



COMPUTER MODELLING
AND
NEW TECHNOLOGIES

2017
VOLUME 21 NO 2

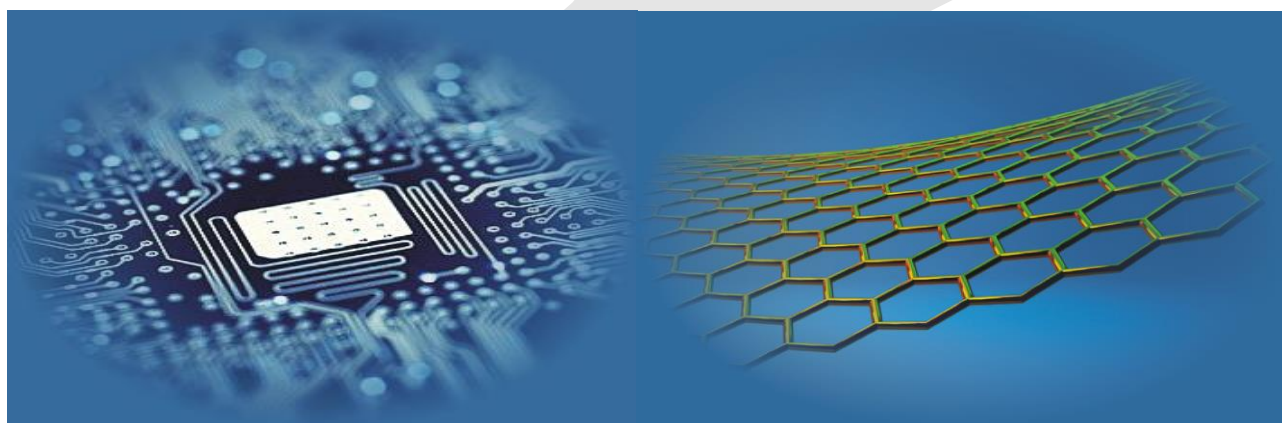
ISSN 1407-5806 ISSN 1407-5814 on-line

Latvian Transport Development and Education Association

Computer Modelling and New Technologies

2017 Volume 21 No 2

ISSN 1407-5806, ISSN 1407-5814 (On-line: www.cmnt.lv)



Riga – 2017

EDITORIAL BOARD

Prof. Igor Kabashkin	Chairman of the Board , <i>Transport & Telecommunication Institute, Latvia</i>
Prof. Yuri Shunin	Editor-in-Chief , <i>University of Latvia, Solid State Physics Institute, ISMA University, Latvia</i>
Dr. Brent Bowen	<i>Embry-Riddle Aeronautical University, United States of America</i>
Prof. Sergey Maksimenko	<i>Institute for Nuclear Problem, Belarus State University, Belarus</i>
Prof. Vladimir Litovchenko	<i>V. Lashkaryov Institute of Semiconductor Physics of National Academy of Science of Ukraine, Ukraine</i>
Prof. Pavel D'yachkov	<i>Kurnakov Institute for General and Inorganic Chemistry, Russian Academy of Sciences, Russian Federation</i>
Prof. Stefano Bellucci	<i>Frascati National Laboratories – National Institute of Nuclear Physics, Italy</i>
Prof. Arnold Kiv	<i>Ben-Gurion University of the Negev, Israel</i>
Prof. Alytis Gruodis	<i>Vilnius University, Lithuania</i>
Dr. Jiri Vacik	<i>Nuclear Physics Institute, Chehia</i>
Dr. Lital Alfonta	<i>Ben-Gurion University of the Negev, Israel</i>
Dr. Amita Chandra	<i>Delhi University, India</i>
Dr. Jacob Kleiman	<i>Toronto University, Canada</i>
Dr. Ian Brown	<i>Lawrence Berkeley National Laboratory, USA</i>
Dr. Nadia Kabachi	<i>Lyone University, France</i>
Dr. Calagero Pace	<i>Calabria University, Italy</i>
Dr. Angelica Strutz	<i>Zurich University, Switzerland</i>
Prof. Michael Schenk	<i>Fraunhofer Institute for Factory Operation and Automation IFF, Germany</i>
Prof. Dietmar Fink	<i>University of Mexico, United Mexican States</i>
Prof. Kurt Schwartz	<i>Gesellschaft für Schwerionenforschung mbH, Darmstadt, Germany</i>
Prof. Eva Rysiakiewicz-Pasek	<i>Institute of Physics, Wroclaw University of Technology, Poland</i>
Prof. Yedilkhan Amirgaliyev	<i>Suleyman Demirel University, Kazakhstan</i>
Prof. Vladimir Barakhnin	<i>Institute of Computational Technologies of SB RAS, Novosibirsk State University, Russia</i>
Prof. Kewen Zhao	<i>Institute of Applied Mathematics & Information Sciences, University of Qiongzhou, Sanya, P.R.China</i>
Guest Editor	Prof. Ravil Muhamedyev, <i>Institute of Information and Computational Technologies MES RK, SDU, Kazakhstan</i>
Contributing Editor	Prof. Victor Gopeyenko, <i>ISMA University, Latvia</i>
Literary Editor	Prof. Tamara Lobanova-Shunina, <i>Riga Technical University, Latvia</i>
Technical Editor , Secretary of Editorial Board	MSc Comp Nataly Burlutskaya, <i>ISMA University, Latvia</i>

Journal topics:	Publisher	Supporting Organizations
<ul style="list-style-type: none"> mathematical and computer modelling computer and information technologies natural and engineering sciences operation research and decision making nanoscience and nanotechnologies innovative education 	Latvian Transport Development and Education Association	Latvian Academy of Sciences Latvian Operations Research Society Fraunhofer Institute for Factory Operation and Automation IFF, Germany

Articles should be submitted in **English**. All articles are reviewed

EDITORIAL CORRESPONDENCE	COMPUTER MODELLING AND NEW TECHNOLOGIES, 2017, Vol. 21, No.2 ISSN 1407-5806, ISSN 1407-5814 (on-line: www.cmnt.lv)
Latvian Transport Development and Education Association	Scientific and research journal The journal is being published since 1996
68 Graudu, office C105, LV-1058 Riga, Latvia Phone: +371 29411640 E-mail: yu_shunin@inbox.lv http://www.cmnt.lv	The papers published in Journal 'Computer Modelling and New Technologies' are included in: INSPEC , www.theiet.org/resources/inspec/ VINITI , http://www2.viniti.ru/ CAS Database http://www.cas.org/ SCOPUS



Editors' Remarks

Lost Time

by Rabindranath Tagore

On many an idle day have
I grieved over lost time.
But it is never lost, my lord.
Thou hast taken every moment of my life
in thine own hands.

I was tired and sleeping on my idle bed
and imagined all work had ceased.
In the morning I woke up and found my
garden full with wonders of flowers.

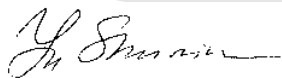
Hidden in the heart of things thou art
nourishing seeds into sprouts,
buds into blossoms, and ripening flowers
into fruitfulness.

Rabindranath Tagore (1861-1941)*


This 21th volume No.2 includes research papers on **Nature Phenomena and Innovative Engineering** and **Mathematical and Computer Modelling**.

Our journal policy is directed to fundamental and applied scientific researches, innovative technologies and industry, which is the fundamentals of the full-scale multi-disciplinary modelling and simulation. This edition is the continuation of our publishing activities. We hope our journal will be of interest for research community and professionals. We are open for collaboration both in the research field and publishing. We hope that the journal's contributors will consider collaboration with the Editorial Board as useful and constructive.

EDITORS



Yuri Shunin



Igor Kabashkin

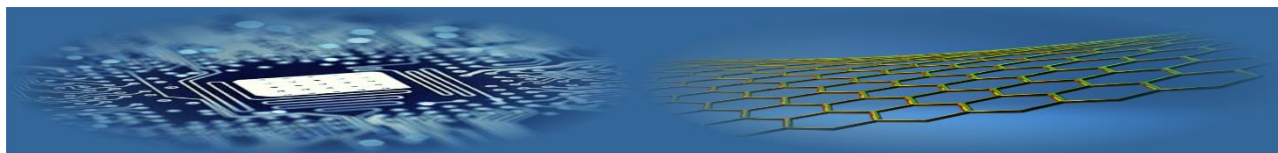
* **Rabindranath Tagore (7 May 1861 – 7 August 1941)**, was a Bengali poet, novelist, musician, painter and playwright who reshaped Bengali literature and music. As author of Gitanjali with its "profoundly sensitive, fresh and beautiful verse", he was the first non-European and the only Indian to be awarded the Nobel Prize for Literature in 1913. His poetry in translation was viewed as spiritual, and this together with his mesmerizing persona gave him a prophet-like aura in the west. His "elegant prose and magical poetry" still remain largely unknown outside the confines of Bengal.



CONTENT

NATURE PHENOMENA AND INNOVATIVE ENGINEERING		
E Dragomeretskaya	Effect of texture on mechanical and magnetic properties of steel from the petroleum distillation column	7
A Zvaigzne, O Bondarenko, A Boiko	Decision support system on the base of genetic algorithm for optimal design of a specialized maritime platform	11
MATHEMATICAL AND COMPUTER MODELLING		
G Beketova, B Akhmetov, A Korchenko, V Lakhno, A Tereshuk	Cyber intelligence systems based on adaptive regression splines and logical procedures of attack recognition	19
G Ramesh, T V Rajini Kanth, A Ananda Rao	Metrics for consistency checking in object oriented model transformations	29
S Manaseer, D Alsoudi, A Aljawawdeh	Border node detection: a new experimental approach	37
P Dileep Kumar Reddy, C Shoba Bindu, R Praveen Sam	Tri-Partite graph: a novel security scheme for cloud data	41
Zh E Aytkhozhaeva, A A Ziro, A Zh Zhaibergenova	Virtualization safety	48
L Tereykovskaya, I Tereykovskiy, E Aytkhozhaeva, S Tynymbayev, A Imanbayev	Improvement of learning efficiency of the neural networks, intended for recognition of graphic images in systems of biometric authentication	54
S Mall, U C Jaiswal	Word sense disambiguation in Hindi applied to Hindi-English machine translation	58
K Tewani	Ant colony optimization algorithm: advantages, applications and challenges	69
K Sekar, M Padmavathamma	Business process re-engineering capability based on ECMM: Efficient Configuration Model and Management	71
A Mishra, D Singh	Handwritten digit recognition using combined feature extraction technique and neural network	80
Y Minghui, Z Sheng	Fuzzy comprehensive evaluation model for mobility model	89
V Srivastava, D Singh	Enhance detecting and preventing scheme for ARP Poisoning using DHCP	93
Authors' Index		100
Cumulative Index		101





Effect of texture on mechanical and magnetic properties of steel from the petroleum distillation column

E Dragomeretskaya

South Ukrainian National Pedagogical University named after K.D. Ushinsky, 26 Staroportofrankovskaya Street, Odessa 65020, Ukraine

**Corresponding author's e-mail: drag_8181@mail.ru*

Received 6 February 2017, www.cmnt.lv

Abstract

Texture, mechanical properties and coercive force of steel 09G2S from the column fragment of petroleum distillation after prolonged use studied. Anisotropy of mechanical properties and coercive force take place. Significant pair wise linear correlations and appropriate regression equations with coefficients reliability of approximation not less than 0.90 were found between magnitudes of the coercive force, tensile strength, yield strength, elongation and texture characteristics. Found correlations may be used for nondestructive mechanical properties control of investigated steel by means of monitoring of coercive force.

Keywords:

texture, anisotropy, mechanical properties, coercive force, correlation

1 Introduction

Low alloy steels of type A515 and A516 are widely used in equipment of refinery complex, in particular for the production of distillation petroleum columns [1]. During the exploitation of the above equipment arise problems of mechanical properties of steel control, as well as the further safe operation estimation. Uniaxial tensile tests, fatigue, experiments on the long-term strength, etc. are carried out to study the mechanical properties [2, 3]. Cutting of samples from appropriate plots of material is necessary for such research. This requires stopping of the equipment operation. Therefore the development of non-destructive monitoring methods of the structural state and properties of the steel is important. The method of coercive force measuring is one of perspective non-destructive monitoring methods of structural condition of steels. Crystallographic texture as well as shape and size of grains, and elastic stresses have a main influence on the coercive force and her anisotropy [4]. Possibility of the structural state evaluation, of accumulated fatigue damage level, value of internal stress by measurement of coercive force was demonstrated in number of studies (e.g., [5-8]). In [6] was found a linear correlation of coercive force with the pole density on inverse pole figures (IPF) as well as with broadening of appropriate X-ray diffraction lines with increasing of the hydraulic pressure in the steel pipeline at testing. However relationship of coercive force (H_c) anisotropy with mechanical and structural characteristics of ferromagnetic construction steels is studied deficiently.

This work aimed to ascertainment of reasons anisotropy coercive force measured by non-destructive method, as well as relationship of coercive force with texture and mechanical characteristics of low-alloy steel of petroleum

distillation column after long-term use.

2 Experimental material and methods

Low alloy steel of type 09G2S thickness of 20 mm from the column fragment of petroleum distillation after long-term use was by material for the study. The studied steel has the following chemical composition: 0.11 wt% C; 1.47 Mn; 0.70 Si; 0.13 Cr; 0.05 Ni; 0.06 V; 0.02 Al; 0.02 P; 0.009 S; 0.05 Cu; 0.04 Nb; 0.03 wt% Mo; Fe balance.

The coercive force H_c was measured non-destructively using a magnetic analyzer (coercimeter) KRM-Ts-MA by overlay of pole tips of the portable measuring device on surface of the test product. The area of the test products between the pole tips of the magnetic converter is periodically magnetized to saturation by current pulses with amplitude of at least 2 A. Automatic compensation of residual magnetization field is then carried out. Value of coercive force is automatically calculated by the current magnitude of magnetic field compensation. Readings of device are dependent only from the metal properties but independent of confounding factors such as the protective coating (paint, film, etc.) to 6 mm on controlled metal or equivalent to this gap the corrosion metal, roughness, curvature etc. The maximum error does not exceed 2%. [13]. Coercive force was measured through every 15° from longitudinal direction (LD) up to the transverse direction (TD) orienting the measuring probe without damaging the product.

Samples for mechanical testing by uniaxial tension (Figure 1) with the diameter of working part of 3 mm were cut from the column fragment through every 15° from longitudinal direction (LD) up to the transverse direction (TD).



FIGURE 1 Sample after the test

The arithmetic average of test results of at least three specimens in every of above directions has been taken as the value of the corresponding mechanical characteristics. Mechanical testing was performed on a setup 1246-R. Velocity of the active grip was 2 mm / min. Mechanical properties were determined according to standard procedures [10].

The X-ray method was used for the study of texture [11]. Scanning θ - 2θ of the sample without texture (which was manufactured from sawdust of investigated steel after recrystallization), as well as of specimens cut out in the ND, DD, and in TD was performed by means X-ray diffractometer DRON-3M by Bragg-Brentano geometry in

the radiation of $K\alpha$ - Mo. Texture was investigated in the ND near the outer convex surface of the column, in the middle of fragment thickness, and near of her inner concave surface. Appropriate surfaces were chemically polished up to 0.1 mm before recording for removing of layer distorted by machining. On obtained data were constructed IPF for respective directions described above. The three-dimensional distribution function of the crystals orientation in space of ideal orientations were calculated by us from the IPF LD and the IPF TD according to the method described in earlier our work [12].

Metallographic structure of end surfaces of samples orthogonal to the RD and TD was examined by the microscope Axioplan 2 of firm KARL ZEISS.

3 Results and Discussions

Results of mechanical tests and measurement of the coercive force H_c are shown in Table 1.

TABLE 1 Mechanical properties and coercive force of steel samples, cut out in different directions from the fragment of distillation petroleum column

Angle with the LD, °	Tensile strength σ_m , MPa	Conditional yield strength $\sigma_{0.2}$, MPa	Relative elongation, $\epsilon = \Delta l / l$, %	Coercive force, H_c , A/cm
0	400±2.0	255±1.4	31.0±0.4	5.9±0.12
15	405±2.2	258±1.8	30.2±0.4	6.1±0.12
30	416±2.3	265±2.1	28.8±0.4	6.3±0.12
45	425±2.2	272±2.3	28.0±0.4	6.5±0.12
60	421±2.0	268±2.0	28.4±0.4	6.5±0.12
75	417±2.0	266±1.5	29.4±0.4	6.4±0.12
90	415±1.8	260±2.5	30.0±0.4	6.2±0.12

Anisotropy of mechanical characteristics and coercive force take place. The minimal values of the strength properties of σ_m , $\sigma_{0.2}$, and the coercive force H_c are observed in the LD. Their maximal values occur in the LD+45°, and in the TD they take an intermediate value. Elongation ϵ shows the opposite behavior.

Anisotropy coefficient η was calculated by the formula

$$\eta = (F_{max} - F_{min} / F_{min}) \cdot 100\% \tag{1}$$

Here F_{max} and F_{min} are maximal and minimal values of the corresponding property.

Anisotropy coefficients of σ_m , $\sigma_{0.2}$, H_c and ϵ amounted respectively 6.25 %, 6.27 %, 10.71 % and 10.17 %.

Strong linear correlations of the H_c value with values of mechanical characteristics σ_m , $\sigma_{0.2}$ and ϵ take place. Corresponding regression equations with high reliability approximation coefficients R^2 have the form

$$\sigma_m = 38.2H_c + 174.8 ; R^2 = 0.92, \tag{2}$$

$$\sigma_{0.2} = 26.1H_c + 99.6 ; R^2 = 0.93, \tag{3}$$

$$\epsilon = -4.6H_c + 58.0 ; R^2 = 0.89. \tag{4}$$

Experimental inverse pole figures obtained by us are presented in Figure 2.

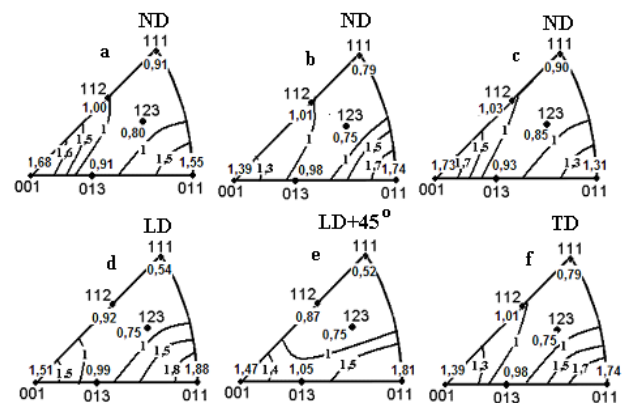


FIGURE 2 IPF's of steel column: a, c correspond to the convex and concave surface of the column respectively; b corresponds to the middle of the thickness of the metal; d-f correspond to the LD, LD+45° and TD

Texture of polycrystalline bodies presents a continuous distribution of crystals by orientations. At the same time there are certain preferable orientations of crystals, which are for clarity usually described using ideal orientations. Important components of the low carbon steel rolling texture are arranged along three fibres orientations [11]:

- 1) α -fiber with the fiber axis $\langle 110 \rangle$ parallel to the rolling direction including the main components of $\{001\} \langle 110 \rangle$, $\{112\} \langle 110 \rangle$ and $\{111\} \langle 110 \rangle$.
- 2) γ -fiber with the fiber axis $\langle 111 \rangle$ parallel to the normal direction including the main components of $\{111\} \langle 110 \rangle$ and $\{111\} \langle 112 \rangle$.
- 3) ϵ -fiber with the fiber axis $\langle 110 \rangle$ parallel to the transverse direction including the main components of $\{001\}$

$\langle 110 \rangle$, $\{111\}$ $\langle 112 \rangle$, $\{554\}$ $\langle 225 \rangle$ and $\{011\}$ $\langle 100 \rangle$.

When referring to ideal orientations $\{hkl\}$ $\langle uvw \rangle$ in the cylindrical sample we mean that planes of family $\{hkl\}$ are located in a plane tangent to the cylindrical surface, and a set of crystallographic directions $\langle uvw \rangle$, owned by $\{hkl\}$, are parallel to cylinder axis.

From Figure 2 it can be concluded that parallel to the side surface of the column metal are arranged families of crystallographic planes $\{001\}$ and $\{110\}$ since their pole density is greater than 1 that corresponds to the state without texture. Crystallographic directions $\langle 110 \rangle$ and $\langle 100 \rangle$ of families mainly coincide with the LD, TD and LD + 45°. A

TABLE 2 The composition and volume content of ideal orientations in the texture of steel of distillation oil column

Ideal orientation	$\{100\}\langle 010 \rangle$	$\{100\}\langle 011 \rangle$	$\{100\}\langle 013 \rangle$	$\{110\}\langle 110 \rangle$	$\{110\}\langle 111 \rangle$	$\{110\}\langle 001 \rangle$
Volume content	0.20	0.12	0.11	0.18	0.14	0.25

Crystallographic texture, shape and size of the grains, and the elastic stresses have a mainly influence on the coercive force and its anisotropy [4] as mentioned above. Metallographic analysis showed that the investigated steel has a typical ferrite – pearlite microstructure with average grain size of 22 μm (Figure 3). This microstructure can hardly be the main cause of the anisotropy of the coercive force.

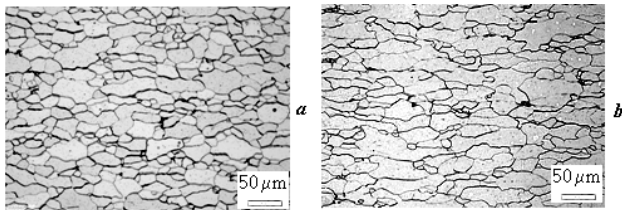


FIGURE 2 Ferrite – pearlite structure of the investigated steel: *a, b* have been photographed from the LD and TD direction

These features are caused by the magnitude of the external magnetizing field, if the steel does not magnetized to saturation. But when the coercive force is measured using the coercimeter, steel is magnetized to saturation, since the magnetizing field is sufficiently large ($B = 1.5$ T) [13]. Therefore, we can assume that the behaviors of domains in the magnetization and demagnetization play a secondary role in formation of the coercive force anisotropy, but the main role belongs to the energy of the magnetic crystallographic anisotropy in the investigated steel.

Let's estimate the energy of the magnetic crystallographic anisotropy in the investigated material. Suppose that the coercive force is associated only with the energy of the magnetic crystallographic anisotropy (external applied mechanical stresses are absent, structure of the investigated steel is homogeneous). Energy of the magnetic crystallographic anisotropy as a first approximation is expressed by the following equation [4] for the material with cubic lattice

$$W_k \approx K_1 \left(\alpha_1^2 \alpha_2^2 + \alpha_2^2 \alpha_3^2 + \alpha_1^2 \alpha_3^2 \right), \quad (5)$$

Here α_1 , α_2 and α_3 are direction cosines of the magnetization with respect to the cube axes; K_1 is anisotropy constant.

Let's will call of function energy of magneto-crystalline anisotropy the expression

three-dimensional ODF was calculated by us in the space of ideal orientations on the base of IPF LD (Figure 2(b)) and IPF RD (Figure 2(d)). Texture can be described as a combination of ideal orientations with the volume content, which are presented in the Table 2 as it was determined by the analysis of the ODF.

Effect of crystals orientation on the coercive force electrical steel previously is investigated in several studies [e.g. 14, 15]. Summarizing, can be concluded that the anisotropy of coercive force in the steel dependent not only from crystal orientation (i.e. from texture) but and from features of the magnetic domains formation.

$$\psi = \left(\alpha_1^2 \alpha_2^2 + \alpha_2^2 \alpha_3^2 + \alpha_1^2 \alpha_3^2 \right). \quad (6)$$

The direction cosines of orientations indicated in Table 2 are presented in Table 3. Numerical values of the function Ψ of magnetic crystallographic anisotropy energy calculated from (3) for combinations of ideal orientations as well as corresponding volume content (Table 2) and considering direction cosines (Table 3) are shown in Table 4.

TABLE 3 Ideal orientations and functions of the magnetic crystallographic anisotropy energy

α_1	α_2	α_3
$\cos \varphi$	$\sin \varphi \cdot \sin 90^\circ$	$\sin \varphi \cdot \cos 90^\circ$
$\cos(\varphi+45^\circ)$	$\sin(\varphi+45^\circ) \cdot \sin 90^\circ$	$\sin(\varphi+45^\circ) \cdot \cos 90^\circ$
$\cos(\varphi+18.43^\circ)$	$\sin(\varphi+18.43^\circ) \cdot \sin 90^\circ$	$\sin(\varphi+18.43^\circ) \cdot \cos 90^\circ$
$\cos(\varphi+90^\circ)$	$\sin(\varphi+90^\circ) \cdot \sin 45^\circ$	$\sin(\varphi+90^\circ) \cdot \cos 45^\circ$
$\cos(\varphi+54.7^\circ)$	$\sin(\varphi+54.7^\circ) \cdot \sin 45^\circ$	$\sin(\varphi+54.7^\circ) \cdot \cos 45^\circ$
$\cos \varphi$	$\sin \varphi \cdot \sin 45^\circ$	$\sin \varphi \cdot \cos 45^\circ$

TABLE 4 The calculated numerical values of function Ψ of the magnetic crystallographic anisotropy energy

Angle with LD, °	0	15	30	45	60	75	90
Ψ	0.09	0.16	0.21	0.24	0.23	0.19	0.15

Function Ψ takes the maximal value in the direction of LD + 45°. The minimal value of Ψ is observed in the LD. Function Ψ has an intermediate value in the TD. This is consistent with the character of coercive force anisotropy (Table 1). A strong linear correlation between the values of Ψ function and H takes place, as showed the correlative analysis conducted by us. The corresponding regression equation with a coefficient of reliability correlation $R^2 = 0.91$ has the form

$$\Psi = 0.23H_c - 1.23. \quad (6)$$

Thus, the character of the observable coercive force anisotropy in the studied steel of the oil distillation column can be explained, mainly by influence of the magnetic crystallographic anisotropy energy. Correlations (2) - (4) can be used to non destructive control the mechanical characteristics of steel 09G2S by measure of the coercive force during operation of oil distillation column.

4 Conclusion

(1) Mechanical properties, the coercive force, and the texture in the fragment of steel petroleum distillation column after prolonged use studied.

(2) Anisotropy of mechanical properties and coercive force take place. Minimal values of strength properties σ_m and $\sigma_{0.2}$ coercive force H_c are observed in the longitudinal direction, their maximum values occur in a diagonal direction, and in the transverse direction abovementioned properties takes the intermediate values. Elongation ε shows the opposite behavior.

(3) Strong linear correlations of the coercive force H_c with mechanical characteristics tensile strength σ_m , proof strength $\sigma_{0.2}$, and elongation ε are found. Reliability coefficients of linear approximations were no less than 0.89.

(4) Energy of the magnetic crystallographic anisotropy that is associated with orientation of crystals (i.e. with the texture), is main factor of coercive force anisotropy in the investigated steel. A strong linear correlation (with a correlation coefficient of at least 0.9) was found between calculated values function of magnetic crystallographic anisotropy energy and experimental values of coercive force.

References

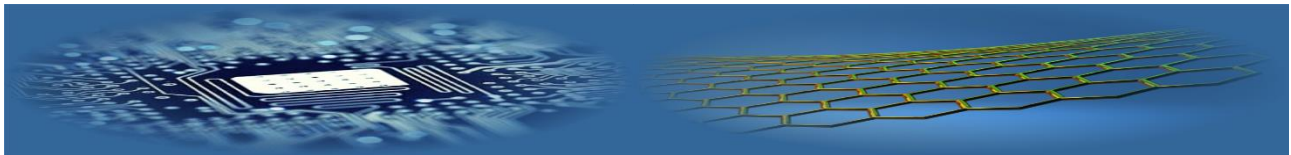
- [1] Towers, columns <http://instruct.uwo.ca/engin-sc/cbe497/Doc/Icarus/ir08.pdf>
- [2] Pressure vessel inspection code: in-service inspection, rating, repair, and alteration <https://law.resource.org/pub/us/code/ibr/api.510.2006.pdf>
- [3] Damage mechanisms affecting fixed equipment in the refining industry http://www.silcotek.com/hs-fs/hub/22765/file-815218612-pdf/docs/api_rp_571.pdf
- [4] Terunobu Miyazaki, Hanmin Jin, 2012 *The physics of ferromagnetism* Springer Berlin Heidelberg
- [5] Zhang Yu, Wang Z, Wang Y, Zhang Z, Zhang Yi 2015 A study on the relationship between hardness and magnetic properties of ultra-high strength steel *Advanced materials research* **106** 78-81
- [6] Solomakha R, Pittala R D, Bezlyudko G, Baskaran B V Practical evaluation of fatigue and stress state, and residual life of metal by non-destructive method for measuring magnetic characteristic "The coercive force" – A case study" <http://www.ndt.net/article/apcndt2013/papers/243.pdf>
- [7] Altawalbeh N A M, Thuneibat S A Al, Olaimat M M 2012 A new electromagnetic technique for controlling stress in metals *Applied physics research* **4**(3) 48-52
- [8] Gorkunov E S, Mitropolskaya S Yu, Zadvorkin S M, Shershneva L S, Vichuzhanin D I, Tuyeva E A *Relation of the magnetic properties of control-rolled pipe steel to its structural anisotropy, internal stresses and damage* http://www.ndt.net/article/ecndt2010/reports/1_01_24.pdf
- [9] Lobanov L M, Nehotyaschiy V A, Rabkina M D, Usov V V, Shkatulyak N M, Tkachuk E N 2010 Anisotropy of the coercive force and texture of deformed steel *Deformation and fracture of material* **10** 19–24 (In Russian)
- [10] *Tensile Testing* 2004 Materials Park, Ohio
- [11] Randle V, Engler O 2000 *Introduction to texture analysis: microtexture, microtexture and orientation mapping* CRC PRESS, Boca Raton, London, New York, Washington
- [12] Usov V V, Tarlovsky V A 1991 The calculation method of the three-dimensional orientation distribution function and integral texture characteristics of cubic polycrystals from inverse pole figures *Zavodskaya Laboratoriya* **7** 25-8 (In Russian)
- [13] *Nomenclature / Instruments and equipment for non-destructive testing / magnetic analyzer KRM-Ts-MA* <http://promsouz.com/pribori14.html> (In Russian)
- [14] Campos M F, De Campos M A, Landgraf F J G, Padovese L R 2011 Anisotropy study of grain oriented steels with Magnetic Barkhausen Noise *Journal of physics: conference series* **303** 012020
- [15] Paltanea V, Paltanea G, Gavrila H 2012 Magnetic anisotropy in silicon iron alloys *Electrical and electronic engineering* **2**(6) 383-8

AUTHORS



Elena Dragomeretskaya, 1981, Ukraine

Current position, grades: Leading specialist of postgraduate and doctoral studies,
University studies: South Ukrainian National Pedagogical University named after K.D. Ushinsky.
Scientific interest: Influence of crystallographic texture on the anisotropy of physical and mechanical properties,
Publications: 14
Experience: more than 10 years



Decision support system on the base of genetic algorithm for optimal design of a specialized maritime platform

Andrejs Zvaigzne^{1*}, Oleksandr Bondarenko², Anzhela Boiko³

¹Latvian Maritime Academy, Riga, Latvia, 12-k Flotes Street, LV-1016

²Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine, Geroev Ukraine Avenue 9, 54025

³Petro Mohyla Black Sea National University, Mykolaiv, Ukraine, 68 Desantnikov street 10, 54003

Corresponding author's e-mail: andrejs.zvaigzne@latja.lv

Received 16 May 2017, www.cmnt.lv

Abstract

The analysis of possibilities of application of the small waterplane area twin hull ships (SWATH) as a specialty (universal) platform is performed. It is shown that the design of the specialized platform with a small waterplane area twin hull is characterized by a large number of parameters to be determined. The optimum relation selection between SWATH dimensions, seaworthiness, cost and efficiency is proposed by solving a multidimensional optimization problem with the use of special methods of searching solutions. The optimization problem of designing a universal platform is formulated. The constraints accounting on SWATH technical characteristics is produced by using the method of penalty functions. To solve the optimization problem, one of modern search methods – genetic algorithm is used. An example of solving the problem of selection the main dimensions of 25 m platform using a genetic algorithm is presented.

Keywords:

SWATH, specialized platform, genetic algorithm, optimization, Mission Module

1 Introduction

For most small countries, as well as countries with limited funding for the maintenance and development of the fleet, the actual problem is to provide the legal regime and protect national interests in the maritime exclusive zone in difficult weather conditions.

One of the possible ways of solving this problem is to use a universal specialized platform with increased seaworthiness and small dimensions. One of the most rational options for such platform could be a small waterplane area twin hull ship (SWATH).

Universal platform type SWATH use has a number of advantages in contrast with other architectural and structural types of ships, as follows:

- large area that allows to accommodate replacement modules, additional equipment to expand the functional capacities;
- high seaworthiness, providing speed loss on seaway and smooth motion;
- high survivability in case of emergency;
- high firmness on a course.

Currently there is an experience of using this ships type as a universal platform in the world. For example, a 25 meter ship project developed by Abeking & Rasmussen's (Germany). The platform has the following characteristics: length overall – 25,65 m; length between perpendiculars – 23,25 m; breadth overall – 13,0 m; depth – 5,9 m; design draught – 2,7 m; vertical clearance – 1,7 m; lower hull length – 26,65 m; lower hull maximum diameter – 2,4 m; lower hull transverse section shape – round. 125–135 tons displacement, depending on the purpose. The platform is

based on a twin hull ship with two struts on each hull. Propulsion plant type is diesel-electric. Currently on the basis of this platform, 19 ships for various purposes are built, they are: 10 pilot boats, 1 research vessel, 6 patrol vessels, 1 for maintenance personnel delivery to offshore wind power plants and 1 pleasure yacht (Figure 1). Two more pilot boats are planned to be delivered in 2017 for the Houston Pilot. It is also possible to expand the ship functions by installing replacement modules (Figure 2) (Grannemann, 2015).

Small waterplane area twin hull platform designing is associated with certain difficulties caused by the following factors:

1. Insignificant design experience.
2. The presence of a large number of parameters that determine the hull shape. For a traditional single hull ship, the hull is determined by nine parameters: length, breadth, draught, depth, three fullness coefficients, center of buoyancy position by ship length and waterplane centroid position by ship length.

For a small waterplane area twin hull ship, there are much more of such parameters as SWATH hull consists of the following structural elements: box, lower hulls, struts and sponsons. Each of these structural elements is characterized by a set of its parameters length, breadth, depth, fullness coefficients. Moreover at the SWATH full hull parametric design, it is necessary to take into account the mutual position and structural elements interaction, which determines the SWATH hydrodynamic characteristics in general. All these factors result in a specialized platform required optimal parameters vector large dimension and technical solutions significant variety.



Pilot Tender



Patrol



Windpark Tender



Hydrographic Vessel



Yacht



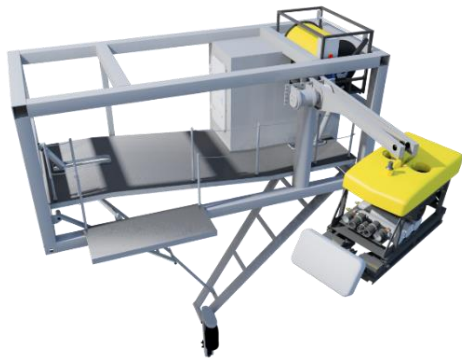
MCMV Demonstrator

FIGURE 1 Different variants of 25 m platform use

Therefore, it is rational to choose the optimal relationship between SWATH dimensions, seaworthiness, cost and efficiency by solving a multidimensional optimization task using special methods of solution search. The multidimensionality of the optimization problem leads to the so-called "curse of dimension" in optimization theory and, as a result, to a significant increase in the volume of computations and the complexity of finding the global optimum. One of the solutions to this problem is the use of a genetic algorithm. Currently, this algorithm is increasingly used in the marine industry, for example different aspects of passenger SWATH design optimization task solution are considered in (Bondarenko et al., 2013). The issues of genetic algorithm application for ship hull optimization are

considered in the following articles (Guha, and Falzarano, 2005), (Zakerdoost et al., 2013) and (Dejhalla, R., Mirsa, Z., Vukovic, S. (2001). Application of genetic algorithm for the design of other types of vessels considered in paper (Sekulski, 2011), (Papanikolaou, 2012), (Boulougouris et al., 2012), (Gammon, 2011) and (Brown, Salcedo, 2003). At the same time, the universal specialized small waterplane area twin hull platform optimization design algorithm is underexplored.

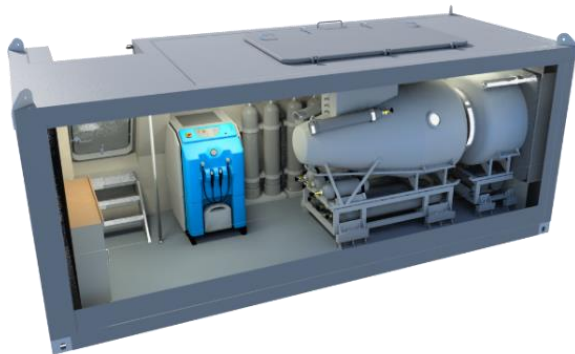
Therefore, the objective of this article is to develop a key element in ship design methodology - special algorithm for selecting the optimal characteristics of universal specialized small waterplane area twin hull platform using a genetic algorithm.



ROV Mission Module



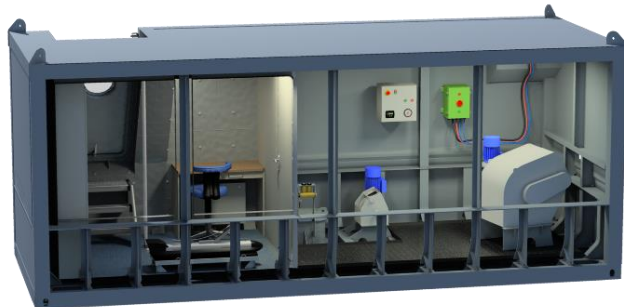
Fire Fighting MissionModule



Diving Mission Module



Gun Mission Module



Sidescan Sonar Mission Module



Oil Spill Recovery Mission Module

FIGURE 2 Mission modules for SWATH

2 Methodology description

2.1 PROBLEM STATEMENT

The optimization task of selecting the optimal characteristics of a universal specialized platform is formulated as follows:

$$F(X,C) \rightarrow \min_{X \in R^n} (\max_{C \in R^m})$$

$$\text{subject to a: } g_j(X) = 0, \quad j = p+1, \dots, k,$$

$$g_j(X) \geq 0, \quad j = 1, \dots, p,$$

$$a_i \leq x_i \leq b_i, \quad i = 1, \dots, n$$

where $X \in R^n$ is the vector of independent variables, $F(X,C)$ is the objective function, $C(C_1, \dots, C_m)$ is a vector of the parameters that form the design task; m is a number of the C vector parameters; n is a number of independent variables; k is a total number of optimization task constraints; p is a number of optimization task constrains in the form of inequalities; R^n is n -dimensional Euclidean space; g is the vector constraints; a_i is lower bounds on the independent variable, b_i is upper bounds on the independent variable; x_i is values of the independent variables, j is index for constrains; i is index for independent variables.

The main questions connected with a universal specialized platform design task formulation were considered in (Zvaigzne and Bondarenko, 2017). One of the

features of this task is the availability of a limitations system to the platform technical qualities as $g_j(X) \geq 0$. To take them into account in this research, the Penalty Functions method (Rao, 2009) is applied. The main idea of the Penalty Functions method is to turn the task of conditional optimization into unconditional by replacing the objective function:

$$F_l(X, C, r_l) = F(X, C) \pm \sum_{j=1}^k \left[\frac{g_j^+(X)}{r_l} \right]^n,$$

where r_l is the penalty coefficient, which value decreases from one stage to another; l is calculating optimization process cycle number; n is the degree, in this research $n = 2$;

$\sum_{j=1}^k \left[\frac{g_j^+(X)}{r_l} \right]^n$ is penalty for limitations violation (penalty function):

$$g_j^+ = \begin{cases} \max \{g_j(X); 0\}, & j \in [p+1, k] \\ |g_j(X)|, & j \in [1, p] \end{cases}.$$

The resulting new objective function $F_l(X, C, r_l)$ hereinafter referred to as fitness function, which corresponds to the terminology used in (Back, 1996), (Davis, 1991), (Rutkovskaya et al., 2006), (Sivanandam and Deepa 2007). The fitness function is minimized (maximized) using the genetic algorithm. While using the genetic algorithm, the independent variables boundary values do not participate in the penalty functions creation, since they are used in encoding/decoding of independent variables (an independent variable will always be in the boundary range).

For example, encoding/decoding real-valued independent variable

$$c = \frac{(x_i - a_i)(2^s - 1)}{(b_i - a_i)};$$

$$x_i = \frac{c(b_i - a_i)}{2^s - 1} + a_i,$$

where

a_i, b_i is lower and upper bounds of the i -th independent variable; s is the number of bits per one element of chromosome (gene); x_i is the decoded real value from bit string of length s . c is the coding representations of x_i .

2.2 ALGORITHM DESCRIPTION

Let us consider in more detail the genetic algorithm nature and the features of its use for the universal (specialized) platform designing.

The genetic algorithm operation is based on the processes of natural selection and evolution occurring in living nature. In nature, the most adapted individuals survive and give offspring, i.e. the principle "the strongest survives" is observed. In terms of optimization, the search of optimal, i.e. the best solution, corresponds to the search of the fittest individual. And the best solution searching iterative process resembles the population evolution in nature. Only in nature

the fittest individuals give offspring, and in the optimization task – they form the task allowable solutions.

The independent variables X vector numerical values are the individual genetic code and should be stored in the computer's memory in the form of a fixed-length line for the selection process realization. There is range of ways to represent numbers (encoding) in genetic operators: decimal, binary, Gray encoding.

At fitness function calculations, as well as at the optimal solution output, the values are decoded, i.e. converted into numerical values.

The optimal solution searching general scheme using a genetic algorithm can be represented as follows (Figure 3):

1. To generate N individuals initial population;
2. To measure chromosomes fitness in the population on the basis of the objective function $F_l(X, C, r_l)$;
3. To perform the selection operation, i.e. for each agent of the new generation to select two parents from the current generation in proportion to fitness;
4. For selected parents to create candidates for the new population creation using genetic operators (mutations, crosses, inversions, mutation);
5. To create a new population;
6. If the criterion for stopping the algorithm is done, then finish the search, otherwise – to do the iteration search next cycle.

To create a new population, the so-called genetic operators are used: selection, crossover, mutation, inversion.

The selection of individuals (parents) involved in the creation of offspring is done using selection operators. There are several options of selection mechanism realization: roulette-wheel selection, tournament selection, ranking selection etc. Detailed information about each of the selection options is given in (Back, 1996), (Davis, 1991), (Rutkovskaya et al., 2006), (Sivanandam and Deepa 2007). In this article the tournament selection in which all populations are divided into subgroups that consist of two individuals is used by authors. Then the individuals with the best fitness are selected in each of these subgroups. The diagram in figure 4 below illustrates the tournament selection method for subgroups that consist of two individuals.

The crossover operator is a language construction that allows creating descendants chromosomes on the basis of the parents chromosomes transformation (crossing) (or their parts). The crossing operator exchanges chromosome parts between two (maybe more) chromosomes in the population. There are different types of crossing, as their structure basically determines the genetic algorithms efficiency (Back, 1996), (Davis, 1991), (Rutkovskaya et al., 2006), (Sivanandam and Deepa 2007).

In this article, single-point crossover is realized. In this case, two individuals are selected, the chromosomes of which are cut into parts at the so-called Crossover point. Two segments are the result. Then, the corresponding segments of different chromosomes are glued together and two genotypes of descendants are obtained.

Crossing does not always apply to all pairs of individuals. Couples are usually chosen randomly, and the probability of crossing is assumed to be equal to any number from 0,6 to 1,0. Crossing is allowed if the random number (obtained with the help of the random number sensor in the range from 0 to 1) is less than the predetermined probability. If the

crossing does not occur, the offspring copy parents exactly. The crossing operator operation is illustrated by the

following example (Figure 5).

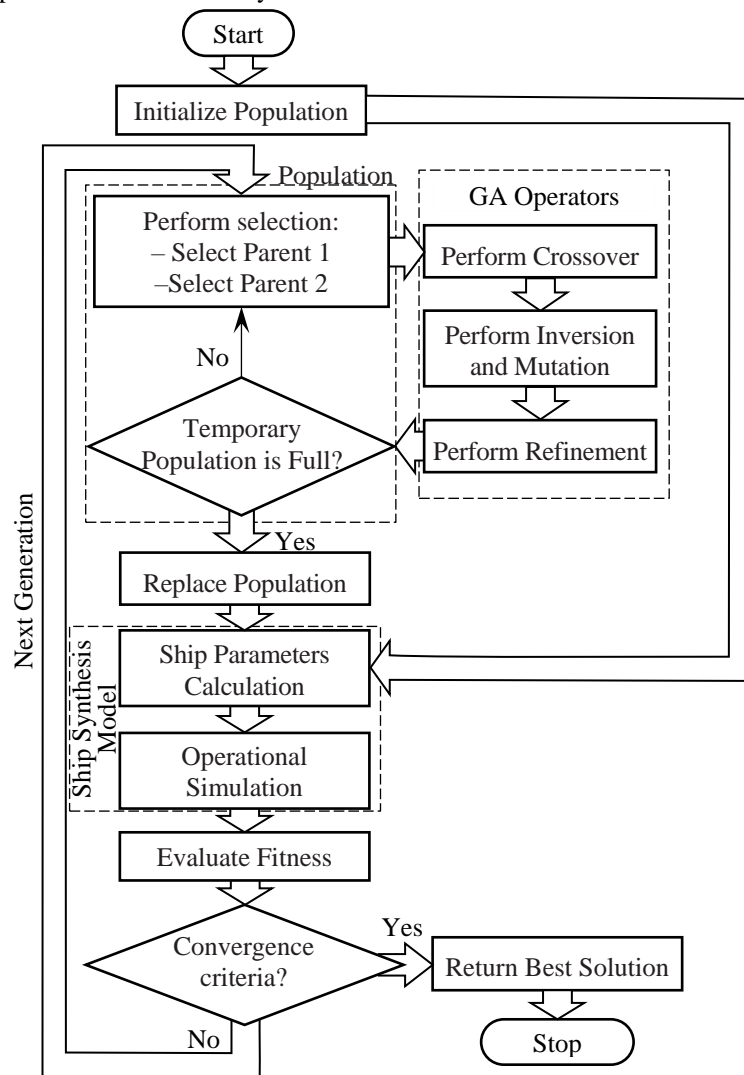


FIGURE 3 Block scheme of platform optimal parameters selection using the genetic algorithm

Fitness	Initial Population		
220	1 0 1 0 1 0 1 0 0 1	Selection	Selected parent string one 1 1 0 0 1 1 0 1 0 1
90	1 1 0 0 1 1 0 1 0 1		
80	1 1 1 1 1 0 1 0 1 1		
700	1 1 1 0 0 1 1 1 1 1		
190	1 1 0 0 1 1 0 1 0 1		
480	1 0 1 1 1 0 1 0 1 1	Selection	Selected parent string two 1 1 1 0 0 1 1 1 1 1
230	1 1 0 0 1 1 0 1 0 1		
380	1 1 1 0 0 1 1 1 1 1		

FIGURE 4 Example GA population solutions and selection operation

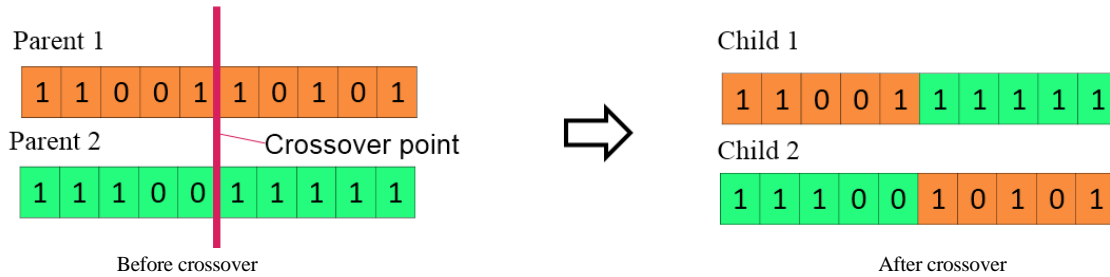


FIGURE 5 Genetic algorithm one point crossover operations

To support the diversity of individuals in a population, a mutation operator is used, a language construction that allows the descendant chromosome creation on the basis of the parent chromosome transformation (or its part). The mutation operator randomly changes each gene in the

chromosome with a little probability P_{mut} (user-defined), with 0 being replaced by 1 and backwards (Figure 6). It is realized with the random number generator help in the same way as crossing.

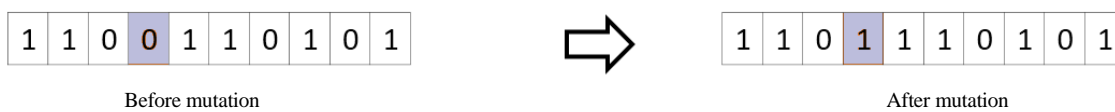


FIGURE 6 Genetic algorithm mutation operations

Inversion (inversion operations) – is piece or full chromosome U -turns. Inversion is performed on a single chromosome; at its realization, the consequence of alleles between two randomly chosen positions in the chromosome

changes (the last gene changes places with the first, the penultimate - with the second, etc.). An example of an inversion is illustrated below (Figure 7).

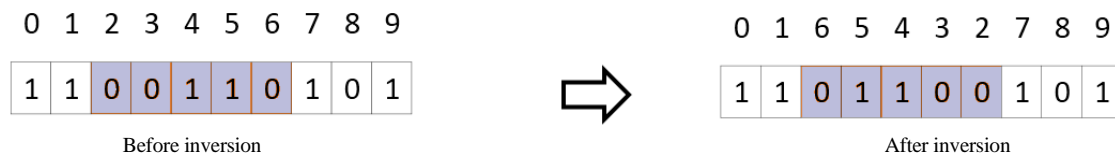


FIGURE 7 Genetic algorithm inversion operations

At new population creation, either a complete replacement or a partial replacement of the previous generation is possible, at which part of the population goes into the next generation without changes, i.e. the chromosomes of this part are not exposed to the crossing and mutation operations (the elitism strategy).

The main task of such a platform - high-speed transportation of passengers and service personal. Change the target purpose of the platform is possible by modifying of the superstructure and plug-in mission modules.

The new population creation corresponds to the genetic algorithm one iteration.

The initial data for the designing: ship speed is 30 knots, seaworthiness is force 4, number of struts is two on each hull, and endurance at maximum speed is 300 miles. As an indicator of economic efficiency, net discounted income was used (Net Present Value – NPV):

As a search completing criterion can be:

- generations given limiting number achievement;
- time set period expiration;
- fitness function values stabilization (lack of fitness function values changes);
- good enough solution getting.

$$NPV = \sum_{t=1}^T \frac{(Pr_t + A_t) - IC_t}{(1 + d)^t},$$

As a result of modeling the evolutionary process with the genetic algorithm help, we get the most adapted individual, i.e. the optimization problem solution.

where Pr_t - the sum of net profit in the t -th period; A_t - the amount of depreciation deductions in the t -th period; IC_t - the amount of investment costs in the t -th period; t - current year of the billing period; T - duration of ship life cycle (assumed to be 15 years); d is the discount rate.

The proposed approach to the solution of the optimization task of selecting the design characteristics of the offshore platform using the genetic algorithm was implemented in the form of the software product “SWATHShips”

To determine the SWATH main characteristics the genetic algorithm was used with the following parameters: the population number is 50 chromosomes, the genes length are 32 bits, the crossover probability is 0,9, the mutation probability is 0,1, the inversion probability is 0,05 the initial value of penalty coefficient r_1 is 0,5, the extremum reaching accuracy is 0,000001. The elitism strategy was used at optimization. These parameters are set experimentally as a

2.3 CALCULATION RESULT

Let us consider the genetic algorithm use for choosing the ship optimal dimensions on the example of small waterplane area twin hull small-specialized platform designing. The

result of many test runs of the program. The search for optimal characteristics was carried out using the software product "SWATHShips".

The optimized variables values and the SWATH main characteristics that are obtained as a result of the optimization program are given in Table 1 and Table 2 respectively.

TABLE 1 The optimized variables values

Independent variable	Material of hull and superstructure		
	Steel	Aluminum	Steel+Aluminum
Relative length of lower hull L_H/D_H	11,883	14,705	12,270
Slenderness coefficient of strut L_S/t_S	22,508	19,910	24,388
Waterplane area strut coefficient C_{WPS}	0,849	0,873	0,853
Relative waterplane area $A_{WPS}/\nabla^{2/3}$	1,310	1,378	1,084
Ratio of the distance between lower hull center-line to the length of the ship B_S/L_{OA}	0,408	0,383	0,400
Ratio of the ship draft to the lower hull diameter d/D_H	1,417	1,617	1,421
Ratio of the lower hull beam to its depth, B_H/H_H	1,301	1,092	1,190
Lower hull prismatic coefficient C_{PH}	0,891	0,868	0,880
Factor of the lower hull nose shape n_f	2,275	3,833	3,818
Factor of the lower hull tail shape n_a	2,244	3,996	2,233
Factor of the lower hull cross section shape n_h	4,617	2,506	4,164
Strut nose and tail shape factor n_s	3,906	2,553	2,508
Strut setback L_{CS}/L_H	-0,019	0,002	0,014

TABLE 2 The SWATH main characteristics obtained as a result of optimization

Description	Hull/Superstructure material		
	Steel/Steel	Aluminum Alloy/Aluminum Alloy	Steel/Aluminum Alloy
Lower hull length, m	25,755	25,123	25,808
Lower hull beam, m	2,472	1,786	2,295
Lower hull depth, m	1,9	1,635	1,928
Hull nose length, m	3,863	3,768	3,871
Hull tail length, m	3,963	9,731	5,887
Strut length, m	26,024	20,888	23,734
Strut thickness, m	1,156	1,049	0,973
Strut height, m	2,885	2,752	2,817
Strut nose length, m	6,506	5,222	5,934
Strut tail length, m	10,852	6,672	8,994
Waterplane area strut coefficient	0,849	0,873	0,853
Box clearance, m	2,092	1,743	2,005
Distance between lower hull center line, m	10,507	9,628	10,328
Ship draft, m	2,693	2,643	2,740
Depth up to the main deck, m	5,79	5,367	5,749
Length overall, m	26,378	25,123	25,808
Box length, m	26,378	25,123	25,808
Box beam, m	12,979	11,413	12,623
Height of cross structure box, m	1,004	0,98	1,004
Displacement, t	250	150	225
Deadweight, t	39,26	34,75	37,95
Main Engines, number × kW	2 × 3460	2 × 2300	2 × 3460
Generator, kW	190	190	190
Payback period, years	9,3	5,9	8,4
Net present value, thousand\$.	2390	4357	2427

3 Conclusions

Practice shows that 80% of the ship's lifecycle costs formed during the ship conceptual design phase (Brown, A., Salcedo, J. 2003), and it may be several million dollars in ship lifecycle years. Small waterplane area twin hull specialized platform designing is characterized by a large number of

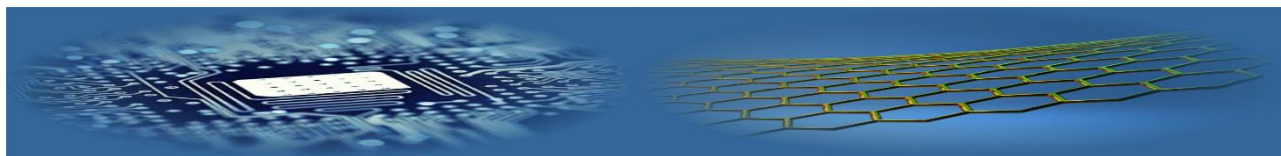
parameters to be determined. The parameters optimal values can be selected using a genetic algorithm, the algorithm help in short time solve complicated problems of optimization tasks by using parallel calculations possibility, providing possible options for the ship preliminary design. Without that information, it will be more complicated to take correct, optimal decisions on maritime platform design.

References

- [1] Back T 1996 *Evolutionary algorithms in theory and practice* New York: Oxford University Press
- [2] Bondarenko O V, Boiko A, Seropyan I 2013 Determination of the main characteristics of the small waterplane area twin hull ships at the initial stage of design *Polish Maritime Research* 20(1) 11–22 DOI: <https://doi.org/10.2478/pomr-2013-0002>
- [3] Boulougouris E, Chontzopoulos D, Papanikolaou A 2012 A conceptual design tool for the multi-criteria optimization of ships by use of Genetic Algorithms In: *4th International Symposium on Ship Operations, Management and Economics*, Athens, Greece, November 2012
- [4] Brown A, Salcedo J 2003 Multiple-objective optimization in naval ship design *Naval engineers journal* 115(4) 49–62
- [5] Davis L 1991 *Handbook of genetic algorithms* New York: Van Nostrand
- [6] Dejhalla R, Mirsa Z, Vukovic S 2001 Application of genetic algorithm for ship hull form optimization *International shipbuilding progress* 48(2) 117–33
- [7] Gammon MA 2011 Optimization of fishing vessels using a multi-objective genetic algorithm *Journal of ocean engineering* 38(10) 1054–64
- [8] Grannemann F 2015 SWATH - A new concept for the safety and security at sea *Ship science & technology* 8(17) 47–56

- [9] Guha A, Falzarano J 2005 Application of multi objective genetic algorithm in ship hull optimization *Ocean systems engineering* **5**(2) 91–107 DOI: <http://dx.doi.org/10.12989/ose.2015.5.2.091>
- [10] Michalewicz Z 1996 *Genetic algorithms + data structures = evolution programs* Berlin-Heidelberg: Springer-Verlag
- [11] Papanikolaou A 2012 Holistic ship design optimization: merchant and naval ships *Journal ship science and technology* **5**(9) 9–26 Available at: <http://www.shipjournal.co/index.php/sst/article/view/48>
- [12] Rao S S 2009 *Engineering optimization: theory and practice* Hoboken: John Wiley & Sons, Inc.
- [13] Rutkovskaya D, Pilinskiy M, Rytkovskiy L 2006 *Neural networks, genetic algorithms and fuzzy systems* Moscow: Goryachaya liniya – Telekom Publ.
- [14] Sekulski Z 2011 Multi-objective optimization of high speed vehicle-passenger catamaran by genetic algorithm *Polish maritime research* **4**(71) 3-13 DOI: <https://doi.org/10.2478/v10012-011-0020-0>
- [15] Sivanandam S N, Deepa S N 2007 *Introduction to genetic algorithms* New York: Springer Publ.
- [16] Zakerdoost H, Ghassemi H, Ghiasi M 2013 Ship hull form optimization by evolutionary algorithm in order to diminish the drag *journal of marine science and application* **12** 170–9 DOI: 10.1007/s11804-013-1182-1
- [17] Zvaigzne A, Bondarenko O 2017 Efficiency estimation of specialized multifunctional ships at optimal designing *Transport and telecommunication* **18**(1) 70–8 DOI: <https://doi.org/10.1515/tj-2017-0007>

AUTHORS	
	<p>Andrejs Zvaigzne, 18 June 1963, Liepaja, Latvia</p> <p>Current position, grades: Vice rector In Latvian Maritime Academy, Associate professor. University studies: M.Sc. in field of navigation (Naval academy-1985; Naval Staff College- 1997; Naval Command College-2004). Scientific interest: Ship building, e- navigation, multifunctional ships. Publications: 12 scientific papers including journal articles and domestic and international conference and congress contributions Experience: 26 years Navy, project development, academic.</p>
	<p>Oleksandr Bondarenko, 20 January 1970, Vinnitsa region (Ukraine)</p> <p>Current position, grades: Director of the Shipbuilding Institute, Associate Professor of theory and ship design department, Ph.D. in Naval Architecture and Marine Engineering (Admiral Makarov National University of Shipbuilding). University studies: M.Sc. in Naval Architecture, Mykolayiv Shipbuilding Institute named after admiral S.O. Makarov (currently: Admiral Makarov National University of Shipbuilding), 1994. Scientific interest: Wind Farm Support Vessel, SWATH and multi-hull ship design, simulation modelling, optimization, genetic algorithms and their application in ship design, development of advanced ship design methodologies, risk-based ship design, safety of ships. Publications: 170 scientific papers including journal articles and domestic and international conference and congress contributions Experience: 4 years of research assistant, 18 years university lecturer. ResearcherID: J-5948-2015 Scopus Author ID: 55767287600 orcid.org/0000-0002-6115-1422</p>
	<p>Anzhela Boyko, 04 April 1970, Nikolayev region (Ukraine)</p> <p>Current position, grades: Associate Professor at the Computer Engineering Department, Ph.D. in Naval Architecture and Marine Engineering (Petro Mohyla Black Sea National University). University studies: M.Sc. in Naval Architecture, Mykolayiv Shipbuilding Institute named after admiral S.O. Makarov (currently: Admiral Makarov National University of Shipbuilding), 1993. Scientific interest: SWATH and multi-hull ship design, simulation modelling, optimization, genetic algorithms and their application in ship design, development of advanced ship design methodologies. Publications: 77 scientific papers including journal articles and domestic and international conference and congress contributions Experience: 15 years university lecturer. Scopus Author ID: 57190417475 orcid.org/0000-0002-3449-0453</p>



Cyber intelligence systems based on adaptive regression splines and logical procedures of attack recognition

Beketova G^{1*}, Akhmetov B¹, Korchenko A², Lakhno V³, Tereshuk A³

¹Kazakh National Research Technical University named after K.I.Satpayev, Kazakhstan

²National Aviation University, Ukraine

³European University, Ukraine

*Corresponding author's e-mail: beketova2111@gmail.com

Received 20 April 2017, www.cmnt.lv

Abstract

The article presents the results of research devoted to the further development of methods, models and algorithms for recognizing cyber threats, as well as the most common classes of cyber attacks and anomalies in critical computer systems (CCS). It is shown that the cyber security process for CCS controlled and analyzed by the values of several parameters of anomalies or signs of cyber attacks. This, in turn, makes it possible to carry out a preliminary assessment of information security with the help of two-stage recognition procedure in which initially used the methodology of adaptive regression splines for the processing of statistical data on the anomalies and cyber incidents in CCS, and then in the second stage are used designed logical recognition procedures based on the signs of matrix surfaces. This minimizes the number of training samples for the detection of objects in the framework, the relevant classes of cyber threats, attacks and anomalies.

