene induction. The simulation steps are shown in Fig. 7. The model gives the expected result.

5.2.5 GATA1 Inhibition

GATA1 can reverse γ -globin gene silencing [19]. To simulate, we may consider initial state as: $\{C, H, S, F, \gamma, M\}$. Then reactions numbered by 1, 4, 8, 10, 12, 14, 15, 16 and 19 are allowed. After first round, the new state is $\{C, K, H, N, S, F, \gamma, M\}$. Then reactions indexed by 1, 2, 4, 8, 12, 14, 15, 16 and 19 are allowed. After second round, new state is $\{C, K, B, H, N, S, F, \gamma, M\}$. Then reactions 1, 2, 4, 12, 14, 15, 16 and 19 are allowed. After third round, new state is the same previous which is the fix point of the system and does not change anymore. Therefore, GATA1 inhibition leads to γ -globin gene expression induction. The simulation steps are demonstrated in Fig. 8.

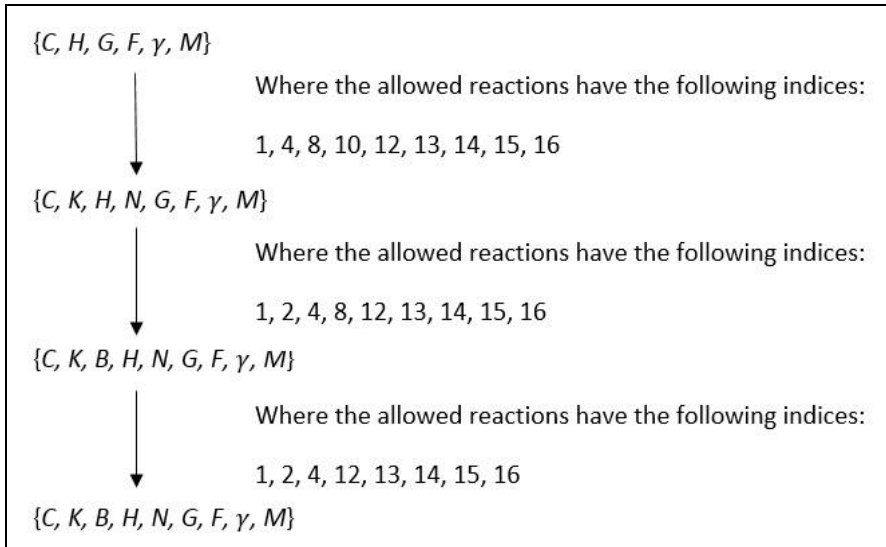


Figure 7

Simulation of steps in Reaction Systems in case of SOX6 down-regulation

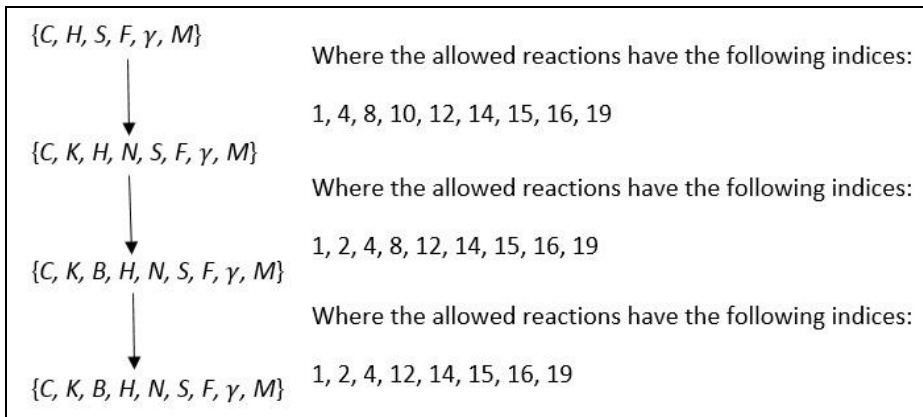


Figure 8

Simulation of steps in Reaction Systems in case of GATA1 inhibition

5.3 Predicting Result of a Novel Strategy in Favor of Gamma-globin Gene Expression Induction

In this subsection we identify a new possibility for treatment of β -thalassemia based on our model.

5.3.1 RNA Interference (RNAi)

A powerful methodology for studying the function of any gene is to experimentally disrupt its expression to examine the resulting phenotype. RNA interference (RNAi) is a naturally occurring mechanism of gene regulation. This can be induced by the introduction of double stranded RNA into a cell. This event can be synthetically utilized to down-regulate expression of specific genes by transfecting mammalian cells with synthetic short interfering RNAs (siRNAs). These siRNAs can be designed to silence the expression of genes of interest having a certain target sequence, and may potentially be presented as a therapeutic strategy for inhibiting transcriptional regulation of genes [15].

5.3.2 Inhibition of FOG1 by RNAi Methodology

To simulate, we may consider initial state as: $\{C, H, G, S, \gamma, M\}$. Then reactions indexed by 1, 4, 8, 10, 12, 13, 15, 16 and 19 are allowed. After first round, new state is $\{C, K, H, N, G, S, \gamma, M\}$. In this state the reactions indexed by 1, 2, 4, 8, 12, 13, 15, 16 and 19 are allowed. After second round, the new state is $\{C, K, B, H, N, G, S, \gamma, M\}$. Then reactions numbered by 1, 2, 4, 12, 13, 15, 16 and 19 are allowed. After third round, the new state is the same as previous one, which is the fix point of the system and does not change anymore. Therefore, FOG1 inhibition has led to γ -globin gene expression induction as a possible. The steps of this simulation process are illustrated in Fig. 9.

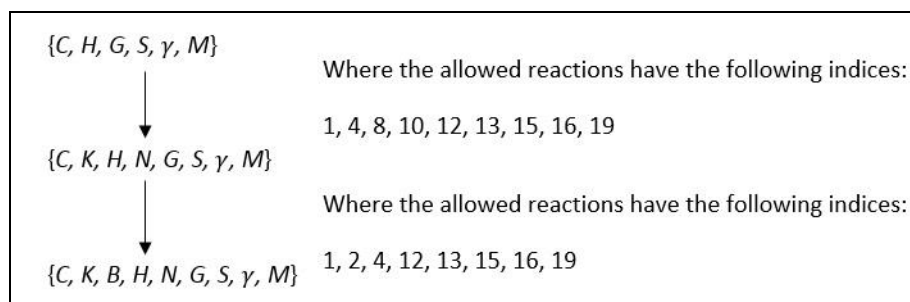


Figure 9

Simulation of steps in Reaction Systems in case of FOG1 inhibition by RNAi

6 Discussion

In this section, we give a brief comparison of our simulation model and other models. Various simulation techniques were used for β -thalassemia, e.g., in [1, 16]. However, in these papers the authors only wanted to simulate and understand

the disease itself, simulating how the gene can produce hemoglobin, and what is wrong in case of β -thalassemia, without proposing or checking any possible treatments.

The nature of the considered problem results relatively simple simulations, after a few steps the system halts in each case. The same phenomenon occurs by simulation with Petri nets [14], the simulation goes without any cycles. In [14] hybrid places (both discrete and continuous) were used instead of only discrete ones. Therefore, instead of a discrete value (integer number of tokens), real numbers were used at each place. In this way, in the simulation Petri times were used, e.g., the simulation program ran for 500 Petri times. The simulation results of Subsections 5.2.1, 5.2.2, 5.2.3 and 5.2.4 are in line with the simulation results with the Hybrid Functional Petri nets model, where HDAC1/2 gene expression was inhibited using ST-20 drug, KLF-1 gene expression was inhibited using MS-275, ST-20, and the combination of Simvastatin and tBHQ drugs, BCL11A gene expression was inhibited using ACY-957 drug and SOX6 gene expression was inhibited using ACY-957 drug, respectively [14]. Our reaction systems model had the advantage to simplify all this. We could answer whether a treatment can be useful or not in just a few steps. The advantage of the binary feature of the reaction systems approach is that it simplifies the system analysis. Such method can be used as a pre-test to identify potential entities which can play an important role to up-regulate or down-regulate a specific component in the system. After identifying the entities which play an important role, it will be possible to extend the model into a model which is appropriate to investigate the role of these components more accurately (in terms of quantities, ratios, concentrations etc.). The disadvantage of our binary approach could be that it cannot investigate the role of each entity in the system quantitatively. Although, this approach shows which manipulation of the system can be useful to up-regulate or down-regulate the target entity, it does not provide any sort of comparison between the efficiency of the possible treatments. To compare the treatments quantitatively, hybrid models can be used such as hybrid functional Petri nets [14] and fuzzy stochastic hybrid Petri nets [12]. We underline that the proposed strategies shown in Subsections 5.2.5 and 5.3 are novel and up to our knowledge they have not been simulated with other methods, nor biological experiments are done to verify them.

Each of the known strategies proposed for treatment of β -thalassemia we have checked with simulations were in line with lab experiments. This model shows also a nice example, how the analysis (through simulation) of biological systems such as hemoglobin switching network can be useful. This approach can be used as a pre-test to analyze more complex biological systems in future studies.

Conclusions

We have demonstrated that “Reaction Systems” is a beneficial model, that simulates pathways, such as, fetal to adult hemoglobin switching developmental stage, in healthy people, thalassemic or sickle cell anemia patients and in case of

various treatments to compensate for the lack of β -globin gene expression via γ -globin gene expression induction by down-regulation or inhibition of gene expressions such as HDAC1/2, KLF1, BCL11A, SOX6, and GATA1. The simulation results in our Reaction Systems model shows that β -globin gene expression is upregulated as expected in adult stage, and downregulated in patients who are suffering from β -globin gene related disorders caused by mutation in β -globin gene. Moreover, simulation results of our proposed model demonstrate that inhibition of HDAC1/2 gene expression decreases the concentration level of NuRD complex, down-regulation of KLF1 gene expression decreases gene expression of BCL11A, down-regulation of BCL11A gene expression decreases the binding rates of γ -globin gene and the multi-protein complex including BCL11A, NuRD, and erythroid transcription factors (GATA1, FOG1, and SOX6), and finally down-regulation of SOX6 gene expression and inhibition of gene expression of GATA1 decreases the concentration of erythroid transcription factors. In all cases, γ -globin gene up-regulated in adult stage, which agree with current known treatments of β -globin gene related disorders via γ -globin gene expression induction. Therefore, since hemoglobin switching process can be represented as binary of gene expression up regulation or down regulation, our proposed reaction systems model can describe and analyze this biological phenomenon.

Moreover, we propose a novel strategy to treat β -thalassemia and sickle cell anemia by inhibiting expression of FOG1 by using methods such as RNAi. RNAi as a naturally occurring mechanism of gene regulation method can be used to inhibit FOG1 gene expression by introducing double stranded RNA into the cell. As the results of our simulation have illustrated, this strategy decreases concentration level of erythroid transcription factors and leads to γ -globin gene expression induction, that is, this proposed strategy could provide potential treatment options. However, we have not performed any laboratory experiments to explore how the proposed strategy would work in a wet lab setting. Thus, further validation of this strategy needs to be performed in our future studies.

Acknowledgement

The authors would like to acknowledge Prof. Dr. Rza Bashirov for his continuous support and encouragement in our scientific endeavors.

References

- [1] AbdulAzeez S, Borgio JF (2016) In-silico computing of the most deleterious nsSNPs in HBA1 gene. *PLoS one* 11(1), e0147702
- [2] Bank A (2006) Regulation of human fetal hemoglobin: new players, new complexities. *Blood* 107(2):435-43
- [3] Bauer DE, Kamran SC, Orkin SH (2012) Reawakening fetal hemoglobin: Prospects for new therapies for the Beta-globin disorders. *Blood* 120(15):2945-53

-
- [4] Breton A, Theodorou A, Aktuna S et al (2016) ASH1L (a histone methyltransferase protein) is a novel candidate globin gene regulator revealed by genetic study of an English family with beta-thalassaemia unlinked to the beta-globin locus. *British journal of haematology* 175(3):525-530
- [5] Cao DJ, Wang ZV, Battiprolu PL et al (2011) Histone deacetylase (HDAC) inhibitors attenuate cardiac hypertrophy by suppressing autophagy. *Proc Natl Acad Sci U S A* 108(10):4123-8
- [6] Ehrenfeucht A, Rozenberg G (2007) Events and modules in reaction systems. *Theoretical Computer Science* 376(1-2):3-16
- [7] Ehrenfeucht A, Rozenberg G (2004) Basic Notations of Reaction Systems, *Developments in Language Theory. LNCS 3340:27-29*
- [8] Galanello R, Origa R (2010) Beta-thalassemia. *Orphanet J Rare Dis* 5(11)
- [9] Grieco AJ, Billett HH, Green NS et al (2015) Variation in Gamma-Globin Expression before and after Induction with Hydroxyurea Associated with BCL11A, KLF1 and TAL1. *PLoS, One* 10(6):e0129431
- [10] Inoue T, Kulkeaw K, Muennu K et al (2015) Herbal drug ninjin'yoeito accelerates myelopoiesis but not erythropoiesis in vitro. *Genes Cells* 19(5):432-40
- [11] Lin YC, Lin JH, Chou CW et al (2008) Statins increase P21 through inhibition of Histone Deacetylase Activity and release of promoter-associated HDAC1/2. *American Association for Cancer Research* 68(7):2378-83
- [12] Liu F, Heiner M, Yang M (2016) Fuzzy stochastic petri nets for modeling biological systems with uncertain kinetic parameters. *PloS one* 11(2), e0149674
- [13] Macari ER, Shaeffer EK, West RJ, Lowrey CH (2013) Simvastatin and t-butylhydroquinone suppress KLF1 and BCL11A gene expression and additively increase fetal hemoglobin in primary human erythroid cells. *Blood* 121(5):830-9
- [14] Mehraei M, Bashirov R, Tüzmen Ş (2016) Target-based drug discovery for β -globin disorders: drug target prediction using quantitative modeling with hybrid functional Petri nets. *Journal of bioinformatics and computational biology* 14(05), 1650026
- [15] Ozcan G, Ozpolat B, Coleman R et al (2015) Preclinical and clinical development of siRNA-based therapeutics. *Advanced drug delivery reviews* 87:108-119
- [16] Paokanta P, Harnpornchai N, Srichairatanakool S et al (2011) The Knowledge Discovery of β -Thalassemia Using Principal Components

Analysis: PCA and Machine Learning Techniques. International Journal of e-Education, e-Business, e-Management and e-Learning, 1(2), 169

- [17] Sankaran VG (2011) Targeted Therapeutic Strategies for Fetal Hemoglobin Induction. Hematology Am Soc Hematol Educ Program 1:459-65
- [18] Suzuki M, Yamamoto M, Engel JD (2014) Fetal globin gene repressors as drug targets for molecular therapies to treat β -globinopathies. Mol Cell Biol 34(19):3560-9
- [19] Yao X, Kodeboyina S, Liu L et al (2009) Role of Stat3 and GATA-1 Interactions in γ -Globin Gene Expression. Exp Hematol 37(8):889-900

A Chaotic Image Encryption Algorithm Robust against Phase Space Reconstruction Attacks

Jakub Oravec, Ján Turán, Ľuboš Ovseník, Tomáš Huszaník

Department of Electronics and Multimedia Communications,
Faculty of Electrical Engineering and Informatics, Technical University of Košice,
Němcovej 32, 040 01 Košice, Slovakia

jakub.oravec@tuke.sk, jan.turan@tuke.sk, lubos.ovsenik@tuke.sk,
tomas.huszanik@tuke.sk

Abstract: This paper describes a modification of a chaotic logistic map which could be exploited in a field of image encryption. After a summary of basic image encryption methods and problems, the paper mentions properties of a modified version of the logistic map. It is shown that the proposed changes help to achieve greater robustness against phase space reconstruction attacks. The paper also describes the usage of a modified map in an image encryption algorithm. Other techniques applied in the proposed algorithm include key diffusion, ciphertext chaining or four step diffusion stage. Evaluation of properties of the proposed algorithm is done by means of commonly used techniques. The numerical results are then compared with values obtained by other published algorithms.

Keywords: image encryption; logistic map; phase space reconstruction

1 Introduction

One of the first encryption algorithms based on the chaotic maps was proposed in 1989 by Matthews [1]. Since then, various chaotic encryption algorithms were designed, including the one described by Fridrich in 1998 [2]. Fridrich's approach could be considered as important, since it introduced an idea of image encryption. Operations created especially for two dimensional matrices allowed simpler and more effective computations which is still the main advantage over conventional encryption algorithms such as Advanced Encryption Standard (AES). Also, the majority of proposals adopted a two stage encryption process which was described by Fridrich. The first stage – confusion changes positions of plaintext image pixels in order to minimize their correlation. The second stage – diffusion calculates the intensities of pixels in the resulting encrypted image.

The development of the chaotic image encryption algorithms continued by removing their drawbacks. Small key space problems were solved by high dimensional chaotic maps [3] or by combinations of various maps [4]. The Dynamic degradation of chaos, present in the discrete versions of chaotic maps, was studied in [5]. An attack capable of revealing permutations of the image pixels was published by Solak *et al.* in 2010 [6]. Especially the last mentioned problem caused the usage of more complicated diffusion stages.

The diffusion stage usually employs a technique called ciphertext chaining for establishing dependencies between the intensities of consecutive image pixels. The dependencies are useful for creating different encrypted images for plaintext images with only small amount of changes, however they could be easily found out by Solak's attack. In order to provide certain level of robustness against this attack, the diffusion stage needs to use another operation. Probably one of the most used operations is an addition of elements from pseudo-random (PR) sequences in modular arithmetic. In this case, the Solak's attack would obtain only the pixel intensities combined with the elements of PR sequences.

However, also the generation of the PR sequences requires some care. In the case that the PR sequences are simply computed by some of the chaotic maps, already calculated elements of the PR sequence could be evaluated by the phase space reconstruction attacks. The basic theory of the phase space reconstruction was given by Takens in 1985 [7]. In the context of the image encryption algorithms, the phase space is a set of values that could be achieved by the used chaotic map.

Approaches that are effective against phase space reconstruction can be divided into two groups. The first group changes parameters of the chaotic maps during computation of the PR sequences. Murillo-Escobar *et al.* [8] described an algorithm where parameter of the used map is changed by values produced by other map. Liu and Miao [9] used binary PR sequences for diffusion by means of a generated code book. The second group of proposals, modifies each sequence element after its computation. Guanghui *et al.* [10] proposed a scheme with modular design which could utilize various chaotic maps, however the results were presented only for the logistic map. Liu *et al.* [11] described a modification of calculated sequence elements by another map. In all of these cases, the robustness against the phase space reconstruction was created by suppressing the dependencies between consecutive elements of the PR sequences.

The algorithm presented in this paper tries to provide certain amount of robustness against both Solak's attack and the phase space reconstruction attacks. Since the elements of the generated PR sequences are combined with results of ciphertext chaining by means of bitwise eXclusive OR (XOR), the Solak's attack would be useful only for obtaining the combined values. As the elements of PR sequences are computed by a modified version of chaotic logistic map which is effective against phase space reconstruction, it is difficult to evaluate the elements of PR sequences that are required for a successful decryption.

The rest of the paper is organized as follows: Chapter 2 deals with the logistic map, its properties and the modification which is proposed in this paper for purposes of the image encryption. Chapter 3 describes the algorithms used for encryption and decryption. An analysis of experimental results is given in Chapter 4 and the lastly in Chapter 5, conclusions the paper are given, by a brief summary of advantages and disadvantages of the proposed solution and plans for the future work.

2 Logistic Map and its Modification

2.1 Logistic Map

Logistic map (LM) can be described as an one dimensional chaotic map that uses one parameter $r \in (0; 4)$. An initial value x_0 and the values of x in all iterations of the map belong to an interval $(0; 1)$. The LM was popularized mainly by a paper of May in 1976 [12]. The values of x in consecutive iterations, called iterates of the map could be calculated by applying (1):

$$x_{n+1} = rx_n(1 - x_n), \quad (1)$$

where n denotes an iteration number.

Chaotic behavior of the LM could be illustrated by its bifurcation diagram. The diagram shows values of x_n that were calculated with various values of the parameter r . An example of the bifurcation diagram that has the initial value x_0 equal to 0.5 and plots 800 values of x_n is shown in Figure 1.

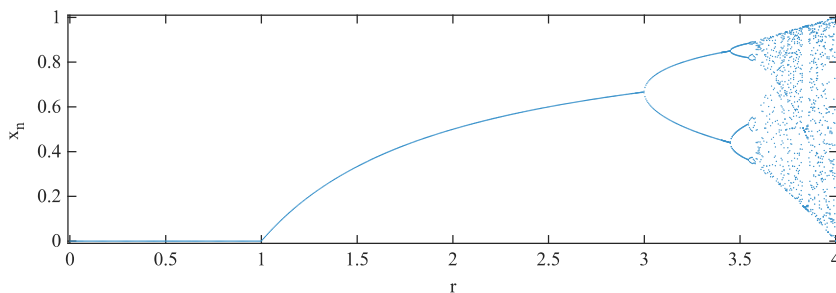


Figure 1

A bifurcation diagram of the logistic map

As it is visible, behavior of the LM is predictable until the parameter r reaches certain value close to 3. At this point the values of calculated iterates start to oscillate between two sets. This property of the LM is known as a bifurcation.

After several other bifurcations, it is quite difficult to see the relations between consecutive iterates. The point where $r \sim 3.56995$ is known also as a start of a chaotic behavior of the LM. However, also some values of r after this point show predictable behavior. These are known as islands of stability. Probably the most notable example is present around r equal to 3.85.

Another important property of the LM is an existence of a transient period. This period contains iterates that are calculated among the first and therefore their values could be predictable. In most cases, the effects of the transient period are suppressed by using some iterates only for modification of the initial value x_0 . Common sizes of the transient period are powers of 10, e. g. 100 or 1,000 iterates.

Because the LM could be considered as a discrete version of a continuous logistic differential equation, also the properties regarding the finite amount of possible iterate values should be investigated. Periodicity testing was performed in a computing environment MATLAB R2015a. In total, 10 sequences were computed, each one consisted of 10^8 iterates. The iterates were represented as double precision values (64 bits). As 52 bits are used for storage of a fractional part of these values, their precision could be expressed as $\log_{10}(2^{52}) \sim 15.6536$. Therefore, this data type provides precision of 15 decimal places.

Each of the 10 sequences used the initial value x_0 set to 0.5 and the transient period with size of 1,000 iterates. Sequences differed by value of the parameter r which was set from $4 \cdot 10^{-14}$ to $4 \cdot 10^{-15}$ with a step of 10^{-15} . The period lengths for the sequences with investigated values of the parameter r are shown in Table 1.

Table 1

Period lengths of the sequences generated by the LM with various values of the parameter r

Value of r	Period length	Value of r	Period length
$4 \cdot 10^{-14}$	10^8	$4.5 \cdot 10^{-15}$	10^8
$4.9 \cdot 10^{-15}$	10^8	$4.4 \cdot 10^{-15}$	10^8
$4.8 \cdot 10^{-15}$	10^8	$4.3 \cdot 10^{-15}$	12,960,875
$4.7 \cdot 10^{-15}$	10^8	$4.2 \cdot 10^{-15}$	33,767,629
$4.6 \cdot 10^{-15}$	10^8	$4 \cdot 10^{-15}$	15,599,659

The results from Table 1 show that the values of the parameter r influence period lengths. However, also the worst shown case ($r = 4.3 \cdot 10^{-15}$) would be sufficient for element-wise processing of approx. 13 million elements. This number of elements is present in a true color image with a resolution of 2,078x2,078 pixels.

2.2 Phase Space Reconstruction Attacks

As it was already mentioned, the Solak's attack and similar known-plaintext attacks could be used for revealing the permutations of image pixels done in the confusion stage of the image encryption algorithms. Therefore, the main part of

security provided by these algorithms is created by the diffusion stage. Since the diffusion stage of image encryption algorithms usually sequentially processes each pixel of the input images, the amount of used operations should be minimal.

Usually, the diffusion stage consists of two operations. The first one, ciphertext chaining is applied for establishing dependencies between consecutive image pixels in the encrypted images. However, the most popular version of the ciphertext chaining could be reversed by anyone who has access to the encrypted image. The second operation used during diffusion stages is the combination of pixel intensities with the elements of a PR sequence. The combination could be done as an addition in modulo 256 or a bitwise XOR. As the ciphertext chaining and the confusion stage could be broken by attackers, the security of some algorithms depends solely on this operation.

The LM could be used for generating the PR sequences, but there are still some relationships between iterates. The relationships are expressed by a Poincaré plot which uses values of two consecutive iterates as the coordinates of plotted points. An example is shown in Figure 2, where the LM with 10^8 iterates used the initial value $x_0 = 0.5$, the transient period with size of 1,000 iterates and the parameter r set to $4 \cdot 10^{-15}$. The plot shows only the first 2,000 points for better readability.

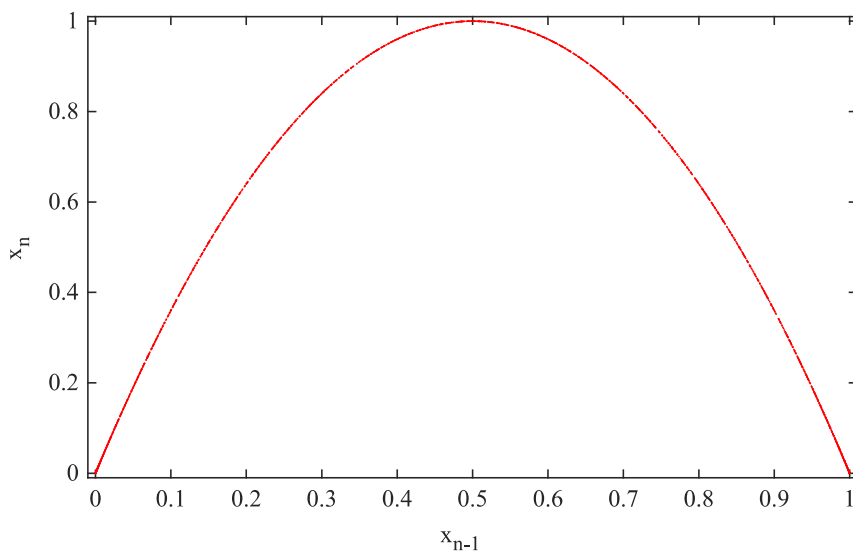


Figure 2

A Poincaré plot for consecutive iterates of the LM

An iterate (x_{n-1}) that was used for the calculation of a current iterate (x_n) could be observed by using x_n as a coordinate on y axis. The plot shows that there are, at most, two points with chosen y coordinate. Therefore, it could be assumed that the iterate x_n was calculated from one of the two possible coordinates on the x axis.

However, this technique has many drawbacks. First of all, the attacker needs to determine the parameter r before construction of plots. Also, the plots do not provide values of x_{n-1} for all possible x_n . For the precision of 10^{-15} , the plots would need to contain $2 \cdot 10^{15}$ points in order to provide all possible relations between consecutive iterates. Finally, if each previous element would be represented by 2 possible values, the reconstruction of sequence with num elements would result in 2^{num} possible solutions. Therefore, the computations needed for the phase space reconstruction only by the Poincaré plots seem to be computationally exhaustive.

There are also some other, more efficient approaches for the phase space reconstruction. One of them is known as a time delay method. Some examples of usage of this method are presented in [13] or [14].

2.3 Proposed Modification of Logistic Map

As it was shown in the previous subchapters, the LM has some drawbacks which could impact the security of the designed image encryption algorithms. In order to suppress some of them, we propose a modification of (1):

$$x_{n+1} = 10^4 \cdot rx_n(1-x_n) \pmod{1} \quad (2)$$

The changes in the equation could seem unimportant, but they cause great differences in its behavior. As a double value with 15 decimal places is multiplied by 10^4 , the first four decimal places move to a left side of the decimal mark. Other 11 decimal places are shifted by 4 places to the left side. Remaining 4 decimal places are created by increasing the amount of decimal places of the double value.

The last four decimal places were originally on 16th to 19th decimal place of the iterate. As the double precision produces only numbers with 15 decimal places, the numbers with more decimal places are chosen to be the closest approximations of the double values. This causes a variation from the continuous chaotic systems, which could be viewed as an effect of the dynamical degradation of chaos [5].

The second operation applied in (2) – a usage of modulo 1 is performed to remove the numbers which are on the left side of the decimal mark. This operation is important for producing values of x_{n+1} that belong to an interval (0; 1).

It is important to point out that the changes done in the map (2) cause appearance of other fixed points. While fixed points of the map (1) are well studied, their analysis for the map (2) is not done yet. There are only some observations, e. g. the initial value $x_0 = 0.5$ produces the same value when the parameter r equals 3.9902 or 3.999. This is caused by fact that the map (1) computes iterate value that has number 5 on the fifth decimal place and it is followed by zeros. After the shift of decimal places done in the map (2), iterate values 0.99755 (for $r = 3.9902$) and 0.99975 (for $r = 3.999$) both become 0.5 which is equal to the initial value x_0 .

However, it is quite safe to mention that the occurrence of the fixed points for values of the parameter $r \geq 3.9999$ is quite rare, because calculated iterates have more decimal places than the initial value x_0 . Hence the probability that the last 11 decimal places of calculated iterate are equal to other iterates is negligible.

2.4 Comparison of the Maps

Following subchapters compare the properties of the LM (1) and the proposed map (2), which is for better readability denoted as MLM (modified logistic map).

2.4.1 Time Consumption

Measurement of the computational time needed for generating sequences used following setup: all sequences generated by the LM (1) and the MLM (2) had the transient period with length of 1,000 elements and the initial value x_0 was equal to 0.5. The number of sequence elements was set as 10^6 , 10^7 or 10^8 elements. Two different values of the parameter r were utilized, $4 \cdot 10^{-14}$ and $4 \cdot 10^{-15}$. Used PC had 2.5 GHz CPU and 12 GBs of RAM and it utilized the computational environment MATLAB R2015a running on Windows 10 OS. The times presented in Table 2 are arithmetic means of 100 repeated measurements.

Table 2
Comparison of the time consumption

Length of sequences [elements]	Value of r	Time needed for the LM [ms]	Time needed for the MLM [ms]
10^6	$4 \cdot 10^{-14}$	15.5829	106.1821
	$4 \cdot 10^{-15}$	15.4199	104.7699
10^7	$4 \cdot 10^{-14}$	170.6184	1185.5463
	$4 \cdot 10^{-15}$	168.2268	1170.6595
10^8	$4 \cdot 10^{-14}$	1642.9382	11209.9446
	$4 \cdot 10^{-15}$	1556.0445	10607.3564

The results presented in Table 2 show that the computational complexity of the MLM is approx. 7 times higher than the one of the LM. This is due to higher amount of operations used for a calculation of each iterate. However, longer computational durations are balanced by advantages that are described in following subchapters. Also it could be stated that different values of the parameter r have only a small impact on the time consumption – the maximal recorded difference was approx. 5 %.

2.4.2 Periodicity Concerns

It could be assumed that the replacement of some decimal places done by the MCM could result in a reduction of the chaotic behavior and therefore also in

smaller period lengths. However, the periodicity that occurred for the LM with certain values of the parameter r ($4.3 \cdot 10^{-15}$, $4.2 \cdot 10^{-15}$ and $4 \cdot 10^{-15}$) was caused by the finite precision of the double values – exactly because there are not any double values between $1 \cdot 10^{-15}$ and 1. Therefore, all iterates which would have values in an interval $[1 \cdot 10^{-15}; 1]$ in a system with an infinite precision, result in one of these two values in the double precision system (with finite precision). As the value of the following iterate depends only on a value of the current iterate, the values $1 \cdot 10^{-15}$ and 1 always produce the same two values. This is the cause of periodicity for the LM with certain values of the parameter r .

Because the MLM changes four decimal places at the end of each iterate, the values before the multiplication in an interval $[1 \cdot 10^{-15}; 1)$ are changed to an interval $[1 \cdot 10^{-11}; 1)$ after the multiplication. This prevents computation of the same values in following iterations. Therefore, the period lengths should be enlarged. The resulting period lengths computed with the same setting as was used for the LM in Table 1 (the initial value x_0 set as 0.5, the transient period with size of 1,000 iterates and the parameter r set in an interval from $4 \cdot 10^{-14}$ to $4 \cdot 10^{-15}$ with a step of 10^{-15}) are shown in Table 3.

Table 3

Period lengths of the sequences generated by the MLM with various values of the parameter r

Value of r	Period length	Value of r	Period length
$4 \cdot 10^{-14}$	10^8	$4.5 \cdot 10^{-15}$	10^8
$4.9 \cdot 10^{-15}$	10^8	$4.4 \cdot 10^{-15}$	10^8
$4.8 \cdot 10^{-15}$	10^8	$4.3 \cdot 10^{-15}$	10^8
$4.7 \cdot 10^{-15}$	10^8	$4.2 \cdot 10^{-15}$	10^8
$4.6 \cdot 10^{-15}$	10^8	$4 \cdot 10^{-15}$	10^8

It could be observed that the MLM produces sequences with period lengths of 10^8 elements (length of the generated sequence) in all described cases. As the LM had shorter period lengths for sequences with the parameter r equal to $4.3 \cdot 10^{-15}$, $4.2 \cdot 10^{-15}$ and $4 \cdot 10^{-15}$, it could be concluded that the MLM achieves better results by means of periodicity.

2.4.3 Robustness against the Phase Space Reconstruction

The replacement of the last four decimal places also improves robustness against the phase space reconstruction attacks. Because the decimal places of iterates are shifted by 4 places to the left side, the relations between values of two consecutive iterates are quite unpredictable. This property is also shown by the Poincaré plot in Figure 3, where the relations between consecutive iterates of the MLM are illustrated. The sequence of 10^8 elements was generated with the initial value x_0 of 0.5, the transient period with size of 1,000 iterates and the parameter r set to $4.4 \cdot 10^{-15}$. The plot shows only the first 2,000 points for better readability.

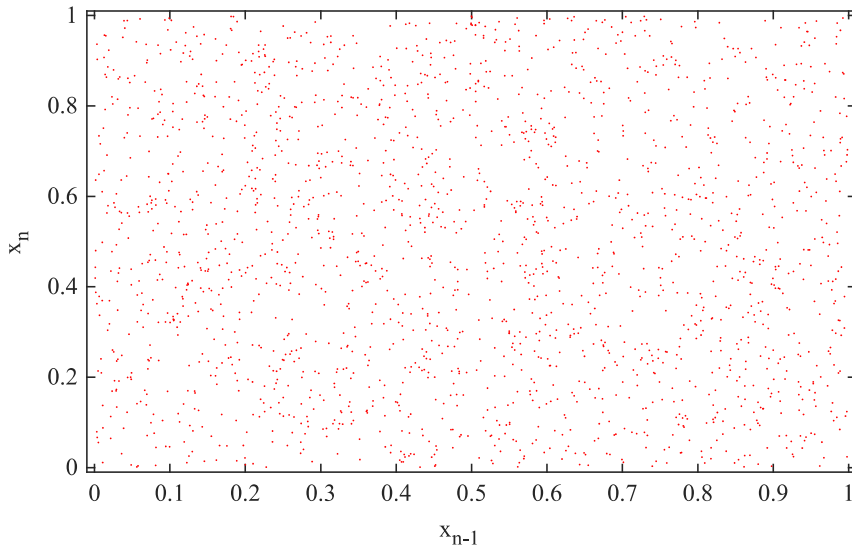


Figure 3

A Poincaré plot for consecutive iterates of the MLM

Because coordinates of the plotted points for the MLM are determined mainly by decimal places close to the fourth place prior to the multiplication, distribution of points in the plot is more uniform. This is due to weak relation between individual decimal places of the iterates before the multiplication. More uniform distribution of the points could be considered as a difficult problem for the phase space reconstruction algorithms, as the number of possible values of previous iterates is not clear (it was 2 for the parabola plotted for the LM in Figure 2). Therefore, complete reconstruction by the Poincaré plots needs to evaluate all possible x coordinates (x_{n-1}) for each current iterate value represented on the y axis (x_n).

3 Proposed Image Encryption Algorithm

The described map (2) shows a set of properties that could be useful for the image encryption algorithms. Therefore, we applied the MLM in an algorithm with usual architecture. Firstly, the confusion stage permutes the image pixels and then the diffusion stage changes their intensities. The proposed algorithm works with images of an arbitrary resolution and color depths of 8 bits (grayscale images) and 24 bits per pixel (true color images). Algorithm in following subchapter is used for the encryption. Different steps done during the decryption are described in subchapter 3.2.

3.1 Encryption Algorithm

Inputs: a plaintext image P , a 16-byte key K

Output: an encrypted image E

Step 1: Height h , width w and the number of color planes num_{cp} of the matrix P are determined. The matrix P is then reshaped to a matrix P_{mat} that has h rows and $w' = w \cdot num_{cp}$ columns. This is done for establishing dependencies between the color planes in true color images.

Step 2: The key K is divided into four subkeys: K_{col} (bytes 1 to 4), K_{row} (bytes 5 to 8), K_{dif1} (bytes 9 to 12) and K_{dif2} (bytes 13 to 16).

Step 3: The subkeys K_{col} , K_{row} , K_{dif1} and K_{dif2} are converted to decimal numbers and they are used for calculation of four values r_{col} , r_{row} , r_{dif1} and r_{dif2} by (3):

$$\begin{aligned} r_{col} &= 3.9999 + \frac{K_{col}}{10^4 \cdot (1+2^{32})} & r_{row} &= 3.9999 + \frac{K_{row}}{10^4 \cdot (1+2^{32})} \\ r_{dif1} &= 3.9999 + \frac{K_{dif1}}{10^4 \cdot (1+2^{32})} & r_{dif2} &= 3.9999 + \frac{K_{dif2}}{10^4 \cdot (1+2^{32})} \end{aligned} \quad (3)$$

where 10^4 and $1+2^{32}$ are constants used to ensure that the values of r_{col} , r_{row} , r_{dif1} and $r_{dif2} \in [3.9999; 4)$.

Step 4: Four sequences seq_{col} , seq_{row} , seq_{dif1} and seq_{dif2} are generated by the MLM (2) with the initial values x_0 equal to 0.5. The transient period has size of 1,000 iterates. The value of parameter r used during and after the transient period depends on a calculated sequence and is shown in Table 4. The lengths of generated sequences are following: seq_{col} has w' elements, seq_{row} has h elements and both seq_{dif1} and seq_{dif2} have $2 \cdot h \cdot w'$ elements.

Table 4
Values of the parameter r used during and after the transient period

Sequence	Value of r used for iterates 1 to 250	Value of r used for iterates 251 to 500	Value of r used for iterates 501 to 750	Value of r used for following iterates
seq_{col}	r_{row}	r_{dif1}	r_{dif2}	r_{col}
seq_{row}	r_{dif1}	r_{dif2}	r_{col}	r_{row}
seq_{dif1}	r_{dif2}	r_{col}	r_{row}	r_{dif1}
seq_{dif2}	r_{col}	r_{row}	r_{dif1}	r_{dif2}

Changing of the parameter r during the transient period creates effect known as a key diffusion. In this case, the change of only one byte in the key K should result in differences in all four generated sequences.

Step 5: Elements of the sequences seq_{col} , seq_{row} , seq_{dif1} and seq_{dif2} are quantized by a set (4). The quantized sequences are denoted by an apostrophe.

$$\begin{aligned}
seq'_{col}(k) &= \lfloor h \cdot seq_{col}(k) \rfloor & seq'_{row}(l) &= \lfloor w' \cdot seq_{row}(l) \rfloor \\
seq'_{dif1}(i) &= round(255 \cdot seq_{dif1}(i)) \\
seq'_{dif2}(i) &= round(255 \cdot seq_{dif2}(i))
\end{aligned} \tag{4}$$

where $k = 1, 2, \dots, w'$, $l = 1, 2, \dots, h$ and $i = 1, 2, \dots, 2 \cdot h \cdot w'$

Step 6: Two sequences seq'_{dif1} and seq'_{dif2} are reshaped to four matrices. The first half of seq'_{dif1} creates a matrix mat'_{dif11} with h rows and w' columns. Sequence elements are stored in the columns of the matrix, starting from the left side. Other half of seq'_{dif1} creates matrix mat'_{dif12} with the same size. The sequence seq'_{dif2} is split and reshaped by the same way to matrices mat'_{dif21} and mat'_{dif22} .

Step 7: The first part of the confusion stage takes place. Columns of the matrix P_{mat} are scanned and their pixels are permuted by a circular shift. The amount of shifting for each column is given by the elements of sequence seq'_{col} . Image with shifted pixels in its columns is stored in an auxiliary matrix A .

Step 8: The second part of the confusion stage is done. Rows of the matrix A are scanned and their pixels are permuted by a circular shift. The amount of shifting done for each row is given by the elements of sequence seq'_{row} . The permuted image is stored in the matrix A .

Step 9: The first part of the diffusion stage takes place. Pixels in rows of the matrix A are diffused from the top to the bottom row (5):

$$A(l,:) = [A(l,:) + A(l-1,:)] \oplus mat'_{dif11}(l,:) \tag{5}$$

where $l = 1, 2, \dots, h$ denotes a row index, $:$ stands for all pixels in columns of the matrix A and \oplus is an operator of bitwise XOR. All additions of elements from matrix A are done modulo 256. The top row of pixels ($l = 1$) uses the bottom row ($l = h$) for the additions.

Step 10: The second part of the diffusion stage is done. Pixels in rows of the matrix A are diffused in the opposite direction from the bottom to the top row (6):

$$A(h-l,:) = [A(h-l,:) + A(h+1-l,:)] \oplus mat'_{dif12}(h-l,:) \tag{6}$$

where $l = 0, 1, \dots, h-1$ denotes the row index. All additions of elements from the matrix A are done modulo 256. The bottom row of pixels ($l = h$) uses the top row ($l = 1$) for the additions.

Step 11: The third part of the diffusion stage is carried out. Pixels in columns of the matrix A are diffused from the leftmost to the rightmost column (7):

$$A(:,k) = [A(:,k) + A(:,k-1)] \oplus mat'_{dif21}(:,k) \tag{7}$$

where $k = 1, 2, \dots, w'$ denotes a column index and $:$ stands for all pixels in rows of the matrix A . All additions of elements from the matrix A are done modulo 256.

The leftmost column of pixels ($k = 1$) uses the rightmost column ($k = w'$) for the additions.

Step 12: The fourth and final part of the diffusion stage is computed. Pixels in columns of the matrix A are diffused in the opposite direction from the rightmost to the leftmost column (8):

$$A(:, w'-k) = [A(:, w'-k) + A(:, w'+1-k)] \oplus mat'_{dif22}(:, w'-k) \quad (8)$$

where $k = 0, 1, \dots, w'-1$ denotes the column index. All additions of elements from the matrix A are done modulo 256. The rightmost column of pixels ($k = w'$) uses the leftmost column ($k = 1$) for the additions.

Step 13: The auxiliary matrix A is reshaped to a matrix E with h rows, w columns and num_{cp} color planes. The matrix E represents encrypted version of the image P .

3.2 Differences in the Decryption Algorithm

The decryption is analogous to the encryption, only the order of operations is reversed. After the generation and the processing of sequences (Steps 1 to 6), the first change is done when the removal of diffusion removal is applied before the removal of confusion. Also the parts of the diffusion stage are used backwards, starting with Step 12 and continuing to Step 9. These steps apply subtractions instead of the additions, the usage of modulo 256 arithmetic remains the same. Since repeated usage of bitwise XOR produces the values before diffusion, this operation is not changed. However, the order of the two operations is reversed, the bitwise XOR is used prior to the subtractions.

The removal of confusion is done also in the opposite order. First, shuffling in image rows is removed by using circular shifts with negative values of elements from sequence seq'_{row} . Then, the rearrangements in image columns are removed by circular shifts given by negative values of elements from sequence seq'_{col} .

4 Analysis and Comparison of Experimental Results

All experiments with the proposed algorithms were performed on a PC with 2.5 GHz CPU, 12 GBs of RAM in the MATLAB 2015a running on the Windows 10 OS. The set of images used for testing is shown in Figure 4. The first two images, *lena* and *lenaG* have resolution of 512x512 pixels and color depths of 24 and 8 bits per pixel, respectively. Images *black1* and *black2* have resolution of 256x128 pixels and color depth of 8 bits per pixel. Image *black2* has a pixel with intensity 255 located on the coordinates [128; 64]. The keys used during experiments are illustrated in Table 5. Differences between similar keys are indicated by bold characters. The images from Figure 4 encrypted by key K_I are shown in Figure 5.



Figure 4
Set of images used for the experiments

Table 5
Keys used for the experiments

Key	Value
K_1	0x746869736973617365637265746B6579
K_2	0x7468697 2 6973617365637265746B6579
K_3	0x74686973697361736563726574 6C 6579

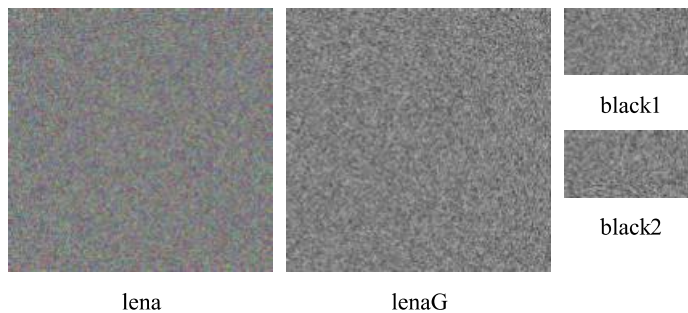


Figure 5
Encrypted versions of the images from the testing set

4.1 Size of the Key Space and the Key Sensitivity

A key space consists of all possible keys that could be used. As the proposed algorithm utilizes 16-byte key K , the total size of the key space is given as $256^{16} = 2^{128}$. If we would consider the time required for a decryption of one true color image with resolution of 512x512 pixels as approx. 550 ms, the brute-force attack on the image with these parameters would take approx. $5.9347 \cdot 10^{30}$ years. Therefore, the brute-force attack could be considered as infeasible.

Sensitivity of the proposed algorithm to used keys could be investigated by an encryption with one key and a decryption with other keys. The images decrypted by an incorrect key should not contain any information about the original plaintext image (which is the same as the image decrypted by a correct key). This experiment is shown in Figure 6, where the image *lena* was encrypted by key K_1 . Then it was decrypted with all three keys.

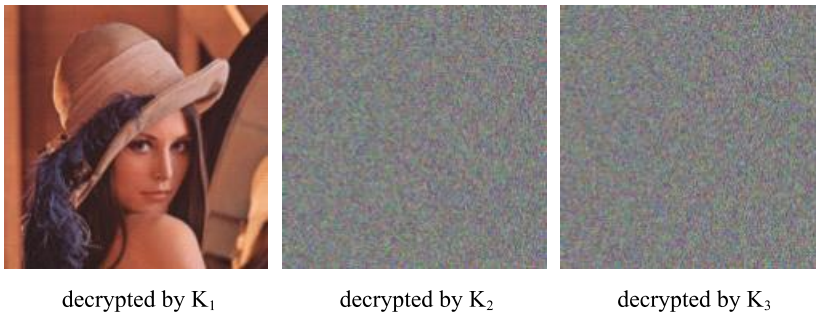


Figure 6

Illustration of the key sensitivity

Please note that the change of a key in portion which is used for the confusion stage (pixel rearrangement) also influences the diffusion stage. This property is result of the key diffusion and it is visible on image decrypted by key K_2 .

4.2 Statistical Attacks

Statistical attacks try to obtain some properties of the image encryption algorithms by comparing known pairs of the plaintext images and corresponding encrypted images. When some properties of the encryption algorithms are known, statistical attacks could be used for breaking the algorithms or their parts. There are several measures that could be used to illustrate robustness against the statistical attacks.

The first measure is histogram comparison. The histogram of the encrypted image should have distribution close to uniform without notable peaks which are present in the histogram of the plaintext image. Histograms of the image *lenaG* before and after encryption by the key K_1 are shown in Figure 7.

Second measure is illustrated by scatter plots that contain points with coordinates given by intensities of two adjacent image pixels. The adjacencies could be horizontal, vertical or diagonal. If the plotted points are close to line $y = x$, it could be concluded that the intensities of adjacent pixels are highly correlated. The distribution of the points for the encrypted images should be close to uniform. The scatter plots for the horizontal adjacencies of 1,000 randomly chosen pixel pairs from the image *lenaG* and its version encrypted by key K_1 are shown in Figure 8.

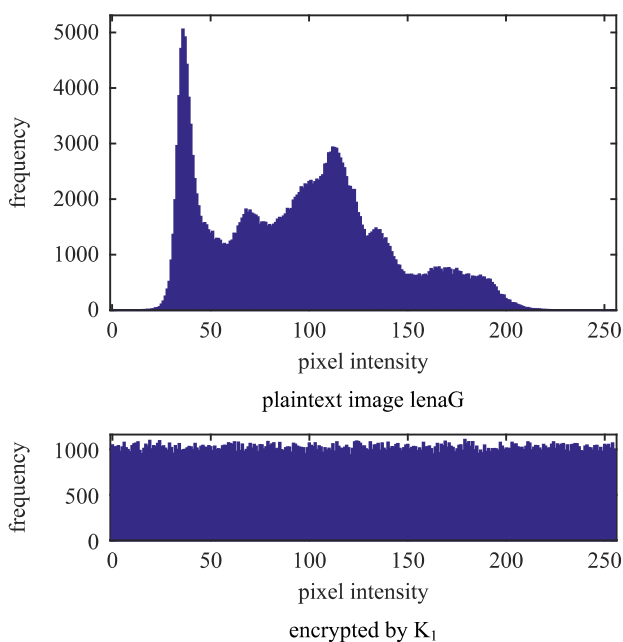


Figure 7
Comparison of the histograms

Third measure are values of correlation coefficients ρ , computed separately for each color plane by (9) for the horizontally (ρ_{hor}), vertically (ρ_{ver}) or diagonally (ρ_{diag}) adjacent image pixel pairs. Vector I_1 contains intensities of the first pixels from the pairs and vector I_2 is created by scanning of the adjacent pixel intensities.

$$\rho = \frac{\sum_{pp=1}^{npp} (I_1(pp) - \bar{I}_1) \cdot (I_2(pp) - \bar{I}_2)}{\sqrt{\sum_{pp=1}^{npp} (I_1(pp) - \bar{I}_1)^2 \cdot \sum_{pp=1}^{npp} (I_2(pp) - \bar{I}_2)^2}} \quad [-] \quad (9)$$

where $pp = 1, 2, \dots, npp$ is an index of pixel pair, npp denotes an amount of the pixel pairs and \bar{I}_x stands for an arithmetic mean of vector I_x .

The last measure is an entropy H which is calculated for each color plane by (10).

$$H = - \sum_{in=0}^{255} p(in) \cdot \log_2(p(in)) \quad [bits/pixel] \quad (10)$$

where $p(in)$ is a probability of occurrence of a pixel with intensity in .

Computed values of the correlation coefficients ρ and entropy H are included with other values in Table 6. The values of the correlation coefficients are arithmetic means of 100 repeated measurements for 1,000 randomly chosen pixel pairs.

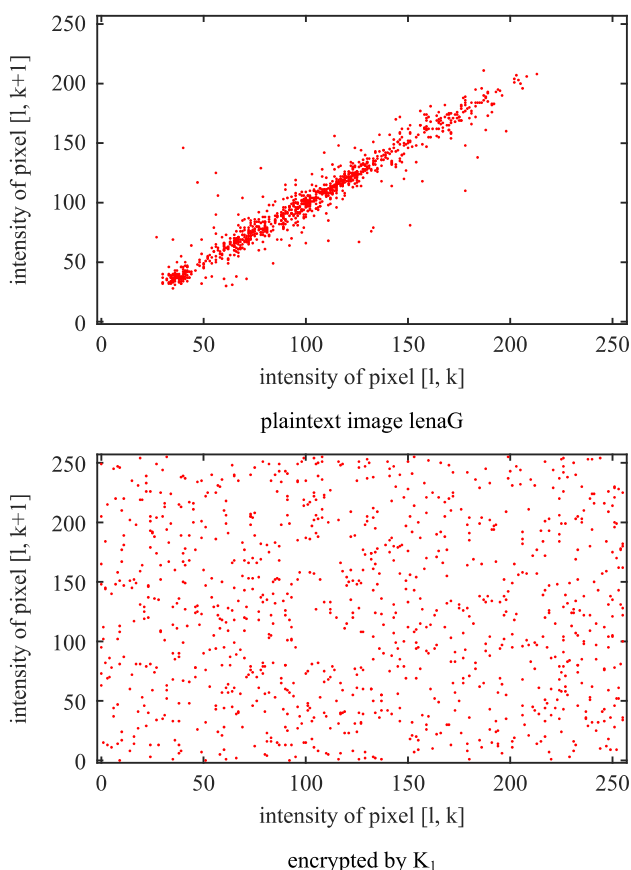


Figure 8

Scatter plots for the horizontally adjacent pixel pairs

4.3 Differential Attacks

Differential attacks reveal the properties of the encryption algorithms by exploring differences in encrypted versions of similar plaintext images. Robustness against the differential attacks is evaluated by two parameters – Number of Pixel Change Ratio (NPCR) and Unified Average Changing Intensity (UACI). Both these parameters use a pair of plaintext images, P_1 and P_2 which differ only in an intensity of one pixel. Furthermore, the size of this difference is minimal (one intensity level). These two images are then encrypted by the same key as E_1 and E_2 . NPCR for individual color planes of image pair E_1 and E_2 is given by (11):

$$NPCR = \frac{100}{h \cdot w} \cdot \sum_{l=1}^h \sum_{k=1}^w Diff(l, k) \quad [\%] \quad (11)$$

where h and w represent height and width of images E_1 and E_2 , l and k denote row and column indexes, $Diff$ is a difference matrix, $Diff(l, k) = 1$ if $E_1(l, k) \neq E_2(l, k)$, otherwise $Diff(l, k) = 0$.

Calculation of UACI for the color planes uses (12):

$$UACI = \frac{100}{h \cdot w} \cdot \sum_{l=1}^h \sum_{k=1}^w \frac{|E_1(l, k) - E_2(l, k)|}{2^L - 1} \quad [\%] \quad (12)$$

where L is a color depth of a color plane in bits per pixel.

A difference between NPCR and UACI is hidden in the way how these parameters evaluate changes in the pair of encrypted images. While NPCR only sums the number of pixel intensity changes, UACI also records the sizes of the changes. Calculated values of NPCR and UACI for the set of images presented in Figure 4 are included in Table 6. These values are arithmetic means of 100 repeated measurements with randomly chosen pixel with modified intensity.

Both NPCR and UACI are given as percentages. Wu et al. [15] described theoretically critical values of NPCR and UACI that depend on the resolution of the encrypted images. For the resolution of 512x512 pixels, the theoretically critical values are 99.6094% for NPCR 33.4635% for UACI.

4.4 Time Consumption and Computational Complexity

Images as a data type could be characterized by high redundancy. Therefore, the speed of encryption or decryption is an important property of the image encryption algorithms. Durations of encryption t_{enc} and decryption t_{dec} are included in Table 6. These times are arithmetic means of 100 repeated measurements.

The computational complexity of image encryption algorithms could be expressed also by amount of performed operations in so-called big O notation (also known as an asymptotic notation). Following paragraph investigates the case of encryption, with the height of a processed image denoted as h , its width w and number of color planes num_{cp} represented by $w' = w \cdot num_{cp}$. The generation of PR sequences by the modified version of the logistic map takes $2hw' + h + w' + 4,000$ operations. Each operation for this stage of algorithm consists of three multiplications, one subtraction and one modulo operation. Quantization of PR sequences requires $2hw' + h + w'$ multiplications and rounding operations. Confusion stage of the proposed algorithm needs $h + w'$ circular shifts of sequences with h or w' elements. The last stage of the algorithm – diffusion computes $2hw'$ additions and XOR additions.

If the complexity of various operations would be considered as the same, the proposed algorithm requires $16hw' + 7(h + w') + 16,000$ operations for one encryption. Therefore, it could be concluded that the computational complexity

for the proposed algorithm is linear and it depends solely on a resolution of the processed image (hw'). As for most cases, the term $16hw'$ is greater than the other two terms, the big O notation of the proposed algorithm could be given as $O(16n)$.

4.5 Comparison of Numerical Results

Resulting values of the numerical measures for the proposed algorithm are shown in Table 6. Characters R , G and B in the brackets denote individual color planes of true color images (Red, Green or Blue). The word “plain” in key column is used for plaintext images which are not encrypted.

