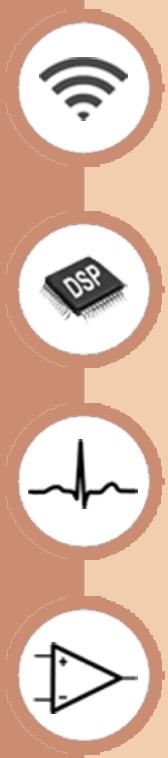
International Journal of Advances in Telecommunications Electrotechnics, Signals and Systems

a publication of the International Science and Engineering Society



Vol. 2, No. 3
2013

ISSN: 1805-5443

www.ijates.org

I J A T E S²

International Journal of Advances in Telecommunications Electrotechnics, Signals and Systems

a publication of the International Science and Engineering Society

Vol. 2, No. 3, 2013 ISSN: 1805-5443

Aims and Scope

The International Journal of Advances in Telecommunications, Electronics, Signals and Systems (IJATES²) is an all-electronic international scientific journal with the aim to bring the most recent and unpublished research and development results in the area of electronics to the scientific and technical societies, and is supported by the ISES (International Science and Engineering Society, o.s.). The journal's scope covers all the aspects of telecommunication, signal processing, theory and design of circuits and systems for electronics.

The IJATES² is ready to publish experimental and theoretical full papers and letters submitted by prospective authors. Paper submitted for publication must be written in English and must follow a prescribed format. All papers are subjected to a critical peer-review prior to publication.

The IJATES² is an open access journal which means that all content is freely available without charge to the user or his/her institution. Users are allowed to read, download, copy, distribute, print, search, or link to the full texts of the articles in this journal without asking prior permission from the publisher or the author. This journal provides immediate open access to its content on the principle that making research freely available to the public supports a greater global exchange of knowledge.

www.ijates.org

CONTENTS

Vol. 2, No. 3, 2013 ISSN: 1805-5443

Announcements

Analysis of Two New Voltage Level Converters With Various Load Conditions

Mahendranath B and Avireni Srinivasulu, *SMIEEE*

***Abstract*—Application of level converter in dual supply voltage circuits is one of the most effective ways to reduce power consumption. To prevent static current, level converter is introduced as an interface at each low-to-high boundary. The design of an efficient level converter with least power consumption and overheads delay is one of the major design constraints. In this paper, two new level converter circuits with low power consumption are proposed for less propagation delay and load adaptability. The proposed level converter circuits are examined using cadence and the design parameters of a 180 nm CMOS process. The simulation results exhibit that proposed level converters can reduce propagation delay and increase speed over the existing circuits available. These level converters are simulated for different loads and operating conditions. The proposed level converters can operate at different values of V_{DDL} as +1 V, +1.8 V, +2 V and V_{DDH} of +3.3 V. The topology reports low sensitivity and has features suitable for VLSI implementation. The proposed circuits are suited for low power design without degrading performance.**

***Keywords*— Dual cascade voltage switch logic, level converter, low power, power dissipation, propagation delay**

I. Introduction

At present semiconductor device market demands the manufacture of storage devices having higher operating potentiality, and this triggers another surge in the switching speed. The power consumption has become a major issue on portable electronic systems. As a result, the level converter should have enough current to meet the requirements of transmission speed. This leads to considerable current fluctuations during a short time (d*i*/d*t*), which may raise switching noises on the power supply lines. In case of large capacitive loads, non-negligible voltage bumps are observed on the power supply lines. These are mainly due to the inductive bond wires, package and board traces, which may induce power supply and ground-bounce (switching noise or *L*d*i*/d*t* noise). This noise can also lead to data transition delay, oscillation at the end of signal transitions and cross-talk between adjacent signal lines. Moreover, it can even cause malfunctioning of the circuits that are connected to the same supply lines [1].

Due to quadratic relation between voltage and power consumption, reduction in the supply voltage is very effective in decreasing power consumption. However, that would be at the expense of the circuit delay. In order to lower the supply voltage without the degradation of system performance, Clustered Voltage Scaling (CVS) has been developed in which critical and non-critical paths of the circuit are grouped together [2]. In the CVS Scheme, by using low voltage (V_{DDL}) in the non-critical paths and using high supply voltage (V_{DDH}) in speed sensitive paths, the whole system of power consumption could be reduced without degrading the performance. Whenever an output from a low V_{DDL} cluster is to drive an input to a high V_{DDH} cluster, Level Conversion is required at the interface. The reason being the output from a low swing voltage (V_{DDL}) block cannot connect to a PMOS in a high swing voltage (V_{DDH}) block directly, since the PMOS cannot shut off with low voltage V_{DDL}. One of the main challenges in the CVS Scheme is to design Level Converters (LCs) with less power and overhead delay to interface low voltage (V_{DDL}) blocks with high voltage (V_{DDH}) blocks [3].

Despite the fact that level converter also consumes power, any formal technique that attempts to formulate the use of dual supply voltages for circuit design most possibly take the delay of and the power consumed into account. In other words, the level conversion must be accomplished by minimal delay and lower power consumption to achieve high performance CMOS circuits. Furthermore, a structure such as "V_{DDL} circuit - V_{DDH} circuit - V_{DDL} circuit - V_{DDH} circuit…" need a lot of level converters to be inserted at each "V_{DDL} circuit - V_{DDH} circuit" interface [4]. Hence both lowering the power consumed in the level converter and reducing the number of level converters as such becomes an important issue for the use of dual supply voltages. Care has been taken to design level converter accurately for lowering power consumption and propagation delay without degrading performance. In a traditional Dual Cascade Voltage Switching (DCVS), large delay has been resulted because of the contentions problem between different transistors on the level shifting path. This contention problem will lead to increase both in delay time and power consumption [5-13]. For practical purposes, a CMOS logic circuit with mixed gates operating on a lower supply voltage V_{DDL} and on a higher supply voltage V_{DDH} may be preferred. However, any gate operating on V_{DDH} and connected after the gate operating on V_{DDL} generates a short-circuit current.

Manuscript received February 28, 2013, revised July 26, 2013.

Mahendranath. B is with the Department of Electronics & Communication Engineering, Vignan University, Vadlamudi, 522213 Guntur, Andhra Pradesh, INDIA.

Avireni Srinivasulu is with the Department of Electronics & Communication Engineering, Vignan University, Vadlamudi, 522213 Guntur, Andhra Pradesh, INDIA. (Mobile: +91 9502223336; Fax No: +91 863 2534468; e-mail: avireni_s@yahoo.com (or) avireni@ieee.org).

doi: 10.11601/ijates.v2i3.45

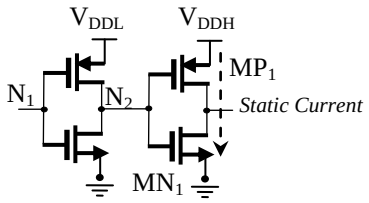


Fig. 1 Direct connection of V_{DDL} circuit and the V_{DDH} circuit

Figure.1 shows a CMOS logic circuit taking into consideration the output of first CMOS inverter is directly connected to the second CMOS inverter. The first CMOS inverter operates on lower supply voltage V_{DDL} and the second one on a higher supply voltage $V_{DDH.}$ If the input node N_1 of V_{DDL} circuit swings from high to low, the output node N_2 retains the reverse that is V_{DDL}. Subsequently, the logical high at node N_2 should turn off the pull-up transistor MP_1 and turn on the pull-down transistor MN_1. Although the voltage at node N_2 is high enough to activate the NMOS transistor MN_1, it cannot turn off the PMOS transistor MP_1 due to the fact that $V_{DDL} < V_{DDH} - |V_{th,P}|$. Therefore, there exist a static current flowing directly from the applied voltage source to ground through the path of MP_1 and MN_1. This static current also consumes power which is not desirable for low power application [6]. To restrict this unwanted power consumption, there should be a level converter circuit placed between V_{DDL} and V_{DDH} circuits. The main advantage of the level converter circuit is to reduce the static power consumption. Level converter transforms a logical high produced by a V_{DDL} circuit to the logical high for a V_{DDH} circuit. Thus, the condition of both networks MP_1 and MN_1 are activated at a time as described in Fig. 1 would not be possible, and the power consumption by the static current therefore gets eliminated. In order to achieve high performance CMOS circuits, the level converter circuit must be designed so as to have minimal delay and power consumption.

II. The Conventional Level Converters

The circuit diagram of conventional level converter is shown in Fig. 2. It is termed as Dual Cascade Voltage Switch (DCVS), interposed between gates operating on different supply voltages in a CMOS logic circuit, to prevent the short-circuit current and reduce power consumption. In this circuit there exists two cross coupled PMOS transistors MP_1 and MP_2 to generate the circuit load. The cross coupled PMOS transistor acts as a differential pair. As the output at one side gets pulled down, followed by opposite PMOS transistor to turn ON, and the output on that side will be plugged high. Below the PMOS load, there are two NMOS transistors MN_1 and MN_2 that are regulated by the input signal [7].

Although the level converter blocks the short circuit current, it consumes relatively large dynamic power while carrying out switching operation. If the CMOS logic circuit must have many level converters; the power consumption therefore would have increased to nullify the effort of decreasing the power consumption by using the two supply voltages V_{DDL} and V_{DDH}. Furthermore, this conventional level converter has relatively large delay as it has to rely on a contention between different transistors on the level conversion [7].

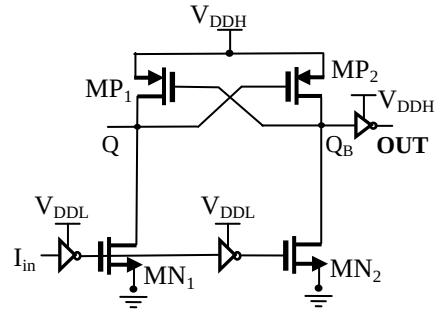


Fig. 2 Conventional level converter (DCVS)

Figure 3 shows the conventional level converter [6], with 12 transistors. This circuit has large power dissipation and propagation delay because it contains more number of transistors.

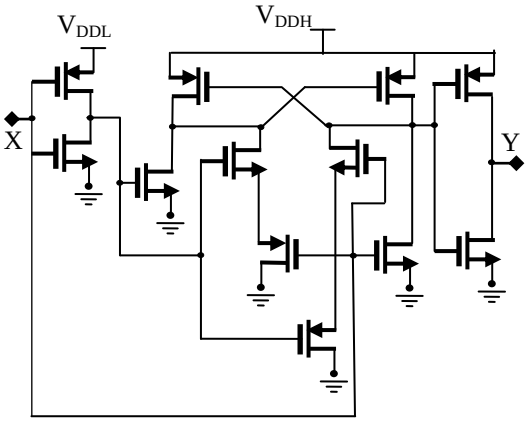


Fig. 3 Conventional level converter [6]

Figure. 4 shows the conventional level converter Standard Dual Cascade Voltage Switch (SDCVS) [7] which endorse the same cross coupled pair of PMOS transistors as DCVS with 14 transistors. The simulation profile of Fig. 4 has indicated the results in the higher propagation delay and power dissipation. To overcome these disadvantages, two new level converters with 10 transistors are proposed in this paper. These proposed circuits reduce the propagation delay and are shown in Fig. 5 and Fig. 6.

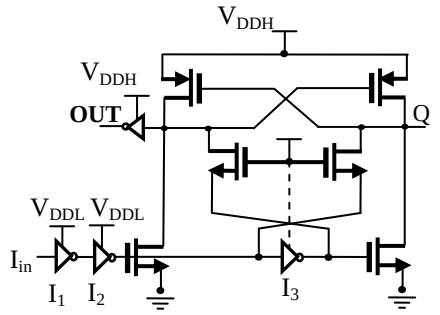


Fig. 4 Conventional level converter SDCVS [7]

The remaining sections of the paper are structured as follows. The new proposed level converter and circuit description is presented in section III. Simulation results and comparative conventional circuits are included in section IV. Finally, conclusion is included in section V.

III. PROPOSED LEVEL CONVERTERS

The proposed level converters are shown in Fig. 5 and Fig. 6. In these circuits, the level conversion circuit converts a signal V_{in} on the lower voltage side to a signal V_{out} on the higher supply voltage side; the signal V_{in} is transmitted to the inverted phase, which inturn generated by inverter constituted by transistors P_1 and N_1, ultimately transmitted to the gates of transistors N_2 and N_4. The respective gates of transistors P_2 and P_3 are cross connected to the drains of transistors P_3 and P_2, while the source of both transistors are connected to the higher supply voltage V_{DDH} and V_{03} is again connected to the gate terminals of P_4 and N_4 in the output stage, which is further connected to the higher supply voltage V_{DDH}. Here in this circuit the V_{in} signal is given as 0/1-V, 0/1.8-V and 0/2-V voltage levels and V_{DDH} is 0/3.3-V voltage level.

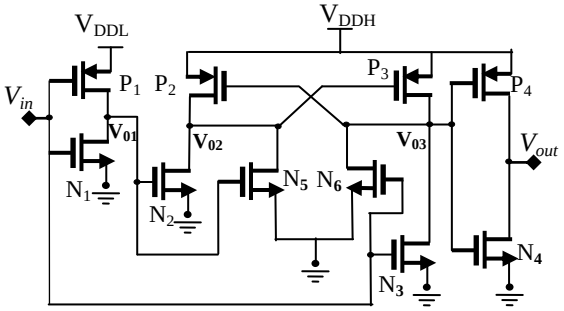


Fig. 5 Proposed level converter-1

The proposed level converter-1 is shown in Fig. 5, composed of inverter, DCVS logic and an output stage. It converts 0/1-V, 0/1.8-V and 0/2-V voltage swing to 0/3.3-V voltage swing. To reduce the body effect in Fig. 5, the bulk of the PMOS and NMOS transistors are connected to their source and drain terminals respectively. The proposed level converter employs the same cross-coupled pair of PMOS transistors as DCVS. Two NMOS transistors N_5 and N_6 are added to achieve high speed operation. These two transistors are connected to the gate terminals of P_2 and P_3 transistors to increase the switching speed of the output transition. In Fig. 5, P_1 and N_1 transistors acts as input inverter with low supply voltage level V_{DDL}. The output stage consists of PMOS P_4 and NMOS N_4. The node voltage V_{03} is connected to the gate terminal of transistor P_4 and N_4.

The circuit operates as follows: when the voltage level of input signal swings from high to low, the output voltage level of input inverter becomes the lower supply voltage V_{DDL}. Therefore, N_2 and N_5 transistors are turned ON; as a result, node V_{02} is then discharged to ground. Thus, P_3 is turned ON and then voltage level of node V_{03} becomes a higher supply voltage V_{DDH}. In this case, P_2, N_3 and N_6 transistors are turned OFF, and there by it has been made possible to prevent a short-circuit current from flowing between the higher supply voltage V_{DDH} and the ground. When the voltage level of the input signal is switched to logic high, the output voltage level of the input inverter becomes logic low. Therefore, N_3 and N_6 transistors are turned ON; as a result node V_{03} is then discharged to ground. In this case, transistor P_2 is turned ON and thereby, transistor P_3 turns OFF. There by transistors N_2 and N_5 are also turned OFF and made it possible to prevent a short-circuit current from flowing between the higher power supply voltage V_{DDH} and ground. It is thus clear that the output level transition in each input signal condition switches fast. These results contribute to faster output transition as well as contention problem on nodes V_{02} and V_{03}. As a result of this propagation delay time of the circuit gets reduced.

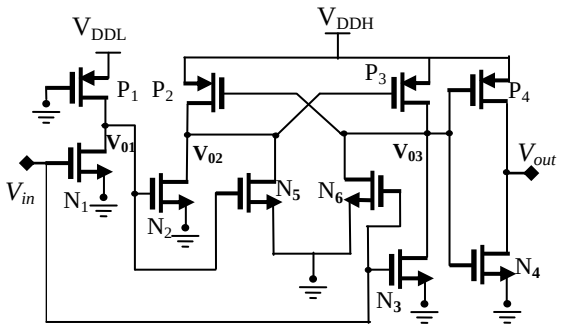


Fig. 6 Proposed level converter-2

Figure 6 display the proposed level converter-2 which is composed of pseudo inverter, DCVS logic and an output stage. It converts 0/1-V, 0/1.8-V and 0/2-V voltage swing to 0/3.3-V voltage swing. To reduce the body effect in Fig. 6 too the bulk of the PMOS and NMOS transistors are connected to their respective source and drain terminals. The proposed level converter employs the same cross-coupled pair of PMOS transistors as DCVS. Two NMOS transistors N_5 and N_6 are introduced to achieve high speed operation. These two transistors are connected to the gate terminals of P_2 and P_3 transistors to increase the switching speed of the output transition. In Fig. 6, transistor P_1 and N_1 transistor acts as input pseudo inverter with low supply voltage level V_{DDL}. The output stage consists of PMOS P_4 and NMOS N_4. The node voltage V_{03} is connected to the gate terminal of transistor P_4 and N_4.

Nextly, operation of the proposed level converter-2 is explained as follows. While the gate of PMOS P_1 transistor is connected to ground it will always be in ON condition. While the voltage level of input signal swings from low to high, the NMOS transistor N_1 turns ON and the voltage at node V_{01} is discharged to ground, so that the output voltage level of input inverter becomes low. Therefore, transistor N_3 and N_6 transistors are turned ON; as a result, node V_{02} is converted for supply of high voltage V_{DDH}. Thus, P_2 is turned ON and then voltage level of node V_{03} becomes a low. In this case, transistors P_3, N_2 and N_5 are turned OFF, and there by it is made possible to prevent a short-circuit current from flowing between the higher supply voltage V_{DDH} and the ground. When the voltage level of the input signal is switched to logic low, the output voltage level of the input inverter attains lower supply voltage V_{DDL}. Therefore, N_2 and N_5 transistors are turned ON; as a result node V_{03} then becomes higher supply voltage V_{DDH}. In this case, transistor P_3 is turned ON, and thereby, transistor P_2 to OFF. Moreover, transistors N_3 and N_6 are also turned OFF, and thereby, it is made possible to prevent a short-circuit current from flowing between the higher power supply voltage V_{DDH} and ground. It is thus clear that the output level transition in each input signal condition switches fast. These results provide faster output transition as well as contention into problems thrusting on nodes V_{02} and V_{03}. Consequently propagation delay time of the circuit becomes lesser than proposed level converter-1, with no short-circuit current flow; therefore it is possible to reduce the power dissipation.

IV. Simulation Results

The circuit in Fig. 5, Fig. 6 and low power level converter [6] and SDCVS [7] were simulated by using cadence and the model parameters of a 180 nm CMOS process. The simulations were carried out with pulse amplitude of +1 V, +1.8 V, +2 V and frequency of 1 MHz.

In order to observe the performance of the proposed level converters under various load conditions, the conventional and proposed circuits are simulated at load capacitance varying from 1 pF to 15 pF for +1 V input voltage and 1 pF to 25 pF for +1.8 V input voltage. Standard DCVS level converter in Fig. 2 suffers from the contention problem; therefore simulation results show maximum power and delay over head. On the other hand low power level converter [6] and standard SDCVS [7] level converter uses more transistors in its operation as compared to proposed level converters, that results in increase in power consumption and propagation delay.

The typical simulated input and output waveforms on the configuration of Fig. 5, with V_{DDL} = +1V, +1.8V, +2V, V_{DDH} = +3.3V and output load of C_L = 10 pF is presented in Fig. 7, 8, 10 respectively. It is evident from Fig. 7 simulated input and output waveforms are with amplitude of +1 V and +3.3V respectively with no voltage drop. For all the level converters the supply rail V_{DDL} of +1 V, +1.8V, +2 V, V_{DDH} of 3.3V and pulse input parameters shown in Table. I was used.

Figure 9 shows the simulated DC Response of proposed voltage level convrter-1 of Fig. 5. From Fig. 9 it is evident that the output voltage is varied with respect to input voltage.

The layout of the proposed voltage level converter-1 of Fig. 5 has been designed using Assura layout XL with 180 nm technology file. Layout of proposed voltage level converter-1 of Fig. 5 is shown in Fig 11. The area of the proposed voltage level converter-1 is 164.64 μm^2.

Fig. 12, 13, 15 shows the simulated waveforms on nodes V_{in}, V_{01}, V_{02}, V_{03} and V_{out} of the proposed level converter-2 with V_{DDL} = +1V, +1.8V, +2V, V_{DDH} = +3.3V and output load of C_L = 10 pF respectively. It is evident from Fig. 10 simulated input and output waveforms are with amplitude of +1 V and +3.3V respectively without any voltage drop. For all the level converters the supply rail V_{DDL} of +1 V, +1.8V, +2 V, V_{DDH} of +3.3V and pulse input parameters shown in Table. I was used. The aspect ratios of Fig. 5 and 6 transistors are given in table II.

Figure 14 shows the simulated DC Response of proposed voltage level convrter-2 of Fig. 6. From Fig. 14, it is evident that the output voltage is varied with respect to input voltage.

The layout of the proposed voltage level converter-2 of Fig. 6 has been designed using Assura layout XL with 180 nm technology file. Layout of proposed voltage level converter-2 of Fig. 6 is shown in Fig 16. The area of the proposed voltage level converter-1 is 170.79 μm^2.

TABLE I – Pulse Input Parameters

Voltage 1	+1 V, +1.8 V, +2V
Voltage 2	0 V
Delay Time	1 ps
Rise Time	2 ps
Fall Time	1 fs

TABLE II – Aspect Ratios of Fig. 5 and 6

Transistor	W (μm)	L (μm)
P_1- P_4	2	0.18
N_1-N_6	2	0.18

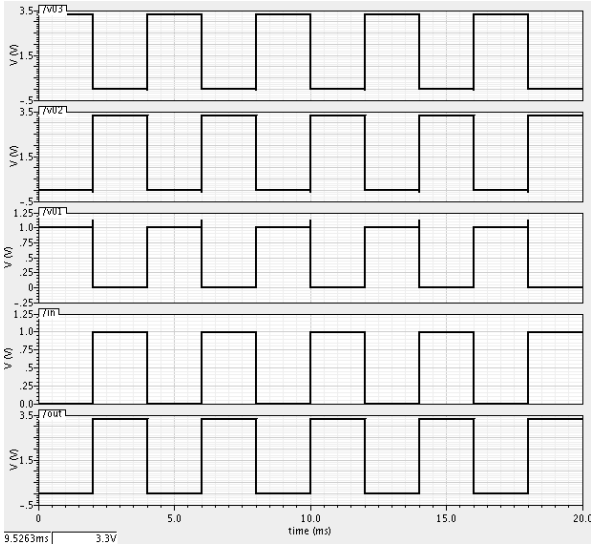


Fig. 7 Simulated input and output waveforms of Fig. 5 with load capacitance C_L= 5 pF (Supply rail voltages V_{DDH} = +3.3V and V_{DDL} =1 V).

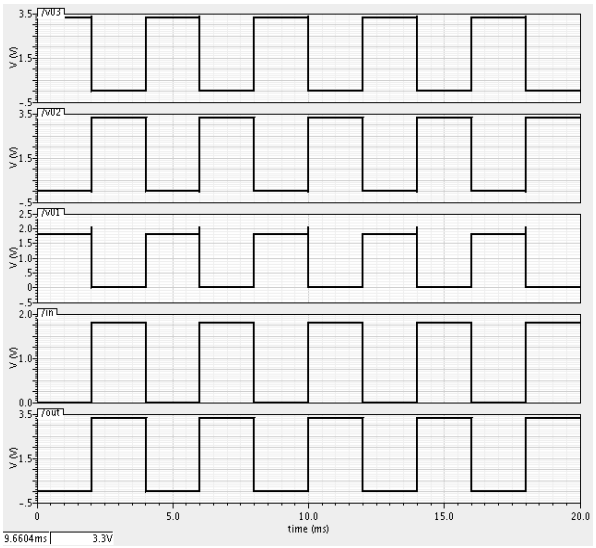


Fig. 8 Simulated input and output waveforms of Fig. 5 with load capacitance C_L= 10 pF (Supply rail voltages V_{DDH}=3.3V and V_{DDL}=1.8 V).

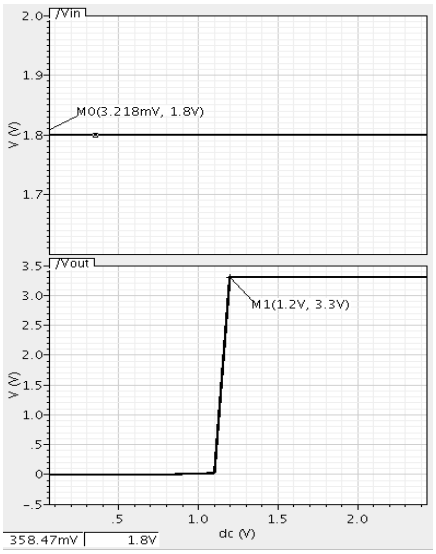


Fig. 9 Simulated DC response of Fig. 5 with load capacitance C_L= 10 pF (Supply rail voltages V_{DDH}= +3.3V and V_{DDL}= +1.8 V).

