# JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY

4/2012



*(Contents Continued on Back Cover)*

ISSN 1509-4553 on-line: ISSN 1899-8852



Warsaw 2012

Circulation: 300 copies

Sowa – Druk na życzenie, www.sowadruk.pl, tel. 22 431-81-40

JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY

# Preface

This issue of the *Journal of Telecommunications and Information Technology* contains ten papers that cover diverse problems related to either information society, socio-economic impacts of telecommunication and Internet, to network control and security, or various issues addressed to design and application wireless communication networks and sensing systems.

The first paper is devoted to social penetration of Information and Communication Technologies (ICT). Jan Grzegorek and Andrzej P. Wierzbicki in the paper entitled *Multiple Criteria Evaluation and Ranking of Social Penetration of Information Society Technologies* consider and compare two approaches to the multiple criteria ranking of social penetration of ICT, i.e., the traditional one and so-called objective ranking. Another issue presented in this paper is the concept of dynamic ranking, a systematic presentation and prediction of the change of ranking in time.

Mariusz Kamola in the paper titled *Who is Asking and for What: WHOIS Traffic Analysis* presents the analysis of WHOIS requests for about one-year period. The focus is put on the popularity of requested names. The author claims that WHOIS traffic can be roughly classified into systematic scanning of domain names and individual low-volume activity, mostly targeting on very popular names.

The recent trends in the regulation of telecom services regarding the diffusion of broadband ones is discussed by Cristine Vianna Rauen in the paper *Competition and Diffusion of Telecommunication Services: The Multimedia Communication Services in Brazil*. The regulations established in such countries like Chile and Korea that force the competition in order to expand the access to new forms of broadband services are compared with the Brazilian scenario. The conclusion is that the regulation mechanisms should be reinforced in the Brazilian broadband market.

The next two papers are devoted to network design and flow control. Michał P. Karpowicz discusses the selected aspects related to the control and security of network systems. The paper entitled *On the Design of the TCP/AQM Traffic Flow Control Mechanisms* deals with the TCP/AQM system design and its influence on the performance of the network. Author presents the way the commonly applied TCP/AQM design procedures may give rise to mechanisms that are prone to attacks, discreetly moving the network traffic flow away from the desired operating point. Moreover, there is a short discussion concerning the countermeasures that can be taken to reduce these effects.

Najib A. Odhah *et al.*, in their paper *Low Complexity Greedy Power Allocation Algorithm for Proportional Resource Allocation in Multi-User OFDM Systems* start with the premise that Multi-User Orthogonal Frequency Division Multiplexing (MU-OFDM) is an efficient technique for achieving high downlink capacity in high-speed communication systems. The novel proportional rate-adaptive resource allocation algorithm, Greedy Power Allocation (GPA), for MU-OFDM is proposed and described. The simulation results presented in this paper confirm that the GPA algorithm performs better than the competitive solution described in the literature.

The profitability of application of simulation methods for topological models to analyze and design of information systems is discussed in the paper *UML Simulation of a Topology Configuration Model*. Zbigniew Zieliński, Andrzej Stasiak and Włodzimierz Dąbrowski claim that using the UML extensions and the UAL language allows not only to build a topological model for a software, but also to perform efficient simulations of topological models. The presented discussion is confirmed through simulation and practical examples.

The next two papers deal with various application of sensors. Wojciech Szynkiewicz in the paper titled *Skill-based Bimanual Manipulation Planning* addresses the issues associated with robot systems equipped with diverse sensors, such as vision, force/torque or tactile sensors. The focus is put on the specification and utilization of manipulation skills to facilitate programming of bimanual manipulation tasks. Manipulation skills constitute an interface between low level constraint-based task specification and high level symbolic task planning. Rubik's cube solving problem is presented as an example of a 3D manipulation task using the two-arm robot system.

Igor Goncharenko, Marian Marciniak, Alexei Konojko and Vitaly Reabtsev in their paper *Optimizing the Structure of Vector Bend and Strain Sensor on the Base of Three-Core Microstructured Fiber* present an optical sensor designed to measure a direction, values and localization of bends and stresses in building structures. The technology and the architecture of the sensor is described. Furthermore, the optimization of the sensitive element parameters depending on the application is proposed and discussed.

The following two papers are devoted to the wireless sensor networking. Anna Felkner in the paper *How the Role-Based Trust Management Can Be Applied to Wireless Sensor Networks* addresses the problem with assuring security in the wireless sensor networks. The Author focuses on the important component of all the security systems – trust management. The application of common Role based Trust management languages (RT) to network formed by wireless sensors is proposed and discussed. These languages are used to implement security policies and credentials in decentralized, distributed access control systems.

The last paper deals with the application of wireless sensor networks to an environmental monitoring, and the education activities related to wireless technologies and applications. Sandro Radicella, Ryszard Strużak and Marco Zennaro in the paper *Educating on Wireless Solutions for Environmental Monitoring* provide the information about the International School "Sustainable Wireless ICT Solutions for Environmental Monitoring" that was organized by the International Center for Theoretical Physics (ICTP) in a collaboration with a few other entities. This school is aimed at exposing young scientists from around the world to the newest wireless solutions for environmental monitoring. The Authors start with the general information about educational activities. Furthermore, they present the program of the school and the conference.

We wish our Readers an interesting reading time.

Ewa Niewiadomska-Szynkiewicz
Guest Editor

Paper

# Multiple Criteria Evaluation and Ranking of Social Penetration of Information Society Technologies

Jan Grzegorek[a] and Andrzej P. Wierzbicki[b]

[a] TP SA, Warsaw, Poland
[b] National Institute of Telecommunications, Warsaw, Poland

**Abstract—The paper presents a comparison of two ways of multiple criteria ranking of social penetration of information society technologies (ICT): the traditional one and the so-called objective ranking, illustrated on the example of Network Readiness Index. It is shown that objective ranking stresses a balanced development more than the traditional one. Another issue presented in this paper is the concept of dynamic ranking, a systematic presentation and prediction of the change of ranking in time. This is illustrated on the example of regions of three European countries: Finland, Italy and Poland.**

***Keywords*—*dynamic ranking, ICT social penetration, multiple criteria ranking, objective ranking.***

## 1. Introduction

Classical ranking is a classification in the order of numerical values assigned to a specific index, criterion or indicator. *Multicriteria* ranking assumes some *aggregation* (called technically scalarization) of many such indexes, criteria or indicators. Usually, such an aggregation is fully subjective; only recently, see [1], an aggregation and *ranking that is as objective as possible* (none can be fully objective) was proposed. Thus, one of the goals of this paper is to compare different aggregations used for multiple criteria ranking of ICT development in different countries and regions. This is illustrated using Network Readiness Index (NRI) [2], [3] of World Economic Forum. The objective ranking might give different conclusions even if data prepared by experts might remain subjective.

However, rankings are usually perceived as static classification – with some repeated rankings changing in time. Therefore, we also illustrate in this paper the issue of dynamic ranking – a systematic presentation and prediction of the change of ranking in time – using the example of regions or provinces of three European countries: Finland, Italy and Poland. Such predictions might give additional arguments for decision-makers to plan and take appropriate action in the future.

### 1.1. Data

Contemporary evaluations of social or *socio-economic penetration of information society technologies*, often called ICT, information and communication technology, are based on many statistical measures, indexes or criterions. Example might be the data from Eurostat portal, e.g.:

- households with access to the Internet at home,
- households with broadband access to the Internet,
- individuals regularly using the Internet,
- individuals who ordered goods or services over the Internet for private use.

World Economic Forum uses much broader set of indexes and their evaluations, described in further sections. However, we shall see that such data supported only by expert evaluations and not by statistics might be biased. In Poland, we can also use data from the Local Data Bank of GUS (Polish Central Statistical Office), e.g.:

- schools equipped with computer laboratories, schools with broadband access to the Internet, etc.,
- households with computers,
- households with the Internet,
- households with mobile telephone.

Such data is typically presented according to a territorial unit classification.

### 1.2. NUTS Clasification

The Nomenclature of Territorial Units for Statistics (NUTS) classification[1] is a hierarchical system for dividing the economic territory of the EU up for the purpose of the collection, development and harmonization of EU regional statistics. It helps in diverse socio-economic analyses of regions, subdivided into three classes.

- NUTS 1: major socio-economic regions,
- NUTS 2: basic regions for the application of regional policies,
- NUTS 3: small regions for specific diagnoses.

[1] See also Regulation (EC) no. 1059/2003 of the European Parliament and of the Council of 26 May 2003 on the establishment of a common classification of territorial units for statistics (NUTS).

Such a classification is also used for framing of EU regional policies:

- Regions eligible for aid from the structural funds (Objective 1) have been classified at NUTS 2 level.
- Areas eligible under the other priority objectives have mainly been classified at NUTS 3 level.
- The Cohesion Report has so far been mainly prepared at NUTS 2 level.

The average size of the NUTS administrative units is limited by the following population thresholds (see Table 1).

Table 1
Population threshold

| Level | Minimum | Maximum |
|---|---|---|
| NUTS 1 | 3 million | 7 million |
| NUTS 2 | 800 000 | 3 million |
| NUTS 3 | 150 000 | 800 000 |

while the smallest member states might be classified as one NUTS territorial unit of a population just larger than that of the state.

### *1.3. Regions Considered in the Paper*

We shall consider regions of Finland, Italy and Poland according to NUTS classification (see Figs. 1 and 2). Italy actually uses two-tier (NUTS 1 and 2) classification of its regions, which is listed in Table 2.

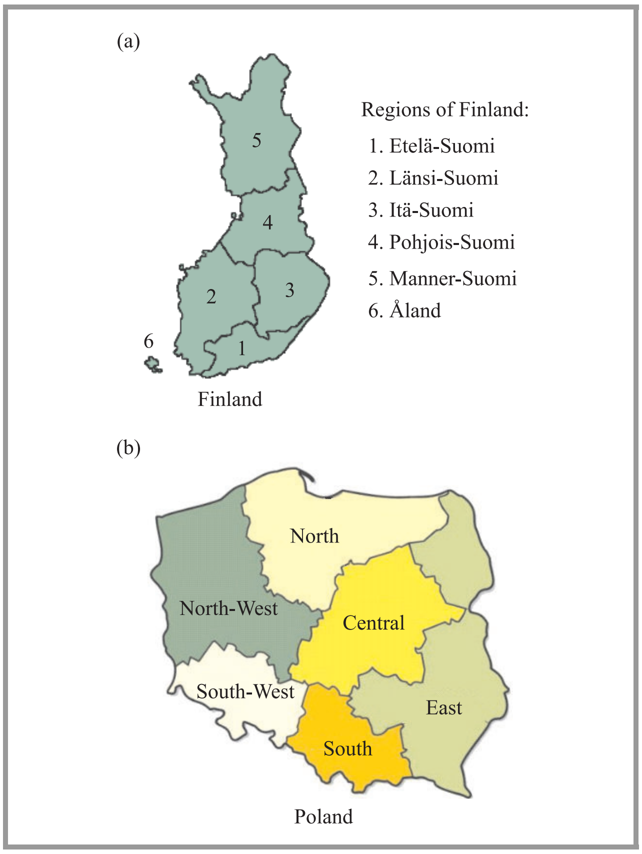



***Fig. 1.*** Regions of Finland (a) and Poland (b).


***Fig. 2.*** Regions of Italy.

Table 2
Regions of Italy by the Eurostat NUTS clasification

| NUTS 1 | NUTS 2 |
|---|---|
| North West | Aosta Valley, Liguria, Lombardy, Piedmont |
| North East | Emilia-Romagna, Friuli-Venezia Giulia, Trentino-Alto Adige/Südtirol, Veneto |
| Centre | Lazio, Marche, Tuscany, Umbria |
| South | Abruzzo, Apulia, Basilicata, Calabria, Campania, Molise |
| Islands | Sardinia, Sicily |

## 2. Traditional Ranking of World Countries

### *2.1. Network Readiness Index*

Network Readiness Index (NRC) is an overall measure of socio-economic penetration of information and communication technology (ICT), quite detailed and comprehensive but based on expert evaluation of diverse aspects of ICT [2]. Thus the relation of these evaluations to statistical data is not fully transparent, see below. However, the evaluations are fully accessible in the Internet, thus they can be analyzed in diverse ways.
These evaluations are the basis of the Global Information Technology Report (GITR) series produced by the World Economic Forum (WEF), as a part of WEF research on competitiveness. In this research, the Networked Readiness Index (NRI) was developed.

The WEF concentrates its analysis on economic and social impact of ICT. Although a positive impact of ICT diffusion on GDP growth of a country has been documented, estimates show that a 10% increase in mobile phone penetration is associated with a 1% growth in GDP (actually, the impact of the Internet penetration is even stronger). There are still open questions about using ICT to improve the condition of each individual, and about the impact of ICT on socio-political relations. Leaving these questions aside, we concentrate here on the methodology used by WEF.

### 2.2. The Method of Calculating the Ranking by WEF

Diverse indicators used in WEF report take diverse counting units and diverse numerical values. Therefore, the report uses normalization of indicators that are statistical or expert opinion based, to the values between 1 and 7, while 7 denotes the best result. For those indicators that are best when they increase (such as GDP), the used transformation is as follows:

$$6 \cdot \left( \frac{country\ score - sample\ \text{minimum}}{sample\ \text{maximum} - sample\ \text{minimum}} \right) + 1 \qquad (1)$$

For those indicators that are best when they decrease (such as unemployment) the used transformation is:

$$-6 \cdot \left( \frac{country\ score - sample\ \text{minimum}}{sample\ \text{maximum} - sample\ \text{minimum}} \right) + 7 \qquad (2)$$

The structure of Network Readiness Indicators is as follows. The following indicators are assessed subjectively by experts, later simply aggregated (by taking most simple averages after normalization):

**Environment**

1. Market environment
2. Political and regulatory environment
3. Infrastructure environment

**Readiness**

4. Individual readiness
5. Business readiness
6. Government readiness

**Usage**

7. Individual usage
8. Business usage
9. Government usage

Thus the method of calculating NRI is simple:

**Networked Readiness Index** = 1/3 **Environment subindex** + 1/3 **Readiness subindex** + 1/3 **Usage subindex**

**Environment subindex** = 1/3 Market environment + 1/3 Political and regulatory environment + 1/3 Infrastructure environment

**Readiness subindex** = 1/3 Individual readiness + 1/3 Business readiness + 1/3 Government readiness

**Usage subindex** = 1/3 Individual usage + 1/3 Business usage + 1/3 Government usage

However, the sub-subindexes called *pillars* in NRI computations consist of an aggregation (again, using simple averaging) of several indicators, selected by experts. Their numbers and changing in time (see the example of their structure in the Appendix). Data on the numbers of indicators used in diverse pillars in 2003, 2009, 2011 are presented in Table 3.

Table 3
Number of variables in the reports NRI 2003, 2009 and 2011

| Pillar | NRI 2003 | NRI 2009 | NRI 2011 |
|---|---|---|---|
| 1. Market environment | 9 | 14 | 10 |
| 2. Political and regulatory environment | 7 | 9 | 11 |
| 3. Infrastructure environment | 5 | 7 | 10 |
| 4. Individual readiness | 10 | 9 | 9 |
| 5. Business readiness | 5 | 10 | 8 |
| 6. Government readiness | 3 | 4 | 3 |
| 7. Individual usage | 4 | 5 | 8 |
| 8. Business usage | 3 | 5 | 8 |
| 9. Government usage | 2 | 5 | 4 |

Thus the final NRI score is an average of the three composing subindex scores, while each subindex score is an average of those of three composing pillars, but pillars are defined as an average of a changing number of normalized indicators. When using such an averaging, it is not only assumed that all index components give a similar contribution to national networked readiness, but also that the evaluation of pillars is done, as objectively as possible by the experts.

### 2.3. Examples of NRI Rankings

While Fig. 3 presents graphically an example of recent NRI ranking, Table 3 shows changes of NRI ranking and the position of Poland in this ranking during the recent years. We see that Poland is classified on the positions between 58 and 69, while Italy, for example, is classified on the positions between 38 and 51. A natural question that will be addressed in a further section is whether Poland has the chance to overtake Italy.

World Economic Forum also uses World Bank classification of countries into income groups. It is significant that the first twenty countries in the NRI ranking all belong to the highest income group [2].

In order to check methodological validity of simple average aggregation of diverse indicators used by WEF in NRI ranking, in the next section we use WEF data, but apply a different and more complex but also more objectively aggregated.

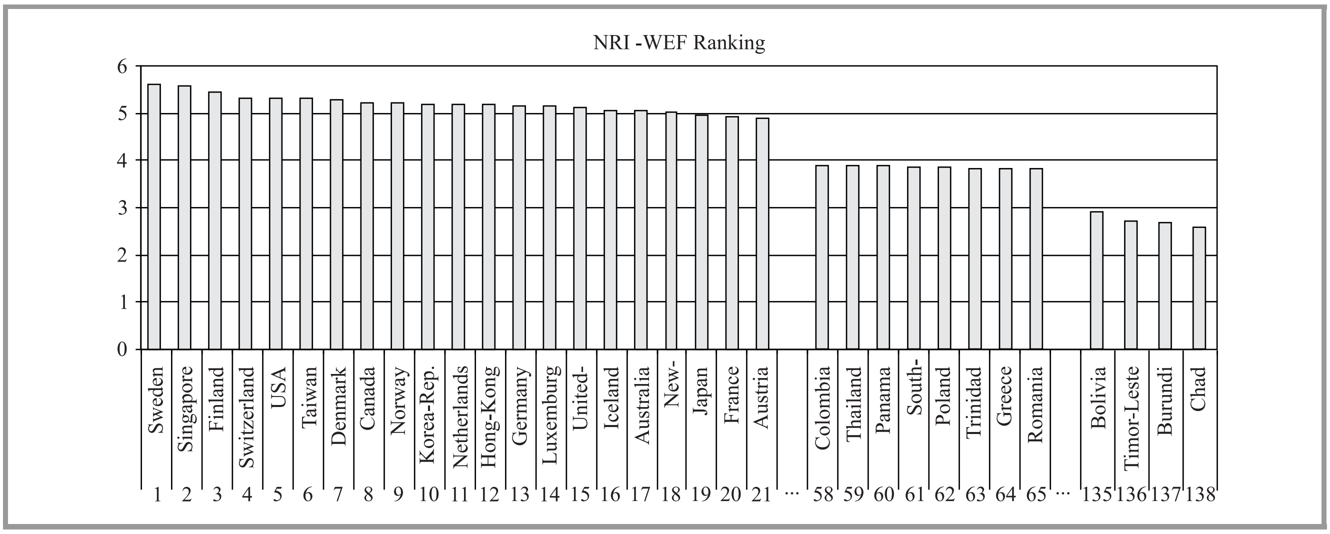



*Fig. 3.* Classification according to the NRI ranking.

Table 4
Position of Poland in NRI ranking in the last few years

| No. | Country | 2006/2007 | 2007/20082 | 2008/2009 | 2009/2010 | 2010/2011 |
|---|---|---|---|---|---|---|
| European Union | | | | | | |
| 1 | Sweden | 2 | 2 | 2 | 1 | 1 |
| 2 | Finland | 4 | 6 | 6 | 6 | 3 |
| 3 | Denmark | 1 | 1 | 1 | 3 | 7 |
| 4 | Netherlands | 6 | 7 | 9 | 9 | 11 |
| 5 | Germany | 16 | 16 | 20 | 14 | 13 |
| 6 | Luxembourg | 25 | 24 | 21 | 17 | 14 |
| 7 | United Kingdom | 9 | 12 | 15 | 13 | 15 |
| 8 | France | 23 | 21 | 19 | 18 | 20 |
| 9 | Austria | 17 | 15 | 16 | 20 | 21 |
| 10 | Belgium | 24 | 25 | 24 | 22 | 23 |
| 11 | Estonia | 20 | 20 | 18 | 25 | 26 |
| 12 | Malta | 27 | 27 | 26 | 26 | 27 |
| 13 | Ireland | 21 | 23 | 23 | 24 | 29 |
| 14 | Cyprus | 43 | 41 | 33 | 32 | 31 |
| 15 | Portugal | 28 | 28 | 30 | 33 | 32 |
| 16 | Slovenia | 30 | 30 | 31 | 31 | 34 |
| 17 | Spain | 32 | 31 | 34 | 34 | 37 |
| 18 | Czech Republic | 34 | 36 | 32 | 36 | 40 |
| 19 | Lithuania | 39 | 33 | 35 | 41 | 42 |
| 20 | Hungary | 33 | 37 | 41 | 46 | 49 |
| 21 | Italy | 38 | 42 | 45 | 48 | 51 |
| 22 | Latvia | 42 | 44 | 48 | 52 | 52 |
| 23 | Poland | 58 | 62 | 69 | 65 | 62 |
| 24 | Greece | 48 | 56 | 55 | 56 | 64 |
| 25 | Romania | 55 | 61 | 58 | 59 | 65 |
| 26 | Bulgaria | 72 | 68 | 68 | 71 | 68 |
| 27 | Slovak Republic | 41 | 43 | 43 | 55 | 69 |
| Other | | | | | | |
| 1 | USA | 7 | 4 | 3 | 5 | 5 |
| 2 | Korea, Rep. | 19 | 9 | 11 | 15 | 10 |
| 3 | Japan | 14 | 19 | 17 | 21 | 19 |
| 4 | China | 59 | 57 | 46 | 37 | 36 |
| 5 | India | 44 | 50 | 54 | 43 | 48 |
| 6 | Croatia | 46 | 49 | 49 | 51 | 54 |
| 7 | Turkey | 52 | 55 | 61 | 69 | 71 |

Source: https://wieloletni.itl.waw.pl

# 3. Objective Ranking

### *3.1. The Concept of Objective Ranking*

This concept was introduced in [1]. Such a ranking consists of:

- counting overall average of a given partial indicator (such as market environment subindex),
- counting the worst under-achievement among all partial indicators,
- correcting slightly this worst under-achievement by the sum of under-achievements (or over-achivements).

In more detail, all partial indicators are transformed into partial achievement indicators (by comparing them to their statistical averages), then these partial achievement indicators are aggregated into an overall achievement indicator determined by the worst under-achievement with a slight correction by the sum of all partial achievements. The aggregated achievement indicators are used for ranking. Thus such ranking assumes, similarly as the NRI ranking, that all indicators are significant, but selects as most significant such indicators that have among all partial indicators the worst underachievement when compared to statistical averages for a given indicator.

No ranking can be absolutely objective, because the choice of the method of ranking is subjective itself. However, the so-called objective ranking is *as objective as possible*, because most parameters of aggregation depend on statistical averages, not on subjectively determined weighting coefficients.

### *3.2. Method of Calculating Objective Ranking*

The method as described in [1] is based on a specification of double reference levels: aspiration level $a_j$ and reservation level $r_j$, for each criterion or indicator. After this specification, the approach uses a nonlinear aggregation of criteria by an achievement function that is performed in two steps.

First, we count achievements for each individual criterion or satisfaction with its values by transforming it (monotonically and piece-wise linearly), for example in the case of maximized criteria as shown in Eq. (3). In this transformation, we can choose its coefficients to have a reasonable interpretation of the values of the *partial* (*or individual*) *achievement function*. In the original objective ranking approach, the of [0; 10] points for eliciting expert opinions was used, but a modification to the range [1; 7] used in NRI calculations is easy.

$$\sigma_j(q_j,a_j,r_j) = \begin{cases} \alpha(q_j-q_j^{lo})/(r_j-q_j^{lo}), & \text{for } q_j^{lo} \le q_j < r_j \\ \alpha+(\beta-\alpha)(q_j-r_j)/(a_j-r_j), & \text{for } r_j \le q_j < a_j \\ \beta+(10-\beta)(q_j-a_j)/(q_j^{up}-a_j), & \text{for } a_j \le q_j \le q_j^{up} \end{cases}, \quad (3)$$

where $q_j$ is the value of $j$-th indicator, $q_j^{lo}$ is the lowest value and $q_j^{up}$ is the highest value of this indicator between all alternatives (countries, regions etc.). The parameters $a_j$ and $r_j$ are defined statistically as in Eq. (5). The parameters $\alpha$ and $\beta$, $0 < \alpha < \beta < 10$ if we use the [0; 10] range, denote correspondingly the values of the partial achievement function for $q_j = r_j$ and $q_j = a_j$. The value $\sigma_{jk} = \sigma_j(q_{jk}, a_j, r_j)$ of this achievement function for a given ranked alternative (country, region, etc.), $k \in \mathbf{K}$ signifies the satisfaction level with the criterion or indicator value for this alternative. Thus, the above transformation assigns satisfaction levels from 0 to $\alpha$ (say, $\alpha = 3$) for criterion values between $q_j^{lo}$ and $r_j$, from $\alpha$ to $\beta$ (say, $\beta = 7$) for criterion values between $r_j$ and $a_j$, from $\beta$ to 10 for criterion values between $a_j$ and $q_j^{up}$.

After this transformation of all criteria values, we might then use the following form of the overall achievement function:

$$\sigma(\mathbf{q},\mathbf{a},\mathbf{r}) = \min_{j\in\mathbf{J}} j_i(q_j,a_i,r_j) + \varepsilon/J \sum_{j\in\mathbf{J}} \sigma_j(q_j,a_j,r_j), \quad (4)$$

where $\mathbf{q} = (q_1,\ldots,q_j,\ldots,q_J)$ is the vector of criteria values, $\mathbf{a} = (a_1,\ldots,a_j,\ldots,a_J)$ and $\mathbf{r} = (r_1,\ldots,r_j,\ldots,r_J)$ are the vectors of aspiration and reservation levels, while $\varepsilon > 0$ is a small regularizing coefficient. The achievement values $\sigma_k = \sigma(\mathbf{q}_k,\mathbf{a},\mathbf{r})$ for all $k \in \mathbf{K}$ can be used either to select the best alternative, or to order the alternatives in an overall ranking list or classification list, starting with the highest achievement value.

A statistical determination of reference levels concerns values $m_j$ that would be used as *basic reference levels*, an upward modification of these values to obtain *aspiration levels* $a_j$, and a downward modification of these values to obtain reservation levels $r_j$. These might be defined as follows:

$$m_j = \sum_{k\in\mathbf{K}} \mathbf{q}_{jk}/K;\ r_j = 0.5(q_j^{lo}+m_j);\ a_j = 0.5(q_j^{up}+m_j),\ \forall j \in \mathbf{J} \quad (5)$$

where $K$ denotes the number of alternative options, thus $m_j$ are just average values of criteria in the set of all alternative options. Aspiration and reservation levels are, therefore, just averages of these averages and the lower and upper bounds, respectively.

### *3.3. Objective Ranking on WEF – NRI Data*

The calculations were performed on WEF data as used for NRI calculations, but using the objective ranking method, and the resulting rankings were compared. We start with illustrating results concerning Poland (Fig. 4).

We see that the essential differences are in the evaluations of Pillar 4 and, therefore, Subindex B: objective ranking identifies weak points and decreases the ranking if weak points are found. See also Table 5 for more detailed numerical data for two yearly rankings: 2009 and 2011. However, the sudden drop of the evaluation of WEF experts of

Table 5
The position of Poland in the world – according to NRI 2009 and 2011 data and two methods of ranking: objective ranking and WEF ranking

| Poland | Ranking | WEF 2009 | Objective 2009 | WEF 2011 | Objective 2011 |
|---|---|---|---|---|---|
| Pillar 1 | Market environment | 87 | 74 | 74 | 68 |
| Pillar 2 | Political and regulatory environment | 100 | 82 | 81 | 59 |
| Pillar 3 | Infrastructure environment | 41 | 36 | 43 | 46 |
| Pillar 4 | Individual readiness | 43 | 35 | 83 | 126 |
| Pillar 5 | Business readiness | 52 | 40 | 54 | 42 |
| Pillar 6 | Government readiness | 103 | 125 | 103 | 107 |
| Pillar 7 | Individual usage | 46 | 43 | 46 | 43 |
| Pillar 8 | Business usage | 69 | 64 | 60 | 55 |
| Pillar 9 | Government usage | 127 | 124 | 93 | 100 |
| Subindex A | Environment component | 68 | 65 | 60 | 63 |
| Subindex B | Readiness component | 62 | 72 | 73 | 112 |
| Subindex C | Usage component | 80 | 79 | 57 | 50 |
| NRI | | 69 | 67 | 62 | 67 |

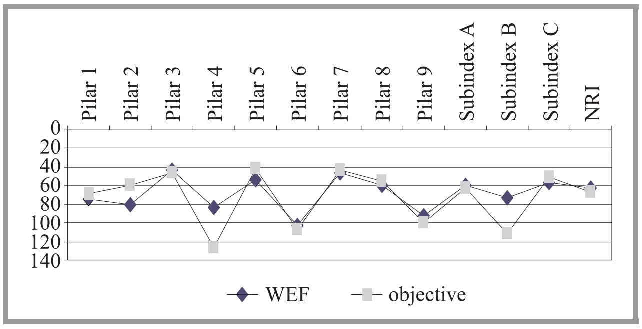



*Fig. 4.* The position of Poland in the world – according to NRI 2011 data and two methods of ranking: objective ranking (grey) and WEF ranking (black).

individual readiness pillar in Poland (from 43 to 83 place in ranking during two years, even more drastic – from 35 to 126 place – when the objective ranking method was applied) indicates that the expert evaluation applied by WEF might be contested.

From the above, we can draw diverse conclusions, but most important are as follows. Since objective ranking concentrates on relatively worst component indexes, it is apt to draw attention to actual weaknesses of a country. In the case of Poland, to generally the readiness component and to its pillars: individual readiness and government readiness. The simple averaging method applied by WEF tends to overlook relative worst cases. It draws an attention to government readiness, but not so much to individual one, and ranks much higher overall readiness component. On the other hand, not very promising assessment of individual readiness in Poland results from the NRI data, and it is only stressed by the method of objective ranking. Other sources of data, e.g., the Eurostat data on the index IOGSI (Individuals who Ordered Goods or Services over Internet), indicate that Poland might be even ahead of some other European countries in individual readiness, see [4]. Thus the assessments of WEF experts might be biased.

Table 6 shows a comparison of rankings of the first 22 countries in the world when using NRI WEF ranking and when using an objective ranking based on WEF data. Classifica-

Table 6
Comparison of first 22 places in WEF NRI ranking and objective ranking based on the same data

| Rank | WEF | Score | Objective | Score |
|---|---|---|---|---|
| 1 | Sweden | 5.60 | Sweden | 5.60 |
| 2 | Singapore | 5.59 | Singapore | 5.50 |
| 3 | Finland | 5.43 | Switzerland | 5.36 |
| 4 | Switzerland | 5.33 | Finland | 5.20 |
| 5 | United States | 5.33 | Iceland | 5.02 |
| 6 | Taiwan China | 5.30 | Luxembourg | 4.93 |
| 7 | Denmark | 5.29 | United States | 4.90 |
| 8 | Canada | 5.21 | United Kingdom | 4.89 |
| 9 | Norway | 5.21 | Germany | 4.89 |
| 10 | Korea Rep. | 5.19 | Canada | 4.76 |
| 11 | Netherlands | 5.19 | Canada | 4.76 |
| 12 | Hong Kong SAR | 5.19 | Denmark | 4.72 |
| 13 | Germany | 5.14 | France | 4.68 |
| 14 | Luxembourg | 5.14 | Netherlands | 4.68 |
| 15 | United Kingdom | 5.12 | Australia | 4.64 |
| 16 | Iceland | 5.07 | New Zealand | 4.64 |
| 17 | Australia | 5.06 | Austria | 4.62 |
| 18 | New Zealand | 5.03 | Hong Kong SAR | 4.61 |
| 19 | Japan | 4.95 | Korea Rep. | 4.56 |
| 20 | France | 4.92 | Taiwan China | 4.55 |
| 21 | Austria | 4.90 | Japan | 4.52 |
| 22 | Israel | 4.81 | Belgium | 4.44 |

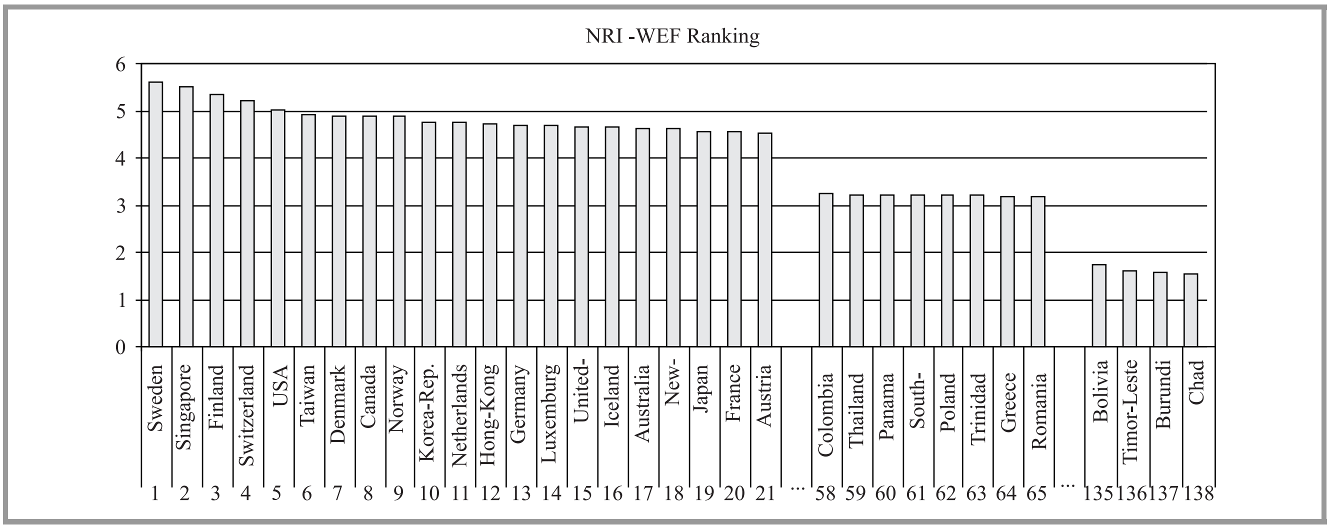



***Fig. 5.*** Classification of countries in the world according to objective ranking based on WEF data.

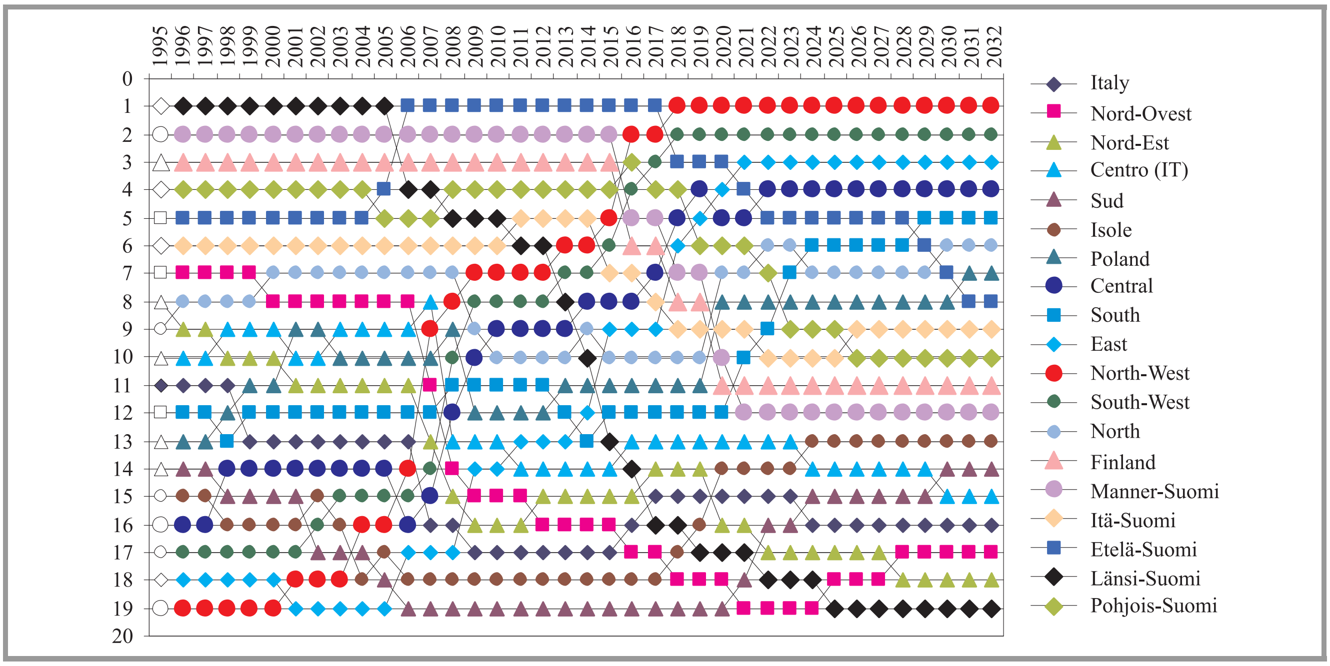



***Fig. 6.*** Dynamic ranking of regions of Finland, Italy and Poland.

tion of other countries, according to objective ranking, is presented in Fig. 5. We see that the overall rankings based on WEF data and applying the averaging method of ranking (NRI WEF Ranking) do not differ substantively form the rankings based on the same data but applying an objective ranking (NRI Objective Ranking). Nevertheless there are astonishing differences. For example, Taiwan is ranked in the 6th place in the world according to NRI WEF Ranking, but only in the 20th place according to NRI Objective Ranking, which indicates that Taiwan has some weak points that were averaged out by the method applied in WEF ranking. Similar comments concern Hong Kong. Objective ranking places countries, such as Iceland or Luxembourg higher as they have more balanced achievements in network readiness indicators.

## 4. Dynamic Ranking of Regions in Finland, Italy and Poland

In order to estimate future ICT development, we selected three European countries: Finland, Italy, and Poland and analyzed Eurostat data on three indicators of information society development in regional focus:

– households with access to the Internet at home,

– households with broadband access to the Internet,

– individuals regularly using the Internet.

We used an aggregation of these three indicators as in objective ranking, hence stressing the weak points. However,

Table 7
Dynamic ranking of ICT penetration in Polish voivodeships (places in ranking)

| Year/Voivodeships | Łódzkie | Mazowieckie | Małopolskie | Śląskie | Lubelskie | Podkarpackie | Podlaskie | Świętokrzyskie | Lubuskie | Wielkopolskie | Zachodniopomorskie | Dolnośląskie | Opolskie | Kujawsko-Pomorskie | Pomorskie | Warmińsko-Mazurskie |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1995 | 11 | 6 | 4 | 2 | 9 | 10 | 3 | 16 | 12 | 5 | 8 | 7 | 14 | 15 | 15 | 13 |
| 2000 | 8 | 3 | 10 | 2 | 11 | 9 | 5 | 16 | 12 | 4 | 6 | 7 | 14 | 15 | 15 | 13 |
| 2005 | 8 | 2 | 14 | 3 | 12 | 6 | 15 | 16 | 10 | 4 | 5 | 7 | 9 | 11 | 11 | 13 |
| 2010 | 11 | 1 | 10 | 6 | 14 | 8 | 16 | 15 | 5 | 9 | 7 | 4 | 3 | 2 | 2 | 13 |
| 2015 | 14 | 4 | 5 | 6 | 15 | 10 | 16 | 11 | 2 | 9 | 7 | 8 | 3 | 1 | 1 | 12 |
| 2020 | 15 | 6 | 4 | 5 | 14 | 11 | 16 | 7 | 2 | 9 | 8 | 10 | 3 | 1 | 1 | 12 |

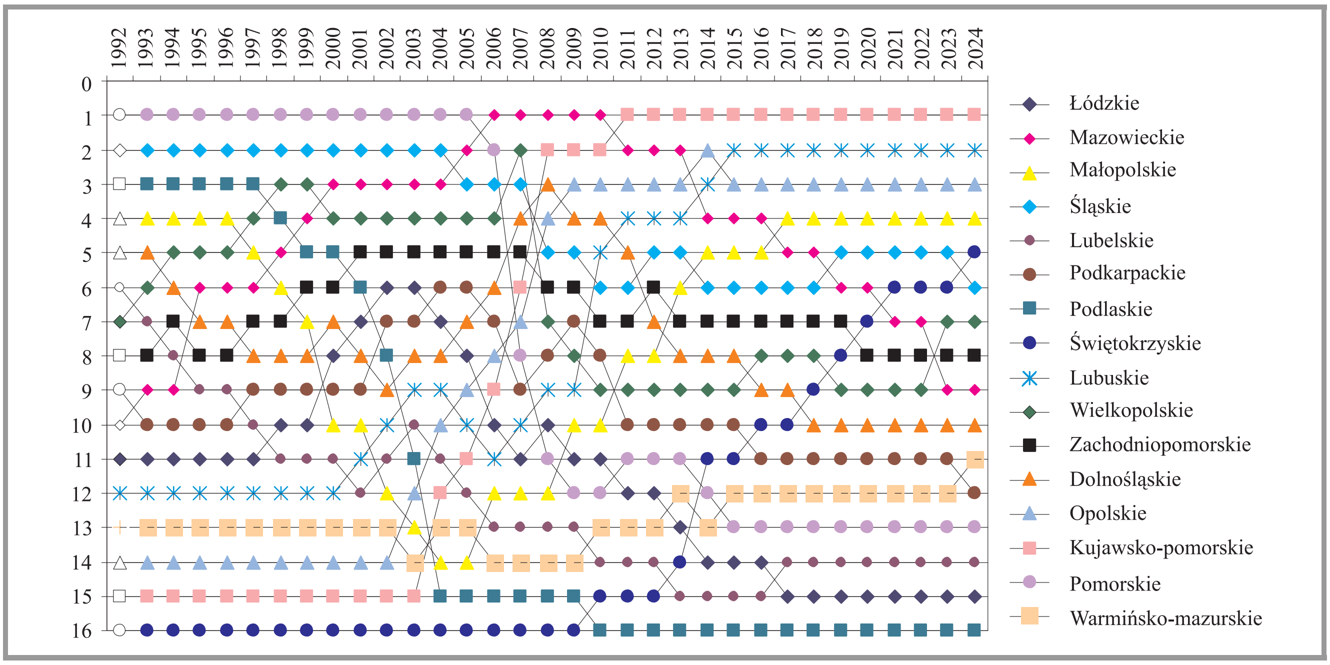



***Fig. 7.*** Dynamic ranking of voivodeships in Poland.

in order to obtain predictions, we estimated the future development of these indicators by fitting statistically sigmoidal curves, which stresses regular changes and eliminates accidental changes. We analyzed first data for NUTS 1 regions of Finland, Italy and Poland (as illustrated in Figs. 1 and 2). The method applying objective ranking on past data and data predicted by statistically estimated sigmoidal curves might be called as *dynamic ranking*. The results are illustrated in Fig. 6.

As we see in Fig. 6, the regions of Finland maintain in majority the relatively high positions, although one of them, Länsi-Suomi, during the years 2005–2011 has lost its leading position and further projections may even indicate a decline to the last place, behind Polish and Italian regions. While the other region of Finland, Etelä-Suomi, from the year 2006 emerged as the first on the ranking position and might maintain it until 2017, further projections indicate a continuation of its relatively high (fifth or eighth) position. Poland and its regions are placed better than Italy and its regions[2] already in the period 2006–2011. Further forecasts indicate that Poland (but not Italy) after 2020 may surpass Finland; around 2018, two Polish regions, North-West (Północno-Zachodni) which began at the last, 19th place, but in 2011 already climbed to the seventh place and the South-West (Południowo-Zachodni) which in 2011 was already on the eighth place will take higher places. Italian

[2]This fact justifies our doubts about the objectivity of the very low evaluation of Poland's individual readiness indicators by WEF experts, since all three indicators used in the dynamic ranking discussed here concern individual network readiness and are based on Eurostat data.

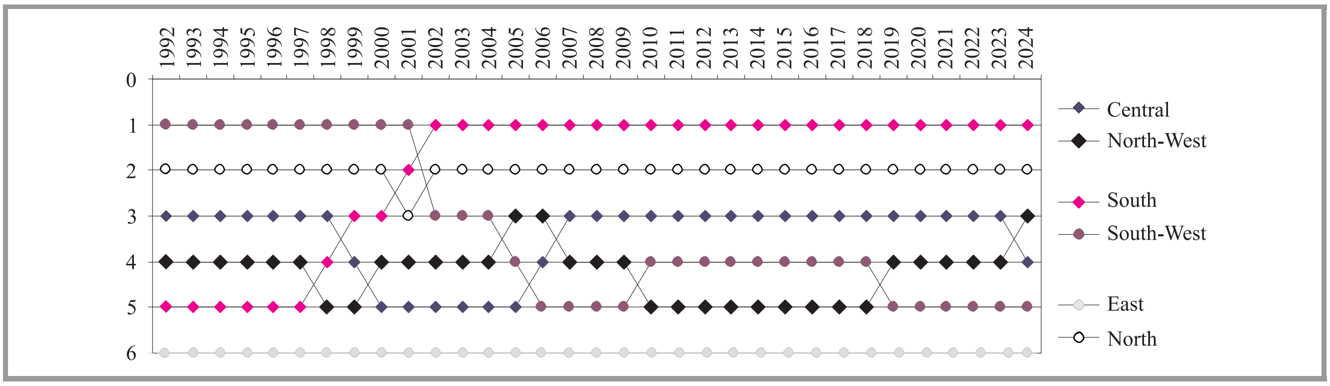



*Fig. 8.* Dynamic ranking of ICT penetration in Polish macro-regions.

regions, which initially occupied better position than Polish regions, already during the period 2006–2011 decline for further spaces. It is significant that the Italian region which has the best long term forecast is Isole region. Naturally, it can be questioned whether such long-term predictions might come true. It is evident that the development of penetration of the information society is an uneven process, not wholly dependent on the wealth of regions, more of their strategic determination, but historical data from years 2006 to 2010 indicate significant changes in the trends and future position of regions.

## 5. Dynamic Ranking of Polish Voivodeships and Regions

In order to illustrate method of dynamic ranking, we used data from GUS (Central Statistical Office of Poland) concerning Polish voivodeships (they are rather NUTS 2 regions, smaller than NUTS 1 macro-regions) on the same indicators, as used in the dynamic ranking of macro-regions above, using past data 1992–2011 and their extrapolations by sigmoidal curves 2012–2024. The results are illustrated by Table 7 and Fig. 7.

We see that Kujawsko-Pomorskie voivodeship has great chances to become the best one, due to a positive and strong dynamics of development in the years 2003–2010, while Podlaskie voivodeship might become the last one (good position until 2001, very negative trend 2002–2005). Eastern Region (Podlaskie, Lubelskie, Podkarpackie voivodeships) is generally low in the ranking. Mazowieckie voivodeship, currently the best one, might loose its leading position in the future. When applied to European macro-regions (NUTS 1) of Poland, see Fig. 8, the dynamic ranking shows more stable positions of regions, but confirms the worst position of the Eastern Region, the Southern Region has already become the best one and will probably keep this position.

Generally, GUS data are not quite consistent with Eurostat data. However, the above examples show that dynamic ranking can give more interesting information than just static one. An alternative way to dynamic ranking is counting delay or advancement times (how many years it takes to achieve the average level of an indicator or of aggregated indicators, see [5].

## 6. Conclusions

There are many indicators of socio-economic penetration of information society technologies, thus any evaluation of them requires multicriteria aggregation.

Simple aggregation using weighting coefficients is subjective and gives less interesting results than the so-called objective ranking.

Much more important conclusions for regional policy can be drawn by applying dynamic ranking – a method of objective ranking using past data and data predicted by statistically estimated sigmoidal curves leading to a change of ranking in time.

Other method of incorporating dynamic phenomena might be to compute delay or advancement times.