Table 6
Numerical results achieved by the proposed algorithm

Value	Key	<i>lena</i> (R)	<i>lena</i> (G)	<i>lena</i> (B)	<i>lenaG</i>	<i>black1</i>	<i>black2</i>
ρ_{hor} [-]	plain	0.973	0.9677	0.953	0.9712	~1	~1
	K_1	0.0094	0.003	-0.0006	-0.0013	0.0156	-0.0011
	K_2	-0.0031	-0.0049	-0.005	0.0044	0.0189	-0.01
ρ_{ver} [-]	plain	0.9737	0.9735	0.956	0.9743	~1	~1
	K_1	0.0066	0.0044	0.001	-0.0005	-0.0071	0.0202
	K_2	-0.0055	-0.0023	-0.0067	-0.0012	0.0049	0.0017
ρ_{diag} [-]	plain	0.9541	0.9536	0.9349	0.9572	~1	~1
	K_1	-0.005	0.0063	0.0005	-0.0034	0.0017	0.0007
	K_2	-0.001	0.0021	-0.0019	0.0018	0.0001	-0.0018
H [bits/ pixel]	plain	7.5883	7.106	6.8147	7.2344	0	0.0005
	K_1	7.9992	7.9992	7.9992	7.9992	7.9944	7.9942
	K_2	7.9992	7.9993	7.9993	7.9993	7.9944	7.9945
$NPCR$ [%]	K_1	99.7456	99.6094	99.6906	99.6792	99.6674	99.6429
	K_2	99.7318	99.6227	99.675	99.7074	99.6307	99.6521
$UACI$ [%]	K_1	33.6223	33.6513	33.6724	33.6274	33.6473	33.6338
	K_2	33.6308	33.6425	33.6356	33.6401	33.6186	33.6788
t_{enc} [ms]	K_1	598.481			191.1818	24.6219	24.2081
	K_2	601.264			193.2376	24.6448	24.1386
t_{dec} [ms]	K_1	552.0765			173.8621	22.1377	21.9269
	K_2	555.3535			174.073	22.1272	21.8784

Values of the correlation coefficients ρ for images *black1* and *black2* are close to 1, as the first image consists only of pixels with zero intensity. The second image has one pixel with different, maximal intensity level (255). Presented encryption times t_{enc} and decryption times t_{dec} for the true color image are obtained for encryption or decryption of all three color planes. A comparison of results with other algorithms which used the same images or color planes is shown in Table 7.

Table 7
Comparison of the numerical results with other algorithms

Value	Proposed algorithm		Ref. [8]	Ref. [10]
	<i>lena (R)</i>	<i>lenaG</i>	<i>lena (R)</i>	<i>lenaG</i>
ρ_{hor} [-]	0.0094	-0.0013	0.0135	-0.0278
ρ_{ver} [-]	0.0066	-0.0005	Unknown	-0.0065
ρ_{diag} [-]	-0.005	-0.0034	Unknown	-0.0074
H [bits/pixel]	7.9992	7.9992	7.9974	7.9895
NPCR [%]	99.7456	99.6792	99.63	99.66
UACI [%]	33.6223	33.6274	33.31	33.57
t_{enc} [ms]	598.481	191.1818	243.2	not reported
complexity [operations]	$O(16n)$		not reported	$O(8n)$

Based on the presented results, it could be stated that our algorithm achieves better values of the correlation coefficients ρ than the two other algorithms. Values of entropy H are also higher. NPCR and UACI are slightly over the critical values and they are also considerably better than the results obtained by other algorithms. The smallest difference between algorithms is present for NPCR value of the image *lenaG*. Also, the values of all mentioned parameters are quite similar for two tested keys. However, the advantages of our proposal are balanced by its slower performance – it is approx. 2.5 times slower than the approach from [8] and it has two times the computational complexity of the algorithm presented in [10]. However, the computational complexity of the scheme [10] depends on length of used feedback, which was chosen as 4 by the authors of algorithm [10].

Conclusions and Future Work

In this paper, we describe a modification of a chaotic logistic map, which was also employs in an image encryption algorithm. Image encryption can be used in various applications, such as an improvement of data security in steganographic systems [16] [17] or for secure transmission and storage of features in biometric systems which utilize images [18]. Because the modified version of the logistic map is more robust, to phase space reconstruction attacks, the encryption algorithm also holds this property. Other required properties of the algorithm are achieved by a combination of several techniques, such as key diffusion, ciphertext chaining or four step diffusion method. However, the number of these techniques causes slower performance of the proposed algorithm.

The proposed image encryption algorithm reaches correlation coefficients, with values < 0.01 for all planes of true color image *lena* and also for grayscale image *lenaG*. The computed results of entropy are close to the theoretical bound of 8 bits per pixel. Also, the arithmetic means of 100 repeated measurements of NPCR and UACI are equal to or higher than the required values reported by Wu et al. [15].

Therefore, the proposed algorithm is robust against all types of attacks commonly used for cryptanalysis of image encryption algorithms.

In the future, we plan to explore other possible techniques which would provide similar properties, with a smaller computational complexity. This goal may involve other modifications of the equation of the logistic map. The solution proposed in this paper multiplies iterates of the logistic map by a constant. This operation and the fact that the following iterate needs to be from an interval (0; 1) cause usage of modular arithmetic. Other operations, which do not change the interval of iterates (e.g. rearrangement of decimal places of iterates or combinations of multiple iterates), could be faster than the two operations utilized in this paper, nonetheless, the properties of other operations regarding phase space reconstruction attacks, needs further investigation.

Acknowledgement

This work was supported by the research grants KEGA 023TUKE-4/2017 and APVV-17-0208.

References

- [1] R. Matthews: On the Derivation of a “Chaotic” Encryption Algorithm. *Cryptologia*, Vol. 13, 1989, No. 1, pp. 29-42
- [2] J. Fridrich: Symmetric Ciphers Based on Two-dimensional Chaotic Maps. *International Journal of Bifurcation and Chaos*, Vol. 8, 1998, No. 6, pp. 1259-1284
- [3] Y. Mao, G. Chen, S. Lian: A Novel Fast Image Encryption Scheme Based on 3D Chaotic Baker Maps. *International Journal of Bifurcation and Chaos*, Vol. 14, 2004, No. 10, pp. 3613-3624
- [4] Z.-H. Guan, F. Huang, W. Guan: Chaos-based Image Encryption Algorithm. *Physics Letters A*, Vol. 346, 2005, No. 1-3, pp. 153-157
- [5] S. Li, G. Chen, X. Mou: On the Dynamical Degradation of Digital Piecewise Linear Chaotic Maps. *International Journal of Bifurcation and Chaos*, Vol. 15, 2005, No. 10, pp. 3119-3151
- [6] E. Solak, C. Çokal, O. T. Yildiz, T. Biyikoğlu: Cryptanalysis of Fridrich’s Chaotic Image Encryption. *International Journal of Bifurcation and Chaos*, Vol. 20, 2010, No. 5, pp. 1405-1413
- [7] F. Takens: On the Numerical Determination of the Dimension of an Attractor. *Dynamical Systems and Bifurcations. Lecture Notes in Mathematics*, Vol. 1125, Berlin Heidelberg: Springer, 1985, pp. 99-106, ISBN 978-3-540-15233-0
- [8] M. A. Murillo-Escobar, C. Cruz-Hernández, F. Abundiz-Pérez, R. M. López-Gutiérrez, O. R. Acosta Del Campo: A RGB Image Encryption

- Algorithm Based on Total Plain Image Characteristics and Chaos. Signal Processing, Vol. 109, 2015, No. C, pp. 119-131
- [9] L. Liu, S. Miao: A New Image Encryption Algorithm Based on Logistic Chaotic Map with Varying Parameter. SpringerPlus, Vol. 5, 2016, No. 1, pp. 289-300
- [10] C. Guanghui, H. Kai, Z. Yizhi, Z. Jun, Z. Xing: Chaotic Image Encryption Based on Running-Key Related to Plaintext. The Scientific World Journal, Vol. 2014, 2014, No. 1, pp. 1-9
- [11] Y. Liu, Y. Luo, S. Song, L. Cao, J. Liu, J. Harkin: Counteracting Dynamical Degradation of Digital Chaotic Chebyshev Map via Perturbation. International Journal of Bifurcation and Chaos, Vol. 27, 2017, No. 3, pp. 1-14
- [12] R. M. May: Simple Mathematical Models with Very Complicated Dynamics. Nature, Vol. 261, 1976, No. 5560, pp. 459-467
- [13] B. Kliková, A. Raidl: Reconstruction of Phase Space of Dynamical Systems Using Method of Time Delay. Proceedings of WDS'11, Prague (Czech Republic), 2011, pp. 83-87, ISBN 978-8-073-78186-6
- [14] Y. Rong-Yi, H. Xiao-Jing: Phase Space Reconstruction of Chaotic Dynamical System Based on Wavelet Decomposition. Chinese Physics B, Vol. 20, 2011, No. 2, pp. 1-5
- [15] Y. Wu, J. Noonan, S. Aghaian: NPCR and UACI Randomness Tests for Image Encryption. Journal of Selected Areas in Telecommunications, Vol. 2, 2011, No. 4, pp. 31-38
- [16] V. Hajduk, M. Broda, O. Kováč, D. Levický: Image Steganography with QR Code and Cryptography. Proceedings of Radioelektronika 2016, Košice (Slovakia), 2016, pp. 350-353, ISBN 978-1-509-01673-0
- [17] J. Oravec, J. Turán: Substitution Steganography with Security Improved by Chaotic Image Encryption. Proceedings of Informatics 2017, Poprad (Slovakia), 2017, pp. 284-288, ISBN 978-1-538-60888-3
- [18] F. Abundiz-Pérez, C. Cruz-Hernández, M. A. Murillo-Escobar, R. M. López-Gutiérrez, A. Arellano-Delgado: A Fingerprint Image Encryption Scheme Based on Hyperchaotic Rössler Map. Mathematical Problems in Engineering, Vol. 2016, 2016, pp. 1-15

SEFRA - Web-based Framework Customizable for Serbian Language Search Applications

Mioljub Jovanović^{*}, Goran Šimić^{}, Milan Čabarkapa^{*}, Dragan Randelović^{***}, Vojkan Nikolić^{****}, Slobodan Nedeljković^{****}, Petar Čisar^{***}**

^{*}Department for Postgraduate Studies, Singidunum University, Danielova 32, 11000 Belgrade, Serbia, mioljub.jovanovic.12@singimail.rs, mcabarkapa@singidunum.ac.rs

^{**}Research Centre for Simulations, University of Defense, Generala Pavla Jurišića Šturma 33, Banjica, 11000 Belgrade, Serbia, goran.simic@va.mod.gov.rs

^{***}Department for Informatics and Computing, Criminalistics and Police University, Cara Dušana 196, 11070 Belgrade, Serbia, dragan.randjelovic@kpa.edu.rs, petar.cisar@kpa.edu.rs

^{****}Ministry of Interior of the Republic of Serbia, Kneza Miloša 101, 11000 Belgrade, Serbia, vojkan.nikolic@mup.gov.rs, slobodan.nedeljkovic@mup.gov.rs

Abstract: This paper presents SEFRA – a web-based framework for searching Web content written in Serbian. SEFRA is an easily customizable hybrid solution that can be a platform for new search applications and/or a service for already existing ones. The proposed architecture solves the problems of indexing, searching and displaying search results adjusted for Serbian. It unifies several web technologies and services into one product suitable for use in the Western Balkan's countries for helping e-Government citizens' services and other public-sector services, private company administration, solving specific search problems for academic institutions and scientific literature publishers, etc. The proposed solution uses advanced Serbian language services accessible over the Web. It is also implementable for any other language where the target language morphology service exists. In other words, architecture is also customizable in this direction. It should be noted that the proposed architecture is optimized from both backend and web front-end perspective. The source code can be pulled from <https://bitbucket.org/mjovanov/pretraga/>. The one application of the proposed architecture is experimentally demonstrated through the search of crime law documents of Serbia. The experimental usage of this implementation shows that the problem of search relevance, is well-solved and easily customizable.

Keywords: web-based architecture; Serbian language text search; software implementation; search results

1 Introduction

An accelerated development of the Internet as a platform and WWW (Web) as the most frequently used service of the Internet, brought access to a huge number of documents on the global network. Moreover, the documents' content is distributed in the same way. The page can also consist of fragments that originate from different hosts. Considering such a complex situation, advanced search application developers face many challenges, such as: collecting all available pages, analyzing the content of the collected material and enabling a quick query, as well as, display relevant documents based on the specified search criteria.

Since it is a fact that English is the most commonly used language on the Web [1], there are many representative search applications and services specialized for English content (e.g. Google, Bing...). In other words, we are in a position where the problem of collecting, analyzing documents, and finding the search results is solved for the English language. Since there are significant differences between the languages of the Western Balkans and the English language, then the question arises – if the data search problem is thoroughly resolved in English, is the problem of indexing and searching documents in our local languages also solved?

This paper suggests the possible approach, through the example of the realization of modern web architecture, to provide the answer to the above question of indexing, searching and displaying the results adjusted for the Serbian language. This work certainly would not be possible without going through various research documents and reports which are mentioned in Section 2. The architecture of the proposed solution along with its used components are discussed in Section 3, while implementation details and most code excerpts are covered in Section 4 of the document. Evaluation of results is given in Section 5, followed by conclusions and potential future work.

2 Related Works

Nikolic *et al.* [2] presented one e-Government services to get quick responses. This service enables citizens to receive answers, in the form of documents in the Serbian language, at any time and in any place, to the questions in the criminal law domain. This service has developed a Question and Answer (Q&A) system, based on Bag of Words (BoW) and Bag of Concepts (BoC), for categorizing text and incorporating background knowledge. The automatic mapping of relevant documents stands out as an important application for automatic question-documents classification strategies. This research presents a contribution to the identification concepts in text comprehension in unstructured documents as a significant step towards clarifying the role of explicit concepts in retrieval

information in general. These authors introduce a new approach to create concept-based text representations and apply it to a text categorization collection in order to create predefined classes in the case of a short document analysis document. In the revolutions of this Q&A system, is a classification-based algorithm for a question matching topic model. The results obtained proved to be satisfactory based on the "golden rule".

In article Martinovic et al. [3] is presented an information retrieval system for Serbian language. Approaches designed and adopted to handle them are depicted and illuminated in this article. As a backbone of this system, they used a SMART retrieval system which they augmented with features necessary to deal with the specifics of the Serbian alphabet. Serbian language is a morphologically rich language that leads to specific implications of the text prefix. During the development a SMART retrieval system, the authors developed two algorithms which increased retrieval precision by 14% and 27%, respectively. Complete testing was conducted using two gigabyte EBART collection of Serbian newspaper articles.

Considering the existing solutions that depend on e-Government requirements, Šimić et al. in [4] proposed focusing on testing in different conditions and improving the ability of adaptation in the next research phases. One of the objectives pursued in this work is to find solutions for the functioning of such a system in multilingual environments and increasing content complexity concerning grammar and dictionaries of different languages, regardless of the area of use.

Kolomiyets et al. [5] represent the Question Answering method as a comprehensive approach that provides a qualitative way for information retrieval. This approach is a system of queries and documents in relation to the possible functions of search to find an answer. This research discusses general questions contained in a complex architecture with increasing complexity and the level of frequency of questions and information objects. These authors represent here a method of how natural language roots are reduced on keyword for search, while knowledge databases, and resources, obtained from natural language questions and answers, are made intelligible.

In addition, there are now research efforts where the authors try to solve a specific language searching problem [6] - [13], but there is no complete software architecture easily customizable for different search applications. In [6] author's give one optimization of the method proposed in [2] where selection of the similarity measure is performed using the principles of redundancy and fault tolerance, in [7] is described one search engine using MySQL as one of cheap option, work [8] presents one architecture which uses different semantic web technologies and builds one prototype of semantic web mashup possibility, paper [9] proposes one novel Italian Sign Language Multi Word Net using process of integration the Multi Word Net lexical database and the Italian Sign Language,

paper [10] describes a novel LInSTSS approach which is suitable for using to create a software tool which is capable to determine the semantic similarity of two presented no large texts, in paper [11], authors propose the use of smoothed n-gram language models to classify tweets as a typical short texts from Twitter in both Portuguese languages - Brazilian and European variants, paper [12] deals with the software architecture which establishing electronic services for searching and presentation in an information system on scientific activities of the Ministry of Education, Science and Technological Development of the Republic of Serbia and work [13] has objective to give a lexicon based algorithm which is able to perform different natural language identification using minimal training data in the obligatory process of machine learning because this step is often the first step in many natural language processing tasks which is normally necessary to make in the shortest possible time. Therefore, we have a strong motive for designing the SEFRA framework – hybrid solution based on existing Web services and technologies (framework source code is available at: <https://bitbucket.org/mjovanov/pretraga/>). Additionally, there is a search application developed for demonstration and testing SEFRA (the implementation available online: <http://88.99.175.85/pretraga/>).

3 Proposed Web Architecture

According to previous researches and already existing implementations [14] [15], there are four processes necessary to obtain relevant search results (Figure 1). During processing, targeted content passes through the two stages: collecting, preparing and indexing belong to the preparation stage while query processing, searching and presenting belongs to the production stage.

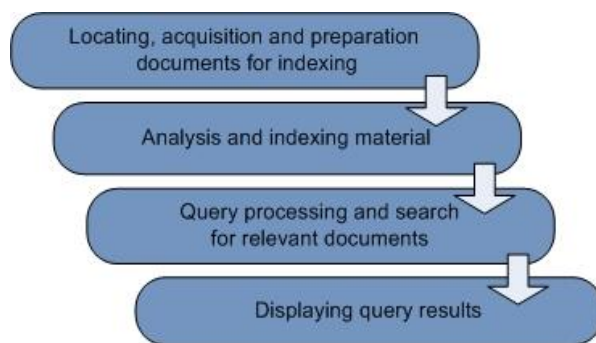


Figure 1

Four processes necessary for Web content searching

Additionally, entering the production stage, the system must run these four processes simultaneously. This is a consequence of permanent changes of the content. For instance, constantly adding, removing, updating and replacing of documents and their references (URI) is a common case on the global network. Such challenges as well as a complex nature of Serbian language directed SEFRA design (Figure 2) to be modular solution based on open components.

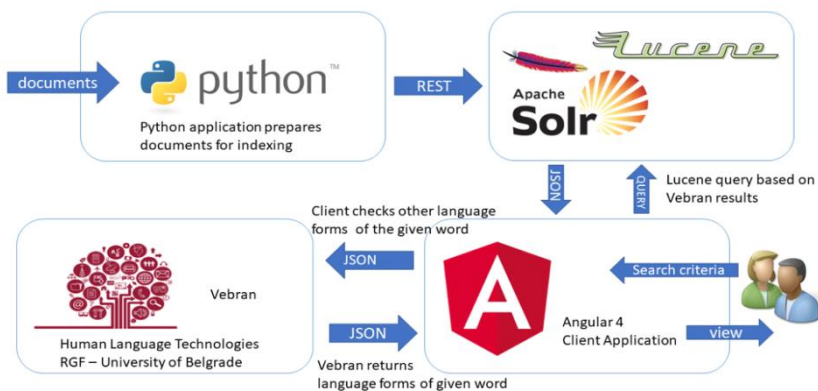


Figure 2

SEFRA architecture - modular solution based on open components

The developed solution is an open and modular framework which consists of four components. For preparation stage, SEFRA uses SolrClient – Python library (solrclient.readthedocs.io) for locating, collecting, and preparing documents. Further documents' analyzing and indexing SEFRA performs by using Apache's Solr and Lucene libraries (lucene.apache.org/solr/).

Solr is a popular fast – search platform based on Lucene technology. Both are developed under Apache Foundation [16] as platforms for full text search written in the Java programming language. Lucene [19] represents a framework with high – performances in full-text search. Designed as a system centered solution, Lucene API is complex for implementing special requests and search customization. For this reason, Solr is a solution dedicated to enable simpler interface and better customization abilities for Lucene with resources accessible locally as well as remotely (through the REST API).

For production stage, SEFRA uses a reach Web client application based on Angular 4 (angular.io) [18] and Bootstrap 4 (v4-alpha.getbootstrap.com) [19] libraries. SEFRA client communicates with the end users as well as with the external services. In concrete scenario, SEFRA uses the Vebran – Serbian language service (hlt.rgf.bg.ac.rs/VeBran) to obtain lemmatization of query (transforming its words into normal form). Then, it sends a prepared query, as a

REST request to the Solr search service. After receiving search results, the SEFRA Client prepares the representation for delivering to the end user.

4 Implementation

4.1 Preparation Stage Processes

As mentioned, SEFRA uses *SolrClient* – Python library for locating, collecting, and preparing documents. *Python* is a modern, easy-to-use programming language, which contains many libraries useful for acquiring documents, analyzing them and creating fields and schemas for indexing as well. In SEFRA the *SolrClient* represents a module which works together with Solr server acting as an interface between Solr and rest of the system as well as with the outer world. In other words, SEFRA uses *SolrClient* instance to retrieve local or remote documents and to prepare them for later indexing and searching. It leverages the potential of the Solr server quickly and easily from the *Python* environment, facilitating the use of the REST interface of the *Solr* platform.

For analyzing of acquired documents *Solr* server by default tries to find out fields and their types based on the content. This approach is interesting when objective is to index text in English as *Solr* uses built in functions (and thesaurus) designed for English language in this process. This automatized indexing unfortunately produces unexpected results for content written in other languages (e.g. Serbian). In this case, *Solr* recognizes the sentences in the Serbian language in the wrong way – meaningless strings often considered as a single string instead at least, a list of words. Therefore, adding new pre-defined fields is necessary for making improvements of indexing process.

SolrClient allows adding of new fields or metadata to a document, specifying the type of field(s) and schema on the server, preparing it for further document processing. There are several *SolrClient* functions for this purpose. In the concrete scenario, *solr.schema.create_field* and *solr.index* methods are used to add desired fields in the scheme and to add indexes to documents. To enable proper processing of words and sentences in Serbian, the default *Solr* behavior has been adapted to manage specific language requirements such as grammatical cases, synonyms, stop words and processing of diacritics.

There is a REST service designed for this purpose. It is accessible over *Solr*'s URI *solr/sd/schema*. It enables remote setup of *Solr* server by using simple JSON formatted messages (Figure 3) sent as a HTTP request (*application/json* type).


```

3  { "add-field-type": { "name": "text_rs", "class": "solr.TextField", "positionIncrementGap": "100",
4    "multiValued": "true",
5    "analyzer": {
6      "charFilters": [ {
7        "class": "solr.PatternReplaceCharFilterFactory",
8        "replacement": "$1$1",
9        "pattern": "[a-zA-Z]\\\\\\\\|+" },
10     "tokenizer": { "class": "solr.StandardTokenizerFactory" },
11     "filters": [ { "class": "solr.StopFilterFactory", "ignoreCase": "true", "words": "stopwords_rs.txt" },
12       { "class": "solr.SynonymFilterFactory", "synonyms": "index_synonimus.txt",
13         "ignoreCase": "true", "expand": "false" },
14       { "class": "solr.LowerCaseFilterFactory" },
15       { "class": "solr.SerbianNormalizationFilterFactory", "haircut": "bald" } ] ],
16     "add-field" : { "name": "tekst", "type": "text_rs", "multiValued": "true",
17       "indexed": "true", "stored": "true" }
18 }

```

Figure 3

Solr server setup for indexing documents written in Serbian

The above figure shows the complete process of creating a *text_rs* field type. This field is based on similar solutions for other languages that are supported in *Solr*, and it has been defined including new entities into *Solr* core: stop-words filter (*stopwords_rs.txt*), filter for synonyms (*index_synonimus.txt*), filter for upper / lower cases (already built-in resource) and filter for the processing of diacritical signs (*SerbianNormalizationFilterFactory*).

SerbianNormalizationFilter is Java based Lucene library which is implemented and leveraged for use in *Solr* custom field “*text_rs*”, as depicted on Figure 3. Since Serbian language uses Cyrillic alphabet as well as Latin it’s possible to perform several language-specific steps in order to render both indexing and querying process more effective. Example implementation in SEFRA uses *haircut=“bald”* which removes all diacritics from letters such as ‘č’, ‘ć’, ‘š’, ‘ž’ or ‘đ’ from the search text which reduces the precision of the query results. However, since *Vebran* is invoked as a backend service before *Solr* query, text in query field entered with diacritics will be correctly returned in all shapes and forms. Finally, these settings are stored in *Solr* server configuration file ready for use.

After creating fields and setup *Solr*, the documents are ready to be loaded into index. Next Python code demonstrates this process (Figure 4). In this example, the origin of documents to be indexed is in *dokumenti/Na1* folder. The system opens each file to read its content and put it in the previously created field named *tekst* for further analysis. After the collection is completed, document is sent to a server by forwarding it as a parameter of *SolrClient (solr) index()* method. Example depicted on Figure 4 demonstrates use of Python code to enrich index by additional fields which may be relevant for such – such as “*clan_id*”, which is the exact article id referenced in the original text search corpus. Python code provides capability to easily extend given code to any number of additional parameters or fields which may be required to properly index the text.

```
1 clanovi = []
2 clanovi_files = os.listdir("dokumenti/Na1")
3 for clan_file in clanovi_files:
4     clan = {}
5     clan_file_path = os.path.join(INPUT_DIR_CLANOVI,clan_file)
6     f = io.open(clan_file_path,"r",encoding='utf8')
7     clan_tekst = f.read()
8     clan["tekst"] = clan_tekst
9     match = re.match("clan_(.*)\.txt", str(clan_file))
10    clan_id = match.group(1)
11    clan['clan_id'] = clan_id
12    clanovi.append(clan)
13    f.close()
14    logging.info("DOCUMENT {} IS LOADED".format(clan_file_path))
15
16 solr.index('sd', clanovi,commit=True)
```

Figure 4

Preparing documents for indexing

4.2. Production Stage Processes

SEFRA performs production stage processes in both of entities – *Solr* server and *Angular* clients. The distribution of functionalities reduces stress to the server side by using powerful client technology. The main reasons in favor of using the Angular 4 as a client platform [20] are:

- Separating of the user interface from the business logic
- Modularity (Enabling flexible design of low coupled forms and logic)
- Asynchronous calls support (Easy to code client-side multithreading)
- Rich user interface support (forms based on templates as well as program-generated forms)
- Reusable and responsive components (support for Angular Material and Bootstrap 4)

The list of features and advantages embedded in *Angular 4* is a quite long. TypeScript is language of choice in our implementation. There are several reasons for choosing TypeScript: modular development support (object oriented language semantic), easy compiling (*transpiling*) into *JavaScript* code and compilation time error detection, etc. Moreover, *Angular* is written in *TypeScript*, which is obviously a huge advantage if one develops application in the same programming language as the platform itself. Typescript is a super-set of JavaScript that will be transpiled into pure JavaScript code [21].

The core production – stage functionality is to process searching queries, prompted by the users, based on existing services specialized for Serbian language. This client-side module provides integration of *Solr* indexing &

searching services with *Vebran – Delafs* services (Vocabulary of word forms with all their morphological properties) [3]. *Vebran* is Serbian linguistic web-based service offering different linguistic capabilities for semantic and morphologic extension of the given phrase. Implementation of *Vebran* service invocation is explained in page 69 and Figure 4. As an example, for input search text “delo”, *getDelafs* method will invoke *Vebran* API “/Vebran/api/delafs/delo” which returns all other morphological shapes of the given input text, which is:

```
<string xmlns="http://schemas.microsoft.com/2003/10/Serialization/">
```

```
delo;delima;delo;delom;delu;дела;делима;дело;делом;делу</string>
```

With *Vebran* it is also possible to leverage other linguistic services, such as synonyms using API “/Vebran/api/sinonimi/”, which for given term “delo” also gives it is synonyms:

```
<string xmlns="http://schemas.microsoft.com/2003/10/Serialization/">
```

```
delo;opus;podvig;rad;duhovnatvorevina;дело;опус;подвиг;рад;духовна  
творевина</string>
```

On one hand, with the Web services used on the server side, SEFRA client communicates asynchronously. On the other hand, there has to be synchronization between Solr and *Vebran* services during this process. Therefore, SEFRA uses *RxJS Observable Library* (*Observer* design pattern) [22] [23] as appropriate one for handling multiple service requests and responses simultaneously. This way it becomes possible to extend the number of services used. Enabling full control of concurrent execution, *Observable* also simplifies the termination of running asynchronous calls if timeout is expired. In SEFRA example implementation *Vebran Delafs* API has been leveraged, yet all other services can be easily involved and processed using *Observable* design pattern as described in the remainder of the paper.

4.3. User Interface

SEFRA uses *Bootstrap Navbar* component (getbootstrap.com) for creating end – user interface. It enables responding to the device size (Figure 5), easy navigation and intuitive interface. There is one menu bar with drop down menus, modelled through the appropriate classes.

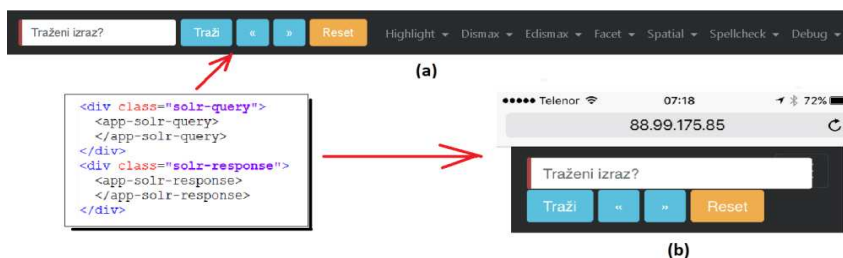


Figure 5

Responding GUI – the same code produces different appearance for desktop (a) and smartphone (b) display

Angular enables clear separating of components and easy arranging of elements on the web page. It also reduces the content of HTML elements combined them with the ones defined in *Angular*. Previous picture illustrates it – GUI contains only *div* HTML elements, while *app-solr-query* and *app-solr-response* are user-defined Angular components. Their definitions are split in three parts - three files generated for each active GUI element in Angular (Figure 6): HTML, CSS and *TypeScript (ts)* files (there is additional *spec.ts* file only for testing purposes).

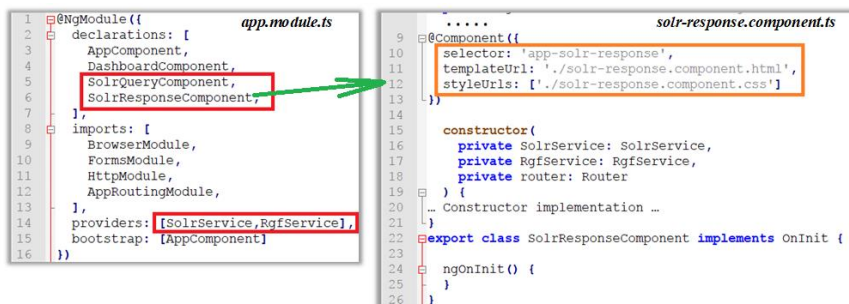


Figure 6

Modular design of Angular application

HTML and CSS files contain details of design, while the *ts* files implement the application logic (interactivity). In addition, a separate file (i.e. *app.module.ts*) contains declarations for each GUI component, imports of built-in library modules and other specifications necessary for running the application. Consequently, the described approach produces low coupling between presentation and functionality behind it enabling if–necessity, or on–demand, easy replacing any of these two.

4.4. The Search Query Processing

App-solr-query is the frontend Angular component (previous section) which receives searching criteria entered by the user (Figure 7) putting it in the variable named *guiQ* (line 14). There is bounded variable of the same name defined in *SolrQueryComponent* class. SEFRA starts searching process when the user clicks the button labeled *Traži* (line 15). This event triggers *searchClanovi* method defined in background class (*SolrQueryComponent* class).

```
solr-query.component.html  X
14 | <input required type="text" class="form-control" [(ngModel)]="guiQ" name="guiQ" #clanN ....
15 | <button class="btn btn-info" (click)="searchClanovi(solrquery);" type="submit" >Traži</button>
```

Figure 7

Fragment of app-solr-query component

Method *SolrQueryComponent.searchClanovi* starts the *RgfService* firstly (Figure 8), calling its observable method *getDelafs* forwarding the user search criteria as created *query* instance to normalize it by finding the basic form of each word in a query. In more details, the system consequently calls additional three observable operators: *map* – calls the function for each (query) element found, *mergeMap* – calls *SolrService* method named *getClanovi* enabling the use of more than one service at the same time and merging their results, and *subscribe* – that enables an observer object to receive items emitted by an observable instance. In the concrete case, the *SolrQueryComponent* is a subscriber. That means that any change in a query string produces a new request to the Solr service and updates the results presented to the user.

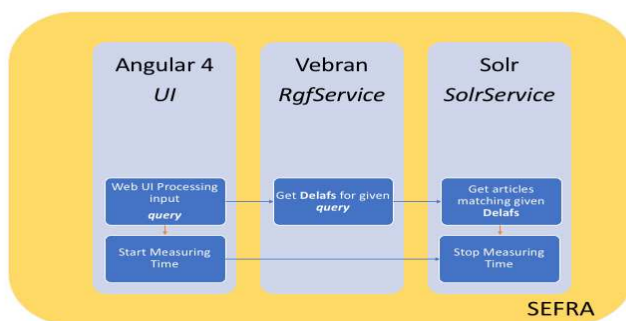


Figure 8

The core function of SolrQueryComponent

In the last statement, *SolrQueryComponent* uses RxJS synchronization to propagate the information through the rest of the system that initiates the search of specified query (*SolrService.announceQueryStart*).

As mentioned, SEFRA uses *Vebran* service to pre-process the search criteria originally sent by the user to obtain regular query expression. It happens through the *RgfService* class (Figure 9). More precisely, running in separate process thread, *getDelafs* method implements it. Due to *getDelafs* method, instances of *RgfService* class become observable for consumer class instance.

```
getDelafs(parameters: query)
  Set q to query value entered in Web UI
  Initialize params object as empty instance of URLSearchParams class
  Set helper array paramsArr to contain all query keys, values used in UI
  Populate all keys, values from paramsArr into params object
  Get Delafs response from Vebran service and map it to response object
  Return response
```

Figure 9

Angular service implementation for Vebran query

SolrQuery is a class that contains everything necessary for performing a search. SEFRA communicates with the Vebran service through the HTTP GET request sending a query as JSON formatted string. The method *getDelafs* returns content received from the Vebran service to the observer (*SolrQueryComponent*).

In the same manner, *SolrQueryComponent* leverages the other service encapsulated in *SolrService* class. After query normalization, *SolrService*'s method *getClanovi* forwards query as a JSON string through the HTTP GET request to the *Solr* server (

```
getClanovi(parameters: query)
  Set q to query value entered in Web UI
  Initialize params object as empty instance of URLSearchParams class
  Set helper array paramsArr to contain all query keys, values used in UI
  Populate all keys, values from paramsArr into params object
  Get Articles response from Solr service and map it to response object
  Return response
```

Figure 10).

```
getClanovi(parameters: query)
  Set q to query value entered in Web UI
  Initialize params object as empty instance of URLSearchParams class
  Set helper array paramsArr to contain all query keys, values used in UI
  Populate all keys, values from paramsArr into params object
  Get Articles response from Solr service and map it to response object
  Return response
```

Figure 10

Sending of search query to the Solr server

Since *getClanovi* is an observable method, it returns content received from the service to the observer (*SolrQueryComponent*). More precisely, it triggers *updateSolrResults* method subscribed to wait this response (Figure 8, line 138) encapsulated in *ResponseSolr* instance for search results (Figure 11). It contains a set of highlighted document fragments that fit the criteria, the links and similarity scores as well.

```
updateSolrResults(parameters: response)
    Cast response parameter as ResponseSolr class
    Get Articles located in response.docs as array of instances of Clan class objects
    Announce that query has finished, so that timers could measure elapsed time
```

Figure 11

The last preparations before presenting search results

After the preparation, *SolrQueryComponent* calls the *SolrService* to announce that the searching is finished, and results are ready for presenting. Further, the system broadcast this information for updating GUI components that present the results.

4.5. Presentation of Search Results

An appropriate binding between GUI components and needed features of the SEFRA provides handling of the user requests and system responses separately because there are several services and asynchronous calls used for this purpose. *SolrService* class is responsible for a mutual synchronization in this complex process. It uses broadcasted events (*observables*) for triggering appropriate component functions. After completing the query task, *SolrQueryComponent* forces *SolrService* to emit this information to the all subscribed objects (Figure 11, line 89). It performs this task by using observable string subject *queryFinishedSource* and observable string stream *queryFinished\$* (Figure 12).

```
32     private queryFinishedSource = new Subject<string>();
33     queryFinished$ = this.queryFinishedSource.asObservable();
34     announceQueryFinish(query: string) {
35         this.queryFinishedSource.next(query);
36     }
```

Figure 12

Broadcasting the event when the search has finished

SolrResponseComponent is a class responsible for presenting searching results. Subscribing the stream named *queryFinished\$* (Figure 13), it receives the event when the search has finished and prepares the results for rendering. *SolrResponseComponent* extracts all the information necessary to perform mentioned task.

```

54 SolrService.queryFinished$.subscribe(
55   q => {
56     this.clanovi = this.SolrService.clanovi;
57     this.response = this.SolrService.response;
58     this.responseJSON = JSON.stringify(this.response);
59     this.responseHeader = this.SolrService.response.responseHeader;
60     this.vebran = this.RgfService.vebran;
61     this.responseHeaderParams = JSON.stringify(this.responseHeader.params);
62     this.queryInProgress = false;
63     if (this.responseHeader.status == 0) {
64       this.success = true;
65     }
66   });

```

Figure 13

SolrResponseComponent subscribed for event that the search is finished

Angular GUI component *app-solr-response* is responsible for presenting the search results to the user (Figure 14). This component is acting along with *SolrResponseComponent* class and they share the same scope. In other words, the variables defined in *SolrResponseComponent* are visible to *app-solr-response* and vice versa.

```

1 <div *ngIf="success" id="response-success" class="bs-callout bs-callout-warning">
2   <...</>
31 </div>
32 <div id="rezultat" class="col-sm-12 col-xs-12">
33   <ul class="clanovi list-inline">
34     <li class="list-inline-item"
35       *ngFor="let clan of clanovi" [class.selected]="clan === selectedClan">
36       <span class="badge">Clan br.{{clan.clan_id}}</span>
37       <span class="badge bg-info">TF-IDF:{{clan.score | number}}</span>
38       <span class="badge bg-success">
39         Relevantnost:{{clan.score / response.response.maxScore | percent}}
40       </span>
41       <p [innerHTML]="clan.tekst"></p>
42     </li>
43   </ul>
44 </div>

```

Figure 14

GUI component

This way a variable *success* is examined by using *ngIf* directive to check the conditions if there are quantitative details of the search process (this part of code is collapsed) which should be shown. The collection *clanovi* represents the search result that is iterated through by using *ngFor* directive. The component represents each element in this collection by using temporary variable *clan* showing its id, absolute and relative score, and textual content.

5 Evaluation

For evaluation purposes, the set of criminal law documents written in Serbian Latin is used as a searching content. SEFRA framework shows a flexible behavior for different test cases. The following examples illustrate it. In the first one, irrespectively whether the search criteria have been written in different alphabets – Latin (Figure 15a) or Cyrillic (Figure 15b), SEFRA obtained the identical results in both cases. The ranking, similarity measures and relevance were the same. Moreover, the example shows that SEFRA responds appropriately on inaccurate written query – *novcana kazna* (eng. *Fine/monetary penalty*). Instead of correcting word *novčana*, word *novcana* is used. In other words, it compensates the case in which the user cannot use specific letters of national alphabet (e.g. mobile devices or keyboard without this kind of support).



Figure 15

The same query written in different alphabets produces the same result

SEFRA also has a flexible search for different word forms found in documents. For concrete searching term *kazna zatvora* (eng. *Prison sentence*), it responds with document ranking that shows the terms are counted regardless of their forms (singular/plural, tenses, grammatical cases etc.). Consequently, there is a minor influence of word forms on a final documents' rank.

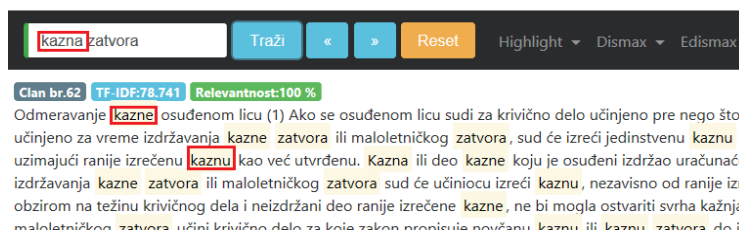


Figure 16

Flexible response on different word forms

For more details in search criteria SEFRA produces better results in document selection and ranking. Next example shows expanded criteria *kazna zatvora za ubistvo* (eng. *Prison sentence for murder*, Figure 17a) related with the previous one *kazna zatvora* (eng. *Prison sentence*, Figure 17b). The best-fit (100% relevancy) document explains a *murder* as a term and time range the jury can punish the accused with, while previously first-ranked document is shifted down the list.



a)



b)

Figure 17

The more details in search criteria the better results in response

Additionally, there is a quantitative analysis performed through two types of measurement. The first one shows the *Solr* service response time (Figure 18a). As expected, service-processing time is 5 to 10 times less than service-delivering time. A satisfactory fact is that the *Solr* aggregate response time varies from 10^{-2} to 10^{-1} seconds.

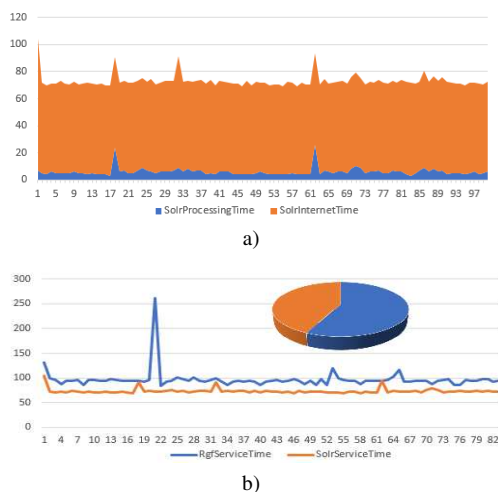


Figure 18

SEFRA quantitative analysis

As SEFRA includes using of outside services, its performances depend on their response time. In the concrete study, *Vebran* and *Solr* services are used (see Section 3). The second chart (Fig. 18b) presents the performance of these two services used together. It shows timelines (x-axis values are in hours) of their responses. It is obvious that (excepting one remarkable peak produced by *Vebran*) there are minor differences between them. On the other hand, pie chart shows that *Solr* consumes 43% while *Vebran* consumes 57% of aggregate response time which varies from 156 to 236 ms with the average response time of 170 ms. Also, in comparison with previous work [2], when consider precision, recall and accuracy metrics, SEFRA has obtained better validity results performance in considerably more advanced testing conditions but in comparison with work [6] which is one optimization of algorithm described in [2] it has poorer results. These results are an excellent starting point for further development.

Table 1
Validity performance results

PERFORMANCE	Work [2]	Work [6]	SEFRA
Precision	75.71%	49.67%	78.3%
Recall	57%	49.67%	57.76%
Accuracy	46.66%	74.83%	48.21%

Conclusions

SEFRA represents one of the rare solutions focused on improving search capabilities of specific (*non-English*) language(s). Regardless of a significant

progress made in natural language processing (NLP) of Serbian and other western Balkan languages (i.e. south Slavic languages group), neither commercial solutions, nor similar public services exist yet. Fortunately, there are Web services (as *Vebran*) that enable further NLP development offering their capacities to the researchers and developers. Moreover, SEFRA brought important contribution as new TypeScript libraries for Angular 4 framework both for Solr as well as Vebran backend services. Therefore, new Solr and Vebran libraries written as a part of SEFRA can be easily reusable by research and development community, so that powerful language and search services can be leveraged by simple instantiation of Classes defined in SEFRA libraries.

From design prospective, SEFRA is a hybrid Web framework with processing power balanced between advanced application delivered to the platforms on the client side and different Web services on server side. At the same time, it represents a proof of concept that it is possible to make reliable and efficient synchronization of different Web services on the client side. This approach which is very similar to *observer* design pattern, enables easy subscribing of new services in the roles of observables (e.g. for pre-processing the search queries, or for displaying results to the end users), or observers (e.g. different Web search services that can be used). This flexibility is also useful to relieve search engines, as SEFRA will not engage search services if there are no results returned from query pre-processing service(s).

During the evaluation, SEFRA satisfied the expectations of various searching tasks preformed. The collection of criminal law documents written in Serbian Latin enabled us a full control during this process (comparison of obtained and expected results). Domain experts agreed almost with all search results, which means SEFRA made reliable selection and ranking of documents. Inability to handle search queries that include different forms, cases and tenses and delivering nothing, or unexpected results, represent the main weaknesses of already existing searching engines for Serbian content. Therefore, such problems became also high-priority evaluation tasks. SEFRA uses *Vebran* service as a part of solution. Specific set up of *Solr* server as a search engine represents the other service being used. The Solr service was prepared by adding new fields, configuring the stop-words set and set of synonyms, and modifying indexing schema. In other words, the issues described above became solvable by using different services combined and synchronized in the joint solution.

As a modular and flexible framework built of low-coupled and easy-to-change components that interact with each other through the standardized services, SEFRA provides conditions for making modifications and improvements permanently. The core component is *Angular* multithread application (SEFRA client) that can manage any number of services involved in the search process. Holding services in the separate threads, SEFRA client synchronizes the API calls and mutual information exchange making them to act as a whole. Alternatively, it

delivers the results to the end user in seconds, hiding a lot of processing performed on various distributed platforms.

There are several ways for future development. Improving the search quality by including new services is one of them. For instance, there are many foreign companies running their business in Serbia. Including Bi/Multi-lingual services can significantly improve SEFRA usability. On the other hand, it is necessary to index as much available content as possible. Increasing quantity of the content results in the need for a content categorization (clustering). Moreover, it implies separate, domain-specific dictionaries for this purpose. Finally, every new (domain-specific) collection requires resetting of search engine(s). As Web content is constantly changing, the researchers and developers face the challenges in the same manner. SEFRA provides well-formed infrastructure for such efforts.

References

- [1] Miniwatts Marketing Group, "Internet World Stats", Miniwatts Marketing Group, [Online]. Available: <http://www.internetworldstats.com/stats7.htm> [Accessed August 2017]
- [2] V. Nikolić, B. Markoski, K. Kuk, D. Randjelović, P. Čisar, "Modelling the System of Receiving Quick Answers for e-Government Services: Study for the Crime Domain in the Republic of Serbia", *Acta Polytechnica Hungarica*, Vol. 14, No. 8, pp. 143-163, 2017
- [3] M. Martinović, S. Vesić and G. Rakić, "Building an Information Retrieval System for Serbian - Challenges and Solutions", *Springer International Publishing*, 2015
- [4] G. Šimić, Z. Jeremić, E. Kajan, D. Randjelović and A. Presnall, "A Framework for Delivering e-Government Support", *Acta Polytechnica Hungarica*, Vol. 11, No. 1, pp. 79-96, 2014
- [5] O. Kolomiyets and M.-F. Moens, "A Survey on Question Answering Technology from an Information Retrieval Perspective", *Information Sciences*, No. 181, pp. 5412-5434, 2011
- [6] S. Nedeljković, V. Nikolić, M. Čabarkapa, J. Mišić, D. Randjelović, "An Advanced Quick-Answering System Intended for the e-Government Service of the Republic of Serbia", *Acta Polytechnica Hungarica*, paper accepted for publishing in 2019
- [7] C. Gyrodi, R. Gyrodi, G. Pecherle and G. Mihai Cornea, "Full-Text Search Engine Using MySQL", *Int. J. of Computers*, Vol. V, pp. 735-743, 2010
- [8] S.-C. Necula, "Implementing the Main Functionalities Required by Semantic", *International Journal of Computers Communications & Control*, Vol. 7, No. 5, 2012

-
- [9] U. Shoaib, N. Ahmad, P. Prinetto and G. Tiotto, “Integrating MultiWordNet with Italian Sign Language Lexical Resources”, *Expert Systems with Applications*, Vol. 41, No. 5, pp. 2300-2308, 2014
- [10] B. Furlan, V. Batanović and B. Nikolić, “Semantic Similarity of Short Texts in Languages with a Deficient Natural Language Processing Support”, *Decision Support Systems*, Vol. 55, No. 3, pp. 710-719, 2013
- [11] D. W. Castro, E. Souza, D. Vitório, D. Santos and A. L. Oliveira, “Smoothed n-gram Based Models for Tweet Language Identification: A Case Study of the Brazilian and European Portuguese National Varieties”, *Applied Soft Computing*, pp. 1568-4946, 2017
- [12] D. Ivanović, D. Surla, M. Trajanović, D. Misić and Z. Konjović, “Towards the Information System for Research Programmes of the Ministry of Education, Science and Technological Development of the Republic of Serbia”, *Procedia Computer Science*, Vol. 106, pp. 122-129, 2017
- [13] A. Selemat and N. Akosu, “Word-Length Algorithm for Language Identification of Under-Resourced Languages”, *Journal of King Saud University - Computer and Information Sciences*, Vol. 28, No. 4, pp. 457-469, 2016
- [14] G. Kowalski and M. Maybury, *Information Storage and Retrieval Systems*, Springer US, 2002
- [15] H. Bast and B. Buchhold, “An Index for Efficient Semantic Full-Text Search”, in *International Conference on Information and Knowledge Management, Proceedings*, 2013
- [16] Apache Solr, “Overview of Searching in Solr”, 2017 [Online] Available: https://lucene.apache.org/solr/guide/6_6/
- [17] M. White, *Enterprise Search*, 2nd Edition ed., O'Reilly Media, Inc., 2015
- [18] Google, Inc, “Angular Fundamentals – Overview”, 2017. [Online]. Available: <https://angular.io/guide/architecture>
- [19] Twitter Bootstrap, “Bootstrap”, 2017 [Online] Available: <http://getbootstrap.com/>
- [20] Y. Fain and A. Moiseev, *Angular 2 Development with Typescript*, Manning Publications, 2016
- [21] “Microsoft Typescript project on GitHub”, 2017 [Online] Available: <https://github.com/Microsoft/TypeScript>
- [22] SourceMaking.com, “Design Patterns”, 2017 [Online] Available: https://sourcemaking.com/design_patterns
- [23] S. Salehi, *Angular Services*, Packt Publishing, 2017

Optimum Control Parameters of Switched Reluctance Motor for Torque Production Improvement over the Entire Speed Range

Mahmoud Hamouda^{1,2}, László Számel¹

¹ Budapest University of Technology and Economics, Department of Electric Power Engineering, Egry József utca 18, H-1111 Budapest, Hungary

² Mansoura University, Electrical Engineering Department, El Gomhouria street 25, Mansoura 35516, Egypt

E-mail: m_hamouda26@mans.edu.eg, szamel.laszlo@vet.bme.hu

Abstract: The switched reluctance motor (SRM) is a powerful candidate for many domestic and industrial applications. However, the double salient structure and discrete commutation process make it very difficult to acquire the analytical model of SRM. The performance optimization of SRM is achieved mainly based on the observation and analysis of its static magnetization characteristics. This paper presents multi-objective optimization of SRM's control parameters for optimum motor operation over wide range of speeds. The optimization aims to achieve the maximum torque production with the lowest copper loss. A searching algorithm is developed to find the base values as they vary for each operating point. The objective-function is calculated using a dynamic/actual simulation model of SRM. For a highly trusted model of SRM, the static magnetization characteristics of tested 8/6 SRM are measured experimentally. Then, the measured data are used to build the model in a MATLAB/Simulink environment. The proposed control is implemented using an artificial neural network (ANN). A series of simulations and experimental results are obtained to show the feasibility of the proposed control.

Keywords: switched reluctance motor (SRM); control parameters; optimization; artificial neural network; MATLAB/Simulink; experimental

1 Introduction

Due to their attractive features, switched reluctance motors (SRMs) have attracted increasing attention over the past few decades. They have a simple and rugged construction, low-cost of manufacturing, high-reliability, wide range of operating speeds, fault-tolerance, and high-efficiency [1]-[5]. Over the last decades, various researchers have been directed to improve the drive performance of SRMs for several applications like electric vehicles [6], aerospace [7], ships [8], wind power generation [9] and household appliances [10]-[11]. However, the double salient

structure and high magnetic saturation are the reasons for the highly nonlinearity of SRM magnetization characteristics. This in turn makes it very difficult to acquire a trusted analytical model of SRM [12]-[15].

The highest torque/ampere ratio over the possible range of operating speeds is an essential approach for many applications [16]-[19]. The develop torque of SRMs can be optimized by machine design and/or applying appropriate control parameters [18]-[22]. The control parameters are reference current, turn-on (θ_{on}) angle and turn-off (θ_{off}) angle. As the reference current is determined by outer loop controller, switching-angles (θ_{on} , θ_{off}) are the main control parameters for SRM torque optimization [18], [23]. For single-phase excitation, the conduction angle (θ_c) is set to a constant value. Hence, the turn-off angle can be calculated directly as $\theta_{off} = \theta_{on} + \theta_c$. Therefore, the turn-on (θ_{on}) angle is the dominant parameter for maximum torque production of SRM drives.

Many researches are interested in the optimum θ_{on} angle that can provide maximum torque production with minimum copper losses. As modeling of SRM is a very difficult task, these researches depend mainly on the analysis and observation of static magnetization characteristics of SRMs. In [24], the conventional approach for optimum turn-on angle is introduced. It assumes a linear inductance profile. It can provide acceptable results till base speed. In [25], the conventional approach is used to obtain an initial value for θ_{on} . Then, within a certain range around the obtained value, an experiment is employed to find the most efficient angles. This method is a time consuming and requires accurate measurements. In [26], a closed loop θ_{on} control is designed to force the first-peak of phase current to occur at the end of minimum inductance region. This method can be used over wide speed range as it uses a closed loop control, but it is much complicated and requires two sub-techniques to monitor first peak of phase current and its position. In [27], θ_{on} is tuned continuously under steady state in order to minimize the total power consumption. In this method, the control strategy is affected by energizing switching angle and requires a complicated process to shorten its searching time. Analytical solutions are used for turn-on angle optimization [18], [28]. They mainly introduce the turn-on angle as a function of multiple variables. These variables can be calculated by the curve-fitting of phase inductance over minimum inductance zone. The turn-on (θ_{on}) angle has also been optimized using field reconstruction method and fuzzy controllers [29]-[31].

In this paper, a multi-objective optimization of SRM control parameters (θ_{on} , θ_{off}) is achieved. The aim of optimization process is to obtain the maximum average torque with the minimum copper losses. Because of the highly nonlinear magnetization characteristics of SRM, the objective function is calculated using a built Simulink model of the tested 8/6 SRM. This model is built based on the experimental measurement of SRM magnetization characteristics. A searching algorithm is used to calculate the base values of objective function as they vary for each operating point. The optimum control parameters are defined for each

operating point. Then, the obtained data are used to train a feed forward artificial neural network (ANN) in order to implement the control algorithm.

The paper is organized as follows: Section II obtains the problem description and basic control principles of SRM. The machine modeling and its performance indices are given in Section III. Section IV involves the optimization problem and the implementation of control algorithm. In addition, Sections V and VI contain the simulation and experimental results respectively. Finally, Section VII covers the conclusions.

2 Problem Description

Figure 1 shows a linear inductance profile and the optimal current waveforms for low and high-speed operation of SRM. The motor coils must be excited in the increasing inductance zone ($dL/d\theta > 0$), and de-energized before negative inductance zone ($dL/d\theta < 0$) to avoid negative torque production [32], [33]. Considering the motor phase inductance and its continuous commutation process, the phase current requires an amount of time to rise/fall. The amount of time depends on motor speed, current magnitude, turn-on instant, and turn-off. At low speeds, the motor current can rise and decay quickly enough to reach its commanded reference level. Therefore, θ_{on} can be delayed to be close to the starting point of increasing inductance zone (θ_m) as illustrated in Figure 1(a). On the contrary, for high speeds, the phase current can't rise or decay quickly enough to reach its reference commanded level. For that reason, the motor phase winding is turned-on early in order to allow phase current to reach its commanded level as shown in Figure 1(b) [18], [28].