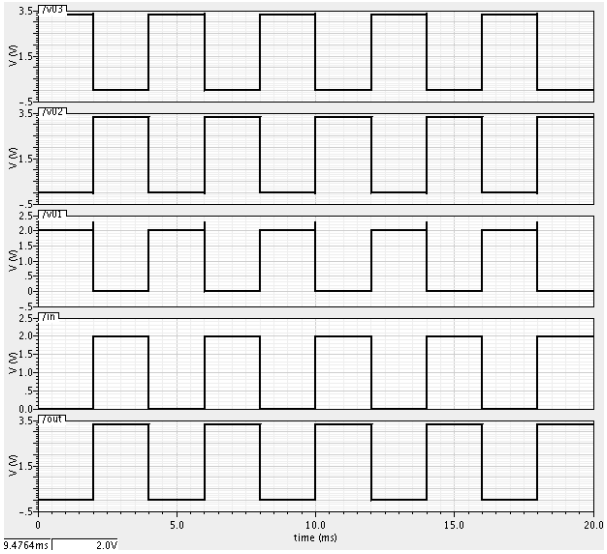


Fig. 10 Simulated input and output waveforms of Fig. 5 with load capacitance C_L= 20 pF (Supply rail voltages V_{DDH}=3.3V and V_{DDL} =2 V).

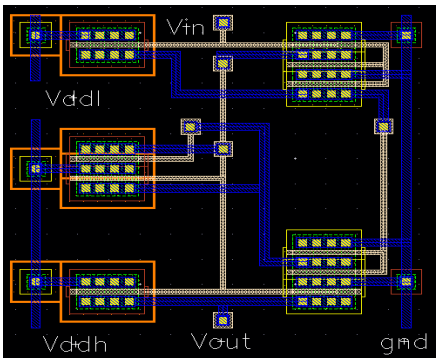


Fig. 11 Layout of proposed Fig 5

Figure 17 empowers the comparison of propagation delay for proposed level converters, low power level converter [6] and conventional SDCVS [7]. In Fig. 17 load capacitance is varied from 1 pF to 25 pF with input pulse amplitude of +1.8 V and frequency of 1 MHz. It is thus clear that the proposed level converters have less propagation delay and load adaptability than the earlier ones. From the simulation results, it is proved that the delay and power consumption in proposed level converters is very less as compared to that of in the standard DCVS, SDCVS and low power level converters. It is also observed that falling delay and rising delay in case of proposed level converter is much less as compared to the existing circuits. Accordingly, the proposed level converter circuits can be effectively applied to LSI high speed input-output circuit, as an interface between internal and external buses such as a server or exchanger, and as an interface circuit between optical devices for optical communication and an LSI, etc.

Table III represents the comparative study of propagation delay for the proposed level converters, low power level converter [6] and conventional SDCVS [7]. In Table II load capacitance is varied from 1 pF to 15 pF with input pulse amplitude of +1 V and frequency of 1MHz. From Table II also it is clear that there is a significant reduction of delay in the proposed level converter compared to the existing circuits.

Table IV shows the simulated values of the propagation delay for the proposed level converter-1 for different supply

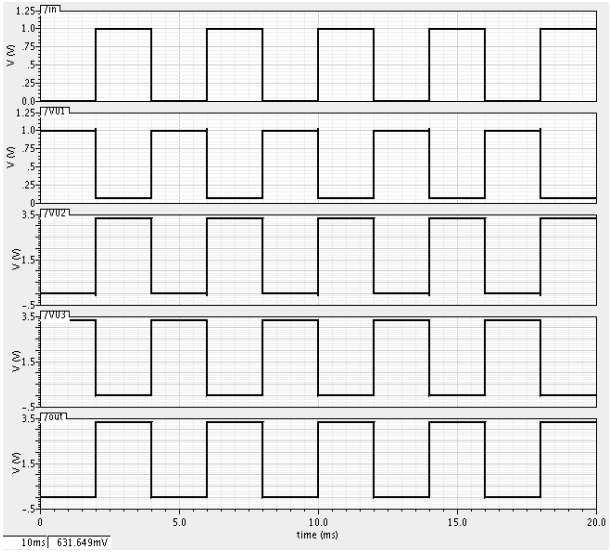


Fig. 12 Simulated input and output waveforms of Fig. 6 with load capacitance C_L= 5 pF (Supply rail voltages V_{DDH}= +3.3V and V_{DDL} = 1 V).

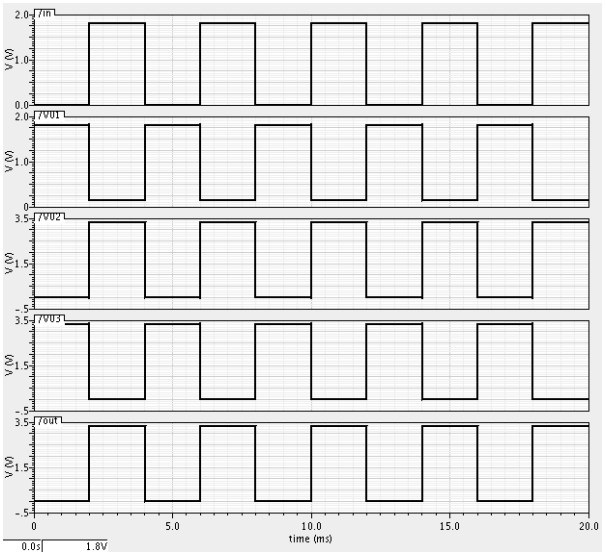


Fig. 13 Simulated input and output waveforms of Fig. 6 with load capacitance C_L= 5 pF (Supply rail voltages V_{DDH}=3.3V and V_{DDL}= 1.8 V).

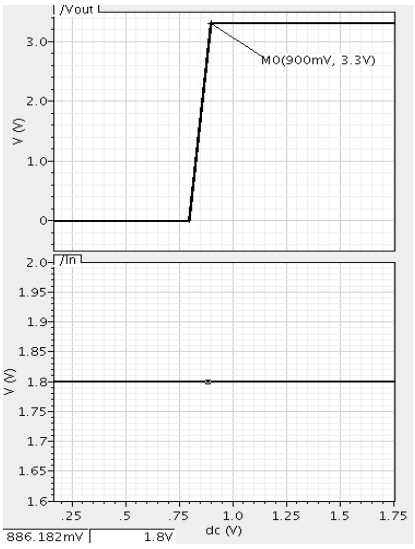


Fig. 14 Simulated DC response of Fig. 6 with load capacitance C_L= 10 pF (Supply rail voltages V_{DDH}= +3.3V and V_{DDL}= +1.8 V).

voltages V_{DDL} of (+1 V, +1.2 V, +1.8 V, and +2 V) and V_{DDH} of (+3.3 V, +5 V) with load capacitor C_L ranging from

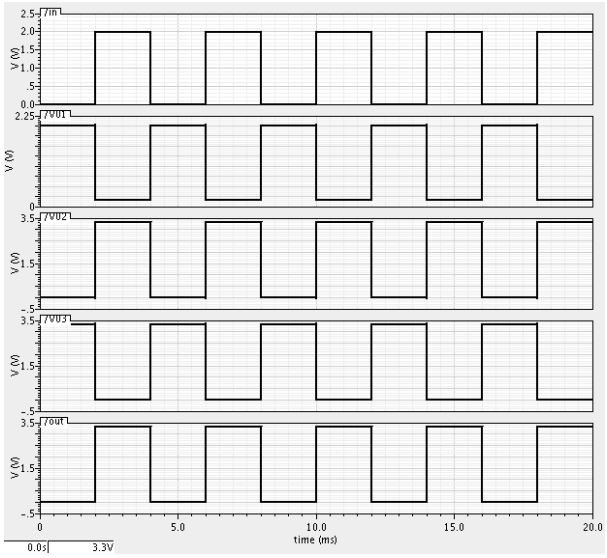


Fig. 15 Simulated input and output waveforms of Fig. 6 with load capacitance C_L= 5 pF (supply rail voltages V_{DDH}=+3.3V and V_{DDL}= +2 V).

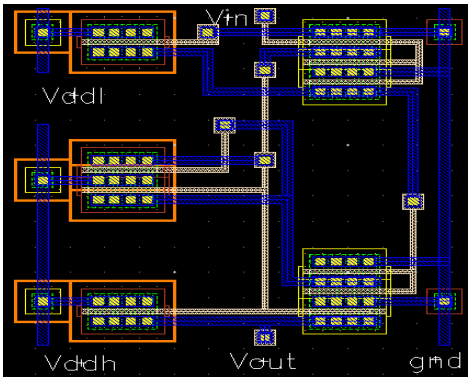


Fig. 16 Layout of proposed Fig 6

from 1 pF to 10 pF. It is also doubly clear that proposed level converters gives less propagation delay for low voltage and high voltage conversion and hence the proposed converter is very useful for low voltage and high voltage application such as PCI-X Interface.

Table V shows the simulated values of the propagation delay for the proposed level converter-2 for different supply voltages V_{DDL} of (+1 V, +1.2 V, +1.8 V and +2 V) and V_{DDH} of (+3.3 V, +5 V) with load capacitor C_L ranging from 1 pF to 10 pF. It is therefore clear that proposed level converter-2 of Fig. 6 gives less propagation delay for low voltage and high voltage conversion compared to proposed level converter-1 of Fig. 5 and hence the former is stated to have an edge over the later in efficiency and performance as well.

V. CONCLUSION

Two new level converters have been tested and designed with 180 nm CMOS Technology with V_{DDL} = +1 V, +1.8 V, +2 V and V_{DDH} = 3.3 V. It has resulted into minimizing the propagation delay and switching noise. The proposed level converters therefore can reduce the contention problem that existed in the conventional DCVS circuit. Also there is a reduction in the rising time and falling time in the proposed level converters when compared to conventional designs. Simulation results however show that the proposed designs have recorded less noise and less propagation delay, compared to previous designs. The topology reports low sensitivity and has features suitable for VLSI implementation.

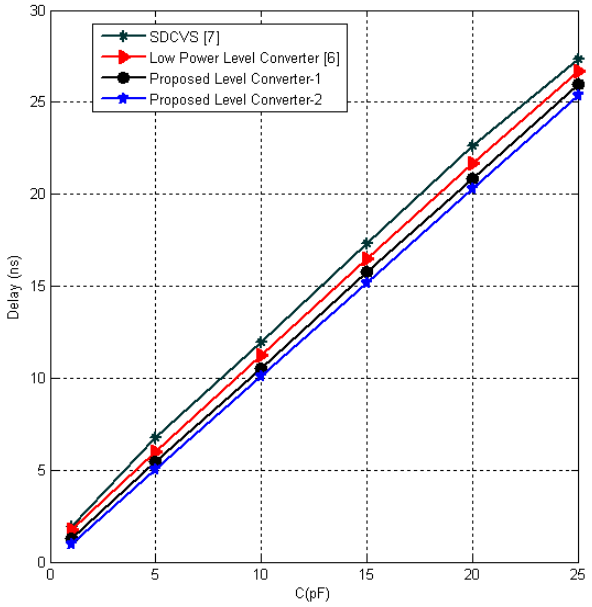


Fig. 17 Proposed level converters and conventional level converters propagation delay comparison for different loading conditions with V_{DDL} = +1.8V and V_{DDH} = +3.3 V.

REFERENCES

[1] Y.Lin, Xuecheng Zou, Zhaoxiao Zheng, W. Huo, Xiaofei Chen and W. Kang, "High-speed, Low Switching Noise and Load Adaptive Output Buffer" *in Proc. the Int.Symp. Integrated circuits 2009*, (ISCI 2009), Singapore, 14-16 Dec, 2009, pp. 280-282.

[2] M. Hamada, M. Takahashi, H. Arakida, A. Chiba, T. Terazawa, T. Ishikawa, M. Kanazawa, M. Igarashi, K. Usami, and T. Kuroda, "A top-down low power design technique using clustered voltage scaling with variable supply-voltage scheme," *in Proc. the IEEE Custom Integrated Circuits Conf.*, 1998, pp. 495–498.

[3] L. Benini, E. Macii, and G. De Micheli, "Designing low power circuits: Practical recipes," *IEEE Circuit Syst. Mag.*, vol. 1, no. 1, pp. 6–25, 2001.

[4] J. Y. Jou, and D. S. Chou, "Sensitisable-path-oriented clustered voltage scaling technique for low power," *in Proc. the IEEE* on *Comput. Digit. Tech.*, vol. 145, no. 4, Jul 1998, pp. 301-307.

[5] J. S. Shor, Y. Afek, and E. Engel, "IO buffer for high performance, low-power application," *in Proc. the IEEE Custom Integrated Circuits Conf.*, 1997, pp. 595-598.

[6] Chin Ping-Yuan, Yu Chien-Cheng "A voltage Level Converter Circuit Design with Low Power Consumption" *in Proc. the IEEE International Conference on ASIC*, Oct. 22-27, 2005, Shanghi, pp. 358-359.

[7] Chien-Cheng Yu, Wei-Ping Wang and Bin-Da Liu "A new level converter for low power applications" *in Proc. the IEEE International Symposium on Circuits syst.*, 2001, vol.1, pp. 113-116.

[8] Mahendranath. B and Avireni Srinivasulu, "Performance analysis of a new CMOS output buffer", *in proc. the IEEE International Conference on Circuit, Power and Computing Technologies* (IEEE ICCPCT-2013), Kumaracoil, India, Mar 21-22, 2013, pp. 752-755. DOI: 10.1109/ICCPCT.2013.6529041.

[9] T. Venkata Rao and Avireni Srinivasulu, "Modified level restorers using current sink and current source inverter structures for BBL-PT full adder", *Radioengineering*, vol. 21, no. 4, pp. 1279-1286, Dec 2012. ISSN: 1210-2512.

[10] Shih-Lun Chen, Ming-Dou ker "An Output Buffer for 3.3 V application in a 0.13 µm 1/2.5-V CMOS Process" *IEEE transaction on circuits and systems –II: express briefs*, vol. 54, pp. 14-18, Jan 2007.

[11] R. J. Baker, H. W. Li, and D. E. Boyce, "CMOS Circuit Design, Layout, and Simulation", New York: IEEE Press, 1997.

[12] Avireni Srinivasulu and Madugula Rajesh, "ULPD and CPTL Pull-Up Stages for Differential Cascode Voltage Switch Logic," *Journal of Engineering* (*Hindawi*), vol. 2013, Article ID 595296, 5 pages, 2013. DOI: 10.1155/2013/595296 ISSN: 2314-4912.

[13] I. Kavitha Madhuri, Avireni Srinivasulu, P. Chandra Shaker and Indira Priyadarsini. S, "Two CMOS Schmitt trigger circuits using current sink and pseudo inverters", *International Journal of Computer Applications*, iC3S - Number 3 , Year of Publication: 2013, pp. 18-21. ISSN: 0975 – 8887.

Mahendranath Bheemanboina was born in Tharlupadu, Prakasam District, (A.P), India in 1989. He received the B.Tech degree in electronics and communication engineering from J.N.T.University, Kakinada in 2010. Currently, he is working towards his M.Tech thesis under the esteemed guidance of Dr. Avireni Srinivasulu. His area of research includes VLSI Design.

Avireni Srinivasulu was born in Thurimella, Andhra Pradesh, India. He received the B.Tech degree in electronics and communication engineering from Sri Venkateswara University, Tirupati in 1986, M.E, degree in power electronics engineering from Gulbarga University, Gulbarga in 1991, M.S, degree in software systems from Birla Institute of Technology and Science (BITS), Pilani in 1998 and Ph.D, degree in electronics and communication engineering (VLSI Design) from Birla Institute of Technology, Mesra, India in 2010.

He worked as a Lecturer, Assistant Professor, Reader and Associate Professor in the Department of Electronics and Communication Engineering, T.G.L.G. Polytechnic, Adoni; Guru Nanak Dev Polytechnic, Bidar; K.S.R.M. College of Engineering, Kadapa; Defence University Engineering College, Debre-Zeit, Ethiopia; Kigali Institute of Science, Technology & Management, Kigali, Rwanda; Birla Institute of Technology, Mesra, Ranchi and working as a Professor in the Department of Electronics and Communication Engineering, Vignan University, Vadlamudi, Guntur, India. He has 25 years of teaching and 15 years of research experience in the Department of Electronics and Communication Engineering.

Dr. Avireni.S is a senior member of IEEE, senior member of IACSIT, life member of I.S.T.E and a member of the Institution of Engineers (India). He has published over 30 articles in international journals and international conference proceedings; his main research areas are microelectronics, VLSI design and analog ASIC.

TABLE. III COMPARISON OF PROPAGATION DELAY FOR FIG. 5 FIG. 6 AND CONVENTIONAL LEVEL CONVERTERS FOR DIFFERENT LOAD CONDITIONS

Level Converters	**Supply Rail Voltages**		**Propagation Delay (ns)**			
	$+V_{DDL}$	$+V_{DDH}$	C_L=1 pF	C_L=5 pF	C_L=10 pF	C_L=15 pF
SDCVS [7]	1 V	3.3 V	4.553	9.767	15.04	19.68
LPLC [6]	1 V	3.3 V	4.627	8.631	14.3	15.9
Proposed Fig.5	1 V	3.3 V	4.868	8.376	13.28	16.54
Proposed Fig.6	1 V	3.3 V	2.998	7.169	12.41	15.13

TABLE. IV COMPARISON OF PROPAGATION DELAY FOR FIG. 5 WITH DIFFERENT LOAD CONDITIONS AND SUPPLY RAIL VOLTAGES

Supply Rail Voltages		**Capacitor Output load (C_L = 1 pF)**			**Capacitor Output load (C_L = 5 pF)**			**Capacitor Output load (C_L = 10 pF)**		
$+V_{DDL}$ (Volts)	$+V_{DDH}$ (Volts)	Rise Time (fs)	Fall Time (ps)	Delay (ns)	Rise Time (ps)	Fall Time (ns)	Delay (ns)	Rise Time (ps)	Fall Time (ns)	Delay (ns)
1	3.3	7.93	515.6	4.187	45.85	2.532	8.376	4.209	5.057	13.61
1.2	3.3	4.44	510	2.404	66.29	2.53	6.59	5.162	5.063	11.82
1.8	3.3	37.54	509.3	1.462	101	2.535	5.677	1.055	5.078	10.9
1.2	5	22.61	690.6	2.487	618	3.397	7.152	1.241	6.861	13.09
1.8	5	55.27	683.8	1.457	936.3	3.402	6.139	498.9	6.85	11.89
2	5	960.8	682.9	1.546	618	3.397	7.152	1.241	6.861	13.09

TABLE. V COMPARISON OF PROPAGATION DELAY FOR FIG. 6 WITH DIFFERENT LOAD CONDITIONS AND SUPPLY RAIL VOLTAGES

Supply Rail Voltages		**Capacitor Output load (C_L = 1 pF)**			**Capacitor Output load (C_L = 5 pF)**			**Capacitor Output load (C_L = 10 pF)**		
$+V_{DDL}$ (Volts)	$+V_{DDH}$ (Volts)	Rise Time (ps)	Fall Time (ps)	Delay (ns)	Rise Time (ps)	Fall Time (ns)	Delay (ns)	Rise Time (ps)	Fall Time (ns)	Delay (ns)
1	3.3	496	516.4	2.998	26.42	2.534	7.169	6.521	5.706	12.41
1.2	3.3	4.332	509.9	1.622	574.3	2.532	5.82	2.348	5.05	11.10
1.8	3.3	1.134	520.7	1.112	719.2	2.534	5.134	7.194	5.057	10.56
1.2	5	53.15	690.5	1.709	570.1	3.404	6.382	1.816	6.865	12.3
1.8	5	665	691.2	1.226	613.6	3.445	5.94	693.2	6.803	11.74
2	5	710.3	696.1	1.148	14.43	3.409	5.837	15.89	6.868	11.75

A hybrid unsupervised and supervised clustering applied to microarray data

Raul Măluţan, Pedro Gómez Vilda, Monica Borda

***Abstract*— Clustering methods have been often applied to large data with the main purpose of reducing the dimension, time computation and identifying clusters with similar behavior. This work presents a state-of-the-art in unsupervised clustering and cluster validation. It proposes a method for hybrid bi-clustering of microarray data combined with a supervised validation for determining the optimal amount of clusters of genes.**

***Keywords* — microarray, clustering, internal validation, external validation, gene shaving.**

I. Introduction

The microarray data processing challenges nowadays consists of how to make it more reliable, easy to use and efficient. In this task other fields of knowledge as Signal and Image Processing, Pattern Recognition, or Statistical Data Analysis [1] may help in yielding their enormous potential in solving problems as microarray image enhancement, segmentation, correction, gridding, data analysis, reliable expression estimation in relation with hybridization dynamics, etc. Others have to see with data interpretation, dimensionality reduction, cluster analysis, etc.

Clustering techniques play an important role in analyzing high dimensional data such as microarray data. An analysis of microarray data is a search for genes that have similar, correlated patterns of expression. This indicates that some of the data might contain redundant information. For example, if a group of experiments were more closely related than it was expected, some of the redundant experiments can be ignored, or use some average of the information without loss of information.

Data analysis methods [2] can be grouped in two categories: supervised and unsupervised. In the unsupervised approach, also known as clustering, data is organized without a priori information.

This paper describes several unsupervised algorithms used mostly for microarray data, like k-means, Partitioning Around Medoids (PAM) and Expectation-Maximization (EM). These algorithms have proven to be a useful when the number of clusters is known or can be estimated. Two of them, k-means and PAM, are based on minimizing the mean squared error, while the third, EM, on mixture modeling.

The algorithms were run on several data sets [3], observing that the quality of the obtained clusters is dependent on the number of clusters specified. To assess the effectiveness of these algorithms, but also to estimate the actual number of clusters, several clustering validation method were implemented. These methods consist in calculating internal and external indices used to estimate the optimal number of clusters for each algorithm.

The microarray data has a particularity compared with other type of data. A two dimension data is in fact a gene expression matrix which usually has the rows corresponding to genes from an experiment and the columns corresponding to different experiments. If one finds that two rows are similar, it can be assumed that the genes corresponding to the rows are co-regulated and functionally related, and by comparing two columns it can found which genes are differentially expressed in each experiment. To perform a comparison a similarity measure between the objects, genes or experiments under comparison, has to be used. Mostly the same algorithm is used for analyzing both the genes and the experiments, *i.e.* bi-clustering of microarray data.

Considering the dimensionality of the data, large amount of genes and small number of experiments, we propose in this paper a hybrid bi-clustering, where we combine unsupervised methods with the purpose of obtaining the optimal combination of clusters.

Besides, as a supplementary validation, a supervised method was used on the same data. Supervised classification represents the issue of identifying the subset to which new observations belong, where the identity of the subset is unknown, on the basis of a training set of data containing observations whose subset is known. Therefore the classification will display a variable behavior which can be analyzed by statistics. It is required for new sample items to be placed into the respective groups based on quantitative information on one or more measurements, attributes or features and based on the training set in which previously decided groupings are already established.

Manuscript received Ortober 28, 2012, revised March 04, 2013. This work was supported by the project "Development and support of multidisciplinary postdoctoral programmes in major technical areas of national strategy of Research - Development - Innovation" 4D-POSTDOC, contract no. POSDRU/89/1.5/S/52603, project co-funded by the European Social Fund through Sectoral Operational Programme Human Resources Development 2007-2013.

R. Malutan is with the Communications Department, Technical University of Cluj-Napoca, 26-28 George Baritiu St., 400027 Cluj-Napoca, Romania, (phone: 004-0264-401564; fax: 004-264-401575; e-mail: raul.malutan@com.utcluj.ro).

P. Gómez Vilda is with Departamento de Arquitectura y Tecnología de Sistemas Informáticos (DATSI), Universidad Politécnica de Madrid, Campus de Montegancedo, s/n, 28660, Boadilla del Monte, Madrid, Spain (e-mail: pedro@pino.datsi.fi.upm.es).

M. Borda is with the Communications Department, Technical University of Cluj-Napoca, 26-28 George Baritiu St., 400027 Cluj-Napoca, Romania (e-mail: Monica.Borda@com.utcluj.ro).

doi: 10.11601/ijates.v2i3.21

II. CLUSTER ANALYSIS

A. Unsupervised algorithms

The unsupervised clustering methods are divided in two categories: hierarchical and non-hierarchical. Hierarchical methods group the objects in an iterative way, generating a hierarchical tree structure, also known as dendogram. On the other hand, the non-hierarchical clustering, or partitioning methods, does the partitioning of a data set into a predefined number of different clusters, without a hierarchical structure. Beside, these algorithms produce an integer number of partitions, and also they optimize a certain criterion function. The partitioning methods classify the data in *k* clusters which must fulfill the following conditions:

- each group should contain at least one element;
- each object must belong to one group only.

For the microarray data the most suitable clustering methods are unsupervised ones, because we cannot observe the (real) number of clusters in the data [4]. From the unsupervised algorithms used in microarray data analysis the k-means, PAM (Partitioning Around Medoids) and EM (Expectation Maximization) are the most frequently used. The first two from the list are using the minimization of the root mean square error and the third one is using the mixing models methods.

The PAM algorithm [2] is based on the search for *k* representative objects or medoids among the objects of the dataset. These objects should represent the structure of the data. After finding a set of *k* medoids, *k* clusters are constructed by assigning each object to the nearest medoid. The goal is to find *k* representative objects which minimize the sum of the dissimilarities of the objects to their closest representative object. The algorithm first looks for a good initial set of medoids. Then it finds a local minimum for the objective function, that is, a solution such that there is no single switch of an object with a medoid that will decrease the objective.

The k-means algorithm [5], an unsupervised learning algorithm, has been used to form clusters of genes in gene expression data analysis. The algorithm takes the number of clusters (*k*) to be calculated as an input. The number of clusters is usually chosen by the user. The procedure for k-means clustering is as follows:

1. First, the user tries to estimate the number of clusters.
2. Randomly choose N points into *k* clusters.
3. Calculate the centroid for each cluster.
4. For each point, move it to the closest cluster.
5. Repeat steps 3 and 4 until no further points are moved to different clusters.