The research on minimizing the amount of training samples of recognizing signs were performed. It is shown that for the recognition of objects within the known class of cyber threats, attacks and anomalies in the use of training facilities matrices used for training a representative set of long 3-5 attributes will allow to achieve maximum efficiency of the algorithm, reaching up to 98%.

Using the proposed method and models has allowed to reduce the amount of required object recognition rules within the class of 2.5-10 times, compared to the widely used in anomaly detection systems and methods of cyber attacks sequential sorting features and statistical algorithms states.

Keywords:

intelligent recognition system, cyber threats, anomalies, signs of cyber attacks, adaptive regression splines, logical procedures, elementary classifier

1 Introduction

The widespread use of computer systems and information and communications technology improves business efficiency, reduce raw material costs, improve product quality, etc. Today, critical computer systems (CCS) play a key role in the deployment, operation and maintenance of information and communication infrastructure (ICI), responsible for the timely delivery to the consumers of energy resources, water, food, transport services and communications. The most important element of ICI are computerized systems and information technology, disruption of which can lead to serious or even explosive social and economic consequences in the country or a particular region, that is caused by a strong system interconnection between the various components CCS and life support systems. To ensure high performance, reliability and security CCS will need to proactively solve problems related to their information security (IS) and cyber defense.

Active extension applications of CCS, especially in the segment of mobile, distributed and wireless information technology, accompanied by the emergence of new threats to IS, as evidenced by the rapid growth in the number of incidents related to IS and cyber defense CCS and identified vulnerabilities in their software. Thus, the relevance of

studies aimed at the further development of models and methods of protection on the basis of intellectual recognition of threats of CCS and providing them by information security is one of the key problems of cyber critical infrastructure of any state.

2 The aim and tasks of the research

The aim of the work is the further development of models and methods of protecting of critical computer systems based on the use of adaptive, self-learning cyber capable of recognition systems, anomalies and attacks.

To achieve this goal it is necessary to solve the following tasks:

1. Develop the method of recognition of threats, anomalies and cyber attacks, allowing to provide cyber defense of CCS based on the application of innovative adaptive cyber defense systems to increase resilience of CCS to cyber attacks.

2. Develop the model of intellectual recognition using of adaptive regression splines and logical procedures for the identification of anomalies and cyber attacks, based on the signs of matrix surfaces (MS) and the concept of an elementary classifier (EC).

3 The review of the previous researches

According to various sources [1-3], for the period from 2009 to 2015 the number of cyber incidents, including cyber attacks directed at information system of countries, entering the torus 20, has grown by an average 15 times, Figure 1. And the tendency of a strong growth of number of cyberincidents and cyber attacks is fixed that, in particular, is explained by growth of quantity of CCS connected to wide area networks.

After, in industrial, energy and transport CCS were identified as complex viruses like the Stuxnet (2010), Duqu (2011), Flame (2012), Careto (2014), there was a sharp jump in interest in the IS critical automated control systems (ACS or SCADA). As a result, during the period from 2011 to 2015, in critical components in SCADA was found more than 130 vulnerabilities [3, 4].

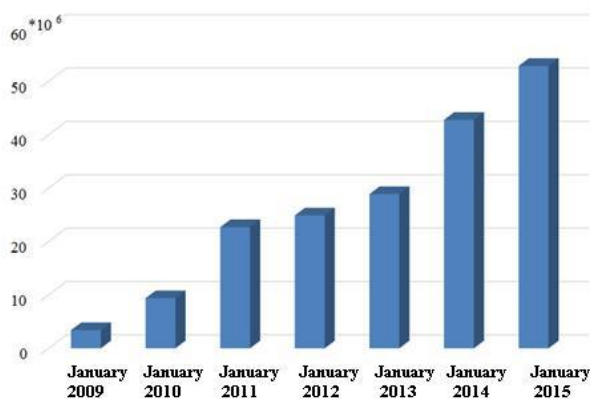


FIGURE 1 Dynamics of cyber incidents in CCS for the period from 2009 to 2015 (Sources: [1-3])

Growth of number of the fixed cyber threats, anomalies and cyber attacks become a powerful stimulus of development of research in the field of analysis and synthesis of various detection systems and detection of anomalies and cyber attacks.

Rather large numbers of works for the last decade are devoted, also to the problems of the development of intelligent intrusion detection systems (IDS). In particular, in works [5-7] presented reviews of anomaly detection methods, offered the principles of classification detection methods based on machine training and the statistical analysis. Overview of modern methods of machine training for systems to detect of cyber attacks (SDCA) is rather fully presented in works [8-10]. However, beyond these publications were some techniques such as k-average method [11] and its modifications [12, 13]. Methods for detection of cyber attacks based on finite state machine (FSM) are rather in detail stated in works [14-16]. Another perspective direction of development of adaptive systems of recognition cyber attacks (ASRCA) is a direction connected with the detection of abuse based on CCS states [17-19].

In work [20] considered a possibility of using in ASRCA IDA-splines, allow to build accurate approximation of behavior of the normal user or attacker on the set parameters.

Methods of computational intelligence, in particular, neural networks (NN) for problems of detection of cyber attacks are described in works [21, 22]. In [17, 23] describes the models and methods of adaptation of genetic algorithms

for the detection of the problem of cyber attacks. In works [24, 25] computing immune system, which can be used for the task of building ASRCA are described.

Typical disadvantage of most SDCA described in [17, 18, 21] works - erroneous operation during recognition. This is in particular due to the use, in most existing systems, a technology for detecting (identifying attacks). According to many authors [19, 20, 24, 25], the most perspective direction of development of cyber attacks and anomalies detection methods is to associate existing approaches in adaptive or hybrid SDCA with the ability to self-training.

Among the methods that are used in SDCA, researchers have identified the following areas: 1) the detection of anomalies in the system (anomaly detection system - ADS); 2) detection of abuses [5, 6, 22]. In works [9, 18] examined the peculiarities of SDCA, which used different methods and models. Applied aspects of commercial SDCA - IDES, NIDES, EMERLAND, JiNao, HayStack, etc., are considered in works [22, 23].

Anomalies detection methods (ADM), offer an opportunity for defense to perform detection with a high degree of accuracy and to make an informed opinion about the cause of changes in the state CCS. To create ADM decided to use: 1) controlled training; 2) uncontrolled training [5, 17]. The difference between the approaches is that the discrete set of features used in a controlled learning and training duration is determined previously. For uncontrolled training set of features usually change over time, and training can continue with the improvement of the system. Today, only a controlled training is used in commercial IDS [24].

Most modern SDCA and ADS based on models and methodologies which founded in pattern recognition theory [26-29]. In accordance with the basic principles of this theory to detect anomalies or cyber attacks, it is necessary to form an image of normal and abnormal behavior of CCS, for example, using expert assessment. So formed image, can be described as a set of values of the evaluation parameters i.e. signs, for convenience will be applied binary forms of describing features that are stored in the repository. If the image changes at some point of time, we can talk about abnormal functioning of the system. After, the anomaly or a cyber attack identified, and also assessed the degree of risk for functional tasks CCS, SDCA or ADS gives a conclusion on the possible cause of the changes. Thus, it is possible one of the following options for this conclusion: the change of state CCS - the result of cyber attacks; change of state - tolerance.

The difficulty in implementing of existing models of SDCA formalized device of recognition theory lays on that a particular informative complex for CCS often including unique software and information files, as well as its own IS subsystem which consists of heterogeneous components. However, carried out within this research, a specification of problems of cyber attacks recognition and application of models, which can to minimize the amount of training samples in the form of matrix signs, as well as elementary classifiers for each simulated class of cyber attacks, will optimize the work SDCA.

4 The use of adaptive regression splines in the intellectual system of cyber defense

Modern cyber attacks have become extremely complex.

Narrowly focused, systematic and specialized (targeted) attacks, able to hide from anti-virus systems, and are not always detected by firewalls and intrusion detection systems. Thus, further research is needed, which directed to develop the methodological and theoretical bases of creation of adaptive capacity to learn cyber attacks recognition systems that include the various technologies of detection and recognition, Figure 2. Within this research will consider the ASRCA, based on two landmark detected threats, cyber attacks and anomalies. The first stage uses the methodology of statistical data processing, which used adaptive splines. And on the second stage of the work of ASRCA are involved logical procedures.

RandomForest algorithm and the use of multivariate adaptive regression splines (multivariate adaptive regression splines - MARS).

RandomForest [20, 22, 28, 29] algorithm - an algorithm of machine learning, the essence of which is to construct a plurality of decision trees for the training sample. Each decision tree is constructed independently of the other as follows:

- For the beginning of the training sample generates a random subsample with repetitions. This procedure called bagging (bootstrap aggregating or bagging);
- Constructed decision tree for classification under this sub-sample, wherein the L signs used a limited subspace signs $l = \sqrt{L}$;
- Construction of decision tree extends to the complete exhaustion of the subsample. The procedure pruning branches (pruning) are not carried out;
- Classification of objects is realized by the majority sampling: every tree set of attributes classified object to one of the classes, and wins the class for which chose the largest number of trees;
- Optimal number of trees selected in such a way as to minimize the error of the classifier on the test sample. In case of default, minimized error evaluation indicators that were classified incorrectly and, as a result, were not included in the sample (out-of-bag).

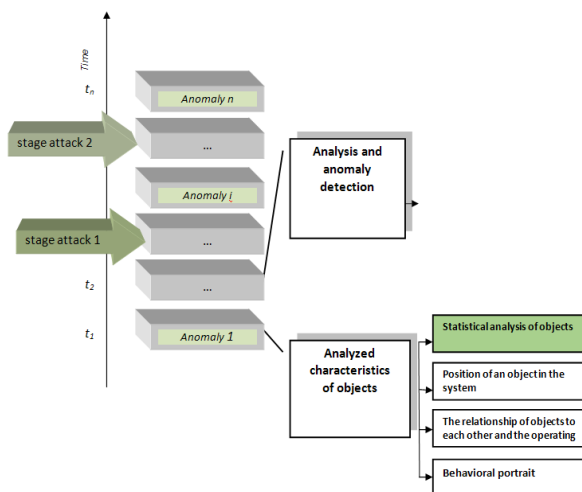


FIGURE 2 Scheme of work of multi-stage recognition adaptive system of threats, anomalies and cyber attacks

The main disadvantage of this method has high computational complexity of $O(NP)$, where P is the number of trees. As a result of this method can be used for

vulnerability detection complex, but not for the problem of intrusion detection in real time.

Instead of decision trees, as a base model, we can use a variety of non-linear models, such as polynomial logistic regression [20] or multivariate adaptive regression splines (MARS).

The statistics of normal (or abnormal) activity are displayed in a sequence of vectors of this space. The task of MARS method is to construct the best approximation of conduct on the statistics in the given form of a training set of vectors, in this case as approximating functions used multivariate adaptive regression splines. Construction of MARS model comes in two approaches: forward and reverse. During the forward stroke criterion for adding vertices in the model the next step is optimal. Vertices to be added until the model reach the maximum level of complexity. In reverse stroke the course irrelevant peaks are removed from the model, leading to its simplification. Built spline is a "template" of attack or normal behavior of systems.

Suppose that a sample is set $\{x_i; y_i\}, i = \overline{1, N}$, while the dependence between y_i and x_i , may be represented as

$$y_i = f(x_i) + \varepsilon, \tag{1}$$

where $f(x)$ - unknown function, ε - approximation error.

MARS algorithm approximates the predicted value of the activity \tilde{f} in the form of an expansion in a row of basis functions

$$\tilde{f} = \alpha_0 + \sum_{k=1}^K \alpha_k F_k(x), \tag{2}$$

where α_0 - shift model; K - the number of basis functions; F_k and α_k is k -th basis function and its coefficient [20].

Let $\delta(y)$ - step function, defines a positive argument

$$\delta(y) = \begin{cases} 1, & \text{if } y \geq 0; \\ 0, & \text{if } y < 0. \end{cases} \tag{3}$$

In the one-dimensional case, the basis functions are selected piecewise-linear function of the form $\delta(\pm(x-z))_+^r$, where z - node coordinates; $r \geq 0$ - the degree of the spline.

The simplest basis functions of MAR-spline order $r = 1$ called reflected pairs. Often, these functions are in the form of

$$(x-z)_+ = \begin{cases} x-z, & \text{if } x \geq z, \\ 0, & \text{if } x < z; \end{cases} \tag{4}$$

$$(z-x)_+ = \begin{cases} z-x, & \text{if } x \leq z, \\ 0, & \text{if } x > z. \end{cases} \tag{5}$$

Example of reflexive pair is shown in Figure 3.

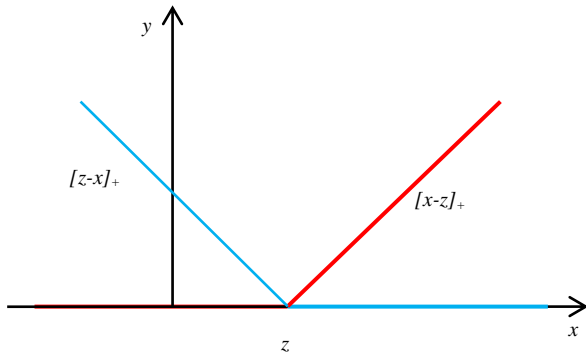


FIGURE 3 Schedule of basis functions of MAR-splines

In the multidimensional case the independent variable is a vector $X = (x_1, x_2, \dots, x_i, \dots, x_s)$. For each values x_i constructed reflective pairs with a node point $z = x_{i,j}$, $i = \overline{1, N}$, $j = \overline{1, s}$. According to the values of data can build a class of basic functions $\Psi = \left\{ (x_j - z)_+^r; (z - x_j)_+^r \right\}$, $z \in \{x_{1,j}; x_{2,j}; \dots; x_{N,j}\}_{j=1}^s$.

As a result, the basis function F_k defined by an equation of the form

$$F_k(x) = \prod_{l=1}^{N_k} \delta \left[\pm(x_{l,k} - z_{l,k}) \right]_+^r, \quad (6)$$

where N_k - number of functions in the class Ψ , which are part of k -th basis function, $x_{l,k}$ - vector coordinate X , which is part of l -th linear function of k -th basic functions, $z_{l,k}$ - the node that corresponds to $x_{l,k}$.

To construct the basis function $F_k(x)$ can be used not only function in the class Ψ , but also functions that are derived from them. To find the coefficients α_k can be used methods of discrepancy minimization, for example, discrete least squares method.

At the forward stroke MARS algorithm is finished. Next, you need to simplify this model by reducing the number of nodes. To do this, in every step of reverse stroke the node is removed, which the lack of that gives the smallest increase in the amount of residual squares.

The next stage of MARS algorithm has the choice of the optimal number of nodes that remain in the model of reverse stroke. To evaluate the value of the quantity K use a method of generalized cross-validation [5].

Let \tilde{f}_k - the function of optimal approximation values y_i .

We introduce the matrix F of dimension $K \times N$ such that $F_{ij} = F_i(x_j)$.

$$T(K) = \text{tr}(F(F^T F)^{-1} F^T) + 1, \quad (7)$$

where $\text{tr}(A)$ is the trace function of the matrix A .

$T(K)$ - the number of parameters, which is necessary to

determine, for its decision we can use the following equation

$$T(K) = q + aY, \quad (8)$$

where q - the number of basic linearly independent functions in the model, Y - the number of nodes that have been added to the model with the forward stroke, a - the parameter that shows estimates of optimization for each basis function. a parameter value are selected equally for two in adaptive model and $a = 3$ for non-adaptive model.

Next, calculate the coefficient GCV , which is proportional to the mean square of residual.

$$GCV(K) = \frac{1}{N} \sum_{i=1}^N \frac{(y_i - \tilde{f}_k(x_i))^2}{\left(1 - \frac{T(K)}{N}\right)^2}. \quad (9)$$

Minimizing coefficient value GCV we find the optimal number of model nodes

$$K^* = \arg \min_K GCV. \quad (10)$$

Thus, in the first phase of the functioning ASRCA using the methodology of MARS allows to build the best approximation conduct on the given statistics of recognizing threats, anomalies, or cyber attacks in the form of a training set of vectors.

5 Using logical procedures in the adaptive system of cyber defense

The use of adaptive regression splines on the first stage of ASRCA allows to formulate a finite set of objects $\{s_{a1}, \dots, s_{am}\}$, which the analyst or system know to what classes of anomalies, attacks or threats they belong (it is precedents, i.e. the objects used for training – ITO). The next task of ASRCA is to identify a particular class of anomalies, threat or an object from a given set of features in the ITO values $\{s_{ax1}, \dots, s_{axn}\}$.

The use of the logical procedures in the second stage of ASRCA, makes it possible to obtain reliable results for situation, when there is no aprioristic information about the distribution function of the available values of threat, cyber attack or anomalies.

When using logical procedures of cyber attacks recognition (LPCAR), we will consider informative fragments which are found in the description of objects in one class of cyber-attacks, but absent in the descriptions of other classes.

In constructing LPCAR used so-called elementary classifiers (EC) [18, 28, 29]. EC is a fragment of a brief describing the object, and is used for training ASRCA. For these facilities (cyber threats, anomalies, vulnerabilities, etc.) (CT_1, \dots, CT_l) constructed the set of EC with predetermined properties.

Algorithms for synthesis of efficient implementations for LPCAR directly dependent on the success of the metric (quantitative) properties research of the set of informative fragments, i.e., signs of cyber attacks (cyber threats,

anomalies, vulnerabilities). It is necessary to turn the unclassified training matrix (UTM) as classified and in the learning mode to construct a clear partition space of recognition features to recognition of classes $CT_m^0 | m = \overline{1, M}$, where M - the power of alphabet classes.

Technically difficult implemented in ASRCA are the following problems. Determination of the asymptotic estimates of the number of deadlock coverings for integer matrix, containing signs of object recognition. Determination of the asymptotic evaluation of the permissible and maximum values of conjunctions of Boolean function, which can be applied to the synthesis circuit solutions of hardware ASRCA for the CCS.

In the article we consider the problem of constructing LPCAR based on the principle of "nonoccurrence" sets of permissible values of cyber attack signs (cyber threats, anomalies, vulnerabilities).

Let: RA - the number of possible targets attacking from the side of CCS; Q - total number of cyber threats to CCS; $\{s_{ax1}, \dots, s_{axn}\}$ - the set of object attributes, such as threats, anomalies, cyber attacks, (signs for convenience represented in binary form); (CT_1, \dots, CT_l) - an integration of disjoint subsets (classes) of cyberthreats to CCS; B_s - the set of numbers of cyber threats, implemented by the attacker to reach of p_a -th goal of cyber attacks; NP_{s_a} - a valid set of discrete signs (threats, anomalies, cyber attacks etc.) $\{s_{a1}, \dots, s_{aQ}\}$ typed.

An algorithm for calculating estimates (ACE) of the importance of a sign for ASRCA is possible as follows. In GIA features system select a set of the type of subsets $NP_{s_a} = \{s_{aj1}, \dots, s_{ajQ}\}$, $r_{p_a} \leq Q$. Presume that selected subsets supporting for ACE. We will designate all their set - ΩQ .

We define the following additional parameters: po_{ss_a} - the significance of the attack target (object) ss_{ai} , $i=1, 2, \dots, PA$; $po_{NP_{s_a}}$ - the significance of an object of a basic set $NP_{s_a} \in \Omega Q$.

For each class of cyber attacks on CCS $CT \in \{CT_1, \dots, CT_l\}$, calculate the estimate $E(ss_a, CT)$ of the object ss_a class TT, which has the form:

$$E(ss_a, CT) = \frac{1}{|LW_{CT}|} \sum_{ss_{ai} \in CT} \sum_{NP_{s_a} \in \Omega Q} po_{ss_a} \cdot po_{NP_{s_a}} \cdot BN, \quad (11)$$

where $|LW_{CT}| = |CT \cap \{ss_{a1}, \dots, ss_{aQ}\}|$, BN - the proximity of objects ss'_a and ss''_a .

The object ss_{am} belongs to the class, possessing the highest ratings $E(ss_a, CT)$. If there is a set of similar classes, then the algorithm refuses the subsequent recognition. In order to increase the correctness of the algorithm is necessary to solve a system of inequalities:

$$\begin{aligned} E(ss_{a1}, CT_1) &> E(ss_{a1}, CT_2), \\ &\dots \\ E(ss_{aQ}, CT_l) &> E(ss_{aQ}, CT_{l+1}). \end{aligned} \quad (12)$$

To solve the system (12) is necessary to choose parameters $po_{ss_{ai}}$ $i = 1, 2, \dots, PA$, and $po_{NP_{s_a}}, NP_{s_a} \in \Omega Q$. In a situation where the system is inconsistent, it is necessary to find the most compatible subsystem for it. Then, from the decision of this subsystem determine the values of $po_{ss_{ai}}$ and $po_{NP_{s_a}}$.

An alternative way to increase a correctness of algorithm work is the way of a selection of reliable basic sets system for object recognition (anomalies, threats, vulnerabilities, and cyber attacks). For example, to choose a selection so that the condition for any ITO. In addition, for each GIA $ss''_a \in CT$, $E(ss''_a, CT) > 0$ inequality was carried out. This can be done as follows. Let $NP_{s_a} = \{s_{aj1}, \dots, s_{ajQ}\}$ - support set. The set of attributes NP_{s_a} will be considered to satisfy the test requirements, if each GIA ss'_a, ss''_a , and thus belonging to different classes, $BN(ss'_a, ss''_a, NP_{s_a}) = 0$ condition satisfied. Thus, our test – is a set (group) of signs on which only any two objects from different classes distinguish [29].

Denote as MC - plurality of EC, which were obtained from a set of attributes $\{s_{ax1}, \dots, s_{axn}\}$, i.e. $MC = (\sigma_{DOP}, NP_{s_a})$, where $NP_{s_a} \subseteq \{s_{ax1}, \dots, s_{axn}\}$, $\sigma_{DOP} = (\sigma_{DOP1}, \dots, \sigma_{DOPr})$, $\sigma_{DOPi} \in NP_{s_{aj}}$, at $i = 1, 2, \dots, r_{s_a}$.

Suppose that a Z series of measurements of controlled features in CCS is carried out, and received the matrix by a sign

$$S = \begin{pmatrix} s_{ax11} & s_{ax12} & \dots & s_{ax1i} & \dots & s_{ax1n} \\ s_{ax21} & s_{ax22} & \dots & s_{ax2i} & \dots & s_{ax2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ s_{axr1} & s_{axr2} & \dots & s_{axri} & \dots & s_{axrn} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ s_{axz1} & s_{axz2} & \dots & s_{axzi} & \dots & s_{axzn} \end{pmatrix}, \text{ for example,}$$

$$S = \begin{pmatrix} 0 & 1 & \dots & 1 & \dots & 1 \\ 1 & 0 & \dots & - & \dots & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ - & 1 & \dots & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & 1 & \dots & - & \dots & 0 \end{pmatrix}.$$

Thus, a set of checked objects, belonging to the class is given the binary signs $\{1001\dots-01\}$. A dash indicates the uncertainty of feature in GIA.

Each algorithm used for recognition in CCS cyber attacks, threats, anomalies or CCS vulnerabilities within the class, denoted - AL . Then, for each class will be considered a subset of $MC^{AL}(CT)$ from set MC .

Let $MC^{AL} = \bigcup_{j=1}^l MC^{AL}(CT_j)$. The object of analysis

sp_{an} is based on determining the value $BN(\sigma_{DOP}, ss_a, NP_{sa})$ of each element (σ_{DOP}, NP_{sa}) of the set $MC^{AL}(CT)$, $CT \in \{CT_1, \dots, CT_l\}$. In addition, for each element $MC^{AL}(CT)$ is performed calculation assessment, which determines ss_a accessory of CT class. Each algorithm AL, in turn, is characterized by set of EC $MC^{AL}(CT)$ and in the way of calculation of assessment $E(ss_a, CT)$.

The classifiers used in algorithms $\sigma_{DOP} = (\sigma_{DOP_1}, \dots, \sigma_{DOP_r})$, are formed informational signs from NP_{sa} . Wherein, each EC should have at least one of the properties listed below: 1) fragment groups (ss'_a, NP_{sa}) , where $ss'_a \in CT$, coincides with $\sigma_{DOP} = (\sigma_{DOP_1}, \dots, \sigma_{DOP_r})$; 2) only a part of the fragments (ss'_a, NP_{sa}) , where $ss'_a \in CT$ coincides with $\sigma_{DOP} = (\sigma_{DOP_1}, \dots, \sigma_{DOP_r})$; 3) fragments of the group (ss'_a, NP_{sa}) , where $ss'_a \in CT$, not match with $\sigma_{DOP} = (\sigma_{DOP_1}, \dots, \sigma_{DOP_r})$.

In models described in works [26, 29], the methodology of EC construction σ_{DOP_i} for a particular class of cyber attacks, threats, anomalies or CCS vulnerabilities, based on the synthesis of matrix covering σ_{DOP_i} , which is formed by GIA descriptions for CT. The use of such models [29] allow to reduce computing costs in work of algorithms in some ways, for example, when inequality is carried out $|CT| < |\overline{CT}|$ (in particular, when a large number of classes of cyber attacks, threats, vulnerabilities or anomalies of CCS - $(CT_1, \dots, CT_l) = (B_{s_{a1}}, \dots, B_{s_{al}})$).

We put the object in compliance EC - (σ_{DOP}, NP_{sa}) , where $\sigma_{DOP} = (\sigma_{DOP_1}, \dots, \sigma_{DOP_r})$, NP_{sa} - a set of signs with numbers $j_1, \dots, j_{r_{sa}}$ elementary conjunction $\mathfrak{R} = s_{axj_1}^{\sigma_{DOP_1}} \dots s_{axj_{r_{sa}}}^{\sigma_{DOP_{r_{sa}}}}$. If $ss_a = (\alpha_{s_{a1}}, \dots, \alpha_{s_{aQ}})$ - the object from a set of PA, therefore, $BN(\sigma_{DOP}, ss_a, NP_{sa}) = 1$ in only case when $(\alpha_{s_{a1}}, \dots, \alpha_{s_{aQ}}) \in NI_{\mathfrak{R}}$, where $NI_{\mathfrak{R}}$ - interval of the validity of elementary conjunction \mathfrak{R} .

During the creation of LPCAR, should be noted that the definition of the set of EC is reduced to finding the EC and permissible maximal conjunctions for the distinctive features of the object class (i.e., cyber threats, anomalies, cyber attacks, and so on). Moreover, this function is two-digit Boolean function that takes on different values of ITO from CT_l and $\overline{CT_l}$.

Then, an object recognition procedure $ss_a = (\alpha_{s_{a1}}, \dots, \alpha_{s_{aQ}})$, such as cyber attacks in SDCA is performed on the basis of the results of elementary

conjunctions \mathfrak{R} calculation. In the study, the results of which are described in works [29] proved that the most economical option is to use the algorithm for calculating the conjunctions to cover the corresponding object class (cyber threats, vulnerabilities or attack). Then, distinctive (characteristic) function of class CT_l - will be presented as a function of the algebraic logic (Boolean function) F_{KL} , which is equal to zero (0) on the informative descriptions of the object $ss_{an} = (\alpha_{s_{an1}}, \dots, \alpha_{s_{anQ}})$ from CT_l and equal to one (1) on the remaining sets of signs from E_{CT}^O . Where E_{CT}^O is a set of signs, having a length r_{s_a} . Then, cover the class will correspond permissible for $F_{\overline{CT}}$ conjunction. Maximum for the $F_{\overline{CT}}$ conjunction corresponds to the deadlock cover. Acceptable \mathfrak{R} in matrices of attributes objects define the particular object $ss_{an} = (\alpha_{s_{an1}}, \dots, \alpha_{s_{anQ}})$ belonging to the class CT_l , if the condition $(\alpha_{s_{a1}}, \dots, \alpha_{s_{aQ}}) \notin NI_{\mathfrak{R}}$ performed.

In this case, getting the abbreviated disjunctive normal form (ADNF) of function reduces to finding ADNF for F_{CT} , which takes the value 0 on the sets from $B_{\overline{CT}}$ and the value 1 on the remaining sets E_{CT}^O . After getting ADNF for $F_{\overline{CT}}$ conjunction \mathfrak{R} that do not possess $NI_{\mathfrak{R}} \cap A_{F_{CT}} \neq 0$ property must be removed from it.

For example, to get logic function ADNF can be achieved by converting the conjunctive function of the form $D_1 \wedge D_2 \wedge \dots \wedge D_u$, where $D_i = s_{ax1}^{\beta_{i1}} \vee s_{ax2}^{\beta_{i2}} \vee \dots \vee s_{axQ}^{\beta_{iQ}}$, $i = 1, 2, \dots, mu$ realizes the function F_{CT} , β_{iQ} - set of elements $B_{\overline{CT}}$.

Let: $s_{ax}^{\alpha} = \bigvee_{\beta_i \neq \alpha_i} s_{ax}^{\beta}$. Then conjunctive function takes the form $D_1^* \wedge D_2^* \wedge \dots \wedge D_u^*$, where $D_i^* = \bigvee_{t \neq \beta_{i1}} s_{ax1}^{\beta_{i1}} \vee \bigvee_{t \neq \beta_{i2}} s_{ax2}^{\beta_{i2}} \vee \dots \vee \bigvee_{t \neq \beta_{iQ}} s_{axQ}^{\beta_{iQ}}$, $i = 1, 2, \dots, u$.

During the recognition proximity of objects $ss'_a = (\alpha_{s'_{a1}}, \dots, \alpha_{s'_{aQ}})$ and $ss''_a = (\alpha_{s''_{a1}}, \dots, \alpha_{s''_{aQ}})$ from RA on a matrix of signs NP_{sa} was estimated by parameter

$$BN(ss'_a, ss''_a, NP_{sa}) = \begin{cases} 1, & \text{if } \alpha_{s'_{ti}} = \alpha_{s''_{ti}} \text{ at } ti = 1, 2, \dots, r_{s_a}, \\ 0 & \text{if else.} \end{cases} \quad (13)$$

Thus, obtaining LPCAR and a set of EC for the simulated class of objects (cyber threats, cyber attacks or anomalies) is as follows: 1) set the distinctive function; 2) find a DNF (or ADNF) that realizes this function; 3) find permissible (maximum) conjunction \mathfrak{R} , which defines the object belonging to the class.

To assess the effectiveness of the training algorithm ASRCA used informative criteria of functional efficiency (ICFE) index:

$$\bar{E}^* = (1/C) \cdot \sum_{c=1}^C \max_{\{w\}} E_c, \tag{14}$$

where E_c - ICFE value of ASRCA training for the implementation of the class of cyber threats, cyber attacks, or anomalies - CT_m^0 ; $\{w\}$ - a set of steps to study ASRCA.

Thus, as a result of research developed a method of intellectual detection of threat, anomaly and attack, the essence of which is to determine the conjunctions for coverings class of object recognition, and which differs from the existing use of adaptive regression splines on the first stage of the statistical analysis of anomalies in CCS, as well as the application of the second stage of discrete treatments using the apparatus of logic functions and matrix signs and elementary classifiers of object recognition, which will allow to create effective analytical, hardware and software solutions for adaptive systems cyber defense of CCS.

6 The simulation results

The results obtained in the simulations led to make a conclusion that objects belonging to different classes of anomalies, threats or cyber attacks is often difficult to separate from each other. Quite number of features (for some classes of cyber attacks to 50%) have a weight of information [29] is almost equal to 0. In the case of using a set of attributes for the formation of GIA is advisable to waive the requirement of its deadlock. This is done to increase the speed of algorithm work. For example, in case of increasing the number of features from 3 to 6, the average number of inspections on the object was from 150 to 800, respectively. The use of representative sets with length 3-5 in GIA matrices allows achieving maximum effectiveness of the algorithm recognition works for the majority of the known anomalies, cyber attacks and threats. In a situation if the features of object class (e.g., cyber attacks) arranged with decreasing informational content, there is a set of features with a large informational content for each object (I) [29], and then, the information content of the group

gradually decreased, Figure 4.

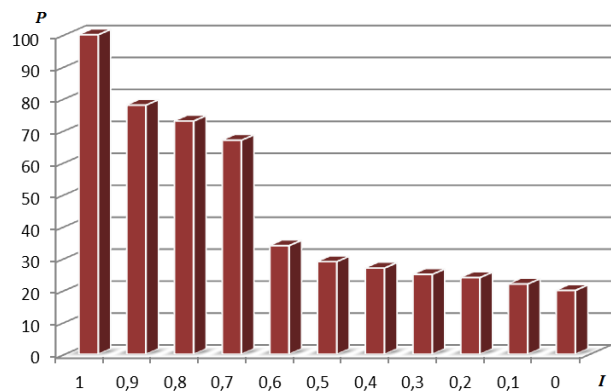


FIGURE 4 Visualize the importance of signs in a training matrix of GIA for network attacks

Thus interesting feature of the matrixes which forming GIA has been revealed, the informative content of control set for some classes of the attacks, for example, Dos/DDos, U2R, R2L or cyber attacks to the GPS system, several times greater than the weight of features forming this set. Thus the level of cyber attack recognition for which GIA training matrix were composed ranged from 25% to 30% for 2 features, 85-87% for 3-5 features, 92-98% for 6-9 features, Figure 5.

In this way GIA described by fragments of 3-4 features, belonging to different object classes is better characterized analyzed class than each of the features separately.

During the studies carried out comparison of the effectiveness of the proposed model by criteria - the average number of rules for training, Table 1.

The information about features of object recognition (cyber attacks) was adopted on the data from different sources (sensors) of software and hardware CCS. In particular, discussed the reports about the attacks generated by complex antiviral agents; analyzed log files, dumps of access memory AWP and PC, reports on the hard disk works, login logs into the system, database queries, etc. Some signs of attack are accepted on works [30, 31].

TABLE 1 The average number of rules, matrices and ASRCA training steps for the recognition of typical classes of cyber attacks in CCS

Object Class Recognition (Cyber attacks) (According to:: [1, 2, 15, 24, 28, 29])	Number of features (features and their informative content on works [29, 30, 31])	The average number of rules, matrices and steps for learning object (Rules / Matrix / learning steps)		
		Models and algorithms for sequential sorting features [10, 12, 16, 24]	Statistical Forecasting model [16, 18, 24]	A model based on the MARS, training samples and EC Class
Network attacks through the corporate system	11	200/30/2000	350/65/2000	60/10/2000
Attacks on standard software components of software CCS	19	350/50/3500	450/35/3500	30/15/1500
Network investigation	15	320/40/2500	120/30/2500	70/20/2000
Attacks aimed at the selection of passwords	12	230/15/1500	180/25/1300	25/20/1500
Attacks such as Man-in-the-Middle	9	300/40/4000	350/30/3000	40/20/2000
DoS/DDoS attacks	9	150/25/2500	170/25/2000	30/15/1500
Virus attacks	21	400/50/2700	400/60/2500	35/25/1700
Attacks on the ERP system through a HARD protocol	5	170/30/2700	210/50/2300	60/35/1900
Attacks on the LAN components	9	260/25/2400	200/40/2500	45/35/2000
Attacks SCADA systems	7	600/70/4000	800/60/3000	150/50/3500
Attacks on the HMI	3	500/50/3000	400/60/3000	70/30/2600
Node substitution attacks ("funnel attack")	15	150/35/1500	100/55/1500	30/15/1500
Compromise of knot of data collection	5	250/30/1700	190/35/1800	30/20/1300
Substitution of the router	11	300/40/2300	380/60/2500	35/20/1700
Removing the data from peripherals	15	150/25/1500	75/20/1400	45/10/1000
Attacks on the satellite navigation system	9	90/30/4000	150/50/4000	20/15/150

Figure 5 shows a histogram according to the maximum value of ICFE for the matrix dictionary of signs of anomalies and cyber attacks on the amount of learning algorithm steps ASRCA - $\{w\}$. Figure 6 shows the dependence of ICFE on the number of steps used for learning ASRCA.

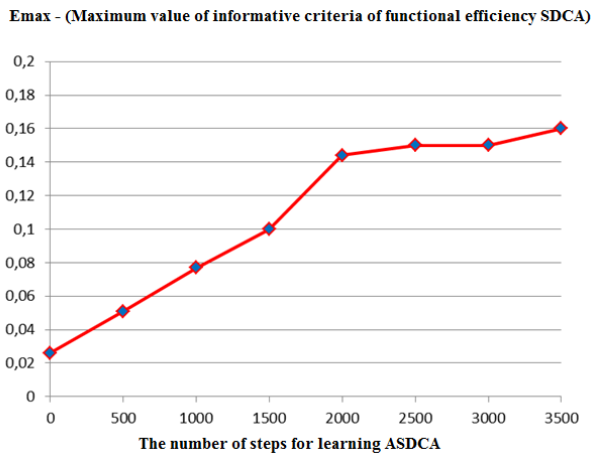


FIGURE 5 Dependence of the maximum value of ICFE for the matrix dictionary of signs of anomalies and cyber attacks on the amount of learning algorithm steps ASRCA

Analysis of the results shown in Figures 5 and 6, led to make a conclusion that quite effective in ASRCA is the use of algorithms with 4-10 features to train the system. In this case, ICFE reaches the maximum value, which gives grounds to speak about the possibility of constructing unmistakable

decision rules and signs of matrices to recognize threats, cyber attacks and anomalies within the class.

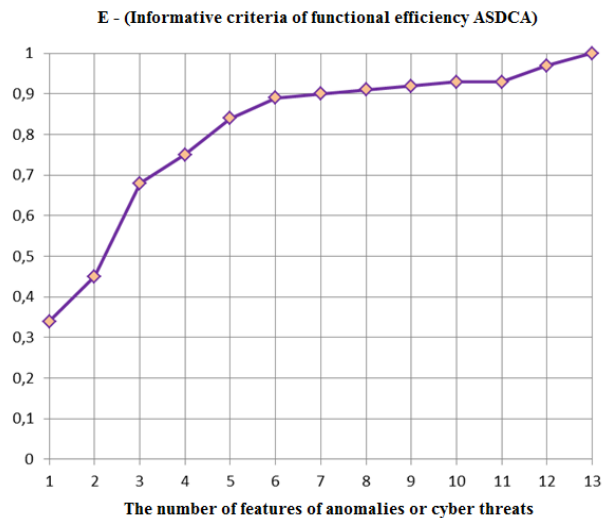


FIGURE 6 Schedule of the dependence of ICFE on the number of steps used for learning ASRCA

If in the algorithm recognition use representative sets of features of greater length ($s_{axi} > 5$), the efficiency of the algorithm provided the same. When using a representative set of features at the reduced length the effectiveness of the algorithm is decreased. To test the effectiveness of proposed model is performed a series of experiments for the major attacks shown in Table 1. Examples of test for attacks aimed at SCADA system shown in Figure 7.

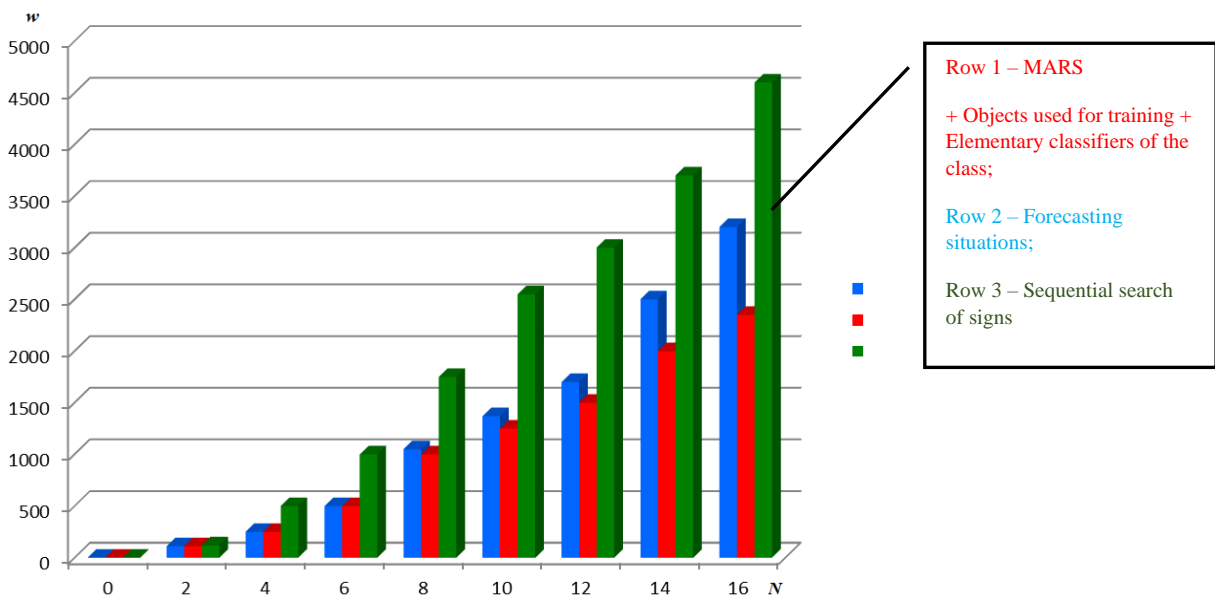


FIGURE 7 Comparative effectiveness of the proposed model for the recognition of attacks on SCADA system (N - number of signs; w - number of steps for learning ASRCA)

Compared to commonly used in ASRCA by sequential sorting features methods and statistical algorithms states, the use of models based on training samples and EC class allow to reduce the amount of required object recognition rules within the class of 2.5-10 times (depending on the class of objects - anomalies, cyber attacks, threats), and thus significantly reduce

the time recognition of anomalies, cyber attacks or threats. In the test training mode of ASRCA for the proposed model is a rational number of training steps of GIA made $w \approx 3000$ for known classes of objects, and $w \approx 3500...4500$ for more complex cyber attacks and anomalies.

7 Discussion of test results model and the prospects for further research

Difficulties of learning of ASRCA with the use of the device of adaptive regression splines, logic functions and elementary classifiers, exclusively associated with the stage of obtaining the disjunctive normal form (DNF) from maximal conjunctions of distinctive functions for each of the classes. However, the developed model compared with the results obtained for the models discussed in section 3, on the basis of finite automation [14-16], random sampling [4, 9], Bayesian networks, neural networks [21, 22], provide significantly a smaller number of relevant features for classifying threats, while reducing the time of ASRCA training.

At this stage of research, testing model made only for certain classes of anomalies, threats to information security and cyber-attacks. This is a definite disadvantage of the work.

Thus, the prospects for further research is to improve the knowledge base signs in the form of their matrix representation, as well as to conduct research model on a larger number of objects stored in the repository of ASRCA.

References

- [1] 2015 Cyber Attacks Statistics (2016). Available at: <http://www.hackmageddon.com/2016/01/11/2015-cyber-attacks-statistics/>
- [2] Cyber Attacks Statistics. Available at: Available at: https://securelist.ru/files/2015/12/KSB_2015_Stats_FINAL_RU.pdf
- [3] MITRE Research Program. Available at: <http://www.mitre.org>
- [4] Raiyn J 2014 A survey of Cyber Attack Detection Strategies *International Journal of Security and Its Applications* **8**(1) 247–56 doi: /10.14257/ijisia.2014.8.1.23
- [5] Jyothshna V, Prasad Rama V V 2011 A review of anomaly based intrusion detection systems *International Journal of Computer Applications* **28**(7) 26–35 DOI: 10.5120/3399-4730
- [6] Baddar S A-H, Merlo A, Migliardi M 2014 Anomaly detection in computer networks: a state-of-the-art review *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications* **5**(4) 29–64
- [7] Gyanchandani M, Rana J ., Yadav R N 2012 Taxonomy of anomaly based intrusion detection system: a review *International journal of scientific and research publications* **2**(12) 1–13 ISSN 2250-3153
- [8] Vinchurkar D P, Reshamwala A 2012 A review of intrusion detection system using neural network and machine learning technique *International Journal of Engineering Science and Innovative Technology (IJESIT)* **1**(2) 54–63 ISSN: 2319-5967
- [9] Tsai C-F, Hsub Y-F, Linc C-Y, Lin W-Y 2009 Intrusion detection by machine learning: a review *Expert Systems with Applications* **36**(10) 11994–2000 doi: 10.1016/j.eswa.2009.05.029
- [10] Omar S, Ngadi A, Jebur H H 2013 Machine learning techniques for anomaly detection: an overview *International journal of computer applications* **79**(2) 33–41 doi:10.5120/13715-1478
- [11] Riadi I, Istiyanto J E, Ashari A, Subanar 2013 Log analysis techniques using clustering in network forensics *International journal of computer science and information security* **10**(7) 1–7
- [12] Ranjan R, Sahoo G 2014 A new clustering approach for anomaly intrusion detection *International journal of data mining knowledge management process (IJDKP)* **4**(2) 29–38 DOI: 10.5121/ijdkp.2014.4.203
- [13] Guan Y, Ghorbani A A, Belacel N 2003 Y-means: a clustering method for intrusion detection *In canadian conference on electrical and computer engineering* **2** 1083–6 DOI: 10.1109/CCECE.2003.1226084
- [14] Li W, Yi P, Wu Y, Pan L, Li J 2014 A new intrusion detection system based on knn classification algorithm in wireless sensor network *Journal of electrical and computer engineering* **2014** DOI: 10.1155/2014/240217
- [15] Ilgun K, Kemmerer R A, Porras P A 1995 State transition analysis: a rule-based intrusion detection approach *IEEE transactions on software engineering* **21**(3) 181–99
- [16] Khan L, Awad M, Thuraisingham B 2007 A new intrusion detection system using support vector machines and hierarchical clustering *The international journal on very large data bases* **16**(4) 507–21 doi: 10.1007/s00778-006-0002-5
- [17] Wu S X, Banzhaf W 2010 The use of computational intelligence in intrusion detection systems: a review *Applied soft computing* **10**(1) 1–35 doi: 10.1016/j.asoc.2009.06.019
- [18] Kabiri P, Ghorbani A A 2005 Research on intrusion detection and response: a survey *International journal of network security* **1**(2) 84–102
- [19] Ameziane El Hassani A, Abou El Kalam A, Bouhoula A, Abassi R, Ait Ouahman A 2014 Integrity-OrBAC: a new model to preserve Critical Infrastructures integrity *International journal of information security* **14**(4) 367–85 doi: 10.1007/s10207-014-0254-9
- [20] Mukkamala S, Sung A H, Abraham A, Ramos V 2006 Intrusion detection systems using adaptive regression splines *Sixth international conference on enterprise information systems Part 3* 211–8 DOI:10.1007/1-4020-3675-2_25
- [21] Al-Jarrah O, Arafat A 2014 Network Intrusion Detection System using attack behavior classification *Information and communication systems (ICICS) 5th International Conference* 1–6 DOI: 10.1109/IACS.2014.6841978
- [22] Selim S, Hashem M, Nazmy T M 2010 Detection using multi-stage neural network *International journal of computer science and information security (IJCSIS)* **8**(4) 14–20
- [23] Pawar S N 2013 Intrusion detection in computer network using genetic algorithm approach: a survey *International journal of advances in engineering technology* **6**(2) 730–6
- [24] Zhou Y P 2009 Hybrid Model Based on Artificial Immune System and PCA Neural Networks for Intrusion Detection. *Asia-Pacific Conference on Information Processing* **1** 21-4 DOI: 10.1109/APCIP.2009.13
- [25] Komar M, Golovko V, Sachenko A, Bezobrazov S 2013 Development of neural network immune detectors for computer attacks recognition and classification *IEEE 7th Intern. Conf. on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS)* **2** 665–8 DOI: 10.1109/IDAACS.2013.6663008

8 Conclusions

During the research, the results which are presented in the article:

Developed recognition model of cyber attacks, anomalies and threats to critical computer systems, which is based on the use of cyber defense of adaptive regression splines in intelligent systems, training samples in the form of a matrix signs and elementary classifiers for each of the modeled classes;

Performed a study on minimizing the number of training samples of recognizing signs. It is shown that for the recognition of objects within the known class of cyber threats, attacks and anomalies, the use of training matrices of GIA representative sets with length 3-5 features allows to achieve maximum efficiency of the algorithm, reaching up to 98%. Compared to commonly used in ASRCA by sequential sorting features methods and statistical algorithms states, the use of models based on training samples and EC class allowed to reduce the amount of required object recognition rules within the class of 2.5-10 times.

[26] Zhan Z, Xu M, Xu S 2013 Characterizing honeypot-captured cyber attacks: statistical framework and case study *IEEE transactions on information forensics and security* **8**(11) 1775–89 DOI: 10.1109/TIFS.2013.2279800






[27] Bartosz Jasiul, Marcin Szpyrka, Joanna Śliwa 2014 Detection and modeling of cyber attacks with petri nets *Entropy* **16**(12) 6602-23 doi: 10.3390/e16126602

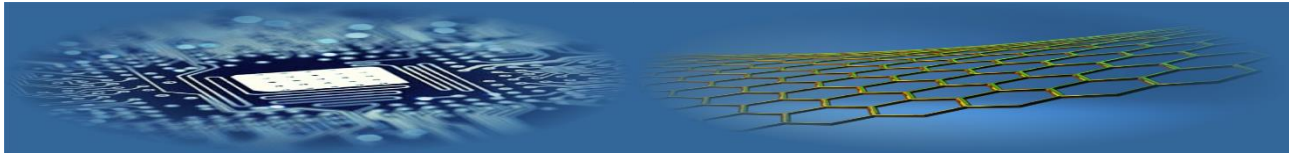
[28] Peddabachigari S, Abraham A, Grosan C, Thomas J 2007 Modeling intrusion detection system using hybrid intelligent systems *Journal of network and computer applications* **30**(1) 114–32 doi: 10.1016/j.jnca.2005.06.003

[29] Lakhno V 2016 Creation of the adaptive cyber threat detection system on the basis of fuzzy feature clustering *Eastern-European journal of enterprise technologies* **Vol. 2**, No 9(80): Information and controlling system 18–25 DOI: 10.15587/1729-4061.2016.66015

[30] Rid T, Buchanana B 2015 Attributing cyber attacks *Journal of strategic studies* **38**(1–2) 4–37 DOI: 10.1080/01402390.2014.977382

[31] Guitton C, Korzak E 2013 The sophistication criterion for attribution *The RUSI journal* **158**(4) 62–8 DOI: 10.1080/03071847.2013.826509

AUTHORS	
	<p>Beketova Gulzhanat</p> <p>Current position, grades: PhD student of Kazakh National Technical University. (Kazakhstan) University studies: Master degree in Computer Science Scientific interests: Information security</p>
	<p>Akhmetov Bakhytzhan</p> <p>Current position, grades: professor of Kazakh National Research Technical University after K.I.Satpayev, Doctor of Technical Sciences. (Kazakhstan)</p>
	<p>Oleksandr Korchenko</p> <p>Current position, grades: professor of National Aviation University, Doctor of Technical Sciences (Ukraine)</p>
	<p>Lakhno Valery</p> <p>Current position, grades: associate professor of European University, Doctor of Technical Sciences (Ukraine)</p>
	<p>Tereshuk Anna</p> <p>Current position, grades: Senior lecturer of Department of Information Systems and Mathematical Disciplines, European University(Ukraine)</p>



Metrics for consistency checking in object oriented model transformations

G Ramesh^{1*}, T V Rajini Kanth², A Ananda Rao¹

¹CSE Department, JNT University Anantapur Ananthapuramu, Andhra Pradesh, India

²CSE Department, Sreenidhi Institute of Science and Technology, Ghatkesar Hyderabad, Telangana, India

*Corresponding author's e-mail: ramesh680@gmail.com

Received 29 March 2017, www.cmnt.lv

Abstract

Model transformation is the cornerstone of Model-Driven Engineering (MDE) as it is crucial in Computer Aided Software Engineering (CASE) towards Object Oriented Analysis and Design (OOAD) and Object Oriented Programming (OOP). It also plays vital role in entity relationship model. Therefore it is indispensable to be treated as traditional software artefacts and assess quality of model transformations. Model-to-model transformations are from Platform Independent Model (PIM I) to Platform Independent Model (PIM II) and from PIM to Platform Specific Model (PSM). The goal of our research in this paper is to make these model transformations measurable. However, it is confined to proposing a set of metrics pertaining to consistency checking. The quality of transformations is measured in terms of consistency. The metrics proposed in this paper are general and can be reused. We evaluate the metrics using our framework named Extensible Real Time Software Design Inconsistency Checker (XRTSDIC) which supports end-to-end transformations of object oriented models. Our empirical study revealed that the proposed metrics add value to our model consistency checker as they quality in model transformations.

Keywords:

Model Driven Engineering (MDE), XRTSDIC, model transformations, consistency checking, quality measures

1 Introduction

Model Driven Approach (MDA) is an important alternative for developing information systems. The underlying principle of this approach is defining abstract models that can be used for implementations. Unified Modelling Language (UML) is widely used to model information systems that are built in object oriented approach. As part of Model Driven Engineering (MDE) design and exploitation of domain models became important in software development. The conceptual models can help understand development process quickly besides ensuring that the productivity is more with Computer Aided Software Engineering (CASE) tools. Model transformations can be part of CASE tool. This research is our ongoing work on consistency checking in model transformations. This paper focuses on deriving metrics for checking consistency of model transformations.

Our prior works [1-5] provide a series of related research efforts in realizing a framework that supports end-to-end approach for model transformations besides detecting and tracking software design inconsistencies. In [1] we defined a framework named Extensible Real Time Software Design Inconsistency Checker (XRTSDIC) which checks model inconsistencies and provide feedback dynamically. The framework is flexible and extensible. It has placeholders for future methods besides having personalized configuration and execution models. In [2] explores the realization of the framework proposed in [1] with consistency rules, provision for tolerance of inconsistencies to support notion of "living with inconsistencies" in the form of a prototype application.

In [3] we improved the framework with rule detector algorithm, consistency checker algorithm, and visualization algorithm. In [4] our framework is further enhanced and evaluated with end-to-end model transformations from Platform Independent Model (PIM) to Platform Specific Model (PSM) often with intermediate PIMs. It focused on class diagram transformation rules, Entity Relationship Diagram (ERD) transformation rules, handling issues with class relationships, and case study to evaluate the work. In [5] the framework is evaluated with UML class diagram to source code of different object oriented languages.

Our contributions in this paper include derivation of metrics for consistency checking of object oriented models and integrating with our framework XRTSDIC to leverage its utility further. A case study is provided to evaluate the framework with the metrics derived. The remainder of the paper is structured as follows. Section II provides review of literature. Section III presents the proposed system in detail. Section IV presents quality attributes. Section V presents the proposed metrics. Section VI shows Evolution methodology and experimental results while section VII concludes the paper.

2 Related works

This sections reviews related works. The reviewed content is categorized into model transformations and metrics used to measure quality of model transformations.

2.1 MODEL TRANSFORMATIONS

Kuzniarz *et al.* (2003) [16] focused on consistency issues in

UML-based software design models. They proposed consistency rules for different transformation models. It has mechanisms for finding inconsistencies in the design models made of UML. Hutchison *et al.* (2009) [8] focused on model-driven software engineering for self-adaptive systems. Paredis *et al.* (2010) [6] model transformations between two languages that are complement to each other. They are known as Modelica and SysML from OMG. SysML is a generalized modelling language while Modelica for analyzing systems with discrete time dynamics. The transformation between them is bi-directional. In [7] Model Driven Interoperability is focused for achieving interoperability transformations in distributed environments.

Biehl *et al.* (2010) [19] made a good review of model transformations. They explored many transformation approaches such as graph-based, template-based, and hybrid approaches besides presenting model transformation languages such as EMF Henshin, ATL, Query/View/Transformation (QVT), SmartQVT, ETL, XSLT and ModelMorf. They opined that synthesis and integration are the two advantages of model transformations broadly. Kessentini *et al.* (2012) [9] focused on search-based model transformation with example. Model Transformation (MT) became very important activity in software engineering as it is supported by Computer Aided Software Engineering (CASE) tools. They proposed an approach that is independent of source and destination formalisms and works for any source model. Model Transformation By Example (MTBE) is the main focus of them. However, they explored different methods for transformation including model transformation based on search.

Rodriguez *et al.* (2010) [10] proposed a method for semi-formal transformation a business process into use case and class diagrams of UML by adapting MDA. They focused on security aspects in the modelling. Towards this end they defined transformation rules to transform business process into class and use case diagrams. Their semi-automated approach could obtain useful artefacts of information systems. Hermann *et al.* (2010) [12] employed triple graph grammars (TGG) for efficient model transformations. Bi-directional model transformations are possible with well known Triple Graph Grammars. Towards this end, they employed Negative Application Conditions (NAC) as well. NACs can improve model transformation specifications. Garcia *et al.* (2012) [18] introduced a semi-automatic process that takes care of model transformation co-evolution. It has two phases namely detection phase and co-evolution phase. The former takes care of detects changes to metamodel while the latter takes care of performing required actions to complete co-evolution process.