# Appendix

## Structure of NRI Indicators 2010–2011

Subindex A. Environment component
1st pillar: Market environment
1.01 Venture capital availability*
1.02 Financial market sophistication*
1.03 Availability of latest technologies*
1.04 State of cluster development*.
1.05 Burden of government regulation*
1.06 Extent & effect of taxation*
1.07 Total tax rate, % profits
1.08 No. days to start a business
1.09 No. procedures to start a business
1.10 Freedom of the press*
2nd pillar: Political and regulatory environment
2.01 Effectiveness of law-making bodies*
2.02 Laws relating to ICT*
2.03 Judicial independence*

2.04 Efficiency of legal system in settling disputes*
2.05 Efficiency of legal system in challenging regs*
2.06 Property rights*
2.07 Intellectual property protection*
2.08 Software piracy rate, % software installed
2.09 No. procedures to enforce a contract
2.10 No. days to enforce a contract.
2.11 Internet & telephony competition, 0-6 (best)
3rd pillar:Infrastructure environment
3.01 Phone lines/100 pop.
3.02 Mobile network coverage, % pop. covered.
3.03 Secure Internet servers/million pop.
3.04 Int'l Internet bandwidth, Mb/s per 10,000 pop.
3.05 Electricity production, kWh/capita.
3.06 Tertiary education enrollment rate, %.
3.07 Quality scientific research institutions*
3.08 Availability of scientists & engineers*
3.09 Availability research & training services*
3.10 Accessibility of digital content*
Subindex B Readiness component
4th pillar: Individual readiness 5.6 16
4.01 Quality of math & science education*
4.02 Quality of educational system*
4.03 Adult literacy rate, %..
4.04 Residential phone installation (PPP $).
4.05 Residential monthly phone subscription (PPP $)
4.06 Fixed phone tariffs (PPP $) .
4.07 Mobile cellular tariffs (PPP $)
4.08 Fixed broadband Internet tariffs (PPP $)
4.09 Buyer sophistication*
5th pillar: Business readiness 4.1 53
5.01 Extent of staff training*
5.02 Quality of management schools*
5.03 Company spending on R&D*
5.04 University-industry collaboration in R&D*
5.05 Business phone installation (PPP $).
5.06 Business monthly phone subscription (PPP $)
5.07 Local supplier quality*
5.08 Computer, communications, & other services imports, % services imports.
6th pillar: Government readiness
6.01 Gov't prioritization of ICT*
6.02 Gov't procurement of advanced tech*
6.03 Importance of ICT to gov't vision* 6.04 Subindex C Usage component
7th pillar: Individual usage 4.7 35
7.01 Mobile phone subscriptions/100 pop.
7.02 Cellular subscriptions w/data, % total
7.03 Households w/ personal computer, %
7.04 Broadband Internet subscribers/100 pop
7.05 Internet users/100 pop.
7.06 Internet access in schools*
7.07 Use of virtual social networks*
7.08 Impact of ICT on access to basic services*
8th pillar: Business usage
8.01 Firm-level technology absorption*
8.02 Capacity for innovation*
8.03 Extent of business Internet use*
8.04 National office patent applications/million pop
8.05 Patent Cooperation Treaty apps/million pop
8.06 High-tech exports, % goods exports
8.07 Impact of ICT on new services and products*
8.08 Impact of ICT on new organizational models*
9th pillar: Government usage
9.01 Gov't success in ICT promotion.
9.02 ICT use & gov't efficiency*
9.03 Government Online Service Index, 0-1 (best)
9.04 E-Participation Index, 0-1 (best).

## References


[1] A. P. Wierzbicki, "The problem of objective ranking foundations, approaches and applications", *J. Telecom. Inform. Technol.*, no. 3, 2008.

[2] "The Networked Readiness Index 2009–2010", World Economic Forum, 2010.

[3] S. Dutta and I. Ma, Eds., *The Global Information Technology Report 2010–2011*. Geneva: World Economic Forum, 2011.

[4] B. Ziewiec, "Diversity of temporal and territorial social penetration rates of information technology in Europe", *J. Telecom. Inform. Technol.* (to be published in 2013).

[5] J. Grzegorek and A. P. Wierzbicki, "New Statistical Approaches in the Systemic Analysis of Regional, Intra-regional and Cross-Regional Factors of Information Society and Economic Development: The Case of Mazovia", *Mazowsze: Studia Regionalne*, no. 3, pp. 117–128, Warszawa, 2009.

[6] J. Grzegorek and A. P. Wierzbicki, "Metody analizy statystycznej rozwoju wskaźników społeczeństwa informacyjnego: aspekty metodologiczne oraz zastosowania na przykładzie regionów Polski i Włoch", *KSTiT 2010*, Wrocław, Poland, 2010 (in Polish).


**Jan Władyslaw Grzegorzek** born 26-08-1952 in Mielcuchy. In 1978 graduated from Warsaw University of Technology, specializing in electronic measuring equipment. In the years 1977–1994 he worked in the National Institute of Telecommunications, initially as a design engineer, next as scientific research assistant and then assistant professor. He participated in the development of fixed-line telephony RSŁA that was distinguished by an award of the Minister of Communications. Since he 1994 works at TP S.A. as a Chief Specialist and since 2007 as an Expert of Support Operations. Since 2007 he works on a doctoral dissertation. The subject is "Use of advanced statistical method to study development of information society in Poland treated as the system of regions".


E-mail: Jan.Grzegorek@orange.com
Telekomunikacja Polska S.A.
Twarda st 18
00-105 Warsaw, Poland

**Andrzej Piotr Wierzbicki** got his M.Sc. in 1960 in telecommunications and control engineering, Ph.D. in 1964 in nonlinear dynamics in control, and D.Sc. in 1968 in optimization and decision science. He worked as the Dean of the Faculty of Electronics, Warsaw University of Technology (WUT), Poland (1975–1978); the Chairman of Systems and Decision Sciences Program of International Institute for Applied Systems Analysis in Laxenburg n. Vienna, Austria (1979–1984). He was elected a member of the State Committee for Scientific Research of Republic of Poland and the Chairman of its Commission of Applied Research (1991–1994). He was the Director General of the National Institute of Telecommunications in Warsaw (1996–2004). He worked as a research Professor at Japan Advanced Institute of Science and Technology (JAIST), Nomi, Ishikawa, Japan (2004–2007). Beside teaching and lecturing for over 45 years and promoting over 100 master's theses and 20 doctoral dissertations at WUT, he also lectured at doctoral studies at many Polish and international universities. Professor Wierzbicki is an author of over 200 publications, including 14 books, over 80 articles in scientific journals, over 100 papers at conferences; the author of 3 patents granted and industrially applied. His current interests include vector optimization, multiple criteria and game theory approaches, negotiation and decision support, issues of information society and knowledge civilization, rational evolutionary theory of intuition, theories of knowledge creation and management.

E-mail: A.Wierzbicki@itl.waw.pl
National Institute of Telecommunications
Szachowa st 1
04-894 Warsaw, Poland

Paper

# Who is Asking and for What: WHOIS Traffic Analysis

Mariusz Kamola

*Research and Academic Computer Network (NASK), Warsaw, Poland*
*Institute of Control and Computation Engineering, Warsaw University of Technology, Warsaw, Poland*

**Abstract—The paper presents analysis of WHOIS requests for 13-month period. Both requestor address and the domain name being requested are analyzed, showing that WHOIS traffic can be roughly classified into systematic scanning of domain names and individual low-volume activity, mostly targeting very popular names. The comparison of requested names with standard dictionary entries reveals typical mutations for registered names, and mutations performed by scanning automata. As most popular names in WHOIS coincide with standard top website ranks, the ways of utilizing WHOIS data for the benefit of Internet community as a whole, are proposed.**

***Keywords—request map, semantic analysis, SEO, WHOIS.***

## 1. Introduction

The term WHOIS refers in broad sense to a protocol [1] designed to query personal details related with various entities found in today's Internet. In this paper we will deal with WHOIS in more narrow and well known sense – as the technology to retrieve Internet domain name registrant's data. These data are made available to the public by appropriate servers maintaining registration databases. There are three main ways to access the registration data: a HTTP interface, a service operating WHOIS protocol on port 43, and bulk datasets obtainable from the registry. Due to the fact that the data may contain personal information as e-mail, phone number and even street address, there have been always discussions in ICANN about privacy issues, and the conflict between data openness in Internet community and privacy law imposed by local governments [2].

The issue is an important one because WHOIS data as such can serve as huge, effective, and legal directory for spamming, hacking and other socially undesirable behaviors. Tackling the matter, ICANN has come up with a series of requirements and recommendations for registries, aiming at preventing misuse of the data. Web access has been mostly equipped with CAPTCHA technology and port 43 service with rate limitations to prevent massive and automated database scanning. Such scanning is still possible on bulk data, under declaration that the results will *not* be used for marketing and alike (cf. eg. [3]). ICANN is monitoring the issues with WHOIS as DNS is evolving; see the relevant memorandum on the occasion of gTLD (Generic Top-Level Domain) release [4] where minimum set of registrant information in different domain classes has been specified; also prior related regulatory activities are mentioned therein.

Naturally, despite regulatory efforts, business wants to make money from the valuable WHOIS information – and the retail requests generated by serious or curious individuals mix with regular database scanning performed by companies. Such is the major outcome of the cursory study on WHOIS requests to NASK register. The major motivation for such a study was to gain insight into how actually the database is used, by whom and, if possible, for what purpose. Investigating business models underlying WHOIS requests made by companies thriving on added Internet services has been considered particularly important. This is not because NASK is going to compete with them; on the contrary, being a supervisor of a large part of Polish web activities, NASK is going to consider utilization of those data to stimulate healthy growth of Internet community in the country – also through educational activities, backed by sound research results reported in scientific papers.

This paper is organized as follows. Section 2 presents the data that have been worked on along with the computing equipment. Then author focuses on the part which is, in his opinion, not present in the literature: to identify requestors of WHOIS data. The sole purpose of such an investigation is to classify WHOIS users into categories and, possibly, into subcategories, based on the traffic generated by them. In Section 3 attention is shifted to the domain names requested. Their similarity and temporal pattern of requests will be examined, focusing on key commercial requestors. Section 4 comes with conclusions, possible exploitation of results and planned future work description.

## 2. Who is Asking

The data being subject to analysis were 180 million WHOIS requests recorded in 13 months since September 2009, stored in WHOIS database in NASK. The register covers `.pl` domain, as well as some other functional and regional domains (e.g. `.gov.pl`, `.edu.pl`, `.poznan.pl`). The selected period is long enough to get rid of any kind of seasonality if one operates on averages. However, it must be emphasized that the stable volume growth biases the results, giving more weight to latter data, cf. Fig. 1.

This section covers the analysis of the *source* of incoming requests, i.e. the IP address of the requester. From now on we will operate on IP addresses with its last byte canceled. This has been done for two reasons. The first is privacy.

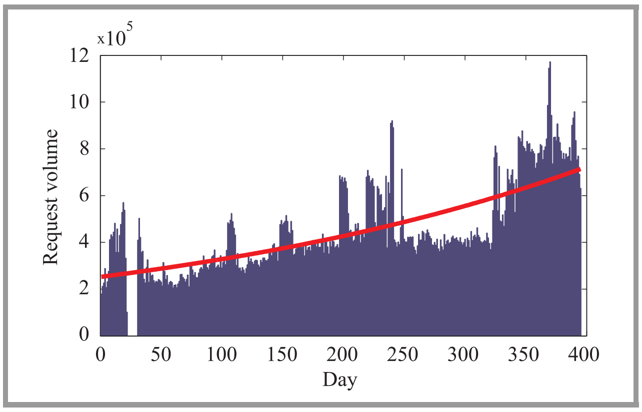



*Fig. 1.* Request volumes in subsequent days versus fitted exponential growth curve.

The second is that we suspect some organizations to use a pool of addresses while scanning WHOIS – and we did not wish to partition such activities artificially. Removal of the last byte reduces the number of unique requestor addresses from the original 3.2 million by two thirds.

The ranking of activity for 1,000 most active source addresses is presented in Fig. 2. Considering logarithmic scale, one can easily notice that the 30 most active ones account for more than 50 percent of the total requests. The first client in the ranking generates as much as 8 percent of the total traffic. Such an activity is clearly a kind of machine to machine communication. Next 100 most active requestors are also too active to be individuals; they may be commercial organizations, as well as gateways of networks with NAT – as, e.g. mobile operators. The activity of smaller requestors follows almost ideally the power law, like in social network node degrees, personal income and many natural phenomena.

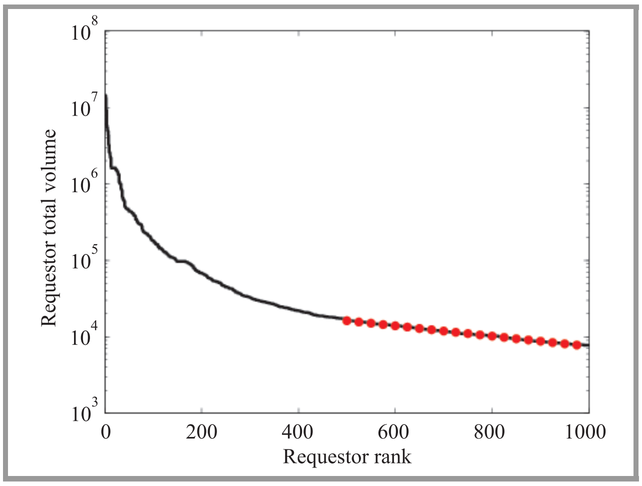



*Fig. 2.* Total number of requests for requestors ranked most active. Power law scaling for smaller clients shown as dotted line.

The most active in the ranking is one company providing Internet-related services as domain registration web and mail hosting, etc. It is located in one of the major Polish cities. Its traffic is generated by twelve equally loaded physical IP addresses. After stable growth in the first half of the analyzed period, its daily volume of request has stabilized around 50,000 a day. This number has not been affected at all by the rapid and stable growth of total WHOIS requests in August 2010, cf. Fig. 1.

Looking for similarities in behaviour of top requestors, let us have a look at activity of the second, third and seventh ones in the ranking (7th biggest requestor is included because of its unique location). Histograms of their daily request volumes have been compared to the overall traffic in Fig. 3. The first thing that "1" and "3" (also "2" to some extent) have in common is they exhibit quite precise limit of requests per day. If there are fluctuations in rates, they are drops, frequently reaching zero. On the contrary, the total traffic shows quite many peaks above the average. The differences in distributions are based in traffic trends, not shown here. All requestors *do not* exhibit any regular request growth in longer (e.g. monthly) window throughout the whole timespan. As regards "7", it resembles the total trend the most, but it does not grow either. Moreover, it switches off just the moment the total number of requests grows rapidly (August 2010).

Let us now study geographical location of request origins. The results may give an idea for whom it could be valuable to have a Polish domain name – or interested in discovering who owns a name. A free geographical location of IP addresses service [5] was used for this purpose. The lookup gives the two-letter country code and a city name. It must be noted that the lookup is not reliable as it could not find the address location in 24 percent of cases. Also, the geographical location obtained is sometimes confusing, e.g. the third largest requestor, obviously registered as Polish company, the location found was Turkey. However, despite of imperfections, the request world map shown in Fig. 4 looks reasonable. Poland is evidently the leader, and the other major requesting countries are either big (China), geographically close (Czech Republic) or with big Polish diaspora (Brazil). Or all of them, as for France, Great Britain and USA. Unfortunately, the map rendering mechanism [6] is not flawless, taking USA state names for the names of the countries, and not displaying requests from Germany and Netherlands, the fifth and eighth in the ranking, respectively.

In this map of interest there is an absence of big countries: Spain, Ukraine, Greece, Lithuania – with their obvious links with Poland. To explain such disinterest and its correlation with other factors, e.g. the amount of foreign investments is, however, beyond the scope of this paper. On the other hand, some countries exhibit surprisingly high rate of requests, as Denmark and, most of all, Iran. Considerably high interest of Polish southern Czech and Slovak neighbors explains easily as they are already actively participating in Polish retain Internet trade and services.

WHOIS requests generated regionally, i.e., in Poland, have also been analysed w.r.t. its origin. Biggest requesting lo-

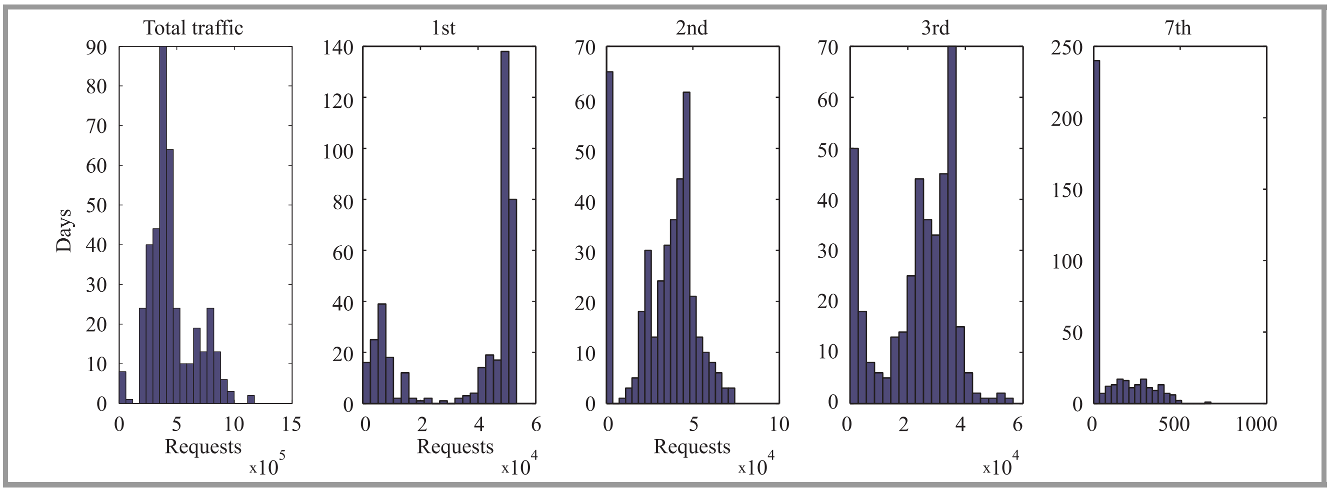



*Fig. 3.* Histograms of daily activity of all users, the three top requestors and the 7th requestor in the ranking.


*Fig. 4.* Request world map.

cations are shown in Fig. 5, vs. the population of those locations. In general, the biggest sources of traffic lie in the biggest cities – which also means that the IP location database in Poland is more accurate than the global one. This rule is represented in the figure by points clustered around the straight line. Note that there appear cities rather too little active for their size ("15"), which may mean that IT there need more development. But much bigger disproportions can be seen on the other side of the straight line: there is a small town ("17") and a village ("7") generating more requests than a half-million metropolis. It is difficult to expect that the latter one is just an ordinary case. A closer look at 7's traffic history reveals that it started all of a sudden in February 2010, which mean of 1,600 requests per day and standard deviation as big as 1,500. The traffic experienced from time to pauses of several days time, and it did not show any growth trend – like for the biggest requestor, and unlike in global statistics.

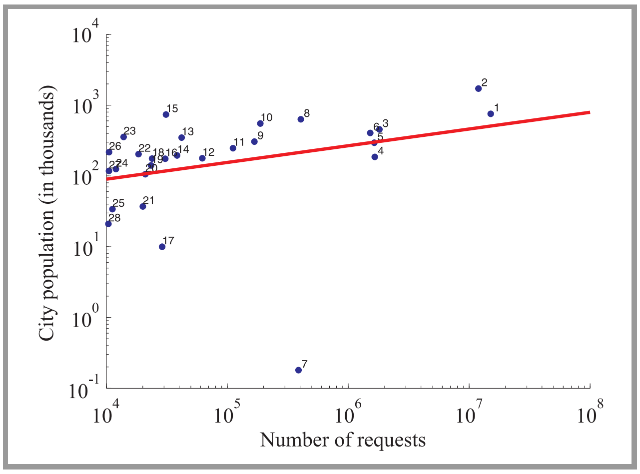



*Fig. 5.* Top Polish requesting cities versus their population. Mark numbers represent city rank w.r.t. generated traffic. Linear regression in log-by-log domain is shown by the red line.

Summing up, we may observe that the biggest traffic is generated by commercial scanners. These are located mainly in big cities, with few exceptions. Commercial requestors usually maintain their daily rate of requests, insensitive to overall WHOIS traffic growth. They use a single address or a pool of addresses, thus distributing their machine loads.

## 3. Requested names analysis

To get a reference point, let us confront our top 20 requested WHOIS names with top 20 visited `.pl` addresses by Alexa [7] as shown in Table 1 . Although Alexa ranking was done a year after the WHOIS data end, one can still see that entries in both columns are similar, especially for the top 10. Therefore WHOIS statistics may serve as a decent measure of company or website popularity – at least this applies for big fish. This interest is risen by individuals who, certainly, are not going to buy such domain names; it is rather curiosity that drives users to check extra info about companies. Further entries are not so well matched: those WHOIS names that are not present in Alexa top 20 are given in italics. Such discrepancies may be due to the lag of Alexa ranking, but it may be also different kind of interest driving users to WHOIS and to the wepage itself. Take for example `platformaobywatelska.org.pl` which is the political party governing right now: it is not shown in Alexa top 20[1] but appears for WHOIS. The reason could be that requestor is interested in who is *personally* involved in domain registration, which amounts to looking for reliable extra information about the party.

Table 1
Polish WHOIS vs. Alexa: top 20

| | **WHOIS** | **Alexa** |
|---|---|---|
| 1 | google.pl | google.pl |
| 2 | onet.pl | onet.pl |
| 3 | wp.pl | allegro.pl |
| 4 | nk.pl | wp.pl |
| 5 | allegro.pl | gazeta.pl |
| 6 | nasza-klasa.pl | interia.pl |
| 7 | gazeta.pl | nk.pl |
| 8 | interia.pl | mbank.com.pl |
| 9 | home.pl | o2.pl |
| 10 | *test.pl* | pudelek.pl |
| 11 | o2.pl | sport.pl |
| 12 | demotywatory.pl | otomoto.pl |
| 13 | *tpnet.pl* | goldenline.pl |
| 14 | *wrzuta.pl* | kwejk.pl |
| 15 | *platformaobywatelska.org.pl* | demotywatory.pl |
| 16 | *nazwa.pl* | ceneo.pl |
| 17 | pudelek.pl | home.pl |
| 18 | *peb.pl* | tvn24.pl |
| 19 | *blox.pl* | filmweb.pl |
| 20 | *ropa.pl* | chomikuj.pl |

However, a domain name in WHOIS estimates interest in smaller companies, ideas or activities as well. Take for example `tiny.pl`, which does not appear in the above list, but is second last frequently asked domain on the 30th Sept. 2010, i.e., the last day of the period analyzed. The service accomplishes domain name abbreviation, like many other ones – and with similar business model behind. Figure 6 illustrates the rapid growth of interest in the name, preceded by a long period of rather poor interest (the name was registered as early as 2004). About 50% of those requests have been made from unique IP addresses, which means in this specific case (as well as for more popular names) most of the traffic is generated by curious individuals rather than commercial scanners.

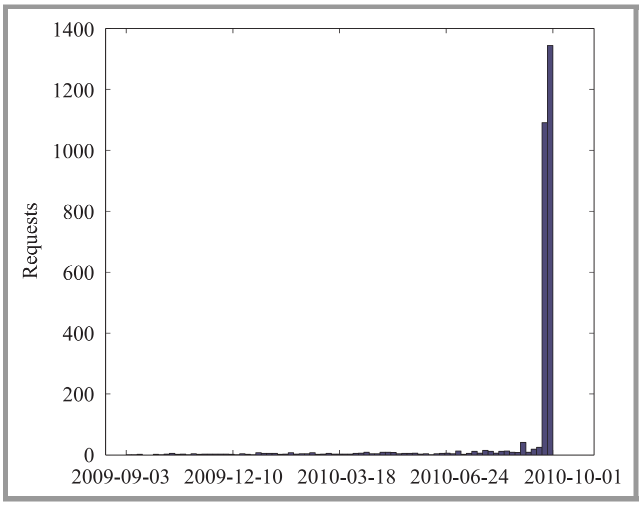



***Fig. 6.*** Number of WHOIS queries for `tiny.pl` name, grouped in 5-day periods.

Such general curiosity about big companies never shows for big requestors. The objects of interest of big requestors are (in the order of their frequency):

- names of objects (not being trademarks or proper names): *shawl*, *office*, *room*, etc.,
- expressions (with words mostly being written together): *new photography*, *stairs from Poland*, *property valuation*, etc.,
- the above names, but located in functional or regional `.pl` subdomains,
- proper names and trademarks.

It is worth noting that verbs in names are rare and, contrary to widespread opinion, names referring to sexuality are few.

It is interesting to examine how requested domain names are related to words commonly used in written or spoken language. In particular:

- which dictionary words are most popular as domain names,
- which dictionary words are not interesting, and why.

The problem is what should be considered the reference dictionary. Language corpora contain lots of words and their variations according to specific language grammar

[1]Municipal and presidential elections in 2011 and parliament election in 2012 were equally good reasons for `platformaobywatelska.org.pl` to appear in Alexa top 20 ranking, however, such domain name did not appear there altogether.

and morphology that are not appealing as domain names (e.g., non-infinitive forms of verbs) or anachronisms. To make our reference dictionary a contemporary one, we decided to use frequency list for Polish words as offered by `wiktionary.org` [8]. It contains only 10,000 Polish words, and probably other alternative sources of data [9] could have turned out to be more useful and suitable.

To measure word popularity, we have to take into account not only exact requests for this word, but also similar requests. The most widespread technology for fuzzy word matching, used also in spell checking and correction, is based on Levenshtein metric [10]: the minimum number of original word elementary transformations that make a dictionary word out of it. Elementary transformations come in four kinds:

- *deletes* – removal of a single letter at position $i$,
- *transposes* – swap of letters at positions $i$ and $i+1$,
- *replaces* – replacement of a letter at position $i$ with another letter,
- *inserts* – insertion of a single letter at position $i$ (shifting letter at $i+1$ one position behind and so on).

Let $A = 'a', 'b', \ldots, 'z'$ be the alphabet and $w$ be a word $w = (w_1, \ldots, w_n)$ where $w_{i \in 1..n} \in A$. Let us denote the set of all words being result of a single transformation of $w$ by $U^{(1)}(w)$. Let $D = d_1, .., d_m$ be the dictionary of words and $F = f_{w_1}, \ldots, f_{w_m}$, $f \in \mathfrak{R}$ their respective frequencies. Then the distance-one fuzzy word matching procedure chooses

$$M^{(1)}(w) = \arg \max_{q \in U^{(1)}(w) \cap D} f_{w_q},$$

i.e. the most frequent dictionary entry reachable by one modification of $w$. Analogously,

$$M^{(2)}(w) = \arg \max_{q \in U^{(2)}(w) \cap D} f_{w_q},$$

is the distance-two matching procedure, where

$$U^{(2)}(w) = \bigcup_{z_i \; U^{(1)}(w)} U^{(1)}(z_i)$$

defines the set of words available by two elementary modifications to $w$. The matching routine proposed in [11] combines distance-zero, distance-one and distance-two algorithms, returning the best match regardless of the number of modifications (0, 1 or 2, respectively) needed.

The above routine has been employed, with further modifications, for matching requested domain names with dictionary entries. The needed modifications are requested domain name preprocessing rules, before the actual matching takes place:

- Cutting domain name to 15 initial characters. This is because matching algorithm performance decreases rapidly for longer words.
- Removing the domain name part after first dot. This makes `.pl` itself and its all functional and geographical subdomains equally important, being the simplest way to detect interest in names registered in subdomains.
- Removing all dashes in names as irrelevant.
- Replacing all numerical substrings with single ‘0’ (zero) character. This way, popular numerical pre- or suffixes to names are easily detectable.

In practice, precise and reliable operation of the above matching approach increases with dictionary word length. Obviously, distance-two modifications of short words give plenty of other dictionary words, obscuring the conclusions. Otherwise, for longer (e.w. 10-letter) words the algorithm detects most of the word’s variations that are requested. For instance, the most popular (by means of related WHOIS requests) 10-letter dictionary word, ‘fotografia’ (*photography*), has 619 distance-two similar domain names. The most popular ones have been listed in Table 2. Even for readers not familiar with Polish language all those names appear as loose variations of the basic term ‘photography’: they point to websites with similar functionality, too. Interestingly enough, the basic form, ‘fotografia’ appears only on the 5th position. These variations are not created by adding random infixes; they all are meaningful: diminutive, imperative, plural, bearing typical suffixes for village names, indicating continuous service. This is also an indication that those domain names are registered and alive.

Table 2
Requests similar to dictionary entry ‘fotografia’ (*photography*), in order of their frequency in WHOIS requests

| |
|---|
| fotografuj.pl |
| fotografijka.pl |
| fotogratis.pl |
| fotografika.pl |
| *fotografia.pl* |
| fotograma.pl |
| fotografie.pl |
| ek-fotografia.pl |
| fotografow.pl |
| foto-grafi.pl |
| fotografie.org.pl |
| fotografia24.pl |

If we rank popularity of 10-letter or longer dictionary words referred to in WHOIS requests, it will be as in Table 3. The number of domain names similar to a dictionary word does not apparently depend on the word rank – but it is always substantial (at least 62, which means that those domain names are valuable). The average Levenshtein distance can be as small as 0.57, meaning that regional or

Table 3
Top 15 dictionary words,with their number of relevant WHOIS requests, number of distance-two similar domain names, average Levenshtein distance and actual Alexa Rank value

| Dictionary word | English translation | Number of requests | No. of similar domains | Average L. distance | Alexa Rank |
|---|---|---|---|---|---|
| fotografia | photography | 21090 | 619 | 1.0388 | 17,298,998 |
| apartament | suite | 19797 | 465 | 1.4710 | 0 |
| akademicki | academic | 17202 | 68 | 1.1912 | 0 |
| gospodarka | economy | 15821 | 79 | 1.7595 | 9,578,089 |
| nieruchomość | real estate | 15303 | 137 | 1.8978 | 1,089,090 |
| certyfikat | certificate | 10906 | 239 | 1.1172 | 0 |
| biblioteka | library | 9251 | 334 | 0.6317 | 7,299,153 |
| elektronika | electronics | 8855 | 261 | 1.1533 | 0 |
| bezpieczeństwo | safety | 7442 | 169 | 0.6331 | 0 |
| dziewczyna | girl | 7438 | 178 | 1.4663 | 12,629 |
| autostrada | motorway | 7295 | 144 | 1.6181 | 897,172 |
| elektroniczny | electronic | 7269 | 243 | 1.6626 | 0 |
| architektura | architecture | 7247 | 285 | 1.4421 | 3,966,672 |
| administracja | administration | 5576 | 156 | 1.1731 | 14,118,094 |
| budownictwo | construction | 5474 | 154 | 0.5714 | 97,740 |
| encyklopedia | encyclopedia | 5197 | 156 | 0.9423 | 175 |
| dominikana | Dominican Republic | 5093 | 62 | 1.2581 | 12,207,332 |
| astrologia | astrology | 5081 | 143 | 1.4336 | 1,147,848 |
| administrator | administrator | 4557 | 149 | 1.3624 | 0 |
| elektrownia | power plant | 4355 | 93 | 1.3656 | 0 |

functional domains are preferred than variations of the base name (cf. *library, safety, construction*). On the other end there are domains with big distance: *real estate, economy, electronic*, i.e. presumably denoting services with national range.

The last column of Table 3 gives *current* Alexa Rank, made available by one of SEO (search engine optimization) services [12]. The first observation is that this rank is much incomplete, missing highly interesting domains (other metrics: Page Rank and link popularity provide even more sparse data). The second is that Alexa Rank is quite inadequate to our rank of the dictionary word. The probable reasons are:

– comparison ignores the 2 years that passed since the end last analyzed data,

– bigger number of name alternatives decreases value of the name `dictionary_word.pl` itself.

Contrary to 'fotografia' (*photography*) keyword case, there are quite interesting domain names that apparently are targeted by scanners. Taking, for example the word 'wyjazd' (*trip*), confronting Table 4, we can see that the names requested are mutations of the base word. Mutation operations include swapping and doubling of pairs of letters, thus following common typographical errors made while entering the domain name. Therefore aim of the activity could be finally to register names similar to existing ones to intercept http requests containing typos and, for example, redirect them to competitive websites. Alternatively, the domain owner may resell the name to the owner of the "correct" domain name. Definitely, none of healthy competitors of the `wyjazd.pl` owner would like to run her business under a name containing a typo. Regarding the time pattern, we see that requests are made at equal 7-second intervals. Discontinuities of this schedule are due to the fact that some of the requests names had too big Levenshtein

Table 4
Scanning activity for the name `wyjazd` performed in a single day from a single IP address

| Time | Name |
|---|---|
| 16:07:37 | e–wyjazd.pl |
| 19:17:07 | e-wjazd.pl |
| 19:17:14 | e-wjyazd.pl |
| 19:19:15 | e-wwyjazd.pl |
| 19:19:29 | e-wyajzd.pl |
| 19:19:43 | e-wyjaazd.pl |
| 19:20:10 | e-wyjazdd.pl |
| 19:20:17 | e-wyjazzd.pl |
| 19:20:23 | e-wyjjazd.pl |
| 19:20:30 | e-wyjzad.pl |
| 19:20:37 | e-wyjzd.pl |
| 19:22:26 | e-wyyjazd.pl |
| 19:31:35 | e-ywjazd.pl |
| 23:13:41 | ee-wyjazd.pl |
| 00:32:23 | eewyjazd.pl |

distance to 'wyjazd' (*trip*) to be detected by our analytic software. Taking into account the geographical location of the requestor, it is located abroad and makes also regular requests for the proper domain name from neighboring IP addresses, but in much longer timescale.

If we consider least frequently queried names that are similar to dictionary entries, we will find that there are surprisingly many being the vocabulary entry itself, with no mutations. For example, in a reverse popularity ranking complementary to Table 3 the first mutation occurs only on the 95th position. Out of 30 first dictionary words least used, 15 are nouns (*vaccine*, *agreement*), 8 are adjectives (*southern*, *fixed*) and 5 are verbs (*apply*, *happen*). The words given in parentheses, which are examples of domain names queried only once, seem not to be uncommon at all – yet they did not gain much popularity. Therefore there is an measurable reason for them to gain popularity (high frequency in dictionary), and at the same time there exist technical possibility to achieve that (registering word mutations as domain names). Making such statistics available to the public may stimulate further growth of registered domain names.

## 4. Conclusion

Analysis of the source of WHOIS requests and their content prove that at systematic domain names scanning activity is commonplace, and that it has a considerable share of the overall WHOIS traffic. Certainly, there must be the business case for that: it may be the detection of unregistered and attractive domains, monitoring of availability of the popular names for registration, retrieving e-mail addresses to send commercial offers or any other reason. Explanation of reasons of scanning would require joint analysis of the domain registration and name querying processes, which lies outside of the scope of this paper. Such analysis would be definitely interesting and worth effort – but, most of all, it needs to have a well defined social purpose.

On the other side, we observe big volume of requests for very popular domains that correlate well with Alexa ranking (cf. Table 1). Therefore we can spot at least two basic classes of requestors: commercial scanners and private users. As for the latter, we may suppose they place requests out of curiosity for a company that is behind a domain name: its real name, location, entry date. In this regard, a WHOIS record is equal to a economic press release, or better, an official register of companies.

Regardless of requestors' motives, WHOIS activity for a domain name can be perceived as reliable metric of the domain importance and – maybe – its true value. WHOIS statistics, when used skillfully, may contribute to overall domain market growth. We believe that such growth is good for country's economy as such – regardless of the benefit of companies already profiting from domain registration or trade processes. Having big number of domains means that Internet users appreciate their Internet identity and – since DNS itself should not be commercial – their freedom. It is only that such identity and freedom should come at adequate and affordable prices. So, we can exploit WHOIS statistic in two ways: 1) to sanitize names trade and 2) to suggest unused domain names that are meaningful and otherwise valuable. As regards the first idea, comparing prices of domains sold at stock (cf. Table 5, with data retrieved from [13]) we see that often quoted prices are strikingly inadequate to the number of registered WHOIS requests. Examples of such names are: `licencje24.pl` (*licences24*), `green-age.pl` or `we-love.pl`[2]. As the mentioned names do not denote trademarks but common terms, their rank should not be affected by 2-year time difference in dates of WHOIS statistic recording and domain stock exchange query. Anyhow, their prices seem to be far exaggerated; we hope that making this sort of comparison publicly available may restore some order and sanitize prices[3]. Suggesting unused domain names for sale is considered now an idea for further discussion.

Utilizing WHOIS as reliable register of domain values has been addressed already in a patent [14] in a complementary context: the author proposed to enrich WHOIS data with a record describing its value, computed on the base on a number of external metrics. The article proves that internal WHOIS statistics themselves can also be considered

Table 5
List of currently most valuable names for sale vs. historical WHOIS number of requests

| Domain name | Stock quote (PLN) | No. of WHOIS requests |
|---|---|---|
| msza.pl | 200,000.00 | 91 |
| jedwab.pl | 130,000.00 | 81 |
| icrm.pl | 100,000.00 | 58 |
| cov.pl | 100,000.00 | 213 |
| ddw.pl | 92,000.00 | 279 |
| goracezrodla.pl | 60,000.00 | 61 |
| pc.com.pl | 50,000.00 | 216 |
| dobrarobota.pl | 50,000.00 | 62 |
| e-kontakt.pl | 40,000.00 | 148 |
| licencje24.pl | 27,000.00 | 7 |
| najwiekszy.pl | 24,000.00 | 43 |
| e-sprzedawca.pl | 10,000.00 | 52 |
| forumeo.pl | 10,000.00 | 53 |
| sciag.pl | 10,000.00 | 77 |
| green-age.pl | 10,000.00 | 0 |
| sondeo.pl | 10,000.00 | 22 |
| highspeed.pl | 10,000.00 | 57 |
| zyjmyzpasja.pl | 10,000.00 | 0 |
| we-love.pl | 10,000.00 | 3 |
| naprawa-serwis.pl | 10,000.00 | 27 |

[2]Names that appear to be registered trademarks and those registered with Polish diacritic signs are excluded from analysis because they might not exists at time when WHOIS samples were registered.

[3]Obviously, such action may result in fake requests made in order to inflate domain popularity; we believe such activity can be filtered out easily.

a good estimate of domain true value. Both approaches share the same point of view that WHOIS database is currently underutilized as a source of reliable information on domain names. It can be improved that without getting involved in privacy issues.

Such preliminary analysis opens way for future activities. The most needed is repeating the research for current WHOIS records to avoid lags between WHOIS and external data like domain market pricing, cf. Table 5. The most powerful of them and the most complicated at the same time is joint analysis of WHOIS requests and domain registration process. Substantial part of temporal [15] and semantic [16] analysis of DNS registration have been already performed in NASK, the latter one focusing on malbehavior detection. Both tasks are strongly related to the approaches presented here; also, the three publications constitute a strong basis for future joint study on WHOIS requests and domain registration, in economic and safety aspects. Other useful directions are providing requestor classification criteria, mastering the algorithm for word match and promoting registration of domains related to popular dictionary entries.

# References


[1] L. Daigle, “WHOIS Protocol Specification”, RFC Draft Standard, 2004 [Online]. Available: http://tools.ietf.org/html/rfc3912

[2] M. Mueller and M. Chango, “Disrupting global governance: the Internet whois service, ICANN, and privacy”, *J. Inform. Technol. Polit.*, vol. 2, no. 3, 2008.

[3] “Obtaining Bulk Whois Data from ARIN”, 2012 [Online]. Available: https://www.arin.net/resources/request/bulkwhois.html

[4] “Thick vs. Thin Whois for New gTLDs”, ICANN Memorandum, 30 May 2009 [Online]. Available: http://archive.icann.org/en/topics/new-gtlds/thick-thin-whois-30may09-en.pdf

[5] “My Address Lookup and Geo Targeting”, 2012 [Online]. Available: https://www.hostip.info

[6] “Visualization: Geomap – Google Chart Tools”, 2012 [Online]. Available: https://developers.google.com/chart/interactive/docs/gallery/geomap

[7] “Alexa Internet – Our Technology”, 2012 [Online]. Available: http://www.alexa.com/company/technology

[8] “Listy frekwencyjne”, 2012 [Online]. Available: http://pl.wiktionary.org/wiki/Kategoria:Listy_frekwencyjne

[9] “Frequency Word Lists”, 2012 [Online]. Available: http://invokeit.wordpress.com/frequency-word-lists/

[10] V. Levenshtein, “Binary codes capable of correcting deletions, insertions, and reversals”, *Soviet Phys. Doklady*, vol. 10, pp. 707–710, 1966.

[11] “How to Write a Spelling Corrector”, 2012 [Online]. Available: http://www.norvig.com/spell-correct.html

[12] “Alexa Rank”, 2012 [Online]. Available: http://seotracker.pl/alexa-rank.html

[13] “Domain Stock Exchange”, 2012 [Online]. Available: http://www.nazwa.pl/gielda-domen/najciekawsze-domeny/top-lista-domen

[14] “Publishing Domain Name Related Reputation in WHOIS Records”, US Patent Application Publication No. 2006/0095459, 4th May 2006.

[15] P. Arabas, P. Jaskóła, M. Kamola, and M. Karpowicz, “Analysis and modeling of domain registration process”, *J. Telecom. Inform. Technol.*, no. 2, pp. 63–73, 2012.

[16] K. Lasota and A. Kozakiewicz, “Analysis of the similarities in malicious DNS domain names”, in *Proc. Sec. Trust Comput., Data Manag. Appl. STA 2011*, Loutraki, Greece, 2011, Springer, vol. 187, pp. 1–6.



**Mariusz Kamola** received his Ph.D. in Computer Science from the Warsaw University of Technology in 2004. Currently, he is an Associate Professor at Institute of Control and Computation Engineering at the Warsaw University of Technology. Since 2002 with Research and Academic Computer Network (NASK). His research area focuses on technologies supporting modern IP network design, monitoring and management, social networks and other complex systems.

E-mail: Mariusz.Kamola@nask.pl
Research Academic Computer Network (NASK)
Wąwozowa st 18
02-796 Warsaw, Poland

E-mail: mkamola@ia.pw.edu.pl
Institute of Control and Computation Engineering
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland

Paper

# Competition and Diffusion of Telecommunication Services: The Multimedia Communication Services in Brazil

Cristiane Vianna Rauen

*Brazilian National Institute of Metrology, Quality and Technology (Inmetro), Rio de Janeiro, Brazil*

**Abstract—This paper aims to discuss the recent trends in the regulation of telecom services regarding the diffusion of broadband. To this matter, it compares the recent international competition regulation experiences to the Brazilian scenario. Considering that the telecom sector passed through huge technological and structural transformations in the last decades (from typical voice communications to convergent services based on broadband, and from voice services monopoly to multiple services oligopoly), its regulatory framework has also changed. Lately, several countries – such as the United States, the Netherlands, Korea and Chile – have established inter and intra-platform competition regulations in order to expand the access to the new forms of communications based on broadband. In Brazil, such mechanisms are not fully explored, and the country has lots of areas in which there is no competition in broadband services (known as multimedia communication services). As a result, Brazil has one of the highest prices of broadband services and one of the lowest levels of broadband penetration. In this way and considering international success in expanding services through competition, the paper defends that such mechanisms should be reinforced in the Brazilian broadband market.**

***Keywords—Brazil, broadband, competition regulation, international experiences.***

## 1. Introduction

The recent changes in telecommunication sector, such as liberalization, privatization and convergence, brought several challenges to policy makers. The recognition that the growing importance and essentiality of broadband has changed the socio-economic dynamics of countries made governments to change their regulatory agendas in order to introduce mechanisms to stimulate broadband diffusion.

Based on this, the competitive regulatory mechanisms have been introduced in the broadband markets in order to reduce the monopoly power of incumbents and to promote the expansion and modernization of broadband infrastructure mainly to unserved areas. Lately, these regulatory mechanisms are applied in order to encourage intra and inter-platform competitions.

This paper aims to discuss these regulatory measures and their effective impacts in the diffusion of broadband in selected international experiences and in Brazil. Section 2 discusses the recent transformations that have happened in the telecom sector, stressing the important role of broadband as a subject of central appeal in the telecom regulatory agendas. Section 3 shows how these transformations mentioned in the first section have affected the pattern of competition in telecom sector, introducing intra and inter-platform competition, and how these kinds of competition affect broadband diffusion. Section 4 presents the experiences of the USA, the Netherlands, Korea and Chile in competition-based regulation and broadband diffusion. Finally, Section 5 presents the Brazilian case, showing the differences in patterns of competition in the broadband markets of its different regions, and how these differences are correlated to the diffusion of broadband. The Conclusion presents the major achievements of the paper and points some possible measures to be applied in the Brazilian scenario.

## 2. Recent Changes in Telecom Sector

Over the past two decades, the telecom sector has been shaken by huge transformations. One is related to the sector's liberalization and privatization processes that took place around the world. These processes have accelerated the movement of global mergers and acquisitions, especially by developed countries searching for scale gains. The other important change that took place in this sector was the digitalization of telecom networks. This innovation gave rise to Next Generation Networks (NGN) that integrate the existing separate voice and data network into an easier and more flexible network mainly using Internet Protocols (IP). So, these NGN are chiefly based on new platforms such as broadband Internet, 3G mobile networks, wireless LANs and digital televisions [1].

This digitalization of telecom networks provides a convergence of services and technologies – enabling a single network to carry a range of voice, audiovisual and data transmission services – modifying the patterns of competition and innovation among firms and countries, and blurring traditional boundaries in a large range of telecom services. Besides, it is important to stress that this revolution in information and telecommunication technologies (ICT) has also induced major social changes related to forms of communication and information access[1].

The digitalization of telecommunication networks is increasingly extending the limits of this industry, previously restricted to equipment and services for voice transmission. Beyond the rapid spread of mobile telephone services since the 1990s, the increasing speed of technological convergence tends to incorporate different types of media and data services for communication networks. Furthermore, the disseminated use of uploads and downloads requiring higher bandwidths transforms broadband Internet access into an indispensable resource for accessing information today.

In this context, there is growing public concern related to the generalization of availability and access to these new forms of communication, especially in the remote areas and/or to low income populations, which could not be supplied only by the private sector. The most common networks to provide broadband access to the mentioned forms of communication are traditional telecommunications access networks, through the ADSL technology, and the cable TV networks, using different versions of coaxial cables.

Based on this most common market structure, regulatory measures focused on stimulate competition between operators have been explored by regulators all around the world in order to expand broadband infrastructure and services. These competitive regulatory measures are analyzed in the next section.

## 3. Competition in the Broadband Market

Technological and structural changes that took place in telecom sector in the latest decades have changed the kind of regulatory interventions in the market. Previously, competitive stimulus between telecom operators was conceived as inefficient, facing the duplication of networks, and as an obstacle to the expansion of infrastructure investments [5]. But, since the market liberalization and the technological and services convergence processes, competition has been seen as an important mechanism for the diffusion of telecom services, especially of broadband, once it allows costs reduction and increase of penetration rates.

There are two major forms of competition between telecom operators: service-based competition (or intra-platform competition) that is known as the competition between entrants and the incumbent in a single network with regulated access, and platform-based competition (or interplatform competition) known as the competition between different kinds of infrastructure for the provision of services [6], [7]).

[1]Indeed, there is vast economic literature addressing the interpretation and consequences of the ICT Revolution in socio-economic development. An interesting discussion on this issue from an evolutionary perspective is presented by [2]–[4].

The classical example of a platform-based competition is the competition between cable TV operators, providing content, broadband and telephony services, and telecom operators, providing the same kind of services. In this way, the promotion of platform-based competition depends on the level of encouragement supported by the regulator in order to attract investments in alternative networks than the already established in a telecom market.

According to Katz [8], the advantages of platform-based competition are that, because it occurs between companies with their own different infrastructures, it promotes a multidimensional form of competition (in price, service and quality), and it stimulates each operator to invest and to innovate in their own networks. In the same way, Prado [9] affirms that the advantages of platform-based competition are the promotion of the expansion and of the modernization of the operator's networks, leading to the cheapening of the technological solutions.