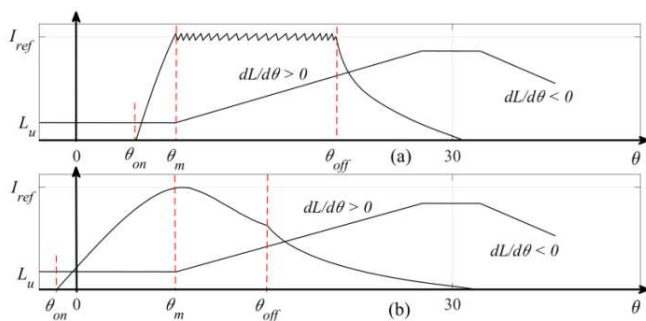


Figure 1

Ideal inductance profile and current waveforms at (a) low speed (b) high speed

Regarding the observation and analysis of static magnetization characteristics of SRM, for a given current, the maximum torque occurs as rotor begins to move out from minimum inductance zone [24]-[28]. Thus, the maximum value of torque/ampere occurs at position (θ_m) as shown in Figure 1(a, b) for low and high

speeds respectively. The main idea behind maximum torque production is to make phase current able to reach its commanded level at/before θ_m . The copper losses will be very high if motor coils are turned-on too advanced. Therefore, in order to optimize motor efficiency, θ_{on} can be calculated backward from θ_m .

It is concluded that the optimum θ_{on} under analysis and observation of SRM static characteristics should satisfy two conditions. First, it should allow motor phase current to reach its reference value. Second, it should force the first-peak of phase current to occur at angle θ_m [26], [28].

The conventional approach assumes linear inductance profile and calculates θ_{on} as follows [24]:

$$\theta_{on} = \theta_m - \left(\frac{L_{min} \cdot I_{ref} \cdot \omega}{V_{dc}} \right) \quad (1)$$

where L_{min} is the minimum inductance. V_{dc} is the supply voltage. ω is the speed of rotor. I_{ref} is the reference current. Equation (1) assumes constant inductance over region $[-\theta_m, \theta_m]$. This approach can give reasonable performance under low speeds (up to base speed) unless θ_{on} becomes less than $-\theta_m$. For speed higher than rated speed, equation (1) starts to break down because of the dominant effect of back-emf voltage.

In order to consider the effect of back-emf voltage, the inductance profile is analyzed and fitted accurately over the minimum inductance zone [18], [19]. After that the optimum θ_{on} is calculated as follows [18]:

$$\theta_{on} = \theta_m + \frac{\omega \cdot L_u}{R + \omega \cdot k_b} \ln \left[1 - I_{ref} \left(\frac{R + \omega \cdot k_b}{V_{dc}} \right) \right] \quad (2)$$

where L_u is the inductance, $k_b = dL/d\theta$ is the inductance derivative according to rotor position θ , and R is the phase resistance.

Equation (2) gives the optimum θ_{on} , but it represents θ_{on} as a function of multiple variables. The accurate determination of optimum θ_{on} depends on the accurate calculation of these variables [18].

A closed loop turn-on angle controller (CL- θ_{on}) is discussed in details through [26]. Its structure is shown in Figure 2. It compares the peak value of phase current (I_{peak}) to its reference level (I_{ref}), and angle of first-peak position (θ_{peak}) to θ_m . The error signal is processed using a PID controller whose output compensates the conventional approach. This controller forces first-peak of phase current (θ_{peak}) to occur at θ_m and also allows phase current to reach its reference level.

Figure 3 shows a comparison between static and dynamic/actual torque curves at different operating speeds for a typical SRM. The difference between torque curves is very clear. It increases as the motor speed increases.

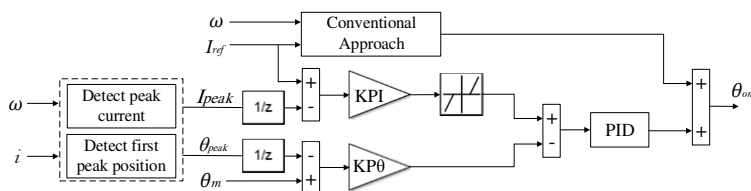


Figure 2

The structure of closed loop turn-on angle (CL- θ_{on}) controller

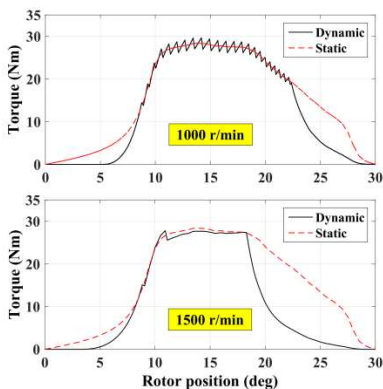


Figure 3

The static and dynamic/actual phase torques at different speeds

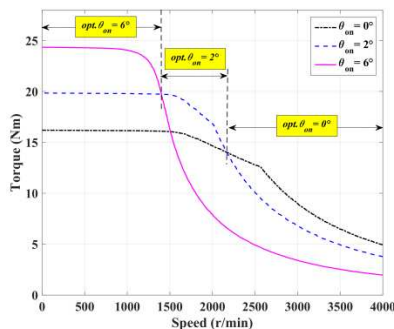


Figure 4

The torque-speed curves with different turn-on angles

Moreover, Figure 4 illustrates the effect of turn-on (θ_{on}) angles variation on SRM torque-speed characteristics. As noted, the developed torque is affected greatly by turn-on angle variation. From zero up to 1400 r/min, the maximum torque production is achieved with $\theta_{on}=6^\circ$, and from 1400 r/min up to 2200 r/min, the maximum produced torque occurs with $\theta_{on}=2^\circ$. Finally, for speeds higher than 2200 r/min, the maximum torque production is obtained with $\theta_{on}=0^\circ$. For these reasons, the observation and analysis of static torque curves is not enough to provide the absolute maximum torque production for SRM drives. The control parameters should be optimized using the accurate dynamic/actual torque-speed curves instead of static torque curves. This in turn requires a trusted simulation model that accurately involves the highly nonlinear characteristics of SRM.

3 Accurate Machine Modeling

Accurate modeling of SRM requires accurate determination of its magnetization characteristics. These characteristics can be calculated by a magnetic equivalent circuit (MEC), finite element method (FEM), and indirect measurements [34]–

[36]. MEC has a very complicated calculation process. The accuracy depends mainly on assumptions. FEM can offer higher accuracy than MEC. But the accuracy depends on accurate SRM dimensions and steel properties that may not be easy to obtain [36]. FEM doesn't consider effects of end-winding. Therefore, the indirect measurement is preferred. It can include the manufacturing processes' introduced imperfections. In addition, the measured data contains the physical effects. Hence, the indirect measurements are employed to estimate the flux linkage $\lambda(i, \theta)$ in every stator pole, phase inductance $L(i, \theta)$ and developed torque $T(i, \theta)$.

Figure 5(a-c) shows the measured torque, flux linkage, and inductance characteristics for the tested 8/6 SRM respectively. As seen, the characteristics are highly nonlinear functions of current (i) and position (θ). The unaligned and aligned positions are defined by $\theta=0^\circ$ and $\theta=30^\circ$ respectively. The dimensional parameters of SRM are given in Table 1. Once accurate magnetization characteristics are obtained, they can be stored in form of lookup tables $\lambda(i, \theta)$ and $T(i, \theta)$ or trained using ANNs [35], [36]. These characteristics are used directly to build a highly reliable/trusted MATLAB simulation model. Simulation of one phase of SRM is shown in Figure 5(d) [37]. The full details about the measuring process and error minimization methods are obtained in previous work [36].

The performance indices for SRM are calculated within the simulation model. The total electromagnetic torque (T_e) is the summation of phases' torques. Its average value (T) can be calculated over one electric cycle (τ) as follows [6], [22].

$$T = \frac{1}{\tau} \int_0^\tau T_e(t) dt \tag{3}$$

The mechanical output power (P_m) is calculated from motor speed (ω) as follows:

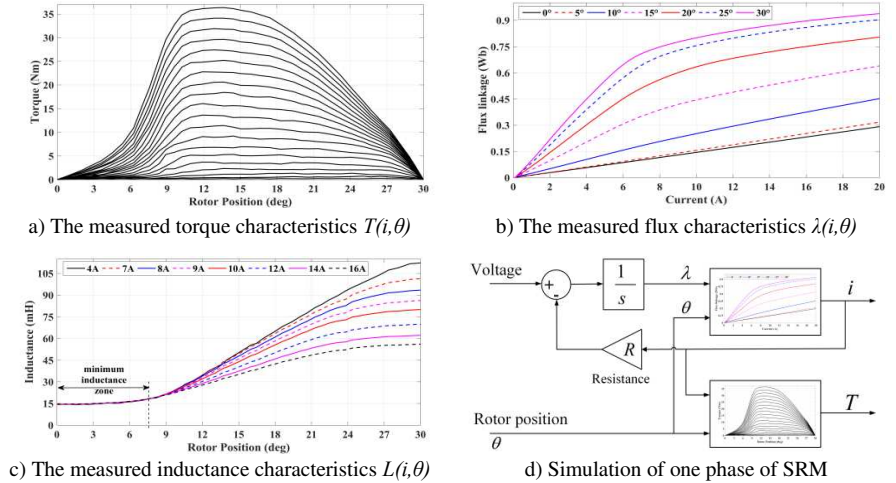


Figure 5
The simulation of SRM with its measured characteristics

$$P_m = \omega \cdot T \quad (4)$$

The supply current (i_s) is a periodical waveform. Its average (I_s) can be calculated from as follows:

$$I_s = \frac{1}{\tau} \int_0^{\tau} i_s(t) \cdot dt \quad (5)$$

The copper losses (P_{cu}) in motor windings are calculated as:

$$P_{cu} = m \cdot I^2 R \quad (6)$$

where m is the phases number, I is the RMS phase current, and R is the resistance.

The RMS value of phase current can be calculated as follows:

$$I = \sqrt{\frac{1}{\tau} \int_0^{\tau} i_{ph}^2(t) \cdot dt} \quad (7)$$

Table 1

The design data of 8/6 SRM in mm

Geometry parameter	Value	Geometry parameter	Value
Output power	4 kW	Stator outside diameter	179.5
Rated speed	1500 rpm	Shaft/Bore diameters	36/96.7
Phase resistance	0.642 Ω	Rotor/stator pole arc	21.5°/20.45°
Air-gap length	0.4	Stack length	151
Height of rotor/stator pole	18.1/29.3	Turns per pole	88

4 The Optimization Problem

The optimization of SRM control parameters aims to achieve the highest average torque with the lowest copper losses. But, it is impossible to obtain the highest average torque with the lowest copper losses simultaneously, as different control parameters (θ_{on} , θ_{off}) are required for each case. Therefore, a two-group multi-objective optimization function is used to attain the desired adjustment between average torque and copper loss.

4.1 Problem Formulation

A single objective optimization problem is obtained from the multi-objective problem by linear combination of average torque and copper losses as follows:

$$F_{obj}(\theta_{on}, \theta_{off}) = \min \left(w_T \frac{T_b}{T} + w_{cu} \frac{P_{cu}}{P_{cub}} \right) \quad (8)$$

$$w_T + w_{cu} = 1 \quad (9)$$

Subject to:

$$\theta_{on}^{\min} \leq \theta_{on} \leq \theta_{on}^{\max} \quad (10)$$

$$\theta_{off} = \theta_{on} + \theta_c \quad (11)$$

where F_{obj} is the objective function, T is the average torque, and P_{cu} is the copper loss. T_b is the base value of average torque. P_{cub} is the base value of copper loss. w_T is the average torque weight factor. w_{cu} is the copper loss weight factor. The control variables are θ_{on} and θ_{off} . For 8/6 SRM, the conduction-angle $\theta_c=15^\circ$.

4.2 Solution Method

The well-known optimization-techniques such as evolutionary-algorithms, genetic-algorithm (GA), particle swarm optimization (PSO) are hardly employed for such a problem because the base values (T_b, P_{cub}) have different values for each operating point [38]-[41]. For that reason, a searching algorithm is developed to calculate the base values and hence the optimum control parameters at each operating point. The flowchart of searching algorithm is shown in Figure 6. For each operating point, a step changing in turn-on angle (θ_{on}) is made. Then, for each step, the average torque and copper loss are calculated within the simulation model. At the end of search, the maximum average torque and the minimum copper losses are defined as the base values (T_b, P_{cub}). The turn-on angle (θ_{on}) is varied from $\theta_{on}^{\min} = -10^\circ$ to $\theta_{on}^{\max} = 10^\circ$ in steps of 0.2° while the current step is taken as 1A. The smaller the variation steps, the better the accuracy.

The weight factors (w_T and w_{cu}) are chosen according to the desired level of optimization. In this paper, greater importance is directed to improve average torque production than to minimize copper losses because the motor has a very small resistance. The weighting factor of average torque is taken as $w_T = 0.95$ while weight factor of copper loss is set to $w_{cu} = 0.05$. For a different level of optimization, different weight factors can be chosen.

Solving equation (8) gives the optimum parameters that fulfill the highest average torque and the lowest copper loss at each of the operation points. The optimized turn-on angles obtained from (8) are given in Figure 7 for different weight factors. For a given motor speed, the turn-on angle is decreased as the reference current increases.

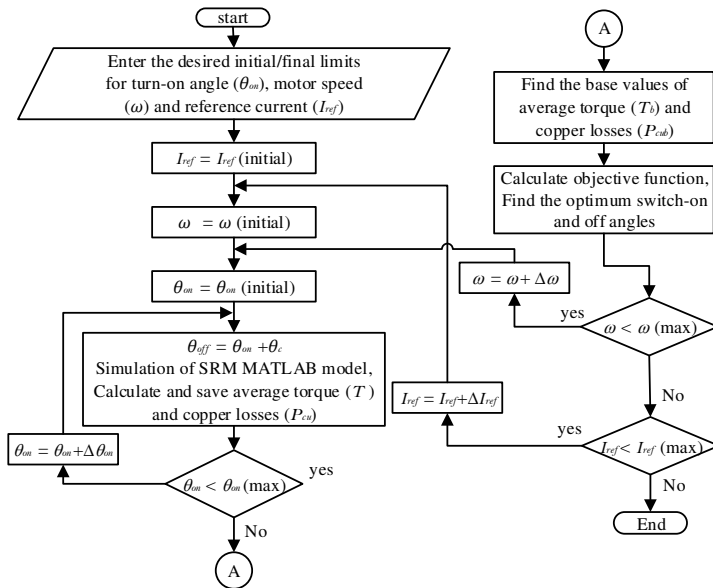


Figure 6

The flowchart of searching algorithm

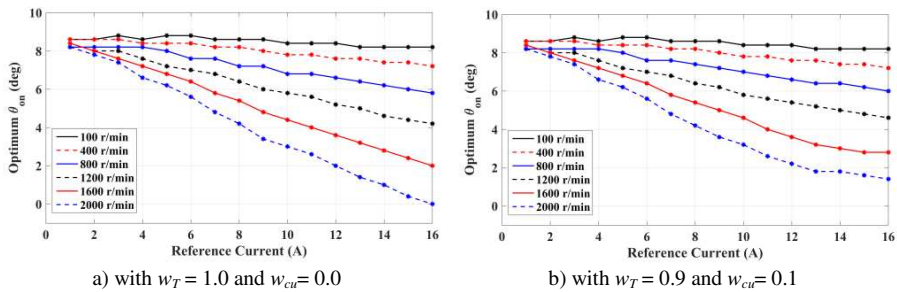


Figure 7

The optimum turn-on angles with different weight factors

The optimization process provides the optimum turn-on angles, as a function of motor speed and reference current $\theta_{on}(w, I_{ref})$, as illustrated in Figure 7. These data are implemented within the control algorithm using a feed forward artificial neural network (ANN). Figure 8 shows the architecture details of trained ANN. The ANN is trained using MATLAB function “*nntool*”. The ANN has two inputs (w, I_{ref}) and one output (θ_{on}). The ANN uses Levenberg-Marquardt technique for training with 10 neurons in the hidden layer. Figure 9(a) shows the linear regression performance with R-value over 0.999 for the total response. Figure 9(b) shows the training performance of ANN. It shows a small means square error. Therefore, the network can work in an efficient way.

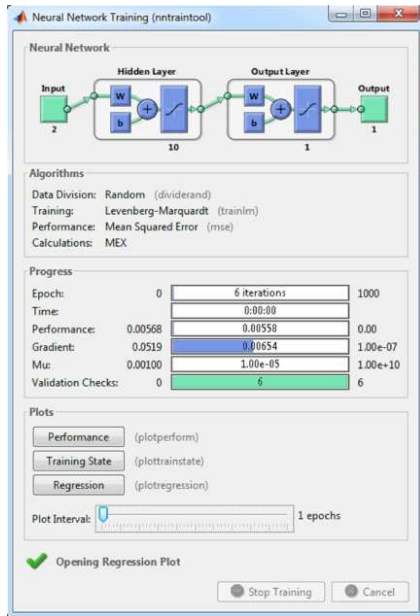


Figure 8
The architecture of ANN

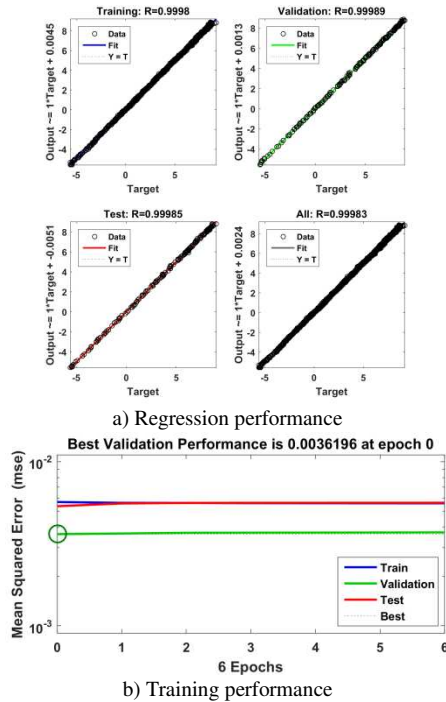


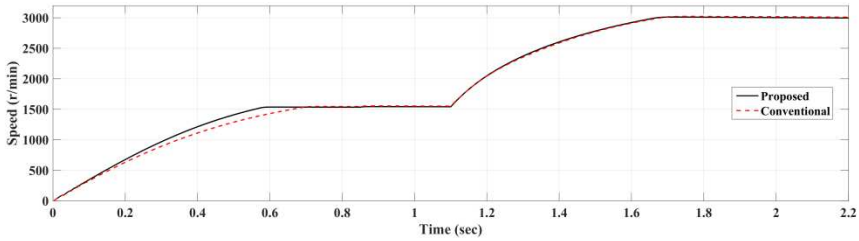
Figure 9
The ANN performance

5 Simulation Results and Discussion

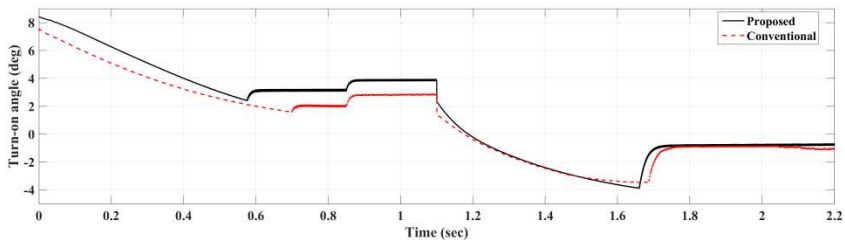
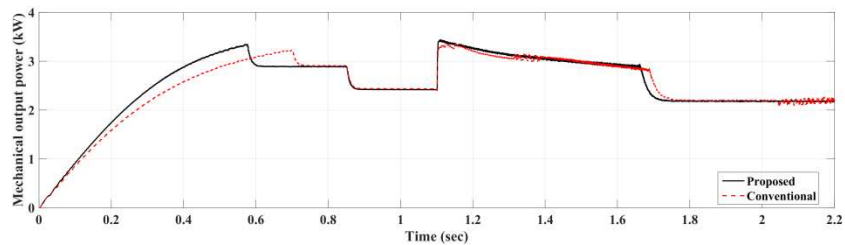
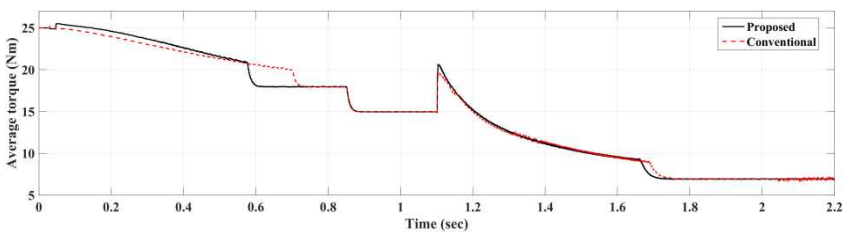
In order to show the effectiveness and feasibility of proposed controller, a closed-loop turn-on angle ($CL-\theta_{on}$) controller is used for the purpose of comparison. The $CL-\theta_{on}$ represents the conventional methods for optimum solution of θ_{on} . Due to its closed loop control, it forces the first-peak of phase-current to occur always at angle θ_m . The controller parameters are $KP\theta=0.5$, $KPI=0.2$ °/A, and PID gains are ($KP=0.5$, $KI=15$, and $KD=-0.006$).

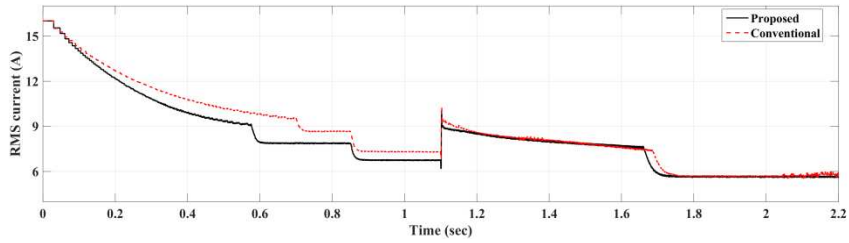
Figure 10 shows the simulation results under sudden-change of commanded speed and loading torque. The reference speed is suddenly changed from 1500 r/min to 3000 r/min at 1.1 sec. The motor started under load torque of 18 Nm. Then, the load torque is changed suddenly from 18 Nm to 15 Nm at 0.85 sec, and from 15 Nm to 7 Nm at 1.1 sec. Figure 10(a) shows the motor speed. As noted, the proposed control can achieve higher dynamic performance compared to conventional controller. It allows motor to reach its reference speed faster. The online variation of θ_{on} with motor speed and load torque/current is illustrated in

Figure 10(b). A fast and adaptive changing of θ_{on} is achieved as the proposed controller uses ANN to implement control algorithm. The mechanical output power and average torque are given in Figure 10(c) and Figure 10(d) respectively. The proposed controller can provide higher average torque and mechanical output power over the entire speed range. Furthermore, it consumes lower supply current and dissipates lower copper losses as shown in Figure 10(e) and Figure 10(f) respectively.

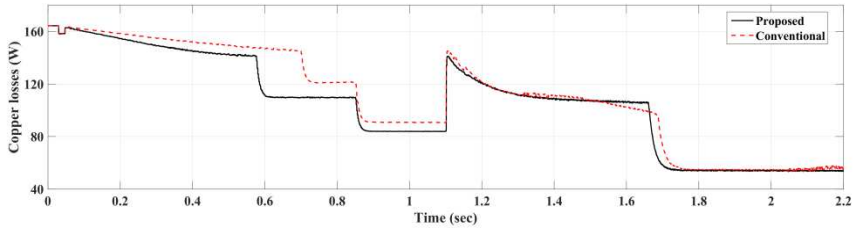


a) The motor speed

b) The turn-on angle (θ_{on})c) The mechanical output power (P_m)d) The average torque (T)



e) The supply current (I_s)

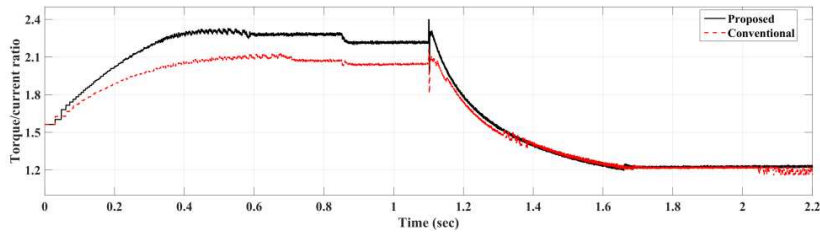


f) The copper losses (P_{cu})

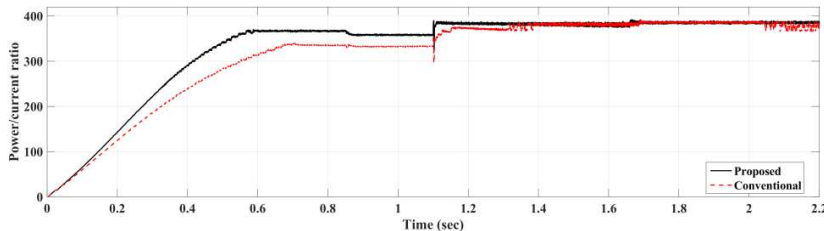
Figure 10

The simulation results with sudden change of reference speed and load torque

Figure 11(a, b) shows the torque/current and power/current ratios respectively. The proposed control provides the highest ratios over the operating range of speeds. For high speeds (over 1500 r/min), the power/current ratio is constant as illustrated in Figure 11(b) after 1.1 sec. This means that all the current flowing in motor windings is an effective current that produces mechanical output power.



a) The torque/current ratio



b) The power/current ratio

Figure 11

The torque/current and power/current ratios

Figure 12 shows the position of first-peak of phase current and peak phase current value. Under the analysis of static torque curves, the conventional controller gives optimum solution for θ_{on} by forcing the first-peak of phase current to occur at angle $\theta_m = 7.5^\circ$ as shown in Figure 12(a). On the other hand, with the proposed controller, the position of first-peak of phase current does not have a constant value. It varies with motor speed and load torque especially at low speeds. As seen, it has a noticeable difference from $\theta_m = 7.5^\circ$ for speeds lower than base speed (1500 r/min) and tends to be very close to $\theta_m = 7.5^\circ$ for speeds higher than rated speed. Furthermore, the proposed controller allows phase current to always reach its reference level (16A) as shown in Figure 12(b).

The current waveforms at different speeds are shown in Figure 13. For both controllers, the position of first-peak of phase current is obvious to have a clear difference under low speeds of 1500 r/min, and have a small difference for higher speed of 3000 r/min. It can be noted that, for the proposed controller, the RMS phase current has a lower value compared to conventional controller especially at lower speeds.

The steady state torque-speed curve is shown in Figure 14. The proposed controller provides higher torque production with lower current consumption. It has approximately **4%** improvement in torque production for speeds up to base speed (1500 r/min).

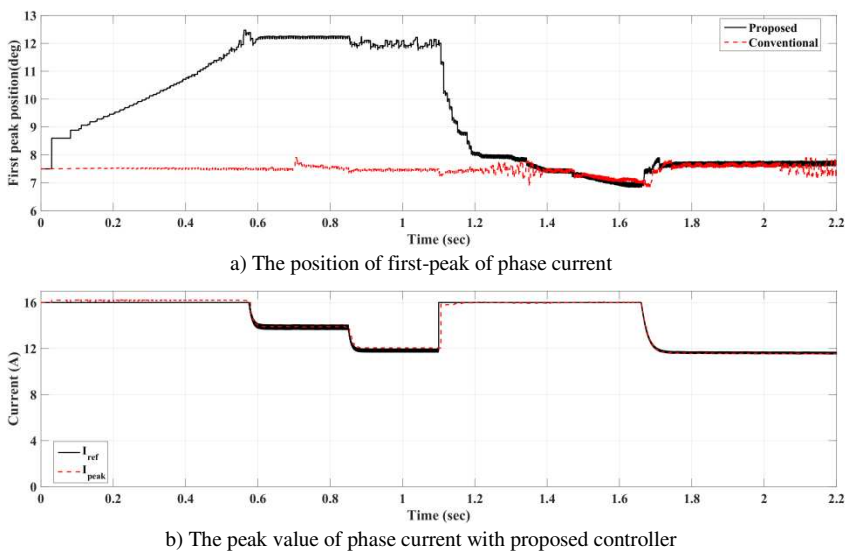


Figure 12

The simulation results under sudden change of speed and load torque

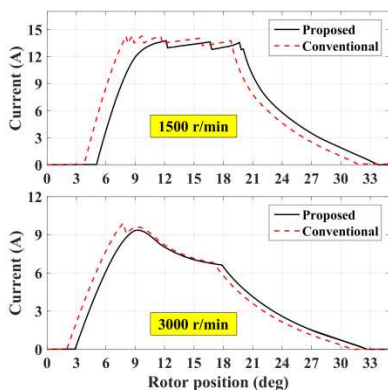


Figure 13

The phase current waveform at different speeds

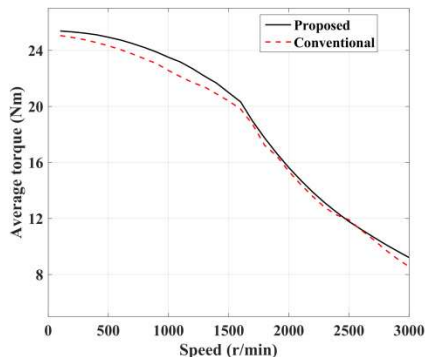
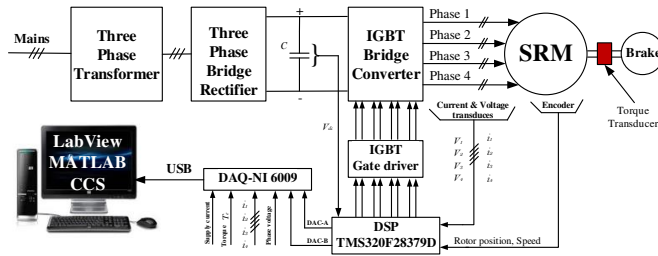


Figure 14

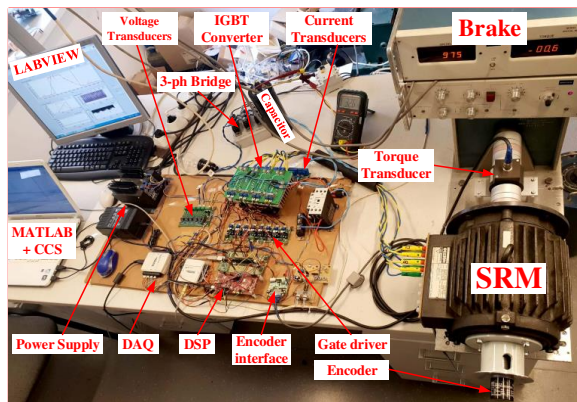
The steady-state torque-speed curve

6 Experimental Verification

The controller performance is experimentally verified with four-phase 8/6 SRM. The SRM specifications are given in the Table 1. The control algorithm is implemented using a Texas Instruments TMS320F28379D digital signal processor (DSP). The SRM is coupled to an electromagnetic brake (MAGTORL model 4605c), which acts as a mechanical load. The shaft torque is measured using a DRBK torque transducer. An incremental-encoder (600 PPR) is used to provide rotor position. The phase currents are measured using high-accuracy and linearity current transducers (LAH50-P). The voltage measurement is achieved using a high-speed and linearity op-amp based circuit. A three-phase transformer, three-phase diode rectifier, and capacitor are used to provide DC power. The data are collected and plotted using a data acquisition board (DAQ NI USB-6009) with LabView software. C2000 microcontroller support package and code composer studio (CCS) are used for DSP programming. The schematic diagram and the practical implementation of measurement platform are shown in Figure 15(a, b) respectively.



a) The schematic diagram of measurement platform

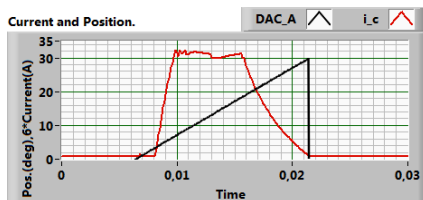


b) The practical implementation of measurement platform

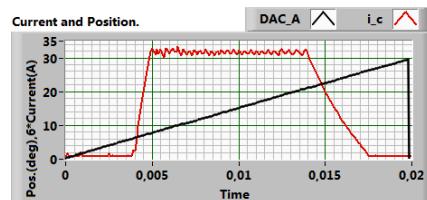
Figure 15

The experimental test bench

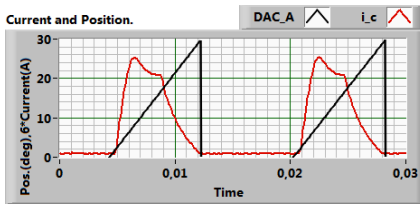
Figure 16 and 17 show the waveforms of phase current and its phase position with the conventional and proposed controllers respectively. It is obvious that the conventional controller forces the first-peak of phase current to occur always at angle $\theta_m = 7.5^\circ$ as shown in Figure 16(a, b) for low and high speeds respectively. On the other hand, with the proposed controller, the first-peak of phase current occurs at different positions. This is clear in Figure 17(b) as the first-peak of phase current occurs at 12° . In addition, the current waveform has a very similar shape compared to obtained simulation results in Figure 13.



a) At speed of 335 r/min



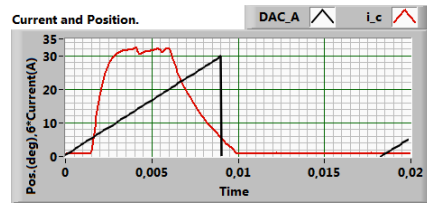
a) At speed of 251 r/min



b) At speed of 621 r/min

Figure 16

The experimental waveforms of phase current and position with the conventional controller

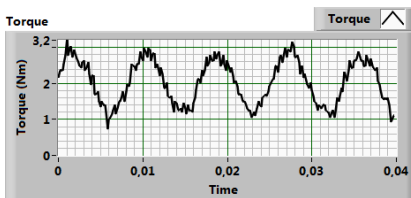


b) At speed of 545 r/min

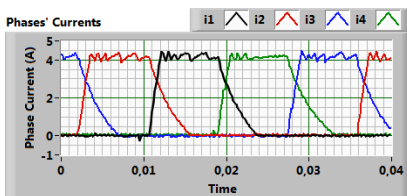
Figure 17

The experimental waveforms of phase current and position with the proposed controller

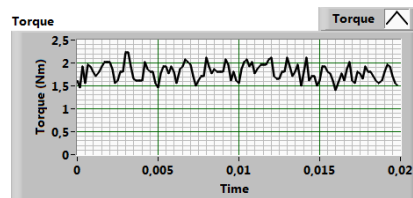
Figure 18 and 19 show the waveforms of total electromagnetic torque, phases' currents and supply current at low speed of 294 r/min and high speed of 806 r/min respectively. The applied supply voltage is 100 V. The inherited drawback of torque ripple for SRM is very clear especially at low speeds as shown in Figure 18(a). The average torque is calculated from the instantaneous torque signal. The motor phases are energized in a certain sequence as illustrated by phases' currents in Figure 18(b). The supply current is given in Figure 18(c). A noticed part of this current is regenerated back to supply because of the chopping process. At higher speed, a single pulse control is employed and phase current becomes much smoother without chopping as shown in Figure 19(b). The supply current becomes almost positive without the negative regenerated part as illustrated in Figure 19(c).



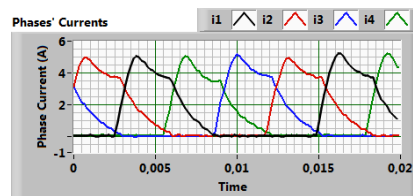
a) The total electromagnetic torque



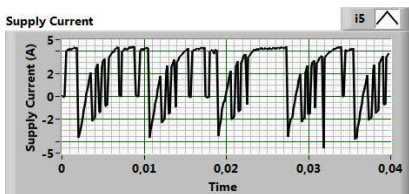
b) The waveform of phases' currents



a) The total electromagnetic torque



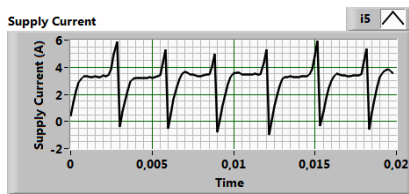
b) The waveform of phases' currents



c) The waveform of supply current

Figure 18

The experimental results at speed of 294 r/min,
 2.05 Nm, $\theta_{on} = 4^\circ$, $\theta_{off} = 19^\circ$

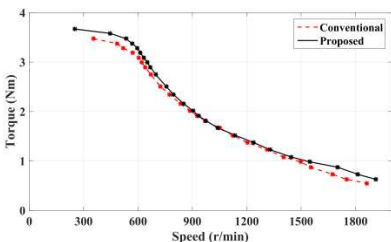


c) The waveform of supply current

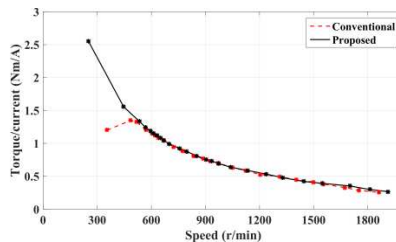
Figure 19

The experimental results at speed of 806 r/min,
 1.8 Nm, $\theta_{on} = 2^\circ$, $\theta_{off} = 19^\circ$

Figure 20(a, b) shows the measured steady state torque-speed curve and torque/current ratio respectively. For the proposed controller, the average torque is improved very clearly especially at low speeds as shown in Figure 20(a). It has a good agreement with the simulation results obtained in Figure 14. The speed difference comes from different voltages as simulation is carried out with 400 V and practical measurements are taken with 100 V. For speeds lower than 600 r/min, the improvement of average torque production is about **5.3%**. With the proposed control, the torque/current ratio is higher all over the speed range especially for lower speeds. As the speed increases the current control becomes very difficult, that is why the torque improvement is very clear for low speed than high speeds.



a) The experimental torque-speed curves



b) The experimental torque/current ratio

Figure 20

The experimental steady state results with $I_{ref}=5A$

Conclusions

This paper presented an optimization based method for SRM control parameters to improve torque production. The proposed control calculates the most efficient θ_{on} according to torque production and copper losses. Instead of the conventional analysis of static torque curves, a trust dynamic machine model is used to ensure the absolute calculation of optimum θ_{on} over the entire range of operating speeds. This model is built using measured data of SRM magnetization characteristics. The objective function is calculated within the Simulink model, using a two-

dimensional search algorithm. The obtained data are implemented using ANN. The proposed controller offers low cost and simple implementation. It provides optimum motor operation over the entire speed range. It has a faster dynamic response. It can provide the highest torque/current and power/current ratios. It consumes lower supply current and dissipates lower copper losses. It does not depend on motor parameters and improves torque production capability by **5.3%** for speeds lower than the rated motor speed.

References

- [1] Z. Yueying, Y. Chuantian, Y. Yuan, W. Weiyan, Z. Chengwen: Design and optimisation of an In-wheel switched reluctance motor for electric vehicles, *IET Intelligent Transport Systems*, 2019, Vol. 13, No. 1, pp. 175-182
- [2] J. J. Wang: Parameter optimization and speed control of switched reluctance motor based on evolutionary computation methods, *Swarm and Evolutionary Computation*, 2018, Vol. 39, pp. 86-98
- [3] Y. Hu, W. Ding, T. Wang, S. Li, S. Yang, Z. Yin: Investigation on a multimode switched reluctance motor: design, optimization, electromagnetic analysis, and experiment, *IEEE Transactions on Industrial Electronics*, 2017, Vol. 64, No. 12
- [4] Z. Zhang, S. Rao, X. Zhang: Performance prediction of switched reluctance motor using improved generalized regression neural networks for design optimization, *CES Transactions On Electrical Machines and Systems*, 2018, Vol. 2, No. 4, pp. 371-376
- [5] J. Zhu , K. W. E. Cheng, X. Xue: Design and analysis of a new enhanced torque hybrid switched reluctance motor, *IEEE Transactions on Energy Conversion*, 2018, Vol. 33, No. 4, pp. 1965-1977
- [6] H. Cheng, H. Chen, Z. Yang: Average torque control of switched reluctance machine drives for electric vehicles, *IET Electrical Power Applications*, 2015, Vol. 9, No. 7, pp. 459-468
- [7] J. B. Bartolo, M. Degano, J. Espina, C. Gerada: Design and initial testing of a high-speed 45-kw switched reluctance drive for aerospace application, *IEEE Transactions on Industrial Electronics*, 2017, Vol. 64, No. 2, pp. 988-997
- [8] G. K. Ptakh, A. P. Temirev, D. A. Zvezdunov, A. A. Tsvetkov: Experience of developing and prospects of application of switched reluctance drives in Russian Navy Fleet, *IEEE Conference and Expo Transportation Electrification Asia-Pacific (ITEC Asia-Pacific) 2014*, pp. 1-4
- [9] T. A. S. Barros, E. Ruppert: Direct power control for switched reluctance generator in wind energy, *IEEE Latin America Transactions*, 2015, Vol. 13, No. 1, pp. 123-128

-
- [10] J. Kim, R. Krishnan: Novel two-switch-based switched reluctance motor drive for low-cost high-volume applications, *IEEE Transactions on Industry Applications*, 2009, Vol. 45, No. 4, pp. 1241-1248
- [11] H. Chen, J. J. Gu: Switched reluctance motor drive with external rotor for fan in air conditioner,” *IEEE/ASME Transactions on Mechatronics*, 2013, Vol. 18, No. 5, pp. 1448-1458
- [12] S. S. Ahmad, G. Narayanan: Linearized modeling of switched reluctance motor for closed-loop current control, *IEEE Transactions on Industry Applications*, 2016, Vol. 52, No. 4, pp. 3146-3158
- [13] W. Uddin, Y. Sozer: Analytical modeling of mutually coupled switched reluctance machines under saturation based on design geometry, *IEEE Transactions on Industry Applications*, 2017, Vol. 53, No. 5, pp. 4431-4440
- [14] C. Li, G. Wang, J. Liu, Y. Li, Yu. Fan: A Novel method for modeling the electromagnetic characteristics of switched reluctance motors, *Applied Sciences*, 2018, Vol. 8, No. 537, pp. 1-14
- [15] D. S. Mihic, M. V. Terzic, S. N. Vukosavic: A New nonlinear analytical model of the SRM with included multiphase coupling, *IEEE Transactions on Energy Conversion*, 2017, Vol. 32, No. 4, pp. 1322-1334
- [16] A. Argeşeanu, E. Ritchie, K. Leban: Torque optimization algorithm for SRM drives using a robust predictive strategy, *IEEE 12th International Conference on Optimization of Electrical and Electronic Equipment (OPTIM)* 2010, pp. 252-257
- [17] X. Wang, Z. Yang, T. Wang, D. He, Y. Huo, H. Cheng, G. Yu: Design of a wide speed range control strategy of switched reluctance motor for electric vehicles, *IEEE International Conference on Information and Automation*, 2015, pp. 294-299
- [18] M. Hamouda, L. Számel: A new technique for optimum excitation of switched reluctance motor drives over a wide speed range, *Turkish Journal of Electrical Engineering and Computer Sciences*, 2018, Vol. 26, No. 5, pp. 2753-2767
- [19] M. Hamouda, L. Számel: Optimum excitation angles for switched reluctance motor drives, *XXXIII. Kando Conference*, 2017, pp. 128-142
- [20] S. R. Mousavi-Aghdam, M. R. Feyzi, N. Bianchi, M. Morandin: Design and analysis of a novel high-torque stator-segmented SRM, *IEEE Transactions on Industrial Electronics*, 2016, Vol. 63, No. 3, pp. 1458-1466
- [21] S. Nakano, K. Kiyota, A. Chiba: Design consideration of high torque-density switched reluctance motor for hybrid electrical vehicle, *IEEE 19th International Conference on Electrical Machines and Systems (ICEMS)* 2016

-
- [22] H. Chen, W. Yan, J. J. Gu, M. Sun: Multiobjective optimization design of a switched reluctance motor for low-speed electric vehicles with a taguchi–CSO algorithm, *IEEE/ASME Transactions on Mechatronics*, 2018, Vol. 23, No. 4, pp. 1762-1774
- [23] B. Anvari, H. A. Toliyat, B. Fahimi: Simultaneous optimization of geometry and firing angles for in-wheel switched reluctance motor drive, *IEEE Transactions on Transportation Electrification*, 2018, Vol. 4, No. 1, pp. 322-329
- [24] B. K. Bose, T. J. E. Miller, W. H. Bicknell, P. M. Szczesny, W. H. Bicknell: *Microcomputer Control of Switched Reluctance Motor*, *IEEE Transaction on Industrial Applications*, 1986, Vol. IA-22, No. 4, pp. 708-715
- [25] C. Mademlis, I. Kioskeridis: Performance optimization in switched reluctance motor drives with online commutation angle control, *IEEE Transactions on Energy Conversion*, 2003, Vol. 18, No. 3, pp. 448-457
- [26] Y. Sozer, D. A. Torrey, E. Mese: Automatic control of excitation parameters for switched-reluctance motor drives,” *IEEE Transactions on Power Electronics*, 2003, Vol. 18, No. 2, pp. 594-603
- [27] S. C. Wang, W. H. Lan: Turn-on angle searching strategy for optimized efficiency drive of switched reluctance motors, 30th Annual Conference of IEEE Industrial Electronics Society (IECON) 2004, Vol. 2, pp. 1873-1878
- [28] Y. Z. Xu, R. Zhong, L. Chen, S. L. Lu: Analytical method to optimise turn-on angle and turn-off angle for switched reluctance motor drives, *IET Electric Power Applications*, 2012, Vol. 6, No.9, pp. 593-603
- [29] C. Lin, B. Fahimi: Optimization of commutation angles in SRM drives using FRM, *IEEE Transportation Electrification Conference and Expo (ITEC) 2012*, pp. 1-6
- [30] H. Chen, X. Wang, X. Zan, X. Meng: Variable angle control for switched reluctance motor drive based on fuzzy logic, the fifth international conference on power electronics and drive systems (PEDS) 2003, Vol. 2, pp. 964-968
- [31] H. M. Cheshmehbeigi, S. Yari, A. R. Yari, E. Afjei: Self-tuning approach to optimization of excitation angles for switched- reluctance motor drives using fuzzy adaptive controller, 13th European Conference on Power Electronics and Applications, 2009, pp. 1-10
- [32] M. Hamouda, A. R. A. Amin, E. Gouda: A drive system design and implementation for switched reluctance motor based on wide range speed control, 17th International Middle East Power System Conference (MEPCON) 2015, pp. 1-8
-

-
- [33] M. Hamouda, A. R. A. Amin, and E. Gouda: Artificial intelligence based torque ripple minimization of switched reluctance motor drives, 18th International Middle East Power System Conference (MEPCON) 2016, pp. 943-948
- [34] S. Song, M. Zhang, L. Ge, "A new fast method for obtaining flux-linkage characteristics of SRM," IEEE Transaction on Industrial Electronics, 2015, Vol. 62, No. 7, pp. 4105-4117
- [35] R. Zhong, Y. Xu, Y. Cao, X. Guo, W. Hua, S. Xu, W. Sun: Accurate model of switched reluctance motor based on indirect measurement method and least square support vector machine, IET Electric Power Applications, 2016, Vol. 10, No. 9, pp. 916-922
- [36] M. Hamouda, L. Számel: Accurate measurement and verification of static magnetization characteristics for switched reluctance motors, IEEE 19th International Middle East Power System Conference (MEPCON) 2017, pp. 993-998
- [37] M. Hamouda, L. Számel: Torque control of switched reluctance motor drives for electric vehicles, Proceedings of the Automation and Applied Computer Science Workshop, 2017, pp. 9-20
- [38] V. Oduguwa, A. Tiwari, R. Roy: Evolutionary computing in manufacturing industry: an overview of recent applications, Applied Soft Computing, 2005, Vol. 5, No. 3, pp. 281-299
- [39] A. Ürmös, Z. Farkas, M. Farkas, T. Sándor, L. T. Kóczy, Á. Nemcsics: Application of self-organizing maps for technological support of droplet epitaxy, Acta Polytechnica Hungarica, 2017, Vol. 14, No. 4, pp. 207-224
- [40] R. E. Precup, S. Preitl, P. Korondi: Fuzzy controllers with maximum sensitivity for servo systems, IEEE Transactions on Industrial Electronics, 2007, Vol. 54, No. 3, pp. 1298-1310
- [41] M. Shams, E. Rashedi, S. M. Dashti, A. Hakimi: Ideal gas optimization algorithm, International Journal of Artificial Intelligence, 2017, Vol. 15, No. 2, pp. 116-130

Parameter Estimation Method for the Unstable Time Delay Process

Radmila Gerov, Zoran Jovanović

University of Nis, Faculty of Electronic Engineering, Aleksandra Medvedeva 14,
18000 Nis, Serbia; gerov@ptt.rs, zoran.jovanovic@elfak.ni.ac.rs

Abstract: The paper considers the evaluation of all the three parameters of the unstable first order process with time delay by using the process data received from the closed loop step response under proportional control. The new method of analysis of parameters identification is presented. One is required to read five parameters from the closed loop step response for the purpose of applying the method. For the selected proportional controller gain and the received process gain, the time constant and time delay of the unstable first-order plus time delay model is received by solving a characteristic system equation using the features of the Lambert W function. The suggested way of parameter estimation is simple and it yields better results than the well-documented methods in literature which the present method is compared with. Simulation results are given for linear system and a nonlinear bioreactor system.

Keywords: Delay system; Model reduction; System identification; Parameters estimations; Unstable system; Nonlinear bioreactor

1 Introduction

System Identification, the common label for all the techniques for receiving the mathematical model of a dynamic system, was developed, based on the observed data in the field of system control (Zadeh 1956). Depending on how one or more input signals affects the behavior of the system, over time, the mathematical models of dynamic systems can be classified in different ways: linear, nonlinear, time continuous, time discrete, parametric and non-parametric, deterministic, stochastic... In literature there are different identification techniques for obtaining them [1].

Although Mathematical statistics is the most present in the process of system model identification, new ideas from other scientific communities have made a significant contribution to the development of new theories and algorithms necessary for the process of system identification [2]. For example, methods that have been developed for the identification and control of nonlinear dynamical

system [3]-[4] (neural networks), neuro-fuzzy state-space model [5] obtained from experimental data acquired from a real robotic arm, Kohonen's self-organizing maps, examinations of the classification of droplet epitaxial nanostructures [6] (machine learning), etc.

The synthesis for parameter identification, i.e. for the estimation of linear systems parameters is most frequently predicated on the Prediction Error Methods (PEM). As the model is considered to be adequate if the errors between the measured exits and their estimated values are sufficiently low, by applying this method the estimation task is treated as an optimization issue. It is well-known, that there are online and offline algorithms, for solving the optimization problem and parameters estimation, and that the transmission function of dynamic systems can be rational and irrational.

For online estimation of parameters of dynamic systems, which are described by using the rational transmission function, recursive algorithms are used such as Recursive Least Squares Method (RLS) and Kalman filtering (KF). Time delay systems (TDS) belong to the group of system which are described by using the irrational transmission function. Beside parameters, time delay needs to be estimated, which represents a new challenge for researchers [7]-[8]. For parameter estimation with an offline method, the optimization problem could be solved by applying e.g. Genetic Algorithm (GA) and Particle Swarm Optimization (PSO).

In the industry, system identification is used for obtaining models for the purpose of control, i.e. for regulation, synthesis, and realization of various controller type such as Model Predictive Controllers (MPC) [9], Linear Quadratic Regulators (LQR) [10], PID controllers [11], PI controllers [12], etc. For finding the model parameters, different methods may be used, which, as a result of the identification, due to the tendency of the mathematical model to satisfy all the dynamic characteristics of the observed process, can produce a high-order model. It is good to know that most of the methods for controller designing are based on low-order models such as the first-order plus time delay model (FOPTD), second-order plus time delay model (SOPTD), integral plus dead time model (IPDT), the unstable first-order plus time delay model (unstable FOPTD), or the unstable second-order plus time delay model (unstable SOPTD) [13]. Unlike the open loop stable processes control, the time delay processes control or the open loop unstable processes control, which are frequent for instance in chemical industry, are much more complex, which is why it is necessary to obtain, as a result of identification, a simple mathematical model that more accurately describes the dynamics of the process.

For model identification, the data obtained by step test which can be applied as an open- or close-loop structure are most regularly used. In addition to the closed-loop step test for model identification, a relay feedback test is widely used [14]. One of the first works with relay feedback test application is the method of estimation of critical gain and critical period for the purposes of automatic regulation of PID controllers [15].

In literature, different techniques of receiving the FOPTD and SOPTD models are known by means of analysis from the data of the step or frequency response, whereby the relay method is most frequently used. For example: in [16] the more precise estimation of FOPTD model parameters has been achieved by using the modified relay auto tune method, adding the equation which realizes more accurate process amplification, and by using a modified method of calculating model parameters at relay feedback method which accounts for higher-order harmonics of the obtained response; in [17], a modified relay feedback control under static load disturbances is given to identify parameters of the integrating plus first-order plus dead time model (IFOPDT) using exact expressions for a limit cycle to occur. The application of the final value theorem to suitably selected control loop signals, where the derivative action of PID controller is applied directly to the process variable instead of the control error, have been proposed for the estimation of the FOPTD model parameters, [18] [19] is a proposed method for estimate up to three frequency response points from a single biased relay test and getting parameters of SOPTD transfer function models.

In some industrial systems, open-loop dynamics may be unstable, so for safety reasons, closed-loop identification is applied. Unlike the identification of the open loop stable models, the identification of the parameters at both unstable SOPTD and the unstable FOPTD model represents a challenge. For a class of unstable systems, closed-loop test provides a reduced order model which can be used for controller design. The identification of the unstable model by using the relay feedback method is one of the methods that appears in literature. Using a single symmetric relay test all three parameters of unstable FOPTD model [20] have been identified, an asymmetrical relay feedback test is introduced along with sinusoidal signal for finding model parameters of various processes, among which is the unstable FOPTD model [21], method for identification of low order unstable process by using of relay with additional delay is given in [22], two identification algorithms using a single biased/unbiased relay feedback waveforms for the identification of unstable FOPTD model have been proposed in [23]. Identification method of parameters of an unstable FOPDT model when a limit cycle exists by using a single relay controllers is given in [24], two different techniques of process identification of unstable FOPTD model are analyzed in [25] by using the PID controller, while the improvements of the existing techniques of identification are [25] and [26], which solves the problems of stability and attains better time delay estimations, are given in [27].

It is known that, due to nonlinearity, there is a possibility that the system has multiple steady states, some of which can be unstable steady state [28]. For the purpose of designing a controller, these types of nonlinear systems are usually linearized and approximated by using the unstable FOPTD model. For example, by using the closed-loop identification methods in [20] [25] [26] [29], parameters for unstable FOPTD model of the nonlinear continuous bioreactor are received and compared with its linearized model.

In this study, the identification of parameters of the unstable FOPTD model by using the closed loop step response under proportional control method is being considered. The new estimation method for all the three parameters of the unstable FOPTD model which includes using five parameters from the recorded closed loop step response is presented. Time delay that renders the characteristic equation, transcendental, with an infinite number of solutions, is most often approximated by using the Pade approximation. This leads to an error during the calculation of the parameters of the unstable FOPTD model. The time delay approximation is not employed with the suggested method, instead, the square roots of the characteristic equation, i.e. closed loop poles, are received by using the Lambert W function [30] [31]. The proposed method can be applied for identifying unstable FOPTD models of nonlinear processes.

The results of the identification received in the proposed way have been compared with the results of the identification of the unstable FOPTD model and the reduction of the unstable SOPTD model into the unstable FOPTD model by using the methods given in [20] [22]-[27]. The findings indicate that the proposed method gives better results with identifying the unstable FOPTD model compared with the rest of the methods provided that it is a first-order process, and better results at the reduction of the unstable SOPTD model into unstable FOPTD model with all the methods except from the biased relay test method given in [23] whose results are similar.

An unstable FOPTD model of a nonlinear continuous bioreactor, obtained by the proposed identification method, is compared with the unstable FOPTD models obtained by other identification methods given in the references in [20] [25] [26] [29].

Mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), mean relative squared error (MRSE), mean absolute percentage error (MAPE) etc. are regularly employed in model validation studies. In this paper, the MAE and RMSE index are used to validate the model obtained.

The paper is divided in the following way: in Chapter II there is a short description of the Lambert W function. Chapter III contains an instruction for identification of the unstable FOPTD model in the suggested way. In Chapter IV, the results of the identification of the unstable FOPTD model are given for different values of the proportional controller gain by applying the proposed method. Validation of the model is given, too. Chapter V show the concurrent results of the identification with other methods used on the unstable FOPTD model, and the results received by identifying the unstable SOPTD model into unstable FOPTD model, respectively. In Chapter VI, the procedure of identification of unstable continual bioreactor into unstable FOPTD model is revealed. In this chapter, the comparison is given of the received model with the linearized bioreactor model and unstable FOPTD models received by using other methods.

