ITHEA

International Journal

INFORMATION TECHNOLOGIES & KNOWLEDGE

2016 Volume 10 Number 2

International Journal
INFORMATION TECHNOLOGIES & KNOWLEDGE

Volume 10 / 2016, Number 2

**International Journal “INFORMATION TECHNOLOGIES & KNOWLEDGE” (IJ ITK)
is official publisher of the scientific papers of the members of
the ITHEA International Scientific Society**

IJ ITK rules for preparing the manuscripts are compulsory.
The **rules for the papers** for IJ ITK are given on www.ithea.org

Responsibility for papers published in IJ ITK belongs to authors.

International Journal “INFORMATION TECHNOLOGIES & KNOWLEDGE” Volume 10, Number 2, 2016
Edited by the **Institute of Information Theories and Applications FOI ITHEA**, Bulgaria, in collaboration with:
Institute of Mathematics and Informatics, BAS, Bulgaria; V.M.Glushkov Institute of Cybernetics of NAS, Ukraine;
Universidad Politechnika de Madrid, Spain; Hasselt University, Belgium;
St. Petersburg Institute of Informatics, RAS, Russia; Institute for Informatics and Automation Problems, NAS of the Republic of Armenia.
Printed in Bulgaria
Publisher ITHEA®
Sofia, 1000, P.O.B. 775, Bulgaria. www.ithea.org, e-mail: info@foibg.com
Technical editor: Ina Markova

Издател: ИТЕА®, София 1000, ПК 775, България, www.ithea.org, e-mail: info@foibg.com

ISSN 1313-0455 (printed) ISSN 1313-048X (online)

A METHOD FOR EVALUATION OF INFORMATIONAL SERVICES - STEP 1: COMPUTING THE HARDWARE PROPORTIONALITY CONSTANTS

Krassimira Ivanova, Emiliya Saranova, Krassimir Markov, Stefan Karastanev

Abstract: *Enhancing the hardware power does not cause linear enhancing of the informational services' performance. To discover the value of growth, one has to test both source and enhanced systems running equal or similar services. If we need to discover the growth of services' performance for different computers' configurations we have to have common basis for comparing one software service with those of other systems, which are tested on different computer configurations. In this paper we outline the first step of a method for solving such problem. This step consists of computing the hardware proportionality constants. Further two papers will present the rest two steps of the method. All examples in the paper are based on results from real experiments presented in the [Markov et al, 2015].*

Keywords: *Evaluation of informational services; computing the hardware proportionality constants.*

ACM Classification Keywords: *H.3.4 Systems and Software - Performance evaluation (efficiency and effectiveness); H.3.5 Online Information Services.*

Introduction

If one needs to discover the growth of services' performance for different computers' configurations he/she has to have common basis for comparing one service with same or other services which are tested on different computer configurations.

Enhancing the hardware power does not cause linear enhancing of the software performance. To discover the value of growth one has to test both source and enhanced systems running equal or similar software. Practically, the computers have different characteristics and operational systems. In addition, the target computers and operational systems may be not available for experiments but some benchmarks may be published.

Let we have to compare loading times for given datasets for different software services in the next conditions:

- Service **X** is tested on two computer configurations: **U** and **W**, where **W** is enhanced configuration in respect of **U**; service **Y** is tested on different computer configuration **V** of the same class and similar characteristics as **U**. We have testing couples **(X,U)**, **(X,W)**, and **(Y,V)**;
- Computer configurations **U** and **W** are not available for testing and all work has to be done on computer configuration **V**;
- **X** has published results from tests on **U** by dataset **S1** with **|S1|** instances and on **W** with similar dataset **S2** with **|S2|** instances; **Y** is tested on configuration **V** by datasets **S1** and **S2**;
- Loading times are respectively: $L_{(X,U,S1)}$, $L_{(X,W,S2)}$, $L_{(Y,V,S1)}$, $L_{(Y,V,S2)}$.

The problem we have to solve is: "What will be the loading time of service **Y** if it will be run on computer configuration **W** with dataset **S2**?" i.e. $L_{(Y,W,S2)} = ?$.

The method for solving this problem consists of three steps:

1. Computing the hardware proportionality constants;
2. Computing the software systems' performance and proportionality constants;
3. Analysis of experiments: Rank-based multiple comparison.

In this paper we outline the first step of the method. This step consists of computing the hardware proportionality constants. Further two papers will present the rest two steps of the method. All examples in the paper are based on results from real experiments presented in the [Markov et al, 2015].

Computing the hardware proportionality constants

Evaluation, comparison, and selection of modern computer and communication systems are complex decision problem. System evaluation techniques can be either qualitative or quantitative [Dujmovi'c, 1996]:

- Qualitative techniques are usually based on a list of features to be analyzed for each competitive system. The list includes technical characteristics, costs, and other components for evaluation. After a study of proposed systems the evaluator creates for each proposal a list of advantages and a list of disadvantages. The lists summarizing advantages and disadvantages are then intuitively compared and the final ranking of proposed systems is suggested. Such an approach is obviously attractive only when the decision problem is sufficiently simple. In cases with many decision criteria it is difficult to properly intuitively aggregate a number of components affecting the final decision, and it is not possible to precisely identify minor differences between similar proposals. In addition, it is extremely difficult to justify whether a given difference in total cost is commensurate to a corresponding difference in total performance. These difficulties can be reduced by introducing quantitative components in the decision process [Dujmovi'c, 1996].

– The aim of quantitative methods is to make the system evaluation process well structured, relatively simple, and accurate, providing global quantitative indicators which are used to find and to justify the optimum decision [Dujmovi'c, 1996].

For purposes of this research we will use simple evaluation system based on traditional scoring techniques. The basic idea is very simple [Dujmovi'c, 1996]: for a set of evaluated systems we first identify n relevant components (performance variables) that are individually evaluated. The results of evaluation are individual normalized scores $E_1, \ldots, E_n$, where $0 \leq E_i \leq 1$ (or $0 \leq E_i \leq 100\%$). The average score is then

$$E = (E_1 + \ldots + E_n)/n.$$

If all components are not equally important then we introduce positive normalized weights, which reflect the relative importance of individual components. $W_1,\ldots,W_n$. Usually, $0 \leq W_i \leq 1$, i = 1, 2,..., n, and $W_1 + \ldots + W_n = 1$.

The global score is defined as a weighted arithmetic mean:

$$E = W_1E_1 + W_2E_2 + \ldots W_nE_n,\ 0 \leq E \leq 1.$$

Let compare a basic hardware configuration with three others. The characteristics we will take in account are Processor (**P**), Physical Memory (**M**) and Hard Disk capacity (**D**).

We assume that the operating systems and service software are equivalent in all cases. For concrete computer systems used in the experiments let we have respectively:

- ✓ **Configuration K** is basic configuration:
  - Processor: Intel Core2 Duo T9550 2.66GHz; CPU Launched: 2009; *Average CPU Mark:* **1810** (P_K=1810) [T9550, 2009];
  - Physical Memory: 4.00 GB (M_K=4);
  - Hard Disk: 100 GB data partition; 2 GB swap (D_K=100);
  - Operating System: 64-bit operating system Windows 7 Ultimate SP1.

Characteristic values of configuration **K** are: P_K=1810, M_K=4, D_K=100.

- ✓ **Configuration A** is benchmark configuration of [Becker, 2008]:
  - Processor: Intel Pentium Dual Core 2.8 GHz; CPU Launched: 2008; Average CPU Mark: 598 (PA =598) [Pentium Dual, 2008];
  - Physical Memory: 1 GB (M_A =1);
  - Hard Disk: 40 GB data partition; 2 GB swap (D_A =40);
  - Operating System: Ubuntu Linux 7.10 64-bit.

Characteristic values of configuration **A** are: P_A =598, M_A =1, D_A =40.

- ✓ **Configuration B**: is benchmark configuration of [BSBMv2, 2008] and [BSBMv3, 2009] DELL workstation:
- Processor: Intel Core2Quad Q9450 @ 2.66GHz, CPU Launched: 2008; *Average CPU Mark:* **3791** (P_B =3791) [Q9450, 2008];
- Physical Memory: 8GB DDR2 667 (4 x 2GB) (M_B =8);
- Hard Disks: 160GB (10,000 rpm) SATA2, 750GB (7,200 rpm) SATA2 (D_B =160 + 750 = 910);
- Operating System: Ubuntu 8.04 64-bit, Kernel Linux 2.6.24-16-generic; Java Runtime: VM 1.6.0, HotSpot(TM) 64-Bit Server VM (build 10.0-b23); Separate partitions for application data (on 7,200 rpm HDD) and data bases (on 10,000 rpm HDD).

Characteristic values of configuration **B** are: P_B =3791, M_B =8, D_B =910.

- ✓ **Configuration C** is benchmark configuration used for LDIF [LDIF Benchmarks, 2013; LDIF, 2013]:
  - Processor: Intel i7 950, 3.07GHz (quad core); CPU Launched: 2009, *Average CPU Mark:* **5664** (P_C =5664) [i7 950, 2009];
  - Physical Memory: 24GB (M_C =24);
  - Hard Disks: 2 × 1.8TB (7,200 rpm) SATA2 (D_C =3600);
  - Operating System: Ubuntu 11.04 64-bit, Kernel: 2.6.38-10; Java version: 1.6.0_22.

Characteristic values of configuration **C** are: P_C =5664, M_C =24, D_C =3600

Global scores of computer configurations

Normalized estimation E_P of processors' power will be computed by formula:

$$E_p = \frac{P_i}{P_K},\ i = A,B,C$$

where P_j, j=K,A,B,C is the processor's average CPU mark.

We assume that the processors' power is very important and because of this we will use processors weight as 0.5, i.e.

$$W_P = 0.5.$$

Normalized estimation E_M of physical memory will be computed by formula:

$$E_M = \frac{M_i}{M_K}, i = A, B, C$$

where M_j, j=K,A,B,C is the size of main memory in Giga bytes.

We assume that main memory is more important than hard disk memory and because of this we will use main memory weight as 0.3, i.e.

$$W_M = 0.3.$$

Normalized estimation E_{HD} of hard disk capacity will be computed by formula:

$$E_D = \frac{D_i}{D_K}, i = A, B, C$$

where D_j, j=K,A,B,C is the size of hard disk memory in Giga bytes.

We assume that the hard disk memory weight as 0.2, i.e.

$$W_D = 0.2.$$

Formula for computing the global score of computer configuration is defined as a weighted arithmetic mean:

$$E_i = W_P\, E_P + W_M\, E_M + W_D\, E_D$$

or

$$E_i = 0.5\, E_P + 0.3\, E_M + 0.2\, E_D$$

Global scores of experimental computer configurations

The global scores of experimental computer configurations are as follow.

- Global score $\mathbf{E_K}$ of configuration **K** is **1:**

P_K=1810; $\mathbf{EK_P}$ = 1810/1810 = **1**

M_K=4; $\mathbf{EK_M}$ = 4/4 = **1**

D_K=100; $\mathbf{EK_D}$ = 100/100 = **1**

$$\mathbf{E_K} = 0.5EK_P + 0.3EK_M + 0.2EK_D = 0.5*1+0.3*1+0.2*1 =$$
$$= 0.5+0.3+0.2 = \mathbf{1}$$

- Global score $\mathbf{E_A}$ of configuration **A** is **0.32**:

P_A=598; $\mathbf{EA_P}$ = 598/1810 = **0.33**

M_A=1; $\mathbf{EA_M}$ = 1/4 = **0.25**

D_A=40; $\mathbf{EA_D}$ = 40/100 = **0.40**

$$\mathbf{E_A} = 0.5EA_P+0.3EA_M+0.2EA_D = 0.5*0.33+0.3*0.25+0.2*0.40 = $$
$$= 0.165+0.075+0.08 = \mathbf{0.32}$$

- Global score $\mathbf{E_B}$ of configuration **B** is **3.465**:

P_B=3791 $\mathbf{EB_P}$ = 3791/1810 = **2.09**

M_B=8 $\mathbf{EB_M}$ = 8/4 = **2**

D_B=910 $\mathbf{EB_D}$ = 910/100 = **9.1**

$$\mathbf{E_B} = 0.5EB_P+0.3EB_M+0.2EB_D = 0.5*2.09+0.3*2+0.2*9.1 = $$
$$= 1.045+0.6+1.82 = \mathbf{3.465}$$

- Global score $\mathbf{E_C}$ of configuration **C** is **10.565**:

P_C=5664; $\mathbf{EC_P}$ = 5664/1810 = **3.13**

M_C=24; $\mathbf{EC_M}$ = 24/4 = **6**

D_C=3600; $\mathbf{EC_H}$ = 3600/100 = **36**

$$\mathbf{E_C} = 0.5EC_P+0.3EC_M+0.2EC_D = 0.5*3.13+0.3*6+0.2*36=$$
$$= 1.565+1.8+7.2 = \mathbf{10.565}$$

Hardware proportionality constants

The ***hardware proportionality constants*** H_i, i = A, B, C, for normalizing our results to be comparable with results received on other computer configurations are as follow:

$\mathbf{K \propto A}$: H_A = E_K/E_A = 1 / 0.32 = **3.125**

$\mathbf{K \propto B}$: H_B = E_K/E_B = 1 / 3.465 = **0.289**

$\mathbf{K \propto C}$: H_C = E_K/E_C = 1 / 10.565 = **0.095**

Conclusion

The goal of this work was to outline the first step of a method for estimating further development of any informational service. This step consists of computing the hardware proportionality constants. All examples in the paper were based on results from real experiments presented in the [Markov et al, 2015]. Enhancing the hardware power does not cause linear enhancing of the informational services' performance. To discover the value of growth one has to test both source and enhanced systems running equal or similar services. If we need to discover the growth of services' performance for different computers' configurations we have to have common basis for comparing one software service with those of other systems, which are tested on different computer configurations. Further two papers will present the rest steps of the method.