The Expectation-Maximization (EM) algorithm [6] is a method for finding maximum likelihood estimates of parameters in statistical models, where the model depends on unobserved latent variables. This is a general method for optimizing likelihood functions and is useful in situations where data might be missing or simpler optimization methods fail.

B. Unsupervised clustering validation

Clustering validation is a technique to find a set of clusters that best fits natural partitions, *i.e.* number of clusters, without any class information. There are two types of clustering techniques [7], [8]: *external validation*, based on previous knowledge about data and *internal validation*, based on the information intrinsic to the data alone.

There are three external indexes which were used in our previous study [3], Rand index, Jaccard coefficient, and Fowlkes and Mallows index, and they are going to be used also in this paper.

In contrast to external validation, internal validation evaluates the clustering without any a priori information. The obtained values are known as internal indexes as they are computed from the data used for clustering. In this paper we evaluate for the microarray data the following internal indexes: silhouette, Calinski-Harabasz index, Krzanowski-Lai index, Hartigan index and Davies-Bouldin index.

Silhouette index calculates the silhouette width for each sample, average silhouette width for each cluster and overall average silhouette width for a total data set:

$$S(i)=\frac{b(i)-a(i)}{\max\{a(i),b(i)\}}, \tag{1}$$

where *a(i)* is the average dissimilarity of *i*-object to all other objects in the same cluster; *b(i)* is the minimum of average dissimilarity of *i*-object to all objects in other cluster. The largest overall average silhouette indicates the best clustering.

Calinski-Harabasz index is defined by:

$$CH(k)=\frac{B(k)/(k-1)}{W(k)/(n-k)}, \tag{2}$$

where *k* denotes the number of clusters, and *B*(*k*) and *W*(*k*) denote the between and within cluster sums of squares of the partition, respectively. An optimal number of clusters is then defined as a value of *k* that maximizes *CH*(*k*).

Krzanowski-Lai index is given by the equation:

$$KL(k)=\left|\frac{DIFF(k)}{DIFF(k+1)}\right|, \tag{3}$$

where $DIFF(k)=(k-1)^{2/p}W(k-1)-(k)^{2/p}W(k)$ and *p* denotes the number of features in the data set. A value of *k* is optimal if it maximizes *KL*(*k*).

Davies-Bouldin index is a function of the ratio of the sum of within-cluster scatter to between-cluster separation:

$$DB=\frac{1}{n}\sum_{i=1}^{n}\max_{i\neq j}\left\{\frac{S_n(Q_i)+S_n(Q_j)}{S(Q_i,Q_j)}\right\}, \tag{4}$$

where *n* is the number of clusters, S_n is the average distance of all objects from the cluster to their cluster centre, $S(Q_i,Q_j)$ is the distance between clusters centres. Consequently, Davies-Bouldin index will have a small value for a good clustering.

For the review, the manuscript is submitted to the IJATES[2] Editorial Board only electronically in PDF format. Do not change any formatting; otherwise you undergo the risk to be directly rejected without review process.

C. Gene Shaving

The method of Gene Shaving is designed to extract coherent and typically small clusters of genes that vary as much as

possible across the samples. According to [9], the algorithm consists of the following steps:

1. Start with the entire expression matrix *X*, each row centered to have zero mean
2. Compute the leading principal component of the rows of *X*
3. Shave off the proportion (typically 10 %) of the genes having smallest absolute inner-product with the leading principal component
4. Repeat steps 2 and 3 until only one gene remains
5. This produces a nested sequence of gene clusters $S_N \supset S_k \supset S_{k_1} \supset S_{k_2} \supset \cdots \supset S_1$, where S_k denotes a cluster of *k* genes. Estimate the optimal cluster size $\hat{k}$ using the gap statistic [10].
6. Orthogonalize each row of *X* with respect to $\bar{x}_{S_k}$, the average gene in $S_{\hat{k}}$
7. Repeat steps 1-5 above with the orthogonalized data, to find the second optimal cluster. This process is continued until a maximum of *M* clusters are found, where *M* is chose a priori.

When implementing this method we made some changings compared with the steps form [9]. But I have made some changes in the algorithm. First of all, we shaved off not *α %* genes, but 1 gene each time. This is because we will lose the precision of the algorithm if using *α%*. For example, supposing we remain with S_k clusters with *k* being 135, 122, ..., 53, 47, etc. genes, and according to the gap statistic step, the algorithm decides to have a cluster with 52 or 47 genes, which is not correct. This is the reason why we have decided to have clusters with ..., 53, 52, 51, 50, 49, 48,... genes, in this way hoping to obtain a maximum *Gap(k)* closer to ideal 50. Also when computing the gap statistics we have made some changes. If we consider all possible permutations and after that finding each *D(k)*, then, in case of 150 genes with 4 characteristics of each gene, it is required to have all possible permutations of the matrix by permuting the elements within each row. This means that each row has from 4 parameters a number of 24 possible permutations, this implies that we have another sets of 24150 matrices, which is too much (we cannot take for example 3rd permutation from gene 1 with 3rd permutation for gene 2, 3rd permutation for gene 3, and so on, because we obtain the same *D(k)*). We have tried to consider random matrices, a number of 5000, but the problem in this case is that the result is varying at every other analysis and they are also different at each simulation. Finally we have decided to use just the first input matrix and get its *D(k)* and *Gap(k)*, without any permutations.

III. Data Clustering

A. Unsupervised clustering

Three unsupervised algorithms were used to cluster microarray data. We used for our study two different datasets from public Affymetrix databases. The first set was the Chowdary database [11] the authors compared pairs of snap-frozen and RNA later preservative-suspended tissue from 62 lymph node-negative breast tumors and 42 colon tumors, with purpose of separating them. The second set [12] contains 24 acute lymphoblastic leukemia (ALL), 28 acute myelogenous leukemia (AML) and 20 mixed-lineage leukemia (MLL) samples.

For the microarray data the clustering was done by a two-way clustering or bi-clustering [13] in which both the samples and the genes are clustered in the same time using the portioning method.

Regarding the conclusions from [3] a useful classification was obtained for microarray data when EM clustered the genes and k-means the samples. In this study we will use the k-means algorithm to cluster the samples and the PAM and respectively the EM algorithm to cluster the genes. Before combining the algorithms we applied the clustering validation methods, both external and internal indexes.

Also, based on these results we combined the unsupervised method with a supervised one. So we clustered the samples by k-means algorithm and we classify the genes using the gene shaving method.

In Table I, the numbers of clusters obtained after using the internal and external indexes are indicated. For the k-means algorithm the number refers to sample partitioning, while for the PAM and EM algorithms the numbers refers to gene clustering. In the EM clustering validation only the external indexes were computed.

TABLE I
THE NUMBER OF CLUSTERS OBTAINED WITH THE CLUSTERING VALIDATION METHODS

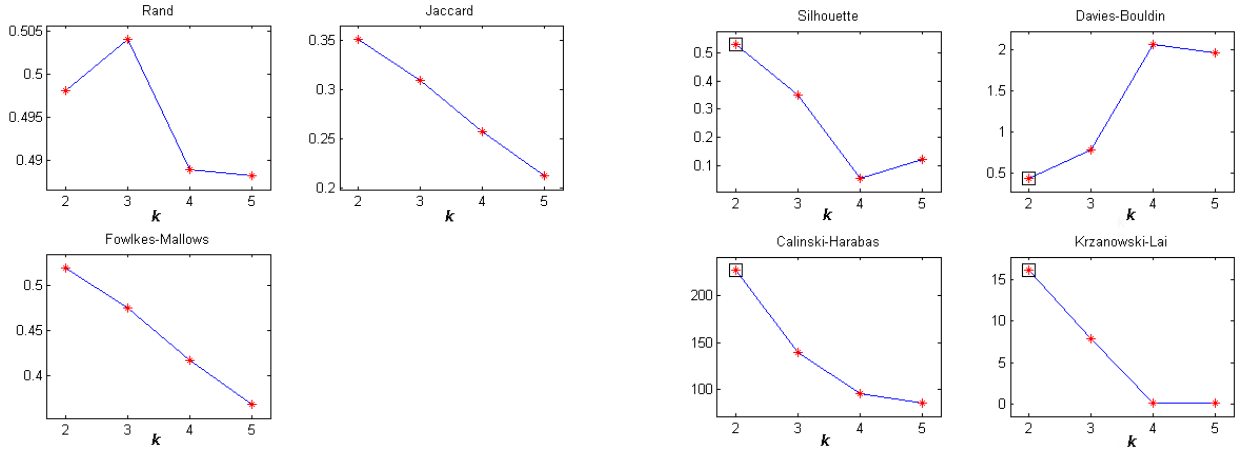
Index	Chowdary database			Leukemia database		
	k-means	PAM	EM	k-means	PAM	EM
Rand	2	3	3	3	3	4
Jaccard	2	2	2	3	3	3
Fowlkes-Mallows	2	2	2	3	3	3
Silhouette	2	2	-	2	3	-
Calinski-Harabasz	2	2	-	3	3	-
Krzanowski-Lai	2	2	-	3	3	-
Davies-Bouldin	2	2	-	2	3	-

After the validation was done we applied the combined clustering for the datasets. Thereby, for the Chowdary dataset the genes were clustered into 2 groups, with the PAM method, Fig 2.a, and then with the EM algorithm, Fig.2.b. The samples were clustered with the k-means algorithm.

The obtained values were compared with the given values from the microarray databases and a similarity between these values was observed.

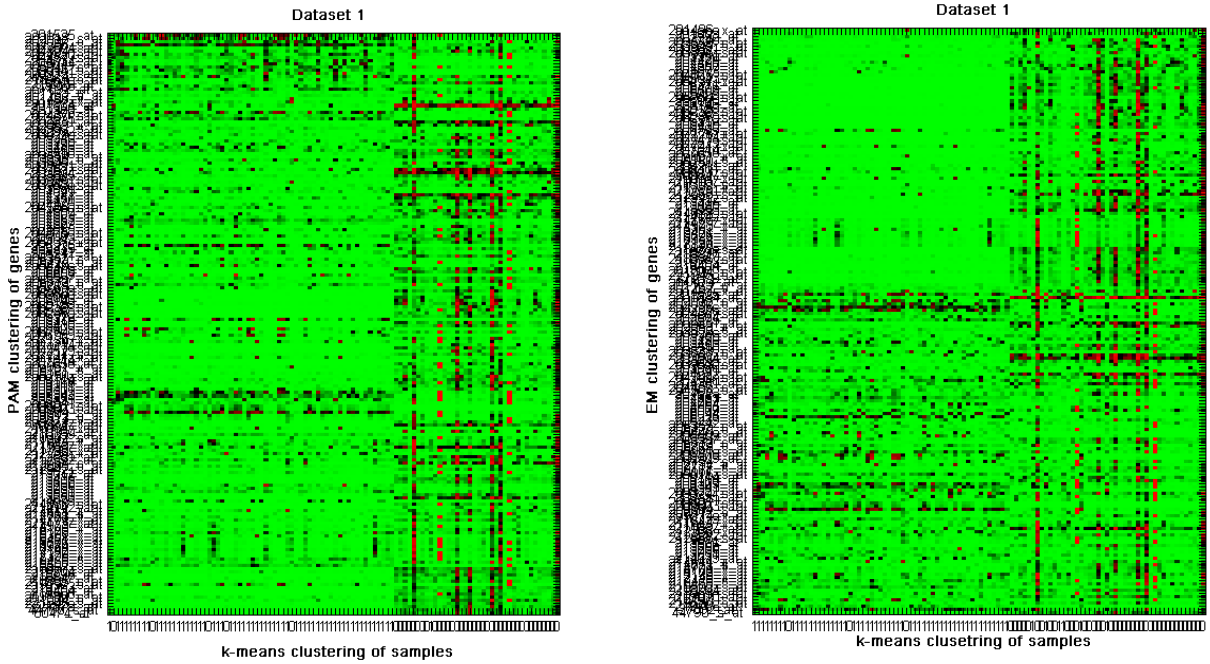
Fig. 1.a and 1.b shows all the computed indexes for the Chowdary database, internal and external, in the case of PAM algorithm. For the external indexes the highest values obtained gave the optimal number of clusters. In the case of internal indexes the optimal value was marked by a square in Fig. 1.b.

According with the optimal number of clusters indicated by the validation indexes, in the case of the leukemia dataset the clustering was done in 3 clusters. Thus the samples were group into three sets by the k-means algorithm, while the genes formed also three groups once with the PAM method



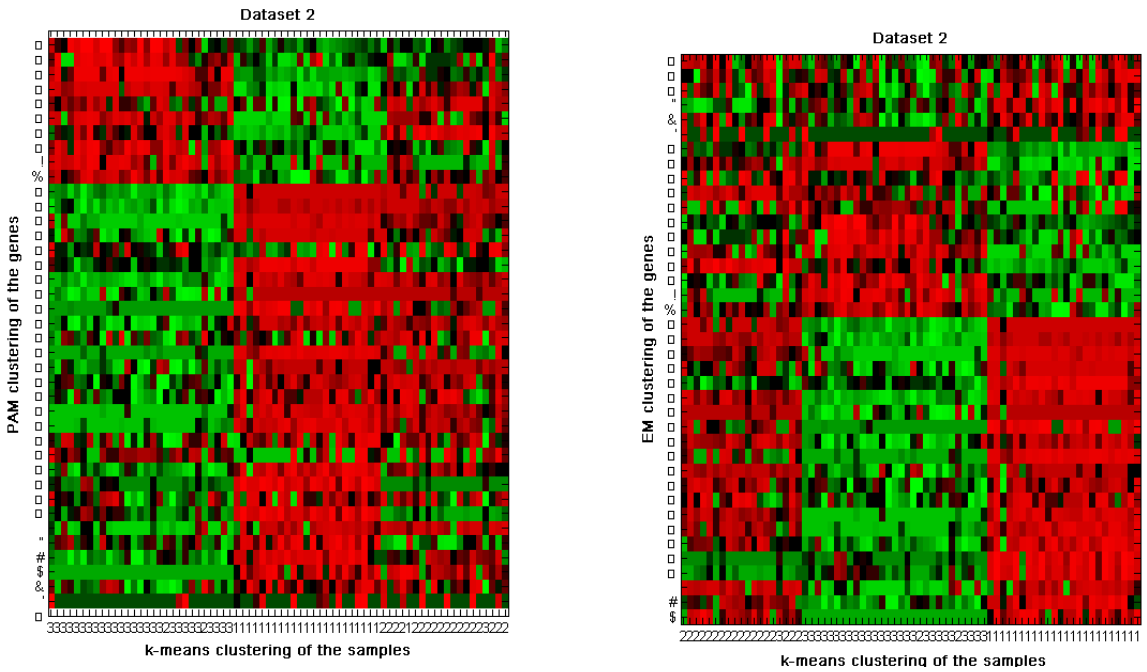
(a) (b)

Fig. 1. Values of the (a) external indexes, (b) internal indexes determined for PAM clustering of genes for Chowdary database



(a) (b)

Fig. 2. Microarray data clustering by the unsupervised algorithms: (a) PAM clustering, (b) EM clustering of the genes for the Chowdary dataset.



(a) (b)

Fig. 3.a Microarray data clustering by the unsupervised algorithms: (a) PAM clustering, (b) EM clustering of the genes for the leukemia dataset.

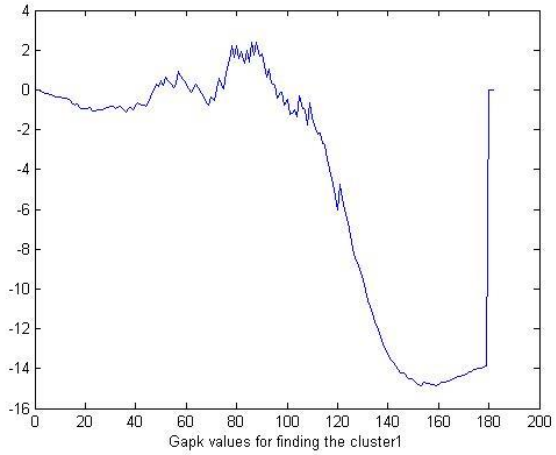


Fig. 4 The values of the Gap(k) function for the Chowdary dataset. The maximum was obtained for a number of 89 genes.

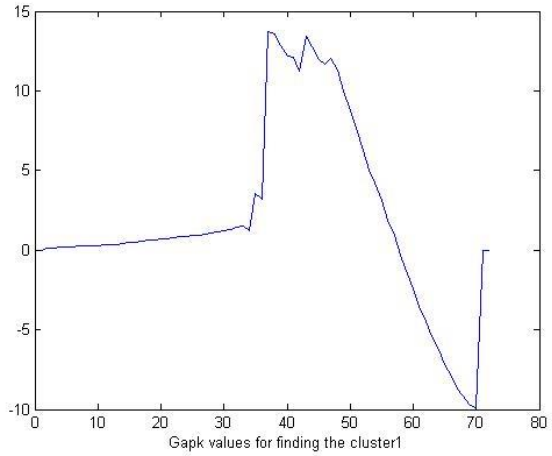


Fig. 5 The values of the Gap(k) function for the leukemia dataset. For the first cluster the maximum was obtained for a number of 37 genes

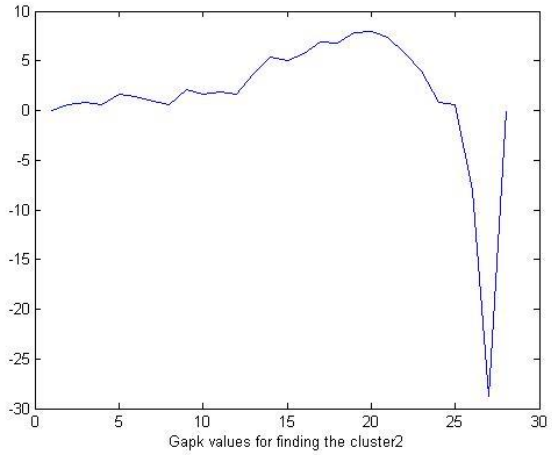


Fig. 6 The values of the Gap(k) function for the leukemia dataset. For the second cluster the maximum was obtained for a number of 21 genes

and then with the EM algorithm, as it can be seen in Fig. 3.a and Fig. 3.b.

For the gene shaving algorithm we used the same number of clusters as the one given by the validation methods. In the case of Chowdary dataset the genes were classified in 2 sets, while in the case of leukemia dataset we applied gene shaving with 3 clusters.

Once applying the gene shaving method we were able to classify the Chowdary dataset in 2 clusters. The number of gene in the first cluster was determined by using the gap statistic method. The value of 89 genes was given by the maximum of the graph from Fig. 4.

In a similar manner we computed the gap statistics for the leukemia dataset and we were able to classify the data into 3 clusters: the first one with 37 genes, the second one with 21 genes and the third one with the remaining 14 genes from a total of 72 genes. The values of the Gap function for this dataset are shown in Fig 5 and 6.

IV. Conclusion

In this paper some unsupervised and supervised clustering methods were applied to microarray dataset in order to obtain a bi-clustering of the data with different algorithms. The selection of the PAM algorithm for clustering the genes, despite of the k-means method was done because of the robustness of the first one. The EM algorithm showed some disadvantages when working with large datasets that is way the data was previously filtered.

Gene shaving, a supervised method was also combined with k-means in order to classify the genes according with the number of clusters given by the validation methods.

For cluster validation we used both internal and external methods by computing some indexes which gave us the optimal number of clusters for each clustering method. As future work we take into consideration the combination of these methods with other data mining algorithms like Independent Component Analysis, which can be used for large datasets.

References

[1] S. González, L. Guerra , V. Robles, JM. Peña, F. Famili, "CliDaPa: A new approach to combining clinical data with DNA microarrays" *Intelligent Data Analysis Journal*, vol 14(2), pp. 207 – 223, 2010

[2] J. Han, M. Kamber, *Data mining: Concepts and techniques*. Morgan Kaufmann, 2000

[3] R. Malutan, B. Belean, P.G. Vilda, M. Borda, "Two way clustering of microarray data using a hybrid approach" in *Proc. of 34th Int. Conf. on Telecommunications and Signal Processing*, Budapest, 2011, pp. 417 - 420

[4] G. McLachlan, K-A. Do, C. Ambroise, *Analyzing Microarray Gene Expression Data*. Wiley-Interscience, 2004

[5] M.B. Zoubi, A. Hudaib, A. Huneiti, B. Hammo, "New efficient strategy to accelerate k-means clustering algorithm" *American Journal of Applied Sciences*, vol 5(9), pp. 1247 – 1250, 2008

[6] G. McLachlan, T. Krishnan, *The EM Algorithm and Extensions*. John Willey & Sons 2008

[7] N. Bolshakova, F. Azuaje, "Cluster validation techniques for genome expression data" *Signal Processing* vol 83, pp. 825 – 833, 2002

[8] K. Wang, B. Wang, L. Peng, "CVAP: Validation for Cluster Analyses" *Data Science Journal* vol 8, pp. 88 – 93, 2009

[9] T. Hastie *et. al.,* "Gene shaving as a method for identifying distinct sets of genes with similar expression patterns", *Genome Biology*, vol. I(2), research 0003, pp. 1-21, 2000

[10] W. L. Martinez, A. R. Martinez, *Exploratory Data Analysis with MATLAB*, CRC Press LLC, 2005

[11] D. Chowdary *et al*, "Prognostic gene expression signatures can be measured in tissues collected in RNAlater preservative" *J Mol Diagn* vol 8(1), pp. 31 – 39, 2006

[12] S. Armstrong *et al,* "MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia" *Nature Genetics* vol 30, pp. 41 – 47, 2001

[13] S. C. Madeira, A. L. Oliveira, "Biclustering algorithms for biological data analysis: A survey" *IEEE/ACM Transactions on Computational Biology and Bioinformatics* vol 1 (1), pp. 24-45, 2004

Neural Networks Based Physical Cell Identity Assignment for Self Organized 3GPP Long Term Evolution

M. Basit Shahab and Abdul Aziz Bhatti

***Abstract*— This paper proposes neural networks based graph coloring technique to assign Physical Cell Identities throughout the self-organized 3GPP Long Term Evolution Networks. PCIs are allocated such that no two cells in the vicinity of each other or with a common neighbor get the same identity. Efficiency of proposed methodology resides in the fact that minimum number of identities is utilized in the network wise assignment. Simulations are performed on a very large scale network, where initially all the cells are without any PCIs assigned. Results of simulations are demonstrated to analyze the performance of the proposed technique. Discussions about the presence of femto cells and PCI assignment in them are also presented at the end.**

***Keywords*— Collision, Confusion, Graph Coloring, Long Term Evolution, Neural Networks, Physical Cell Identity (PCI)**

I. INTRODUCTION

With an increase in the number of mobile users utilizing cellular networks and thus the wireless technology, demands on systems and their management have increased up to a very high level. Auto configuration, optimization and healing of radio parameters using self-organized networks [1], is a key feature in meeting the requirements of the future mobile cellular networks. Therefore the New Generation Mobile Networks Forum [2] has listed some key parameters that should be handled by adaptive algorithms. 3GPP Long Term Evolution, referred to as LTE and marketed as 4G LTE, is a standard for wireless communication for mobile phones and data terminals. It aims at high data rates [3][4], low delay/latency, high spectral efficiency, high performance broadcast services, cost effective network design, automatic neighbor relation, coverage and capacity optimization, energy saving, interference reduction, physical cell identity automatic configuration, mobility load balancing optimization and random access channel optimization. These goals result in reducing capital expenditures (CAPEX) and operational expenditures (OPEX), by replacing most of the manual work with adaptive techniques that auto tune the network according to the environment changes. One of these parameters is the PCI.

M. Basit Shahab is a Lecturer at Electrical Engineering Department, University of Management and Technology, Lahore, Pakistan (Phone: +92-333-9299532; Email: basit.shahab@umt.edu.pk).

Abdul Aziz Bhatti (PhD) is the Dean of School of Science and Technology, University of Management and Technology, Lahore, Pakistan (Phone: +92-305-4440629; Email: drabhatti@umt.edu.pk).

II. PREVIOUS WORK

In LTE, each base station named eNodeB broadcasts a distinguishing signal, a fingerprint that helps the user equipment to differentiate between the serving and the neighboring cells [5]. In some circumstances, a conflict occurs among the PCIs of neighboring cells, which results in handover failures when the user equipment being mobile, moves across the boundaries of these regions. This is due to the fact that mobile units can't differentiate between two same broadcasted PCI's, irrespective of their signal strengths. A solution to this problem needs to be explored based on Neighbor cell relation lists [6][7], in order to improve the efficiency of future networks, in the sense that they will have a lesser number of call drops, better quality, improved coverage and efficient mobility handling.

Various types of cases in the PCI assignment issue have been explored [8], according to which it is quite clear that two types of conflicts can occur while assigning the PCIs in a cellular network; when two cells in the vicinity of each other or two cells having a common neighbor are assigned the same PCI. This causes collision in the former, and confusion in the later case. The situations are depicted in Figure 1.

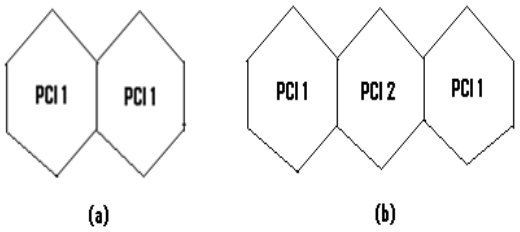


Figure 1: (a) Collision, (b) Confusion

In the first case; Figure 1 (a), two adjacent cells have the same PCI. When user equipment in either of these two cells moves across the cell boundary towards the adjacent cell with same PCI, a collision is detected and may result in handover failure. In the second case Figure 1 (b), two outer cells having the central cell as a common neighbour have same PCI, resulting in a confusion kind of conflict, which causes handover failure during mobility from either one of them towards the central cell. In order to improvise mobility handling quality throughout the cellular network, these conflicts must be resolved.