2 Lambert W Function

Lambert W Function $W(z)$, where z belongs to a set of complex numbers C , is the solution of the equation

$$W(z)e^{W(z)} = z \quad (1)$$

The function has an infinite number of branches $W_k(z)$ where $k \in (-\infty, \infty)$, as well as an infinite number of solutions. Only two branches of the function, principal branch $W_0(z)$ where $k=0$ and $W_{-1}(z)$ for $k=-1$ can have real values. The range of the branches for z belongs to a set of real numbers R given in Figure 1.

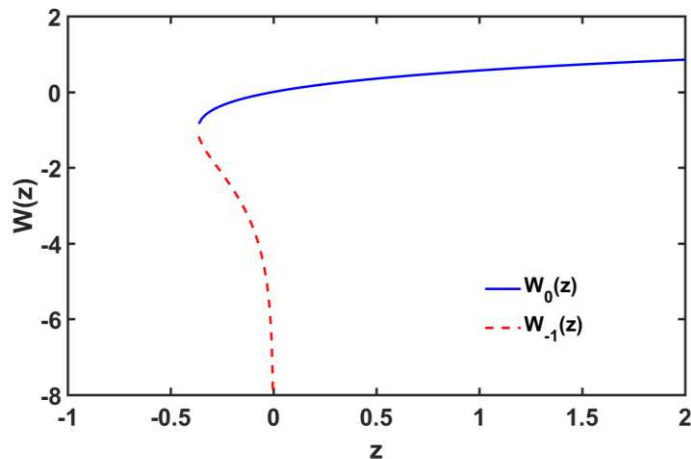


Figure 1

Two main branches of the Lambert W function for z belong to a set of real numbers R

It can be clearly seen in Figure 1 that the values of the principal branch $W_0(z)$, belong to the set $(-1, \infty)$, if z is a real number and if z takes values from the set $(-e^{-1}, \infty)$. The branch $W_{-1}(z)$ can have real values from the set $(-\infty, -1)$, only if z belongs to the set $(-e^{-1}, 0)$. Therefore, it is obvious that the equation (1) has two solutions $W_0(z)$ and $W_{-1}(z)$ if z belongs to a set of real numbers from $(-e^{-1}, 0)$.

A more detailed explanation of the method of solution (1), the branch range $W_k(z)$ and the conditions of convergence into C , can be read in [30].

3 Proposed Method of Parameter Estimation

Let the unstable first order plus time delay system, where K is the plants' gain, T is the time constant and θ is the time delay, be described by transfer function model

$$G(s) = \frac{K}{Ts - 1} e^{-qs} \quad (2)$$

The closed loop transfer function of the unstable FOPTD model stabilized by proportional controller, gain coefficient K_p , where $y(t)$ is the output and $r(t)$ is a reference step input amplitude R , becomes

$$W(s) = \frac{y(s)}{r(s)} = \frac{KK_p e^{-qs}}{Ts - 1 + KK_p e^{-qs}} \quad (3)$$

The selection of the controller gain needs to be undertaken, in such way, so that the underdamped system with the transfer function given in (3) is received, which equals to request that $0 < \zeta < 1$, where ζ is a damping ratio.

If, in the equation (3), time delay from a denominator is approximated by a Pade approximant, where τ and τ_0 are time constants defining poles and zero of the system transfer function, respectively, and K_i gain, the output of the closed loop system can be written down in the form

$$y(s) = \frac{K_i(t_0 s + 1)}{t s^2 + 2xt s + 1} e^{-qs} r(s) \quad (4)$$

Thus the observed system can be considered as the second order plus time delay processes with dynamic numerators.

The time transient closed loop step response is

$$y(t) = K_i R \left[1 - A e^{-xw_n(t-q)} \sin(w_d(t-q) + f) \right] \mathbb{1}_{t-q} \quad (5)$$

where natural frequency is $\omega_n = \tau^{-1}$, A coefficient which depends on a damping ratio ζ and time constants τ and τ_0 , ω_d damping frequency and ϕ starting phase of the system. The dependence of the damped frequency of the non-stoked frequency is given by the following equation:

$$w_d = w_n \sqrt{1 - \zeta^2} \quad (6)$$

A typical closed loop step response of the system (3) is shown in Figure 2.

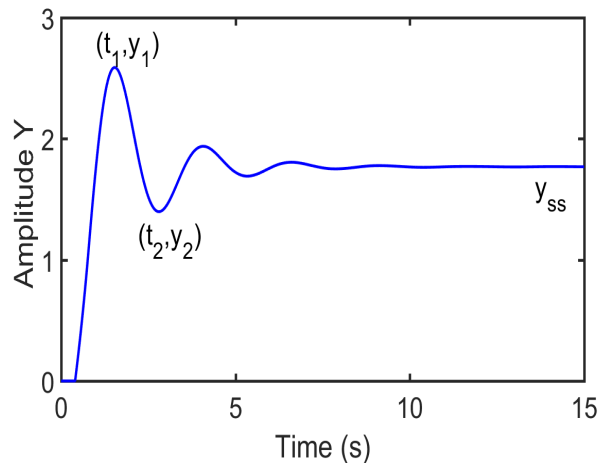


Figure 2

Closed loop step response of unstable FOPTD model under P control

In Figure 2, y_{ss} - is the value of output when time stretches to infinity i.e. steady state value, t_1 - is the time required for the output to reach its first maximum value, y_1 - is the first maximum value of output, t_2 - is the time required for the output to reach its first minimum value and y_2 - is the first minimum value of the output.

To apply the suggested method of the parameter estimation of the unstable FOPTD system it is necessary to determine all the already mentioned parameters.

By applying the final value theorem,

$$y_{ss} = \lim_{t \rightarrow \infty} y(t) = \lim_{s \rightarrow 0} sW(s) \frac{R}{s} = \frac{KK_p R}{KK_p - 1} \quad (7)$$

steady state value of output is received.

From (7), the gain K of unstable FOPTD model (2) is obtained

$$K = \frac{y_{ss}}{K_p (y_{ss} - R)} \quad (8)$$

Overshoot (OS) can be approximately calculated in the following way

$$OS = \frac{y_{ss} - y_2}{y_1 - y_{ss}} = e^{\frac{-px}{\sqrt{1-x^2}}} \quad (9)$$

wherefrom the damping ratio received is

$$x = \frac{-\ln(OS)}{\sqrt{p^2 + \ln^2(OS)}} \quad (10)$$

The time difference required for the output to reach its first maximal and minimal value equals the half the oscillation period whose frequency corresponds to the damped frequency ω_d

$$w_d = \frac{P}{t_2 - t_1} \quad (11)$$

so the un-damped frequency is

$$w_n = \frac{w_d}{\sqrt{1 - x^2}} \quad (12)$$

The characteristic equation of the system described by equation (4)

$$s^2 + 2xw_n s + w_n^2 = 0 \quad (13)$$

has conjugate-complex poles if the controller gain is selected so as for the underdamped closed loop system is received, whose values are

$$s_{1/2} = -xw_n \pm jw_d = -xw_n \pm jw_n \sqrt{1 - x^2} \quad (14)$$

The characteristic equation of the closed loop transfer function of the unstable FOPTD model stabilized by P controller (3)

$$Ts - 1 + KK_p e^{-qs} = 0 \quad (15)$$

has an infinite number of solutions which are received by applying Lambert W Function. Equation (15) can be converted into a Lambert W form

$$\theta(s - \frac{1}{T}) e^{\theta(s - \frac{1}{T})} = -\frac{K_p K}{T} \theta e^{-\frac{\theta}{T}} \quad (16)$$

wherefrom

$$\theta(s - \frac{1}{T}) = W_k(-\frac{K_p K}{T} \theta e^{-\frac{\theta}{T}}) \quad (17)$$

where k stands for an ordinal number of the Lambert W function branch. From (17) what follows is

$$s_k = \frac{1}{\theta} W_k(-\frac{K_p K}{T} \theta e^{-\frac{\theta}{T}}) + \frac{1}{T} \quad (18)$$

Considering that (4) is an approximation of the closed loop transfer function of the unstable FOPTD model stabilized by P controller (3), it is clear that it may be thought that the solutions of the characteristic equation (13) and (15) received by using relations (14) and (18) are identical. Because of this, the solution (18) for all the major branches does not need to be identified, but only for those which give dominant poles, and it has been illustrated that the latter are received by using the principal branch $W_0(z)$ and $W_{-1}(z)$.

This means that the solutions (18) assume a form of

$$\begin{aligned} s_1 &= -\xi\omega_n + j\omega_n\sqrt{1-\xi^2} \\ s_2 &= -\xi\omega_n - j\omega_n\sqrt{1-\xi^2} \end{aligned} \quad (19)$$

For the known poles (19), the unknown time constant T and time delay θ , the unstable FOPTD system (2) are received by solving a system of two equations

$$\begin{aligned} s_1 &= \frac{1}{\theta} W_0 \left(-\frac{K_p K}{T} \theta e^{-\frac{\theta}{T}} \right) + \frac{1}{T} \\ s_2 &= \frac{1}{\theta} W_{-1} \left(-\frac{K_p K}{T} \theta e^{-\frac{\theta}{T}} \right) + \frac{1}{T} \end{aligned} \quad (20)$$

whereby all the parameters of the unstable FOPTD system have been estimated.

The proposed way of the unstable FOPTD system parameter estimation follows these steps:

Step 1. For the selection gain K_p of the P controller and the selected amplitude R of the reference step input record the closed loop step response. If the received response does not have characteristics of the underdamped step response increase the controller gain and record the closed loop step response.

Step 2. From the received closed loop step response find (read, i.e. measure) the values of the necessary parameters y_{ss} , t_1 , y_1 , y_2 and t_2 for applying the proposed method.

Step 3. By applying the first part of the equation (9), on the basis of the measured values y_{ss} , y_1 , y_2 determine the overshoot (OS). The received value OS replace in (10) and calculate the damping ratio ξ .

Step 4. Based on the measured values t_1 and t_2 from closed loop step response (Step 2), determine the damped frequency ω_d by using (11) and then calculate the un-damped frequency ω_n by applying (12).

Step 5. For the received values ξ , ω_n and ω_d , determine the closed loop poles s_1 and s_2 by applying (14).

Step 6. By applying (8), for the measured value y_{ss} , the applied amplitude step input R and applied proportional controller gain K_p , determine gain K of unstable FOPTD model.

Step 7. Replace the applied value of the controller gain K_p , the received gain K and received closed loop poles s_1 and s_2 into (20). By solving a system of two equations (20) the time constant T and the time delay θ are received.

Step 8. Evaluating model performance. For performance indicators can be used

$$MAE = \frac{1}{n} \sum_{i=0}^n |y - y_m|$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=0}^n (y - y_m)^2}$$
(21)

The MAE and RMSE index (21) are used to measure the Mean Absolute Error and Root Mean Squared Error between y - the real process output and y_m - the output produced by the model. MAE and RMSE of 0, indicates a perfect model.

4 Simulation Study and Numerical Examples

In literature, this kind of process (2) is usually considered alongside parameters $K=1$ and the relation between the time delay and time constant within the range $\theta=(0.1-0.8)T$.

Considered the unstable FOPTD model with $K=1$, $T=1$ and $\theta=0.1T$ studies in the reference [26].

For illustrating the proposed method, different proportional controller gain values have been used. Closed loop step responses of the received models with different gains K_p are provided in Figure 3.

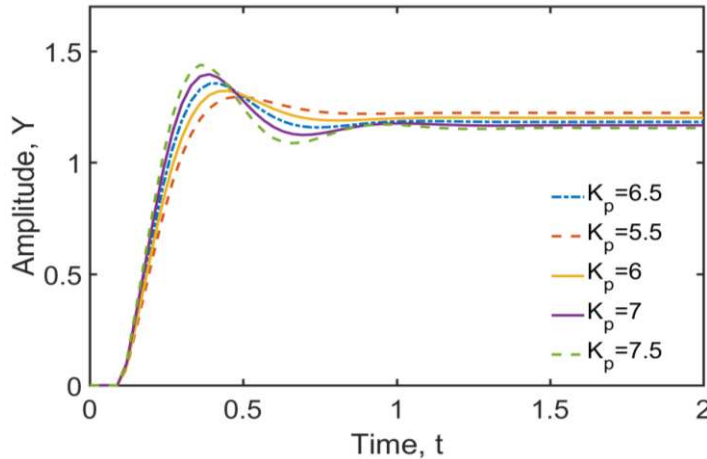


Figure 3

Closed loop reaction curve for various K_p values

Matlab code and Simulink model for example $K_p=6$ are available on the following link <https://drive.google.com/open?id=1OrrcjCePghquJe5X2jqWiKcA0KgVxd2F>

The read values of the closed loop step response and the obtained parameters of the unstable FOPTD model for three different proportional controller gains are given in Table 1.

Frequency responses of the real process and the received models, with three different proportional gains K_p , are provided in Figure 4.

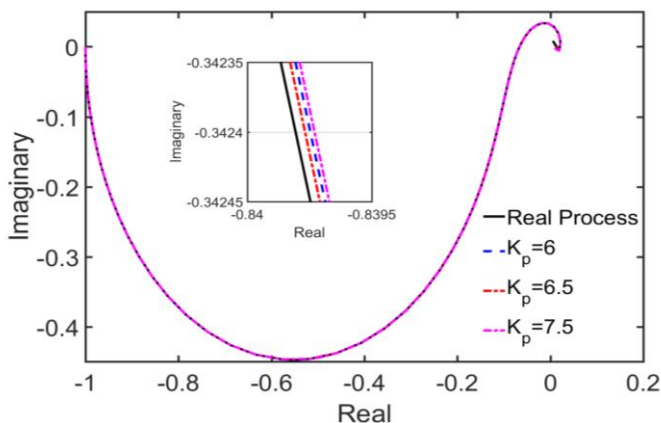


Figure 4

Nyquist plots of real process and identified models for various K_p values

Table 1

Received Parameters of unstable FOPTD model for different value of proportional gain of P controller

	$K_p=6$	$K_p=6.5$	$K_p=7.5$
y_{ss}	1.2	1.182	1.1538
y_1	1.3207	1,3546	1.4376
y_2	1.188	1.1569	1.086
t_1	0.45	0.42	0.36
t_2	0.84	0.78	0.69
OS	0.09899	0.1439	0.2388
ξ	0.5928	0.5251	0.4148
ω_d	8.0554	8.7266	9.5199
ω_n	10.0027	10.2541	10.4626
$s_{1/2}$	$-5.9299 \pm 8.0554j$	$-5.3845 \pm 8.7266j$	$-4.3400 \pm 9.5199j$
Identified model	$K=1$ $T=1.0719$ $\theta=0.1074$	$K=1$ $T=1.0813$ $\theta=0.1083$	$K=0.9999$ $T=1.1238$ $\theta=0.1122$

The validation of the received models has been carried out by applying (21), i.e. by finding MAE and RMSE index. As the considered processes and the received process are unstable, for model validation in the time domain in (21), for y and y_m

the closed loop step response output of real process and identified model has been selected, with the proportional controller gain K_{pv} . For validating the model in the frequency domain, in (21) $y=|K_{pv}G_m(j\omega)|$ is the magnitude of the open loop system with real process response, and $y_m=|K_{pv}G_m(j\omega)|$ is magnitude of the open loop system with identified model response.

For the model validation the proportional controller gain $K_{pv}=4$ has been used. Closed loop step and frequency response has been simulated with the software, Matlab/Simulink. Specifications are: solver ODE5, $R=1$, duration 30 s, step size 0.03 s, frequency range for frequency response (0.01-phase crossover frequency)=(0.01-15)rad/s.

The received values of the response errors in the time domain (TD) and frequency domain (FD) of the given MAE and RMSE indices, during the identification (proportional controller gain K_p) and validation (proportional controller gain K_{pv}), are given in Table 2.

Table 2

MAE and RMSE received by identification and validation in the time and frequency domain

	Model ($K_p=6$)		Model ($K_p=6.5$)		Model ($K_p=7.5$)	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
TD-identif.	0.000546	0.004121	0.000705	0.005043	0.001449	0.009360
TD-validat.	0.000626	0.004316	0.000714	0.004898	0.001108	0.007458
FD-identif.	0.052512	0.075997	0.064088	0.092725	0.110478	0.159653
FD-validat.	0.035008	0.050665	0.039438	0.057060	0.058922	0.085148

It can be observed, based on the results in Table 2, that the considered unstable FOPTD process has been adequately identified, and that the model received by using the proportional controller gain $K_p=6$ with parameters $K=1$, $T=1.0719$ and $\theta=0.1074$ has the lowest MAE and RMSE index, i.e. it represents the dynamic of the researched process most faithfully.

5 Illustrations of the Examples of Comparison with Other Methods

For testing the quality of the received results the comparison of the proposed method with other methods from two reference works has been presented. First of the examples show the comparison of the results of identification in the low order model and the second shows the identification in the second order model which needs to be classified as the unstable first-order model whereby one comparison has been done for the closed loop identification and the other for the methods based on relay use.

Example 1

Consider the unstable first-order process studies in the references [20], [22] and [23] with the parameters $K=1$, $T=1$ and $\theta=0.4$.

Step 1: Closed loop step response with gain $K_p=1.5$ and $R=1$. Step 2: From closed loop step response received parameters are: $y_{ss}=3$, $t_f=3.27$, $y_1=3.0536$, $y_2=2.991$, $t_2=6.27$. Step 3: obtained parameters $OS=0.01765$, $\xi=0.7892$. Step 4: calculated $\omega_d=1.0472$, $\omega_n=1.7051$. Step 5: calculated closed loop poles $s_{1/2}=-1.3457\pm 1.0472j$. Step 6: identified model gain $K=1$. Step 7: identified model time constant and time delay $T=1.0036$, $\theta=0.4015$. Step 8: The obtained performance index with $K_p=1.5$ for identification are: in time domain $MAE=0.000326$, $RMSE=0.001157$; in frequency domain $MAE=0.000655$, $RMSE=0.001019$. Model validation with $K_{pv}=3$ and frequency range (0.01-3.16) rad/s. The obtained performance index for validation are: in time domain $MAE=0.011578$, $RMSE=0.015213$ in frequency domain $MAE=0.001309$, $RMSE=0.002039$.

The estimation of the three parameters by using the proposed closed loop step response under proportional control method, with control gain $K_p=1.5$, the process with parameters $K=1$, $T=1.0036$, $\theta=0.4015$ has been successfully identified.

For this low order model, by using of relay with additional delay in [22] received the model with parameters $K=0.928$, $T=0.757$ and $\theta=0.395$. By using the advanced symmetric relay feedback test method, this unstable process is identified in [20] with the following parameters $K=0.9841$, $T=1.1332$ and $\theta=0.4372$. The same process is identified with the following parameters $K=1.0001$, $T=0.9954$ and $\theta=0.4$ by performing a biased relay method in [23].

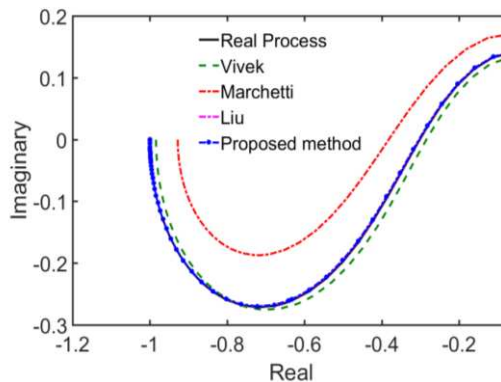


Figure 5
Nyquist plots of identified models for Example 1

After comparing the received results, by inspecting the frequency characteristics shown in Figure 5, it can be inferred that with the suggested identification method better results have been received compared with those shown in [20] and [22], but same as those shared in [23].

Example 2

Considered the unstable second-order process studies in the references [20] [23]-[25], and [27]

$$G(s) = \frac{1}{(2s - 1)(0.5s + 1)} e^{-0.5s}$$

Step 1: Closed loop step response with gain $K_p=1.5$ and $R=1$. Step 2: $y_{ss}=3.00028$, $t_1=6.06$, $y_1=3.5097$, $y_2=2.9135$, $t_2=11.6999$. Step 3: $OS=0.1703$, $\xi=0.4908$. Step 4: $\omega_d=0.5570$, $\omega_n=0.6393$. Step 5: $s_{1/2}=-0.3138 \pm 0.5570j$. Step 6: identified $K=0.9999$. Step 7: identified $T=2.3106$, $\theta=1.1507$. Step 8: Performance index with $K_p=1.5$ for identification are: in time domain $MAE=0.048533$, $RMSE=0.106096$ in frequency domain $MAE=0.015099$, $RMSE=0.022390$. Model validation with $K_{pv}=1.7$ and frequency range (0.01-1.26) rad/s. The gain of the controller cannot be changed much because the closed loop system would become unstable. The obtained performance index for validation are: in time domain $MAE=0.062554$, $RMSE=0.113107$ in frequency domain $MAE=0.017112$, $RMSE=0.025375$.

The identified unstable FOPTD model by using proposed method for the unstable SOPTD process has parameters: $K=0.9999$, $T=2.3106$, $\theta=1.1507$.

The improved closed loop step response identification of the process with PID regulator has been given in the [27] where two methods of identification are put forward. In [25] the observed process has been identified by using a PID controller. The received parameters of the suggested way of identification and other methods [25] [27] have been given in Table 3.

The results of the comparison presented by Nyquist diagram are shown in Fig. 6. It can be clearly seen from the figure that the suggested method, compared with the closed loop step response methods, gives incomparably better results.

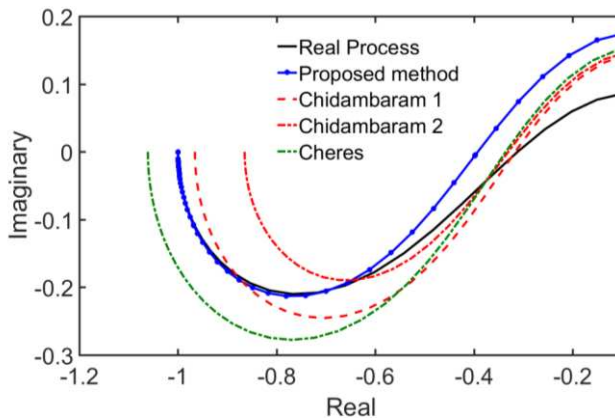


Figure 6

The Nyquist fitting of identified unstable FOPTD models with parameters given in Table 3

Table 3
Parameters of identified models for Example 2

Method	Gain K	Time constant T	Time delay θ
Sree and Chidambaram I [26]	0.9567	2.4278	1.0416
Sree and Chidambaram II [26]	0.8649	2.0615	1.0051
Cheres [24]	1.061	2.545	1.06
Proposed method	0.9999	2.3106	1.1507

In the study [20] the results for relay and improved relay method of identifying this process within the unstable first order model are given. For the same unstable SOPTD process, an unstable FOPTD model was obtained by the proposed method in [23] and using one relay controller in [24].

Figure 7 shows the Nyquist fitting of identified unstable FOPTD models, whereas the parameters of the identified models in [20] [23] [24] and parameters of proposed identified model have been indicated in Table 4.

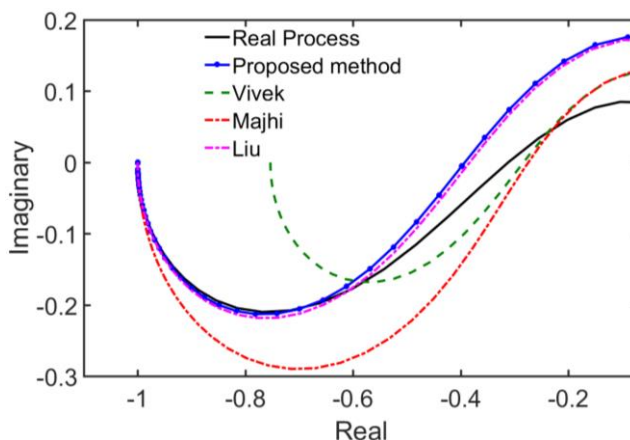


Figure 7

The Nyquist fitting of identified unstable FOPTD models with parameters given in Table 4

Table 4
Parameters of identified models

Method	Gain K	Time constant T	Time delay θ
Vivek and Chidambaram [19]	0.7534	2.1642	1.0412
Liu and Gao [22]	1.0001	2.1459	1.0486
Majhi and Atherton [23]	1	2.875	1.061
Proposed method	0.9999	2.3106	1.1507

From the frequency characteristics it can be concluded that the results received in [20] and [24] indicate the greatest deviations and that the proposed method and the method suggested in [23] yield positive and similar results.

6 Simulation Study of a Continuous Bioreactor

A nonlinear continuous bioreactor exhibits output multiplicity behavior. Considered bioreactor model used in [20], [25]-[26], [28]-[29] with substrate inhibition

$$\begin{aligned}\frac{dx_1}{dt} &= (\mu - D)x_1 \\ \frac{dx_2}{dt} &= (x_{2f} - x_2)D - \frac{\mu x_1}{\gamma} \\ \mu &= \frac{\mu_{\max} x_2}{K_m + x_2 + K_i x_2^2}\end{aligned}\quad (22)$$

where $\mu_{\max}=0.53 \text{ h}^{-1}$, $K_m=0.12 \text{ g/l}$, $K_i=0.4545 \text{ l/g}$, $x_{2f}=4.0 \text{ g/l}$, $\gamma=0.4 \text{ g/g}$ and controlled variable x_1 – biomass (cell) concentration (g/l), x_2 – substrate concentration (g/l), manipulated input D – dilution rate (h^{-1}), disturbance input x_{2f} – substrate feed concentration (g/l), μ – specific growth rate constant (h^{-1}), μ_{\max} – maximum specific growth rate constant (h^{-1}), γ – the yield of cell mass (g/g), K_m – substrate saturation constant (g/l) and K_i – substrate inhibition rate constant (l/g). The steady state dilution rate $D=0.3 \text{ h}^{-1}$.

There are three steady state solutions for biomass and substrate: washout (trivial) stable solution $x_{1s}=0$, $x_{2s}=x_{2f}=4.0$, unstable solution $x_{1s}=0.995103$, $x_{2s}=1.512243$ and stable solution $x_{1s}=1.530163$, $x_{2s}=0.174593$.

Step 1: The dilution rate is taken as the manipulated variable in order to control the biomass (cell) concentration x_1 at the unstable steady state. A delay of 1h is considered in the measurement of x_1 . The nonlinear model equations are solved along with the proportional controller $K_p=-1.1$. A step change from 0.995103 to 1.144368 is introduced in the biomass concentration reference and the closed loop response is obtained using Matlab/Simulink. Specification: solver ode45, duration 80h, frequency range (0.01-4.68)rad/s for validation.

Step 2: From recorded output the following value are obtained: $y_{ss}=1.173249$, $t_1=2.1204$, $y_1=1.4171$, $y_2=0.9936$, $t_2=4.6728$. From Step 3 to Step 5: $OS=0.7365$, $\xi=0.0969$, $\omega_d=1.2309$, $\omega_n=1.2367$, $s_{1/2}=-0.1198 \pm 1.2309j$. Step 6: Considering that the reference input signal changes from 0.995103 to 1.144368, the change of the referent biomass concentration is $\Delta r=0.149265$. This modulation is in line with the change of the biomass signal concentration from 0.995103 (initial condition) to 1.173249, $\Delta y=0.174186$. Therefore, the equation (8) is transformed into

$$K = \frac{Dy}{K_p(Dy - Dr)} \quad (23)$$

From the estimated value of the amplification gain $K=-5.60756$. Step 7: identified $T=5.5423$, $\theta=1.0818$.

The identified model can be compared on the model obtained by linearization around the operating point. For the given condition of the unstable operating point, the local linearized unstable FOPTD model for the unstable bioreactor is

$$G_{linear.}(s) = \frac{Dx_1(s)}{DD(s)} = \frac{-5.8604}{5.8893s - 1} e^{-s} \quad (24)$$

Step 8: Performance index for identification are: in time domain MAE=0.0454, RMSE=0.0866 in frequency domain MAE=0.1113, RMSE=0.1581.

Table 5 shows the values of the parameters of unstable FOPTD models given in [20], [25], [26] and [29] as well as MAE and RMSE indices received based on the frequency response of the unstable FOPTD model and linearized model of the unstable bioreactor for the frequency range (0.01-4.68)rad/s where 4.68 rad/s is the phase crossover frequency

$$MAE = \frac{1}{n} \sum_{i=0}^n \left| |G_{linear.}(j\omega)| - |G_{ident.}(j\omega)| \right| \quad (25)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=0}^n (|G_{linear.}(j\omega)| - |G_{ident.}(j\omega)|)^2}$$

The Nyquist fitting of linearized model and identified unstable FOPTD models for the unstable bioreactor are shown in Figure 8.

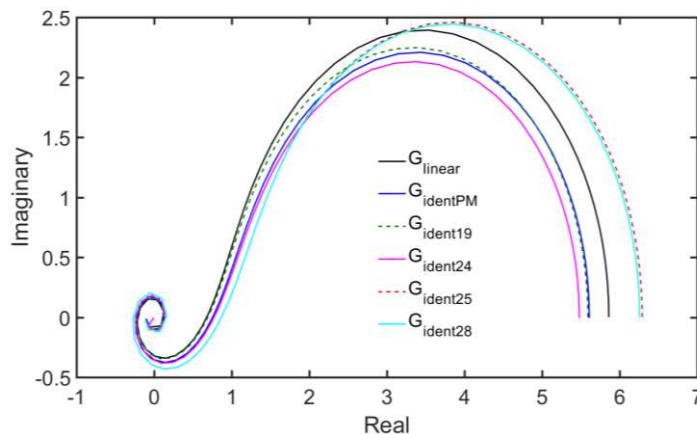


Figure 8

The Nyquist fitting of linearized and identified models with parameters given in Table 5

Table 5

Parameters of identified unstable FOPTD models for the bioreactor and MAE and RMSE index

Model from reference	K	T	θ	MAE	RMSE
[20]	-5.5903	5.6125	1.0152	0.1118	0.1586
[25]	-5.48	4.51	0.92	0.2352	0.2628
[26]	-6.29	5.001	1.0	0.4010	0.4409
[29]	-6.257	5.36	1.076	0.3151	0.3501
Proposed Method	-5.60756	5.5423	1.0818	0.1011	0.1437

The output results indicate that the proposed method of the unstable FOPTD model, for the unstable bioreactor, shows the lowest level of deviation from the model received by linearization, in comparison with the models received by using other methods.

Conclusions

The proposed, Closed Loop Step Response method of identification, using a proportional controller, yields good results, regardless of the fact that there is a large offset of the output signal. In future work, the method can be improved using the PI or PID controller, whereby the unknown parameters would be calculated by using the Lambert W Function, to identify not only the unstable FOPTD process, but also to identify other types of process.

Acknowledgement

This paper was realized as a part of the projects III 43007 and TR 35005, funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

References

- [1] L. Ljung: System Identification-Theory for the User, Prentice-Hall, Englewood Cliffs, New Jersey, 1987
- [2] L. Ljung: Perspectives on system identification, Annual Reviews in Control, Vol. 34 (1), April 2010, pp. 1-12
- [3] K. S. Narendra, K. Parthasarathy: Identification and control of dynamical systems using neural networks, IEEE Transactions on Neural Networks, vol. 1(1), 1990, pp. 4-27
- [4] J. Saadat, P. Moallem, H. Koofgar: Training Echo State Neural Network Using Harmony Search Algorithm, International Journal of Artificial Intelligence, Vol. 15, No. 1, 2017, pp. 163-179
- [5] A. Chatterjee, R. Chatterjee, F. Matsuno, T. Endo: Augmented stable fuzzy control for flexible robotic arm using LMI approach and neuro-fuzzy state space modeling, IEEE Transactions on Industrial Electronics, Vol. 55, No. 3, 2008, pp. 1256-1270

-
- [6] A. Ürmös, Z. Farkas, M. Farkas, T. Sándor, L. T. Kóczy, Á. Nemcsics: Application of self-organizing maps for technological support of droplet epitaxy, *Acta Polytechnica Hungarica*, Vol. 14, No. 4, 2017, pp. 207-224
- [7] J. Kozłowski and Z. Kowalczyk: On-line parameter and delay estimation of continuous-time dynamic systems, *International Journal of Applied Mathematics and Computer Science*, Vol. 25(2), 2015, pp. 223-232
- [8] Y. Orlov, I. V. Kolmanovskiy, O. Gomez: Delay estimation in linear systems using output feedback, *Proceedings of the 45th IEEE Conference on Decision and Control*, San Diego, CA, USA, Dec. 2006, pp. 858-863
- [9] S. J. Qin, T. A. Badgwell: A survey of industrial model predictive control technology, *Control Engineering Practice*, Vol. 11(7), 2003, pp. 733-764
- [10] R. Zhang, F. Gao, Z. Cao, P. Li: Design and implementation of an improved linear quadratic regulation control for oxygen content in a coke furnace, *IET Control Theory and Applications*, Vol. 8(14), 2014, pp. 1303-1311
- [11] T. Haidegger, L. Kovács, R. E. Precup, B. Benyó, Z. Benyó, S. Preitl: Simulation and control for telerobots in space medicine, *Acta Astronautica*, Vol. 81(1), 2012, pp. 390-402
- [12] R. Gerov, Z. Jovanović: Synthesis of PI Controller with a Simple Set-Point Filter for Unstable First-Order Time Delay Processes and Integral plus Time Delay Plant, *Elektronika ir Elektrotehnika*, Vol. 24(2) 2018, pp. 3-11
- [13] K. J. Åström, T. Hägglund: *PID Controllers: Theory, Design, and Tuning*, Research Triangle Park, North Carolina, Instrument Society of America, 1995
- [14] T. Liu, Q. G. Wang, H. P. Huang: A tutorial review on process identification from step or relay feedback test, *Journal of Process Control*, Vol. 23 (10), 2013, pp. 1597-1623
- [15] K. J. Åström, T. Hägglund: Automatic tuning of simple regulators with specification on phase and amplitude margins, *Automatica*, Vol. 20, 1984, pp. 645-651
- [16] K. Srinivasan, M. Chidambaram: An Improved Autotune Identification Method, *Chemical and Biochemical Engineering Quarterly*, Vol. 18(3), 2004, pp. 249-256
- [17] I. Kaya: Parameter Estimation for Integrating Processes Using Relay Feedback Control under Static Load Disturbance, *Industrial & Engineering Chemistry Research*, Vol. 45(13), 2006, pp. 4726-4731
- [18] M. Veronesi, A. Visioli: Process Parameters Estimation, Performance Assessment and Controller Retuning Based on the Final Value Theorem: Some Extensions, *IFAC-PapersOnLine*, Vol. 50(1), 2017, pp. 9198-9203

- [19] M. Hofreiter: Alternative Identification Method using Biased Relay Feedback, *IFAC PapersOnLine*, Vol. 51(11), 2018, pp. 891-896
- [20] S. Vivek, M. Chidambaram: An improved relay auto tuning of PID controllers for unstable FOPTD systems, *Computers and Chemical Engineering*, Vol. 29, 2005, pp. 2060-2068
- [21] D. Kishorea, K. Anand Kishorea, R. C. Panda: Identification and control of process using the Modified asymmetrical Relay Feedback method, *Procedia Computer Science*, Vol. 133, 2018, pp. 1029-1034
- [22] G. Marchetti, C. Scali, D. R. Lewin: Identification and control of open-loop unstable processes by relay methods, *Automatica*, Vol. 37(12), 2001, pp. 2049-2055
- [23] T. Liu, F. Gao: Identification of integrating and unstable processes from relay feedback, *Computers and Chemical Engineering*, Vol. 32, 2008, pp. 3038-3056
- [24] S. Majhi, D. P. Atherton: On-Line Tuning of Controllers for Unstable FOPTD Processes, *IEEE Proceedings Control Theory and Applications*, Vol. 147 (4), 2000, pp. 421-427
- [25] E. Cheres: Parameter estimation of an unstable system with a PID controller in a closed loop configuration, *Journal of the Franklin Institute*, Vol. 343 (2), 2006, pp. 204-209
- [26] I. Ananth, M. Chidambaram: Closed-loop identification of transfer function model for unstable systems, *Journal of the Franklin Institute*, Vol. 336, 1999, pp. 1055-1061
- [27] R. P. Sree, M. Chidambaram: Improved closed loop identification of transfer function model for unstable systems, *Journal of the Franklin Institute*, Vol. 343 (2), 2006, pp. 152-160
- [28] B. Wayne Bequette: *Process Control: Modeling, Design and Simulation*, Prentice Hall, 2003
- [29] S. Pramod, M. Chidambaram: Closed loop identification of transfer function model for unstable bioreactors for tuning PID controllers, *Bioprocess Engineering*, Vol. 22(2), 2000, pp. 185-188
- [30] R. Corless, G. Gonnet, D. Hare, D. Jeffrey, D. Knuth: On the Lambert W function, *Advances in Computational Mathematics*, Vol. 5, 1996, pp. 329-359
- [31] S. Yi, P. Nelson, A. G. Ulsoy: Analysis and Control of Time Delayed Systems via the Lambert W Function, *IFAC Proceedings Volumes*, Vol. 41(2), 2008, pp. 13414-13419

Vehicle Dynamic-based Approach for the Optimization of Traffic Parameters of the Intelligent Driver Model (IDM) and for the Support of Autonomous Vehicles' Driving Ability

Tamás Péter¹ and István Lakatos²

¹ Department of Control for Transportation and Vehicle Systems, Budapest University of Technology and Economics; Stoczek u. 2, H-1111 Budapest, Hungary; peter.tamas@mail.bme.hu

² Széchenyi István University SZE KVJT and JKK Egyetem tér 1, H-9026 Győr, Hungary, lakatos@sze.hu

Abstract: The research identifies the dynamical parameters in the area of mechanics within the base of traffic model parameters used by the Intelligent Driver Model (IDM), which are the highest acceleration parameters set by the vehicle, the desired speed parameters of the vehicle and the distance-keeping parameters of the vehicle. All this facilitates the automatic control of autonomous electric vehicles in certain vehicle groups.

Keywords: IDM; dynamics-based approach; support for autonomous vehicles' driving

1 Introduction

The aim of the research is the longitudinal dynamic optimization of the driving processes. This means the support of an optimally attenuated, non-hectic traffic process that results in energy savings and noise reduction of vehicles in the vehicle group and in addition to this, emission reduction of conventional vehicles [14, 15]. During the course of the ride, minimal oscillations in speed occurs. Therefore, it reduces speed changes and the number of excessive braking. In the analysis the acceleration capabilities are vehicle characteristics. In the optimization task the variables are tracking distances and desired speed values. An important requirement is the definition of minimum distances for the former and the upper limit for the latter. In this paper, we investigate a group of vehicles with almost identical characteristics in terms of acceleration properties, but this analysis can naturally be extended to variable-component vehicle groups as well.

The research material discusses a highly complex problem, i.e. the re-formulation of the IDM models into mechanical-dynamic systems, and thus conducts studies in a physical parameter space and system in which significant knowledge in the field of physics and dynamics has accumulated.

It can be said that the theoretical foundation and application [12] of IDM models is very good, however, the complexity [7, 11] of real-world traffic processes poses serious problems to professionals in the applications. On the one hand, greater error tolerance than the accuracy of traffic parameters is to be expected than in many other dynamical systems. On, the other hand, it can be stated that complexity is extremely high [16, 28, 29, 30] either when assessing the entirety of effects of road traffic on a single vehicle, or only the information [1] to be processed by the driver or the autopilot. Similarly, the system is very complex if we consider only the surface traffic processes as a whole over a large-scale network [21, 22, 23, 24, 25]. The reality is, however, much more complex, since the above elements form a single large-scale dynamic system during their operation and are in constant interaction with each other. This complex system in its physical reality consists, on the one hand of the multitude of vehicle-dynamical systems [8] (which is a multitude of man-autopilot-machine systems) and, on the other hand, of the multitude of static and dynamic traffic network elements. Finally, all the above is surrounded by a very complicated dynamic external environment that, in addition to seasonality, also has different geographic, meteorological, economic and cultural characteristics [5, 6, 13].

In connection with the above, we will briefly review the traffic-related applications of IDM, as well as the relationship between the IDM model and the large-scale networks. We investigate the longitudinal dynamic properties of the vehicles during the catch-up with the leader vehicle.

For this complex dynamic process we interpret Lehr's relative attenuation factor δ and ω eigenfrequency from the field of mechanics as dynamic characteristics with well-known effects. The purpose of the analysis is to optimize traffic parameters and to support the driving of autonomous vehicles using the above introduced dynamic characteristics.

2 The IDM Model and Its Traffic-related Applications

Adaptive Cruise Control (ACC) is a vehicle system that allows the vehicle to adjust its speed to the environment. The Intelligent Driver Model (IDM) is an adaptive cruise control (ACC) model that is widely used in transportation research to model longitudinal movement. Treiber, Hennecke and Helbing developed the Intelligent Driver Model (IDM) (1), which is being used by the car company BMW, in 2000 at the transport laboratory of Dresden Technical University [33, 34].

$$\dot{v}_k = a_k \left[1 - \left(\frac{v_k}{v_k^0} \right)^4 - \left(\frac{s^*(v_k, \Delta v_k)}{s_k} \right)^2 \right] \quad (1)$$

Where:

a_k is the maximal acceleration of the k th vehicle,

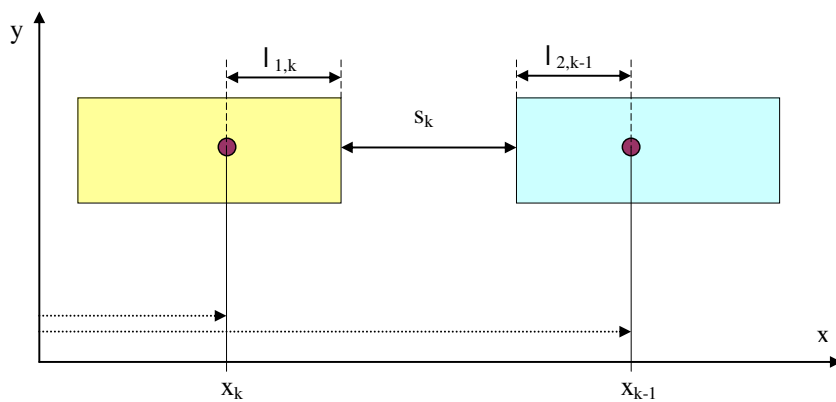
$v_k = \dot{x}_k$ is the speed of the k th vehicle,

v_k^0 is the desired speed of the k th vehicle,

s_k is the distance between the k th and the preceding vehicle,

$\Delta x_k = x_{k-1} - x_k$ the difference between the positions of the centre of gravity of the $(k-1)$ th vehicle and that of the k th vehicle,

$s_k = \Delta x_k - l_k = (x_{k-1} - x_k) - l_k$; (see Fig. 1)



$$s_k = (x_{k-1} - l_{2,k-1}) - (x_k + l_{1,k}) = (x_{k-1} - x_k) - (l_{2,k-1} + l_{1,k}) = (x_{k-1} - x_k) - l_k$$

Figure 1

The s_k distance between consecutive vehicles

$$s^*(v_k, \Delta v_k) = s_k^0 + T_k \cdot v_k + \frac{v_k \cdot \Delta v_k}{2\sqrt{a_k b_k}}; \quad (2)$$

Where:

s_k^0 is the congestion speed of the k th vehicle,

The IDM model is used for modeling continuous traffic flows in simulations of highway and city traffic. As a vehicle tracking model, IDM describes the dynamics of the position and speed of each vehicle. In the case of Multi-model Open Source Traffic Simulator, [32] use IDM to simulate the longitudinal movement of the vehicle and this simulator also introduces a lane change strategy. Model-based single-lane traffic inhomogeneity is studied by [35].

The work of [36] studies vehicle stability and IDM parameter sensitivity. The work of [10] proposed extending the driver parameters of the IDM model. They study the impact of vehicles equipped with IDM on traffic flow and travel times as bottlenecks. The work of [9] also uses the IDM model and examines the impact of Adaptive Cruise Control on traffic flows. The results of the above work show that increasing the proportion of ACC vehicles will result in increased traffic efficiency by reducing travel times. The work of [37] used IDM to study instability in congested traffic.

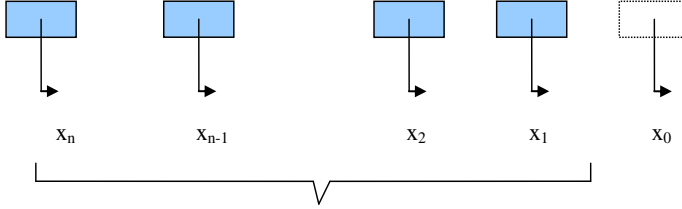
The IDM model has many advantages over other ACC models from a calibration and intuitive parameters point of view, and also modeling requires simple simulation. However, there are also disadvantages in respect of assuring the proper features of the vehicle and the driver. The IDM is a collision-free model. Therefore, in critical accident situations, the desired minimum distance is no longer sufficient to guarantee driver safety and, in the event of an emergency braking, it tends to overshoot the actual deceleration of the vehicle.

The works of [2, 3] developed a proposal for a more accurate operation of the IDM model and studied possible modifications to IDM, taking into account driver's safety and the real capabilities of the vehicle. As a result of this amendment, the driver has to take into account the behavior of the following vehicles, and thus a modified IDM model has been developed and tested with a microscopic simulator considering string stabilization. This modified IDM model already highlights the proper vehicle capabilities.

Based on this, the IDM model is already providing greater performance in driver security by following real reactions in near-collision critical situations. The paper shows the modification and the state-of-the-art operation of the intelligent driver model in connection with the proper capabilities of the vehicle.

Modeling and research work encompasses a complex area and includes approaches of both microscopic and macroscopic modeling, [4]. The complex macroscopic traffic environment is generated by the large-scale network model, in which the microscopic traffic simulation model provides the individual vehicle movement in traffic on the sections of the defined trajectories. However, this microscopic model must properly reproduce dynamic traffic processes and must also be validated. Accordingly, at this stage of our work we rely on Intelligent Driver Model research and development of [33, 34, 4].

The features of the classical IDM model are the following: a single system of differential equations that analyses the case of n vehicles traveling on a single lane. The microscopic model describes a chain model-like longitudinal dynamics. Each driver looks only forward and aims to keep an appropriate distance. There is no overtaking, the vehicles keep their order and the first vehicle has a dominant role, as do the slow-moving vehicles in the group.



The n element vehicle group

Figure 2

The n-element vehicle group and the environment determining their movement

According to the above the classical IDM model is written with separate differential equations, member by member. The works of [2, 3], are summarized in the following system of differential equations (4), where the current position of the i th vehicle is described by function $x_i(t)$. The parameters and functions used in the model are as follows:

a_i is the maximal acceleration of the i th vehicle,

v_i is the desired speed of the i th vehicle,

s_i is the required distance between the i th and the preceding vehicle ($i=1,2, \dots, n$),

$$\langle \underline{\underline{A}} \rangle^{-1} \ddot{x}(t) + \langle \underline{\underline{V}} \rangle^{-1} \underline{f}_1(\dot{x}(t)) + \langle \underline{\underline{S}} \rangle \underline{f}_2(x(t)) = \underline{1} \quad (4)$$

$$\langle \underline{\underline{A}} \rangle^{-1} = \left\langle \frac{1}{a_1}, \frac{1}{a_2}, \dots, \frac{1}{a_n} \right\rangle; \langle \underline{\underline{V}} \rangle^{-1} = \left\langle \frac{1}{v_1^4}, \frac{1}{v_2^4}, \dots, \frac{1}{v_n^4} \right\rangle; \langle \underline{\underline{S}} \rangle = \langle s_1^2, s_2^2, \dots, s_n^2 \rangle$$

$$s_i = s_{0i} = \text{const.}, \text{ or: } s_i = s_i(\dot{x}_{i-1}, \dot{x}_i) \quad (i=1,2, \dots, n).$$

$$\underline{f}_1(\dot{x}(t)) = \begin{bmatrix} \dot{x}_1^4 \\ \dot{x}_2^4 \\ \dots \\ \dot{x}_n^4 \end{bmatrix}, \quad \underline{f}_2(x(t)) = \begin{bmatrix} \frac{1}{(x_0 - x_1)^2} \\ \frac{1}{(x_1 - x_2)^2} \\ \dots \\ \frac{1}{(x_{n-1} - x_n)^2} \end{bmatrix}, \quad \underline{1} = \begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix}$$

Detailed description of the above is provided by [2, 3]. This model, using a function $h(t)$ also takes into account the fact that drivers monitor the movement of the following vehicles as well, see Fig. 3.

$$\langle \underline{\underline{A}} \rangle^{-1} \ddot{x}(t) + \underline{\underline{V}} f_1(\dot{x}(t)) + \underline{\underline{S}} f_2(x(t)) = \underline{\underline{1}}(t) + \underline{\underline{h}}(t) \quad (5)$$

$$\langle \underline{\underline{A}} \rangle^{-1} = \left\langle \frac{1}{a_1}, \frac{1}{a_2}, \dots, \frac{1}{a_n} \right\rangle;$$

$$\underline{\underline{V}} = \begin{bmatrix} \frac{1}{v_1^4} & \frac{h_1}{v_2^4} \\ & \frac{1}{v_2^4} & \frac{h_2}{v_3^4} \\ & & \frac{1}{v_i^4} & \frac{h_i}{v_{i+1}^4} \\ - & - & - & - \\ & & & \frac{1}{v_n^4} \end{bmatrix}; \quad \underline{\underline{S}} = \begin{bmatrix} s_1^2 & h_1 s_2^2 & & & \\ & s_2^2 & h_2 s_3^2 & & \\ & & s_i^2 & h_i s_{i+1}^2 & \\ - & - & - & - & - \\ & & & & s_n^2 \end{bmatrix};$$

$$\underline{\underline{h}}(t) = \begin{bmatrix} h_1(t) \\ h_2(t) \\ \dots \\ h_n(t) \end{bmatrix}; \quad h_i(t) = h f_i(t) \cdot \frac{a_{i+1}}{a_i}; \quad (i=1,2, \dots, n-1); \quad h_n(t) = 0.$$

Applying $h_i(t)$, we assume that the driver takes into account the follower vehicle behavior. Where “ h ” is a human factor (dimensionless parameter), which is a calibrated parameter according the driver characteristic.

$$s_i = s_{0i} = \text{const.}, \text{ or: } s_i = s_i(\dot{x}_{i-1}, \dot{x}_i) \quad (i=1,2, \dots, n).$$

$$\underline{\underline{f}}_1(\dot{x}(t)) = \begin{bmatrix} \dot{x}_1^4 \\ \dot{x}_2^4 \\ \dots \\ \dot{x}_n^4 \end{bmatrix}; \quad \underline{\underline{f}}_2(x(t)) = \begin{bmatrix} \frac{1}{\varepsilon_1^2 + (x_0 - x_1)^2} \\ \frac{1}{\varepsilon_2^2 + (x_1 - x_2)^2} \\ \dots \\ \frac{1}{\varepsilon_n^2 + (x_{n-1} - x_n)^2} \end{bmatrix}; \quad \underline{\underline{1}}(t) = \begin{bmatrix} 1(t) \\ 1(t) \\ \dots \\ 1(t) \end{bmatrix};$$

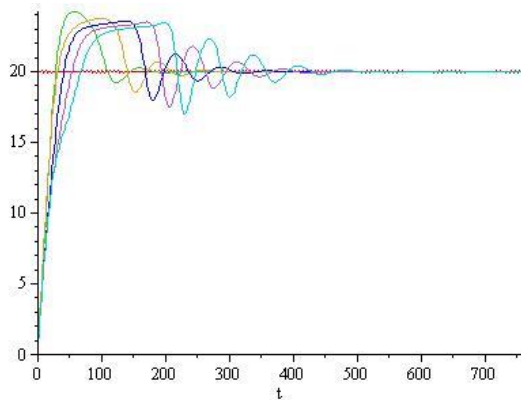


Figure 3

Setting of the stabilized speed state after the starting of a vehicle group

3 Relationship between the IDM Model and the Large-Scale Network

The speed of a given vehicle and the distance kept are determined by the driver. Their decision, however, depends on their own perceptions, on signals that are transmitted by the physical environment and received by the vehicle and on the local and general effects of network traffic [17, 18, 19, 20, 21, 22]. Physical impacts resulting from road quality, meteorological and visibility conditions at a given vehicle density determine a selectable speed range. The modified IDM model discussed in the previous section can be used to describe the dynamic traffic connections originating from forward-moving vehicle-vehicle effects in a given section.

At the same time, the dynamics of the movement of the IDM model group is not arbitrary. It is determined by control speeds formed in the large-scale network or network sections. The vehicles slow down if a congestion occurs, stop when the traffic light switches to red, but after the reaction delay time, they will accelerate to the maximum permitted speed if the road section ahead is free. **This is indicated in Fig. 2 by the control speed function $x_0(t)$ defined by the large-scale macroscopic network processes for each trajectory.**

4 The Impact of the Vehicles' maximal Acceleration Parameter a on the Motion Process during Catch-up to the Leader Vehicle

The following diagrams show simulated route-time and speed-time diagrams for different a acceleration capabilities. In our case, stationary vehicles in the same lane start from different starting points at $t_0 = 0$ and follow the movement of the leader vehicle (leading point).

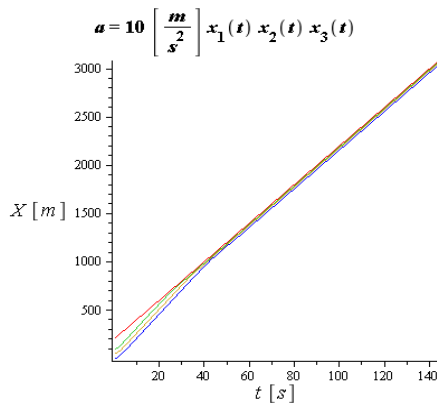


Figure 4

The catch-up motion process at $a=10$ [m/s²]

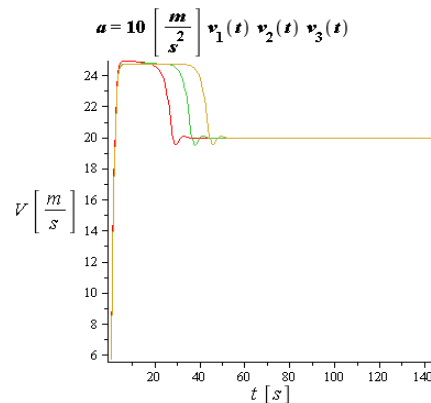


Figure 5

The catch-up speed process at $a=10$ [m/s²]

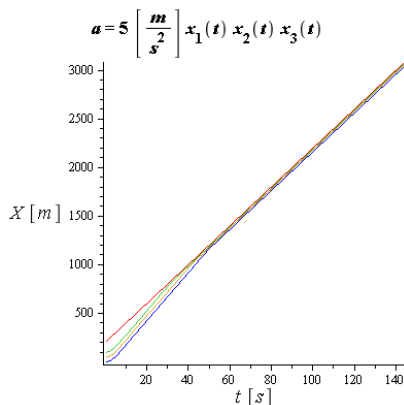


Figure 6

The catch-up motion process at $a=5$ [m/s²]

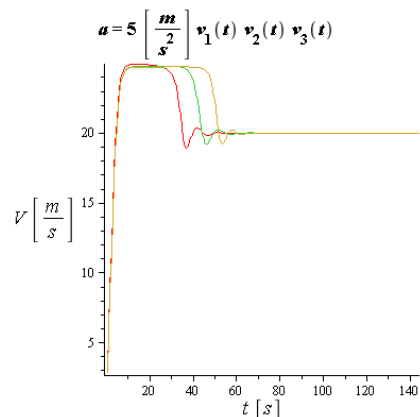


Figure 7

The catch-up speed process at $a=5$ [m/s²]

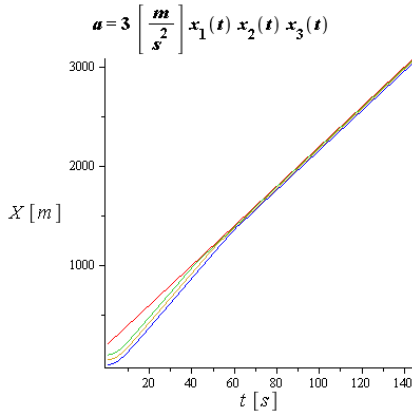


Figure 8

The catch-up motion process at $a=3 \text{ [m/s}^2\text{]}$

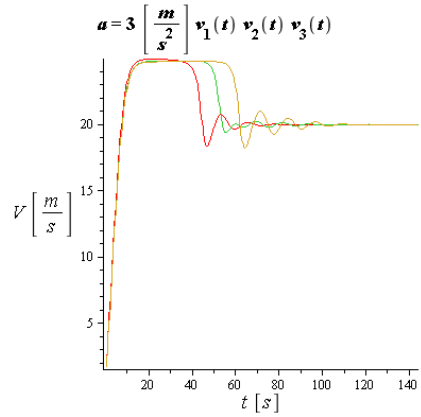


Figure 9

The catch-up speed process at $a=3 \text{ [m/s}^2\text{]}$

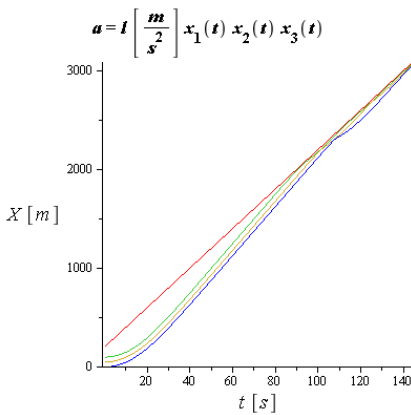


Figure 10

The catch-up motion process at $a=1 \text{ [m/s}^2\text{]}$

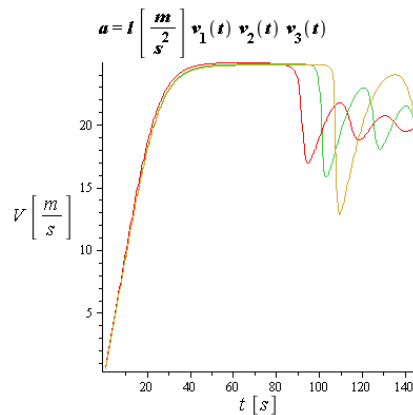


Figure 11

The catch-up speed process at $a=1 \text{ [m/s}^2\text{]}$

The simulations in this case included three successive vehicles from the vehicle group outside the leader vehicle. For vehicles with less acceleration, slower movements can be seen. Acceleration capabilities have an effect on the overshoot, and on the time of speed stabilization. All of these features naturally determine the movements of other participants in the traffic, energy consumption, emissions and traffic noise, so it is worth taking a deeper analysis that will support the smoother traffic processes in the vehicle groups.