Acknowledgement

The paper is published with partial support by the Project "Methods for modeling and evaluation of informational services" of the University of Telecommunications and Posts, Sofia, Bulgaria.

Bibliography

[Becker, 2008] Christian Becker, "RDF Store Benchmarks with Dbpedia", Freie Universität Berlin, 2008, http://wifo5-03.informatik.uni-mannheim.de/benchmarks-200801/ (accessed: 05.04.2013)

[BSBMv2, 2008] Berlin SPARQL Benchmark Results, V2 2008, http://wifo5-03.informatik.uni-mannheim.de/bizer/berlinsparqlbenchmark/results/V2/index.html (accessed: 31.07.2013)

[BSBMv3, 2009] Berlin SPARQL Benchmark Results, V3, 2009, http://wifo5-03.informatik.uni-mannheim.de/bizer/berlinsparqlbenchmark/results/V3/index.html (accessed: 31.07.2013)

[Dujmovi'c, 1996] Jozo Dujmovi'c, "A Method for Evaluation and Selection of Complex Hardware and Software Systems", In: CMG 96 Proceedings, 1996, pp. 368-378 http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.48.4388 (accessed: 31.07.2013)

[i7 950, 2009] Intel i7 950 @ 3.07GHz (quadcore); CPU Launched: 2009; http://www.cpubenchmark.net/cpu.php?cpu=Intel+Core+i7+950+%40+3.07GHz&id=837 (accessed: 31.07.2013)

[LDIF Benchmarks, 2013] LDIF - Benchmark Results, http://ldif.wbsg.de/benchmark.html (accessed: 31.07.2013)

[LDIF, 2013] LDIF – Linked Data Integration Framework, http://ldif.wbsg.de/ (accessed 09.04.13).

[Markov et al, 2015] Krassimir Markov, Krassimira Ivanova, Koen Vanhoof, Vitalii Velychko, Juan Castellanos, „Natural Language Addressing", ITHEA® Hasselt, Kyiv, Madrid, Sofia, IBS ISC No.: 33, 2015, ISBN: 978-954-16-0070-2 (printed), ISBN: 978-954-16-0071-9 (online), 315 p.

[Pentium Dual, 2008] Intel Pentium Dual Core CPU @ 2.8 GHz; CPU Launched: 2008; http://www.cpubenchmark.net/cpu.php?cpu=Intel+Pentium+D+2.80GHz&id=1126 (accessed: 31.07.2013)

[Q9450, 2008] Intel Core 2 Quad Q9450 @ 2.66GHz, CPU Launched: 2008; http://www.cpubenchmark.net/cpu.php?cpu=Intel+Core2+Quad+Q9450+%40+2.66GHz&id=1046 (accessed: 31.07.2013)

[T9550, 2009] Intel® Core2 Duo CPU T9550 @ 2.66GHz; CPU Launched: 2009, http://www.cpubenchmark.net/cpu.php?cpu=Intel+Core2+Duo+T9550+%40+2.66GHz&id=1011 (accessed: 31.07.2013)

Authors' Information

***Krassimira Ivanova** – Assoc. prof. Dr.; University of Telecommunications and Posts, Sofia, Bulgaria; Institute of Mathematics and Informatics, BAS, Bulgaria; e-mail: krasy78@mail.bg;*

***Major Fields of Scientific Research:** Software Engineering, Business Informatics, Data Mining, Multidimensional multi-layer data structures in self-structured systems*

***Emiliya Saranova** - Assoc. Prof. PhD; Vice-rector of University of Telecommunication and Post, Sofia, Bulgaria; Institute of Mathematics and Informatics – Bulgarian Academy of Science, Sofia, Bulgaria; e-mail: emiliya@cc.bas.bg*

Major Fields of Scientific Research: Mathematical Modeling in Telecommunications, Informational Modeling

***Krassimir Markov** – Institute of Mathematics and Informatics, BAS, Acad. G.Bontchev St., bl.8, Sofia-1113, Bulgaria; ITHEA Institute of Information Theories and Applications, P.O. Box: 775, Sofia-1000, Bulgaria; e-mail: markov@foibg.com*

***Major Fields of Scientific Research**: General theoretical information research, Multi-dimensional information systems; Software Engineering, Business Informatics, Data Mining.*

***Stefan Karastanev** – Assist. prof.; Institute of Mechanics, BAS, Bulgaria;*

e-mail: stefan@imbm.bas.bg

***Major Fields of Scientific Research:** Software Engineering, Data Processing and Mining, Data Structures in Information Systems.*

A COHESIVE TECHNO-BUSINESS VISION FOR FUTURE WIRELESS NETWORKING

Ivan Ganchev

***Abstract**: This paper proposes a cohesive vision for future wireless networking, based on a novel generic techno-business model, which – while different from the evolving legacy one – is both feasible and attractive to stakeholders, and could drive the '5G and beyond' vision for future generations of wireless communications. The technical implications of the proposed model are explained and the key technological innovations required to support it are outlined.*

***Keywords**: Consumer-Based techno-business Model (CBM), Always Best Connected and best Served (ABC&S), Ubiquitous Consumer Wireless World (UCWW), Third-Party Authentication, Authorization and Accounting (3P-AAA), Consumer Identity Module (CIM) card, Personal IPv6 (PIPv6) address.*

***ACM Classification Keywords**: C.2.1 Network Architecture and Design – Wireless communication, C.2.2 Network Protocols – Protocol architecture, D.2.2 Design Tools and Techniques – Evolutionary prototyping, D.4.6 Security and Protection – Authentication.*

1. Introduction

This paper puts forward views on evolving fifth generation wireless world and how that evolution may be directed for the benefit of the consumers and other stakeholders. It argues that an infrastructural re-think on the way Authentication, Authorization and Accounting (AAA) service is supplied is the key to this evolution. At a high level this may be described as a business plan for the supply of mobile services, especially with the wireless access service component being founded on a consumer-based structure rather than on a subscriber-based one. The novel Consumer-Based techno-business model (CBM) will enable a loose dynamic (even casual) consumer-type association between mobile users (MUs) and access network providers (ANPs). Innovations required to support this include a pivotal role of a third-party AAA (3P-AAA) service provider entity.

The view on the next (5G) generation of wireless communications, employed in this paper, is the one which encompasses all existing, planned and future mobile and fixed wireless networks, both terrestrial and satellite. The goal is to sell services to a great market of users in overlapping local, regional, and global domains; to create a Ubiquitous Consumer Wireless World (UCWW), where connectivity will be available anywhere-anytime-anyhow and services will be rapidly deployed on-demand, customized to the user's needs, and adapted to the current user context and network context, in the best possible way independent of the user's movement across heterogeneous access networks. This vision requires unprecedented levels of autonomy, application service adaptability, and network element integration at all levels including mobile devices, access networks (ANs), and core networks. Other terms that characterize this 5G vision include dynamic quality of service (QoS) provisioning, and network- and device dynamic reconfigurability. From access networks point of view, especially that based on the existing Subscriber-Based techno-business Model (SBM), the challenges this 5G vision raises are convergence, integration, and interworking of existing and emerging heterogeneous wireless networks, in order to provide users with all their desired services in a seamless Always Best Connected and best Served (ABC&S) way.

For the mobile user, the experience of ABC&S communications services should preferable move towards having consumerist-type characteristics where, through user-friendly interfaces, ABC&S decisions are user-driven and user-executed. ABC&S scenarios should enable mobile users to move seamlessly between different wireless access networks according to their own criteria (e.g. on the basis of price/performance ratios), while maintaining active service sessions, i.e. without interrupting service sessions, restarting applications, or losing data. The CBM model, promoted and further explained in this paper, would foster this new type of wireless network environment, called the UCWW. It is much less constrained than the SBM model widely used today, which in fact would militate against many aspects of it. Section 2 provides more details on the novel CBM model in comparison to the existing SBM model.

A real drive towards an open ABC&S UCWW paradigm has the potential to gradually restructure the existing subscriber-based business realization of mobile communications, transforming it into a consumer-based one. In this it raises important challenges for existing ANPs (i.e. mobile operators) and opens new opportunities for new ANPs, aiming to fill niche markets. For a new ANP entrant, it would mean the possibility of ease of entry and of having dynamic (even casual) consumerist-like relationships with users, i.e. offering and providing services to them without any prior business relationship and subscription. Significant re-thinking and breakthroughs in the way traditional AAA services are provided are necessary. One of the main challenges is the creation of a framework for strategic infrastructure and protocol development for independent autonomous provision of AAA services by third parties (3P-AAA), which are not network providers. This concept, the foundation stone for a re-structured business plan for the provision of wireless services, is presented in Section 3.

The CBM model utilizes a novel concept of a personal IPv6 (PIPv6) address, which enables advanced mobility, i.e. in ways not presently possible, and continued participation in various evolving communications scenarios. Through an enhanced AAA functionality, the PIPv6-based CBM model also has the potential to enable commercially viable ad-hoc and/or open mesh-networking solutions, where a mobile node (object) acting as a gateway (or relay) may offer (or facilitate) wireless Internet access services casually or persistently to other mobile nodes/objects and be paid for this service, e.g. through a 3P-AAA service provision. Realization of this would bring about a radical change to the access network business, and add many new ways whereby mobile users will be able to gain access to network services. More details on the PIPv6 address mechanism along with a corresponding generic communication scenario are presented in Section 4.

A typical implementation approach for the identification and authentication of mobile users (MUs) on mobile devices today is by means of a subscriber identity module (SIM) card inserted in the corresponding device currently used. In the UCWW, this type of card will be replaced by a new type – a smart consumer identity module (CIM) card – which will contain the user's credit card details, or a specific authentication code acceptable/provided by a 3P-AAA service provider (3P-AAA-SP). This way each service charge, incurred by the user, will be paid indirectly through the 3P-AAA-SP, who will then arrange relevant payment to the corresponding service provider. Section 5 deals with the CIM card aspects.

2. Techno-Business Models

Important in the process of bringing about this ABC&S UCWW reality are the questions of feasible techno-business models:

- What kind of generic model(s) would be attractive to stakeholders and would drive this 5G vision?
- What are the technical implications of such models?

The existing and widely used SBM model is founded on the necessity of a mobile user (MU) having a contract (an account) with at least one ANP before s/he may start using mobile services. For the MU, this ANP is called the home ANP. The user is a **subscriber** and as such has a mobile number(s) tied to that home ANP. Even though the mobile phones may have multiple SIM cards inserted, equivalent to having a number of subscriber contracts with different ANPs, this is not user friendly and the mobile numbers are under the control of home ANPs, even where number portability is facilitated. Relative to the home ANP, all other ANPs are termed foreign or visited ANPs. Prior roaming agreements between foreign and home ANPs are required in order for MUs to camp on them. Similarly, the mobile service providers (xSPs) are able to offer their services (excluding those, of course, that could be accessed/obtained directly on/from the Internet) through ANP networks under respective prior business agreements with them directly or indirectly through value-added service providers (VASPs*). The key idea of the SBM model is that the user is being locked-in as a subscriber to a particular ANP for a long time. The general trend today is to continue with this model, which mostly benefits the existing ANPs (i.e. the dominant cellular mobile operators) and converts into a pipe-dream the entire idea of using the access network simply as a section of service-independent transport pipes [O'Droma, 2004b].