2.2 METRICS

Chidamber and Kemerer (1994) [13] focused on a suite of metrics that can be used for improving object oriented design (OOD). Hutchinson *et al.* [14] provided an approach for assessing MDE. Generally MDE promotes software development with advantages such as interoperability, maintainability, portability and productivity. The maturity of MDE is assessed with automation. The degree of code generation is from 65% to 100%. Amstel and Brand (n.d) [22] studied model transformations made using ATL. They assessed quality of transformations using metrics. They classified metrics into different categories. They are rule

metrics, helper metrics, dependency metrics and miscellaneous metrics. They concluded that metrics alone are not adequate to assess quality of model transformations. Moreover those metrics are to be associated with quality attributes so as to relate with quality model of transformations. The quality assessment provided by their metrics and manual assessment is compared to know the error rate in quality assessment of chosen model transformations.

Amstel *et al.* (2008) [25] studied possible measures for quality transformations. They proposed many consistency related metrics such as number of code clones, number of unused variables, number of different types per variable name, and different variable names per type. Kapova *et al.* (n.d) [23] explored code metrics on model-to-model transformations for evaluating maintainability. They used automated metrics such as transformation size metrics, relational metrics, consistency metric and inheritance metrics. Apart from these metrics, they employed manually gathered metrics such as similarity of relations, and number of relations that follow certain design pattern. The computation of metrics is made using QVT transformations and metrics support availability.

Vignaga (2009) [24] applied metrics to measure ATL model transformations. There are many quality metrics such as conciseness, consistency, completeness, modularity, reuse, reusability, modifiability and understandability. The unit metrics available with ATL include Number of Imported Libraries (NIL), Total Number of Imported Libraries (TIL), Number of Helpers (NH), Number of Helpers without Parameters (NHP), Balance of a Unit (BOU) and Number of Helpers per Context (NHC). Other metrics available are categorized into module metrics, library metrics, rule metrics, matched rule metrics, lazy matched metrics, called matched rule metrics, and helper metrics.

Testing model transformations is an important activity in MDA. Baudry *et al.* (2010) [11] identified barriers to systematic testing of model transformations. They considered model transformation example of converting hierarchical state machine to flattened state machine the hierarchical state machine has many incoming and outgoing transmission. The states are of many types namely simple states, initial states and final states. Apart from these, composite states are also available. The barriers identified for model transformations include heterogeneity of transformation languages, lack of tools for model management, and complexity of inputs and outputs.

Kessentini *et al.* (2011) [20] focused on model transformation testing using two steps known as selection of test cases and finding test oracle functions. Their approach also focuses on finding the risk of detected faulty candidates and sorts them in the order of risk. They used immune system metaphor of biological science in order to achieve this. They defined precision and recall measures to evaluate the transformations. Pean (2012) focused on change metrics to measure incrementally built model transformations. They defined language feature metrics and change metrics based on abstract syntax difference model.

Arendt and Taentzer (2013) [21] used Eclipse Modelling Framework and explored it for quality assurance. They employed 6C goals such as correctness, completeness, consistency, comprehensibility, confinement, and changeability. They explored project specific quality

assurance techniques. Therefore it is made possible to specify such techniques based on the need of the project. Their specification supports model smells detection using metrics and anti-patterns. They also employed Domain Specific Modelling Language (DSML) known as SimpleClassModel (SCM) for demonstrating quality assurance of models.

Chitra and Sherly (2016) [15] used graph based models for verification of software design models. The process of model verification is used for observing behaviour preservation. With verification it is possible to have refactoring. Here graph isomorphism is the property utilized for model verification. Rosenberg and Hyatt (n.d) discussed software quality metrics for systems built on object orientation. Then they evaluated metrics using certain criteria such as testability, maintainability, reusability, understandability, complexity and efficiency. The metrics covered by them include cyclomatic complexity, size, comment percentage, weighted methods per class, response for a class, lack of cohesion of methods, coupling between object classes, depth of inheritance tree, and number of children. Amstel *et al.* (n.d) [26] proposed metrics for assessing ASF+SDF model transformations. Their metrics are related to different quality attributes such as understandability, modularity, modifiability, reusability, completeness, and consistency.

3 Our framework: XRTSDIC

Our framework defined in [1] is known as Extensible Real Time Software Design Inconsistency Checker (XRTSDIC). It is presented in Figure 1. It gives an overview of the generic approach for model inconsistency checking with provision for personalized configuration and execution model. The framework allows modelling tool selection, consistency rule language selection and visualization approach selection. These are pertaining to personalization which does mean that the models drawn by users are associated with such users and their configurations are retained.

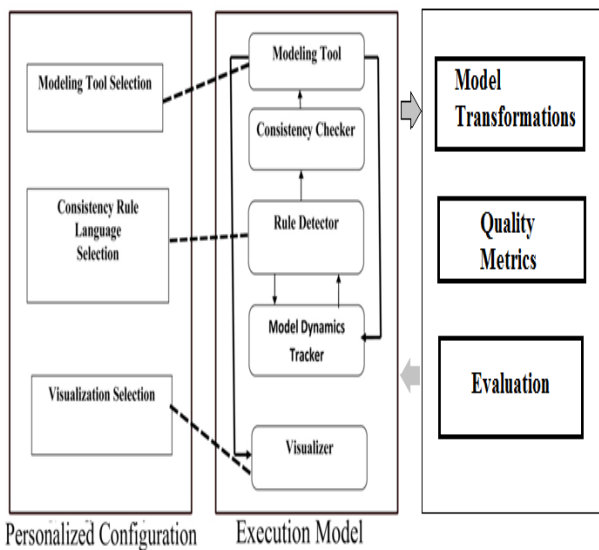


FIGURE 1 Overview of our Framework XRTSDIC

This framework was implemented in [1] and made further enhancements in [2-5]. In this paper we focused on improving

it further to facilitate measures for checking quality of model transformations. Quality attributes and consistency metrics are discussed in section 4 and 5. We considered only model consistency metrics that check the quality of model transformations. The proposed metrics are applied to a case study where model transformations are made from class diagram (PIM) to sequence diagram (PIM). And then the class diagram is transformed into source code (PSM). The metrics are useful to discover any inconsistencies in the way of model transformations from source to target. The source is reused number of times in model transformations as the same source is transformed into multiple targets.

The execution model of the framework helps developers to make use of a modelling tool to build models and then visualize any model inconsistencies. The tool also supports rectification of inconsistencies besides presenting them in chosen phenomenon. The execution model is based on the algorithm 1 presented here.

As shown in algorithm 1, the execution model pseudo code provides useful logic that helped in building the tool. The tool here is enhanced with proposed metrics presented in section 5. However, the consistency rules are taken from our previous work [3] where case study and evaluation of model inconsistencies are demonstrated. In this paper we focused on non only inconsistency checking but also measure quality of model transformations.

```

1 Initialize context vector C
2 Initialize rules vector R
3 Initialize model dynamics vector MD
4 Initialize inconsistencies vector INC
5 Do
6 IF MC is true THEN
7   Notify MDT
8 END IF
9 IF MDT has fresh notification THEN
10  MD = model dynamics
11  R = RuleDetector(MD)
12 END IF
13 IF r!=NULL THEN
14  INC = ConsistencyChecker(R, MD)
15  IF INC !=NULL THEN
16    Update C with INC data and metadata
17    Visualizer(C)
18  END IF
19 While(drawing==true)
    
```

Algorithm 1: Flow of Execution Model [3]

Class diagram to sequence diagram transformation rules

- Class Name → Instance of Class
- Consistency and transformation rule: If(new instance is created) then it should have a corresponding class in class diagram
- Class Method → Interaction in sequence diagram
- Consistency and transformation rule: The operation invoked by source should really exist in destination

Listing 1: Transformation Rules from Class Diagram to Sequence Diagram

Class diagram to source code transformation rules

Class diagram contains class name, attributes and methods. The class diagram is transformed to corresponding source code (classes) according to the object oriented language selected.

<p>Class Name → Class Name Consistency and transformation rule: if(a new class is created then) the class name should be unique and should be available in class diagram</p> <p>Class Attribute → Class Instance Variable Consistency and transformation rule: if(a new attribute is created then) the attribute name should be unique and should be available in the class attributes</p> <p>Class Attribute Type → Class Attribute Type Consistency and transformation rule: if(attribute type is determined then) the attribute type should match or compatible with that of class attribute</p> <p>Class Method → Class Method This method should match or compatible with that of class method.</p> <p>Class Method Arguments → Class Method Arguments The arguments in the generated classes should match arguments of method. However it is subject to the support in UML notation of class diagram.</p> <p>Class Method Return Type → Class Method Return Type The return type of method should have same or compatible type in generated class</p>

Listing 2: Consistency and transformation rules (Class Diagram → Source code)

These rules are applied when the transformation takes place. Again the generated source code is based on the functionality of corresponding dialect chosen. The dialect can provide accurate source code generation.

4 Quality attributes

With respect to model transformations, many quality attributes are identified. These quality attributes can be applied to many software artefacts. Particularly attributes that can be applied to model transformations are described here.

Understandability: This attribute refers to the amount of effort needed for user to understand model transformation. It also promotes reusability and modifiability. As understanding can help in modifications and reusability, it plays important role in model transformations. Model transformation is source→target model and its syntax and semantics are to be easy to understand.

Modifiability: Model transformations can be adapted to different context or altered to have additional functionalities. Changing requirements may force a model transformation to be modified. Another reason for the change is the language. When language needs to be changed, it warrants changes in model transformations. This attribute refers to the amount of effort required for alter model transformation in order to accommodate new requirements.

Reusability: It is the attribute that refers to the extent to which a model transformation or a part of it can be reused in other model transformations without making changes to the model being reused. Thus this attribute differs from modifiability attribute which causes modifications to model transformations. Especially, the reusability attribute comes into picture when a source is transformed to different target

and vice versa.

Reuse: It is somewhat related to reusability. However, it refers to the extent to which a model transformation is actually reused. It is best practice to reuse model transformations as much as possible instead of reinventing the wheel. Moreover MDE advocates reuse. Reuse in model transformations is common as source is common for many transformations. Therefore reuse can be considered as a measure which indicates how best a model can adhere to the principles of MDE.

Modularity: This attribute refers to the extent to which a given model transformation is built systematically. Systematic structure is essential to have modularity and every module in the model transformation should have its own purpose. Again modularity is pertaining to reusability. When functionality is repeated across modules, it is possible to reuse model or part of model transformations. Therefore the number of steps involved in the model transformation also can relate to modularity.

Completeness: This attribute refers to the extent to which model transformation is built fully. A model transformation is said to be complete when it has all parts of source model are completely transformed to target model according to specifications. In other words, the model transformation is made with all functionalities. An incomplete transformation results in target model which is not complete.

Consistency: This is the attribute which refers to extent to which a model transformation is without conflicts surfaced. According to Boehm [28] there are two kinds of consistencies. They are known as internal and external consistencies respectively. When uniform notation is maintained across the model transformation, it is known as internal consistency. It is often related to understandability. Internal inconsistencies can lead to target model in model transformations. External consistency refers to the extent to which model transformation adheres to given specifications.

Conciseness: This attribute refers to the extent to which model transformation has lack of superfluous information such as unused function parameters, code clones and so on.

5 Proposed metrics

This section provides metrics we have defined for measuring consistency and conciseness of model transformations. The main focus of these metrics is to measure consistency in model transformations in general. There are several measures possible. However, we like to define measures that are tool independent. Therefore the consistency measures defined by us are number of signatures with improper arguments (RSIA), number of unused variables (ROUV), number of code clones (NOCC), Population of Clone Class (POP), and Ratio of Non-Repeated Token Sequences (RNRS).

5.1 RATE OF SIGNATURES WITH IMPROPER ARGUMENTS (RSIA)

Model transformations from class to corresponding source code (PIM→PSM) of target language can exhibit inconsistencies. Every function modelled in the class diagram needs to be transformed into a function signature

with appropriate arguments. When it does not happen properly, it results into inconsistency. This kind of inconsistency is measured using NSIA.

$$RSIA = \text{FUN}_{\text{improper}}(A) / \text{FUN}_{\text{whole}}(A)$$

$\text{FUN}_{\text{improper}}(A)$ indicates the number of functions with improper signature and $\text{FUN}_{\text{whole}}(A)$ indicates all the functions that have been transformed. This measure is used to discover inconsistencies in model transformations in terms of number of functions containing improper signature. In other words it finds number of functions that are not consistency in terms of arguments.

5.2 RATE OF UNUSED VARIABLES (ROUV)

This measure is used to know the number of variables which are not used in the model transformations. The unused variables can affect conciseness quality.

$$\text{ROUV} = \text{VAR}_{\text{unused}} / \text{VAR}_{\text{all}}$$

Here $\text{VAR}_{\text{unused}}$ refers to the number of variables that are declared but not used in the application. VAR_{all} refers to all the variables declared in the application.

5.3 NUMBER OF CODE CLONES (NOCC)

Due to signatures with near similar arguments code repetition can occur in model transformations. Therefore this measure is relevant to know model inconsistencies.

$$\text{NOCC} = \text{COUNT}(\text{CC})$$

Where CC refers to code clones and the $\text{COUNT}(\text{CC})$ returns the number of code clones. Number of code clones or duplicate pairs of code is a good measure which may help to discovery model inconstancies.

5.4 POPULATION OF CLONE CLASS (POP)

The number of clone elements in a clone is measured using POP. A clone class is a class that may contain at least one clone pair. Clone pair is two pieces of code that are identical. The increase in POP reflects increase in clones in system.

$$\text{POP} = \text{Elements}_{\text{clone}} / \text{Elements}_{\text{all}}$$

Here $\text{Elements}_{\text{clone}}$ refers to the count of elements in the code clones while the $\text{Elements}_{\text{all}}$ refers to the number of elements.

5.5 RATIO OF NON-REPEATED TOKEN SEQUENCES (RNRS)

Ratio of non-repeated token sequences refers to the ratio of non-repeated token sequences in a given clone set. Higher rate of RNRS indicates the presence of more non-repeated token sequences in code clone. This metric is computed as follows.

$$\text{RNR}(S) = \frac{\sum_{i=1}^n \text{LOS}_{\text{non-repeated}}(c_i)}{\sum_{i=1}^n \text{LOS}_{\text{while}}(c_i)}$$

$\text{LOS}_{\text{non-repeated}}(c_i)$ refers to length of the non-repeated

token sequence of code clone c_i . In the same fasion, $\text{LOS}_{\text{while}}(c_i)$ refers to while token sequence of code clone c_i . LOS stands for Length of token Sequence.

6 Evaluation methodology

The tool implemented by us is Extensible Real Time Software Design Inconsistency Checker (XRTSDIC). It is used to perform model transformations and consistency checking. For evaluating metrics discussed in this paper along with the performance of the tool, we invited five industry experts who are aware of software engineering and model transformations well. They spent their valuable time on our request to provide ground truth for the case study described in this paper. The ground truth is evaluated with the system generated values with respect to metrics that are used to evaluate the quality of model transformations done by the tool. An average is computed for the independent values given by the experts. The average values are considered to be the ground truth and used in comparison.

6.1 CASE STUDY AND RESULTS

Our framework XRTSDIC with prototype application is used to have a case study which helps in model transformations with consistency checking. Besides it helps in using the metrics presented in this paper to know quality of model transformations. UML class diagram is transformed into corresponding ERD. This process is done by using transformation and consistency rules. There is intermediate result in the form of XML file that encapsulates classes in the PIM. Then the class diagram is transformed into source code. The case study class diagram considered is related to Hospital Management System (HMS). The class diagram is as shown in Figure 2.

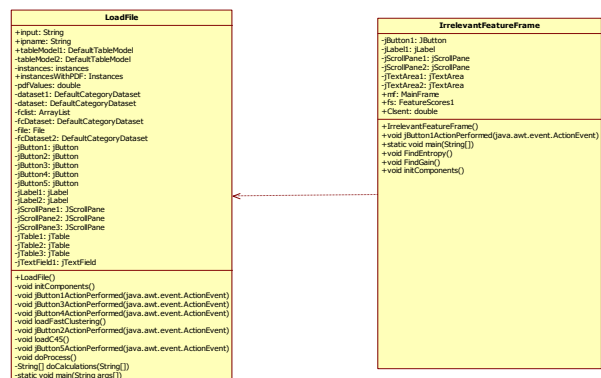


FIGURE 2 UML class diagram for FSC project case study (PIM I)

The class diagram is drawn using our framework. The model transformation is done with two experiments. In the first experiment, the class diagram is transformed into ERD. Then the class diagram is also transformed to source code using Java syntax and semantics. In either case, the model transformation rules and consistency rules are employed. The generated ERD is presented in Figure 3.

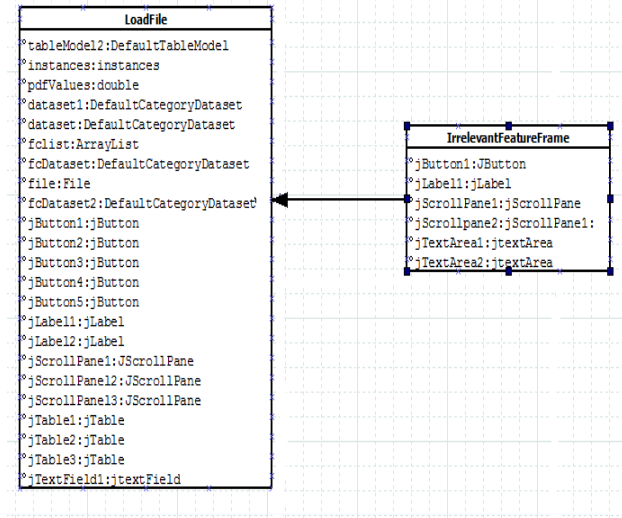


FIGURE 3 Transformed ERD (PIM II)

We made another empirical study on model transformations. The UML class diagram (PIM) is first of all transformed into another PIM model known as sequence diagram. Afterwards the class diagram is transformed into PSM known as source code of object oriented programming languages like C++, Java and C#. We proposed a Dialect hierarchy in Java language to handle transformation semantics for Java, C# and C++ [5]. The transformation dialect is a class that takes care of syntactical and semantic differences based on the target language chosen. The model transformation procedure and its flow are in our prior work [5]. In this paper our focus is more on checking quality of model transformations. Our enhanced framework XRTSDIC is used to apply metrics to measure quality of model transformations. Section 4 and 5 provided more details on quality attributes and proposed metrics for quality of consistency in model transformations. Following are the details of metrics applied to know the quality of transformations.

6.2 RESULTS AND DISCUSSION

We considered the case study pertaining to a data mining application named FSC (Feature Selection and Classification). Out of this project two important classes are considered for empirical study. LoadFile and IrrelevantFeatureFrame are the two classes presented in the class diagram shown in Figure 2. These two classes in the diagram are transformed into corresponding ERD as shown in Figure 3. This is achieved by generating some intermediate file in XML format. With the XML file, the model transformation is verified for correctness. Then the class diagram is transformed to source code using Java. This too was verified for consistency. Then the source code is implemented and subjected to metrics proposed in section 5.

6.3 RATE OF SIGNATURES WITH IMPROPER ARGUMENTS (NSIA)

This metric applied to the source code in Java that is LoadFile class. $FUN_{improper}(A)$ value obtained is 0 and the value for $FUN_{whole}(A)$ is 12. RSIA is computed as follows.

$$RSIA = 0/12 = 0$$

6.4 RATE OF UNUSED VARIABLES (ROUV)

This metric when applied to LoadFile class of PSM, the unused variables (VAR_{unused}) obtained is 0 and all variables in the class (VAR_{all}) is 27. The ROUV is finally computed as follows.

$$ROUV = 0/27 = 0$$

The result of ROUV metric is 0.24 which indicates rate of unused variables.

6.5 NUMBER OF CODE CLONES (NOCC)

This metric is applied to LoadFile class in the source code. The result obtained by the tool is 13. It is the count of code clones which is the functionality of our tool which detects clones and visualizes the same.

$$NOCC = 13$$

6.6 POPULATION OF CLONE CLASS (POP)

This metric when applied to LoadFile, the tool has returned values for two variables involved in the metric. Number of elements in clone $Elements_{clone}$ has got 13 while the total number of elements $Elements_{all}$ has got 313. The result of the metric is as given below.

$$POP = 13/313 = 0.041533$$

Here $Elements_{clone}$ refers to the count of elements in the code clones while the $Elements_{all}$ refers to the number of elements.

6.7 RATIO OF NON-REPEATED TOKEN SEQUENCES (RNRS)

This metric is applied to LoadFile class using our tool. The tool obtained the sum of length of the non-repeated token sequence of code clones and the sum of token sequence of code clones. The results are as shown below.

$$LOS_{non-repeated}(c_i) = 11$$

$$LOS_{while}(c_i) = 13$$

When these values are substituted into the metric, the result is as shown below.

$$RNR = 11/13 = 0.846153$$

High RNR value indicates ratio of non-repeated token sequences is more while lesser value indicates the repeated token sequences is more. According to the methodology described in section 6, the group truth is obtained from human experts and the results are presented in Table 1.

TABLE 1 Results of metrics compared with ground truth

Measures	Ground Truth	Tool Result
RSIA	1	0
NOCC	1	0
ROUV	1	0
POP	0.9	0.041533
RNR	1.2	0.846153

From Table 1, it is evident that the results of metrics computed by our tool and the results of metrics computed by human experts are presented.

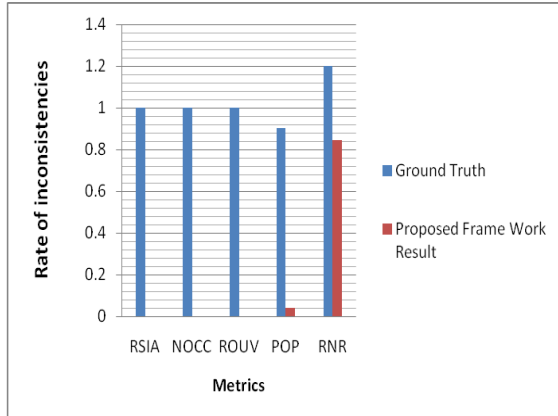


FIGURE 4 Results of quality metrics

The metrics are applied to the model transformations and the results are compared with the ground truth. The results are evaluated using metrics such as RSIA, NOCC, ROUV, POP, and RNR. A new metric is derived from the aforementioned metric. The details are as follows.

$$NewMetric = \frac{(P_1 + A_1) + (P_2 + A_2) + (P_3 + A_3) + \dots + (P_n + A_n)}{n}$$

$$A_n = P_n \times W_n \%$$

$$P_n = final\ value$$

$$W_n \% = weight\ percentage$$

$$A_n = weight\ value$$

RSIA

$$A_1 = P_1 \times W_1 \% = 1 \times (30/100) = 0.3$$

$$P_1 + A_1 = 1 + 0.3 = 1.3$$

ROUV

$$A_2 = P_2 \times W_2 \% = 1 \times (25/100) = 0.25$$

$$P_2 + A_2 = 1 + 0.25 = 1.25$$

POP

$$A_3 = P_3 \times W_3 \% = 0.9 \times (25/100) = 0.225$$

$$P_3 + A_3 = 0.9 + 0.225 = 1.125$$

RNR

$$A_4 = P_4 \times W_4 \% = 1.2 \times (20/100) = 0.24$$

$$P_4 + A_4 = 1.2 + 0.24 = 1.44$$

Substitute all these values in new metric equation
Then we get = (1.3+1.25+1.125+1.44)/4=1.27

TABLE 2 Result Comparison

Tool	Result of new metric
Solid SDD	0.82
ConQAT	0.91
XRTSDIC	1.27

As shown in Table 2, it is evident that the derived metric which provides overall performance of the framework in terms of finding quality of transformations is compared with other tools such as SolidSDD and ConQAT. The XRTSDIC shows better performance.

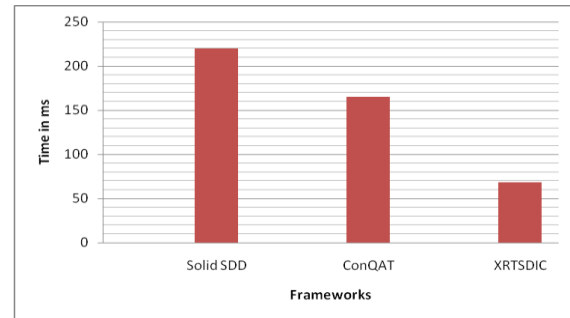


FIGURE 5 Performance comparison


From Figure 5, it is evident that the performance of XRTSDIC is better when compared with SolidSDD and ConQAT. SolidSDD and ConQAT are tools have inconsistency metrics including code clones. However, they do not have model transformation capabilities. XRTSDIC thus shows superior performance.

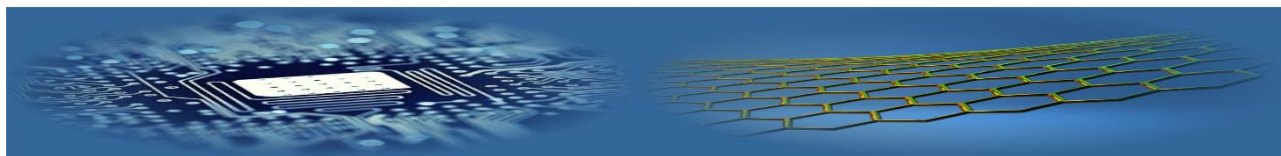
7 Conclusions and future work

This paper presented our research made on Model Driven Engineering (MDE) in terms of proposing metrics for measuring quality of model transformations. It is our ongoing research and the framework we built earlier [1] named Extensible Real Time Software Design Inconsistency Checker (XRTSDIC) which supports end-to-end transformations of object oriented models. In this paper we focused on defining consistency metrics meant for measuring quality of model transformations. The metrics are pertaining to model consistency as our research was focusing on this area. Model-to-model transformations are from Platform Independent Model (PIM I) to Platform Independent Model (PIM II) and from PIM to Platform Specific Model (PSM). The goal of our research in this paper is to make these model transformations measurable. Towards this end we proposed different metrics namely number of signatures with improper arguments (RSIA), number of unused variables (ROUV), number of code clones (NOCC), Population of Clone Class (POP), and Ratio of Non-Repeated Token Sequences (RNR). We enhanced our tool [1] to demonstrate the proof of concept of the application these metrics to know quality of model transformations. Our empirical study revealed that the proposed metrics add value to our model consistency checker as they quality in model transformations.

References

- [1] Ramesh G, Rajini Kanth T V, Ananda Rao A 2016 XRTSDIC: Towards a flexible and scalable framework for detecting and tracking software design inconsistencies *Proceedings of the first A.P. Science Congress*
- [2] Ramesh G, Rajini Kanth T V, Ananda Rao A 2016 Extensible real time software design inconsistency checker: a model driven approach *Proceedings of the International multi conference of engineers and computer scientists 2016 Vol I, IMECS Hong Kong*
- [3] Ananda Rao A, Rajini Kanth T V, Ramesh G 2016 A model driven framework for automatic detection and tracking inconsistencies *Journal of software* **11**(6) 538-53
- [4] Ramesh G, Rajini Kanth T V, Ananda Rao A 2016 An extended model driven framework for end-to-end consistent model transformation *Indian journal of computer science and engineering (IJCSSE)* **7**(4) 118-32 Aug-Sep 2016
- [5] Ramesh G, Rajini Kanth T V, Ananda Rao A 2016 XRTSDIC: Model Transformation from PIM to PSM *Communicated to IET Software*
- [6] Paredis C J J, Bernard Y, Burkhart R M, de Koning H-P, Friedenthal S, Fritzson P, Rouquette N F, Schamai W 2010 An overview of the sysml-modelica transformation specification *IEEE* 1-14
- [7] Bézin J, Soley R M, Vallecillo A 2010 Editorial to the proceedings of the first international workshop on model-driven interoperability *ACM* 1-112
- [8] Cheng B H C, Lemos R de, Giese H, Inverardi P, Magee J 2009 *Software engineering for self-adaptive systems* Springer 1-270
- [9] Kessentini M, Sahraoui H, Boukadoum M, Omar O B 2012 Search-based model transformation by example *Softw Syst Model* 209-26
- [10] Rodríguez A, García-Rodríguez de Guzmán I, Fernández-Medina E, Piattini M 2010 Semi-formal transformation of secure business processes into analysis class and use case models: An MDA approach *Information and Software Technology* **52** 945-71
- [11] Baudry B, Ghosh S, Fleurey F, France R, Traon Y L, Mottu J-M 2010 Barriers to systematic model transformation testing *IEEE* 1-12
- [12] Hermann F, Ehrig H, Golas U, Orejas F 2010 Efficient analysis and execution of correct and complete model transformations based on triple graph grammars *ACM* 1-10
- [13] Chidamber S R, Kemerer C F 1994 A metrics suite for object oriented design *IEEE transactions on software engineering* **20**(6) 1-18
- [14] Hutchinson J, Whittle J, Rouncefield M, Kristoffersen S 2011 Empirical assessment of MDE in industry *ACM* 1-10
- [15] Chitra M T, Sherly E 2016 Verification of behavior preservation in uml sequence diagrams using graph models *Indian journal of computer science and engineering* **7**(4) 1-6
- [16] Kuzniarz L, Huzar Z, Reggio G, Sourrouille J L, Staron M 2003 Workshop on Consistency Problems in UML-based Software Development II *IEEE* 1-89
- [17] Rosenberg L H, Hyatt L E 2010 Software quality metrics for object-oriented environments *IEEE* 1-6
- [18] García J, Diaz O, Azanza M 2013 *Model transformation co-evolution: a semi-automatic approach* Springer 144-53
- [19] Biehl M 2010 Literature study on model transformations *Royal institute of technology* 1-28
- [20] Kessentini M, Sahraoui H, Boukadoum M 2011 Example-based model-transformation testing *Autom Softw Eng*, 199-224
- [21] Arendt T, Taentzer G 2013 A tool environment for quality assurance based on the eclipse modeling framework *Autom Softw Eng* 141-84
- [22] van Amstel M F, van den Brand M G J 2010 Quality assessment of ATL model transformations using metrics *IEEE* 1-15
- [23] Kapova L, Goldschmidt T, Becker S, Henss J 2011 Evaluating maintainability with code metrics for model-to-model transformations *ACM* 1-16
- [24] Vignaga A 2009 Metrics for measuring ATL model transformations *IEEE* 1-15
- [25] van Amstel M F, Lange C F J, van den Brand M G J 2008 Metrics for analyzing the quality of model transformations *ACM* 1-11
- [26] van Amstel M F, Lange C F J, van den Brand M G J 2009 *Using metrics for assessing the quality of ASF+SDF model transformations* Springer 239-48
- [27] Paen E 2012 Measuring incrementally developed model transformations using change metrics *Queen's University* 1-125
- [28] Boehm B W, Brown J R, Kaspar H, Lipow M, Macleod G J, Merrit M J 1978 Characteristics of software quality North-Holland

AUTHORS	
	<p>G. Ramesh</p> <p>University studies: received B. Tech Degree in Information Technology from RGMCE, Nandyal, Kurnool Dist. Andhra Pradesh, M. Tech Degree in Software Engineering from JNTUA college of Engineering, Ananthapuramu, Andhra Pradesh, India, Perusing Ph. D at JNTUA, Anantapuramu, Andhra Pradesh, India</p> <p>Scientific interests: Software Engineering and Big Data</p> <p>Publications: several papers in various International Journals/ Conference</p>
	<p>Dr. T.V. Rajinikanth</p> <p>University studies: received M.Tech degree in Computer Science & Engineering from Osmania University Hyderabad, Andhra Pradesh, India and he received PhD degree from Osmania University Hyderabad, Andhra Pradesh, India. He is Professor of Computer Science & Engineering Department, SNIST, Hyderabad, Andhra Pradesh, India.</p> <p>Publications: more than 50 publications in various National and International Journals/Conferences. Organised and Program Chair 2 International Conferences, 2 grants received from UGC, AICTE. Editorial Board Member for several International Journals.</p> <p>Best Paper Award: "Design and Analysis of Novel Similarity Measure for Clustering and Classification Of High Dimensional Text Documents" in the Proceedings of 15th ACM-International Conference on Computer Systems and Technologies (CompSysTech-2014), pg: 1-8, 2014, Ruse, Bulgaria, Europe. His main research interest includes Image Processing, Data Mining, Machine Learning.</p>
	<p>Dr. Ananda Rao Akepogu</p> <p>University studies: received B.Tech degree in Computer Science & Engineering from University of Hyderabad, Andhra Pradesh, India and M.Tech degree in A.I & Robotics from University of Hyderabad, Andhra Pradesh, India. He received PhD degree from Indian Institute of Technology Madras, Chennai, India. He is Professor of Computer Science & Engineering Department and currently working as Director Academic and Planning, of JNTUA College of Engineering, Anantapur, Jawaharlal Nehru Technological University, Andhra Pradesh, India.</p> <p>Publications: more than 100 publications in various National and International Journals/Conferences.</p> <p>Best Research Paper award for the paper titled "An Approach to Test Case Design for Cost Effective Software Testing" in an International Conference on Software Engineering held at Hong Kong, 18-20 March 2009. Received Best Paper Award: "Design and Analysis of Novel Similarity Measure for Clustering and Classification Of High Dimensional Text Documents" in the Proceedings of 15th ACM-International Conference on Computer Systems and Technologies (CompSysTech-2014), pg:1-8,2014, Ruse, Bulgaria, Europe. Also received Best Educationist Award, Bharat Vidya Shiromani Award, Rashtriya Vidya Gaurav Gold Medal Award, Best Computer Teacher Award and Best Teacher Award from the Andhra Pradesh chief minister for the year 2014. His main research interest includes software engineering and data mining.</p>



Border node detection: a new experimental approach

Saher Manaseer^{1*}, Dua Alsoudi², Asmaa Aljawawdeh²

¹King Abdullah School for Information Technology, Computer Science Department, The University of Jordan, Amman, 11942, Queen Rania Street, Jordan

²The University of Jordan, Amman, 11942, Queen Rania Street, Jordan

*Corresponding author's e-mail: saher@ju.edu.jo

Received 13 April 2017, www.cmnt.lv

Abstract

This paper aims at sensing the network, and detects the border nodes, the researcher use NS2, in order to represent, simulate and calculate the delivery ratios of the distributed packets which accordingly will help to detect the border nodes. The importance of this research comes from detecting the border nodes without depending on other resources, since Ad hoc networks coordinates are virtual. The researchers analysed the results of the trace file that came as an output of carrying out simulations in Network simulator (NS2) for the evaluation of the ratios. The methodology of this experiment depends on using the IEEE 802.11 MAC protocol. Flooding technique was used to send data packets through three scenarios: First, 5% of the nodes are randomly chosen to send their data packets per minute. In the second and third scenarios, the percentages of nodes that flood their data are 25% and 50% respectively.

Keywords:

MANETs,
Broadcast,
NS2,
IEEE
MAC,
Flooding

802.11,

1 Introduction

The wireless network is a system of nodes (sensors, laptops, etc.) where each node collects and exchanges the information without any infrastructure (bridge and access point). Ad Hoc and Infrastructure are two classifications of the wireless network [1].

Ad Hoc networks are used in many fields of interests such as mobile and wireless Ad Hoc networks, wireless Local and Personal Area Networks, Quality-of-Service Issues, performance of protocols and many other issues [1].

Coordination of the messages flow through the network in Ad hoc networks does not depend on physical base station or routers, nodes of the network sends messages to each other. The original Latin word "Ad hoc" means "for this", leading to the purpose of this network "For this purpose" [1].

Ad hoc networks may contains many types of nodes, this paper aims at detecting the border nodes for if the node is a border node, it is not essential to broadcast all messages it receives; and this will enable saving energy used in messages transmission. Moreover, the border node is threatened to leave the network area at any time. Therefore, it has to be given higher priority when it has data to send. In addition, this unstable existence of the node is crucial for some network applications such as routing. Routes based on this node are considered weak or unstable.

2 Related works

Ad Hoc is a decentralized network consists of nodes that interface and exchange information between each other without any need for infrastructure. Each node has two main systems: the computing and communication systems. Mobile Ad hoc Network (MANET) beginning was in The

1990s, and introduced as a special case of Ad Hoc network. It has been considered as one of the biggest challenges since its beginning in wireless networks field. The main difference between Ad Hoc and MANET was the mobility of the nodes [14]. MANETs became popular research topic since the mid-1990s due to the high growth of laptops and 802.11/Wi-Fi wireless networking.

For achieving the main purpose of MANET network some obstacles need to be overtaken like the mobility that leads to dynamic topology which causes changing routes, losing packets, breaking the links [2]. Also, there are other challenges such that the limitation of the wireless transmission range, battery lifetime, bandwidth, the variable capacity links, the high cost of money, the low security and at the end the self-organization. Despite all these obstacles, MANET is used frequently in many significant fields like; smart cities, ambient intelligence, pervasiveness, monitoring, controlling, mobile social networking data dissemination, Road safety, traffic efficiency, infotainment, rescue, home network and much more [6, 7].

The nodes that know their exact locations are located at fixed points or have Global Positioning System (GPS) [16]. However, the use of GPS is not always available or facilitated in the case of MANETs. Also, for more accuracy some nodes depend also on two measurements; distance and angle measurements. Distance measurement has attracted many attentions since it relies on measuring the range wireless signal transmission by using many methods like RSSI, TOA and TDOA [15].

Link reliability, Noise and hard environment conditions may reduce the RSSI accuracy while the problem that faces TOA and TDOA is the non-rigidity of the graph. The second measurement is the angle which requires more cost due to the multiple receivers or the antenna array on nodes. Angle

of Arrival (AOA) is an example of angle measurement. Many other approaches are proposed like anchor based localization or centroid method which is also known as a reference node [5, 8, 15]

A homogenous network is the one with the nodes that have the same characteristics and communication capabilities, while heterogeneous network contains different types of nodes. Although heterogeneous networks are hard to deploy due to the different communication capabilities and resources, they are more realistic than the homogeneous networks [9]. Military fields are an obvious example of heterogeneous networks when helicopter, army vehicles, and ambulances communicate with each other. Border node or gateway node is a node that has at least one or more neighbours that belong to different networks. It is considered important to secure the network from any attack and help intra-cluster and inter-cluster routing. Service Border Nodes (SBNs) offer many services between the different MANETs like forward and store the service discovery information. Moreover, they can do the aggregation of service information and routing information about the MANETs networks. One of the main functions of the border node is to prevent or allow the nodes or agents to be accessed between the MANETs too [3, 9].

Figure 1 demonstrates a show case of the importance of border node or gateway node is obviously clear in the clustering algorithms. Clustering is a natural arrangement of nodes in different groups [12]. Within each cluster there is a cluster head that aggregates and collects the data from all node members then sends the information to the base station. However, the cluster heads are connected with each other directly or by using border nodes or gateway nodes. The cooperation between cluster heads and gateway nodes form a backbone that lead to provide a high scalability in the large networks, and prolong the network lifetime [12].

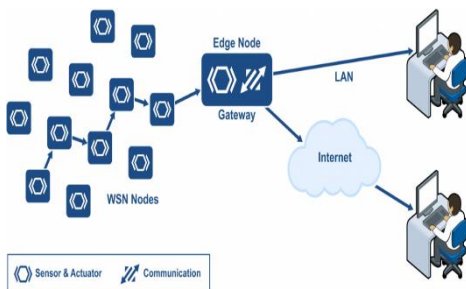


FIGURE 1 The use of border Nodes as Gateways

Border nodes can play a major role in Routing [1]. The Routing protocol is defined as a combination of roles for selecting the data packet path from a source to any destination across the network. In spite of determining the path in decentralized environment it is a bit difficult with the dynamic topology. Many approaches were proposed over the last decade. Two classifications were introduced, the topology based routing protocols and the position based routing protocols. Topology based routing protocols need to make the routing decisions by the information about the virtual links. While, the position based routing protocols need the physical locations of the nodes by using the location service. Position based routing protocols become more significant because there is no need for maintenance or establishing the route, which reduces the network

overhead in general. Thus, the only information that the sender node needs to know is the position of the destination node [15].

A location service is a software application that proposes knowledge about where nodes are located. Many routing protocols apply the location services with very restricted conditions. Most Forward within Distance (MFR) is a routing protocol that orders each node to broadcast beacons message to announce its position in a periodic way [10]. Besides that, many other routing protocols have their restricted conditions about the location services like DREAM routing protocol that is proposed by Basagni et al. [4]. DREAM uses the information from GPS systems to complete the communication process. Location-Aided Routing (LAR) is a position based routing protocols that is proposed by K. Young and V. Nitin [11]. As DREAM routing protocol GPS is used for obtaining the nodes geographical information. Using GPS has many advantages and disadvantages, reducing the network overhead and bandwidth utilization are obvious advantages. On the other hand, GPS service is not granted by all networks as a facility, in addition; this service consumes the battery life rapidly [1].

3 The proposed method

Text should be produced within the Scenarios modelled Network Simulator is a simulate network platform that plays a major role for testing, developing and evaluating any network type. Usually, it is highly used in academic studies, development and research. However, some important performance metrics are used like time, packet loss ratio, speed, delay, throughput, energy, and bandwidth. NS2 is an open source network simulator that uses many programming languages like, Tcl, Tk, C++, and Otcl. NS2 supports MANET and Ad Hoc network types too [13].

The Flooding is a technique that is used to send the same message to all over the network. To execute this technique there are some common steps that should be followed. First of all, the node should send the data packet to its neighbors, the second step, each node that receives the data packet has to forward it to its neighbors until the data packet reaches all the nodes in the network [15]. However, to avoid forwarding the data packet more than once, a sequence number is used. The Flooding technique has some advantages and disadvantages. Although, the simplicity and the high reliability of delivering the data are advantages, the high overhead is a very significant disadvantage. This technique is somehow useful when there is a rapid change in the topology with a small data packet.

The simulations model contains 100 nodes with a non-uniformly distributed over a 1000m x 1000m. The node movements are based on the random-waypoint model (RWP) [2]. The IEEE 802.11 MAC protocol uses a simulation time of 800 seconds and a pause time of 1 second is also applied in the RWP model. The node speed which is 10 meters per second is applied. The flooding technique is used. There are three scenarios. In the first scenario, in each minute 5% of the nodes have been chosen randomly for sending their data packets. In the second and third scenarios, the percentages of nodes that flood their data are 25% and 50% respectively.

TABLE 1 Simulation parameters

Value	Parameter
100	Total nodes
random-waypoint model	Movement model
10 m/sec	maximum speed
1 second	Pause time
IEEE 802.11	MAC protocol
1000m * 1000m	Simulation area

3.1 ALGORITHM

Python language is used to analyze the NS2 result file. The python code builds two matrices:

-The first matrix or ratio matrix calculates the ratio of the sent data packets (s) over the received ones (p); s / p . This equation is calculated for each single node over the whole 800 seconds.

I.e: the delivery ratio used in this methodology to detect the border nodes is the ratio of: (#packets received (receiver)/ number of packets transmitted (sender)).

The second matrix contains three steps.

A- The coordination as (x, y) is written for each single node over the whole 800 seconds.

B- If the x and y values for a specific node are under this condition;

($x < 25.0$ or $x > 975.0$) or ($y < 25.0$ or $y > 975.0$) it is considered as a Border Node (B). Otherwise, the node is not a border Node (NB).

C- The number of times that the nodes are defined as a border node over the total numbers: $(B) / (B + NB)$

4 Experimental results

After comparing the two matrices with each other.

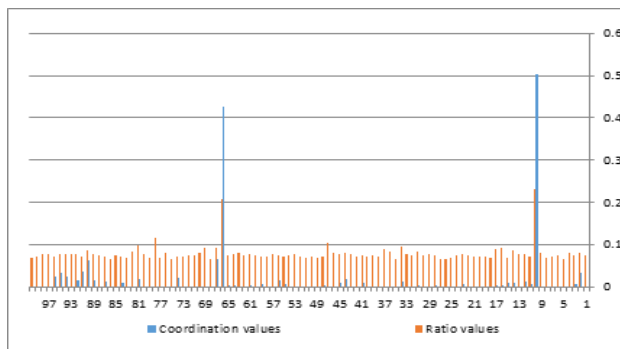


FIGURE 2 Ratio with 5% of the nodes number broadcast

The line chart illustrates two lines, the coordination and the ratio values. The x axis presents the nodes number. From the charts there is an obviously relation between the

References

[1] Manaseer S 2010 *Back off mechanisms for wireless mobile ad hoc networks* Doctoral dissertation, University of Glasgow
 [2] Manaseer S 2016 The choice of parameter values for simulation based experiments on mobile ad hoc networks *International Journal of Communications, Network and System Sciences* 9(04) 90
 [3] Al Amri H, Abolhasan M, Wysocki T 2010 Scalability of MANET routing protocols for heterogeneous and homogenous networks *Computers & Electrical Engineering* 36(4) 752-65
 [4] Bagni S, Chlamtac I 1998 A distance routing for mobility (DREAM) *International Conference on Mobile Computing and Networking*,

coordination value and the ratio value. Nodes number 10 and 66 have considerably a peak in their ratio values. As a result if the ration value is high this means that the node is a border node.

The first line chart shows the scenario with 5% of the nodes number that send data each minute. While the second and third line charts show the sent data from 25% and 50% of the total number of nodes.

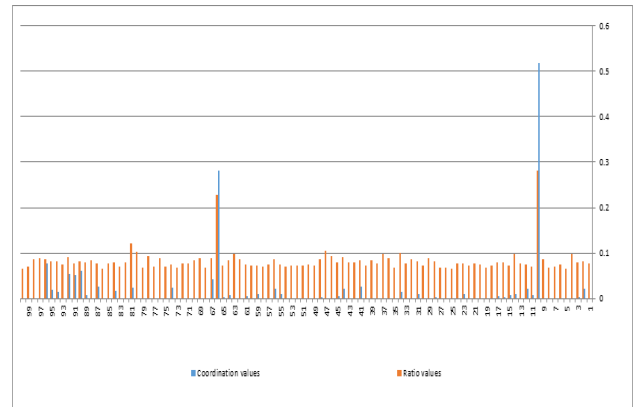


FIGURE 3 Ratio with 25% of the nodes number broadcast

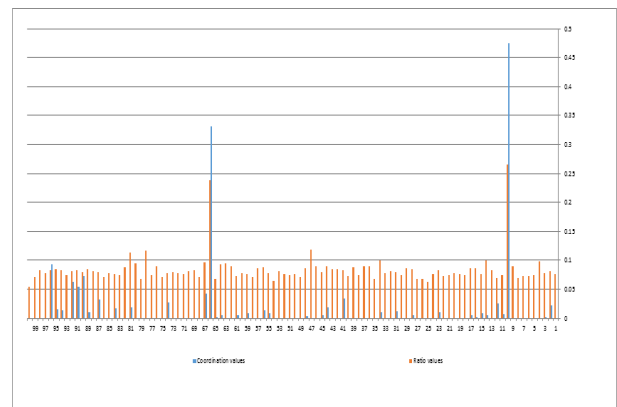





FIGURE 4 Ratio with 50% of the nodes number broadcast

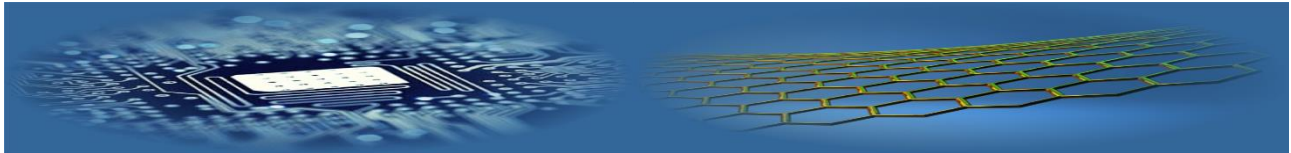
5 Conclusion

In this paper, the proposed algorithm applied a sender/receiver ration to detect border nodes. NS2 as a simulator platform is used to create MANET network and apply the ratio. The simulation results show a positive relationship between the coordination and the sender/receiver ration of the nodes. The border nodes without relying on any extra resources or even over heading the network by broadcasting messages can be detected.

Proceedings of the 4th annual ACM/ IEEE international conference on Mobile computing and networking 76-84
 [5] Bulusu N, Heidemann J, Estrin D 2000 GPS-less low-cost outdoor localization for very small devices *IEEE: Journal of Personal Communications* 7(5) 28-34
 [6] Chlamtac I, Conti M, Liu J J N 2003 Mobile ad hoc networking: imperatives and challenges *Ad hoc networks* 1(1) 13-64
 [7] Conti M, Giordano S 2014 Mobile ad hoc networking: milestones, challenges, and new research directions *Communications Magazine* 52(1) 85-96

- [8] Eren T, Goldeingberg O, Whiteley W, Yang Y R, Morse A S, Anderson B D O, Belhumeur P N 2004 Rigidity, computation, and randomization in network localization *IEEE NFOCOM conference* 2673-84, Hong kong
- [9] Harju J, Heijnen G, Langendörfer P, Siris V (Eds.) 2008 Wired/Wireless internet communications *6th International Conference, WWIC 2008 Tampere*, Finland, May 28-30, 2008 Proceedings (Vol. 5031). Springer
- [10] Kaur S, Gupta A K 2012 Position based routing in mobile ad-hoc networks: an overview *IJCST* 3(4)
- [11] Ko Y B, Vaidya N H 2000 Location-aided routing (LAR) in mobile ad hoc networks *Wireless networks* 6(4) 307-21
- [12] Pal S, Singh S P 2013 Mobility based clusterhead & gateway selection algorithm in MANET *International Journal of Engineering Research and Technology* 2(1) (January-2013) ESRSA Publications
- [13] Rao B N, Sri B R, Sumanjali K, Sai C, Raju A S R 2014 Performance analysis for routing protocols in MANETS by using NS2 (Network Simulator) *International Journal of Computer Science and Information Technologies* 5(1) 724-7
- [14] Selim B, Yeun C Y 2015 Key management for the MANET: A survey *Information and Communication Technology Research (ICTRC), International Conference on* (pp. 326-329) IEEE. (2015, May)
- [15] Stoica P, Sharman K C 1990 Maximum likelihood methods for direction of arrival estimation *Journal of Transactions on Acoustics, Speech, and Signal Processing IEEE* 38(7)
- [16] Wang J, Ghosh R K, Das S K 2010 A survey on sensor localization *Journal of Control Theory and Applications* 8(1) 2-11

AUTHORS	
	<p>Dr. Saher Manaseer</p> <p>Current position, grades: an assistant professor at the University of Jordan.</p> <p>University studies: PhD in Computer Science from the Department of Computing Science at the University of Glasgow.</p> <p>Scientific interests: Computer Networks and Embedded Systems. Currently, Dr. Manaseer is an active researcher in the field of Mobile Ad Hoc Networks.</p>
	<p>Dua Alsoudi</p> <p>Current position, grades: is Master student in Computer Science the University of Jordan.</p> <p>Scientific interests: Computer Networks. Currently, Miss Alsoudi is an active researcher in the field of WSN. More specifically, her research is focused on developing Routing protocols. She got his undergraduate degree in computer science at Al-Zaytoonah University in 2013</p>
	<p>Asmaa M. Aljawawdeh</p> <p>Current position, grades: a graduate student in computer science Department (Master Degree) at the University of Jordan, Amman, Jordan.</p> <p>University studies: Bachelor degree in computer engineering from Al-Hussein Bin Talal University (AHU), Jordan in 2010.</p> <p>Scientific interests: cloud computing, parallel computing, Bioinformatics, and image processing.</p>



Tri-Partite graph: a novel security scheme for cloud data

P Dileep Kumar Reddy^{1*}, C Shoba Bindu¹, R Praveen Sam²

¹CSE JNTUA College of Engineering Ananthapuramu

²CSE GPREC Kurnool

*Corresponding author's e-mail: dileepreddy503@gmail.com

Received 17 April 2017, www.cmnt.lv

Abstract

Cloud data security is the most concentrated feature of the cloud computing technology. Many cloud computing techniques like cloud data partitioning emerged reflecting new heights of providing data security by defining data priorities. The proposed method presents a novel scheme of maintaining owners prioritized data, while equally ensuring the security for whole portion of the data. The proposed method uses a tripartite graph for securely managing the prioritized data at various levels.

Keywords:

Encryption,
Authentication,
tripartite graph,
Hash,
MAC

1 Introduction

Today's most elegant data service model is the cloud computing technology that offers tenants with varied shared pool of resources. Wide spread with economical benefits cloud computing succeeded in attracting the major enterprises and totally emerged as widely recognized computing technology. The cloud is intended to provide various hardware and software services where in a user can access these services from anywhere in the world. The applications providing these services need a model of computation, a storage model and a model for communication. The model of computation ensures quality of service for better cloud management. Addressing the important aspect of quality of service is the cloud data security which has always emerged as a vital service by the cloud. To ensure data security the minimum policies expected from the cloud include: a strong encryption mechanism to safeguard the cloud data; prevention of unauthorized access to the cloud using strong authentication schemes ensuring cloud data integrity.

Present research is setting its trend with its innovative, more secured, proactive methods to ensure data security. More than 70% of the cloud providers use either SSL level encryptions or various symmetric key encryption methods for data privacy. Indexes are being geared up to make the search easy over encrypted data in the cloud. Raising the strength of security levels some of the approaches even encrypt these indexes.

A study showed many cloud applications generated the needed keys for data encryption. After data encryption MAC code is generated and appended to the encrypted data to be transmitted along. A thorough study of the prevailing literature showed more than 50% of the cloud service providers followed single mechanism on the whole data. But there may arise a case where the owner at a time is interested with a small portion of his whole data. Even then the cloud mechanism forwards his full encrypted data. This is degrading performance with increased communication cost

and increased application cost.

For efficient data management and accessing the prevailing technique followed by many cloud providers is cloud data partitioning. The owner's whole data is divided into small units and each unit is encrypted and signed so as to raise the data security. Works proved the data partition technique increase the storage efficiency of the cloud as well as error free data retrieval.

The whole of owner data is of not equally important to him. There may be cases where he frequently queries a common portion of data and some parts of the data may be less queried and sometimes not queried at all. This is even apt in cases where a single user data is prioritized as: 1) private- which is the owners sensitive data and in need of strong security mechanisms to protect it; 2) public-which can be accessed by any authorized user and where in case the strength of security mechanisms may be decreased. In the scenarios of clouds, with prioritized data categorizations it is better to have separate security policies running on top of each prioritized level.

The main drawback with cloud data partitioning with prioritization is owners trust. In data partitioning with prioritized data the owner has the provision to query his most prioritized partition. When such data is requested, only that data is securely being forwarded by the cloud. At this moment the owner may feel unsecured with other parts of the data. Though he has proved the data integrity of the queried part he may be more concern with other parts of the data. If there is a mechanism which constantly provides the data integrity of the parts that are not being queried by the owner then it would have raised the owners trust on partitioned data.

This paper presents a novel scheme for maintaining prioritized data of the cloud where a new data structure called the tripartite graph is taken advantage. Using the tripartite graph the strength of security is varied at different levels. The main contributions of the work are:

- A mechanism for easy maintenance of prioritized data.
- Usage of tripartite graph to increase the complexity

of data security.

- Provision to check the integrity of other un-queried data parts.

The organization of the paper is as follows: Section 2 presents the literature reviewed. Section 3 presents the basic preliminaries required to escalate the proposed method. Section 4 presents the proposed approach. Section 5 discusses the performance study and finally Section 6 concludes the paper.

2 Literature

The need for organizational growth has made the enterprise to outsource their storage and computing needs. The advent of cloud as a data storage has made the data owner to ponder on how secure the data within the cloud is. To address these security concerns the cloud provider assured security services like:

1. **Owners Data encryption:** So that the storage provider does not learn about owners important data.
2. **Integrity check:** Using which owner's data modification by the provider can be detected.
3. **Secure data sharing:** Authorized users can securely access the enterprise data from the cloud.
4. Efficient data retrieval techniques.

Works discussed in [1] showed the performance of various symmetric key data encryption schemes and showed AES has highest data security capabilities. But the drawback is the data owner faced problems to analyze their own data in the encrypted form. As data owners are naïve to encryption mechanisms they feel that large data encryption makes data loss if not correctly decrypted and so many owners prefer no encryption scheme while storing their data in the cloud. But unencrypted data in the cloud may be a threat. Cloud data security with homomorphism encryption by Adriana and Eran [2] boosted the standards of encryption schemes, where owner can by own analyze his encrypted data. Works presented in [3] discussed various issues of owner on security schemes like:

1. Is all data correctly encrypted.
2. How strong are the encryption algorithms?
3. How strong is the key maintenance mechanism?
4. Even if data is stored encrypted since I am a lame man I can think encrypted data can be decrypted by the cloud and can be changed. How can the cloud raise my trust to address this.

Data integrity is where unauthorized data modification to be identified. Works discussed in [4] showed how an appended MAC code can check data integrity. On data retrieval the owner tries to calculate the MAC and compares with appended MAC; if both are equal the data is not changed. Ensuring data integrity using a third party auditor (TPA) is discussed in [5]. The TPA who is on behalf of owner performed well in verifying the integrity of the dynamic cloud data. Data integrity verifiability using bilinear maps is discussed in [6].

Secure data sharing is where an authorized user can access the cloud data using some kind of user verifiability schemes. A key aggregate scheme for secure data sharing is discussed in [7]. The work discussed an approach of secret key sharing between various data sharing parties where in later stage these keys are used to identify the owner's accountability. Secure cloud data sharing using Tokens is

discussed in [8], wherein a user with appropriate token is only allowed to search the encrypted data of the cloud.

For efficient data retrieval from the cloud techniques like data partitioning are discussed in [9]. An index method is used on the partitioned data for efficient data retrieval. The efficiency of distributed hash indexing on data partitions is discussed in [10]. The other implemented technique for efficient data retrieval is prioritizing the data stored. Works discussed in [11] showed a classification of cloud data levels as public, private there by restricting the private data access with more strong security specifications.

3 Background and preliminaries

Cloud technology has envisioned as the present generation cream of hard research. A cloud is a storage area where massive data owners can securely store and can access their data whenever needed. Today cloud computing is more appealing with its high quality applications and services. Though cloud computing is excelling with its computing capabilities, it still faces some new born threats concern to data storage. This section discusses some threats as a part of background study and approaches to counter them. The section also discusses the preliminaries needed to move our proposed work.

3.1 PROBLEMS ON HUGE DATA TRANSFER

Huge data transfer between the owner and the cloud is always a bottleneck. Today many of the cloud providers are using link up clouds to transfer owner's data from cloud to owner. These link up clouds are vulnerable to attacks and may raise to huge reliability problems. The Azure cloud of Microsoft underwent with a serious outrage accident where in the owners secured data is lost at the vendors link ups [].

The other problem can be high cost of huge data transfer. Though the owner has stored huge data of his in the cloud sometimes he is interested to only a portion of the whole data. But this is not possible as owners whole data is placed encrypted with his public key and the whole data only can be decrypted by the owner's private key.

3.2 PROBLEMS ON HUGE DATA ENCRYPTIONS

Many of the service providers provide no encryption mechanisms as owners feel querying their encrypted data is an increased overhead, the owner may not trust the strength of encryption algorithm. Since the encryption mechanism is done as a part of cloud service; making the owner to have less trust on cloud providers. Since whole data is being encrypted owners are at a fear of losing the entire data if not correctly decrypted and most of the owners prefer to place unencrypted data in the cloud as accessing their direct readable data is easy.

3.3 PROBLEMS ON HUGE DATA AUTHENTICATION

Data authentication is a mechanism where the owner is given provision to check the integrity of their stored data in the cloud. A known mechanism for data authentication is appending a MAC to the data. At the verification side the MAC is generated on the data and if appended MAC and generated MAC are

equal the owner's data is authenticated, implicitly showing the data is not altered. Today most of the owners are every time checking the integrity of whole data as their trust on the cloud provider is less. But working the authentication mechanisms even on the unnecessary parts of owner's data is waste of time and raises the cost of the application.

All these problems have risen because of considering the mechanism on the owner's whole data. The proposed mechanism addresses these issues using a naïve structure: a tripartite graph.