5 The Relationship between the Dynamical Characteristics in the Area of Mechanics and the Traffic Parameters

The IDM model investigates the longitudinal movement of vehicle groups in traffic, thus the essence of the analysis is a longitudinal dynamic traffic analysis.

The parameters of the IDM basic system are traffic system parameters with uncertainties. Therefore, let us look at the following interpretation and appropriately transcription of the mathematical model that is useful for further analysis of dynamical properties and for exploring new relationships.

In addition to the following model considerations, the applied model structure can be used to carry out an analysis of a suitable multi-mass dynamical model by using purely dynamics and vibrations concepts and deducting conclusions from these in respect to the original traffic system parameters.

$$\langle \underline{\underline{A}} \rangle^{-1} \ddot{x}(t) + \langle \underline{\underline{V}} \rangle^{-1} \underline{f}_1(\dot{x}(t)) + \langle \underline{\underline{S}} \rangle \underline{f}_2(x(t)) = \underline{1} \quad (6)$$

The system of differential equations (6) contains dimensionless members on the left-hand side because the dimensions at the multipliers are reciprocal to each other. It follows from the above that the vector on the right-hand side is necessarily dimensionless. The numerical values of the solutions of the system of differential equations do not change if the following physical dimensions are applied to each element of the matrix and of the vectors: A^{-1} [kg]; $V^{-1} f_1$ [N], $S f_2$ [N] and $1(t)$ [N]; $x(t)$ [m]; (consequently: $\dot{x}(t)$ for the first derivative [m/s]; $\ddot{x}(t)$ for the second derivative [m/s²]); The importance of this approach is that while the mathematical model is equivalent in the two cases, the physical interpretation is completely different. In the first case, we examine a traffic dynamics model and in the second case a mechanical dynamics one. However, such conclusions can be drawn based on the parameters of the second physical model, that cannot be drawn based on the first model. On that basis in the model (6) the mass matrix $\underline{\underline{M}}$, the non-linear attenuation vector $\psi(\dot{x}(t))$ and the non-linear spring force vector $\varphi(x(t))$ can be interpreted, which defines the classical nonlinear vehicle dynamics model (7). ($\underline{\underline{M}} \in \mathfrak{R}^{n \times n}$; $\psi(\dot{x}(t)) \in \mathfrak{R}^n$; $\varphi(x(t)) \in \mathfrak{R}^n$;

$$\begin{aligned} \underline{\underline{M}} &= \langle \underline{\underline{A}} \rangle; \\ \psi(\dot{x}(t)) &= \langle \underline{\underline{V}} \rangle^{-1} \underline{f}_1(\dot{x}(t)); \\ \varphi(x(t)) &= \langle \underline{\underline{S}} \rangle \underline{f}_2(x(t)); \end{aligned}$$

The elements of the diagonal matrix $\underline{\underline{M}}$ take values of $m_i = \frac{1}{a_i}; (i = 1, 2, \dots, n)$.

$$\underline{\underline{M}}\ddot{\underline{x}}(t) + \underline{\psi}(\dot{\underline{x}}(t)) + \underline{\varphi}(\underline{x}(t)) = \underline{1} \quad (7)$$

The special feature of the IDM system is that if the leading point takes constant v_0 speed, the movement of the members of the vehicle group ($i=1, 2, \dots, n$) is asymptotically set to a stable state:

$$\begin{aligned} \ddot{x}_i(t) &\rightarrow 0 \\ \dot{x}_i(t) &\rightarrow v_i^* = \text{const!} \\ x_{i-1}(t) - x_i(t) &\rightarrow s_i^* = \text{const!} \end{aligned} \quad (8)$$

In formulas (8), v_i^* is the desired exact speed and s_i^* is the desired exact distance for the i th vehicle, which parameters the drivers want to maintain based on the longitudinal dynamics of the vehicles. Since they make mistakes and can differ from these, the above is taken into account with the coefficients $\alpha > 0$ and $\beta > 0$, thus $s_i = \alpha s_i^*$ and $v_i = \beta v_i^*$ are considered to be the actual operating points. The setting of the coefficients is performed during validations, measurements, or using simulation results. Then we examine the movements around the operating point, because in this way system-specific parameters can be derived, which can be interpreted well from previous vehicle dynamics studies and important information can be obtained in the analysis of more complex IDM parameters.

Let us study the attenuation factor k_i , the spring rigidity coefficient S_i and the mass m_i of the i th element of the system linearized around the working point:

$$k_i = \left[\frac{d}{d\dot{x}_i} \left(\frac{\dot{x}_i}{v_i} \right)^4 \right]_{\dot{x}_i=v_i} = \frac{4}{v_i}; \quad (9)$$

$$S_i = \left[\frac{d}{dx_i} \left(\frac{s_i}{x_{i-1} - x_i} \right)^2 \right]_{x_{i-1}-x_i=s_i} = \frac{2}{s_i}; \quad (10)$$

$$m_i = \frac{1}{a_i}; \quad (11)$$

On the basis of (9), (10) and (11) we can introduce the eigenfrequency ω_i and Lehr's relative attenuation factor δ_i around the operating point, which are the following based on the original IDM model parameters:

$$\omega_i^2 = \frac{S_i}{m_i} = 2 \frac{a_i}{s_i}; \quad (12)$$

(13) and (14) applies to the relative attenuation factor δ_i based on the definition and on (9) and (11), respectively:

$$\frac{k_i}{m_i} = 2 \cdot \delta_i \cdot \omega_i \quad (13)$$

$$\frac{k_i}{m_i} = \frac{4}{\frac{1}{a_i}} = 4 \frac{a_i}{v_i} \quad (14)$$

Based on (13) and (14):

$$\delta_i = \sqrt{2} \frac{\sqrt{a_i \cdot s_i}}{v_i} \quad (15)$$

The above formula was determined based on the IDM model parameters. The value of δ can be approximated by simulation or measurement by applying the appropriate logarithmic decrement in a way that the attenuated oscillation process of the velocity function around the axis $v=v_0$ is modelled with the function $\dot{x}_i(t) = A \cdot \cos(\omega \cdot t) \cdot e^{-2 \cdot \delta \cdot \omega \cdot t}$. In this case, we determine the amplitudes $A_1, A_2, A_3, \dots, A_k$, according to local extremes $t_1, t_2, t_3, \dots, t_k$. We also utilize that $|\cos(\omega \cdot t_1)| = |\cos(\omega \cdot t_2)| = |\cos(\omega \cdot t_3)| = \dots = |\cos(\omega \cdot t_k)|$, consider the following arbitrary times $t_i \neq t_j (t_1 \leq t_i < t_j \leq t_k)$ and apply the following formulas:

$$\frac{A_i}{A_j} = \frac{A \cdot \cos(\omega \cdot t_i) \cdot e^{-2 \cdot \delta \cdot \omega \cdot t_i}}{A \cdot \cos(\omega \cdot t_j) \cdot e^{-2 \cdot \delta \cdot \omega \cdot t_j}} = \frac{e^{-2 \cdot \delta \cdot \omega \cdot t_i}}{e^{-2 \cdot \delta \cdot \omega \cdot t_j}} \quad (16)$$

$$\ln A_i - \ln A_j = 2 \cdot \delta \cdot \omega \cdot (t_j - t_i) \quad (17)$$

$$\delta = \frac{\ln A_i - \ln A_j}{2 \cdot \omega \cdot (t_j - t_i)} \quad (18)$$

For practical calculations, the first 3-4 elements of the series $t_1, t_2, t_3, \dots, t_k$ can be used as a result of the rapid decrease in the amplitudes and the increase in the relative measurement error, therefore the amplitude of the highest value and the subsequent ones with acceptable measurement error can be taken into account.

Based on the above, according to the acceleration parameter a , the eigenfrequency $\omega = \omega(a)$ and relative attenuation factor $\delta = \delta(a)$ functions can be determined with a linearization method around the operating point.

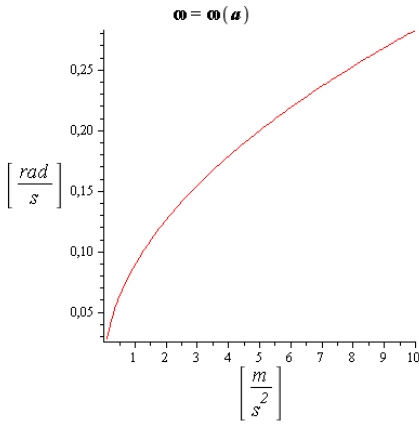


Figure 12

$\omega = \omega(a)$ function calculated based on the IDM model-parameters

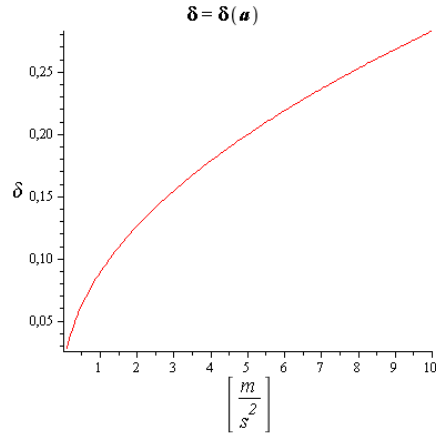


Figure 13

$\delta = \delta(a)$ function calculated based on the IDM model-parameters

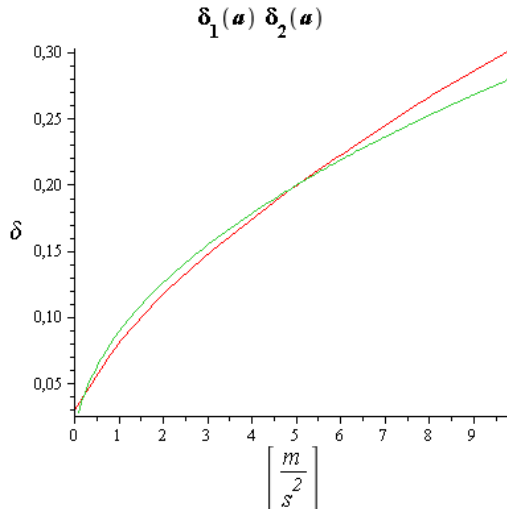


Figure 14

Validation of the relative attenuation function for different acceleration capabilities. The functions $\delta_1 = \delta_1(a)$ and $\delta_2 = \delta_2(a)$ were determined based on the IDM model parameters and on the basis of the logarithmic decrement, respectively

Based on the above, for each vehicle, the relative attenuation factor δ can be defined by a closed formula with the linearization of the nonlinear characteristics of the differential equation system around the operating point. At the same time, attenuation values can be measured in the dynamic vehicle system by examining the attenuating speed processes. Although the phenomenon is non-linear (the oscillation is anharmonic), based on our findings the above linearization can be validated surprisingly well using the logarithmic decrement method. The application of both methods together is very useful and important as it highlights the extent of the actual range to be taken into account in the linearization around the operating point, which means the determination of a related coefficient value. Lehr's relative attenuation factor δ derived from measurement or simulation (the notation D_L is commonly used in the literature) is located in a well-defined range ($1 < \delta < 1$). This physical parameter characterizing the attenuation of the system checks the δ value calculated by linearization during our process. The relative error values calculated for the relative attenuation factors δ during the validation are shown in the table below for the functions seen in Figure 17.

Table 1

a [m/s²]	1	2	3	5	6	8	10
Relative error [%]	8.59	6.65	4.50	0.13	1.77	5.53	7.94

6 The Variation of Relative Attenuation Factor δ at Acceleration $a=5 \text{ m/s}^2$ and Fixed $\omega =0.2$

It can be stated that in the case of a low relative attenuation factor ($\delta=0.16-0.18$) the speed overshoot is high and instead of 20 m/s it can reach up to 30 m/s, therefore, significant braking is needed so that the speed drops close to 0-10 m/s. Thus, the setting of vehicle speeds is very hectic and the speed oscillation can last for 1-1.5 minutes.

As the relative attenuation increases to $\delta=0.2$, the value of the speed overshoot and the setting time of the tracking speed decreases, e.g. at $\delta=0.2$ $v=25$, and the stability time is 1 min. At relative attenuation values $\delta=0.2-0.24$ the speed overshoot further decreases and the speed functions become smooth and oscillation-free, but the stability time starts to increase.

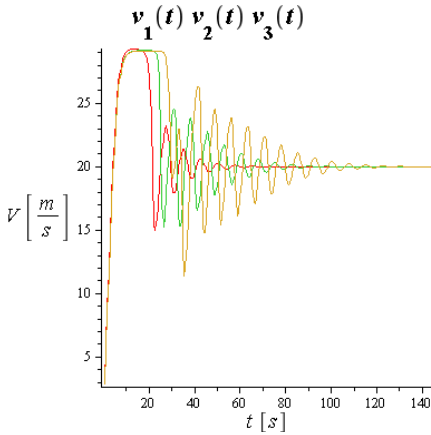


Figure 15
 $a=5 \text{ [m/s}^2\text{] and } \delta=0.17$

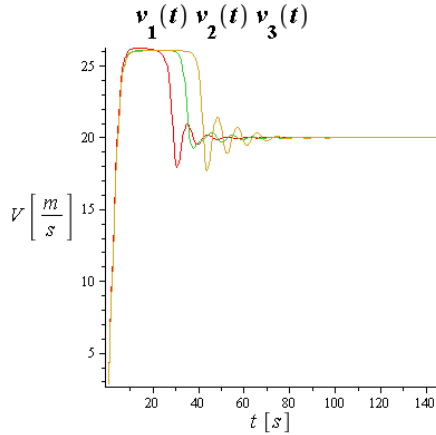


Figure 16
 $a=5 \text{ [m/s}^2\text{] and } \delta=0.19$

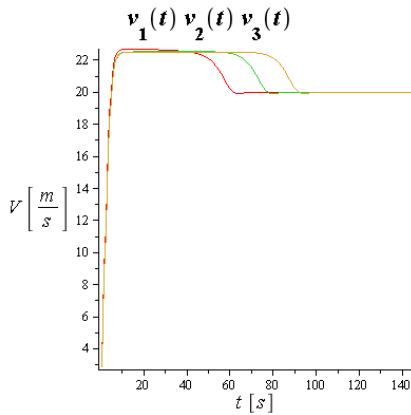


Figure 17
 $a=5 \text{ [m/s}^2\text{] and } \delta=0.22$

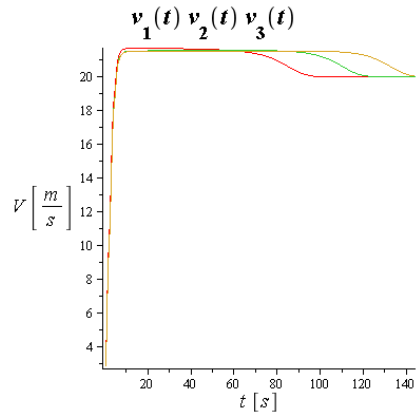


Figure 18
 $a=5 \text{ [m/s}^2\text{] and } \delta=0.23$

7 Variation of ω at Acceleration $a=5 \text{ m/s}^2$ and Fixed Relative Attenuation Factor $\delta=0.2$

It can be stated that in the case of low eigenfrequency range ($\omega < 0.17$) the speed overshoot is high and instead of 20 m/s it can reach up to 29 m/s, therefore, significant braking is needed so that the speed drops close to 13-15 m/s. Thus, the setting of vehicle speeds is very hectic and the speed oscillation can last for 90 seconds. As the eigenfrequency increases to $\omega=0.2$, the value of the speed

overshoot and the setting time of the tracking speed decreases, e.g. at $\omega=0.2$ $v=25$, and the stability time is below 1 min. At eigenfrequency values $\omega=0.2-0.24$, 24 the speed overshoot further decreases ($v=21$) and the speed functions become smooth and oscillation-free, but the stability time starts to increase.

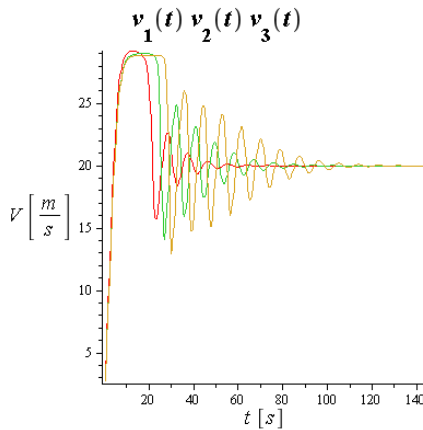


Figure 19
 $a=5$ [m/s²] and $\omega=0.17$

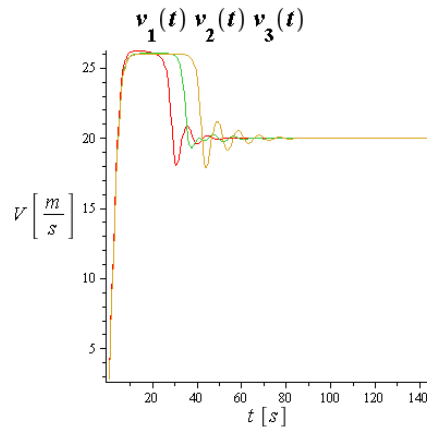


Figure 20
 $a=5$ [m/s²] and $\omega=0.19$

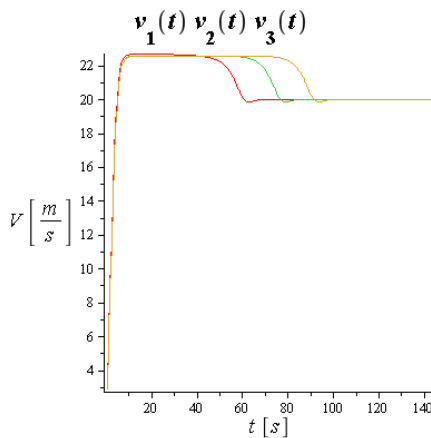


Figure 21
 $a=5$ [m/s²] and $\omega=0.22$

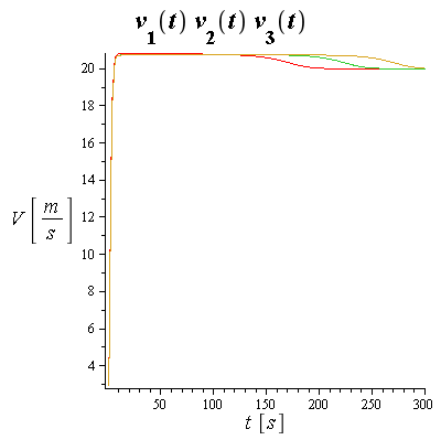


Figure 22
 $a=5$ [m/s²] and $\omega=0.24$

Tables 2 and 3 below show the calculated IDM parameters "s" and "v" related to ω and δ , while the last column shows the amount of specific energy consumption pertinent to the movement of 1 kg mass of the vehicle.

Table 2
At fixed $a=5\text{m/s}^2$ and $\omega=0.2$ rad/s values

δ	s, dist. [m]	v [m/s]	Specific energy [Nm]
0,16	10	31,25	689,31
0,18	10	27,77	673,29
0,20	10	25,00	666,71
0,22	10	22,73	660,50
0,23	10	21,74	656,37

At optimal value $\delta=0.23$ a decrease of 4.779% in the energy consumption of the vehicle group occurred until the stable speed had been reached compared to the one calculated with the initial value $\delta=0.16$.

Table 3
At fixed $a=5\text{m/s}^2$ and $\delta=0.2$ values

ω [rad/d]	s, dist. [m]	v [m/s]	Specific energy Nm]
0,17	13,84	29,41	673,45
0,18	12,35	27,78	670,08
0,20	10,00	25,00	666,71
0,22	8,26	22,73	662,98
0,24	6,94	20,83	632,74

At optimal value $\omega=0.24$ a decrease of 6.045% in the energy consumption of the vehicle group occurred until the stable speed had been reached compared to the one calculated with the initial value $\omega=0.17$.

Conclusions

The parameters of the IDM base-system are traffic system parameters that can cause uncertainties due to the specificity or the hectic nature of traffic processes. This research investigated the vehicle-dynamical properties of the IDM model by the suitably chosen interpretation and transcription of the mathematical model. In order to further analyse the dynamical properties, it has led to the exploration of newer relations. The chosen model structure and the considerations used are suitable for analysing the multi-mass dynamical model using purely dynamics and vibrations concepts. Based on this, one can obtain important information on the more complex IDM system parameter structure and optimization.

The study takes into account the speed of the vehicle to be followed and determines the optimal maximum speed and the optimal distance for each of the specified acceleration capabilities. This affects both the rate of speed overshoot and the setting time of the stable speed state, as well as the optimization of the motion energy. All of these optimum properties have an impact on the other participants in the traffic, on the emissions and traffic noise. This facilitates smoother traffic processes in the vehicle groups and the automation of traffic-

dependent optimum decisions by automatically adjusting optimal parameters in the case of autonomous vehicles [26, 27, 31].

Acknowledgement

The research presented in this paper was carried out as part of the EFOP-3.6.2-16-2017-00016 project in the framework of the New Széchenyi Plan. The completion of this project is funded by the European Union and co-financed by the European Social Fund

References

- [1] Csiszár, Cs., Földes, D. (2018) System Model for Autonomous Road Freight *Transportation, Promet - Traffic&Transportation*, Vol. 30, No. 1, February 2018, pp. 93-103, <https://doi.org/10.7307/ptt.v30i1.2566>
- [2] Derbel, O.; Peter, T.; Zebiri, H.; Mourllion, B.; Basset, M. (2012) Modified intelligent driver model, *Periodica Polytechnica Transportation Engineering* 40(2): 53-60, <https://doi.org/10.3311/pp.tr.2012-2.02>
- [3] Derbel, O.; Peter, T.; Zebiri, H.; Mourllion, B.; Basset, M. (2013) Modified intelligent driver model for driver safety and traffic stability improvement, *IFAC Proceedings Volumes* 46(21): 744-749, <https://doi.org/10.3182/20130904-4-JP-2042.00132>
- [4] Derbel, O., Péter, T., Mourllion B., & Basset M. (2017) Generalized Velocity–Density Model based on microscopic traffic simulation, *Transport*, DOI: 10.3846/16484142.2017.1292950 To link to this article: <http://dx.doi.org/10.3846/16484142.2017.1292950> ISSN: 1648-4142 (Print) 1648-3480 <http://www.tandfonline.com/loi/tran20>
- [5] Farooq, A., Xie, M., Williams, E., Gahlot, V., Yan, D. and Yi, Z. (2018) “Downsizing Strategy for Cars, Beijing for People Not for Cars: Planning for People”, *Periodica Polytechnica Transportation Engineering*, 46(1), pp. 50-57, doi: <https://doi.org/10.3311/PPtr.10851>
- [6] Ghadi, M., Török, Árpád and Táncczos, K. (2018) “Study of the Economic Cost of Road Accidents in Jordan”, *Periodica Polytechnica Transportation Engineering*, 46(3), pp. 129-134, doi: <https://doi.org/10.3311/PPtr.10392>
- [7] Iordanopoulos, P., Mitsakis, E. and Chalkiadakis, C. (2018) “Prerequisites for Further Deploying ITS Systems: The Case of Greece”, *Periodica Polytechnica Transportation Engineering*, 46(2), pp. 108-115, doi: <https://doi.org/10.3311/PPtr.11174>
- [8] Istenes, G., Szauter, F., Rödönyi, G. Vibration analysis of a suspension system subject to high level of measurement noise (2017) *2017 4th International Conference on Control, Decision and Information Technologies (CoDIT)* 5-7 April, 2017 Barcelona, Spain. DOI: 10.1109/CoDIT.2017.8102707

- [9] Jerath, K. (2010) Impact of adaptive cruise control on the formation of self-organized traffic jams on highway. Master's thesis, The Pennsylvania State University The Graduate School. Department of Mechanical and Nuclear Engineering
- [10] Kesting, A., Treiber, M., Helbing, D. (2008) Agents for traffic simulation. *Physics and Society* 11, 325-356
- [11] Koryagin, M. (2018) "Urban Planning: a Game Theory Application for the Travel Demand Management", *Periodica Polytechnica Transportation Engineering*, 46(4), pp. 171-178, doi: <https://doi.org/10.3311/PPtr.9410>
- [12] Kovács, T., Bolla, K., Gil, R., A., Csizmás, E., Fábíán, Cs., Kovács, L., Medgyes, K., Osztyényi, J., Végh A. (2016) Parameters of the intelligent driver model in signalized intersections *Technical gazette*, Vol. 23, No. 5, October 2016, pp. 1469-1474, ISSN 1330-3651 (Print), ISSN 1848-6339 (Online) DOI: 10.17559/TV-20140702174255
- [13] Lakatos, A., Mándoki P. (2017) Quality evaluation of the long-distance bus and train transportation in Hungary. *Transportation Research Procedia*, Volume 27, 2017, pp. 365-372, ISSN: 2352-1465
- [14] Lukács P., (2017) Development of Material and Energetic Usage Solutions for Problematic Fractions Originated from ELV's in Hungary (2017) 17th International Automobile Recycling Congress - IARC 2017, March 22-24, 2017, Berlin, Germany
- [15] Lukács, P., Gombkötő, I., (2014) Critical Elements in the Today's and Future Vehicle Technology In: Gombkötő, Imre (szerk.) 18th International Conference on Waste Recycling Miskolc, University of Miskolc (2014) pp. 1-6, 6 p
- [16] Mihály, A., Németh, B. and Gáspár, P. (2018) "Real-time Look-ahead Cruise Control Simulator", *Periodica Polytechnica Transportation Engineering*, 46(1) pp. 11-16, doi: <https://doi.org/10.3311/PPtr.9896>
- [17] Péter T, and Bokor J. (2010.1) Research for the modelling and control of traffic, In: Scientific Society for Mechanical Engineering, 33rd *Fisita-World Automotive Congress: Proceedings*, Budapest, Hungary, May 30-June 4, 2010, Budapest: GTE, 2010, pp. 66-73 (ISBN:978-963-9058-28-6)
- [18] Péter T, and Bokor J. (2010.2) Modeling road traffic networks for control. *Annual international conference on network technologies & communications: NTC 2010*, Thailand, 2010.11.30-2010.11.30. pp. 18-22, Paper 21 (ISBN:978-981-08-7654-8)
- [19] Peter, Fülep and Bede (2011) The application of a new principled optimal control for the dynamic change of the road network graph structure and the analysis of risk factors, 13th *EAEC European Automotive Congress*, June 13-16, 2011, Valencia – SPAIN Society of Automotive Engineers (STA), 2011, pp. 26-36 (ISBN:978-84-615-1794-7)

- [20] Péter T. and Bokor J. (2011) New road traffic networks models for control, *GSTF International Journal on Computing*, Vol. 1, Number 2. pp. 227-232, DOI: 10.5176_2010-2283_1.2.65 February 2011
- [21] Péter, T. (2012) Modeling nonlinear road traffic networks for junction control, *International Journal of Applied Mathematics and Computer Science (AMCS)*, 2012, Vol. 22, No. 3, pp. 723-732, DOI: 10.2478/v1006-012-0054-1
- [22] Péter T, Fazekas S. (2014) Determination of vehicle density of inputs and outputs and model validation for the analysis of network traffic processes *Periodica Polytechnica Transportation Engineering*, 42:(1) pp. 53-61, (2014) (Budapest University of Technology and Economics)
- [23] Pokorádi, L. (2018) Graph model-based analysis of technical systems *IOP Conf. Series: Materials Science and Engineering* 393 (2018) 012007, pp. 1-9, doi:10.1088/1757-899X/393/1/012007
- [24] Pokorádi, L., (2018) Methodology of Advanced Graph Model-based Vehicle Systems' Analysis In: Szakál, Anikó (ed.) IEEE 18th International Symposium on Computational Intelligence and Informatics (CINTI 2018) Budapest, IEEE Hungary Section (2018) pp. 325-328, 4 p.
- [25] Pokorádi, L., Gáti, J., (2018) Markovian Model-based Sensitivity Analysis of Maintenance System In: Anikó, Szakál (ed.) IEEE 16th International Symposium on Intelligent Systems and Informatics : SISY 2018 Budapest, IEEE Hungary Section (2018) pp. 117-121, 5 p.
- [26] Rövid, A., Szeidl, L., Várlaki, P., (2014) Reconstruction of Inner Structures Based on Radon Transform and HOSVD pp. 311-319, In: János, Fodor; Robert, Fullér (eds.) *Advances in Soft Computing, Intelligent Robotics and Control* New York, London, Heidelberg, Springer, (2014)
- [27] Rövid A., Szeidl L., Várlaki P., (2015) Integral Operators in Relation to the HOSVD-Based Canonical Form *ASIAN JOURNAL OF CONTROL* 17 : 2 pp. 459-466, 8 p. (2015)
- [28] Takács, Á., Drexler, D., A., Galambos, P., Rudas, I., J., Haidegger, T. (2018) Assessment and Standardization of Autonomous Vehicles 2018 *IEEE 22nd International Conference on Intelligent Engineering Systems (INES)* 21-23 June, Las Palmas de Gran Canaria, Spain, pp. 185-192, ISSN: 1543-9259, DOI: 10.1109/INES.2018.8523899
- [29] Szabó, G., Szabó, K. and Zerényi, R. (2004) Safety Management Systems in Transportation: Aims and Solutions. *Periodica Polytechnica Transportation Engineering* 32(1):123-134
- [30] Szabó K., Szabó G., Renner P. (2009) Emberi hibamodellezés alkalmazása a légiközlekedési kockázatelemzésekben *Közlekedéstudományi Szemle* 59:(5) pp. 29-35 (2009)

- [31] Szalay, Z., Tettamanti, T., Esztergár-Kiss, D., Varga, I. and Bartolini, C. (2018) "Development of a Test Track for Driverless Cars: Vehicle Design, Track Configuration, and Liability Considerations", *Periodica Polytechnica Transportation Engineering*, 46(1), pp. 29-35, doi: <https://doi.org/10.3311/PPtr.10753>
- [32] Treiber, M., Helbing, D. (2002) Realistische mikrosimulation von straenverkehr mit einem einfachen modell. In: *Symposium "Simulationstechnik ASIM*
- [33] Treiber, M.; Hennecke, A.; Helbing, D. (2000a) Congested traffic states in empirical observations and microscopic simulations, *Physical Review E* 62(2): 1805-1824, <https://doi.org/10.1103/PhysRevE.62.1805>
- [34] Treiber, M.; Hennecke, A.; Helbing, D. (2000b) Microscopic simulation of congested traffic, in D. Helbing, H. J. Herrmann, M. Schreckenberg, D. E. Wolf (Eds.). *Traffic and Granular Flow'99: Social, Traffic, and Granular Dynamics*, 365-376, https://doi.org/10.1007/978-3-642-59751-0_36
- [35] Treiber, M., Hennecke, A., Helbing, D. (2004) Microscopic simulation of congested traffic. *Physical Review E* 62, 1805-1824
- [36] Treiber, M., Hennecke, A., Helbing, D. (2006) Delays, inaccuracies and anticipation in microscopic traffic models. *Physica A* 360, 71-88
- [37] Treiber, M., Kesting, A. (2011) Evidence of convective instability in congested traffic flow: A systematic empirical and theoretical investigation. *Procedia Social and Behavioral Sciences* 17, 698-716

Systematic Overview of Password Security Problems

Viktor Taneski, Marjan Heričko, Boštjan Brumen

Faculty of Electrical Engineering and Computer science, University of Maribor,
Koroška cesta 46, 2000 Maribor, Slovenia

viktor.taneski@um.si, marjan.hericko@um.si, bostjan.brumen@um.si

Abstract: Alphanumeric passwords are the first line of defense in security for most information systems. Morris and Thompson identified passwords as a weak point in an Information System's security, 35 years ago. Their findings showed that 86% of the passwords were too short, contained lowercase letters only, digits only, were easily found in dictionaries and/or easily compromised. The objective of this paper is to perform a systematic literature review in the area of passwords and passwords security, in order to determine whether alphanumeric passwords are still weak, short and simple. The results show that only 42 out of 63 relevant studies propose a solid solution to deal with the identified problems with alphanumeric passwords, but only 17 have statistically verified it. We find that only 3 studies have a representative sample, which may indicate that the results of the majority of the studies cannot be generalized. We conclude that users and their alphanumeric passwords are still the "weakest link" in the "security chain". Careless security behavior, involving password reuse, writing down and sharing passwords, along with an erroneous knowledge concerning what constitutes a secure password, are the main problems related to the issue of password security.

Keywords: authentication; password security; password security problems; systematic literature review

1 Introduction

The rapid growth of the Internet technology and the widespread use of the World Wide Web (WWW) has changed the way people operate nowadays. The increased number of online services, online social networks (e.g. Facebook, Twitter, etc.) and other websites that have content that is tailored to the users' interests, has increased the need for authentication mechanisms. Authentication is the core of today's Web experience [1]. Online services, social networks and websites require an authentication so that users can create a profile, post messages and comments, and tailor the website's content so it can match their interests.

In an information security sense, authentication is the process of verifying someone's identity. Typically, authentication can be classified into three main categories: *knowledge-based authentication* - "what you know" (e.g., textual or graphical passwords), *biometrics authentication* - "what you are" (e.g., retina, iris, voice, and fingerprint scans), and *token-based authentication* - "what you have" (e.g., smart cards, mobile phones or other tokens). Lately, another alternative authentication method is becoming more available - the two-step verification. The problems with these alternative authentication methods are not related to the security itself, in fact, these methods also provide excellent security for the system. Instead, the weaknesses of these authentication methods are that they can be expensive (biometrics, smart cards), they must be carried around at all times when access to the system is required (smart cards, two-step verification), they are difficult to implement on a large scale, and they are not widely accepted by the users. Single Sign-On (SSO) is another method for authentication that is recently becoming more available, that provides access to many resources once the user is initially authenticated. However, a recent study [2] found that SSO solutions impose a cognitive burden on web users, and users have significant trust, security, and privacy concerns, which hinders the wide acceptance and usage of SSOs.

We focus on the textual passwords and their security simply because the username-password combination used to be [3] [4] and still is the most widely used method for authentication [5]. Even though passwords suffer from a number of problems, they continue to be one of the most common control mechanisms to authenticate users in information systems, due to their simplicity and cost effectiveness. The problems related to textual passwords and password security are not new. Morris and Thompson [6] were first to identify textual passwords as a weak point in information system's security. More than three decades ago, they conducted experiments about typical users' habits about how they choose their passwords. They reported that many UNIX-users have chosen passwords that were very weak: short, contained only lower-case letters or digits, or appeared in various dictionaries. Zviran and Haga [7] had similar findings in their study conducted 20 years later. They came to the conclusion that users are one of the biggest threat to information system's security. In their study, almost half of the users created passwords composed of five or fewer characters, 80% had only alphanumeric characters, and 80% never changed their password.

The objective of this paper is to perform a systematic literature review of studies related to textual passwords and textual passwords security. There are three reasons for conducting the review in this specific field. The first reason is to identify any problems that may arise in creating or managing textual passwords. The second reason is to assess the current situation of passwords with respect to password strength, password management and password memorability. Finally, the third reason is to find out whether the users are still considered the "the weakest link?" in information security.

The paper is based on our previous work [8] where we presented the preliminary results of our systematic literature review. We extend our preliminary work by including additional papers that were published in the period from 2014 to 2018. We improve our systematic literature review in a more detailed and strict manner. We also perform quality assessment for the relevant studies and categorize the data extracted from the studies in order to answer our research questions more effectively.

2 The Review Methodology

A systematic literature review (SLR) is “a means of identifying, evaluating and interpreting all available research relevant to a particular research question, or topic area, or phenomenon of interest.” [9]. Most research studies begin with the process of literature review. The extent and the properties of the review are not necessarily fully consistent with the research methodology of systematic literature review. If a literature review is not conducted in a thorough and proper way, its scientific value is low. To this end, we need a systematic literature review carried out in accordance with a pre-defined search strategy. When performing our systematic literature review, we took in consideration the guidelines by Kitchenham and Charters [9] for performing SLR in software engineering. These guidelines propose carrying out the systematic literature review in the form of three major phases: *planning the systematic literature review*, *conducting the review* and *reporting the review (presenting the results)*. The tasks performed in each phase are described in more detail in the following subsections.

2.1 Planning The Systematic Literature Review

A review protocol should be defined prior to conducting the systematic literature review in order to reduce the researcher’s bias [9]. The protocol prescribes pre-review activities and, in our case, includes *defining the research questions*, *defining the search strategy*, *defining the study selection procedure*, *defining the quality assessment checklist* and *data extraction strategies*.

2.1.1 Research Questions

We used a modified version of the Population, Intervention, Comparison, Outcomes, Context (PICOC) [9] [10] structure, in order to construct well-formulated research questions. This structure contains the attributes that can help us define the research questions. The *population* is represented by textual passwords, while the *intervention* is represented by different approaches (methods, strategies or techniques) used for creating and managing textual passwords. The

attribute *comparison* is not applicable, because we do not compare textual passwords with other types of passwords, since our subject of interest are strictly textual passwords. The *outcome* refers to the problems that may arise when creating and using (or managing) textual passwords. The *context* are the user-selected textual passwords and the relationship between the users and their textual passwords.

Considering the above structure, we formulate the following research questions:

- **RQ1:** What are the major problems with creating and managing textual passwords?
- **RQ2:** What is the current situation of textual passwords with respect to the password strength, password management and password memorability?
- **RQ3:** What is the relationship between users and textual passwords?
 - **RQ3.1:** Are the users still “the weakest link”?

We defined the RQ1 to get a better view if the past, and already known problems related to creating and managing textual passwords, still exist today. Please note that the term “managing” is referred to the way users use and store their passwords, and manage the aspects surrounding it (e.g. how often do they change it, do they reuse the password on several other services and accounts etc.).

In addition, we are interested in the current situation of textual passwords with respect to the password length, password management and password memorability. This issue is covered by RQ2 and helps us assess the current situation of textual passwords so we can make a comparison with earlier findings in this area.

Furthermore, RQ3 helps us assess the current relationship between the users and their textual passwords. Users were already identified earlier as “the weakest link” because they used very weak passwords (short, contained only lower-case letters or digits, or appeared in a dictionary). Confirming that this statement still holds, combined with the answers to the RQ2, can help us outline directions for future research that can be used in aiding the users when selecting and managing their textual passwords.

2.1.2 Search Strategy

The important phases of our literature search strategy are: *initial search* and *reference search*. The initial search was performed over digital libraries. When selecting the digital libraries, we followed the recommendations in [9]. We also took into account our knowledge and practical experience, and the fact that we do not have access to *all* digital databases. We carried out the search through the following databases: *IEEEExplore*, *ScienceDirect*, *SpringerLink* and *ACM Digital*

Library. The Google Scholar database was also considered, but was not included. Later in the paper we will explain the reasons for not including this database. We used these databases since they provide for many of the leading publications in the Computer Science field. Furthermore, these databases allow searching by keywords. We restricted the initial search to articles in journals, conference papers and books/book chapters written in English that were peer-reviewed and published since 1979, i.e., the year when the first article in the area of password security was published. We conducted this search in 2018. Therefore, this systematic literature review includes studies that were published before and including 2018.

We composed a search string for searching through the digital databases. The search string contains major search terms from our research questions connected by using Boolean OR:

*(“password security” OR “password strength” OR “password memorability”
OR “password cracking” OR “password management”).*

During the search of the digital databases it was necessary to slightly modify the search string and to modify it in such a way so it could fit the syntax requirements and capability of the search engine of each digital database used.

After the completion of the initial search, we performed a reference search by reviewing the reference lists of studies found in the previous step in order to identify additional studies that are relevant to our review.

2.1.3 Study Selection Procedure

We performed the search by using the search string and the search result was a set of documents in which the search string appears partially or entirely. We excluded irrelevant studies and publications, and select those that are relevant to our study and may very likely provide answers to our research questions. We systematically selected the relevant studies by applying the following steps:

1. We examined the paper titles and excluded the papers and publications that were clearly irrelevant to our search focus.
2. We examined the abstracts and keywords in the remaining studies to select relevant studies.
3. For filtering the remaining studies, we used inclusion and exclusion criteria given in Table 1. To carry out the selection in an objective manner and to reduce the likelihood of bias, we defined the inclusion and exclusion criteria during the definition phase of the review protocol (with a possibility of later adjustments during the search).

The titles and abstracts of the documents do not always provide clear information whether the document meets the specified criteria. If this was the case, we took a further step and read the whole document to determine whether it meets the inclusion and exclusion criteria, which is presented in Table 1.

Table 1
Inclusion and Exclusion Criteria

Inclusion criteria	Exclusion criteria
1. Studies that focus mainly on textual passwords	1. Studies that focus on graphical passwords or any other type of user authentication (biometrics, tokens or smart cards, etc.)
2. Studies that focus on password security	2. Studies that deal with computer security or cryptography in general
3. Studies that present method(s) for password creation	3. Studies that are not peer-reviewed
4. Studies that deal with issues or problems with password use, password management or password memorability	

2.1.4 Quality Assessment Checklist

The quality assessment is a means of weighting the importance of the relevant studies and relates to the extent to which the study minimizes bias [9]. We evaluated the quality of the selected relevant studies and we based our quality assessment on a *quality instrument* which is a checklist that needs to be evaluated for each relevant study. Our quality assessment checklist comprises of three main questions, each of which directly corresponds to one of our main research questions. The questions are answered with 'Yes', 'No' or 'Partially' to which values '1', '0' or '0.5' are assigned, respectively. Each of the quality assessment questions also contains additional sub-questions. The scores for the sub-questions are divided so that the overall score of each question would range between '1' (very good) and '0' (very poor). For example, the first question has four sub-questions, which can be answered with 'Yes', 'No' or 'Partially' for which values '0.25', '0' or '0.125' are associated, respectively. The three quality assessment questions are:

1. Does the study address any problem related to creating or managing textual passwords?
 - a. Is it clear what problem is identified?
 - b. Is the problem clearly defined?
 - c. Is there a solution proposed for solving the identified problem?
 - d. Is the proposed solution experimentally or statistically verified?
2. Is the approach towards acquired data for password strength, password management or password memorability sound?
 - a. Is the data retrieved through a questionnaire?
 - b. Is the sample size known?

- c. Is the sample representative?
 - d. Is the data retrieved through an experiment?
 - e. Is the experiment set in a realistic setting using real data?
3. Does the study address the issue of users being “the weakest link”?
 - a. Is the approach that addresses the issue well-defined?
 - b. Is there a proposed solution for improving the relationship between users and their textual passwords?

The academic studies about password security and usability can be divided into two major categories: studies of real world passwords (e.g. leaked/cracked password lists like the RockYou or MySpace password databases) and user studies [11]. Furthermore, the most common choices for a user study are *online study (in a form of an online survey)* and *laboratory studies* [11]. When it comes to such password studies an important issue is the *ecological validity*. Ecological validity refers to whether or not the findings of a research study are able to be generalized from observed behavior in the laboratory to real-life settings [12] i.e. do the participants of the study behave the way users would in real life. Ecological validity is very important in user studies, since it is believed that the description of the study can influence user behavior from the beginning of the study. The authors in [11] explored the impact user study setups actually have on the ecological validity of these studies. They came to a conclusion that participants are biased and their behavior changes due to the fact that they are participating in a password study.

The terms “experiment”, “realistic setting”, and “real data” are closely related to the context of ecological validity. In our case, the term “experiment” can be defined as a laboratory study where users are not in their natural environment or an analysis done over leaked password lists. The term “realistic setting” can be defined as an environment where users are not aware that they are being studied (e.g. at home, at work etc.). The last term “real data” represents real world passwords that users are using in their everyday life.

An experiment about users’ passwords, conducted over real world passwords (real data) or in an environment where users are not aware that they are being studied (realistic setting), can significantly reduce the potential bias. Furthermore, a combination of an experiment conducted in a realistic setting, using real password data, with a survey can additionally increase the value and the quality score of the study.

2.1.5 Data Extraction Strategy

In addition to the quality assessment check list, we need to extract relevant information from the selected studies for answering the research questions. To that end, and also to make sure that the task is performed in an accurate and consistent

manner, we used a data extraction form based on the research questions. Again, in order to prevent bias, it is important that this form is defined during the definition phase of the review protocol. Table 2 shows the data extraction form used for retrieving relevant data from the selected studies.

Table 2
The Data Extraction Form

Data Item	Description
Basic information about the study	
Title	The title of the study
Author(s)	The author(s) of the study
Venue	Venue where the study is published
Type	Type of the article (journal/conference/book section)
Year	Year of publication
Analysis of the abstract	
Problem	The problem statement in the abstract
Idea	The idea of the paper described in the abstract
Research data	
Domain	The domain of textual passwords
Research methodology	The research methodology used for retrieving the results (experiment, questionnaire or both)
Sample	The type and the size of the sample (if there is one)
Realistic setting	Is the survey or the experiment conducted in a realistic setting using real data?
Identified problems	Identified problems related to password use, password management or password memorability
Proposed solutions	Proposed solution for the identified problems
Interpretation of results	
Conclusions	Conclusions and findings from the research
Main results	Main results of the research
Future work	Future work stated in the study

2.2 Conducting The Review

2.2.1 Initial and Reference Search Results

Table 3 lists the results from the initial (keyword based) search. The first column represents the electronic databases. The second column shows the number of studies found in each database, while the number of studies that have already occurred in another digital database is presented in the third column. The last column shows the total number of relevant studies (excluding the duplicates). The search resulted in relevant studies published in journals, conference proceedings

or book titles. When selecting the relevant studies, we used the predetermined and detailed inclusion/exclusion criteria, presented in subsection 2.1.3 Study Selection Procedure. The study selection in the initial search, in our case, was performed by one of the co-authors. After selecting the relevant studies from the initial search, the co-author consulted and discussed included and excluded papers with the other co-authors. We acknowledge that there is still a possibility of researcher bias in the process of study selection.

Table 3
Summary of Found and Selected Studies

Electronic database	Found studies	Duplicates	Relevant studies
IEEEExplore	192	20	27
ScienceDirect	88		9
SpringerLink	1676		17
ACM Digital Library	144		29
Total	2100	20	82

The initial search found a total of 2,100 candidate studies. We first examined all 2,100 studies by reading the paper titles and removed studies that were unrelated to our research focus. The next step included reviewing the abstract of each remaining paper to exclude additional studies that are not relevant to our research. In some situations, the abstract did not provide enough information to determine whether a study is relevant to our research. In this case, we reviewed the introduction and the conclusions of the article, as well. Next, we examined the content of the remaining studies by reading the whole documents, and filtered them by applying the inclusion and exclusion criteria. After applying these three steps, 101 studies remained. As we were searching through different search engines, we encountered some duplicates, i.e., studies that already appeared in more than one digital database. In this sense, 20 studies were excluded because they were duplicates. In the end, we have **82** relevant studies from the search into 4 electronic databases. Table 3 shows that the ACM Digital Library contributed the largest number of relevant studies (35.36%), while ScienceDirect contributed the smallest number of relevant studies (11.11%). Figure 1 shows the distribution of published relevant studies per year.

We identified the relevant studies that provide the needed information for answering our research questions through a keyword based search in four digital databases. When using keyword based search, there is a potential risk of incomplete identification of relevant studies. There is a possibility that there are some relevant studies that do not explicitly mention the keywords that we use. Hence, there is always a risk that we might have missed some relevant studies during the initial search. When performing the initial search for relevant studies, we did not include the Google Scholar digital library because of a few reasons. The first reason is that Google Scholar only supports keyword search by title and full text and does not support keyword search within paper keywords and abstract.

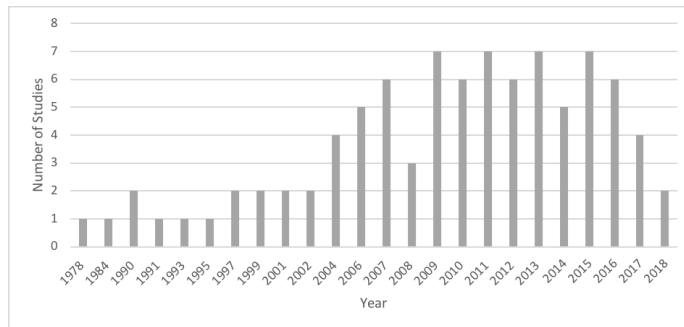


Figure 1

Distribution of relevant studies published per year

The second reason is that this increases the number of hits and complicates the selection of relevant papers. The third reason for not including this digital library is that we performed a test search by using our search string, and while reviewing the first 10 pages (according to a study made in 2013 the first 10 pages in Google search results are the most significant and the most visited [13]), we came to a conclusion that they were containing articles that we have already found in other digital libraries, that we have included in our study. We found no additional relevant papers with this search.

In addition, the search of the references of the primary studies found 8 additional articles that were not found by the initial search of the electronic databases. Thus, we used a total of **90** articles as relevant studies for our systematic literature review.

2.2.2 Quality Assessment

We performed quality assessment on the relevant studies based on three quality assessment questions presented in subsection 2.1.4. Due to the insufficient or unclear data in some of the papers (e.g., unknown sample size, unclear procedure for performing the study, unclear whether real data are used when performing the experiment or whether the experiment is set in a realistic setting, etc.), there is a slight possibility of bias when answering these quality assessment questions. This bias can later affect the data extracted from the relevant studies. In order to reduce the likelihood of bias, when fulfilling the quality assessment checklist and assessing the relevant studies, and to provide better quality assessment results, the quality assessment task was performed by two independent evaluators. The results were statistically compared in order to find out if there is an inter-rater agreement between the two evaluators. We can achieve this if we compare the scores of the two evaluators using the Kappa coefficient, so we can find out if there is any inter-rater agreement. Kappa coefficient measures inter-rater agreement for qualitative (categorical) items and takes into account the agreement occurring by chance.

Since there were only two evaluators/ratters per subject we used the Cohen's kappa instead of Fleiss' kappa, which is used in case there are more than two evaluators/ratters [14] [15].

Each evaluator assessed each of the 90 relevant papers by answering the quality assessment questions. This resulted in two sets of scores, which come from the same pool of relevant studies. The two evaluators/ratters were independent (i.e., one ratter's judgement did not affect the other ratter's judgement) and physically apart from each other. With these precautions we removed the potential for bias from the quality assessment evaluation as much as possible.

The statistical comparison of the quality assessment scores of the two reviewers gives us the following Kappa Coefficients: $k = 0.806$ ($p < 0.0005$), $k = 0.844$ ($p < 0.0005$), and $k = 0.796$ ($p < 0.0005$) for each quality assessment question pair respectively. These coefficients show that there is excellent agreement beyond chance [15]. Furthermore, since $p \leq 0.000$ (or $p < 0.0005$), our kappa (k) coefficient is statistically significantly different from zero.

The further classification of the relevant studies will follow the organization of the quality assessment. It is important to note that we only used the average quality assessment scores to organize the relevant studies, which are suitable for answering each research question, into tabular form so as to provide concrete and concise answers to our research questions. Our intention is not to objectively assess the quality of the studies, or in any other way to criticize any of the studies, since that is not the purpose of this research.

3 Results

In this section, we provide answers to our research questions, defined in Section 2.1.1. We took a comprehensive analysis of the relevant papers to extract the necessary data from the selected relevant studies. The data extracted and used to answer each research question is organized in a tabular form. Only the studies that are associated with a score higher than '0' (quality score '0' means that the study is not relevant for answering the corresponding research question) for the corresponding quality assessment question are taken into account. The studies are sorted by multiple criteria: a descending order with respect to their average quality score and an ascending order with respect to the year that the study was published. For every research question, we present a summary of the results and a discussion.

3.1 RQ1: What are the major problems with creating and managing textual passwords?

In order to answer this research question, we analyzed the relevant studies regarding the identified problems and proposed solutions for those problems. 27 relevant studies have a quality score of '0.5' or lower. Such studies are further not taken into account, since provide an incomplete answer to this research question because the addressed problems are either not clearly defined in the study or there is no proposed solution.

We analyzed the relevant studies to identify the most common problems and most common proposed solutions related to creating and managing textual passwords. We identified 11 different categories of problems related to creating and managing textual passwords: *Human limitations*, *Multiple passwords*, *Weak passwords*, *Password reuse*, *Information overload*, *Password writing down*, *Users lack security knowledge*, *Strong password policies*, *Password sharing*, *Poor password management*, *Outdated password strength metrics*. The studies that 1.) do not belong to any of these 11 categories, or 2.) neither identify a common problem nor specify a solution, are classified under the category *Other*. Due to the nature and the interrelationship of the problems related to creating and managing textual passwords, some problems were address by multiple studies.

Figure 3 shows the number of studies for each identified category of the most common problems.

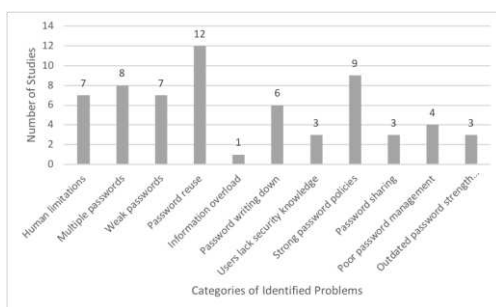


Figure 2

Identified categories of most common problems

We found that almost all (86 out of 90) relevant studies address a problem related to creating or managing textual passwords. The most common problems are related to *password reuse* and are addressed by 12 out of 63. Reusing the same password for more than one account can cause serious damage and can compromise other accounts in the system. Enforcement of *strong password policies* is the second most common problem and is addressed by 9 out of 63 studies. The problems related to users having *multiple passwords* to maintain are

addressed by 8 out of 63 studies. Further down the list are the problems related to users choosing *weak passwords* (7 out of 63 studies), *human limitations* (7 out of 63 studies), users *writing their passwords* down (6 out of 63 studies), *poor password management* (4 out of 63), *password sharing* (3 out of 63 studies) and *outdated password strength metrics* (3 out of 63), the problems related to *users lacking security knowledge* (3 out of 63 studies) and the *information overload* (1 out of 63 studies) as a reason for users having problems to remember and manage all their passwords.