In this ANP-centric SBM model, the home ANP is placed at the center as both the effective manager of the user's wireless communications and AAA activities, and the supplier of part of the wireless communication services (Figure 1). It may be well argued that this uniquely strong position of the home ANPs, with its potential to constrain the users' freedom and independence, by the fact of their being 'subscribers' more than 'consumers', in seeking better value for money, is inimical to the user's best interests [O'Droma, 2004a].

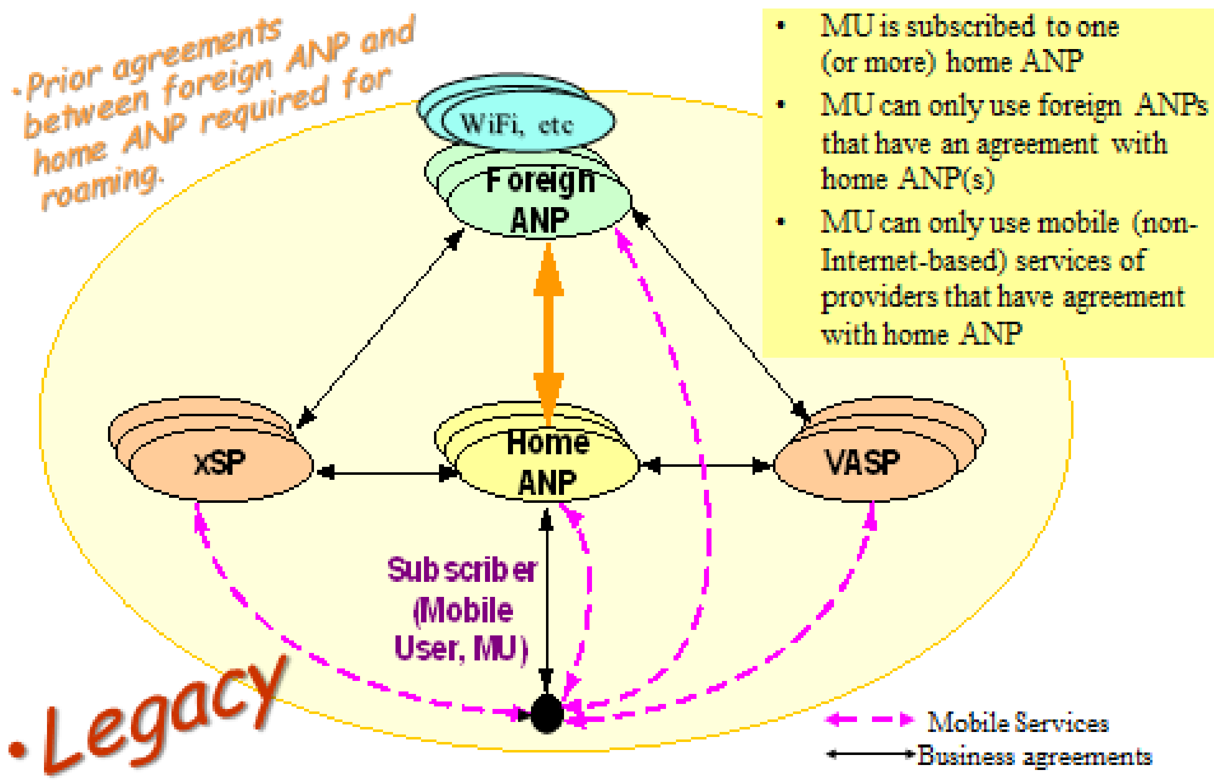

Fig. 1. The SBM with the home ANP at the center

* VASPs provide additional services related to, e.g. ordinary service deployment and adaptation to user/device/network profiles, special service configurations, transparent network and mobile device reconfigurations, service adaptations to satisfy special service requests, etc.

It will always be difficult to escape the constraints and limitations effected by the home ANPs for their own business reasons including those serving dominance goals within the AN market, and even within the market of the mobile service providers (xSPs). It will also be a difficult business place for new ANP entrants as the business agreements with other ANPs for interoperability, interworking, AAA etc., and with many crucial xSPs, will always be a critical and necessary component for their entry and survival, as well as having to have in place customer administrative and management support before they might hope to start seeking 'user (subscriber) accounts' and make inroads into the market, and face the task of convincing great numbers of potential users, who are probably already subscribers to competing long standing ANPs, to become their subscribers. This is the inherent nature of the SBM model [O'Droma, 2004a].

Naturally the problems with the SBM will put a brake on fast deployment and flexible provision of new services and ANs, as well as result in significant insecurity for (barrier to) new xSP entrants because they will need prior xSP-ANP business agreements, which is slow and time-consuming process. In addition, it will cause slow, constrained ABC&S 5G evolution due to the fact that the mobile user must be a subscriber to an ANP before may access their ABC&S offerings, and will find him/herself constrained to these. While in the short term this path of 5G evolution may benefit existing cellular wireless license holders, in the long term it will be to the detriment of all, with the likely appearance of disruptive wireless technologies leading to the specter of serious fragmentation in the absence of the core elements of a technological foundation for a structured business plan, through which such technologies can be brought into easy access for customers. Thus the SBM is considered as a legacy techno-business model [O'Droma, 2004a].

Within the context of the shortcomings of the SBM, the alternative CBM model comes with many attractions for the evolving ABC&S UCWW. In this model, the mobile users (MUs) do not have formal subscriber relationships with ANPs and thus may act as **consumers**, not as subscribers. This is much like other consumer services, e.g. the shopping of goods on a street or in the mall. Moreover in the CBM, the MU owns his/her mobile numbers (e.g. IPv6 addresses) as of right and this is not subject to any 'subscriber agreement' as the consumer-user does not have a subscriber contract with any ANP. This allows him/her to request different types of services from different types of (mobile) service providers (xSP), which services are provided through multiple service-specific ABC&S wireless access connections which best match the consumer's profile/role. In the CBM, there is no distinguishing between home ANPs and foreign/visited ANPs; they are simply called 'ANPs'. They provide access network infrastructure and transport medium. Examples include cellular (2G/3G/4G) mobile operators, Wireless Local Area Network (WLAN) providers (utilizing the IEEE 802.11 standard) with corresponding Wi-Fi hotspots, WiMax providers (based on the IEEE 802.16 standard), etc.

A key CBM element is that it separates out the administration and management of users' one-stop-shop authentication and accounting system from the business of supplying a wireless access network service, and locates it with a third-party AAA service provider (3P-AAA-SP), who is not traditionally a stakeholder in the wireless communications business. Through business agreements with such 3P-AAA-SPs, all types of providers (ANPs and xSPs) will be able to offer their charge- or fee-based services to MUs who have credit arrangements with one or more 3P-AAA-SP, just as they have one or more credit cards today, and similarly through this entity will receive periodic itemized bills for all services where they incur costs, which have been paid through this 3P-AAA-SP. This way all ANPs and xSPs charges will be paid indirectly through 3P-AAA-SPs. In the case of a MU using multiple 3P-AAA-SPs, the MU's choice of 3P-AAA-SP at any time for any service will be dictated by decision processes similar to those occurring today when deciding which credit card to use for a particular bill. However, this will be done mostly transparently to the user, based on his/her predefined preferences [O'Droma, 2004a]. In this case, the 3P-AAA-SP becomes the central player. An explanatory example of this consumer-based business plan structure is illustrated in Figure 2.

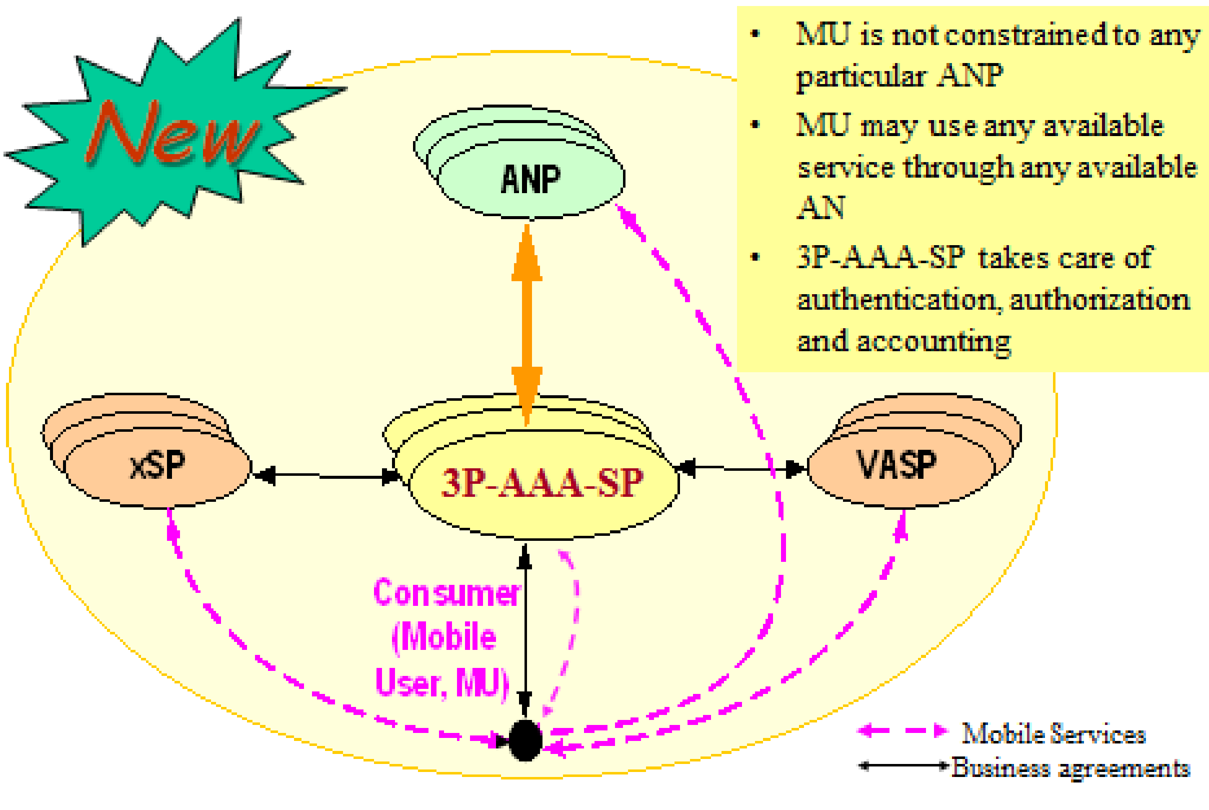

Fig. 2. The CBM with a third-party AAA service provider (3P-AAA-SP) at the center

In the UCWW, established on this CBM model, the MU will be identified by a smart consumer identity module (CIM) card, replacing the legacy SIM card. The CIM will contain the user's credit card details and/or a specific authentication code acceptable, or even provided by, the 3P-AAA SP. This way each service charge (e.g. to ANP or xSP) may be paid indirectly through a 3P-AAA-SP, indicated by the payee. Financial institutions, such as present-day credit-card companies, would probably be the most suitable contenders for the 3P-AAA-SP business [O'Droma, 2004a].

Besides giving the MU much greater freedom of movement, through the creation of a more open access-network and mobile-service market with easier and fairer access, this network-independent business model foundation and facility would be particularly attractive to new ANP entrants, and to existing ANPs and xSPs trying to extend their market share, streamline their business, fill niche and specialized AN service provision, and so forth. It has the potential to open the wireless communications market thus facilitating the new entrants. It will provide wider range of freedom and autonomy for ANPs (especially new ones), levelling the AN playing pitch and fostering real ANP competition. For MUs, it will provide a lot of more benefits in terms of greater range of ABC&S offerings as an important business driver for the evolution of ABC&S 5G networking. Other pros include fast deployment and flexible provision of new services, preventing scalability problems, no real differentiation between home ANPs and foreign ANPs, leading to reduction and even elimination of roaming charges, which is completely in line with the EU directives (i.e. a local call will always be a local call regardless of where the user has roamed to!) [O'Droma, 2004b].