3.4 CLOUD DATA PARTITIONING

For efficient accessing of the large data, partitioning of bigger data is what followed. Partitioning makes data storage easy at the cloud. When the data is fed at the client module then itself partitioning is done and each partition is encrypted and stored at the cloud. While data is being partitioned many of the cloud providers follow partitioning with various priorities. Data is generally prioritized as sensitive and Un-sensitive.

Sensitive data is where owners most important files reside. The frequency of accessing these files may be more. Un-sensitive data is not of that important to the owner. The frequency of owner accessing these files may be less when compared to sensitive files. Sensitive data may be frequently queried by the owner. Un-sensitive data may be rarely queried or even not queried by the owner. When data is partitioned with such priorities owners trust on his stored data is very much concerned. When he frequently queries his sensitive data, then the data integrity of this portion can be frequently checked there by raising his trust on the cloud. But what about the portions of the data that are not queried or rarely queried. The owners trust on the integrity of these portions may be zero or less. This is the main disadvantage of partitioning with priorities. To address this, the proposed method uses a tripartite mechanism of checking the integrity of un-queried data so as to raise owner trust on the cloud.

3.5 PRELIMINARIES

This section discusses some preliminaries required to escalate the proposed work.

Owner: The owner is the person who created the data and willing to store the data in the cloud. Much of the research has taken the Owner as a lame man without the knowledge of how actually the privacy mechanisms run. He is at a constant thrive for confidentiality and integrity of his data in the cloud.

Prioritized data: The whole data of the owner is not of equal importance to him. There can be priorities of data of his interest. The owner may be more interested with a portion of his data which he frequently queries the cloud. This portion of data of his whole data is prioritized.

Graph: A graph is structure which can be used to store data or sometimes used to implement the logic of the procedure. A graph is a collection of nodes.

Tripartite graph: A tripartite graph is whose vertices can be divided into 3 independent sets and having an edge between every pair of vertices from independent sets. The tripartite graph divides the whole vertex set under three levels P1, P2, P3. Each level may include any number of

vertices. In our proposed method these vertices are the data files/partitions. Tripartite graphs with 2 vertices in each independent set are denoted by $K(2,2,2)$ and is as shown in Figure 1:

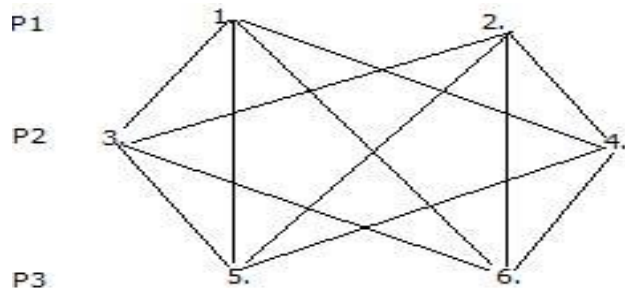


FIGURE 1 Tripartite graph

2-Partitioning: If each level of the tripartite graph includes 2 vertices we say the graph is 2-partitioned.

3-cycle: We call a 3-cycle as a cycle with its 3 distinct nodes in each of the vertex partitions.

Clearly from the tripartite graph there are 3-cycles joining points of P1, P2, P3. Cycle 1-3-5-1 has its 3 distinct nodes from each vertex partitions. In our proposed work (1,3,5) representing the distinct ends of this 3-cycle is taken as the unique id of this 3-cycle.

Clearly in Figure 1 there are $2*(2+2)= 8$, 3-cycles. For a $K(3,3,3)$ there are $3*(3+3)= 18$, 3-cycles. In general in $K(n,n,n)$ there are $n*(n+n)$ cycles. An algorithm to generate the possible distinct 3-cycles with unique id is shown.

Algorithm 1: Generate 3-Cycles

Input: Tripartite

Output: Distinct 3-Cycles.

Step 1: read the tripartite.

Step 2: for each P_i ($i=1,2,3$): Start at an arbitrary partition F_j ($j=1,2$): identify the cycle with its three distinct ends in each P_i . Give an identity which is the end nodes of the cycle.

Step 3: repeat step 2 for all P_i, F_j .

Step 4: Store these 3-cycle ids.

4 Proposed approach

The proposed method uses a tripartite graph to enhance cloud security at finer granules. A tripartite graph partitions its vertices to be at three levels L1, L2, L3. The proposed method allows the owner to prioritize his data into these three levels, with priority labels P1, P2, P3. Level P1 holds data of high priority like owners sensitive and most personal data; and P3 holds data of less priority. With high priority we mean owner's frequency of querying the cloud for that data is more. Data with low priority P3 is less queried or even not queried. Here we consider a tripartite graph $K(2,2,2)$; shows each leveled data is 2-partitioned with labels F as shown in Figure 2. Owners frequency of accessing F1, F2 of P1 is more when compared to other data. In our proposed approach such a prioritized leveled data is more secured with strongest security specifications. More over our proposed approach provides a way to check the data integrity of un-queried data of the owner.

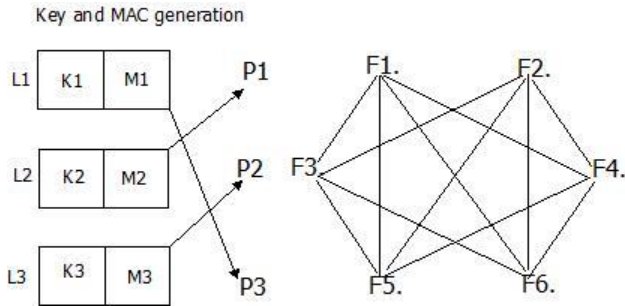


FIGURE 2 Tripartite graph: Varied Specifications

4.1 LEVELS OF SECURITY

The proposed approach discusses levels of security as how, where the strength of security specifications are varying. The proposed work introduced the concept of levels of security as owner always has the high frequency of requesting his high prioritized data at level P1. Which indicates the high prioritized data is frequently communicated in the channel; there by a need for strengthening the security specifications at prioritized level has risen.

The strength of security standards vary as the data priority vary among the 3 levels: P1, P2, P3. Here to increase the complexity of security mechanism, the specifications are used in a way the tripartite edge is connected between the key generation module and data partitioning F. A tripartite edge between L1 and P3 shows the specifications at L1 are used to secure the data at P3 i.e F5,F6. A tripartite edge between L2 and P1 shows the specifications at L2 are used to secure the data at P1. A tripartite edge between L3 and P2 shows the specifications at L3 are used to secure the data at P2.

Since level L1 is connecting less prioritized data of P3, L1 uses less stronger security specifications. Since level L2 is connecting to high prioritized data of P1, L2 uses stronger security specifications. As we know the stronger security specifications are like strong encryption algorithms, strong keys and strong authentication mechanisms.

4.2 SECURITY SPECIFICATIONS

We categorize the security specifications as various mechanisms that provide the data security. The most concentrated security specifications are: Encryptions, Keys and Data Authentication Code.

Stronger specifications: A security specification is strong if:

- The algorithm is very strong with complex modules.
- Difficult to analyze.

Weaker specifications: A security specification is weak if:

- The algorithm need not be that strong, but can solve the purpose.
- Difficult to analyze.

4.2.1 Encryption schemes at various levels

Since level 1 tripartite edge is connecting less prioritized data P3, for data encryption at level 1 less stronger public key cryptosystem, RSA is used. Here F5, F6 are the partitioned data of P3. For better data accessibility, in the proposed approach RSA encryption is done separately on

data of F5 and on data of F6. We denote these encryptions at level 1 as E1(F5), E1(F6).

Since level 2 tripartite edge is connecting high prioritized data P1, for data encryption at level 2 more stronger symmetric key cryptosystem, blowfish is used. Here F1 and F2 are the partitioned data of P1. Blowfish encryption is done separately on F1 and on F2. We denote these encryptions at level 2 as E2(F1), E2(F2). At level 3 for data encryption RSA is used and on the partitions F3, F4 of P2 and are denoted as E1(F3),E1(F4).

Now specifically the encryption specifications are E1-RSA at level 1, 3; E2-Blowfish at level 2.

4.2.2 Data Authentication schemes at various levels

Since level 1 tripartite edge is connecting less prioritized data P3, the defined data authentication specifications at level 1 is MAC: M1; where MAC is of less strong when compared to hash.

Since level 2 tripartite edge is connected to high prioritized data of P1, the defined specifications at level 2 is a hash generated by SHA-512: M2. Since level 3 tripartite edge is connected to P2 of medium priority the authentication specifications defined at level 3 is MD5: M3.

Now specifically the integrity check specifications are M1-MAC at level 1, M2-SHA512 at level 2, M3-MD5 at level 3.

4.2.3 Keys at various levels

To enhance the security of prioritized data the proposed method uses varying key at each prioritized level.

At level 1 as RSA encryption is used, the used key pair is (PR1, PU1). At level 2 since blowfish a symmetric key encryption is used, we use a secret key K1. At level 3 since RSA encryption is used the used key pair is (PR3, PU3). Table 1 shows the security specifications (S) at various levels.

TABLE 1 Security specifications at various levels

Level	S	Encryption	Authentication	Keys
L1	S1	E1-RSA	M1-MAC	PR1,PU1
L 2	S2	E2-Blowfish	M2-HASH-SHA512	K1
L 3	S3	E1-RSA	M3-HASH-MD5	PR3,PU3

4.3 THE APPROACH

This section discusses various storage structures at client and cloud and how data is retrieved from cloud.

Stage 1: Client Storage:

When the owner uploads his data then the client application tripartite the whole data with owners choice of priorities and partitions each accordingly as K(2,2,2). Encrypts each partitioned data as discussed below:

In the proposed approach initially different random numbers are stored at each level. Ensuring less possibility of key cracking these random numbers are assumed to be large. The random number at level 1 is subjected to Fermats Test[] and Miller-Rabin tests[] to generate two primes. These two primes are given input to RSA algorithm to generate PR1, PU1 at level1. Similarly PR3, PU3 at level 3 are generated. Next we have to generate K2 of level 2. Here while K2 is being generated at level 2 the data at P2 and P3 are parallelly encrypted with already generated key pairs of level1 and

level 3 using RSA.

The random number at level 2 is used as a key to run the blowfish algorithm to generate the secret key. The output of the blowfish algorithm is taken as the secret key K2 for encryption at level 2. Once K2 is generated the data at P1 is now encrypted using blowfish algorithm.

Since at the cloud owners encrypted data is stored, owners concern that even this encrypted data can be changed by the cloud is more. To raise owners trust on data integrity, in the proposed approach the integrity check is imposed on top of encrypted data.

The hash, MAC are generated on each encrypted partitions F1, F2 etc and the MAC is appended to each encrypted partitioned data.

A. Tripartite hash Ring: Hash is generated on each encrypted partition by giving encrypted data as input to respective hash algorithms at each level as shown in Table 1. This hash is called the tripartite hash and is denoted as $M(E(F))$ where $M=M1,M2,M3$ as discussed in section 4.2(B) and $E=E1,E2$ as discussed in section 4.2(A). All these tripartite hashes of various data partitions at various levels are stored in a client tripartite hash ring as shown in Figure 3.

Now the client stores these tripartite hashes within its tripartite hash ring. The client identifies the possible unique 3-cycles from this tripartite using algorithm 1 and generates a unique 3-cycle id for each and stores with it.

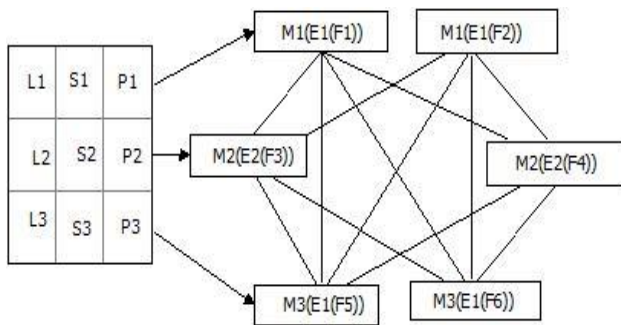


FIGURE 3 Client tripartite hash ring

Stage 2: Cloud Storage:

The client module forwards the tripartite to the cloud which includes: encrypted data appended with the MAC and various MAC specifications used at various levels and is shown in Figure 4.

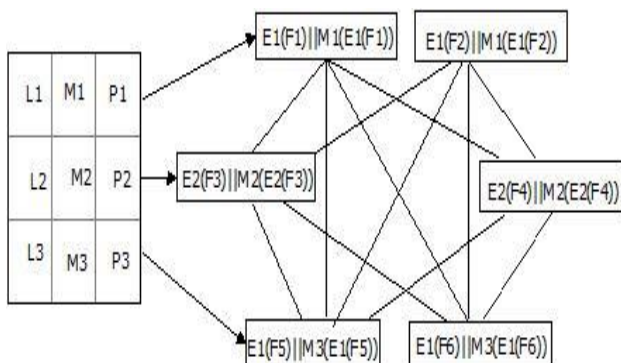


FIGURE 4 Tripartite: forwarded by client to cloud

A. Cloud verification: On receiving the tripartite as shown in Figure 4 the cloud performs an initial verification on data modification during the communication from client.

For this the cloud performs hash on $E_i(F_j)$ and generates $M_k(E_i(F_j))$; $k=1,2,3; i=1,2; j=1$ to 6. If generated $M_k(E_i(F_j))$ is same as client appended $M_k(E_i(F_j))$ then the cloud authenticates data integrity as forwarded by the client.

After verifying the data integrity as forwarded by the client the cloud now constructs its tripartite called the cloud tripartite. It removes each appended hash and stores only each of encrypted data. The cloud tripartite also includes the hash specifications ($M1,M2,M3$) as forwarded by the client and is shown in Figure 5.

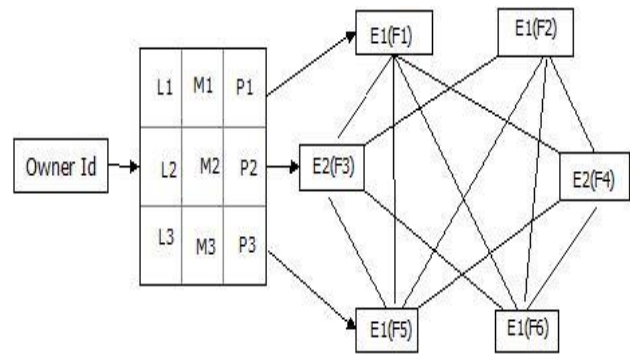


FIGURE 5 Cloud tripartite

Stage 3: Data Processing

When the owner queries for the file F1 then the client module sends the following to the cloud:

Client → Cloud: $L1||P1||E(F1) || Owner_id$.

The cloud on receiving this, first process the owner_id, authenticates the id and indexes to L1, fetches $E1(F1),M1$ from the storage structure. Using the hash specification defined in M1, the cloud generates hash on $E1(F1)$ i.e $M1(E1(F1))$. The cloud now sends the following to the client:

Cloud → Client:

$1.E1(F1)||M1(E1(F1))||M2(E2(F3))||M3(E1(F5))||(1,3,5)$

Where $M2(E2(F3)), M3(E1(F5))$ are the hashes generated by the cloud on the encrypted un-queried partitions of the owner. Here $F1,F3,F5$ are the ends of 3-cycle of the tripartite where these ends $F1,F3,F5$ are connecting each prioritized level in the tripartite for which $F1$ is the queried data. So on request of data from priority level the cloud even forwards two hashes from other un-queried less prioritized levels. Here $(1,3,5)$ is the 3-cycle id forwarded by the cloud.

The cloud may also forward other 3-cycles with $F1$ as initial vertex; where the first hash refers to the queried data and the rest of the two hashes refer to un-queried data. The other possible 3-cycles starting from $F1$ are:

2. $E1(F1)||M1(E1(F1))||M2(E2(F3))||M3(E1(F6))||(1,3,6)$
3. $E1(F1)||M1(E1(F1))||M2(E2(F4))||M3(E1(F5))||(1,4,5)$
4. $E1(F1)||M1(E1(F1))||M2(E2(F4))||M3(E1(F6))||(1,4,6)$.

The selection of 3-cycle from the possible 3-cycles is based on the frequency. If a selected 3-cycle is exceeding its threshold of selection, then another 3-cycle is selected with requested file as the initial point.

A. Integrity check of the queried partition:

The owners queried partition is $F1$. Now on receiving 1,

the client first process the 3-cycle id (1,3,5), identifies the 3-cycle from its tripartite, the client compares forwarded $M1(E1(F1))$ with that stored in the client tripartite hash ring. If both hashes are equal $F1$ is not modified. The client then decrypts $E1(F1)$ using $K2$ and finds $F1$.

B. Integrity check of un-queried partitions:

The client on receiving the other two hashes:

$M2(E2(F3))||M3(E1(F5))$, compares within its tripartite 3-cycle. Identifies the un-queried data parts from the 3-cycle. If these hashes are equal then integrity of un-queried parts $F3, F5$ is checked, raising the owners trust on cloud for the un-queried portions. The communication between client and the cloud is shown in Figures 6, 7.

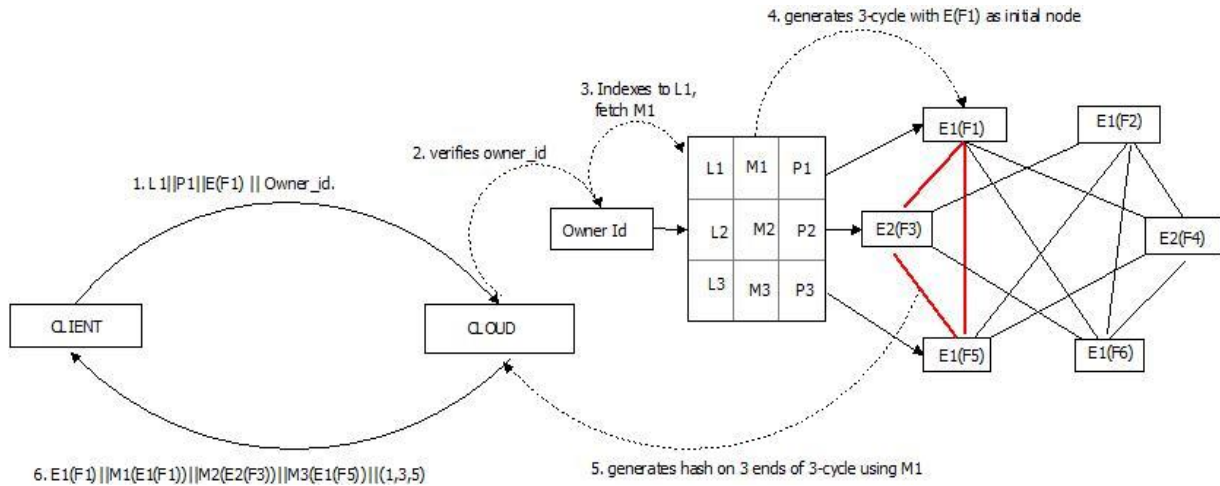


FIGURE 6 Processing by cloud

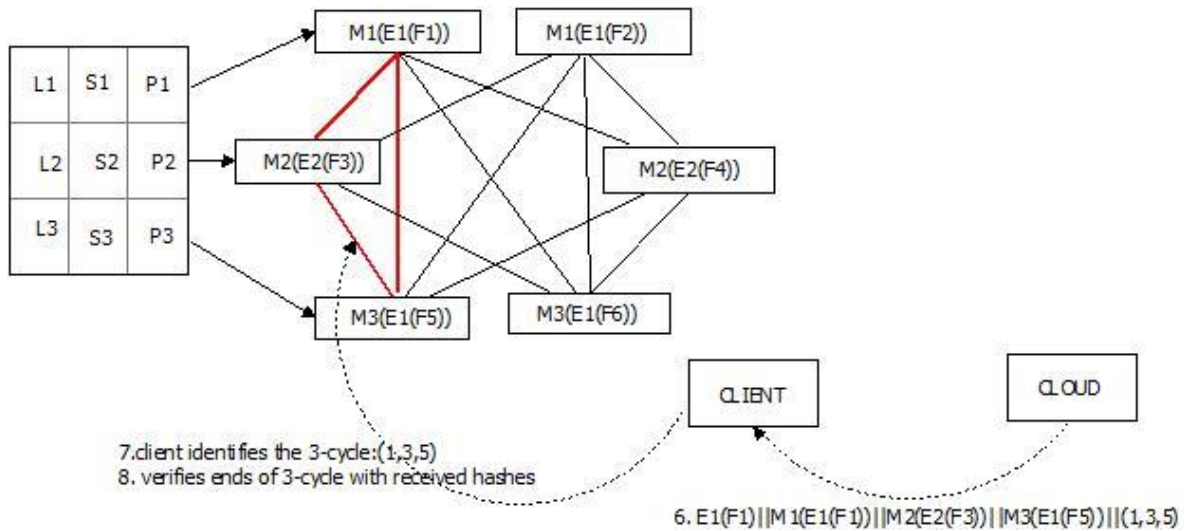


FIGURE 7 Processing by client

5 Performance comparisons

The performance study is done under two cases a) Without Partitioning b) With Partitioning.

A Without Partitioning

Here the whole data is considered as a single unit. Even data is not prioritized; Owners un-important data is equally treated as that of important data. The encryptions are done on the whole data and the encrypted data is placed in the cloud. If the Owner is interested with just a portion of the data then the mechanism has to decrypt the entire data. Since decryption of data happens at owners client module, the cloud has to pass the entire encrypted data to the client. This may raise communication cost because even the data which is not needed by the owner is passed.

Moreover the time taken to run the encryption and decryption algorithm is high as the whole data is being

considered. A single encryption specification is used for the whole data. Using single encryption specification may increase the threat rate. Even owners un-important data is encrypted at a cost same as important data. This is waste of encryption cost on un-important data.

B With Partitioning

The proposed method uses partitioning concept where the whole data is divided into parts under three priorities. There is a provision to query these partitions separately. The frequently queried data of the owner is given top priority. Since the data at this level is important stronger algorithms are used to secure the data at this level. The proposed method then varied the strength of the specifications as the data priorities varied. This type of specification variations at each prioritized level increases the complexity of the method and lowers the threat rate.

The usage of tripartite graph by the proposed method

increased the trust rate of the owner on the cloud. The tripartite mechanism forwarded two un-queried data hashes along with the queried data. With one queried partition the owner can cross check the integrity of two partitions there by the increasing the trust rate by 2% (if there are 100 partitions). Whereas in the usual data partition methods only the queried portion is passed and hence the owner is constantly pondering over the integrity of un-queried partitions thereby reducing his trust on the cloud. Figure 8 shows the trust rate of un-queried partitions.

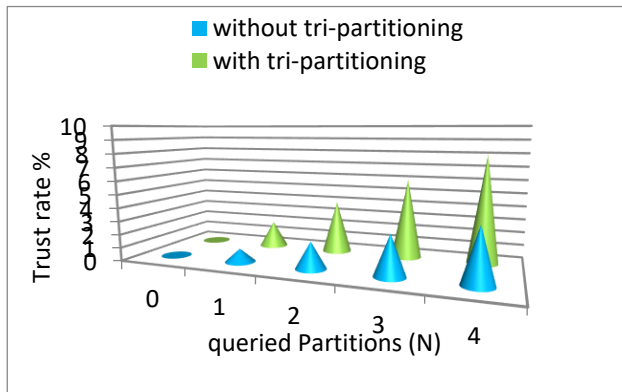


FIGURE 8 Owners trust rate on un-queried partitions.

6 Conclusion

Cloud technology has emerged as a boon to enterprise data. A cloud is a storage area where massive data owners can securely store and can access their data whenever needed. Today cloud computing is more appealing with its high quality applications and services. Though cloud computing is excelling with its computing capabilities, it still faces some new born threats concern to data storage and retrieval.




The whole of owner data is of not equally important to him. There may be cases where he frequently queries a common portion of data and some parts of the data may be less queried and sometimes not queried at all. To address this, what followed by many clouds is data partitioning. The owner’s whole data is divided into small units and each unit is encrypted and signed so as to raise the data security. Some clouds partitioned data with various priorities.

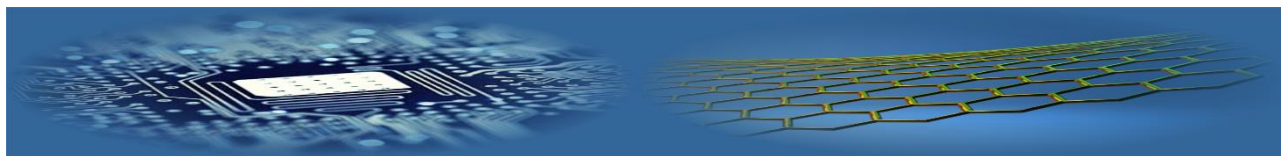
The main drawback with cloud data partitioning with prioritization is owners trust. In data partitioning with prioritized data the owner has the provision to query his most prioritized partition. This partition may be frequently queried.

This paper presents a novel scheme for maintaining prioritized data of the cloud where a new data structure called the tripartite graph is taken advantage. Using the tripartite graph the strength of security is varied at different levels there by raising the complexity of security mechanism. Further using the tripartite mechanism the owners trust of un-queried partitions is considerably raised.

References

- [1] Mohamed E M, Abdelkar H S 2012 Enhanced data security model for cloud computing *IEEE INFOS* 12-6
- [2] Adriana L, Eran T, Vinod V 2012 On-the-fly multiparty computation on the cloud via multikey fully homomorphic encryption *ACM-STOC* 1219-34
- [3] Kandukuri B R, RamaKrishna V 2009 Cloud security issues *IEEE-SCC* 517-20
- [4] Sood S K 2012 A combined approach to ensure data security in cloud computing *Network and computer applications* 35(6) 1831-8
- [5] Wang Q, Wang C, Li J, Lou Ren 2009 *Enabling public verifiability and data dynamics for storage security in cloud computing* Springer 5789 355-70
- [6] Wang C, Ren K, Lou W, Li J 2013 Storing shared data in the cloud via security - Mediator *IEEE ICDCS* 124-33
- [7] Chu C K, Chow S S M, Tseng W G, Zhou J 2013 Key aggregate cryptosystem for scalable data sharing in cloud storage *IEEE, parallel and distributed systems* 25(2) 468-77
- [8] Ren K, Wang C, Wang Q 2012 Security challenges for the public cloud *IEEE Internet computing* 69-73
- [9] Khedkar S V, Gawande A D 2014 Data partitioning technique to improve cloud data storage security *IJCSIT* 5(3) 3347-50
- [10] Ye Y, Xiao L, Yen I L, Bastani F 2010 Cloud storage design based on hybrid of replication and data partitioning *IEEE, ICPADS* 415-22
- [11] Kaufman L M 2009 Data security in the world of cloud computing *IEEE security and privacy* 7(4) 61-4

AUTHORS	
	<p>P. Dileep Kumar Reddy</p> <p>Current position, grades: Lecturer in Department of Computer Science & Engineering, JNTUA College of Engineering, JNT University, Anantapur.</p> <p>University studies: B.Tech, M.Tech in JNTUA College of Engineering, Anantapur. His areas of specialization are cloud computing, Network Security. He is in teaching since 2010.</p> <p>Publications: presented papers at National and International Conferences and published articles in National & International journals.</p>
	<p>Dr. C. Shoba Bindu</p> <p>Current position, grades: Professor and Head of the Department in Computer Science & Engineering, JNTUA College of Engineering, JNT University, Anantapur. She has guided many external and internal projects and has good contributions in many of reputed journals.</p> <p>University studies: Ph.D degree in computer science and engineering from JNT University, anantapur.</p> <p>Scientific interests: mobile and adhoc networks, network security, data mining and cloud computing.</p> <p>Experience: around 16 years of experience in teaching and Research.</p>
	<p>Dr. R. Praveen Sam</p> <p>Current position, grades: Professor in Computer Science & Engineering, G.Pulla Reddy Engineering College, Kurmool.</p> <p>University studies: received his Ph.D degree in computer science and engineering from JNT University, anantapur.</p> <p>Scientific interests: mobile and adhoc networks, network security, data mining and cloud computing. She has around 13 years of experience in teaching and Research.</p> <p>Publications: presented papers at National and International Conferences and published articles in National & International journals. He is life member in CSI, ISTE, IAENG, IE.</p>



Virtualization safety

Zh E Aytkhozhaeva, A A Ziro, A Zh Zhaibergenova*

Kazakh National Research Technical University, associated professor, Satpayev Str. 22, Almaty, Kazakhstan

**Corresponding author's e-mail: zhanshuak@gmail.com*

Received 13 March 2017, www.cmnt.lv

Abstract

Article considered virtualization technologies, their types, advantages and disadvantages. Attention to specific risks and information security threats in case of virtualization platforms is paid. The main risks of virtualization platforms are defined. Potential internal vulnerabilities of virtualization platforms can be revealed only by testing for penetration which user-friendly and available instrument for implementation is specialized by Kali Linux OS. The attacks to the virtual machines with use of the Kali Linux tools were organized. As a result of experiments is Kali Linux allows revealing and analyzing vulnerabilities at the channel, network and transport levels. For detection of problems at the level of applications that is urgent for virtualization of platforms, it is necessary to use commercial products of ethic hacking in addition.

Keywords:

virtualization platforms,
risks,
penetration testing

1 Introduction

Virtualization technologies, along with cloud computing, take key positions among the advanced and perspective trends in IT area since 2009 (according to the analytical company Gartner). Taking into account that virtualization technologies are the fundament of cloud computing, it is possible to give with confidence a prize-winning place of virtualization without which cloud computing are unrealizable. Much of advantages and disadvantages of virtualization technologies automatically reflected in cloud computing.

Virtualization technologies passed the already considerable way of the development from purely scientific interest and decisions for insulation of computing environments of different tasks within one mainframe before creation of the virtual area networks and program containers encapsulating a complete set of the virtual hardware resources [1].

In a general view there is the type of virtualization: virtualization of resources. Virtualization technologies resources historically gained earlier development and recognition- the multiprocessor systems, clustering of computers, grid computing, the virtual area networks, etc. Virtualization technologies platforms began to develop later. Now actively develop and progress, have a set of different types of implementation are cornerstones of cloud computing.

Deployment on one physical server of a set of the virtual servers which ensure functioning of any operating system gives big advantages and absolutely new opportunities [2]. Prospects of technology of virtualization of platforms are also defined by it. But wide recognition and application restrains existence enough serious shortcomings of this technology [3].

Experience shows that many projects on virtualization were comprehended by failure. Nearly a half of the companies (44%) which made virtualization attempts can't claim about their successful completion. This is due to the difficulty of assessment of resetting of investments, and to

complexity and high cost of deployment and support of corporate virtual infrastructure.

2 Virtualization disadvantages

Appearance of new poorly studied and low-probed risks and security risks of information when using virtualization including in cloud computing also belongs to shortcomings of technology. This weakness is compounded by the fact that different types of virtualization of platforms bear specific risks and threats due to specifics of the implementation.

Physical, logical and program structures are a defining factor for appearance of risks and threats of information. In architecture of the computing systems used in the virtual decisions nothing changed. Therefore the basic conventional principles of support of information security shall be observed also in such systems. Virtualization not to a lesser extent needs protection and a judgment on its bigger safety thanks to the most structure doesn't respond the reality. On the contrary, specific risks and security risks of information in case of virtualization of platforms in addition take place, as well as in case of any new technology. The main problem constraining development and implementation of virtualization of platforms is the problem of protection of such systems. Use of standard checked methods and security features in case of virtualization platforms, in that look in what they exist now isn't enough. In case of virtualization of platforms of property of a physical medium exist in the form of program settings. And it is simpler to change program settings illegally.

For example, the standard security reference monitor (a resident component of safety) controlling loading processes can't perform the functions when loading the virtual machines. At the same time the resident component of safety shall be present and have access to the controlled environment. For the virtual environment there shall be specific mechanisms of monitoring. By operation in the virtual environments of

virtualization platforms the situation changes very quickly. New components are quickly created and together with them also new potential threats are created. At the time of creation of the virtual machine it isn't protected in any way. Constant control of a situation is necessary.

The solution is complicated by the fact that architecture of systems of virtualization different for different types of virtualization platforms. For example, depending on a type of virtualization platforms, the hypervisor, being the manager (monitor) of the virtual machines (compact highly specialized OS), is set or on "bare iron", or/and on OS. There are also such types of virtualization (virtualization at the level of an operating system) when the hypervisor isn't used.

The expression that virtualization ensures the best information security, is based only that on the virtual machines it is simpler to set rules of a network access. The statement, that vulnerability of a hypervisor and probability of the attack to it very low, is based that there is no information on the attacks to hypervisors. It isn't confirmation of either absence of the attacks, or high safety of hypervisors.

There is no accuracy of an exception of risks and threats of virtualization of platforms exist. Besides, now virtualization technologies begin to be applied to violation of confidentiality, integrity and accessibility of information more and more actively. The research Blue Pill project which visually showed how technologies of the hardware virtualization can be applied in the espionage purposes (the Blue Pill program including a hypervisor) is known. Now Blue Pill is a code class name of root kits (the programs hiding presence at system of malicious software) based on use of the hardware virtualization.

3 Risks and threats of virtualization

The research of risks and information security threats when using virtualization is an urgent problem, both in respect of safety of the virtual infrastructures, and in the theoretical and practical development plan and advances of technologies of virtualization.

It is necessary to define the main risks of virtualization platforms:

1. a uniform point of a failure in a failure mode of the physical server;
2. a uniform point of a failure in a hypervisor failure mode (in case of its existence) and/or hosts OS (in case of its existence);
3. risk of a compromise of a hypervisor of the virtual machines (in case of its existence) and/or hosts OS (in case of its existence);
4. implementation of a hypervisor in the form of software module is more vulnerable to the attacks, than hardware or software implementation;
5. risk of a compromise of data by memory transmission (local storage) from one virtual machine to another;
6. violation of insulation of processes (virtual machines), basic principle of virtualization;
7. risk of a nonadjustable data migration of limited access;
8. a possibility of network attacks between the virtual machines (virtual servers) located on one physical server;

9. the known decisions on program virtualization aren't provided with protection at the hardware level on TPM technology (the specification describing the trustable module which makes available to an operating system guaranteed safe services). Even in new technologies of the hardware virtualization a part of the mechanism of virtualization is implemented by the software of a hypervisor.

There is a nonzero probability of existence of the hidden or functional vulnerabilities of a hypervisor and possibility of carrying out the attack against it:

1. the risks connected to increase in mobility and use of the virtual machines in architecture of cloud computing;
2. the unrolled virtual systems often don't conform to requirements of a corporate policy of information security.

Operations in the field of minimization of new risks and threats of virtualization technologies of platforms are carried in different directions. Some of them are provided below.

Good hypervisors contain the virtual switchboards and firewalls which settle down between physical interfaces of the server and the virtual interfaces of the virtual machines. Protection against network attacks between the virtual machines (virtual servers) located on one physical server is provided (in the absence of a compromise of a hypervisor of the virtual machines).

For the hypervisors Microsoft, IBM, Citrix and VMware companies gives a guarantee of insulation of processes and data of the virtual machines from each other. It is considered that it provides a possibility of safe information processing of different level of privacy on single physical device.

The Symantec Company offers the Symantec NetBackup™ platform with V-Ray technology which on the basis of patent decisions provides evident representation of the virtual machines and applications on physical and virtual servers. The technology of backup and restoration for the environments VMware and Microsoft Hyper-V, quickly selective recovery of data from the applications working under control of hypervisors of VMware and Hyper-V is implemented.

The AMD company developed AMD-V technology in which a special protect mode of start of the monitor (hypervisor) of the virtual machines is realized. The Intel Company used the previous TPM 1.2 specification (the last TPM 2.0 specification) for increase in security in one of the chipsets (technology of safety LaGrande/TXT).

Even this short list of operations shows, risks of virtualization of platforms are how various. At the enterprise using technologies of virtualization it is necessary to clarify - what risks threaten business processes at present and to estimate these risks [4]. In case of estimation of risks assessment of probabilities of events as the risk is a combination of probability of an event and its consequences are executed.

$$R = S * E, \quad (1)$$

where R is the risk, S is the extent of damage, E means probability of an event.

For detection of potential internal vulnerabilities and assessment of probability of an unauthorized event, including in case of virtualization platforms, it is possible to use penetration tests. During testing the tester (auditor)

models actions of the malefactor, trying to break information security of a subject to protection. Search of vulnerabilities of system of protection and their subsequent use is executed. Now testing for penetration is one of the information security systems recognized around the world as method in case of the active audit. There is a standard on conducting testing for penetration [5]. There are different programs for conduct testing for penetration (ethic hacking), including specialized Kali Linux OS [6]. The Kali Linux tools allow executing search of operation of vulnerabilities in Web servers, wireless protocols, communication links, mobile devices, applications. Having set Kali Linux on single virtual machine, it is possible to attack other virtual machines, as for the purpose of testing for penetration, and unauthorized obtaining information.

4 Detection of vulnerabilities by means of testing

In an experiment the virtual machines created in VMware Workstation were used. Having set Kali Linux on one virtual machine, the attacks to other virtual machines for the

purpose of testing for penetration and unauthorized obtaining information were organized. Results of some of them are given below.

One of the methods of detection of vulnerabilities is port scanning of the virtual machines by means of the network Wireshark analyzer. Wireshark analyzer is the application of the Kali Linux. Wireshark analyzes the traffic passing through the network interface of the computer that allows viewing completely contents of the transferred packets at all levels. Wireshark listens to all network traffic and captures it.

In a Figure 1 the results received by Wireshark by the analysis of the traffic passing on wires through ports of one of the virtual machines are provided. In a Figure 1 the flow of network packets which can be analyzed is visible. It is visible that the system of virtualization VMware is used. Contents of the packet both sent, and received can be opened, having executed click on it. We obtain information on port and the IP address of the sender, port and the IP address of assignment, the transfer protocol, lifetime of a packet, etc.

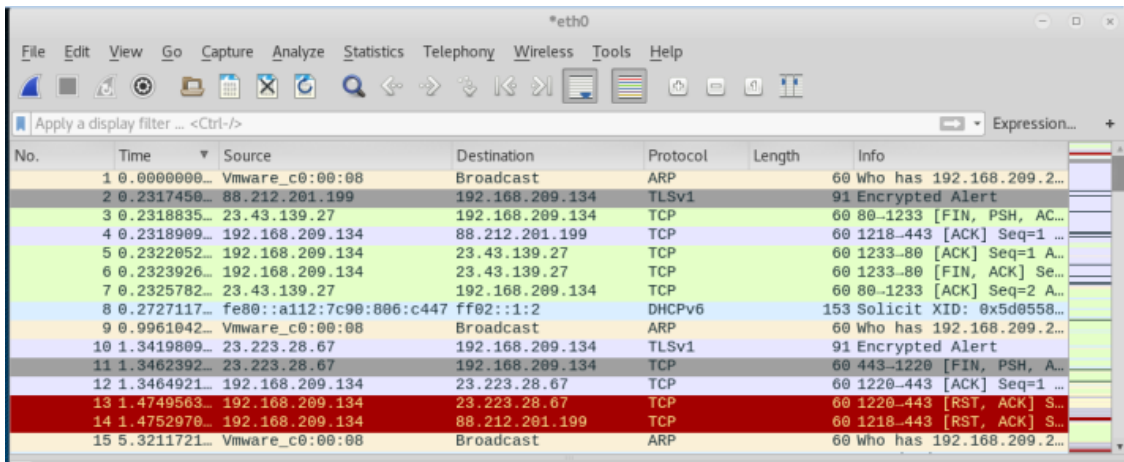


FIGURE 1 An example of the scanned traffic

Packets can be filtered in a set of parameters. It is possible to set the filter for receiving the traffic meeting certain requirements. By means of the "http.request.method == "POST"" filter it is possible to intercept login and the password, to obtain information on a frame, the version the protocol Internet, the data transfer protocol, the hypertext transfer protocol. In a Figure 2 in the upper part the selected

packet from the intercepted flow of packets which contents reveal below is shown. It is visible that the packet of HTTP is encapsulated in a packet of TCP (transport layer), the packet of TCP is encapsulated in IP (network layer), and IP is in turn encapsulated in Ethernet. In the lower part of a Figure 2 the HEX code is shown.

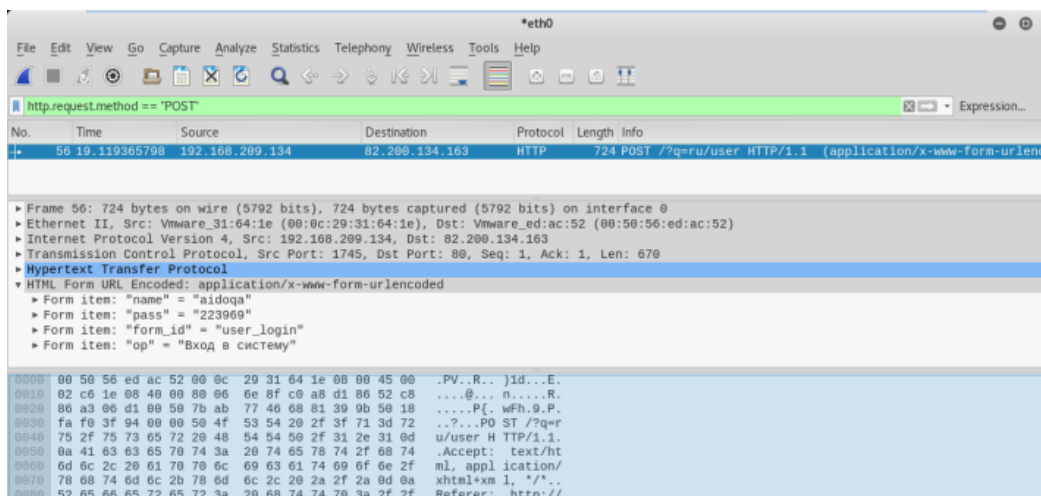


FIGURE 2 An example. Interception of data

Disclosing the level of each protocol, we obtain the detailed information from each level. In a Figure 2 information of a packet 56 of HTML forms is shown. The login and the password entered by the user are defined. If to open the most top line, it is possible to obtain general

information about a frame.

Information on a packet of 2572 from the Ethernet level is shown in a Figure 3. Information about destination, source address and type of IP (IPV4) is presented.

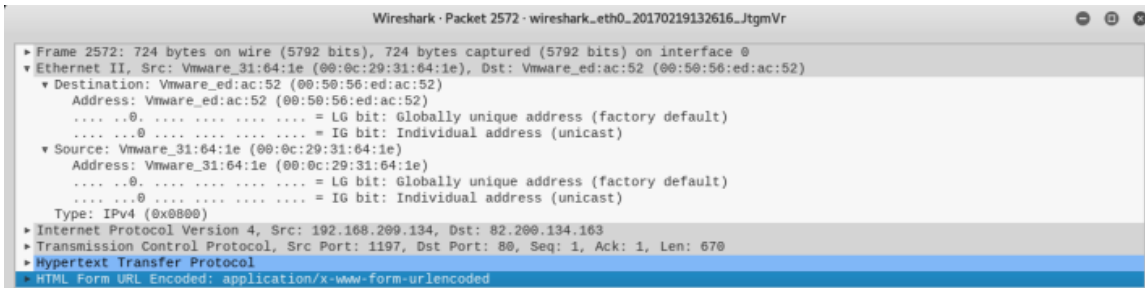


FIGURE 3 An example of the intercepted data (from the Ethernet level)

In a Figure 4 is shown information on a packet 2572 from the IP (IPV4) level. The figure includes information

about differentiated services, flags, destination GeoIP.

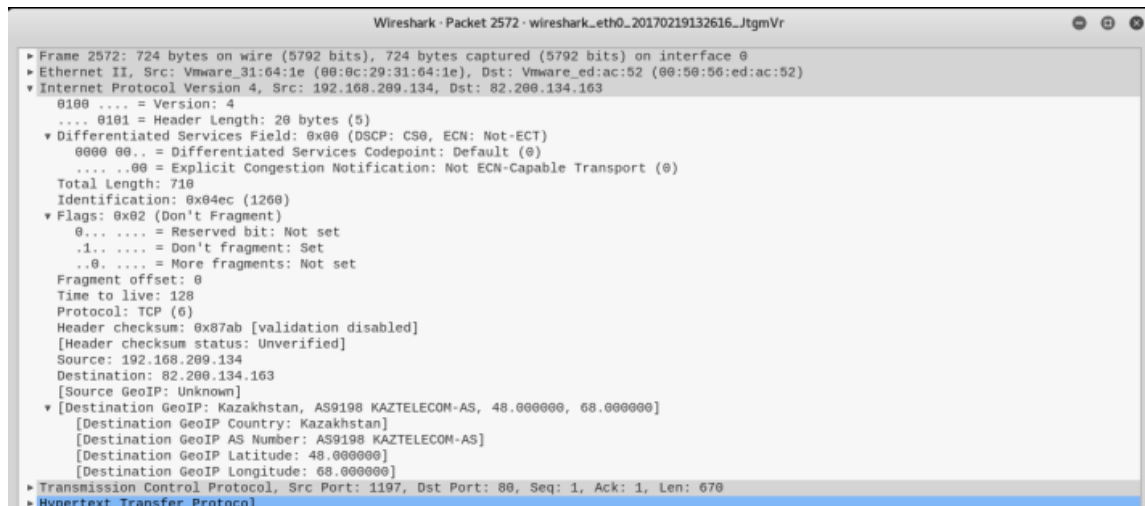


FIGURE 4 An example of the intercepted data (from the IP level)

Information on a packet of 2572 from the TCP level is shown in a Figure 5. Figure defined source port, destination

port, flags and etc.

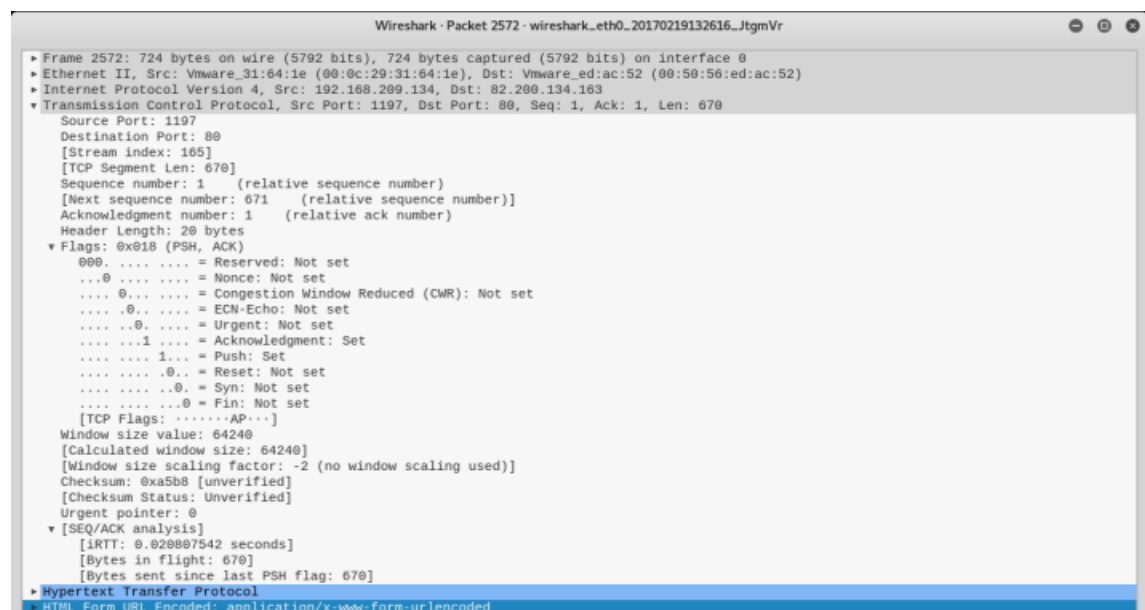


FIGURE 5 An example of the intercepted data (from the TCP level)

Information on a packet of 2572 from the HTTP level is shown in a Figure 6. Information about request method, host, content-length and cookie is presented.

In case of use of cryptography protocols (in this case TLS), a part of the intercepted information will be ciphered

and the level of the ciphered sockets will be visible (Figure 7). Figure defined TLSv1, handshake protocol.

In case of interception of the ciphered data they can be decrypted by using the appropriate settings.

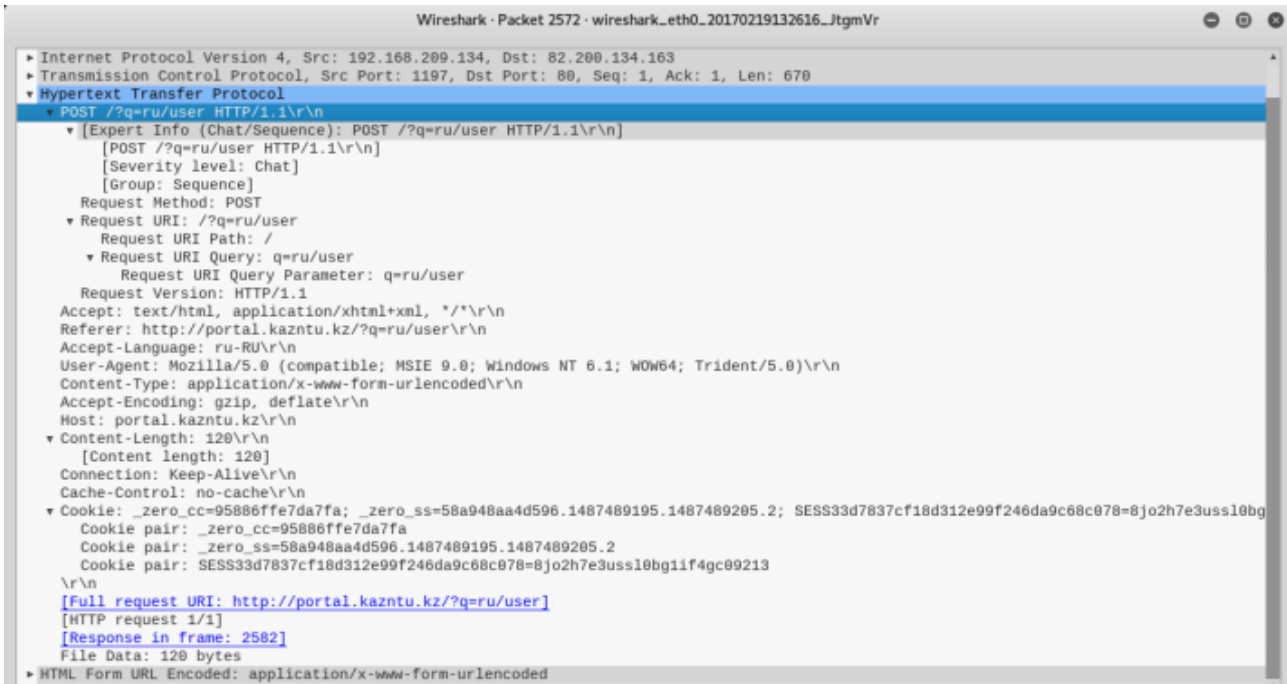


FIGURE 6 An example of the intercepted data (from the HTTP level)

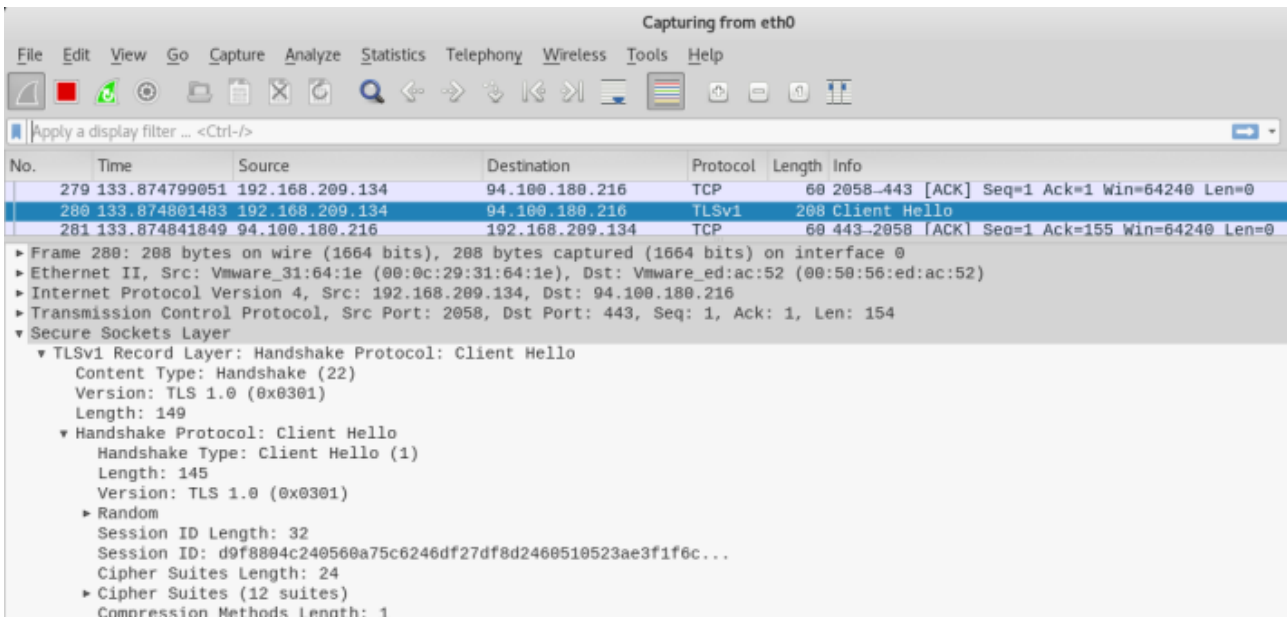


FIGURE 7 An example. Interception of the ciphered data

5 Conclusions

Virtualization is not only perspective strategic technology which has specific risks and threats. The vulnerabilities of platform virtualization need to be identified considered and minimized, including by means penetration testing. Free products of penetration testing are easy to use. They allow identifying vulnerabilities on the channel, network and transport levels, having built-in expert systems. When




solving problems at the application level, these products unusable. In this case necessary to use commercial solutions. Commercial products of the penetration testing use more advanced technologies and have extended capabilities compared to free ones. This provides additional opportunities in assessing risks and threats. Platform virtualization works on the application level. Therefore, penetration testing should be performed on both the channel, network and transport layers, and at the application level.

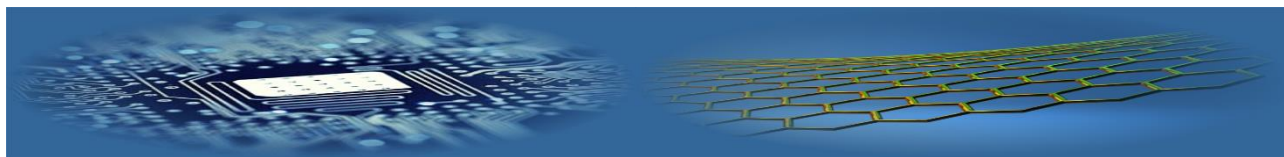
It should be remembered that virtualization platform technologies can be applied also in illegal purposes to violation of information security. Virtualization gives new

opportunities, but also place of IT professionals more demands, both in respect of professional level, and in respect of level of responsibility and ethics.

References

- [1] Portnoy M 2016 *Virtualization Essentials* John Wiley & Sons: Indianapolis p. 309
- [2] *The Advantages and Disadvantages of Virtualization* Milner 2015 <http://milner.com/company/blog/technology/2015/07/14/the-advantages-and-disadvantages-of-virtualization/> 15 Feb 2017
- [3] Ziro A A, Aytkhozhaeva E Zh 2017 Trend virtualizatscii i ego osobennosti III *Mezhdunarodnaya nauchno-practicheskaya konferentsiya: Fundamental'nye Nauchnye Issledovaniya: Teoreticheskie i Practicheskie Aspekty ZapSibNTS: Kemerovo pp 206-9 (in Russian)*
- [4] ISO/IEC 27005:2011 Information technology Security techniques Information security risk management *International Organization for Standardization* 2011 http://www.iso.org/iso/catalogue_detail?csnumber=56742/ 1 Feb 2017
- [5] Information Supplement: Penetration Testing Guidance *PCI Security Standards Council* 2015 https://www.pcisecuritystandards.org/documents/Penetration_Testing_Guidance_March_2015.pdf 1 Mar 2017
- [6] Official Kali Linux Documentation *Kali* 2014 <https://www.docs.kali.org/kali-linux-documentation/> 9 Mar 2017

Authors	
	<p>Evgeniya Aytkhozhaeva, 1947/02/01, Republic of Kazakhstan</p> <p>Current position, grades: associated professor of the Department of Information Security, Candidate of Technical Sciences University studies: St. Petersburg State Electrotechnical Institute (Technical University "LETI"), Russia Scientific interest: Information Security, Databases Systems, Hardware of Cryptography Publications (number or main): over 160 Experience: more than 30 years of scientific and pedagogical experience</p>
	<p>Aasso Ziro, 1992/01/01, Republic of Kazakhstan</p> <p>Current position, grades: tutor of the Department of Information Security University studies: ITMO University, Russia Scientific interest: Information Security Publications (number or main): 7 Experience: 1 year</p>
	<p>Zhanshuak Zhaibergenova, 1993/06/17, Republic of Kazakhstan</p> <p>Current position, grades: tutor of the Department of Information Security University studies: ITMO University, Russia Scientific interest: Information Security Publications (number or main): 6 Experience: 1 year</p>



Improvement of learning efficiency of the neural networks, intended for recognition of graphic images in systems of biometric authentication

**L Tereykovskaya¹, I Tereykovskiy², E Aytkhozhaeva³,
S Tynymbayev³, A Imanbayev^{3*}**

¹Institute of Solid State Physics, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", 14-a Polytechnichna str., 03056, Kyiv, Ukraine

²Kyiv National University of Construction and Architecture, 31, Povitroflotsky Avenue, Kyiv-037, 03680 Ukraine

³Kazakh National Research Technical University, Satpayev Str. 22, Almaty, Kazakhstan

*Corresponding author's e-mail: azekee_92@mail.ru

Received 29 March 2017, www.cmnt.lv

Abstract

Article is devoted to a problem of use of neural network technologies in the field of biometric authentication of users. It is shown that one of important the shortcomings of application of neural networks technology on the basis of a multi-layer perceptron for recognition graphic images in systems of biometric authentication of users is insufficient quality of processing of statistical data which are used when forming parameters of educational examples. It is offered to increase quality of educational examples due to use of the procedure of neural network coding of value of the expected output signal of educational examples which allows consider closeness of standards of the recognized classes in this signal. The coding procedure of the expected output signal providing use of a probable neural network is developed. The appropriate mathematical devices are created. As a result of numerical experiments it is shown that application of the developed procedure allows reduce the number of the computing iterations necessary for achievement of the given error of training by 30-50%. It specifies prospects of use of the proposed solutions for improvement of learning efficiency of the neural networks, intended for recognition of graphic images in systems of biometric authentication.

Keywords:

neural network,
information security,
learning,
biometric authentication

1 Introduction

The proved ability of neural network models (NNM) to effectively recover unknown multivariate table valued functions provided their broad application during creation of systems of recognition in different fields of science and technology [1, 2]. Application of NNM is especially urgent when the task of recognition is poorly formalized, and the result of its decision is highly responsible. A characteristic example of such tasks is application of NNM for recognition of graphic images in systems of biometric authentication of users of information systems. Though the practical experience of use of the known neural network systems and the analysis of sources [1, 2] also specifies rather powerful scientific and practical practices in this direction, but the same analysis specifies also insufficient learning efficiency of NNM on the basis of a multi-layer perceptron which are a basis of the specified systems. Because of this shortcoming time of creation of neural network system improvement and the accuracy of recognition of unknown input images decreases. Let's mark that NNM like a multi-layer perceptron is adjusted for training by the method "with the teacher" are considered. Today in technical systems generally such models are read the most approved. The main advantage of the multi-layer perceptron is high computational capability and as the shortcomings refers

complexity and duration of training. Key parameters which define learning efficiency of NNM is the time and an error of training [2]. Values of these parameters directly depend on quality of educational examples which in case of the given statistical selection shall be provided due to different processing procedures of statistical data.