III. RELATED WORK

Numerous techniques and algorithms have been proposed to solve these conflicts. Graph coloring is considered as a better approach. In this technique, each cell is considered as a vertex of a graph, and all the cells which are direct neighbours to each other, represented by these vertices are connected through an edge. Trick towards solution is to assign colours to the vertices in such a way, that both cases are properly handled. Solution to Figure 1(a), and 1(b), has been demonstrated in Figure 2(a) and Figure 2(b).

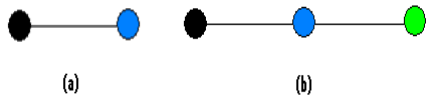


Figure 2: (a) Collision free assignment, (b) Confusion free assignment

So far, various graph coloring techniques and methods of graph coloring optimization have been proposed [9] [10] [11]. Graph coloring is used in many applications like VLSI Channel Routing [12], Subcarrier allocation in Cognitive Radio [13], and has been implemented using neural networks [14]. Any graph coloring technique can be applied to solve the PCI assignment task, but the major goal is to resolve the conflicts using minimum number of colours/PCIs. One method, dealing with a four-color-map problem is explored [15], followed by graph coloring using neural networks [16]. A neural network parallel algorithm for channel assignment problems in cellular radio networks is a step that utilizes neural networks in the telecommunication area [17].

IV. MATHEMATICAL MODELLING

Consider a case of "n" cells in a cellular network and "m" available PCIs or colours to be assigned to these cells with no collision or confusion conflicts. The very famous Hopfield neuron model is considered as shown in Figure 3. Two types of amplifiers are involved in the circuit. An ordinary amplifier and its inverting version are present. Impulse is the input applied to neurons. Feedback paths are also shown. Output is passed through the sigmoid function to determine whether it is 0 or 1.

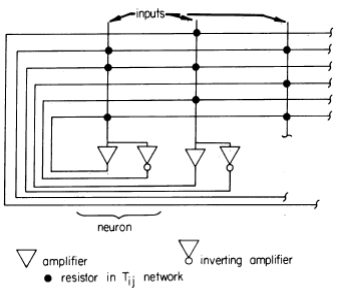


Figure 3: Hopfield neuron model

Three restrictions are applied on the Hopfield network.

1) Inputs and outputs are binary (exclusively ones or zeros). Considering this, it should be quite clear that the nodes produce only binary results.
2) There should be no weights.
3) A condition on the activation function applied such that it is always a unit step.

This model is also applied in artificial neural networks for four-coloring problems and k colorability [18], an NP complete problem [19], in which colors are assigned to different regions in a map, so that no two regions having a common boundary should get the same color. In four-coloring problem and k colorability, after all the mathematical work involved, final energy equation for an i^{th} neuron in any X^{th} region is

$$E = \frac{A}{2}\sum_{X=1}^{n}\left[\sum_{i=1}^{m} V_{xi} - 1\right]^2 + B\sum_{X=1}^{n}\sum_{Y=1}^{n}\sum_{i=1}^{m} D_{xy}\ V_{xi}\ V_{yi} \quad (1)$$

Where

The first term makes sure that only one neuron's output is 1 in each region. This is done due to the fact that each neurons output corresponds to a particular color. If multiple neurons' outputs are 1 in a particular region, then more than one color will be allocated to this region, which is to be avoided. The second term makes sure that no two adjacent/ neighboring cells get the same color. For any two regions X and Y, the adjacency matrix Dxy describes the adjacency between these two regions. If X and Y are direct neighbors to each other, then the value corresponding to their relationship will be 1 in the matrix Dxy. Thus for this case, if the values of Vxi and Vyi are same which means that both neurons have the same output, there will be an energy penalty equal to the value of the constant B. As this equation always tries to converge to the 0 energy point, therefore, such condition will be avoided. In short, this second term makes sure that the color allocation is collision free.

This is followed by motion equation, which is actually the partial derivative of (1) with respect to neuron's output.

$$\frac{dU_{xi}}{dt} = -A\left[\sum_{i=1}^{m} V_{xi} - 1\right] - B\sum_{Y=1}^{n} D_{xy}\ V_{yi}\sum_{k=1}^{n} D_{yk} + C.h\left(\sum_{i=1}^{m} V_{xi}\right)\left(C1\sum_{k=1}^{n} D_{xk} + C2\frac{\sum_{k=1}^{n}\sum_{Y=1}^{n} D_{xy}\ D_{yk}}{\sum_{k=1}^{n} D_{xk}}\right) \quad (2)$$

Colors to the map are allocated by applying sigmoid function on the final motion equation.

$$Sig\left(\begin{array}{c} -A\left[\sum_{i=1}^{m} V_{xi} - 1\right] - B\sum_{Y=1}^{n} D_{xy}V_{yi}\sum_{k=1}^{n} D_{yk} \\ + C.h\left(\sum_{i=1}^{m} V_{xi}\right) \\ \left(C1\sum_{k=1}^{n} D_{xk} + C2\frac{\sum_{k=1}^{n}\sum_{Y=1}^{n} D_{xy}\ D_{yk}}{\sum_{k=1}^{n} D_{xk}}\right) \end{array}\right) \quad (3)$$

This equation is written for the map coloring case, but it can also be used in solving collision free PCI assignment problem. Second case that remains unsolved is the confusion free assignment. This paper extends Yoshiyasu Takefuji's neural networks k-colorability map coloring approach in the form of a single equation for both collision free and the unresolved confusion free PCI assignment. One important factor that needs to be taken care of simultaneously is the upper and lower bounds on chromatic number, which in the other sense means an assignment with minimum number of PCIs utilized.

The proposed technique starts with writing an energy equation like (3) for some ith PCI in an Xth cell. The first thing to be kept in mind is that only one neuron's output should be 1 in any cell, otherwise more than one colors/PCIs will be assigned to a single cell. For that reason, one term in the energy equation, same as the term proposed for collision free assignment, should be present, to deal with the issue. This has been done in (4).

$$E1 = \frac{A}{2}\sum_{X=1}^{n}[\ \sum_{i=1}^{m} V_{xi} - 1]^2, \qquad (4)$$

where A is a constant, used to represent an energy penalty. V_{xi} is the output of ith neuron in Xth cell. Second term needs to deal with the case of collision and confusion free assignment, as depicted in (5).

$$\mathrm{E2} = B\sum_{X=1}^{n}\left[\sum_{Y=1}^{n}\sum_{Z=1}^{n}\sum_{i=1}^{m} D_{xy}\ D_{xz}\ V_{xi}V_{yi}\ V_{zi}\right], \qquad (5)$$

where B is a constant that will take the energy away from converging, if a boundary violation takes place. D_{xy} and D_{xz} are elements of the adjacency matrices between X, Y, and X, Z. $V_{xi,}$ V_{yi} and V_{zi} are outputs of the ith neuron in Xth, Yth and Zth cell respectively.

D_{ij} = {1 if any region i and j are adjacent
{0 otherwise

(4) And (5) together result in (6)

$$E = \frac{A}{2}\sum_{X=1}^{n}[\ \sum_{i=1}^{m} V_{xi} - 1]^2 + B\sum_{X=1}^{n}\left[\sum_{Y=1}^{n}\sum_{Z=1}^{n}\sum_{i=1}^{m} D_{xy}\ D_{xz}\ V_{xi}\ V_{yi}\ V_{zi}\right] \qquad (6)$$

Motion equation for an ith PCI in any Xth cell is the partial derivative of Energy equation with respect to neuron's output

$$\frac{dU_{xi}}{dt} = -\frac{\partial \mathrm{E}}{\partial V_{xi}} \qquad (7)$$

Because interest lies in the output of some ith neuron in any Xth Cell, applying partial derivative on the energy equation with respect to V_{xi}, results in equation (8).

$$\frac{dU_{xi}}{dt} = -\frac{\partial}{\partial V_{xi}}\left[\frac{A}{2}\sum_{X=1}^{n}[\ \sum_{i=1}^{m} V_{xi} - 1]^2\ \right] - \frac{\partial}{\partial V_{xi}} B\sum_{X=1}^{n}\left[\sum_{Y=1}^{n}\sum_{Z=1}^{n}\sum_{i=1}^{m} D_{xy}D_{xz}\ V_{xi}\ V_{yi}V_{zi}\right] \qquad (8)$$

Expanding the 1st term i-e E1 with respect to X and taking the partial derivative for some specific value, gives

$$\frac{\partial E1}{\partial V_{xi}} = A[\sum_{i=1}^{m} V_{xi} - 1] \qquad (9)$$

Writing (5) in terms of X and i, gives

$$\begin{aligned}\frac{\partial E2}{\partial V_{xi}} = B\frac{\partial}{\partial V_{xi}}\Big[&\Big[\left[\sum_{Y=1}^{n}\sum_{Z=1}^{n} D_{1y}\ D_{1z}\ V_{11}\ V_{y1}\ V_{z1}\right] + \\ &\left[\sum_{Y=1}^{n}\sum_{Z=1}^{n} D_{1y}\ D_{1z}\ V_{12}\ V_{y2}\ V_{z2}\right] + \\ &\ldots\left[\sum_{Y=1}^{n}\sum_{Z=1}^{n} D_{1y}\ D_{1z}\ V_{1m}\ V_{ym}\ V_{zm}\right]\Big] + \\ &\Big[\left[\sum_{Y=1}^{n}\sum_{Z=1}^{n} D_{2y}\ D_{2z}\ V_{21}\ V_{y1}\ V_{z1}\right] + \\ &\left[\sum_{Y=1}^{n}\sum_{Z=1}^{n} D_{2y}\ D_{2z}\ V_{22}\ V_{y2}\ V_{z2}\right] + \\ &\ldots\left[\sum_{Y=1}^{n}\sum_{Z=1}^{n} D_{2y}\ D_{2z}\ V_{2m}\ V_{ym}\ V_{zm}\right]\Big] + \\ &\Big[\left[\sum_{Y=1}^{n}\sum_{Z=1}^{n} D_{3y}\ D_{3z}\ V_{31}\ V_{y1}\ V_{z1}\right] + \\ &\left[\sum_{Y=1}^{n}\sum_{Z=1}^{n} D_{3y}\ D_{3z}\ V_{32}\ V_{y2}\ V_{z2}\right] + \\ &\ldots\left[\sum_{Y=1}^{n}\sum_{Z=1}^{n} D_{3y}\ D_{3z}\ V_{3m}\ V_{ym}\ V_{zm}\right]\Big] + \\ &\ldots\Big[\left[\sum_{Y=1}^{n}\sum_{Z=1}^{n} D_{ny}\ D_{nz}\ V_{n1}\ V_{y1}\ V_{z1}\right] + \\ &\left[\sum_{Y=1}^{n}\sum_{Z=1}^{n} D_{ny}\ D_{nz}\ V_{n2}\ V_{y2}\ V_{z2}\right] + \\ &\ldots\left[\sum_{Y=1}^{n}\sum_{Z=1}^{n} D_{ny}\ D_{nz}\ V_{nm}\ V_{ym}\ V_{zm}\right]\Big]\Big] \end{aligned} \qquad (10)$$

Partial derivative is applied for some specific X and i, but in general, (10) becomes

$$\frac{\partial E2}{\partial V_{xi}} = B\left[\sum_{Y=1}^{n}\sum_{Z=1}^{n} D_{xy}\ D_{xz}\ V_{yi}\ V_{zi}\right] \qquad (11)$$

Combining (9) and (11), gives the motion equation

$$\frac{\partial E}{\partial V_{xi}} = A[\sum_{i=1}^{m} V_{xi} - 1] + B\left[\sum_{Y=1}^{n}\sum_{Z=1}^{n} D_{xy}\ D_{xz}V_{yi}V_{zi}\right] \qquad (12)$$

But $\frac{dU_{xi}}{dt} = -\frac{\partial E}{\partial V_{xi}}$, so

$$\frac{dUxi}{dt} = -A[\sum_{i=1}^{m} V_{xi} - 1] - B\left[\sum_{Y=1}^{n}\sum_{Z=1}^{n} D_{xy}D_{xz}V_{yi}V_{zi}\right] \qquad (13)$$

Because this is an NP complete problem, there are many solutions to it. In the line of finding the best one, a hill climbing term needs to be included in the motion equation. Once the hill climbing term has been added, sigmoid function is applied on it. This has been done in (14).

$$sig\big(-P1 - P2 + (P3 \text{ x } P4)\big) \qquad (14)$$

Where

P1= $A[\sum_{i=1}^{m} V_{xi} - 1]$

P2= $B\left[\sum_{Y=1}^{n}\sum_{Z=1}^{n} D_{xy}\ D_{xz}\ V_{yi}\ V_{zi}\sum_{k=1}^{n} D_{yk}\ D_{zk}\right]$

P3= $C3.h(\sum_{i=1}^{m} V_{xi})$

P4= $\left(C1\sum_{k=1}^{n} D_{xk}\ C2\frac{\sum_{k=1}^{n}\sum_{Y=1}^{n}\sum_{Z=1}^{n} D_{xy}\ D_{xz}\ D_{yk}\ D_{zk}}{\sum_{k=1}^{n} D_{xk}}\right)$

C1, C2 and C3 are constants, used for normalization. $\mathbf{h(\sum_{i=1}^{m} V_{xi})}$ Is the hill climbing function, which together with normalization constants in the motion equation, forces the state of the system to escape from local minimum and converge to global minimum.

H(x) = {1 if x=0
{0 otherwise

Hill climbing term performs the excitatory force only when all of V_{xi}'s are 0 or in other words when PCIs are assigned to all the cells. This means that the algorithm repeats itself unless the best result is achieved.

V. FEMTOCELLS AND PCI ASSIGNMENT

The position and number of femtocells are unknown by the operator, and because of the fact that they can be switched on/off or moved at any time by the users; configuration and optimization of a femtocell network cannot be based on classic network design. Secondly, femtocells must be plug-and-play devices with a significant degree of self-configuration because of non-technical expertise of the femtocells customers. This increases the demand for efficient self-organization techniques for successfully deploying and managing a large femtocell layer over the existing macro cell network. In this way, femtocells are able to react to the changing conditions of the network, traffic and channel, and mitigate cross- and co-layer interference [20]. A lot of studies have been made on femtocells [21] and their efficiency and deployment for providing optimized coverage [22].

Numerous works have been done in providing algorithms and techniques for confusion free PCI assignment in heterogeneous networks. One related work is the extended synchronization signals for eliminating PCI confusions [23]. Distributed PCI assignment in LTE based on consultation mechanism [24], graph coloring based physical cell identity assignment for LTE networks [25] and physical cell identity self-organization for home eNodeBs deployment in LTE [26] are some more famous researches in this area.

The proposed neural networks based technique for PCI assignment can be modified to accommodate the femtocells present inside a macro cell layer. Two basic things need to be taken care of, and the femtocells will be easily configured with PCIs without any confusion:

1. A check for the macro cell's PCI in which the femtocell is present, and then assignment of an extended PCI amongst the extended PCIs available for femtocells.
2. Femtocells within a macro cell needs to be checked for collision or confusion with the femtocells in a particular macro cell, and also with the femtocells of the neighboring macro cell.

More research work is under process to accommodate the femtocells with PCIs using the same but a bit modified neural networks based PCI assignment equations. The new equation needs to have another term which places a condition to check whether the cell considered as a candidate for PCI assignment is a macro or a micro/femto cell, and then assign PCIs to it from the list of available PCIs. The conditions or the new form of equation will simply implement a nested if-else condition (in programming sense), and the task will be achieved.

VI. SIMULATIONS AND RESULTS

In order to check the performance of proposed technique, a cellular network is considered for assigning Physical Cell Identities. A small part of that network is shown in Figure 4.

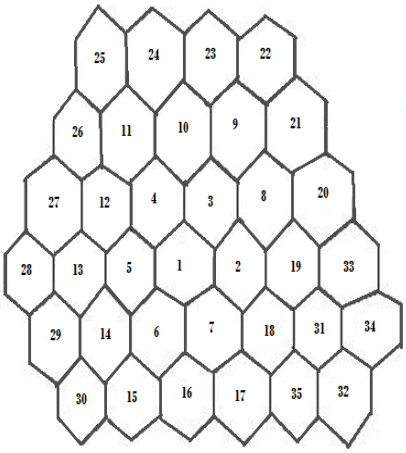


Figure 4: Cellular network model for simulations

Initial simulations are performed for the case of collision free assignment in accordance with equation (3). Results of these simulations describe how the map coloring technique can be used to perform collision free PCI assignment. After that simulations are performed for the proposed technique which accomplishes both the tasks of collision free and confusion free assignment utilizing equation (14). Simulations are performed in following steps.

1. The first step starts with setting values for constants and variables. A, and B being penalty constants are assigned values such that both the terms are optimized. Penalty constants are selected in such a way to vary the impact of constraints for achieving stable results. Hopfield did not propose any proper selection criteria, however, since then multiple strategies have been proposed, among which the one that got great appreciation is Convergence properties of a modified Hopfield-Tank model [27]. This whole strategy was then used to find shortest path in Travelling Salesman problem. The term, which is subjected to more minimization, is multiplied with larger constant and vice versa. Along with penalty constants, 'm' and 'n' are also initialized.
2. Second step is to define an adjacency matrix for the cellular system. The adjacency matrix considered for any cell X with all the other cells is given by D_{xy}. In actual case, it is present only in the form of a list in the eNodeBs, much easier to manipulate for use in the PCI assignment algorithms.
3. After that, an input vector U_{xi} for all the neurons is defined. U_{xi}'s are generated randomly within the range given below initially. It is simply a row vector of order (1 x m).

$$-0.002 \leq U_{xi} \leq 0.002$$

4. Following that is the initialization of the output matrices, and then the implementation of motion equation.

At first, simulations are performed for the map coloring equation (3) based methodology, utilizing it to perform the collision free assignment. A, and B are assigned values 200 and 250 respectively, whereas C=C1=C2=1. X, Y and 'n' are equal to 35 because they all represent the number of cells. Number of PCIs to be used is calculated using upper and lower bounds on the chromatic number which actually is the minimum number of colors needed to color the graph. 3 PCIs/colors are used to perform proper assignment, thus 'm' = 3. Adjacency matrix D_{xy} of order (35 * 35) is defined. After that the motion equation is implemented. All the summations are implemented using loops in the programming. Output of simulations is a matrix V_{xi} of order (3 * 35) as shown below.

$$V_{xi} = \begin{bmatrix} 1\,0\,0\,0\,0\,0\,0\,1\,0\,1\,0\,1\,0\,1\,0\,1\,0\,0\,0\,1\,0\,0\,1\,0\,0\,1\,0\,0\,0\,1\,1\,0\,0 \\ 0\,1\,0\,1\,0\,1\,0\,0\,1\,0\,0\,0\,1\,0\,0\,0\,1\,0\,0\,1\,0\,0\,0\,1\,0\,1\,0\,0\,0\,1\,1\,0\,0\,0\,0 \\ 0\,0\,1\,0\,1\,0\,1\,0\,0\,0\,1\,0\,0\,0\,1\,0\,0\,0\,1\,0\,1\,0\,1\,0\,0\,0\,1\,0\,1\,0\,0\,0\,0\,1\,1 \end{bmatrix}$$

This matrix is the output of an i^{th} neuron in any X^{th} cell. Output array is always an (m x n) matrix, with "m" being the number of colors/PCIs used, and "n" representing the number of cells to which the colors/PCIs are to be assigned. This means the element $V_{3,\ 13}$ of resultant matrix is output of 3^{rd} neuron in the 13^{th} cell. If 3^{rd} neuron represents PCI 3, then this PCI is assigned to 13^{th} cell if the value of $V_{3,\ 13}$ in output matrix is 1. On the basis of values in the matrix Vxi, PCIs have been shown in the corresponding cells in Figure 5, where each number represents a particular PCI and a specific color.

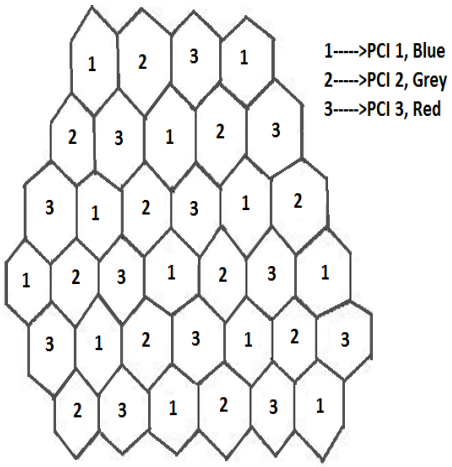


Figure 5: Collision Free PCI Assignment

PCI assignment in figure 5 can easily be analyzed to check whether the map coloring technique is sufficient to provide a good collision free allocation of PCIs or not. It is clearly demonstrated by the figure that no two direct neighbor cells are having the same PCI, thus comfortably achieving the required goal. The numbers 1, 2 and 3 actually represent the PCI number, or a specific color in the sense of graph coloring. If 1 stands for PCI 1 or color Blue, then this PCI/color will be assigned to all the cells having number 1 mentioned on them.

The same set of steps is followed for the simulations of proposed technique. A, B, C, C1, C2, and n are assigned the same values as for the map coloring simulations, but m=9. Because the proposed technique performs both collision and confusion free, another adjacency matrix D_{xz} of the order (35*35) is defined along with D_{xy}. This is done because the PCIs of two cells Y and Z which are direct neighbors to X are to be checked for confusion. Result of simulations is a matrix of order (9 x 35). This matrix is also named $V_{xi.}$

$$\begin{bmatrix} 1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,1\,0\,0\,0\,1\,1\,0\,0\,0\,0 \\ 0\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,1\,0\,0\,0\,0\,0\,0\,1\,0 \\ 0\,0\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0\,0\,1\,0\,0\,0 \\ 0\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,1\,0\,0\,1\,0\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0 \\ 0\,0\,0\,0\,1\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0\,1\,0\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1 \\ 0\,0\,0\,0\,0\,0\,1\,0\,1\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0 \\ 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,0\,1\,0\,0\,1\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0 \end{bmatrix}$$

This matrix like the first matrix is the output of an i^{th} neuron in any X^{th} cell. Number of rows is 9 because 9 colors are used to color the considered network with collision and confusion free constraints. Each row contains the output of a particular neuron in all the cells. On the basis of values in the matrix V_{xi}, PCIs have been shown in the corresponding cells in Figure 6, where each number like the previous case represents a particular PCI and a specific color. If PCI allocation according to the output is considered, only one neuron's output is 1 in each cell, so that no cell has more than one color or PCI assigned to it. This is done by the 1^{st} term of equation (14). The 2^{nd} term results in a collision and confusion free assignment. Moreover 3^{rd} and 4^{th} terms extract the best output out of all the possible solutions. If simulations are performed without the hill climbing function, there can be collision or confusion conflicts at some locations, but still they will be negligible, compared to the number of cells considered for assignment. Also it can result in large number of PCIs/colors utilization for accomplishing the same task that has been done with the minimum number used.

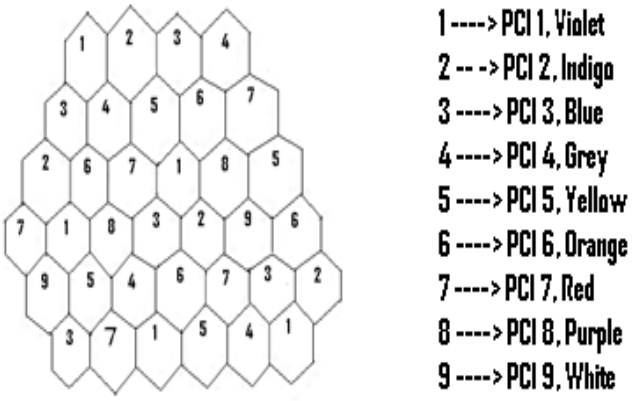


Figure 6: Collision Free PCI Assignment

State of the system always converges to global minimum, supported by almost 750 simulation runs. Figure 7 depicts a

plot between frequency and number of iteration steps for the proposed technique. Horizontal axis represents the number of iterations required to achieve a satisfactory output, and vertical axis represents frequency; repetition of that particular value of iterations. Reason behind these recursions/iterations is that energy/motion equations are optimization functions, in which different required terms are to be optimized with the help of energy penalty constants. The final output of neurons is the result of these iterations that take place updating values after each iterative step. Optimization is done with the help of constants which offer a penalty increase in energy, if something goes wrong. A lot of recursions take place until these optimization functions converge to the lowest energy value and become stable to give a satisfactory output.

For PCI assignment in each cell, a number of recursions are required to finally find the best assignment, which can be same or different compared to the recursions required for assignment in any other cell. Figure 7 actually depicts the reuse factor of a specific number of iterations during whole network wise assignment for the proposed technique. Average number of iteration steps can be calculated from the graph using any mathematical technique. In this case, average number of iterations is around 980 for about 200 cells.

Simulations were also performed for equation (3); collision free assignment case, for which the average number of iterations were 820 for 200 cells. This difference is because of an additional factor of confusion free assignment along with the collision free case in the motion equation (14), resulting in an increase in recursive steps. If cells in the structure are increased, a corresponding increase in the average number of iteration steps occurs, but this increase is very less, compared to the increase in problem size. This means that the proposed technique is not very much dependent on the problem size.