The CBM has inherent business attributes to drive forward the evolution of ABC&S networking, as both competition, interoperation, and collaboration advantages will be underpinned by efforts to provide a wide range of QoS offerings with greater flexibility and with a wider range of price/performance ratio options in the efforts of providers to attract in greater numbers of service users. With this novel approach a real accomplishment and true meaning of the word 'mobile' will be reached in relation to 'mobile devices' as these will no more be tied (locked-in) to any access network / ANP. The realization of the full mobility potential will also have a great social impact.

The CBM also has the potential to kick-start the user-driven Integrated Heterogeneous Networking (IHN). The inherent consumers' ability to access whatever services from whatever networks (within whose footprints they are present) whenever they want, is already a basic form of IHN. Decision-making on choices is part of the ABC&S functionality. As IHN evolves, ABC&S-based functionality will support user-driven switching of live connections seamlessly among homogeneous or heterogeneous access networks, in ways largely transparent to the access networks themselves, referred to as 'hot' access network change (HAC). As the signaling involved is between users and the target network and not network-to-network it is a 'paid-for' service (i.e. not a network cost overhead as in network-driven

IHN). Some IHN-type decisions might be taken in collaboration with the mobile service providers (xSPs). In general, HAC-IHN will require supporting functionality at both ends of the service connection (user and xSP), e.g. through the use of the multi-homed functionality and dynamic address reconfiguration of the Mobile Stream Control Transmission Protocol (mSCTP), [O'Droma, 2008].

The CBM also facilitates user-driven IHN for asymmetric service provision, which is network-transparent. For instance the user and the mobile service provider (xSP) may 'collude' to communicate user requests through one type of access-network connection (e.g. 2G/3G) and download content requested through another access network (e.g. Wi-Fi/WiMax). In respect of this, neither access network needs to know about the existence of a part of the connection path through the other. HACs (under the control of the user, the xSP or both collaborating) may occur independently on both parts of such connections [O'Droma, 2008].

The trend from the SBM to the CBM is seen even today (Figure 3).

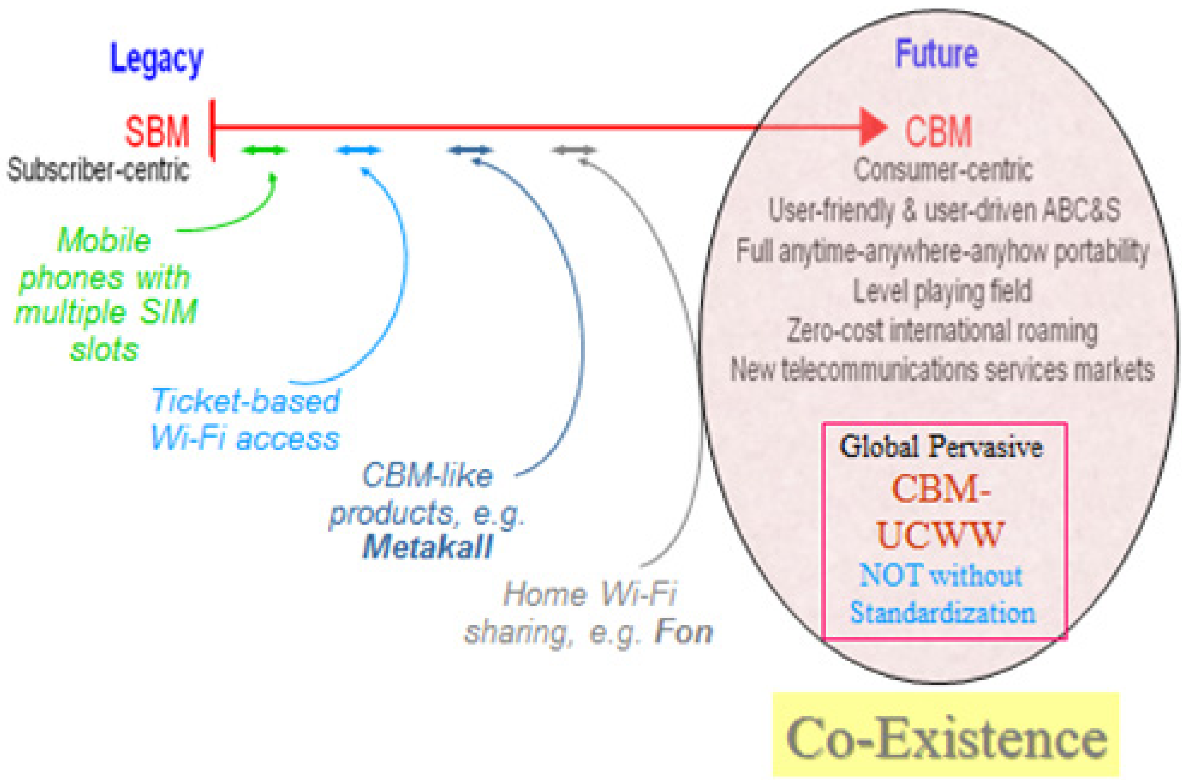

Fig. 3. The transition trend from the SBM to the CBM

The first step, which is widely used by many users who travel a lot around the world in order to avoid high roaming charges, involves the use of dual (or even multiple) SIM active phones, which are capable of receiving calls on both (all) SIM cards. The next step is to utilize ticket/voucher-based Wi-Fi access, again for the same reason to avoid the high cost of mobile access to the Internet while roaming abroad. The third step includes CBM-like products, such as Metakall, which enables Android users to make phone calls from hot-spots operated by a range of Wi-Fi providers around the world without a need for subscription; the users just pay for the Wi-Fi service they use to make a call. Another (closer to CBM) step involves sharing of (home) Wi-Fi access with other users. A primarily example here is Fon† (https://fon.com)– a system of dual access wireless networks – claiming to be the largest Wi-Fi network in the world, with over eight million hotspots as at July 2013. Fon members are allowed to share a part of their Internet connection with other Fon members; otherwise users who choose not to share their Internet connection can buy Wi-Fi access passes or credit from Fon. Fon members, whose Wi-Fi hotspots are used to access the Internet by a paying customer, can receive part of the revenue.

A possible transition solution (from the SBM moving towards the CBM) could be to enable users to avail of 3P-AAA services concurrently with present (SBM) procedures. In the future, mobile phones will be bought in a shop without any binding to a particular network (provider). The user's number (i.e. a personal IPv6 address, c.f. Section 4) will be bought separately by the user and assigned to him/her (e.g. by adding it to his/her CIM card, c.f. Section 5). The key feature here is that the consumer-user will really 'own' his/her personal IPv6 address(es) and that this address will not be bound to a particular network and may be moved from one mobile device to another.

The new type of wireless communications environment UCWW, established on the CBM model, is in harmony with the International Telecommunication Union's Telecommunication Standardization Sector (ITU-T) Next Generation Networks (NGN) objectives [ITU-T, 2004]. Its focus are packet-based communications and networks (primarily IPv6). For this, it utilizes a new 'personal IPv6 address' class described further in this paper. In the UCWW, as in NGN, "service-related functions are independent from underlying transport-related technologies". The UCWW facilitates "unfettered access for users to networks and to competing service providers and/or services of their choice"; the users are not tied to any particular ANP. The UCWW also caters for "generalized mobility which will allow consistent and ubiquitous provision of services to users". Mobility in the UCWW is end-to-end controlled and executed, e.g. via a Hot Access network Change (HAC) with service session continuity, and is primarily user-driven (and also supported by the mobile service providers), and facilitated by a full number portability. Interworking with legacy networks is supported via open interfaces. For this, four new open 3P-AAA

† Nowadays Fon is working mostly with mobile operators, telecommunications companies, and service providers by providing global Wi-Fi access and technology solutions to them.

interfaces are defined and explained in the next section. The UCWW supports unified service characteristics for the same service as perceived by the mobile user. It also provides decoupling of service provision from the network. Furthermore, the UCWW supports a variety of identification schemes, which can be resolved to IP addresses for the purposes of routing in IP networks. The main identification scheme is based on the novel concept of the personal IPv6 address.

A high-level view of the UCWW is depicted in Figure 4.

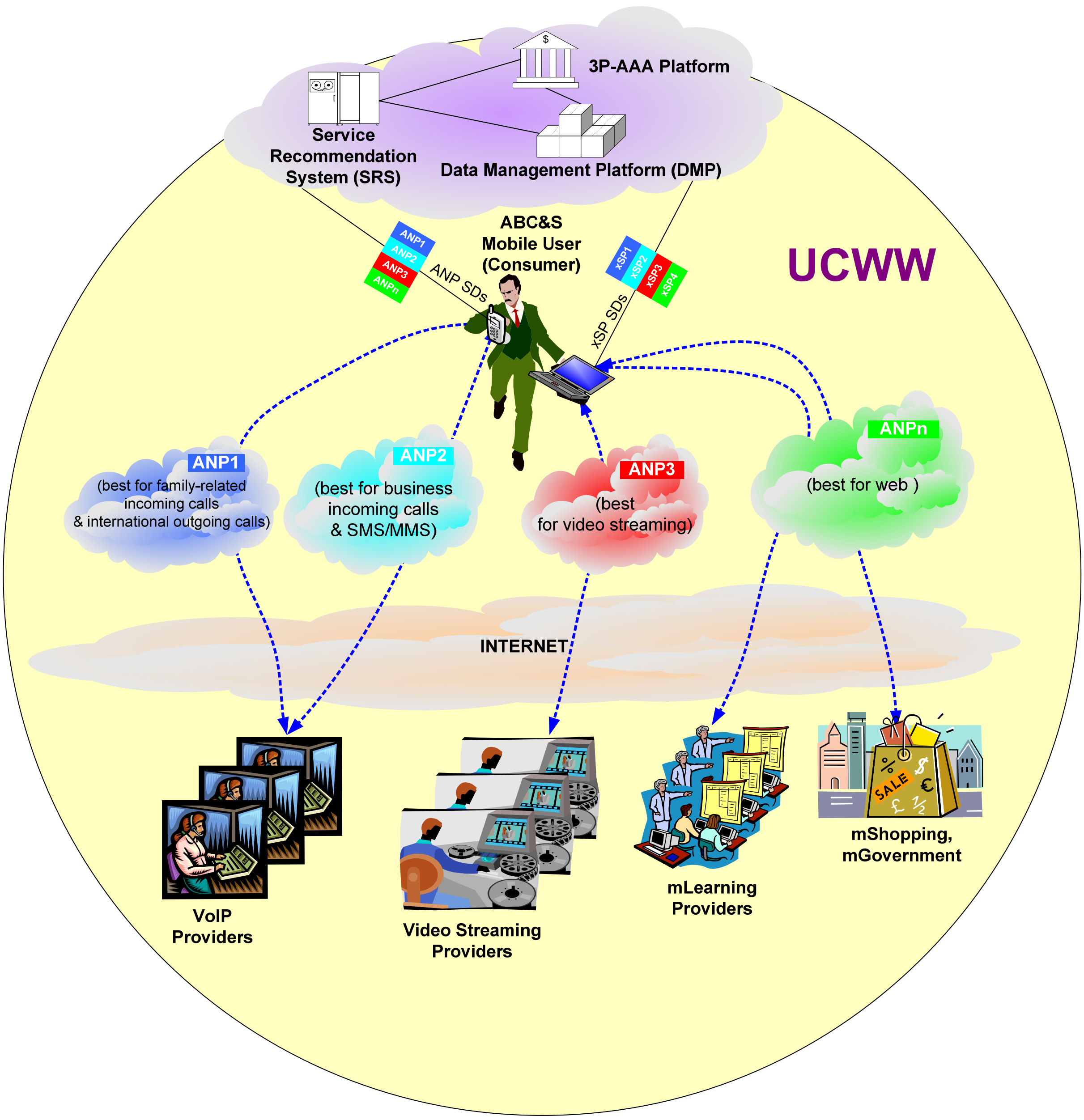

Fig. 4. A high-level view of the UCWW

By utilizing a distributed cloud-based Service Recommendation System (SRS) as proposed in [Ganchev, 2015b], the 'best' service instances are found and recommended to mobile users in a highly personalized and customized way to suit each of them thus facilitating the access to those services through the 'best' available wireless connection under the ABC&S communications paradigm. Moreover, it is a true form of a user-driven ABC&S, which is facilitated and supported by the mobile service providers (xSPs). From a functional point of view, in addition to the SRS, another important UCWW infrastructural component is the Data Management Platform (DMP) [Ganchev, 2016a], which acts as a machine learning platform for turning raw data into actionable analytic dataset, i.e., user behavior profiles, including user preferences, content consumption preferences, shopping preferences, interest preferences, app usage, etc., abiding by the user-privacy principles. For this, it utilizes real-time user's profiling algorithms and off-time data processing algorithms [Ganchev, 2016b].