After identifying the most common problems related to creating and managing textual passwords, we went further analyzing whether those studies have proposed some solution for coping with the identified problems. The evidence from these relevant studies helped us identify the most common solutions proposed by the reviewed studies, for creating and managing textual passwords. By the studies we identified 13 different categories of proposed solutions related to creating and managing textual passwords: *Mnemonic passwords*, *Password meter / password rule presentation*, *Cognitive passwords*, *Proactive password checker*, *A “user-centered” approach*, *Password policy*, *Persuasive technology*, *Information security training*, *Associative passwords*, *Password manager*, *New password strength metrics*, *New password security scheme*, *Recommendations*. The studies that do not belong to any of these 13 categories are classified under the category *Other*. Figure 3 shows the number of studies for each category of most commonly proposed solutions.

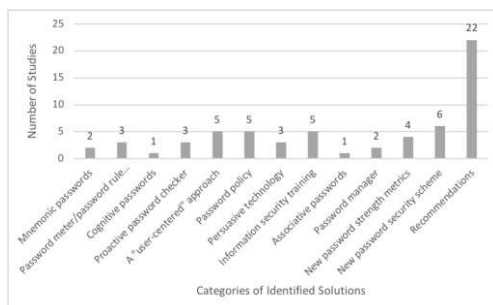


Figure 3

Identified categories of most common solutions

We found that 42 out of 63 relevant studies propose a solid solution for an identified problem related to creating or managing textual passwords. The most common solutions are *a new password security scheme* (addressed by 6 out of 63 studies). A *“user-centered” approach* (addressed by 5 out of 63 studies) and *password policies* (5 out of 63 studies) are the second most common proposed solutions. Following are *information security training* (5 out of 63), and *new password strength metrics* (4 out of 63) as an approach for encouraging users to choose stronger passwords. Further down the list are *password meter or other*

presentation styles (addressed by 3 out of 63 studies), *proactive password checker* (3 out of 63), and *persuasive technology* (3 out of 63) also proposed as solutions for encouraging users to choose stronger passwords, *mnemonic passwords* (addressed by 2 out of 63 studies), *password manager* (2 out of 63), and finally *cognitive passwords* (1 out of 63), and *associative passwords* (1 out of 63). Only 17 of these studies have statistically or experimentally verified their proposed solutions. 22 studies propose a set of (unverified) recommendations, while 26 out of 90 propose neither a solution nor a recommendation.

3.1.1 Discussion on Research Question 1

Morris and Thompson [6] are the first authors that addressed the issue of password security in 1979. The goal of their experiments was to determine typical users' habits related to choosing passwords. They found that users very often choose short and simple passwords that are constructed from a restricted character set (e.g., alphanumeric password with all lower-case letters) and can be found in dictionaries. To increase the difficulty of password cracking and to prevent the fast, simple attacks it is important that systems implement certain password policies that will require the passwords to contain a certain amount of entropy [16].

Unfortunately, users do not always comply with password policies. Basically, human limitations are one of the most common problems related to information security, that is still addressed today [17]. Research has proven that despite the recommendations by information system professionals and their efforts to educate users about secure password policies, users still tend to choose weak passwords that are easy to remember. Very often these passwords are based on user's personal data, or a combination of meaningful details [7]. These problems can be related to users' lack of security motivation and understanding of password policies. In our previous work [18] we performed a questionnaire where we tried to motivate and educate users about frequent password change. We analyzed the effect of password security training on user's practices regarding password creation, frequent password changes and their consciousness about security and the importance of creating strong and hard-to-guess passwords. We found that educating users about password security and assisting them with creating secure passwords can raise the security consciousness of system users and can help achieve greater security. Unfortunately, in order to provide better password security, some security systems incorporate stricter password policies that are forcing users to create stronger passwords with higher amount of entropy. Entropy is a two-edged sword, since higher entropy increases the difficulty of the user to memorize a password [19]. One should be careful with creating a password policy. One "side-effect" of strong password policies is that users tend to circumvent password restrictions for the sake of convenience [20]. Another "side-effect" of strong password policies noted in the literature is users writing down their passwords [21] [22]. Furthermore, due to the increased number of online

services requiring password based authentication, the number of different passwords for different accounts that one user has to maintain is increasing [23]. We expect users to follow common recommendations that say different passwords should be used for every account in order to prevent their other accounts and the accounts of other users in the system to be compromised [24]. Such recommendations are not in line with the issues related to human limitations that we mentioned earlier. This results in the users having many different accounts to maintain and many different passwords to remember, and tend to reuse their passwords.

One of the most common vulnerability related to password security is the password reuse (using the same password or a very similar one for multiple accounts or secure items) [21] [25]. In a system where one password is used for authentication for more than one different accounts, password reuse can cause serious damage, if the password is successfully cracked for a single (not-so-important) account. Information may be revealed that can aid the hackers in infiltrating into other accounts (including the ones that are far more secure than the first one) [24].

Since there is 35 years of research in this area, we were expecting that a proper and useful solution for solving the trade-off between memorability and security would have already been found. The above results give evidence that this is not the case. Contrary to our expectations, we find that only about half (42 out of 90) of the relevant studies propose a solid solution for better coping with the identified problems with textual passwords. Most of the relevant studies propose recommendations for better password creation and password management. We summarize the most common recommendations to the following:

- A secure password should not appear in dictionaries, should not be too short and should not contain personal data
- The use of special characters is strongly advised in order to increase the password security
- Some of the studies recommend a strategy which consists of creating different passwords suitable for different accounts regarding their level of security (e.g. simpler passwords should be used for accounts that contain less important information)
- Associative passwords (i.e. passwords based on associations) combined with guidelines for categorized passwords can ease the construction of strong and easy-to-remember passwords
- The use of enterprise single sign-on is advised as a coping mechanism with password overload and eliminates the need for users to remember multiple passwords

Overall, this is a very low number of proposed solutions, given the fact that these problems have been known for some time now. Furthermore, in the next section we present that only a small number of these studies have verified their solutions. Despite all these solutions and recommendations, there is no common solution proposed or verified by the academic world, and businesses are far from a standardized solution.

3.2 RQ2: What is the current situation of textual passwords with respect to password strength, password management and password memorability?

We searched for attributes in our relevant studies that can help us assess the current situation of textual passwords with respect to their strength, management and memorability. We were interested in what type of research methodology is used (e.g., questionnaire, experiment), whether a realistic setting and real data are used and whether the sample is representative or not.

We found that 70 out of 90 studies are relevant for answering our second research question. For this analysis we took into account all of the studies that had a research methodology. We only excluded the studies with score of '0'. By examining the relevant studies found that half of the studies (35 out of 70, or 50%) are neither conducted in a realistic setting nor use real data (real textual passwords that users use in their daily life).

By analyzing the relevant studies, we identified 2 research methodologies that are used. All 70 studies have retrieved the data either through a survey (questionnaire), an experiment or a combination of a questionnaire and an experiment. A questionnaire as a research methodology was a choice for 23 out of 70 (32.86%) studies, while an experiment was used in 30 out of 70 (42.86%) studies. Both research methodologies were used in 17 studies. A surprising fact is that only 3 out of 70 studies have a representative sample (as claimed by authors).

3.2.1 Discussion on Research Question 2

The findings presented in the previous subsection may indicate that the data related to textual passwords, collected by these studies, may not be accurate or may not reflect the reality. Conducting laboratory experiments and surveys in which participants are aware that they are being monitored, may lead to biased, less accurate or fake data. For example, participants may create fake passwords in order to protect their real ones or to quickly conclude the survey; or create stronger passwords if they are expecting additional effort. The fact that all 70 studies have retrieved the data either through a questionnaire, an experiment or a combination of both, but only 3 out of 70 studies have a representative sample, may indicate that the retrieved results are neither statistically significant, nor

represent the population, or both. This is very important, since this increases the standard error and we cannot conclude whether the results are reliable or can be generalized. By analyzing the relevant studies for this research question we noticed that users are becoming more aware about the threats related to password security and the importance of creating passwords that are strong and hard-to-guess [26]. Also, passwords became more secure over time regarding password length. The average password length has increased to slightly less than 7-8 characters [25] [27]. Despite that, almost nothing has changed regarding password composition and password management. User-selected passwords are still weak (composed only of lower-case letters, upper-case letters or numbers), users still tend to write their passwords down, so they can easily remember them, still tend to share their passwords with their friends, and they also rarely change their passwords [28] [29].

Because of the ease with which random passwords are recoverable offline, we can expect that, in the future, the security of any information system that is based on passwords will be related to the availability of the material for passwords disclosure and not on how random and strong passwords are [30]. Therefore, stronger passwords may not be always the right solution, as long as the security mechanisms and protocols are well designed (e.g., freezing the account for a time if the wrong password is entered for a certain number of times) [31]. This can be more useful for smaller institutions with hundreds of users where more complex security protocols can be easily applied. Users can also be encouraged to design strong passwords using elements associated with a given service together with a personal factor [32]. We discussed in subsection 3.1.1 that the growth of Web-based services will bring additional challenges for the users, since they will have to memorize even more passwords in the future. This can develop the need for some other usable alternatives to textual passwords in the future [27]. On the other hand, as discussed in subsection 3.1.1, it is very important to prevent users from entering weak passwords into the system, since this can lead to compromising other users' accounts. In order to reach that goal a certain number of new password strength metrics and password meters have been developed [33]–[36] [37]. Nevertheless, due to the widespread use of the World Wide Web and the increased number of Web accounts that a user has to maintain, it is debatable whether these solutions can help users to cope with the large number of accounts and passwords.

3.3 RQ3: What is the relationship between users and their textual passwords? (Are the users still “the weakest link”?)

The issue of users being the weakest link in password security is addressed by 13 out of 90 studies. By analyzing the 13 relevant studies we came to a conclusion that the user behavior is a common issue in the security of information systems. User are often treated by the security departments as a security risk that needs to

be controlled, consequently creating security mechanisms whose usability is rarely investigated [38] [39]. From what has been presented and discussed so far we can argue that users usually are not aware about the security threats and their importance in the security of any information system. The lack of communication between users and organizations (or their security departments) is still present and often leads to the development of useless security mechanisms because they are badly matched to users' capabilities and their tasks [38] [39]. Therefore, if we want more usable security mechanisms for the users, then maybe we should use a "user-centric" approach for designing "usable security" (i.e., human factors should be given priority over technological factors) [40] [41]. Some preliminary studies even imply that "nudges" using multiple psychological effects could serve as important design cues towards making users to perform the intended behavior more easily [17]. On the other hand, Vidaraman et al. [42] claim that users are nonetheless the enemies of the system and different security policies should be tailored for different types of users. They divide the users to *ignorant* and *non-compliant* users. They argue that the solution to cope with ignorant users is to educate them about security mechanisms, and the solution to cope with non-compliant users is to persuade them to follow the security best practices. Users and their textual passwords will continue to be "the weakest link" in any password system. Security departments should consider implementing a "user-centered" design in order to motivate the users to behave in a secure manner [20] [38]. Users have to be treated as partners in the endeavor to secure organization's systems, not as the enemy within.

Conclusions

This paper presents the results of a systematic review of 90 relevant studies in the area of password use and password security. To the best to our knowledge, this is the first systematic literature review about password security problems. We identified the most common problems related to creating and managing textual passwords. We also outlined the various solutions proposed and used over the years. Because passwords continue to be one of the most common authentication mechanisms, we expected to find a considerably high number of relevant studies in the area of password use and password security. Contrary to what we expected, we found only 90 relevant studies, out of 2201 potential search results. Almost all of them (86), address a problem regarding to creating or managing textual passwords, but only 42 propose a solution for coping with the identified problems, which is a very low number of proposed solutions, given the fact that these problems have been known for almost 35 years. Furthermore, only 17 studies have statistically verified their solutions and used real data in their surveys or experiments, which may raise a suspicion that the retrieved data in the remaining studies is biased or may not reflect the reality. Finally, the most important finding is that only 3 studies have a representative sample, which may indicate that the results of the majority of the studies are not statistically significant and cannot be

generalized to the population. In other words, only the results of 3 studies can be regarded as scientifically acceptable.

Overall, our results demonstrate that not much has changed in password management in almost 35 years. For example, an average user has 6.5 passwords (each of which is shared across 3.9 different websites), resulting in users, to very often, write them down, so they can easily remember them [43]. Lax security behavior involving password reuse, writing down and sharing passwords still exists, along with a lack of, or erroneous knowledge, about what constitutes a secure password. The main weakness in any password system is the end user, because they often choose weak and easily guessed passwords: dictionary words, names, birthdates, etc., only because they are easy to remember. Users' awareness about the consequences of their password choice is not at a high level and a common solution regarding to password problems has not been proposed.

In order to solve many password-related problems, much more research into the matter should be conducted. One way to increase password strength and decrease password "guessability" is devising future security policies, guidelines and education, in such a way, that will take into account human capabilities and strategies for dealing with password overload. A password manager could be used as a way of dealing with password overload. It could greatly reduce the need to remember or write down a password. The problem with common password managers is that they have a number of critical vulnerabilities (e.g. authorization vulnerabilities, user interface vulnerabilities, Web vulnerabilities etc.) [44]. They are also a single point of failure of the system, which is not quite recommended for achieving better security. Another way is to restrict the passwords that are entered in the system, by using a password checker, that filters out weak and easily guessed passwords. Most of the password checkers that we encountered during our systematic literature review, basically check for password length, perform a brute force or a dictionary check of the password, or entropy based checking for presence of non-alphabetic and upper-case characters [45] [46]. Lately, new ways of checking password strength are incorporated into password meters or password checkers [34] [35], that also check the probability of a given password to be chosen by the user. This means that meaningless but pronounceable passwords (which are easier to remember) should take precedence, thus, sacrificing some strength for usability. Understandably, such password checkers should be supported by an appropriate password policy.

We have noted that stricter password policies can pose an additional burden to the users. There is a possibility that this kind of thorough and prudent proactive password checker that forces users to choose complex passwords, can add some additional difficulty for users, when selecting their passwords. Hence, our future research will focus on creating flexible password policies tailored specifically for certain types of users, following the recommendations from [22]. Furthermore, we want to combine flexible password policies with a proactive password checker, based on Markov models. Such a password checker could check the probability of

a given password to be chosen by the user. This approach could help users create strong and easy-to-remember passwords.

Acknowledgement

The authors acknowledge the financial support from the Slovenian Research Agency (research core funding No. P2-0057).

References

- [1] S. M. Taiabul Haque, M. Wright, and S. Scielzo, "Hierarchy of users beware: passwords: Perceptions, practices and susceptibilities," *Int. J. Hum. Comput. Stud.*, Vol. 72, No. 12, pp. 860-874, Dec. 2014
- [2] S.-T. Sun, E. Pospisil, I. Muslukhov, N. Dindar, K. Hawkey, and K. Beznosov, "What Makes Users Refuse Web Single Sign-on?: An Empirical Investigation of OpenID," in *Proceedings of the Seventh Symposium on Usable Privacy and Security*, 2011, pp. 4:1-4:20
- [3] K. D. Loch, H. H. Carr, and M. E. Warkentin, "Threats to Information Systems: Today's Reality, Yesterday's Understanding," *MIS Q.*, Vol. 16, No. 2, pp. 173-186, 1992
- [4] W. Tzong-Chen and S. Hung-Sung, "Authenticating passwords over an insecure channel," *Comput. Secur.*, Vol. 15, No. 5, pp. 431-439, Jan. 1996
- [5] S. Creese, D. Hodges, S. Jamison-Powell, and M. Whitty, "Relationships between Password Choices, Perceptions of Risk and Security Expertise," in *Human Aspects of Information Security, Privacy, and Trust*, Vol. 8030, L. Marinou and I. Askoxylakis, Eds. Springer Berlin Heidelberg, 2013, pp. 80-89
- [6] R. Morris and K. Thompson, "Password security: a case history," *Commun. ACM*, Vol. 22, No. 11, pp. 594-597, 1979
- [7] M. Zviran and W. J. Haga, "Password security: an empirical study," *J. Manag. Inf. Syst.*, Vol. 15, No. 4, pp. 161-185, 1999
- [8] V. Taneski, M. Heričko, and B. Brumen, "Password security — No change in 35 years?," in *2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2014, pp. 1360-1365
- [9] B. Kitchenham and S. Charters, "Guidelines for performing Systematic Literature Reviews in Software Engineering," 2007
- [10] M. Petticrew and H. Roberts, *Systematic Reviews in the Social Sciences - A Practical Guide*, 1 edition. 2006
- [11] S. Fahl, M. Harbach, Y. Acar, and M. Smith, "On the Ecological Validity of a Password Study," in *Proceedings of the Ninth Symposium on Usable*

- Privacy and Security*, 2013, pp. 13:1-13:13
- [12] M. A. Schmuckler, "What Is Ecological Validity? A Dimensional Analysis," *Infancy*, Vol. 2, No. 4, pp. 419-436, 2001
- [13] Chitika Insights, "The value of Google result positioning." p. 10, 2013
- [14] J. Cohen, "A Coefficient of Agreement for Nominal Scales," *Educ. Psychol. Meas.*, Vol. 20, No. 1, pp. 37-46, 1960
- [15] J. L. Fleiss, B. Levin, and M. C. Paik, *The Measurement of Interrater Agreement*. John Wiley & Sons, Inc., 2004
- [16] D. Feldmeier and P. Karn, "UNIX Password Security - Ten Years Later," *Adv. Cryptol. — CRYPTO' 89 Proc. SE - 6*, Vol. 435, No. November 1988, pp. 44-63, 1990
- [17] S. Kankane, C. DiRusso, and C. Buckley, "Can We Nudge Users Toward Better Password Management?: An Initial Study," *Ext. Abstr. 2018 CHI Conf. Hum. Factors Comput. Syst.*, p. LBW593, 2018
- [18] V. Taneski, M. Heričko, and B. Brumen, "Impact of security education on password change," in *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO) 2015*, pp. 1350-1355
- [19] P. Cisar and S. M. Cisar, "Password - A form of authentication," *5th Int. Symp. Intell. Syst. Informatics, SISY 2007*, pp. 29-32, 2007
- [20] A. Adams, M. A. Sasse, and P. Lunt, "Making Passwords Secure and Usable," in *People and Computers XII*, 1997, pp. 1-19
- [21] A. S. Brown, E. Bracken, S. Zoccoli, and K. Douglas, "Generating and remembering passwords," *Appl. Cogn. Psychol.*, Vol. 18, No. 6, pp. 641-651, 2004
- [22] P. G. Inglesant and M. A. Sasse, "The true cost of unusable password policies," *Proc. 28th Int. Conf. Hum. factors Comput. Syst. - CHI '10*, p. 383, 2010
- [23] G. Notoatmodjo and C. Thomborson, "Passwords and perceptions," *Conf. Res. Pract. Inf. Technol. Ser.*, Vol. 98, No. Aisc, pp. 71-78, 2009
- [24] B. Ives, K. R. Walsh, and H. Schneider, "The domino effect of password reuse," *Commun. ACM*, Vol. 47, No. 4, pp. 75-78, 2004
- [25] S. Egelman, A. Sotirakopoulos, I. Muslukhov, K. Beznosov, and C. Herley, "Does my password go up to eleven? The impact of password meters on password selection.," *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, pp. 2379-2388, 2013
- [26] A. Moallem, "Did You Forget Your Password?," in *Design, User Experience, and Usability. Theory, Methods, Tools and Practice*, 2011, pp.

29-39

- [27] E. Von Zezschwitz, A. De Luca, and H. Hussmann, "Survival of the shortest: A retrospective analysis of influencing factors on password composition," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, Vol. 8119 LNCS, No. PART 3, pp. 460-467, 2013
- [28] S. Riley, "Password security: what users know and what they actually do," *Usability News*, Vol. 8, No. 1, pp. 2833-2836, 2006
- [29] L. Tam, M. Glassman, and M. Vandenwauver, "The psychology of password management: a tradeoff between security and convenience," *Behav. {&} Inf. Technol.*, Vol. 29, No. 3, pp. 233-244, 2010
- [30] L. St. Clair et al., "Password Exhaustion: Predicting the End of Password Usefulness," *Inf. Syst. Secur. Lect. Notes Comput. Sci.*, pp. 37-55, 2006
- [31] J. Yan, B. Alan, R. Anderson, and A. Grant, "Password memorability and security: Empirical results," *IEEE Secur. Priv.*, Vol. 2, No. 5, pp. 25-31, 2004
- [32] K. Helkala and N. K. Svendsen, "The security and memorability of passwords generated by using an association element and a personal factor," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, Vol. 7161 LNCS, pp. 114-130, 2012
- [33] M. Aliasgari, N. Sabol, and A. Sharma, "Sesame: A secure and convenient mobile solution for passwords," in *2015 1st Conference on Mobile and Secure Services, MOBISECSERV 2015*, 2015, pp. 1-5
- [34] Y. Guo and Z. Zhang, "LPSE: Lightweight password-strength estimation for password meters," *Comput. Secur.*, Vol. 73, pp. 507-518, 2018
- [35] D. Wang, D. He, H. Cheng, and P. Wang, "FuzzyPSM: A new password strength meter using fuzzy probabilistic context-free grammars," *Proc. - 46th Annu. IEEE/IFIP Int. Conf. Dependable Syst. Networks, DSN 2016*, pp. 595-606, 2016
- [36] M. Dell'Amico and M. Filippone, "Monte Carlo strength evaluation: Fast and reliable password checking," *Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Secur.*, pp. 158-169, 2015
- [37] P. Y. Lee and Y.-Y. Choong, "Human Generated Passwords -- The Impacts of Password Requirements and Presentation Styles," in *Human Aspects of Information Security, Privacy, and Trust: Third International Conference, HAS 2015, Held as Part of HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015. Proceedings*, T. Tryfonas and I. Askoxylakis, Eds. Cham: Springer International Publishing, 2015, pp. 83-94
- [38] A. Adams and M. A. Sasse, "Users Are Not the Enemy," *Commun. ACM*, Vol. 42, No. 12, pp. 40-46, Dec. 1999

-
- [39] M. A. Sasse, S. Brostoff, and D. Weirich, "Transforming the 'weakest link' - A human/computer interaction approach to usable and effective security," *BT Technol. J.*, Vol. 19, No. 3, pp. 122-131, 2001
- [40] C. A. Fidas, A. G. Voyiatzis, and N. M. Avouris, "When security meets usability: A user-centric approach on a crossroads priority problem," *Proc. - 14th Panhellenic Conf. Informatics, PCI 2010*, pp. 112-117, 2010
- [41] M. Adeka, S. Shepherd, and R. Abd-Alhameed, "Resolving the password security purgatory in the contexts of technology, security and human factors," *Int. Conf. Comput. Appl. Technol. ICCAT 2013*, Vol. 2013, 2013
- [42] S. Vidyaraman, M. Chandrasekaran, and S. Upadhyaya, "Position: The User is the Enemy," in *Proc. of the 2007 Workshop on New Security Paradigms*, 2008, pp. 75-80
- [43] D. Florencio and C. Herley, "A large-scale study of web password habits," *Proc. 16th Int. Conf. World Wide Web - WWW '07*, p. 657, 2007
- [44] Z. Li, W. He, D. Akhawe, and D. Song, "The Emperor's New Password Manager: Security Analysis of Web-based Password Managers.," in *USENIX Security Symposium*, 2014, pp. 465-479
- [45] M. Bishop and D. V. Klein, "Improving system security via proactive password checking," *Comput. Secur.*, Vol. 14, No. 3, pp. 233-249, 1995
- [46] J. J. Yan, "A Note on Proactive Password Checking," *Proc. 2001 New Secur. Paradig. Work.*, pp. 127-135, 2001

Motion Detection and Face Recognition using Raspberry Pi, as a Part of, the Internet of Things

Zoltán Balogh¹, Martin Magdin¹, György Molnár²

¹Constantine the Philosopher University in Nitra, Faculty of Natural Science, Tr. Andreja Hlinku 1, Nitra, Slovakia, zbalogh@ukf.sk, mmagdin@ukf.sk

²Department of Technical Education, Budapest University of Technology and Economics, Budapest, Hungary, molnar.gy@eik.bme.hu

Abstract: Face recognition and motion detection are described in the context of the construction of a system of intelligent solutions, for use in the home, that can be used as a single free standing functional unit or as an element of a bigger system, connected to the Internet of Things. To create a complex system, a micro PC, Raspberry Pi 3, was used together with an application capable of recognizing faces and detecting motion. The outputs are saved into cloud storage, for further processing or archiving. The monitoring system, as designed, can be used autonomously; with the use of a battery and a solar panel, it is possible to place it anywhere. Furthermore, it could be used in various fields, such as health care, for the real-time monitoring of patients, or in the tracking of the spatial activity of people and/or animals. Thus, the system created, may be regarded as a part of the IoT.

Keywords: Computer Vision; Raspberry Pi; Face detection; OpenCV; Motion detection

1 Introduction

Technological progress has made possible the development of methods of communication between people and objects, facilitating information flow in terms of both speed and security. The Internet of Things (IoT) uses this kind of progress and integration of new elements. As a part of information systems, remote access and various system controls, are enabled.

In accordance with the prevailing philosophy of ever-present communication, including machine to machine communication, the IoT may be defined as a set of technologies designed to allow the connection of heterogeneous objects through various networks and methods of communication. The main objective is to position intelligent devices in different locations to capture, store and manage information, making it is accessible, to anyone anywhere in the world [1].

The phrase “Internet of Things” was first used by Kevin Ashton, the founder of the Auto-ID Center in 1999. Although there is no standard definition of “Internet of Things”, various technologies (Radio Frequency Identification (RFID), Bar Codes and Global Positioning Systems (GPS)) are used in the implementation of the IoT [2].

The IoT may be regarded as a general term for various technologies, connecting, monitoring and controlling devices such as tags, sensors, actuators, mobile phones, home appliances, vehicles, industrial devices, robots and even medical devices over a data network. There are a number of definitions of and viewpoints regarding the IoT [3] [4].

The concept of the IoT is based on the use of a large number of smart devices, objects and sensors [4] [5]. The sustainable cost and ease of deploying these smart devices is an ever-more critical issue. The lifetime costs of a local IoT installation, “a smart thing”, can be divided into three main components: first, the cost of the smart thing itself, hardware and software, and second, the cost of connectivity. Last is the cost of deployment [6].

At first sight, it may seem obvious that the IoT is represented by physical objects. It should, however, be noted that the IoT’s value propositions are equally based on the software that runs these. This could present in many forms, for example, embedded, middleware, applications, service composition logic, and management tools [7] [8].

The Internet of Things resulted from new technologies and several complementary technical developments. It provides capabilities that collectively help to bridge the gap between the virtual and physical worlds [1]. These capabilities may include communication and cooperation, addressing capability, identification, perception, information processing, location and user interfaces.

On the basis of the previous claims, the Internet of Things could be considered a simple data network [9]. If this data network is applied to an area, for example, a smart home [10], then it is increasingly being used as a cheap and efficient sensor for simple cameras (IP cameras) [11].

The Internet of Things is the connection of all physical objects based on the Internet infrastructure with the aim of exchanging information; devices and objects are no longer disconnected from the virtual world, but can be remotely controlled and act as points of access to services [1].

Currently, various smart device solutions (gadgets) are being developed. An example might be Intel’s new smart bracelet, a system based on the principle of using ultra-sound. The bracelet is designed to detect if the wearer has a fall, and is primarily designed for sick patients with mobility problems). However, one problem with this solution is of its difficulty in detecting that a wearer is paralyzed and needs help (for example, in the case of upper and lower limb paralysis).

Therefore, at present, a great deal of attention is being paid to solutions that focus on complex motion detection - i.e., to detect limb movement, and also face detection.

According to Mano [12], an increasing number of patients are receiving home treatment. Home treatment has its benefits - especially in the field of psychological support. This is especially true of older people, who often live alone. Older people, however, do not have the same level of immunity or resistance to various diseases as young people. However, they are not willing to leave the home environment and get to the hospital. Therefore, they are often treated within the framework of home care, but this service often requires 24/7 readiness, in the form of a nurse. This is, however, for most people financially not possible. However, if it is only necessary to keep an eye on the sick person, it is possible to apply elements of the Internet of Things (IoT) with some hope of success. These might be very simple, efficient and inexpensive devices including wireless communication devices and cameras, smartphones or built-in devices e.g. Raspberry Pi. Such technologies are called Health Smart Homes (HSH). HSH is still an area relatively unexplored in the context of the Internet of Things [12].

For all these reasons, we consider the use of Health Smart Homes (HSH) very important [13] [14] [15]. The proposal for the use of HSH results from a combination of the medical environment and the possibilities of implementing information systems in the form of an intelligent home. Such a home is then equipped with a variety of remote sensing devices. The principle of such sensing devices is very simple. The sensing devices send data either at regular intervals and/or in a non-standard (critical) situation. The HSH system is described in detail, for example, in [24] [17], in which the authors propose technological solutions that help caregivers to monitor people in need. The goal is fully fledged medical supervision in the home environment while reducing financial costs [12].

This paper describes the ways of creating a monitoring and detection system with the use of computer vision and Raspberry Pi as a part of the IoT. The IoT may be of service in various fields of everyday life and may be used when detecting motion, face and person recognition. The use of Raspberry Pi and Raspberry Pi with a full HD camera module to create an IoT system is also discussed, with reference to systems that might be employed in healthcare, security, the education system, industry, etc. The aim of the paper is to create an RPi integrated intelligence with a HD camera that can process and evaluate the scanned image and detect motion. An application has been created that is able to monitor and record the activity of the surroundings with a camera. The monitoring system is capable of recognizing changes in the surroundings.

2 Related Work

The scientific area of computer vision, is currently most popular in the implementation of various applications in daily life. It is already a standard to recognize or localize one or more faces or individual objects.

In the recognition process there are 3 phases: detection, extraction and classification. Detection is used to localize objects in images. Real-time detection of the face with many other objects in the background is not a simple problem. There are situations in which it is not possible to capture the face of an object due to face rotation of more than about 30°, or to a change in the lighting, other factors which might inhibit facial appearance, such as, beards, glasses [18] [19]. The second phase is extraction, and this is determined by certain specifications and requires particular features to detect an object. These features are later used for classification. Classification is the last phase of the recognition process, in which, it is mainly used in the classification of emotional states when the subject of detection is a human face.

In recent years, a lot of literature has appeared dealing with and detailing the various phases of the recognition process [20] [21] [22], or for another view [23].

The research area of computer vision is widely used in HSH. Computer vision in the form of HSH provides various solutions in the form of health and safety services. Recently, computer vision has also become a synonym for marketing and entertainment (Kinect, though Microsoft has halted its further development) [24] [25].

However, the two fields of face recognition and the IoT are not currently connected. With regard to literature dealing with face detection in IoT, there are very few relevant sources [14] [26] [27]. The problem has a big potential, since the issues of the IoT are currently a 'hot topic' of discussion. Therefore, at least a partial overview of the most recent research works is provided.

According to Koliass et al. [28], for automatic surveillance of a scene in HSH, multiple cameras need to be installed in a given area. However, they must be resistant to current limitations such as noise, lighting conditions, image resolution and computational cost. To overcome such limitations and to increase recognition accuracy, Koliass et al. use a Radio Frequency Identification technology, which is ideal for the unique identification of objects. They examined the feasibility of integrating RFID with hemispheric imaging video cameras. The advantages and limitations of each technology and their integration none the less indicates that their combination could lead to a robust detection of objects and their interactions within an environment. The present work concludes with a presentation of some possible applications of such integration.

Mano et al [12] point to a problem with HSH; according to them, the use of camera and image processing on the IoT is a new research area. The article discusses not only face detection, but also the classification of emotion observable

on patients' faces. Mano et al. also discuss the existing literature, and show that most of the studies in this area do not put the same effort into monitoring patients. In addition, there are only relatively few studies that take the patient's emotional state into account, and this may well be crucial to their eventual recovery. The result is prototype which runs on multiple computing platforms. Indeed, the present paper was inspired by the ideas formulated by Mano et al.

According Kitajima et al. [11], people today use various smart sensors in the IoT, e.g. wristbands, because these can provide significant amounts of information about health. However, the majority of the information is personal and there are significant concerns that individual privacy could be compromised. If these wristbands are connected to the Internet as an element of the IoT, there is a potential concern about the unintentional leaking of user data.

Most of the studies [14] in the literature focus on the use of body sensors (Galvanic Skin Response (GSR), Electroencephalography (EGG) and other) and the sending of data for future processing using the IoT. The main concern is the loss of data and/or to the data falling into the hands of third parties. A few studies [26] use cameras as smart devices for improving the analysis of the environment (e.g. detecting possible hazardous situations) in HSH. These two different approaches (various sensors implemented in smart wristbands or IP cameras) clearly have advantages and disadvantages and should be used with specific objectives [12].

3 Materials and Methods

There is an example of the use of cameras in conjunction with the IoT in the context of the home healthcare presented by Mano in 2016 [12]. A classical wireless architecture for Personal Area Network (WPAN) is used to create a Smart Architecture for In-Home Healthcare (SAHHc) consisting of sensors (IP cameras) and a router with a server (Decision Maker).

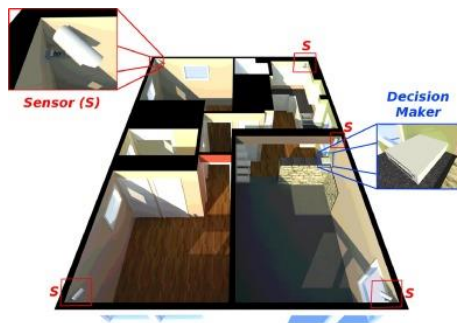


Figure 1
Example of simple SAHHc [19]

This solution is standard, but from the point of view of security, it carries various risks. Therefore, a classic HD webcam is used, connected it to a Raspberry Pi 3 micro-controller.

The microcontroller Raspberry Pi 3 was the main decision element, the Decision Maker. The communication route is illustrated in the following block scheme:

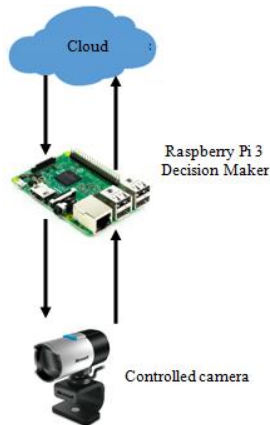


Figure 2
Block scheme of communication

The structure of this system consisted of:

- Simple devices (e.g. in this case HD cameras) that act as sensors and are capable of collecting information about the patient's health and the environment if necessary.
- A decision element, that is a decision maker with more powerful computing resources, in this case, the Raspberry Pi 3.

Sensors can detect the non-standard movement or falls and also capture their faces. When a person enters the room, the sensors detect activity. Depending on how the camera is rotated and the person's face is captured, the person can be identified (conditions depend on the degree of face rotation, light conditions, camera distance, etc.). If the identified person is a patient, the person's activity will start to be monitored. [12].

Face detection, tracking and motion detection is an important and popular research topic in the area of image processing. Computer vision as a scientific and technological study deals with the ability of electronic devices to gather information from a digital image, "to understand a situation" and thus make a decision as to whether to carry out the task or not. This quality is present in technology that is used in all fields of industry and research. Here, the Raspberry Pi 3 micro-controller and the OpenCV pack is described. OpenCV had already been used in previous systems. On the other hand, the hardware design could be

described together with the software code written in Python, and its main function is person recognition and motion detection. Three programs have been created for facial detection and their functions can be compared on the basis of the algorithms used. The speed and precision of facial and motion recognition can then be compared as well.

3.1 Raspberry Pi Characteristics

The Raspberry Pi is a one-chip PC comparable to a (weaker) desktop computer. It contains a port for the screen (HDMI) and via a USB port it is possible to connect it to a keyboard and mouse. Multiple generations of this computer have already been developed, all differing in performance and intended use. The micro CPU is from the ARM family, so it is comparable to a common smartphone. On Raspberry Pi, it is possible to operate various distributions of Linux, Windows 10 or the IoT Core, from Microsoft. Unlike PC Arduino, it is possible to use Raspberry Pi not only for various device control (with GPIO contacts), but also for the relevant application development itself. For facial and motion detection, the newest model of Raspberry Pi 3, model B is used, which was launched in February 2016. It is the first 64-bit Raspberry Pi, and is supplied with in built Wi-Fi and Blue-Tooth (Figure 3).

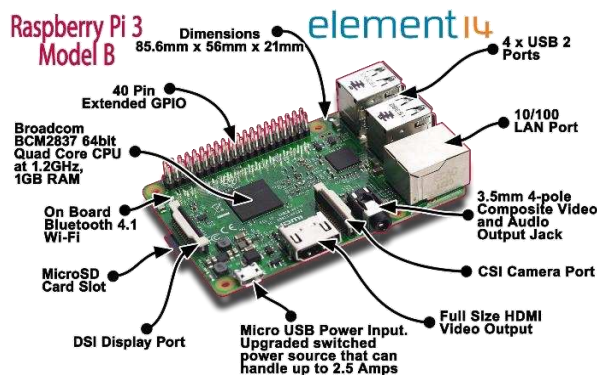


Figure 3
Raspberry Pi 3 Model B

Specifications:

- Quad core 64-bit ARM Cortex A53 with frequency 1.2 GHz, approx. 50% faster than Raspberry Pi 2
- 802.11n Wireless LAN
- Bluetooth 4.1 (including Bluetooth Low Energy)
- 400MHz Video Core IV multimedia

- 1 GB LPDDR2-900 SDRAM (900 MHz)
- 100 Mb/s Ethernet port, 4x USB 2.0, HDMI output

To create an autonomous system, the Raspberry Pi camera module v2 is used. The camera has a high quality 8 megapixel Sony IMX219 image scanner. From the point of view of static images, the camera is capable of making 3280 x 2464 image points of static pictures, it also supported 1080p in 30 fps, resolution of 720p in 60 fps and 640x480p in a 60 or 90 fps video.

Specifications:

- 8 megapixel native resolution
- High quality Sony IMX219 image scanner
- 3280 x 2464 image points of static pictures
- Supports 1080p in 30fps, resolution of 720p in 60 fps and 640x480p in a 60 or 90 fps video
- The camera is supported in the newest version of Raspbian
- 1.4 μm x 1.4 μm pixels with technology OmniBSI for high performance (high sensitivity, low crosstalk, low noise)
- Optical size $\frac{1}{4}$
- Size: 25 mm x 23 mm x 9 mm
- Weight (camera module + connecting cable): 3.4 g [29]

3.2 Visual Detection

Visual detection and facial and object recognition has been one of the biggest challenges in the field of computer vision over the last decade. The potential use of detection systems and object recognition is rather wide, from security systems (identification of authorized users), medical techniques (detection of tumors), industrial applications (visual inspection of products), robotics (navigation of a robot in an inaccessible terrain), to augmented reality, and many others.

Each phase of the recognition process uses various methods and techniques. Detection, extraction and classification have been solved by these methods. Over time, not only new methods have been developed but also algorithms have developed from the original methods, until currently, there are over 200. According to this approach and the its various uses, detection methods can be divided into four categories [30]:

- Knowledge-based methods
- Feature invariant approaches

- Template matching
- Appearance-based methods

The division of the methods is not uniform, because the methods for all three phases of the recognition process could be used. From the point of view of speed and simplicity of use, an OpenCV library that includes a powerful Viola-Jones detector based on Haar cascades might be suggested (for more information see [31]).

3.3 OpenCV

Open CV (Open-source Computer Vision, opencv.org) might be called the Swiss army knife of computer vision. It has a range of modules by which it is possible to solve a number of problems in computer vision. The most useful part of OpenCV may well be its architecture and memory management. It provides a framework within which it is possible to work with images and videos in any number of ways. The algorithms to recognize faces are accessible in OpenCV library and are the following:

- FaceRecognizer.Eigenfaces: Eigenfaces, also described as PCA, first used by Turk and Pentland in 1991
- FaceRecognizer.Fisherfaces: Fisherfaces, also described as LDA, invented by Belhumeur, Hespanha and Kriegman in 1997 [32]
- FaceRecognizer.LBPH: Local Binary Pattern Histograms, invented by Ahonen, Hadid and Pietikäinen in 2004 [33]

The choice of Fisherfaces has been made because it was based on the LDA algorithm. Rashmi [34] claims that when comparing different algorithms, a 95.3% success was achieved with LDA, while the time needed for detection was compared to other algorithms. The method used in Fisherfaces is taught from a class transformation matrix. Unlike the Eigenfaces method, it does not record the intensity of lighting. The discriminatory analysis finds the facial traits needed for person comparison. It is necessary to mention that the Fisherfaces' performance is influenced to a great extent by input data. In practical terms, if in a specific case Fisherfaces is taught using a well-lighted image, then in experiments with bad lighting, there will be a higher number of incorrect results. This is logical, as the method does not have a chance to capture the lighting on the images. The Fisherfaces algorithm is described below.

X is a random vector from the c class:

$$X = \{X_1, X_2, \dots, X_c\} \quad (1)$$

$$X_i = \{X_1, X_2, \dots, X_c\} \quad (2)$$

The scattering of the matrices S_B and S_W are calculated as:

$$S_B = \sum_{i=1}^c N_i (\mu_i - \mu)(\mu_i - \mu)^T \quad (3)$$

$$S_W = \sum_{i=1}^c \sum_{x_j \in X_i} (x_j - \mu_i)(x_j - \mu_i)^T \quad (4)$$

Where μ is the total mean:

$$\mu = \frac{1}{N} \sum_{i=1}^N X_i \quad (5)$$

And μ_i is the mean of class $i \in \{1, \dots, c\}$:

$$\mu_i = \frac{1}{|X_i|} \sum_{X_j \in X_i} X_j \quad (6)$$

Fisher's standard algorithm searches for the projection of W that maximizes the criterion of divisibility of the given class [35] [32]:

$$W = W_{fld}^T W_{pca}^T \quad (7)$$

These kinds of solutions have been used in various industrial applications, in educational projects, and more often in intelligent households. Facial recognition and motion detection have been used to construct a system of intelligent solutions in households that can be used as a separate functional unit or as an element of a bigger system connected to a technology of the Internet of Things.

4 An experiment and a discussion of Face Recognition and Motion Detection

The OpenCV library was used for face recognition and motion detection. It is a multi-platform library of programming functions related to computer vision. First, a proper choice of algorithm for face recognition must be determined. OpenCV contains two popular methods for facial recognition, the Haar Cascade and Local Binary Patterns (LBP). The Haar Cascade method is based on the principle of machine learning. The cascade function is trained using multiple images and is then used to detect objects in other images. The LBP Method (local binary pattern) is used to describe the characteristics of images with various attributes that characterize the image. When creating an attribute, it goes through each pixel of the image and using assessing features arrives at a value. In this case, the LBP

method was because its functions were simpler and faster than the Haar Cascade method.

Both algorithms use XML files to record the properties of the object which needs to be detected. The OpenCV library already contains some XML files for face recognition and body detection. Using this, a particular LBP or Haar Cascade XML file is created via training. It is possible to train the classifier to detect any object that may be required. Three programs working in different ways were created to detect faces. This means that multiple methods were tried out in order to compare the programs with respect to speed and accuracy.

4.1 Face Recognition

Face_recognition.py

The first program was slow (5 - 6 FPS), but worked very reliably. It uses a single core processor for processing images taken by the Raspberry Pi camera module. The algorithm was programmed using the OpenCV library, and employed a cascading file "lbpcascades_frontalface.xml" to recognize the faces in the picture (Figure 4).

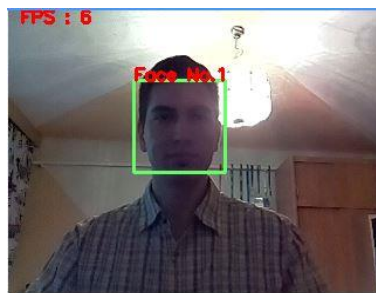


Figure 4

Face detection by "lbpcascades_frontalface.xml"

As the program runs, when it recognizes a face, this is indicated by a green square appearing around the face. Given that only one core was used, FPS was quite slow, showing only 5-6 frames per second and using about 53 percent of system resources.

Object_recognition.py

The second program could recognize more faces and other objects, such as mobile phones, at the same time. The algorithm worked in the same way as the previous face_recognition.py program but another cascade file cascade.xml was used here for the detection of mobile phones. It should be emphasized that the XML file used to detect phones was generated by Radames Ajna and Thiago Hersan. The output of the program is shown in the figure below (Figure 5).

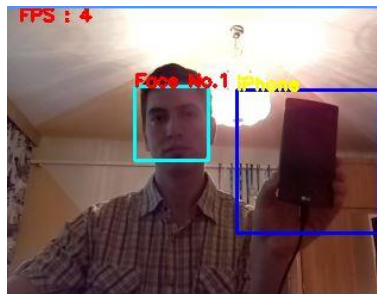


Figure 5
Detection of mobile phones

In this program, two different cascading files are used for identification: `cascade.xml` for phone detection and `lbpcascades_frontalface.xml` for face detection. This means that the search time is doubled and the FPS rate drops to 3 - 4 images per second, while the program uses about 61 percent of system resources.

Multi_core_face_recognition.py

The previous two programs were slowed down, achieving only four to six FPS and therefore are not the best solution to the problem. In this case, more cores could have been used, seeing as the Raspberry Pi 3 contains four cores. In cases where it is necessary for the program to work faster, the algorithms for facial recognition need to be made to operate in parallel. A library for Python can be imported so that multiple processes can operate at once, as library features could be used to support parallel processing (multiprocessing). This program loads four images at once (one for each core) to be processed using a face recognition algorithm and it finally evaluates the results. Using the parallel processing, the program reached rates of up to 15 FPS, about three times faster than in the two previous programs, and used about 91 percent of system resources. The output of the program is shown in the figure below (Figure 6).

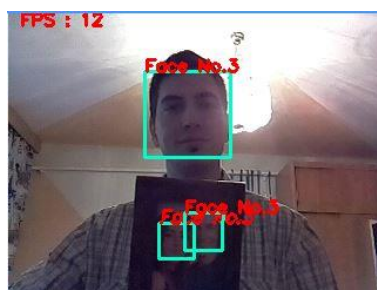


Figure 6
Face detection using four cores of Raspberry Pi 3

This program is similar to the first one (`face_recognition.py`), with the exception that in this case all four cores are used and thus the program works faster and more smoothly.

Table 1
The use of system resources

Program name	The number of used cores	FPS	System resources
<code>face_recognition.py</code>	1	6	53.00%
<code>object_recognition.py</code>	1	4	45.00%
<code>multi_core_face_recognition.py</code>	4	15	91.00%

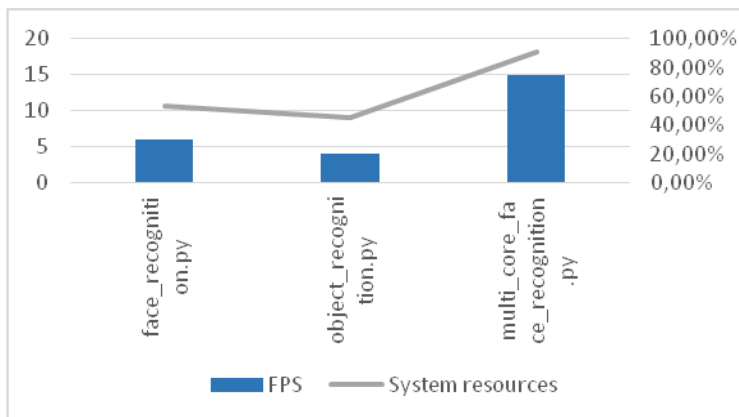


Figure 7

The use of system resources

In Table 1 and Figure 7, the FPS value and use of system resources were directly proportional, which means that if more FPS were needed, then Raspberry had to use more system resources (cores).

4.2 Motion Detection

For detecting motion, two programs were created. The first program (`detektor_pohybu.py`) works without using OpenCV library and when it perceives a movement it creates an image and stores it in a folder. The second program (`monitorovaci_system.py`) uses an OpenCV library and after detecting motion it draws a green box around the detected area and records the file into Dropbox.

The programs were tested in internal and external environments for 5 hours in each case. The motion detector saved 72 images outside and 56 inside, 128 altogether. There were 109 good images of all the results on which real movement could be detected. The other 19 pictures showed distorted results, for example due to changing light conditions. In the course of the testing of the monitoring system, it saved 64 images outside and 48 inside, 112 altogether. There were 107 good images of all the results and 5 distorted ones. The results may be seen in the table and graph below.

Table 2
Testing of motion detection and monitoring system

	Time	Inside	Outside	Together	accurate	inaccurate
motion detector	5 hour	56	72	128	109	19
monitoring system	5 hour	48	64	112	107	5

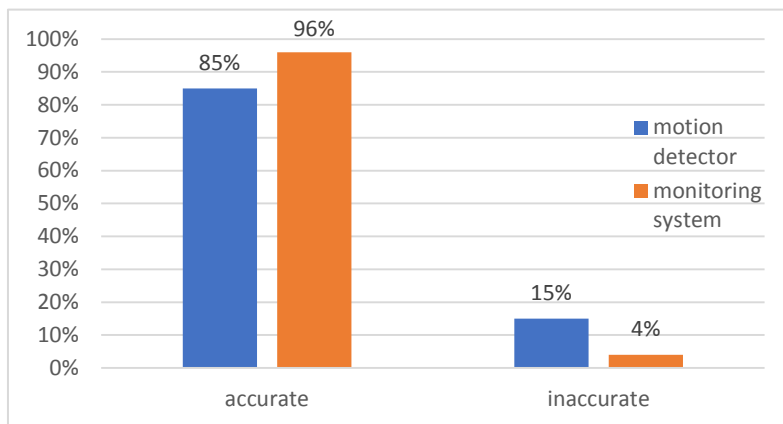


Figure 8
Percentage of accurate and inaccurate image while recording

As may be seen in Figure 8, the monitoring system achieved better results because it took the weighted mean of images for comparison, instead of the previous images. Thus, if the lighting conditions change, the algorithm still works precisely.

A functioning system for monitoring internal and external environments was designed, and was capable of being implemented in practice as an intelligent security system or to monitor different rooms, objects, people or animals in nature. The camera recorded only when motion was perceived and only then stored or sent images to cloud storage. In this way, space on the hard drive could be saved,

since instead of watching long videos, pictures were taken, and if an important image (for example, showing a burglary) needed to be found, it would be enough to search by date and see what happened in the monitored area.

Conclusions

The Internet of Things is a new trend in informatics that connects different devices to the Internet, usually wirelessly, using Wi-Fi or Bluetooth. The monitoring system designed here, could also be used autonomously, meaning that with the help of a battery or a solar panel, it could be placed anywhere and would operate without any outside control. It could be used in various fields, such as healthcare, monitoring patients in hospitals and checking their status in real time. Motion activity could be monitored, for example, in parking lots, allowing access to the lot only if there are vacancies. Then the activity and behavior of animals can be observed and described, using the obtained data as detailed in previous publications of the author of the present and other related paper [36] [37] [38] or as an intelligent safety and mobile system, whether to monitor different rooms or animals in the wild.

However, even though cameras can provide significant amounts of information, the vast majority is personal data, and there are significant concerns that individual privacy could be compromised. Furthermore, since home appliances are increasingly being connected to the Internet via the IoT, it has become possible for user images to leak out unintentionally. With these concerns in mind, there is a need to propose a human detection method that protects user privacy by using intentionally blurred images. In this method, the presence of a human being is determined by dividing an image into several regions and then calculating the heart rate detected in each region. This proposal was realized first by Kitajima et al. in 2017. In overall performance evaluation, the proposed method showed favorable performance results when compared with an OpenCV based face detection method, and was confirmed to be an effective method for detecting human beings in both normal and blurred images [11]. Therefore, in future research, the focus will be on the application of the method, in practice.