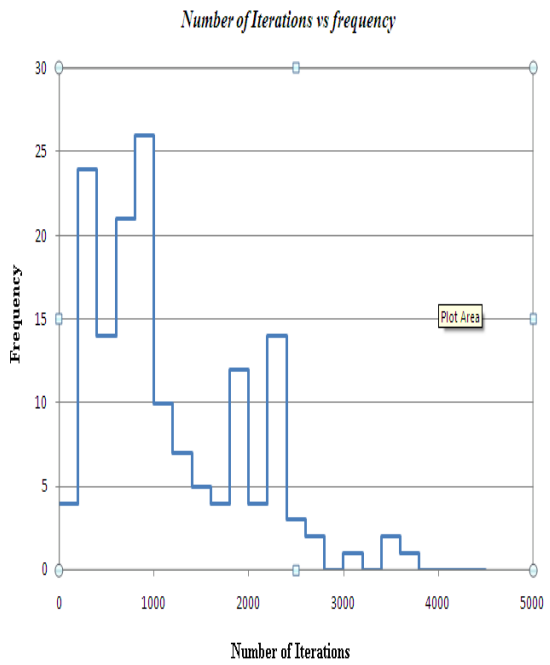


Figure 7: No of iterations and Frequency for Proposed Technique

VII. CONCLUSION

This paper proposes an algorithm that automatically assigns PCIs to a complete cellular network, during initial network wise configuration, with the least chromatic number [28]. In the same manner, a slight modification in the proposed technique can be used to assign PCIs during incremental growth of an already present network and to accommodate the closed subscriber groups. Another important scenario is the PCI assignment when a conflict takes place. Further improvements to the proposed solution can also resolve PCI conflicts, whenever they occur within a network.

Previous graph coloring techniques solve the collision type of conflict, which mean that they only deal with coloring of adjacent vertices of a graph, and do not deal with assignment of colors without confusion. But in this case both collision and confusion conflicts have been resolved. Secondly neural networks based implementation increases the processing speed because it deals with parallel processing, so the algorithm achieves a better time complexity. Moreover the hill climbing concept improves the algorithm in the sense that the best solution out of all possible ones is extracted at the end, and then it is finally implemented on the whole network.

Having developed and successfully shown the performance of the method, it is believed that the results will have a significant positive impact while dealing with such a problem.

REFERENCES

[1] Sujuan Feng, Eiko Seidel, 2008, "Self-Organizing Networks (SON) in 3GPP Long Term Evolution," Nomor Research GmbH, Munich, Germany.

[2] Hamid Akhavan et al., 2006 "Next Generation Mobile Networks - Beyond HSPA & EVDO - Whitepaper," Tech. Rep, NGMN Ltd.

[3] E Dahlman, S. Parkvall, J. Sköld, and P. Beming, 2007, "3G Evolution –HSPA and LTE for Mobile Broadband," Elsevier.

[4] 3GPP TR 25.913, 2008. "Technical Specification Group Radio Access Network; Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN)," Release 7.

[5] Mehdi Amirijoo, Pål Frenger, Fredrik Gunnarsson, Harald Kallin, Johan Moe, Kristina Zetterberg, 2008, "Neighbor Cell Relation List and Physical Cell Identity Self Organization in LTE," Wireless Access Networks,Ericsson Research, Ericsson AB, Sweden.

[6] M. et al and Amirijoo, 2008, "Neighbor cell relation list and measured cell identity management in lte," Network Operations and Management Symposium, NOMS 2008. IEEE.

[7] NGMN, 2007, "Operator Use Cases related to Self Organizing Networks," ver. 1.53.

[8] Tobias Bagndh, Georg Carle, 2009, "Graph Coloring Based Physical-Cell-ID Assignment for LTE Networks," Proceedings of the International Conference on Wireless Communications and Mobile Computing.

[9] Koljonen, J.-M, 2010, "Distributed generalized graph coloring," IEEE Conference on Self-Adaptive and Self-Organizing Systems (SASO).

[10] Mittal, A., 2011, "Graph Coloring with Minimum Colors: An Easy Approach," International Conference on Communication Systems and Network Technologies (CSNT).

[11] Feizi, S., 2010, "Cases where finding the minimum entropy coloring of a characteristic graph is a polynomial time problem," IEEE International Symposium on Information Theory Proceedings (ISIT).

[12] Shenshen Gu, 2004, "A chaotic neural network for the graph coloring problem in VLSI channel routing," International Conference on Communication, Circuits and Systems.

[13] Guobin Zhang, 2010, "Subcarrier allocation algorithms based on graph-coloring in Cognitive Radio NC-OFDM system," IEEE International Conference on Computer Science and Information Technology (ICCSIT).

[14] Di Blas, A. 2002, "Energy function-based approaches to graph coloring," IEEE Transactions on Neural Networks.

[15] Appel K and Haken, 1977, "the solution of the four-color-map problem," Scientific American, pp.108-121.

[16] Philipsen, W.J.M.; Stok, L, 1991 "graph coloring using neural networks," IEEE International Symposium on circuits and systems.

[17] Funabiki, N, 1992, "A neural network parallel algorithm for channel assignment problems in cellular radio networks," IEEE transactions on Vehicular technology.

[18] Takefuji Y and Lee K. C, 1991, "Artificial neural networks for four-coloring problems and k colorability problems," IEEE Trans, on Circuits and Systems, 38, 3, pp.326-333.

[19] Garey M.R., and Johnsan D.S., (1979), Computers and Intractability: A Guide to the Theory of NP-Completeness. (W.H. Freeman and Company).

[20] Ladányi, A. ; López-Pérez, D. ; Jie Zhang, 2012, "Self-organization for LTE enterprise femtocells," GLOBECOM Workshops 6-10 Dec, 2012.

[21] H. Claussen, L. T. W. Ho, and L. G. Samuel, "An overview of the femtocell concept," Bell Labs Technical Journal, vol. 3, no. 1, pp. 221–245, May 2008.

[22] H. Claussen, L. T. W. Ho, and L. G. Samuel, "Self-optimization coverage for femtocell deployments," in Wireless Telecommunications Symposium, ser. 24–26, California, USA, April 2008, pp. 278–285.

[23] Zahran, A.H., 2012, "Extended synchronization signals for eliminating PCI confusion in heterogeneous LTE," WCNC 1-4 April 2012.

[24] Y. Liu, W. Li, H. Zhang, and L. Yu, "Distributed PCI Assignment in LTE Based on Consultation Mechanism," in Proc. 6th Int Wireless Communications Networking and Mobile Computing WiCOM Conf, pp. 1–4, 2010.

[25] T. Bandh, G. Carle, and H. Sanneck, "Graph coloring based physical-cell- ID assignment for LTE networks," in Proc. of the 2009 Intl Conf. on Wireless Communications and Mobile Computing, IWCMC '09, pp. 116–120, New York, NY, USA: ACM, 2009.

[26] Y. Wu, H. Jiang, Y. Wu, and D. Zhang, "Physical Cell Identity Self- Organization for Home eNodeB Deployment in LTE," in Proc. 6th Int Wireless Communications Networking and Mobile Computing (WiCOM) Conf, pp. 1–6, 2010.

[27] Bizzarri AR, 1991," Convergence properties of a modified Hopfield-Tank model," Biological Cybernetics, volume 64, number 4

[28] Sen, A.; Roxborough, T.; Medidi, S, 1998, "Upper and lower bounds of a class of channel assignment problems in cellular networks," Seventeenth Annual Joint Conference of the IEEE Computer and Communications Societies, Proceedings, IEEE.

An Overview of Warehouse Optimization

Jan Karásek

***Abstract*—This paper presents an overview related to warehouse optimization problems. The problems are divided in to several groups. First, the basic technical structure of warehouse is described. Second, the standard operational and organizational framework of warehousing company is characterized, which is in special attention in this work. Third, the coordinating and controlling systems for warehouse operations is briefly mentioned, and typical warehousing operations dependent on technical and operational structure are described. The main contribution of this paper is to show the current state of the art in optimization in mentioned three groups of interest in logistic warehouses and distribution centers.**

***Keywords*—Logistics, Optimization, Warehouse.**

I. Introduction

Modern logistic warehouses and distributions centers are designed on the basis of dozens optimization studies. In consequence of that, Warehouse Management Systems (WMS) become important and more complex and users find it quite hard to manage. The software market offers large variety of solutions with different system requirements and possibilities, and to choose the suitable system for every company is not quite an easy task, because it is influenced by many aspects which must be considered, and one of this aspects are optimization methods based on automated processes for tasks dynamically changeable in time. The WMSs which drive logistic warehouses and distribution centers are core elements of the material and goods flow in logistic chain and they will be subjected to further investigation in following text related to optimization of technical and operational structure.

According to the [1] the activities of the warehousing optimization can be divided into three groups. First, the basic technical structure of warehouse. Second, the operational and organizational framework, to which a special attention is paid in this work. Third, the coordinating and controlling systems for warehouse operations. The main contribution of this paper is to show the current state-of-the-art in optimization in mentioned three groups of interest, and to help researchers with orientation in logistic warehouse optimization problems.

The rest of the paper is organized as follows. The second section gives an overview of optimization related to technical structure of logistic warehouses and distribution centers. The third section is related to optimization of operational structure. The fourth section very briefly describes the WMSs which are the core of every warehouse. The fifth section describes the typical warehousing operations which are tightly related to optimization of technical and operational structure of warehouses. The last section gives a conclusion of the paper.

Jan Karásek is with the Department of Telecommunications, Faculty of Electrical Engineering and Communication, Brno University of Technology, Technická 12, 616 00 Brno, Czech Republic, e-mail: karasekj@feec.vutbr.cz.

Manuscript received October 21, 2013; revised October 31, 2013.

II. Optimization of Technical Structure

The basic technical structure involves e.g. the layout design of the logistic warehouse or whole distribution center, the choice and dimensioning of conveyors and warehouse equipment, the design of the physical interfaces to neighboring systems and other attributes related to technical structure.

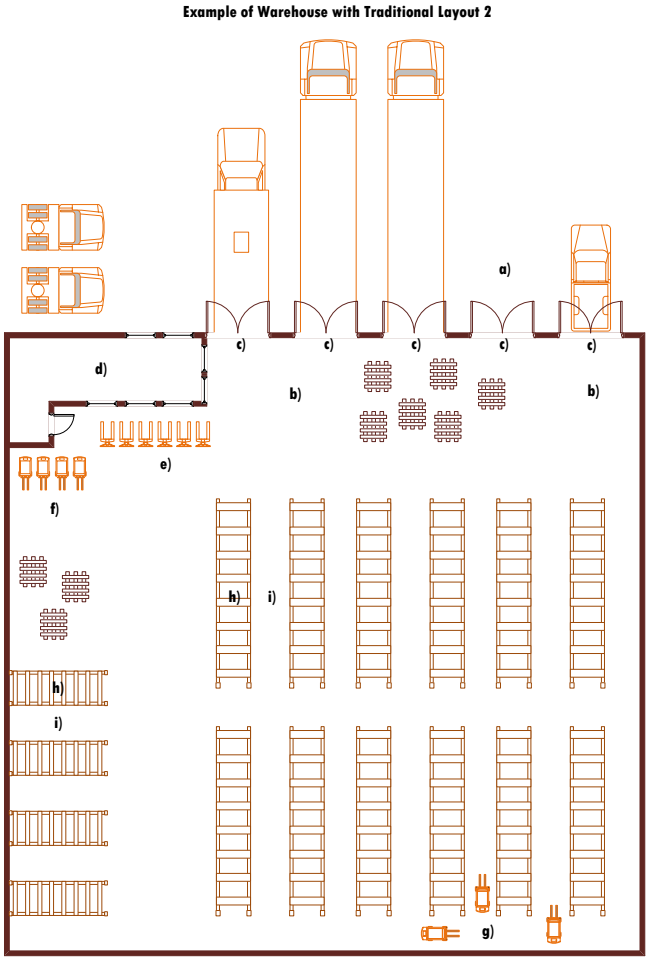


Fig. 1. The Example of Traditional Rectangular Warehouse Layout

The layout design of the warehouse [2] is a key component of further optimization tasks and has a significant impact on order-picking and traveling distances in the warehouse. In [3] it was found out that the layout design has more than 60% effect on the total travel distance, and three basic types of warehouse layout were presented. In [4] and [5], the application of parallel cross aisles in the warehouse was presented, and it was considered as a significant improvement. The layout is usually of rectangular shape and based on pallet manipulation [6], see Fig. 1. According to [7] and [8] there is

a few factors to be considered in the layout design, such as: the number of blocks; the length, the width, and the number of picking aisles; the number and shape of cross aisles if they are present; the number of rack levels; and the position of input and output gates in the warehouse. In [9] has been presented a new *Flying-V* and *Fishbone* design of cross aisles, which offers a 10% – 20% reduction of traveling distance. In [10] has been introduced an analysis of dual-commands and in [11] has been introduced adapted *Fishbone* design for dual-commands. In dual-commands environment the worker loads the goods in a pickup and deposit location and travels to a storage location and then travels to another location from which he picks goods and returns back. In [12] has been proposed a more developed *Flying-V* design of cross aisle and *Inverted-V* design. This improvement brings another 3% saving of traveling distance. The warehouse layout is also connected with the aisle design [13], [14]. The layout is mostly narrow-aisle-like, which increases the space utilization with minimal costs, but it can lead to higher operational costs and more congestions among workers.

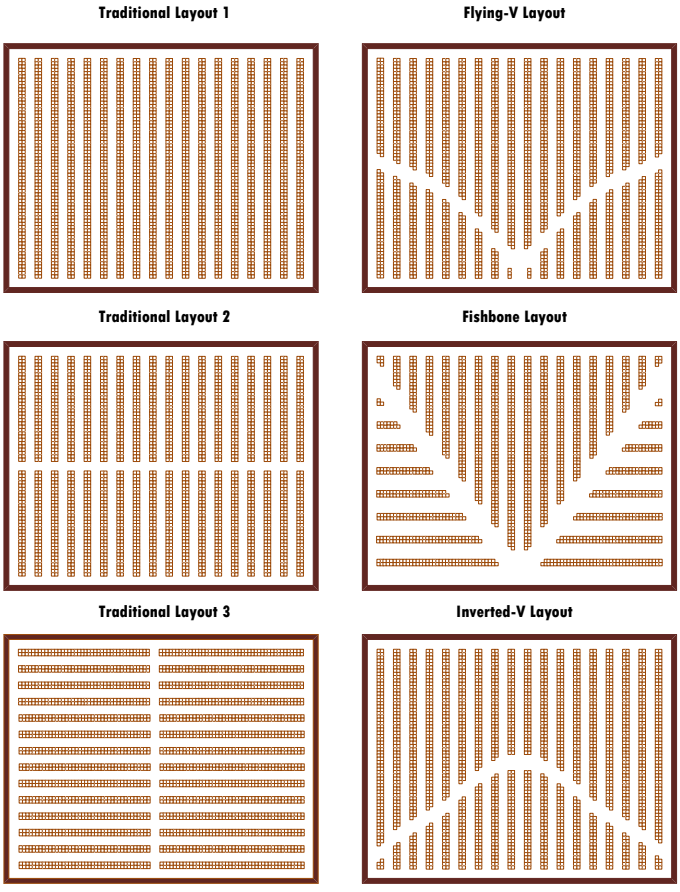


Fig. 2. The Warehouse Layout Examples, see [9] and [12]

There are many types of warehouse equipment, especially the equipment which should reduce labor cost and increase its utilization. Common storage models cover pallet racks, cartoon flow racks for high-volume picking, and shelving for lower-volume picking. All this equipment is standardized according to the dimensions, but the standardization is mostly only for a specific continent. While pallet manipulation is demanded in all types of warehouses, conveyors are not used everywhere. Conveyors divide the warehouse into zones, move a material through a given path, and also restricts the movement of workers and saves their energy. With the deployment of conveyors the sortation system is quite often installed. Sortation is mostly based on some scanning technology of Bar-Codes, RFID chips, Magnetic strips or Machine vision. The system works on a few common principles, e.g. *push sorter* pushes a passing carton to alternative path from the main conveyor, *tilt-tray sorter* works on the principle of tilting a tray and the object slides into the collecting bin, and others. Cranes are used to move materials over a variable paths in restricted area, e.g. jib crane, bridge crane, gantry crane, and stacker crane. Positioning equipment is used to handle material at a single location, e.g. hoists, balancers, and other manipulators. The automation in this area often covers systems such as Carousels, A-frames, and Automated Storage and Retrieval System (ASRS).

Carousel is a shelf rotating in the circle. Instead of the picker traveling, the storage location is moving. A simple rotation pattern on how to quickly find the shortest way to pick the order was introduced in [15] and [16]. Large orders in carousel environment have been studied in [17]. The carousels with multiple order-picking have been studied in [18]. Optimal storage locations have been investigated in [19]. An *A-frame* is an automated dispensing machine dropping items onto a conveyor. A-frame is used when a product is picked in very high volumes, the labor is expensive and is used only to refill A-frames. In-aisle cranes, so called ASRSs, replace the humans using trucks with simple robotic devices, moving in horizontal and vertical direction in the full extent of aisle. The design and performance of such models as well as travel time models have been investigated in [20], [21], [22], and [23]. Despite of all these inventions, the typical model of warehouse with pickers and various models of trucks are still quite popular. Other design and performance models are described in [24], [25], [26].

III. Optimization of Operational Structure

The operational and organizational framework combines different aspects from many areas, e.g. business management, inventory management, organization management, transportation management and many other areas of management. There are two basic slotting strategies (storing assignment policies [27]): *random* and *dedicated*. While random strategy allows to store a pallet on an arbitrary empty location with the same probability [28] or on the closest empty location [7], the dedicated strategy allows to store a pallet only on specified locations. The storage locations are often organized somehow. The organization can be e.g. *class-based storage*, where the goods are clustered according to the frequency of orders. This policy assigns the most frequently requested goods to the best (closest) locations from input/output gates. Another possibility is to use *family grouping*, where the goods are clustered according to relations or similarities between products or orders [29], [30].

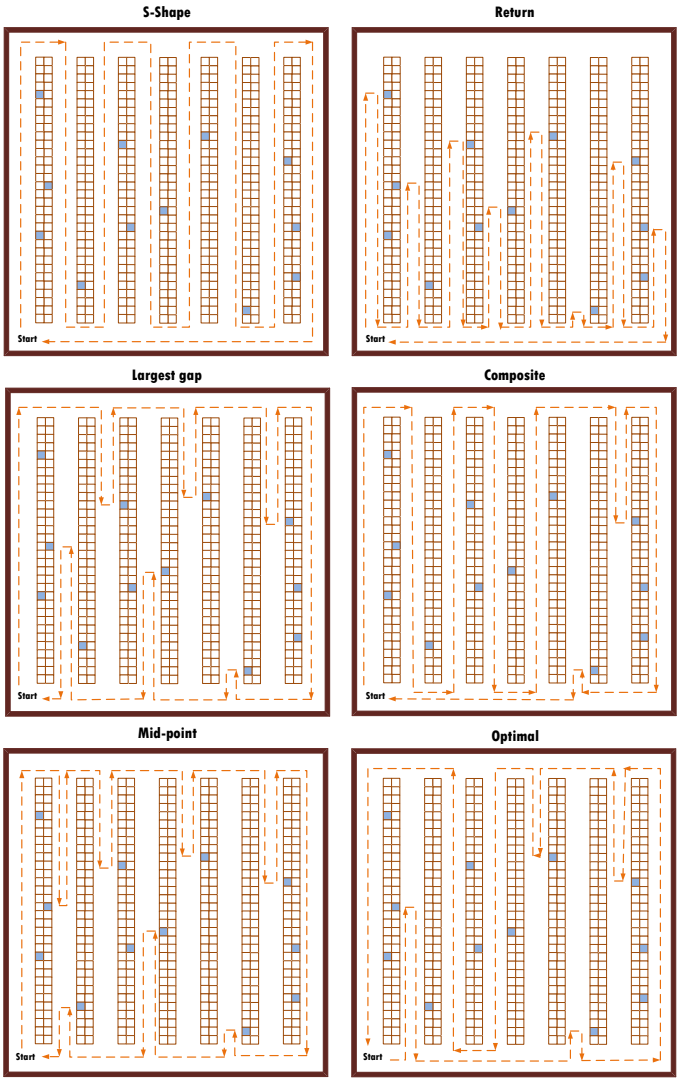


Fig. 3. The Common Routing Methods described in [5]

Single order-picking is the strategy where pickers pick only one order at a time and it is one of the most used picking policies. A Stock Keeping Unit (SKU) is tightly related to order-picking. SKU represents the smallest physical unit of a product with which a company manipulates, e.g. a box, a kind of case or carton consisting of inner packs and individual pieces of product, but it can also be only a pallet in huge distribution centers.

The routing policies should ensure an optimal travel path through the warehouse for order-picking. One of the first algorithms for optimal order-picking path design was introduced in [31]. Since the algorithm can be applied only for conventional warehouses, the problem is mostly solved by heuristic methods. The common routing methods, see Fig. 3, described in [5] are: *S-shape*, *Return policy*, *Mid-point strategy*, *Largest gap strategy*, *Composite heuristic*, *Optimal routing*. All the methods were primarily developed for single-block warehouses. Modified methods for multiple-block warehouses were proposed in [4].

If the order is small and is far from exceeding the picking capacity, it is possible to pick more orders in a single order-picking tour. This is known in the literature as *order batching* or simply *batching*. Since this is a job with sub-tasks (a picking tour with several orders) it is considered a NP-hard problem. It was proved in [32] that *batching* has a significant impact on the performance of order-picking. Therefore, researchers pay attention to the problem of *batching* and the heuristic methods are still under investigation [33], [34]. It is also possible to divide an order-picking process into zones. Goods belonging to the same product group are stored in the same zone. In comparison with batching, *zoning* does not have a significant impact on the performance of the order-picking system [35]. The advantage of zoning lies in reducing the congestion in the aisles and when the goods are really in one small area, the traveling is also reduced. The main disadvantage is the consolidation of order when it is completed by more pickers from different zones.

IV. Optimization of Warehouse Management

The coordinating systems and control are of special importance. The WMSs are used to control and optimize the warehouse and all typical warehousing operations (see section V), to know every detail about goods and their actual storage location all the time, the utilization of workforce, orders, and they also orchestrate the flow of people, machines, and goods. Such systems have many interfaces to adjacent systems in a company, e.g. merchandizing systems, information systems, production and enterprise resource planning systems, material flow and warehouse controlling systems and other systems related to a business-to-business or business-to-consumer.

Why is all the optimization being applied? Everything is based on customer demands. The main reasons to optimize are to increase the performance of the company regarding the demand-driven production (pull system), to ensure the productivity (based on just-in-time delivery) and minimize the stock along the supply chain, provide additional services, and reduce the transportation costs.

V. Typical Warehousing Operations

The basic processes in warehouse are receiving, storing, put-away, picking/retrieving and shipping goods. The shipping operation can also consist of many sub-tasks such as consolidation of goods if the batching, grouping or zoning is applied, checking the order according to its completeness, packing and, of course, shipping. The literature also mentions cross-docking as a special warehouse operation. The cross-docking is described at the end of this sub-section.

Receiving is the first operation in the warehouse. This process starts by notification of the arrival of goods. Then begins process of unloading, counting, identifying, quality control, and goods acceptance (incoming inspection) related to a type and quantity by unloading staff according to the company rules. When the goods are accepted, the receipt is issued. The acceptance depends on the delivery status – the delivery date, the quality of delivery, the planned schedule which should also minimize a truck waiting time.

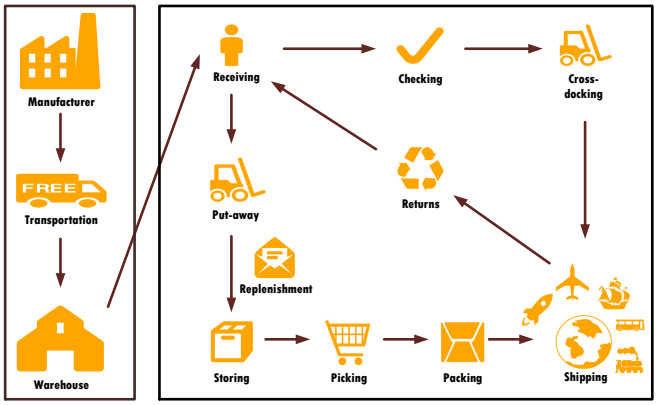


Fig. 4. The Logistic Process Flow in Warehouse

The product is then accepted, marked e.g. by a bar code and registered in the information system, and staged for put away. The receiving process takes about 10% of operating costs [6]. The paper [36] discusses the role of goods receiving and shipping in warehouse environments. A formal notation of schedules is proposed and the specific analytical examples are shown in this paper.

Storing operations consist of the distribution of goods to storage areas (transportation to a storage place or cross-docking, which is the transportation directly to the shipping department), identification (if it was not done during acceptance), assignment of the storage bin and put-away which is a simple determination of a storage bin concerning the physical dimensions and the weight of goods; storage monitoring is also a part of management systems – to know which goods are available and where.

Put-away is a process which requires a strictly determined storage location. This is very important, because the information system has to know all the time what storage locations are available, what is the location of a specific type of goods and where each particular pallet is stored. This information is also used for an efficient design of a pick-list. This process requires about 15% [6] of operating costs, because this covers a lot of transfers from the gate to the storage place.

Picking (also called **Retrieval**) is a process which covers lots of issues. First, a pick-list is given to an employee. The picking takes about 55% of the warehouse operating costs [6] (according to [7] it is 50% – 70%) and consists of: Traveling (55%), Searching (15%), Extracting (10%), and Paperwork (20%). Picking can be of two types, homogeneous and heterogeneous. Homogeneous picking is quite simple, the picker operates simply with a whole pallet. In heterogeneous picking the picker is told where and what to pick, in what quantity and units. Due to customer needs, the heterogeneous picking is logically more frequent. The disadvantage of heterogeneous picking is that a smaller unit means higher costs. The pick-list is still quite often a sheet of paper, but in some warehouses the pickers use the smart embedded devices such as the Bar-Code Reader, Personal Digital Assistant etc. The advantage of using such equipment is reviewed in [37], [38].