Naturally new architectural entities, technical and standardization innovations, and internationally agreed protocol structures are required for a managed CBM-UCWW revolution as to support many aspects of this new concept, such as secure protocol interfaces for 3P-AAA-SPs, user identification by a tamper-resistant smart CIM card containing the user's credit card details and providing these when needed as a 'secure AAA ticket', etc. These and other CBM aspects are considered in the next sections.

3. 3P-AAA

As already stated, the 3P-AAA service providers (3P-AAA-SPs) are new business entities, playing a central role in the UCWW, established on the CBM model. Through these entities, all wireless communications- and mobile services' purchasing transactions are made. For the user, this process is accomplished by means of a smart CIM card installed on his/her mobile device. The 3P-AAA facilitates reaching the main CBM goal of separation of the administration and management of users' AAA activity from the supply of a wireless access network service. The 3P-AAA-SPs are (access) network-independent, autonomous, and trusted business entities. This is to prevent their unfair access, in the role of an ANP, to a very wide database of consumers and xSP market information, which could disadvantage other ANPs. This is a strong distinction between the CBM and the SBM with significant socio-economic implications. It may require international regulation, though more likely with time consumer and competition dynamics will make this unnecessary [O'Droma, 2008].

Analogous to the credit-card payment systems, each service charge incurred by a user may be paid indirectly through the user's 3P-AAA-SP, who will periodically send out itemized bills to him/her. This will create also new business development opportunities through the expansion into all areas of purchasing via this smart CIM card and the development of new "mobile money" / "wireless wallet" mobile apps for payment. Examples of potentially suitable 3P-AAA-SPs are financial institutions, such as present-day credit-card companies.

Some proposal details on security, the functional model, and signalling protocols for the 3P-AAA architecture are presented below.

The 3P-AAA interface infrastructure's and signaling protocols' standardization program to enable global provision of 3P-AAA services is outlined in [O'Droma, 2010]. Standards proposed for agreement should respect the need for these services to be scalable, hierarchical, and cognizant. Such attributes may be seen in global 3P-AAA service solutions where service providers deploy their AAA servers regionally, in multiple hierarchical layers, reflecting the dimensions and characteristics of their customer bases. Service to customers may also be improved, e.g., better response times achieved, by porting a copy of a customer's data to the AAA server in the region closest to where a customer is at the moment.

Figure 5 presents a schematic of the key features of the 3P-AAA functional model. In addition to the third-party aspect of the AAA provision, another novel element here is the installation of an AAA client directly on the mobile device. This is a radical distinction from the traditional Diameter/COPS models. Operationally-wise, a smart CIM card (e.g. based on the 'Java card' technology [Oracle, 2015]), inserted into the mobile device, could carry the 3P-AAA client application for AAA communication with various service providers. The AAA servers in the ANP and xSP domains also have 3P-AAA clients for communication with the 3P-AAA-SPs' servers, e.g. for the exchange of accounting information, and charging and billing (C&B) information for vendor transactions to consumers.

The 3P-AAA functional model is based on the following four interfaces, all new to the wireless communications world:

(a) User↔ANP/xSP;

(b) User↔3P-AAA-SP;

(c) ANP/xSP↔3P-AAA-SP;

(d) 3P-AAA-SP_x↔3P-AAA-SP_y.

Corresponding to foreseeable major market sectors, three specialized 3P-AAA-SP classes are posited in [O'Droma, 2010], namely class A for ANPs, class B for xSPs and VASPs, and class C for consumers. It is not intended with this division to exclude other forms or even that a single 3P-AAA-SP would cater for all three markets. The advantage of having different classes of 3P-AAA-SP is that each of them can focus only on one group of service functions so that these functions can be made much more sophisticated. Also with establishing such classes any consequent effects on the new interface protocols may be more easily handled. This specialization option implicates the requirement for a signaling protocol for interaction between these types of 3P-AAA-SP. For this, an Inter-3P-AAA-SP signaling protocol is required to operate over the ***d*** interface.

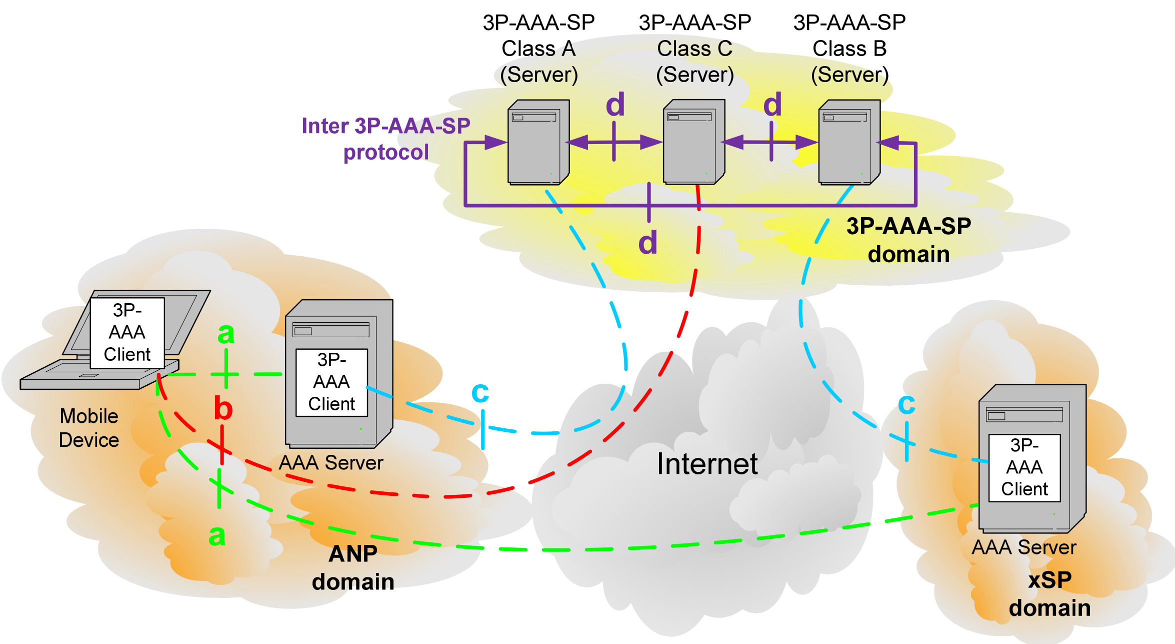

Fig. 5. The 3P-AAA functional model schematic (with four new application-layer interfaces ***a***, ***b***, ***c***, ***d***).

There will be similarities but also distinctions in the services provided within each class. For ANPs, for instance, services will include accounts and related AAA policies, C&B policies, pricing and rating functions, charging detail records generation, account balances, and the like. For consumers (besides accounts), these may include various types of credit top-up services, billing system configuration functionality, functionalities to enable customer-retention discount and promotional schemes, user-definable account format and layout, etc.

On the newly defined interfaces, the existing Internet Engineering Task Force (IETF) Diameter protocol [Calhoun, 2003] has suitable attributes for carrying 3P-AAA signaling. Being conceived for the SBM environment, however, some adjustments for 3P-AAA will be required. One potential solution is to extend its base protocol so as to support 3P-AAA functionality via the addition of new commands and/or attribute value pairs (AVP); another is to define a new 3P-AAA signaling application. The latter seems more attractive standardization route because exploiting the Diameter design, whereby it accepts new autonomous applications which run on its core, has the advantage of not constraining the Diameter core from evolving independently [O'Droma, 2010].

Through standardized protocols, the 3P-AAA client on the user's mobile device interacts with AAA servers of the ANPs and xSPs for mutual authentication and exchange of security credentials. A part of these standardized protocols will be the signaling to establish authentication securely prior to any service purchase. An authentication scheme, based on the ITU-T's X.509 recommendation for a public key infrastructure (PKI) [ITU-T, 2012], can support strong secure mutual authentication and trusted relationship establishment between communicating parties with a minimum number of protocol exchanges.

An example of the consequences of this kind of methodology, being applied to the 3P-AAA infrastructure, is an amendment proposed in [O'Droma, 2010] to the ITU-T's authentication architecture for interworking in NGN [ITU-T, 2008] which to enable user-driven IHN. Figure 6 represents the ITU-T's graphical illustration of an NGN authentication architecture for interworking among heterogeneous wireless networks operating within the SBM. Its objective is to enable a subscriber of one (home) network, e.g., a 3G cellular network, to gain wireless access services as a roamer from another heterogeneous network, e.g., a WLAN or WiMax network, with payment executed through the home network. The approach also may form a basis for subscriber-transparent network-driven handover while roaming among collaborating heterogeneous networks. It uses a four-layer architecture, with the user equipment (UE) at the bottom. The network attachment control function entities are the authenticator (AM-FE), acting as an AAA client, and the home-ANP's AAA server (TUP-FE/TAA-FE) positioned at the third and fourth layers, respectively. The second layer (AR-FE in the access network) acts as an enforcement point filtering packets and allowing through only packets exchanged for initial authentication or subsequently authenticated data packets. The key communication to allow roaming on a heterogeneous network is on the 'roaming interface', (marked 'R' in Figure 6) between the AAA servers in the different network domains.

The proposed modifications of this architecture for the UCWW environment are overlaid on this ITU-T illustration. In the UCWW, the 'home network' attribute no longer applies to access networks. Also the roaming interfaces 'R' are not required. Instead 3P-AAA server entities are present, with which the network-specific AAA servers communicate over standardized 3P-AAA interfaces of type ***c***. Similar type interfaces are needed between 3P-AAA servers and the service-specific AAA servers (SAA-FE). In addition, the AAA client is shifted from the network to the mobile device (UE), and its communication with the ANP's AAA server and 3P-AAA server is performed over the proposed standardized 3P-AAA interfaces of type ***a*** and ***b***, respectively. The balance of the heterogeneous networking decision-making power is now much more in the hands of mobile users and away from networks.

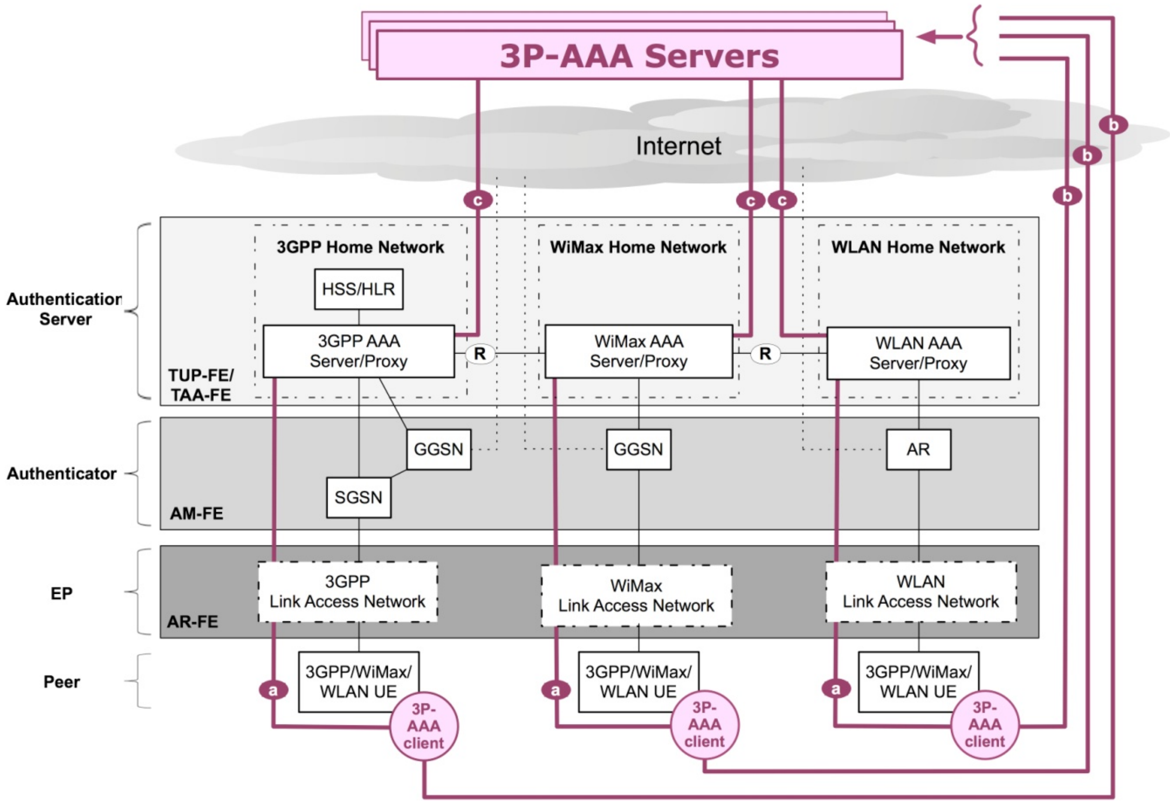

Fig. 6. The ITU-T's SBM authentication architecture for interworking in NGN

and the proposed modification (in purple color) for the UCWW, established on the CBM.