References

- [1] J. I. R. Molano, D. Betancourt, G. Gómez, "Internet of things: A prototype architecture using a raspberry Pi," *Lecture Notes in Business Information Processing*, Vol. 224, 2015, pp. 618-631
- [2] L. G. Guo, Y. R. Huang, J. Cai, L. G. Qu, "Investigation of architecture, key technology and application strategy for the internet of things," *Proceedings of 2011 Cross Strait Quad-Regional Radio Science and Wireless Technology Conference*, CSQRWC 2011, 2011, pp. 1196-1199
- [3] L., Atzori, A. Iera, G. Morabito, "The Internet of Things: A survey. *Computer Networks* 54(15), 2010, pp. 2787-2805, doi:10.1016/j.comnet.2010.05.010

- [4] J. Katona, T. Ujbanyi, G. Sziladi, A. Kovari: Electroencephalogram-Based Brain-Computer Interface for Internet of Robotic Things, Cognitive Infocommunications, Theory and Applications, Springer 2018, pp. 253-275
- [5] J. Katona, P. Dukan, T. Ujbanyi, A. Kovari: Control of incoming calls by a windows phone based brain computer interface, Proceedings of the 15th IEEE International Symposium on Computational Intelligence and Informatics, Budapest, Hungary, 2014, pp. 121-125
- [6] A. Iivari, J. Koivusaari, H. Ailisto, "A rapid deployment solution prototype for IoT devices," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 10070 LNCS, 2016, pp. 275-283
- [7] C. Ebert, C. Jones, "Embedded software: Facts, figures, and future," *Computer*, 42(4), 2009, pp. 42-52, doi:10.1109/MC.2009.118
- [8] S. Stastny, B.A Farshchian, T. Vilarinho, "Designing an application store for the internet of things: Requirements and challenges," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 9425, 2015, pp. 313-327
- [9] I. Farkas, P. Dukan, J. Katona, A. Kovari: Wireless sensor network protocol developed for microcontroller based Wireless Sensor units, and data processing with visualization by LabVIEW, Proceedings of the IEEE 12th International Symposium on Applied Machine Intelligence and Informatics, Herlany, Slovakia, 2014, pp. 95-98
- [10] P. Dukan, A. Kovari: Cloud-based smart metering system, Proceeding of the 14th IEEE International Symposium on Computational Intelligence and Informatics, Budapest, Hungary, 2013, pp. 499-502
- [11] T. Kitajima, E. A. Y. Murakami, S. Yoshimoto, Y. Kuroda, & O. Oshiro, "Human detection using biological signals in camera images with privacy aware," *Proceedings of the 16th International Conference on Intelligent Systems Design and Applications*, Volume 557, 2017, pp. 175-186, doi:10.1007/978-3-319-53480-0_18
- [12] L. Y. Mano, B. S. Faiçal, L. H. V. Nakamura, P. H. Gomes, G. L. Libralon, R. I. Meneguete, J. Ueyama, "Exploiting IoT technologies for enhancing health smart homes through patient identification and emotion recognition," *Computer Communications*, 89-90, 2016, pp. 178-190, doi:10.1016/j.comcom.2016.03.010
- [13] G. Riva, "Ambient intelligence in health care," *Cyberpsychology and Behavior*, 6(3), 2003, pp. 295-300, doi:10.1089/109493103322011597
- [14] E. Romero, A., Araujo, J. M. Moya, J.-M. de Goyeneche, J. C. Vallejo, P. Malagon, D. Villanueva, D. Faga, "Image processing based services for ambient assistant scenarios," *Distributed Computing, Artificial Intelligence*,

- Bioinformatics, Soft Computing, and Ambient Assisted Living. Lecture Notes in Computer Science*, 5518, Springer, Berlin, Heidelberg, 2009, pp. 800-807
- [15] A. N. Siriwardena, "Current state and future possibilities for ambient intelligence to support improvements in the quality of health and social care," *Quality in Primary Care*, 17(6), 2009, pp. 373-375
- [16] V. Rialle, F. Duchene, N. Noury, L. Bajolle, J. Demongeot, "Health Smart home: information technology for patients at home," *Telemed. J. e-Health*, 8 (4), 2002, pp. 395-409
- [17] J. A. Stankovic, Q. Cao, T. Doan, L. Fang, Z. He, R. Kiran, S. Lin, S. Son, R. Stoleru, A. Wood, "Wireless sensor networks for in-home healthcare: potential and challenges," *High Confidence Medical Device Software and Systems Workshop (HCMDSS)*, 2005
- [18] Y. L. Tian, T. Kanade, C. J. F. *Handbook of Face Recognition*. Springer, 2005
- [19] M. Magdin, M. Turcani, L. Hudec, "Evaluating the Emotional State of a User Using a Webcam," *International Journal of Interactive Multimedia and Artificial Intelligence*. 4(1), Special Issue: SI, 2016, pp. 61-68
- [20] K. Bahreini, R. Nadolski, & W. Westera, "Towards multimodal emotion recognition in e-learning environments," *Interactive Learning Environments*, no. Ahead-of-print, 2014, pp. 1-16
- [21] U. Bakshi, & R. Singhal, "A Survey of face detection methods and feature extraction techniques of face recognition," *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, 3(3), 2014, pp. 233-237
- [22] V. Bettadapura, "Face Expression Recognition and Analysis: The State of the Art," *arXiv: Tech Report*, (4), 2012, pp. 1-27
- [23] A. Ollo-López, & M. E. Aramendía-Muneta, "ICT impact on competitiveness, innovation and environment," *Telematics and Informatics*, 29 (2), 2012, pp. 204-210
- [24] S. Marcutti, G. V. Vercelli, "Enabling touchless interfaces for mobile platform: State of the art and future trends," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 9769, 2016, pp. 251-260
- [25] A. Chaudhary, R. Klette, J. L. Raheja, X. Jin, "Introduction to the special issue on computer vision in road safety and intelligent traffic," *Eurasip Journal on Image and Video Processing*, 2017(1), 2017, doi:10.1186/s13640-017-0166-5
- [26] J. Augusto, P. McCullagh, V. McClelland, J. Walkden, "Enhanced healthcare provision through assisted decision-making in a smart home

- environment,”*IEEE Second Workshop on Artificial Intelligence Techniques for Ambient Intelligence, IEEE Computer Society*, 2007
- [27] S. Helal, W. Mann, J. King, Y. Kaddoura, E. Jansen., “The gator tech smart house: a programmable pervasive space,”*Computer*, 38 (3), 2005, pp. 50-60
- [28] V. Koliás, I. Giannoukos, C. Anagnostopoulos, I. Anagnostopoulos, V. Loumos, & E. Kayafas, “Integrating RFID on event-based hemispheric imaging for internet of things assistive applications,” Paper presented at the *ACM International Conference Proceeding Series*, 2010, doi:10.1145/1839294.1839367
- [29] H. Tubman, Raspberry Pi Camera Board. <https://www.adafruit.com/products/13671>, 2016
- [30] M. Yang, D. J. Kriegman & N. Ahuja, “Detecting faces in images: A survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1), 2002, pp. 34-58. doi:10.1109/34.982883
- [31] P. Viola & M. J. Jones, “Robust real-time face detection,” *International Journal of Computer Vision*, 57(2), 2004, pp. 137-154, doi:10.1023/B:VISI.0000013087.49260.fb
- [32] P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman, “Eigenfaces vs. fisherfaces: Recognition using class specific linear projection,” *IEEE Trans Pattern Anal Mach Intell*, 19(7), 1997, pp. 711-720, doi:10.1109/34.598228
- [33] D. Baggio, S. Emami, D. Escrivá, K. Ievgen, N. Mahmood, J. Saragih & R. Shilkrot, “Mastering OpenCV with practical computer vision projects,” *Packt Publishing Ltd*, 2012
- [34] R. Rashmi, D. Namrata, “A Review on Comparison of Face Recognition Algorithm Based on Their Accuracy Rate,” *International Journal of Computer Sciences and Engineering*, 3(2), 2015, pp. 40-44
- [35] P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection,” In: Buxton, B., Cipolla, R. (eds.) *4th European Conference on Computer Vision, ECCV 1996*, Vol. 1064, 1996, pp. 45-58
- [36] Z. Balogh, R. Bízik, M. Turčáni, Š. Koprda, “Proposal for Spatial Monitoring Activities Using the Raspberry Pi and LF RFID Technology,” In: Zeng, Q.-A. (ed.) *Wireless Communications, Networking and Applications*, Vol. 348, Lecture Notes in Electrical Engineering, 2016, pp. 641-651
- [37] Z. Balogh, M. Turčáni, “Complex design of monitoring system for small animals by the use of micro PC and RFID technology,” In: Oualkadi, A. E., Moussati, A. E., Choubani, F. (eds.) *Mediterranean Conference on*

Information and Communication Technologies, MedCT 2015, Vol. 380, 2016, pp. 55-63

- [38] R. Pinter, S. Maravić Čisar, “Measuring Team Member Performance in Project Based Learning” *Journal of Applied Technical and Educational Sciences*, ST Press, 2018, Volume 8, Issue 4, pp. 22-34

The Use of an Adjusted Transportation Model, for Optimizing Provision of International Help, in Case of Emergency Situations

Dalibor Kekić¹, Miloš Milenković², Aleksandar Čudan¹

¹ University of Criminal Investigation and Police Studies, Department of Police sciences, Cara Dušana 196, 11080 Zemun, Serbia; dalibor.kekic@kpu.edu.rs, aleksandar.cudan@kpu.edu.rs

² University of Belgrade, Faculty of Organizational Sciences, Jove Ilića 154, 11000 Belgrade, Serbia; mijatov51804@fon.bg.ac.rs

Abstract: This paper focuses on finding a model to optimize the provision for international emergency help, for emergencies caused by natural or man-made disasters. Nowadays, natural and man-made disasters occur more often than before, possibly, due to climate change, industrial activity, urbanization and migration of people. The national institutions for protection and rescue, in many cases, when the emergency situation is declared, cannot often cope and need help. The international organizations have recognized needs to develop mechanisms, which can be used to help affected countries. Examples are European mechanisms for civil protection and numerous guidelines from the United Nations Office for the Coordination of Humanitarian Affairs (OCHA). If the affected country cannot solve the problems and minimize risks for their citizens, material and cultural heritage, Governments send an official request to International Organizations, to obtain different kinds of international help as quickly as possible. However, the problems can appear with costs and other needed resources for providing international help, in terms of country's distance which could provide help or duplication of resources that should be available. Currently, International Organizations do not have the documents, guidelines or software to be used for an emergency, when they have to make optimal decisions, concerning which country will provide help. This can be recognized as a main research gap, which is addressed by this paper. This paper uses operational research, to develop an adjusted transportation model, for optimizing the provision of international help in emergency situations. The main goal of this paper is to find useful solutions for those responsible for emergency management in making decisions for providing help to affected countries. Moreover, we aim to develop a model that will facilitate the appropriate disposition of human and material resources to an affected country in experiencing a disaster. The applied method involves an application of an adjusted transportation model for the case study, based on a real emergency situation during the May floods of 2014, in the Republic of Serbia. Having this in mind, the authors try to provide general results and a model, with a recommendation of how the model can be applied to any emergency situation in the world. The applicability is obvious for the activities of international organizations responsible for emergency management.

Keywords: costs; model; disasters; emergency situations; international help

1 Introduction

Natural and man-made disasters generate huge problems for many countries all around the world. The scope and damage of actual emergency situations, caused by natural and man-made disasters, is greater than the past. The number of casualties and victims rise during recent years, commensurate with population densities of affected areas. Also, material damage creates life conditions which need a lot of money and other resources for a recovery phase. But, one of the main characteristics of the actual natural and man-made disasters, is that National borders are not to be officially used during emergency situations. Floods are a good example of a natural disaster, which includes international or regional cooperation, between the affected countries, through which, a river passes. Also, the huge impact of emergency situations toggles countries to think globally, more than locally. If the affected country cannot cope with disasters, it will try to get help from an international source [1].

The international organizations and institutions, such as the European Union and the United Nations, recognize need to unit resources from many member states. The resources are united within the developed mechanisms and frameworks. In 2001, the European Union Civil Protection Mechanism was established, in order to have better cooperation among national civil protection authorities across Europe. Whenever the sheer scale of an emergency overwhelms the response capabilities of a country, the EU Civil Protection Mechanism enables coordinated assistance from its participating states. These include all EU Member States, as well as, Iceland, Macedonia, Montenegro, Norway, Serbia and Turkey. The Mechanism is there to protect EU citizens and extend solidarity outside Europe's borders to people who are affected by disasters and need help. Any country in the world, including, the UN and its Agencies, and International Organizations can request assistance, through the EU Civil Protection Mechanism [2]. On the other side, the United Nations Office for the Coordination of Humanitarian Affairs (OCHA) was established in 1991. OCHA is the part of the United Nations Secretariat responsible for bringing together humanitarian actors to ensure a coherent response to emergencies. OCHA also ensures there is a framework within which, each actor can contribute to the overall response effort [3]. These are the best examples of the organizations and institutions that provide international help, in case of emergency situations declared, due to natural or man-made disasters [4].

However, in practice there are obstacles when non-affected countries want to help to affected country, within developed mechanisms and frameworks. The affected country, needs help as soon as possible. Otherwise, the distance and availability of resources could be potential threats for the organization providing international help. The additional problem occurs when more than one country is affected by natural or man-made disasters. That means that the emergency situation hits the whole region at same time. It is important to make adequate calculations in order to satisfy all of the included parties in chain of providing and obtaining different kinds of help [5]. The adjusted transportation model of operational research

should be used in practice to reduce costs of logistics and improve organization of providing international help to affected country in order to minimize consequences during natural and man-made disasters. It is important to highlight the different types of natural or man-made disasters. Natural hazards and disasters can be split into three categories: Hydro-meteorological, Geophysical and Biological hazards. Hydro-meteorological disasters are natural processes or phenomena of atmospheric, hydrological or oceanographic nature that may cause loss of life or injury, property damage, social and economic disruption or environmental degradation. These include: floods, droughts, landslides, storms, hurricanes and tidal waves. Geophysical disasters are natural earth processes or phenomena that may cause loss of life or injury, property damage, social and economic disruption or environmental degradation. These include: earthquakes, tsunamis and volcanic eruptions. Biological disasters are processes of organic origin or those conveyed by biological vectors, including exposure to pathogenic micro-organisms, toxins and bioactive substances, which may cause loss of life or injury, property damage, social and economic disruption or environmental degradation. These include: epidemic and insect infestation [6]. On the other side, man-made disasters are caused by human impact and include fires, chemical and technical accidents in industry or traffic. Finally, there are many definitions of emergency situations. For the purpose of this paper, maybe the best definition of emergency situation, that is mostly used in Serbia, is the part of Law, on Emergency Situations mentioning that, “emergency situation is the condition, in which, risks and threats or consequences of catastrophes, extraordinary incidents and other hazards threatening the population, environment and material goods, are of such a volume and intensity, that their occurrence or consequences cannot be prevented or eliminated by regular action on by the authorities and/or services in charge, due to which, it is necessary to deploy special measures, forces and means together with an enhanced work regime, in order to mitigate or eliminate them” [7]. In other words, emergency situations are coming after natural or man-made disasters when there is lack of recourses to provide adequate response. This paper attempts to develop an adjusted general transportation problem, which should be used for any type of emergency situation when affected country does not have enough resources and request international help. The difference of adjusted general transportation problem, when it is used for different types of emergency situations, only appears in terms of the located requested resources. For example, Sweden has resources to help other countries, in case of wild and forest fires, but does not have resources for earthquakes. Conversely, France can help in cases of emergency situations caused by earthquakes. So, the Model, will be in a small extent, changed, according to the different types of emergency situations. Moreover, the Model should be used by someone in the international environment, who is in charge of making optimal decisions concerning recourses to the affected country. Optimal decisions, means decisions that save money and resources and, at the same, time provide all the needs for the affected country.

After the introduction, this paper continues with a theoretical background and methodology, which is based on a case study of the May floods 2014, in the Republic of Serbia. Authors in the first paragraph of the methodology section will

develop a general model, which after will be used in the formal case study. Finally, results of using a transportation model will be presented and discussed, in order to make final conclusions.

2 Theoretical Background

In many areas, where logistic services take an important role, reducing costs of transportation is a constant goal. In order to minimize transportation costs researchers in the area of operational research try to find the model which will propose optimal solutions. Each area has its own characteristics that affect model behavior. However, defining transportation problems and the model, should be useful tools for optimizing the provision of international help, in case of emergency situations.

Humanitarian logistics is defined as “the process of planning, implementing and controlling the efficient, cost effective flow and storage of goods and materials as well as related information from the point of origin to the point of consumption for the purpose of alleviating the suffering of vulnerable people” [8]. One of the main characteristics of humanitarian logistics, is time. When disasters occur, relief needs to happen, as soon as possible. So, the participants of humanitarian logistic chain have to react and take action, very fast. In the meantime, economic criteria plays an important role in the decision process. On the other side, various definitions of logistic can be found in the scientific literature, one says that “Logistics is defined as the planning, organization, and control of all activities in the material flow, from raw material until final consumption and reverse flows of the manufactured product, with the aim of satisfying the customer’s and other interest party’s needs and wishes i.e., to provide a good customer service, low cost, low tied-up capital and small environmental consequences” [9]. Maybe the better definition, for the purpose of this paper, is that “Logistics is defined as those activities that relate to receiving the right product or service in the right quantity, in the right quality, in the right place, at the right time, delivering to the right customer, and doing this at the right cost (The even R’s)” [10]. The latter definition emphasizes the importance of seven things, that have to be in the right place, during the provision of international help to countries affected by natural or man-made disaster. Humanitarian logistics differs from the logistics operations in the commercial supply chains, because of uncertainties in route selection, changing facility capacity, changing demand, safety issues, unused routes and other challenges, like disrupted communication systems, limited availability of resources and the need for efficient and timely delivery [11]. Many unknown factors increase logistic costs, especially those one which refers to transport.

The transportation problem involves finding the lowest-cost plan for distributing stocks of goods or supplies from multiple origins to multiple destinations that demand the goods. The transportation model can be used to determine how to allocate the supplies available from the various factories to the warehouses that

stock or demand those goods, in such a way that total shipping cost is minimized (i.e., the optimal shipping plan). Usually, analysis of the problem will produce a shipping plan that pertains to a certain period of time (day, week), although once the plan is established, it will generally not change unless one or more of the parameters of the problem (supply, demand, unit shipping cost) changes [12]. The questions are, on which this paper tries to give answer, is it possible to use this kind of model to optimize provision of international help in case of emergency situations?

During the preparation of this paper authors found some guidelines on how to manage humanitarian assistance in disaster situations. For example, the Pan American health organization developed one guideline for effective aid. Humanitarian assistance is beneficial to disaster victims and can play an important role in the development of the country if it is properly coordinated and responds to real needs. Both donors and authorities in disaster-prone countries should keep in mind the several principles for effective humanitarian assistance [13]. This guideline provides a lot of principles that should be used during provision of the international help. But, there is no any tool how to optimize this process in order to reduce cost for sending country or not to duplicate resources in the affected country. Chan with her research team worked on a real-time optimization for disaster response, using a mathematical programming approach. Their mathematical model focuses on providing different kinds of humanitarian aid, but in term of a local response. They proposed a real-time decision support tool, with the use of optimization, to aid post-disaster decision making. They adopted a mathematical programming approach to model the problem, where the decisions are the shipments of commodities from emergency supplies storage facilities to affected communities. They also considered multiple types of commodities and heterogeneous vehicles for transportation. The objective function of their model involves the cost of shortage and a piece-wise linear cost of travel time, to penalize the delays within different time intervals [14]. There are also other models in this area, such as, one developed by Baraka, Yadavalli and Singh. This group of authors worked on a transportation model for an effective disaster relief operation in the SADC region. SADC consists of 15 countries: Angola, Botswana, Democratic Republic of Congo, Lesotho, Madagascar, Malawi, Mauritius, Mozambique, Namibia, Seychelles, South Africa, Swaziland, Tanzania, Zambia, and Zimbabwe [15]. In their paper and with the SADC real-life cases studied, the linear programming model targeted a cost-effective route from origins to supplies, while the spanning tree-based genetic algorithm solved the shortest delivery route, by minimizing both time and cost. The models mentioned in this paragraph try to optimize the humanitarian aid in case of emergency situation. The main targets of optimization are costs and time. But, the existing models do not focus on the provision of global international help in case of large scale disasters when a lot of countries want to help the affected country. This reflects the main aim of this paper, more precisely how to optimize the humanitarian aid and operations in the large international environment using the adjusted transportation model and operational research.

Namely, some international legislative defines that costs of humanitarian aid providing to affected country will be paid accordingly to agreement between the state which request and state which provide help. Moreover, the practice shows that usually the providing country takes care about transportation and other logistic costs. The main reason is situation of affected country or countries, if the natural or man-made disaster hit the huge territory or the whole region, which probably needs a lot of money for recovery phase, so not need the additional expenses. Besides that, the providing country has the restricted funds for this purpose. There is necessity to find model which will optimize the transport route between the providing and the affected country or countries in case of emergency situations, with the primary goal to minimize the transportation costs. The transportation model should be an option for solving this problem [16].

3 Methodology

In this paper, the adjusted transportation model for optimizing provision of international help in case of emergency situations caused by natural is suggested. Many constraints affect the final decision on which country or countries will provide operational and humanitarian aid to one where emergency situation is declared. The humanitarian aid includes human, as well as, material resources. In this paper will be consider deployment and scheduling of rescuers as human resources that help to the domestic search and protection capabilities. One of the most influential constraints is the incompatibility between demand and offer of humanitarian aid. Moreover, a limiting factor should be the budget for providing humanitarian aid and sending rescuers to affected countries. Currently, many countries cope with limited available funds and budget. Otherwise, during natural or man-made disasters principle of solidarity becomes actual. Practice shows that many wants to provide help. The goal is to optimize the number of rescuers which will be deployed to affected countries and at same time to minimize their transportation and other costs related to stay in the affected country. Nevertheless, equipment used by international rescuers should be also limiting factor and constraint. The interoperability between domestic and international rescuers is very important. During the decision-making process it is necessary to take into account that rescuers with the corresponding equipment can be sent, regardless of the costs of sending. Adjusted transportation model will be used for minimization of deployment costs of rescuers. In addition of minimization of transportation costs, model will be developed to optimize international help in terms of satisfying all requests that must be fulfilled when sending international rescue forces. The first request is that all domestic rescue forces should be deployed before international teams arrive in the affected country. The second request is teams will rarely be separated. So, beside the costs, during the optimization aspects, that also should be considered, are constraints that all domestic rescue power has to be used, as well as, that the international teams cannot be separated.

Also, the catastrophic floods in the territory of the Republic of Serbia, during May 2014, would be used as a case study for creating the model. During the third week of May, exceptionally heavy rains fell on Serbia which was caused by a low-pressure system ('Yvette') that formed over the Adriatic. Record-breaking amounts of rainfall were recorded, more than 200 mm of rain fell in western Serbia, in a week's time, which is the equivalent of 3 months of rain under normal conditions. Overall the floods affected some 1.6 million people, living in 38 municipalities/cities, mostly located in central and western Serbia. Two cities and 17 municipalities were severely impacted. In reaction to the severe flooding and ensuing landslides, on 15 May, the Government of Serbia declared a state of emergency for the entire territory. At the same time, in order to maximize the effectiveness of the response to the emergency, a request for assistance was sent to the international community, notably to the Governments of the European Union (EU) Member States, EU Candidate Countries in the region, the Russian Federation, the European Commission (EC) and the United Nations (UN). In response, the European Commission immediately activated the EU Civil Protection Mechanism, to call on Member States resources and staff [17].

3.1 Model Definition

In order that the affected country satisfied the needs, regarding help from abroad, in the case of an emergency situation, the model has to be properly defined, with all of the actual constraints. First, the affected country will try to solve problem with their own resources. But, when domestic resources are not enough to cope with natural or man-made disasters, the Government of the affected country will send an official request to international organizations, which will immediately consider it and take action. This request will be send to all member countries of the international bodies, of civil protection and humanitarian aid. Then, all countries will answer with possibilities of help. Mostly, they will offer help in terms of human resources with equipment. It is very important that the affected country or countries precisely define what is needed. For example, they need four water rescue teams with boats and engines or divers with all necessary equipment. Finally, the international organizations will answer the affected country as to what they can provide, as international help. Practice shows that many countries want to help on the basis of social and human responsibility.

The model will focus on providing international help, in term of rescuers with appropriate equipment. A general model will be developed and presented, that then can be adopted.

Table 1
Example of table form of the request for the international help

	Country that providing assistance 1 (CPA1)	Country that providing assistance 2 (CPA2)	Country that providing assistance 3 (CPA3)	Country that providing assistance 4 (CPA4)	Number of rescuers - Demand
Affected country – Area 1 (A1)	c11	c12	c13	c14	D1
Affected country – Area 2 (A2)	c21	c22	c23	c24	D2
Affected country – Area 3 (A3)	c31	c32	c33	c34	D3
Affected country – Area 4 (A4)	c41	c42	c43	c44	D4
Affected country – Area 5 (NWD)	c51	c52	c53	c54	D5
Number of rescuers - Offer	O1	O2	O3	O4	

For the purpose of development, a general model is used in the previous table. Columns of the table represent m countries (in this case, four) which have a possibility to provide help. Countries are marked with CPA1, CPA2, CPA3 and CPA4 whose offers are expressed with number of rescuers and labelled as O1, O2, O3 AND O4. Rows of the table represent n areas of the affected country that need help in terms of rescuers with specialized equipment (points of demand) labelled as A1, A2, A3, A4 and A5. Needs of the affected areas are expressed by known numbers of demand D1, D2, D3, D4 and D5 respectively. Having in mind that one of the goals is to minimize costs of engagement of international rescuers, costs of daily engagement (daily allowances) are taken into account. Daily allowances per one rescuer, from different countries, which should be deployed, marked with c_{ij} from each point CPA $j = 1, 2, \dots, 4$ to any point A1, A2, A3, A4 and A5, $i = 1, 2, \dots, 5$. The daily allowances of rescuer engagement include different costs, such as, transportation costs, food costs, accommodation costs and costs for using specialized equipment. The transportation costs include the expenses of transportation from domestic to affected country and back, divided on whole days of stay in the affected country, plus daily transportation costs. Also, the number of rescuers which will be deployed from some country to the affected country will be labelled as x_{ij} .

We want to minimize the total cost of deploying international help, in terms of rescuer, with specialized equipment and at the same time, to satisfy request from affected country. Moreover, the optimization is also based on constraints that all domestic rescue forces should be used and that international teams cannot be separated. Before defining the model, it is important to define the difference between opened and closed transportation problems. If the demand and offer are equal, the problem will be closed. Otherwise, as in our case, if the demand and offer are not equal the transportation problem will be open and it can be closed by adding fictitious points. The fictitious points will be domestic rescue powers.

So, regarding the topic of this paper and previously defined minimization goal, the model will be defined as:

$$F = \sum_{i=1}^m \sum_{j=1}^n cij \quad (1)$$

subject to:

$$x_{11} + x_{12} + \dots + x_{14} + ffp_{15} = D_1 \quad (2)$$

constraint to demand of the first area of the affected country

$$x_{21} + x_{22} + \dots + x_{24} + ffp_{25} = D_2 \quad (3)$$

constraint to demand of the second area of the affected country

$$x_{31} + x_{32} + \dots + x_{34} + ffp_{35} = D_3 \quad (4)$$

constraint to demand of the third area of the affected country

$$x_{41} + x_{42} + \dots + x_{44} + ffp_{45} = D_4 \quad (5)$$

constraint to demand of the fourth area of the affected country

$$x_{51} + x_{52} + \dots + x_{54} + ffp_{55} = D_5 \quad (6)$$

constraint to demand of the fifth area of the affected country

$$x_{11} \text{ or } x_{21} \text{ or } x_{31} \text{ or } x_{41} \text{ or } x_{51} = O_1 \quad (7)$$

constraint to offer of the first country including teams cannot be separated

$$x_{12} \text{ or } x_{22} \text{ or } x_{32} \text{ or } x_{42} \text{ or } x_{52} = O_2 \quad (8)$$

constraint to offer of the second country including teams cannot be separated

$$x_{13} \text{ or } x_{23} \text{ or } x_{33} \text{ or } x_{43} \text{ or } x_{53} = O_3 \quad (9)$$

constraint to offer of the third country including teams cannot be separated

$$X_{14} \text{or} X_{24} \text{or} X_{34} \text{or} X_{44} \text{or} X_{54} = O4 \quad (10)$$

constraint to offer of the fourth country including teams cannot be separated

$$fp_{15} + fp_{25} + fp_{35} + fp_{45} + fp_{55} = FP \quad (11)$$

constraint to engagement of domestic rescue powers

$$x_{ij} \geq 0 \quad (12)$$

The objective function (1) is to minimize the total costs of deploying rescuers as one kind of international help to the affected country. The constraints (2, 3, 4, 5, 6) shows the potential demand satisfaction of the affected areas of country i ($i=1,2,\dots, 5$). These constraints include domestic rescue powers as fictitious points (fp_1, fp_2,\dots,fp_5). Constraints (7, 8, 9, 10) show the potential offer of country that express their will and possibility to help the affected country and areas. So, X_{ij} should be equal with O_j and there is no possibility to have sum of x_{ij} which finally will be equal with offer from different countries. Practically, these constraints include fact that international rescue team cannot be separated and all rescuers should be deployed in same affected area. The constraint (11) refers to request that all domestic rescuers should be used. Finally, the natural constraint (12) refers to rule that the number of rescuers cannot be less than zero.

For solving this transportation problem it is important to predefine the costs in euro, per day, of rescuer deploying and working in the affected country. Then, the model will seek to find an optimal solution, that meets the previously defined constraints. That means, the minimized costs of international help and deployment of rescue teams, engagement of all domestic rescue forces and no single separated international team.

4 Results

The General Adjusted Transportation Problem, which was developed in the previous chapter, will be adopted through the case study, made from the example of the May 2014 flood, in the Republic of Serbia. During these catastrophic floods, the Republic of Serbia could not cope with the severe consequences and the Serbian Government made the decision to send the official request to international organizations and the European mechanism of civil protection. Also, the help was requested from the nearest countries based on bilateral and multilateral agreements. Immediately, these requests were considered and many countries, through different channels, offered help. It was good news for the Serbian authorities. But, the problem occurred, due to large number of countries that wanted to help. Decision makers had complex tasks in making selection of teams with different specializations and from different parts of the world.

In this chapter, using the general adjusted transportation model for optimizing provision of the international help, in case of emergency situations, we will show the usefulness of this model, for making optimal decisions. Finally, tables shown in the following part of the paper represent the application of the adjusted transportation problem.

In the beginning of the case study, it should be noted that the countries which have possibilities to provide help, in terms of rescuers with specialized equipment, previously defined the exact costs of deploying the teams. In this case, the biggest influence on total costs will have the distance between the affected Serbian areas and country that will provide help.

The offer and demand for rescuers, as a kind of the international help and the daily allowance, of deploying and rescuing, in euro are shown in Table 2.

Table 2

Offer, demand and daily allowance in case of providing international help in the Republic of Serbia during May floods 2014

	Country that providing assistance 1 – Russian Federation	Country that providing assistance 2 - Montenegro	Country that providing assistance 3 - Romania	Country that providing assistance 4 - Turkey	Number of rescuers - Demand
Affected country Serbia– Macva area – (MA)	80	90	90	90	250
Affected country Serbia – Pcinja area – (PA)	40	60	70	60	220
Affected country Serbia – Kolubara area (KA)	70	90	90	70	200
Affected country Serbia – Sumadija area (SUA)	70	100	100	110	250
Affected country Serbia – Srem area (Sa)	90	120	70	80	200
Number of rescuers - Offer	250	150	180	200	

Before the beginning of the implementation of the transport model, it is necessary to define what the specific fields in the table mean. For example, in the cell which merges the Russian Federation, as a providing country, and Macva area as the affected part of Serbian territory, is written 80. This means that daily allowance of one Russian rescuer engagement, in the Macva area in Serbia is 80 euro per day. The main goal of the exploited model, is to satisfy the demand for rescuers with predefined offer and to minimize total costs, respecting all constraints. Constraints will be same as in the general model and refer to fact that international teams cannot be separated and that all domestic rescue forces have to be used. Having in mind different types of transportation, Romanian and Montenegrin rescuers used land transport and the Russian and Turkish rescuers used the airport in the city of Nis. This first table in the case study represents the opened transportation problem. The difference is 340 rescuers between demand and offer. More precisely, demand is 340 rescuers higher than the international offer. In next table will be shown the first results of the use of the adjusted transportation model for finding the optimal schedule of providing the international help.

Table 3

The first results of using the transportation model

	Russian Federation (number of offered rescuers 250)	Montenegro (number of offered rescuers 150)	Romania (number of offered rescuers 180)	Turkey (number of offered rescuers 200)	F (difference 340 rescuers) – host nation capacities	The row difference
Macva area – (MA) – 250 rescuers need	80	90	90	90	0	10,10,0,0
Pcinja area – (PA) 220 rescuers need	40	60	70	60	0	20
Kolubara area (KA) 200 rescuers need	70	90	90	70	0	0,0,20
Sumadija area (SUA) 250 rescuers need	70	100	100	110	0	30,30,0,0
Affected country – Srem area (SA) 200 rescuers need	90	120	70	80	0	10,10,10,50
The column difference	30,0	30,0,0,10	0,20,20,20	10,10,10		

The next step is, the application of transportation model procedure. First, we have to find the two smallest numbers, in each row and each column, which relate to the amount of the daily engagement. When we find that, the next step is to make difference between those numbers. For example, in the first row, which relate to the Macva area – (MA) the two smallest numbers are 80 and 90. The difference between them is 10. This is the first number which is written in the first cell of column “the row difference”. The procedure will be repeated for all other rows and columns. Than we have to find the biggest number between those that relate to “row and column” difference. In the previous example, it is 30, but on few positions. In this case, we choose to solve the first column. In the first column, we are looking for the lowest number. This is 40. Finally, we calculate how to satisfy demand and offer it in a cell where the daily engagement is the lowest number, because the model goal is minimization. Pcinja area needed 220 rescuers, and the Russian Federation could provide 250. So, it is possible to satisfy whole demand and deploy 220 rescuers from Russian Federation to Pcinja district. As a difference, it would leave 30 rescuers from Russian Federation, free. The demand of Pcinja district is fully satisfied. The previously described process will be continued until the demand and offer are not the same. In cases, where the demand was bigger than the offer, we have to add one more column, marked with F, which describe that the difference and will be solved with the host nation capacities and Serbian rescue power. The constraint related to column F is that all domestic rescue teams have to be engaged. But, the problem is that team cannot be separated, as is in this case with Russian rescuers. So, this result cannot be used as optimal and we have to continue the process.

The next step is to define the level of rank. This is also part of standard solving of transportation problem. A transportation problem’s solution has $m+n-1$ basic variables, (where ‘ m ’ and ‘ n ’ are the number of rows and columns respectively) which mean that the numbers of occupied cells in the initial basic solution are one less than the number of rows and number of columns. When the number of occupied cells in an initial basic solution is less than $m+n-1$, the solution is called a degenerate solution [18]. The rank is calculated as:

$$R = m + n - 1 \quad (13)$$

where m is number of countries which provide help plus domestic capacities, and n is number of the affected areas.

In this case, the rank is 9. It is necessary, in order to have optimal result, to check if the rank is equal with numbers which are marked in squares and that is popularly called “stones” in the transportation problem. So, we have 8 stones and the rank is 9. They are not equal and we have to continue the process to find the optimal solution. Before next step it is important to include the \mathcal{E} . This is an additional tool, without which, we cannot find the optimal result. With the aim to solve degeneracy, the general transportation model needs to allocate an infinitesimally small amount \mathcal{E} to one of the independent cells. More precisely,

to allocate a small positive quantity ϵ to one or more unoccupied cells that have the lowest transportation costs. Also, the value of ϵ is approximately zero.

Table 4
The second results of using the transportation model T0

	Russian Federation (number of offered rescuers 250)	Montenegro (number of offered rescuers 150)	Romania (number of offered rescuers 180)	Turkey (number of offered rescuers 200)	F (difference 340 rescuers) – host nation capacities	U _i
Macva area– (MA) – 250 rescuers need	80 10	90	90 20	90 0	0	0
Pcinja area – (PA) 220 rescuers need	40	60 150	70	60	100	-30
Kolubara area (KA) 200 rescuers need	220 20	90 20	90 40	70 ϵ	0 20	-20
gSumadija area (SUA) 250 rescuers need	70	100 10	100 30	100 20 200	0	0
Affected country – Srem area (SA) 200 rescuers need	30	120 30	70 180	80 -10	220 20	0
V _j	70	90	70	90	20	

Now we have to assign the potential of the affected areas. It will be U_i where i=1,2,3..5. After, the potential of the providing countries is V_j where j=1,2,3..5. Both potentials are calculated in the same way as in the previous table. U_i potential is same as the “row difference” in the previous table. V_j potential is same as, ”column difference” in the previous table. The task is to define the values of base variables:

$$c_{ij} = U_i + V_j \tag{14}$$

and non-base variables

$$d_{ij} = c_{ij} - U_i - V_j \tag{15}$$

In this case, the base variables are the daily allowances for international rescuer engagements in some of the affected areas. The values d_{ij} are non-base variables as well as numbers below of diagonal in cells. When this process is finished we have to look are there and d_{ij} value less than zero. In this case there is d₅₄ which is -10. None of the d_{ij} values can be less than zero. So, the result is not optimal and

we have to continue the transportation problem. Before the next table, we need to connect stones or base variables and to find solutions so all non-base variables, are greater than zero. In next table, we will change the negative value of d_{54} with \mathcal{E} and repeat the same process as in previous table.

Table 5
The second results of using the transportation model T1

	Russian Federation (number of offered rescuers 250)	Montenegro (number of offered rescuers 150)	Romania (number of offered rescuers 180)	Turkey (number of offered rescuers 200)	F (difference 340 rescuers) – host nation capacities	U _i
Macva area – (MA) – 250 rescuers need	80 10	90 150	90 20	90 0	0	70
Pcinja area – (PA) 220 rescuers need	40	60	70 30	60 20	100 30	40
Kolubara area (KA) 200 rescuers need	70 10	90 10	90 20	70	0 10	60
Sumadija area (SUA) 250 rescuers need	70	100 10	100 30	110 20	200 0	70
Srem area (SA) 200 rescuers need	90 20	120 30	70	80 \mathcal{E}	0 220	70
V _j	0	20	180	10	-7 20	

Now, all non-base variables are positive. So, we have the first optimal solution. But, because one of the non-base variables is equal with zero ($d_{22}=0$) we have multiple optimal solution and we will once more repeat process in previous table. The problem with this result is again Russian team's separation on two parts which is contrary to constraints.

Table 6
The second results of using the transportation model T2

	Russian Federation (number of offered rescuers 250)	Montenegro (number of offered rescuers 150)	Romania (number of offered rescuers 180)	Turkey (number of offered rescuers 200)	F (difference 340 rescuers – host nation capacities)	U _i
Macva area – (MA) – 250 rescuers need	80 10	90 0	90 20	90 20	0 250	70
Pcinja area – (PA) 220 rescuers need	40 70	60 150	70 30	60 10	0 20	40
Kolubara area (KA) 200 rescuers need	70 10	90 10	90 30	70 200	0 10	60
Sumadija area (SUA) 250 rescuers need	70 180	100 10	100 30	110 30	0 70	70
Srem area (SA) 200 rescuers need	90 20	120 30	70	80 ε	0	70
V _j	0	20	180	10	20	

Finally, the last table shows the recommended decision. But, constraints of general model are not satisfied. The Russian team is separated and domestic capacities (column F) are not fully engaged. In next step we have to merge the Russian team. Having in mind all constraints, the Russian team will be connected in the Sumadija area, but not at full capacity. 40 Russian rescuers from Pcinja area

will be merged with rest of team in the Sumadija area. Demand in the Pcinja area will be satisfied by the Montenegro team, as well as, with domestic rescuers from Pcinja area (20 rescuers), Kolubara area (10 rescuers) and from Sumadija area (40 rescuers). Demand of Sumadija area will be satisfied with 220 Russian rescuers and 30 domestic rescuers. In this way, all constraints are satisfied. Moreover, the Russian Federation will deploy fewer people, 220 rescuers instead of the 250 offered, because all the domestic forces are engaged.

5 Discussion

The final result in table 6 and explanation in the last paragraph of previous chapter illustrates the aim of this research, where all demands for rescuers, in the affected districts of the Republic of Serbia are satisfied. Also, we use all offered rescuers, from minimal countries and deployed them with minimal costs, which also was one of the goals. Only the Russian offer was not completely used considering that 220 rescuers will be deployed instead of the 250 offered. According to the adjusted transportation model, goal function and constraints, which are used as part of the case study, we obtained optimal solutions. The Russian Federation, according to the last solution, deployed 220 rescuers to Sumadija district, with the daily costs of 70 euro per day, which is not lowest, but is only 10 euro more than the lowest daily price of engaging Russian rescuers. Moreover, in this case, is the satisfied constraints, that teams cannot be separated and deployed to two different sides. One of the main obstacles, when some countries send their rescuers to help somewhere abroad, is the dividing them on two different sides. Then, Romania deploys all of rescuers to Srem district with the lowest daily price, exactly 70 euro. Montenegro deploys all of rescuers Pcinja district with the lowest daily price – 60 euro. Finally, Turkey will deploy all of rescuers to the Kolubara district with daily price which is not lowest, but is only 10 euro more than the lowest daily price of engagement of the Turkish rescuers. Now, by comparing the last solution and the first solution in the table 3, we can conclude that all of constraints are satisfied. Teams will not be separated and domestic rescue teams are fully engaged.

In the case where many countries want to help an affected country with rescue teams and equipment, using an adjusted transportation problem will facilitate this process. Comparing the first and last option, the costs are the same – 51,000 euros. However, the final solution satisfies all of constraints.

Manual use of a linear programming or transportation problem, without any software, for decision makers in this area, will be too complex. So, the one solution for both obstacles, either for “force-device” calculation model or manual use, is an appropriate software tool. LINDO and LINGO, are examples of software products that can be a useful base for making an appropriate program, or probably sufficient for reducing costs of providing international help in the case of emergency situations [19].

Conclusion

The purpose of this paper is to give a recommendation for using transportation problem solving tools to create a model for providing international help, when natural or man-made disaster occur. In this case, in terms of the rescuers with specialized equipment, with the main goal, to minimize costs. Moreover, the main contribution, as well as, the novelty, is that through this paper, is developed an adjusted transportation model, that should be used by those responsible, to make optimal decisions on provision of international help, in case of emergency situations, with emphasis on deployment of rescue teams with equipment. A special part of this paper belongs to the practical use of a transportation model, which is shown through one case study, belonging to the May floods of 2014, that hit the Republic of Serbia. The one of main finding is that a model is applicable for any kind of emergency situation, caused by natural or man-made disaster. In all cases, is possible to compare requests from an affected country, with offered help. The difference appears only in the numerical indicators for the number of offered rescuers and for the needs of the affected country. Lessons learnt after using the adjusted transportation model, is the need to include all of the practical rules, when the international help is provided. First, all of the domestic teams, in the affected country, have to be engaged. Only in this case, when the national resources are overwhelmed, can the affected country send requests for international help. Secondly, international rescue teams cannot be separated during their work in the affected country. Both of these two constraints are used in the adjusted transportation model, developed in this paper. Maybe the only obstacle now is how to adopt this operational research, as a unique helpful tool, for this topic. In practice it is only possible to recommend, in some general guidelines, and present it at conferences and seminars, to decision makers of the international environment and emergency management community.

Future work will focus on special software, which can be made for this purpose, in connection with the adjusted transportation model. This software should be developed to include existing data, for possible provision of international help, such as, information of currently available teams and equipment. Then, this software, based on predefined constraints, will automatically solve deployment problems and provide optimal results.

Acknowledgement

Paper is result of research within the project No. 47017 Security and protection of organization and functioning of the educational system in the Republic of Serbia (basic precepts, principles, protocols, procedures and means) realized on the Faculty of Security Studies in Belgrade and financed by Ministry of Education and Science of the Republic of Serbia.

References

- [1] Mladjan, D. and Kekić, D.: Emergency: A contribution toward conceptual determination of security (orig. Vanredna situacija – prilog konceptualnom odredjenju bezbednosti), NBP – Journal of Criminalistics and Law, Vol. 12, No. 3 (2007) pp. 61-83

-
- [2] The European Commission: EU Civil Protection, Echo Factsheet, 2017
- [3] The United Nations: OCHA Brochure, The United Nations Office for the Coordination of Humanitarian Affairs, New York (2016)
- [4] Milenković, M. and Kekić, D.: INSARAG, Natural Disasters and Emergencies (orig. Elementarne nepogode i vanredne situacije) Institute of Comparative Law, Academy of Criminalistic and Police Studies, Belgrade, (2015) pp. 46-68
- [5] Pálfi, J. and Holcsik, P.: Emergency Situations Management with the Support of Smart Metering, Acta Polytechnica Hungarica, Vol. 13, No. 3 (2016) pp. 195-206
- [6] Jan Sørensen; Trond Vedeld; Marit Haug; Natural hazards and disasters Drawing on the international experiences from disaster reduction in developing countries, Norwegian Institute for Urban and Regional Research (NIBR) (2006) pp. 16-17
- [7] Law on Emergency Situations, Official Gazette of RS, No. 111/2009, 92/2011 and 93/2012
- [8] Safer, M., S. P. Anbuudayasankar, S. P., Balkumar, K., Ganesh, K.: Analyzing transportation and distribution in emergency humanitarian logistics, 12th Global congress on manufacturing and management, Elsevier Ltd. (2014) pp. 2248-2258
- [9] Jonsson,P., Mattsson,S.: Läran om effektiva materialflöden, Lund Studentlitteratur (2005)
- [10] Shapiro, D. R., Heskett, L. J.: Logistics Strategy: Cases and Concepts, St.Paul, Minn: West (1985)
- [11] Balcik, B. and Beamon, B. M.: Facility location in humanitarian relief, International Journal of Logistics Research and Applications, Vol. 11 (2008) pp. 101-121
- [12] Stevenson, W., and Ozgur, C.: Introduction to Management Science with Spreadsheets, New York: McGraw-Hill (2006)
- [13] Pan American Health Organization: Humanitarian Assistance in Disaster Situations A Guide for Effective Aid (1999)
- [14] Chan, H. H. et al. Real-time optimization for disaster response: A mathematical programming approach, International Journal of Big Data (ISSN 2326-442X) Vol. 2, No. 2 (2015)
- [15] Baraka, M., Yadavalli, S., Singh, R.: A transportation model for an effective disaster relief operation in the sadc region, The South African Journal of Industrial Engineering, Vol. 28, No. 2 (2017) DOI:10.7166/28-2-1311
- [16] Afroz, S. and Hasan, B. M.: A Computer Oriented Method for Solving Transportation Problem, Dhaka University Journal of Science, Vol. 63, No. 1 (2015) pp. 1-7

- [17] UNDP, European Commission and World Bank: Serbia floods (2014)
- [18] The Institute of Chartered Accountants of India: Advanced Management Accounting, The Transportation Problem, Chapter 11 (2008)
- [19] Stevanović, O., Kekić, D., Kónya, V., Milenković, M.: The Use of Linear Programming for Determining Number of Fire - Fighters on Shifts in Case of Special Events, *Acta Polytechnica Hungarica*, Vol. 13, No. 5 (2016) pp. 155-167

Interdisciplinary Survey of Fault Localization Techniques to Aid Software Engineering

Árpád Beszédés

University of Szeged, Department of Software Engineering
Árpád tér 2, H-6720 Szeged, Hungary
beszedes@inf.u-szeged.hu

Abstract: Fault localization (narrowing down the cause of a failure to a small number of suspicious components of the system) is an important concern in many different engineering fields and there have been a large number of algorithmic solutions proposed to aid this activity. In this work, we performed a systematic analysis of related literature, not limiting the search to any specific engineering field, with the aim to find solutions in non-software areas that could be successfully adapted to software fault localization. We found out that few areas have significant literature, in this topic, that are good candidates for adaptation (computer networks, for instance), and that although some classes of methods are less suitable, there are useful ideas in almost all fields that could potentially be reused for software fault localization.

Keywords: faults/defects/failures; fault localization; software fault localization; literature review; method assessment

1 Introduction

Our everyday lives are driven by complex systems; we are directly interacting with some of them, while others support background technologies in diverse industrial areas [1]. These complex systems may be mechanical, electrical, software-driven, or any combination thereof, and are developed and produced by the respective engineering disciplines. These systems are often mission, safety or business critical, and every effort is made to avoid failures in them. Failures can cause damage to the environment, people's health and lives, or the operation of businesses and governments. Hence, failures and the underlying faults are a high priority concern.

Among the many different engineering areas that deal with complex systems, there is one common subtopic, the central theme of this article, *fault localization*. Without loss of generality, fault localization means identifying components (parts, modules, software code parts, etc.) of the system that are responsible for a specific

observed failure. Fault localization as a discipline is given a high priority in many fields, especially in the case of highly critical systems.

In this paper, we explore semi-automatic fault localization techniques from various domains, and aim at producing an interdisciplinary analysis of the area. Our goal is specific, though. The background area is software engineering, and our research agenda deals with enhancing existing techniques and providing new approaches in the field of *software fault localization* [2] [3]. To this end, the primary goal of this survey is to provide a systematic analysis of fault localization techniques from non-software domains and discuss their possible adaptation to and implementation in software fault localization.

In any of the mentioned engineering areas, systems tend to be large and complex, and they are often connected to each other, forming even more complex systems-of-systems [4]. This has the implication that, upon occurring failures, it may be very difficult to localize their source (root cause). Hence, various fields have developed algorithmic approaches to automate the fault localization process.

Naturally, each field deals with its peculiarities and many of the techniques are domain-dependent, yet we found out that there are some similarities across disciplines. Furthermore, some of the methods are generic and could be applied, theoretically, to any engineering field and fault localization problem.

Software fault localization is a relatively young area compared to, for instance, aerospace or electronics. Yet, there is already a large literature covering many different subtopics [2] [3]. A lot of research has been performed to design effective fault localization algorithms and propose their use in different phases of the software process, most notably debugging. However, related research suggests that the practical applicability of research results in this area is still limited [5], and further research is needed to achieve more widespread use of automatic software fault localization by practitioners.

It is noticeable that existing software fault localization techniques concentrate around a relatively small number of fundamental approaches with little overlap between them [2]. This motivated the present work: to investigate other engineering fields and find out if they employ techniques that could be adapted to software and hence advance the state-of-the-art in this field.

This paper is a first attempt to investigate the applicability of fault localization methods to software from other fields; we are not aware of any similar research. Our preliminary investigations show that there are promising related approaches, but we also found that in some cases there are barriers to the adoption of such techniques. This is due to fundamental differences in how these systems (software and non-software) are described and handled (for example, if a detailed behavioral model is required). In many other cases, however, the techniques or some underlying ideas could be successfully adapted to software.

The paper is organized as follows. In Section 2, we briefly overview the terms used in the remaining parts of the paper. Section 3 deals with the assessment criteria we used for the analysis of the literature. The assessment results are presented in Section 4, while Section 5 contains their evaluation. Section 6 concludes this work.

2 Background

One of the main difficulties in a cross-disciplinary analysis of a specific topic is the diversity of the used terms. Often, the same concepts are referred to by different terms, and specific terms may have different meanings in different technological areas. In this work, we came across the following areas: software technology, computer networks, electric engineering, aerospace, among others. In the following, we overview the main constituents of a general fault localization approach, and the terminology we will use to describe it.

System and its components. Since this paper deals with many different areas, a system may refer to any complex artifact that performs a specific task [1]. It may either be a mechanical, electrical, chemical, computer software, etc. system that is composed of specific, interacting components. Often, a complex system includes components of different types, e.g. interacting mechanical and electrical, or computer based using hardware and software components. A system is often described using a domain specific *model*, which is then used in the fault localization process.

Fault. Without loss of generality, in this paper, fault refers to a defective component (or a set of defective components) of a system [6]. A fault may be defined at different granularity levels, depending on the domain and fault localization method. A fault may be present due to a design or implementation error made by a human or other external entity, or may be developed during operation by natural wear or physical damage. (This, of course, does not apply to software, for instance.)

Fault identification. This refers to the (systematic or incidental) process of discovering that there is a fault in a system. This process merely proves that there is at least one fault, and does not necessarily shows its exact location and context.

Execution and observation. A fault in a system may be identified by merely analyzing the system's components by automatic or manual means (we call this a *static* approach), or by executing (using) it and observing its behavior. Execution, in a general sense, means using the system in its intended or test environment and usage scenarios, either in its entirety or using only some of its sub-components. A fault identified in such a way will be referred to as using the *dynamic* approach. Execution and observation may mean diverse things in the case of different

systems, such as real-time observing a working system in live environment, running software test cases, probing a network with test packages, etc.

Test. An individual test will mean any atomic execution of the system whose behavior can be observed, measured and interpreted. Alternatively, a system may be statically tested by analyzing the components. This, again, can be very diverse in the different domains.

Intended (or expected) behavior. This will refer to a type of execution of the system, which conforms to a set of explicit or implicit behavioral requirements. In other words, it is the behavior when all of the system's components work correctly. Some parts of the intended behavior are defined by a *behavioral model* (documentation, or formal model), while in other cases undesired behavior is documented (such as possible failure modes), or it may even refer to implicit, undocumented, expected behavior.

Failure. Based on the previous, a failure of a system should mean any observed behavior which is different from the intended one [6]. Note, that failure may mean many different things and can be classified according to severity starting from minor glitches, through functional and non-functional issues (for example, performance) to serious malfunctions. (The static fault identification does not require the manifestation of a failure.)

Fault localization. Finally, fault localization refers to any automated or semi-automated process whose goal is to select a sub-component or set of components of a complex system, which are most probably responsible for a set of observed failures or identified (but not yet localized) faults.

In the case of various domains, fault localization may mean different concrete things but a basic approach is to perform a set of tests on the system, observe its behavior and, based on the failures, use an algorithm to narrow down the possible causes to specific sub-components of the system. In this process, a behavior model may or may not be required, and in some cases the tests may be performed statically, as discussed above.

The different fault localization approaches can have various properties that determine its effectiveness and usage efficiency. In this context, *effectiveness* means how successful the method is in localizing the fault (successfulness can, in turn, mean different things but usually refers to how many of them and how precisely the location of the faults are found). *Efficiency*, on the other hand, means any practical property of the method that determines its execution time, complexity, storage requirement, or any other aspect which is important for its usability.

3 Assessment Criteria

The process for identifying the corresponding research reports and their selection was the following. In the first phase, we used general and research oriented search engines and research repositories, which included google, google scholar, ResearchGate, Mendeley, and Scopus. We did not use generic search terms like “fault localization” alone because these produced too much irrelevant results. Instead, we added specific keywords that we expected to be relevant fields for our search: networks, electronics, engineering, operations, systems, etc. We also applied different variations and synonyms to the term, which included localizing faults, failure diagnosis, problem diagnosis, error localization and similar terms.

We then restricted the search results to publicly available full-text scientific publications. We aimed at limiting the results to publications that appeared in peer-reviewed journals or conferences, however there were few exceptions such as doctoral theses and technical reports. The next filtering, we applied was to limit the list to papers that correspond to some of the following categories: software-related, generic algorithms, methods in engineering fields that we expected to be relatively easy to adapt to software-related artifacts. For example, pure mathematical methods, methods used in programming education, or approaches in non-related scientific branches like biomedicine, navigation, linguistics or other, were removed.

In the next phase, we performed a lightweight “snowballing” with the identified papers: considering the referenced works for new candidates. Finally, we consolidated the results by organizing the works by specific research groups or authors and concentrating on a few relevant reports by the same team.

In the next phase, we started the classification of the papers based on the criteria set forth in this section. In this phase, several papers also dropped out because they were difficult to categorize according to the criteria (mainly due to the *fundamental area* category as described below). Also, the criteria had to be modified slightly during this phase.

Fundamental area. The main classification direction was the fundamental area in which the method is applied. To enable easy further processing of the methods, we decided to use a very simple classification in this respect. We have the following categories: *software*, *networking*, *other engineering* and *various/generic*. The description of the methods in Section 4 is organized along these categories.

Since our goal was to identify potential approaches from other areas different from software faults, the methods we include belonging to the *software* category are only the most important, basic approaches, which are provided for reference.

We soon realized that there exists a large amount of publications that deal with fault localization in *computer networks*, hence we established a separate category for this area.

The *other engineering* category includes all methods that belong to a specific engineering field other than software or networks. In the corresponding table in Section 4, we will denote the specific field in question.

Finally, there are some approaches that are not limited to any specific field (although some of them include one or more example applications); in this sense, they are *generic*. We used the same category to denote methods belonging to some other *various* fields.

The other classification criteria we used for each method are the following:

Base method. This refers to the fundamental approach (mathematical model, algorithm) on which the method is based on. Of course, many methods are using complex solutions and it is difficult to categorize them into a single approach, but we managed to classify most of the methods into one of the following: *Machine Learning* including any subfield thereof, *Statistics*, which are based on statistical analysis of the failures, tests, etc., *Entropy*, a special case of statistics which also includes probabilistic approaches. Finally, *Model* refers to model-based approaches that include various types of models such as mathematic structures or engineering descriptions of the systems. In some cases, a combination of the previous was applied in which case we used *Combined*. Finally, if the base method could not be determined or would be very different than the mentioned ones, we used *Other*.

Faults. It is an important property of a method if it relies on an assumption that there is a single fault in the system, or it can handle (or is designed to handle) multiple faults occurring at the same time. Therefore, we use the *Single* and *Multiple* categories for this aspect.

Base Data. The next category we used is the basic type of data the method relies on for performing the fault localization computation. We found that most of the approaches are using either a *Graph* representation of the elements, probes, tests, etc., or they are represented in a *Matrix* format (such as rows containing the probes and columns the elements on which localization is to be performed with test results in the cells). In a number of cases, the base data is much more complex, in which case we used *Complex*. Finally, some approaches use a *Domain specific* data representation.

Behavior model. This category deals with the question if a behavioral model is required to perform fault localization. Such a model describes the expected behavior of the system. In simple cases, the tests (or probes) are providing simple pass/fail answers, but in other cases, a more complex model is needed. We used *Yes* or *No*.

Empirical. This category classifies the methods according to whether they include empirical measurements, and if yes, what kind of. The *Theory* category means that only theory is described, *Simulated* refers to a case when simulation data were used in the experiments, while in the case of *Real*, real data was used.

Data set. If the method included any kind of experiments, this category will provide the amount of data they were executed on. *Example* means that only toy examples were used, *Small* refers to a realistic but small data set, while *Large* includes any real data that can be treated large but is limited to a small number of projects or sets. Finally, *Mass* was used when an automated method was used to collect mass amounts of data from some repositories.

Availability. This category deals with the availability of the underlying information of the method. Namely, if only the *Implementation* or the measurement *Data* are available, *Both* of these or *None*.

For each of the criteria from above, if it cannot be interpreted for a specific method, we will use *N/A* to denote this situation.

4 Methods by Areas

In this section, we present the results of our assessment of fault localization techniques literature. We list the identified papers along with the properties following the categorization presented in the previous section. This section is organized into subsections by the *Fundamental area* category defined above. Each subsection is composed of a table of the same structure: we list the papers with their authors and publication year noted to help easier identification, and make a brief note of the assessment results for each classification aspect. An exception is the *Other Engineering Fields* category, in which case an additional column is used to indicate the specific field.

4.1 Software

Research related to fault localization in computer software is a large and diverse area. It is not the purpose of the present paper to provide a comprehensive overview of this literature, as the goal is to identify method *not related* to software. For an interested reader, we refer to the excellent surveys of Wong *et al.* [2] and Parmar and Patel [3]. Nevertheless, we include several works related to this area (Table 1), which we think are important representatives of the field. These approaches are diverse enough to serve as examples of the main techniques for software fault localization.

The basic goal of any software fault localization approach is to identify the location of software defect(s) in the source code given one or more faulty executions of the system. In software testing, one just shows that there is a defect somewhere in the system, and it is the task of fault localization to identify the exact point of the fault, typically in the source code.

A fundamental approach to software fault localization is to observe the behavior of distinct test cases and, based on their outcomes and their interaction with the system, compute the most suspicious code elements to contain the defects.

Table 1
Software fault localization techniques

<i>Paper</i>	<i>Base Method</i>	<i>Faults</i>	<i>Base Data</i>	<i>Behav. model</i>	<i>Empirical</i>	<i>Data set</i>	<i>Availability</i>
Abreu et al., 2007 [7]	Combined	Multiple	Complex	No	Real	N/A	Both
Abreu et al., 2009 [8]	Combined	Multiple	Complex	No	Simulated	Example	Implementation
Artzi et al., 2010 [9]	Model	Multiple	Matrix	No	Simulated	Example	None
Christ et al., 2013 [10]	Other	N/A	Domain specific	No	Theory	N/A	Implementation
Pearson et al., 2017 [11]	N/A	N/A	N/A	N/A	Real	Large	Data
Ravindranath et al., 2014 [12]	Model	Multiple	Matrix	Yes	Real	N/A	Data
Renieris et al., 2003 [13]	Mach. learn.	Single	Complex	No	Simulated	Small	Data
Wang et al., 2011 [14]	Mach. learn.	Multiple	Domain specific	No	Simulated	Example	None

4.2 Networking

Fault localization in computer networks is a large and important area as networking technologies are becoming more and more complex as well as the internet itself, and the reliability of computer networks is increasingly important.