The classical example of the service-based competition is the competition between entrants (telecom operators that do not have their own networks, and that, because of this, rent a portion of the incumbent's network) and incumbents through the same platform but providing different services. This kind of competition has emerged after the convergence advent. The major regulatory mechanisms applied to guarantee service-based competition are: network interconnection, unbundling, wholesale and structural/functional separation.

The network interconnection is the connection of compatible telecom networks in order to provide communication or service access between users of different networks. The unbundling is the mechanism by which telecom operators are able to buy or rent portions of incumbent's networks to provide services to their own subscribers. The wholesale is the mechanism by which telecom operators are able to acquire network services, by wholesale prices, to resale in retail. Through the structural separation between services and platforms (also known as functional separation), the company that owns the infrastructure is prevented from offering services, in order to avoid vertical integration. Finally, through the structural separation between services, the regulator determines the establishment of different business units to a unique telecom operator. In such a way, incumbents that own fixed telephony networks should only offer telephony services, and not internet access services, for example[2]. All of these regulatory mechanisms aim at avoiding anticompetitive behaviors based on the historical market power of the incumbents. The major advantage of the promotion of service-based competition is the diminishing of the incumbents' market power through the provision of the access for the entrants to the already established networks (especially through the unbundling mechanism). But, according to Bourreau and Dogan [10], although it stimulates competition within existing networks,

[2]This kind of mechanism is rarely used because, by essence, it goes against the convergence trend of telecom sector.

service-based competition could discourage investments in the expansion of alternative networks.

In this way, contrary to the situation of a platform-based competition – that satisfies the objective of the dynamic efficiencies (such as the stimulus in investments in infrastructure and in innovation) [8] – in a service-based competition situation, this stimulus is much lower since competitors (entrants) benefit from investments made by networks owners (incumbents).

As one can imagine, competition between telecom operators has an important role on broadband diffusion. International experiences show that platform-based competition has been responsible for the expansion of broadband access to Internet, especially between cable and ADSL platforms. The next section shows some international experiences of how the regulatory competitive mechanisms are increasing broadband penetration rates.

## 4. Experiences of Competition and Broadband Penetration

Policies and regulations concerning the diffusion of broadband vary among the world's leading ICT countries. However, there is now consensus regarding the growing importance of the role of government in the promotion of a competition-based regulation.

With regard to broadband diffusion, empirical research emphasizes that inter-platform competition drives broadband adoption, while intra-platform competition in DSL is estimated to play a less significant role. The benefits of unbundling are emphasized for countries with a lack of alternative infrastructure, like some Member States of European Union, and less broadband penetration.

As it is presented in this section, DSL and cable are the two most common broadband access technologies nowadays. DSL is by far the dominant broadband access technology in the majority of the cases, with exception of the United States of America. Furthermore, cases show that facility-based competition is, lately, considered by regulators the most effective mechanism to promote broadband diffusion, mainly due to the fact that this kind of competition leads to decreases in prices and to stimulus in investments in the infrastructure expansion and modernization.

### *4.1. The case of the United States of America*

According to Denni and Gruber [11], in the case of the USA, in the beginning of United States Market liberalization, regulatory mandates towards interconnection and unbundling were established to reduce incumbent monopoly power and to stimulate competition in order to increase the provision of broadband services. Initially, intra-platform competition seemed to have a positive impact only on the rate of diffusion but then dissipates. So, the outputs obtained in terms of increasing of broadband penetration rates, new investments and innovation were bellow than expected. Based on this fact, in 2003, the FCC reoriented its policy priorities towards equal access conditions to networks incumbent wireline firms reducing the regulatory effort and in favour of investment incentives that promote inter-platform competition. Two years later, the FCC decided to deregulate the US broadband market totally, eliminating regulatory tariffs and unbundling obligations, allowing mergers and acquisitions, and reinforcing inter-platform competition.

Nowadays, competition in the US broadband market is happening almost exclusively between telecommunications and cable operators and it is stimulating new investments on the expansion of infrastructure and on innovation. As positive results derived from the stimulus to inter-platform competition in the US, the following are [8] the diminishing of 80% in the broadband services prices between 2001 to 2005 (from USD 80.00 per month to USD 15.00 per month on average) and the increasing of the offered average speed. Moreover, data from ITU show that from 2003 (mark of the beginning of encouragement of inter-platform competition in the US) to 2010, the penetration of broadband services had a significant increasing of 190%, from 9.5% to 27.6% (as shown in Table 1).

So, it is possible to consider that in the US experience the inter-platform competition had a much more important role in driving the rate of diffusion for the longer term, and the US model of inter-platform competition was considered the basis for this kind of competition in the world [8].

However, as considered by Picot and Wernick [7], by relying solely on the benefits of platform competition, the US is pursuing a different path than Korea and Europe. Despite its attempts in spurring the broadband development by public initiatives on the local level, the role of the US government in furthering broadband deployment can be interpreted as rather passive. So, the passive role of the US regulation with regard to supply- and demand-side activities linked with deregulatory efforts in market regulation have furthered massive investments by incumbent operators in NGN, but not yet succeeded in bridging the lack in relation to penetration rates in leading broadband economies.

### *4.2. The case of the Netherlands*

As well as in the case of the USA, the broadband market of the Netherlands has two major operators: telecommunications (KPN representing the major incumbent) and cable. Regulation in the Netherlands is based on the rules related to unbundling and open access, but the major form of competition is inter-platform.

This competitive dynamic has encouraged operators to increase their investments on infrastructure expansion, as well as on the modernization of networks based on NGN services, which allowed the Netherlands to achieve the highest level of broadband penetration in the world (38%, in 2010, as shown in Table 1).

Differently from the US broadband market – in which the cable operator has the largest share (54%) – in the Netherlands, KPN is still dominant in the market (44%). This is

Table 1
USA, Korea, Netherlands and Chile: Competition and impacts on broadband market

| Country | Kind of competition encouraged | Market share | Broadband penetraton rate (2010) | Monthly minimum broadband price (USD PPP) |
|---|---|---|---|---|
| United States | From Service-based to Platform-based | Cable: 54%<br>Telecom 1: 20%<br>Telecom 2: 12% | 27.6% | USD 15.00 |
| Korea | Platform-based | Telecom 1: 45%<br>Telecom 2: 26%<br>Telecom 3: 10%<br>Cable: 19% | 35.7% | USD 31.00 |
| Netherlands | Platform-based | Telecom 1: 44%<br>Cable: 39% | 38.1% | USD 8.00 |
| Chile | From Service-based to Platform-based | Telecom 1: 50%<br>Cable: 40% | 10.5% | USD 55.00 |
| Source: [8], [13]. | | | | |

due to the fact that the incumbent has always been committed to invest on the modernization and expansion of its networks, although obliged to share access of its infrastructure with entrants [8].

This situation of inter-platform competition combined to increasing of investments from the both sides (telecom operators and cable operators) has led to decreasing of prices and increasing of broadband penetration rates. In relation to prices, it is shown that the minimum monthly price of broadband subscription in the Netherlands is USD 8.00 (PPP), one of the lowest prices in the world [8].

As one can notice, the Netherlands could be considered as the best succeeded case of outputs of broadband based on inter-platform competition.

### *4.3. The case of Korea*

Based on the fact that in Korea the model of regulation was focused on the encouragement of inter-platform competition, as well as in the government intervention (through the protection of national groups and the election of the "national champions"), it is suggested that it is possible to classify the case of Korea as a hybrid model of competition regulation [8]. This model of "administered competition" is responsible to transform Korea in one of the most important players of telecom sector in the world and, at the same time, it has enabled the country to achieve one of the highest levels of broadband penetration (35.7%, as shown in Table 1).

The leading position of Korea has been furthered by platform competition between DSL and cable modem [7]. While LLU played a negligible role, open access obligations for cable owners were important for new entrants to compete on a playing field level.

In this way, platform competition between the incumbent, KT (45% of market share, as shown in Table 1), offering broadband by DSL, and Thrunet and Hanaro (26% of market share, as shown in Table 1), being dependent on leasing Cable TV networks, at least in the early phase after market entry, contributed significantly to the launch of broadband markets in Korea [12]. Lately, based on the sanction of regulator for the provision of bundled services (multiple services packages), a price war has been established between Korean major operators [8]. So, it is possible to notice that inter-platform competition in Korea led to rapidly decreasing price levels of broadband, which furthered broadband adoption, as well as to the increasing of infrastructure expansion and modernization investments.

### *4.4. The case of Chile*

As well as in the case of the US, in Chile, the regulator has also reoriented its strategy towards the encouragement of competition. After the liberalization of Chilean telecom market, in 1990's, regulatory measures, such as unbundling and wholesale through the incumbent's infrastructure were used in order to promote intra-platform competition. In 2004, the regulator proposed the adoption of the structural separation between services and platforms, a mechanism that prohibits the operators that offer telecom services to use their own infrastructure to this matter. This mechanism reinforced the option of Chilean regulation towards the encouragement of intra-platform competition.

However, as it is presented in [8], from 2000 to 2005, the level of infrastructure investments on broadband networks dropped significantly. This fact influenced the Chilean regulator to change, in 2006, the regulation strategy towards the encouragement of inter-platform competition. As one can notice, the change of the regulatory orientation that took place in Chile is the same that the one observed in the US – from service-based competition to platform-based competition.

Nowadays, the Chilean broadband market has two major competitors: the incumbent operator, Telefonica, that is responsible for 50% of the broadband market, and the cable operator, VTR, that is responsible for 40% of it (Table 1).

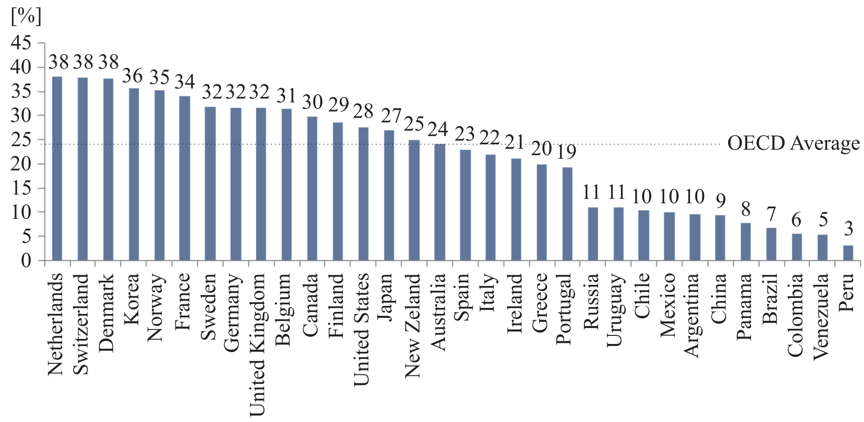



***Fig. 1.*** Selected countries: Broadband penetration rates (2010) [13].

Although close to a duopoly, the structure of Chilean broadband market has intense competition, what is confirmed by important indicators, such as, penetration level of broadband (10.5%, as shown in Table 1 – the highest in Latin America) and the drop of broadband prices (50%, from 2002 to 2008) [8].

## 5. The Brazilian Broadband Market

After the process of market liberalization that occurred at the end of 1990's, the General Telecommunications Law [14] established some regulatory mechanisms in order to stimulate competition between operators (such as unbundling and interconnection obligations). After several processes of mergers and acquisitions, today, three major competitors are responsible for 80% of the provision of multimedia communication services, i.e., broadband services[3]. Oi, the major telecom incumbent, is responsible for 34.4% of the market; NET, the major cable operator, for 22.65%; and Telefonica, the telecom incumbent authorized to explore the richest region of the country – the São Paulo state – is responsible for 21.77% of the market (see Table 2).

Table 2
Market Shares of the majors multimedia communication services providers in Brazil

| Operator | Market Share |
|---|---|
| Oi | 34.4% |
| NET | 22.65% |
| Telefonica | 21.77% |
| Source: [15]. | |

The Brazilian market structure is quite similar to those of the cases mentioned above (the USA, the Netherlands, Korea and Chile), and the kind of competition achieved, inter-platform competition – between DSL and cable, is the same as the successful cases indicate as the most effective in broadband diffusion. So, why the Brazilian broadband penetration level, prices and infrastructure investments are one of the worst in the world?

Brazil has one of the smallest rates of broadband penetration in the world (6.81%, in 2010, according ITU database). It represents the 15th place in penetration rates of Latin America (behind countries like Uruguay – 10.91%, Chile – 10.45%, Mexico – 9.98%, Argentina – 9.56%, and Panama – 7.84%), and it is far beyond the average of 24% of broadband penetration in OECD countries (as shown in Fig. 1).

Moreover, the Brazilian monthly prices of broadband subscription are also one of the highest in the world. As shown in Fig. 2, the relative price of broadband connection in Brazil (the minimum basket of the service divided by GDP per capita) is nine times higher than in the United States, eight times higher than in Denmark and Switzerland, six times higher than in Canada, five times higher than in Sweden, the Netherlands and Finland, four times higher than in Japan and Germany, and three times higher than in Korea. The broadband connection index in Brazil is even higher than in some developing countries, such as Mexico and Venezuela, but lower than in other developing countries like India, Chile and Colombia.

However, as a country of continental size, with a huge range of dispersion of income and population, it is important to analyze the Brazilian case through the differences between its regions. As shown in Table 3, the concentration ratio (CR1) of the largest telecom company in each Brazilian region is high, especially in North, Northeast and Center-West regions. This means that the market power of a single company is significantly high, which threatens competition and, as a consequence, the diffusion of broadband.

In the North region, it is possible to observe the worst situation in terms of competition. The level of CR1 is higher than 80%, which means that a single company, the telecom incumbent Oi, is responsible for almost the totality

[3]ANATEL, the Brazilian telecom sector regulator, determines that companies interested in providing broadband services can only do so with the authorization issued by ANATEL to explore the multimedia communication services.

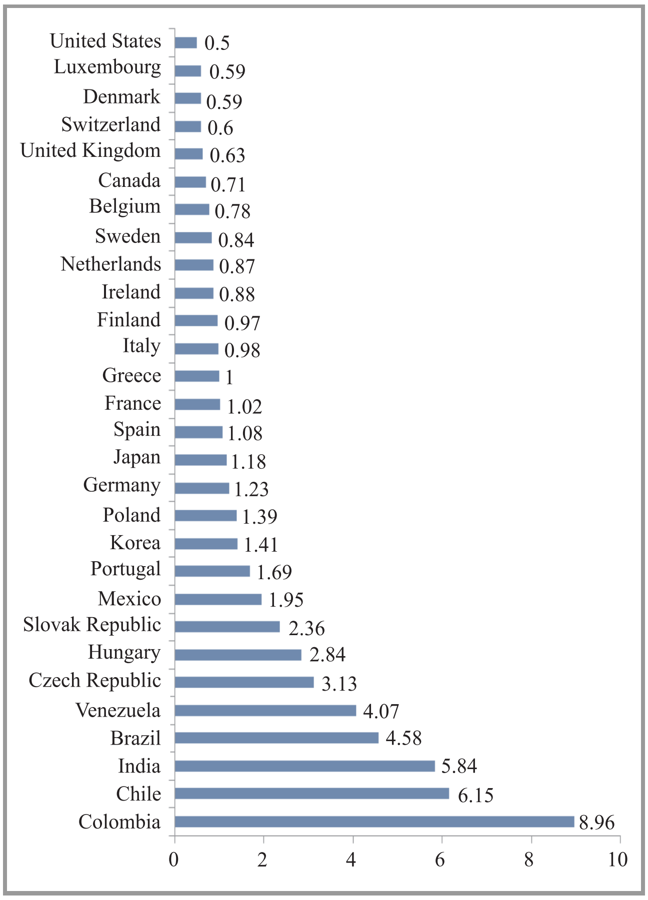



*Fig. 2.* Selected countries: Relation between broadband monthly subscription cost and GDP per capita (2009) [16].

of the market. Besides the fact the Northern region is the most concentrated broadband market of the country, the predominant kind of competition is the intra-platform, or service-based, competition (see Annex I). The exception is the Amazonas state, in which it is possible to observe inter-platform competition between the cable company NET (52.52% of market share) and the telecom company Oi (36.32% of market share). In this way, the North region presents the lowest percentage of cities with broadband access in the country (62%), and a level of penetration of broadband services that is far below the Brazilian average (3.75%) – comparable to the penetration levels of countries like Peru (3.14%) and Suriname (2.99%) [13].

Table 3
Brazil: Broadband market by region

| Region | CR1 [%] | Broadband penetration [%] | Cities with broadband access [%] |
|---|---|---|---|
| North | 83.2 | 3.75 | 61.7 |
| Northeast | 65.1 | 1.46 | 66.4 |
| Center-West | 63.9 | 6.97 | 96.6 |
| South | 47.8 | 8.27 | 90.5 |
| Southeast | 56.8 | 11.24 | 91.2 |
| Source: [15], [17]. | | | |

The situation of the Northeast region in terms of competition is similar to the one observed in the North region and is highly concentrated. Oi, the largest company, has 65% of the Northeast market share; in almost half of the Northeast states the largest company has more than 70% of market share, and in one third of the Northeast states the major kind of competition between telecom operators is intra-platform (see Annex I). Because of this, the Northeast region presents the worst level of broadband penetration in the country (1.46%) – one of the lowest in the world – and only 66.4% of the cities of this region have access to broadband networks.

Although the Center-West region also presents a highly concentrated market structure (the largest company, Oi, has 64% of market share), the major form of competition between broadband companies is inter-platform, especially between the telecom company – Oi and the cable companies, GVT and NET. This situation has led to levels of broadband infrastructure access and penetration in the Center-West region that are above the Brazilian average (96.6% and 6.97%, respectively).

Nevertheless, the best situations in the Brazilian broadband market are presented in the South and Southeast regions. In the South region, the level of broadband penetration is about two percentage points higher than the average of the country; the level of CR1 is below 50% (which represents the best competitive situation of inter-platform competition in the Brazilian broadband market – especially between the telecom operator Oi, and the two cable companies, NET and GVT), and the percentage of cities with access to broadband infrastructure is higher than 90%.

Finally, it is possible to observe that, in terms of market concentration, the Southeast has two major telecom companies that dominate the market[4]: Telefonica, in São Paulo state, and Oi, in Rio de Janeiro state, Minas Gerais and Espírito Santo that together have 56.8% of market share. But, as well as observed in the South region, in the Southeast region there is also a strong inter-platform competition between these two telecom companies and the cable companies NET and GVT (see Annex I). This strong presence of inter-platform competition has led the Southeast region to achieve the highest level of broadband penetration in Brazil (11.24 %, comparable, for instance, to some Latin American countries like Chile, 10.45%, and Uruguay, 10.91%, and to Russia, 10.98%) and the highest percentage of cities with broadband infrastructure access in the country (91.2%).

As noticed in this section, although in the aggregate level Brazil presents a situation of low level of concentration in the structure of broadband market and inter-platform competition (especially between telecom and cable companies),

[4]Based on the General Concession Plan [18], the country was divided in regions to be explored by the private company utilities. Due to the fact that São Paulo state is the richest and most populated region of the country, it was considered an isolated concession region, and Telefonica won the concession to explore it. The concession to explore the other states that compound the Southeast region (Rio de Janeiro state, Minas Gerais and Espírito Santo) was won by Oi.

the penetration rate of broadband is still low, even in comparison to other Latin American countries, and the prices of broadband monthly subscription are one of the highest in the world. The low level of penetration and high prices of broadband are the symptoms of a fragile competitive structure. So, facing the fact that Brazil is a huge country, with regions of different levels of density, income and competition, this paper analyzed the different patterns of competition between regions of the North, Northeast, Center-West, South and Southeast, in order to explain the country's lag in a broadband diffusion.

As observed, in the North and Northeast regions – that present high levels of CR1 and cases of intra-platform competition – one can find the lowest levels of broadband penetration and the lowest percentage of cities with access to broadband infrastructure. On the other hand, in the Center-West, South and Southeast regions, where inter-platform competition and lower levels of CR1 are presented, one can find the highest percentages of cities with access to broadband infrastructure and levels of broadband penetration above the average of the country.

## 6. Conclusion

It is often claimed among practitioners and policy makers that broadband adoption can be stimulated more effectively by promoting competition between different platforms (inter-platform competition), rather than focusing on the market for DSL services (intra-platform competition). The international cases and the analysis of Center-West, South and Southeast Brazilian regions presented in this paper can confirm such a claim.

The international experiences show that, although elected in the early stages of telecom markets liberalization, the regulatory mechanisms directed towards reducing incumbent's market power (such as unbundling, interconnection, wholesale and structural/functional separation) turned out to be insufficient in promoting broadband diffusion, drop in prices and infrastructure expansion and modernization. Nevertheless, while platform competition seems to have much impact on a high deployment rate, especially in metropolitan areas, LLU can contribute to broadband diffusion in regions and countries lacking of alternative infrastructure.

In the Brazilian case, in which broadband penetration diffusion is far behind in comparison to several countries – including countries in Latin America – the regional analysis confirmed that, in the regions where there is inter-platform competition, higher levels of broadband diffusion were achieved.

In this way, this paper states that the strengthening of regulatory and political measures towards the increasing of inter-platform competition is mandatory in order to eliminate the digital divide between regions, as well as to increase the level of broadband penetration of the country as a whole.

Finally, it is important to stress that Brazilian hindrances to the expansion of broadband penetration are much bigger and more striking than those faced by developed countries, such as: huge income concentration, educational and instruction shortcomings and unavailability of some essential services in remote areas of the country.

Thus although public and regulatory initiatives appear to be in line with those taken by some OECD countries, it is necessary to consider that in the context of developing countries like Brazil the government involvement in directing the private sector and in planning and implementing wise public policies is of paramount importance to successfully mitigating the digital divide, and to achieve real conditions for broadband diffusion.

# Appendix

## Brazil: Broadband market structure by region and state

| NORTH REGION | | |
|---|---|---|
| States | | |
| ***Acre*** | | |
| Company | Network | Market share [%] |
| Oi | DSL | 93.28 |
| Embratel | DSL | 2.93 |
| ***Amazonas*** | | |
| Company | Network | Market share [%] |
| Oi | DSL | 52.52 |
| Embratel | DSL | 36.32 |
| ***Ampá*** | | |
| Company | Network | Market share [%] |
| Oi | DSL | 67.95 |
| Embratel | DSL | 24.15 |
| ***Pará*** | | |
| Company | Network | Market share [%] |
| Oi | DSL | 87.59 |
| Embratel | DSL | 6.93 |
| ***Rondônia*** | | |
| Company | Network | Market share [%] |
| Oi | DSL | 95.15 |
| Embratel | DSL | 2.15 |
| ***Roraima*** | | |
| Company | Network | Market share [%] |
| Oi | DSL | 92.54 |
| Embratel | DSL | 6.74 |
| ***Tocantins*** | | |
| Company | Network | Market share [%] |
| Oi | DSL | 93.02 |
| Embratel | DSL | 1.97 |

| NORTHEAST REGION | | |
|---|---|---|
| States | | |
| ***Maranhão*** | | |
| Company | Network | Market share [%] |
| Oi | DSL | 76.23 |
| Embratel | DSL | 13.85 |

| | | |
|---|---|---|
| ***Piauí*** | | |
| Company | Network | Market share [%] |
| Oi | DSL | 88.47 |
| Embratel | DSL | 5.34 |
| ***Ceará*** | | |
| Company | Network | Market share [%] |
| Oi | DSL | 57.63 |
| GVT | Cable | 22.16 |
| Videomar | Radio | 9.48 |
| ***Rio Grande do Norte*** | | |
| Company | Network | Market share [%] |
| Oi | DSL | 87.59 |
| Cabo Serviços de Telecom ltda. | Cable | 6.93 |
| ***Paraiba*** | | |
| Company | Network | Market share [%] |
| Oi | DSL | 47.55 |
| NET | Cable | 27.96 |
| GVT | Cable | 13.47 |
| ***Sergipe*** | | |
| Company | Network | Market share [%] |
| Oi | DSL | 84.32 |
| Embratel | DSL | 4.34 |
| ***Pernambuco*** | | |
| Company | Network | Market share [%] |
| Oi | DSL | 51.94 |
| GVT | Cable | 31.12 |
| Embratel | DSL | 4.66 |
| ***Bahia*** | | |
| Company | Network | Market share [%] |
| Oi | DSL | 56.2 |
| GVT | Cable | 16.08 |
| Embratel | DSL | 4.17 |
| CENTER-WEST REGION | | |
| States | | |
| ***Goiás*** | | |
| Company | Network | Market share [%] |
| Oi | DSL | 55.84 |
| GVT | Cable | 19.43 |
| NET | Cable | 15.08 |
| ***Mato Grosso*** | | |
| Company | Network | Market share [%] |
| Oi | DSL | 76.16 |
| GVT | Cable | 14.62 |
| ***Mato Grosso do Sul*** | | |
| Company | Network | Market share [%] |
| Oi | DSL | 59.59 |
| NET | Cable | 15.30 |
| GVT | Cable | 14.02 |
| SOUTHEAST REGION | | |
| States | | |
| ***Minas Gerais*** | | |
| Company | Network | Market share [%] |
| Oi | DSL | 51.35 |
| NET | Cable | 14.74 |
| CTBC | Optical Fiber | 11.75 |
| GVT | Cable | 5.82 |
| ***Espírito Santo*** | | |
| Company | Network | Market share [%] |
| Oi | DSL | 51.59 |
| GVT | Cable | 24.89 |
| NET | Cable | 13.44 |
| ***Rio de Janeiro*** | | |
| Company | Network | Market share [%] |
| Oi | DSL | 68.30 |
| NET | Cable | 21.49 |
| ***São Paulo*** | | |
| Company | Network | Market share [%] |
| Telefonica | DSL | 55.87 |
| NET | Cable | 33.31 |
| SOUTH REGION | | |
| States | | |
| ***Paraná*** | | |
| Company | Network | Market share [%] |
| Oi | DSL | 38.37 |
| GVT | Cable | 28.61 |
| NET | Cable | 12.75 |
| ***Santa Catarina*** | | |
| Company | Network | Market share [%] |
| Oi | DSL | 62.28 |
| NET | Cable | 17.34 |
| GVT | Cable | 12.24 |
| ***Rio Grande do Sul*** | | |
| Company | Network | Market share [%] |
| Oi | DSL | 42.85 |
| NET | Cable | 26.44 |
| GVT | Cable | 18.03 |
| Source: [18]. | | |

# References


[1] "The Implications of Convergence for Regulation of Electronic Communications". Directorate for Science, Technology and Industry, OECD, 2004.

[2] C. Freeman and F. Louçã, A*s Time Goes By: From Industrial Revolutions To Information Revolution*. Oxford, England: Oxford University Press, 2002.

[3] B. Verspagen, "Structural change and technology: a long view", *Révue Économique*, vol. 55, no. 6, pp. 1099–1125, 2004.

[4] "OECD Information Technology Outlook 2002", OECD, Directorate for Science, Technology and Industry, 2002.

[5] C. Blackman, "Universal service: obligation or opportunity?", *Telecomm. Policy*, vol. 19, no. 3, pp. 171–176, 1995.

[6] M. Falch, "Penetration of broadband services – the role of policies", *Telematics and Informatics*, vol. 24, no. 4, pp. 246–258, 2007.

[7] A. Picot and C. Wernick, "The role of government in broadband access", Telecommunications Policy, vol. 31, no. 10–11, pp. 660–674, 2007.

[8] R. Katz, *Competencia entre plataformas: teoria y resultados*. Madrid: ENTER, 2008.

[9] L. C. D. Prado, "A convergência tecnológica e seus impactos concorrenciais" (Technology convergence and its impacts on competition). Conselho Administrativo de Defesa Econômica. Audiência Pública, Processo Administrativo no. 08700.001638/2007-67, 2007.

[10] M. Bourreau and P. Dogan, "Regulation and innovation in the telecommunications industry", *Telecom. Policy*, vol. 25, pp. 167–184, 2001.

[11] M. Denni and H. Gruber, "The diffusion of broadband telecommunications: the role of competition". Universitá Degli Studi Roma Tre, Dipartimento di Economia Working, paper no. 60, 2006.

[12] *Global Broadband Battles – Why the US and Europe Lag While Asia Leads*, M. Fransman, Ed. Stanford, CA: Stanford Business Books, 2006.

[13] International Telecommunication Union [Online]. Available: www.itu.int

[14] "Lei Geral das Telecomunicações" (General Telecommunications Law). Law no. 9.472 of July, 16th of 1997, Brasil, 1997.

[15] Sici/Anatel [Online]. Available: www.anatel.gov.br

[16] "Measuring the Information Society – The ICT Development Index", ITU, Geneva, Switzerland, 2010.

[17] "Barômetro Cisco de Banda Larga – Estimativas de junho de 2010. (Cisco Barometer of Broadband – Estimates for June 2010), IDC, (2010), accessed in June 2010 [Online]. Available: http://www.cisco.com/web/BR/assets/docs/Barometro_1H10_01Set10_Atualizado_Site.pdf.

[18] "Plano Geral de Outorgas" (General Concession Plan). Decree no. 2.534 of April, 2nd of 1998, Brasil, 1998.


**Cristiane Vianna Rauen** B.A. in Economics (Campinas College FACAMP, Brazil, 2003), Master in Science and Technology Policy (State University of Campinas UNICAMP, Brazil, 2007), Ph.D. candidate in Economics (State University of Campinas, Brazil) and Doctoral Visiting Scholar at Columbia University (Columbia Institute for Tele-Information, USA, 2011). She is currently a government employee at the Brazilian National Institute of Metrology, Quality and Technology (INMETRO), where she holds the position of Researcher in Technology and Innovation and is the head of INMETROñS Business Incubator. She has academic and research experience in the Economics field, with emphasis in Industrial Economics: Industrial and Technology Economics, Industrial Competitiveness, Economic Regulation and Public Policy.


E-mail: crisrauen@gmail.com
Brazilian National Institute of Metrology, Quality and Technology (INMETRO)
Santa Alexandrina St, 416
Rio Comprido – Rio de Janeiro
RJ, Brazil

Paper

# On the Design of the TCP/AQM Traffic Flow Control Mechanisms

Michał P. Karpowicz

*Research and Academic Computer Network (NASK), Warsaw, Poland*

**Abstract—Several aspects of the TCP/AQM system design are discussed that may affect performance of the network. Namely, due to decentralized structure of the network traffic flow control system in which traffic rate control tasks are delegated to autonomous agents, it may be possible for the agents to profitably re-engineer the TCP congestion control algorithm at the cost of the overall performance of the network. In this paper it is shown how the commonly applied TCP/AQM design procedures may give rise to mechanisms that are prone to attacks discreetly moving the network traffic flow away from the desired operating point. Furthermore, a short discussion is presented concerning the countermeasures that can be taken to reduce these effects.**

***Keywords*—*game theory, network security, TCP/AQM, traffic flow control.***

## 1. Introduction

The transmission control protocol (TCP) congestion control algorithm adjusts packet transmission rate of a TCP node to locally observed network congestion signals, usually related to the rate of incoming transmission acknowledgment (ACK) messages. Since the congestion signals close the feedback loop between the interacting network nodes, thereby providing information about the utilization of the network's links, it is natural to interpret TCP and active queue management (AQM) mechanisms as elements of a distributed control system optimizing performance of the network.

In this brief paper the above point of view is taken to address a specific question regarding the expected performance of the popular TCP/AQM designs. Namely, a discussion is presented concerning the implications that *information asymmetry* between the interconnected network nodes may have on the overall performance of the network. The discussion is supported with the formal arguments that we have derived and studied in [1], [2].

The notion of information asymmetry which is being referred to here describes the setting in which the information about the control goals that are addressed at the level of a given node is not available to other nodes, but is essential for the efficiency of the interactions. It should be pointed out that this is a typical setting for the networked systems in general. Indeed, it results quite naturally from the decentralized structure of the network in which control tasks are delegated to autonomous elements of the system. It is the delegation of the tasks that creates an *information gap* in the system and severely limits the ability to control locally taken actions.

To be somewhat more specific about the addressed question, let us for a moment think of the network as of a distributed system of interacting plants, each one being controlled by an autonomous and intelligent (active) agent. An active agent should be identified with a network node administrator capable of implementing an arbitrary operating configuration of the node. Suppose next that we are faced with the problem of designing a distributed mechanism that leads the interacting plants, managed by the agents, to a solution of a given system performance optimization problem. Clearly, in the considered TCP/AQM networking context this amounts to designing traffic rate control and queue management algorithms that, if installed in the network's nodes, support reaching a desired operating point of the network. The performance optimization problem represents preferences defined with respect to the operating point, which in turn is specified by the performance index being optimized. This brings us to the observation that draws our attention here. A common engineering practice is to evaluate the system's performance by means of a scalarizing function, aggregating individual performance indices of the interacting plants. Indeed, under sufficient regularity conditions optimization of the scalarizing function leads to Pareto-optimal outcome of the addressed multiobjective problem, i.e., the problem in which each plant tries to reach its optimal performance subject to the interaction constraints. The question of our interest is related to the possibility of reaching the solution to the performance optimization problem. Namely, in the addressed asymmetric information environment the only way to calculate the value of the network-wide performance index, aggregating agents' performance, is to process information that the agents reveal in the course of their interactions. In other words, performance of the system can only be optimized if the preferences of the agents are known to the mechanism that coordinates the agents' interactions. However, since the agents remain autonomous in their decisions, they may find a way to reveal a profitably modified information on their performance, thereby taking advantage of the monopoly they have on the information regarding their individual goals. It, therefore, becomes clear that in such a case efficient performance of the system can be questioned.

In the following sections the above problem is addressed in the context of the TCP/AQM control system design. It is demonstrated that commonly applied design procedures may give rise to mechanisms that are prone to a specific kind of actions, or even intentional attacks, moving the network traffic flow away from the desired operating point.

These attacks may seem to resemble the well known TCP SYN flooding technique, exploiting the way of establishing a new TCP connection [3], [4]. However, as it is to be illustrated, in the considered case the architecture of attack is different and may result in *degradation of service* rather that in *denial of service* (DoS). In the following section the countermeasures are also discussed that can be taken by the designer of the TCP/AQM mechanisms to reduce the potentially adverse effects of distributed control under asymmetric information. Our results are built upon the models studied in [5]–[7], and recently in [1], [2]. Technical details concerning implementation of TCP and AQM algorithms can be found, e.g., in [8], [9]. For a survey of results on information asymmetry see e.g., [10].

## 2. Traffic Flow Control Mechanism Design

Consider a network that consists of $E$ links shared by transmission streams originating from $S$ sources. Each source $s = 1,\dots,S$ is associated with a connection that is realized between a specified source-destination pair of the network's nodes. The connection can be established by a collection of $P_s,\ s = 1,\dots,S$, paths. Multi-path routing, which is admitted in the considered setting, is defined by routing matrices $\mathbf{A}_s,\ s = 1,\dots,S$, consisting of elements $a_{spe} = 1$ if link $e$ is used by the $p$-th route originating from source $s$ and $a_{spe} = 0$ otherwise.
With each source there is associated a transmission performance index $U_s,\ s = 1,\dots,S$. The value $U_s(x_s)$ can be interpreted as a utility that transfer rate $x_s \geq 0$ has to the source. Similarly, with each link $e = 1,\dots,E$ there is associated a performance index $C_e$ defining value of cost $C_e(y_e)$ at which network link serves flow $y_e \geq 0$ of incoming data.
A two-step design procedure will be now discussed that is usually applied to engineer a stable and efficient process of network traffic flow control. In the first step, a reference point for the design is defined to represent a preferred outcome of the network performance optimization problem. In the second step, based on the desired performance conditions satisfied by the selected reference point, a dynamic system is constructed that is guaranteed to converge to a neighborhood of the reference point. The definition of the dynamic system is then used as a design guideline for the algorithms implemented in the network.
Let us illustrate the above procedure. The basic problem that underlies current designs of TCP/AQM algorithms, mostly due to [5], [11], is defined as follows:

$$\texttt{SYSTEM}(\mathbf{U},\mathbf{A},\mathbf{c}):\quad \left|\begin{array}{ll} \text{maximize} & \sum_s U_s(x_s) \\ \text{subject to} & \mathbf{A}^T\mathbf{x} \leq \mathbf{c} \\ \text{over} & x_s \geq 0,\ s = 1,\dots,S, \end{array}\right.$$

where $\mathbf{c}^T = [c_1,\dots,c_e]$ is a vector of fixed capacities of the links. A problem SYSTEM defines the reference point for the design of distributed network flow control system. As it can be easily seen, the basic problem is *multi-objective*, with the objectives defined by the performance indices of the network elements given by functions $U_s,\ s = 1,\dots,S$, and $C_e,\ e = 1,\dots,E$. In the above formulation, efficient (undominated) solutions are searched for by means of a simple (utilitarian or maxisum) scalarization. For other interesting approaches see, e.g., [12]–[15]. Notice that a single path routing matrix $\mathbf{A}$ is used here. Furthermore, it follows that $C_e(y_e) = 0$ for $0 \leq y_e \leq c_e$ and $C_e(y_e) = +\infty$ whenever $y_e > c_e$.
The following sequence of arguments supports the construction of a distributed traffic flow control system that is intended to lead the network to a neighborhood of the reference point defined above. Suppose that each source submits to the network a bid $\theta_s \geq 0$ denoting willingness to pay for the traffic rate $x_s = \theta_s/\lambda_s \geq 0$, where $\lambda_s \geq 0$ can be regarded as a charge per unit traffic flow. Let us also assume that each source, *taking* $\lambda_s = \sum_e a_{se}\mu_e > 0$ *as given*, chooses $\theta_s$ that maximizes *payoff* related to the assigned transfer rate, i.e., it solves the problem:

$$\texttt{USER}_s(U_s,\lambda_s):\quad \left|\begin{array}{ll} \text{maximize} & U_s(\theta_s/\lambda_s) - \theta_s \\ \text{over} & \theta_s \geq 0. \end{array}\right.$$

Next, suppose that given the vector $\boldsymbol{\theta} = (\theta_1,\dots,\theta_S)$ of bids, the network calculates prices (Lagrange multipliers, congestion signals) $\mu_e \geq 0,\ e = 1,\dots,E$, and rates $x_s \geq 0,\ s = 1,\dots,S$, solving the problem:

$$\texttt{NETWORK}(\mathbf{A},\mathbf{c},\boldsymbol{\theta}):\quad \left|\begin{array}{ll} \text{maximize} & \sum_s \theta_s \log(x_s) \\ \text{subject to} & \mathbf{A}^T\mathbf{x} \leq \mathbf{c} \\ \text{over} & x_s \geq 0,\ s = 1,\dots,S. \end{array}\right.$$

The reason for application of the specific form of utility function, $U_s(x_s) = \theta_s \log(x_s)$, will become clear in a moment. Theorem 1, presented below, shows that under the assumption that the sources take the feedback signals $\bar{\boldsymbol{\lambda}}$ *as given*, a feasible network traffic flow can be found, which is arbitrarily close to solution of the problem $\texttt{SYSTEM}(\mathbf{U},\mathbf{A},\mathbf{c})$. Namely, the flow can be reached by a distributed algorithm that solves the problem $\texttt{NETWORK}(\mathbf{A},\mathbf{c},\boldsymbol{\theta}(t))$ at instant $t$ and, on a larger time scale, drives $\boldsymbol{\theta}(t)$ to $\bar{\boldsymbol{\theta}}$ defining optimal solution to $\texttt{SYSTEM}(\mathbf{U},\mathbf{A},\mathbf{c})$.

*Theorem 1:* Suppose that $U_s$ is an increasing, strictly concave and continuously differentiable function over $x_s \geq 0$ for $i = 1,\dots,n$. There exist vectors $\bar{\boldsymbol{\lambda}} = (\bar{\lambda}_1,\dots,\bar{\lambda}_n)$, $\bar{\boldsymbol{\theta}} = (\bar{\theta}_1,\dots,\bar{\theta}_n)$ and $\bar{\mathbf{x}} = (\bar{x}_1,\dots,\bar{x}_n)$ such that:

– $\bar{\theta}_s$ solves $\texttt{USER}_s(U_s,\bar{\lambda}_s)$, $s = 1,\dots,S$;

– $\bar{\mathbf{x}}$ solves $\texttt{NETWORK}(\mathbf{A},\mathbf{c},\bar{\boldsymbol{\theta}})$ and $\texttt{SYSTEM}(\mathbf{U},\mathbf{A},\mathbf{c})$;

– $\bar{\theta}_s = \bar{\lambda}_s\bar{x}_s$, $s = 1,\dots,S$;

*Proof:* For the proof see, e.g., [5]. ■

Based on the above result, a distributed traffic flow control algorithm can be derived from the following system of differential equations:

$$\begin{aligned}&\mathtt{RATE}(\mathbf{A},\mathbf{q},\boldsymbol{\theta}):\\ &\left|\begin{array}{l}\dfrac{\mathrm{d}x_s}{\mathrm{d}t}(t)=\kappa\left[\theta_s-x_s(t)\sum_e a_{se}\mu_e(t)\right],\ s=1,\ldots,S,\\ \mu_e(t)=q_e(\sum_s a_{se}x_s(t)),\ e=1,\ldots,E.\end{array}\right.\end{aligned}$$

System $\mathtt{RATE}(\mathbf{A},\mathbf{q},\boldsymbol{\theta})$ describes a network-wide traffic rate adjustment process, $x_s(t),\ s=1,\ldots,S$, with feedback provided to each network node by congestion signals $\mu_e(t)$, $e=1,\ldots,E$. By construction, the system exploits the properties of the equilibrium point described in Theorem 1, applying function $\bar{U}_s(x_s)=\theta_s\log(x_s),\ s=1,\ldots,S$, as a model of preference indicator applied by each traffic source. Precisely, the construction of the desired distributed control algorithm rests on the following result. Given a fixed signal $\theta_s,\ s=1,\ldots,S$, and a well-behaved and suitably designed function $q_e,\ e=1,\ldots,E$, system $\mathtt{RATE}(\mathbf{A},\mathbf{q},\boldsymbol{\theta})$ converges to minimum of function:

$$\sum_e\int_0^y q_e(s)\mathrm{d}s-\sum_s\theta_s\log x_s,\ \text{with } y=\sum_s a_{se}x_s. \qquad (1)$$

It follows that by letting the sources periodically update signals $\theta_s$, a neighborhood of solution to problem $\mathtt{SYSTEM}(\mathbf{U},\mathbf{A},\mathbf{c})$ can be reached. Indeed, suppose that each source applies the following control rule:

$$\theta_s(t)=x_s(t)U_s'(x_s(t)),\ s=1,\ldots,S, \qquad (2)$$

where $U_s'\equiv\mathrm{d}U_s/\mathrm{d}x_s$. It can be demonstrated that the above evolution of $\boldsymbol{\theta}(t)$ defines TCP traffic rates $\mathbf{x}(t)$ that converge to a stable point $\bar{\mathbf{x}}$ minimizing:

$$\sum_e\int_0^y q_e(s)\mathrm{d}s-\sum_s U_s(x_s),\ \text{with } y=\sum_s a_{se}x_s. \qquad (3)$$

For a suitable choice of function $q_e,\ e=1,\ldots,E$, representing an AQM policy, the above function arbitrarily closely approximates objective in $\mathtt{SYSTEM}(\mathbf{U},\mathbf{A},\mathbf{c})$. Thus vector $\bar{\mathbf{x}}$ of traffic rate solves relaxation of the network performance optimization problem. Details concerning practical implementations of the above algorithm can be found, e.g., in [16]–[22].

***Example: FAST TCP algorithm***

To control the rate at which packets are transmitted, the FAST TCP algorithm updates the amount of data transmitted into the network at a given time, based on the observed average round-trip time and average queuing delay [23]–[18]. The amount of transmitted data is defined by the value of the TCP networking stack variable denoting the congestion window size, $w_s\geq 0$. Precisely, a FAST TCP node $s$ adapts control variable $w_s$ according to the following rule:

$$w_s(t+1)=\gamma\left(\frac{d_s w_s(t)}{d_s+\lambda_s(t)}+\theta_s\right)+(1-\gamma)w_s(t), \qquad (4)$$

where $d_s\geq 0$ denotes the round-trip propagation delay, $\lambda_s\geq 0$ denotes the round-trip queuing delay observed by source $s$ and $\gamma\in(0,1]$. The algorithm can be proved to converge to the following operating point of the network:

$$\bar{w}_s=\bar{\theta}_s+\bar{x}_s d_s,\ \bar{\lambda}_s=\bar{\theta}_s/\bar{x}_s=\bar{U}_s'(\bar{x}_s). \qquad (5)$$

In simple terms, in the above equilibrium point source $s$ maintains $\bar{\theta}_s=\bar{x}_s\bar{\lambda}_s$ packets in the buffers along its path and $\bar{x}_s d_s$ packets in the transmission lines. It should also be noticed that for the equilibrium to be practically implementable, it is necessary that the total amount of buffering in the network be at least $\sum_s\bar{\theta}_s$, i.e., the transmission delay (or, so called, budget) balancing condition must be satisfied.

# 3. Anticipative Flow Control: Design and Countermeasures

By the above description, the TCP/AQM flow control system defines target equilibrium conditions for the network. Precisely, these conditions are built into the TCP/AQM control rules in order to lead the network of interacting nodes to the solution of problem $\mathtt{NETWORK}$, desired by the control system designer. It will be now argued that in the considered distributed environment the TCP/AQM control system may be prone to *attacks* if the concept of information asymmetry is not taken into account in the control system design.

The TCP/AQM mechanisms are commonly known to the network users managing their network nodes. At the same time, the network users are autonomous in choosing their protocol implementations and are capable of modifying the parameters of the applied traffic flow control algorithms. In other words, there exists a natural information asymmetry between the network users. In particular, such an asymmetry exists between the network users adjusting their transmission rates and the network manager allocating traffic forwarding resources. Let us stick to this setting to illustrate how the information asymmetry may create incentives for the network users to implement *feedback-anticipating* strategies in their rate control algorithms.

The idea which underlies the construction of such control rules is quite simple: the network users should reveal to the network a suitably reduced level of demand for traffic flow.

Indeed, since the network is required to satisfy the observed demand, its reduced level may give rise to lower levels of congestion signals that provide network-wide feedback on the utilization of resources. As a result, the network users may expect to receive an improved level of payoffs (individual performance indices) from the traffic rates admitted by the network.

To implement the above anticipative rate control strategy, a network user may modify the rule according to which $\theta_s(t)$ is updated in the original TCP design. Namely, source $s$ may submit signal:

$$\theta_s(t) = x_s(t)\tilde{U}_s'(x_s(t)), \quad (6)$$

where $\tilde{U}_s'(x_s(t)) < U_s'(x_s(t))$. If the above bidding rule is applied, the network is informed that marginal performance gain that source $s$ receives from the rate at which it sends that traffic is lower than it is in reality. Precisely, this information is propagated in the network through the congestion window size forming burstiness of the traffic. As a consequence the network's operating point is modified.

The above conclusion can be clearly illustrated with the example of equilibrium point conditions given by Eq. (5). Under anticipative rate control one may expect a neighborhood of the following operating point to be reached:

$$\hat{w}_s = \hat{\theta}_s + \hat{x}_s d_s, \ \hat{\lambda}_s = \hat{\theta}_s/\hat{x}_s = \tilde{U}_s'(\hat{x}_s) < \bar{U}_s'(\bar{x}_s) = \bar{\theta}_s/\bar{x}_s. \quad (7)$$

This observation shows that under anticipative rate control the network traffic flow should be expected to deviate from the solution to `NETWORK` problem. Also, notice that by the above description the strategy may be quite easily implemented by the administrator of a network node.