2 Analysis of the known approaches to processing of statistical data

On the basis of data [1, 2] is defined that the majority of the known approaches to processing of statistics assume performing procedures which realize centering, normalization, scaling and/or scaling of input and output parameters of educational examples. For output parameters the main objective of the specified procedures is coercion of definition boundaries of variables of a real object to a certain interval. When using sigmoidal function of neurons activation of an output layer this interval is restricted to zero and unit, and when using a hyperbolic tangent limits of an interval from -1 to +1.

Thus, the listed procedures only adapt output parameters of educational examples to the look suitable for application in neural network models, but aren't intended for impact on time and an error of training. At the same time results [3] are specified to reduce time and an error of training it is possible

due to reflection in the expected output signal of educational examples of closeness of standards of the recognized classes. In the same operation [3] it is shown that it is possible to realize such display by means of the procedure of expert assessment of closeness of the specified standards. However use of the offered procedure is related to need of attraction of highly enough qualified experts for specific application area use of neural network system. In many cases it is impossible. Also it is possible to apply algorithmic criterion for evaluation of closeness of standards for implementation of display. However development of qualitative criterion requires considerable efforts and its existence levels needs use of NMM. At the same time the task of assessment of closeness of a limited set of standards of classes can be considered in a perspective use of low-resource neural networks for prospecting data analysis [2, 5] that allows assuming prospects of neural network coding of the expected output signal of educational examples of a multi-layer perceptron.

3 Formulation of the problem

The purpose of the real research is development of the procedure of application of low-resource neural networks for coding an output signal of educational examples of a multi-layer perceptron which at the expense of the accounting of closeness of standards of the recognized classes allows increase efficiency of its training.

3.1 DEVELOPMENT OF CODING PROCEDURE OF THE EXPECTED OUTPUT SIGNAL

We detail the task of reflection in an output signal of educational examples of closeness of standards on a specific example of neural network recognition of uppercase printing letters of the Ukrainian alphabet which are displayed in black color on a white background. The classical coding procedure of parameters of standards of letters illustrated with Figure 1 consists in sequential implementation of five stages.

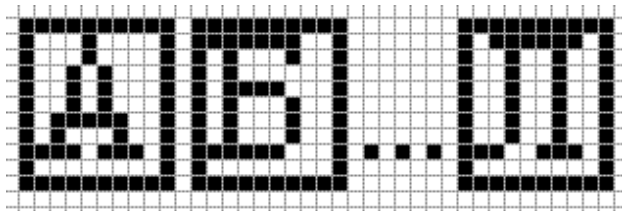


FIGURE 1 Illustration of coding procedure

1 stage. Every letter accommodates in a separate rectangle of the fixed size which is partitioned into cells by a grid chart. The quantity of these cells determines the number of input parameters of neural network model. In Figure 1 this quantity is equal:

$$K = a \times b = 11 \times 10 = 110, \quad (1)$$

where a and b – height and width of a rectangle which describes a letter. Within a separate rectangle each cell assigned certain number which corresponds to number of the entering parameter of neural network model.

2 stage. For a separate letter i value of input parameter is equal 1 if appropriate i a cell is filled in black color, and is

equal 0 in case it is filled in white color.

3 stage. Letters register and numbered in alphabetical order. For example, as it is shown in figure 1, to a letter A there corresponds $n = 1$, to a letter B there corresponds $n = 2$, and a letter L - $n = 13$.

4 stage. Number of a letter in the alphabet defines the raw value of the expected output signal of neural network model. Thus, for a letter A the raw value of the expected output signal is equal 1, for a letter B - 2, and for a letter L - 13.

5 stage. For each letter the raw value will be transformed to the look suitable for use in neural network model. Let's consider basic model with sigmoidal function of activation. Depending on structure of neural network model two options of conversion are possible.

Option 1. The output signal of neural network model is defined by one output neuron. In this case for an educational example of a standard of n -y of a letter the expected output signal decides on the help of the following expression:

$$y(n) = \bar{y}(n) / N = n / N, \quad (2)$$

where $\bar{y}(n)$ – the raw value of an output signal of a standard n -oh letters, n – number of a letter in the alphabet, N – quantity of letters in the alphabet.

Option 2. The output signal of neural network model is defined by a set of output neurons which amount is equal to quantity of letters in the alphabet. At the same time number of output neuron is equal to number of the appropriate letter in the alphabet. Therefore for an educational example of a standard n -oh of a letter the expected output signal is defined so:

$$\begin{cases} y_n(n) = \bar{y}(n) / n = 1 \\ y_k(n) = \bar{y}(n) - n = 0, k = 1, \dots, N, k \neq n \end{cases}, \quad (3)$$

where k - number of output neuron.

Expression (3) can be interpreted as follows, n -oh of a letter for n -go of output neuron the expected output is equal in an educational example of a standard 1, and for all remaining neurons the expected output is equal to 0. Let's mark that concerning the first option of conversion the second option more general. Therefore only the second option will be considered further.

The basic lack of the described coding procedure is that the value of the raw expected output signal defined at the fourth stage badly corresponds to geometrical closeness of the recognized images. It is obvious that the image of a letter A is much more similar to the image of a letter L, than to the image of a letter B. At the same time the classical option of coding contradicts this fact. Afterwards in case of implementation of the 5th stage this error doesn't allow to consider geometrical similarity of images correctly.

For elimination of this shortcoming it is offered to use low-resource NMM which training doesn't require determination of the expected output signal in a numerical look for assessment of closeness of standards. First of all, this type of network includes the NMM which is capable self-learning. Classical representatives of this type are networks on the basis of Kohonen's card and Boltzmann's machine. However a hindrance to their application is the low generalizing ability which involves an erratic possibility of reference of different standards to one cluster. More perspective is application of the PNN network in which

educational examples the expected output signal represents not number, but the name of a class [2]:

$$\{x\}_K \rightarrow Name_Y, \tag{4}$$

where $\{x\}_K$ – a set of input parameters, $Name_Y$ – the name of a class to which this educational example belongs.

For an example in Figure 2 the structure of the PNN network which is intended for correlation of unknown black-and-white graphic images to one of three classes - A, B or C is shown. It is supposed that each of graphic images is placed in a separate rectangle by the size of $a \times b$ pixels. Thus, the amount of input neurons (input parameters) correspond to number of signs of a class and decide on the expression help (1).

The network consists of four layers of neurons: input - Ln_{in} , images - Ln_0 , adding - Ln_s and a day off - Ln_{coll} . The quantity of elements of a layer of images is equal to quantity of educational images. The input layer and a layer of images make full-meshed structure. The quantity of elements of a layer of summing is equal to quantity of classes. The element of a layer of images is connected only to that element of a layer of summing to which corresponds the class of an image.

Generally both the quantity of classes and quantity of educational examples can be arbitrary number. For the example shown in Figure 2, the amount of neurons of a layer of adding is equal to 3. To each neuron of a layer of adding there correspond two neurons of a layer of images, that is to each class there correspond two standards - educational an example.

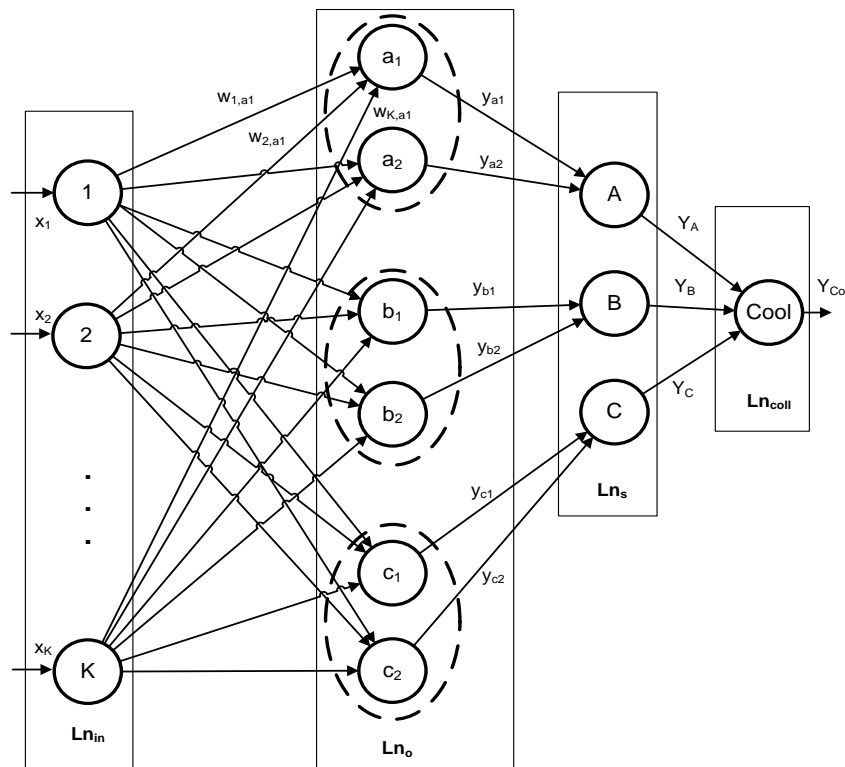


FIGURE 2 Example of structure of the PNN network

For the communications entering neuron of a layer of images, weight factors are set same as components of the appropriate educational vector of an image. So, if for an educational example a1 a set of parameters $\{1, 0, \dots\}_K \rightarrow A_s$, that $w_{1,a1} = 1$, and $w_{2,a1} = 0$. Entering of this example into a network is realized soconnection between neuron a_1 and neuron A an adding layer is established:

1. the new neuron a1 is added to a image layer;
2. weight factors of the entering communications are set;
3. connection between neuron a1 and neuron A an adding layer is established.

Let's mark that weight factors of the communications entering neurons of a layer of summing and an output element are equal to 1. Thus, all PNN parameters directly are defined by educational data, and training of such NNM happens rather quickly. The important positive point of training activity of the PNN network is presence only of one

controlling parameter of training which value is selected by the user. Actually this parameter is the radius of function of Gauss which value not considerably influences quality of recognition. The output signal of an arbitrary j is neuron of a layer of images is calculated like:

$$y_j = \sum_{k=1}^K \exp\left(\frac{-(w_{k,j} - x_k)^2}{\sigma^2}\right), \tag{5}$$

where x - unknown image, x_k - k of a component of an unknown image, $w_{k,j}$ - weight factor of communication between k input neuron and j neuron of a layer of images, K - quantity of components of an input image, σ - radius of function of Gauss.

In neurons of a layer of summing the linear function of activation is used. The output signal of n-go of neuron of a layer of summing (Y_n) is calculated so:

$$Y_n = \frac{\sum_{i=1}^I y_i}{I}, \tag{6}$$

where I – amount of neurons of a layer of the images connected to n neuron of a layer of summing, y_i – activity i neuron of the layer of images connected to n summing layer neuron.

Value of activity of neuron of a layer of summing is equal to probability of reference of an input image to a class which corresponds to this neuron.

The task of an output element is only determination of neuron of a layer of summing with the maximum activity. Therefore in practice the output element can be realized without use of neural network technologies. Let's mark that though only the name of the most probable class is result of recognition of the PNN network, but values of output signals of a layer of summing are specified probability of belonging of an unknown input example to one of the recognized classes.

The offered procedure of use of the PNN network for coding of the expected output signal is as follows:

1. by means of expression (4) the set of educational examples which correspond to a set of standards of the recognized classes is created;
2. learning of a network is implemented;
3. on an input of the trained network standards of the recognized classes sequentially move. For each standard the help of expressions (5, 6) values of output signals of neurons of a layer of summing are calculated. If necessary these values should be scaled. For this purpose it is possible to apply results [2, 4]. The scaled values will also be the expected output signal of a multi-layer perceptron for educational examples of the appropriate class.

Let's consider application of the developed coding procedure on a specific example of recognition of 5 abstract black-and-white figures shown in Figure 3.

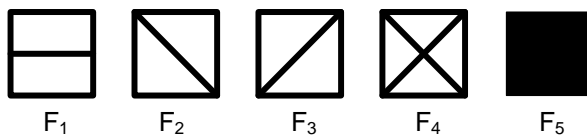


FIGURE 3 Recognized figures

Each figure is written in a square 3x3 i.e. the number of input parameters of NNM is equal to 9. Each figure is accepted in the form of a standard of the recognized class. The set of input and output parameters of educational examples of the specified standards created based on expression (4) is provided in table1. Let's mark that 3 boundaries of squares shown in figure in educational examples aren't considered.

TABLE 1 Parameters of educational examples

Input parameter	Name of reference				
	F ₁	F ₂	F ₃	F ₄	F ₅
x ₁	0	0	1	1	1
x ₂	0	0	0	0	1
x ₃	0	1	0	1	1
x ₄	1	0	0	0	1
x ₅	1	1	1	1	1
x ₆	1	0	0	0	1
x ₇	0	1	0	1	1
x ₈	0	0	0	0	1
x ₉	0	0	9	1	1

The structure of the built PNN network is shown in Figure 4. Accept $\sigma = 0,5$.

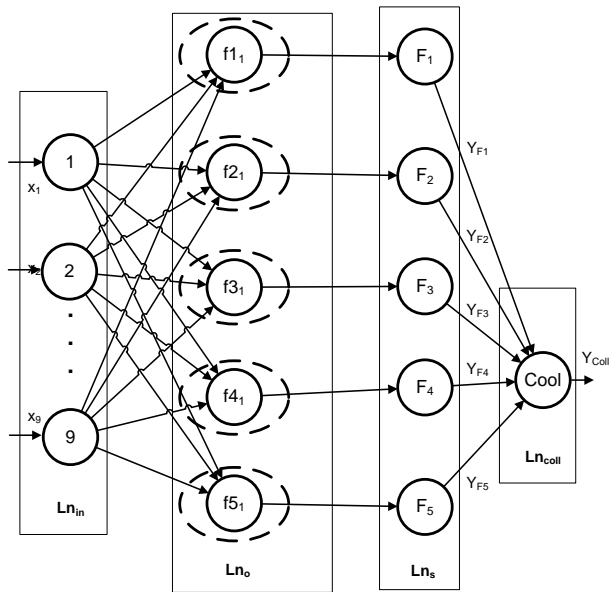


FIGURE 4 Structure of the PNN network for recognition of the abstract figures

Application of the built network allowed is to carry out coding of an output signal for each of standards. So, for example, for F1 figure standard not scaled expected output signal $Y_{F1} = \{9; 5,073263; 5,054947; 3,109894; 3,109894\}$. After scaling the expected values of an output signal are used when forming educational examples for a two-layer perceptron. The made comparative experiments showed that use of such educational examples allows reduce by 30-50% the number of computing iterations concerning examples in which the well-known coding is used. Thus, in a basic case, proved can read prospects of application of the developed coding procedure.

4 Conclusion

It is shown that one of important shortcomings of application of neural networks technology on the basis of a multi-layer perceptron for recognition of graphic images in systems of biometric authentication of users is insufficient quality of processing of statistical data which are used when forming parameters of educational examples.

It is offered to increase quality of educational examples due to use of the procedure of neural network coding of value of the expected output signal of educational examples which allows to consider closeness of standards of the recognized classes in this signal.

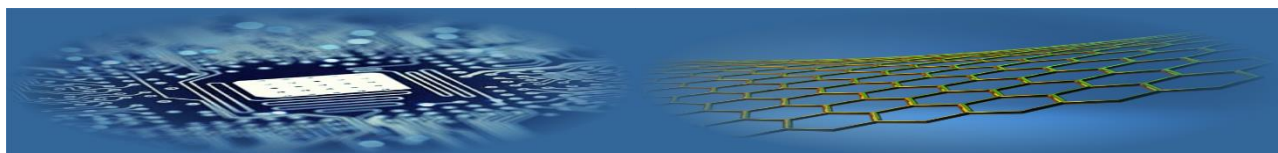
The coding procedure of the expected output signal providing use of a probable neural network is developed. The appropriate mathematical apparatus is created.

As a result of numerical experiments it is shown that application of the developed procedure allows reducing the number of the computing iterations necessary for achievement of the given error of training by 30-50%. It specifies prospects of use of the proposed solutions for increase in learning efficiency of the neural networks intended for recognition of graphic images in systems of biometric authentication.

References

- [1] Korchenko A, Tereykovskiy I, Karpinskiy N, Tynymbayev S 2016 *Neyrosetevye modeli, metody i sredstva otsenki parametrov bezopasnosti Internet-oriyentirovannykh informatsionnykh sistem* TOV NashFormat: Kiev 275 s (in Russian)
- [2] Rudenko O G, Bodyans'kiy E V 2006 *Shtuchni neyronni merezhi* TOV Kompaniya SMIT: Kharkiv 404 s (in Ukrainian)
- [3] Tereykovs'ka L O 2016 *Neyromerezhevi modeli ta metody rozpiznavannya fonem v golosovomu signali v sistemi distantsiynogo navchannya*: dís. kand.tekhn. nauk: 05.13.06 Kyev 312 s. (in Ukrainian)
- [4] Tereykovs'kiy I 2007 *Neyronnimerzhi v zasobakh zakhistu komp'yuternoi informatsii* PoligrafKonsalting: Kyev 209 s (in Ukrainian)
- [5] Tereykovskaya L, Petrov O, Aleksander M 2015 Prospects of neural networks in business models *TransComp. 30 November – 3 December, 2015* Zakopane 1539-45

AUTHORS	
	<p>Lyudmila Tereykovskaya, 20.12.1975, Kuibishevskiy, Kazakhstan</p> <p>Current position, grades: Ph.D., associate professor (KNUCA) University studies: State Academy of Light Industry of Ukraine, Kiev Scientific interest: data mining, development of neural network systems, recognition of voice signals, the construction of distance education systems, recognition of cyber attacks. Publications (number or main): more 50 Experience: more than 10 years of scientific and pedagogical experience</p>
	<p>Igor Tereykovskiy, 20.11.1967, Ternopil, Ukraine</p> <p>Current position, grades: Doctor of Science, professor (NTUU "Igor Sikorsky KPI") University studies: MSc (Master's degree), Specialty – Maintenance of aircraft and engines. Kyiv International University of Civil Aviation (Kyiv, Ukraine) Scientific interest: information security, cyber attacks neural network recognition systems, voice recognition signals Publications (number or main): over 100 Experience: more than 20 years of scientific and pedagogical experience</p>
	<p>Evgeniya Aytkhozhaeva, 1947/02/01, Republic of Kazakhstan</p> <p>Current position, grades: associated professor of the Department of Information Security, Candidate of Technical Sciences University studies: St. Petersburg State Electrotechnical Institute (Technical University "LETI"), Russia Scientific interest: Information Security, Databases Systems, Hardware of Cryptography Publications (number or main): over 160 Experience: more than 30 years of scientific and pedagogical experience</p>
	<p>Sakhybay Tynymbayev, , Republic of Kazakhstan</p> <p>Current position, grades: associated professor of the Department of Information Security, Candidate of Technical Sciences University studies: Bauman Moscow State Technical University, Russia Scientific interest: Information Security, Hardware of Cryptography Publications (number or main): over 150 Experience: more than 45 years of scientific and pedagogical experience</p>
	<p>Imanbayev Azamat, 1992/06/30, Republic of Kazakhstan</p> <p>Current position, grades: tutor of the Department of Information Security University studies: ITMO University, Russia Scientific interest: Information Security Publications (number or main): 4 Experience: 1 year</p>



Word sense disambiguation in Hindi applied to Hindi-English machine translation

S Mall, U C Jaiswal*

Madan Mohan Malaviya University of Technology Gorakhpur, India

**Corresponding author's e-mail: shachimall@gmail.com*

Received 9 March 2017, www.cmnt.lv

Abstract

The Word Sense Disambiguation for Hindi Language is one of the biggest challenges faced by Natural Language Processing. In this paper we discuss issues in reducing ambiguity in Word Sense Disambiguation for Hindi Language. The concepts are induced in two modules Parsing and Word Sense Disambiguation for Hindi Language. Parsing is an extension of our previous work on shallow parser method that creates groups word which are essential for Machine Translation. Monolingual Hindi and English corpora are used. Following this we used machine learning technique such as supervised approach, unsupervised approach and domain specific sense with the help of Knowledge based methods. Knowledge based method uses Hindi and English WordNet tools. Supervised method is used to disambiguate the multiple tags in the context label with the correct tag. Unsupervised method is used to update the sentence with the correct sense and parts of speech tag. There are various websites which provide the facility of translation of Hindi language to English language such as Google Translator and Babefish Translator but these translators fail to resolve polysemy words in Hindi sentences the result is discussed in this paper. The accuracy result of part of speech tagging generated by our system is 92.09%. The accuracy results generated by our system for Chunk are window-3, window 2 and window1 are: 94.45%, 81.23%, and 81.11% respectively. We modify and develop Lesk algorithm which uses WordNet tools for Word Sense Disambiguation. We compare the system's performance with the website Google Translator. We also examine errors made by Google Translator for given input Hindi sentence. Our system generates correct translation with Word Sense Disambiguation for given input Hindi sentence as shown in the Figure12.

Keywords:

Domain specific sense, Word Sense Disambiguation, Morphological analysis, Part of speech tagging and Parsing

1 Introduction

Word Sense Disambiguation in Hindi Language is difficult problem for finding the correct sense of a word in a context, when the word is polysemy. To identify the correct sense for machine translation is very difficult problem. This problem categorized in the field of Natural Language Processing [1]. There are many ongoing approaches is used to resolve Word Sense Disambiguation in machine translation for Hindi Language. We apply Rule based approach as well as machine learning techniques, which performs both unsupervised approach and supervised approach.

The objective is to resolve this ambiguity problem in Word Sense Disambiguation for Hindi language and produce correct translation in English Language for example **उस कबूतर के पर कतर दो** Here Hindi word **पर** has two synsets meaning. Synsets for adjective grammar **पर** means other and Synsets for noun grammar **पर** means wing. To find out Synsets meaning we have used English WordNet [5] and Hindi WordNet [6]. WordNet contain synset set of synonyms and ontological categories. Ontological categories consist of syntactic category like Noun, Pronoun, verb etc. We propose Lesk algorithm that uses WordNet tools for disambiguation has modified.

Development of good quality of machine translation for Hindi to English language using limited resource is challenging task. There are many website available for translation of Hindi language to English language such as Google translator [3] and Babefish Translator [4] but they

are fail to resolve polysemy Hindi word. The output result of Google translator is discussed in Table 1 and Table 2 discussed the result of Babefish Translation from Hindi to English language. Google Translation [3] and Babefish Translation [4] fail to resolve word sense disambiguation. Our developed system discussed in Table 3 resolves word sense disambiguation and produce correct translation from Hindi sentence to English sentence

TABLE 1 Google Translator

S. No.	Google Translator from Hindi language to English language
Input Sentence	उस कबूतर के पर कतर दो
Output Sentence	Two doves on the Qatar

TABLE 2 Babefish Translator

S. No.	Babefish Translator
Input Sentence	उस कबूतर के पर कतर दो
Output Sentence	The pigeon at Qatar two

TABLE 3 Our system generated translation from Hindi language to English language

S. No.	Our system generated translation from Hindi language to English language
Input Sentence	उस कबूतर के पर कतर दो
Output Sentence	The pigeon at Qatar two

The proposed work to develop Machine Translation System is divided into the following modules:

1. Morphological Analyzer

2. Parts of Speech Tagging
3. Chunk
4. Parsing
5. Word Sense Disambiguation
6. Hindi to English Translation

In our previous work [2] we had developed a system for first three modules. In this paper we developed for later three modules. In section 2, we describe different module definition in Section 3. Related work in Section 4. Proposed work in Section 5 Simulation Result and Analysis and Section 6 Conclusion and Future work

2 Preliminaries

Word sense disambiguation (WSD) is a primal problem for different Indian Languages Technology such as Machine Translation. We have developed Word sense disambiguation using parsing method for Hindi language and used this method for Hindi to English Translation. This paper proposed following methods:

2.1 DICTIONARY TO SENSE SEMANTIC CATEGORY

We use Hindi WordNet and English WordNet for disambiguation and translation. WordNet contain dictionary definition for each word and label with unique frequency ids. Disambiguation is performing on sentence by sentence basis. The frequency is manually tagged in domain table. Domain table is corpora contains with list of words and meaning of each word with their domain name. Domain name contain information of words belong to which category in given Hindi WordNet. In the Hindi WordNet word and their meanings are given it consist of following characteristics:

1. Synset
2. Gloss
3. Position in Ontology
4. Hyponymy and Hypernymy

Detail description of Hindi WordNet for Hindi word सोना
word = ".सोना".decode('utf-8', 'ignore')

while True:

if word2Synset.has_key(word):

synsets = word2Synset[word]

print "Word -->", ". सोना "

for pos in synsets.keys():

print "POS Category -->", pos

for synset in synsets[pos]:

print "\t\tSynset -->", synset

if synonyms.has_key(synset):

print "\t\t\tSynonyms -->", synonyms[synset]

if synset2Gloss.has_key(synset):

print "\t\t\tSynset Gloss", synset2Gloss[synset]

if synset2Onto.has_key(synset):

print "\t\t\tOntological Categories", synset2Onto[synset]

if synset2Hypernyms.has_key(synset):

print "\t\t\t\t\tHypernym\t\t\t\t\tSynsets",

synset2Hypernyms[synset]

if synset2Hyponyms.has_key(synset):

print "\t\t\t\t\tHyponym\t\t\t\t\tSynsets",

synset2Hyponyms[synset]

word = raw_input("Enter a word:").decode("utf-8",

"ignore")

2.2 MORPHOLOGICAL ANALYZER

Hindi language is morphology rich and free order in nature. Morphological information is used to constructed basic meaning units called morphemes. We identify the morphological information from tokenize words. The feature structure of Morphological Analyser is given below:
<fsaf = 'root, lcat, gend, num, pers, case, vibh, suff'>

These eight cases are mandatory for the morph 'fs' is feature structure which contains 'af' is a composite attributes consisting of root of the word, Lexical category of the root, Gender of the word, Number corresponding to the word form, Person of the word, Case (Direct / Oblique), case name, Specificity Marker, Emphatic Marker, Dubitative Marker, Interjection Marker, Conjunction Marker, Honorific Marker, Gender of the agreeing noun, Number of the agreeing noun, Person of the agreeing noun. Form of suffix and prefix representing we take input from text file then apply suffix smoothing and prefix smoothing is done for example Identification of prefix token we extract feature of token. Token character are extracted up to character length seven (7) for example consider Hindi token अरे now each character of this token extracted अ1 अर2 अरे3 NULL4 NULL5 NULL6 NULL 7 feature extraction for suffix token characters up to length three (3) अरे रे1 अरे2 NULL3 Total word length is 3, this method is used to identify root word in the given context. The result of feature extraction is discussed in our previous paper [2] Morphological Analyzer example for Hindi token हिन्दी((NP<fsaf='हिन्दी,n,f,sg,3,d,0,0'head='hinxI'>हिन्दीNN P<fsaf='हिन्दी,n,f,sg,3,d,0,0' name='hinxI'>)).

2.3 PARSING

Parsing uncover the hidden structure of Hindi text input it can provides structural description that can identifies the break intonation and analyse a given sentence to determine its syntactical structure according to the part of speech tag and chunk. In natural language processing the syntactic analysis of Hindi language can vary from low level such as Part of speech tagging methodology has been discussed in our previous paper [2] Part of speech tagging is a process of labelling tags to each token with their related parts of speech such as nouns, verbs, adjectives, adverbs etc. to each word in the given sentence. We consider 19 Parts of speech class for Hindi language Table 4 shows the abbreviation classes of Parts of speech. To remove ambiguity in multiple tag for a single word we use Hidden Markov is also called Maximum Likelihood Tagger [5] for Parsing to identify the dependency between each predicate in a given input sentence. We use Viterbi approximation in equation (12) to choose the most probable tag sequence for given input Hindi sentence. To estimate we read off count from the training corpus and then computer the maximum likelihood. Firstly we calculate Transition matrix we have a set of words in a given sentence.

TABLE 4 Abbreviation classes of Parts of speech

S.No.	Symbol	Parts of speech
1	NN	Noun
2	NNS	Noun Plural
3	NST	Noun denoting spatial and temporal expressions
4	NNP	Proper Nouns
5	PRP	Pronoun
6	DEM	Demonstratives
7	VM	Verb Main
8	VAUX	Verb Auxiliary
9	JJ	Adjective
10	RB	Adverb
11	PSP	Postposition
12	RP	Particle
13	CC	Conjuncts
14	WQ	Question Words
15	QF	Quantifiers
16	QC	Cardinals
17	QO	Ordinals
18	SYM	Special Symbol
19	NEG	Negative Words

Figure 1 shows the flow chart of parts of speech tagging with following step

- User input Hindi sentence. The sentence is converts the input file into Shakti Standard Format (SSF) to Trigram to Shakti Standard Format (TnT).
- Build Transition Count Matrix and Build Emission count matrix. Build a hash of the tag sequence and its frequency calculated by equation 1
- N-grams smoothing technique is used as discussed in equation 8,9,10 and 11. The tag sequence of a given word sequence.
- Convert the output generated by part of speech tagger which is in TnT format to SSF format.

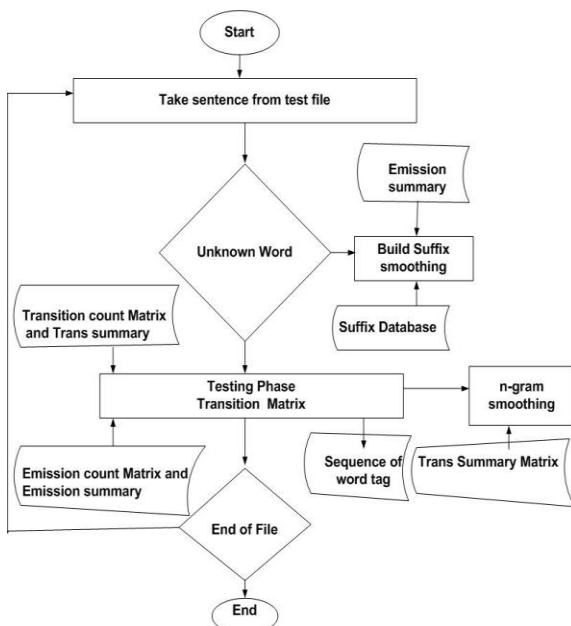


FIGURE 1 Flowchart and Description Process of Part of Speech Tagging

To remove ambiguity in multiple tag for a single word we use Hidden Markov is also called Maximum Likelihood Tagger [5] for Parsing to identify the dependency between each predicate in a given input sentence. We use Viterbi approximation in equation (12) to choose the most probable tag sequence for given input Hindi sentence. To estimate we read off count from the training corpus and then computer the maximum likelihood. Firstly we calculate Transition

matrix we have a set of words in a given sentence $W_1-----W_T$ represents the sequence of the word. T is the probable tag sequence $T = t_1, t_2, \dots, t_n$

$$\hat{T} = \arg \max_{T \in \tau} P(T / W) . \tag{1}$$

Equation (1) is used to choose the sequence of tags that maximizes $\{P(T)P(W / T)\} / P(W)$

$$\hat{T} = \arg \max_{T \in \tau} \{P(T)P(W / T)\} / P(W) . \tag{2}$$

We make use of N-gram method for modelling the probability of word sequences. From the chain rule probability:

$$P(T)P(W/T) = \prod_{i=1}^n P(w_i/w_1t_1 \dots w_{i-1}t_{i-1})P(t_i/w_1t_1 \dots w_{i-1}t_{i-1}) . \tag{3}$$

Thus we are choosing the tag sequence that maximizes:

$$P(t_1)P(t_2/t_1) \prod_{i=3}^n P(t_i/t_{i-2}t_{i-1}) \prod_{i=1}^n P(w_i/t_i) , \tag{4}$$

$$\arg \max \left[\prod_{i=1}^T P(t_i/(t_{i-1}, t_{i-2}))P(w_i/t_i) \right] P(t_{T+1}/t_T) . \tag{5}$$

We use Maximum likelihood estimation from relative frequency to estimate these probabilities:

$$P(t_i/t_{i-2}t_{i-1}) = c(t_i - 2t_{i-1}t_i) / c(t_i - 2t_{i-1}) , \tag{6}$$

$$P(w_i/t_i) = c(w_i, t_i) / c(t_i) . \tag{7}$$

Equation 2, 3, 4 and 5 calculate the probabilities of N-gram smoothing technique. This technique is used to resolve the issues of multiple tags for same word.

$$u \text{ nigrams} = \hat{P}(t_3) = \frac{f(t_3)}{N} , \tag{8}$$

$$B \text{ igras} = \hat{P}(t_3/t_2) = \frac{f(t_2, t_3)}{f(t_2)} , \tag{9}$$

$$T \text{ rigrams} = \hat{P}(t_3/(t_1, t_2)) = \frac{f(t_1, t_2, t_3)}{f(t_1, t_2)} , \tag{10}$$

$$Lexical = \hat{P}(w_3/t_3) = \frac{f(w_3, t_3)}{f(t_3)} . \tag{11}$$

Let us consider an example Input Hindi Text: यह एशिया की सबसे बड़ी मस्जिदों में से एक है ।

The abbreviation of the Parts of speech tag is given in Table 1.Let user input Hindi sentence denoted by X and w is word tokenize from the sentence $X = w_1, w_2, \dots, w_n$.

Let each word in the given Hindi sentences is sequentially label with their corresponding Parts of speech tag $T = T_1, T_2, \dots, T_n$, where $T_i \in T (i \leq i \leq n)$.

Let S is a set of sequence tagging word with related tag

$$\begin{aligned} & (u_1, u_2 \dots u_n, v_1 \dots v_n) \in S \quad P(u_1 \dots u_n, v_1 \dots v_n) \geq 0 \\ & \sum P(u_1 \dots u_n, v_1 \dots v_n) = 1 \quad (u_1 \dots u_n, v_1 \dots v_n) \in S \end{aligned}$$

Hence $(u_1 \dots u_n, v_1 \dots v_n)$ is a probability distribution over a pair of sequence set S. our approach Trigram Hidden Markov consists of a finite set W of possible words, and a finite set T of possible tags, together with the following parameters:

Sequence of pair $(u_1 \dots u_n, v_1 \dots v_{n+1})$ such that $n \geq 0, u_i \in W$ For $i = 1 \dots n, u_i \in T$ For $T_i = 1 \dots N$
 $v_{n+1} = Stop$
 Probability for any sequence

$$\begin{aligned} & (u_1, u_2 \dots u_n, v_1 \dots v_{n+1}) \in S \quad \text{as } P(u_1 \dots u_n, v_1 \dots v_{n+1}) = \\ & \prod_{i=1}^{n+1} q(v_i / v_{i-2}, v_{i-1}) \\ & \prod_{i=1}^n e(u_i / v_i) \end{aligned}$$

where we assume u is a sentence and v is a tag

$$\begin{aligned} v_0 &= v_{-1} \\ n &= 11 \\ u_i &\dots u_{11} \end{aligned}$$

यह एशिया की सबसे बड़ी मस्जिदों में से एक है ।
 $q \langle JJ \rangle * \langle NP \rangle * \langle PSP \rangle * \langle QF \rangle * \langle JJ \rangle * \langle NNP \rangle * \langle PRP \rangle * \langle PSP \rangle * \langle QC \rangle * \langle VM \rangle * \langle SYM \rangle$ Here we use second order Markov Model
 एशिया $\langle NP \rangle$ की $\langle PSP \rangle$ सबसे $\langle QF \rangle$ बड़ी $\langle JJ \rangle$ मस्जिदों $\langle NNP \rangle$ में $\langle PRP \rangle$ से $\langle PSP \rangle$ एक $\langle QC \rangle$ है $\langle VM \rangle$ । $\langle SYM \rangle$ let a set of sentence $u_1 \dots u_n$ and paired with sequence of tag $v_1 \dots v_n$.

Define $l(a, b, c, d, e, f, g, h, i, j, k)$ to be the number of times to the sequence of 11 state is seen in training data $\langle JJ, NP, PSP, QF, JJ, NNP, PRP, PSP, QC, VM, SYM \rangle$. Similarly, define $l(a; b)$ to be the number of times the tag bigram (a; b) is seen. Define $l(s)$ to be the number of times that the state S is seen in the corpus can be interpreted as conditional probability where u is the sentence, यह एशिया की सबसे बड़ी मस्जिदों में से एक है । and v is a tag Maximum likelihood estimate are $q(S/a, b) = l(a, b, \dots, n) / l(a, b)$, where a, b, ... n are number of words in the given sentence.

$$e(u/S) = c(s \rightarrow u) / c(S),$$

where $c(s \rightarrow u)$ = number of time state S is seen paired with observation in the corpus $c(v \rightarrow यह)$ would be the number of times the word यह is seen paired with the tag v

$$q(JJ, NP, PSP, QF, JJ, NNP, PRP, PSP, QC, VM, SYM) = c(JJ, NP, PSP, QF, JJ, NNP, PRP, PSP, QC, VM, SYM) / c(JJ, NP, PSP, QF, JJ, NNP, PRP, PSP, QC, VM, SYM)$$

$$e(यह / v) = c(v \rightarrow 2Tag) / c(v)$$

To estimate we read off count from the training corpus and then computer the maximum likelihood. Firstly we calculate Transition matrix we have a set of words in a given sentence.

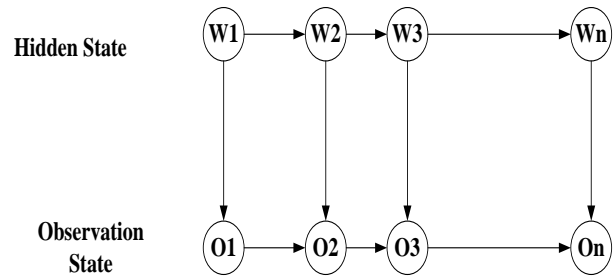


FIGURE 2 Transition matrix method

एशिया $\langle NP \rangle$ की $\langle PSP \rangle$ सबसे $\langle QF \rangle$ बड़ी $\langle JJ \rangle$ मस्जिदों $\langle NNP \rangle$ में $\langle PRP \rangle$ से $\langle PSP \rangle$ एक $\langle QC \rangle$ है $\langle VM \rangle$ । $\langle SYM \rangle$

The process starts from one word to another word. Each move is called a step. If the chain is currently in state X i then it moves to state X i at the next step with a probability denoted by P i,j, and this probability does not depend upon which states the chain was in before the current state this method is known as Transition matrix Figure 2 elaborate the method to calculate Transition matrix $Q_i = [W_i = i]$. Suppose we have N-state Hidden Markov Model parameterized by (E, Q, R) where emission probability represents E, Q is an initial probability and transition probability matrix represent R. Let rows of R identical and given by vector r, the joint probability of the hidden states and observations over a sequence of length O can be $O = O1, O2, O3 \dots On$ Calculated as:

$$\begin{aligned} & Z(U, V) / E, Q, R) \\ & = Z u_1 / Q \prod_{o=2}^O Z(u_i / R(u_{i-1}, :)) Z(v_o / u_o, E) \\ & = Z(u_1 Q \prod_{o=2}^O Z(u_o, r) Z(v_o / u_o, E) \quad (12) \end{aligned}$$

Maximum likelihood can calculate the sequence $\lambda = (W_{i,j}, O_{i,j}, Q_i)$
 Output POS Tag: यह $\langle JJ \rangle$ एशिया $\langle NP \rangle$ की $\langle PSP \rangle$ सबसे $\langle QF \rangle$ बड़ी $\langle JJ \rangle$ मस्जिदों $\langle NNP \rangle$ में $\langle PRP \rangle$ से $\langle PSP \rangle$ एक $\langle QC \rangle$ है $\langle VM \rangle$ । $\langle SYM \rangle$

Through the above calculation we find tag for other words in a given sentence and input for the process of Chunk. Figure 11 shows the snapshot of Parsing with Parts of speech tagging for given Hindi sentence. Chunking [3] is an important process to identifying and segmenting the text into syntactically correlated chunk tag such as is NP chunk label the word in the sentence start with different Phrases, we label the word with boundary marker B represents - Beginning phrase and I represent as Inside phrase for example we input Hindi sentence: दफ्तर के सभी लोग अपनेअपने घरों को जाने की जल्दी में थे।

दफ्तर NN B-NP के PSP I-NP सभी QF B-NP लोग NN I-NP अपने PRP B-NP SYM I-NP अपने RDP I-NP घरों NN B-NP

The sentence is individually tokenize by the delimiter “?” in sentence start with $\langle Sentence\ id=? \rangle$ chunk start with assigning chunk number “((chunk phrase=Hindi word, the features of the word and chunk Table 5 shows the abbreviation of chunk symbols. Chunk is an arbitrating step towards parsing. In Figure 3 Head computation is used for functional specification to

compute the phrase with heads of different phrases of groups such as noun, verb groups etc. Chunk head provides the sufficient information for further processing of the sentence. Figure 7 shows the output result of Hindi token label with related parts of speech tagging and chunk.

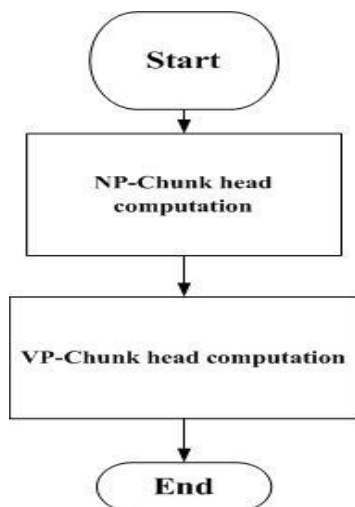


FIGURE 3 Chunk computation

2.4 WORD SENSE DISAMBIGUATION

Hindi words have more than one meaning in the context. for example Hindi word 'सोना' hold English meaning gold and सोना hold English meaning sleep we have use English WordNet synset and Hindi WordNet synset. Synset contains list of synonyms. Hindi WordNet is an ontological category. Ontological categories are coarse grained distinctions of word senses. Our methods for word sense disambiguation are following

- Unsupervised approach: Unsupervised method used to find the correct word in polysemy word and update the sentence.
- Supervised approach: Supervised method used to sense from a sense tagged corpora
- Domain Information for Sense Disambiguation: We use WordNet for domain information.
- Modified Lesk algorithm: Modified Lesk algorithm used for the target word's synset which has maximum overlap of its gloss, its hypernymy gloss and its hyponymy gloss with the words in the context of target word is chosen as the sense of the word

Ambiguity can be resolved with syntactic information. Word Sense ambiguities [3] disambiguate the senses of word with the meaning of multi-sense words using Distributed Domain approach by analyzing the context in which sense the multi-sense words and produce correct output. In Hindi language contains multi-sense words in its corpus. This is based on supervised and unsupervised approach [4]. Supervised approach [5] is used to identify the correct meanings in multi-sense words in Hindi languages. If the database does not contain sufficient information then it cannot sense the ambiguous word. Unsupervised approaches use dictionary for learning and shows the correct result and update the database. This can be done by using WordNet [6,7,8,9] as lexical database resource to identify the correct meanings in multi-sense words in Hindi languages. Domain specific Modified Lesk's algorithm [10]

is based on dictionary definitions which are also called glosses. This technique is a resource as a key factor to word senses in corpora. Where sense disambiguation is performed using the overlap between the contexts in which a word appears in the discourse.

2.5 HINDI TO ENGLISH TRANSLATION

We find a mapping to an English word for each sense in Hindi and predict the translation of a polysemy Hindi word in an English context. A mapping is available between Syset identifier in English Wordnet and Hindi sense dictionary.

3 Related work

A number of Indian researchers have carried out their work related to machine translation for Indian languages. Little work has been carried out in Hindi language. Sense disambiguation is done by using dictionary method [11] without using prior annotated data. New linear time algorithm for lexical chaining [12] is used for word sense disambiguation. Sense dependency and selective dependency [13] using this combination word sense disambiguation problem was resolved. Various technique have been proposed by researchers to resolve the issues like Unsupervised Weighted Graph [14] the result was 54.9 % in sensaval-3 and 60.2 % in sensaval-2 dataset. Dependency parsing approach is a dynamic programming based algorithm [15]. Knowledge based [16] is a multilingual joint approach for Word Sense Disambiguation. Graph based Word Sense Disambiguation is used to improve the performance of both monolingual Degree and PLength, and compete with the state of the art on all disambiguation tasks. [17] Word Sense Disambiguation use dominant senses of words in specified domains. The accuracy values of approximately 65%. They presented the methodology for Word Sense Disambiguation based on domain 23information [18]. The drawback of the algorithm is that it can disambiguate a word provided it has only one sense per domain. The work on English French Cross-lingual Word Sense Disambiguation [19] French to English translation the target English word depending on the context by identifying the nearest neighbours of the test sentence from the training data using a pairwise similarity measure. The performance of the system was less than the baseline by around 3% it outperformed the baseline system for 12 out of the 20 nouns [20].

4 Proposed work

This paper is extension of my previous paper [2] in which morphological analysers and parts of speech tagging is completed Using these modules we develop we developed algorithm for Parsing and Word Sense Disambiguation which are as below:

4.1 PARSING

Algorithm 1 Parsing

1. for $i \leftarrow 0$ to length words
2. do
3. for each word is Chunk with Noun phrase then
4. Select parent head word "B"
5. Select part of speech H

6. Select voice of H
7. Select position of H (left, right)
8. Else if word is a verb then
9. Select nearest word N to the left word such that word is the parent head word of "I"
10. Select nearest word r to the right of word such that word is the parent head word of r
11. Select part of speech of l
12. Select part of speech r
13. Select the part of speech word
14. Select voice of word
15. Else if word is adjective then
16. Select parent head word head
17. Select part of speech of head
18. end

Parsing is used to estimate the number of useful probability concerning and its syntactical structure of the sentence the method is explained in section 2 parsing. In the parsing algorithm we develop some identification rule are as follows:

- In case of NN most of the time ambiguity is in case marking (direct, oblique). We can decide the case on the basis of following PSP.
1. Rule 1: If NN is just followed by PSP, then we will take only the feature structures having oblique case. Else we will take the direct case.
 - In case of JJ the case (d/o), should agree with the noun it is modifying.
 2. Rule 2: If JJ has multiple morph analysis then we will look for noun it is modifying and we will take the morph analysis of JJ having case marked same as that of modified noun and eliminate the rest.
 - In case of PSP the pruning module is giving multiple morph analysis for 'ke' and 'ki'.
 3. Rule 3: We will look for the noun to which our PSP is related and will keep the morph analysis having gender and case agreeing with the gender and case of the noun to which our PSP is related and eliminate the rest.

Most of the time the noun is related with PSP is found in the next chunk to chunk containing PSP. Then most probably head of the chunk is NP.

4.2 WORD SENSE DISAMBIGUATION

Algorithm 2 Modified Lesk Algorithm

Input: Text with only meaningful words

Output: Actual sense of ambiguous words

1. Loop Start for all dictionary definition of the ambiguous word
2. Ambiguous word is selected
3. Each word is selected from preliminary input texts.
4. Gloss of ambiguous word is obtained from typical WordNet.
5. Intersection is performed between the meaningful words from the input text and the glosses of the ambiguous word.
6. Loop End
7. If the counter value is mismatched with all other values, then associated sense is considered as the disambiguated sense.
8. Else, Bag-of-Words fails to disambiguate the sense.
9. If occurrence of an unmatched word in anticipated database having a particular sense crosses the

threshold value, then the word is moved to the related bag of words database.
10. Stop.

4.1.1 Definitions

Let window of context is, $2t+ 1$ with the grammar R. Were list of the word in WordNet is define $T_i, 1 \leq i \leq R$. compare the list of in the WordNet is less than $2t + 1$, if all the list of words in WordNet belong to the context. Were T_i is list of words contain more than two meaning in the gloss, list of words are assign with unique synset having a unique sense tag. Were T_i is the lists of sense tag are represented by $|T_i|$. We evaluate sense tag for each pair of words in the context of the window. Were $R = \sum |T_i|$ represents combinations of words this is referred as candidate combination.

4.1.2 Process

When user input the sentence then each word are tokenize and label with grammatical tag. If the tag is multiple then it depends on tokens of the input sentence. Words which is polysemy is labeled with multiple tag their related parts of speech here we use N gram technique. We divide the context in three windows window 1, window 2 and window 3. Window 1 applies unigram method take only right word next to the polysemy word and finds the candidate combination as given in equation 2. Window 2 applies bigram method take only left word next to the polysemy word and finds the candidate combination as given in equation 3. Window 3 applies unigram method take right and left both words next to the polysemy word and finds the candidate combination as given in equation 3. we use overlap technique discussed in the flowchart of Figure 3. Overlap technique find overlap between pair of word with polysemy word in each window. We map each combination with gloss if the combination scores high then we consider that word given in the gloss with their related tag and update the list with the help of unsupervised approach.

For example Input Hindi sentence: उस कबूतर के पर क़तर दो.

Two glosses can have more than one overlap where each overlap covers as many words as possible. Each gloss compares with pairs of words divide the sentence in three size window. Each window has pair of relation we find individual score by comparing the combination score with particular candidate as shown in the Table 4. After comparing highest score window will winner. Winner word will be chosen as correct sense for given polysemy word पर Hindi sentence. The target word, are specified by WordNet. The window of the sentence would be पर क़तर दो and उस ,के token are separated as shown in Table 5, Table 6, and Table 8. Table 8 shows all word from Table 5, Table 5 and Table 7 possible pair Finally, two senses of the keyword "पर" have their counter readings (refer Table 5) as follows:

$$P \text{ counter, } PC = E' + F' + U' + S'$$

$$Y \text{ counter, } YC = E'' + F'' + U'' + S$$

TABLE 5 Sense for Token पर

Keyword	sense
पर	P
	Q

TABLE 6 Sense for token कतर

Context Word	sense
कतर	E
	F

TABLE 7 Sense for token दो

Context Word	Likely sense
दो	U
	S

TABLE 8 Candidate pairs of the word in the given sentences

Candidate pairs	General word in the sentence
P and A	E
P and B	F'
Q and A	E''
Q and B	F''
P and U	U'
P and S	S'
Q and U	U''
Q and S	S''
P and E	E'
P and F	F'
Q and E	E''

The Hindi word पर meaning is nothing more in a sentence and other meaning of पर is wing of a bird. The list of words is stored and their meaning is stored in the list of words can find the correct sense of a word having different meaning due to different contexts. The list of words sense the disambiguate word which is considered as keyword. The general words are separated from the sentence only keyword and context word is compared with each word of each "sense" list of words searching for the maximum frequency of words in common. The above algorithm is based on the learning set. In initial stage, if word is not present in the learning set, then it will not participate for disambiguation. Though, its probable meaning would be stored in the database. When the number of occurrences of the particular word with a particular sense crosses specific threshold value, the word is inserted in the learning set to take part in disambiguation procedure. Therefore, the efficiency of the disambiguation process is increased by this auto increment property of the learning set. Output Hindi sentence translated in English sentence: Slice/cut/chop of the wings

TABLE 9 Precision, Recall and F-score for parts of speech tagging

Abréviation of parts of speech	Precision %	Recall%	F-Score	Accuracy
CC	95.75	98.235294	97.909091	94.33333333
DEM	63.157895	92.307692	75	60
INJ	80	36.363636	50	33.33333333
JJ	69.230769	66.176471	67.669173	51.13636364
NEG	100	100	100	100
NN	76.352705	96.455696	85.234899	74.26900585
NNP	100	27.272727	42.857143	27.27272727
NST	100	100	100	100
PRP	91.891892	80	85.534591	74.72527473
PSP	97.154472	95.6	96.370968	92.99610895
QC	85.714286	100	92.307692	85.71428571
QF	75	81.818182	78.26087	64.28571429
RB	100	42.857143	60	42.85714286
RP	84.090909	88.095238	86.046512	75.51020408
SYM	100	100	100	100
VAUX	91.715976	85.635359	88.571429	79.48717949
VM	80.269058	85.238095	82.678984	70.47244094
WQ	89.473684	89.473684	89.473684	80.95238095
OVERALL SYSTEM	89.655647	92.862734	91.717502	94.97284249

of that pigeon.

4.2 HINDI TO ENGLISH TRANSLATION

Algorithm 3 Translation of Hindi sentence to English sentence

Input: Hindi word tagged with corresponding English word
 Output: Translation of Hindi word in English word in the context.

1. Each word in the context is in the root form.
2. Extract each word from the sentence.
3. Generate unique ids to each Hindi word and store in the database
4. Map each Hindi word with the English WordNet dictionary.
5. Return the label of the selected English translation of the target Hindi word in the context.
6. Stop.

For translation of Hindi sentence into English sentence we use the concept of mapping between English WordNet and Hindi sense dictionary. The given Hindi word is searched in the Hindi sense dictionary for synset frequency ids of all the synset. We use these sense ids to query the English WordNet to label all the synset. The set of all English words are mapped with the unique ids of the Hindi words. The English words contained in the English synset ids for translation of Hindi words to English words.

5 Simulation result and analysis

The simulations have been carried out using Python language to obtain the accuracy results of parts of speech tagging and Chunk. Chunking is a method used for parsing the Hindi sentences. The evaluation result of Precision, Recall and F-score for parts of speech tagging is discussed in Table 9 and Chunk evaluation results is discussed in Table 10, 11 and 12. The output result of system generated parts of speech tag and Chunk are compared with Gold Standard parts of speech tag and Chunk [2]. Gold standard contains correct output of the parts of speech tag and Chunk for the given words. The total Hindi token was 1657 tokens with 990 phrases. Label each token with related parts of speech.

5.1 PARTS OF SPEECH TAGGER

This is simulation result of parts of speech tagger. The data set value is taken for 1657 tokens within 990 phrases and found correct parts of speech tag for 1024 tokens within 919 phrases. Figure 4 shows the graph plot for Parts of speech tagger with the help of evaluation results of Precision, Recall, F score and accuracy data is as below:

- Precision 'P' = Correct POS / (Correct POS + False POS)
- Recall 'R' = Correct POS / (Correct POS + False POS)
- F- Score or measure it evaluate the test accuracy by computing the value of both the precision P and the recall R:
- F-score 'FB1' = 2 *RP/(R+P)
- Accuracy 'A' = R*P/(R+P+F)

Results for each part of speech were calculated in confusion matrix. Confusion matrix is shown in Figure 5 shows the correct match by system generated with Gold standard. The comparison result of system generated with Gold standard part of speech is given below:

TP: 1408 Counter({'NN': 381, 'PSP': 239, 'VM': 179, 'SYM': 166, 'VAUX': 155, 'PRP': 68, 'JJ': 45, 'RP': 37, 'CC': 30, 'DEM': 24, 'NNP': 18, 'WQ': 17, 'NST': 15, 'NEG': 12, 'QF': 9, 'QC': 6, 'INJ': 4, 'RB': 3, 'QO': 0, 'INJC': 0, 'RDP': 0, 'QCC': 0, 'NNPC': 0, 'RBC': 0, 'VMC': 0, 'JJC': 0, 'ECH': 0, 'PRPC': 0, 'NNC': 0})

FN:248 Counter({'NN': 118, 'VM': 44, 'JJ': 20, 'VAUX': 14, 'DEM': 14, 'PSP': 7, 'RP': 7, 'PRP': 6, 'NNC': 6, 'QF': 3, 'NNPC': 3, 'CC': 2, 'WQ': 2, 'INJ': 1, 'QC': 1, 'NNP': 0, 'QO': 0, 'NEG': 0, 'RB': 0, 'INJC': 0, 'RDP': 0, 'NST': 0, 'QCC': 0,

'RBC': 0, 'VMC': 0, 'JJC': 0, 'ECH': 0, 'PRPC': 0, 'SYM': 0})
 FP: 248 Counter({'NNP': 48, 'VM': 31, 'NNC': 29, 'VAUX': 26, 'JJ': 23, 'PRP': 17, 'NN': 14, 'PSP': 11, 'INJ': 7, 'RDP': 6, 'RP': 5, 'RB': 4, 'CC': 4, 'NNPC': 4, 'INJC': 3, 'QO': 2, 'QF': 2, 'WQ': 2, 'DEM': 2, 'VMC': 'NEG': 0, 'NST': 0, 'QC': 0, 'SYM': 0})Accuracy: 92.09%; precision: 84.76%; recall: 89.29%; F-score: 86.97%.