The signaling between 3P-AAA servers and local AAA servers of ANPs (and xSPs) could lead to significant network traffic. This would be the case if the charging and billing (C&B) implementation includes a requirement for continuous processing of the frequently clocked records coming from individual charging functional entities in order to calculate the correct service charge. This is a typical implementation for C&B of phone calls, and other non-flat-rate services, in the SBM environment today. To address this, the concept of a C&B agent could be utilized. Downloaded in advance from the 3P-AAA-SP and deployed in the metering domain of the ANP (or xSP), this agent would perform all associated C&B functions there. For this, the agent would come with a budget for the service to be supplied to the consumer and any other relevant consumer-account details. This agent would have the functionality to manage the budget, e.g., expending it in response to the provider's (ANP or xSP) metering triggers, sending budget replenish requests to the 3P-AAA server when a budget depletion threshold is crossed, and at the end of the service session informing the 3P-AAA server of the total charge, [O'Droma, 2010]. More aspects of the corresponding C&B, including details of evaluation of possible 3P-AAA policy-based accounting models, the possible authorization framework, the proposed generic third-party C&B (3P-C&B) architecture along with supporting protocol candidates, the corresponding rating scenarios for different 3P-AAA-SP classes involved, etc., are presented in [Jakab, 2014]. For instance, the protocol analysis conducted there confirms that for the 3P-AAA infrastructure some extensions of the conventional AAA protocol (e.g. Diameter) will be required, e.g. for Network Access Server (NAS), mobility support, resource brokering, etc.

The novel 3P-C&B approach has the potential of creating an innovative environment for service creation with the following strengths [Jakab, 2009]:

- Providers (ANP, xSP, VASP) do not need to invest in their own C&B system;
- Providers can focus better on their services;
- Providers get the C&B system for free;
- Newly created services (access network communications services or mobile services) are automatically exposed to consumers after their registration with the 3P-AAA-SP;
- Delivery of a 'fair 5G system':
    - The consumer chooses which provider (ANP/xSP/VASP) and which service instance s/he is going to use (so each provider has equal chance for success!);
    - The consumer can freely seek for 'value for money' services;
- Creation of sophisticated C&B systems:
    - 3P-AAA-SPs can invest in their charging services (for new charging schemes or flexible charging solutions) and these will automatically be available to each provider, which would certainly raise more interest and bring new clients.

4. New 'Personal Address' Scheme

For the proposed user address ownership, standardization of a separate class of globally-significant, network-independent 'personal' IPv6 (PIPv6) addresses, as proposed in [Ganchev, 2007], is needed. The PIPv6 address is a static, permanent, and unique address, which is managed and allocated by a global address supplier. Its uniqueness will eliminate the need for duplicated address detection, which is compulsory in IPv6 networks with stateless address auto-configuration (SLAAC).

This new PIPv6 address will enable real consumer-number ownership and full 'anywhere-anytime-anyhow' portability for future generations of mobile users empowered to opt out of their long-term subscriptions with access network providers, and use advertised communication services from any consumer-centric wireless access network present to them. The PIPv6 address can also give more flexibility to set up and operate Wireless Networks of Moving Objects (WiNeMO) [Ganchev, 2014a] because a node (object) can use the same address (identity) in every case and in any communication scenario. It could be used as a long-term identity solution that can prevent impersonation, and Sybil, whitewashing and similar attacks in WiNeMO, and be useful in schemes to deter other types of security attacks.

A new IPv6 address class should be identified for this new PIPv6 address by appropriately assigned *Class Prefix*. Figure 7 shows a possible format, with the space including this field and three other fields, described below. A further small *version* field may also be advisable to allow greater restructuring flexibility in the future.

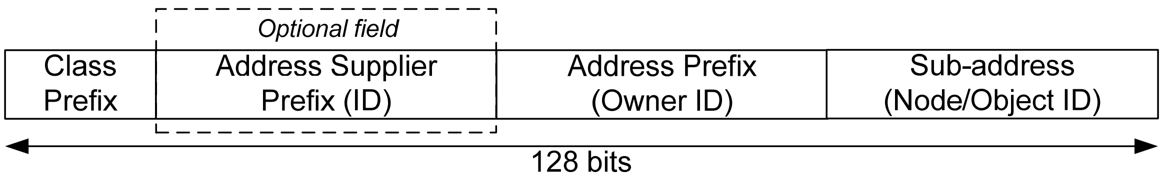

Fig. 7. The format of the new personal IPv6 (PIPv6) address.

The *Address Prefix* is the primary field in the PIPv6 address which could be used to identify the owner (user) of the address. Having the length of the *Owner ID* ranging from 34 to 37 bits will allow addressing of 17 to 137 billion owners. This may seem plenty in a world population context of 7 billion. However, perhaps a longer length, such as 40 bits, would be advisable to increase the duration before a long-lease address automatically reverts to the pool, and to reduce the cost (e.g., of enforcing leases), stress or necessity on returning addresses over a few generations. It is probably important to be generous on this number as the whole PIPv6 concept is to serve the goal of personal addressing and having an unlimited range (or as near as seems like that) is in the best interests of that goal. An additional *Sub-address* field is owner/user assignable and could be used by the owner for a range of sub-addresses (each for use in a separate transition scenario or developing wireless scenario). The assignable sub-address part may also be used as a *Node/Object ID* to facilitate its smooth participation in Mobile Ad hoc NETworks (MANETs), Vehicular Ad hoc NETworks (VANETs), and other WiNeMO types. The length of this field should be sufficiently large to allow addressing of hundreds of nodes/objects belonging to the same owner. For instance, allowance can be made for narrowcast addresses which may find use in corporations and various community and social groupings.

Key to any network-independent personal address is the prevention of duplicates, whether by accident or (malicious) design. A second issue is the eventual return of unused addresses or addresses whose use has ceased or become defunct. In the case of the PIPv6 address proposal, this could be achieved by a centralized purchased scheme through authorized address suppliers, each of which owns a portion/subset of this new IP address class' space and is identified by an optional *Address Supplier ID* field and/or by characteristics in the *Owner ID* field in the address. The selling of PIPv6 addresses within a 'renewable lease-based' system would also facilitate unused or defunct addresses being returned to the pool of available addresses [Ganchev, 2014b].

Obtaining PIPv6 addresses would be a commercial transaction. In addition, as there is no reason why owners might not engage in address trading, the commercial legal arrangements should allow for this, e.g. ownership should be legally verifiable and transferable without difficulty. Perhaps this responsibility would ultimately fall to an Internet Assigned Numbers Authority (IANA) / Internet Corporation for Assigned Names and Numbers (ICANN)-type organization. Address trading would also incentivize use or return of addresses. There would be privacy concerns with this permanent PIPv6 address employed by users for node/object identification and addressing, authentication, authorization and network access admission. These reflect on possible compromise of privacy related to the potential for tracking of, and gathering statistics about, a user/node/object as s/he/it moves through different locations. However, some of the existing mechanisms for privacy protection may still be used in this case, e.g., encrypting the traffic at different communication layers, use of temporary or changing "pseudonyms" as identifiers, etc., [Ganchev, 2014b].

There is also a need for this new PIPv6 address to be securely 'locked' to enable the user/node/object to be uniquely identified and authenticated during communication. This is a key attribute. It could be achieved by embedding the PIPv6 address into a X.509 public-key digital certificate. The ITU-T's X.509 authentication framework [ITU-T, 2012] defines a good model for strong secure authentication with a minimum number of exchanges. The authentication is performed through simple automatic exchange of X.509 digital certificates between communication parties (network nodes, objects, entities, etc.). It seems reasonable to employ the three-way option for mutual authentication, as it does not require the communication parties to have synchronized clocks. The exchange of certificates will enable trusted relationship and secure payment of (micro) transactions in the UCWW.

The extensions defined in version 3 of the X.509 standard (X.509v3) provide methods for associating additional attributes to carry information unique to the owner of the certificate. In particular, the *Subject Unique ID* field (Figure 8), which allows additional identities –e.g. e-mail address, Domain Name System (DNS) name, IP address, Uniform Resource Identifier (URI), etc.– to be bound to the owner, can accommodate the proposed PIPv6 address. This, however, must be clearly marked as a critical X.509v3 extension in order to be used in a general context [ITU-T, 2012]. Because the *Subject Unique ID* is definitively bound to the public key, all parts of it (including the PIPv6 address) can be verified by the corresponding certificate authority (CA).

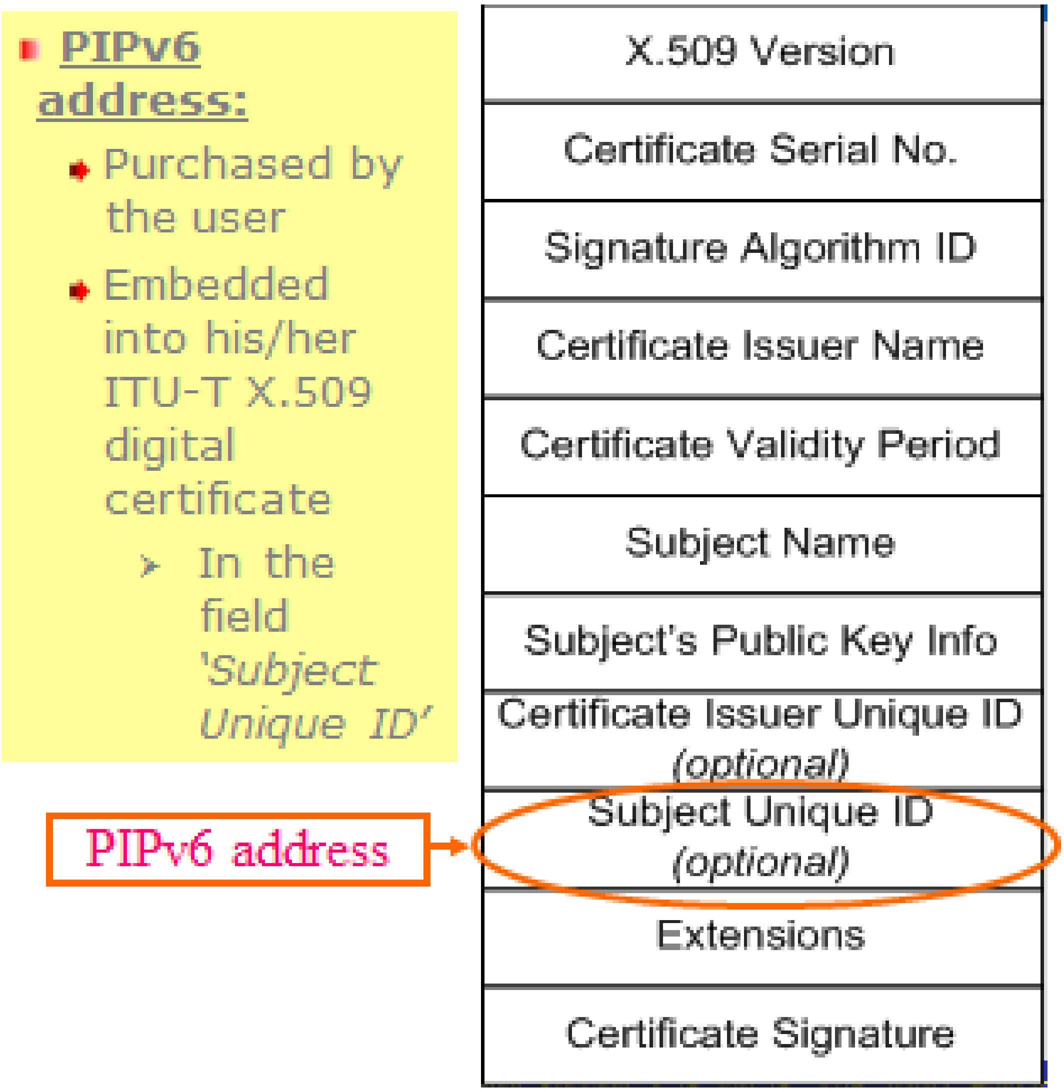

Fig. 8. The PIPv6 address, embedded in the user's X.509 digital certificate.