Clearly, an experimental study is necessary to identify performance characteristics of the discussed process, especially under traffic demand shaped by a stochastic process. Nonetheless, formal studies suggest that one may expect to observe substantial performance variations. An important game-theoretic model of outcomes reachable under the discussed strategy can be found in [6], [7]. Namely, it is demonstrated that the network-wide efficiency of outcomes may fall by approximately 1/3 relative to the reference solution of problem `SYSTEM`. In [2] the proposed model is further developed to investigate the outcomes that are reachable *individually* by each agent. In particular, referring to the reference point properties, in equilibrium of the traffic flow control process with feedback-anticipating agents competing for a single link:

- each agent communicates reduced demand to the network, which leads to reduced charge per unit traffic flow and reduced utilization of the link capacity;
- some of the agents (but not all) may be allocated improved rates, in case of which they also receive improved payoffs;
- some of the agents (but not all) may receive improved payoffs with reduced rates;
- commonly applied truthful preference revelation cannot be strictly dominated by commonly applied anticipative strategy with respect to the traffic rates individually received by each agent.

In light of the above predictions, an immediate question arises whether it is possible to reduce the potentially adverse effects of the feedback-anticipation and keep the operating point of the traffic flow in a neighborhood of the reference point. This issue is addressed in [1]. The environment in which the network performance optimization problem is addressed here, is assumed to be characterized by the following properties. First, the problem can be decomposed with respect to *control* (independent) variables $x_i, \ i = 1,\dots,n$, and *interaction* (dependent) variables $y_j, \ j = 1,\dots,m$. Second, the problem of calculating $x_i$ is delegated to a designated agent $i = 1,\dots,n$, whereas the interaction variables $y_j, \ j = 1,\dots,m$, remain managed by the network manager. Finally, in order to calculate control inputs $x_i, \ i = 1,\dots,n$, the agents actively exploit the first-order optimality conditions satisfied by the commonly known reference point, i.e., a solution to the network performance optimization problem. The following theorem is proved.

*Theorem 2:* Consider the following problem:

$$\begin{array}{l} \mathtt{P}(f_j, \ j=0,\dots,m): \\ \left| \begin{array}{ll} \text{minimize} & f_0(\mathbf{x},\mathbf{y}) \text{ over } (\mathbf{x},\mathbf{y}) \in \mathbb{R}^n \times \mathbb{R}^m \\ \text{subject to} & f_j(\mathbf{x},\mathbf{y}) = 0, \ j = 1,\dots,m, \\ & n,m > 0, \ f_j \in \mathscr{C}^2, \ j = 0,\dots,m. \end{array} \right. \end{array}$$

Suppose that $f_j(\mathbf{x},\mathbf{y}) = \sum_i f_{ji}(x_i) + g_j(\mathbf{y})$ for $j = 0,\dots,m$. Let $\bar{\mathbf{z}} = (\bar{\mathbf{x}},\bar{\mathbf{y}})$ be a point for which the second-order necessary optimality conditions for P hold with $\det \nabla_{\mathbf{y}}\mathbf{F}(\bar{\mathbf{x}},\bar{\mathbf{y}}) \neq 0$. Suppose also that for $\mathbf{x} \in \mathbb{B}_\varepsilon(\bar{\mathbf{x}}) = \{\mathbf{x}\colon \|\bar{\mathbf{x}} - \mathbf{x}\| \leq \varepsilon, \ \varepsilon > 0\}$ functions $\eta_i, \ i = 1,\dots,n$, are defined by:

$$\begin{cases} \dfrac{\partial \eta_i}{\partial x_i}(\mathbf{x}) = \sum_j p_j(\mathbf{x}) \dfrac{\partial f_{ji}}{\partial x_i}(x_i), \ i = 1,\dots,n, \\ \mathbf{p}(\mathbf{x}) \equiv -\left(\dfrac{\partial \mathbf{F}^T}{\partial \mathbf{y}}(\mathbf{x},\mathbf{Y}(\mathbf{x}))\right)^{-1} \dfrac{\partial f_0}{\partial \mathbf{y}}(\mathbf{x},\mathbf{Y}(\mathbf{x})), \\ \mathbf{F}(\mathbf{x},\mathbf{Y}(\mathbf{x})) \equiv 0, \end{cases} \quad (8)$$

and for any $\mathbf{v} \in \mathbb{R}^n \setminus \{0\}$ the following condition holds:

$$\mathbf{v}^T \frac{\partial \mathbf{Y}^T}{\partial \mathbf{x}}(\bar{\mathbf{x}}) \frac{\partial^2 H}{\partial \mathbf{y}^2}(\bar{\mathbf{x}},\mathbf{p}(\bar{\mathbf{x}})) \frac{\partial \mathbf{Y}}{\partial \mathbf{x}^T}(\bar{\mathbf{x}})\mathbf{v} > 0, \quad (9)$$

where $H(\mathbf{x},\mathbf{p}(\mathbf{x})) = f_0(\mathbf{x},\mathbf{Y}(\mathbf{x})) + \mathbf{p}(\mathbf{x})^T\mathbf{F}(\mathbf{x},\mathbf{Y}(\mathbf{x}))$. Then $\bar{\mathbf{z}}$ is an isolated solution to problem P and also a unique solution to system:

$$\begin{array}{l} \mathtt{PAYOFF}_i(f_{0i},\bar{x}_i,\mathbf{x}_{-i}), \ i = 1,\dots,n: \\ \left| \begin{array}{ll} \text{minimize} & J_i(x_i,\mathbf{x}_{-i}) = f_{0i}(x_i) + \eta_i(\mathbf{x}) \\ \text{over} & x_i \in \mathbb{B}_\varepsilon(\bar{x}_i), \end{array} \right. \end{array}$$

where $\mathbf{x}_{-i} = (x_1,\dots,x_{i-1},x_{i+1},\dots,x_n)$.

*Proof:* Presented in [1]. The proof is based on the classic elimination approach, according to which the constrained problem is reduced to an unconstrained one with $m$ dependent (interaction or basic) variables expressed in terms of $n$ independent (control) variables. At the same time, however, it should be emphasized that in the analyzed environment the applied approach serves as a *model of reasoning* that is followed by intelligent agents exploiting the commonly available information on the performance optimization goal. ■

Theorem 2 specifies sufficient conditions for implementation of a regular local solution to problem P in a Nash equilibrium point of a noncooperative feedback-anticipation game. Precisely, the theorem shows how to design payoffs $J_i,\ i=1,\dots,n$, in order to reach solution to problem P in a Nash equilibrium point. By construction the payoffs remain invariant with respect to the first-order variations of the feedback variables, modeled by $\mathbf{p}(\mathbf{x})$. This property maintains compatibility of the performance optimization goal with the private goals of the interacting agents. Furthermore, the theorem describes a game (or control mechanism) design procedure that exploits local properties of a feasible solution for problem P in order to define a *collection* of games (mechanisms) that coordinate interactions of the feedback-anticipating agents in the neighborhood of the feasible solution.

Detailed analysis of the equilibrium point defined by a solution to system $\mathtt{PAYOFF}_i,\ i=1,\dots,n$, supports the following collection of conjectures:

- if all agents apply the feedback-anticipating strategy to calculate the controls optimizing payoffs $J_i,\ i=1,\dots,n$, then they implement the performance optimization goal by autonomously acting in their own best interest (incentive compatibility condition);
- since no profitable distortions of the feedback variables (i.e. misrepresentations of preferences) are reachable under the applied rules of the game, it is optimal for the agents to calculate their controls by taking the best responses to the observed *values* of feedback variables;
- in equilibrium the feedback (dual) variables are assigned values equal to those that would be obtained if the network had enough information to solve P in a centralized manner.

The above conclusions provide motivation for the search of iterative and distributed procedures that may lead the strategically interacting agents to the operating point, maximizing performance of the network under asymmetric information. Indeed, since the non-anticipative control is equivalent to the optimal anticipative one in equilibrium point of the game defined by system $\mathtt{PAYOFF}_i,\ i=1,\dots,n,$, the direct preference revelation strategy, given by Eq. (2), may be used at the network traffic source level to control the rate of transmission. However, for this strategy to be implemented by the feedback-anticipating agents, it is necessary for the network to provide sufficient rate control incentives to the agents. By Theorem 2 these incentives can be given the following form:

$$\bar{\eta}_s(\lambda_s,x_s)=x_s\lambda_s-\bar{h}_s(x_s),\ i=1,\dots,n, \tag{10}$$

$$\bar{h}_s(x_s)=\sum_e\int_0^{x_s}a_{se}s\mathrm{d}q_e(s+\sum_{k\neq s}a_{ke}x_k)-b_s(\mathbf{x}_{-s}), \tag{11}$$

where $\lambda_s=\sum_e a_{se}q_e(y_e),\ s=1,\dots,S$, and $b_s(\mathbf{x}_{-s})$ is the transmission delay (or budget) balancing component. Thus in equilibrium source $s$ transmits packets at rate $\bar{x}_s$ if it incurs transmission cost $\bar{\eta}_s(\bar{\lambda}_s,\bar{x}_s)$. The corresponding charge per unit traffic flow for $i$ should, therefore, be defined by:

$$\lambda_s^*=\max\{\bar{\lambda}_s-\bar{h}_s(\bar{x}_s)/\bar{x}_s,0\}. \tag{12}$$

This result can be given the following interpretation. Suppose that $\lambda_s^*>0$ with $\bar{h}_s(\bar{x}_s)>0$ for some $\bar{x}_s>0$. In such a case, source $s$ is motivated to adjusts its rate to $\bar{x}_s$ if it *observes* average delay of $\lambda_s^*<\bar{\lambda}_s$, which would correspond to $\theta_s^*=\bar{\theta}_s-\bar{h}_s(\bar{x}_s)<\bar{\theta}_s$ packets maintained in the buffers along the routing paths for $w_s^*=\bar{w}_s-\bar{h}_s(\bar{x}_s)<\bar{w}_s$. Hence, the network motivates the source to optimally adjust its rate by providing to it the quality of service (QoS) parameters that are improved in comparison to those arising as a solution to NETWORK problem. This implies that the network must be capable of providing differentiated services to the interacting sources, for example by applying suitable active queue management (AQM) techniques.

Although the above requirements can be supported by the currently available networking technology, it is clear that the traffic engineering cost imposed by the mechanism may be substantial. Indeed, game theory shows that for a wide class of problems it is impossible to avoid the costs of Nash equilibrium design; for details see, e.g., [25]–[28]. Namely, gains from reaching a desired solution to the performance optimization problem need not balance the losses corresponding to the introduction of incentives that make this solution attainable in a noncooperative game, i.e., under asymmetric information. These costs must be incurred by the coordinator through violation of the *balancing* condition, or by the agents through violation of the *rational participation* constraint. Theorem 2, presented above, can be applied to give quantitative characterization of theses costs as well. In general, in the considered networking context Theorem 2 implies that:

- under anticipative traffic flow control it may be more desirable for the network manager to reach a suboptimal traffic flow than the optimal one requiring additional coordination costs;
- costs of counterspeculation can be managed, at least to some extent, by a proper design of coordination instruments and by a choice of interaction variables in the network, where the interaction variables represent capacities of the actively managed queues in the network;
- cost of reaching the reference point may discourage some of the agents from participation in the game, i.e., some of the sources may receive payoffs that are below their expectations;
- counterspeculation may be an option for the network manager only if the balancing condition can be satisfied.

The above costs of reaching the desired operating point of the network can be intuitively related to the information monopoly that exists in the considered class of distributed systems. Since it is not possible to fully eliminate the costs of enforcing incentive compatibility, these costs may play the key role in the control system design decision-making process.

# 4. Design Framework For Traffic Flow Control Mechanisms

The discussion in the previous sections brings us to the concept of a procedure that can be applied to engineer a distributed process of traffic flow control under asymmetric information. The idea that underlies the design is to refer to the constructions described by Theorem 2 in order to harmonize interactions of the feedback-anticipating agents and incentivize them to reach the reference solution to the network performance optimization problem. Let us now give an overview of the procedure. The question of algorithm design and implementation is omitted here, since several related remarks and suggestions have already been made in the previous sections.

First, a reference point for the design should be defined to describe a desired performance profile of the network. The following (quite standard) optimization problem can be proposed to serve as a model of preferences defined with respect to the operational performance of the network:

$$\texttt{TRAFFIC}(\mathbf{U},\mathbf{C},\mathbf{A}):$$
$$\left|\begin{array}{ll} \text{maximize} & \sum_s U_s(\sum_p x_{sp}) - \sum_e C_e(\sum_s \sum_p x_{sp} a_{spe}) \\ \text{over} & x_{sp} \geq 0,\ s = 1,\ldots,S,\ p = 1,\ldots,P_s. \end{array}\right.$$

Thus the reference point for the design is defined as a solution $\bar{\mathbf{x}}$ to problem $\texttt{TRAFFIC}(\mathbf{U},\mathbf{C},\mathbf{A})$. Notice that the multipath routing problem is considered here in which flows originating from source $s$ may follow more than one route to destination.

As we have already argued, due to decentralized nature of the network, specifications of the utility functions $U_s,\ s = 1,\ldots,S$, and the cost functions $C_e,\ e = 1,\ldots,E$, are known only locally. To engineer a network-wide control process in the addressed asymmetric information environment we propose to follow a coordination-based approach in which an attempt is made to decompose the reference solution and to build it into the local control rules at the level of networking elements. For this purpose, two sets of auxiliary problems can be defined associated with the network links and traffic sources, respectively.

### *4.1. Link-Control Problem*

The first set of problems is given the following form:

$$\texttt{LINK}_e(y_e),\ e = 1,\ldots,E:$$
$$\left|\begin{array}{ll} \text{maximize} & \eta_e(z,y_e) - Q_e(z) \\ \text{over} & z \geq y_e. \end{array}\right.$$

Problem $\texttt{LINK}_e$ is intended to provide a description of the following link behavior pattern: based on a observed incoming data transfer rate $y_e \geq 0$ select a feasible operating service rate $z \geq 0$ such that QoS targets are met. The corresponding mechanism design problem is to construct functions $\eta_e$ and $Q_e$ for each $e = 1,\ldots,E$ that make the above behavior pattern implementable.

### *4.2. Transfer Rate Control*

The second set of optimization problems is given the following form:

$$\texttt{SOURCE}_s(\boldsymbol{\mu}),\ s = 1,\ldots,S:$$
$$\left|\begin{array}{ll} \text{maximize} & U_s(z_s(\mathbf{x}_s)) - \bar{\eta}_s(\boldsymbol{\mu},\mathbf{x}_s) \\ \text{subject to} & x_{sp} \geq 0,\ p = 1,\ldots,P_s. \end{array}\right.$$

Problem $\texttt{SOURCE}_s(\boldsymbol{\mu})$ models behavior of source $s$ adjusting its data transfer rate $z_s(\mathbf{x}_s)$, distributed according to $\mathbf{x}_s$ along the set of routing paths, to the observed coordination signals $\boldsymbol{\mu}$ providing information about utilization of resources in the network. In this case the corresponding design problem is to construct functions $U_s$, $z_s$ and $\bar{\eta}_s,\ s = 1,\ldots,S$ that support efficient adaptation of transfer rates over optimized set of paths, and lead the sources to the reference solution of problem $\texttt{TRAFFIC}(\mathbf{U},\mathbf{C},\mathbf{A})$.

# 5. Summary

Our intention here was to point out several aspects of the TCP/AQM system design that may affect its performance. It has been argued that due to decentralized structure of the network traffic flow control system, in which traffic rate control tasks are delegated to autonomous agents and coordinated by means of the network congestion signals, it may be possible (for the agents managing the traffic sources) to profitably re-engineer the TCP congestion control algorithm at the cost of the overall performance of the network. Since the formulated conjectures have been derived from our theoretic considerations, we find it necessary to verify them experimentally. From this point of view this paper can therefore be taken as a proposal for further studies.

# Acknowledgements

The author is grateful to Dr. Adam Kozakiewicz for helpful remarks on the network security issues.

# References

[1] M. P. Karpowicz, "Nash equilibrium design and price-based coordination in hierarchical systems", *Int. J. Appl. Mathem. Comp. Sci.*, vol. 22, no. 4 (to appear in 2012).

[2] M. P. Karpowicz and K. Malinowski, "Price-based coordinability in hierarchical systems with information asymmetry", *Contr. Cybernet.*, vol. 42, no. 3 (to appear in 2013).

[3] W. M. Eddy, "Defenses against TCP SYN flooding attacks", *Cisco Inter. Prot. J.*, vol. 8, no. 4, pp. 2–16, 2006.

[4] W. M. Eddy, "TCP SYN flooding attacks and common mitigations", Techn. rep., The IETF Trust RFC 4987, 2007.

[5] F. P. Kelly, “Charging and rate control for elastic traffic”, *Eur. Trans. Telecom.*, vol. 8, pp. 33–37, 1997.

[6] R. Johari and J. N. Tsitsiklis, “Efficiency loss in a network resource allocation game”, *Mathem. Oper. Res.*, vol. 29, no. 3, pp. 407–435, 2004.

[7] R. Johari, S. Mannor, and J. N. Tsitsiklis, “Efficiency loss in a network resource allocation game: the case of elastic supply”, *IEEE Trans. Autom. Contr.*, vol. 50, no. 11, pp. 1712–1724, 2005.

[8] W. Stallings, *High-Speed Networks*. Prentice Hall, 1998.

[9] W. R. Stevens and G. R. Wright, *TCP/IP Illustrated*, vol. 1–3. Addison-Wesley Professional, 1994.

[10] M. P. Karpowicz, “Designing auctions: a historical perspective”, *J. Telecom. Inform. Technol.*, no. 3, pp. 114–122, 2011.

[11] F. P. Kelly *et al.*, “Rate control for communication networks: shadow prices, proportional fairness, and stability”, *J. Oper. Res. Soc.*, vol. 49, pp. 237–252, 1998.

[12] W. Ogryczak and A. P. Wierzbicki, “On multi-criteria approaches to bandwidth allocation”, *Contr. Cybernet.*, vol. 33, no. 3, pp. 427–448, 2004.

[13] W. Ogryczak, M. Pióro, and A. Tomaszewski, “Telecommunications network design and max-min optimization problem”, *J. Telecom. Informa. Technol.*, no. 3, pp. 43–56, 2005.

[14] W. Ogryczak, “Multicriteria models for fair resource allocation”, *Contr. Cybernet.*, vol. 36, no. 2, pp. 303–332, 2007.

[15] J. Lubacz *et al.*, *Mechanizmy Aukcyjne i Giełdowe w Handlu Zasobami Telekomunikacyjnymi*. Warsaw: WKŁ, 2011 (in Polish).

[16] S. H. Low and D. E. Lapsley, “Optimization flow control, I: Basic algorithm and convergence”, *IEEE/ACM Trans. Netw.*, vol. 7, no. 6, pp. 861–874, 1999.

[17] R. J. La and V. Anantharam, “Charge-Sensitive TCP and Rate Control in the Internet”, in *Proc. 19th Ann. Joint Conf. IEEE Comp. Commun. Soc. IEEE INFOCOM 2000*, Tel-Aviv, Israel, 2000.

[18] J. Mo and J. Walrand, “Fair end-to-end window-based congestion control”, *IEEE/ACM Trans. Netwo.*, vol. 8, no. 5, pp. 556–567, 2000.

[19] K. Malinowski, “Optimization network flow control and price coordination with feedback: Proposal of a new distributed algorithm”, *Comp. Commun.*, vol. 25, pp. 1028–1036, 2002.

[20] S. H. Low, “A duality model of TCP and queue management algorithms”, *IEEE/ACM Trans. Netw.*, vol. 11, no. 4, pp. 525–536, 2003.

[21] S. H. Low, F. Paganini, J. Wang, and J. C. Doyle, “Linear stability of TCP/RED and a scalable control”, *Comp. Netw.*, vol. 43, no. 5, pp. 633–647, 2003.

[22] R. Srikant, *The Mathematics of Internet Congestion Control*. Boston: Birkhäuser Boston, 2003.

[23] S. H. Low *et al.*, “Understanding TCP Vegas: a duality model”, *J. ACM (JACM)*, vol. 49, no. 2, pp. 207–235, 2002.

[24] D. X. Wei, C. Jin, S. H. Low, and S. Hegde, “FAST TCP: Motivation, architecture, algorithms, performance”, *IEEE/ACM Trans. Netw.*, vol. 16, no. 6, pp. 1246–1259, 2006.

[25] *Information, Incentives, and Economic Mechanisms: Essays in Honor of Leonid Hurwicz*, T. Groves, R. Radner, and S. Reiter, Eds. University of Minnesota Press, 1987.

[26] J. R. Green and J.-J. Laffont, *Incentives in Public Decision-Making.* North-Holland Publishing Company, 1979.

[27] A. Mas-Colell, M. D. Whinston, and J. R. Green, *Microeconomic Theory*. Oxford University Press, 1995.

[28] V. Krishna, *Auction Theory*. Academic Press, 2002.

---

**Michał P. Karpowicz** received his Ph.D. in Computer Science from the Warsaw University of Technology, Poland, in 2010. Currently, he is an Assistant Professor at Research and Academic Computer Network (NASK). His research interests focus on game theory, network control and optimization.

E-mail: michal.karpowicz@nask.pl
Research and Academic Computer Network (NASK)
Wąwozowa st 18
02-796 Warsaw, Poland

Paper

# Low Complexity Greedy Power Allocation Algorithm for Proportional Resource Allocation in Multi-User OFDM Systems

Najib A. Odhah, Moawad I. Dessouky, Waleed E. Al-Hanafy, and Fathi E. Abd El-Samie

*Faculty of Electronic Engineering, Menoufia University, Menouf, Egypt*

**Abstract—Multi-User Orthogonal Frequency Division Multiplexing (MU-OFDM) is an efficient technique for achieving high downlink capacity in high-speed communication systems. A key issue in MU-OFDM is the allocation of the OFDM subcarriers and power to users sharing the channel. In this paper a proportional rate-adaptive resource allocation algorithm for MU-OFDM is presented. Subcarrier and power allocation are carried out sequentially to reduce the complexity. The low complexity proportional subcarriers allocation is followed by Greedy Power Allocation (GPA) to solve the rate-adaptive resource allocation problem with proportional rate constraints for MU-OFDM systems. It improves the work of Wong *et al*. in this area by introducing an optimal GPA that achieves approximate rate proportionality, while maximizing the total sum-rate capacity of MU-OFDM. It is shown through simulation that the proposed GPA algorithm performs better than the algorithm of Wong *et al.*, by achieving higher total capacities with the same computational complexity, especially, at larger number of users and roughly satisfying user rate proportionality.**

***Keywords—GPA, MU-OFDM, proportional resource allocation, sum-rate capacity.***

## 1. Introduction

MU-OFDM is a technology that has become increasingly important for wireless systems over the past decade. It is considered as one of the two cornerstone technologies of the next generation wireless networks, with Multiple-Input Multiple-Output (MIMO) systems being the other one. MU-OFDM is an expansion upon the basic principles of the Single-User Orthogonal Frequency Division Multiplexing (SU-OFDM) technology, and is designed to accommodate wideband multi-user systems.

SU-OFDM is one of the promising signal processing techniques to provide a high performance physical layer. It has been widely adopted in standards by wireless industry such as IEEE 802.11a and IEEE 802.11g Wireless Local Area Networks (WLAN), IEEE 802.16 fixed Wireless Metropolitan Area Networks (WMAN) which was later extended in IEEE 802.16e (WiMAX) to satisfy high speed mobility and to support both fixed and mobile user stations [1], [2].

SU-OFDM is based on multicarrier transmission in which the broadband channel is divided into $N$ narrowband subchannels, each with a bandwidth much smaller than the coherence bandwidth of the channel. The high rate data stream is then split into $N$ substreams of lower rate data which are modulated into $N$ OFDM symbols and transmitted simultaneously on $N$ orthogonal subcarriers [3]. The low bandwidth of the subchannels along with the frequency spacing between them are necessary to have flat fading orthogonal subcarriers with approximately constant channel gain during each transmission block.

In SU-OFDM, the user can use the total power to transmit on all $N$ subcarriers. The system is then optimized by exploiting the frequency selectivity of the channel and dynamically adapting the modulation type and power on each subcarrier. These dynamic power allocation schemes [4], [5] have shown significant performance gain in terms of throughput compared to static schemes.

In an MU-OFDM system there is a need for a multiple access scheme to allocate the subcarriers and the power to the users. In static subcarrier allocation schemes, each user is assigned to predetermined time slots or the frequency bands respectively, regardless of the channel status. In other words, in non-adaptive fixed subcarrier allocation schemes, an independent dimension is allocated to each user without considering the channel status. In such systems, the optimization problem of maximizing the total throughput of the system reduces to only power allocation or bit loading on the subcarriers. On the other hand, since the fading parameters for different users are mutually independent, the probability that a subcarrier is in deep fade for all users is very low. Each subcarrier is likely to be in a good condition for some users in the system. This is the principle of MU-OFDM with adaptive power allocation in which subcarrier allocation itself plays a very significant role in maximizing the total throughput by using multiuser diversity. A survey of the adaptive MU-OFDM system design problems including an overview of physical layer, Medium Access Control (MAC), and radio resource allocation design issues are provided in [6], [7].

The problem of allocating the base station resources (subcarriers, rates, and powers) to the different users in an MU-OFDM system has been an area of active research over the past several years. This problem has been studied from two perspectives; schemes that minimize the amount of transmit power (Margin Adaptive (MA) approaches [8]–[11]), and those that maximize the sum-rate capacity of the system (Rate Adaptive (RA) approaches [12]–[19]).

In an MU-OFDM system with adaptive subcarrier and power allocation, since the fading parameters for different users are mutually independent, the probability that a subcarrier is in deep fade for all users is very low, and thereby each subcarrier is likely to be in a good condition for some users in the system. This allocation plays a very significant role in maximizing the total sum-rate capacity by using a multi-user diversity [7].

In [16] the RA problem was investigated, wherein the objective was to maximize the total sum-rate capacity over all users subject to power and Bit Error Rate (BER) constraints. It was shown that in order to maximize the total sum-rate capacity, each subcarrier should be allocated to the user with the best gain on it, and the power should be allocated using the waterfilling algorithm across the subcarriers. However, no fairness among the users was considered in [16]. This problem was partially addressed in [12] by ensuring that each user would be able to transmit at a minimum rate, and also in [13] by incorporating a notion of fairness in the resource allocation through maximizing the minimum user's data rate. In [18] the fairness was extended to incorporate varying priorities. Instead of maximizing the minimum user's capacity, the total sum-rate capacity was maximized subject to user rate proportionality constraints. This is very useful for service level differentiation, which allows flexible billing mechanisms for different classes of users. However, the algorithm proposed in [18] involves solving non-linear equations, and this requires computationally expensive iterative operations and is thus not suitable for a cost-effective real-time implementation. The authors in [19] developed a subcarrier allocation scheme that linearizes the power allocation problem, while achieving approximate rate proportionality. The resulting power allocation problem is thus reduced to a solution of simultaneous linear equations.

This paper uses the subcarrier allocation algorithm proposed in [19] and simplifies the power allocation using the GPA algorithm to optimally allocate the transmit power. In simulation, the proposed algorithm achieves a higher total sum-rate capacity than that achieved by [19] and satisfies the same computational complexity, while achieving an acceptable rate proportionality.

The rest of this paper is organized as follows. In Section 2, the system model of MU-OFDM with adaptive subcarrier and power allocation is presented. The proposed resource allocation algorithm is described in Section 3. Simulation results are given in Section 4. Finally, conclusions are given in Section 5.

## 2. System Model

The block diagram of the downlink of MU-OFDM system is shown in Fig. 1. At the base station transmitter, the bits for each of the different $K$ users are allocated to

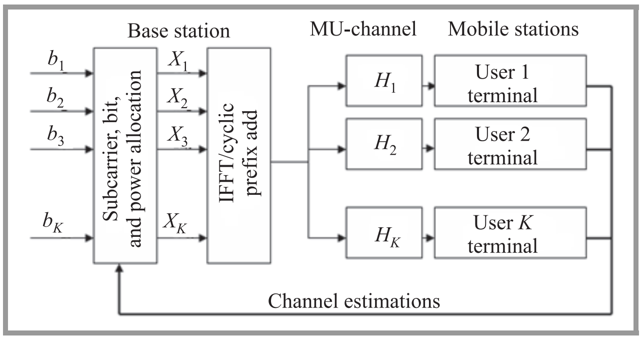



***Fig. 1.*** System model of the downlink MU-OFDM.

the $N$ subcarriers, and each subcarrier $n$ $(1 \leq n \leq N)$ of user $k$ $(1 \leq k \leq K)$ is allocated a power $p_{k,n}$. It is assumed that subcarriers are not shared by different users. Each of the user's bits are then modulated into $N$ M-level Quadrature Amplitude Modulation (QAM) symbols, which are subsequently combined using the Inverse Fast Fourier Transform (IFFT) into an MU-OFDM symbol. This is then transmitted through a slowly time-varying, frequency-selective Rayleigh channel with bandwidth $B$. The subcarrier allocation is known to all users through a control channel; hence each user needs only to decode the bits on his assigned subcarriers. It is assumed that each user experiences independent fading, and the channel gain of user $k$ in subcarrier $n$ is denoted as $g_{k,n}$, with Additive White Gaussian Noise (AWGN) with power $\sigma^2 = N_0 B/N$ where $N_0$ is the noise power spectral density. The corresponding subcarrier Signal-to-Noise Ratio (SNR) is thus denoted as $h_{k,n} = g^2_{k,n}/\sigma^2$ and the $k$th user's received SNR on subcarrier $n$ is $\gamma_{k,n} = p_{k,n} h_{k,n}$. The slowly time-varying channel assumption is crucial, since it is also assumed that each user is able to estimate the channel perfectly, and these estimates are known to the transmitter via a dedicated feedback channel. These channel estimates are then used as input to the resource allocation algorithms. In order that the BER constraints are met, the effective SNR has to be adjusted, accordingly. The BER of a square M-level QAM with Gray bit mapping as a function of the received SNR $\gamma_{k,n}$ and the number of bits $r_{k,n}$ can be approximated to within 1 dB for $r_{k,n} \geq 4$ and BER $\leq 10^{-3}$ [20].

$$BER_{MQAM}(\gamma_{k,n}) \cong 0.2 \exp\left(\frac{-1.6\gamma_{k,n}}{2^{r_{k,n}} - 1}\right) \tag{1}$$

Solving for $r_{k,n}$

$$r_{k,n} = \log_2\left(1 + \frac{\gamma_{k,n}}{\Gamma}\right) = \log_2\left(1 + p_{k,n} H_{k,n}\right) \tag{2}$$

where $\Gamma = -\ln(5BER)/1.6$ is a constant SNR gap, and $H_{k,n} = h_{k,n}/\Gamma$ is the effective subcarrier SNR.

The problem of the MU-OFDM resource allocation with proportional rate constraints is formulated as follows.

Objective:

$$\max_{c_{k,n},\, p_{k,n}} \frac{B}{N} \sum_{k=1}^{K} \sum_{n=1}^{N} c_{k,n} \log_2\left(1 + p_{k,n} H_{k,n}\right).$$

Subject to

$$\begin{cases} C_1: & c_{k,n} \in \{0,1\},\ \forall\, k,n \\ C_2: & p_{k,n} \geq 0,\ \forall\, k,n \\ C_3 & \sum_{k=1}^{K} c_{k,n} = 1,\ \forall\, n \\ C_4: & \sum_{k=1}^{K}\sum_{n=1}^{N} p_{k,n} \leq P_t \\ C_5: & R_i : R_j = \phi_i : \phi_j,\ \forall\, i,j \in \{1,2,\ldots,K\},\ i \neq j \end{cases} \tag{3}$$

where the objective is to maximize the total sum-rate capacity within the total power constraint of the system, while maintaining rate proportionality among the users indicated in $C_5$. Here, $c_{k,n}$ is the subcarrier allocation indicator such that $c_{k,n} = 1$ if and only if subcarrier $n$ is assigned to user $k$, and $P_t$ is the total transmit power constraint.
In $C_5$,

$$R_k = \frac{B}{N}\sum_{n=1}^{N} c_{k,n}\, r_{k,n} = \frac{B}{N}\sum_{n=1}^{N} c_{k,n} \log_2(1 + p_{k,n}H_{k,n}) \tag{4}$$

is the total data rate for user $k$ and $\phi_1, \phi_2, \ldots,$ and $\phi_K$ are the normalized proportionality constants where $\sum_{k=1}^{K} \phi_k = 1$. Note that constraints $C_1$ and $C_2$ in Eq. (3) ensure the correct values for the subcarrier allocation indicator and the power, respectively. $C_3$ imposes the restriction that each subcarrier can only be assigned to a single user, $C_4$ and $C_5$ are the power and proportional rate constraints, respectively.

# 3. The Proposed Resource Allocation (PRA) Algorithm

The resource allocation problem of the MU-OFDM system is divided into two stages; subcarrier allocation and power allocation stage. In the first stage, the number of subcarriers to be allocated to each user is first determined before the actual subcarrier assignments are chosen, whereas the second one is concerned with subsequent power allocation. The proposed resource allocation algorithm depends on the first stage and uses the GPA algorithm to optimally allocate the total transmit power to subcarriers that were previously proportionally allocated by the greedy subcarrier allocation [19].
The subcarrier and power allocation algorithm proposed in [19], which is referred to in our simulations as Ian Resource Allocation (IRA) algorithm, is summarized in the following steps.

**Step 1**: Determine the number of subcarriers $N_k$ to be initially assigned to each user;

**Step 2**: Assign the subcarriers to each user in a way that ensures rough proportionality;

**Step 3**: Assign the total power $p_k$ for user $k$ to maximize the capacity, while enforcing the proportionality;

**Step 4**: Assign the powers $p_{k,n}$ for each user's subcarriers subject to his total power constraint $p_k$.

The proposed algorithm considers the first two steps to optimally and proportionally allocate the subcarriers to the users, but uses GPA to avoid the mathematical complexity of the power allocation in [19].

### *3.1. Greedy Subcarrier Allocation*

The subcarrier allocation stage is described as follows.

**Step 1: Subcarriers Allocation**

The subcarriers of each user in the MU-OFDM system are proportionally allocated to satisfy $N_1 : N_2 : \cdots : N_K = \phi_1 : \phi_2 : \cdots : \phi_K$ that relaxes the proportionality constraint $C_5$ in Eq. (3) to simplify the solution of the resource allocation problem of the MU-OFDM system. The proportion of subcarriers assigned to each user is approximately the same as their eventual rates after power allocation, and thus would roughly satisfy the proportionality constraints, i.e., $N_k = \lfloor \phi_k N \rfloor$. This may lead to $N^* = N - \sum_{k=1}^{K} N_k$ unallocated subcarriers.

**Step 2: Subcarriers Assignment**

This step allocates the per-user subcarriers $N_k$, and then the remaining $N^*$ subcarriers in a way that maximizes the overall sum-rate capacity, while maintaining a rough proportionality with a greedy algorithm, which is a modification of the one used in [19], as described below:

**(a)** For $k = 1$ to $K$
  $R_k = 0$
  for $n = 1$ to $N$
    $c_{k,n} = 0$
  end
end
$p_{k,n} = P_t/N$
$\mathbf{N} = \{1, 2, \ldots, N\}$

In this step, all the variables are initialized. $R_k$ is the capacity for each user, and $\mathbf{N}$ is the set of unallocated subcarriers.

**(b)** For $k = 1$ to $K$
  $\tilde{n} = \arg\max_{n \in \mathbf{N}} |H_{k,n}|$
  $c_{k,\tilde{n}} = 1$
  $N_k = N_k - 1$
  $\mathbf{N} = \mathbf{N} \backslash \{\tilde{n}\}$
  $R_k = R_k + \frac{B}{N}\log_2(1 + p_{k,\tilde{n}}H_{k,\tilde{n}})$
end

In this step, the unallocated subcarrier that has the maximum gain for each user is assigned for that user. Note that the probability of assignment of the same subcarrier for more than one user will be very low, when $N \gg K$.

**(c)** While size $(\mathbf{N}) > N^*$

$$\mathbf{K} = \{1, 2, \ldots, K\}$$

$$\tilde{k} = \arg\min_{k \in K} \frac{R_k}{\phi_k}$$

if $N_{\tilde{k}} > 0$

$$\tilde{n} = \arg\max_{n \in N} |H_{\tilde{k},n}|$$

$$c_{\tilde{k},n} = 1$$

$$N_{\tilde{k}} = N_{\tilde{k}} - 1$$

$$N = N \backslash \{\tilde{n}\}$$

$$R_{\tilde{k}} = R_{\tilde{k}} + \frac{B}{N} \log_2(1 + p_{\tilde{k},\tilde{n}} H_{\tilde{k},\tilde{n}})$$

else

$$\mathbf{K} = \mathbf{K} \backslash \{\tilde{k}\}$$

end

end

In this step, the subcarriers are assigned to each user according to the greedy policy that the user, who needs a subcarrier most in each iteration, chooses the best subcarrier. Since proportional rates are enforced, the need of a user is determined by the user who has the least capacity divided by its proportionality constant.

**(d)** $K = \{1, 2, \ldots, K\}$

for $n = 1$ to $N^*$

$$k = \arg\max_{k \in \mathbf{K}} |H_{k,n}|$$

$$c_{\tilde{k},n} = 1$$

$$R_{\tilde{k}} = R_{\tilde{k}} + \frac{B}{N} \log_2(1 + p_{\tilde{k},n} H_{\tilde{k},n})$$

$$\mathbf{K} = \mathbf{K} \backslash \{k\}$$

end

In this step, the remaining $N^*$ subcarriers are assigned to the best users.

### 3.2. Greedy Power Allocation

After the available subcarriers are proportionally allocated to the users, a GPA algorithm is used to optimally allocate the transmit power between the allocated subcarriers to maximize the sum-rate capacity of the MU-OFDM under the proportional user rate constraint as in Eq. (3). The GPA algorithm of the MU-OFDM system can be described by the following steps.

**Step 1**: Calculate the power of each user in the MU-OFDM system as follows:

$$p_k = \sum_{n=1}^{N} c_{k,n} p_{k,n} \,. \tag{5}$$

**Step 2**: The problem of subcarriers and power allocation for the MU-OFDM system in Eq. (3) is simplified into the problem of power allocation for a single-user OFDM system as follows [21]:

$$\max_{p_{k,n}} \frac{B}{N} \sum_{n=1}^{N} c_{k,n} \log_2(1 + p_{k,n} H_{k,n}), \;\; \forall\, k \in \{1, 2, \ldots, K\} \tag{6}$$

with total user power budget $p_k$, target BER, and maximum permissible QAM modulation order. These constraints can be formulated as:

$$\sum_{n=1}^{N} c_{k,n} p_{k,n} \le p_k, \quad p_{b,n} = BER, \quad r_n \le r^{\max} \tag{7}$$

where $p_{k,n}$ is the amount of power allocated to the $n$th subcarrier of $k$th user to achieve a BER $p_{b,n}$, and $r^{\max}$ is the maximum number of permissible allocated bits per subcarrier. Note that the BERs are assumed equal, i.e., $p_{b,n} = BER$, $\forall\, n \in \{1, 2, \ldots, N\}$, and therefore the subscript n will be dropped from the BER notation to become $p_b$.

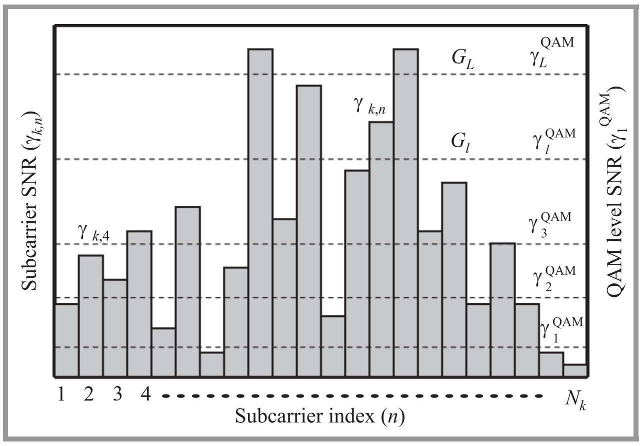



***Fig. 2.*** An illustrative example for grouping of subcarriers.

**Step 3**: The problem of power allocation for subcarriers in the single-user OFDM system is solved by the GPA algorithm that is summarized as follows:

**(a)** Calculate $\gamma_l^{QAM}$ for all $M_l$, $1 \le l \le L$, and $P_b = BER$ using Eq. (1), where $\gamma_l^{QAM}$ is the SNR of the QAM level that is permissible by the transmission system, i.e., $M_l = 2^{r_l^{\max}}$.

**(b)** Redistribute subcarriers according to their SNRs $\gamma_{k,n}$ into QAM groups $G_l$, $1 \le l \le L$ bounded by QAM levels $\gamma_l^{QAM}$ and $\gamma_{l+1}^{QAM}$ with $\gamma_0^{QAM} = 0$ and $\gamma_L^{QAM} = +\infty$ as shown in Fig. 2, i.e.,

$$\gamma_l^{QAM} \le \gamma_{k,n} < \gamma_{l+1}^{QAM} \,. \tag{8}$$

**(c)** For $n = 1$ to $N_K$

Find $l_n$ that satisfies Eq. (8)

if $l_n = 0$

$$r_{k,n} = 0 \;\; p^{up}_{k,n} = \frac{\gamma^{QAM}_{l_n}}{H_{k,n}}$$

else if $l_n < L$

$$r_{k,n} = \log_2 M_{l_n},\;\; p^{up}_{k,n} = \frac{\gamma^{QAM}_{l_n+1} - \gamma^{QAM}_{l_n}}{H_{k,n}}$$

else

$$r_{k,n} = \log_2 M_{l_n},\;\; p^{up}_{k,n} = +\infty$$

end

end

**(d)** Collect power difference from total budget:

$$p_d = \sum_{n=1}^{N_k} \frac{\gamma_{k,n} - \gamma^{QAM}_{l_n}}{H_{k,n}}$$

Initiate greedy bit allocation to

$$r^{gpa}_{k,n} = r_{k,n},\; \forall n \in \{1, 2, \ldots, N_k\},$$

$$p^{gpa}_{d} = p_d$$

while $p^{pga}_{d} \geq \min(p^{up}_{k,n})$ and $\min(l_n) < L, 1 \leq n \leq N_k$

$$j = \arg \min_{1 \leq n \leq N_k} (p^{up}_{k,n})$$

$$l_j = l_j + 1,\; p^{gpa}_{d} = p^{gpa}_{d} - p^{up}_{k,j}$$

if $l_j = 1$

$$r^{gpa}_{k,j} = r^{gpa}_{k,j} + \log_2 M_{l_j},\;\; p^{up}_{k,j} = \frac{\gamma^{QAM}_{l_j+1} - \gamma^{QAM}_{l_j}}{H_{k,j}}$$

else if $l_j < L$

$$r^{gpa}_{k,j} = r^{gpa}_{k,j} + \log_2 \left(\frac{M_{l_j}}{M_{l_j-1}}\right),\;\; p^{up}_{k,j} = \frac{\gamma^{QAM}_{l_j+1} - \gamma^{QAM}_{l_j}}{H_{k,j}}$$

else

$$r^{gpa}_{k,j} = r^{gpa}_{k,j} + \log_2 \left(\frac{M_{l_j}}{M_{l_j-1}}\right),\;\; p^{up}_{k,j} = +\infty$$

end

end

$$R^{gpa}_{k} = \sum_{n=1}^{N_k} r^{gpa}_{k,n}.$$

**Step 4**: Repeat the GPA algorithm for all users to calculate the sum-rate capacity of the MU-OFDM system:

$$R^{gpa} = \sum_{k=1}^{K} R^{gpa}_{k}. \tag{9}$$

# 4. Simulation Results

The frequency selective multipath channel was modeled as consisting of six independent Raleigh multipaths, with an exponentially decaying profile. A maximum delay spread of 5 $\mu$s and a maximum Doppler shift of 30 Hz were assumed. The channel information was sampled every 0.5 ms to update the subcarriers and power allocation. The total transmit power was assumed as 1 W, the total bandwidth as 1 MHz, and total subcarriers as 64. The average subcarrier SNR = 20 dB, and BER = $10^{-3}$, giving an SNR gap $\Gamma = -ln(5 \times 10^{-3}/1.6 = 3.3$. This constant is used in the calculation of the rate $r_{k,n}$ of user $k$ in subcarrier n given in Eq. (2).

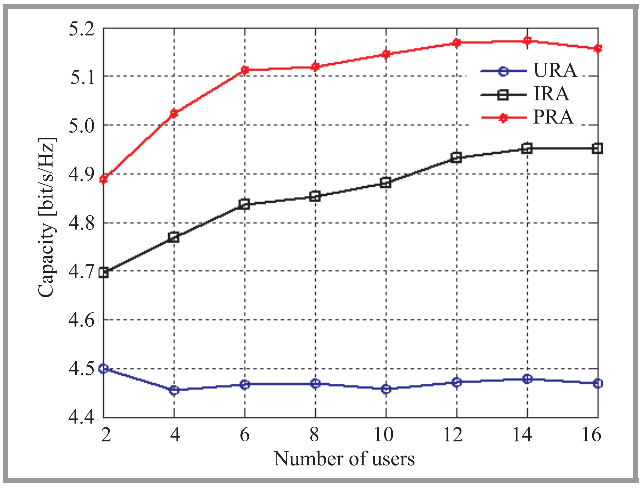



***Fig. 3.*** Average sum-rate capacity versus number of users in an MU-OFDM system.

In Fig. 3, the average sum-rate capacity versus the number of users is depicted for the case of $N = 64$ and SNR = 20 dB. This figure shows that the average sum-rate capacity for the two dynamic resource allocation algorithms; the IRA algorithm and the PRA algorithm is increased significantly with the number of users, while the average sum-rate capacity for the Uniform Resource Allocation (URA) algorithm remains approximately constant. In the URA algorithm, the resources (subcarriers and power) are allocated equally to the users of the MU-OFDM system regardless of their channel gains. So, the multi-user diversity is not exploited and the average sum-rate capacity of the URA algorithm is approximately constant at 4.48 bit/s/Hz. On the other hand, the capacities of the IRA and PRA algorithms are increased as the number of users is increased. This is the effect of multi-user diversity gain which is more prominent in systems with larger number of users.

Figure 4 depicts the total sum-rate capacity versus the average SNR for various radio resource allocation algorithms. This figure shows that the total sum-rate capacity for all the algorithms is increased with the increase in the average SNR and the two dynamic resource allocation algorithms; IRA and PRA outperform the URA algorithm. Both of the two dynamic resource allocation algorithms have approximately the same performance although the transmit power allocation strategies are different for them.

Figure 5 shows the normalized proportions of the capacities for each user for the case of 16 users averaged over 10000 channel realizations. The normalized user capacity is given by $R_k$ divided by $\sum_{k=1}^{16} R_k$, and is observed for both IRA and PRA algorithms. This is compared to the normalized random generated proportionality constraints $\phi_1, \phi_2, \ldots, \phi_K$ that are referred to in Fig. 5 by Gamma. The IRA and PRA satisfy an acceptable proportionality among the users for MU-OFDM systems.