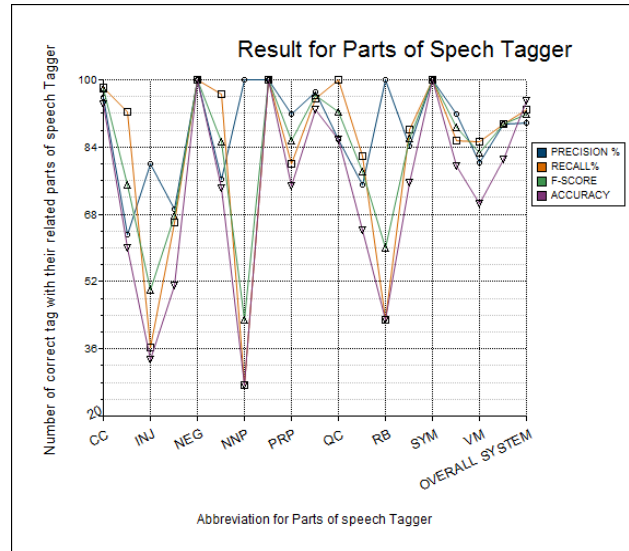


FIGURE 4 Precision, Recall, F-score & Accuracy for parts of speech tagger

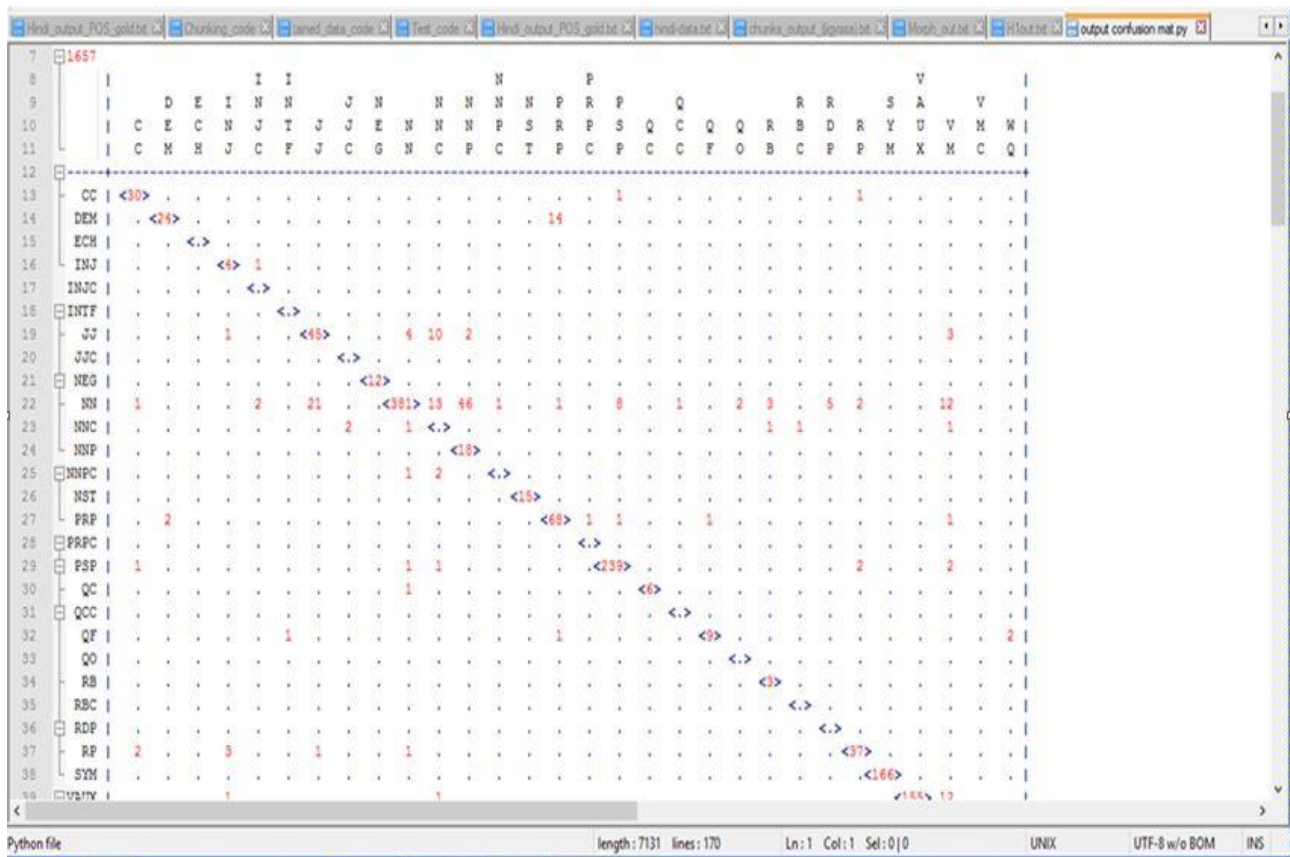


FIGURE 5 Precision, Recall, F-score & Accuracy for parts of speech tagger evaluated through confusion matrix

5.2 CHUNK

We created confusion matrix for Chunk as same way as we created for parts of speech tag. Figure 6 shows the graph and Figure 7 shows the output snapshot of chunk from given input Hindi sentence. The evaluation results for chunk is calculated for Precision, Recall, F-score and Accuracy for chunk are divided into three windows such as window 1, window 2, and window 3 in which we calculate Precision, Recall, F-score and Accuracy for each window. The data is as below:

TABLE 10 Window 1

	PRECISION	RECALL	FB1	ACCURAC
BLK	95.97	98.62	97.28	
CCP	100	100	100	
FRAGP	0	0	0	
JJP	62.96	58.62	60.71	
NEGP	50	100	66.67	
NP	91.51	95.65	93.53	
RBP	81.25	86.67	83.87	
VGf	87.5	97.47	92.22	
VGNF	78.57	61.11	68.75	
VGNN	46.67	43.75	45.16	
OVERALL SY	89.75	92.83	91.26	94.45

TABLE 11 Window 2

	PRECISION	RECALL	FB1	ACCURAC
BLK	95.97	98.62	97.28	
CCP	100	100	100	
FRAGP	0	0	0	
JJP	65.52	65.52	65.52	
NEGP	100	100	100	
NP	92.63	95.65	94.12	
RBP	76.47	86.67	81.25	
VGf	90.06	97.47	93.62	
VGNF	83.33	69.44	75.76	
VGNN	66.67	62.5	64.52	
OVERALL SY	91.24	93.64	92.42	95.17

TABLE 12 WINDOW 3

	PRECISION	RECALL	FB1	ACCURAC
BLK	95.33	98.62	96.95	
CCP	100	100	100	
FRAGP	0	0	0	
JJP	65.52	65.52	65.52	
NEGP	100	100	100	
NP	91.96	95.29	93.59	
RBP	81.25	86.67	83.87	
VGf	89.53	97.47	93.33	
VGNF	80	66.67	72.73	
VGNN	64.29	56.25	60	
OVERALL SY	90.67	93.23	91.93	94.75

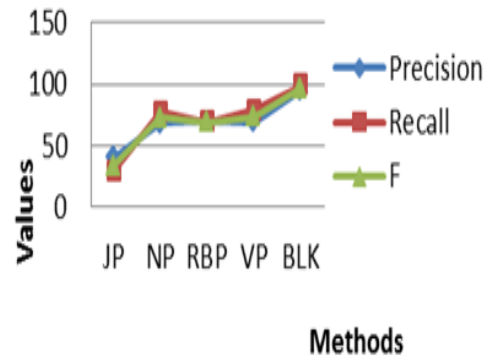


FIGURE 6 Accuracy result for Chunk for Window 1, 2 and 3

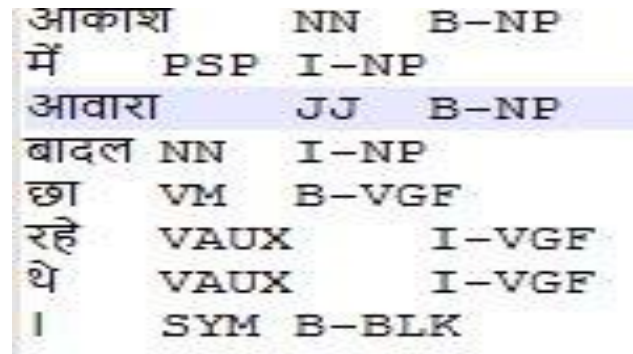


FIGURE 7 Chunk for Hindi sentence

5.3 WORD SENSE DISAMBIGUATION

For word sense disambiguation and translation we compare our system generated output of Hindi sentence to English sentence translation with translating website Google translator [21] as shown in the Figure 9 we input Hindi sentence तुम्हारे लिए मेरा दर खुला रहेगा। Here word दर in the given Hindi sentence is polysemy word which has two meaning Rate and Door. We compare our output result with translating website Google translator we input same Hindi sentence as shown in Figure 9. Here Hindi word दर is translated as Rate in the English sentence but correct translation is Door for the given sentence. Figure 8 shows that Google translator is failed to translate correctly but our system generates correct translation for the given Hindi word दर is translated as Door in the English sentence as shown in the Figure 12. We have input 100 Hindi sentence with polysemy words The result accuracy WSD systems generated output is compared with gold standards created by human annotators we have develop simulation method in which three candidate files in which first file collect the output result of system generated WSD as shown in Figure 12, second gold file which contains correct translation with WSD for given Hindi input sentence and the third file contain the output of Google Translator. Figure 8 shows the comparison graph with the system generated file and Google file with gold file to calculate the accuracy of WSD.

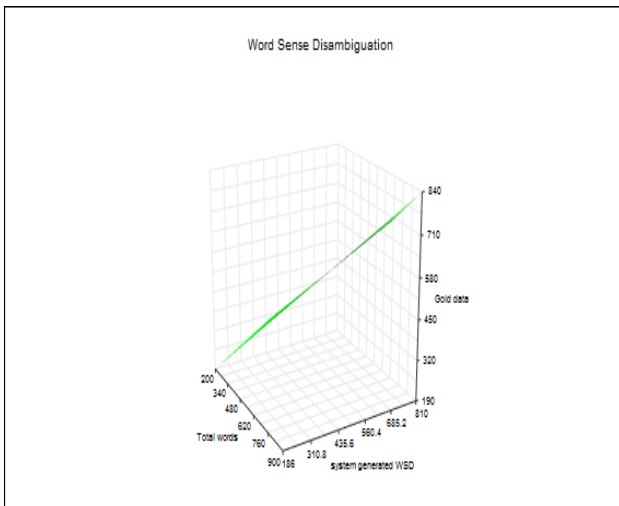


FIGURE 8 WSD systems generated output is compared with gold standards



FIGURE.9 Google output of Hindi to English Translation

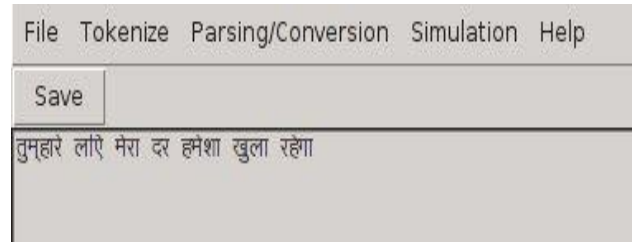


FIGURE 10 Input Hindi sentence with polysemy word दर

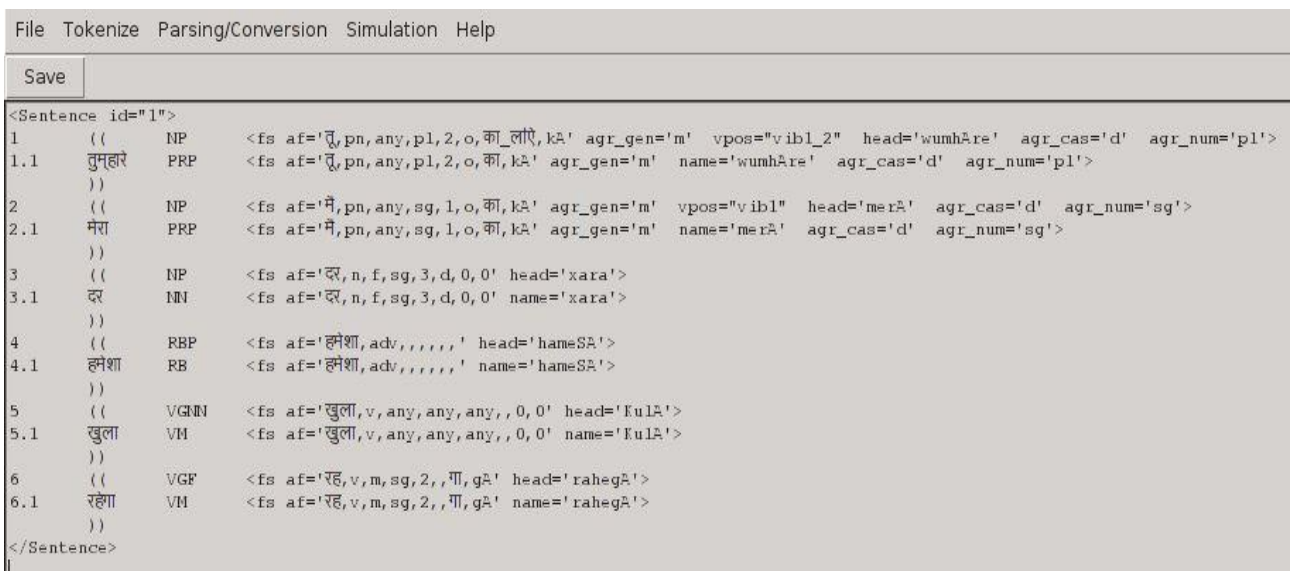


FIGURE 11 snapshot of Parsing for given Hindi sentence

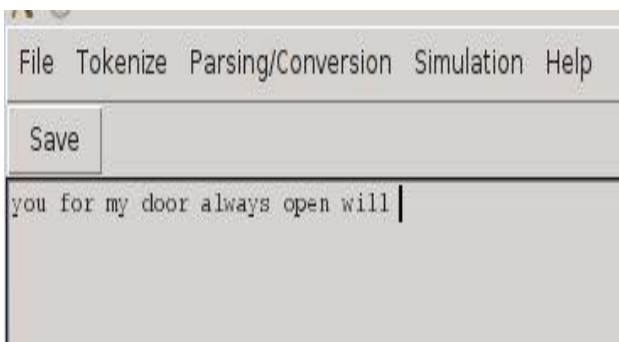


FIGURE 12 system generated output snapshot of Hindi to English Translation

6 Conclusion and future work

This work is carried out to evaluate how parsing is used for machine translation for Hindi language to English language. Figure 11 shows the output snapshot of Hindi sentence is parsed. No previous attempts have been reported in



literature, which analyze the Hindi Language translation into English translation for Indian language. This work confirms that the Parts of speech tagging algorithm obtain 92.09% accuracy result. The accuracy results for chunk are evaluated for three windows. Window-3, window 2 and window1 are: 94.45%, 81.23%, and 81.11% respectively. We enhance the Modified Lesk algorithms in which overlap is finding between three pieces of words in a given context to find the correct word sense by counting word overlaps between glosses of the words in the context. All the glosses of the key word are compared with the glosses of other words. The sense for which the maximum number of overlaps occur, represents the desired sense of the of the polysemy word. We use Hindi and English WordNet which is used in lexical knowledge. The Modified Lesk algorithm improves word sense disambiguation and the system generated result as shown in figure 12 is compared with Google translator website as shown in the Figure 9. This work shows that Google Translator cannot handled word sense disambiguation but our system can resolve word sense

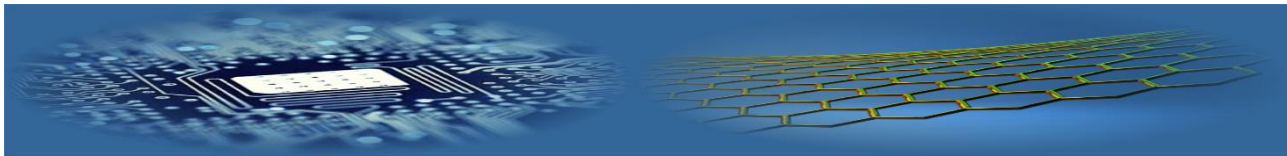
disambiguation. We compare our system generated output with available translating website such as Google Translator. The result accuracy WSD systems is compared with gold standards created by human annotators we have develop simulation method in which three candidate files in which first file collect the output result of system generated WSD, second gold file which contains correct translation with WSD for given Hindi input sentence and the third file contain the output of Google Translator. We compare the

system generated file and Google file with gold file to calculate the accuracy of WSD. Our system can resolves word sense disambiguation and generate translation for each Hindi polysemy word in the given sentence without word alignment. This work can be further extracted by resolving the issues of language in which subject object and verb appear. Hindi language is subject verb object and English language is subject object verb.

References

- [1] Chan Yee Seng, HweeTou Ng, David Chiang 2007 Word sense disambiguation improves statistical machine translation *In Annual Meeting-Association for Computational Linguistics* 45(1) 33
- [2] Mall Shachi, Umesh Chandra Jaiswal 2016 Evaluation for POS tagger, chunk and resolving issues in word sense disambiguate in machine translation for Hindi to English languages *In Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on*, pp. 14-18.IEEE
- [3] Radu F, Cucerzan S, Schafer C, Yarowsky D 2002 Combining classifiers for word sense disambiguation *Natural Language Engineering* 8(04) 327-41
- [4] Eneko A, Edmonds P, eds. 2007 Word sense disambiguation: Algorithms and applications 33 Springer Science & Business Media
- [5] Mallapragada Pavan Kumar, Rong Jin, Anil K. Jain, Yi Liu 2009 Semiboost: Boosting for semi-supervised learning *IEEE transactions on pattern analysis and machine intelligence* 31(11): 2000-2014
- [6] Redkar Hanumant Harichandra, Sudha Baban Bhingardive, Diptesh Kanojia, Pushpak Bhattacharyya 2015 World WordNet Database Structure: An Efficient Schema for Storing Information of WordNets of the World *In AAAI* 4290-1
- [7] F Radu, S Cucerzan, C Schafer, D Yarowsky 2002 Combining classifiers for word sense disambiguation *Natural Language Engineering* 8(04) 327-41
- [8] Lesk M 1986 Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone *In Proceedings of the 5th annual international conference on Systems documentation ACM* 24-6
- [9] Fellbaum C 1998 A semantic network of English verbs *WordNet: An electronic lexical database* 3 153-78
- [10] Resnik P 1995 Using information content to evaluate semantic similarity in a taxonomy *arXiv preprint cmp-lg/9511007*
- [11] Gaume B, Nabil Hathout, P Muller 2004 Word sense disambiguation using a dictionary for sense similarity measure *In Proceedings of the 20th international conference on Computational Linguistics* 1194 Association for Computational Linguistics
- [12] Galley M, McKeown K 2003 Improving word sense disambiguation in lexical chaining *In IJCAI* 3 1486-8
- [13] Chaplot Devendra Singh, Pushpak Bhattacharyya, Ashwin Paranjape 2015 Unsupervised Word Sense Disambiguation Using Markov Random Field and Dependency Parser *In AAAI* 2217-23
- [14] Hessami Ehsan, Faribourz Mahmoudi, Amir Hossien Jadidinejad 2011 Unsupervised Weighted Graph for Word Sense Disambiguation *In 2011 World Congress on Information and Communication Technologies*
- [15] Li Zhenghua, Min Zhang, Wanxiang Che, Ting Liu, Wenliang Chen, Haizhou Li 2011 Joint models for Chinese POS tagging and dependency parsing *In Proceedings of the Conference on Empirical Methods in Natural Language Processing* 1180-91 Association for Computational Linguistics
- [16] Navigli R, Ponzetto S P 2012 Joining forces pays off: Multilingual joint word sense disambiguation *In Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning* 1399-410 Association for Computational Linguistics
- [17] Khapra Mitesh, Pushpak Bhattacharyya, Shashank Chauhan, Soumya Nair, Aditya Sharma 2008 Domain specific iterative word sense disambiguation in a multilingual setting *In Proceedings of International Conference on NLP (ICON 2008), Pune, India*
- [18] Kolte Sopan Govind, Sunil G Bhirud 2008 Word sense disambiguation using wordnet domains *In 2008 First International Conference on Emerging Trends in Engineering and Technology* 1187-91 IEEE
- [19] Mahapatra Lipta, Meera Mohan, Mitesh M Khapra, Pushpak Bhattacharyya 2010 OWNS: Cross-lingual word sense disambiguation using weighted overlap counts and wordnet based similarity measures *In Proceedings of the 5th International Workshop on Semantic Evaluation* 138-41 Association for Computational Linguistics
- [20] Sawhney Radhike, Arvinder Kaur 2014 A modified technique for Word Sense Disambiguation using Lesk algorithm in Hindi language *In Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference on 2745-9 IEEE*
- [21] <https://translate.google.co.in/?hl=en>

AUTHORS	
	<p>Shachi Mall, 23-01-1986, India</p> <p>Current position, grades: Teaching Cum Research Scholar fellowship PhD[(CSE) [Thesis Submitted] M.M.M.U.T., Gorakhpur, M Tech(CSE) M.M.M.E.C., Gorakhpur, B. Tech(CSE)I.T.M., Gorakhpur Scientific interest: Natural Language Processing Publications: 12 Experience: 06</p>
	<p>Umesh Chandra Jaiswal, 01-06-1967, India</p> <p>Current position, grades: Associate Professor University studies: PhD(CSE),M Tech(CSE) IITD, BE(CE) M.M.M.E.C., Gorakhpur Scientific interest: Natural Language Processing, Design and Analysis of Algorithms, Operating Systems, Computer Networks Publications: 20 Experience: 25</p>



Ant colony optimization algorithm: advantages, applications and challenges

Kavita Tewani

Institute of Technology and Management Universe, Vadodara Gujarat, 391510

Corresponding author's e-mail: kavitatewani012@gmail.com

Received: 19 April 2017, www.ckmnt.lv

Abstract

Ant Colony optimization is a technique for optimization that was introduced in early 1990's. ACO algorithm models the behaviour of real ant colonies in establishing the shortest path between food sources and nests and this technique is applied on number of combinatorial optimization problem, communication networks and robotics. This paper introduces the advantages of using the ACO algorithms with the help of some problem examples and the challenges faced for solving the problems. Initially, the paper discusses about the biological inspiration and behaviour of ant colony and then relates with the real life problems.

Keywords:

Ant colony optimization (ACO),
pheromone,
Travelling Salesman Problem (TSP)

1 Introduction

In the real world, ants are able to find the shortest path from the source to the destination (food) without using any visual cues. While moving from source to destination the ants deposit a chemical called pheromone on their way, and each ant prefers to follow the direction which is rich in pheromone. By using this behaviour of ants, the researchers have related the meta heuristic algorithm to define methods applicable to a wide set of different problems. Ant Colony Optimization (ACO) is part of larger field of research termed ant algorithms or swarm intelligence that deals with algorithmic approaches that are inspired by the behaviour of ant colonies and other insects. In ACO, artificial ants build solution to the considered optimization problem at hand and exchange information on the quality of these solutions via a communication that is reminiscent of the one adopted by real ants. The aim of this paper is to introduce ant colony optimization and to survey its application.

Problems to be discussed: TSP, quadratic assignment problem. Job-shop scheduling problem

2 ACO to solve NP Complete Problem like Travelling Salesman Problem

The travelling salesman problem can be considered as a complete graph with n vertices, we can say that a salesman wishes to make a tour by visiting each city exactly once and finishing at the city he starts from. There is an integer cost $c(i,j)$ to travel from city i to city j , and the salesman wishes to make tour whose total cost is minimum, where the total cost is the sum of the individual costs along the edges of the tour.

In ant colony optimization the problem is tackled by simulating a number of ants moving on a graph that encodes the problem itself.

The artificial ants are having following properties:

1. Give individual ants limited amount of memory
2. Record trip to destination

3. Upon reaching destination, scan to eliminate loops
4. Retrace steps on the return trip, ignore pheromone trail
5. Apply pheromone only on the return trip.

3 ACO Algorithm overview

1. Initialize all of the arcs with a uniform pheromone level, initially ($p=0$).
2. Randomly place ants on the grid.
3. Progress Forward.
4. Eliminate the loop in the path traced.
5. Retrace steps
6. Globally update the trail by evaporating a portion of the pheromone according to parameters.
7. Apply the amount of pheromone to retrace arc.
8. Loop or exit

Consider the situation in the Fig.1, when searching for food ants initially explores the area surrounding their nest in a random manner. While moving, ants leave a chemical pheromone trail on the ground. Ants can detect the pheromone and the path with strongest pheromone is chosen by the other ants. Therefore the pheromone trails will guide other ants to the food source. This communication between the ants via pheromone trails-known as stigmergy- enables them to find the shortest path between their nest and the destination.

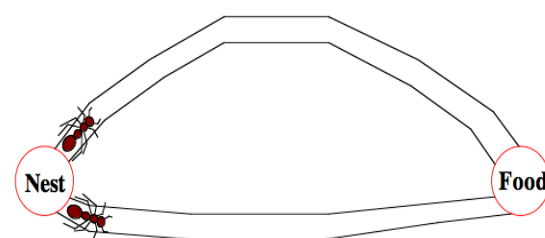


FIGURE 5 Ant selects the random path to go from source to destination

While considering the computation using ACO, we assume that initially the probabilities of the ants for selecting the path are equal and they do not leave pheromone while going from nest to food. We assume that the time taken by each ant per unit distance is constant, as can be seen from Figure 2 at time= t , one of the ants has already reached the destination whereas the other is still on the way back to the nest. If the situation repeats itself, the deposition of pheromone on the shortest path will be strong and after some interval of time all the ants leaving from nest to food will choose the shortest path as shown in Figure 3 and hence will be considered as the optimized path to reach from nest to food.

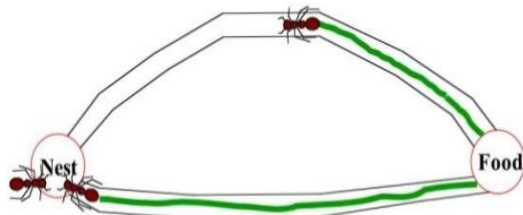
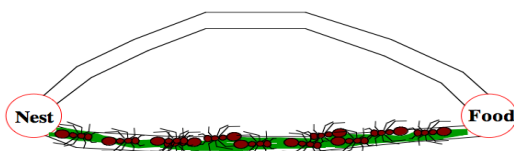


FIGURE 6 During return journey ant leaves behind the traces of pheromones



The scenario illustrate the situation when there are two ants moving from one point to another, now consider the real time TSP in which there are i cities. During construction of a feasible solution, ants select the following city to be visited by probabilistic decision rule. The probability rule between two nodes j , called Pseudo-Random-Proportional Action Choice Rule, and it depends upon two factors: the

heuristic and metaheuristic.

$$p_{ij} = [T_{ij}]^a [\eta_{ij}]^b,$$

$$\sum [T_{ij}]^a [\eta_{ij}]^b,$$

where T_{ij} is the amount of pheromone on the edge (i,j) ; a is the parameter to control the influence of T_{ij} ; η_{ij} is the desirability of edge (i,j) ; b is the parameter to control the influence of η_{ij} .

4 Applications

Many NP problems can be solved by using ant colony optimization; one of them is Graph colouring problem. The problem can be described as follows: given m colours, one has to find a way of colouring the vertices of a graph such that no two adjacent vertices are of same colour. The feasible solution to this problem is solved by ACO which is referred in paper. Other applications include Scheduling problem, routing in telecommunication network, traffic dispersion routing etc.

5 Conclusions

ACO is a recently proposed metaheuristic approach for solving hard combinatorial optimization problems. Artificial ants implement a randomized construction heuristic which makes probabilistic decisions. ACO is a class of algorithms, whose prime member called ant system was initially proposed by Colomni, Dorigo and Maniezzo. The idea was inspired by the behaviour of real ants and their way of approaching to the destination (food) from the source with the help of pheromone.

Acknowledgments

I would like to acknowledge all the colleagues and seniors for helping me for my research work and for being continuous support to complete this paper.

References

- [1] Dorigo M, Gambardella L M 1997 Ant Colony System: A Cooperative Learning Approach to the Traveling Salesman Problem *IEEE Transactions on Evolutionary Computation* 1(1) 53-66
- [2] Dorigo M, Maniezzo V, Colomni A 1996 Ant System: Optimization by a colony of cooperating agents *IEEE Transactions on Systems, Man, and Cybernetics – Part B* 26(1) 29–41
- [3] Dorigo M, Stutzle T 2004 *Ant Colony Optimization* MIT Press, Cambridge, MA
- [4] Dorigo M, Di Caro G 1999 *The Ant Colony Optimization Metaheuristic* New Ideas in Optimization, D. Corne et al., Eds., McGraw Hill, London, UK 11-32
- [5] Zhang Yi, Pei Zhi-li, Yang Jin-hui, Liang Yan-chun *An Improved Ant Colony Optimization Algorithm Based on Route Optimization and Its Applications in Travelling Salesman Problem* BIBE.2007.4375636
- [6] Nada M A, Salami Al 2009 Ant Colony optimization Algorithm *UbiCC Journal* 4(3)
- [7] Ibraheem Sapna Katiyar, Ansari Abdul Quaiyum 2015 Ant colony optimization: A Tutorial Review *MR International Journal of Engineering and Technology* 7(2) December 2015
- [8] Bonabeau E, Dorigo M, Theraulaz G *Swarm Intelligence - from natural to Artificial System A volume in the Santa Fe Institute studies in the sciences of complexity*
- [9] Dorigo M, Stutzle T 2009 *Ant Colony Optimization: Overview and Recent Advances* IRIDIA –Technical Report Series Technical Report No. TR/IRIDIA/2009-013May
- [10] Doerr B, Ashish Ranjan Hota, Timo Kotzing 2012 *Ants easily solve stochastic problems* GECCO'12 July 7–11, 2012, Philadelphia, Pennsylvania, USA

AUTHOR



Kavita Tewani, 12-09-1990, Gujarat, India

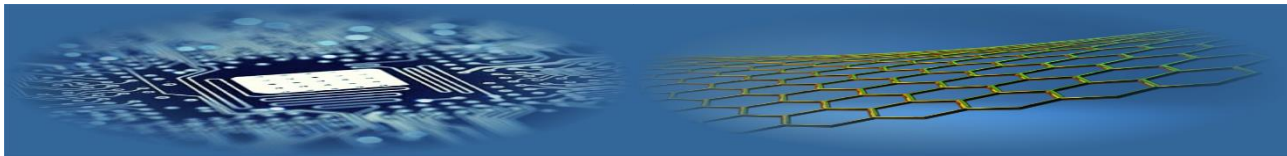
Current position, grades: Assistant Professor

University studies: M.Tech in Computer Science

Scientific interest: Algorithms, Theory of Computation, Data Structures, Natural Language Processing

Publications: 3

Experience: Currently working as Assistant Professor in Department of Computer Engineering, ITM Universe, Vadodara, with three years of experience.



Business process re-engineering capability based on ECMM: Efficient Configuration Model and Management

K Sekar^{1*}, M Padmavathamma²

¹Associate Professor, Research Scholar, S.V.Engineering College for Women, Tirupati

² Professor, Research Supervisor, S.V.U College of Commerce Management & Computer Sciences

*Corresponding author's e-mail: sekhar.k@svcolleges.edu.in

Received 8 April 2017, www.cmnt.lv

Abstract

Most business process Companies are interested for new solutions and techniques in organisations. Relating to the big data to achieve that business process must be reengineered. Reengineering of business process can be done based on six sigma activities like Define, Measure, Analyze, Improve, Control and Report. In Business Process Reengineering, the two constants of any organisation are people and process. If individuals are motivated and working hard, here the business process are compressive and organisational process will be poor and posses high failure rate. In order to overcome these effects Business Process Reengineering must have some assessing capabilities which is referred as Desired Organizational Capabilities (DOC) and total quality management (TQM) to increasing the efficiency of reengineering and makes the manufacturing of logistical systems more scientific.

Keywords:

Six Sigma activities,
DOC,
Business Process Reengineering,
TQM

1 Introduction

Reengineering is the fundamental rethinking and radical redesign of business process to achieve dramatic improvements in critical, contemporary measures of performances such as cost, quality, service and speed. The keywords in the preceding to finishing or utilise once. The BPR advocates that enterprises go back to the basic and re-examining their favour routes. Reengineer should focus on process and should not be limited to thinking about the organisation. After all the organisation only aspect as a group of process but it is a single process. Business process is a series and a step design to produce a product or a service.

2 How to Reengineer?

Planning and preparation are vital factors for any activity or event to be successful but reengineering is known exception. Before attempting reengineering the question is which business process is necessary there should be a significant process to be reengineered. The justification of this needs marks beginning of the activity. Some of the researchers argue that the original concept of reengineering can be traced back to the management theories of the nineteenth century where people, data and technical logic must be considered.

The main objectives of the BPR to be considered in the present technology are:

1. Customer focus: The main aim is to eliminate the customer complaints.
2. Speed: The dramatic compression of time it takes key business.
3. Processes: For instances if processes BPR every cycle time 5 hours to cut down to half an hour.

4. Compression: The operation level to reduce cost, flexibility, adaptive processes and structures to change conditions and competitions.
5. Quality: A session to the superior service value to the customer and the level of quality is always a scheme control and monitor by the processes and not depend mainly as the person who serving the customer.
6. Innovation: The leadership through imaginative change to the providing organisation for competitive advantage.
7. Productivity: Improve drastically, effectiveness and efficiency.

3 BPR project Implementation/Alternative techniques

The six stages of BPR are to be implemented in the drastic change of companies' outcomes:

The Envision stage: The companies' reviews existing strategy and business process and based targeted an IT opportunities are identified.

The Initiation stage: The project team are sign perform goes project planning and employee notification are assigned.

The Diagnosis stage: The documentation process take place of attributes, activities, resources, communication, rules, IT and cost.

The redesign stage: The new process design is develop by device of processing alternatives brainstorming and creativity techniques.

The reconstruction stage: The management technique changes and smooth migration to the new process responsibilities and human resources roles.

The evaluation stage: The new process is monitor to determine goals and methods to examine total quality programs.

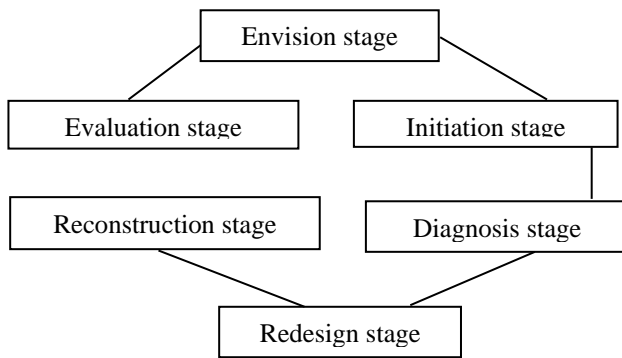


FIGURE 1 BPR Implementation of six stages

The BPR could be implemented to all firms (manufacturing firms, retailers, services etc) and public organisation to satisfy the following criteria:

- Minimum no of employees: 20 (at least 4 in management positions)
- Strong management commitment to new ways of working and innovation.
- Well-formed IT infrastructure.

Business process reengineering could be applied to companies that content problems such as: High operational cost, Low quality offered to customers, high level of "bottleneck" processes at pick reasons, poor performances of middle level managers, Inappropriate distribution of resources and jobs in order to achieve maximum performance etc. The BPR is achieving dramatic process performance improvements through radical change in organisations processes, remodelled of business and management processes. It involves redrawing of the organisational, boundary, the reconstruction of jobs and skills. The use of six sigma to reduce temporary labour expenses are define, measure, analyse, improve, control. The main advantages of business process reengineering are satisfaction, growth of knowledge and increases of proof.

The component that are essential to BPR achieve include:

1. BPR cooperative unit report.
2. Business necessarily synthesise work.
3. Allowable IT infrastructure.
4. Powerful [change management].
5. Current continuous betterment.

The prospect of a BPM attempt that are changed to consider organizational design, Consultant systems, employee duties and execution measurements, motivator systems, attainment development, and the use of IT. BPR can possible affect every look of how business is arranged today. Sweeping changes can cause effect from desirable success to complete failure. If self-made, a BPM enterprise can result in developed character, client service, and aggressive willingness to compete, as well as step-down in cost or cycle time. Even if, 50-70% of redesigning projects are either failures or do not achieve important welfare.

There are many causes for business processes which include:

1. One section may be use best at the sacrifice of another.
2. Lack of time to point on rising business process.
3. Lack of acknowledgment of the extends of the problem.
4. Lack of preparation.
5. People engaged use the optimal tool they have at

their administration which is usually rule to fix problems.

6. Lacking substructure.
7. Overly bureaucratic process.
8. Lack of needs.

More defeated BPR cause may have been due to the confusion surrounding BPR, and how it should be execute. Organizations were well sensible that changes necessary to be made, but did not know which region to change or how to change them. As a result, processes redesigning is a management construct that has been formed by track and error or, in other words, realistic go through. As more and more businesses reengineer their process, knowledge of what stimulate the successes or failures is becoming evident. To draw permanent welfare, companies must be prepared to analyse how scheme and reengineering full complement each other by acquisition to measure system in terms of cost, milestone, and timetables by exceptive property of the strategy throughout the organization, by evaluate the organization's present capabilities and process realistic manner, and by associate strategy to the make a budget process. Other than, Business process reengineering is only a short-term efficiency workout.

4 Literature Survey

Goldstein, D. & Hilliard, R et.al. Proposed that in every organization resources and capabilities are very important because resources can be treated as available factors that are controlled by the firm and capabilities are treated as firm's capacity to deploy resources for a desired result. Basically, Development capabilities and Deployment capabilities are the two types of capabilities which are very important for every organization. These capabilities help to get intent and deliberation and also considered as important elements in routine.

Feline, T., Foss, et.al. Proposed that whenever new organizations or existing organizations are entered into the market they need to develop new capabilities and routine according to the organization or else alter the existing ones. Routine can be stated as an approach for organizational actions in organization and strategic research literature. Normally Observation theory and Theory Based Intervention are two theories that are claimed to be affected by the routine actions. There is integration between routines and individual organization process which termed as Habits. Habits are difficult to change because it is interlinked with the basic knowledge and theme of the organization.

PELAEZ V., HOFMANN, et.al. Proposed that Dynamic capabilities are important for every organization because it helps to provides higher level of competence that makes organizations to best use of their internal and external capabilities and reduce the future difficulties. In order to get this organization must align resources both inside as well as outside. However new organizations are failed to establish routines and capabilities which effects on natural growth rate so, they need to identify where do routines come from and their relationship with capabilities and other forms of behaviours.

Firms or organizations can have various types of capabilities such as managerial and technological based on various situations which can be called as Desired Organizational Capabilities. Much of the empirical investigation and research survey explains that it is difficult

to measure the capabilities that are independent of the result or outcome of filed. Desired Organizational Capabilities help to the organization as long as they build routines in an organization. So, every organization needs to learn how to develop and construct new capabilities over time which will in turn ensure a smooth functioning of the organization. Implementing the routines according to the basic theme of organization also plays a vital role in boosting the efficiency level of organization. Furthermore these routines and desired organizational capabilities helps in investigating or verifying the extensive organizational research in order to identify the vest and beneficial organization in the market.

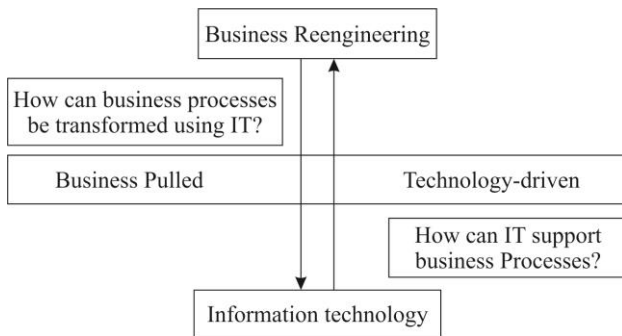


FIGURE 2 Identify enabling IT & generate alternative process redesigns

5 Risk factors and barriers to process reengineering

The rope of reengineering business processes has been underscored by the high powered Narshimhan committee on restructuring of Indian banks and also by the vasudevan committee on technology issues. The issues are financial risks, technical risks, general project risk (implementing solution may not perform to the desire level), functional risk (which organisation is engaged), political risk (support and commitment of top management due to change in perception).

BPR in E-business and E-commerce, these two concepts are often mixed up

E-business electronically connects in multiple ways in many organisations internally or externally. There are three types of perspective in BPR

Communication perspective: Delivery of information or payment over telephones lines or computer networks or any other electronics.

Business process perspective: Automation of business transactions workflow

Service perspective: Consumers and managed to art service cost while proving the quality of goods and increase the service deliver.

Online perspective: Buy and sell products through online services.

5.1 BUSINESS PROCESS REENGINEERING VERSUS BUSINESS PROCESS IMPROVEMENT

The purpose of both business process reengineering and business process improvement is to make business efficient, effective and flexible. They need to provide the moderate variables where industry type information and intensity of the industry are two moderator variables. The extent to which product and services of divisions are dependent on informative. Data integration and process improvement trust by recognition central management process, data coding

activity often introduces new ways as much as sequential process should be handled.

Desired organizational capabilities are important for BPR because DOC helps BPR in order to fill the gap in the literature on decreasing the risk in BPR. DOC helps to organization combines people and requirements along with resources together in order to complete any kind of work successfully. It also distinguishes the difference between what is good at doing and consequences of bad work effect in growth of organization. DOC is stable and difficult to copy for competitors because DOC can form according to the organization policies

DOC emerges when a company delivers on combined abilities and competencies of its individuals. Whenever an employee will have an ability to show a leadership skill but the company as a whole may not have an ability to maintain the same leadership skill as an individual. So, DOC can be formed and maintained according to the employee abilities towards the organization. In addition to that DOC can enable a company implies its technical activities to run any kind of work effectively and it will show good effect in results of organization.

DOCS are not predefined and suitable for every organization. However, there are some general and basic desired organization capabilities for every organization.

Capability 1: Employees must and should have the capabilities to convince a customer according to business project. Committed employees must deploy their skills regularly according to the task. So, this step helps to measure productivity, check static's and conduct surveys' through observation.

Capability 2: Organization must have an ability to react quickly to grab and recognize opportunities in order to exploit new market and implement changes in product according to the market, must acquire new employees and implement new business process according to the business problem. Managers and important people in organization must have an ability to take decisions according to the market statistics

Capability 3: Every employee in the organization must feel free enough to share an idea regarding the changes needed in order to improve the market status. Every team must consider the top three things to implement in future in order to satisfy customers regarding the product. The next step is to take customers feedback on brand identity.

Capability 4: Organization must be establish in such a way that it can grab the good performance from employees in order to provide best results in the market. Performance of an employee can be counted as a success measure for every organization. Every employee must consider the goals of the organization and work according to that in order to meet the goals in a less time and provide best results to increase the share in the market.

Capability 5: Every organization must set a goal to ensure both efficiency and leverage. Organization must work as a whole to gain efficient results for every business problem. Every task must be completed through pooling of services or some technologies or by sharing ideas to get more efficient output. There must collaboration between organization and the teams working under the organization in order to increase the organization's strength

Capability 6: Organization must respect employees'

new ideas in order to increase the value of organization at market. Organization must ready to do experimenting on the basis of ideas generated by employees and ready to face the risk. Organization must adopt new ideas and leave old practices for solving business problems.

Capability 7: Organization must elect best leaders in order to get growth of results because best leadership skills must have a capability get best results. Leaders of organization must have the capability to know how to do and what to do in order to compete with competitors in the market. Organization must track the leadership skills of employees and monitor the pool of future leaders. Organization must maintain a backup for employees and stand in any kind of situation

Capability 8: Organization must and should maintain a good relationship with customers and ensure trust for customers. Involvement of customer in solving a business problem by considering requirements for problem. In order to have this capability every organization must maintain a customer service centre and ready to solve their problems at any time. Frequent customer surveys must be conducted to get to know about customer opinion on the process of execution.

Capability 9: Every company must and should follow some strategy in order to solve business problems and maintain organization in unambiguous way to deal with problems. Organization must notice how employees respond to customer and make sure employees must follow organization strategy to avoid ambiguity in solving problems for customers. Organization must take suggestions from employees in order to form a default strategy for organization.

Capability 10: Organization must innovate something whether in products or in strategy process to possess an effective outcome. Every organization must focus on future success and forget about past results. Efficiency is key to success so leaders must concentrate on costs related to solve problems and try to get efficient results in order to grow the top line in market. Efficiency may be the easy capability for every organization which is linked to employees working in the organization

These are the some of the key capabilities that every organization must and should consider as desired organizational capabilities in business process reengineering. It will helps for every organization to solve any kind of business problem after reengineering and it will acts as a blueprint for new organizations who are ready to compete in the business market.

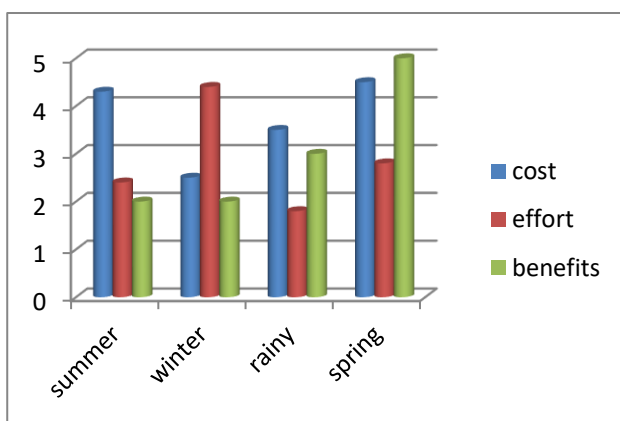


FIGURE 3 Differentiating Individual competence vs Organisation capability

	Individual	Organisation
Technical	An individual functional competence	An organisation core competence
Social	An individual leadership ability	An organisation desired capabilities

6 Auditing capabilities

After setting own desired organizational capabilities in an organization, next step is to auditing the capabilities in order to crosscheck how the capabilities are helping to get growth in results as well as in business market. The capability auditing must be done to check leadership behaviours and monitor organizational assets. It will highlight which is more important in given organization history and strategy. It will measure how well company delivers on the capabilities and lead to implementing a plan for improving results based on history.

Next, to evaluate the organizations performance on these capabilities. Performance reflects the results for the business problems and helps to grow in the market. Performance of employees according to working strategy and leaders to get involve in guiding employees to work correctly. Auditing helps to know about capabilities well and have a ability to choose best capabilities that most effect the ability to deliver effective results in market.

The leaders of every organization must discuss about the survey at an off-site meeting. Meetings will help to address the gap between the strategy statements and procedures that focus on services. Before implementing overall improvement plan define the capabilities that would be most important to execute the strategy after completing the auditing. Once the auditing is done don't choose the capabilities with low scores in performance. The leaders are ready to invest in further developing capabilities which would lead to get success. In particular they must focus on marketing skills and hiring employees who will suitable for organization. Effort for creating a leadership brand and forming a new team who will give a high performance is most difficult part after auditing.

While auditing is going on it is important to understand that which capabilities depend on one another. So, even though we target on one capability it will link the depended capability which is important to audit. Most important capabilities must be combined with one another based on any aspect common between both the capabilities. If we do combining it will help to improve two capabilities at same time. A leader must built each capability in organization by considering the main factors of organization so, working on any one of the capability may helps to build leadership in terms of assessing factors which will helps to get success for organization.

Finally, auditing helps to assess strength in support of leaders in organization and starting with the organization with a new essence of implementing plans to get effective results for every business problem based on capabilities. It is not necessary to boost weak capabilities but to identify and build capabilities that will have the strong impact on execution of strategy in

Every organization Auditing also helps to know at what capabilities would essential for future success, and assessing desired capabilities in terms of organization requirements.

TQM=CQI through+ customer focus+ process

improvement +total involvement

Total quality management is continuously improving quality (CQI) by focusing on customer requirements, improving the processes to relate these expectations and involving everyone in the improvement.

TQM is an overall philosophy and management system.CQI can be useful for the organised to access and to enforce TQM.

Auditing for capabilities must be done in every business unit in the entire organization. Every part of strategy must be audited according to the plan. Auditing can be done starting with the core modules of the organization by checking how well a capability will set to a particular module at any situation.

The first step in auditing is to know about the critical areas in the organization which will meet the goals of organization. The audit process started with collection of feedback from multiple sources on critical areas and capabilities involved in organization. Based on above mentioned capabilities which are generic for every organization the auditing can be taken place. Business requirements must be adapted and not to be changed when auditing is going on.

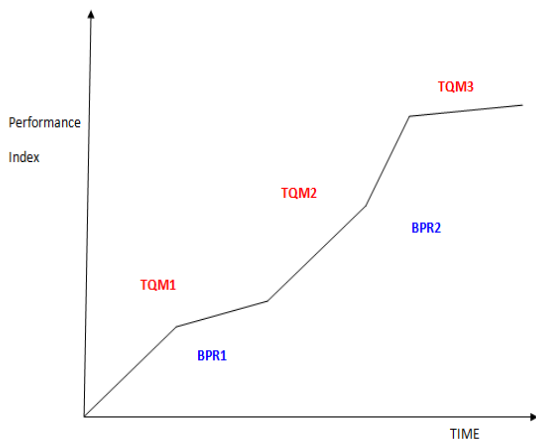


FIGURE 4 TQM and BPR based on performance index and time

Cost: Design and implementing the business process, Hire and train employee, develop supporting information system, bought of other instruments and opportunity.

Benefits: Customer satisfaction, related goals, performance evaluation, queries.

Risk: Modern facilities availability and related goals, time needed for design and implementation, serving curve, cost and time table must be followed.

Why business process management is necessary?

ERP implementations result in significant changes to a business and its organisations.

1. Identifies process and organisational changes.
2. Highlights organisational change management issues.
3. Improve the process based on needs of the company vs the needs of software
4. Business process model is compulsory for all ERP techniques.
 - Reduces overall risk to the project.
 - Increases the return on investment and identifies other cost saving opportunities.

- Ensures key differentiating business processes and logic are not lost during implementation.
5. Protects and maintains company competitive edge.

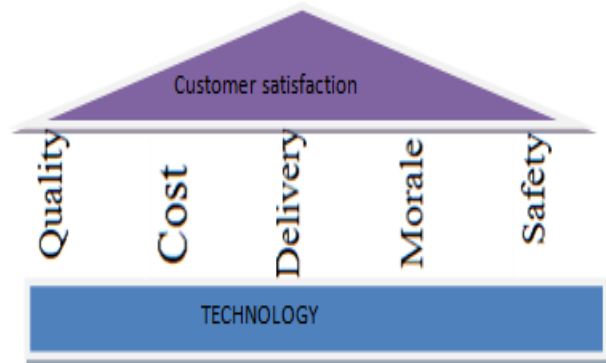


FIGURE 5 Before and After of TQM

Before implementing of TQM	After implementing of TQM
1.Working as an individual	1.Working as team
2.Focus on results	2.Focus on continuous improvement and changes
3.Adhoc decision making	3.Fact based decision making
4.Fixed wages for assigned work	4.Rewards and recognition



FIGURE 6 Stages of ERP at market planning

Profit key is Enterprise resource planning or ERP is a software solution that provides manufacturers with the information necessary to effectively manage their business processes.ERP is a fully integrated real time system giving the information that we need to grow to our business.ERP enables by connecting information for everyone, every department and every process throughout your entire enterprise. It begins up to the movement seamless communication and accurate picture of our valuable resource management and way to use for better project planning & leading company business decisions really in the manufacturing process. This is true business intelligence with complete quote to bill management capabilities and streamlined office automation. A capability ERP solution is a manufacturing execution system or quality management module. Profit key is proprietary for rapid response manufacturing of industries first fully integrated ERP and MES software solution for small to mid-sized manufacturers.

7 Business Process Reengineering (BPR) and Total Quality Management (TQM):

The TQM and BPR interfaces a two-sided role level played. The total Quality investigates to explain rapidly changing or growth exchange and gradual increasing of process, while proposed of reengineering often seek to again redesign for radical incremental of process. The quality management related as continuous increasing, means programs and techniques, which evaluate incremental improvement in work process, and outputs over an open-closed period of time. In adding, reengineering also known as business process redesign or process initiation, refers to pretend initiatives intended to achieve radically redesigned and improved work processes in a specific time period. In related to continuous improvement & TQM, BPR implies on a different way of thinking.

The extract difference between continuous process improvement and business process reengineering lies in where start from and also the magnitude and rate of resulting changes. In particular period of time, many derivatives of radical, breakthrough improvement and continuous improvement have emerged to address the difficulties of implementing major changes in corporations. Leadership is most important for effective BPR deployment, and successful leaders use leading styles to suit the particular situation and perform their tasks, giving the importance to both people and hard work. Business process is essentially value engineering applied to the system to bring forth, and sustain the product with an emphasis on information flow. By mapping the functions of the business processes, low value functions can be identified and eliminated, thus reducing cost. The priority of effective responsibility or total quality management did not be disclosed. They should provide the valued resources to work evaluate their active support for the team, set the every stage for reengineering by considering core business techniques, and by identifying the project purpose and problems area. The two techniques of BPR should also give the importance to provide effective findings, set motive standards as well as supported others to be realising to their innovative ideas. More business process projects fail to be succeeded or do not reach end-line business outputs. Because of this, BPR 'success factors' has become an most important area to realise. It is way to think of a particular group structure area in 3 solutions: third parties, relative team, and existing team.

The stakeholders are key business leaders ultimately accountable for the success of the project. Their role is to provide high-level guidance to the team, help remove barriers, and provide funding. The core team is the group responsible for the design and implementation of the solution. Your extended team includes other people in the organization contributing to the project on an as-needed basis. These extended-team members include subject-matter experts. A well-rounded team includes a mix of people and skills. Such a team may include individuals who thoroughly understand the current process, who actively use the process and also work closely with customers, technical experts, and consultants, if necessary. But the main criterion is that the entire team should work together for the project to succeed.

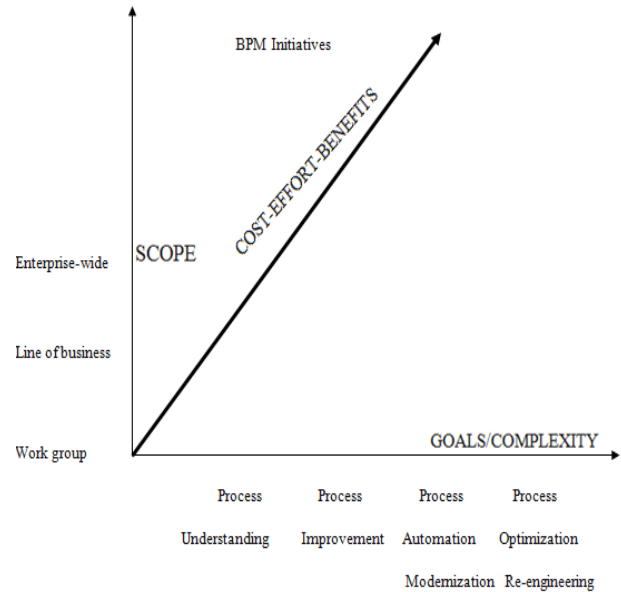


FIGURE 7 The role of consultants in BPR projects

New reengineering teams typically employ the assistance of a consultant for their project. Consultants can play a valuable role in BPR projects. They are objective and immune to internal politics. Having followed the processes before, they provide valuable information and best practices from a wide range of experience. Consultants can also serve as good communication bridge between the team and management, write project documentation, lead the project and facilitate meetings, make presentations to stakeholders and associates, and last but not the least, contribute subject-matter expertise in your organization's work processes.

9 BPR and IT:

BPR has quickly formulated toward a recent management belief. The inbuilt business processes preferences modified the position of world-wide management from a functional to that of a process view. The redesigning of business is only way of the management of business process. In specific, the re-engineering of world-wide business processes needs particular tending, because the various structure of multinational entity growing the complexity of business processes, there by examine the choices for reengineering.

BPR has rapidly developed regarding a recent management philosophy based upon forerunner like TQM, Overhead Value Analysis. Business processes can be redesigning by reengineering the steps, by changing the ordered and secular arranged of the steps, or by changing any other feature of the process. Exponent of data systems prefer the view that the recent technology is an enabled of process re-engineering. IT has to be supervising always to check whether it can create new process designs or put up to the performance of a business process. The discovery of BPR is nearly attached with IT, which opens new measurements of process reorganization.

However, those who take the enterprise in process improvement/redesign, regulating the role of IT. If the data processing sections the process change, then IT will have more of a source purpose for new process redesigns. On the other hand, the top management sets off the change process,

then the process will be first redesigned and after optimized through IT. BPR involves the basic redesign of effect business processes to reach spectacular improvements in productivity, cycle times and quality.

In BPR, companies start with a blank sheet of paper and rethink existing processes to deliver more value to the customer. They typically follow recent value system that places rapidly accent on customer needs. Companies shrink organizational levels and transformed unproductive activities in two important areas. First, they rebuild functional organizations into cross-functional teams. Second, they use technology to improve data quality and decision making.

10 Companies use BPR to:

Companies use BPR to increase performance well on key processes that strong effect customers.

- **Make less costs and cycle time.** BPR reduces costs and cycle times by reject unprofitable actions and the employees who execute them. Reengineered by team's decreases the need for management levels, speed up information rate of flow, and reducing the errors and rework caused by multiple ways.
- **Improve quality.** BPR improves character by reducing the separating into few areas of work and constitute clear ownership of process. Employees profit responsible for their output and can quantify their performance based on motivate resubmit.
- TQM refers to a constant effort of management along with the employees of a particular organization to improve the quality of goods and services. Businesses need to accent on quality of their products rather than quantity to endure the intense competitor. Recollect in today's extension, there is no famine of competitors in the market. Quality is an important argument for each and every business and should not be avoided at any cost.
- TQM works on a very simple rule: The responsibility of delivering quality goods and services to clients lies on every single individual who is even remote assort with the organization. It is not only the management but also employees regardless of their identification, providers, clients, customers who need to come up with increasing ideas to make unailing systems and process to deliver quality products which meet and exceed the expectations of end- users.

11 Results

The two techniques Six Sigma and Total Quality Management are rapid tools for increasing the quality management but very rare line of gap does exist both of them. Consider the methods and procedures involved in between the two appear almost same but there are certain major differences.

The Six-Sigma is a commonly new concept of Total Quality Management but not exactly its presence. The major difference between Total Quality Management and Six Sigma is that TQM delivers superior and ordered quality manufactured goods and services where as six sigma on the other hand output in better results. Total Quality management refers to continuous effort by employees to evaluate high quality products to customers. The process of

Six Sigma presences many small changes in the systems to ensure effective results and better customer satisfaction.

Total Quality Management involves redesigning and developing new systems and processes and ensures effective interfaces among various departments. New Processes are developed based on various customer feedbacks and researches.

The main focus of Total quality management is to maintain existing quality standards whereas Six Sigma primarily focuses on making small necessary changes in the processes and systems to ensure high quality.

The process of Total quality management does reach to an intensity level after a certain period of time. After reaching the ending stage, no further improvements in quality can be made. Six Sigma on the other hand rarely reaches the saturation stage by originate another level reference process.

The procedure of Total quality management involves regard in present policy and operation to check more quality. **Six-Sigma direction on raising quality by understate and finally eliminating flaws from the system.** The purpose of TQM guarantee that each and every single member relate with the arrangement is functioning near the increasing of existing process, systems, services and work refinement for outlook character of goods/services. Six Sigma, on the other hand focuses on first identifying and eventually removing various defects and obstacles which might come in the way of organization's success. In a business model total quality management underline on raising the present policies and making appropriate changes in the systems to check senior quality goods and services. Organizations practicing Six Sigma are focused on removing errors and flaw to assure high quality intersection.

TQM is a less complex process than Six Sigma. Six-Sigma involves particularly disciplined individuals whereas TQM does not expect comprehensive training. The process of Six Sigma creates special levels for employees who are only competent to implement the same. Employees prepared for Six Sigma are often certified as "Green Belts" or "Black Belts" depending on their level of proficiency. Six-Sigma requires involving of only certified professed whereas total quality management can be pertain to a part time activity which does not require any special training. Six-Sigma can be implemented by devoted and well trained professionals.

Six-Sigma is famous to present improved and efficient results as liken to total quality management. The process of Six Sigma is established on customer feedback and is more exact and result destined. Customer feedbacks play a measurable part in Six Sigma. Experts predict that six sigma will highlight TQM in overdue course of time.

Clients and their feedbacks are the based of every TQM model. In simplex words, TQM begins with agreement clients, their needs and what they expect from the establishment. Pattern unailing processes and systems to collect customer data, information to further study, synthesize and act accordingly. Such action not only help you understand your aim for customers but also anticipate client behaviour.

As a business marketer, you need to know the age group of your target customers, their preferences and needs. Workers need to know how their products or services can accomplish customer needs and demands.

TQM model needs precise planning and research. Every TQM design incorporate customer feedbacks with relevant information and plans accordingly to design impressive

scheme to achieve high quality products.

Strategies developed to generate better quality products need to be measured and re-examined from time to time. Recollect, clients are fulfilled only when products meet their expectations, action their needs and are value for money. Their overall receive with the organization needs to be pleasing for them to be happy and return to the organization even the next time.

Continuous improvements, changes and adjustments in the existing processes according to customer prospect are necessary to effort higher profits. Processes can't be same always. If a client complaints about a particular product of yours, find out the root cause of problem. Understand and implement necessary TQM frameworks to evaluate the problem, remove the fault for a high quality product.

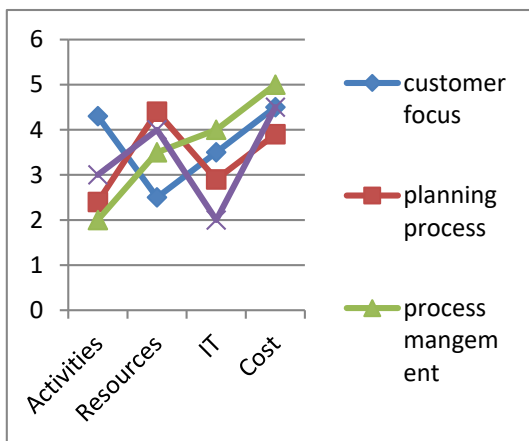
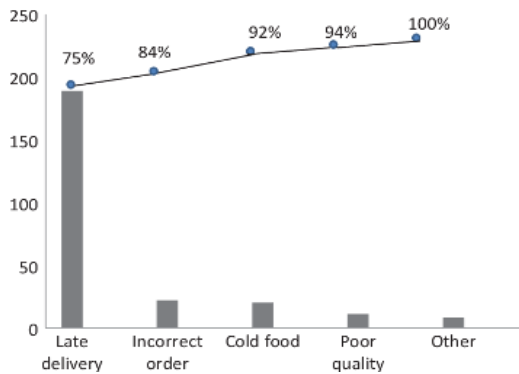


FIGURE 8 Six sigma related to business sector before implementing the TQM

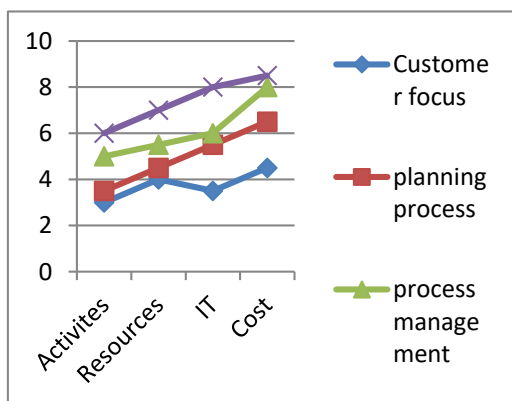


FIGURE 9 Six sigma related to business sector after implementing the TQM

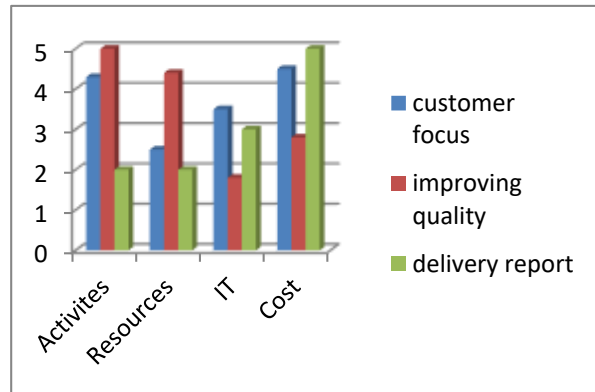


FIGURE 10 Six sigma related IT business sector

According to six sigma activities leading the business sectors now-a-days. In this TQM increases the efficiency of business process based on six sigma where as quality must be improved TQM organizations use the techniques of process management to develop cost-controlled processes that are stable and capable of meeting customer expectations.