In networking, the goal of fault localization is to identify faulty networking elements (“nodes”) such as routers, etc. This is typically done by probing the network with network packages, and based on the responses from the nodes and the routes taken, the faulty nodes are identified.

Table 2 contains the results of our assessment of methods in the computer networking area.

Table 2
Networking fault localization techniques

<i>Paper</i>	<i>Base Meth.</i>	<i>Faults</i>	<i>Base Data</i>	<i>Behav. model</i>	<i>Empirical</i>	<i>Data set</i>	<i>Availability</i>
Aghasaryan et al., 1997 [15]	Model	N/A	Complex	N/A	Theory	N/A	None
Aghasaryan et al., 1997 [16]	Model	Multiple	Complex	N/A	Theory	N/A	Implementation.

Alekseev et al., 2014 [17]	Model	Multiple	Complex	No	Theory	N/A	None
Brodie et al., 2002 [18]	Model	Multiple	Complex	No	Theory	N/A	None
Chao et al., 1999 [19]	Model	Multiple	Domain specific	No	Theory	N/A	Implement.
Chen et al., 2004 [20]	Mach. learn.	Multiple	Domain specific	Yes	Real	N/A	None
Deng et al., 1993 [21]	Mach. learn.	N/A	Domain specific	No	Theory	Example	Implement.
Fecko et al., 2001 [22]	Combined	N/A	Complex	No	Simulated	Example	None
Garshasbi et al., 2013 [23]	Other	Multiple	Matrix	No	Theory	Example	Data
Hood, 1997 [24]	Mach. learn.	Multiple	Domain specific	No	Simulated	Example	None
Kant et al., 2003 [25]	Model	N/A	Complex	No	Theory	N/A	None
Katzela et al., 1995 [26]	Model	Multiple	Domain specific	No	Simulated	Example	None
Kompella et al., 2005 [27]	Model	Multiple	Matrix	No	Simulated	Example	Implement.
Lu et al., 2013 [28]	Model	Multiple	Complex	No	Simulated	Example	Data
Natu et al., 2006 [29]	Other	N/A	Matrix	No	Theory	N/A	Implement.
Natu et al., 2007 [30]	Model	Multiple	Domain specific	No	Theory	N/A	Implement.
Natu et al., 2007 [31]	Statistics	Multiple	Matrix	No	Simulated	Example	Both
Rish et al., 2004 [32]	Other	N/A	Complex	No	Real	N/A	Implement.
Steinder et al., 2004[33]	Model	N/A	Complex	Yes	Simulated	Example	Implement.
Steinder et al., 2004[34]	Model	N/A	Complex	Yes	Simulated	Example	Implement.
Tang et al., 2005 [35]	Model	Multiple	Complex	Yes	Simulated	Example	Both
Traczyk, 2004 [36]	N/A	Multiple	Matrix	No	Simulated	Example	None
Wang et al., 2012 [37]	Combined	Multiple	Complex	No	Simulated	Example	Both
Zhang et al., 2011 [38]	N/A	N/A	N/A	Yes	Theory	N/A	None

4.3 Other Engineering Fields

This category deals with different engineering fields in which some form of automated fault localization is investigated. Faults are possible and need to be avoided or identified in virtually any automatic system, whether it is mechanical,

electrical, logical (software), or even chemical or biological. Some systems are complex and composed of different components of the mentioned types.

Automatic fault localization is used to various degree in these areas, typically based on the criticality of the system. Some areas are particularly notable in this respect, which have a relatively large literature on fault localization. These areas include the aerospace industry (detecting faults in aircraft systems), power electronics (detecting faults and source of outages in electrical networks), electronics (detecting faults in hardware components of computer systems or other electronic devices, most typically in the digital domain). Other areas we encountered include mechanical engineering (detecting faults of rotary machines), oil pipelines (detecting leakage points) and chemistry (detecting faults in chemical plants that implement complex chemical reactions).

We are certain that there may be many other areas that encounter similar issues and have domain-specific solutions to fault localization, but the domains we list in this section illustrate the diversity of approaches used. Interestingly, there are many common basic approaches used in these diverse areas (such as entropy-based and neural networks), which means that they might be good candidates in reusing the methods to software fault localization.

Table 3 contains the results of our assessment of other engineering field methods.

Table 3
Other engineering fault localization techniques

<i>Paper</i>	<i>Base Method</i>	<i>Faults</i>	<i>Base Data</i>	<i>Behav. model</i>	<i>Em-pirical</i>	<i>Data set</i>	<i>Avail</i>	<i>Field</i>
Adamovits et al., 1993 [39]	Model	Multiple	Domain spec.	Yes	Theory	N/A	None	Aero-space
Balaban et al., 2007 [40]	Model	Multiple	Domain spec.	Yes	Theory	N/A	None	Aero-space
Benbouzid et al., 1999 [41]	N/A	N/A	N/A	N/A	Theory	Example	None	Power electr.
Beschta et al., 1993 [42]	Model	Single	Complex	No	Theory	N/A	None	Power electr.
Digernes, 1980 [43]	Model	Single	Complex	No	Simulated	Example	None	Oil pipelines
Dries, 1990 [44]	Model	Multiple	Domain spec.	N/A	Theory	N/A	Im-plem.	Aero-space
Pálfi et al., 2017 [45]	Other	Multiple	Domain	N/A	Real	Small	None	Power electr.
Poon, 2015 [46]	Model	Multiple	Domain spec.	Yes	Simulated	Example	Data	Power electr.
Peischl et al., 2006 [47]	Model	N/A	Graph	N/A	Simulated	Example	None	Elec-tronics
Tanwani et al., 2011 [48]	Model	N/A	Complex	N/A	Simulated	Example	Im-plem.	Power electr.

Tóth et al., 2013 [49]	Other	Multiple	Domain	No	Simulated	Example	None	Mech. eng.
Venkatasubramania et al., 1990 [50]	Model	Multiple	Domain	Yes	Simulated	Example	None	Chemistry
Yan et al., 2014 [51]	Other	Multiple	Domain	No	Real	Example	None	Mech. eng.

4.4 Various and Generic Methods

During the assessment of the identified literature, we encountered several works that introduce a fault localization algorithm, which is theoretically application independent. To a certain degree, these generic methods could be applied to any field, including software. Many of these publications are illustrating the use of the approach in a specific field, but it is generally not discussed to what degree is the method generalizable to other areas.

Some methods listed in this category are purely theoretical and advance a certain mathematical subfield, with no obvious practical application. Hence, the applicability of the methods listed in this section should be carefully investigated to any particular field, notably software faults.

Table 4 contains the associated results of our assessment.

Table 4
Various other fields fault localization techniques

<i>Paper</i>	<i>Base Method</i>	<i>Faults</i>	<i>Base Data</i>	<i>Behav. model</i>	<i>Empirical</i>	<i>Data set</i>	<i>Availability</i>
Frank, 1996 [52]	N/A	N/A	Complex	Yes	Theory	N/A	None
Gertler, 1991 [53]	Machine learning	Multiple	Matrix	No	Theory	N/A	None
Isermann, 1984 [54]	N/A	N/A	N/A	N/A	Theory	N/A	None
Kleer, 2009 [55]	Entropy	Multiple	Domain specific	Yes	Simulated	Large	None
Kleer et al., 1987 [56]	Entropy	Multiple	Domain specific	Yes	Theory	N/A	None
Lerner et al., 2000 [57]	Model	Multiple	Complex	No	Simulated	Example	Implementation
Massoumnia et al., 1986 [58]	Model	Multiple	Complex	No	Theory	N/A	None
Mehra et al., 1971 [59]	Statistics	Multiple	Complex	No	Theory	N/A	None
Olivier-Maget et al., 2009 [60]	Combined	Multiple	Complex	No	Theory	N/A	None

Shchekotykhin et al., 2016 [61]	Model	Multiple	Domain specific	N/A	Simulated	Example	Implementation
Tidiri et al., 2016 [62]	N/A	N/A	N/A	Yes	Theory	N/A	None
Varga, 2003 [63]	Statistics	N/A	Domain specific	No	Theory	Example	None

5 Evaluation

The main goal of the paper was to identify potential approaches from non-software domains that can be successfully adapted to software faults and fault localization. Based on the summaries in the previous chapter, it is not easy to pinpoint only a few candidate methods, rather many of them may provide interesting ideas, even if not the complete method is adapted. In particular, we found the following. Figure 1 contains the overview of the various fields we investigated in this article. The arrows from specific areas to software bugs indicate the level of their applicability (dashed lines = moderate, solid lines = probable).

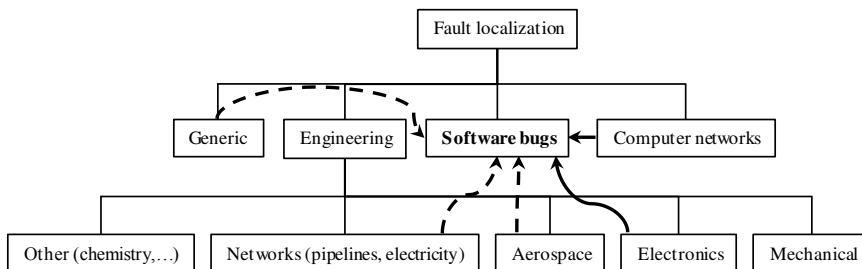


Figure 1

Overview of the investigated fault localization areas and their relation to software faults

5.1 Networking

The most promising techniques for adapting to software faults is the *probing method* in computer networks [36]. A probe is a program that executes on a particular network node and sends commands or transactions to the other elements of the network. Then, the responses are observed and their various properties are measured. From this information, various network issues, bottlenecks and faulty nodes can be estimated. Steinder and Sethi provide a survey of fault localization techniques in computer networks [64].

An interesting property of such network fault localization methods is that an almost direct analogy can be drawn to software fault localization: a network node

corresponds to a software component, a probe can be seen as a test case, and the responses from the network can be identified as the dynamic behavior of the system by executing the test cases. Thus, the traditional *spectrum-based fault localization* methods in software [2, 3, 7] may benefit from advances in probe-based networking fault localization.

For instance, the approaches by Brodie *et al.* [18], and Natu *et al.* [29, 30, 31], provide various optimizations to the basic probing approaches, which are good candidates for adaptation to software faults.

Another common element of network fault localization is the use of probabilistic approaches (such as conditional probabilities and Bayes networks) [17, 19, 34, 35], among others, as well as machine learning [20, 21, 24]. These can be probably adapted to software.

5.2 Other Engineering Fields

Overall, the techniques used by other engineering fields are typically not directly applicable to software faults because of the big differences in the domains. Often, reliable behavioral models are the basis for these approaches which is in many cases difficult to obtain with software. The probabilistic approach used often in some areas may, however, be considered to enhance fault localization in software. Indeed, there are already several enhanced methods in software fault localization that employ conditional probabilities and entropies, such as Abreu *et al.*'s method [7] (also see [2] [3]).

In the *aerospace* industry, use of artificial intelligence, in particular, model based reasoning, seems to be prevalent [39, 40, 44]. Although these approaches seem quite advanced, their application to software fault localization may be limited due to the difficulty of producing a reliable model of the software.

The situation is similar with the *power electronics* area [41, 42, 46], these also frequently utilize various models describing the system. However, they seem to be less complex and more similar to computer networks, hence their applicability may be easier.

Some approaches in fault localization in *electronic circuits* may almost directly be applied because the description of the hardware is done in a similar way to computer software source code [47]. However, often simulation is done based on the circuit model, which is more difficult to employ on software. It is interesting to note, that some techniques that we categorized as "Generic methods" (see next section) have their main application in electronic circuits (Kleer *et al.* [55] [56]), which are based on entropy minimization and probabilistic approach (as with many methods in computer networks).

The other areas we investigated also often use simulation and probabilistic approaches [43], or machine learning with neural networks [50], but in these cases

a model of the system is required as well. Often, advanced concepts are applied in these areas such as Kalman filters to increase the accuracy of fault estimates.

A notable field is that of machine fault diagnosis in mechanical engineering [49] [51]. This concerns of finding faults in machine elements, most specifically in rotating machinery. This area is only remotely related though, as the methods used are very specific to the field, and include spectral and waveform analysis of vibration signals. Reference [51] provides an overview of the field with specific emphasis on wavelets for fault diagnosis of rotary machines.

5.3 Generic Methods

The common property of most generic methods is that they rely on a behavioral model of the system. Many of these model the system as a process, and hence process analysis approaches are used from control theory [50, 54, 58, 59]. This is often applied to fault tolerant systems. Often, these are called Model-Based Diagnosis techniques, which aim at finding the fault of an observed system based on knowledge about the system's expected behavior [52, 55, 56, 61]. The mentioned entropy based and probabilistic approaches are typically used.

Tidiri *et al.* [62] combine model based approaches with data driven methods (which process a large amount of data from the system's output and are based on training data for a correctly working system). This may be a good candidate to be applied to software fault localization, because in this case often the model is not available but the operational data from software executions is easily obtainable through system logs. This publication refers other related work in this area, which can be useful sources for more information about this set of techniques.

Conclusions

This paper presented the results of our interdisciplinary analysis of fault localization techniques. As this was a preliminary study, our goal was to find related publications in various engineering fields, initially evaluate the proposed methods and assess their usability to our central topic, software fault localization. We found that, among the many different engineering fields, computer networks, aerospace, (power) electronics and some other areas are the most promising to help advance software fault localization.

The detailed analysis results presented in Section 4 could provide a starting point for further analyzing the techniques. Based on the various properties of the method, we provided (fault types, base data, empirical results, etc.), the most promising approaches could be selected for further consideration. Section 5, on the other hand, could be used to pin-point specific topics (with references to the main articles) to be used to enhance software fault localization.

Although we performed a systematic Literature Analysis, we cannot claim any completeness thereof. Based on the identified and here referenced works, further

publications could be searched by investigating the references, authors and research groups, etc. Also, scientific venues (conferences, journals) of specific engineering areas could be further analyzed to discover additional results.

Nevertheless, we believe that the survey in its present state is suitable for us to continue our quest for enhancing software fault localization, and for other readers to obtain a wider view of this important and diverse topic.

In future work, we will evaluate the most promising approaches in more detail and eventually implement the findings, for software fault localization.

Acknowledgement

This work was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences. The author would like to help the supporting work of Ottó Eötvös and Brúnó Ilovai.

References

- [1] Frank Schweitzer (editor-in-chief): *Advances in Complex Systems – A Multidisciplinary Journal*, World Scientific, ISSN (print): 0219-5259, ISSN (online): 1793-6802
- [2] W. E. Wong, R. Gao, Y. Li, R. Abreu and F. Wotawa: A survey of software fault localization, *IEEE Transactions on Software Engineering*, 42(8): 707-740, August 2016
- [3] P. Parmar and M. Patel: *Software Fault Localization: A Survey*, *International Journal of Computer Applications*, 154(9):6-13, 2016
- [4] Held, J. M.: *The Modelling of Systems of Systems*, PhD Thesis, University of Sydney, 2008
- [5] Tien-Duy B. Le, Ferdian Thung, David Lo: *Theory and Practice, Do They Match? A Case with Spectrum-Based Fault Localization*, 2013 IEEE International Conference on Software Maintenance, pp. 380-383
- [6] P. Kavulya, Soila, Joshi, Kaustubh, Giandomenico, Felicita and Narasimhan, Priya: *Failure Diagnosis of Complex Systems, Resilience Assessment and Evaluation of Computing Systems*, 2012, pp. 239-261
- [7] R. Abreu, P. Zoetewij and A. J. C. van Gemund: *Spectrum-based multiple fault localization*, In *Proceedings of IEEE/ACM International Conference on Automated Software Engineering*, pp. 88-99, November 2009
- [8] R. Abreu, W. Mayer, M. Stumptner and A. J. C. van Gemund: *Refining spectrum-based fault localization rankings*, In *Proceedings of the 2009 ACM Symposium on Applied Computing*, pp. 409-414, 2009
- [9] Sh. Artzi, J. Dolby, F. Tip and M. Pistoia: *Practical fault localization for dynamic web applications*, *Proceedings of the 32nd ACM/IEEE*

- International Conference on Software Engineering - Volume 1, pp. 265-274, 2010
- [10] J. Christ, E. Ermis, M. Shäf and T. Wies: Flow sensitive fault localization, In Proceedings of Verification, Model Checking, and Abstract Interpretation: 14th International Conference, VMCAI 2013, pp. 189-208, January 2013
- [11] S. Pearson, J. Campos, R. Just, G. Fraser, R. Abreu, M. D. Ernst, D. Pang and B. Keller: Evaluating and improving fault localization, Proceedings of the 39th International Conference on Software Engineering, pp. 609-620, 2017
- [12] L. Ravindranath, S. Nath, J. Padhye and H. Balakrishnan: Automatic and scalable fault detection for mobile applications, Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services, pp. 190-203, 2014
- [13] M. Renieris and S. P. Reiss: Fault localization with nearest neighbor queries, In Proceedings of 18th IEEE International Conference on Automated Software Engineering, pp. 130-139, 2003
- [14] Sh. Wang, D. Lo, L. Jiang, Lucia and H. Ch. Lau: Search-based fault localization, Proceedings of the 2011 26th IEEE/ACM International Conference on Automated Software Engineering, pp. 556-559, 2011
- [15] A. Aghasaryan, E. Fabre and C. Jard: A Petri net approach to fault detection and diagnosis in distributed systems I., Proceedings of the 36th IEEE Conference on Decision and Control, pp. 720-725, December 1997
- [16] A. Aghasaryan, E. Fabre and C. Jard: A Petri net approach to fault detection and diagnosis in distributed systems II, Proceedings of the 36th IEEE Conference on Decision and Control, pp. 726-731, December 1997
- [17] D. Alekseev and V. Sayenko: Proactive fault detection in computer networks, In Proceedings of 2014 First International Scientific-Practical Conference Problems of Infocommunications Science and Technology, pp. 90-91, 2014
- [18] M. Brodie, I. Risha and Sh. Ma: Intelligent probing: A cost-effective approach to fault diagnosis in computer networks, IBM Systems Journal, 41(3):372-385, 2002
- [19] C. S. Chao, D. L. Yang and A. C. Liu: An automated fault diagnosis system using hierarchical reasoning and alarm correlation, Proceedings of 1999 IEEE Workshop on Internet Applications (Cat. No.PR00197), pp. 120-127, August 1999
- [20] M. Chen, A. X. Zheng, J. Lloyd, M. I. Jordan and E. Brewer: Failure diagnosis using decision trees, In Proceedings of International Conference on Autonomic Computing, pp. 36-43, May 2004

-
- [21] R. H. Deng, A. A. Lazar and W. Wang: A probabilistic approach to fault diagnosis in linear lightwave networks, *IEEE Journal on Selected Areas in Communications*, 11(9):1438-1448, December 1993
- [22] M. A. Fecko and M. Steinder: Combinatorial designs in multiple faults localization for battlefield networks, In *Proceedings of 2001 MILCOM Communications for Network-Centric Operations: Creating the Information Force (Cat. No.01CH37277)*, pp. 938-942, 2001
- [23] M. S. Garshasbi and Sh. Jamali: A new fault detection method using end-to-end data and sequential testing for computer networks, *Reliability Engineering & System Safety*, 114(1):45-51, June 2013
- [24] C. S. Hood: Proactive network fault detection, *IEEE Transactions on Reliability*, 46(3):333-341, September 1997
- [25] L. Kant, A. S. Sethi and M. Steinder: Fault localization and self-healing mechanisms for FCS networks, *Proc. 23rd Army Science Conference*, January 2003
- [26] I. Katzela and M. Schwarz: Schemes for fault identification in communication networks, *IEEE/ACM Trans. Netw.*, 3(6):753-764, December 1995
- [27] R. R. Kompella, J. Yates, A. Greenberg and A. C. Snoeren: IP fault localization via risk modeling, *Proceedings of the 2nd Conference on Symposium on Networked Systems Design & Implementation - Volume 2*, pp. 57-70, 2005
- [28] L. Lu, Zh. Xu, W. Wang and Y. Sun: A new fault detection method for computer networks, *Reliability Engineering & System Safety*, 114(Supplement C):45-51, 2013
- [29] M. Natu and A. S. Sethi: Active probing approach for fault localization in computer networks, In *Proceedings of 2006 4th IEEE/IFIP Workshop on End-to-End Monitoring Techniques and Services*, pp. 25-33, April 2006
- [30] M. Natu and A. S. Sethi: Efficient probing techniques for fault diagnosis, In *Proceedings of Second International Conference on Internet Monitoring and Protection (ICIMP 2007)*, pp. 20-20, July 2007
- [31] M. Natu and A. S. Sethi: Probabilistic fault diagnosis using adaptive probing, In *Proceedings of Managing Virtualization of Networks and Services*, pp. 38-49, 2007
- [32] I. Rish, M. Brodie, N. Odintsova, Sh. Ma and G. Grabarnik: Real-time problem determination in distributed systems using active probing, In *Proceedings of 2004 IEEE/IFIP Network Operations and Management Symposium (IEEE Cat. No.04CH37507)*, pp. 133-146, April 2004

- [33] M. Steinder and A. S. Sethi: Non-deterministic fault localization in communication systems using belief networks, *IEEE/ACM Trans. Netw.*, 12(5):809-822, October 2004
- [34] M. Steinder and A. S. Sethi: Probabilistic fault localization in communication systems using belief networks, *IEEE/ACM Transactions on Networking*, 12(5):809-822, October 2004
- [35] Y. Tang, E. S. Al-Shaer and R. Boutaba: Active integrated fault localization in communication networks, In *Proceedings of 2005 9th IFIP/IEEE International Symposium on Integrated Network Management*, 2005, IM 2005, pp. 543-556, 2005
- [36] W. Traczyk: Probes for fault localization in computer networks, *Journal of Telecommunications and Information Technology*, 3:23-27, 2004
- [37] B. Wang, W. Wei, W. Zeng and K. R. Pattipati: Fault localization using passive end-to-end measurements and sequential testing for wireless sensor networks, *IEEE Transactions on Mobile Computing*, 11(3):439-452, March 2012
- [38] X. Zhang, Z. Zhou, G. Hasker, A. Perrig and V. Gligor: Network fault localization with small TCB, In *Proceedings of 2011 19th IEEE International Conference on Network Protocols*, pp. 143-154, October 2011
- [39] P. J. Adamovits and B. Pagurek: Simulation (model) based fault detection and diagnosis of a spacecraft electrical power system, *Proceedings of 9th IEEE Conference on Artificial Intelligence for Applications*, pp. 422-428, March 1993
- [40] E. Balaban, S. Narasimhan, H. N. Cannon and L. S. Brownston: Model-based fault detection and diagnosis system for NASA Mars subsurface drill prototype, In *Proceedings of 2007 IEEE Aerospace Conference*, pp. 1-13, March 2007
- [41] M. E. H. Benbouzid, M. Vieira and C. Theys: Induction motors' faults detection and localization using stator current advanced signal processing techniques, *IEEE Transactions on Power Electronics*, 14(1):14-22, January 1999
- [42] A. Beschta, O. Dressler, H. Freitag, M. Montag and P. Struß: Model-based approach to fault localization in power transmission networks, *Intelligent Systems Engineering*, 2:3-14, February 1993
- [43] T. Digernes: Real-time failure detection and identification applied to supervision of oil transport in pipelines, *Modeling, Identification and Control*, 1(1):39-49, 1980
- [44] R. W. Dries: Model-based reasoning in the detection of satellite anomalies, MS Thesis, AFIT/GSO/ENG/90D-03, School of Engineering, Air Force Institute of Technology, 1990

-
- [45] Judith Pálfi, Miklós Tompa and Péter Holcsik: Analysis of the Efficiency of the Recloser Function of LV Smart Switchboards, *Acta Polytechnica Hungarica*, Volume 14, Number 2, 2017, pp. 131-150
- [46] J. Poon: Model based fault detection and identification for power electronics systems, Technical Report No. UCB/EECS-2015-238, University of California, Berkeley, 2015
- [47] B. Peischl and F. Wotawa: Automated source-level error localization in hardware designs, *IEEE Design Test of Computers*, 23(1):8-19, January 2006
- [48] A. Tanwani, A. D. Domínguez-Garcia and D. Liberzon: An inversion based approach for fault detection and isolation in switching electrical networks, *IEEE Transactions on Control Systems Technology*, 19(5):1059-1074, September 2011
- [49] Lajos Tóth and Tibor Tóth: On Finding Better Wavelet Basis for Bearing Fault Detection, *Acta Polytechnica Hungarica*, Volume 10, Number 3, 2013, pp. 17-35
- [50] V. Venkatasubramanian, R. Vaidyanathan and Y. Yamamoto: Process fault detection and diagnosis using neural networks, *Computers & Chemical Engineering*, 14(7):699-712, 1990
- [51] Ruqiang Yan, Robert X. Gao and Xuefeng Chen: Wavelets for fault diagnosis of rotary machines: A review with applications, *Signal Processing*, Volume 96, Part A, 2014, pp. 1-15, Elsevier
- [52] P. M. Frank: Analytical and Qualitative Model-based Fault Diagnosis – A Survey and Some New Results, *European Journal of Control*, 2(1):6-28, 1996
- [53] J. Gertler: Analytical redundancy methods in fault detection and isolation, *IFAC Proceedings Volumes*, 24(6):9-21, September 1991
- [54] R. Isermann: Process fault detection based on modeling and estimation methods - A survey, *Automatica*, 20(4):384-404, 1984
- [55] Johan de Kleer: Diagnosing Multiple Persistent and Intermittent Faults, In *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI 2009)*, pp. 733-738, 2009
- [56] Johan de Kleer and Brian C. Williams: Diagnosing multiple faults, *Artificial Intelligence*, 32(1):97-130, 1987
- [57] U. Lerner, R. Parr, D. Koller and G. Biswas: Bayesian fault detection and diagnosis in dynamic systems, In *Proc. AAAI*, pp. 531-537, 2000
- [58] M. A. Massoumnia, G. C. Verghese and A. S. Willsky: Failure detection and identification in linear time-invariant systems, *Technology*, No. July, 1986

- [59] R. K. Mehra and J. Peschon: An innovations approach to fault detection and diagnosis in dynamic systems, *Automatica*, 7(5):637-640, 1971
- [60] N. Olivier-Maget, S. Negny, G. Hétreux and J. M. Le Lann: Fault diagnosis and process monitoring through model-based and case based reasoning, In *Proceedings of 19th European Symposium on Computer Aided Process Engineering*, pp. 345-350, 2009
- [61] K. Shchekotykhin, T. Schmitz and D. Jannach: Efficient sequential model-based fault localization with partial diagnosis, In *Proceedings of IJCAI'16 Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pp. 1251-1257
- [62] Khaoula Tidiri and Nizar Chatti and Sylvain Verron and Teodor Tiplica: Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: A review of researches and future challenges, *Annual Reviews in Control*, Volume 42, pp. 63-81, 2016
- [63] A. Varga: On computing least order fault detection using rational nullspace bases, *IFAC Proceedings Volumes*, 36(5):227-232, 2003
- [64] M. Steinder and A. S. Sethi: A survey of fault localization techniques in computer networks, *Science of Computer Programming*, 53(2):165-194, 2004

Modeling the Intuitive Decision-Maker's Mindset

Jolán Velencei, Ágnes Szeghegyi

Keleti Faculty of Business and Management, Óbuda University
Tavaszmező u. 15, H-1083 Budapest, Hungary
velencei.jolan@kgk.uni-obuda.hu; szeghegyi.agnes@kgk.uni-obuda.hu

Zoltán Baracska, Beatrix Bókayné Andráska

Doctoral School of Regional Sciences and Business Administration
Széchenyi István University, Egyetem tér 1, H-9026 Győr, Hungary
baracska.zoltan@sze.hu; bokayne.andrasko.beatrix@sze.hu

Abstract: Today, the term, Working Memory, is closely associated with intelligence. We propose that in addition to improving and speeding-up analysis, Artificial Intelligence (AI) can also be useful as a supplement to Working Memory. It is generally accepted that working memory plays a crucial role in cognition and models by computers, can help us understand the human mind. Building an artificial working memory can bring further benefits; for example, it can separate retrieval from reasoning and therefore, can acquire new concepts. The aim of this research is to solve the capacity shortage problem of Working Memory, by using AI as a supplement. In order to develop our argument, we characterize the ID3 algorithm as a way of looking for a consistent solution in the existing Case Based Graph; as the ID3 algorithm builds it from an empty graph, to an increasingly complex one. Methodologically, our study is based on observation of several Digital Natives (DNs) playing different games at Mobilis Interactive Exhibition Center in Győr, Hungary. The aim is to explore the behavior of the DN generation. By identifying the different mindset patterns of DNs, we will be able to observe how different DNs can be facilitated, to enjoy the games, rather than being bored, anxious or even, becoming addicted.

Keywords: Artificial Intelligence; Knowledge-based System; Machine Learning

1 Introduction

A published overview of expert systems shows the kinds of articles published in the field [1]. Even though intuitive Decision-Makers emphasize that the knowledge bases of their tools cannot have more knowledge than the experts

whose knowledge has been represented, sometimes the illusion still arises. The knowledge base in the expert system will not be able to think differently than the decision maker who was the source of that knowledge. As Liao [2] said, the development of methodological approaches in expert systems shows expert-orientation in ICT-related disciplines, and suggests that there is a possibility of a different orientation in human and social studies. One of the novelties of our Doctus Knowledge-based System [3] is its ability to show the informativity of the attributes of the Decision-Maker through the ID-3 algorithm. The intuitive Decision-Maker's mindset can be discovered through the informativity of these attributes. At the Mobilis Interactive Exhibition Center, we observed Digital Natives during play and built up a knowledge base of their behavior to illustrate the functional novelty of the Doctus Knowledge-based System. We argue for a transdisciplinary approach, in which the two otherwise parallel research paths may meet. Transdisciplinarity examines what lies beyond the different disciplines. It seeks to have an overall picture, an integration of a fuller understanding [4]. To understand the reality of decision making, one has to "pick and choose" from the fields of Philosophy, Cognitive Psychology, Cybernetics and Artificial Intelligence. In this article we aim to provide a demonstration of the ID-3 algorithm-based, Doctus Knowledge-based System, through a case, where the attributes, as indicated by the descriptor, are classified and a graph is developed. The descriptive indicator is a statistical value, which is called entropy, in information theory.

A contemporary Decision-Maker can only work together, with a smart tool, if the model created by the latter, distorts the thinking of the former, only minimally. In this study, we show how to map Working Memory, through the inductive reasoning of the Doctus Knowledge-based System, to create an artificial Working Memory.

2 Background

Daniel Kahneman states, "fast thinking includes both variants of intuitive thought – the expert and the heuristic – as well as, the entirely automatic mental activities of perception and memory, the operations that enable you to know there is a lamp on your desk or retrieve the name of the capital of Russia" [5]. Not having understood the intuitive Decision-Maker's mindset, expert systems have not yet found their domain of validity. "There were many published cases of systems that did not go beyond the basic validation of the application rules and so this pulled down the overall averages" [6].

Knowledge gathered in the knowledge-based system always comes from the memory of the intuitive Decision-Maker. The mind is not tuned for arithmetic, but to the memories of experience. We not only tell stories when we decide we are

going to tell stories. Our memory is also telling us stories, in other words, what we have kept from our experiences is the story. As Daniel Kahneman says in his talk entitled “The riddle of experience vs. memory” at the TED2010 Conference, “We actually don't choose between experiences, we choose between memories of experiences. And even when we think about the future, we don't think of our future normally as experiences. We think of our future as anticipated memories. And, basically, you can look at this, you know, as a tyranny of the remembering self, and you can think of the remembering self-sort of dragging the experiencing self through experiences that the experiencing self doesn't need” [7].

If we examine cognitive psychology, from a meta-level, we find a vast amount of results from just as numerous experiments. It is not the aim of this paper to predict when cognitive psychology will present a few theories, nor if that is even possible. We interpret this situation on the basis of Karl Popper, who declared that the research of human-created organizations does not have its own Galilei. Both Popper and we hope that it will always be so, because the understanding of human organizations is different than that of physical or biological ones. With the efforts of Galilei and Newton, the successes of physics have surpassed all expectations, and so physics leapt far ahead of all other disciplines. Ever since Pasteur appeared as the Galilei of biology, biology has also been almost as successful [8].

In the wake of George Armitage Miller's idea of “The Magical Number Seven, Plus or Minus Two”, published in 1956, the research results of Working Memory experiments have been just as defining for cognitive psychology [9]. “The proposal of the episodic buffer clearly does represent a change within the Working Memory framework, whether conceived as a new component, or as a fractionation of the older version of the central executive. By emphasizing the importance of coordination, and confronting the need to relate WM and LTM [long-term memory], it suggests a closer link between our earlier multi-component approach and other models that have emphasized the more complex executive aspects of WM. The revised framework differs from many current models of WM in its continued emphasis on a multi-component nature, and in its rejection of the suggestion that WM simply represents the activated portions of LTM. It also rejects the related view that slave systems merely represent activations within the processes of visual and verbal perception and production. Although WM is intimately linked both to LTM and to perceptual and motor function, it is regarded as a separable system involving its own dedicated storage processes” [10].

Howard Gardner defined ten types of intelligence [11] and is one of the people who has spent the most effort on defining the concept of intelligence. In his newest book, co-authored with Katie Davis, they examine the interaction between Apps and the human mind. “The second opportunity entails the capacity to make use of diverse forms of understanding, knowing, expressing, and critiquing – in terms that Howard has made familiar, our multiple forms of intelligence. Until recently, education was strongly constrained to highlight two forms of human intelligence: linguistic and logical-mathematical. Indeed, until the end of the

nineteenth century, linguistic intelligence was prioritized; in the twentieth century, logical-mathematical intelligence gained equal if not greater importance” [12]. Nothing guarantees that the intuitive Decision-Maker behaves according to mathematical intelligence. It is impossible to prove, that mathematical intelligence leads to better decisions than other forms of intelligence.

This might indeed be at the core of the difficulty in understanding the intuitive Decision-Maker's mindset; the different disciplines are captive in their respective cages. Developers of machine learning held to their own concepts and methods, occasionally looking to cognitive psychology. Cognitive psychologists, for example Amos Twersky and Daniel Kahneman [13] have occasionally considered decision-making. Researchers in decision-making, often looked to cognitive psychology, but almost never paid attention to machine learning. To make matters worse, all three disciplines neglected philosophy, especially the problem of induction [14] [15]. Whatever may have happened, it is now clear that we must free ourselves from the cages of disciplines and hope to reach another result through meta-knowledge and a transdisciplinary approach. In this approach we must also decide on what level we wish to examine reality: through models, methods or tools. “We describe decision making with the following three levels of reality: (1) Models of decision makers' behavior, (2) Methods used to support intuitive decision makers, (3) Tools we use to implement the support of intuitive decision makers” [16].

3 Rejuvenating Machine Learning

For laymen, a computer is a machine that 'computes', that is, calculates faster than a human. This is still the basic approach, even though humans 'compute' very little. On trains and in pubs, we see people use the machine, but we do not see them calculating with it. The rejuvenation of machine learning, if it was rejuvenation at all, did not bring a paradigm shift. Based on the work of Thomas Kuhn [17], if two people stand in the same place and look in the same direction, then, avoiding solipsism, we conclude that they receive the same stimuli. If their eyes could be in the same place, the stimuli would be identical. However, people do not see stimuli. People have impressions and feelings, and nothing requires that we make the assumption that the two observers' impressions are the same.

At the time of the rejuvenation of machine learning, we are still not able to rethink it. We still tell digital natives what the digital outsiders believed. It is a matter of debate who first introduced the concepts of digital natives and digital immigrants. According to Marc Prensky, it was he himself, that used it first in 2001 [18], but that is beside the point right now. Our narrow field of vision and lack of courage allows us to see only what others have accepted. “What we refer to with the 'meta-' is a very high level of abstraction, something that we can call meta-level.

At a high level of abstraction, where the details of reality dissolve, such knowledge loses direct touch with reality. However, it can be 'concretized' by zooming into reality, and in this 'concretization' the meta-knowledge can take radically different forms. For instance, it may take the form of some knowledge with reference to one reality and some different knowledge with reference to some other reality. For this reason, meta-knowledge does not consist of concepts but of meta-concepts, which are extremely high-density essences of many concepts" [19].

Nick Bostrom in his book, *Superintelligence* [20] said that we cannot expect Artificial Intelligence to be motivated by love or hate or pride or other such common human sentiments. Let us first emphasize, that if we have understood reality at the level of the individual, then the modeling of the intuitive Decision-Maker's mindset can be represented with an algorithm. We also posit that the ID3 algorithm, originally developed by J. Ross Quinlan would be fit for that purpose. If it were not, we could not choose any other existing algorithm, we would have to create a new one. In the next chapter, we will demonstrate that the ID3 algorithm is suitable for adequately describing the intuitive Decision-Maker's mindset. Developed by us and actively applied for two decades, the inductive reasoning of the Doctus Knowledge-based System is based on the aforementioned ID3 algorithm. A tool is a tool, which grows more effective as its validity domain narrows. One must never search for the problem matching the tool, one must search for the most suitable, or least inadequate tool, for the problem.

4 Informativity in Mindset Patterns

Digital Natives are trained as if they would need the same tools as the Digital Immigrants and their ancestors. This new generation knows a little about everything, which is not necessarily a bad thing [21]. If we arouse their attention, they can deepen their knowledge easily, because knowledge is just 'a click away'. This generation of Digital Natives do not need to be specialized in a strict way but rather become de-specialized, with the ability to search for knowledge efficiently and thus to become competitive. The capacity of long term memory, or what can be called meta-knowledge, defines the personal level of knowledge and experience acquired, the levels of which can be gained through many learning hours: novice (10 hours), expert (100 hours), master (1,000 hours) or grand master (10,000 hours). Short-term Memory or Working Memory, however, can contain and hold only 7 plus or minus 2 items.

Due to technological acceleration, the use of the term 'content specific knowledge' has grown significantly in recent decades, in large part because educators now commonly use the term as shorthand to articulate a useful technical distinction between knowledge and skills. It refers to the body of knowledge and

information that teachers teach and that students are expected to learn in a given subject or content area, generally referring to the facts, concepts, theories, and principles that are taught and learned in specific academic courses, rather than to related skills – such as reading, writing, or researching – that students also learn in school. It is incontestable that Working Memory plays a crucial role in cognition and that models created with computers can help us understand the human mind. On the other hand, building an artificial Working Memory can result in many other positive outcomes: it can for example separate retrieval from reasoning and therefore can acquire new concepts. “By placing Working Memory between an agent’s sensors and its decision-making element, we can give it the ability to recognize existing contexts, and reason using precedents – even analogies. This, in turn, allows the designer to focus on the agent’s heuristics. Another reason to create an artificial Working Memory system is that doing so will also give us a framework within which to investigate different types of similarity, measures of uncertainty, and knowledge bases” [22]. In order to examine our topic, we observed several Digital Natives playing different games at Mobilis Interactive Exhibition Center in Győr, Hungary. Figure 1 depicts the attributes of meaningful play. The selection of observed attributes is based on gamification literature [23, 24, 25, 26, 27].

The screenshot shows a software window titled "Doctus Knowledge Based System - [MOBILIS_GYOR]". The menu bar includes "File", "Edit", "View", "Search", "Knowledge Management", "Window", and "Help". Below the menu is a toolbar with various icons. The main area has a tabbed interface with "Attributes" selected. A table lists the following attributes:

Name
MEANINGFUL PLAY
Practice
Comfort
Immersion
Control
Pleasure of learning
Extension of identity
Extension of senses
Tribal conveying values
Social Temptation
EQ
Thinking process
Belief
Fun
Sensation: Game as sense-pleasure
Fantasy: Game as make-believe
Narrative: Game as drama
Challenge: Game as obstacle course
Fellowship (multiplayer): Game as social framework
Discovery: Game as uncharted territory
Expression: Game as self-discovery
Submission: Game as masochism
Physics
Economy
Progression
Tactical maneuvering

Figure 1

Observed attributes (Source: Screenshot by Authors from Doctus)

The aim was to explore the behavior of the DN generation by observing how the structure of the games affected their viability. By understanding the different mindset patterns of DNs, we would be able to observe how different DNs can be facilitated to enjoy the games, rather than getting bored, getting anxious or becoming dependent. We have selected five primary categories based on feedback from randomly selected players on the overall experience of the game, grade M1 became the lowest and grade M5 was the highest ranking in connection to the meaningfulness of the game. We then faced the following question: Which attribute should be first examined, in other words, which has the greatest descriptive power? We have chosen inductive reasoning as the categories are used when we would like to predict the value of an attribute with a discrete value based on a given situation and the knowledge we have on values of other descriptive attributes. Thus, we have a Case Based Graph, which is based on the previously described examples for a similar simulation we want to observe, hence we will be able to provide the expected value of the requested attribute. The ID3 (inductive learning) algorithm creates (or learns) the Case Based Graph based on the examples provided [28], which are built up from the bottom to the top. The basic

idea of this machine learning algorithm is to select an attribute which we are interested in – this will be the target function, at first a binary attribute. Then, we find the additional attributes which best define the output value of the target function – this will give the root of the Case Based Graph and the possible values of each attribute will be the branches. We continue this process for the remaining levels and for each attribute until complete. Then ID3 classifies the attributes based on the descriptor and builds the graph – the descriptive indicator is a statistical value, which is called entropy in information theory. We shall characterize the ID3 algorithm by looking for a consistent solution in the existing Case Based Graph – as the ID3 algorithm builds the tree of decision (hypothesis) from an empty graph to an increasingly complex one. We use the Formula (1), where S is the set of examples, B is the binary target attribute, S_{plus} is positive, S_{minus} is a set of examples with negative target attributes, therefore $s \in S_{plus}$, if $B(s) = \uparrow$, and $s \in S_{minus}$, if $B(s) = \downarrow$. For each Entropy count, $\log 0$ should be 0.

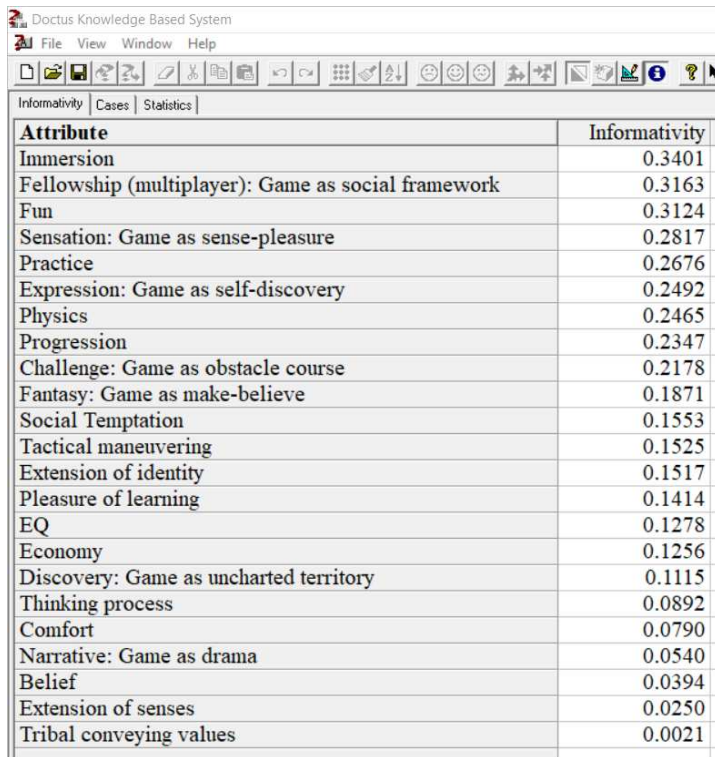
$$Entropy(S) = - \left(\frac{|S_{plus}|}{|S|} \log_2 \frac{|S_{plus}|}{|S|} + \frac{|S_{minus}|}{|S|} \log_2 \frac{|S_{minus}|}{|S|} \right) \quad (1)$$

The Entropy (S) specifies the minimum number of bits in an encoded bit sequence for a given example. If it is 0, then the target function in S is the same, so it does not have to be encrypted as we know what it was. If this is 1, it cannot be compressed and encoded, as positive and negative examples are equally likely. If it is a number between 0 and 1 (e.g. 0.7), then at least 0.7 bits must be used when encoding. We got the following values in Figure 2.

$$Entropy(S, Immersion) = 0.3401$$

$$Entropy(S, Fellowship) = 0.3163$$

$$Entropy(S, Fun) = 0.3124$$



The screenshot shows the 'Doctus Knowledge Based System' interface. At the top, there is a menu bar with 'File', 'View', 'Window', and 'Help'. Below the menu is a toolbar with various icons. The main window displays a table with two columns: 'Attribute' and 'Informativity'. The table lists 20 different attributes and their corresponding informativity values, sorted in descending order.

Attribute	Informativity
Immersion	0.3401
Fellowship (multiplayer): Game as social framework	0.3163
Fun	0.3124
Sensation: Game as sense-pleasure	0.2817
Practice	0.2676
Expression: Game as self-discovery	0.2492
Physics	0.2465
Progression	0.2347
Challenge: Game as obstacle course	0.2178
Fantasy: Game as make-believe	0.1871
Social Temptation	0.1553
Tactical maneuvering	0.1525
Extension of identity	0.1517
Pleasure of learning	0.1414
EQ	0.1278
Economy	0.1256
Discovery: Game as uncharted territory	0.1115
Thinking process	0.0892
Comfort	0.0790
Narrative: Game as drama	0.0540
Belief	0.0394
Extension of senses	0.0250
Tribal conveying values	0.0021

Figure 2

Informativity of Attributes (Source: Screenshot by Authors from Doctus)

As Immersion, Fellowship and Fun had the strongest explanatory force, the root of the Case Based Graph will be Immersion and the resulting edges will be matched to its possible values. Subdivisions that fit into new branches will not be built on the whole set of S , but only with the examples in which the Immersion attribute takes the value corresponding to that branch. We could characterize the ID3 algorithm by looking for a consistent solution in the existing Case Based Graph. Based on the ID3 algorithm, we developed the following graphical model seen in Figure 3.

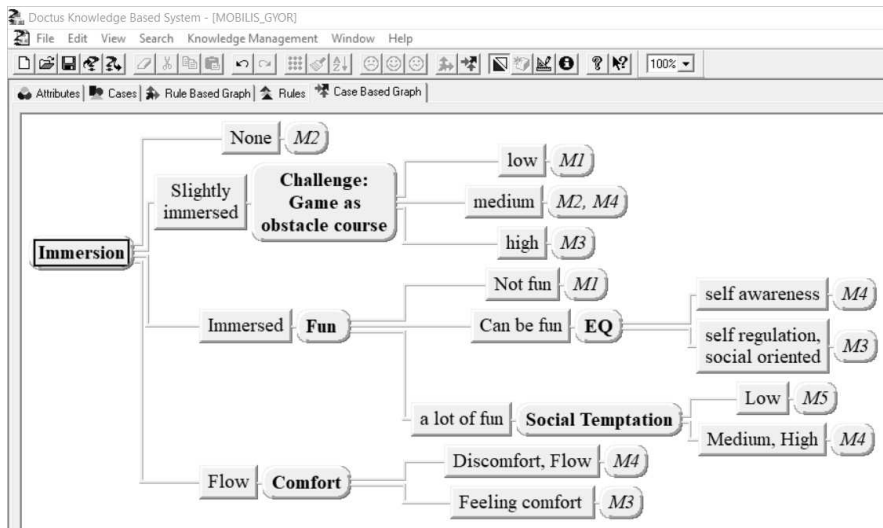


Figure 3

Case Based Graph (Source: Screenshot by Authors from Doctus)

As the Case Based Graph is founded on the statistics for all given attribute values in our examples and because we may have been using faulty data, we will need to modify the termination condition of the machine learning algorithm, to make it work better – we have to accept some imperfect consistencies. Experience shows that the more examples we work on, for a particular situation, produce more precise Case Based Graph. This observation is also valid for problem scenarios: The bigger the graph is, the more precisely we can set the value of the target attribute, but, surprisingly, after approximately 25 leafs, the accuracy decreases constantly.

5 Conceptual Model

Based on Nassim Taleb [29] those living in Mediocristan were satisfied with using arithmetic-based, Multiple Criteria Decision Analysis (MCDA) systems; they were perhaps afraid of new knowledge and the losses that come with change. Those living in Extremistan [29], are practically waiting for some new knowledge to challenge the current knowledge. If we say that it is currently possible to model the Working Memory of the intuitive Decision-Makers, then we will appear frightening in Mediocristan. We may cause a lot of trouble, if we rob someone of their belief in numbers. Life in Mediocristan is nice and calm, if we believe that Artificial Intelligence will bring the faster recalculation of the past. Very few intuitive Decision-Makers live in Extremistan, but they are always happy to see the knowledge representations modeled by knowledge-based systems.

The weak point of Doctus Knowledge-based System is that it is only able to show the mindset patterns of those who have them. In other words, machine learning is only able to help those who have natural intelligence. Perhaps even the concept of intelligence will need further clarification. Let us not dream of a world where every puzzle can be solved by a crutch. Puzzles were created for people who, every now and then, succeed with a good shot. Not everyone has to be a puzzle solver, that is, to try to reach the one true solution quickly.

It is possible that knowledge-based expert systems have to pull themselves out of their predicament. Bootstrapping is a commonly used phrase today. Bootstrapping in Artificial Intelligence and machine learning, according to one definition, is a technique used to interactively improve performance, in other words, recursive self-improvement. For example, Ryan Smith talks about why every start-up should be a bootstrap [30]. The new wave of Artificial Intelligence can start the third S-curve found in Figure 4.

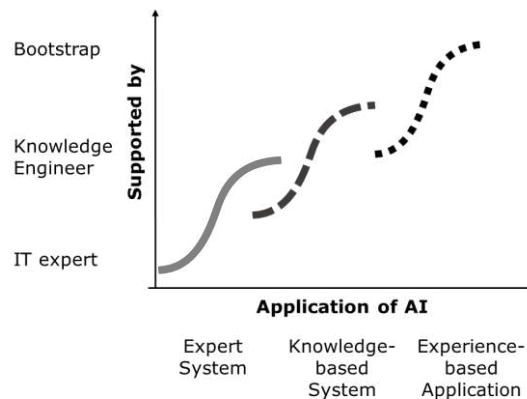


Figure 4

The third S-curve (Source: Drawn by Authors)

The harmony between the model and algorithm presented herein, can lead to the development of a tool that conforms to the already established user habits of digital natives who behave according to bootstrap patterns. This generation is already on the stage in the business world as well, but is often still forced to use the legacy tools of digital outsiders. Let us not forget, however, that the generation of the future, who grew up on computer games, does not want to read help files and does not really want to attend a course where they will only learn the use of one tool.

Conclusions

Although more and more data is being analyzed in the world, the decision-making process will not become smarter. Smartly prepared business decisions are born, based on knowing. An intuitive Decision-Maker can only be in balance with a smart tool if the invisible model between them, distorts as little as possible. In this

paper, we demonstrated that Working Memory can be mapped to artificial Working Memory. This means that the tool does not replace the intuitive Decision-Maker by making a decision for them; it simply frees up their memory capacity limits. In this case, nothing stands in the way of a Decision-Maker, if they want to use further, new attributes, in a new decision. The tool can, however, help make the new attributes consistent and congruent with those already established. Knowledge-based systems, if they can be rejuvenated, will always remain tools, they will never become scary monsters that overcome humans. Just like all other tools, these can also help augment and expand the capacity limits of Working Memory.

Acknowledgement

We would like to give special thanks to Mobilis Interactive Exhibition Center at Győr, Hungary for making it possible for us to make observations.

References

- [1] Sahin, S., Tolun, M. R. and Hassanpour, R.: "Hybrid expert systems A survey of current approaches and applications", *Expert Systems with Applications*, 39(4), 2012, pp. 4609-4617
- [2] Liao, S.: "Expert system methodologies and applications – a decade review from 1995 to 2004", *Expert systems with Applications*, 28(1), 2005
- [3] Baracskaï, Z., Dörfler, V. and Velencei, J.: "Reductive Reasoning", *Montenegrin Journal of Economics*, 1(1), 2005, pp. 59-66
- [4] Nicolescu, B.: "Methodology of Transdisciplinarity: Levels of Reality, Logic of the Included Middle and Complexity", *Transdisciplinary Journal of Engineering & Science*, 1(1), 2010, pp. 19-38
- [5] Kahneman, D.: *Thinking, Fast and Slow*, UK: Penguin Books Ltd, 2012, p. 13
- [6] Wagner, W. P.: "Trends in expert system development: A longitudinal content analysis of over thirty years of expert system case studies", *Expert Systems with Applications*, 76, 2017, p. 89
- [7] Kahneman, D.: *The riddle of experience vs. memory*, TED talk, https://www.ted.com/talks/daniel_kahneman_the_riddle_of_experience_vs_memory, accessed date 10.06.2018
- [8] Popper, K. R.: *The Poverty of Historicism*, London: Routledge, 1961
- [9] Miller, G.: "The magical number seven, plus or minus two: Some limits on our capacity for processing information", *The Psychological Review*, 63, 1956, pp. 81-97
- [10] Baddeley, A.: "The episodic buffer: a new component of working memory?", *Trends in Cognitive Sciences*, 4 (11), 2000, p. 419

-
- [11] Gardner, H.: *Multiple Intelligences: New Horizons in Theory and Practice*, Basic Books, 1993
- [12] Gardner, H. and Davis, K.: *The App Generation – How today’s youth Navigate Identity, Intimacy and Imagination in a Digital World*, Yale University Press, 2013
- [13] Tversky, A., Kahneman, D. “Judgment under uncertainty: Heuristics and biases”, *Science*, 185, 1974, pp. 1124-1131
- [14] Popper, K. R.: *The Logic of Scientific Discovery*. New York: Harper & Row, 1968
- [15] Russell, B. A.: *History of Western Philosophy*. London: Routledge, 1946
- [16] Baracscai, Z. and Dörfler, V.: “An essay concerning human decisions”, *The Transdisciplinary Journal of Engineering & Science*, 8, 2017, p. 72
- [17] Kuhn, T. S.: *The Structure of Scientific Revolutions*, Chicago, IL: The University of Chicago Press, 1970
- [18] Prensky, M.: *Digital Game-based Learning*, McGraw-Hill, 2001
- [19] Baracscai, Z. and Dörfler, V.: “An essay concerning human decisions”, *The Transdisciplinary Journal of Engineering & Science*, 8, 2017, p. 74
- [20] Bostrom, N.: *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press, 2014
- [21] Carr, N.: *The Glass Cage: How our Computers are changing us*, New York: W. W. Norton Company, 2014
- [22] Dreisiger, P.: “Artificial Working Memory: A Psychological Approach”, 16th School of Computer Science & Software Engineering Research Conference, Yanchep, Western Australia, 12–13 June 2008
- [23] Rouse, R.: *Game design: Theory & Practice*, Jones & Bartlett Learning, 2004
- [24] Adams, H.E. and Dormans, J.: *Game Mechanics: Advanced Game Design*, New Riders Games, 2012
- [25] Kiilia, K., de Freitas, S., Arnabb, S. and Lainema, T.: “The Design Principles for Flow Experience in Educational Games”, *Procedia Computer Science*, 15, 2012, pp. 78-91
- [26] Wolf, M. J. P. and Perron, B.: *The Routledge Companion to Video Game Studies*, Hoboken: Routledge, 2014
- [27] Garrelts, N.: *Understanding Minecraft: Essays on Play, Community and Possibilities*, McFarland, 2014
- [28] Quinlan, J. R.: *The Induction of Decision Trees*, *Machine Learning*, 1(1), 1986, pp.: 81-106. DOI: 10.1023/A:1022643204877

- [29] Taleb, N. N.: *The Black Swan*, Random House Trade Paperbacks, 2012
- [30] Smith, R.: "Why Every Startup Should Bootstrap", *Harvard Business Review*, 94(3), 2016, pp. 48-61