The planning of the picking process is based on orders and supported by picker routing methods. Basic routing methods are described in [5]. Picking of goods can be done by many ways. The special case of picking is order-picking which is a consolidation of a customized quantity of one or more articles related to a specific order. Sharing of order-picking is also a quite frequent way of picking. The sharing of an order is very related to batching, grouping, and zoning. The order-picking was also designed by algorithm based on Traveling Salesman Problem (TSP) heuristics introduced in [39], which performs better for multiple picking than routing methods. In [40] it was discovered that appropriate sequencing of picking is one of the crucial factors to achieve a high efficiency of picking. Since the traveling is the most time-consuming part of the process, the scientists paid attention mostly to this part of the problem.

The travel time is an increasing function of the travel distance, which was investigated in many papers and considered one of the primary optimization conditions. In [41] are described three analytical models of expected travel distance for return policy, traversal policy, and midpoint policy. To allay blocking and/or congestion, the order-picking strategies can be used, or the layout of the warehouse can be adjusted in the meaning of wide-aisles, or zoning can be applied [42]. The congestion has also been investigated in [7] by waiting time of a picker queuing to enter the warehouse. The analytical and simulation models of order-picking systems in [13] were developed to discover the system behavior with different activity levels. The result was that the congestion among workers can be a significant issue if the space is highly utilized. In [34] was provided an analytical approach for the expected system throughout time approximation. In [43] and [44] was determined a relationship between pick density and throughput, which has demonstrated the significance of blocking. Inasmuch as the models of picking are mostly investigated only as single-picker operations, and are consequently suitable to evaluate the order-picking efficiency by travel distance, an aisle congestion never takes place in such models. In real-world situations the congestion is a normal everyday situation in systems with multiple order-picking and dozens of workers. The throughput analysis for order-picking with multiple pickers and aisle congestion is investigated in [45] and [46]. In [47] were investigated heuristic methods and it was proved that the storage assignment policies in a multiple picker warehouse environment are outperformed by the proposed heuristics policies in this paper.

The batching in the narrow-aisle order-picking system with the picker blocking consideration was investigated in [48]. The paper has proposed strategies to control picker blocking with 5% – 15% reduction in the total retrieval time. In [49] has been proposed an Ant Colony Optimization (ACO) routing algorithm for two order pickers with the consideration of congestions. The paper analyzes the warehouse layout and its impact on the order-picking system performance and proves the good performance in dealing with the congestions if two pickers are used simultaneously. In [50] has been proposed two new batch construction heuristics called K-means Batching and Self-Organization Map Batching which minimize the total travel distance and the average picking truck utility.

Also the online version of the multiple picking agents for the warehouse management has been focused on in [51]. The rescheduling of the buffer of orders, to which new orders arrive randomly while old orders are being picked, was investigated and two real time algorithms were proposed. The solution performs good when dynamism is low or moderate, but when it is high, the solution tends to fail. The dynamic order-picking and cost reduction generated by optimal policies is discussed in [52]. This research used a Markov Decision Process based heuristics which is compared to some naive heuristics and an improvement in the range of 7% – 99% is depicted. Other heuristic for online batching based on offline batching is introduced in [53] and [54]. The algorithms are evaluated in a series of experiments and it is shown that the choice of an appropriate batching method can lead to significant minimization of the maximum completion time. In [55] have been proposed A*-algorithm for the routing and Simulated Annealing-algorithm for the batching. For batches of tree customers, the proposed algorithm produces results with an error of less than 1.2% compared to the optimal solution.

Consolidation of an order is a process of completion of a customer's order in case that it was picked by more than one pickers. The paper [56] has proposed the design and the operation of an order-consolidation warehouse. The paper provides a simulation model and shows its application. When the order is consolidated, the process of checking follows.

Checking of an order is a process that checks if the order is complete and accurate.

Packing ensures that the picked and consolidated goods, also checked for the completeness of an order, are packed for transportation and given to the shipping department. Packing can also be ensured by an autonomous packing department in the warehouse, then the consolidation and checking are usually part of this department.

Shipping ensures that the packed consignment is provided by transport destinations, assigned to the truck and optimally loaded on the truck. The shipping process is ensured by the shipping department that can also secure three preceding jobs.

Cross-docking is a process which minimizes the storage and order-picking time while the receiving and shipping operations are still allowed to a full extent. The basic idea is to transfer goods directly from the incoming to the outgoing department without any other warehouse operations in between. In [57] the problem was handled as a Vehicle Routing Problem. In [58] have been presented a solution developed for multiple delivery centers by many sub-optimizations of single centers based on Neighborhood Search and Tabu Search algorithms. The scheduling of trucks in the cross-docking system with five meta-heuristic algorithms (Genetic Algorithms, Simulated Annealing, Tabu Search, Electromagnetism-like Algorithm, Variable Neighborhood Search) with respect to minimization of the total operation time is described in [59]. The analytical models for pre-distribution cross-docking (on the side of a manufacturing company) and post-distribution cross-docking (on the side of warehousing company) proposed in [60] had been compared with a traditional distribution center system. Analytical results showed a pre-distribution cross-docking as a preferred solution for centers with shorter supply lead time and lower uncertainty of demand, but in general the preference depends highly on business environment [60].

VI. Conclusion

The optimal operation of a warehouse is achieved when each customer is satisfied completely according to his order, in due time and when all warehouse and logistic processes are done in the shortest possible time, with minimal cost and optimal utilization of resources under dynamically changing conditions. The literature presented in this paper gives great ideas of warehousing optimization, but only some of them are really applied in real-world warehouses and used. The problem of warehouse layout lies mainly in the effective use of space so the typical rectangular warehouses with narrow-aisles are most utilized. There is also a critical pressure on effective utilization of equipment and labor and its minimal quantity in the warehouse, which can also save the costs significantly.

The dedicated assignment based on the frequency of manipulation with goods is broadly used, but some big and well-known companies, e.g. *Amazon*, use the chaotic assignment system and it seems to be a good solution as well. The routing methods supporting order-picking and picking itself have been investigated for single picking tours, but the batching seems to be standard for many companies. Moreover, the most of the scientific papers do not take into account the real conditions as the blocking and congestion are, but there are dozens of workers working simultaneously in the real warehouse or multiple-block warehouse and the congestion or even a collision must be taken into account in everyday operation.

The result of this review of warehouse optimization is that the real-world conditions should be applied instead of their relaxation for the sake of application. The work can be shared by many more employees than two, which was mostly investigated in the papers. The idea of how to apply this optimization is to utilize the shop scheduling techniques combined with Vehicle Routing Problems solving techniques.

The shop scheduling techniques can be employed when the work is scheduled in the warehouse, even when the work must by scheduled dynamically. The machines in shop scheduling problem are represented in the warehouse by any equipment needed for each job, such as trucks driven by workers (fork-lift hand pallet truck, fork-lift low truck, fork-lift high truck), checking units (workers), packing units (workers with special equipment) and others. The operations in the warehouse, called jobs in the scheduling, represent a single assignment given to the worker by operational manager e.g. the employee has to unload a pallet from a lorry, go through the warehouse and store it on a shelf. The job is composed of sub-operations called tasks. A task represents a single operation of job, e.g. receiving, unloading, put-away, moving and storing etc. The tasks can be done by several workers. So, the job is spread in few machines working in sequence in the language of shop scheduling problems. Transports, moving and routing of trucks in warehouse could be inspired by Automated Guided Vehicles (AGV) techniques transformed from open space Vehicle Routing Problem (VRP) techniques to warehouse environment. Application of these methods could further reduce the blocking and congestions as well as collisions of trucks.

ACKNOWLEDGMENT

This research work is funded by projects SIX CZ.1.05/2.1.00/03.0072.

REFERENCES

[1] M. T. Hompel and T. Schmidt, *Warehouse Managmenet - Automation and Organisation of Warehouse and Order Picking Systems*. Springer-Verlag Berlin Heidelberg, 2007, iSBN-13: 971-3-540-35218-1.

[2] J. Gu, M. Goetschalckx, and L. F. McGinnis, "Research on warehouse design and performance evaluation: A comprehensive review," *European Journal of Operational Research*, vol. 203, no. 3, pp. 539–549, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0377221709005219

[3] F. Caron, G. Marchet, and A. Perego, "Optimal layout in low-level picker-to-part systems," *International Journal of Production Research*, vol. 38, no. 1, pp. 101–117, 2000. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/002075400189608

[4] K. J. Roodbergen and R. d. Koster, "Routing methods for warehouses with multiple cross aisles," *International Journal of Production Research*, vol. 39, no. 9, pp. 1865–1883, 2001. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/00207540110028128

[5] ——, "Routing order pickers in a warehouse with a middle aisle," *European Journal of Operational Research*, vol. 133, no. 1, pp. 32–43, 2001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0377221700001776

[6] J. J. Bartholdi and S. T. Hackman, *Warehouse & Distribution Science*. Georgia Institute of Technology, School of Industrial and Systems Engineering, The Supply Chain and Logistics Institute, August 22 2011, latest release: version 0.95.

[7] R. d. Koster, T. Le-Duc, and K. J. Roodbergen, "Design and control of warehouse order picking: A literature review," *European Journal of Operational Research*, vol. 182, no. 2, pp. 481–501, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0377221706006473

[8] S. S. Heragu, *Facilities Design*, 3rd ed. CRC Press, June 19 2008, iSBN-13: 978-1420066265.

[9] K. R. Gue and R. D. Meller, "Aisle configurations for unit-load warehouses," *IIE Transactions*, vol. 41, no. 3, pp. 171–182, 2009.

[10] L. M. Pohl, R. D. Meller, and K. R. Gue, "An analysis of dual-command operations in common warehouse designs," *Transportation Research Part E: Logistics and Transportation Review*, vol. 45, no. 3, pp. 367–379, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1366554508001191

[11] ——, "Optimizing fishbone aisles for dual-command operations in a warehouse," *Naval Research Logistics*, vol. 56, no. 5, pp. 389–403, 2009.

[12] K. R. Gue, G. Ivanovič, and R. D. Meller, "A unit-load warehouse with multiple pickup and deposit points and non-traditional aisles," *Transportation Research Part E: Logistics and Transportation Review*, vol. 48, no. 4, pp. 795–806, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1366554512000117

[13] K. Gue, R. Meller, and J. Skufca, "The effects of pick density on order picking areas with narrow aisles," *IIE Transactions*, vol. 38, no. 10, pp. 859–868, 2006. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-33746452243&partnerID=40&md5=f9fb583c292344e9560d0aa020e6e9ef

[14] M. Napolitano, "Real dc stories: Low cost deep impact," *Logistics Management*, vol. 48, no. 1, pp. 46–49, 2009.

[15] J. J. Bartholdi and L. K. Platzman, "Retrieval strategies for a carousel conveyor," *IIE Transactions*, vol. 18, no. 2, pp. 166–173, 1986. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/07408178608975344

[16] J. B. Ghosh and C. E. Wells, "Optimal retrieval strategies for carousel conveyors," *Mathematical and Computer Modelling*, vol. 16, no. 10, pp. 59–70, 1992. [Online]. Available: http://www.sciencedirect.com/science/article/pii/089571779290060X

[17] N. Litvak, "Optimal picking of large orders in carousel systems," *Operations Research Letters*, vol. 34, no. 2, pp. 219–227, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167637705000532

[18] J. P. V. D. Berg, "Multiple order-pick sequencing in a carousel system: A solvable case of the rural postman problem," *The Journal of the Operational Research Society*, vol. 47, no. 12, pp. 1504–1515, December 1996.

[19] R. G. Vickson and A. Fujimoto, "Optimal storage locations in a carousel storage and retrieval system," *Location Science*, vol. 4, no. 4, pp. 237–245, 1996. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S096683499700003X

[20] Y. A. Bozer and J. A. White, "Travel-time models for automated storage/retrieval systems," *IIE Transactions*, vol. 16, no. 4, pp. 329–338, 1984. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/07408178408975252

[21] ——, "Design and performance models for end-of-aisle order picking systems," *Management Science*, vol. 36, no. 7, pp. 852–866, July 1990.

[22] Y.-H. Hu, S. Y. Huang, C. Chen, W.-J. Hsu, A. C. Toh, C. K. Loh, and T. Song, "Travel time analysis of a new automated storage and retrieval system," *Computers & Operations Research*, vol. 32, no. 6, pp. 1515–1544, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0305054803003575

[23] T. Lerher, I. Potrč, M. Šraml, and T. Tollazzi, "Travel time models for automated warehouses with aisle transferring storage and retrieval machine," *European Journal of Operational Research*, vol. 205, no. 3, pp. 571–583, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0377221710000330

[24] B. R. Sarker and P. S. Babu, "Travel time models in automated storage/retrieval systems: A critical review," *International Journal of Production Economics*, vol. 40, no. 2–3, pp. 173–184, August 1995. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0925527395000752

[25] J. P. V. D. Berg, "Analytic expressions for the optimal dwell point in an automated storage/retrieval system," *International Journal of Production Economics*, vol. 76, no. 1, pp. 13–25, 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925527301001499

[26] I. Potrč, T. Lerher, J. Kramberger, and M. Šraml, "Simulation model of multi-shuttle automated storage and retrieval systems," *Journal of Materials Processing Technology*, vol. 157–158, no. 0, pp. 236–244, 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0924013604010787

[27] J. L. Haskett, "Cube-per-order index - a key to warehouse stock location," *Transportation and Distribution Management*, vol. 3, no. 1, pp. 27–31, 1963.

[28] C. G. Petersen, "An evaluation of order picking routeing policies," *International Journal of Operations & Production Management*, vol. 17, no. 11, pp. 1098–1111, 1997.

[29] H. Brynzer and M. I. Johansson, "Storage location assignment: Using the product structure to reduce order picking times," *International Journal of Production Economics*, vol. 46, no. 1, pp. 595–603, December 1996. [Online]. Available: http://ideas.repec.org/a/eee/proeco/v46-47y1996i1p595-603.html

[30] R. J. Mantel, P. C. Schuur, and S. S. Heragu, "Order oriented slotting: A new assignment strategy for warehouses," *European Journal of Industrial Engineering*, vol. 1, no. 3, pp. 301–316, January 1 2007. [Online]. Available: http://inderscience.metapress.com/content/A262442816131833

[31] H. D. Ratliff and A. S. Rosenthal, "Order-picking in a rectangular warehouse: A solvable case of the traveling salesman problem," *Operations Research*, vol. 31, no. 3, pp. 507–521, May–June 1983.

[32] G. Clarke and J. Wright, "Scheduling of vehicles from a central depot to a number of delivery points," *Operations Research*, vol. 12, no. 4, pp. 568–581, 1964.

[33] R. d. Koster, E. S. V. d. Poort, and M. Woltersa, "Efficient orderbatching methods in warehouses," *International Journal of Production Research*, vol. 37, no. 7, pp. 1479–1504, 1999.

[34] I. V. Nieuwenhuyse and R. B. d. Koster, "Evaluating order throughput time in 2-block warehouses with time window batching," *International Journal of Production Economics*, vol. 121, no. 2, pp. 654–664, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925527309000723

[35] H. d. Ruijter, "Improved storage in a book warehouse," Master's thesis, University of Twente, Enschede – The Netherlands, October 2007.

[36] T. Ambroziak and K. Lewczuk, "A method for scheduling the goods receiving process in warehouse facilities," *Total Logistic Management*, vol. 5, no. 1, pp. 7–14, 2008.

[37] X. Zhu, S. K. Mukhopadhyay, and H. Kurata, "A review of rfid technology and its managerial applications in different industries," *Journal of Engineering and Technology Management*, vol. 29, no. 1, pp. 152–167, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S092347481100049X

[38] M. K. Lim, W. Bahr, and S. C. H. Leung, "Rfid in the warehouse: A literature analysis (1995–2010) of its applications, benefits, challenges and future trends," *International Journal of Production*

Economics, vol. 145, no. 1, pp. 409–430, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925527313002314

[39] C. Theys, O. Braysy, W. Dullaert, and B. Raa, "Using a tsp heuristic for routing order pickers in warehouses," *European Journal of Operational Research*, vol. 200, no. 3, pp. 755–763, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0377221709000514

[40] K. Moeller, "Increasing warehouse order picking performance by sequence optimization," *Procedia - Social and Behavioral Sciences*, vol. 20, no. 0, pp. 177–185, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877042811014030

[41] H. Hwang, Y. Oh, and Y. Lee, "An evaluation of routing policies for order-picking operations in low-level picker-to-part system," *International Journal of Production Research*, vol. 42, no. 18, pp. 3873–3889, 2004. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-4944240916&partnerID=40&md5=73cb6af28941f0e4ab9ea72f412233e1

[42] M. B. M. d. Koster and M. Yu, "Minimizing makespan and throughput times at aalsmeer flower auction," *Journal of Operational Research Society*, vol. 59, no. 9, pp. 1182–1190, September 2008. [Online]. Available: http://dx.doi.org/10.1057/palgrave.jors.2602474

[43] S. Hong, A. L. Johnson, and B. A. Peters, "Analysis of picker blocking in narrow-aisle batch picking," in *Proceedings of 2010 International Material Handling Research Colloquium (IMHRC)*, K. Ellis, K. Gue, d. R. Koster, R. Meller, B. Montreuil, and M. Oglep, Eds. The Material Handling Institue, Charlotte, NC, USA, 2010.

[44] P. J. Parikh and R. D. Meller, "A note on worker blocking in narrow-aisle order picking systems when pick time is non-deterministic," *IIE Transactions*, vol. 42, no. 6, pp. 392–404, 2010. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-77951121237&partnerID=40&md5=6d6f6b92c672c42e12c419901b3428af

[45] J. C.-H. Pan and P.-H. Shih, "Evaluation of the throughput of a multiple-picker order picking system with congestion consideration," *Computers & Industrial Engineering*, vol. 55, no. 2, pp. 379–389, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S036083520800003X

[46] J. C.-H. Pan and M.-H. Wu, "Throughput analysis for order picking system with multiple pickers and aisle congestion considerations," *Computers & Operations Research*, vol. 39, no. 7, pp. 1661–1672, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0305054811002838

[47] J. C.-H. Pan, P.-H. Shih, and M.-H. Wu, "Storage assignment problem with travel distance and blocking considerations for a picker-to-part order picking system," *Computers & Industrial Engineering*, vol. 62, no. 2, pp. 527–535, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S036083521100324X

[48] S. Hong, A. L. Johnson, and B. A. Peters, "Batch picking in narrow-aisle order picking systems with consideration for picker blocking," *European Journal of Operational Research*, vol. 221, no. 3, pp. 557–570, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0377221712002706

[49] F. Chen, H. Wang, C. Qi, and Y. Xie, "An ant colony optimization routing algorithm for two order pickers with congestion consideration," *Computers & Industrial Engineering*, vol. 66, no. 1, pp. 77–85, 2013.

[50] L.-F. Hsieh and Y.-C. Huang, "New batch construction heuristics to optimise the performance of order picking systems," *International Journal of Production Economics*, vol. 131, no. 2, pp. 618–630, 2011. [Online]. Available: http://EconPapers.repec.org/RePEc:eee:proeco:v:131:y:2011:i:2:p:618-630

[51] J. I. U. Rubrico, T. Higashi, H. Tamura, and J. Ota, "Online rescheduling of multiple picking agents for warehouse management," *Robot. Comput.-Integr. Manuf.*, vol. 27, no. 1, pp. 62–71, February 2011. [Online]. Available: http://dx.doi.org/10.1016/j.rcim.2010.06.011

[52] Y. Bukchin, E. Khmelnitsky, and P. Yakuel, "Optimizing a dynamic order-picking process," *European Journal of Operational Research*, vol. 219, no. 2, pp. 335–346, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0377221711011350

[53] S. Henn, "Algorithms for on-line order batching in an order picking warehouse," *Computers & Operations Research*, vol. 39, no. 11, pp. 2549–2563, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0305054812000020

[54] S. Henn and V. Schmid, "Metaheuristics for order batching and sequencing in manual order picking systems," *Computers & Industrial Engineering*, vol. 66, no. 2, pp. 338–351, 2013, iSSN: 0360–8352. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360835213002015

[55] M. Matusiak, R. Koster, L. Kroon, and J. Saarinen, "A fast simulated annealing method for batching precedence-constrained customer orders in a warehouse," *European Journal of Operational Research*, vol. ?, no. ?, p. ?, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0377221713004670

[56] A. E. Gray, U. S. Karmarkar, and A. Seidmann, "Design and operation of an order-consolidation warehouse: Models and application," *European Journal of Operational Research*, vol. 58, pp. 14–36, 1992.

[57] G. Mosheiov, "Vehicle routing with pick-up and delivery: Tour-partitioning heuristics," *Computers and Industrial Engineering*, vol. 34, no. 3, pp. 669–684, 1998.

[58] G. Barbarosoglu and D. Ozgur, "A tabu search algorithm for the vehicle routing problem," *Computers and Operations Research*, vol. 26, no. 3, pp. 255–270, 1999. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-0033102291&partnerID=40&md5=f165e5689ada1c5df72d37d8aaffa60b

[59] B. Vahdani and M. Zandieh, "Scheduling trucks in cross-docking systems: Robust meta-heuristics," *Computers & Industrial Engineering*, vol. 58, no. 1, pp. 12–24, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360835209001776

[60] H. Yan and S.-l. Tang, "Pre-distribution and post-distribution cross-docking operations," *Transportation Research Part E: Logistics and Transportation Review*, vol. 45, no. 6, pp. 843–859, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1366554509000830

Jan Karásek received MSc degree in Telecommunications and Informatics in 2009 and MSc degree in Corporate Management and Economics in 2010 from the Brno University of Technology, Czech Republic. Currently, he is a PhD candidate at the Department of Telecommunications at the same university. The theme of his PhD thesis is the Optimization of Distributed Processes Using Evolutionary Algorithms. At the same time, he is employed with the Department of Telecommunications as a project manager and he is also involved in teaching of Theoretical Informatics. His research interests include: artificial intelligence, machine learning, evolutionary computation techniques, process optimization and automation, genetic algorithms & genetic programming and their real-world application, and parallel programming.

A method for automatically constructing the initial contour of the common carotid artery

Yara Omran, and Kamil Riha

***Abstract*—In this article we propose a novel method to automatically set the initial contour that is used by the Active contours algorithm. The proposed method exploits the accumulative intensity profiles to locate the points on the arterial wall. The intensity profiles of sections that intersect the artery show distinguishable characterstics that make it possible to recognize them from the profiles of sections that do not intersect the artery walls. The proposed method is applied on ultrasound images of the transverse section of the common carotid artery, but it can be extended to be used on the images of the longitudinal section. The intensity profiles are classified using Support vector machine algorithm, and the results of different kernels are compared. The extracted features used for the classification are basically statistical features of the intensity profiles. The echogenicity of the arterial lumen, and gives the profiles that intersect the artery a special shape that helps recognizing these profiles from other general profiles. The outlining of the arterial walls may seem a classic task in image processing. However, most of the methods used to outline the artery start from a manual, or semi-automatic, initial contour. The proposed method is highly appreciated in automating the entire process of automatic artery detection and segmentation.**

***Keywords*—SVM, Ultrasound imaging, Segmentation, Carotid artery.**

I. Introduction

ULTRASONIC imaging is a widely used non-invasive medical imaging procedure since it is economical, comparatively safe, portable and adaptable. However, one of its main weaknesses is the poor quality of images, making the processing of these images more challenging. Segmentation of Ultrasound images is an initial step for the analysis and diagnosis of the data in the image. Our goal in this article is to locate points on the arterial wall, the considered object is the artery (carotis communis) in a transverse section captured by a Sonographic device. The artery in this section has an approximately circular shape, and the artery wall appears as bright areas. This step of detecting the artery is important for making other processes fully automated; such a process is tracking the movement of the artery in a video sequence to obtain useful information from it, or to predict the existence of some diseases or irregularities. Circular pattern detection is one of the classic tasks in image processing; many approaches to achieve this goal were proposed such as (CHT) [2], correlation [3], maximal-likelihood [4] or Genetic Algorithms which are computationally expensive [5]. However, due to the special nature of medical Sonographic images with the high level of noise and the blurred edges, we are in need to new approaches.

Authors are with the Department of Telecommunications, Brno University of Technology, Czech Republic (yara.omran@phd.feec.vutbr.cz, riha k@feec.vutbr.cz).

Many approaches to detect objects in medical images were presented such as [9], or trainable image segmentation [10]. Other approaches dealt specifically with detecting a circular object in medical images like [11], [12], with the exploitation of ellipse fitting. Golemati et al. [13], proposed a system for automatically extracting lines and circles from sequences of B-mode ultrasound images of longitudinal and transverse sections of the carotid artery, their methodology was based on the Hough Transform (HT). They used the CHT for automatically detecting the artery in the transverse section, and estimating the Arterial Distension Waveforms (ADW) as well as the diastolic and systolic diameters. For the longitudinal section, they used the HT for estimating the Intima-Media thickness (IMT). Mao et al. [14], proposed an interactive method based on dynamic contour model initialized manually with one seed, the used model combined geometric constraints, image gradient, and contrast features to control the contour and thus to optimize the segmentation process. In [11], a system for localizing the artery based on motion features was proposed, were the pulsatile character of the artery wall was used as information for localizing it after compensating other global movements. The method is based on the optical flow technique to detect the region of interest that contains the artery, followed by CHT (Circular Hough Transform) to estimate the center and radius of the artery. Golemati et al. [16] proposed an active contour based method to automatically extract lines and circles from longitudinal and transverse sections; the active contour was initialized by HT to detect the circle. This method used gradient vector flow to deform the contour, resulting in a random shaped boundary which follows more closely the actual wall - lumen interface than a circle. In [15], authors used a method based on watershed segmentation algorithm, their technique consisted of four major stages that are: preprocessing, watershed segmentation, region merging, and boundary extraction. In this article we assume that the ROI containing the artery is already defined by means of an object detector or any other methodology, our aim is to locate points on the arterial wall. The main contribution of the article is automatically finding the points that will form the initial contour to be passed to the Active Contour algorithm in order segment the artery in more detail.