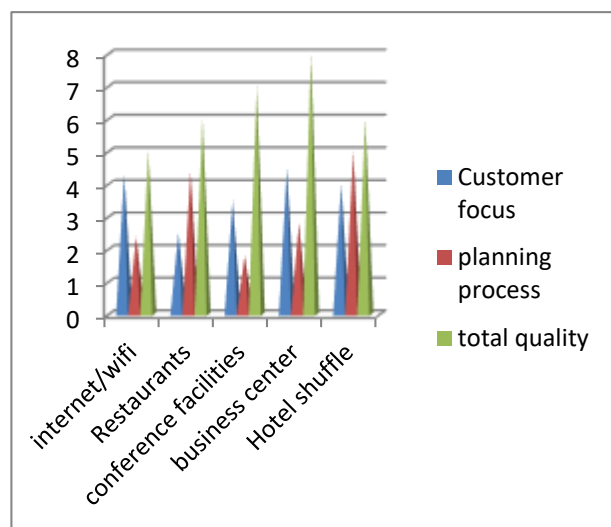
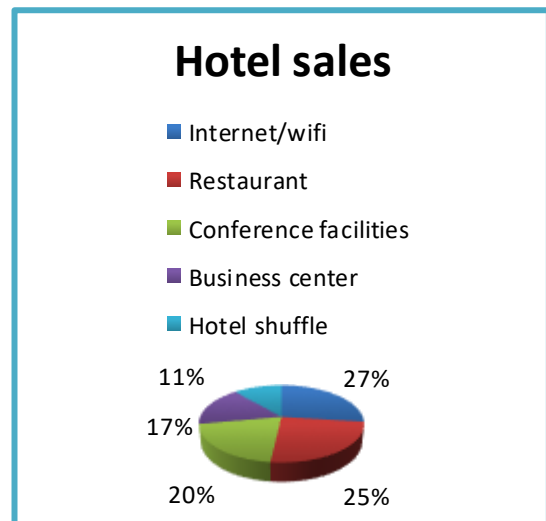


FIGURE 11 six sigma related to TQM in business sector for quality improvement



12 Conclusion

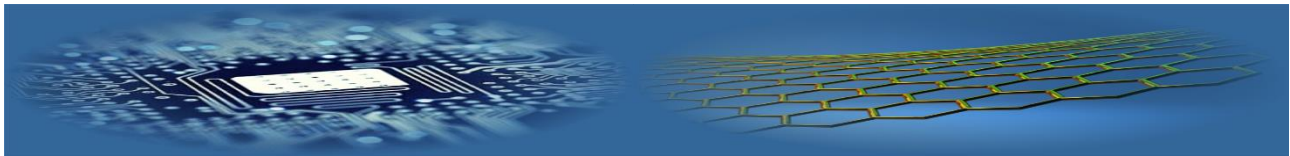
In this paper, Business process and data models are extend based on Desired Organization Capabilities. The world wide scale of the economy and the relaxation of the trade markets have developed new rules in the market place which are characterised by imbalance and characterised by a high degree contest in the business process environment. Business Processes are qualified by three elements: the inputs, the preparing of the data materials and the outputs. Planning and preparation are critical factors for any specific

behaviour or even to be successful and redesigning is no exclusion. The strong effect of the environmental changes that serve as the impulsion for the redesigning effort must also be studying in establishing guidelines for the reengineering project based on TQM. The results that we obtained show that our proposed data model is more efficient in terms of management and organization capabilities and we can also extend in a wide manner by using the PQM (process quality management) to meet the demands of your customers and also improve the quality of your deliverables in a useful way.

References

- [1] Aghdasi M, Albadvi A, Ostadi B 2009 Desired organizational capabilities (DOCs): mapping in BPR context *International Journal of Production Research*
- [2] Albadvi A, Keramati A, Razmi J 2007 Assessing the Impact of information technology on firm performance for considering the role of intervening variables: organizational In infrastructures and business processes reengineering
- [3] Sutcliffe N 1999 Leadership behaviour and business process Reengineering (BPR) outcomes: an empirical analysis of 30 BPR Projects *Information & Management* 36(5) 273–86
- [4] Attaran M 2004 Exploring the relationship between information technology and business process reengineering *Information & Management* 41(5) 529–684
- [5] Laguna M, Marklund J 2004 *Business Process Modeling, Simulation and Design* Prentice Hall
- [6] Zigiari Sotiris 2000 *Business Process Re-engineering BPR Report produced for the EC funded project INNOREGIO: dissemination of innovation and knowledge management techniques*
- [7] Milgrom P, Roberts J 1992 *Economics, Organization and Management* Prentice Hall
- [8] McGee A M, Wang R Y 1993 Total Data Quality Management (TDQM): Zero Defect Data Capture *The CIO Perspectives Conference, Tucson, Arizona*
- [9] Aalst W M P, Vander & Hee, Van 1996 Business Process Redesign: A Petri-Net-Based Approach *Computers in Industry* 29(1-2) 15-26
- [10] Bellassai G, et al 1995 SISCO: A tool to Improve Meeting Productivity *International Workshop on Groupware, CRIWG'95* 149-61, Portugal
- [11] Yu Eric S K et al. 1996 Modelling the Organization: New Concepts and Tools for Reengineering *IEEE Expert, AI Models for Business Process Reengineering* 16-23
- [12] Feline T, Foss N J 2009 Organizational routines and capabilities: Historical drift and a course-correction toward micro foundations *Journal of Management Studies* 25(2) 157-67
- [13] Goldstein D, Hilliard R 2004 Creating organisational capabilities: learning by doing *Journal of Management Studies* 304-18
- [14] Altinkemer K, Chaturvedi A, Kondarredy S 1998 Business Process Reengineering and Organizational Performance: An Exploration of Issues *International Journal of Information Management* 18(6) 381-92
- [15] Altinkemer K, Ozcelik Y, Ozdemir Z 2007 Productivity and Performance Effects of IT-enabled Reengineering: A Firm-Level Analysis *European Conference of Information Systems (ECIS) 2007, St. Gallen, Switzerland*
- [16] Davenport T 1993 *Process Innovation: Re-engineering Work through Information technology* Harvard Business School Press, Boston, USA
- [17] Dedrick J, Gurbaxani V, Kraemer K 2003 Information Technology and Economic Performance: A Critical Review of Empirical Evidence *ACM Computing Surveys* 35(1) 1-28
- [18] Devaraj S, Kohli R 2000 Information Technology Payoff in the Health-Care Industry: A longitudinal Study *Journal of Management Information Systems* 16(4) 41-65
- [19] Grover V, Teng J, Segars A H, Fiedler K 1998 The Influence of Information Technology Diffusion and Business Process Change on Perceived Productivity: the IS Executive's Perspective *Information and Management* 34 141-59
- [20] Earl M 1987 Information systems strategy foundation In *Critical Issues in Information Systems Research*, R. Boland and R.Hirscheim, eds. London: John Wiley & Sons,
- [21] Geisler E 1991 Introduction to the special issue: Management of information technologies-bridging theory and practice *ZEEE Transactions on Engineering Management* 38(4) 290-2

AUTHORS	
	<p>Koneti Sekar, Date of Birth: 02-09-1974, Tirupati</p> <p>Current Position: Associate Professor in the Department of Computer Science and Engineering, S.V.Engineering College for Women, Tirupati.</p> <p>University Studies: He obtained his Bachelor Degree in Computer Science from Sri Venkateswara University. The he obtained his Masters Degree from University of Madras and pursuing Ph.D in Sri Venkateswara University.</p> <p>Scientific Interest: His Specializations include Software Engineering, Computer Programming, Computer Security, Computer Organization and Object Oriented Programming.</p> <p>Publications: He has published 6 International Papers and Presented 2 Papers in National Conference and 2 Papers in International Conference.</p> <p>Experience: He has 17 Years of Experience.</p>
	<p>Mokkal Padmavathamma, Date of Birth: 23-06-1963, Chittoor District</p> <p>Current Position: Head, Department of computer science, S.V. University, Andhra Pradesh, India.</p> <p>University Studies: She received M.Sc, M.Phil, M.Ed, Ph.D from S.V.University, Tirupathi and M.S (Software Systems) from BITS PILANI.</p> <p>Scientific Interest: Her research interests lie in the areas of Number theory, Cryptography, Network Security, Distributed Systems and Data Mining.</p> <p>Publications: She has published 35 research papers in national/International journals and conferences. She published TWO text books as one of the author. Also she is life member of cryptology Research Society of India (CRSI) and Andhra Pradesh Association Mathematical Teachers (APAMT).</p> <p>Experience: She has 28 Years of Experience.</p>



Handwritten digit recognition using combined feature extraction technique and neural network

Ankita Mishra*, Dayashankar Singh

Madan Mohan Malaviya University of Technology, Gorakhpur, India

**Corresponding author's e-mail: m.ankita3011@gmail.com*

Received 10 April 2017, www.cmnt.lv

Abstract

Handwritten digit recognition is established and emerging problem in pattern recognition and computer vision. A very few volume of work related to research has been done in this field till now. Handwritten digit recognition is very useful in cheque processing in bank, form processing systems and many more. In this paper, a robust and novel technique has been introduced for handwritten digit recognition which is tested on well-established MNIST dataset. Histogram of oriented gradient technique and wavelet transform technique is used for feature extraction. Radial basis function neural network and back-propagation neural network have been used as classifier. Experimental analysis has been carried out and result shows that RBF yields good recognition accuracy as compared to back-propagation neural network.

Keywords:

Handwritten digit recognition (HDR), Back propagation Neural Network, Radial Basis Function, Histogram of Oriented Gradient (HOG)

1 Introduction

This Handwritten digit recognition is one of the main challenging and emerging problem of research in pattern recognition, computer vision and machine learning. Many researchers have done a lot of research work in the area of handwritten digit recognition; still there is a lot of scope to enhance the recognition accuracy of handwritten digits. Handwritten digit recognition is getting increasingly attention due to its wide application areas. Digit recognition has been used in several applications such as, processing of bank-checks, reading bank slips, distinct type of forms like loan, health insurance forms, tax, postal addresses, sorting post mail, examination paper, script recognition etc.

Handwritten Digit Recognition (HDR) is basically an art that detects and recognizes digits from scanned input images and then converts it into appropriate machine editable forms. However, handwritten digit recognition is improving the interface between machine and men in several applications. As we know due to wide variation in personnel writing styles, handwritten digits do not look same length, form, style, position, thickness and coordination, so it emerges as a major challenge in the field of pattern recognition. The first kind of difficulties is due to high variability in the digit shape by individuals and uncertainty in writing styles. Not only for the reason of that there is a huge variation in shape and pattern of digits, but also of interconnection and overlaying of neighboring digits [1]. Recognition of digits is a very common and easy task for the human but in case of machines it gives a serious problem especially in that case when there are digits having ambiguity, great similarity in shapes like 1, 7 and 8, 9 etc.

The performance of handwritten digit recognition system is highly depending upon two things: First it depends on feature extraction techniques which is used to increase the performance of the system and improve the recognition

accuracy rate and the second is the neural network approach which takes lots of training time and automatically infer the rule for matching it with the correct pattern Recognizing handwritten digits by computer causes an intent problem because of the large variability in the digit shapes by individual [3, 4]. To solve this problem system should be designed in such a way that it should have capability to read the handwritten digits and provide appropriate results.

This paper consists of six sections, where section 1 contains a brief introduction about HDR where as a brief survey has been discussed in section 2. Background related work to HDR is explained in section 3. Section 4 describes the proposed methodology. In section 5, experimental result has been shown and in section 6, conclusion of the paper is described.

2 Background and related works

D.K. Patel et.al [5] In this paper author proposed a multi-resolution method which is DWT and used Euclidean distance measurement metric to get better recognition rate. Discrete wavelet transform is used to extract features at appropriate level of multi-resolution and for getting minimum classification time class of pattern is described through mean vector. EDM is used to calculate distance of each input vector till every mean vector. Input pattern character is determined by the minimal distance calculated. By using proposed scheme result obtained with 90% recognition accuracy which is good. Recognition accuracy may improve by using other methods.

Malik et.al [7] In this paper author described a HCR method using Wavelet Transform and for classification purpose hop-field network is used. Relevant features are extracted from the images by using wavelet transform and image is decomposed at appropriate level for extracting suitable features. Evaluation has been done by using all the various distortion levels for 26 patterns. By using this method

result shows that at distortion level of 30% system identified all the characters but at 40% distortion level it recognized only some characters. So, the problem is that system was not able to identify the characters at above 40% of distortion levels.

M.C.Padma et.al [6], author used quad-tree based technique for feature extraction and KNN classifier is used to classify digit. By using this feature extraction input image is partitioned up to second level and divided into 4 quadrants Q1, Q2, Q3, Q4. It is further decomposed in 4 partition which is known as zone. To obtain consistent feature value a dataset of 3600 samples are used to train the classifier and tested on a dataset of 1200 samples. After all these training and testing result shows that the overall accuracy of the system is 85.43%.

Swapnil et.al [8], in this paper author used GRNN classifier and pre-processing techniques such as binarization and normalization to obtain accurate result. Positional feature extraction technique is used in this paper which depends on positional characteristics of particular pixels found in input images. All sample of image matrix are added, final outcome matrix is divided with sum of added matrix i.e. average matrix, after that minus it from particular sample image matrix which gives ultimate features. To obtain projection vector matrix singular value decomposition technique is used which gives better accuracy in results. The proposed method results 82.89% accuracy for devanagari character and 85.62% for kannada character.

Pasha et.al [2], in this paper author introduced a new technique for efficient feature extraction and classification of digits. The main idea behind this technique is to extract suitable and relevant feature which have some significance in output. Here structural features like Aspect ratio, Correlation, corner detection etc. and for extracting global features wavelet transform method is used. ANN classifier is applied to recognize the handwritten kannada numeral and character. Wavelet features and structural features are combined into a single feature set. For training dataset samples of 4351 are passed down to find out steady features and tested on 1450 samples. The result shows that overall accuracy obtained for kannada character is 91% and for numeral it is 97.60%.

S.horata et.al [10], in this paper histogram of oriented gradient is used for feature extraction and recognition ability of two classifiers compared with each other. DFBNN and ELM are used as classifiers. Three kind of handwritten datasets Thai, bangla and devanagari numerals are used and each dataset is divided into two parts i.e. with or without hog features. The experimental result shows that recognition rate of both the classifiers are improved by using hog feature. However, DFBNN classifier provides slightly better recognition rate than ELM classifier with all the three datasets.

Akhtar et.al [11] in this paper author proposed two acceptable approaches for feature extraction in digit recognition. Wavelet transform and wavelet packet Transform are used as feature extraction technique for extracting relevant features. Author used KNN and SVM classifiers for classification purpose and tested on MNIST dataset. The overall accuracy on wavelet transform by applying K nearest neighbour is 84.53% and with Support vector classifier the accuracy is 89.51%. Correspondingly on Wavelet Packet Transform by applying KNN the accuracy achieved is 96.24% and with SVM the accuracy

percentage is 96.29% on training the dataset. The obtained outcomes by applying these two classifiers are equivalent with each other. In this paper accuracy may improve by using other sub-bands of wavelet transform.

Lauer [12] et.al suggested a trainable extraction technique for features which is based on LeNet5 architecture. The proposed method for HDR has been tested on well-known database MNIST. This paper introduced two classifiers which are LeNet5 CNN and Support Vector machine. To maximize the generalization efficiency of LeNet5, SVM classifier is used. To get better recognition rate elastic distortion and affine transformation based new training set are generated. However, system outperforms with both the classifiers and performances of both classifiers are comparable with each other. Moreover, combining these two algorithm results into higher complexity, this is the drawback of the proposed method in certain cases.

B.EL.Quancy et.al [13], In this paper author proposed four F.E. approaches which is basically related to discrete continuous transform. The main four approaches used in this paper are: DCT upper left coefficients (ULC), block based DCT coefficients, DCT zigzag coefficients and block-based DCT coefficients. SVM classifier is used to evaluate the performance of DCT variant. MNIST database are used in two variant i.e. raw data and pre-processed data. Based on classification, accuracy, all the four approaches are compared and it has been analyzed that block based DCT zigzag feature extraction technique provides better performance than all its supplements.

3 Background work

3.1 PATTERN RECONITION

Pattern recognition consists of two stages, first one is the feature extraction stage and second one is the classification stage. Feature extraction stage is mainly used for dimensionality reduction and for obtaining relevant features for the application. Second one is classification which is the most important and essential phase for decision making. Pattern recognition is an emerging area of study which is well known since many years of research, especially when we talk about the field of digit recognition. HDR is considered as a large-scale challenge in the field of pattern recognition and getting more attention towards researchers.

3.2 NEURAL NETWORK

A neural network is basically a knowledge refining unit which is highly motivated by biological neurology system functioning as brain. It consists of a huge number of powerful inter-connected refining elements called as neurons works in coordination manner to solve distinct problems. The main objective of neural network is to process information and solve problem in the same way as the human brain does. It is used in various potential application areas like data classification areas, pattern recognition, identifying learning rate and recognition rate [14].

In neural network, hidden layers are having a vital role in generating outcome of recognition task. Hidden layer size matters a lot to classify the digits. Output layer of the neural network is highly depended upon the hidden layer and the input layer.

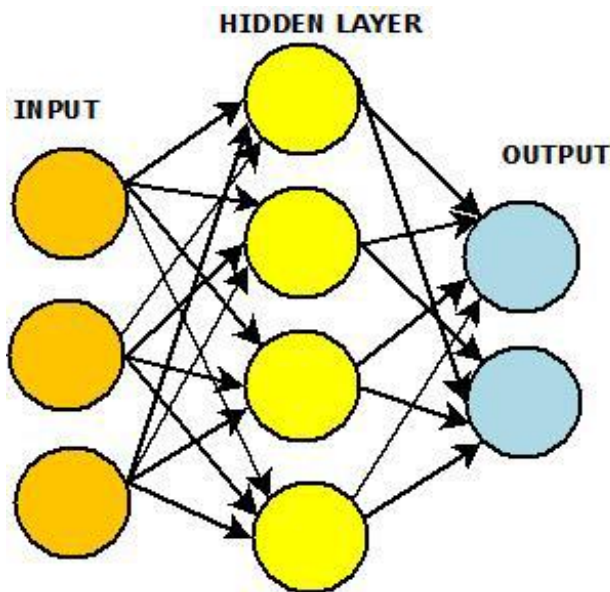


FIGURE 1 Neural network architecture

In the above diagram, there are 3 layers input layer, hidden layer and output layer which are interconnected with each other and work efficiently in a proper manner. In neural network information flows in parallel and knowledge is distributed between the neurons. In this paper, Radial Basis function and BPNN is used as a classifier for recognizing the digits.

3.3 RADIAL BASIS FUNCTION

It is a kind of artificial NN which uses radial basis function as its activation function. In this there are three layers which are as follows:

- Input layer: These are the origin nodes which join to the network to its surroundings.
- Hidden layer: These are the hidden units which provide a set of basis function and High dimensionality.
- Output layer: These are simply the linear combination of hidden functions.

It is sort of supervised neural networks which are having a shorter learning time [23]. There are several training parameters which affiliated with traingd: epochs, goal, show, min-grad, time and lr. The learning rate should be kept small for the better convergence.

3.4 BACK PROPAGATION NEURAL NETWORK

A neural network is a network with nodes of processing elements, and connections as data carriers. The architecture of neural network determines how inputs of the neural network are transformed into an output [22]. It is the most popular supervised learning multilayer feed-forward NN algorithm recommended by Rumelhart, Williams and Hinton [21]. This algorithm is the modification of least mean square algorithm by doing modification in network weights, it minimizes the mean square error between the real outcome and the target outcome. Back-propagation consists of three phases as follows:

- Forward phase
- Backward phase
- Weight update phase

3.5 STEP INVOLVED IN HANDWRITTEN DIGIT RECOGNITION

First step is to scan handwritten document, then transform it to processed image via several pre-processing techniques such as noise removal, slant angle correction, median filter etc. After pre-processing the next crucial stage is image segmentation where image will be decomposed into several sub-images, characters etc. The next step is the most important stage in any recognition process i.e. features extraction. This step is all about extracting suitable and appropriate feature because these features play an important role in training as well as testing the classifier, then the next is to classify the suitable class for input image and finally the last phase is post processing phase. This step is optional stage but sometimes improves recognition accuracy.

3.5.1 IMAGE ACQUISITION

It is also called as image scanning or image digitization. For recognition system input image acquires a scanned image through digital scanner or by capturing photograph or by directly using stylus. This captures input image may be gray scale, binary or colored but have a specific format like jpg, PNG, bmp etc.

3.5.2 PRE-PROCESSING

Pre-Processing is one of the major step and first phase in the recognition process. In this stage, raw image is transformed into a processed image and enhance the image to make it appropriate for the next stage. The raw image involves several operations or pre-processing stages like binarization, noise removal, skew detection or correction, normalization, thinning, slant angle correction. The main objective of this stage is to remove those elements which are not useful for recognition process. The main steps in Pre-Processing are as follows:

- RGB TO GRAYSCALE CONVERSION. In this process scanned images which are stored in different formats (jpeg, bmp.tiff, png) are converted into gray scale format in matrix representation form.
- BINARIZATION. This process convert coloured or gray scale image into black and white or binary image.
- NOISE REMOVAL. After scanning some images may contain various types of noises like gaps in lines, disconnected line segments, bumps, filled and unwanted types of loop which have no significance in output. So, it becomes necessary to remove such type of noise element. The main aim of this sub-step is to remove unwanted type of noise.
- NORMALIZATION. It is simply a procedure of transforming image of odd size into an image of accepted sized [2]. If image size is too large or too small then in this stage manage the image size so that all the ambiguity related to normalization of image can be removed and appropriate results have been generated for next stage.
- THINNING AND SKELETONIZATION. In this process, binary valued images which contain regions, should be reduced to pixel lines. The main objective of skeletonization is to reduce the digit area into one pixel line and produce the skeleton of the pixel.

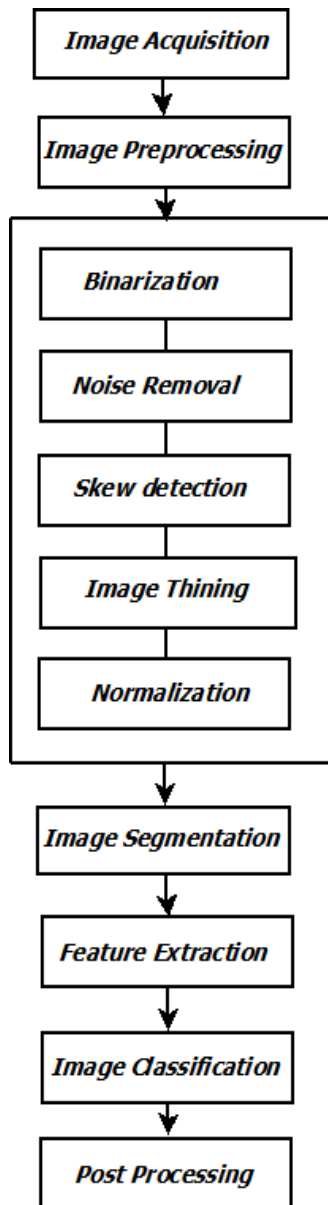


FIGURE 2 Steps of HDR

3.5.3 IMAGE SEGMENTATION

In this stage, an image which is present in the form of sequence of digits is decomposed into sub-images of respective digit. These sub-images provide information that how many number of digits contain in the image and for doing so labeling process is used in this stage by assigning a number to each digit. Accuracy of segmentation process is having a necessary role in the accuracy of final outcome in recognition process.

3.5.4 FEATURE EXTRACTION

It is the most influential steps in digit recognition process sometimes also called as the hearts of recognition process. It is most essential and crucial stage because final outcome of recognition process is very much depending upon this stage.

The main essential task in digit recognition is to extract features and then finally selecting appropriate feature from

raw data. Choosing relevant features and affiliated parameters leads to moderate errors during the classification stage. In this stage, a feature vector is assigned to every digit, so that it can be easily identified. These feature vectors are used to discriminate the digit from other digits. To acquire newness in feature extraction it can be classified into 3 categories which are as follows:

- Statistical features
 - Structural features
 - Global Transformation and series expansion
1. **STATISTICAL FEATURES.** Statistical features are generally depending upon hypothesis. Probability theory, Zoning, loci, distance are the main Statistical features.
 2. **STRUCTURAL FEATURES.** Structural features involve that type of features which contain information related to the structure of character like loops, joints, branches, curve, crossing points, aspect ratio, strokes etc.
 3. **GLOBAL TRANSFORMATION AND SERIES EXPANSION.** Global transformation features implicate global deformations like rotation, translation etc. Hough transform, Gabor transform, wavelet transform is based on global transformation features.

3.5.5 CLASSIFICATION

In each digit recognition classification is the most essential decision making stage to analyze or identify the digits. As per to preset rules it has used the features extracted in the earlier stage to determine the digits. Several types of classifiers are used to classify the digits. These classifiers correlate the input feature with the pattern which are previously stored and identify the best appropriate class for the input.

3.5.6 POST PROCESSING

The main objective of this stage is to reduce or remove such errors which contain irrelevant information and also to find out the system even if it gives required outcomes or not, although this step is not so compulsory step, but sometimes this stage helps to improve the accuracy of the recognition processes.

4 Methodology

In this paper, a novel method is proposed that is we have combined Histogram of oriented gradient and Haar wavelet together for extracting of features and after that we have applied the RBFNN and BPNN as a classifier.

4.1 HISTOGRAM OF ORIENTED GRADIENT

Histogram of oriented gradients feature descriptors are extensively used in image processing and computer visions for the recognition purpose. It was first recommended by Dalal and Trigg's for detection of human body but having a great advantage over other descriptor it is now used in the area of computer vision and pattern recognition.

The main aim behind Hog descriptor is that local or confined shape information and appearance of object within an image can be determined by intensity gradient distribution or edge directions. As the Hog feature produce on localized

block, it justifies invariance to photometric and geometric transformation makes slight changes if they are much smaller than the orientation bin size or local spatial [18].

It decomposes image into some small sub-images often known as cells, these cells can rectangular or circular in shape i.e. R-HOG and C-HOG respectively. For each cell histogram of gradient direction is computed or histogram of edge orientation for the pixels within the cell is computed. The combine histograms are used as feature vector for describing the object. To get appropriate accuracy local histogram are normalized based on contrast, that's why Hog is stable on illumination variation. The main reason to use this feature is that it extracts histogram feature more efficiently which will further used in classification and recognition stage. In handwritten digit recognition, it can capture edge or gradient structure and gives information about the shape of the digit.

4.1.1 Use of HOG

- The main reason to use this feature is that it extracts histogram feature more efficiently which will be further used in classification and recognition stage.
- In comparison to SIFT and LBP, Hog is a quick and fast type of descriptor and being simple in computation Hog features are successful descriptor.
- In handwritten digit recognition, it can capture edge or gradient structure and gives information about the shape of the digit.

Hog feature extraction process consist several several steps.

First a scanned image is taken as input then normalization, smoothing of images are done in next step if images are too large or too small then apply resizing of images. In next step, gradient computation is done after that full image is divide into small regions, these regions are called as cells. After dividing images into cells compute the Hog features of each cell.

Then histogram of each cell is computed. For generating feature vector of each cell the computed histogram of each cell is combined. After all that above processes feature vector is generated.

These feature vectors contain necessary and useful information related to the image. From these feature vectors appropriate and relevant information should be extracted that can be very useful in outcome of digit recognition and having a great significance in final outcome.

4.2 WAVELET

Wavelets are the mathematical functions, which are most widely used in the area of signal analysis and image processing. The major advantage of using wavelet is that it cut up image or signals into distinct frequency components and after that it interpret and brief study about each component with a resolution equivalent to its scale.

Wavelets are used as the basis of multi-resolution process. Wavelet transform are of many types, the most basic wavelet transform is HAAR wavelet transform. Wavelets transforms are achieved through quadrature mirror filters. Two types of filters are used one is the high pass filter and other is the low pass filter. These filters are applied to every row/column of an image to decompose it

into the four-appropriate frequency sub-bands.

As shown in the figure 4: A1 represents approximation coefficient, H1 represents horizontal, V1 represents vertical and D1 represents diagonal coefficients.

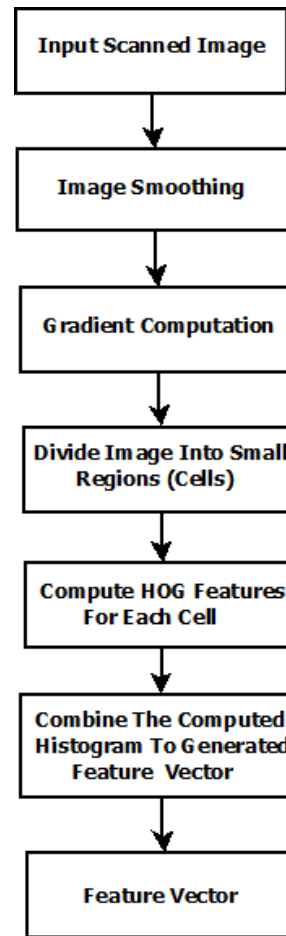


FIGURE 3 HOG feature extraction process

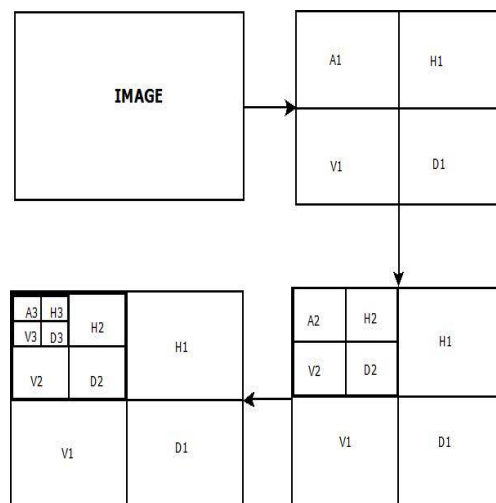


FIGURE 4 Wavelet decomposition of an image

Wavelet decomposition divides an image into approximation and detail coefficients. In case of wavelet decomposition approximation coefficient is further divided into other approximation and detail coefficients but detail coefficient remains same in this case. In case of wavelet packet decomposition approximation and detail both

coefficients are further divided. After wavelet decomposition, hierarchical type structure is obtained because approximation coefficients are divided again and again as levels are increases. Down-sampling is used in wavelet decomposition in each and every level. It will help in wavelet decomposition to decompose the images at every stage.

4.2.1 HAAR WAVELET TRANSFORM

Haar wavelet is one of the straightforward and smooth types of wavelet. It will provide a good foundation for understanding the more sophisticated wavelet transform. The main function of Haar is to compress signals and remove noises. Haar wavelet is a procession of rescaled 'square shaped' function which form a wavelet basis or wavelet family after combining. Haar is also known as Db1 wavelet. To approximate a function Haar used a square pulse as a wavelet.

Any continuous real function with compressed support can be approximated consistently by linear combinations of their shifted function. Haar wavelet transform is a process of decomposition or transform which uses recursive averaging and differencing. The overall data is shown through the resultant coefficients which are required for an image construction. Based on these outputs, are known as detail coefficients and these outputs helps in building the wavelet basis function. The main function of this wavelet is that if it necessary to reconstruct previous levels then construction of any decomposition is possible using averaging process. By using the process of averaging levels can be reconstructed easily.

- In haar the input and output length are same, though the length should be a power of 2, i.e. $N=2^n$, $n \in \mathbb{N}$.
- It can be used to analyze the localized feature of signals.
- No requirement for multiplication, it needs only additions and in HAAR matrix there are many elements with zero value, that's why it will take less computation time.

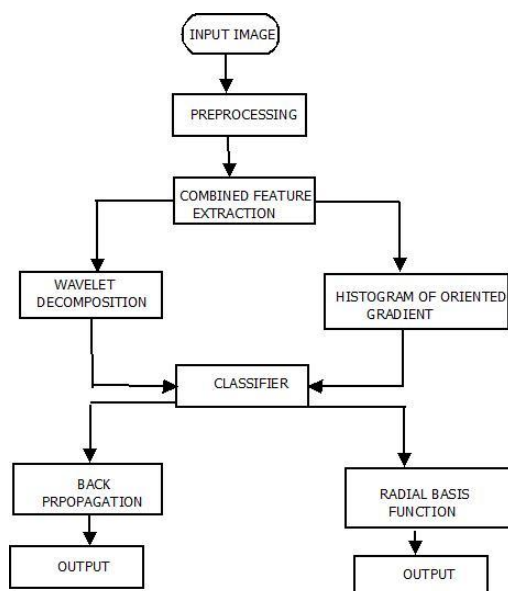


FIGURE 5 Flowchart of proposed method

Handwritten digit recognition learning model is implemented as supervised learning model. In supervised learning model, there are mainly two phases in which it highly depends, first one is training phase and the second

one is testing phase. First of all, from handwritten digits' images i.e. training set images features are extracted using haar wavelet decomposition and histogram of oriented gradients. Wavelet decomposition decomposes images into 4 coefficients which are horizontal (h), vertical (v), approximation (a), diagonal (d) and similarly hog features of digits are obtained. After that these feature vector is taken as input to radial basis function NN and back-propagation network, so that weights of network get optimized. All this is done in training phase. In testing phase, the features of handwritten digits are extracted and feature vector of test image is given to decision unit. This unit is important for making decision and recognizing handwritten digits.

4.2 STEPS FOR PROPOSED WORK

Handwritten digit recognition steps are as follows:

Input: Image containing handwritten digits

Output: Recognized handwritten digits

Step 1: Input image is obtained from the pre-processed image.

Step 2: Perform the feature extraction process by using histogram of oriented gradient feature and haar wavelet transform method.

Step 3: Pre-processed image contains Hog feature vector and wavelet coefficients.

Step 4: Perform classification for recognizing digits using radial basis function and back-propagation classifier by using these features.

5 Result and analysis

This section represents the detailed implementation being conducted by the recommended method on MNIST database. This dataset is an extensively revolved benchmark, which dwell 42000 images of training data set and 28000 images of testing data set. Data files which consists of zero to nine hand-drawn digits are represent in the form of gray-scale images. In this every image consist of 28*28 pixels so a total 784 pixels are there in this dataset. A single pixel value that ranges from 0-255 is contained by each pixel corresponding with it, which specifies the darkness or lightness of that pixel.

We have taken 100 samples images of MNIST dataset and then train and test the dataset in matlab2016a environment. One of the sample images of MNIST dataset is shown below in the following figure.

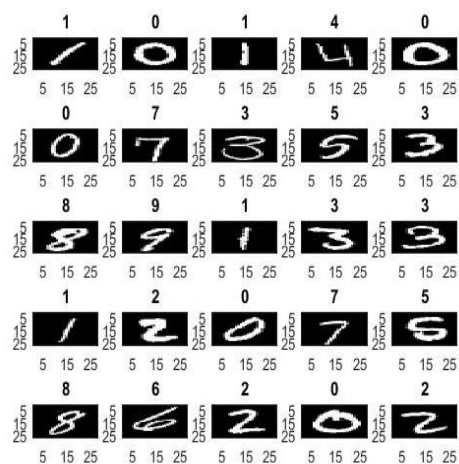


FIGURE 6 Sample image of MNIST dataset

Hog and wavelet features of digits are obtained after applying feature extraction method which are histogram of oriented gradient and HAAR wavelet transform. Extracted HOG features of some digits are shown below in the figure.

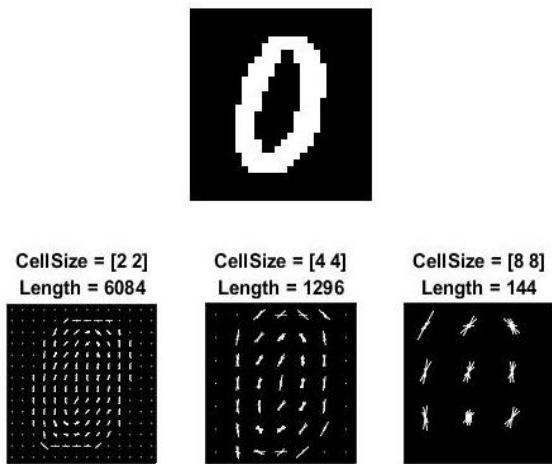


FIGURE 7 HOG feature of digit 0

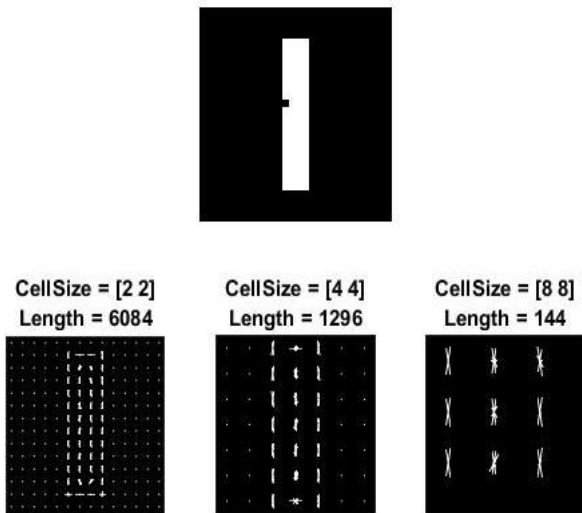


FIGURE 8 HOG feature of digit 1

Wavelet feature of some digits are shown below in the figure in which the four wavelet coefficients approximation coefficient, diagonal coefficient, horizontal coefficient, vertical coefficient are shown.

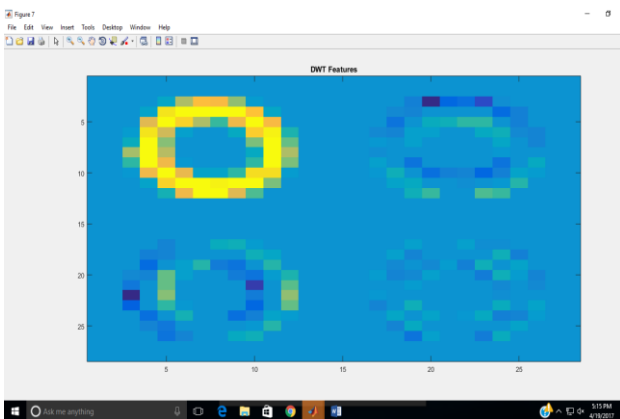


FIGURE 9 Wavelet feature of digit 0

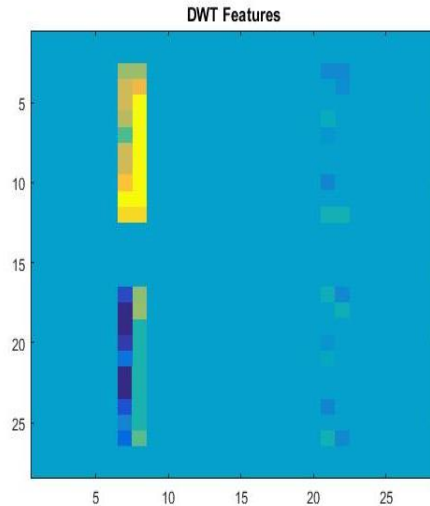


FIGURE 10 Wavelet feature of digit 1

After extracting relevant information from suitable Hog and HAAR features, appropriate classifiers such as back-propagation and RBF neural network are applied to find out the accuracy of the system.

We obtained following results of back-propagation neural network by using nntoolbox.

Regression plot, performance plot, error histogram plot and training state plot are as follows:

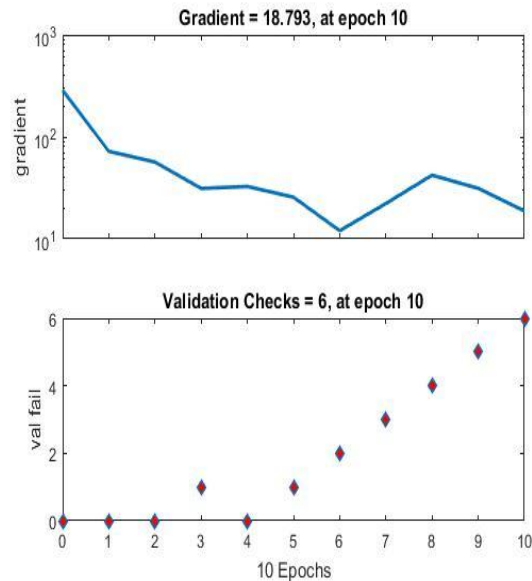


FIGURE 11 Plot of training state

In table 1 training time of both the classifier is given and from this it is clear that training time of radial basis function neural network is less as compare to back-propagation neural network.

TABLE 2 Analysis of both classifiers

Neural Network Classifier	Training Time	Accuracy
Back-Propagation neural network	13.4952	83.66%
Radial Basis Function neural network	13.3962	98.26%

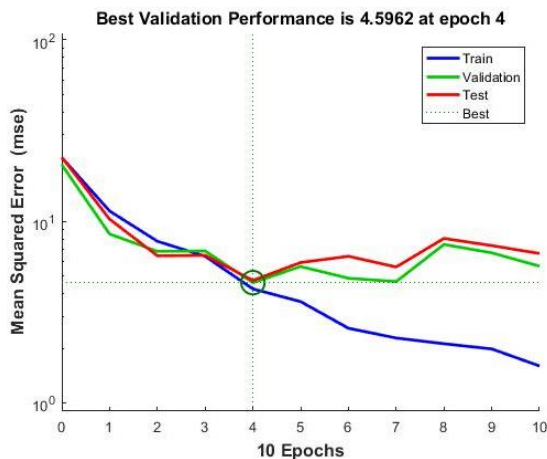


FIGURE 12 Validation performance plot

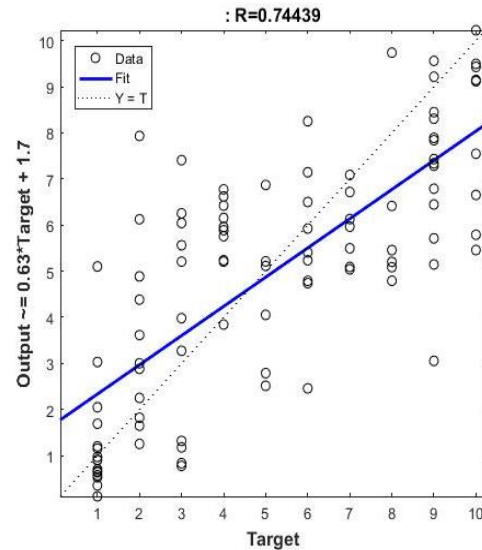


FIGURE 13 Regression plot

6 Conclusion

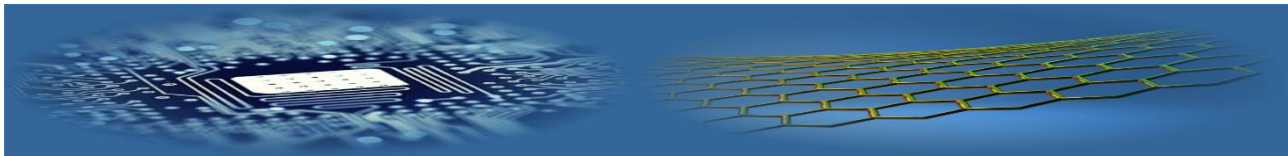
In this paper, Haar wavelet transform and Histogram of oriented gradient is used for feature extraction technique and classification task is handled using Radial Basis Function NN and back-propagation neural network. Main goal of the proposed work is to make the system which recognizes the

References

- [1] Dhande C K, Mahajan P M 2014 A survey on evaluating neural network and hidden markov model classifiers for handwriting word recognition. *International Journal of Advances in Engineering & Technology* 6(6) 2427
- [2] Pasha S, Padma M C 2015 December. Handwritten Kannada character recognition using wavelet transform and structural features *In Emerging Research in Electronics, Computer Science and Technology (ICERECT), 2015 International Conference on (pp. 346-51) IEEE*
- [3] Yong Haw Tay', Pierre-Michel Illicad, Marzuki Khalid', Christian Viard-Gaudin Stefan Knerr 2001 *Offline Handwritten Word Recognition Using A Hybrid Neural Network And Hidden Markov Model (ISSPA)*, Kuala Lumpur, Malaysia, 13 - 16 August, 2001
- [4] Fu Chang, Chin-Chin Lin and Chun-Jen Chen *Applying a hybrid method to handwritten character recognition* Institute of Information Science, Academia Sinica, Taipei, Taiwan
- [5] Patel D K, Som T, Singh M K 2013 Multiresolution technique to handwritten English character recognition using learning rule and Euclidean distance metric *International Conference On Signal Processing And Communication (ICSC), Noida* 207-12
- [6] Padm, M C, Pasha S 2014 Quadtree based feature extraction technique for recognizing handwritten Kannada characters *In Emerging Research in Electronics, Computer Science and Technology* 725-35 Springer, India
- [7] Malik P, Dixit R 2013 Handwritten character recognition using wavelet transform and hopfield network *In Machine Intelligence and Research Advancement (ICMIRA), 2013 International Conference* 125-9 IEEE
- [8] Vaidya S A, Bombade B R 2013 A novel approach of handwritten character recognition using positional feature extraction *International Journal of Computer Science and Mobile Computing* 2(6) 179-86
- [9] Romero D J, Seijas L, Ruedín A 2007 Directional continuous wavelet transform applied to handwritten numerals recognition using neural networks *Journal of Computer Science & Technology* 7
- [10] Iamsa-at S, Horata P 2013 Handwritten character recognition using histograms of oriented gradient features in deep learning of artificial neural network *In IT Convergence and Security (ICITCS), 2013 International Conference* 1-5 IEEE
- [11] Akhtar M S, Qureshi H A 2013 Handwritten digit recognition through wavelet decomposition and wavelet packet decomposition *In Digital Information Management (ICDIM), 2013 Eighth International Conference* 143-8 IEEE
- [12] Lauer F, Suen C Y, Bloch G 2007 A trainable feature extractor for handwritten digit recognition *Pattern Recognition* 40(6) 1816-24
- [13] Qacimy B El, Ait kerroum M, Hammouch A 2014 Handwritten digit recognition based on DCT features and SVM classifier *Second World Conference on Complex Systems (WCCS), Agadir* 13-6
- [14] Kusum Gupta 2015 Neural network handwritten digit recognition using various neural network approaches sakshica1 *IJARCCCE* 4(2)
- [15] Devijver P A, Kittler J 1982 *Pattern recognition, a statistical approach*, Prentice Hall London 480
- [16] Kim K K, Kim J H, Suen C Y 2002 Segmentation-based recognition of handwritten touching pairs of digits using structural features *Pattern Recognition* 23(1) 13-24
- [17] Gattal A, Chibani Y 2015 SVM-Based Segmentation-Verification of Handwritten Connected Digits Using the Oriented Sliding Window *International Journal of Computational Intelligence and Applications (IJCIA)* 14(1) 1-17
- [18] Dalal N, Triggs B 2005 Histograms of oriented gradients for human detection. *In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on* 1 886-93. IEEE

- [19] Mallat S 2008 *A wavelet tour of signal processing* Academic Press, Burlington
- [20] Gattal Abdeljalil, Chawki Djeddi, Youcef Chibani, Imran Siddiqi 2016 Isolated handwritten digit recognition using oBIFs and Background features *12th IAPR Workshop on Document Analysis Systems (DAS)*
- [21] Rumelhart D E, Hinton G E, Williams R J 1986 *Learning internal representations by error propagation*
- [22] Singh D, Singh S K, Dutta M 2010 Handwritten character recognition using twelve directional feature input and Neural Network *International journal of computer applications* 3 975-8887
- [23] Singh D, Saini J P, Chauhan D S 2015 Hindi character recognition using RBF neural network and directional group feature extraction technique *In Cognitive Computing and Information Processing (CCIP), 2015 International Conference on (pp. 1-4) IEEE*

AUTHORS	
	<p>Ankita Mishra</p> <p>Current position, grades: Post graduation student. University studies: Madan Mohan Malaviya University of Technology Gorakhpur, India Scientific interests: Image Processing, Neural Network, Cloud Computing</p>
	<p>Dayashankar Singh</p> <p>Current position, grades: Assistant Professor, CSED, M.M.M. University of Technology, Gorakhpur University studies: Panjab University, Chandigarh, India Scientific interests: Neural Network, Artificial Intelligence, Image Processing, Database, Network Security</p>



Fuzzy comprehensive evaluation model for mobility model

Yao Minghui, Zhang Sheng*

School of Information Engineering, Nanchang Hangkong University, 696 Fenghe South Ave., Nanchang, China, 330063

**Corresponding author's e-mail: zwxzs168@126.com*

Received 1 March 2017, www.cmnt.lv

Abstract

Evaluation of mobility model is an important means to ensure the quality and design level. At present, many mobility models are proposed for opportunistic networks. But, there is no practical quantitative evaluation system to evaluate the mobility models. Firstly, this paper put forward a comprehensive evaluation index system of mobility model based on the analysis of the main factors affecting the quality of mobility model and the relationship between them. Secondly, based on the theory of fuzzy comprehensive evaluation, this paper put forward a fuzzy comprehensive evaluation model for mobility model (FCEM). In this model, the membership function of fuzzy mathematics is used to deal with the fuzzy evaluation of each index of the mobility model. The model realizes the quantitative evaluation of mobility model. This model not only provides new ideas and methods for mobility model evaluation, but also provides help and guarantee for mobile node modelling. Finally, the application of the model is demonstrated through the evaluation of the random waypoint (RWP) model.

Keywords:

mobility model,
evaluating indicator,
membership function,
fuzzy comprehensive evaluation

1 Introduction

Opportunistic network is a mobile ad hoc network that does not require a complete link between the source node and the target node, and uses the opportunity of the meeting to communicate [1]. Mobility model is the basis of network protocol, network topology and network security. Different mobility models have different effects on network performance. The rationality of model plays an important role in the design of protocol parameters. Model evaluation is an important means to evaluate the rationality of a model. Therefore, more and more researchers pay attention to the evaluation methods of mobility model. The evaluation of mobility model can provide objective index and evaluation method for the construction and analysis of mobility model. The evaluation of mobility model can also guide the application of the model. The application of mobility model is diverse, and different scenes have different requirements for mobility model. Therefore, it is very difficult to evaluate the quality of mobility model. Many factors need to be considered in the design and evaluation of mobility models. At present, many mobility models are proposed for opportunistic networks. Most of the models are evaluated by comparative analysis. On the one hand, this evaluation method is only a single index evaluation and it is not comprehensive. On the other hand, this kind of evaluation method is very fuzzy. There is no practical quantitative evaluation system to evaluate the mobility models.

The rest of the paper is organized as follows: In Section 2, recent work on evaluation of mobility model is reviewed. The evaluation index of mobility model is introduced in detail in Section 3. In Section 4, the applicability of the evaluation model is proved through an application example. We conclude the paper and point out future work in Section 5.

2 Related work

At present, there are few researches on evaluation of mobility model. There is not a comprehensive and complete evaluation model to evaluate the mobility model. In the analysis of routing algorithm in the paper [2], evaluation system of the mobility model is proposed based on the physical characteristics, topological characteristics and network performance. However, the system did not present the evaluation indicators and did not do a detailed analysis. In the study of the group mobility model, the group mobility model is evaluated from the physical characteristics of the nodes in the paper [3]. In the literature [4], the calculation model of link duration is introduced in detail. It evaluates the model based on network link duration. In the literature [5], the mobility model is evaluated from the aspects of node velocity distribution, node distribution, node connectivity, and node motion trajectory. The evaluation model is based on the meeting time, the time interval of the meeting and the controllability of the parameters in the paper [6]. These methods are not universal and comprehensive.

At the same time, fuzzy comprehensive evaluation has a very sound theoretical system and has been successfully applied to many fields. Such as: the quality of software [7], quality of the paper [8], information systems [9, 10] and so on. However, there is no application of fuzzy comprehensive evaluation in mobility model.

In this paper, based on the theory of fuzzy comprehensive evaluation and combined with the characteristics of mobile nodes, the evaluation index system of mobility model is proposed. A practical evaluation model of mobility model is designed based on the membership function of fuzzy mathematics theory. This model not only provides new ideas and methods for evaluation of mobility model, but also provides help and guarantee for mobile node modelling.

3 Evaluation indexes

In order to provide an effective evaluation on mobility model, the main factors affecting the mobility model must be determined first. Then, it is necessary to establish a systematic, comprehensive index system according to the divided layers of these factors. As you know, mobility model can be evaluated from many factors or main indexes, which are further composed of some sub-indexes. After balancing seriously among all factors affecting mobility model, a general comprehensive evaluation index system with two-layer indexes is illustrated in figure 1.

Some notations are introduced: the evaluation objective, mobility model, is denoted by M ; the index set $U =$ (authentic (u_1), space-time (u_2), connectivity (u_3), routing (u_4)); in the first layer, and in the second layer, $u_1 =$ (similarity (u_{11}), parameter controllability (u_{12})), $u_2 =$ (node distribution (u_{21}), spatial dependence (u_{22}), velocity distribution (u_{23}), velocity dependence (u_{24}), temporal dependence (u_{25})), $u_3 =$ (average number of link changes (u_{31}), connection duration (u_{32}), time interval (u_{33})), $u_4 =$ (successful delivery ratio (u_{41}), average latency (u_{42})).

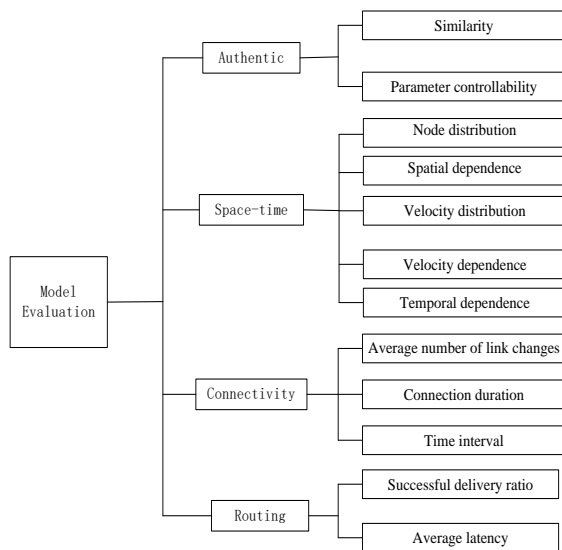


FIGURE 1 Evaluation indexes of mobility model

Obviously, the two-layer comprehensive evaluation index system in figure 1 is characterized by the multi-granular space. The higher layer is, the coarser granularity is. As the evaluation process moves from high layer to low layer, the information granularity to the comprehensive evaluation becomes finer. Therefore, the two-level evaluation index is comprehensive and complete.

4 FCEM

Due to the diversity and complexity of mobility models, there are many uncertain influencing factors. At the same time, these factors may be related to each other, so that we can not accurately determine the quality of the mobility model. So, there is no practical quantitative evaluation system. Fuzzy mathematics is usually used to study fuzzy

problems. Fuzzy comprehensive evaluation method is a comprehensive evaluation method based on Fuzzy Mathematics. This method is to make a comprehensive evaluation of various evaluation factors. It can transform the qualitative evaluation into quantitative evaluation based on the theory of membership degree. This method has the following advantages: 1) the evaluation result is clear. 2) The system is strong. 3) It can solve the problem of fuzzy and difficult to quantify.

4.1 GENERAL PROCESS

Generally, the fuzzy comprehensive evaluation model includes the following factors: factor set, evaluation set, weight assignment set, evaluation matrix. In order to describe, according to the basic concept of fuzzy mathematics, the terms are defined as follows.

1) **Factor set (U)**. Factor set is a collection of various factors that influence the evaluation object.

$$U = \{u_1, u_2, \dots, u_m\},$$

Where u_i ($i = 1, 2, \dots, m$) is the factor affecting the mobility model.

2) **Evaluation set (V)**. The evaluation set is a collection of evaluation results.

$$V = \{v_1, v_2, \dots, v_n\},$$

where v_j ($j = 1, 2, \dots, n$) is the result of evaluation.

Based on the existing Fuzzy comprehensive evaluation research, the mobility model grade is divided into grades of A, B, C, D, E respectively, corresponding to the mobility model grade is best, better, general and worse, worst the rank score is set to 100 points, the grade of mobility model as shown in the table below.

TABLE 1 The grade of mobility model

Grade	A	B	C	D	E
Rank	Best	better	General	Worse	Worst
Point	100~90	90~80	80~70	70~60	60~0

3) **Weight assignment set (A)**. The weight assignment set is the collection of the proportion of each factor in the evaluation.

$$A = \{a_1, a_2, \dots, a_m\},$$

where a_i ($i = 1, 2, \dots, m$) is the proportion of i factors in

model evaluation. $\sum_{i=1}^m a_i = 1, 0 < a_i < 1$.

We determine the weights using the analytic hierarchy process (AHP) uses qualitative and quantitative systematically analysis methods. At present, it has been widely used in many fields [11, 12]. Its key steps are as follows. The judgment matrix (A') is given by expert or decision maker according to the scale of judgment, and then construct comparison judgment matrix to calculate the weights. In judgment matrix the eigenvector of the maximum eigenvalue is the weight vector of the system.

4) **First level fuzzy evaluation (R)**. The evaluation of each factor set is a fuzzy mapping. Different factors will have different evaluation results. The evaluation matrix is constructed from the mapping of the factor set to the evaluation set.

$$R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & r_{mn} \end{bmatrix}$$

5) **Multilevel fuzzy evaluation.** Multi level fuzzy comprehensive evaluation means that fuzzy evaluation can be divided into several grades. The results of the fuzzy evaluation of the upper level fuzzy evaluation vector are normalized to synthesize the evaluation matrix. As shown in the following figure 2.