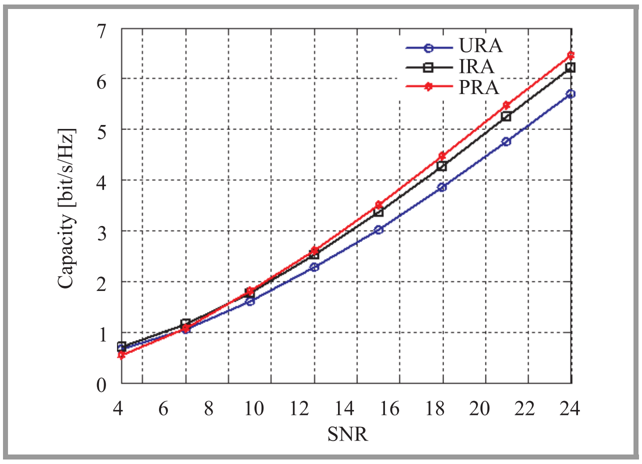



***Fig. 4.*** Average sum-rate capacity versus SNR in an MU-OFDM system.

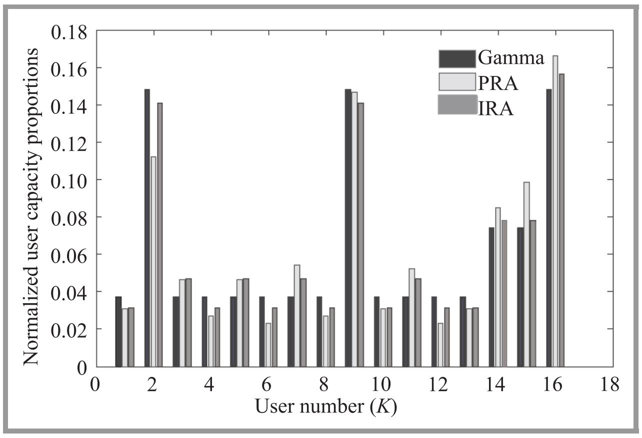



***Fig. 5.*** Normalized user rate proportions versus user index in an MU-OFDM system.

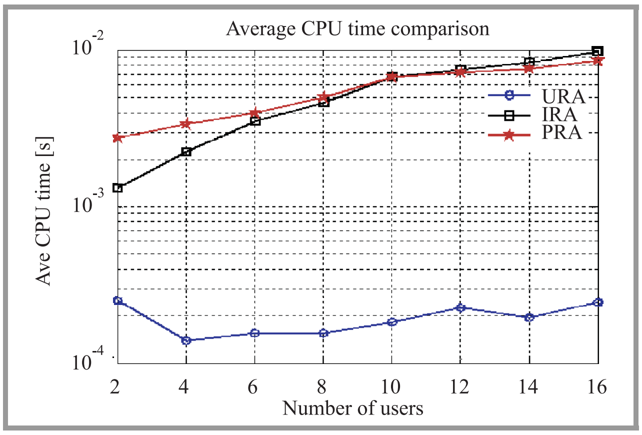



***Fig. 6.*** Average CPU time for compiled Matlab code versus users' number in an MU-OFDM system.

Figure 6 shows a comparison of the computational complexity between the three resource allocation algorithms. These algorithms were run on a Pentium-4 3 GHz personal computer running Windows XP Professional. The simulation experiments used floating-point arithmetics. The simple URA algorithm exhibits less computational complexity with bad capacity performance, whereas the computational complexity of the two dynamic algorithms; IRA and PRA gets high with the number of users, and the PRA algorithm achieves less complexity at larger number of users.

## 5. Conclusion

This paper proposed an optimal solution for the rate-adaptive resource allocation problem with proportional rate constraints for MU-OFDM systems. It benefited from the greedy subcarriers allocation of [19] to proportionally allocate the available subcarriers to the users. After that, the resource allocation problem is dealt with as a single-user resource allocation problem which is solved by the GPA algorithm to maximize the sum-rate capacity of the system, while achieving approximate rate proportionality. This solution avoids the mathematical complexity of [19], while achieving a higher sum-rate capacity. It has been shown through simulations that the PRA algorithm achieves less complexity for higher numbers of users.

## References


[1] "IEEE standard for local and metropolitan area networks part 16: Air interface for fixed and mobile broadband wireless access systems", IEEE, Tech. rep. 802.16, Oct. 2004.

[2] "IEEE standard for local and metropolitan area networks part 16 and amendment 2", IEEE, Tech. rep. 802.16e , Feb. 2006.

[3] H. Schulze and C. Lueders, *Theory and Applications of OFDM and CDMA Wideband Wireless Communications*. John Wiley, 2005.

[4] P. Chow, J. Cioffi, and J. Bingham, "A practical discrete multitone transceiver loading algorithms for data transmission over spectrally shaped channels", *IEEE Trans.Commun.*, vol. 43, pp. 773–775, 1995.

[5] R. Fischer and J. Huber, "A new loading algorithm for discrete multitone transmission", in *Proc. IEEE Global Telecom. Conf. Globecom 1996*, London, England, 1996, vol. 1, pp. 724–728.

[6] M. Sternad *et al.*, "Towards systems beyond 3G on adaptive OFDMA transmission", *Proc. IEEE Inst. Electr. Electron. Eng.*, vol. 95, iss. 12, pp. 2432–2455, 2007.

[7] S. Sadr, A. Anpalagan ,and K. Raahemifar, "Radio resource allocation algorithms for the downlink of multiuser OFDM communication systems", *IEEE Commun. Surveys & Tutorials*, vol. 11, no. 3, pp. 92–106, 2009.

[8] C. Wong, R. Cheng, K. Lataief, and R. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation", *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1747–1758, 1999.

[9] D. Kivanc, G. Li, and H. Liu, "Computationally efficient bandwidth allocation and power control for OFDMA", *IEEE Trans. Wireless Commun.*, vol. 2, pp. 1150–1158, 2003.

[10] L. Xiaowen and Z. Jinkang, "An adaptive subcarrier allocation algorithm for multiuser OFDM system", in *Proc. IEEE VTC 2003*, Orlando, Florida, USA, 2003, vol. 3, pp.1502–1506.

[11] G. Zhang, "Subcarrier and bit allocation for real-time services in multiuser OFDM systems", in *Proc. Int. Conf. Commun. IEEE ICC 2004*, Paris, France, vol. 5, pp. 2985–2989.

[12] H. Yin and H. Liu, "An efficient multiuser loading algorithm for OFDM based broadband wireless systems", in *Proc. IEEE Global Telecom. Conf. Globecom 2000*, San Francisco, USA, 2000, vol. 1, pp. 103–107.

[13] W. Rhee and J. M. Cioffi, "Increase in capacity of multiuser OFDM system using dynamic subchannel allocation", in *Proc. 51st IEEE Veh. Technol. Conf. IEEE VTC 2000*, Tokyo, Japan, 2000, vol. 2, pp. 1085–1089.

[14] G. Song and Y. G. Li, "Adaptive subcarrier and power allocation in OFDM based on maximizing utility", in *Proc. 57th IEEE Semian. Veh. Technol. Conf. IEEE VTC 2003-Spring*, Seoul, Korea, 2003, vol. 2, pp. 905–909.

[15] G. Song and Y. Li, "Cross-layer optimization for OFDM wireless networks-part ii: algorithm development", *IEEE Trans. Wireless Commun.*, vol. 4, pp. 625–634, 2005.

[16] J. Jang and K. Lee, "Transmit power adaptation for multiuser OFDM systems", *IEEE J. Select. Areas Commun.*, vol. 21, pp. 171–178, 2003.

[17] C. Mohanram and S. Bhashyam, "A sub-optimal joint subcarrier and power allocation algorithm for multiuser OFDM", *IEEE Commun. Lett.*, vol. 9, pp. 685–687, 2005.

[18] Z. Shen, J. G. Andrews, and B. L. Evans, "Optimal power allocation in multiuser OFDM systems", in *Proc. IEEE Global Telecom. Conf. Globecom 2003*, San Francisco, USA, 2003, vol. 1, pp. 337–341.

[19] I. Wong, Z. Shen, B. Evans, and J. Andrews, "A low complexity algorithm for proportional resource allocation in OFDMA systems", in *Proc. IEEE Worksh. Sig. Proces. Syst. SIPS 2004*, Austin, Texas, USA, 2004, pp. 1–6.

[20] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[21] W. Al-Hanafy and S. Weiss, "Greedy power allocation for multicarrier systems with reduced complexity", in *Proc. 27th Nat. Radio Sci. Conf. NRSC 2010*, Menouf, Egypt, 2010, vol. 27, pp. 16–18.

**Najib A. Odhah** received the B.Sc. and M.Sc. from the Faculty of Electronic Engineering, Menoufia University, Menouf, Egypt, in 2002 and 2009, respectively. He is currently working towards the Ph.D. degree in Electrical Communications Engineering at Menoufia University, Egypt. From 2003 to 2006, he worked as a demonstrator in the Department of Electronics, Faculty of Engineering and architecture, Ibb University, Ibb, Yemen. His research areas of interest include Multicarrier Communication Systems, Multiple-Input Multiple-Output (MIMO) Systems, Digital Signal Processing, Digital Communications, Interference Cancelation, Channel Equalization and Channel Estimation.

E-mail: Najib_Odhah@yahoo.com
Faculty of Electronic Engineering
Menoufia University
Menouf, Egypt

**Moawad I. Dessouky** received the B.Sc. (Honors) and M.Sc. degrees from the Faculty of Electronic Engineering, Menoufia University, Menouf, Egypt, in 1976 and 1981, respectively, and the Ph.D. from McMaster University, Canada, in 1986. In 1986 he joined the teaching staff of the Department of Electronics and Electrical Communications, Faculty of Electronic Engineering, Menoufia University, Menouf, Egypt. He has published more than 180 scientific papers in national and international conference proceedings and journals. He is currently the head of the Dept. Electronics and Electrical Communications, Faculty of Electronic Engineering, Menoufia University. He has received the most cited paper award from Digital Signal Processing journal for 2008. His current research areas of interest include spectral estimation techniques, image enhancement, image restoration, super resolution reconstruction of images, satellite communications, and spread spectrum techniques.

E-mail: Dr_Moawad@yahoo.com
Faculty of Electronic Engineering
Menoufia University
Menouf, Egypt

**Waleed E. Al-Hanafy** received the B.Sc. (Honors) and M.Sc. degrees from the Faculty of Electronic Engineering, Menoufia University, Menouf, Egypt, in 1996 and 2002, respectively, and the Ph.D. from Strathclyde University, Glasgow, UK, in 2010. In 2010 he joined the teaching staff of the Department of Electronics and Electrical Communications, Faculty of Electronic Engineering, Menoufia University, Menouf, Egypt. He has published more than 20 scientific papers in national and international conference proceedings and journals. His current research areas of interest include the signal processing for communications in particular precoding and equalization methods of MIMO systems, adaptive power/bit loading approaches, Multi-User MIMO, OFDM systems, and resource allocation of wireless communication systems.

E-mail: Waleed_Alhanafy@yahoo.com
Faculty of Electronic Engineering
Menoufia University
Menouf, Egypt

**Fathi E. Abd El-Samie** received the B.Sc. (Honors), M.Sc., and PhD. from the Faculty of Electronic Engineering, Menoufia University, Menouf, Egypt, in 1998, 2001, and 2005, respectively. In 2005 he joined the teaching staff of the Department of Electronics and Electrical Communications, Faculty of Electronic Engineering, Menoufia University, Menouf, Egypt. He is a co-author of about 140 papers in national and international conference proceedings and journals. He has received the most cited paper award from Digital Signal Processing journal for 2008. His current research areas of interest include image enhancement, image restoration, image interpolation, super resolution reconstruction of images, data hiding, multimedia communications, medical image processing, optical signal processing, and digital communications.

E-mail: Fathi_Sayed@yahoo.com
Faculty of Electronic Engineering
Menoufia University
Menouf, Egypt

Paper

# UML Simulation of a Topology Configuration Model

Zbigniew Zieliński[a], Andrzej Stasiak[a], and Włodzimierz Dąbrowski[b]

[a] *Military University of Technology, Warszawa, Poland*

[b] *Warsaw University of Technology, Polish-Japanese Institute of Information Technology, Warsaw Poland*

**Abstract—The article presents the application of simulation methods for topological models to analyze and design information systems. By using UML extensions and the UAL language it is possible not only to build a topological model for software, but also to perform efficient simulations of topological models. Additionally, it is possible to take into account the restrictive conditions stored in UAL and OCL languages. To execute the simulation the authors used an simulator from IBM. These concepts and methods are illustrated by examples.**

***Keywords—configuration topology, system modeling, UML model simulation.***

## 1. Introduction

Methods for modeling information systems using UML is now widely used and can be regarded as a standard in modeling systems [1]. For some time, modeling in UML is complemented with the possibility of executing simulation models in real time [2]–[4]. The benefits of executing simulation models are undeniable. By carrying out simulation models it is possible to: better understand the dynamics of the modeled systems and processes, the early detection of errors in modeling the information system artifacts, the early validation of models to the specifications of the designed systems, the analysis of different variants of constructed solutions, the selection of the best solution in a given situation topology, and the detection and prevention of deadlocks and other undesirable factors in the processes of communication between system artifacts. The use of simulation is conducive to the usage of the UAL language (UML Action Language) to describe the models. Due to this approach a more precise formal description can be applied [2], [4], [5]. A more formal description of the action and semantics facilitates the automatic verification of the correctness of models. In the work [6], thanks to the integration of security models with the architecture models (described in UML) of a specialized system with multilevel security (MLS-type system) and simulation of these models, the verification of many security problems concerning the designed system was possible at the modeling stage. Among the recently developed project models there is a model of a system topology. In this paper a discussion will be conducted on the use of simulation methods for design artifacts expressed in UML, in particular, for topological models.

## 2. Topology Models

The concept of topology is understood as the type of model that illustrates the dependency between the resources of the modeled information system [7]–[9]. An important feature of the topological approach is the ability to plan and then verify the modeled deployment scenarios.

Planning the deployment architecture is a difficult process that can lead to the emergence of a large number of errors. The risk of error can be reduced by applying the following recommendations at the design stage [10], [11]:

- the introduction of the architecture deployment planning process at an early stage of manufacturing applications;
- the communication link between the structure and the architecture deployment by dividing and the re-use of architecture topology templates;
- the use of early deployment architecture validation scenario, and thus the identification of artifacts incompatibility of the model.

Deployment modeling is a bridge between modeling and application development. The topology model defines the elements of infrastructure and their location and dependencies between them.

Application of modeling using the topology shortens the life cycle of the application as follows:

- the application developer can check whether the application meets the requirements of the deployment;
- the architectural engineer can be sure that the established requirements of architecture deployment, the application can be run correctly;
- the analyst can implement best practices and company standards for architecture deployment and test scenarios to reproduce the proper architecture deployment.

By the concept of topology, we mean the deployment architecture model. This model is composed of artifacts called elements. A single element represents an application fragment or fragments of deployed infrastructure (including servers, server software, databases, operating systems) and

be linked with other elements. An element contains information in its structure on the requirements and limitations that must be fulfilled by it and the associated element.
Modeling deployment architecture involves the construction of models at several different levels of abstraction. These levels are as follows [7], [8], [12]:

– application structure level,

– logical model level,

– Physical model level,

• application deployment level.

## 3. Model Simulation

The UML action language (UAL) is a part of the standard maintained by the Object Management Group (OMG) which is an extension of the UML language standard, as a semantics language of the UML action language. UAL is used to construct executable models that are independent from the language, as well as platforms and tools. In earlier practices, as defined in the MDD (Model Driven Development), system designers create "executable code" using notation, a specific implementations language (Platform Specific Implementation (PSI), ex.: C, C++, Java, C#). In addition, developers have to track changes in each of them and tools for their implementation. Currently, the Object Management Group (OMG) developed a language of action semantics and created a standard fUML (Foundational Subset for Executable UML Models) – the fundamental subset for executable UML models and the Action Language for Foundational UML (Alf).
UAL is a subset of definitions defined in the Alf standard, which allows to define a complete system at a higher level of abstraction, and then allows to simulate the model (created in UAL), its debugging and code generation. With the standardization of UAL, the designers will not have to learn many new programming languages – till now related to specific instruments, and they are only limited to the ability to transform models stored in UAL. With UAL, system designers can build it, independently from the target technology platform (PSM and PSI), and the created model can be debugged, simulated, and then started in order to find and correct the errors. Developed models are also independent from the target platform implementation and allows for its determination through choosing the appropriate transformation (UAL2PSI).

## 4. Simulating Models

Currently, it is possible to simulate the behaviors described in UML models. All kinds of behaviors described in UML are namely supported: activity, interaction, state of the machines, and behaviors described in UAL. During simulation, the performed UML diagrams are animated – providing information, such as: another element to perform, elements already performed, the current position of the tokens, active states, etc. The traditional functions of the debugger are also available, such as traps, restart, suspend and renewal. One can also "inject defined events" in event-driven simulation models.
The model simulation can be used together with the configuration planning (Deployment Planning). It allows to visualize the execution of interaction, described in UML, between elements of the topology model – units of implementation. Communication between them will be animated and it is possible to visualize the history of communication with numbered arrows overlapping topology diagrams.
To be able to simulate models UML and topology should be equipped with Rational Software Architect with an additional package: Rational Software Architect Simulation Toolkit.
The solution gives the following benefits:

– enables early understanding of the system so that one can remove the potential drawbacks of their behavior (even during modeling);

– allows to understand how the behavior affects the static structure of the model after developing a diagram of the complex structure;

– allows to understand how the behavior affects the distribution and also to understand the potential impact of the available infrastructure of the built application;

– simulations can work on UML models further specified by the Action Language (UAL), if we assume building strict specifications; this means that simulations can be performed at a very early design stage; then one can try to eliminate any serious design errors and problems, particularly in relation to the availability of infrastructure and networks, as well as later, in order to identify logical errors in behavior.

The current version of the simulator allows:

– creating (stopping/restarting) a session of the executed model;

– selecting paths of implementation in model execution session;

– inserting traps in the model execution session;

– variables modification in the model execution session;

– directing the model execution session to a specific element of the UML diagram;

– execution of models written in UAL (currently it is a textual representation and not pictographic);

– running multiple model execution sessions;

– deleting events that are sent to the model execution session;

– reviewing the history of messages (e.g., for comparing scenario execution paths);
– selection of events and signals in the model execution session;
– animating topology models.

## 5. Simulation Environment

The IBM Rational Software Architect (RSA) environment is a set of integrated tools that support the manufacturing processes (design and the construction of the software), using the technique of modeling in UML [1], [10], [11]. Such an integrated application allows to unify all activities related to the mentioned stages of software engineering. Consolidating multiple functionalities in one tool allows to significantly increase the productivity, resulting in higher work efficiency of a team of analysts, designers, and programmers. RSA makes it possible to verify the architecture design, which allows for easier change management, contributing to raising the quality of the created software.
Within RSA the following modules can be specified [2], [3]:

- Rational Software Modeler (a tool for visual modeling and designing applications);
- Rational Web Developer (visual tool designed for web services and web application developers);
- Rational Aplication Developer for WebSphere Software (an integrated environment that allows to design, create, analyze, test, profile, and deploy web applications and portals, applications in Java and J2EE technologies, and applications that use Web Services);
- Rational Software Architect Simulation Toolkit – an additional package that provides functionality for simulation and animation of UML and topological models.

RSA is a tool that supports the Model Driven Development approach that is focused on the production of models and their transformations. Rational Software Architect enables the use of forward engineering methods (e.g., transformation from UML to C++) and reverse engineering methods (creation of UML models based on existing application code such as C++).

## 6. Example

As an example of simulation models we will consider a process fragment of manufacturing software support recruitment activities at one of the Polish universities on the first degree studies. The recruitment process involves the following entities:

- Candidate – the person applying for the right to study, who has successfully completed the matriculation examination;
- Maturity Exam Committee – (in our system) is the authority that certifies compliance of maturity exam results with those obtained by the candidate (the Regional Examination Commission forwards the results to the Central Examination Board);
- Recruitment Commission – is an institutional body, appointed by the rector who carries out the process of recruiting candidates. The commission announces the results of the recruitment process to the candidates;
- Recruitment Department – at the university, it performs the direct support of candidates, i.e., taking documents from candidates and issuing documents to students.

Prior to the recruitment process at each university, an algorithm is determined, according to which the Recruitment Commission (RC) will conduct the recruitment.
In our further considerations we will confine ourselves to the model of recruitment activities for undergraduate studies. With some simplification we can assume that this algorithm is defined as the weighted average of the results of the candidate's maturity exam from a set of subjects $[i]$, taken at a level of $[j]$ where $\mu_{i,j}$ denotes the weight

$$\text{number of points} = \sum_{i=1,j=1}^{m,n} (\mu_{i,j} * \text{result}_{i,j}) .$$

RC determines, on the basis of results introduced by the candidates, the threshold of eligibility for studying $\gamma[i,j]$ (independently for each faculty, (and/or) direction and type of study) – i.e., the minimum number of points entitling the candidate for enrollment, and then the list specifying the place (rank) of each candidate.
The modeling process begins with the creation of requirements for a system which for the need of the publication will be restricted to the presentation of the modeled functionality (Fig. 1).
The domain model, developed on the basis of the requirements model, allows the formulation of the rules on the

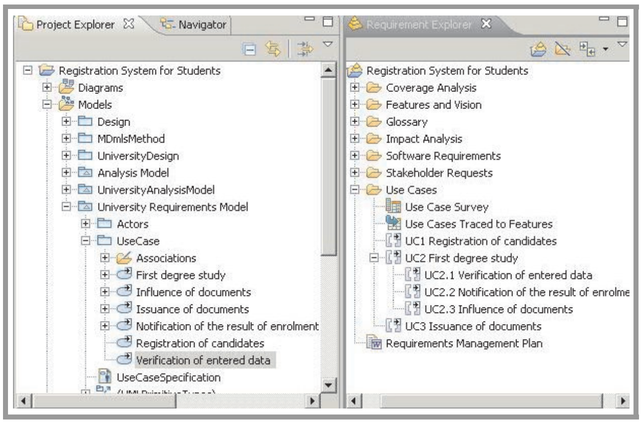



*Fig. 1.* Association of the verbal form with the visual form of the functional requirements.

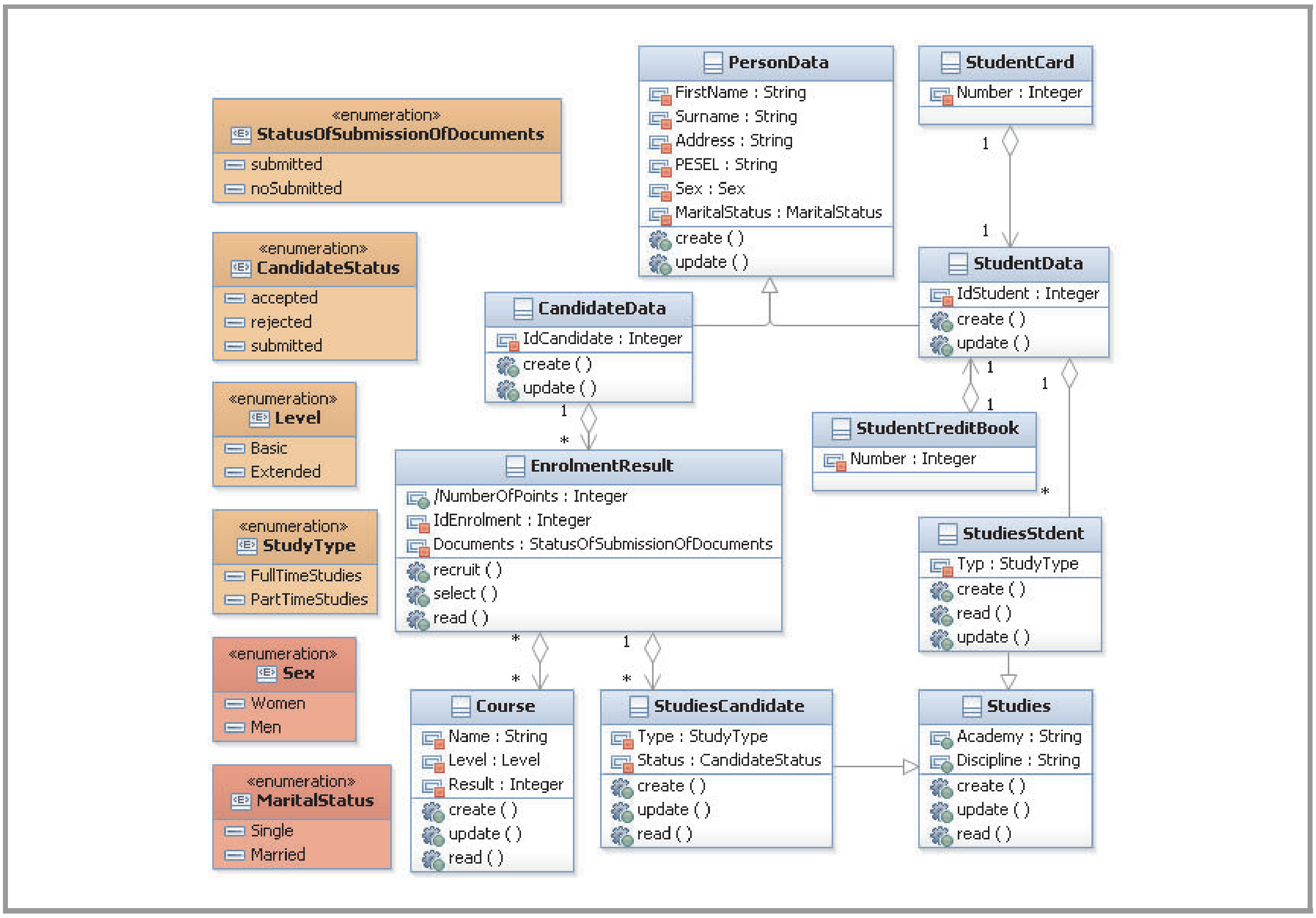



*Fig. 2.* An example of the domain model.

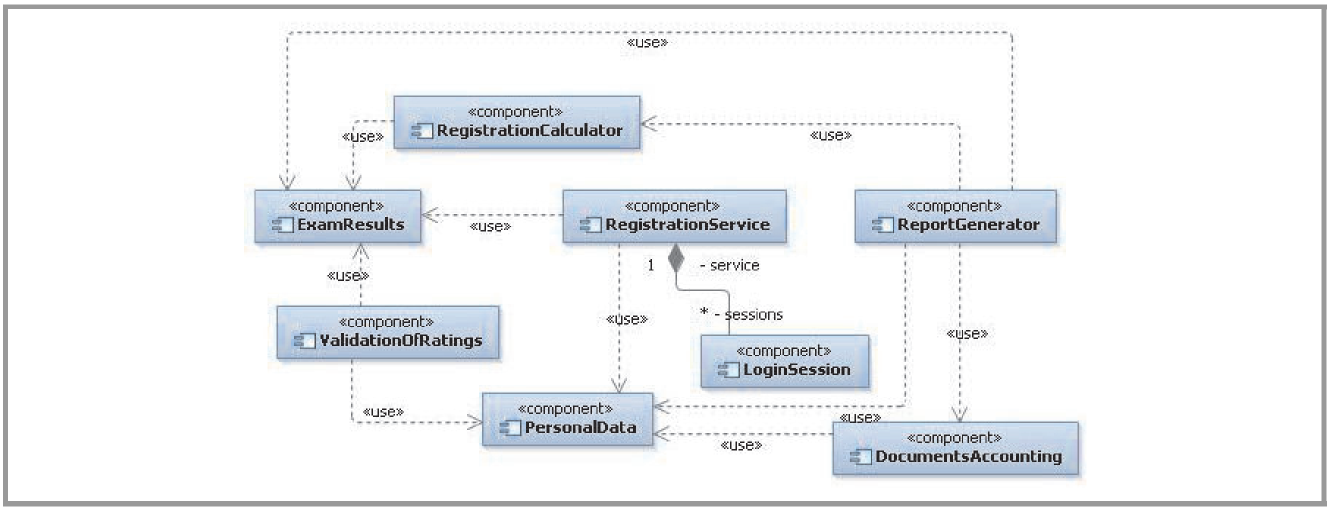



*Fig. 3.* The components model for the student registration system.

recruitment process, and it is shown in Fig. 2. However, not all rules can be described using UML models. For a more complete formalization of the model, it is necessary to use OCL and UAL languages.

An example of a service resulting from the requirements of the system is the authorization service. This service is defined by the LoginSession component (which will be appealed to all services that require authorized access to them).

For example, the student registration service (RegistrationService component) is related to many of their sessions (Fig. 3), for each user a separate object session, maintaining the state only when the user is logged in (Loggedin state (Fig. 4)).

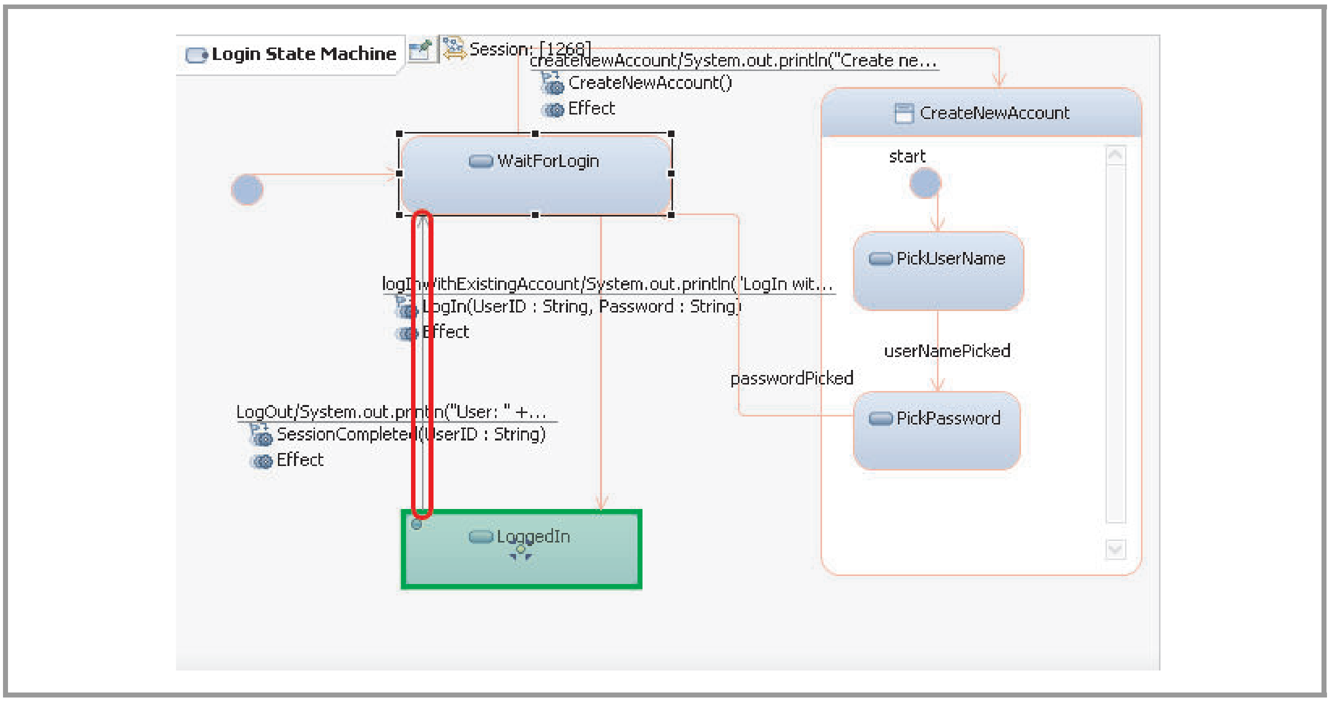



*Fig. 4.* Signaling the history of states (and transitions).

After changing the state forced by the sessionCompleted operation, the object session is destroyed, and the service goes into the WaitForLogin state.
An important part of the planned condition model simulation process is determination of the list of messages introduced into the console, which will be implemented in UAL. The code corresponding to the handling of subsequent messages is as follows:

- System.out.println("Create new user and yours password");
- System.out.println("LogIn with userId =" + msg.UserID + "and password =" + msg.Password);
- System.out.println("User:" + msg.UserID + "session is completed").

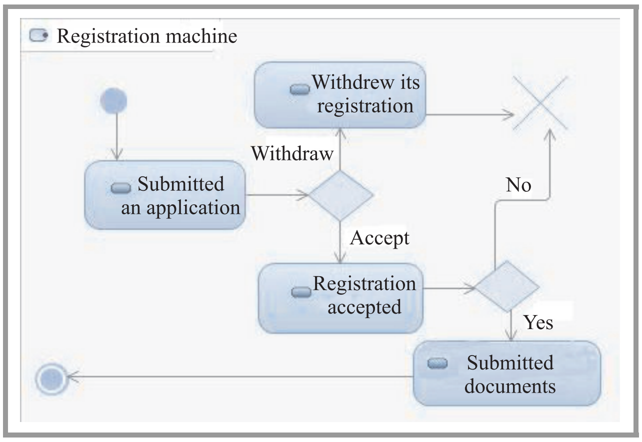



*Fig. 5.* The result of the machine state service animation – registration of students.

After taking into account the principles of building simulation models [2], [3], it is possible to obtain the signaling of history of states and transitions between them in the simulated model. An exemplary signaling of state history is shown in Fig. 4.
After changing the state forced by the sessionCompleted operation, the object session is destroyed, and the service goes into the WaitForLogin state.
Similar model test coverage can be observed by analyzing Fig. 5.

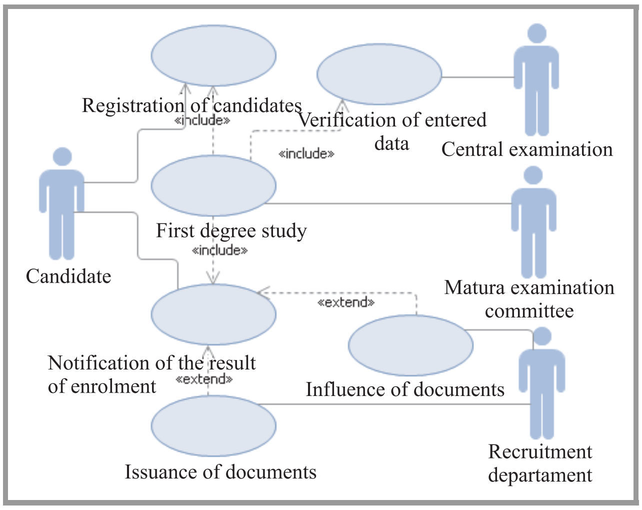



*Fig. 6.* A fragment of the use case model.

The service of registering candidates is an important functionality of the system being built (Fig. 6) and initiates the recruitment process, which is why the following part

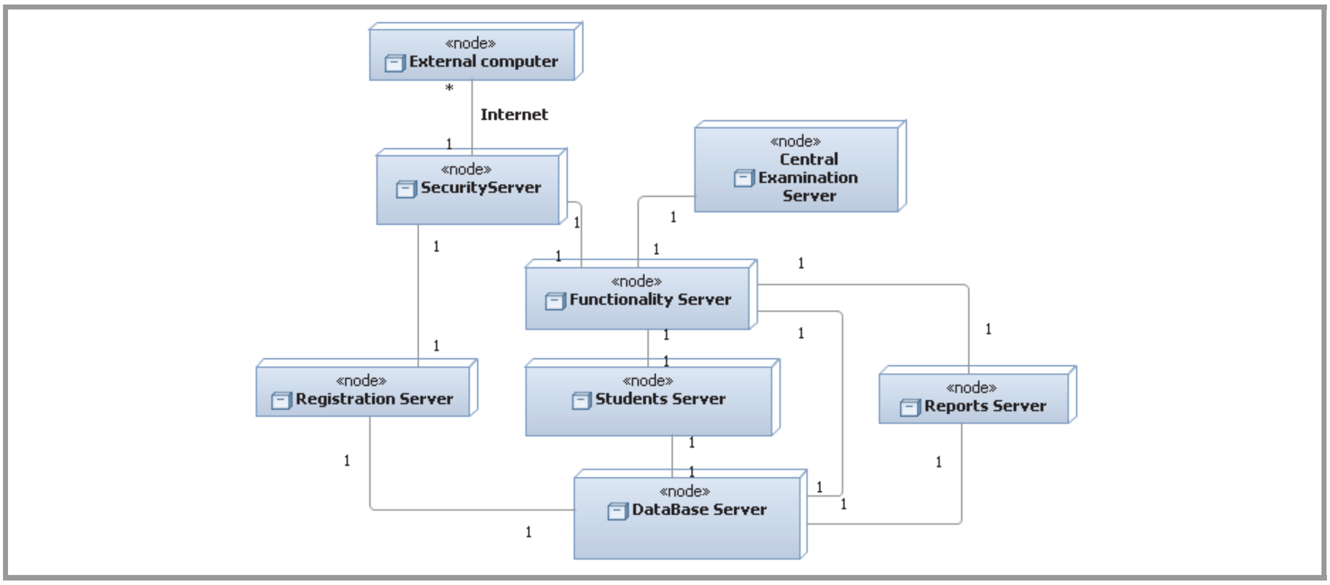



*Fig. 7.* The deployment diagram for the student registrations service.

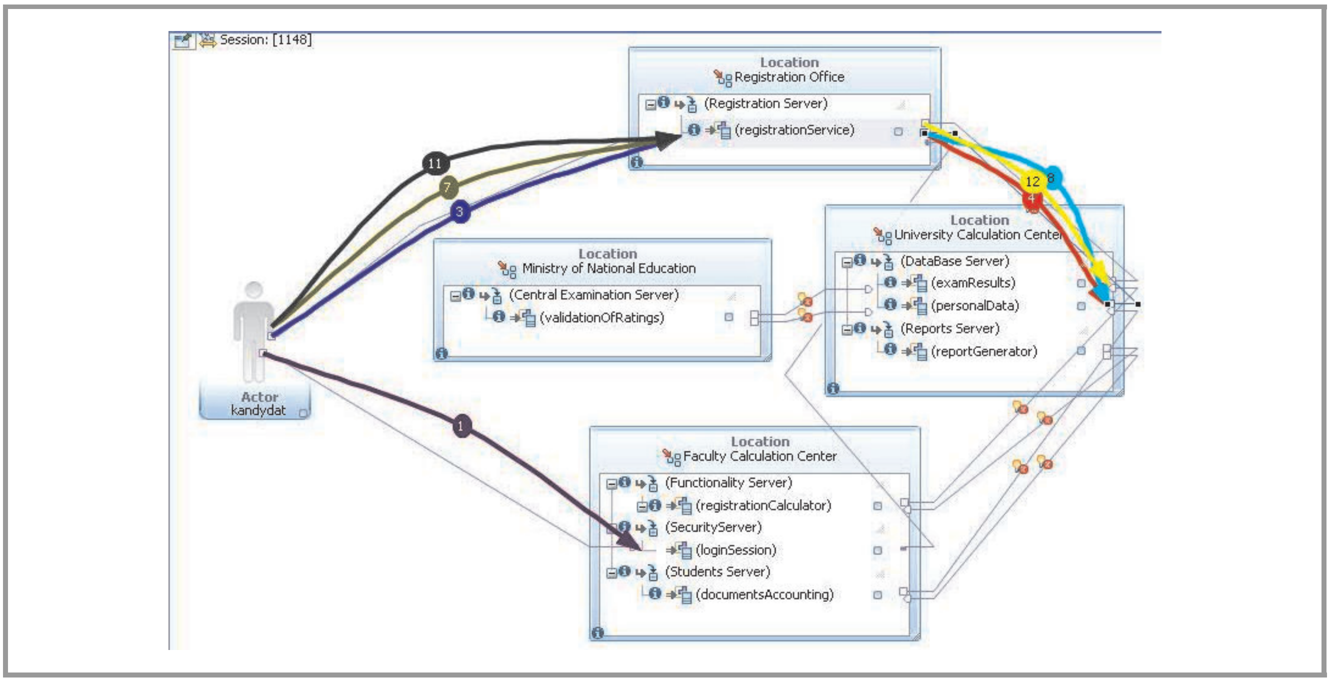



*Fig. 8.* The result of the state machine animation of the student registration service.

of this article will be limited to the presentation of its detail.

For the selected set of use cases (in the service of registering candidates) the deployment diagram was proposed (Fig. 7), which next has been transformed to the more detailed form of topology model.

It would be noticed that using UML language for deployment specification (Fig. 7) is not necessary, because a topology model extends a set of information concerning a system implementation as nodes, components, etc., and additional properties defining its location as hardware and software platforms.

Simulation model of the topology for the discussed fragment of the system, i.e., the candidate registration service is shown in the diagram (Fig. 8).

This model is a component of the system (Fig. 3) running on the nodes placed in specific (named) locations in the structure of the educational institution and relationships between them that define the flow path of messages.

The flows of messages are shown directly on the topology model in the form of numbered arrows.

It is worth noting here that during the process of animation, only those components of the system model from Fig. 3 participate, which are directly involved in the sce-

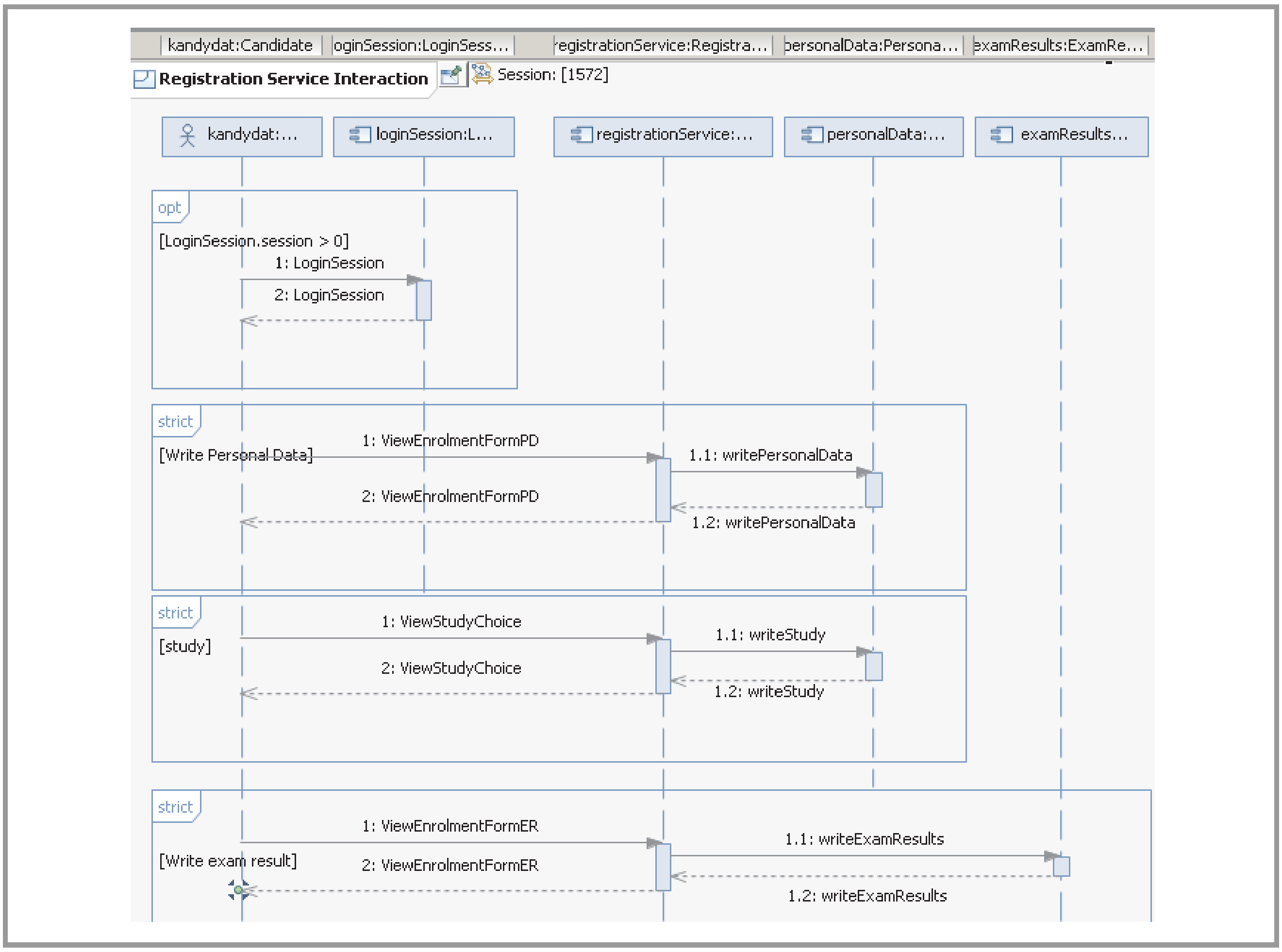



*Fig. 9.* Registration Service Interaction.

nario depicted in Fig. 9 (others are used in other scenarios).

The results show that the essential parameter of the process simulation is the behavioral diagram, which is to be animated. The animation can comprise not only on the presentation of the paths of starting the model (by indicating the place of the token, anticipating the next place, and highlighting the paths with a color, in which the token already was in), but we can also animate the messages flowing in from an established source to target.

## 7. Conclusion

This paper presents the essential capabilities of the simulation process of UML models, which affect the animation capabilities of topological models.

It is worth noting here that UML models significantly reduce the complexity of the topology modeling process, allowing to automatically identify links between elements of the topology model based on relationships defined in behavioral models, and the structure of the built system (relationships in the topology model from Fig. 7 were automatically created after assigning components to nodes from Fig. 3) working in their environment of which behavior is defined in Fig. 8. We can say that the relations in UML models provide guidance to their mappings (via the drag-and-drop method) on topology models.

Nowadays, the modern CASE tools (including Rational Software Architect used in this article) allow to change the elaborate approach to a transformational approach (according to the MDA concept), automating the software development processes that nevertheless requires the use of language semantics of actions to build executable UML models, which may provide an effective basis for building models of system implementation.

## References

[1] W. Dąbrowski, A. Stasiak, and M. Wolski, *Modelowanie systemów informatycznych w języku UML 2.1*. Warszawa: PWN, 2007 (in Polish).

[2] M. Mohlin, "Model Simulation in Rational Software Architect: Simulating UML Models", IBM, 2010.

[3] M. Mohlin, "Model Simulation in Rational Software Architect: Communicating Models", IBM, 2010.

[4] E. Anders, "Model Simulation in Rational Software Architect: Activity Simulation", IBM, 2010.

[5] "OMG, Action Language for Foundational UML (Alf)" [Online]. Available: http://www.omg.org/spec/ALF/1.0

[6] Z. Zieliński, A. Stasiak, and W. Dąbrowski, "Zastosowanie metod symulacji modeli UML do analizy i projektowanie komputerowych systemów specjalizowanych przetwarzających danych z ochroną wielopoziomową", *Przegląd Elektrotechniczny*, no. 2, pp. 120–125, 2012 (in Polish).

[7] "Planning deployment with the topology editor", IBM Tutorial, 2008.

[8] "Modeling deployment topologies", IBM Tutorial, 2008.

[9] S. Willard, *General topology*. Mineola, N.Y.: Dover Publications, 2004.

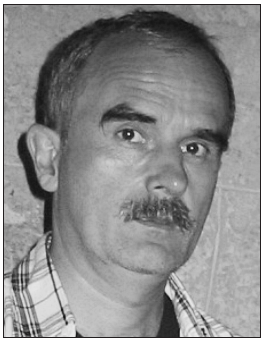
[10] M. Narinder, "Anatomy of a topology model used in IBM Rational Software Architect Version 7.5, Part 2: Advanced concepts", IBM, 2008

[11] M. Narinder, "Anatomy of a topology model in Rational Software Architect Version 7.5: Part 1: Deployment modeling", IBM, 2008.

[12] W. Dąbrowski, A. Stasiak, and K. Markowski, "Modelowanie systemów IT z wykorzystaniem topologii konfiguracji", *Przegląd Elektrotechniczny*, no. 9, pp. 239–242, 2010 (in Polish).