II. Detecting the points on the arterial wall

As mentioned in the introduction, many successful approaches for segmenting the artery used Active contour algorithms, the most weak point in these methods, is that they were initialized manually or semi-automatically. In this article we exploit the fact that the image intensity of the lumen area is approximately

Fig. 1. Ultrasound image of CCA, extracted by an object detector, after thresholding and dilattion

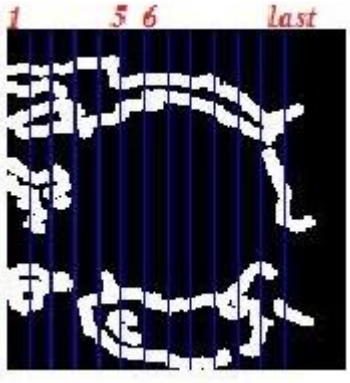


Fig. 2. Ultrasound image of CCA, image profiles are taken according to the vertical lines

uniform (it appears as a black area) since the lumen only contains blood, while the other areas are usually composed of variety of different tissues, and have a non-uniform intensity, which makes the image profiles of areas containing the artery distinguishable from those which don't. The intensity profile in the vertical direction shown in Fig. 2 demonstrates the tissue variation. This variation allows us to analyze the variance of the profile data in different cuts to identify the location of points on arterial wall. As Fig. 3 shows, the intensity profiles of areas which contain the artery share a similar pattern (i.e. the gap in the profile which represents the lumen area), while intensity profiles of areas which don't contain the artery have an arbitrary shape. In this article we classify these profiles into two groups, those which cross the artery, and those which do not. In this way we detect accurately the peaks in the profile which correspond to the edges of the artery wall see Fig. 3. We used the SVM classifier with different kernels, and we compared the performance of each one of these kernels in classifying the intensity profiles.

A. Training and testing data

A set of 29 B-mode ultrasound images, captured by Sonix OP, have been used for this experiment. All images are an outcome of an object detector as in [14], and they contain the CCA in transverse section (see Fig. 1). The images were scanned with different settings of acquisition hardware (frequency, depth, gain) and different positioning of a probe the size of each image is 142×142 pixels.

B. SVM classifier

The classification was based on image local features obtained from a neighborhood of particular pixels, in addition to the pixels' intensity itself. The appropriate selection of features is crucial for the performance of the classifier. It is important to select features that separate both classes. In this implementation, the selected features were- the mean value in the 8^{th} neighborhood, the standard deviation, median value, the maximal and the minimal intensity, and others. The classifier was trained on 249 intensity profiles extracted from the previously described images; these profiles were manually labeled into two groups, the first one is the positive group, which is the group of profiles of areas that cross the artery, and the second group is the negative profiles, which is the group of profiles for areas outside the artery. The profiles were taken in the vertical direction, and the starting point of each profile is shifted by 10 pixels from the adjoining profiles. Each profile consists of intensity values of pixels along the image, between the start point, at the first row, and he end point, at the last row of the image see Fig. 3.

TABLE I
THE STATISTICAL MEASURES FOR EACH KERNEL

Polynomial kernel		
accuracy	presicion	recall
87.13%	86.78%	90.92%
Radial kernel		
accuracy	presicion	recall
92.77%	95.79%	90.64%
Multi-quadric kernel		
accuracy	presicion	recall
52.60%	52.61%	100%
Anova kernel		
accuracy	presicion	recall
95.58%	95.27%	96.86%
Dot kernel		
accuracy	presicion	recall
91.57%	94%	90.83%

III. EXPERIMENTAL RESULTS

We trained the SVM classifier using 5 different kernels (Polynomial, radial, anova, multi-quadric, and dot). The statistical measures: accuracy, precision, and recall were calculated for each used kernel, see equations 1. 2. and 3. The classifier performance was evaluated using the Cross-validation method; this method partitions the set of data into a number of complementary subsets, and performs the training on a group of subsets (called the training set), and validates the analysis on the other subsets (usually only one subset is used for validation, and it is called the testing set). To reduce the

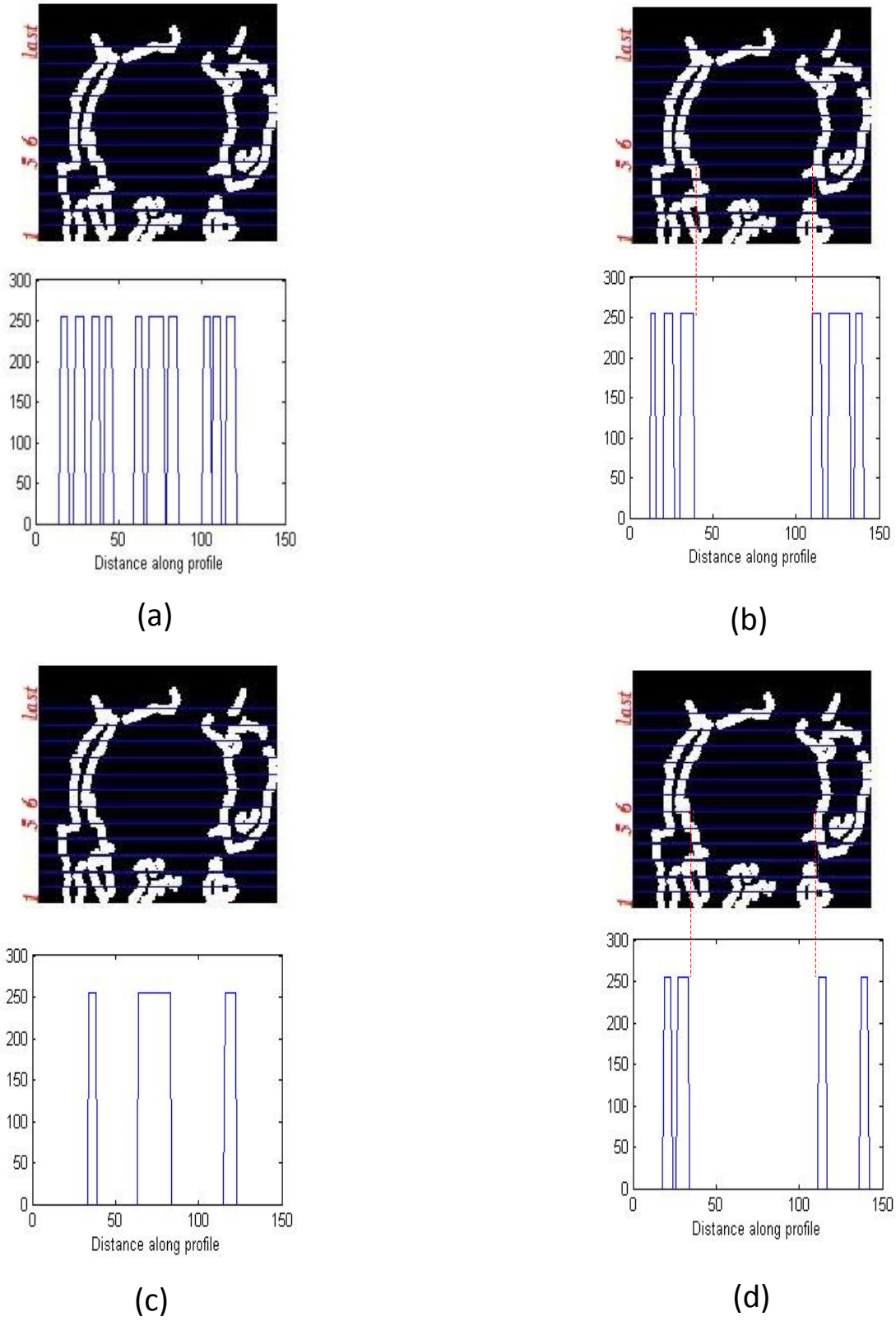


Fig. 3. (a) image profile according to line number 1, (b) image profile according to line number 5, (c) image profile according to the last line, (d) image profile according to line number 6. The artery images are rotated by 90 degrees for illustration purpose

variability, multiple rounds of cross-validation are performed using different partitions (making the training and testing sets change roles), and the validation results are averaged over the rounds. In our case we used 10 subsets for training and testing. The results show a good performance of the SVM classifier for different kernels, except the multiquadric kernel which had a poor performance in classifying the profiles.

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$presicion = \frac{TP}{TP + FP} \quad (2)$$

$$recall = \frac{TP}{TP + FN} \quad (3)$$

IV. Conclusion

In this article, we presented a new method for automatically constructing the initial contour that is then given to the Active contours algorithm. The proposed method is based on the classification of the intensity profiles of ultrasound images of the common carotid artery. The input of the method is the area containing the artery. The used classification algorithm is SVM. Different kernels are used, and their performance is compared. The proposed method is highly appreciated in systems that deal with artery detection and segmentation, as most of the methods used for segmentation are initialized manually or semi-automatically. The proposed methosd is applied on ultrasound images of the transverse section of the arterial wall, but it can be extended to be used on images of the longitudinal wall.

Acknowledgements

Acknowledgement: This research work is funded by the project: SIX CZ.1.05/2.1.00/03.0072, EU ECOP EE.2.3.20.0094, CZ.1.07/2.2.00/28.0062, Czech Republic. Supported by the Ministry of industry and trade of the Czech republic number TI-FR3/170.

References

[1] C. Loizou, C. Pattichis, M. Pantziaris, T. Tyllis, and A. Nicolaides, *Quality evaluation of ultrasound imaging in the carotid artery based on normalization and speckle reduction filtering*, Medical and Biological Engineering and Computing, Volume 44, Issue 5, 2006, Pages 414-426, ISSN 0140-0118, DOI:10.1007/s11517-006-0045-1.

[2] C. P. Loizou, C. S. Pattichis, M. Pantziaris, and A. Nicolaides, *An Integrated System for the Segmentation of Atherosclerotic Carotid Plaque*, Information Technology in Biomedicine, IEEE Transactions on, vol.11, no.6, pp.661-667, Nov. 2007. doi: 10.1109/TITB.2006.890019.

[3] F. Molinari, G. Zeng, and J. S. Suri, *A state of the art review on intima–media thickness (IMT)measurement and wall segmentation techniques for carotid ultrasound*, Computer Methods and Programs in Biomedicine, Volume 100, Issue 3, December 2010, Pages 201-221, ISSN 0169-2607, DOI:10.1016/j.cmpb.2010.04.007.

[4] G. Treece, R. Prager, and A. Gee, *Ultrasound attenuation measurement in the presence of scatterer variation for reduction of shadowing and enhancement*, Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on, vol.52, no.12, pp.2346-2360, Dec. 2005 doi: 10.1109/TUFFC.2005.1563279.

[5] I. B. Casella, C. Presti, R. M. P. Porta, C. R. D. Sabbag, M. A. Bosch, and Y. Yamazaki, *A practical protocol to measure common carotid arteryintima-media thickness*, Clinics, Volume 63, Issue 4, 2008, ISSN 1807-5932.

[6] J. E. Wilhjelm, M.- L. M. Gronholdt, B. Wiebe, S. K. Jespersen, L. K. Hansen, and H. Sillesen, *Quantitative analysis of ultrasound B-mode images of carotid atherosclerotic plaque: correlation with visual classification and histological examination*, Medical Imaging, IEEE Transactions on, vol. 17, no. 6, pp. 910-922, Dec. 1998, DOI: 10.1109/42.746624.

[7] T. Elatrozy, A. Nicolaides, T. Tegos, AZ. Zarka, M. Griffin, and M. Sabetai, *The effect of B-mode ultrasonic image standardization of the echodensity of symptomatic and asymptomatic carotid bifurcation plaque*, International Angiology, September 1998, Volume 17, Issue 3, Pages 179–186.

[8] E. Kyriakou, MS. Pattichis, C. Christodoulou, CS. Pattichis, S. Kakkos, M. Griffin, and A. Nicolaides, *Ultrasound imaging in the analysis of carotid plaque morphology for the assessment of stroke*, 2005, PMID: 15923744.

[9] R. Benes, M. Hasmanda, K. Riha, *Object localization in medical images*, Telecommunications and Signal Processing (TSP), 2011 34th International Conference on, vol., no., pp.559-563, 18-20 Aug. 2011, DOI: 10.1109/TSP.2011.6043667.

[10] . D. Gill, H. M. Ladak, D. A. Steinman, and A. Fenster, *Segmentation of ulcerated plaque: A semi-automatic method for tracking the progression of carotid atherosclerosis,* in Proc. World Congr. Med. Phys. Biomed. Eng., Chicago, IL, 2000, pp. 1–4, 2000.

[11] . Zhang, C. R. McKay, and M. Sonka, *Tissue characterization in intravascular ultrasound images,* IEEE Trans. Med. Imag., vol. 17, no. 6, pp. 889–899, Dec. 1998.

[12] . R. Cardinal, J. Meunier, G. Soulez, E. Th'erasse, and G. Cloutier, *Intravascular ultrasound image segmentation: A fast-marching method,* in Proc. MICCAI, LNCS 2879, 2003, pp. 432–439.

[13] . Brusseau, C. L. De Korte, F. Mastick, J. Schaar, and A. F. W. Van der Steen, *Fully automatic luminal contour segmentation in intracoronary ultrasound imaging: A statistical approach,* IEEE Trans. Med. Imag., vol. 23, no. 5, pp. 554–566, May 2004.

[14] OMRAN, Y.; ŘÍHA, K.; DUTTA, M. Automatic Estimation Of The Arterial Parameters In Ultrasound Video Sequence. In 36th International Conference on Telecommunications and Signal Processing (TSP). 2013. s. 10-14. ISBN: 978-1-4799-0403- 7.

Analysis of Network Parameters Influencing Performance of Hybrid Multimedia Networks

Pavel Masek[1], Dominik Kovac[2], Jiri Hosek[3], Mariya Pavlova[4], Ondrej Krajsa[5]

***Abstract* — Multimedia networks is an emerging subject that currently attracts the attention of research and industrial communities. This environment provides new entertainment services and business opportunities merged with all well-known network services like VoIP calls or file transfers. Such a heterogeneous system has to be able to satisfy all network and end-user requirements which are increasing constantly. Therefore, there is a need for simulation tools enabling deep analysis to find key performance indicators and factors which influence the overall quality for specific network service. This paper provides a study on the network parameters like communication technology, routing protocol, QoS mechanism, etc. and their effect on the performance of hybrid multimedia network. The analysis was performed in OPNET Modeler environment and the most interesting results are discussed at the end of this paper.**

***Keywords* — Multimedia networks, Network KPI, OPNET Modeler, Performance analysis, PON, QoS, Routing system.**

I. Introduction

The most recent trend in the field of communication networks clearly leads to a convergence of all types of services under a single communication platform known as a hybrid multimedia network. Such a network must be able to efficiently handle traditional voice calls, emerging real-time videoconferencing applications or generic data which consist of a large group of network services (email, web, remote access, etc.). Moreover, end-users demands for high quality of network services are constantly growing over the time. Therefore, multimedia networks must be able to meet all strict performance criteria.

Even though modern multimedia networks differ in many parameters, i.e. the topology, type of provided services, link utilization, used communication technologies, etc., they are mostly built on TCP/IP platform which provides a number of advantages already discussed many times. On the other hand, general architecture presents a number of interfaces or subsystems that can potentially affect the network KPIs (Key Performance Indicators), e.g. throughput, end-to-end delay, jitter, etc.). One of the most important factors that significantly affect network KPIs and thus the overall quality of service perceived by end-users is the routing subsystem. However, it is not just about choosing the most suitable routing protocol.

Especially for the communication networks operating the multimedia data transmission in real-time, complementary mechanisms and algorithms closely related to the routing protocol are very important. These mechanisms include methods to ensure QoS (Quality of Service) which aims to provide preferential treatment for specific type of network over others. Due to this fact, it is possible to achieve the desired KPIs for various types of multimedia services.

When building a communication network or implementing new services, it is always very important to perform a thorough analysis and testing before production operation. Testing in a real network is in most cases very expensive and therefore variety of network simulators and emulators are used to verify the correct functioning of newly developed technologies and protocols.

In this work, the analysis and testing of routing subsystem and related mechanisms in order to provide maximal performance of hybrid multimedia network is performed. For this purpose, several simulation scenarios were developed in the OPNET Modeler environment.

II. Optimization of Network Techniques for Hybrid Multimedia Networks

With the gradual development of applications working in real-time, the end-users demands for compliance with predefined parameters during data transfer through network infrastructure grow. The most important of KPIs, also recognized as QoS parameters, are the end-to-end delay, jitter or network throughput. Specific values of these QoS parameters are needed to be ensured during data transfers carried out in the hybrid multimedia network to provide demanded quality perceived by end-user.

To achieve excellent KPIs in multimedia networks, various methods, algorithms and technologies are available. This study includes the evaluation of some of these methods, like the optimization of routing protocol parameters, the classification of network traffic into VLANs (Virtual Local Network) or an implementation of QoS mechanism.

A. Routing protocols in multimedia networks

Selecting a suitable routing protocol and precise setting of its parameters is very important to get an efficient data transmission in the network.

The routing is defined as a process that performs special routing devices called routers to find the most efficient route to deliver the packet to the destination node along this route. The routers are governed by destination IP address which they read out from the IP packet header. To make their

Manuscript received November 28, 2013. This work has been supported by the project n. FR-TI4/580.

Pavel Masek[1], Dominik Kovac[2], Jiri Hosek[3] and Ondrej Krajsa[5] are with Department of Telecommunications, Brno University of Technology, Brno, Czech Republic (e-mails: xmasek12@phd.feec.vutbr.cz[1], xkovac23@phd.feec.vutbr.cz[2], hosek@feec.vutbr.cz[3], krajsao@feec.vutbr.cz[5]).

Mariya Pavlova[4] is with the Department of Communication Networks and Telecommunication Systems, Kalashnikov Izhevsk State Technical University, Izhevsk, Russia (e-mail: marya.m.pavlova@gmail.com[4]).

doi: 10.11601/ijates.v2i3.69

decision about next packet routing accurate and correct, the routers must learn about routes to other networks. There are two methods how the router can learn a new route to other networks: statically and dynamically [1].

The static routing is typically used in small networks or in networks where the topology changes are not so frequent. It is therefore obvious that for the hybrid multimedia networks, the dynamic routing must be used. In case of dynamic routing, the information sharing between routers is realized through a routing protocol. The classification of dynamic routing protocols is shown in Fig. 1.

The information exchanged among routers (including the distance to other networks) enables to create and update routing tables on every router in the network [1].

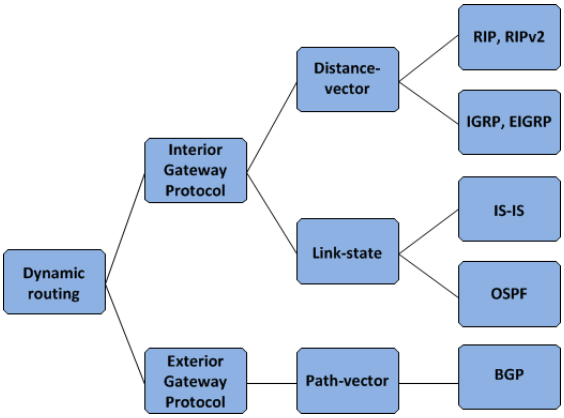


Fig. 1. Dynamic routing protocols

The essential part of the dynamic routing is the routing algorithm. According to the used routing algorithm, the routing protocols are divided into two large groups:

- **Distance-vector protocols**: These protocols are also referred to as Bellman-Ford or Ford-Fulkerson routing protocols.
- **Link-state protocols**: These routing protocols generate complete network map called Link State Database. On the top of this database, the best route is calculated by the help of SPF (Shortest Path First) algorithm.

In this paper, we are focusing on two the most famous routing protocols for local networks – RIP (Routing Information Protocol) and OSPF (Open Shortest Path First) to satisfy the requirements of hybrid multimedia networks.

1) RIP (Routing Information Protocol)

The RIP (Routing Information Protocol) protocol belongs to the class of routing protocols which use a distance vector as a key parameter. The operation of RIP is based on the periodic exchange (so-called triggered updates) of complete routing tables between neighboring routers. The exchange is performed every 30 seconds or when the network topology is changed.

A router that detects a change sends immediately the triggered update to the neighboring routers and this information about the topology changes is disseminated further [2]. There are two versions of the RIP protocol. The so-called RIPv1 is formally defined in RFC 1058 [3]. The second expanded version, RIPv2, is described in RFC 2453 [4].

2) OSPF (Open Shortest Path First)

The OSPF (Open Shortest Path First) protocol is an open protocol which is more suitable for larger networks [5]. Its specification is available as RFC (Request For Comments). The first version of OSPF is described in RFC 1131 [6]. The second version (OPSFv2) is described in RFC 2328 [7]. Within the introduction of IPv6 (Internet Protocol version 6), the third version (OSPFv3) was defined in RFC 2740 [8].

The functional principle of OSPF protocol is based on the flood-based dissemination of LSA (Link State Agreement) messages which contain information about the networks to which the router is connected [2].

B. Quality of Service in hybrid multimedia networks

The requirement for ensuring demanded service quality is directly related to the concept of QoS which means an effort to comply with predefined parameters during the data transfer. QoS assurance is very important, especially, for VoIP or videoconferencing services which are strongly time-sensitive and their popularity is still growing.

There are different solutions and network architectures that provide a classification of network traffic into certain classes with specific priority. Quality of service can be implemented in different network layers. The most commonly used QoS implementations are working on data-link and/or network layer [9].

The data-link layer QoS solutions are connected with the ATM (Asynchronous Transfer Mode) or Ethernet technologies. Considering the QoS implementation on network layer, there are two well-known mechanisms – Integrated Services (IntServ) and Differentiated Services (DiffServ) [10].

Since the IntServ has several practical limitations, it has never been widely implemented. Therefore, currently the most successful QoS technology in IP networks is the DiffServ. One of the fundamental characteristics of DiffServ mechanism is that all network traffic processing, which includes classifying and measuring the intensity of the network stream, packet marking and sorting and packet forwarding, is made at the edge routers [11]. Fundamental principle of DiffServ mechanism, where individual data flows are aggregated into several classes based on the service type, is shown in Fig. 2.

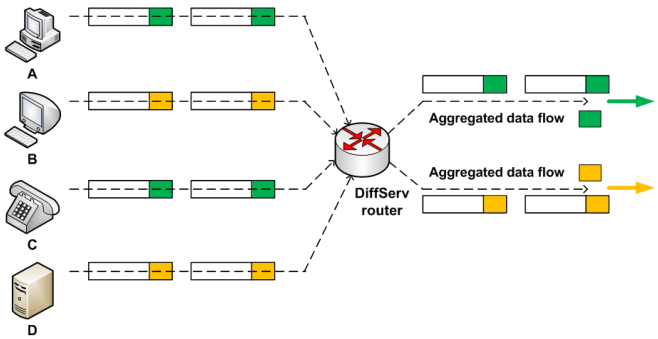


Fig. 2. Aggregation of network flows within DiffServ technology

Each packet entering the network is signed by a special mark – DSCP (DiffServ Code Point) stored in IP header, which expresses how other routers should process each specific packet. This marking process is performed only at

the network entrance (edge router) [12].

The DSCP identifier is sufficient to classify network traffic into classes, so there is no need for a signaling protocol to create a reservation status, as the IntServ mechanism does. Furthermore, since the traffic is aggregated and then processed within classes, the need for a complex sorting and planning of network resources for the individual data streams is eliminated [13, 14].

The simplicity and low overloading of network devices are among main advantages of DiffServ [15] mechanism and also a reason why this technology has been quite popular so far [16].

C. *Virtual Local Area Networks*

Another way how to classify network traffic is by the help of VLAN (Virtual Local Area Network). Virtual network is a group of network devices that are arbitrarily interconnected in computer networks and behave as if they are on the same physical medium which is separated from the rest of network [17, 18]. The VLAN concept is shown in Fig. 3.

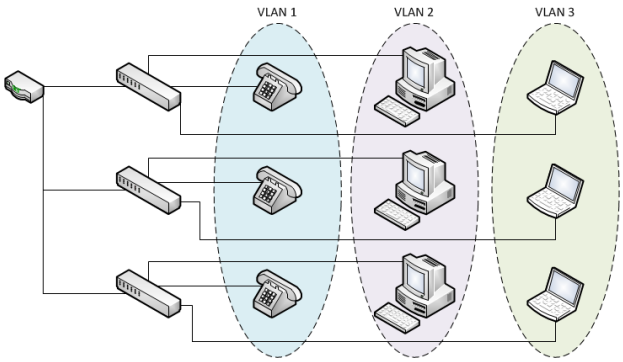


Fig. 3. Network traffic classification into VLANs

Virtual networks allow to interconnect the remote devices and separate these devices from others virtual networks at the same time. The LAN segments where the broadcast traffic is spread are not limited by physical connection to the router but may be defined arbitrarily as needed. Virtual networks can divide a large network into logical units that act as independent local networks. The segmentation of networks is necessary for the limitation of the broadcast domain and also because of security and administrative reasons. A suitably divided network is easier to handle the amount of traffic and network load [17, 18].