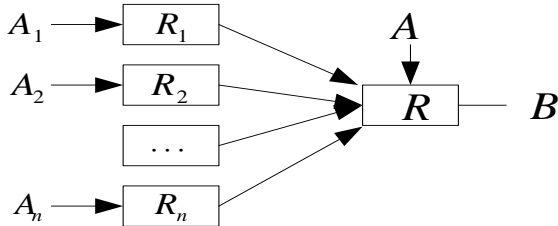


FIGURE 2 Sketch map of two-level fuzzy comprehensive evaluation model

6) **Evaluation results (B).**Based on the above analysis,

TABLE 2 weights and evaluation indexes

Index	Weight	Index	Weight	Evaluation set				
				A	B	C	D	E
u_1	0.5	u_{11}	0.7	0.1	0.5	0.2	0.1	0.1
		u_{12}	0.3	0.3	0.3	0.2	0.1	0.1
u_2	0.2	u_{21}	0.3	0.1	0.4	0.3	0.2	0
		u_{22}	0.2	0	0.1	0.2	0.5	0.2
		u_{23}	0.1	0	0.2	0.4	0.2	0.2
		u_{24}	0.1	0	0.1	0.2	0.4	0.3
		u_{25}	0.3	0	0.2	0.3	0.3	0.2
u_3	0.2	u_{31}	0.5	0.1	0.3	0.3	0.2	0.1
		u_{32}	0.2	0	0.2	0.5	0.2	0.1
		u_{33}	0.3	0	0.3	0.5	0.1	0.1
u_4	0.5	u_{41}	0.6	0.2	0.4	0.2	0.1	0.1
		u_{42}	0.4	0.3	0.3	0.2	0.1	0.1

Based on the above indexes information, the weight and evaluation vector is constructed. First level weight matrix is as follow. The weight vector of u_1 is $A_1 = (0.7, 0.3)$. From u_2 to u_4 , the weight vectors are $A_2 = (0.3, 0.2, 0.1, 0.1, 0.3)$, $A_3 = (0.5, 0.2, 0.3)$, $A_4 = (0.6, 0.4)$. The weight of U is $A = (0.5, 0.2, 0.2, 0.1)$. The evaluation matrix is as follows from u_{ij} to V .

$$R_1 = \begin{bmatrix} 0.1 & 0.5 & 0.2 & 0.1 & 0.1 \\ 0.3 & 0.3 & 0.2 & 0.1 & 0.1 \end{bmatrix}$$

$$R_2 = \begin{bmatrix} 0.1 & 0.4 & 0.3 & 0.2 & 0 \\ 0 & 0.1 & 0.2 & 0.5 & 0.2 \\ 0 & 0.2 & 0.4 & 0.2 & 0.2 \\ 0 & 0.1 & 0.2 & 0.4 & 0.3 \\ 0 & 0.2 & 0.3 & 0.3 & 0.2 \end{bmatrix}$$

we can get the final vector of the multilevel fuzzy evaluation. There are two kinds of methods to judge the evaluation results, the maximum membership principle and the weighted average principle. In this paper, the maximum membership principle is used to evaluate the results.

$$B = A \circ R = [a_1, a_2, \dots, a_m] \circ \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & r_{mn} \end{bmatrix} = [b_1, b_2, \dots, b_n] \cdot (1)$$

Sign (\circ) is a fuzzy synthesis operator. We take the matrix multiplication.

4.2 ANALYSIS OF APPLIED EXAMPLES

Based on the above research results, this paper takes the random waypoint mobility model as an example to illustrate the effectiveness of the proposed method. The secondary indicators in mobility model evaluation index conclude both quantitative indicators and qualitative indicators, so need to combine expert consultation method and fuzzy membership function to calculate the membership. The results are shown in the following table.

$$R_3 = \begin{bmatrix} 0.1 & 0.3 & 0.3 & 0.2 & 0.1 \\ 0 & 0.2 & 0.5 & 0.2 & 0.1 \\ 0 & 0.3 & 0.5 & 0.1 & 0.1 \end{bmatrix}$$

$$R_4 = \begin{bmatrix} 0.2 & 0.4 & 0.2 & 0.1 & 0.1 \\ 0.3 & 0.3 & 0.2 & 0.1 & 0.1 \end{bmatrix}$$

Because all of the evaluation matrixes of u_{ij} are obtained, rank the evaluation matrixes corresponding to each index in u_{ij} . According to the formula (1), we can get the results of the first level fuzzy evaluation.

$$B_1 = A_1 * R_1 = (0.37 \quad 0.44 \quad 0.2 \quad 0.1 \quad 0.1), \quad (2)$$

$$B_2 = A_2 * R_2 = (0.03 \quad 0.23 \quad 0.28 \quad 0.31 \quad 0.15), \quad (3)$$

$$B_3 = A_3 * R_3 = (0.05 \ 0.28 \ 0.4 \ 0.17 \ 0.1), \quad (4)$$

$$B_4 = A_4 * R_4 = (0.24 \ 0.36 \ 0.2 \ 0.1 \ 0.1), \quad (5)$$

Based on the above results, we can get the second level evaluation matrix.

$$R = \begin{pmatrix} B_1 \\ B_2 \\ B_3 \\ B_4 \end{pmatrix} = \begin{bmatrix} 0.37 & 0.44 & 0.2 & 0.1 & 0.1 \\ 0.03 & 0.23 & 0.28 & 0.31 & 0.15 \\ 0.05 & 0.28 & 0.4 & 0.17 & 0.1 \\ 0.24 & 0.36 & 0.2 & 0.1 & 0.1 \end{bmatrix}$$

According to the formula (1), the fuzzy comprehensive evaluation result for evaluation objective M .

$$B = A * R = (0.225 \ 0.358 \ 0.256 \ 0.156 \ 0.11). \quad (6)$$

Finally, we use the weighted sum method to transform (B) into a concrete numerical value. That is an average score set $S = (95, 85, 75, 65, 30)$ is assigned for V , then the weighted sum of scores is as follows.

$$P = B \circ S^T = 840445. \quad (7)$$

From the above comprehensive evaluation value, we can draw a conclusion that the RWP model is better.

More examples about fuzzy comprehensive evaluation are not listed here because of limitations of paper length. However, all the results show that the FCEM can be effectively used to the comprehensive evaluation mobility model.

Reference

[1] Xiong Y P, Sun L M, Niu J W, et al. 2009 Opportunistic networks *Journal of Software* **20**(1) 124-37
 [2] Bai F, Sadagopan N, Helmy A 2003 The important framework for analyzing the impact of mobility on performance of routing for ad hoc networks *Ad Hoc Networks* **1**(4) 383-403
 [3] Hou Yan-shun, Sun Jia-qi, Wang Xiao-bo 2014 Research of representative group mobility models *Computer Science* **41**(s2) 174-7 (in Chinese)
 [4] Tian Guang-li, Cai Wan-dong, Wang Wei 2008 Calculating model of link duration in mobile ad hoc networks *Computer Engineering* **34**(12) 82-4 (in Chinese)
 [5] Kim K, Choi H 2010 A mobility model and performance analysis in wireless cellular network with general distribution and multi-cell model *Wireless Personal Communications* **53**(2) 179-98
 [6] Gao Yuan, Wang Shu-min, Sun Jian-fei 2015 Node mobility model based on user interest similarity *Journal of Computer Applications* **35**(9) 2457-60 (in Chinese)
 [7] Wei L, Xiao L, Wuyi Y, et al. 2008 The research and appliance of

5 Conclusions

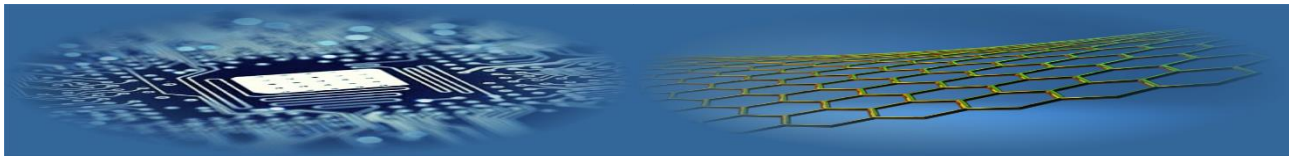
Mobility model evaluation is an important issue in mobility model research. Firstly, this paper studies the characteristics of mobility model and the characteristics of model evaluation. In this paper, we propose the mobility model evaluation system, which is based on the temporal and spatial characteristics of mobility model nodes, the network topology, the authenticity of the model and the influence of the model on the network performance. The model evaluation structure is introduced in detail, and the detailed definition and calculation method of each index are given. Secondly, on the basis of this, the paper puts forward a fuzzy comprehensive evaluation model by using the fuzzy evaluation theory of fuzzy mathematics. This model not only provides new ideas and methods for mobility model evaluation, but also provides help and guarantee for mobile node modelling. Finally, the application of the model is demonstrated through the evaluation of the random waypoint (RWP) model.

Acknowledgements

This work is partially supported by the National Natural Science Foundation of China (61162002, 61661037), the Jiangxi province Natural Science Foundation (20151BAB207038), the Nanchang Hangkong University graduate innovation special foundation (YC2016012).

multilayer fuzzy comprehensive evaluation in the appraisal of software quality *IEEE International Symposium on Knowledge Acquisition and Modeling Workshop, Kam Workshop. IEEEExplore, 2008* 617-20
 [8] Wang L X, Cen T T, Yu J 2008 A multilevel fuzzy comprehensive evaluation model for test paper quality *IEEE International Conference on Granular Computing. IEEE, 2008* 616-9
 [9] Li E, Ma Y, Xu G 2000 Fuzzy and analytic hierarchy process models for comprehensive evaluation of information systems *Journal of the China Society Forentific & Technical Information*
 [10] Xiao L, Dai Z K 2004 Model of multilevel fuzzy comprehensive risk evaluation of information system *Journal of Sichuan University* **36**(5) 98-102
 [11] Wang Jingbo, Liu Lijuan 2011 The Coal Enterprise's Performance Measure Based on the Balanced Scorecard *Value Engineering* **30**(10) 107-9
 [12] An Lihua, Xu Qianjun 2012 Core Enterprise Performance Evaluation Based on Fuzzy Comprehensive Evaluation *Logistics Technology* **9** 299-301

AUTHORS	
	<p>Minghui Yao, March 1991, China</p> <p>Current position, grades: student at Nanchang Hangkong University, China University studies: Nanchang Hangkong University, China Scientific interest: wireless sensor networks and opportunistic networks Publications: 3 Experience: more than 2 years</p>
	<p>Sheng Zhang, December 1968, China</p> <p>Current position, grades: researcher at Nanchang Hangkong University, China. University studies: PhD degree in Geodesy and Survey Engineering from Institute of Geodesy and Geophysics, Chinese Academy of Sciences in 2006. Scientific interests: GPS/GIS, artificial intelligence, cyber-physical systems and wireless sensor networks Publications: 60 Experience: more than 10 years</p>



Enhance detecting and preventing scheme for ARP Poisoning using DHCP

Vidya Srivastava, Dayashankar Singh*

M. Tech Student. Deptt. Of CSE MMMUT, Gorakhpur (UP), India

**Corresponding author's e-mail: dss_mec@yahoo.co.in*

Received 13 April 2017, www.cmnt.lv

Abstract

The client which is using LAN for mapping network address connected to its corresponding MAC address is done by Address Resolution Protocol, which is a primary protocol. It is well known that ARP is determined and works properly in case there is no malignant client in the network but in practical scenario it is not possible. The primary motive of an attacker is always tried to find a strategy which is further accomplished to launch various attacks. ARP gives this accountability – the unsubstantiated and stateless characteristics of the protocol which accredit the attacker to conduct biggest level attacks. In this paper, an attempt is made to resolve out or minimize the attempt of attacker by providing a validation using DHCP server. By the introduction of DHCP (Dynamic host control protocol) such that if an attacker applies the IP of host not in network can be prohibited. The simulation result has been shown in the dissertation report. By the response of DHCP correct matching of IP and MAC could only respond and thus poisoning can be detected and protected successfully.

Keywords:

Address Resolution Protocol,
Network security,
MiTm

1 Introduction

In network layer address resolution protocol is described by RFC [1] (Request for comment) resides within data link layer. For resolving the logical address into physical address. In second layer of OSI that is data link layer and network layer ARP works like an interface for finding the address of any node. Process is done when a specific information send to destination node, these information consist IP and MAC address. Generally ARP messages include ARP request and reply message. ARP request message used for sending MAC (physical address) corresponding to their logical address. Response message is used for retrieval information from host. And when host receive the response message the upgrade their primary cache with their IP-MAC binding. For communication purpose host use IP address of destination host. Logical address is responsible for the purpose of communication over an interface. In LAN environment address resolution protocol plays an important role. But due to the limitation of ARP called loopholes it becomes a serious attack such as MiTm, denial of services attack, bombing attack [2] etc. A host reject the communication to make dupe host. Attacker that are placed inside the network are very harmful as compare to external intruder because they know very well where data is placed .So in LAN address resolution protocol becomes a more risky attack. This paper proposed a validate method for detecting and preventing ARP spoofing. For detecting the ARP spoofing we use primary and secondary cache after detecting send packet directly to the DHCP (dynamic host control protocol) server. Sending the data to DHCP server it reduce the network overload, congestion problem. For Echo request and Echo reply ping command is used for ICMP. Here we used 3 system main aim of sending these system for transferring the ICMP and ARP packet, with three system

backward compatibility.

Rest of the paper is organize as follow ARP spoofing and other context described in section 2. Approaches for ARP poisoning detection and prevention define in section 4, 5. Proposed mechanism described in section 6. Performance analysis and experimental is described in section no.7. Finally conclusion is described in section no, 8.

2 Background

2.1 ADDRESS RESOLUTION PROTOCOL

Suppose A want to established a communication with host B then for the purpose of communication knowing the B's MAC address is important. so first A's search the B's physical address in primary cache then after in secondary cache because in primary cache validity of data is only for 20 min. but in secondary cache data is store for a long duration. Request of ARP is shown in figure 1 [2]. Host B send its MAC address when it receive a request from host A. Reply of host B is shown in figure 2 [3]. Host A start binding of <IP-MAC > after receiving the response. ARP request message are generally related to broadcasting because it fetch the MAC address to destination node.

Host send a unicast REPLY with his MAC address. After 20 minute when the data is removed then host uses secondary cache. All request are received inside a subnet. Binding is always store in volatile form so it always updated at a regular interval of duration for deleting the rushed or invalid entries.

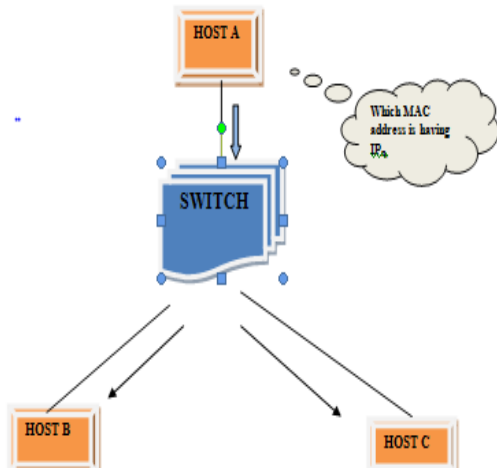


FIGURE 1 ARP request is broadcasts by host A to Host

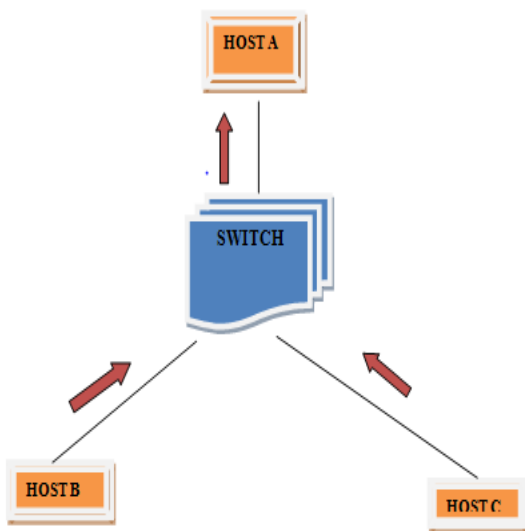


FIGURE 2 unicast reply is send by Host B

2.2 ARP CACHE POISONING

ARP is a protocol that have no state and no any authentication mechanism in figure 3. Host C behave a man-in-the-middle attacker, send a forged message to A by using B IP's address. Same as send a forged message to B using A's IP address belonging to same MAC address C.

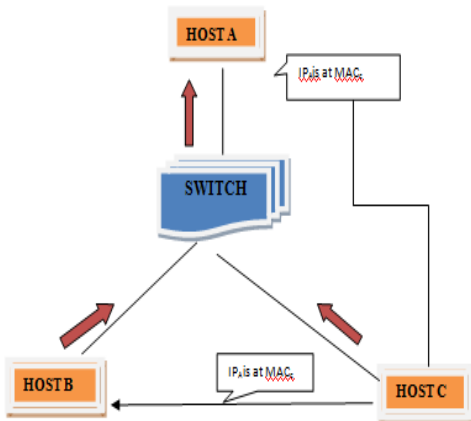


FIGURE 3 Host C perform ARP spoofing on A and B

Number of prevention, detection and mitigation techniques have been proposed till now for the solution of ARP Spoofing.

- Prevention Technique

The techniques falling under this category generally modifies the ARP and follow new set of rules. So, these techniques are resistant to ARP cache poisoning but are not backward compatible because these techniques interfere with the standard OSI model. Examples of such technique include MITM Resistant Address Resolution Protocol [4], Secure ARP [5] etc.

- Detection Technique

It not provide overall solution of spoofing only tends to reduce the chances detection, example of such technique include scheme of detection Trabelsi and Rahmani Technique [6].

- Mitigation Technique

The technique are backward companionable which are included spoofing criteria that are perform presence of attacker. When the spoofing had completed then attacker is detected .The main drawback of this technique is its processing time for categorization of attacker, because it is exceptionally high. It includes Ticket Based address resolution protocol (TARP [7])

3 Various Attack of ARP poisoning [8]

3.1 MAN- IN- THE- MIDDLE ATTACK

When intruder manipulates in between two devices then this attack are arise, It is type of dynamic eavesdropping attack, also called session hijacking attack. Attacker silently sited in between the source host and destination host, but both host are think that they are communicating with each other after extracting the sensitive data(e.g. id, password) from source send information to the destination host, but the host believes data which are received are original data. With MITM attack he can modify the data being send.

3.2 DENIAL OF SERVICE (DOS) ATTACKS

Every packet that is send by host is directly send to intruder, because an intruder spoof the all entry that are exist in ARP table, or intruder send forged packets with fake MAC address, By this way intruder blocks all the way by which host complete his communication.

3.3 THE BOMBING PACKETS ATTACK

It is mainly related with buffer overflow, data overflow in which many of the system spend a lot of time to maintain the ARP cache, Arises when a malicious host send a spoofed message map to a source host frequently.

3.4 MAC, IP CLONING ATTACKS

In Linux system without using of spoofing software, <physical address, logical address > can be changed easily, intruder automatic assign IP, MAC address of host computer. Since physical address is a unique address that is assigned by company when it is manufactured. Host will disconnect his interface once it identify the duplicate in IP, MAC.

4 Literature review

Prerna Arote et. al. Detection and Prevention Against ARP Poisoning Attack Using Modified ICMP and Voting [9] Propose a technique based on ICMP and voting that is backward compatible. In LAN environment physical address that transfer the data at data link layer .ICMP is used by the ping command including echo request and echo reply. It is also called as network protocol that not only store sensitive information also tells about status of system. It included two type of packet i.e. ARP and ICMP central server play an important role other system in network can work efficiently in case of failure of any system. There are two types of table i.e primary and secondary table. Central server maintain secondary table in which data is store for a long period of time. It has several advantage require less cost because of a few system in the network. Ettercap, SSL strip and client side implementation is the main module of this approach. But host for which static entry is not saved it does not provide the MITM solution.

Geojinhua et.al ARP spoofing Detection Algorithm Using ICMP Protocol [10] propose a scheme for detecting ARP poisoning using ICMP packet. On the basis of response packet it collect packet detect the malicious host. During the attack map the real data without disturbing activity of host. It dynamically map IP address into MAC address. For detecting the poisoning it uses the following module i.e sniffer module, detection module, response module. Using a cross layer In ARP and Ethernet header examine a secure consistency in source and destination host. There is a minimum time delay in Capturing and detecting spoofing attack, on the internet when any packet is detected trap ICMP ping is send frequently that reduce the network minimal overhead. Main drawback of this approach it not completely removed the problem of spoofing due to conflicting MAC address.

Nikhil Tripathi, BM Mehtre [3] Analysis of Various ARP Poisoning Mitigation technique: A Comparison proposed a schema in which important fact of several technique that are considered as limitation to the proposed scheme that is based on the cryptography. In LAN environment Attack is sponsored then these fact are derived from that scenarios where the attack is possible. In case of making more efficient scheme these fact are considered as valuable phenomena. In the area of computing every interface is assigned to MAC and IP address. Due to the problem of loop holing and its nature (un-authentication, stateless) intruder launch a very dangerous attack that exploit the vulnerability of ARP. Factor that are included they are:

- Flooding of ARP data.
- Compatibility with alias name
- Single point of failure.
- Main problem of this scheme with it only consider the facts and mostly that fact which are derive in LAN environment. Main limitation of this approach is extra administrative cost.

Nikhil Tripathi and B.M Mehtre [11] AN ICMP based secondary cache approach for the detection and prevention of ARP poisoning- Proposed a feasible technique that reduce the multiple entry of IP and MAC addresses by using secondary cache. In which data is store for a long period of time by using ICMP protocol. Secondary cache ensure that

there is only one entry of IP address corresponding with MAC address, that make the solution is backward compatible. First use of primary cache which are update time to time for deleting entry that are no longer used. Text file is main element of secondary cache that are maintaining at every host and make this technique distributed in nature, backward compatible. Several scenes are present in this scheme either intruder attack at starting stage or it is quite possible that the intruder are already present in network. Though a lot no. of message exchange in this algorithm that make an expansive solution for any confidential function.

Somnuk and Massusai [12] Static <IP, MAC> binding scheme proposed aimed to update all the static entry that are available host cache table. Main drawback of this scheme it increase operating system overhead due to the large no of host.

Gauda et. al [13] proposed a mechanism based on the central server. Request-reply and invite-accept are two protocols that are used by this scheme. On the several registration of IP-MAC should be done in case of new host enter in network by using second protocol that are mention above. Both detection and prevention are perform in this technique. Limitation of this approach is it suffers from single site breakdown, it could lead to be poisoning attack successfully, if intruder itself hack the server .this require modification in existing ARP and do not use cryptography.

Dynamic detection scheme [14] that is completely based on the snort tool. Snort is a type of detection System used to detect attack that is performs by intruder. It has an ability to analyse real time packets on a particular logical address. But due to containing false warning it generate virtual reports to administration. Further lots of technique proposed for detecting poisoning at network layer by which most of functioning of firewall are grouped together with routers, by which problems of false warning approximately reduced. Main limitation of using this scheme, unable to differentiate between intruder and real victim. If we focus towards the complexity of such mechanism resulting a setup found with the very high cost at installation. This is a main reason of not capable using such concept.

5 Requirement for an ideal solution:

- Solution should be cost effective.
- It should be effective for preventing the attack.
- It minimizes the network traffic.
- It reduces the network overload.
- Any changes should not occur in existing protocol model.

6 Proposed mechanism:

We proposed a scenario for reducing network overload. This mechanism are backward compatible and less complex because we do not use cryptography. By using the concepts of DHCP (Dynamic host control protocol) try to reduce overload. This scheme use a centralized approach .In our assumption min 3 no. of host that are available in the network that are maintain primary and secondary table that are permanently Store the data until we not deleted. Data is stored in the form of text in secondary table. Once validity of data is complete primary cache is updated according to validation. Our main aim is to reduce the network overload

and congestion problem after complete the validation phase if any problem occur to identify the data the send a message to DHCP server, that automatically assign IP address to system. In design of current system we have 3 systems that

are connected over LAN. Host will maintain two table primary and secondary table. DHCP server uses only secondary table. Algorithm for detection and prevention of ARP is as follows, that are described as follows:

Step 1:	If a host want to communicate with other host then broadcast request to other host , with its IP address
Step 2:	Other host receive a request send a reply message. After that source host check its entry in primary cache.
Step 3:	If entry found in primary cache then update<IP, MAC> the binding. Else check secondary table.
Step 4:	There are two case arise Case 1:Entry found in secondary cache Case 2: Entry not found in secondary cache.
Step 5:	Case 1: If the binding is not found in primary cache mean the entry would have expired thus secondary table is checked for the entry. If the binding is found to be same as stored in secondary cache, both local ARP primary cache and secondary table is updated.
Step 6:	Case 2: The host send request to DHCP server for IP at a time interval to obtain reply of any one, preventing any flooding attack. If the reply received from more than one host is the chance that the Reply is sent by a malicious host to poison the ARP cache. In this case a alarm is generated.
Step 7:	DHCP server send a response. If(response>1) Then Again send unicast packet to all host Else Update the cache. Else if (reply > 1) Generate a alarm nominate as legitimate.

Flowchart:

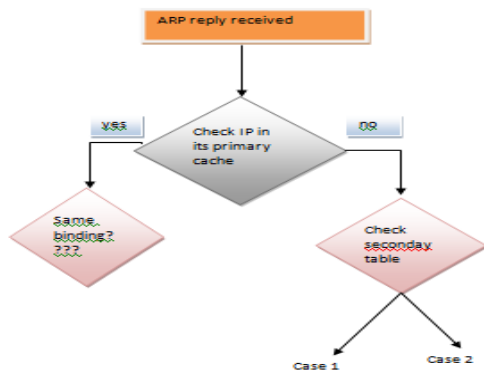


FIGURE 5 check the mapping in primary and secondary table

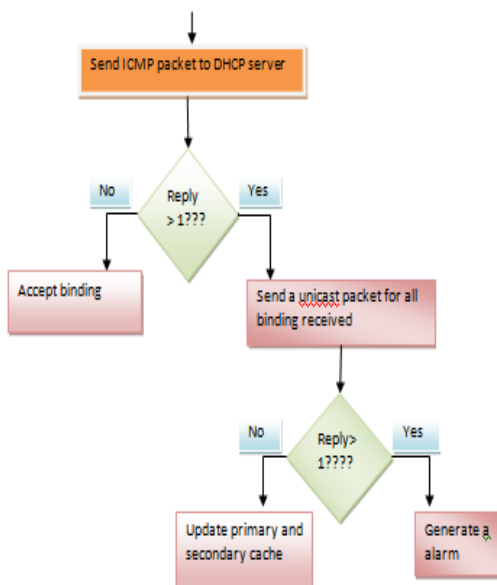


FIGURE 6 find the binding from DHCP server

Case 1: Entry found with same <IP, MAC> association
If the binding is not found in primary cache mean the entry would have expired thus secondary table is checked for the entry. If the binding is found to be same as stored in secondary cache, both local ARP primary cache and secondary table is updated.

Case 2: Entry not found in Secondary table
In case the entry is unavailable in Secondary table too, the host sends ARP request packets to DHCP server IPX at a time interval of to obtain reply of any one, preventing any flooding attack [3]. If the reply received from more than one host is the chance that the Reply is sent by a malicious host to poison the ARP cache. If reply is received from more than one host then send ICMP probe packets to each host from whom reply is received. If the reply is received, accept the binding else discard the entry from local cache. In case ARP reply is received from only one host; the binding is accepted and updated in both primary and secondary table. Figure 7 represents the case if entry not found in secondary table.

7 Implementation and result

We have implemented the scheme using three hosts with IP 172.18.5.190, 172.18.5.191 and 172.18.5.192 respectively. The attacker has IP, 172.18.5.190 communicates normally then tries the attack by pretending to be 172.18.5.191. The IP and MAC of hosts are shown in Figure 7, 9 which is stored as secondary table holding IP-MAC binding. When the script which is saved with extension “.py” on terminal is run, broadcasts the ARP request. The attacker using packEth generates a reply and send it to host. Then the secondary table is searched for that binding. If found same, the secondary table is further updated with display of message no issue. But in case of mismatch ICMP ping packet is sent to DHCP server. If reply is received from previous packet an alarm is raised and the entry is removed from ARP cache. Here festival tool is used which is used to convert text to speech. For new host whose binding is not found in

secondary cache first broadcasts ARP request defined with count and timeout. If reply is received then the entry is stored in both secondary table stored in form of table.txt and primary cache.

```

Open [ ]
172.10.5.190      90:e7:f4:31:9e:07
172.18.5.191     84:8f:69:d3:52:e8
172.10.5.192     50:fb:04:26:6b:c3

```

FIGURE 7 Secondary table stored

```

[33]: Stopped arping 172.18.5.190
root@uapl:~/Inspiron-5420/home/sect04/plug_172.18.5.190
PING 172.18.5.190 (172.18.5.190) 56(84) bytes of data.
64 bytes from 172.18.5.190: icmp_seq=1 ttl=64 time=0.427 ms
64 bytes from 172.18.5.190: icmp_seq=2 ttl=64 time=0.452 ms
64 bytes from 172.18.5.190: icmp_seq=3 ttl=64 time=0.458 ms
64 bytes from 172.18.5.190: icmp_seq=4 ttl=64 time=0.465 ms
64 bytes from 172.18.5.190: icmp_seq=5 ttl=64 time=0.423 ms
64 bytes from 172.18.5.190: icmp_seq=6 ttl=64 time=0.379 ms
64 bytes from 172.18.5.190: icmp_seq=7 ttl=64 time=0.435 ms
64 bytes from 172.18.5.190: icmp_seq=8 ttl=64 time=0.438 ms
64 bytes from 172.18.5.190: icmp_seq=9 ttl=64 time=0.437 ms
64 bytes from 172.18.5.190: icmp_seq=10 ttl=64 time=0.432 ms
64 bytes from 172.18.5.190: icmp_seq=11 ttl=64 time=0.483 ms
64 bytes from 172.18.5.190: icmp_seq=12 ttl=64 time=0.439 ms
64 bytes from 172.18.5.190: icmp_seq=13 ttl=64 time=0.456 ms
64 bytes from 172.18.5.190: icmp_seq=14 ttl=64 time=0.448 ms
64 bytes from 172.18.5.190: icmp_seq=15 ttl=64 time=0.443 ms
64 bytes from 172.18.5.190: icmp_seq=16 ttl=64 time=0.437 ms
64 bytes from 172.18.5.190: icmp_seq=17 ttl=64 time=0.448 ms
64 bytes from 172.18.5.190: icmp_seq=18 ttl=64 time=0.452 ms
64 bytes from 172.18.5.190: icmp_seq=19 ttl=64 time=0.424 ms

```

FIGURE 8 Unicast reply from 172.18.5.190

```

File Edit View Go Capture Analyze Statistics Telephony Tools Internals Help
Filter:
Frame 14: 42 bytes on wire (336 bits), 42 bytes captured (336 bits) on interface 0
Ethernet II, Src: 98:e7:f4:31:9e:07, Dst: TendaTec_5d:f8:40 (c8:3a:35:5d:f8:40)
Duplicate IP address detected for 172.18.5.118 (98:e7:f4:31:9e:07) - also in use by 50:fb:04:26:6b:c3 (frame 13)
Duplicate IP address detected for 172.18.5.185 (c8:3a:35:5d:f8:40) - also in use by 98:e7:f4:31:9e:07 (frame 13)
Address Resolution Protocol (rpl)

```

FIGURE 9 Duplicate IP alert generated by wire shark

```

File Edit View Go Capture Analyze Statistics Telephony Tools Internals Help
Filter:
Frame 14: 42 bytes on wire (336 bits), 42 bytes captured (336 bits) on interface 0
Ethernet II, Src: 98:e7:f4:31:9e:07, Dst: TendaTec_5d:f8:40 (c8:3a:35:5d:f8:40)
Duplicate IP address detected for 172.18.5.118 (98:e7:f4:31:9e:07) - also in use by 50:fb:04:26:6b:c3 (frame 13)
Duplicate IP address detected for 172.18.5.185 (c8:3a:35:5d:f8:40) - also in use by 98:e7:f4:31:9e:07 (frame 13)
Address Resolution Protocol (rpl)

```

FIGURE 10 Expert info in Wire shark by DHCP server

After checking the expert info by wire shark send different type of packet like TCP,ICMP,ARP packet and check it is prevented from poisoning or not.

```

user@UPLC:~$ sudo wireshark
Gtk-Message: GtkDialog mapped without a transient parent. This is discouraged.
root@UPLC:~# sudo wireshark
Gtk-Message: GtkDialog mapped without a transient parent. This is discouraged.
root@UPLC:~# sudo scapy
INFO: Can't import python gnuplot wrapper . Won't be able to plot.
INFO: Can't import PyX. Won't be able to use psdump() or pdfdump().
WARNING: Failed to execute tcpdump. Check it is installed and in the PATH
WARNING: No route found for IPv6 destination :: (no default route?)
Welcome to Scapy (2.2.0)
>>> pkt=IP(dst="172.18.5.191")/ICMP("hello")
Begin emission:
.....WARNING: Mac address to reach destination not found. Using broadcast.
Finished to send 1 packets.
...C
Received 13 packets, got 0 answers, remaining 1 packets
>>> pkt
<-Results: TCP:0 UDP:0 ICMP:0 Other:0, <Unanswered: TCP:0 UDP:0 ICMP:1 Other:0>
>>> pkt.show()
Traceback (most recent call last):
File "<console>", line 1, in <module>
AttributeError: 'tuple' object has no attribute 'show'
>>>

```

FIGURE 11 Send ICMP packet

```

File Edit View Go Capture Analyze Statistics Telephony Tools Internals Help
Filter: icmp
No. Time Source Destination Protocol Length Info
...
Frame 21: 88 bytes on wire (704 bits), 88 bytes captured (704 bits) on interface 0
Ethernet II, Src: 98:e7:f4:31:9e:07, Dst: Broadcast (ff:ff:ff:ff:ff:ff)
Internet Protocol Version 4, Src: 172.18.5.190, Dst: 172.18.5.191
Internet Control Message Protocol
...

```

FIGURE 12 Packet analysed by wire shark according to server response

In above scenario we observe that when we send ICMP (Internet Control Message protocol) just type a text hello. After sending the message analyse by wire shark, obtain only one reply from host there is no duplicate entry is found. It insures that packet are free from poisoning attack. Similarly send TCP packet for checking ARP poisoning is detected or not. In following figure send TCP (transmission control protocol) and observe response.

```

user@UPLC:~$ sudo scapy
SyntaxError: EOL while scanning string literal
>>> [13]+ Stopped sudo scapy
root@UPLC:~# sudo scapy
INFO: Can't import python gnuplot wrapper . Won't be able to plot.
INFO: Can't import PyX. Won't be able to use psdump() or pdfdump().
WARNING: Failed to execute tcpdump. Check it is installed and in the PATH
WARNING: No route found for IPv6 destination :: (no default route?)
Welcome to Scapy (2.2.0)
>>> pkt=IP(dst="192.168.10.3")/TCP(dport=23)
Begin emission:
Finished to send 1 packets.
.....C
Received 10 packets, got 0 answers, remaining 1 packets
>>> pkt
<-Results: TCP:0 UDP:0 ICMP:0 Other:0, <Unanswered: TCP:1 UDP:0 ICMP:0 Other:0>
>>> pkt.show()
Traceback (most recent call last):
File "<console>", line 1, in <module>
AttributeError: 'tuple' object has no attribute 'show'
>>>

```

FIGURE 13 TCP packet is send wait for response

```

File Edit View Go Capture Analyze Statistics Telephony Tools Internals Help
Filter: tcp
No. Time Source Destination Protocol Length Info
...
Frame 4: 88 bytes on wire (704 bits), 88 bytes captured (704 bits) on interface 0
Ethernet II, Src: 98:e7:f4:31:9e:07, Dst: Broadcast (ff:ff:ff:ff:ff:ff)
Internet Protocol Version 4, Src: 172.18.5.190, Dst: 192.168.10.3 (192.168.10.3)
Transmission Control Protocol, Src Port: 54321, Dst Port: 23, Seq: 10000, Len: 0
...

```

FIGURE 14 Analyse response by wire shark according to response of DHCP server

```

user@UPLC:~$ sudo wireshark
Gtk-Message: GtkDialog mapped without a transient parent. This is discouraged.
root@UPLC:~# sudo wireshark
Gtk-Message: GtkDialog mapped without a transient parent. This is discouraged.
root@UPLC:~# sudo scapy
INFO: Can't import python gnuplot wrapper . Won't be able to plot.
INFO: Can't import PyX. Won't be able to use psdump() or pdfdump().
WARNING: Failed to execute tcpdump. Check it is installed and in the PATH
WARNING: No route found for IPv6 destination :: (no default route?)
Welcome to Scapy (2.2.0)
>>> a=sendp(Ether(dst="84:8f:69:d3:52:e8")/ARP(op="is-at",hwdst="84:8f:69:d3:52:e8",hwsrc="98:e7:f4:31:9e:07",pdst="172.18.5.191",psrc="172.18.5.190"))
Send 1 packets.
>>> a.summary()
Traceback (most recent call last):
File "<console>", line 1, in <module>
AttributeError: 'NoneType' object has no attribute 'summary'
>>> a=sendp(Ether(dst="84:8f:69:d3:52:e8")/ARP(op="is-at",hwdst="84:8f:69:d3:52:e8",hwsrc="98:e7:f4:31:9e:07",pdst="172.18.5.191",psrc="172.18.5.190"))
Send 1 packets.
>>> a.summary()
Traceback (most recent call last):
File "<console>", line 1, in <module>
AttributeError: 'NoneType' object has no attribute 'summary'
>>>

```

FIGURE 15 code send to server DHCP server

```

user@UPLC: ~
File Edit View Search Terminal Help
root@UPLC[user]#sudo scapy
INFO: Can't import python gnuplot wrapper . Won't be able to plot.
INFO: Can't import PyX. Won't be able to use psdump() or pdfdump().
WARNING: Failed to execute tcpdump. Check it is installed and in the PATH
WARNING: No route found for IPv6 destination :: (no default route?)
Welcome to Scapy (2.2.0)
>>> srp(Ether(dst="84:8f:69:d3:52:e8")/ARP(op="who-has",pdst="172.18.5.191",psrc="172.18.5.190",hwdst="84:8f:69:d3:52:e8",hwsrc="98:e7:f4:31:9e:87"))
Begin emission:
Finished to send 1 packets.
.....C
Received 15 packets, got 0 answers, remaining 1 packets
<-Results: TCP:0 UDP:0 ICMP:0 Other:0>, <-Unanswered: TCP:0 UDP:0 ICMP:0 Other:1>
>>> srploop(Ether(dst="84:8f:69:d3:52:e8")/ARP(op="who-has",pdst="172.18.5.191",psrc="172.18.5.190",hwdst="84:8f:69:d3:52:e8",hwsrc="98:e7:f4:31:9e:87"))
fail 1: Ether / ARP who has 172.18.5.191 says 172.18.5.190
fail 1: Ether / ARP who has 172.18.5.191 says 172.18.5.190
fail 1: Ether / ARP who has 172.18.5.191 says 172.18.5.190
fail 1: Ether / ARP who has 172.18.5.191 says 172.18.5.190
send...
Sent 4 packets, received 0 packets, 0.0% hits.
<-Results: TCP:0 UDP:0 ICMP:0 Other:0>, <-PacketList: TCP:0 UDP:0 ICMP:0 Other:4>
    
```

FIGURE16 Receive response about packet

TABLE 1 Comparative analysis of previous and proposed approach

Previous mechanism Author	Mechanism for validation of ARP Reply	Centralized Scheme	Flooding attack possibility	IP Exhaustion Problem
G. Jinhua and X. Kejian [15]	Probing mechanism by Central serve	Yes	Yes	Yes
N. Tripathi and B. M. Mehtre [1]	2 algorithm used using ICMP packet	No	Yes	No
P. Pandey [4]	2 ICMP probe packets	No	Yes	Yes
P. Arote [8]	Central server validation	Yes	No	Yes
Vidya Srivastava and Dayashankar Singh (Proposed Work)	DHCP server	Yes	No	No

9 Conclusion

Mechanism that are proposed in dissertation are attempting to detect and prevent ARP poisoning. Attacker can send fake binding that can be deal with other type of attack such that man-in-the middle attack, Denial of services attack. This mechanism provides a solution for detection and prevention of ARP poisoning. Secondary table that is long term storage of data use to validate the entry of data, and by using DHCP server for a new binding checking binding is valid or no. The mechanism can lead to asynchronous behaviour, without consisting any periodic monitoring.

Before proposing a mechanism a criteria that should be necessary for requirement of an ideal solution is always kept in mind, whatever any mechanism proposed but it no change the existing model, and reduced network traffic. ARP resides at data link layer. Attacks are possible over local area network. We present a small scenario where

Can't demonstrate DHCP server. We assume host as a server. For the full completion of scenario need a large host. Some modification are made at few sites and demonstrate DHCP server, work will expanded in future.

Main aim of using DHCP server there is find an intruder because server provide a valid authentication for any other

References

[1] Plummer D 1982 An Ethernet address resolution protoco *RFC* 826
 [2] Kumar S, Tapaswi S A Centralize Detection and Prevention Technique against ARP Poisoning 259–64
 [3] Tripathi N, Mehtre B M 2014 Analysis of various ARP poisoning mitigation techniques: A comparison *International Conference on*

8 Performance evolution

Let assume there are N no. of node present in network. And a unique <IP, MAC >is assign for each ARP packet. Host verify all entry with respect to long term cache (Secondary table) just for checking match status. Complexity will be O (log n) for such type of step. Send the request to DHCP server for finding actual pair. There are one more possibility arises if in worst case. If entry not found then broadcast its request and obtain reply. Complexity will be 1 for such type of step. That is O(1). If we are going on worst case complexity then it will be O(logn). On the basis of proposed approach a comparative analysis is performed with previous technique.

host. That also reduce the network traffic and overhead.

In future main aim is expanding By taking all the scope of network and all the scenario that are existing in local area network with all possibilities to pilfering.



10 Acknowledgements

I take the opportunity to express my heartfelt adulation and gratitude to my supervisor Mr. Dayashankar Singh (Assistant Professor, Department of Computer Science and Engineering, Madan Mohan Malaviya University of Technology, Gorakhpur) for his unrevised guidance, constructive suggestions, thought providing discussions and unabashed inspiration in the nurturing work. It has been benediction for me to spend many opportune moments under the guidance of perfectionist at the acme of profession. He was always there to listen and give advice. He showed me the different ways to approach a research problem and a need to be persistent to accomplish any goal. He taught me how to write academic paper, had confidence in me whenever I doubted myself, and brought out new ideas in me. The present work is testimony to his activity, inspiration and ardent personal interest taken by him during his work in its present form.

Control, Instrumentation, Communication and Computational Technologies (ICCICCT), Kanyakumari 125-32
 [4] Nam S Y, Kim D, Kim J 2010 *Enhanced ARP: Preventing ARP Poisoning-Based Man-in-the-Middle Attacks* 14(2) 187–9
 [5] Bruschi D, Ornaghi A, Rosti E 2003 S-ARP: a secure address

- resolution protocol in *Proceedings of 19th Annual Computer Security Applications Conference, IEEE* 66-74
- [6] Pandey P 2013 Prevention of ARP spoofing: A probe packet based technique *3rd IEEE International Advance Computing Conference (IACC), Ghaziabad* 147-53
- [7] Lootah W, Enck W, McDaniel P 2007 Tarp: Ticket-based address resolution protocol *Elsevier* **51**(15) 4322-37
- [8] Salim H, Li Z, Tu H, Guo Z 2012 *Preventing ARP Spoofing Attacks through Gratuitous Decision Packet* 295-300
- [9] Arote P, Arya K V 2015 Detection and Prevention against ARP Poisoning Attack Using Modified ICMP and Voting *International Conference on Computational Intelligence and Networks, Bhubaneswar* 136-14
- [10] Jinhua G, Kejian X 2013 ARP spoofing detection algorithm using ICMP protocol *Int. Conf. Comput. Commun. Informatics, ICCCI* 0-5
- [11] Tripathi N, Mehtre B M 2013 An ICMP based secondary cache approach for the detection and prevention of ARP poisoning *IEEE International Conference on Computational Intelligence and Computing Research, Enathi* 1-6
- [12] Puangpronpitag S, Masusai N 2009 An Efficient and Feasible Solution to ARP Spoof Problem *6th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, ECTICON2009* ISBN: 978-1-4244-3387
- [13] Gouda M, Huang C T 2003 A secure address resolution protocol *The International Journal of Computer and Telecommunications Networking, Elsevier North-Holland, Inc. New York, NY, USA* **41**(1) 57-71
- [14] Hou X, Jiang Z, Tian X 2010 The detection and prevention for ARP Spoofing based on Snort *In Proceedings of Computer Application and System Modeling, IEEE Int. Conf. V5-137-V5-139*

AUTHORS

	<p>Vidya Srivastava</p> <p>Current position, grades: graduate student. University studies: Madan .Mohan Malviya University of Technology Gorakhpur Scientific interests: Computer Network, Operating System</p>
	<p>Dayashankar Singh</p> <p>Current position, grades: Assistant professor department of computer science and engineering Universisty studies: Punjab University Chandigarh, India Scientific interests: Computer Network, Information Security, Database management system</p>

AUTHORS' INDEX	
Akhmetov B	19
Aljawawdeh A	37
Alsoudi D	37
Ananda Rao A	29
Aytkhozhaeva E	54
Aytkhozhaeva Zh E	48
Beketova G	19
Boiko A	11
Bondarenko O	11
Dileep Kumar Reddy P	41
Dragomeretskaya E	7
Imanbayev A	54
Jaiswal U C	58
Korchenko A	19
Lakhno V	19
Mall S	58
Manaseer S	37
Minghui Y	89
Mishra A	80
Padmavathamma M	71
Praveen Sam R	41
Rajini Kanth T V	29
Ramesh G	29
Sekar K	71
Sheng Z	89
Shoba Bindu C	41
Singh D	80, 93
Srivastava V	93
Tereshuk A	19
Tereykovskaya L	54
Tereykovskiy I	54
Tewani K	69
Tynymbayev S	54
Zhaibergenova A Zh	48
Ziro A A	48
Zvaigzne A	11

NATURE PHENOMENA AND INNOVATIVE ENGINEERING

Effect of texture on mechanical and magnetic properties of steel from the petroleum distillation column

Elena Dragomeretskaya

Computer Modelling & New Technologies 2017 21(2) 7-10

Texture, mechanical properties and coercive force of steel 09G2S from the column fragment of petroleum distillation after prolonged use studied. Anisotropy of mechanical properties and coercive force take place. Significant pair wise linear correlations and appropriate regression equations with coefficients reliability of approximation not less than 0.90 were found between magnitudes of the coercive force, tensile strength, yield strength, elongation and texture characteristics. Found correlations may be used for nondestructive mechanical properties control of investigated steel by means of monitoring of coercive force.

Keywords: texture, anisotropy, mechanical properties, coercive force, correlation

Decision support system on the base of genetic algorithm for optimal design of a specialized maritime platform

Andrejs Zvaigzne, Oleksandr Bondarenko, Anzhela Boiko

Computer Modelling & New Technologies 2017 21(2) 11-18

The analysis of possibilities of application of the small waterplane area twin hull ships (SWATH) as a specialty (universal) platform is performed. It is shown that the design of the specialized platform with a small waterplane area twin hull is characterized by a large number of parameters to be determined. The optimum relation selection between SWATH dimensions, seaworthiness, cost and efficiency is proposed by solving a multidimensional optimization problem with the use of special methods of searching solutions. The optimization problem of designing a universal platform is formulated. The constraints accounting on SWATH technical characteristics is produced by using the method of penalty functions. To solve the optimization problem, one of modern search methods – genetic algorithm is used. An example of solving the problem of selection the main dimensions of 25 m platform using a genetic algorithm is presented.

Keywords: SWATH, specialized platform, genetic algorithm, optimization, Mission Module

MATHEMATICAL AND COMPUTER MODELLING

Cyber intelligence systems based on adaptive regression splines and logical procedures of attack recognition

G Beketova, B Akhmetov, A Korchenko, V Lakhno, A Tereshuk

Computer Modelling & New Technologies 2017 21(2) 19-28

The article presents the results of research devoted to the further development of methods, models and algorithms for recognizing cyber threats, as well as the most common classes of cyber attacks and anomalies in critical computer systems (CCS). It is shown that the cyber security process for CCS controlled and analyzed by the values of several parameters of anomalies or signs of cyber attacks. This, in turn, makes it possible to carry out a preliminary assessment of information security with the help of two-stage recognition procedure in which initially used the methodology of adaptive regression splines for the processing of statistical data on the anomalies and cyber incidents in CCS, and then in the second stage are used designed logical recognition procedures based on the signs of matrix surfaces. This minimizes the number of training samples for the detection of objects in the framework, the relevant classes of cyber threats, attacks and anomalies.

The research on minimizing the amount of training samples of recognizing signs were performed. It is shown that for the recognition of objects within the known class of cyber threats, attacks and anomalies in the use of training facilities matrices used for training a representative set of long 3-5 attributes will allow to achieve maximum efficiency of the algorithm, reaching up to 98%.

Using the proposed method and models has allowed to reduce the amount of required object recognition rules within the class of 2.5-10 times, compared to the widely used in anomaly detection systems and methods of cyber attacks sequential sorting features and statistical algorithms states.

Keywords: intelligent recognition system, cyber threats, anomalies, signs of cyber attacks, adaptive regression splines, logical procedures, elementary classifier

Metrics for consistency checking in object oriented model transformations

G Ramesh, T V Rajini Kanth, A Ananda Rao

Computer Modelling & New Technologies 2017 21(2) 29-36

Model transformation is the cornerstone of Model-Driven Engineering (MDE) as it is crucial in Computer Aided Software Engineering (CASE) towards Object Oriented Analysis and Design (OOAD) and Object Oriented Programming (OOP). It also plays vital role in entity relationship model. Therefore it is indispensable to be treated as traditional software artefacts and assess quality of model transformations. Model-to-model transformations are from Platform Independent Model (PIM I) to Platform Independent Model (PIM II) and from PIM to Platform Specific Model (PSM). The goal of our research in this paper is to make these model transformations measurable. However, it is confined to proposing a set of metrics pertaining to consistency checking. The quality of transformations is measured in terms of consistency. The metrics proposed in this paper are general and can be reused. We evaluate the metrics using our framework named Extensible Real Time Software Design Inconsistency Checker (XRTSDIC) which supports end-to-end transformations of object oriented models. Our empirical study revealed that the proposed metrics add value to our model consistency checker as they quality in model transformations.

Keywords: Model Driven Engineering (MDE), XRTSDIC, model transformations, consistency checking, quality measures

Border node detection: a new experimental approach

Saher Manaseer, Dua Alsoudi, Asmaa Aljawawdeh

Computer Modelling & New Technologies 2017 21(2) 37-40

This paper aims at sensing the network, and detects the border nodes, the researcher use NS2, in order to represent, simulate and calculate the delivery ratios of the distributed packets which accordingly will help to detect the border nodes. The importance of this research comes from detecting the border nodes without depending on other resources, since Ad hoc networks coordinates are virtual. The researchers analysed the results of the trace file that came as an output of carrying out simulations in Network simulator (NS2) for the evaluation of the ratios. The methodology of this experiment depends on using the IEEE 802.11 MAC protocol. Flooding technique was used to send data packets through three scenarios: First, 5% of the nodes are randomly chosen to send their data packets per minute. In the second and third scenarios, the percentages of nodes that flood their data are 25% and 50% respectively.

Keywords: MANETs, Broadcast, NS2, IEEE 802.11, MAC, Flooding

Tri-Partite graph: a novel security scheme for cloud data

P Dileep Kumar Reddy, C Shoba Bindu, R Praveen Sam

Computer Modelling & New Technologies 2017 21(2) 41-47

Cloud data security is the most concentrated feature of the cloud computing technology. Many cloud computing techniques like cloud data partitioning emerged reflecting new heights of providing data security by defining data priorities. The proposed method presents a novel scheme of maintaining owners prioritized data, while equally ensuring the security for whole portion of the data. The proposed method uses a tripartite graph for securely managing the prioritized data at various levels.

Keywords: Encryption, Authentication, tripartite graph, Hash, MAC

Virtualization safety

Zh E Aytkhozhaeva, A A Ziro, A Zh Zhaibergenova

Computer Modelling & New Technologies 2017 21(2) 48-53

Article considered virtualization technologies, their types, advantages and disadvantages. Attention to specific risks and information security threats in case of virtualization platforms is paid. The main risks of virtualization platforms are defined. Potential internal vulnerabilities of virtualization platforms can be revealed only by testing for penetration which user-friendly and available instrument for implementation is specialized by Kali Linux OS. The attacks to the virtual machines with use of the Kali Linux tools were organized. As a result of experiments is Kali Linux allows revealing and analysing vulnerabilities at the channel, network and transport levels. For detection of problems at the level of applications that is urgent for virtualization of platforms, it is necessary to use commercial products of ethic hacking in addition.

Keywords: virtualization platforms, risks, penetration testing

Improvement of learning efficiency of the neural networks, intended for recognition of graphic images in systems of biometric authentication

L Tereykovskaya, I Tereykovskiy, E Aytkhozhaeva, S Tynymbayev, A Imanbayev

Computer Modelling & New Technologies 2017 21(2) 54-57

Article is devoted to a problem of use of neural network technologies in the field of biometric authentication of users. It is shown that one of important the shortcomings of application of neural networks technology on the basis of a multi-layer perceptron for recognition graphic images in systems of biometric authentication of users is insufficient quality of processing of statistical data which are used when forming parameters of educational examples. It is offered to increase quality of educational examples due to use of the procedure of neural network coding of value of the expected output signal of educational examples which allows consider closeness of standards of the recognized classes in this signal. The coding procedure of the expected output signal providing use of a probable neural network is developed. The appropriate mathematical devices are created. As a result of numerical experiments it is shown that application of the developed procedure allows reduce the number of the computing iterations necessary for achievement of the given error of training by 30-50%. It specifies prospects of use of the proposed solutions for improvement of learning efficiency of the neural networks, intended for recognition of graphic images in systems of biometric authentication.

Keywords: neural network, information security, learning, biometric authentication

Word sense disambiguation in Hindi applied to Hindi-English machine translation

S Mall, U C Jaiswal

Computer Modelling & New Technologies 2017 21(2) 58-68

The Word Sense Disambiguation for Hindi Language is one of the biggest challenges faced by Natural Language Processing. In this paper we discuss issues in reducing ambiguity in Word Sense Disambiguation for Hindi Language. The concepts are induced in two modules Parsing and Word Sense Disambiguation for Hindi Language. Parsing is an extension of our previous work on shallow parser method that creates groups word which are essential for Machine Translation. Monolingual Hindi and English corpora are used. Following this we used machine learning technique such as supervised approach, unsupervised approach and domain specific sense with the help of Knowledge based methods. Knowledge based method uses Hindi and English WordNet tools. Supervised method is used to disambiguate the multiple tags in the context label with the correct tag. Unsupervised method is used to update the sentence with the correct sense and parts of speech tag. There are various websites which provide the facility of translation of Hindi language to English language such as Google Translator and Babefish Translator but these translators fail to resolve polysemy words in Hindi sentences the result is discussed in this paper. The accuracy result of part of speech tagging generated by our system is

92.09%. The accuracy results generated by our system for Chunk are window-3, window 2 and window1 are: 94.45%, 81.23%, and 81.11% respectively. We modify and develop Lesk algorithm which uses WordNet tools for Word Sense Disambiguation. We compare the system's performance with the website Google Translator. We also examine errors made by Google Translator for given input Hindi sentence. Our system generates correct translation with Word Sense Disambiguation for given input Hindi sentence as shown in the Figure 12.

Keywords: Domain specific sense, Word Sense Disambiguation, Morphological analysis, Part of speech tagging and Parsing

Ant colony optimization algorithm: advantages, applications and challenges

Kavita Tewani

Computer Modelling & New Technologies 2017 21(2) 69-70

Ant Colony optimization is a technique for optimization that was introduced in early 1990's. ACO algorithm models the behaviour of real ant colonies in establishing the shortest path between food sources and nests and this technique is applied on number of combinatorial optimization problem, communication networks and robotics. This paper introduces the advantages of using the ACO algorithms with the help of some problem examples and the challenges faced for solving the problems. Initially, the paper discusses about the biological inspiration and behaviour of ant colony and then relates with the real life problems.

Keywords: Ant colony optimization (ACO), pheromone, Travelling Salesman Problem (TSP)

Business process re-engineering capability based on ECMM: Efficient Configuration Model and Management

K Sekar, M Padmavathamma

Computer Modelling & New Technologies 2017 21(2) 71-79

Most business process Companies are interested for new solutions and techniques in organisations. Relating to the big data to achieve that business process must be reengineered. Reengineering of business process can be done based on six sigma activities like Define, Measure, Analyze, Improve, Control and Report. In Business Process Reengineering, the two constants of any organisation are people and process. If individuals are motivated and working hard, here the business process are compressive and organisational process will be poor and posses high failure rate. In order to overcome these effects Business Process Reengineering must have some assessing capabilities which is referred as Desired Organizational Capabilities (DOC) and total quality management (TQM) to increasing the efficiency of reengineering and makes the manufacturing of logistical systems more scientific.

Keywords: Six Sigma activities, DOC, Business Process Reengineering, TQM

Handwritten digit recognition using combined feature extraction technique and neural network

Ankita Mishra, Dayashankar Singh

Computer Modelling & New Technologies 2017 21(2) 80-88

Handwritten digit recognition is established and emerging problem in pattern recognition and computer vision. A very few volume of work related to research has been done in this field till now. Handwritten digit recognition is very useful in cheque processing in bank, form processing systems and many more. In this paper, a robust and novel technique has been introduced for handwritten digit recognition which is tested on well-established MNIST dataset. Histogram of oriented gradient technique and wavelet transform technique is used for feature extraction. Radial basis function neural network and back-propagation neural network have been used as classifier. Experimental analysis has been carried out and result shows that RBF yields good recognition accuracy as compared to back-propagation neural network.

Keywords: Handwritten digit recognition (HDR), Back propagation Neural Network, Radial Basis Function, Histogram of Oriented Gradient (HOG)

Fuzzy comprehensive evaluation model for mobility model

Yao Minghui, Zhang Sheng

Computer Modelling & New Technologies 2017 21(2) 89-92

Evaluation of mobility model is an important means to ensure the quality and design level. At present, many mobility models are proposed for opportunistic networks. But, there is no practical quantitative evaluation system to evaluate the mobility models. Firstly, this paper put forward a comprehensive evaluation index system of mobility model based on the analysis of the main factors affecting the quality of mobility model and the relationship between them. Secondly, based on the theory of fuzzy comprehensive evaluation, this paper put forward a fuzzy comprehensive evaluation model for mobility model (FCEM). In this model, the membership function of fuzzy mathematics is used to deal with the fuzzy evaluation of each index of the mobility model. The model realizes the quantitative evaluation of mobility model. This model not only provides new ideas and methods for mobility model evaluation, but also provides help and guarantee for mobile node modelling. Finally, the application of the model is demonstrated through the evaluation of the random waypoint (RWP) model.

Keywords: mobility model, evaluating indicator, membership function, fuzzy comprehensive evaluation

Enhance detecting and preventing scheme for ARP Poisoning using DHCP

Vidya Srivastava, Dayashankar Singh

Computer Modelling & New Technologies 2017 21(2) 93-99

The client which is using LAN for mapping network address connected to its corresponding MAC address is done by Address Resolution Protocol, which is a primary protocol. It is well known that ARP is determined and works properly in case there is no malignant client in the network but in practical scenario it is not possible. The primary motive of an attacker is always tried to find a

strategy which is further accomplished to launch various attacks. ARP gives this accountability – the unsubstantiated and stateless characteristics of the protocol which accredit the attacker to conduct biggest level attacks. In this paper, an attempt is made to resolve out or minimize the attempt of attacker by providing a validation using DHCP server. By the introduction of DHCP (Dynamic host control protocol) such that if an attacker applies the IP of host not in network can be prohibited. The simulation result has been shown in the dissertation report. By the response of DHCP correct matching of IP and MAC could only respond and thus poisoning can be detected and protected successfully.

Keywords: Address Resolution Protocol, Network security, MiTm