**Zbigniew Zieliński** received the M.Sc. in Computer Sciences from the Cybernetics Faculty of Military University of Technology, Warsaw, Poland in 1978, and the Ph.D. degree in Computer Systems in 1988. He is currently an Assistant Professor of Computer Systems in the Institute of Teleinformatics and Automation of Cybernetics Faculty, Military University of Technology. His current research interests are in the areas of computer systems dependability, processors network diagnosis methods, fault-tolerant systems, as well as virtualization and system security.
E-mail: zzielinski@wat.edu.pl
Faculty of Cybernetics
Military University of Technology
Gen. S. Kaliskiego st 2
00-908 Warsaw, Poland


**Andrzej Stasiak** is an expert in the field of design of information systems. From years he is a member of program committees of conferences on Software Engineering and Real Time Systems. From 1987 is an Assistant Professor of Computer Systems in the Institute of Teleinformatics and Automation of Cybernetics Faculty, Military University of Technology. He gained his professional experience directing some complex IT projects.
E-mail: astasiak@wat.edu.pl
Faculty of Cybernetics
Military University of Technology
Gen. S. Kaliskiego st 2
00-908 Warsaw, Poland



**Włodzimierz Dąbrowski** received the M.Sc. in Electrical Engineering from the Warsaw University of Technology, Warsaw, Poland in 1992, and the Ph.D. degree in Control Systems Theory from the Warsaw University of Technology in 1999. He is currently an Assistant Professor in the Institute of Control and Industrial Electronics, Faculty of Electrical Engineering, Warsaw University of Technology. From 2000 he is Assistant Professor in Polish-Japanese Institute of Informatics, Software Engineering Department, Warsaw. His current research interests are in the areas of design methods for computer systems, software engineering, and software project management.
E-mail: w.dabrowski@ee.pw.edu.pl
Faculty of Electrical Engineering
Warsaw University of Technology
Koszykowa st 75
00-662 Warsaw, Poland

Paper

# Skill-Based Bimanual Manipulation Planning

Wojciech Szynkiewicz

*Institute of Control and Computation Engineering, Warsaw University of Technology, Warsaw, Poland*

**Abstract—The paper focuses on specification and utilization of manipulation skills to facilitate programming of bimanual manipulation tasks. Manipulation skills are actions to reach predefined goals. They constitute an interface between low-level constraint-based task specification and high level symbolic task planning. The task of the robot can be decomposed into subtasks that can be resolved using manipulation skills. Rubik's cube solving problem is presented as an example of a 3D manipulation task using two-arm robot system with diverse sensors such as vision, force/torque, tactile sensors.**

***Keywords—bimanual manipulation planning, manipulation skills, robot programming.***

## 1. Introduction

Robots employed in human-centered environments have to be equipped with manipulative, perceptive and communicative capabilities necessary for real-time interaction with the environment and humans. Up to now, robot systems have been only able to deal with the high complexity and the wide variability of everyday surroundings to a limited extent. In this paper we are focused on planning dual-arm/hand manipulation tasks for service robots working in such environments.

In our everyday lives we perform many operations in which our two hands cooperate to manipulate diverse objects. The goal of our research is to understand the nature of two-handed cooperative manipulation and enhancing the manipulative skills of the dual-arm robot. Two cooperative hands, if properly utilized, are capable of grasping and manipulating a much larger class of objects, including long, flexible, irregularly shaped, or complex objects (e.g., with internal degrees of freedom).

Object manipulation tasks typically involve a series of action stages in which objects are recognized, grasped, moved, brought into contact with other objects and released. These stages are usually bound by mechanical events that are subgoals of the task. These events involve the making and breaking of contact between either the hands and the grasped object or the object held in hand and another object or surface. Importantly, these contact events usually produce discrete and distinct sensory events. To simplify a solution of the overall problem, we usually tend to divide the task into a sequence of clearly separated subtasks, each of which accomplishes a specific subgoal. In this case, a task planning focuses on deciding what operations will be needed to execute a particular manipulation task, and in what order the operations should be performed. The operations are considered at an abstract level, i.e., sensory operations, gross and fine motion operations, grasping and releasing operations. In terms of representation, it denotes the smallest entity which is used for describing an action.

Many studies have been devoted to single-handed manipulation, for overview see [1]. Recently, a bimanual manipulation has also attracted more attention, especially in unstructured environments (see for example [2]–[7]). Also, the literature pertaining to the analysis of bimanual operations performed by humans is quite extensive, especially in the field of a human-computer interaction, e.g., [8]. Although many solutions for single-handed manipulation can be easily adopted for bimanual manipulation, the whole potential of two cooperative hands cannot be fully utilized without a deeper understanding of their unique characteristics. In general, two-handed manipulation can be classified into uncoordinated and coordinated tasks [8]. The latter can be further subdivided into symmetric and asymmetric ones.

In this paper we focus on specification and utilization of manipulation skills to facilitate programming of bimanual manipulation tasks. The proposed method uses a hierarchical approach for the decomposition of manipulation skills. Manipulation skills are compositions of basic robot operations to reach some predefined goals. They can serve as an interface between low-level geometric task specification and high level symbolic task planning. If the skills are well-defined and robust, then manipulation planning is simplified because is performed in the "space" of skills rather than in the high dimensional configuration/operational space. The task of the robot can be approximated by a set of parameterized manipulation skills. The approach presented in this paper focuses on tasks, rather than motions, and uses manual programming rather than learning techniques to determine the set of manipulation skills. It should be noted that the concept of using skills to create complex actions is a well-studied topic covering many areas, thus, only its applicability to robot manipulation, especially two-handed manipulation is discussed. In our approach, each individual skill is represented as a hybrid finite state automaton in which each state runs one basic operation, and each violation of a operation can give rise a transition. Each transition is the outcome of the basic operation. Skills can have a set of parameters, which can be used to adapt each skill to a particular use case. Skills are the components of the intermediate level between symbolic and geometric levels. In this paper we focus on solving the coordinated bimanual tasks, and we propose the solution of the Rubik's cube manipulation as an example of such a task. To implement dual-arm manipulation utilizing vision and force sensing, the `MRROC++` robot programming framework is used [9], [10].

The paper is organized as follows. In Section 2 an overview of most representative manipulation planning approaches is given. Section 3 describes the hierarchical representation of manipulation tasks. In Section 4 Rubik's cube problem solving is discussed as an example of two-handed manipulation.

## 2. Related Work

Manipulation planning is an extension to the classical robot motion planning problem. The robot is able to interact with its environment by manipulating movable objects, while it has to avoid self-collisions or collisions with obstacles. Traditionally, manipulation planning concerns the automatic generation of the sequence of robot motions allowing to manipulate movable objects among obstacles. Existing research in manipulation planning has focused mainly on the geometric aspects of the task, while greatly simplifying the issues of grasping, stability, friction, and uncertainty [4], [11]. Symbolic planning algorithms have typically assumed perfect models of both the environment and the robot, not only at an abstract level but at every level of control. This is a quite reasonable assumption in well-structured and fully controlled environments. However, in everyday environments this is not often the case, which makes that most of the proposed theoretical solutions are not directly applicable. The real world does not behave as expected, and in fact it does not behave predictably.

Most of the research in manipulation planning deals with the creation of the manipulation graph and extraction of a manipulation path from this graph [12]–[15]. The concept of a manipulation graph was introduced by Alami *et al*. [12] for the case of one robot and several movable objects manipulated with discrete grasps and placements. In this case, the nodes of the graph represent discrete configurations and the edges correspond to robot motions moving the grasped object (*transfer path*), or leaving it at rest to get to another grasp position (*transit path*). A solution to the manipulation planning problem is now given by a manipulation path in this graph. This path is solved using PRM (Probabilistic Roadmap Method) planners [13], [14].

In most of the existing algorithms it is assumed that a finite set of stable placements and of possible grasps of the movable object are given in the definition of the problem (e.g., [12], [13]). Consequently, a part of the manipulation task decomposition is thus done by the user since the initial knowledge provided with these finite sets has to contain the grasps, and the intermediate placements required to solve the problem. In [14] the authors proposed a general manipulation planning approach capable of addressing continuous sets for modeling, both the possible grasps and the stable placements of the movable object. The nodes of the manipulation graph (i.e., the places where the connections between the feasible transit and transfer paths have to be searched) correspond to a set of sub-manifolds of the composite configuration space, as opposed to discrete configurations. Cambon *et al*. [15] proposed a specialized integration of a symbolic task planning and geometric motion, and manipulation planning. They extended classical action planning formalism based on a STRIPS-like description where manipulation planning problems in configuration space are introduced.

One of the most intuitive ways to acquire new task knowledge is to learn it from the human user via demonstration and interaction. This approach to task learning is known as Programming by Demonstration (PbD) [2]. It is one of the most often used programming paradigms of two-arm manipulation for humanoid robots [7]. PbD systems generally try to decompose the observed task execution of the human demonstrator into a sequence of tasks that are performable by the robot. Typically, tasks are recorded from human demonstrations, segmented, interpreted and stored using some data representation. Several programming systems and approaches based on human demonstrations have been proposed during the last years, e.g., [2], [3].

In [16] an architecture which uses primitive skills that combine to form a skill, which in turn form a complete task is presented. Each primitive skill is selected by heuristic selection out of many possible primitive skills, based on the sensor signals. A neural network is used to detect the change between the skills. Each primitive skill is executed by a separate controller.

## 3. Manipulation Planning

Manipulation planning is a very challenging problem in robotics research as it consists of a number of subproblems that themselves are still open issues and subject to ongoing research. Typical manipulation tasks accomplish relative motions and/or dynamic interactions between various objects. Typically, manipulation planning involves motion and grasp planning. The most effective robot motion planning strategies today are built upon sampling-based techniques, including the PRM [13] and Rapidly-exploring Randomized Trees [17], and their variants. Robot motion planning can also be viewed as an optimization problem that aims to minimize a given objective function [5], [11]. To solve such a problem in efficient way appropriate tools have to be used, e.g., [18]. Grasp planning is also an area of intensive research [19].

Several basic components of manipulation task can be distinguished (Fig. 1).

Using intended subgoals as a criterion, three different classes of manipulation tasks can be distinguished [20].

1. *Transport operations*: the simplest class of robot manipulation. This kind of task can be easily distinguishably by the change of the external state (pose) of the manipulated object. Various types of transport

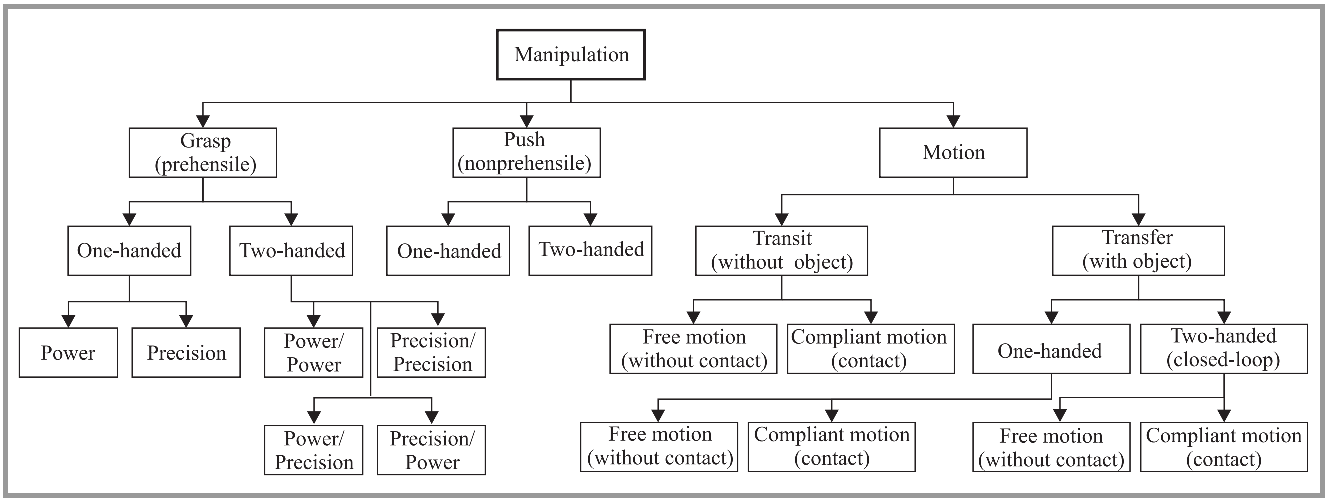



*Fig. 1.* Components of the manipulation task.

tasks such as pick-and-place or fetch-and-carry are a component part of almost all manipulation operations. Accordingly, the trajectory of the manipulated object has to be considered and modeled in transport actions models.

2. *Object handling*: a more specialized class of manipulation tasks deals with changing the internal state of objects without influencing other objects (like opening doors, pushing a button, manipulating the Rubik's cube, etc.). This class of tasks consists of every task changing an internal state of an object without manipulating another object. In the object-handling tasks, transition actions changing an internal state have to be modeled. Moreover, the object models need to incorporate an adequate description of their internal state.

3. *Tool handling*: the most typical characteristic for this type of actions is the interaction between two objects, usually a tool and a workpiece. Interaction is related to the functional role of objects used or the correlation between the functional roles of all objects included in the manipulation, respectively. The object model thus should contain a model of the possible interaction modalities or functional roles the object can take. According to different modalities of interaction, considering contacts, movements, etc., a diversity of handling methods has to be modeled.

We consider a manipulation task to be an activity involving the composition and coordination of an existing set of manipulative skills in order to accomplish a given set of goals. Two representations of robot manipulation skills/tasks can be distinguished symbolic and non-symbolic.

### *3.1. Manipulation Skills*

We make a crucial distinction between tasks, manipulation skills and basic skills or primitive actions in this work. Task is a function to be performed. Manipulation Skill (MS) is a pattern of activity which describes an ability that achieves or maintains a particular goal. A manipulation skill can be defined as an abstraction of a set of basic skills that follow the same control strategy. Basic Skill (BS) is an action that abstract a sensory-motor coupling such as skill motion types (e.g., motion trajectory generators), concrete grasping and releasing strategies, direct and inverse kinematics for the specific robots, etc. The set of the BS serves as an application programming interface. The task of the robot can be represented by a set of parameterized manipulation skills. Therefore the overall planning and control system has a layered hierarchical structure as shown in Fig. 2. It should be noted that hierarchy can exist at all layers. Task is the highest level of abstraction, representing a semantically meaningful task such a solving scrambled Rubik's cube. The task consists of a sequence of MS's, which represent subtasks, such as turning a single face of the Rubik's cube. Skills consist of basic skills or primitive actions which are the lowest level of control in the proposed architecture. Each BS is implemented using a single low-level controller, Control Program (CP) which is responsible for the control of robot hardware.

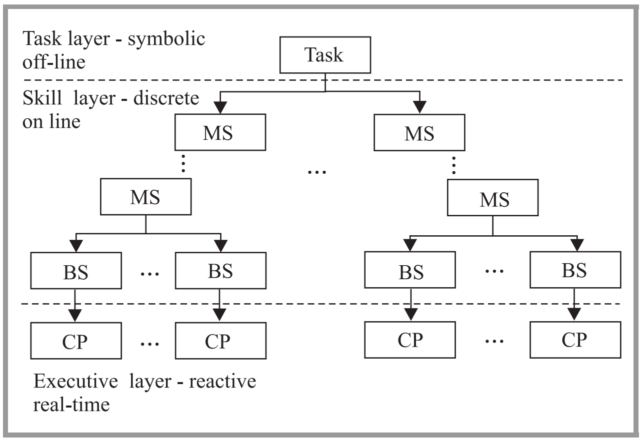



*Fig. 2.* Planning and control three-layered structure.

To model MS we use a hybrid automaton. A hybrid automaton is a dynamical system that describes the evolution in time of the valuations of a set of discrete and continuous variables. A hybrid automaton $H$ [21], [22] is a collection $H = (Q, X, f, Init, D, E, G, R)$, where

- $Q$ is a set of discrete states;
- $X = \mathbb{R}^n$ is a finite set of continuous states;
- $f(\cdot,\cdot) : Q \times X \rightarrow \mathbb{R}^n$ is a vector field;
- $Init \subseteq Q \times X$ is a set of initial states;
- $Dom(\cdot) : Q \rightarrow 2^X$ is a domain;
- $E \subseteq Q \times Q$ is a set of edges (events);
- $G(\cdot) : E \rightarrow 2^X$ is a guard condition;
- $R(\cdot,\cdot) : E \times X \rightarrow 2^X$ is a reset map (relation).

Each state of the automaton has its own low-level controllers and transitions to other states. The proper selection of the set of manipulation skills is a critical step in our approach. The following manipulation skills have been defined to solve the Rubik's cube problem: *Localize*, *Reach*, *Turn*, *Grasp*, *Release*.

*Localize* – robotic manipulation of an object requires that this object must be detected and located first. If vision is used as the robot's primary source of information about the environment, the object of interest must be identified in the image and subsequently localized in 3D space. Generally, in cluttered environments, detecting a certain object is not an easy problem. Recognition and localization of a known object in the image is based on matching its certain previously defined features such as: shapes, sizes, colors, texture, etc. The choice of features and the matching algorithm is arbitrary and it depends primarily on the specification of the object and it will not be discussed here. This task becomes much more difficult if we want not only to localize the object in the scene (2D localization), but also to find its 3D pose (6D localization) in relation to the camera frame or to the world frame. Typical method used for 6D object localization is to calculate the pose based on the correspondence between 3D model and image coordinates from camera image. Most of the works on grasping and manipulation planning have assumed the existence of a database with 3D models of objects encountered in the robot surroundings and a 3D model of the robot itself [2], [9].

*Reach* – for reaching an object the *Reach* skill uses motion planning to compute a collision-free trajectory for moving the robot arm from its current pose to one that allows grasping of a specified object with a hand. If both arms are free, then *Reach* can employ each of the arm to move to the vicinity of the object. If one arm is currently grasping an object, *Reach* can be used for the other arm to prepare for grasping. *Reach* skill requires closed-loop execution to permit interaction with the environment. We utilize a position based, end-effector open loop visual servo with stand-alone camera to perform reaching operation [23].

*Grasp* – this skill is used for grasping objects for manipulation. Grasps are a special subset of manipulation skills that aimed at constraining the mobility of the object. *Grasp* should allow to perform different types of grasps depending on the hand structure. The parameters of the *Grasp* skill are: *grasp type*, *grasp starting point*, *approaching direction*, *hand orientation*. Grasp type determines the grasp execution control, namely, the hand preshape posture, the way the hand approaches the objects, the hand control strategy. For approaching the object, the hand is positioned at point in the vicinity of the object. The approaching line is determined by the grasp starting point and the approaching direction.

*Turn* – this skill is equivalent to *Reach* with an object or tool grasped as the end-effector, rather than the hand. Given an object grasped by the robot one hand or two hands, *Transfer* skill utilize motion planning to move the robot to configuration such that the object is at target pose. *Transfer*, like *Reach* requires closed-loop execution.

*Release* – this skill performs an action opposite to the *Grasp*, it simply release the object in the current configuration.

## 4. An Example of Two-Handed Manipulation

As an example of the task for two-handed manipulation we chose the manipulation of Rubik's cube puzzle. We used a Rubik's cube as an object to be identified, localized and manipulated. Rubik's cube combinatorial puzzle was invented by Ernő Rubik of Hungary in 1974. The standard $3 \times 3 \times 3$ version of the Rubik's cube consists of 27 sub-cubes, or cubies, with different colored stickers on each of the exposed sides of the sub-cubes. In its goal state each of the six faces has its own color. The total state space for solving a scrambled Rubik's cube is sized at $(3^{8-1} \cdot 8!) \cdot (2^{12-1} \cdot 12!)/2 = 43,252,003,274,489,856,000 \approx 4.3 \times 10^{19}$. Obviously, this number of states is prohibitively large for any sort of a brute force search technique, which is why specialized algorithms are needed to solve the Rubik's cube puzzle. However, the presentation of the algorithms for recognizing and solving Rubik's cube are not discussed in this paper, some information about these algorithms can be found in [9].

In this particular case we are interested in a coordinated manipulation in which both hands are manipulating the same object, thus creating a closed kinematic chain [5]. This task was chosen as it closely resembles the tasks that service robots have to execute. The process of manipulation of the Rubik's cube involves all aspects of visual serving to the vicinity of the cube, alignment of robot arms with the cube, grasping it with the grippers, and finally rotating the adequate face of the cube. The last three actions are repeated as many times as the number of moves is required to solve the scrambled cube. Here, we assume that from the high-level task planning system, i.e., Rubik's cube solver, a sequence of the turns of the faces is obtained. The goal is to plan a proper sequence of hand movements and grasping actions for both arms.

### *4.1. Problem Formulation*

The task of solving Rubik's cube needs several sensor-based operations such as:

- recognizing the cube in the image and localizing it in the robot workspace – *Localize* skill,
- approaching the cube while avoiding collisions by using visual information – *Reach*,
- grasping the cube using force/torque measurements for stiffness control and eye-in-hand and tactile sensors mounted in the jaws – *Grasp*,
- re-grasping the cube in order to identify the cube's initial state – *Grasp* → *Release* sequence,
- turning the faces of the cube while avoiding jamming using information from force/torque sensor for implementing interaction control – *Grasp* → *Turn* → *Release* sequence performed $n$ times, where $n$ is the number of moves required to solve the cube.

To support the programmer with task specification, object frames and feature frames are introduced, as well as suitable local coordinates to express the relative pose between these frames. Figure 3 presents the geometrical structure of the two-arm manipulation system, and coordinate frames $F_i$ attached to the appropriate components of the system together with the distribution of sensors.

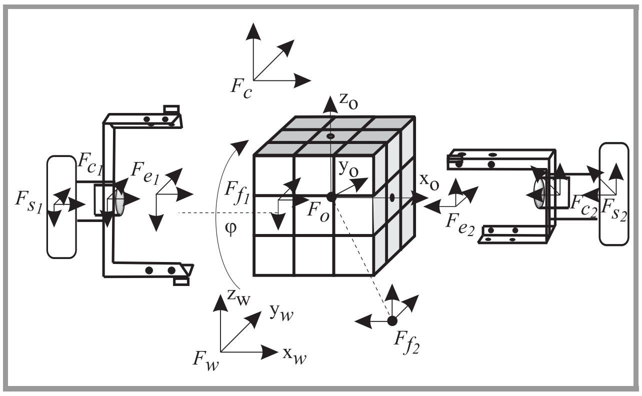



***Fig. 3.*** Coordinate frames attached to the two-handed robotic system.

Coordinate frames $F_{e_i}$, $i = 1,2$ are attached to the grippers and sensor frames $F_{s_i}$ and $F_{c_i}$ attached to the force sensors and to the eye-in-hand cameras, respectively. Given the view of the scene, the robot should be able to recognize the cube and localize it in the robot workspace. As a result of the visual localization, the position and orientation of the object frame $F_o$ attached to the cube with respect to (w.r.t) world frame $F_w$ is computed (as described earlier).

To describe the task for each hand two feature frames $F_{f_i}, i = 1,2$ are introduced, as shown in Fig. 3. These frames are used to plan manipulation skills such as approach trajectories for both hands, grasp and release operations, and hand movements to turn the cube faces.

The $4 \times 4$ matrix ${}^i_jT$ is a homogenous transformation matrix ${}^i_jT \in \mathbb{SE}(3)$ (where $\mathbb{SE}(3)$ is a special Euclidean group of a rigid body motions in $\mathbb{R}^3$ [24]) is a linear operator used in a mapping between the appropriate coordinate frames. Matrix ${}^i_jT$ may be interpreted as the representation of the pose of the frame $F_j$ w.r.t. frame $F_i$. Left-hand superscript is omitted (i.e., ${}_jT$) when the reference frame is evident from the context, e.g., it is the world frame $F_w$.

Pre-computed sequence of turns of the faces can be described in the $F_o$ coordinate frame as a sequence of rotations about unit vectors $\hat{\boldsymbol{x}}_o, \hat{\boldsymbol{y}}_o, \hat{\boldsymbol{z}}_o$ of its axes:

$$ {}^oR(\boldsymbol{u}, \varphi) = \mathrm{Rot}(\boldsymbol{u}, \varphi), \tag{1} $$

where $\boldsymbol{u} = \hat{\boldsymbol{x}}_o, \hat{\boldsymbol{y}}_o$ or $\hat{\boldsymbol{z}}_o$, and $\varphi = -\frac{\pi}{2}, -\pi, \frac{\pi}{2}$ or $\pi$.

The desired grasp configurations w.r.t. coordinate frame $F_o$ are described by the following matrices:

- ${}^o_{f_1}T$ – to grasp a single slice,
- ${}^o_{f_2}T$ – to grasp two slices simultaneously.

Locations of the possible contact regions on the cube are imposed by the specific shape of the gripper jaws. The shape of each of the jaws of the gripper matches the form of the corner of the cube. The cube is being grasped diagonally in such a way that either one or two layers are immobilized, where the corner pieces of one layer define the diagonal.

Now, we have to plan such a sequence of admissible grasps $({}^o_{f_1}T, {}^o_{f_2}T)$ that enable each single turn of the face without re-grasping:

$$ {}_{f_1}T = {}_oT\,{}^o_{f_1}T; \quad {}_{f_2}T = {}_oT\,{}^o_{f_2}T \tag{2} $$

The conditions of grasp feasibility are as follows:

$$ {}_{f_1}T = {}_{b_1}T\,{}^{b_1}_{e_1}T(\boldsymbol{q}_1)\,{}^{e_1}_{o}T\,{}^o_{f_1}T; \quad {}_{f_2}T = {}_{b_2}T\,{}^{b_2}_{e_2}T(\boldsymbol{q}_2)\,{}^{e_2}_{o}T\,{}^o_{f_2}T, \tag{3} $$

or equivalently

$$ {}_{f_1}T = {}_{b_2}T\,{}^{b_2}_{e_2}T(\boldsymbol{q}_2)\,{}^{e_2}_{o}T\,{}^o_{f_1}T; \quad {}_{f_2}T = {}_{b_1}T\,{}^{b_1}_{e_1}T(\boldsymbol{q}_1)\,{}^{e_1}_{o}T\,{}^o_{f_2}T, \tag{4} $$

where ${}_{b_i}T$, $i = 1,2$ is the homogenous transformation matrix from the world frame $F_w$ to the robot base coordinate frame $F_{b_i}$. Matrix ${}^{b_i}_{e_i}T(\boldsymbol{q}_i)$, $i = 1,2$ represents direct kinematics of the robot arm $i$, and $\boldsymbol{q}_i$ is the vector of joint coordinates of the arm $i$.

In this case grasp stability conditions are of a geometric nature and grasp synthesis is reduced to the choice of four contact regions on the cube (two for each gripper) from the given set of contacts and computing desired poses of both grippers, i.e., ${}_{e_1}T$ and ${}_{e_2}T$ which guarantee firm grasps. In fact, grasp synthesis comes down to the proper positioning of the grippers. Therefore grasp configurations can be described in the operational space as well as in the joint space.

When both grippers firmly hold the cube the closed kinematic chain is established. Now the motion planning problem is complicated by the need to maintain the closed loop structure, described by the loop closure constraint.

$$ {}^{b_1}_{e_1}T(\boldsymbol{q}_1)\,{}^{e_1}_{o}T(\varphi) - {}^{b_1}_{b_2}T\,{}^{b_2}_{e_2}T(\boldsymbol{q}_2)\,{}^{e_2}_{o}T = \mathbf{0} \qquad (5) $$

However, in our case, the motion of the closed chain linkage can be described in the $F_o$ coordinate frame as a single rotation about its axes (i.e., the elementary turn of the cube's face). For the frame $F_o$ chosen as it is shown in Fig. 3 these moves are rotations around its axes described in (1).

These moves can be easily transformed to the motions of the grippers. However, due to kinematic calibration errors, the two robot arms cannot be position controlled while executing the turns. This would cause excessive build-up of force in a rigid closed kinematic chain due to small misalignments. Therefore at this stage the motions have to be executed in position-force control mode.

### *4.2. Implementation of the Two-Handed Manipulation in the MRROC++ Framework*

The control system of the two-handed system equipped with special end-effectors, each composed of an electric gripper and diverse sensors, was implemented by using the MRROC++ [1] robot programming framework.

MRROC++ is a robot programming framework, thus it provides a library of software modules (i.e., classes, objects, processes, threads and procedures) and design patterns according to which any multi-robot system controller can be constructed. This set of ready made modules can be extended by the user by coding extra modules in C++ [9], [23], [25]. MRROC++ based controllers have a hierarchic structure composed of processes (Fig. 4) (some of them consisting of threads) supervised by the QNX Neutrino real time operating system. The underlying software is written in C++.

From the point of view of the executed task MP is the coordinator of all effectors present in the system. It is responsible for trajectory generation in multi-effector systems where the effectors cooperate tightly – as is the case in the presented system. The manipulation planning system contained in the MP transforms the solution obtained from Rubik's cube solver into a proper sequence of manipulation skills. In the MRROC++ framework these skills are implemented as motion generators, which are used by the Move instructions. Therefore the MP is responsible both for producing the plan of the motions of the faces of the cube and subsequently the trajectory generation for both manipulators. This trajectory can be treated as a crude reference trajectory for both arms. At a later stage this trajectory is modified by taking into account the force readings.

[1]The name is derived from the fact that this programming framework is the basis for the design of Multi-Robot Research-Oriented Controllers and that the underlaying software is coded in C++.

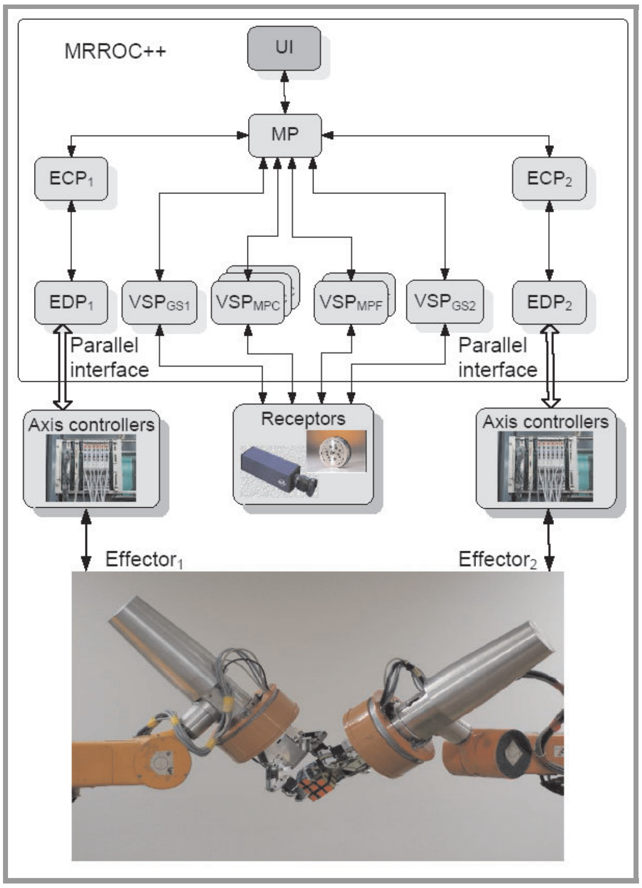



*Fig. 4.* MRROC++ based controller for the two-arm system.

Each effector has two processes controlling it: Effector Control Process ECP and Effector Driver Process EDP. The first one is responsible for the execution of the user's task dedicated to this effector (in our case the task is defined by the MP – it is defined by the reference trajectory that is to be executed by the manipulator), and the other one for direct control of this effector. The EDP is responsible for direct and inverse kinematics computations, as well as for both position and force servo-control.

### *4.3. Experiments*

The overall experimental setup consists of two 6 degree of freedom (dof) modified IRp-6 robot arms, each with a parallel jaw gripper Fig. 5.

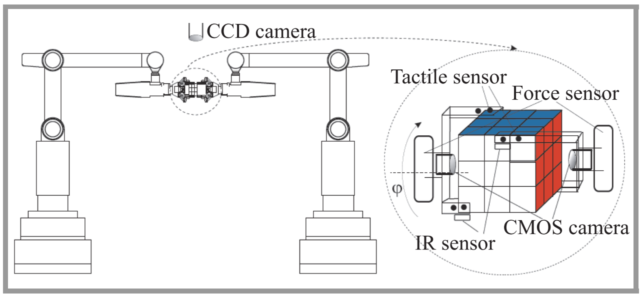



*Fig. 5.* Sensors used to locate and manipulate the Rubik's cube.

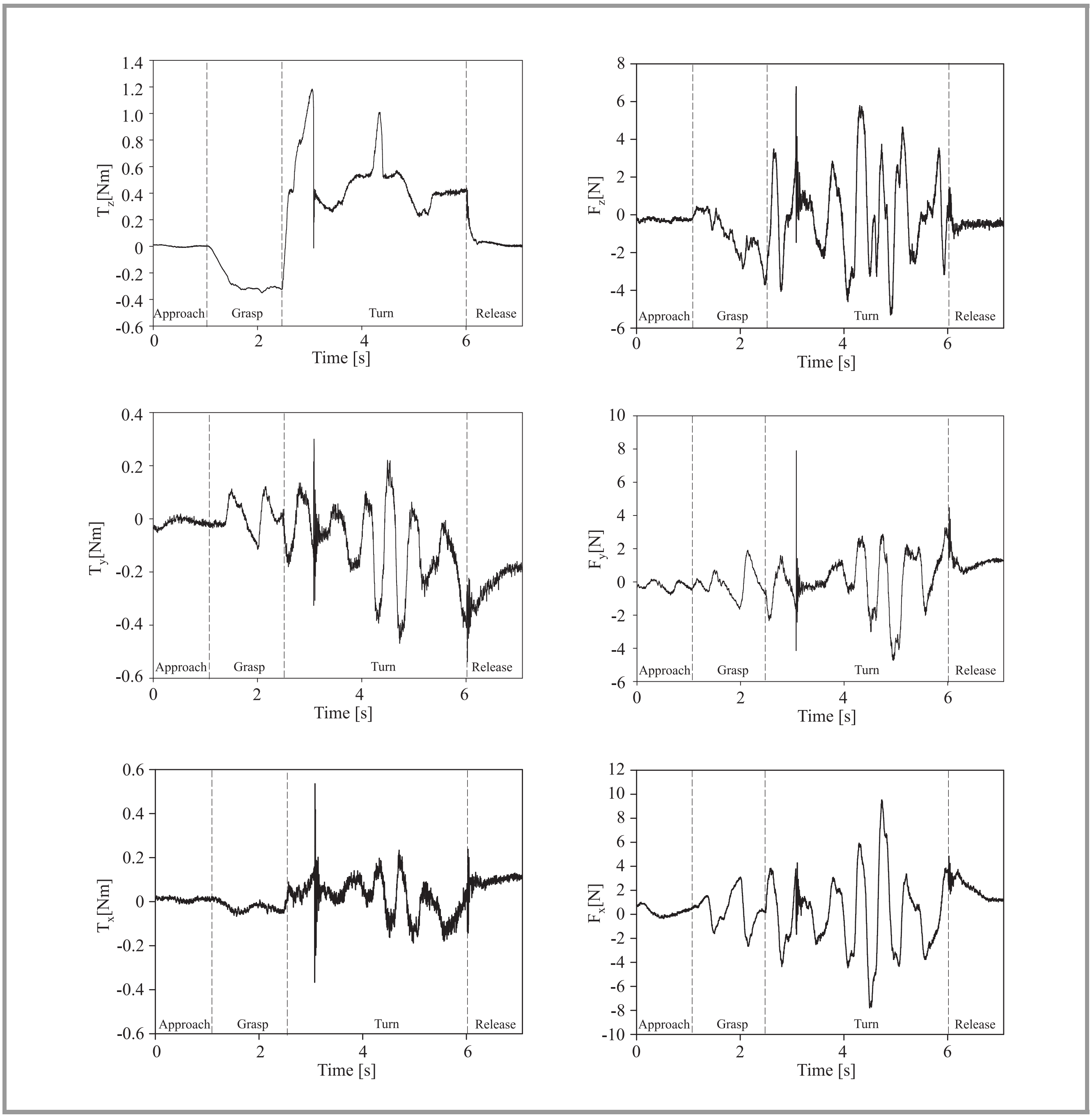



*Fig. 6.* Measured force and torque components while manipulating the Rubik's cube.

Each jaw was instrumented with tactile sensors which detect the presence of contacts with the grasped object. Moreover, each hand was equipped with a wrist-mounted six-axis force-torque sensor, and an eye-in-hand miniature CCD color camera [9]. Additionally, a global vision system with fixed-mount color camera and Digital Video Processor for fast image acquisition and realtime processing of the incoming data was used.

During the task execution either pure position control or position–force control is used, depending on the current task execution stage. Typically, these execution stages are position controlled in which there in no simultaneous contact between the two end-effectors and the cube, or between one of the end-effectors and the cube held by the operator. The stages, where such contact is present or expected to occur, are position–force controlled.

Cube grasping starts with one of the manipulators initiating the closing of the gripper jaws to catch the cube already held by the other manipulator or the operator. The manipulator currently holding the cube is commanded to keep the current position, hence it is position controlled. Figure 6 presents the force and torque plots for three stages of manipulation for the second manipulator, which is currently force-controlled.

Force/torque (F/T) sensors provide information about the magnitude and direction of the forces and torques that appear when the robot arms and the object are in contact. Free motion can be observed in the first phase (reaching the object), this stage occurs when one of the manipulators is currently holding the cube and the second one is approaching to gain a direct contact with the other side of the cube. Then, after the contact, grasping phase begins. The visible oscillations occur due to arms and Rubik's cube compliance. Once the cube is grasped firmly the torque stabilizes (Fig. 6). The rapid change in torque appears when the rotation of the cube face is initiated (turn phase), because initially the rotated face was jammed – this can be seen from the plot. In the release phase the gripper is opened, and the closed kinematic chain is disjoined.

## 5. Conclusion

In this paper, a framework for the description of two-arm/hand manipulation task based on the definition of a priori specified manipulation skills was proposed. The whole task was decomposed into a set of subtask each of which is resolved by a set of manipulation skills. To manage the task or environment variations the skills were parameterized. The parameters are generally related to the task variations, such as: type of a motion, grasping rule, an initial and final points, etc. Rubik's cube solving problem was used as a 3D manipulation task using two-arm robot system with diverse sensors such as vision, force/torque, tactile sensors. The manipulation task was implemented in the MRROC++ framework.

## References


[1] A. M. Okamura, N. Smaby, and M. R. Cutkosky, "An overview of dexterous manipulation" in *Proc. IEEE Int. Conf. Robot. Autom. ICRA 2000*, San Francisco, USA, 2000, pp. 255–262.

[2] T. Asfour, P. Azad, N. Vahrenkamp, K. Regenstein, A. Bierbaum, K. Welke, J. Schröder, and R. Dillmann, "Toward humanoid manipulation in human-centred environments", *Robot. Autonom. Sys.*, vol. 56, no. 1, pp. 54–65, 2008.

[3] A. G. Billard, S. Calinon, and F. Guenter, "Discriminative and adaptive imitation in uni-manual and bi-manual tasks", *Robot. Autonom. Sys.*, vol. 54, no. 5, pp. 370–384, 2006.

[4] Ch. Smith, Y. Karayiannidis, L. Nalpantidis, X. Gratal, P. Qi, D. V. Dimarogonas, and D. Kragic, "Dual arm manipulation – a survey", *Robot. Autonom. Sys.*, vol. 60, no. 10, pp. 1340–1353, 2012.

[5] W. Szynkiewicz and J. Blaszczyk, "Optimization-based approach to path planning for closed-chain robot systems" *Int. J. Appl. Mathem. Comp. Sci.*, vol. 21, no. 4, pp. 659–670, 2011.

[6] N. Vahrenkamp, D. Berenson, T. Asfour, J. Kuffner, and R. Dillmann, "Humanoid motion planning for dual-arm manipulation and re-grasping tasks", in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Sys. IROS 2009*, St. Louis, MO, USA, 2009, pp. 2464–2470.

[7] R. Zöllner, T. Asfour, and R. Dillmann, "Programing by demonstration: Dual-arm manipulation tasks for humanoid robots", in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Sys. IROS 2004*, Sendai, Japan, 2004, vol. 1, pp. 479–484.

[8] Y. Guiard, "Asymmetric division of labor in skilled bimanual action: the kinematic chain as a model" *J. Motor Behav.*, vol. 19, no. 4, pp. 486–517, 1987.

[9] C. Zieliński, W. Szynkiewicz, T. Winiarski, M. Staniak, W. Czajewski, and T. Kornuta, "Rubik's cube as a benchmark validating MRROC++ as an implementation tool for service robot control systems", *Industr. Robot: An Int. J.*, vol. 34, no. 5, pp. 368–375, 2007.

[10] C. Zieliński and T. Winiarski, "Motion generation in the MRROC++ robot programming framework", *The Int. J. Robot. Res.*, vol. 29, no. 4, pp. 386–413, 2010.

[11] S. M. LaValle, *Planning Algorithms*. Cambridge, U.K.: Cambridge University Press, 2006.

[12] R. Alami, T. Siméon, and J. P. Laumond, "A geometrical approach to planning manipulation tasks. The case of discrete placements and grasps", in *Proc. Fifth Int. Symp. Robot. Res.*, Cambridge, MA, USA, MIT Press, pp. 453ℓ-463, 1990.

[13] Ch. Nielsen and L. Kavraki, "A two-level fuzzy prm for manipulation planning", in *Proc. EEE/RSJ Int. Conf. Intell. Robots Sys. IROS*, Takamatsu, Japan, 2000, pp. 1716–1721.

[14] T. Siméon, J.-P. Laumond, J. Cortés, and A. Sahbani, "Manipulation planning with probabilistic roadmaps", *The Int. J. Robot. Res.*, vol. 23, no. 7–8, pp. 729–746, 2004.

[15] S. Cambon, R. Alami, and F. Gravot, "A hybrid approach to intricate motion, manipulation and task planning", *Int. J. Robot. Res.*, vol. 28, pp. 104–126, 2009.

[16] G. Milighetti, H. B. Kuntze, C. W. Frey, B. Diestel-Feddersen, and J. Balzer, "On a primitive skill-based supervisory robot control architecture", in *Proc. IEEE Int. Conf. Robot. Autom. ICRA 2005*, Barcelona, Spain, 2005, pp. 141–147.

[17] J. J. Kuffner and S. M. LaValle, "RRT-connect: An efficient approach to single-query path planning", in *Proc. IEEE Int. Conference Robot. Autom. ICRA 2000*, San Francisco, USA, 2000, pp. 995–1001.

[18] E. Niewiadomska-Szynkiewicz and M. Marks, "Software Environment for Parallel Optimization of Complex Systems", in *Applied Parallel Scientific Computing, Lecture Notes in Computer Science LNCS 7133*, K. Jonasson, Ed. Springer, 2012, pp. 86–96.

[19] A. Bicchi and V. Kumar, "Robotic grasping and contact: a review", in *Proc. IEEE Int. Conf. Robot. Autom. ICRA 2000*, San Francisco, CA, USA, 2000, pp. 348–352.

[20] M. Pardowitz, S. Knoop, R. Dillmann, and R. D. Zoellner, "Incremental learning of tasks from user demonstrations, past experiences, and vocal comments", *IEEE Trans. Sys. Man, Cybernet., Part B*, vol. 37, no. 2, pp. 418–432, 2007.

[21] T. A. Henzinger, "The theory of hybrid automata", in *Proc. Eleventh Ann. IEEE Sympo. Logic Comp. Sci. LICS'96*, New Brunswick, USA, 1996, pp. 278–292.

[22] M. S. Branicky, V. S. Borkar, and S. K. Mitter, "A unified framework for hybrid control: Model and optimal control theory", *IEEE Trans. Autom. Contr.*, vol. 43, pp. 31–45, 1998.

[23] W. Szynkiewicz, C. Zieliński, W. Czajewski, and T. Winiarski, "Control Architecture for Sensor-Based Two-Handed Manipulation", in *CISM Courses and Lectures – 16th CISM-IFToMM Symposium on Robot Design, Dynamics and Control, RoManSy'06*, T. Zielińska and C. Zieliński, Eds., no. 487, pp. 237–244. Wien, New York: Springer, 2006.

[24] R. M. Murray, Z. Li, and S. S. Sastry, *A Mathematical Introduction to Robotic Manipulation*. CRC Press, 1994.

[25] C. Zieliński, W. Szynkiewicz, and T. Winiarski, "Applications of MRROC++ robot programming framework", in *Proc. Fifth Int. Worksh. Robot Motion Control RoMoCo'05*, Dymaczewo, Poland, 2005, pp. 251–257.

**Wojciech Szynkiewicz** received Ph.D. degree in Robotics in 1996 from the Warsaw University of Technology (WUT). He is an Assistant Professor employed in the Institute of Control and Computation Engineering of WUT. From 1999 to 2003, he was the Deputy Director and Secretary to the Scientific Council of the Research Center for Automation and Information-Decision Technology-CATID. His research activities concentrate on multi-robot/multi-agent systems, motion planning, autonomous mobile robots, robot controller structures, and real-time and distributed systems. He works on sensor-based motion planning and control algorithms for multi-robot systems, including service, personal and mobile robots.
E-mail: W.Szynkiewicz@elka.pw.edu.pl
Institute of Control and Computation Engineering
Warsaw University of Technology
Nowowiejska st 15/19
00–665 Warsaw, Poland

Paper

# Optimizing the Structure of Vector Bend and Strain Sensor on the Base of Three-Core Microstructured Fiber

Igor Goncharenko[a], Marian Marciniak[b], Alexei Konojko[c], and Vitaly Reabtsev[a]

[a] Institute for Command Engineers of the Ministry of Emergencies, Minsk, Belarus
[b] National Institute of Telecommunications, Warsaw, Poland
[c] Institute of Physics, National Academy of Sciences, Minsk, Belarus

**Abstract—In the paper the optical sensor allowing measuring a direction, values and localization of bends and stresses in building structures is described. The sensitive element of the sensor is the microstructured fiber with three cores. The use of three-core fiber makes it possible to define the direction of deformation. Distribution of mode fields in fiber cores depending on fiber structure and bend value and direction is analyzed. The optimization of the sensitive element parameters depending on the application is proposed.**

***Keywords—fiber bend, method of lines, microstructured fiber, mode field distribution, multicore fiber, optical sensor.***

## 1. Introduction

Buildings, bridges, tunnels, dams, cranes and other constructions require effective maintenance to keep them working safely. For these purposes it is necessary to control continually their conditions, i.e., to measure their bend, stress, strain, temperature, vibration, formation of cracks and uniformity of constructional elements. The information about controlled variables makes it possible to calculate a settlement of footing and internal stresses, and strains in building structures, to obtain notion about structural elements displacement taking into account a rotation angle, and to draw a conclusion about degree of construction maintenance safety on the base of numerical modeling, and comparison of measured and master data. Optical fiber information-measuring systems are very promising and attractive instruments for monitoring building structures of different functions [1]–[3]. Sensors on the base of optical fibers have important technical advantages over traditional instruments, such as high mechanical strength, tolerance to high temperature, vibration and other environmental activities, immunity to electromagnetic interference, chemical neutrality, ability to carry out noncontact and remote sensing. Fiber-based sensors do not drift over time, so recalibration is unnecessary. By embedding optical fiber sensors into the structure of a building during construction, engineers can check the building's behavior throughout its lifetime and collecting at a central monitoring station a large number of continuous measurements.

By convention, optical fiber sensors can be divided into three types: point, distributed and quasi-distributed sensors [1], [3]–[5]. The point sensors measure and control parameters in particular points of the object. Usually, such devices are of a small size and high precision. The most frequently used point sensors are based on fiber Bragg gratings, long period gratings and interferometers. They can be utilized as local temperature indicators, strain gages, pressure sensors, accelerometers, etc.

The unquestionable advantage of the distributed sensors is the possibility to control parameters along the length (volume) of the object in any point where sensor fiber is introduced. The mode of operation of such sensor systems bases on the analysis of the parameter changes along the fiber length and on nonlinear optical effects. The disadvantage of the distributed measurement of the fiber parameter along the length is relatively small accuracy of definition of perturbation localization (several meters lengthwise), and relatively modest accuracy of value measurement. Distributed sensors can be used for control of wide areas, as the sensors of radiation and temperature, for analyzing, for example, temperature gradients.