A generic WiNeMO communication scenario using the PIPv6 address is depicted in Figure 9. The scenario imagines a mobile node (object) seeking and finding a gateway (GTW) among or through those mobile nodes (MNs) available to it as relays either directly or through other mobile nodes in a WiNeMO network. The GTW is defined as an access point to connect directly to the Internet and through it – to a particular correspondent node (CN). First, a mutual authentication procedure is executed between the object and all supporting relay nodes in this scenario, including the GTW. This being successfully completed, the GTW decides to allow (or not) the object to use its Internet connection for a particular period of time. Then the GTW accepts the PIPv6 address supplied by the object and stores it in its Network Address Translation (NAT) table along with the corresponding IPv4 address to be used for this new Internet session for the duration of communication between the object and CN. Then GTW confirms to the object that it may start using the Internet for communication with CN. After that, following the standard NAT IPv6-to-IPv4 (NAT64) procedure, each IPv6 packet originating from the object will carry its PIPv6 address in the *Source Address* field. When this packet reaches the GTW, the PIPv6 address of the object (used only locally) will be translated into the public IPv4 address allocated to the GTW for global routing on the Internet. In other words, as the IP traffic passes from this WiNeMO to the Internet, the GTW translates 'on the fly' the source address in each packet from the PIPv6 address of the particular object engaged in communication to (one of) its own public IPv4 address(es). The reverse address translation is performed in the opposite direction of communication.

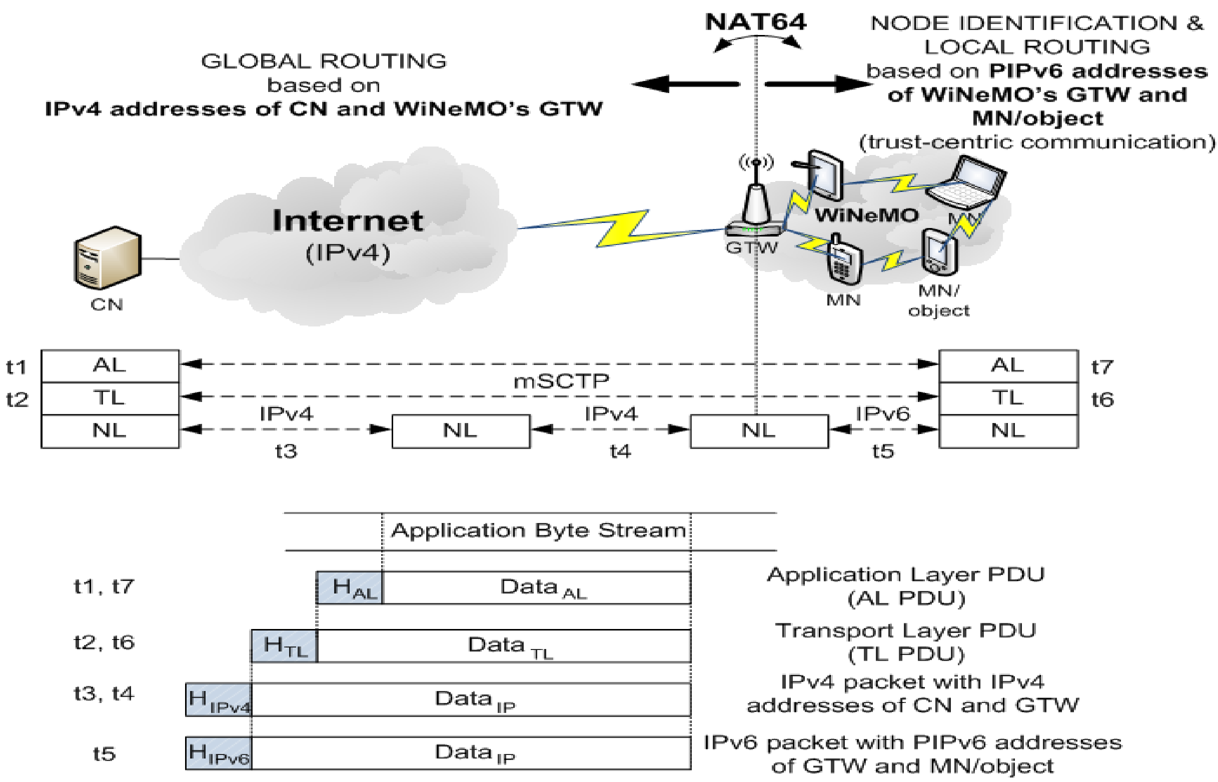

Fig. 9. A generic WiNeMO communication scenario using the PIPv6 address.

5. Smart CIM Card

This new type of card is proposed in [O'Droma, 2007] for the mobile users so as to behave as consumers in the UCWW with whatever mobile device chosen, i.e. to obtain and securely pay for services from any ANP/xSP, anywhere-anytime-anyhow, thus achieving advanced user mobility. With CBM and through their CIM card, consumer-users, wherever they are, would always appear as 'local users' to whatever networks they roam in. Implicit in being a 'local user' is that roaming charges will disappear! Through their universal X.509-based smart CIM cards, consumers would own their personal globally significant, network-independent PIPv6 address(es). By means of relevant CAs' public key infrastructures (PKIs), the validity of the certificates of all parties to a transaction may be mutually checked as required.

The CIM card could be developed by using the 'Java Card' technology [Oracle, 2015], which provides highly secure, market-proven, and widely deployed open-platform architecture for the rapid development and deployment of smart-card applications meeting the real-world requirements of secure system operations. The Java-based CIM card may typically be a plastic card containing an embedded chip. It will enable the client applications to run on a single virtual machine, in order to maintain the user profiles, credit card information, PIPv6 address(es), 3P-AAA data, X.509 certification, etc. As these applications will be required to communicate with each other using shared interface objects (SIO), a firewall must be defined for each application to provide application-level security. A sample CIM card architecture is presented in [Ganchev, 2007] and depicted on Figure 10.

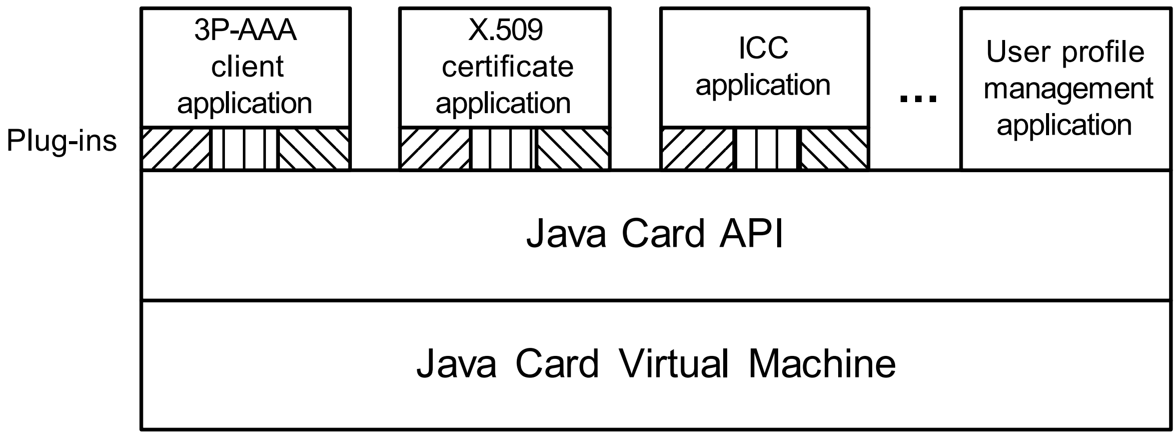

Fig. 10. A sample CIM card structure

The multi-application processor CIM card inserted in the mobile device will have one or more 3P-AAA application clients installed. This devolution of application programs and processing power to the CIM is a significant change from the legacy SIM card operations. Simultaneous accounts with different 3P-AAA-SPs will be possible. CIM cards, for instance, could come with generic 3P-AAA clients pre-installed and several free spaces on the card chip leasable to specific 3P-AAA-SPs for installing (and upgrading) their own specific 3P-AAA client, designated for that consumer.

Different other applications/applets could be deployed on the CIM card, e.g. an application for managing consumer’s X.509 certificates, an ICC application for facilitating the incoming call connection (ICC) service provision, an application for managing the user profile(s), etc. Additionally installed plug-ins will allow further personalization of each application, e.g. in the case of 3P-AAA the user may install a separate plug-in for each individual 3P-AAA-SP with whom s/he has an agreement for AAA and C&B of services (the use of each particular 3P-AAA-SP is activated depending on the current user context as specified in the user profile). The commercial relation between the CIM card issuer and the application provider is independent of the platform technology, ensuring a really open market space. For instance in the 3P-AAA case, the business agreements are between the CIM card issuer and the 3P-AAA application provider, where the issuer can specify the extent of its responsibility for the overall card security and the 3P-AAA application provider can assume its own responsibility for the secure 3P-AAA operation of its in-card business logic implemented in the loaded 3P-AAA client application [Ganchev, 2007].

The Java Card platform provides a secure execution environment with a firewall between different applications on the same card. Each application can encapsulate sensitive data and algorithms within objects, which have provable behavior and increased security. Further security enhancements, such as transaction atomicity and cryptographic classes, are also provided. In addition the dynamic download capability of the card ensures that applications can be securely managed, i.e. tamper-proof downloaded, installed, configured, updated, and removed after the card has been issued.

The Java Card Virtual Machine (JCVM) separates applications from the underlying hardware and operating system. “Split virtual machine” architecture is used: one part is executed on the user’s mobile device, preparing the code for execution in the other part of the virtual machine, on the card. The split JCVM design is intended to reduce the size of the applet image downloaded to the card and to minimize run-time memory requirements [Oracle, 2015]. A standardized API provides a uniform interface to applications and extension Java packages.

The approach towards the UCWW personalized and multi-serving client applications is seeking designs on a smart programmable handheld mobile devices to interoperate seamlessly with the CIM. Given the expected dominance (greater than 50%) of the Android operating system, the development of a client application working within a Google Android environment is primarily targeted [Ganchev, 2015a]. Consistent with this is also the integration of a SIMAlliance Open Mobile API specification [SIMAlliance, 2014] into the Android platform to enable mobile devices to communicate with secure elements in the CIM.

Developing the CIM card as a **virtual card** is another promising direction which deserves better attention in the future.

6. Conclusion

Technical foundations for an effective techno-business model which will make the evolution to a truly Always Best Connected and best Served (ABC&S) fifth generation (5G) wireless world possible have been addressed in this paper. The flaws in the current Subscriber-Based techno-business Model (SBM) for this evolution have been highlighted. An argument for a Consumer-Based techno-business Model (CBM) has been made and a view on the changes and innovations needing to be made to the network and protocol technological structure to achieve this has been put forward. This includes an infrastructural re-think on the way Authentication, Authorization and Accounting (AAA) service is supplied, proposing the creation of a third-party AAA service provider entity (3P-AAA-SP), together with a number of new technological supports requiring global standardization, including a novel personal IPv6 (PIPv6) address mechanism and a new smart Consumer Identity Module (CIM) card utilizing X.509v3 digital certificate security. The new globally significant, network-independent PIPv6 address will enable real number ownership and full ‘anywhere-anytime-anyhow’ portability. It has been proposed and envisaged that in future generations of wireless networks, nodes (objects) will have a unique PIPv6 address, which may serve also as a means of long-term node identity in the network.

While initially the CBM infrastructure may seem to threaten the establish large subscriber-based access network provider (ANP) market share, it has been argued rather that it would greatly open, and ease entry into, the ANP- and mobile service provider (xSP) markets for new entrants, for ‘disruptive’ and creative technologies, and stimulating competition and new and improved services, and all in a way that will be of benefit to all stakeholders.