Each virtual network is defined by the identifier (so called color). The packets may be transmitted from one segment to another only if the packet contains of the same identifier. The great advantage of VLANs is a possibility to change the logical network topology without any need to modify the physical position of the devices (routers, switches) or their addresses. The segmentation of LANs into multiple logical segments can be performed according to their functionalities, user type, operated services etc.

The VLAN identifier (tag) is attached to Ethernet frame. The tag is used by a destination device for the detection if the source device belongs to the same VLAN network.

The quality of service in VLAN networks is provided through the QoS Match VLAN mechanism. This mechanism allows to split and to prioritize the network traffic based on VLAN identifier. E.g. individual VLAN can be created for VoIP or VoD service and prioritize these VLANs. The prioritization of selected VLANs guarantees that required quality of service is assured [18]. The most commonly used VLAN standard is the tagging protocol IEEE 802.1Q (see the structure IEEE 802.1Q frame in Fig. 4) [19].

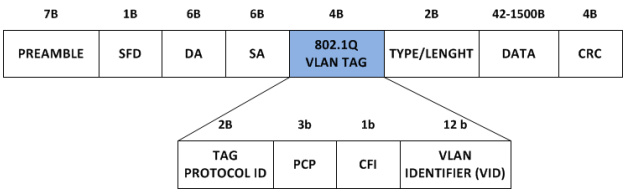


Fig. 4. Structure of framework IEEE 802.1Q

III. SIMULATION MODEL

To verify the influence of the transmission technologies, routing protocols and QoS mechanisms on the performance of the multimedia network and the final quality of the service provided, the OPNET Modeler ver. 17.5 simulation environment was chosen.

A. *Simulation of PON*

PON (Passive Optical Network) is one of communications technologies which has very good performance characteristics and thus is a good candidate for hybrid multimedia networks. Even though the costs for PON's deployment are higher in comparison with traditional IP networks, the popularity of these networks is increasing rapidly [20].

The main goal of the simulation of passive optical network was to verify the ability to serve as a communication platform for hybrid multimedia networks. The demonstration was compiled in a single passive optical network. The network model developed in OPNET Modeler (see Fig. 5) was built with respect to a logical context. Therefore, it does not reflect real distances between nodes. It was based on a simple tree topology where the network consists of one OLT (Optical Line Termination), one passive optical splitter and two ONUs (Optical Terminal Units).

Network nodes are interconnected with an optical link with maximal throughput 1 Gbps. The parameters of data sources connected to the ONUs were varied during the simulation as needed. However, the access network topology remained the same.

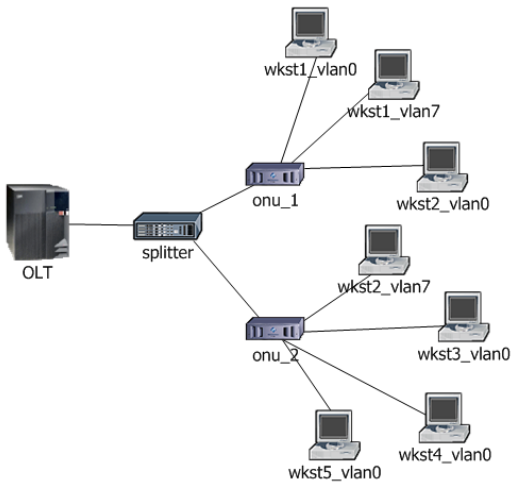


Fig. 5. PON topology

1) OLT

The primary task of OLT is to calculate the size of the grant for each ONU unit in each cycle. The number of units is manually set. The size of grants allocated is calculated either statically or dynamically. In static mode, the same grant's size is set for all ONUs. In dynamic mode, the size is calculated on the basis of requests sent by ONUs. The physical address of OLT is set to 99. The OLT may connect up to eight terminal units [21, 22].

2) Splitter

Passive optical splitter is intended to simulate the behavior of such a device in practice. The splitter is receiving the packets on point-to-point interfaces and forwarding them via all point-to-point transmitters to the destination node. This procedure is simulated by two processor blocks: *proc_tx* and *proc_rx*. The processors contain simple process models where only two states *init* and *receive_send* take care about the functionality of the splitter. The splitter is designed to provide a split ratio of 1:8, but can be easily expanded to 1:N. The splitter itself is generating a certain traffic delay. This is due to the calculations for copying the packet to the individual transmitters. It is necessary to consider this delay when the attenuation for specific connection between OLT and splitter is calculated. The delay is in the range of 1 to 4 microseconds. The value depends on current data stream [21].

3) ONU

ONU receives all data sent from the OLT. Based on the data type and their addresses, the ONU will decide whether they are going to be processed or discarded/forwarded. In the developed simulation model, ONU receives only the GATE messages which inform about the start time of the grant and its length. It is needed to set the attributes of the physical address for each ONU. The ONU model includes eight queues. Each queue processes received data within defined priority (weight), see the parameters for all traffic classes in TABLE I. Data classification is performed based on different point-to-point transmitters/receivers marked as *eth_rx_vlanX* or *eth_tx_vlanX*. In practice, this classification procedure is ensured by a specialized device (packet classifier) or the VLANs can be used [21].

TABLE I
TRAFFIC CLASS FOR EACH QUEUE

Queue	Weight	Type traffic
queue 0	**0.30**	**real-time services - telephony and video-tele.**
queue 1	**0.28**	**real-time services - video streaming**
queue 2	**0.11**	**signalization**
queue 3	**0.09**	**network management**
queue 4	**0.07**	**critical data**
queue 5	**0.06**	**critical data**
queue 6	**0.05**	**best-effort**
queue 7	**0.04**	**less important services**

The TABLE II shows the basic set of simulation parameters. The length of one cycle was set to 188 µs. This time is the tradeoff between the cycle length that can handle larger amounts of data and low delay. The data rate of all links in the network was set to 1 Gbps which is in accordance with the standard EPON. In order to provide higher simplicity, optical links generate no delay. BER (Bit Error Rate) for all simulated optical fibers was set to 10^{-10} which is one of the typical values.

TABLE II
NETWORK PARAMETERS IN THE SIMULATION

Parameter	Value
Number of ONU	**2**
Transmission speed of optical link	**1 Gbps**
Transmission speed of LAN	**1 Gbps**
Optical delay of line	**0 s**
BER of optical fiber	**10^{-10}**
Number of classes for QoS	**8**
Size of cycle	**188 µs**
Guard interval between grants	**1 µs**
Guard interval between cycles	**7 µs**

B. Simulation of TCP/IP Network

As an alternative to the passive optical network, classic TCP/IP network has been selected. This type of network forms the platform for the vast majority of current networks, hence it is also a good candidate for a hybrid multimedia network. Low costs for deployment and operation are the main advantage of TCP/IP- based networks. Moreover, it is a reliable and well recognized communication platform.

The goal of performed simulation was to identify the network parameters that have the greatest impact on the performance of multimedia network operated over TCP/IP platform. Another task was to find optimal settings of these parameters. A side target was to define some recommendations to achieve the best network KPIs.

In more detail, the influence of the selected routing protocol, the implementation of QoS mechanism and the level of core links utilization were analyzed. As the output, the network characteristics (such as end-to-end delay, jitter and throughput) were discussed. The topology of developed simulation model is shown in Fig. 6.

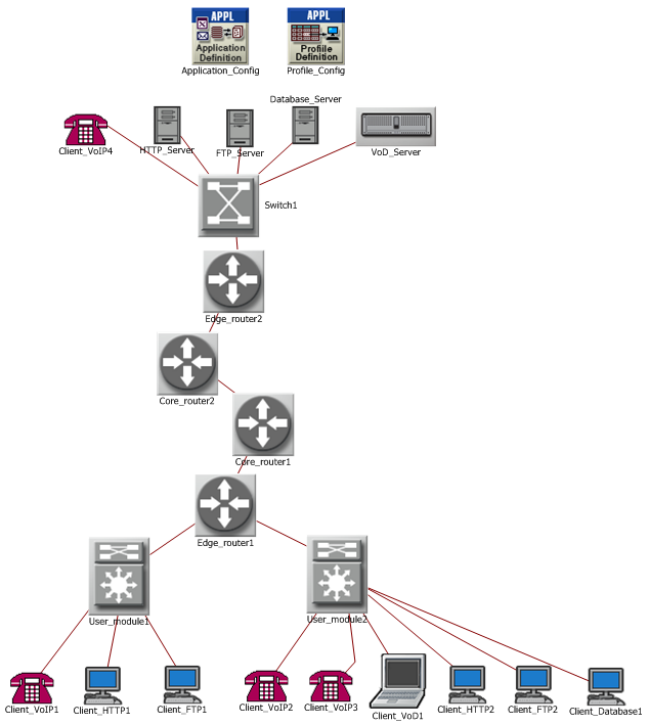


Fig. 6. TCP/IP based hybrid multimedia network

The backbone of the designed topology is composed of four routers (two edge routers and two core routers) that simulate general WAN (Wide Area Network) to provide the connections between user workstations and application

servers. The end terminals are connected to the access network via the subscriber module, whose task is to process incoming or outgoing calls. In addition, this module also provides a functionality of standard switch.

The structure of applications and services operated in current multimedia networks is very variable, so to provide highly realistic scenario, the simulation model defines the following applications:

- voice calls (VoIP),
- videoconferencing,
- web browsing (HTTP),
- database access (SOAP),
- file transfer (FTP).

All network traffic is started in 11 seconds after the beginning of simulation. This time offset is long enough to complete all initial configuration processes that could affect the simulation results. Separate profile that defines the amount of data generated and the time characteristics, was defined for each network application. The entire simulation took 10 minutes which is sufficient time to recognize the required network characteristics.

IV. Analysis of Simulation Results

As already mentioned in the previous chapter, the simulation of hybrid multimedia network was divided into two parts – PON and TCP/IP.

A. Discussion on Simulation Results – PON

The goal of the first part of performed simulation was to evaluate the delay of transmitted data and throughput in the designed passive optical network. Because of the high computational demands, the simulation time was set to 0.04s and data traffic has been generated by the stations for 0.01 seconds from the beginning of the simulation. Despite such a short simulation time, this setting is sufficient for the required verification of monitored parameters.

Fig. 7 shows the throughput of the formed network, respectively throughput between the ONU and a client station. The graph shows that optical network is utilized at 100% as the line speed was set to 1Gbps.

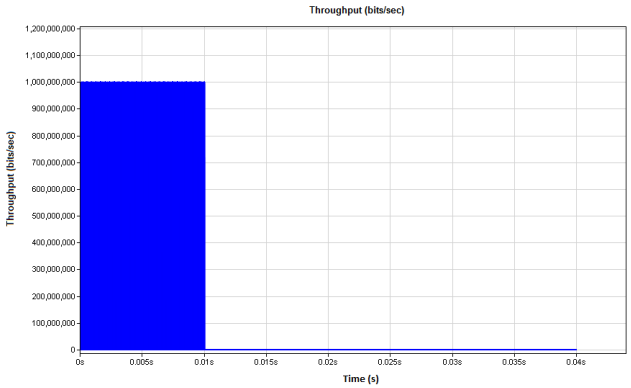


Fig. 7. Throughput between ONU and the client station

The communication delay for data stream between the client and OLT is shown in Fig. 8. The client station generates the traffic for simulation time of 0.01 seconds. The largest delay contributors are the splitter and ONU. The splitter spends certain time by copying all incoming packets and forwarding them to all transmitters. ONU has to process each packet and decide about the subsequent procedure (forward/drop). On the other hand, the optical link's contribution is very low.

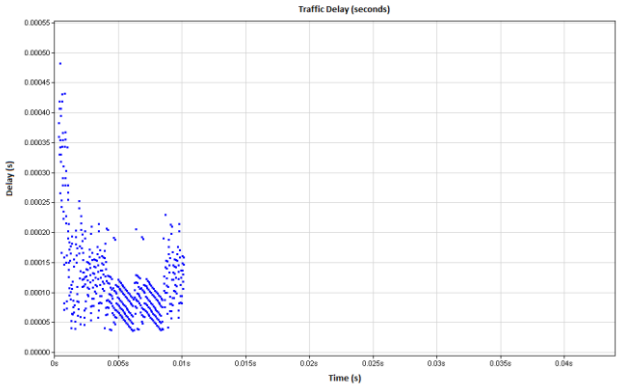


Fig. 8. Traffic delay

The simulation of passive optical network can be summarized so that this communication platform provides very low end-to-end delay, even for maximal utilization of network link. Therefore, it can be said that PON is suitable for the transmission of multimedia data that have strict requirements on delay and jitter.

Despite the very good results obtained from the simulations, PON has several significant drawbacks:

- High deployment costs – comparing to IP networks, the costs of construction and operation are much higher.
- Security (wiretap) – all wavelengths are propagated to all end units which can be tapped for the attacker. Therefore, it is necessary to implement the encryption method.
- Relatively high attenuation – attenuation caused by passive optical splitters in the distribution network.

B. Discussion on Simulation Results – TCP/IP

In order to clearly present the influence of various mechanisms on network performance, the generated traffic is set in the way that the core-network links are utilized by 80% in average. It is generally understood that in lightly loaded networks it is not necessary to solve performance analyzes or implement additional mechanisms (e.g. QoS), as the implementation of these mechanisms has practically no impact.

1) Analysis of routing protocols

The influence of routing protocols on the network traffic characteristics was analyzed as the first. We have developed two completely identical scenarios (see Fig. 6) where the routing protocol (RIP or OSPF) was the only difference. The obtained results clearly showed that even though OSPF generates higher traffic overhead and greater procedural delays within the network nodes (see Fig. 9), it achieves significantly better results in the case of high-loaded network. It is mostly because it can better solve the network congestion issue by the integrated network load balancing mechanism and faster convergence. So, the OSPF allows achieving higher throughput and lower delay for network services.

Fig. 10 shows the progress of the total download time for the 10MB file from the FTP server to the client station. The graph clearly shows, that in case of RIP, the download time is increasing almost linearly as the network load is growing.

On the contrary, the OSPF routing protocol can provide almost constant download response time even at high network load.

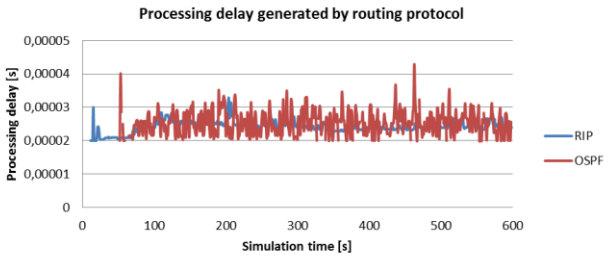


Fig. 9. Processing delay generated by routing protocols on Core_router2

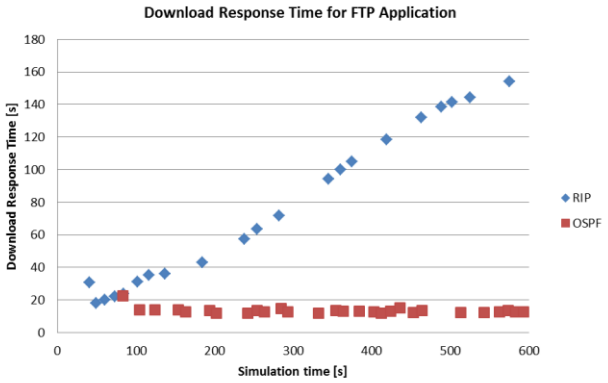


Fig. 10. Download response time for FTP application on Client_FTP1

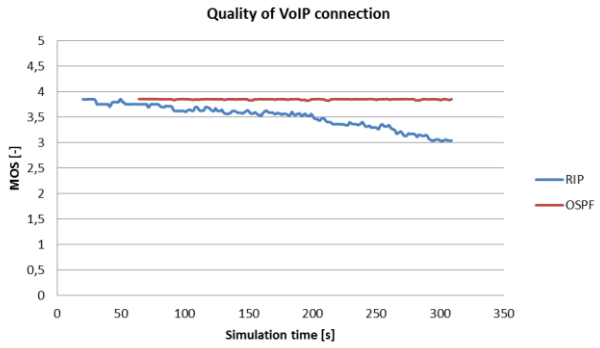


Fig. 11. Call quality between terminals Client_VoIP3 and Client_VoIP4

The OSPF protocol also provides better results for VoIP service. Fig. 11 shows the MOS (Mean Opinion Score) based quality evaluation of VoIP call between the VoIP terminals Client_VoIP3 and Client_VoIP4. The RIP scenario provides evident quality degradation when the network load is high. On the other hand, OSPF can maintain the connection quality at a very stable level. The maximal value MOS value of 3.85 corresponds to a very good quality considering used codec G.729.

2) Analysis of QoS Mechanism

Although the OSPF clearly showed better performance compared to the RIP protocol, to meet the strict requirements of real-time traffic (voice, video), the network extension by QoS mechanisms is needed. This implementation requirement brings certain contribution only for highly loaded networks for sure.

Basic scenario with OSPF protocol was duplicated and the DiffServ mechanism was implemented into this new scenario. Fig. 12 shows the end-to-end delay for VoIP call realized between the VoIP_Client3 and VoIP_Client4. In the RIP scenario, the E2E delay is increasing rapidly and approximately in the middle of the call reaches a critical value of 150ms for which the end-user quality is very degraded. Among scenarios, OSPF and OSPF + QoS is not so noticeably different, but with closer examination was found out that the implementation of QoS mechanisms to achieve greater stability values of delay and jitter value is for scenarios with lower QoS mechanism, which is very important for the realization of high-quality multimedia transmissions in real-time.

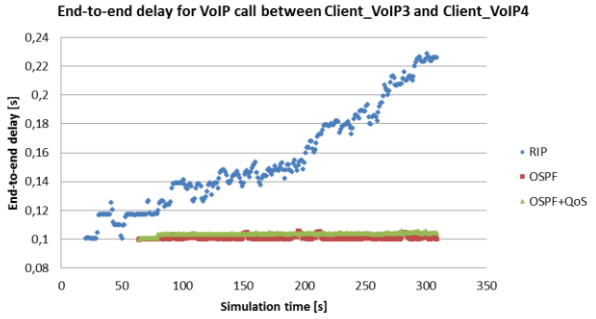


Fig. 12. E2E delay for VoIP call between VoIP_Client3 and VoIP_Client4

Implemented QoS mechanism does not have a favorable effect only on E2E delay or jitter, but also on the overall network throughput. With the classification process and separate treatment of each traffic class according to their requirements, the network resources are efficiently utilized. This prevents the cumulative burden of network nodes and links from retransmissions of packets dropped inside the congested routers.

This phenomenon is clearly shown in Fig. 13, where the utilization of the link between the routers Core_Router1 and Core_Router2 is depicted. In case of the RIP scenario, all available network resources are exhausted which affects performance of all services. In contrast, DiffServ technology most effectively manages network resources and allows keeping the link utilization below 80%.

V. Conclusion

In order to define appropriate communication technology for hybrid multimedia network, a routing protocol and any other supporting mechanisms ensuring the required quality of service, the performance analysis was carried out in a simulation environment OPNET Modeler.

Two communication technologies – passive optical network and classic TCP / IP network were evaluated. Even though both these technologies are quite different and cannot be directly compared, the emphasis was put on the identification of the key network elements or parameters that have the greatest effect on the basic QoS characteristics like E2E delay, jitter and throughput.

Moreover, in case of TCP/IP network, the influence of routing protocol and implementation of QoS mechanism on the overall network efficiency has been studied. Achieved simulation results and their analyses have been discussed in the previous chapter.

Overall, the performed analysis can be summed up so the proper choice of the routing subsystem for multimedia networks is influenced by many factors. It is possible to define a set of core network elements and parameters/subsystems that contribute to the overall network

performance the most, but, at the same time, it is clear that particular network topology, level of network load and composition of operated applications have to be taken into account. Moreover, the total deployment and maintenance costs play a significant role.

REFERENCES

[1] Kalyan, G., P., S., Prasad, D., V., V.: Optimal selection of Dynamic Routing Protocol with real time case studies. 2012 International Conference on Recent Advances in Computing and Software Systems [online]. IEEE, 2012, s. 219-223 [cit. 2014-01-13]. DOI: 10.1109/RACSS.2012.6212727. Available on: http://goo.gl/Rm3LGM.

[2] Thorenoor, S. G. (2010). Dynamic Routing Protocol Implementation Decision between EIGRP, OSPF and RIP Based on Technical Background Using OPNET Modeler. 2010 Second International Conference on Computer and Network Technology, 191–195. doi:10.1109/ICCNT.2010.66.

[3] Hendrick, C. (1988). Routing Information Protocol - RFC 1058, 1–33. URL: http://tools.ietf.org/rfc/rfc1058.txt.

[4] Malkin, G. (1998). RIP Version 2 - RFC 2453, 1–40. URL: http://tools.ietf.org/rfc/rfc2453.txt.

[5] Fitigau, I.; Toderean, G., "Network performance evaluation for RIP, OSPF and EIGRP routing protocols," International Conference on Electronics, Computers and Artificial Intelligence (ECAI), 2013 , vol. 1, no. 1, pp.1-4, 27-29 June 2013.

[6] Moy, J. (1989). OSPF - RFC 1131, 1–107. URL: http://tools.ietf.org/pdf/rfc1131.pdf.

[7] Moy, J. (1998). OSPF Version 2 - RFC 2328, 1–245. URL: http://tools.ietf.org/pdf/rfc2328.pdf.

[8] Colton, R., Ferguson, D., Moy, J. (1999). OSPF for IPv6 - RFC 2740, 1–93. URL: https://tools.ietf.org/rfc/rfc5340.txt.

[9] Lan, L., Li, L., & Jianya, C. (2012). A Multipath Routing Algorithm Based on OSPF Routing Protocol. 2012 Eighth International Conference on Semantics, Knowledge and Grids, 269–272. doi:10.1109/SKG.2012.7.

[10] Szigeti, T., Hattingh, C. End-to-End QoS Network Design: Quality of Service in LANs, WANs, and VPNs. Cisco Press, 2004.

[11] Hosek, J. Nové metody zajištění kvality služeb v datových sítích: doktorská práce. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 159 s., 2011.

[12] Park, I. K. QoS in Packet Networks. Springer, 2005.

[13] Velte, J. T. Cisco: A Beginner's Guide. McGraw-Hill, 2001.

[14] Flannagan, M. E. Administering Cisco QoS for IP Networks. Syngress Publishing, 2001.

[15] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., Weiss, W.: An Architecture for Differentiated Service, RFC2475, 1998.

[16] Marchese, M. QoS Over Heterogeneous Networks. Wiley Publishing, 2007. ISBN 9780470017524.

[17] Cisco Systems Inc. Catalyst 4500 Series Switch Cisco IOS Software Configuration Guide [online]. 2004. Available on: http://www.cisco.com/en/US/docs/switches/lan/catalyst4500/12.2/25ew/configuration/guide/OL_6696.pdf.

[18] Cisco Systems Inc. QoS: Classification Configuration Guide, Cisco IOS XE Release 3S [online]. 2013. Available on: http://goo.gl/VCKRHy.

[19] Cisco Systems Inc. Inter−Switch Link and IEEE 802.1Q Frame Format [online]. 2006. Available on: http://www.cisco.com/image/gif/paws/17056/741_4.pdf.

[20] Sifta, R., Munster, P., Krajsa, O., Filka, M. Simulation of bidirectional traffic in WDM- PON networks. Przeglad Elektrotechniczny, 2014, roč. 90, č. 1, s. 95-100. ISSN: 0033- 2097.

[21] Lafata, P., Vodrazka J. Pasivní optická síť GPON. Access server [online]. 23.05. 2009, č. 2009050002. ISSN 1214-9675. Available on: http://access.feld.cvut.cz/view.php?cisloclanku=2009050002.

[22] ITU-T: G.983.1 - Broadband optical access systems based on Passive Optical Networks (PON). [online], Internet: http://www.itu.int/rec/T-REC-G.983.1-200501-I/. ITU-T, January 2005.

SIX research centre
sensor, information and communication systems

enterprise europe network
Business Support on Your Doorstep

Horizon 2020 Partnership Establishment Workshop

in cooperation with

Enterprise Europe Network

The SIX Research Centre was established to support the innovation potential of companies which actively exploit communication, information and sensor technologies in different areas of life.

The SIX Research Centre exploits the background of the university to obtain new knowledge and to create novel solutions by its own fundamental research. New knowledge is subsequently applied to novel products and novel services of partner companies.

Using the potential of Horizon 2020 programme we would like to bring together and also look for partners that are interested in mutual cooperation in the following areas of research:

- Audio and video processing in communication network,
- Interactive gestural human-machine interfaces,
- Development of biometric systems,
- Processing of biomedical signals,
- 3D acquisition and display,
- Design and measurement of electroacoustic and audio devices,
- Communication, information and telemetric systems,
- Development of new algorithms and protocols for data networks,
- IP telephony and multimedia applications,

and other related areas.

All participants are invited to present their research activities and ideas during the 37th International Conference on Telecommunications and Signal Processing (TSP), which will be held during **July 1-3, 2014**, **Berlin**, **Germany** (tsp.vutbr.cz).

For further information please contact:
Assoc. Prof. Dr. **Jaroslav Koton**
Coverged Systems, SIX Research Centre, vice-head
Technicka 12
616 00 Brno, Czech Republic
email: **koton@feec.vutbr.cz**
web: **six.feec.vutbr.cz**

ISSN 1805-5443