Measuring systems on the base of quasi-distributed sensors are the attempt to combine the advantages of the both types described above. The quasi-distributed sensor comprises an array of point sensor elements connected by a single fiber. Each element has the unique characteristics and thus permits to analyze its state, independently from other sensor elements. The accuracy of such systems is determined by the accuracy of a separate sensor, and the array can include more than 100 elements. Sensor arrays allow monitoring the complex objects, engineering construction, bridges, tunnels, hull of ships and aircrafts, oil wells, etc., analyzing gradient of distribution of temperature, loadings, pressure, controlling a large number of point objects.

Nevertheless, the quasi-distributed sensors are not able to control the object state along the whole length of the fiber and hence unable to substitute the distributed sensors entirely. Moreover, the wavelength-division multiplexing (WDM) technique and set of photodetectors have to be used for transmitting the data from sensor array in a single fiber. Consequently, the number of available WDM channels limits the number of sensor elements.

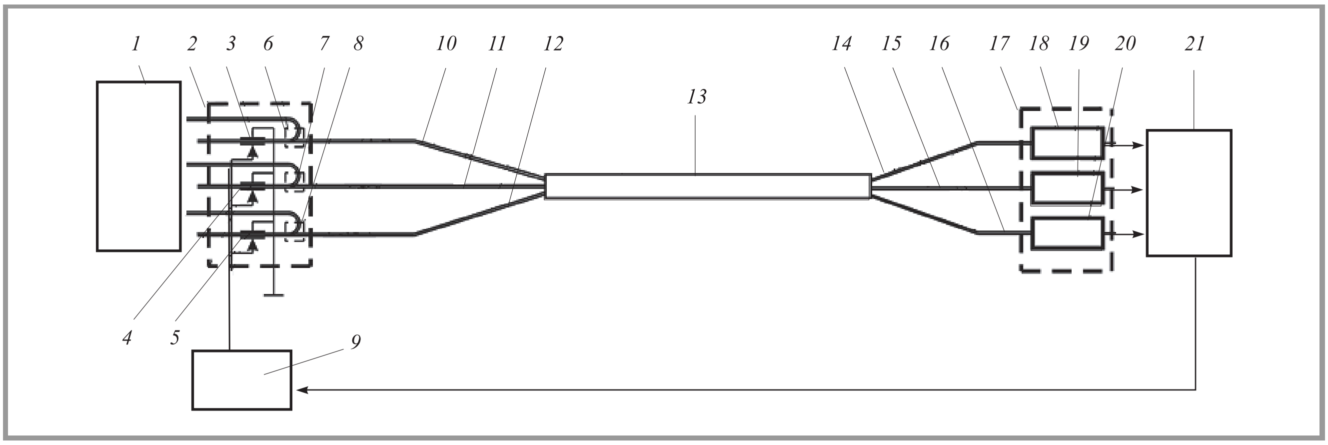



*Fig. 1.* Structural diagram of the optical fiber vector bend and strain sensor device. Explanations: *1* – broadband radiation source, *2* – controlled spectral filter, *3–5* – fiber Bragg gratings, *6–8* – *Y*-shaped couplers, *9* – control voltage block, *10–12* – input optical fibers, *13* – microstructure fiber (sensitive element), *14–16* – output optical fibers, *17* – photodetector array, *18–20* – photodetectors, *21* – measuring device.

Therefore there is still a practical interest to develop and optimize the structure of distributed sensors for continuous measurement of the parameters of the objects under control. One way to increase the distributed sensor accuracy is to measure the parameters on multiple wavelengths and then to average the logged data. The fibers have to operate in a single mode regime for all used spectral regions. From that point of view, the use of microstructured fibers as a sensing elements seems to be very promising. Microstructured fibers have enormous potential in achieving exotic microstructures with relative ease of manufacturing, and they can also be made as single-mode over a wider range of wavelengths in contrast to conventional fibers [6]–[9]. For sensor applications, two or more guiding cores are required, rather than just a single one [10]–[17]. By manipulating air hole diameter $d$, holes pitch $\Lambda$, number of holes $N$, and the distance between guiding cores, it is possible to vary the properties of microstructured fiber such as dispersion, leakage loss, single-mode regime, numerical aperture and effective-mode area, and control the mode field distribution, and therefore modifying the characteristics of the sensor.

Existing sensors are not able to define the bend direction. On the other hand, the ability to determine direction of the deformation can be very important for sensors used, for instance, for monitoring the condition of bridges, cranes, blades of wind turbines, etc. Therefore there is a need to develop sensors enabling to measure both value and direction of the deformation along the whole length of the controlled object, with sufficient accuracy for practical purposes.

In this paper we describe the vector bend and stress sensor based on a three-core microstructured fiber. In Section 1, we consider the structural diagram and operation principles of the sensor. Such a device can be used for precise measurements of the value and direction, both of the bends and displacements of building units, and determining their internal stresses and strains. Therefore the optimization of sensor element parameters depending on the purpose of use has to be carried out. For that, in Section 2, we have calculated the distribution of the mode fields in fiber cores and analyze it depending on the fiber structure and bend.

## 2. Structural Diagram and Principle of Operation of the Sensor

Figure 1 shows the structural diagram of the proposed sensor device. The device consists of a broadband radiation source in the form of the array from three light-emitting diodes, controlled spectral filter based on controlled fiber Bragg gratings (FBG), sensing cell, control voltage block, photodetector array and measuring device. As the sensing element, we propose to use a three-core microstructured fiber. Three cores correspond to hexagonal symmetry of the microsructured fiber structure and enable to calculate the bend direction by the simplest algorithm [16], [17].

Light from the broadband radiation source simultaneously enters into the corresponding cores of the microstructure fiber through the input fibers. Bending the microstructure fiber leads to redistribution of mode power between cores [14], [15]. The difference between mode amplitudes in different cores increases with bend radius decrease. For this reason, by comparing the measured power in microstructure fiber cores it is possible to define the bend radius of the fiber. The bend direction is determined on the base of the ratio of amplitudes of radiation of separate wavelength bands in different cores of the microstructured fiber.

Controlled fiber Bragg gratings are used for measuring the frequency components of the optical signals coming to the photodetectors. In order to increase the accuracy of the measurement of the bend value and direction, the set of signals on separate wavelengths is injected into each core

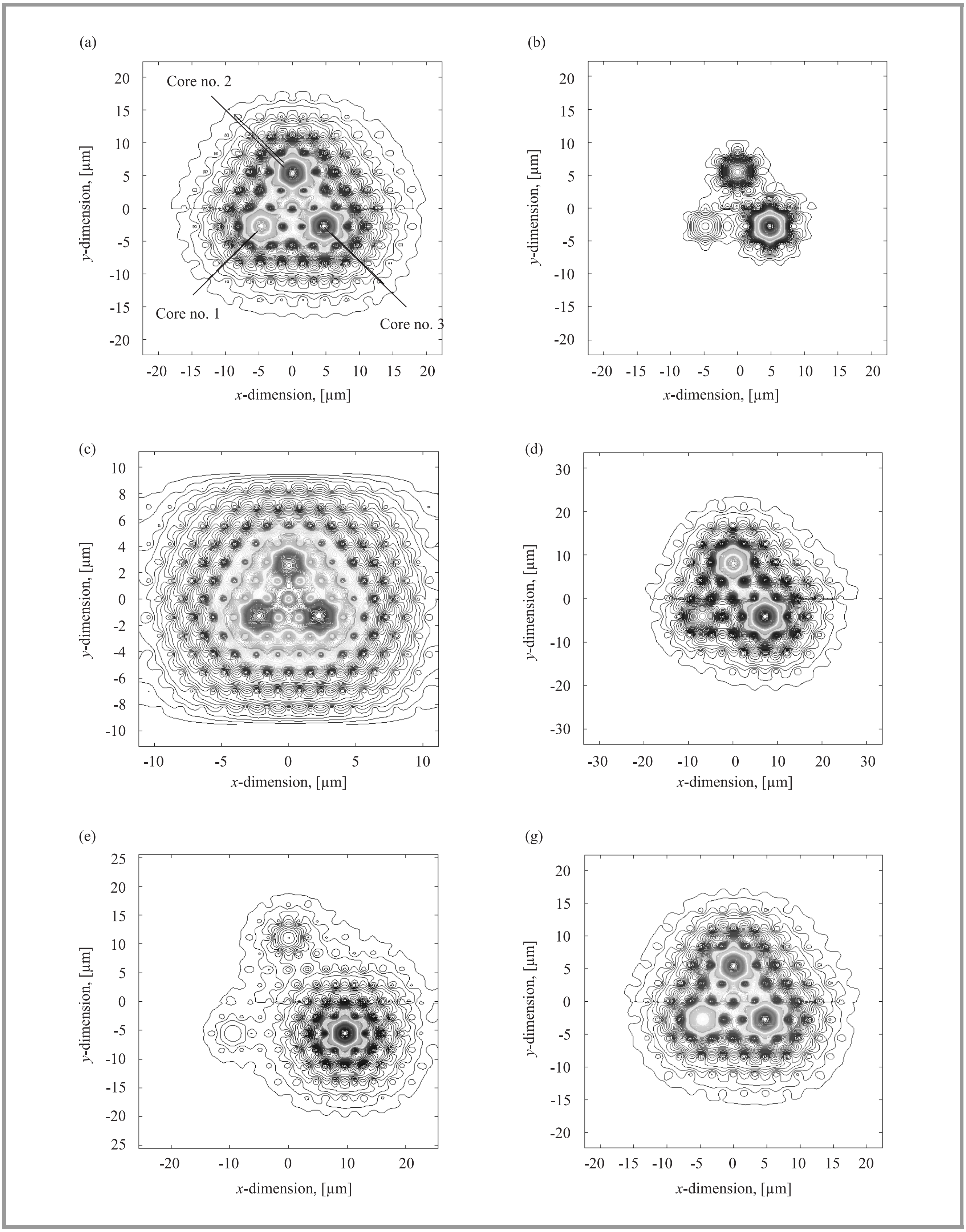



*Fig. 2.* Transverse distribution of the optical radiation intensity in cores of microstructure fibers bent with $R = 10$ cm and core separation in two holes, $d/\Lambda = 0.2$, $\Lambda = 3.2$ $\mu$m (a); $d/\Lambda = 0.4$, $\Lambda = 3.2$ $\mu$m (b); $d/\Lambda = 0.2$, $\Lambda = 1.6$ $\mu$m (c); $d/\Lambda = 0.2$, $\Lambda = 4.8$ $\mu$m (d); core separation in five holes, $d/\Lambda = 0.2$, $\Lambda = 3.2$ $\mu$m (e) for optical wavelength $\lambda = 1.5$ $\mu$m and core separation in two holes, $d/\Lambda = 0.2$, $\Lambda = 3.2$ $\mu$m, $\lambda = 1.3$ $\mu$m (g).

of microstructure fiber. Spectral ranges of the signals for each core are different. When the fiber is bent, the optical power on different wavelengths redistributes over all cores. A control voltage supplied into the FBG's electrodes changes the refractive index of the electrooptical material. That results in variation of optical wavelength, at which the FBG reflection is maximal (Bragg wavelength). Therefore by supplying the variable control voltage into the FBG's electrodes, it is possible to scan sequentially the spectrum of the signals coming into the photodetectors. By averaging the signals coming from the sensing element on different wavelengths, it is possible to determine the bend value and direction with high accuracy.

## 3. Calculating the Mode Fields and Optimizing the Sensor Design

We use the algorithm based on the Method of Lines [14], [15], [18]–[21] modified for the structure under investigation, for calculation of the mode field distribution and dispersion parameters of multicore microstructure fibers.

By using the developed algorithm we have calculated transverse field distribution of the propagating fiber modes for different bend values and different fiber parameters, like an air hole diameter $d$, hole pitch $\Lambda$, core separation, etc.

The results of calculations are shown in Figs. 2–4. Figure 2 presents contour pictures of intensity distribution of optical radiation at $\lambda = 1.5$ $\mu$m wavelength in cores of bent microstructure fibers with core separation in two holes, $d/\Lambda = 0.2$, $\Lambda = 3.2$ $\mu$ m (a), $d/\Lambda = 0.4$, $\Lambda = 3.2$ $\mu$m (b), $d/\Lambda = 0.2$, $\Lambda = 1.6$ $\mu$m (c), $d/\Lambda = 0.2$, $\Lambda = 4.8$ $\mu$m (d), core separation in five holes, $d/\Lambda = 0.2$, $\Lambda = 3.2$ $\mu$m (e) and core separation in two holes, $d/\Lambda = 0.2$, $\Lambda = 3.2$ $\mu$m, $\lambda = 1.3$ $\mu$m (g). Bend radius of all fibers is $R = 10$ cm.

It follows from the figures that bending the fiber causes a redistribution of the mode power between fiber cores. The ratio of optical power transmitting in two cores located along the bend direction defines the bend value, while the relative shares of optical power in each of the three cores depend on bend direction. Thus by measuring the ratio of optical radiation intensity in three fiber cores, it is possible to determine both direction and value of the fiber bend. For instance, in Fig. 2 the bend direction corresponds to the line connecting cores 1 and 3, and optical power redistribution caused by the bend occurs mainly between these cores. When the bend direction changes by 60°, the power redistribution takes place mainly between cores 1 and 2 and so forth.

Figure 3 shows the maximal values of mode field amplitudes in fiber cores no. 1 and 3 located on bend axes as a function of bend radius. Amplitude values are normalized to the mode amplitude in core 2. Numbers of cores are indicated in Fig. 2(a). Figure 3(a) shows the mode amplitudes of radiation at $\lambda = 1.5$ $\mu$m wavelength for different core separation and ratio $d/\Lambda$ values. Curves *1* and *2* represent respectively the mode amplitudes in cores 1 and 3 of the fiber with $d/\Lambda = 0.2$ and core separation in two holes, curves *3* and *4* show the same in cores of the fiber with $d/\Lambda = 0.4$, core separation in two holes, curves *5* and *6* are related to a fiber with $d/\Lambda = 0.2$ and core separation in five holes. For all curves the hole pitch is

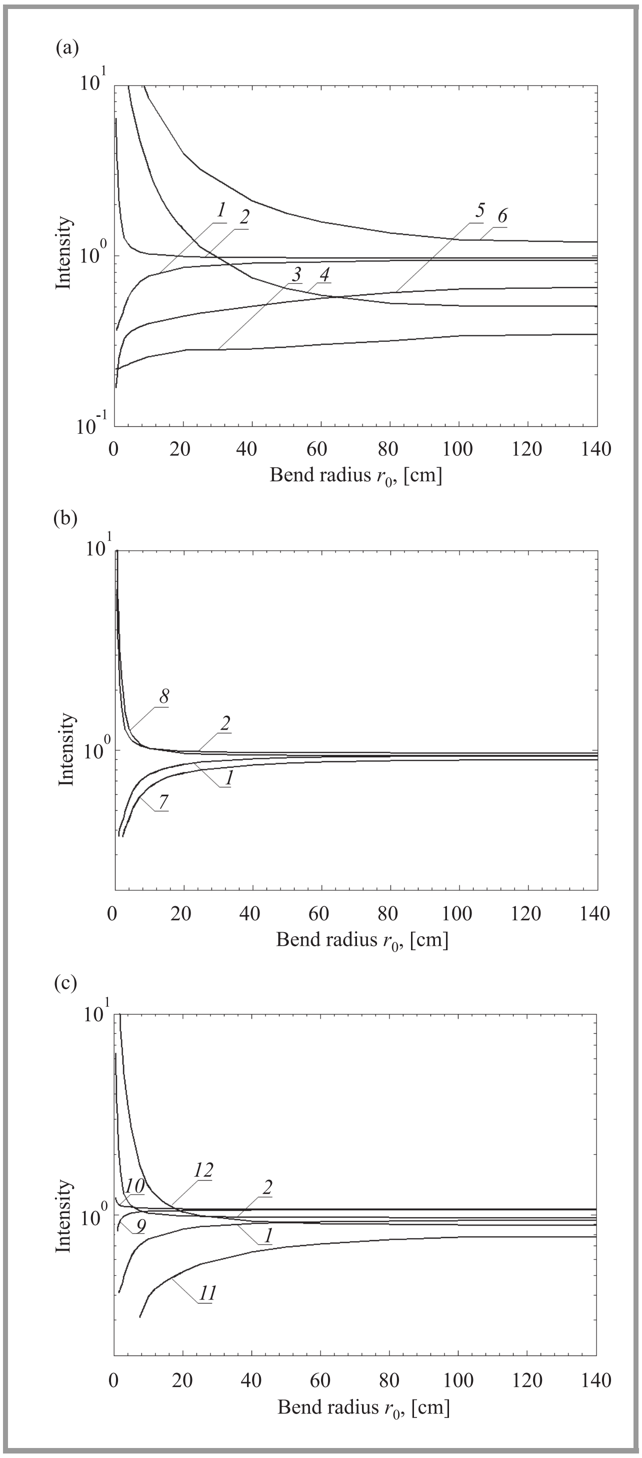



*Fig. 3.* Maximal values of mode field amplitudes in fiber cores located on bend axes versus bend radius for different values of core separation and ratio $d/\Lambda$ (a), different radiation wavelengths (b), and different values of hole pitch (c).

$\Lambda = 3.2$ $\mu$m. Curves *7* and *8* (Fig. 3(b)) relate to the fiber with parameters the same as curves *1* and *2*, but for $\lambda = 1.3$ $\mu$m. The comparison of the dependencies of mode amplitudes on bend radius for different values of hole-to-hole separation is presented in Fig. 3(c). Curves *9* and *10* respectively represent the mode amplitudes in cores 1 and 3 of the fiber with core separation in two holes, $\Lambda = 1.6$ $\mu$m, curves *11*, *12* relate to the fiber with $\Lambda = 4.8$ $\mu$m. Optical wavelength is $\lambda = 1.5$ $\mu$m and parameter $d/\Lambda = 0.2$. Figure 4 shows the ratio of the mode field amplitudes in right and left accordingly to the bend cores in dependence on bend radius for fibers, with core separation in two holes, $\Lambda = 1.6$ $\mu$m and $d/\Lambda = 0.2$ (curve *1*), $\Lambda = 3.2$ $\mu$m, $d/\Lambda = 0.2$ (curve *2*), $\Lambda = 4.8$ $\mu$m, $d/\Lambda = 0.2$ (curve *3*), $\Lambda = 3.2$ $\mu$m, $d/\Lambda = 0.4$ (curve *4*), core separation in five holes, $\Lambda = 3.2$ $\mu$m and $d/\Lambda = 0.2$ (curve *5*) for $\lambda = 1.5$ $\mu$m wavelength. Curve *6* is plotted for the fiber having the same parameters as *2*, but for $\lambda = 1.3$ $\mu$m.

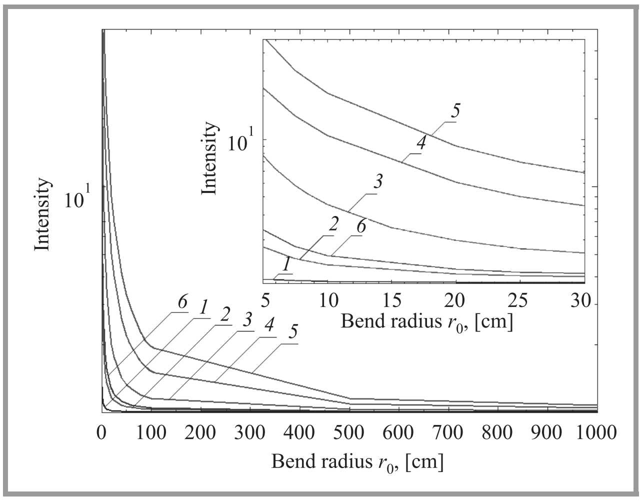



***Fig. 4.*** Dependence of relative mode field amplitudes in fiber cores located on bend axes on bend radius..

As it follows from the figures, fibers with a larger ratio of the air hole diameter $d$ to hole separation $\Lambda$, i.e., with larger air filling, as well as fibers with larger hole pitch and/or core separation are more sensitive to the bend. In such fibers, the mode fields are stronger concentrated in separate cores, and any break of the steady state conditions leads to substantial transmitting optical power from one core to another. In fibers with smaller ratio $d/\Lambda$, $\Lambda$ and/or core separation the mode fields in different cores interact stronger with each other, and such an arrangement is likely to be more stable to an external influence. For instance, for fibers with $d/\Lambda = 0.2$ and core separation in two holes the appreciable optical power coupling between cores takes place for bend radii of 500 cm. When the core separation is increased up to five holes, the same effect occurs for $R$ smaller than 900 cm. The sharp rise of the mode concentration in one core that is unusable for measuring and associated large radiation losses occur if such fibers are bent with radiuses around 3 and 10 cm, respectively. For fibers with $\Lambda$ equal to 1.6 and 4.8 $\mu$m a noticeable increase of power exchange between cores takes place for bend radiuses 300 and 600 cm, respectively. The sharp rise of power concentration in one core occurs for $R$ equal to 1 and 15 cm.

The reduction of the optical wavelength $\lambda$ increases the sensor sensitivity as well. However the wavelength variation is possible only within a limited range. Thus only slight enhancement of the mode power transmitting from one core to another with the fiber bend can be achieved in that way.

## 4. Conclusions

In the paper we have proposed the concept and operation principles of distributed optical vector sensor of bend and deformation based on three-core microstructured fiber. The sensor accuracy is higher than the one of conventional distributed sensors, due to measurement at several wavelengths and following average of the registered signal. This is achieved by using a microstructured fiber as the sensitive element which has a wide spectral range of single mode operation. The use of three-core fiber makes it possible to define the direction of deformation.

The optimization of the sensing element is carried out on the base of the numerical calculations of the mode parameters, and field distribution over the fiber cores, depending on fiber structure and bend value. The numerical simulation shows that the sensor devices based on the fibers with small core separation, small hole pitch and small air filling (small $d/\Lambda$) are preferable for measuring the flexible building construction liable to large bend. Such sensors possess wider measurement range of the bend value. For instance, the fiber with $d/\Lambda = 0.2$, $\Lambda = 3.2$ $\mu$m and core separation in two holes allows bend measuring up to $R = 3$ cm. The fibers with larger $d/\Lambda$, $\Lambda$ and/or core separation are more sensitive to the bend. Therefore, it is practically useful to apply them for measuring the small bends and deflections of the rigid building constructions, i.e., those effects that produce the internal stresses and strains in such units, and can cause their failure. For instance, fibers with $d/\Lambda = 0.4$ allows measuring the bends starting from $R = 800$ cm. Sensors based on such fibers en-sure high accuracy of the measurement of small bends and deflections of the constructive parts. It is necessary to notice that if the core separation or air filling of the fiber photonic crystal cladding is increased substantially, then the cores become completely isolated from each other. The modes of such structure correspond to the modes of separate fibers, and mode power transfer between cores wouldn't take place.

The small correction of the sensor sensitivity can be obtained by changing the wavelength of the used optical radiation.

Special attention has to be paid to regular laying the fiber when installing the sensor on the surface of controlled object. That is necessary in order to avoid the fiber twist that can cause incorrect determination of the bend direction.

## Acknowledgement

The authors thank the European Project COST Action MP0702 "Toward Functional Sub-Wavelength Photonic Structures" for stimulating interactions.


## References


[1] Yu. V. Guliaev, S. A. Nikitov, V. T. Potapov, and Yu. K. Chamorovski, "Optical fiber technologies, devices, sensors and systems", *Photon-Express – Science*, no. 6, pp. 114–127, 2005.

[2] V. B. Garmash *et. al.*, "Possibilities, tasks and trends of optical fiber measuring systems in modern instrument engineering", *Photon-Express – Science*, no. 6, pp. 128–140, 2005.

[3] C. Connolly, "Structural monitoring with fibre optics", *Europhotonics*, no. 2–3, pp. 16–18, 2009.

[4] E. Udd (Ed.), *Fiber Optic Sensors: An Introduction for Engineers and Scientists*. New York: Wiley, 2006.

[5] D. Inaudi and A. del Grosso, "Fiber optic sensors for structural control", in *Proc. 14th World Conf. Earthquake Engin.*, Beijing, China, 2008.

[6] J. Broeng, D. Mogilevtsev, S. E. Barkou, and A. Bjarklev, "Photonic crystal fibers: a new class of optical waveguides", *Optical Fiber Technol.*, no. 5, pp. 305–330, 1999.

[7] J. Broeng, T. Sondergaard, S. E. Barkou, P. M. Barbeito, and A. Bjarklev, "Waveguidance by the photonic bandgap effect in optical fibres", *J. Opt. A.: Pure Appl. Opt.*, vol. 1, no. 4, pp. 477–482, 1999.

[8] P. Russell, "Photonic crystal fibers", *Science*, vol. 299, pp. 358–362, 2003.

[9] P. St. J. Russell, "Photonic-crystal fibers", *J. Lightwave Technol.*, vol. 24, pp. 4729–4749, 2006.

[10] P. J. Roberts and T. J. Shepherd, "The guidance properties of multicore photonic crystal fibres", *J. Opt. A: Pure Appl. Opt.*, vol. 3, no. 6, pp. 133–140, 2001.

[11] L. Zhang and C. Yang, "A novel polarization splitter based on the photonic crystal fiber with non-identical dual cores", *IEEE Photon. Technol. Lett.*, vol. 16, pp. 1670–1672, 2004.

[12] A. Mafi and J. V. Moloney, "Shaping modes in multicore photonic crystal fiber", *IEEE Photon. Technol. Lett.*, vol. 17, pp. 348–350, 2005.

[13] J. H. Rothwell *et al.*, "Photonic sensing based on variation of propagation properties of photonic crystal fibres", *Opt. Express*, vol. 14, pp. 12445–12450, 2006.

[14] I. A. Goncharenko, "Radiation and bending loss in multi-core microstructure fibers", *Bull. Fund Fundamental Res.*, no. 3, pp. 91–98, 2006.

[15] I. A. Goncharenko and M. Marciniak, "Analysis of propagation of orthogonally polarized supermode in straight and curved multicore microstructured fibres", *J. Telecomm. Inform. Technol.*, no. 4, pp. 63–69, 2007.

[16] J. G. Burnett, P. M. Blanchard, and A. H. Greenaway, "Optical fibre-based vectoral shape sensor", *Strain*, vol. 36, no. 3, pp. 127–133, 2000.

[17] P. M. Blanchard *et al.*, "Two-dimensional bend sensing with a single, multi-core optical fibre", *Smart Material Struct.*, no. 9, pp. 132–140, 2000.

[18] R. Pregla, "The method of lines as generalized transmission line technique for the analysis of multilayered structures", *AEÜ Int. J. Electron. Commun.*, vol. 50, no. 5, pp. 293–300, 1996.

[19] R. Pregla, "The method of lines for the analysis of dielectric waveguide bends", *J. Lightwave Technol.*, vol. 14, pp. 634–639, 1996.

[20] S. F. Helfert and R. Pregla, "The method of lines: a versatile tool for the analysis of waveguide structures", *Electromagnetics*, vol. 22, pp. 615–637, 2002.

[21] I. A. Goncharenko, S. F. Helfert, and R. Pregla, "Radiation loss and mode field distribution in curved holey fibers", *AEÜ Int. J. Electron. Commun.*, vol. 59, pp. 185–191, 2005.



**Igor A. Goncharenko** graduated from the Physics Department of the Byelorussian State University in 1981. He received the Ph.D. degree in Physics and Mathematics from the USSR Academy of Sciences (Moscow) in 1985 and Doctor of Sciences (Habilitation) degree from the National Academy of Sciences of Belarus (Minsk) in 2001. Since September 2007 he takes up the position of Professor of the Department of Natural Sciences in the Institute for Command Engineers of the Ministry of Emergencies of the Republic of Belarus and part time Professor position in Belarussian State University. He is a member of advisory committee of the Supreme Certifying Commission of the Republic of Belarus. His fields of investigation include theory of complicated optical fibre and waveguide devices, optical information processing. He has contributed to a number of National and International Conferences and published more than 230 papers in scientific journals including 2 monographs and 53 patents.

E-mail: Igor02@tut.by
Institute for Command Engineers
Ministry of Emergencies of the Republic of Belarus
Mashinostroiteley st 25
220118 Minsk, Belarus



**Marian Marciniak** graduated in solid state physics from Marie Curie-Skłodowska University in Lublin, Poland, in 1977. He holds a Ph.D. degree in optoelectronics (1989), and a Doctor of Sciences (Habilitation) degree in physics/optics (1997). Actually he is a professor at the Department of Transmission and Optical Technologies at the National Institute of Telecommunications, and at Kielce Technical University, Faculty of Electrotechnics, Automatics, and Information Science. He authored over 300 publications, including a number of invited conference presentations. He serves as a Honorary International Advisor to the George Green Institute for Electromagnetics Research, University of Nottingham, UK. He was the originator (1999) and the leading organizer of the International Conferences on Transparent Optical Network ICTON series. He serves as the Chairman of COST Action MP0702 "Towards functional sub-wavelength photonic structures" (2008–2012). He serves, together with Prof. Trevor Benson, as The Editor in Chief of Optical and Quantum Electronics journal.

E-mail: M.Marciniak@itl.waw.pl
E-mail: marian.marciniak@ieee.org
National Institute of Telecommunications
Szachowa st 1
04-894 Warsaw, Poland

**Alexei I. Konojko** graduated from the Engineering and Physics Department of the Leningrad Institute in Precise Mechanics and Optics in 1977. He received the Ph.D. degree in Physics and Mathematics from the National Academy of Sciences of Belarus in 1990. At present he hold a position of leading research fellow in Institute of Physics, National Academy of Sciences of Belarus and part time associate professor position in Belarussian State University of Informatics and Radio Electronics. His fields of investigation include optical information processing and interaction of hetero quality hierarchical systems.
E-mail: kon54@yandex.ru
E-mail: kanojka@inel.bas-net.by
Institute of Physics
National Academy of Sciences of Belarus
Nezavisimosti Av. 68
220072 Minsk, Belarus

**Vitaly N. Reabtsev** graduated from the Engineering Department of the Institute for Command Engineers of the Ministry for Emergencies of the Republic of Belarus in 2010. In 2011 he received the M.Sc. degree in Prevention and Elimination of Emergencies from the Institute for Command Engineers and takes up the position of faculty member. The field of his research is the development of optoelectronic sensors for safety control of building and engineering constructions.
E-mail: V.Reabtsev@gmail.com
Institute for Command Engineers
Ministry of Emergencies of the Republic of Belarus
Mashinostroiteley st 25
220118 Minsk, Belarus

Paper

# How the Role-Based Trust Management Can Be Applied to Wireless Sensor Networks

Anna Felkner

*Research and Academic Computer Network (NASK), Warsaw, Poland*

**Abstract—Trust plays an important role in human life environments. That is why the researchers has been focusing on it for a long time. It allows us to delegate tasks and decisions to an appropriate person. In social sciences trust between humans was studied, but it also was analyzed in economic transactions. A lot of computer scientists from different areas, like security, semantic web, electronic commerce, social networks tried to transfer this concept to their domains. Trust is an essential factor in any kind of network, whether social or computer. Wireless sensor networks (WSN) are characterized by severely constrained resources, they have limited power supplies, low transmission bandwidth, small memory sizes and limited energy, therefore security techniques used in traditional wired networks cannot be adopted directly. Some effort has been expended in this fields, but the concept of trust is defined in slightly different ways by different researchers. In this paper we will show how the family of Role-based Trust management languages (RT) can be used in WSN. RT is used for representing security policies and credentials in decentralized, distributed access control systems. A credential provides information about the privileges of users and the security policies issued by one or more trusted authorities.**

***Keywords—access control, role-based trust management, trust, wireless sensor networks.***

## 1. Introduction

The concept of trust can be understood in quite various ways. Generally it can be based on personal experience, reputation or recommendation. A lot of work connected with trust has been done by sociologist, economists, psychologists and lately also by computer scientist. It has become very important in the late years as a consequence of the growth of fields such as Internet transactions or electronic commerce.

Establishing trust in a network gives two important benefits:

- it helps to make traditional security mechanisms more robust and reliable,
- it can solve the problems that can not be solved through traditional security mechanisms.

Wireless sensor networks are becoming increasingly important due to the growing range of their capabilities. The range of applications of WSN is so wide that it tends to invade our everyday life. The services offered by wireless sensor networks can be classified into four major categories: monitoring, alerting, providing information, and actuating. Their significance is more and more important, especially in relation to gathering information, in fields such as health care, defence, environmental and structural monitoring, homeland security, industry control, intelligent green aircrafts, smart roads and others. There are many applications which are intended to monitor physical and environmental phenomena, such as ocean and wildlife, pollution, earthquakes, and water quality. The main purpose of these WSNs is to provide physical information such as temperature, light, radiation, and others to a computer system and it offers efficient solutions in a great variety of application domains. The network can modify the state of an external system (e.g., barriers, traffic lights, irrigation system) according to the data, going beyond its sensing capabilities. In the future, a sensor network will survey our health, home, the roads we follow, the office or the industry we work in or even the aircraft we use, in an attempt to enhance our safety.

It is a quite young technology with many interesting research problems. One of the issues is security, and trust is a part of it. Very often applications which use WSN require high dependability. Also, networks which provide more sophisticated services require more effective security mechanism. Unfortunately, not all security solutions suitable for traditional networks are appropriate for WSN, because of their resource constraints.

Traditional trust management schemes that have been developed for wired and wireless ad-hoc networks are not well suited for wireless sensor networks, due to their higher consumption of resources such as memory and power. The sensor nodes are highly constrained in terms of communication bandwidth, processing resources, computational capabilities, memory space, and battery capacity.

Some of the approaches adopted for WSN try to imitate those for ad-hoc or peer-to-peer networks, but this is not always possible due to the difference in the features of these networks (like the computational power, energy-constraint and also the size of the networks). In this work we will try to show how our approach to the concept of trust management can be adjusted to wireless sensor networks.

There are some works connected with trust used in sensor networks to increase their security and reliability. Most of these works are based, or take into consideration, the concept of reputation. Quite often the trust is obtained as a function of reputation. Reputation is the opinion of one person about the other, in WSN it can be the opinion of

one node about another. It can be built over time based on the history of behavior of the node.
Guaranteeing that confidential data and services offered by a computer system are not made available to unauthorized users is an increasingly significant and challenging issue, which must be solved by reliable software technologies that are used for building high-integrity applications. The data, whether in electronic, paper or other form must be properly protected. The traditional solution to this problem is access control techniques by which users are identified, and granted or denied access to a system, data and other resources, depending on their individual or group identity. This approach fits well into closed, centralized environments in which the identity of users is known in advance. However, access control in such a resource constrained WSN provides significant challenges, and in an ongoing area of research and trust management is a specific kind of access control in which decisions are based on credentials issued by multiple principals.
The paper is organized as follows: An overview of the work related to trust management in wireless sensor networks is given in Section 2. Section 3 shows the overview of the family of Role-based Trust management languages, including syntax and inference system over *RT* credentials. Section 4 describes time validity in *RT* languages with inference system. Final remarks are given in Conclusions.

## 2. Related Work

Trust has been the focus of researchers for a long time. It origins derives from social sciences where trust between humans was studied. Since Marsh [1] introduced a computational model for trust in his thesis, trust mechanism has gradually obtained more and more researcher's ([2], [3], and so on) interest for its flexibility and extensibility. Numerous trust models were proposed in social network, distributed network, peer-to-peer computing, ad-hoc network, and so on.
Although intuitively easy to conceive, the notion of trust has not been formally defined unanimously. Trust in wireless sensor network is yet to adopt a formal definition. A dictionary definition states that trust is a belief or confidence in the honesty, goodness, skill or safety of a person, organization or thing [4]. It means that such a belief is based on explicit assessment of trustworthiness of the trusted party.
There is a large diversity in the understanding of the concept of trust. The concept of trust management in distributed systems was first defined in 1996, and the approach presented by authors of this paper is based on this definition. Along with the notion of trust, comes that of reputation, which is the opinion of one person about the other, of Internet buyer about an Internet seller, and one node in wireless sensor network about another. Also, reliability is connected with the trust concept. It was originally a measure of how long a machine can be trustworthy. Trust can be understood as a derivation of the reputation of an entity. Based on a reputation, a level of trust is granted upon an entity. The reputation itself has been built over time based on that entity's history of behavior, and may be reflecting a positive or negative assessment.
There are not many publications connected with the area of trust management systems for wireless sensor networks. Most of the work in this field has been made in the last few years (e.g., Reputation-based Framework for Sensor Networks (RFSN) [5], Agent-based Trust and Reputation Management (ATRM) [6], and Parameterized and Localized Trust Management Scheme (PLUS) [7]). However, big efforts have been made in related areas such as introducing the concept of trust management schemes for increasing security and reliability in peer-to-peer networks [8], [9] and ad-hoc networks [10]–[16]. There are some other works available in the literature, e.g., [17]–[22], and so forth, that discuss trust in WSN but not in much detail.
Very often in the literature, trust has been used in WSNs for assessing the availability, reliability, or security property of a node (e.g., whether a node is malicious or not) based on past interaction experiences [5], [23].
Ganeriwal *et al.* [5] were among the first who defined comprehensive trust management scheme for sensor networks. They propose a reputation-based framework for high integrity sensor networks based on a bayesian formulation (more specifically, a beta reputation system) where nodes maintain reputation for other nodes, and use it to evaluate their trustworthiness. The architecture of the framework consists of a watchdog mechanism, reputation, second hand information, trust, and behavior. In this framework each sensor node maintains reputation metrics which both represent past behavior of other nodes, and are used as an inherent aspect in predicting their future behavior. Reputation is stored in a table where the entries are built by the nodes through the watchdog mechanism. Nodes not only use their own direct observations, but they also exchange information with other nodes (second hand information). Reputation is calculated by using the beta reputation distribution and trust is obtained as a function of reputation. Then the behavior of a node is given according to whether the trust values are respectively above or below a given threshold.
A watchdog mechanism is also used in Chen *et al.* [17]. In their work reputation is similarly used in order to define a trust management system for wireless sensor networks. Their model uses probability, statistics and mathematical analysis. They consider the concept of certainty for trust. The first-hand information is aggregated by using a watchdog mechanism. A reputation space is defined considering the positive and negative outcomes, and trust space is defined from the reputation space. In [13] reputation is also considered as a way for building trust.
In [22] Shaikh *et al.* propose a lightweight group based trust management scheme (GTMS) for distributed wireless sensor networks in which the whole group will get a single trust value. Instead of calculating individual trust, in some cases it is much more appropriate to calculate the trust for the entire group. GTMS uses a hybrid trust management scheme instead of using centralised or distributed schemes,

which helps in keeping minimum resource utilisation at the sensor nodes.

In [24] Yao *et al*. introduce a framework similar to existing approaches for ad-hoc networks where trust values are assigned to each node. A trust evaluation process is performed based on the localised trust model and two kinds of knowledge: personal reference gained by interaction with the evaluated node (suspect node) and reputation sent by the juries (specific nodes).

Yao *et al*. [7] also propose a parametrised and localised trust management scheme for WSN security, especially for secure routing where each node maintains highly abstracted parameters to evaluate its neighbours.

Aivaloglou and Gritzalis [23] show a hybrid trust and reputation management protocol for WSNs by combining certificate-based and behaviour-based trust evaluations.

Zhiying *et al*. [25] find distributed trust models appropriate for large-scale sensor network security design, because each node focuses on the trustworthiness of its neighbours and can assess if these nodes comply with agreed security policies. Authors propose an appropriate security framework with different security schemes. Unfortunately, their work does not take into consideration the resource limits of nodes in sensor networks.

Zia in [26] proposes a security framework where integrating the reputation and trust management mechanism is used to provide a comprehensive security solution against well-known threats. In this work nodes monitor their neighboring nodes and rank the neighbors to execute a trust vote.

Momani *et al*. [27] also introduce a trust model and a reputation system for WSNs based on sensing continuous data. Chen *et al.* [28] propose a distributed agent-based trust management scheme where each agent node monitors the behavior of the nodes within its radio range, and broadcasts their trust ratings.

As it was shown just above, there is a large diversity in the understanding of the concept of trust, also in wireless sensor networks. The term trust management was first applied in the context of distributed access control in [2] and the approach presented here is based on this definition.

Traditional access control systems usually rely on Role-Based Access Control model [29], [30] which groups the access rights by the role name and limits the access to a resource to those users who are assigned to a particular role.

The first trust management application described in the literature was PolicyMaker [31] which defined a special assertion language capable of expressing policy statements, which were locally trusted, and credentials that had to be signed using a private key. The next generation of trust management languages were KeyNote [32], which was an enhanced version of PolicyMaker, SPKI/SDSI [33] and a few other languages [34]. All these languages allowed assigning privileges to entities and used credentials to delegate permissions from its issuer to its subject. What was missing in those languages was the possibility of delegation based on attributes of the entities and not on their identity.

Responding to this need, a family of Role-based Trust management languages has been introduced in [35]–[38], and practical application using the *RT* language to control access to virtual machines was presented in [39]. These languages have a well-defined syntax and semantics, which made them easy to extend in order to apply them to different needs. One of the extensions is the use of time validity constraints of the credentials, which made the languages of the RT family more realistic, because in the real world permissions are usually given just for a limited period of time. Time-dependant credentials were introduced in [40] but only for $RT_0$ language. Because $RT^T$ language is more complex, powerful and allows to express security policies more suited to real needs, we decided to develop extensions to this specific language, which has not been done before. The complex time-dependant inference system with necessary proofs was introduced in [41].

## 3. Role-Based Trust Management

Role-Based Access Control (RBAC) model [29], [30] is the most flexible type of access control policy. It uses user role to control which users have access to particular resources. Access rights are grouped by the role name and access to resources is restricted to the users who are assigned to appropriate roles. The meaning of roles in RT captures the notion of groups of users in many systems and has been borrowed from RBAC approach. This type of access control works well in a large-scale centralized system and is often used in enterprise environments. Quite different challenges arise in decentralized and open systems where the identity of users is not known in advance and the set of users can change. It is also different in a wireless sensor network where sets of sensors can change rapidly. The identity of a user itself does not help in making decisions about their rights. What is needed to make such decisions is information about the privileges assigned to the user by other authorities, as well as trust information about the authority itself.

The term of *trust management* was introduced in 1996 by Blaze et al. in [2] who defined it as a unified approach to specify and interpret security policies, credentials and trust relationships. In a trust management system an entity's privilege is based on its attributes instead of its identity. An entity's attributes are demonstrated through digitally signed credentials issued by multiple principals. A *credential* is an attestation of qualification, competence or authority issued to an individual by a third party. Examples of credentials in real life include identification documents, driver's licenses, membership cards, keys, etc. A credential in a computer system can be a digitally signed document. Such a concept of trust management has evolved since that time to a much broader context of assessing the reliability and developing trustworthiness for other systems and individuals [42]. In this paper, however, we will use the term trust management only in a meaning restricted to the field of access control.

The potential and flexibility of trust management approach stems from the possibility of *delegation*: a principal may transfer limited authority over a resource to other principals. Such a delegation is implemented by means of an appropriate credential. This way, a set of credentials defines the access control strategy and allows deciding on who is authorized to access a resource, and who is not. The concept of delegation can also be used in a WSN, especially in routing structures that is why we will try to show how the permissions can be delegated from one sensor to another.

RT languages combine trust management and RBAC features. To define a trust management system, a language is needed for describing entities (principals and requesters), credentials and roles which the entities play in the system.

The core language of RT family is $RT_0$, described in detail in [37]. It allows describing localized authorities for roles, role hierarchies, delegation of authority over roles and role intersections. All the subsequent languages add new features to $RT_0$, they are progressively increasing in expressive power and complexity. $RT_1$ introduces parameterized roles, which can represent relationships between entities. $RT_2$ extends $RT_1$ with logical objects, which can be used to represent permissions given to entities with respect to a group of logically related objects (resources). These extensions can help in keeping the notation concise, but do not increase the expressive power of the language, because each combination of parameters in $RT_1$ and each permission to a logical object in $RT_2$ can be defined alternatively as a set of separate roles in $RT_0$. The most powerful language in the family is $RT^T$, as it provides useful capabilities not found in any other languages: manifold roles to achieve both agreement of multiple principals from one set and from disjoint sets and role-product operators, which can express threshold and separation of duties policies. Similar to a role which defines a set of principals a manifold role defines a set of principal sets, each of which is a set of principals whose cooperation satisfies the manifold role. A threshold policy requires a specified minimum number of entities to agree on some fact, i.e., it requires agreement among $k$ out of a set of entities that satisfy a specified condition, e.g., in a requirement that two different bank cashiers must authorise a transaction. Separation of duties policy requires a set of entities, each of which fulfils a specific role, to agree before access is granted.

$RT^D$ provides mechanism to describe delegation of rights and role activations, which can express selective use of capacities and delegation of these capacities. In many scenarios, an entity prefers not to use or delegate all his rights. For example, if an entity $D$ activates the role $A.r$ to use it in a session $B$, it can take the form of delegation credential, as a:

$$D \xrightarrow{D\ as\ A.r} B,$$

where $D\ as\ A.r$ is called a role activation. $B$ can further delegate this role activation to $C$ by issuing the credential,

$$B \xrightarrow{D\ as\ A.r} C.$$

An entity can issue multiple delegation credentials to another entity and also, several role activations can be delegated in one delegation credential.

The features of $RT^T$ and $RT^D$ can be combined together with the features of $RT_0$, $RT_1$ or $RT_2$. A more detailed treatment of RT family can be found in [36].

The languages have a precise syntax and semantics definition. A set-theoretic semantics, which defines the meaning of a set of credentials as a function from the set of roles into the power set of entities, has been defined for $RT_0$ [40], [37] and we defined relational semantics which apply also to other members of the family up to $RT^T$ in [43]. The logic-programming semantics of $RT_0$ credentials was first introduced in [36], a modified version of this semantics was shown in [40] and the semantics of all the other languages up to $RT^T$ was described in [44]. The member sets of roles can also be calculated in a more convenient way using an inference system which defines an operational semantics of $RT$ languages. An inference system consists of an initial set of formulae that are considered to be true, and a set of inference rules that can be used to derive new formulae from the known ones. The operational semantic was described in [45] and [40].

Table 1
Supported features of RT languages

| RT language | Supported features |
|---|---|
| $RT_0$ | – localized authorities for roles,<br>– role hierarchies,<br>– delegation of authority over roles,<br>– attribute based delegation of authority,<br>– role intersections. |
| $RT_1$ | features of $RT_0$ plus:<br>– parameterized roles,<br>– attribute-relationship based delegation,<br>– attribute-field constraints. |
| $RT_2$ | features of $RT_1$ plus:<br>– logical objects. |
| $RT^T$ | features of $RT_0$ plus:<br>– manifold roles,<br>– threshold policies,<br>– separation-of-duty policies. |
| $RT^D$ | features of $RT_0$ plus:<br>– selective use of role membership,<br>– dynamic credential delegation. |

A summary of the features supported by particular RT languages is shown in Table 1.

### 3.1. The Syntax of RT Family Languages

Basic elements of RT languages are entities, role names, roles and credentials. *Entities* represent principals that can define roles and issue credentials, and requesters that can make requests to access resources. An entity can, e.g., be a person or program identified by a user account in a computer system or a public key. *Role names* represent permissions that can be issued by entities to other entities, or groups of entities. *Roles* represent sets of entities that have particular permissions granted according to the access control policy. *Credentials* define roles by appointing a new member of the role or by delegating authority to the members of other roles.

There are six types of credentials in $RT^T$ (first four can also be used in $RT_0$, $RT_1$, and $RT_2$) which are interpreted in the following way:

- $A.r \leftarrow B$ – *simple membership*: entity $B$ is a member of role $A.r$.
- $A.r \leftarrow B.s$ – *simple inclusion*: role $A.r$ includes (all members of) role $B.s$. This is a delegation of authority over $r$ from $A$ to $B$, because $B$ may cause new entities to become members of the role $A.r$ by issuing credentials that define $B.s$. The hierarchy of roles is also possible.
- $A.r \leftarrow B.s.t$ – *linking inclusion*: role $A.r$ includes role $C.t$ for each $C$, which is a member of role $B.s$. This is a delegation of authority from $A$ to all the members of the role $B.s$. The expression $B.s.t$ is called a *linked role*.
- $A.r \leftarrow B.s \cap C.t$ – *intersection inclusion*: role $A.r$ includes all the entities who are members of both roles $B.s$ and $C.t$. This is a partial delegation from $A$ to $B$ and $C$. The expression $B.s \cap C.t$ is called an *intersection role*.
- $A.r \leftarrow B.s \odot C.t$ – role $A.r$ can be satisfied by a union set of one member of role $B.s$ and one member of role $C.t$. A set consisting of a single entity satisfying the intersection role $B.s \cap C.t$ is also valid.
- $A.r \leftarrow B.s \otimes C.t$ – role $A.r$ includes one member of role $B.s$ and one member of role $C.t$, but those members of roles have to be different entities.

### *3.2. Inference System over RT Credentials*

*RT* credentials are used to define roles which are used to represent permissions. The semantics of a given set $\mathscr{P}$ of *RT* credentials defines for each role $A.r$ the set of entities which are members of this role. The member sets of roles can also be calculated in a more convenient way using an inference system, which defines an operational semantics of *RT* language. An inference system consists of an initial set of formulae that are considered to be true, and a set of inference rules that can be used to derive new formulae from the known ones.

Let $\mathscr{P}$ be a given set of *RT* credentials. The application of inference rules of the inference system will create new credentials, derived from credentials of the set $\mathscr{P}$. A derived credential $c$ will be denoted using a formula $\mathscr{P} \succ c$ which should be read: credential $c$ can be derived from a set of credentials $\mathscr{P}$.

*Definition 1:* The initial set of formulae of an inference system over a set $\mathscr{P}$ of *RT* credentials are all the formulae: $c \in \mathscr{P}$ for each credential $c$ in $\mathscr{P}$. The inference rules of the system are the following:

$$\frac{c \in \mathscr{P}}{\mathscr{P} \succ c} \qquad (W_1)$$

$$\frac{\mathscr{P} \succ A.r \leftarrow B.s \qquad \mathscr{P} \succ B.s \leftarrow X}{\mathscr{P} \succ A.r \leftarrow X} \qquad (W_2)$$

$$\frac{\mathscr{P} \succ A.r \leftarrow B.s.t \qquad \mathscr{P} \succ B.s \leftarrow C \qquad \mathscr{P} \succ C.t \leftarrow X}{\mathscr{P} \succ A.r \leftarrow X} \qquad (W_3)$$

$$\frac{\mathscr{P} \succ A.r \leftarrow B.s \cap C.t \qquad \mathscr{P} \succ B.s \leftarrow X \qquad \mathscr{P} \succ C.t \leftarrow X}{\mathscr{P} \succ A.r \leftarrow X} \qquad (W_4)$$

$$\frac{\mathscr{P} \succ A.r \leftarrow B.s \odot C.t \qquad \mathscr{P} \succ B.s \leftarrow X \qquad \mathscr{P} \succ C.t \leftarrow Y}{\mathscr{P} \succ A.r \leftarrow X \cup Y} \qquad (W_5)$$

$$\frac{\mathscr{P} \succ A.r \leftarrow B.s \otimes C.t \qquad \mathscr{P} \succ B.s \leftarrow X \qquad \mathscr{P} \succ C.t \leftarrow Y \qquad X \cap Y = \phi}{\mathscr{P} \succ A.r \leftarrow X \cup Y} \qquad (W_6)$$

There could be a number of inference systems defined over a given language. To be useful for practical purposes, an inference system must exhibit two properties. First, it should be sound, which means that the inference rules could derive only formulae that are valid with respect to the semantics of the language. Second, it should be complete, which means that each formula which is valid according to the semantics should be derivable in the system. Both properties have been shown in [45], proving that the inference system provides an alternative way of presenting the semantics of *RT* languages.

## 4. Time Validity in *RT*

Inference rules with time validity for $RT_0$ were originally introduced in a slightly different way in [40]. In this paper, we will show the extention of other languages, up to $RT^T$ (by putting time validity constraints into this language). In this case credentials are given to entities just for some fixed period of time. It is quite natural to assume that permissions are given just for a fixed period of time, not forever.

The ability to infer credentials with incomplete information is a significant advantage of Role-based Trust management in distributed systems. However, practical applications are limited by the fact that in real life permissions can rarely be given forever. The need to revoke a credential may not be frequent, but when it occurs, it is crucial. Unfortunately, revocation of credentials is not a simple extention to the method – the system becomes non-monotonous. In this case access rights cannot be correctly inferred without complete information about credentials or at least knowledge which credentials have been explicitly revoked, and which should be invalidated as inferred from the revoked ones. Effectively this ruins the system's scalability.

A complete solution of the credential revocation problem is beyond the scope of this paper. However, it can be partially addressed by limiting the validity of credentials to fixed periods of time. As will be shown, this does not affect the system's ability to work with incomplete information and limits the potential impact of credentials that would otherwise be revoked. An additional effect is the ability to automatically identify outdated credentials, avoiding the problem of unlimited growth of the credential database.

The restricted validity of credentials can also be used to create a system enabling certificate revocation with an arbitrary, but non-zero reaction time. Credentials valid for long periods of time would not be used directly in this case – instead they would be used to create periodically (or on request) new credentials with short validity periods. Revocation of a credential would then be a local action, no more short-term credentials would be created and the revocation would be guaranteed to be effective as soon as the last short-term credential becomes invalid.

Time dependent credentials take the form: $c$ **in** $v$, meaning "the credential $c$ is available during the time $v$". Finite sets of time dependent credentials are denoted by $\mathscr{CP}$ and the new language is denoted as $RT_{+}^{T}$ (as an extention of the most powerful $RT^{T}$ language) To make notation lighter we write $c$ to denote "$c$ **in** $(-\infty, +\infty)$". This type of time constraints can satisfy the need of negation in non-monotonic systems.

Time validity can be denoted as follows: $[\tau_1,\tau_2]$; $[\tau_1,\tau_2)$; $(\tau_1,\tau_2]$; $(\tau_1,\tau_2)$; $(-\infty,\tau]$; $(-\infty,\tau)$; $[\tau,+\infty)$; $(\tau,+\infty)$; $(-\infty,+\infty)$; $v_1 \cup v_2$; $v_1 \cap v_2$; $v_1 \backslash v_2$ and $v_1$, $v_2$ of any form in this list, with $\tau$ ranging over time constants.

Time dependant credentials in wireless sensor networks can be used in a form of credential templates. Credential templates know the precise time validity of credentials and specific credentials know about narrowed period of time (for example one day). When sensor need to use a credential, it does not have to ask each time (what consume some resources) about the validity of credentials.

### 4.1. Inference System over $RT_{+}^{T}$ Credentials

Now, we can adapt inference system over $RT$ credentials to take time validity into account. Let $\mathscr{CP}$ be a given set of $RT_{+}^{T}$ credentials. The application of inference rules of the inference system will create new credentials, derived from credentials of the set $\mathscr{CP}$. A derived credential $c$ valid in time $\tau$ will be denoted using a formula $\mathscr{CP} \succ_{\tau} c$, which should be read: credential $c$ can be derived from a set of credentials $\mathscr{CP}$ during the time $\tau$.

*Definition 2:* from [46] The initial set of formulae of an inference system over a set $\mathscr{CP}$ of $RT_{+}^{T}$ credentials are all in the form: $c$ **in** $v \in \mathscr{CP}$ for each credential $c$ valid in time $v$ in $\mathscr{CP}$. The inference rules of the system are the following:

$$\frac{c \textbf{ in } v \in \mathscr{CP} \qquad \tau \in v}{\mathscr{CP} \succ_{\tau} c} \qquad (CW_1)$$

$$\frac{\mathscr{CP} \succ_{\tau} A.r \leftarrow B.s \qquad \mathscr{CP} \succ_{\tau} B.s \leftarrow X}{\mathscr{CP} \succ_{\tau} A.r \leftarrow X} \qquad (CW_2)$$

$$\frac{\mathscr{CP} \succ_{\tau} A.r \leftarrow B.s.t \qquad \mathscr{CP} \succ_{\tau} B.s \leftarrow C \qquad \mathscr{CP} \succ_{\tau} C.t \leftarrow X}{\mathscr{CP} \succ_{\tau} A.r \leftarrow X} \qquad (CW_3)$$

$$\frac{\mathscr{CP} \succ_{\tau} A.r \leftarrow B.s \cap C.t \qquad \mathscr{CP} \succ_{\tau} B.s \leftarrow X \qquad \mathscr{CP} \succ_{\tau} C.t \leftarrow X}{\mathscr{CP} \succ_{\tau} A.r \leftarrow X} \qquad (CW_4)$$

$$\frac{\mathscr{CP} \succ_{\tau} A.r \leftarrow B.s \odot C.t \qquad \mathscr{CP} \succ_{\tau} B.s \leftarrow X \qquad \mathscr{CP} \succ_{\tau} C.t \leftarrow Y}{\mathscr{CP} \succ_{\tau} A.r \leftarrow X \cup Y} \qquad (CW_5)$$

$$\frac{\mathscr{CP} \succ_{\tau} A.r \leftarrow B.s \otimes C.t \qquad \mathscr{CP} \succ_{\tau} B.s \leftarrow X \qquad \mathscr{CP} \succ_{\tau} C.t \leftarrow Y \qquad X \cap Y = \phi}{\mathscr{CP} \succ_{\tau} A.r \leftarrow X \cup Y} \qquad (CW_6)$$

### 4.2. Inferring Time Validity of Credentials

This inference system evaluates maximal time validity when it is possible to derive the credential $c$ from $\mathscr{CP}$. It enhances formula $\mathscr{CP} \succ_{\tau} c$ to $\mathscr{CP} \succ\!\succ_{v} c$, specifying that at any time $\tau \in v$ in which $\mathscr{CP}$ has a semantics, it is possible to infer the credential $c$ from $\mathscr{CP}$. To make notation lighter we write $\succ\!\succ$ to denote $\succ\!\succ_{(-\infty,+\infty)}$. The inference rules of the system are the following:

$$\frac{c \textbf{ in } v \in \mathscr{CP}}{\mathscr{CP} \succ\!\succ_{v} c} \qquad (CWP_1)$$

$$\frac{\mathscr{CP} \succ\!\succ_{v_1} A.r \leftarrow B.s \qquad \mathscr{CP} \succ\!\succ_{v_2} B.s \leftarrow X}{\mathscr{CP} \succ\!\succ_{v_1 \cap v_2} A.r \leftarrow X} \qquad (CWP_2)$$

$$\frac{\mathscr{CP} \succ\!\succ_{v_1} A.r \leftarrow B.s.t \qquad \mathscr{CP} \succ\!\succ_{v_2} B.s \leftarrow C \qquad \mathscr{CP} \succ\!\succ_{v_3} C.t \leftarrow X}{\mathscr{CP} \succ\!\succ_{v_1 \cap v_2 \cap v_3} A.r \leftarrow X} \qquad (CWP_3)$$

$$\frac{\mathscr{CP} \succ\!\succ_{v_1} A.r \leftarrow B.s \cap C.t \qquad \mathscr{CP} \succ\!\succ_{v_2} B.s \leftarrow X \qquad \mathscr{CP} \succ\!\succ_{v_3} C.t \leftarrow X}{\mathscr{CP} \succ\!\succ_{v_1 \cap v_2 \cap v_3} A.r \leftarrow X} \qquad (CWP_4)$$

$$\frac{\mathscr{CP} \succ\!\succ_{v_1} A.r \leftarrow B.s \odot C.t \qquad \mathscr{CP} \succ\!\succ_{v_2} B.s \leftarrow X \qquad \mathscr{CP} \succ\!\succ_{v_3} C.t \leftarrow Y}{\mathscr{CP} \succ\!\succ_{v_1 \cap v_2 \cap v_3} A.r \leftarrow X \cup Y} \qquad (CWP_5)$$

$$\frac{\mathscr{CP} \succ\!\succ_{v_1} A.r \leftarrow B.s \otimes C.t \qquad \mathscr{CP} \succ\!\succ_{v_2} B.s \leftarrow X \qquad \mathscr{CP} \succ\!\succ_{v_3} C.t \leftarrow Y \qquad X \cap Y = \phi}{\mathscr{CP} \succ\!\succ_{v_1 \cap v_2 \cap v_3} A.r \leftarrow X \cup Y} \qquad (CWP_6)$$

$$\frac{\mathscr{CP} \succ\!\succ_{v_1} c \qquad \mathscr{CP} \succ\!\succ_{v_2} c}{\mathscr{CP} \succ\!\succ_{v_1 \cup v_2} c} \qquad (CWP_7)$$

## 5. RT in Wireless Sensor Networks

Sensors have a limited source of power and it is hard to replace or recharge, for example, sensors in the battle field or sensors in a large sea or forest. That is why it is so important to save these resources. On the other hand, in some cases the security of sensor network is crucial and we can use some resources to protect WSN.

Hierarchical routing, which is proposed to prolong the lifetime of WSNs, is one of the areas where it is possible to use $RT$ languages. Another important area may be delegation of permissions in mobile networks, where $RT^D$ language can be useful.

### 5.1. Hierarchical Routing

The hierarchical routing protocols classify sensor nodes according to their functionalities. The main purpose of such a division is to reduce the energy consumption. It is easy to delegate the privileges between nodes which are similar. The network is divided into groups (or clusters) with a leader sensor (or cluster node). The leader coordinates the activities within the group and communicates with sensors outside the own group. The different schemes for hierarchical routing mainly differ in how the leader is selected and how the sensors behave in the inter and intra-group domain.

Hierarchical routing is one of the fields where a delegation of permission from the Role-based Trust management family can be applied. For example, in a one-way communication scenario, the group leader can broadcast the message that his resources are running out, so he would like to delegate its permissions to another sensor. It can be assumed that he would do this on the condition that the potential sensor has the proper credentials. What is needed to make such decisions is information about the privileges assigned to the potential sensor by other authorities, as well as trust information about the authority itself. If the above conditions are met, the leader can delegate its permissions (and even role activation) to perform its role to another sensor which is authorized to do that.

### 5.2. Permissions Delegation in a Mobile Networks

Mobile sensor networks are incredibly valuable, especially in situations where traditional arrangement mechanisms fail, or are not suitable. Also, in some application scenarios such as ocean monitoring, sensors move with the ocean currents. The coverage of a mobile sensor network depends not only on the initial network congurations, but also on the mobility behavior of the sensors.

The locations covered by sensors change over time, they can regroup in order to cover the range of the new area. In this case, it is good idea to use one of the RT family languages, $RT^D$, which provides mechanisms to describe delegation of role activations and selective use of role membership. Sensors changing its location can delegate their permissions to other sensors. Moving from one place to another, they can change their roles and activate new ones. It is also possible to delegate some of their rights to sensors towards which they change their position. They may give up their role in favor of other sensors. They can interact with some sensors at specified periods of time, and with others in other periods of time – depending on time validity of their permissions.

## 6. Conclusions

Trust and trust management is an important issue in distributed wireless sensor networks. That is why it is more and more often the subject of research scientists. Because it can increase the security of the network, they can be used more widely. The concept of trust and trust management in wireless sensor network is defined in a different way, because it is used in a different cases. As it was shown above, the languages from the family of Role-based Trust management can be applied to WSN. Because of the character of this kind of network it is not suitable to use it in a small WSN where just simple low-resource wireless sensors are used, but in networks where the security is crucial. It is also possible to use RT in a wireless sensor and actuator networks.

## Acknowledgements

I would like to thank Krzysztof Lasota for his contribution to this work. I would also like to thank an anonymous reviewer for very valuable and detailed comments.

## References

[1] S. P. Marsh, "Formalising Trust as a Computational Concept", Ph.D. thesis, Dept. of Computing Science and Mathematics, University of Stirling.

[2] M. Blaze, J. Feigenbaum, and J. Lacy, "Decentralized trust management", in *Proc. 17th IEEE Symp. Secur. Priv.*, Oakland, CA, USA, 1996, pp. 164–173.

[3] A. Perrig, R. Szewczyk, V. Wen, D. Culler, and J. D. Tygar, "SPINS: Security protocols for sensor networks", in *Proc. 7th Int. Conf. Mob. Comput. Netw. MobiCom 2001*, Rome, Italy, 2001, pp. 189–199.

[4] "Cambridge Advanced Learner's Dictionary" [Online]. Available: http://dictionary.cambridge.org/

[5] S. Ganeriwal and M. B. Srivastava, "Reputation-based framework for high integrity sensor networks", in *Proc. 2nd ACM Worksh. Secur. Ad Hoc Sensor Netw.*, Washington, DC, USA, 2004, pp. 66–77.

[6] A. Boukerche, X. Li, and K. EL-Khatib, "Trust-based security for wireless ad hoc and sensor networks", *Comp. Commun.*, vol. 30, pp. 2413–2427, 2007.

[7] Z. Yao, D. Kim, and Y. Doh, "PLUS: Parameterized and localized trust management scheme for sensor networks security", in *Proc. 3rd IEEE Int. Conf. Mob. Ad-Hoc and Sensor Syst. MASS-2006*, Vancouver, Canada, 2006, pp. 437–446.

[8] M. Gupta, P. Judge, and M. Ammar, "A reputation system for peer-to-peer networks", in *Proc. 13th Int. Worksh. Netw. Operating Sys. Support Digit. Audio Video NOSSDAV 2003*, Monterey, CA, USA, 2003, pp. 144–15.

[9] L. Xiong and L. Liu, "Peer trust: Supporting reputation-based trust for peer-to-peer electronic communities", *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 7, pp. 843–857, 2004.

[10] S. Buchegger and J. L. Boudec, "Performance analysis of the CONFIDANT protocol", in *Proc. 3rd ACM Int. Symp. Mob. Ad Hoc Netw. Comput. MobiHoc 2002*, Lausanne, Switzerland, 2002.

[11] S. Buchegger and J. Y. L. Boudec, "A robust reputation system for peer-to-peer and mobile ad-hoc networks", in *Proc. 2nd Worksh. Econom. Peer-to-Peer Sys. P2PEcon 2004*, Cambridge, MA, USA, 2004.

[12] P. Michiardi and R. Molva, "CORE: a collaborative reputation mechanism to enforce node cooperation in mobile ad hoc networks", in *Advanced Communications and Multimedia Security*, B. Jerman-Blazic and T. Klobucar, Eds. Kluwer 2002, pp. 107–121.

[13] Y. Rebahi, V. E. Mujica-V, and D. Sisalem, "A reputation-based trust mechanism for ad-hoc networks", in *Proc. 10th IEEE Symp. Comp. Commun. ISCC 2005*, Murcia, Spain, 2005, pp. 37–42.

[14] Y. L. Sun, W. Yu, Z. Han, and K. J. R. Liu, "Information theoretic framework of trust modeling and evaluation for ad hoc networks", *IEEE J. Selec. Areas in Commun.*, vol. 24, no. 2, pp. 305–317, 2006.

[15] G. Theodorakopoulos and J. S. Baras, "On trust models and trust evaluation metrics for ad hoc networks", *IEEE J. Selec. Areas in Commun.*, vol. 24, no. 2, pp. 318–328, 2006.

[16] Z. Yan, P. Zhang, and T. Virtanen, "Trust evaluation based security solutions in ad hoc networks", in *Proc. 7th Nordic Worksh. Secur. IT Syst.*, Gjovik, Norway, 2003.

[17] H. Chen, H. Wu, X. Zhou, and C. Gao, "Reputation-based trust in wireless sensor netwoks", in *Proc. Int. Conf. Multim. Ubiq. Engin. MUE 2007*, Seoul, Korea, 2007.

[18] K. Daniluk and E. Niewiadomska-Szynkiewicz, "A survey of energy efficient security architectures and protocols for wireless sensor networks", *J. Telecom. Inform. Technol.*, no. 3, pp. 64–72, 2012.

[19] K. Lasota, E. Niewiadomska-Szynkiewicz, and A. Kozakiewicz, "Adaptacja rozwiązań honeypot dla sieci czujników", in *Proc. Konferencja Sieci Komputerowe SK-12*, Szczyrk, Poland, 2012, *Studia Informatica*, vol. 33, no. 3A, pp. 139–148 (in Polish).

[20] K. Lasota, E. Niewiadomska-Szynkiewicz, and A. Kozakiewicz, "Mobilny honeypot dla sieci sensorycznych", *Przegląd Telekomunikacyjny*, no. 8–9, pp. 699–704, 2012 (in Polish).

[21] M. Momani, S. Challa, and K. Aboura, "Modelling trust in wireless sensor networks from the sensor reliability prospective", in *Innovative Algorithms and Techniques in Automation, Industrial Electronics and Telecomm.*, T. Sobh *et al*., Eds. Heidelberg: Springer, 2007, pp. 179–189, pp. 317–321.

[22] R. A. Shaikh, H. Jameel, S. Lee, S. Rajput, and Y. J. Song, "Trust management problem in distributed wireless sensor networks", in *Proc. 12th IEEE Conf. Embedd. Real-Time Comput. Syst. Appl. RTCSA 2006*, Sydney, Australia, 2006, pp. 411–414.

[23] E. Aivaloglou and S. Gritzalis, "Hybrid trust and reputation management for sensor networks", *Wirel. Netw.*, vol. 16, no. 5, pp. 1493–1510, 2010.

[24] Z. Yao, D. Kim, I. Lee, K. Kim, and J. Jang, "A security framework with trust management for sensor networks", in *Proc. Worksh. of the 1st Int. Conf. Secur. Priv. Emerg. Areas in Commun. Netw. SecureComm 2005*, Athens, Greece, 2005, pp. 190–198.

[25] Y. Zhiying, K. Daeyoung, L. Insun, K. Kiyoung, and J. Jongsoo, "A security framework with trust management for sensor networks", in *Proc. 1st Int. Conf. Secur. Priv. Emerg. Areas in Commun. Netw. SecureComm 2005*, Athens, Greece, 2005.

[26] T. A. Zia, "Reputation-based Trust Managmenet in Wireless Sensor Networks", in *Proc. Int. Conf. Intell. Sensors, Sensor Netw. Inform. Proces. ISSNIP 2008*, Sydney, Australia, pp. 163–166.

[27] M. Momani and S. Challa, "Trust management in wireless sensor networks", in *Proc. 5th ACM Conf. Embedded Netw. Sensor Syst.*, Sydney, Australia, 2007.

[28] H. Chen, H. Wu, X Zhou, and C. Gao, "Agent-based trust model in wireless sensor networks", in *Proc. 8th ACIS Int. Conf. Softw. Engin. Artif. Intell., Netw. Parallel/Distrib. Comput.*, Qingdao, China, 2007.

[29] D. F. Ferraiolo, R. S. Sandhu, S. I. Gavrila, D. R. Kuhn, and R. Chandramouli, "Proposed NIST standard for role-based access control", *ACM Trans. Inf. Syst. Secur.*, vol. 3, pp. 224–274, 2001.

[30] R. S. Sandhu, E. J. Coyne, H. L. Feinstein, nad C. E. Youman, "Role-based access control models", *IEEE Computer*, vol. 2, pp. 38–47, 1996.

[31] M. Blaze, J. Feigenbaum and M. Strauss, "Compliance checking in the policymaker trust management system", in *Proc. 2nd Int. Conf. Finan. Cryptogr.*, London, UK, 1998, pp. 254–274.

[32] M. Blaze, J. Feigenbaum, J. Ioannidis, and A. Keromytis, "The role of trust management in distributed systems security", in *LNCS Secure Internet Programming: Security Issues for Mobile and Distributed Objects*, J. Vitek and C. D. Jensen, Eds. Springer, 1999, pp. 185–210.

[33] D. Clarke, J.-E. Elien, C. Ellison, M. Fredette, A. Morcos, and R. L. Rivest, "Certificate chain discovery in SPKI/SDSI", *J. Comp. Secur.*, vol. 9, pp. 285–322, 2001.

[34] P. Chapin, C. Skalka, and X. S. Wang, "Authorization in trust management: features and foundations", *ACM Comp. Surv.*, vol. 3, pp. 1–48, 2008.

[35] N. Li and J. Mitchell, "RT: A Role-Based Trust-Management Framework", in *Proc. 3rd DARPA Inform. Survivabil. Conf. Expos.*, Washington, DC, USA, IEEE Computer Society Press, 2003, pp. 201–212.

[36] N. Li, J. Mitchell, and W. Winsborough, "Design of a role-based trust-management framework", in *Proc. IEEE Symp. Secur. Priv.*, Oakland, CA, USA, 2002, pp. 114–130.

[37] N. Li, W. Winsborough, and J. Mitchell, "Distributed credential chain discoveryin trust management", *J. Comp. Secur.*, vol. 1, pp. 35–86, 2003.

[38] M. R. Czenko, S. Etalle, D. Li, and W. H. Winsborough, "An introduction to the role based trust management framework RT", Tech. Rep. TR-CTIT-07-34, Centre for Telematics and Information Technology University of Twente, Enschede, 2007.

[39] K. Lasota and A. Kozakiewicz, "Model of user access control to virtual machines based on RT – family trust management language with temporal validity constrains – practical application", *J. Telecom. Inform. Technol.*, no. 3, pp. 13–21, 2012.

[40] D. Gorla, M. Hennessy, and V. Sassone, "Inferring dynamic credentials for role-based trust management", in *Proc. 8th ACM SIGPLAN Conf. Princip. Pract. Declarat. Program. PPDP 06*, Venice, Italy, 2006, pp. 213–224.

[41] A. Felkner and A. Kozakiewicz, "$RT_{+}^{T}$ – time validity constraints in $RT^{T}$ language", *J. Telecom. Inform. Technol.*, no. 2, pp. 74–82, 2012.

[42] W. M. Grudzewski, I. K. Hejduk, A. Sankowska, and M. Wańtuchowicz, *Trust Management in Virtual Work Environments: A Human Factors Perspective*. CRC Press, 2008.

[43] A. Felkner and K. Sacha, "The semantics of role-based trust management languages", in *Advances in Software Engineering Techniques*, T. Szmuc, M. Szpyrka, J. Zendulka, Eds., *LNCS 7054*. Heidelberg: Springer, 2012, pp. 179–189.

[44] A. Felkner and A. Kozakiewicz, "Kontrola dostępu w rozproszonych systemach – trzy semantyki języka $RT^{T}$", in *Proc. II Konferencja i3: internet – infrastruktury – innowacje*, Wrocław, Poland, 2010 (in Polish).

[45] A. Felkner and K. Sacha, "Deriving $RT^{T}$ credentials for role-based trust management", *e-Inform. Software Engin. J.*, vol. 4, no. 1, pp. 9–19, 2010.

[46] A. Felkner and A. Kozakiewicz, "Time validity in role-based trust management inference system", in *Proc. Int. Worksh. Sec. Trust Comput., Data Manag., and Appl. IWCS-11*, Loutraki, Greece, 2011, *Communications in Computer and Information Science*, Springer, 2011, vol. 187, pp. 7–15.

**Anna Felkner** graduated from the Faculty of Computer Science of Białystok University of Technology (M.Sc., 2004) and the Faculty of Electronics and Information Technology of Warsaw University of Technology (Ph.D., 2010). At present she is an Assistant Professor at Network and Information Security Methods Team in NASK Research Division. Main scientific interests concern the security of information systems, especially access control and trust management.

E-mail: anna.felkner@nask.pl
Research and Academic Computer Network (NASK)
Wąwozowa st 18
02-796 Warsaw, Poland

Information

# Educating on Wireless Solutions for Environmental Monitoring

Sandro Radicella[a], Ryszard Strużak[b], and Marco Zennaro[a]

[a] *International Center for Theoretical Physics (ICTP), Trieste, Italy*
[b] *National Institute of Telecommunications (NIT), Wrocław, Poland*

**Abstract—Environmental monitoring is of growing interest around the globe and there is a clear need for better education in this area. However, the progress is incredibly fast here and wireless sensor networking is a relatively new area of research, so no standard curricula are available yet. In 2012, the International School "Sustainable Wireless ICT Solutions for Environmental Monitoring" was organized by the International Center for Theoretical Physics (ICTP) in collaboration with a few other entities. The school aimed at exposing young scientists from around the world to the newest wireless solutions for environmental monitoring. The school was followed by an awareness conference, which aimed at identifying issues that would benefit from a wider scientific collaboration. This paper begins with general information about the ICTP educational activities related to wireless technologies and applications. Then it focuses on the program of the school and conference. It shows what topics were included (and what were disregard) and what approaches were applied in order to achieve the best results with the (very) limited resources available. Details are omitted here as the complete materials of the both events are available at the Internet (http://wireless.ictp.it) under the Creative Commons.**

***Keywords—DIGI, Fundación Escuela Latinoamericana de Redes, EsLaRed, International Center for Theoretical Physics, National Institute of Telecommunications, Network Startup Resource Center, NIT, wireless sensor networks, WSN.***

## 1. Introduction

Wireless sensor networks are attracting more and more attention. Collection of empirical data has been always enabling advances in science and contributing to improvements of the quality and safety of life. However, until recently, especially in environmental applications, it is based mainly on expensive equipment and wired infrastructures necessary to collect the data in one point where they could be further processed. The data collection was a costly and difficult task, limited to a relatively small number of fixed, sparsely distributed locations, and maintained by organizations with large budgets. As a result, the data gathered are often incomplete, especially when concerning developing countries and remote areas. Wireless sensor networks (WSN) could change that radically. It is a low-cost and low-power technology that does not require any pre-existing infrastructure and can be deployed in most the remote regions. The vast range of sensors that can be connected to the devices makes them flexible for many different applications, such as air-quality, water-quality and soil-moisture monitoring.

This paper describes the 2012 school on "Sustainable Wireless ICT Solutions for Environmental Monitoring" and "International Awareness Conference on Sustainable Wireless Solutions for Environmental Monitoring" where the participants could discus their cases with the participation of invited experts. These events were organized at the International Centre for Theoretical Physics (ICTP) in Trieste, Italy, from February 6 to 24. That was a common activity of ICTP and the International Telecommunication Union (ITU) Development Bureau, with collaboration of the International Telecommunications Union (ITU), national research entities such as the National Institute of Telecommunications (Poland), Fundación Escuela Latinoamericana de Redes – EsLaRed (Venezuela), and the Network Startup Resource Center – NSRC (Oregon, USA). This short report offers some details on the school and on the case studies discussed there and starts with a brief information about the ICTP and its activities in the field of radio.

## 2. ICTP

The mission of Abdus Salam International Centre for Theoretical Physics (ICTP) is to foster advanced studies and research in developing countries. Founded by Abdus Salam (Nobel Laureate in Physics), ICTP operates under a tripartite agreement among two United Nations Agencies, UNESCO and IAEA, and the Italian government. While the name of the Centre reflects its beginnings, its activities today encompass most areas of theoretical and applied sciences, including information and communications technologies (ICT).

ICTP embraces a large community of scientists worldwide. Since its creation, the Centre has received about 120,000 scientists, half of whom have come from the developing world. Visitors have represented some 180 nations and 40 international organizations. In recent years, more than 6,000 scientists visit ICTP annually to participate in its research and training activities, and to conduct their own research. The ICTP team is convinced that knowledge sharing, open access, and know-how transfer are critical for sustainable development. Appropriate training is fundamental and the most cost-effective long-term investment and ICTP has been playing a leading role in the field. ICTP con-

***Fig. 1.*** The ICTP Miramare campus (courtesy ICTP).

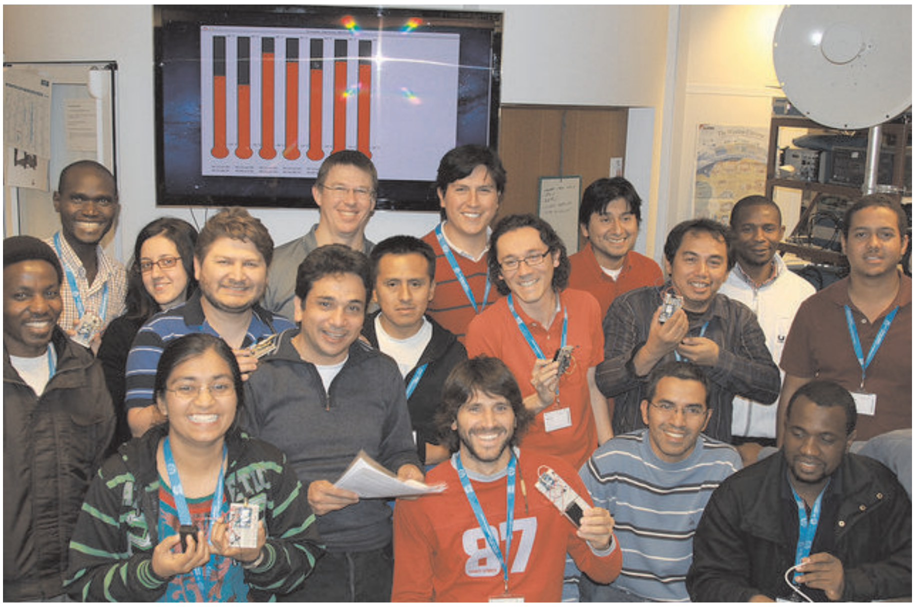


***Fig. 2.*** Group of the participants in the 2012 WSN School at the ICTP Gulliemo Marconi Laboratory. Some of them with wireless sensor nodes in hands (Photo: M. Zennaro).

siders the "training trainers" at academic level as its major mission.

Since 1996, it has established extensive *in-situ* training programs on wireless communications technologies to facilitate Internet access to unconnected academic and other institutions. Its Telecommunications/ICT for Development Laboratory (the former Aeronomy and RadioPropagation Laboratory) has held some 40 training activities, attracting more than 1600 participants from Africa, Asia, Oceania, Europe, and the Americas, as well as from international organizations.

Starting from 1998, the ICTP has offered an annual school, focused on wireless networking for developing countries for local-area networks in academic campuses. Topics include theoretical and practical training on low-cost radio techniques, planning, installation, and maintenance of short- and medium-distance point-to-point digital radio links. The schools have used modern technologies and teaching techniques. Emphasis is put on hands-on laboratory sessions and in-the-field practical exercises. In addition to activities held at ICTP headquarters in Italy, training has been also given in developing countries. From 2000 to 2011,

*Fig. 3.* Exercises with programming of wireless sensor nodes: (a) hard work and (b) joy of success (Photo: M. Zennaro).

Table 1
Topics included in the program

| General topic lectures (10% of total time) | Specific topics & case studies (40% of total time) | Hands-on exercises (50%of total time) |
|---|---|---|
| • dB Math | • Planning a radio link | • Antenna Building |
| • Telecommunications | • Intro to RadioMobile | • Antenna Testing |
| • Protocols | • Wireless network Protocols | • Cable testing |
| • IPv6 | • Introduction to WiFi | • WiFi links |
| • Cloud networks | • Intro to IEEE 802.22 | • WSN planning |
| • Bb Internet | • Configuration | • WSN programing |
| • Wireless Communities | • Outdoor installations | • WSN deploying |
| • Wireless Sensor Networks (WSN) Applications | • Off-grid powering | • AP and client configuration |
| | • Mesh Networking | • Link Budget planning |
| • RF Channel & Modulation | • Network monitoring Tools | • RadioMobile tool |
| • Transmission lines | • Wireless Mesh Networks | • Indoor links |
| • Antennas | • Low Power WiFi | • Indoor campus networks |
| • Radio propagation | • WT KIT | • Outdoor & Long Distance WiFi |
| • Radio Measurements | • Community networking | • Mesh Networks |
| • EM Co-existence | • Testbeds | • Network monitoring |
| • Wireless Security | • Simple and Advanced WSN | • Channel configuration |
| • Spectrum management | • XBee, ZigBee, iDigi | • Spectrum use |
| • Regulatory issues | • Gateways | • Outdoor Links |
| | • Building Environmental Sensor Network | • Doorbell Project |
| | | • Solar power |
| | | • Sensing events |
| | | • Sensing light |
| | | • Sensing temperature |
| | | • Sensing moisture |

the ICTP group has been involved in training activities on wireless for broadband connectivity and computer networking in Benin, Cameroon, Ghana, India, Indonesia, Kenya, Nigeria, Peru, Sudan, Venezuela, and Zimbabwe, to name a few countries.

Starting from 2007, the school's program included the topic of "Wireless Sensor Networking". The potential applications of this technology included water-quality monitoring, intelligent irrigation, and disaster warning, the topics extremely relevant to rural and remote areas. The Radio Science Bulletin has published selected presentations from the Previous ICTP School held in 2011 [1], and training issues related to wireless environmental monitoring have been discussed in [2] .

## 3. The 2012 Activities

An international team of some 10 lecturers presented at the school and led practical exercises. They came from Denmark, Germany, Italy, Spain, Switzerland, United States,

*Fig. 4.* Experimenting with wireless sensor networks (Photo: M. Zennaro).

Poland, South Africa, and Venezuela. Their names are listed at the ICTP homepage (http://www.ictp.it/).

The school consisted of lectures, individual programming of wireless sensor nodes, experimenting with wireless sensor networks in the laboratory and in the field, as well as collective discussions on specific topics. Table 1 lists major topics included in the School program. The experiments took about a half of the total time. The Company "DIGI" generously donated some of the materials for the workshop and provided with two lecturers. The students will take back home what they've learned to build a wide variety of environmental systems for agriculture, solar schoolhouses, water quality, radiation sensing, energy, emergency response and other purposes. The detailed list of lectures is available at the ICTP homepage, together with all the material presented there.

A total of 30 participants from more than 20 countries participated. The students were educators, scientists and engineers from Malawi, Nicaragua, India, Ecuador, Venezuela, West Gambia, Philippines, USA, South Africa, Tanzania, Jamaica, Columbia, Ukraine, Argentina and Albania. Many of them presented practical case studies from their countries. These covered the following wireless sensor networks in environmental and health applications in remote areas using radio:

- WSN in water quality monitoring (Jamaica)
- WSN at breeder reactor facilities (India),
- Environmental Monitoring (Italy),
- Sensing & control in bioreactors (S. Africa),
- Wireless Sensor Network (Italy),
- VSAT & WiLD networks in Amazon jungle (Peru),
- SMS for remote data transmission (USA),
- WSN at a University (Zimbabwe),
- SMS for health applications (USA),
- Smart metering in (Colombia),
- Nigerian Experience (Nigeria),
- Integrating WSN with conventional system (Argentina),
- Monitoring nuclear reactor (Ukraine).

## 4. Conclusions

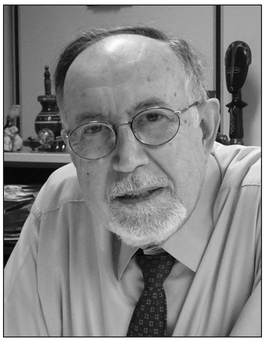
Environmental monitoring could enable important advances in science, agriculture, forestry and disaster mitigation. However, only limited research is currently carried out in many countries using modern technologies. The School has exposed participants to new wireless technologies, amenable also for Developing Regions. These technologies can work with low power and are not dependent on any existing electrical network. These are low- cost sustainable solutions. Once the environmental data are collected in remote areas (volcanoes, mountains, lakes, etc.), they have to be transmitted to a central database for processing and further analysis. The cost of providing the wiring infrastructure to support such transmissions is simply out of the question. In the last couple of years we have witnessed an incredible evolution of wireless technologies that can provide a viable solution. The 2012 School presented the use of various low-cost and sustainable wireless technologies. A compound of these technologies is probably the best solution not only to build environmental sensor networks but also to allow scientists to communicate. Training at the level of Academic and Research institutions on the subject of the School is essential to develop sustainable human capacity. In the Awareness Conference, workshop participants and Telecom Administrations, Regulators, Operators, Researchers and international experts presented their cases and discussed advanced wireless solutions for environmental monitoring. The 2012 events have been well received by all participants. On that basis, the organizers have planned similar activities for 2013.

The full documentation of the 2012 events (available under the Creative Commons license) can be found at: http://wireless.ictp.it.

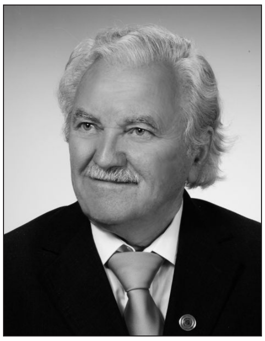
## References

[1] Radio Science Bulletin, International Union of Radio Science (URSI), Special issue on Wireless Networks, no. 339, Dec. 2011, ISSN 1024-4530 [Online]. Available: http://www.ursi.org/files/RSBissues/RSB_339_2011_12.pdf

[2] M. Zennaro, A. Bagula, and M. Nkoloma," From Training to Projects: Wireless Sensor Networks in Africa", in *Proc. IEEE Global Human. Technol. Conf. GHTC 2012*, Seattle, Washington, USA, 2012.



**Sandro Radicella** is the Head of the Telecommunications/ICT for Development Laboratory (former Aeronomy and Radiopropagation Laboratory) of the Abdus Salam International Centre for Theoretical Physics (ICTP). He has published more than 1400 papers in the fields of aeronomy and radiocommunications. The most important achievement of his recent scientific production is the development of models of vertical distribution of electronic density in the ionosphere in collaboration with colleagues from the ICTP and of the University of Graz, Austria. One of these models is being used by the European Space Agency in areas related to the use of EGNOS and the new GALILEO satellite system. He has organized and directed since 1989 a series of Colleges, Schools and Workshops, for participants mainly from developing countries, in the fields of Ionospheric Physics, Radiocommunications and Information and Communication Technology. He has been awarded in 2001 with the Doctor Honoris Causa degree from the University of Bucharest, Romania, and in 2005 with the Doctor of Science degree Honoris Causa from the Obafemi Awolowo University in Ile-Ife, Nigeria.
E-mail: rsandro@ictp.it
Telecommunications/ICT Development Laboratory
Abdus Salam International Centre for Theoretical Physics
Strada Costiera 11
34014 Trieste, Italy



**Ryszard Strużak** is a Professor at NIT, a Co-Director of ICTP Schools on Wireless Networking, a Life Fellow of IEEE. He authored/co-authored many papers as well as 10 patents in the areas of radiocommunications, spectrum management, and electromagnetic compatibility. He was a consultant to ITU, IUCAF, PWC, UNOCHA, WB, and to industry and government agencies in several countries including Switzerland, UK and USA. He served as the Head of Technical Department and Acting Assistant Director at CCIR-ITU, a Visiting Professor at Institut National Polytechnique de Toulouse, a Professor at the University of Information Technology and Management at Rzeszow, and at the Wrocław University of Technology, and the Editor-in-Chief and Chair of Editorial Board, Global Communications. He co-founded and chaired the Electromagnetic Compatibility (EMC) Laboratories at NIT and the International Wrocław Symposium on EMC (now EMC Europe). He was elected to leading positions in international organizations: ITU, URSI, IEC and CISPR; among others he served as the Vice-Chairman of ITU Radio Regulations Board. He was elected a member of New York Academy of Science and an Academician of the International Telecommunication Academy. He was honored by the ITU Silver Medal, by the IEEE EMCS Special Symposium Recognition Award, by the Prize Paper Award at International EMC Symposia (twice), and by the highest national awards and decorations.
E-mail: struzak@gmail.com
National Institute of Telecommunications (NIT)
Swojczycka st 38
51-501 Wrocław, Poland

**Marco Zennaro** received his M.Sc. degree in Electronic Engineering from University of Trieste in Italy. He defended his PhD thesis on "Wireless Sensor Networks for Development: Potentials and Open Issues" at KTH-Royal Institute of Technology, Stockholm, Sweden under the supervision of Prof. Bjorn Pehrson. His research interest is in ICT4D, the use of ICT for Development. In particular, he is interested in using Wireless Sensor Networks in Developing Countries. He is one of the authors of "Wireless Networking in the Developing World", which has been translated in 6 languages, and of "Science Dissemination using Open Access", available in English and Spanish. He has been giving lectures on Wireless technologies in more than 20 different countries. He is the editor of wsnblog.com.
E-mail: mzennaro@ictp.it
Telecommunications/ICT Development Laboratory
Abdus Salam International Centre for Theoretical Physics
Strada Costiera 11
34014 Trieste, Italy

# Information for Authors

***Journal of Telecommunications and Information Technology*** **(JTIT)** is published quarterly. It comprises original contributions, dealing with a wide range of topics related to telecommunications and information technology. **All papers are subject to peer review**. Topics presented in the JTIT report primary and/or experimental research results, which advance the base of scientific and technological knowledge about telecommunications and information technology.

JTIT is dedicated to publishing research results which advance the level of current research or add to the understanding of problems related to modulation and signal design, wireless communications, optical communications and photonic systems, voice communications devices, image and signal processing, transmission systems, network architecture, coding and communication theory, as well as information technology.

Suitable research-related papers should hold the potential to advance the technological base of telecommunications and information technology. Tutorial and review papers are published only by invitation.

**Manuscript.** TEX and LATEX are preferable, standard Microsoft Word format (.doc) is acceptable. The author's JTIT LATEX style file is available:
http://www.nit.eu/for-authors

Papers published should contain up to 10 printed pages in LATEX author's style (Word processor one printed page corresponds approximately to 6000 characters).

The manuscript should include an abstract about 150–200 words long and the relevant keywords. The abstract should contain statement of the problem, assumptions and methodology, results and conclusion or discussion on the importance of the results. Abstracts must not include mathematical expressions or bibliographic references.

Keywords should not repeat the title of the manuscript. About four keywords or phrases in alphabetical order should be used, separated by commas.

The original files accompanied with pdf file should be submitted by e-mail: redakcja@itl.waw.pl

**Figures, tables and photographs.** Original figures should be submitted. Drawings in Corel Draw and PostScript formats are preferred. Figure captions should be placed below the figures and can not be included as a part of the figure. Each figure should be submitted as a separated graphic file, in .cdr, .eps, .ps, .png or .tif format. Tables and figures should be numbered consecutively with Arabic numerals.

Each photograph with minimum 300 dpi resolution should be delivered in electronic formats (TIFF, JPG or PNG) as a separated file.

**References.** All references should be marked in the text by Arabic numerals in square brackets and listed at the end of the paper in order of their appearance in the text, including exclusively publications cited inside. Samples of correct formats for various types of references are presented below:

[1] Y. Namihira, "Relationship between nonlinear effective area and mode field diameter for dispersion shifted fibres", *Electron. Lett.*, vol. 30, no. 3, pp. 262–264, 1994.

[2] C. Kittel, *Introduction to Solid State Physics*. New York: Wiley, 1986.

[3] S. Demri and E. Orłowska, "Informational representability: Abstract models versus concrete models", in *Fuzzy Sets, Logics and Knowledge-Based Reasoning*, D. Dubois and H. Prade, Eds. Dordrecht: Kluwer, 1999, pp. 301–314.

**Biographies and photographs of authors.** A brief professional author's biography of up to 200 words and a photo of each author should be included with the manuscript.

**Galley proofs.** Authors should return proofs as a list of corrections as soon as possible. In other cases, the article will be proof-read against manuscript by the editor and printed without the author's corrections. Remarks to the errata should be provided within one week after receiving the offprint.

**Copyright.** Manuscript submitted to JTIT should not be published or simultaneously submitted for publication elsewhere. By submitting a manuscript, the author(s) agree to automatically transfer the copyright for their article to the publisher, if and when the article is accepted for publication. The copyright comprises the exclusive rights to reproduce and distribute the article, including reprints and all translation rights. No part of the present JTIT should not be reproduced in any form nor transmitted or translated into a machine language without prior written consent of the publisher.
For copyright form see: http://www.nit.eu/for-authors

A copy of the JTIT is provided to each author of paper published.

---

*(Contents Continued from Front Cover)*



INSTYTUT ŁĄCZNOŚCI
PANSTWOWY INSTYTUT BADAWCZY

Editorial Office

National Institute
of Telecommunications
Szachowa st 1
04-894 Warsaw, Poland

tel. +48 22 512 81 83
fax: +48 22 512 84 00
e-mail: redakcja@itl.waw.pl
http://www.nit.eu