Mobile users today need ever more choice and customization in the provision of (mobile) services with more flexible service delivery and an ability to move/migrate quickly to more competitive providers who can provide better price/performance options, a wider selection of service offerings, etc. The Ubiquitous Consumer Wireless World (UCWW) is a creative proposal requiring the implementation of the CBM model for wireless communications. This, in turn, requires the underpinning of some strategic standardization as highlighted in the paper. Once the new standardized elements are in place, the UCWW will begin to take shape and grow along an evolutionary path (as happened with, and in parallel with, the existing wireless world founded on the SBM model) yielding social, economic and policy benefits for users, ANPs (mobile operators), hardware and software suppliers, the full range of mobile service providers and new business entities, such as the cloud service providers. The novel 3P-AAA-SP concept seems especially attractive for the emerging inter-cloud service providers' business.

Acknowledgements

This work is funded and supported by the Telecommunications Research Centre (TRC), University of Limerick, Ireland and by the IT15-FMIIT-004 project of the Scientific Fund of the Plovdiv University "Paisii Hilendarski", Bulgaria.

The author wishes to specially thank the TRC Director, Dr. Máirtín O'Droma, for the fruitful discussions and inspiring thoughts on the subject through many years.

Bibliography

[Calhoun, 2003] P. Calhoun et al., "Diameter Base Protocol", IETF RFC 3588, 2003.

[Ganchev, 2007] I. Ganchev and M. O'Droma. "New personal IPv6 address scheme and universal CIM card for UCWW". Proc. 7th Int. Conf. on Intelligent Transport Systems Telecommunications (ITST 2007), Pp. 381-386. 6-8 June 2007, Sophia Antipolis, France. IEEE CN-07EX1765.

[Ganchev, 2014a] I. Ganchev, M. Curado, A. Kassler (Eds.): Wireless Networking for Moving Objects - Protocols, Architectures, Tools, Services and Applications, LNCS 8611, pp. 301. Springer International Publishing, Switzerland. September 2014. ISBN: 978-3-319-10833-9. DOI: 10.1007/978-3-319-10834-6.

[Ganchev, 2014b] I. Ganchev and M. O'Droma. 2014. "A New Techno-Business Model based on a Personal IPv6 Address for Wireless Networks of Moving Objects". In: Wireless Networking for Moving Objects - Protocols, Architectures, Tools, Services and Applications. I. Ganchev, M. Curado, A. Kassler (Eds.). Springer International Publishing, Switzerland. September. ISBN: 978-3-319-10833-9. Pp. 3-13. DOI: 10.1007/978-3-319-10834-6.

[Ganchev, 2015a] I. Ganchev, Z. Ji, M. O'Droma, C. Dai. 2015. "A CIM System for Use in the UCWW". Proc. of the 2015 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC 2015). Pp. 72-75, 17-19 September, Xi'an, China. ISBN 978-1-4673-9200-6/15. DOI 10.1109/CyberC.2015.63.

[Ganchev, 2015b] I. Ganchev, Z. Ji, M. O'Droma. 2015. "A Distributed Cloud-based Service Recommendation System". Proc. of the 2015 International Conference on Computing and Network Communications (CoCoNet'15). Pp. 212-215. 16-19 December, Trivandrum, India. ISBN: 978-1-4673-7309-8/15. DOI: 10.1109/CoCoNet.2015.7411189. Library of Congress: CFP15C74-USB.

[Ganchev, 2016a] I. Ganchev, Z. Ji, M. O'Droma. "The Creation of a Data Management Platform for Use in the UCWW". Proc. of 2016 SAI Computing Conference. Pp. 585-588. 13-15 July 2016. London, UK. ISBN: 978-1-4673-8460-5/16.

[Ganchev, 2016b] I. Ganchev, Z. Ji, M. O'Droma. "A Conceptual Framework for Building a Mobile Services' Recommendation Engine". Proc. of the IEEE International Conference 'Intelligent Systems' (IEEE IS 2016). Pp. x1-x5. 4-6 September 2016. Sofia, Bulgaria.

[ITU-T, 2004] ITU-T Recommendation Y.2001: General overview of NGN. In Series Y: Global Information Infrastructure, Internet Protocol Aspects and Next-Generation Networks (Next Generation Networks – Frameworks and Functional Architecture Models). Dec. 2004.

[ITU-T, 2008] ITU-T Draft Recommendation Q.3202.1 (Q.nacf.auth1), "Authentication Protocols based on EAP-AKA for Interworking among 3GPP, WiMax, and WLAN in NGN", 2008.

[ITU-T, 2012] ITU-T Recommendation X.509: Information technology – Open Systems Interconnection – The Directory: Public-key and attribute certificate frameworks. In Series X: Data Networks, Open System Communications and Security Directory. Oct. 2012.

[Jakab, 2009] J. Jakab. "New Charging and Billing Models and Mechanisms for the Ubiquitous Consumer Wireless World (UCWW)". Transfer Report (from Masters to PhD). University of Limerick, Ireland. 2009.

[Jakab, 2014] J. Jakab, I. Ganchev, M. O'Droma. "Third-party AAA and corresponding Charging and Billing of Services", Presented at the COST IC1304 3rd MCM on "Autonomous Control for a Reliable Internet of Services (ACROSS)". Larnaca, Cyprus, 23-24 October 2014.

[O'Droma, 2004a] M. O'Droma and I. Ganchev. "New Access Network Techno-Business Model for 4GWW". Proc. of 4th ANWIRE International Workshop on Wireless Internet and Reconfigurability (held in conjunction with the 3rd IFIP-TC6 Networking Conference), Pp. 75-81. 14 May 2004. Athens, Greece.

[O'Droma, 2004b] M. O'Droma and I. Ganchev. 2004. "Techno-Business Models for 4G" (invited paper), Proc. of the International Forum on 4th Generation Mobile Communications, Pp. 3.5.1-30, 20-21 May, King's College London, London.

[O'Droma, 2007] M. O'Droma and I. Ganchev. "Towards a Ubiquitous Consumer Wireless World". IEEE Wireless Communications, Feb. 2007, Pp. 2-13. ISSN: 1536-1284.

[O'Droma, 2008] M. O'Droma and I. Ganchev. "Strategic Innovations through NGN Standardisation for a Ubiquitous Consumer Wireless World". Proc. of the 1st ITU-T Kaleidoscope Academic Conference "Innovations in NGN". Pp. 135-142, 12-13 May 2008. Geneva, Switzerland. ISBN 92-61-12441-0. DOI 10.1109/KINGN.2008.4542259. (Best Paper Award Nomination)

[O'Droma, 2010] M. O'Droma and I. Ganchev. "The Creation of a Ubiquitous Consumer Wireless World through Strategic ITU-T Standardization" (invited paper). IEEE Communications Magazine, Vol. 48, Issue 10, Pp. 158-165. October 2010. ISSN: 0163-6804. DOI: 10.1109/MCOM.2010.5594691.

[Oracle, 2015] Specifications for the Java Card 3 Platform Version 3.0.5, Classic Edition. Oracle Release Notes. October 2015. https://docs.oracle.com/javacard/3.0.5/JCSRN.pdf

[SIMAlliance, 2014] "Open Mobile API specification: V2.05". SIMAlliance. February 2014. http://simalliance.org/wp-content/uploads/2015/03/SIMalliance_OpenMobileAPI2_05_release-Feb143.pdf

Author's Information

***Ivan Ganchev** – DipEng (summa cum laude), PhD, SMIEEE, ITU-T (Invited Expert), IJTMCC Regional Editor (Europe).*

TRC Deputy Director, University of Limerick, Limerick, Ireland & Associate Professor, Plovdiv University "Paisii Hilendarski", Bulgaria; e-mail: Ivan.Ganchev@ul.ie

Major Fields of Scientific Research: novel telecommunications paradigms, future networks and services, smart ubiquitous networking, context-aware networking, mobile cloud computing, Internet of Things (IoT), Internet of Services (IoS), Ambient Assisted Living (AAL), Enhanced Living Environments (ELE), trust management, Internet tomography, mHealth and mLearning ICT solutions.

METHOD FOR CALCULATING THE EQUIPMENT OF INFORMATION NETWORK BASED ON THE MESSAGE STREAM MODEL

Galina Gayvoronska, Maxim Solomitsky

***Abstract**: Usage results of an authors' unified method for calculating the equipment of information network subject to these networks' modern features are reported. This method is based on an authors' model of message streams circulating in convergent network and a model of the information networks' load. It's indicated that usage of the proposed method assures the 18…22% accuracy enhancement of the calculation results. All the reported results are confirmed by the model usage in activity of National Commission for the State Regulation of Communications and Informatization, the methods usage for real design of Ukrainian information networks by "Diprozvyazok" PJSC and the results evaluation by "Ukrtelecom" LLC (Odessa branch).*

***Keywords**: information network, load model, information stream, method for calculating the equipment, message stream, distribution function, service node.*

***ACM Classification Keywords**: C. Computer Systems Organization – C.2 COMPUTER-COMMUNICATION NETWORKS – C.2.1 Network Architecture and Design; G. Mathematics of Computing – G.3 PROBABILITY AND STATISTICS – Distribution functions, Queueing theory, Stochastic processes.*

"Things imagined by us don't exist in themselves as we imagine them and their interrelations are quite unlike as they are for us."

Peter D. Ouspensky

Introduction

While researching information networks (IN), the information conversion seems to be very important process, delivering nontrivial tasks, which solution demands an applying new approach and up-to-date mathematical methods. Process of information conversion when it's transferring across the IN is the subject of a teletraffic theory (TT) research. For a long time TT represented fundamental tool for telecommunication networks (TN) research. However, analysis of scientists research results of last decades [Leland et al., 1994], [Duffy et al., 1994], [Crovella, Bestravros, 1996], [Giroux, Ganti, 1999], [Borella, Brewster, 1998], [Oliveira et al., 2003], [Ilnickis, 2004], [Min, Ould-Khaoua, 2004], [Stallings, 2002] indicates the impossibility of existing TT tools' direct usage for the IN synthesis and design at the modern stage of infocommunications development since they don't match these networks features, aroused last years [Gayvoronska, 2013 (1)]. At that, there are no alternative solutions in most of known publications – there are certain groundworks, but they are particularly dedicated.

One of the factors, which demands review of common approach for the IN synthesis, is fundamental change of view on quality of service (QoS). Previously, all issues of the TN synthesis, design and development were solved in conditions where demand for these networks' services constantly and significantly outdistanced supply. At that the task was to derive the compromise between the QoS characteristics and cost of network design. Quality indicators were normalized by international or national standards, and user had no affect on this process. Such situation continued for decades, but now it has changed radically. A network user became the most important one. User defines necessary services and grade of their rendering quality, which is willing to be paid. A number of authors' works [Gayvoronska, Britsky, 2016], [Gayvoronska, 2015], [Gayvoronska et al., 2014], [Gayvoronska, 2013 (2)], [Gayvoronska et al., 2012] devoted to analysis of modern situation and users requirements to the IN.

Degree of the IN users satisfaction by rendered information-communications services (ICS) is determined by the QoS. According to the International Telecommunication Union definition, quality of service is totality of characteristics of ICS that bear on their ability to satisfy stated and implied needs of the services user [ITU-T, 2008]. It depends on a network and defines level of user satisfaction when the IN systems are highly loaded. Assurance of necessary quality is determined first by an adequacy of accounting the load characteristics and features of its distribution across the network when designing and operating the IN.

Recent analysis of the quality of real IN functioning pointed that methods used for calculating the network equipment give incorrect values of required number of service nodes (SN) [Gannitsky, 2012]. Therefore, in order to enhance the QoS it's necessary first to take into account features of the load generated by different information types, co-transmitted across the IN. Thereto it's necessary to have maximally accurate information about the network load and characteristics of information streams (IS), generating this load. The load parameters and the IS circulating in the TN are subjects of the research of large number of scientists, both in our country and abroad: Kharkevich A.D., Basharin H.P., Roginsky V.M., Lazarev V.G., Livshits B.C., Ionin G.L., Sedol J.J., Ločmelis J.J., Prokofiev V.A., Shneps M.A., Kornyshev Y.N., Duz' V.I., Chumak N.A., Popovsky V.V., Bezruk V.M. and many others. However, due to the fact that the modern IN can't be directly referred to any of networks models used in existing calculation methods, and taking into account essential changes in the networks' situation on the whole, results of these researches should be clarified and corrected.

The IN design based on reliable initial data and usage of calculation methods, which take into account characteristics of the IS and distribution laws of its basic parameters, provides tangible savings on cost of the IN development and functioning, improves the QoS and network bandwidth on the whole. Consequently, while designing and operating the IN, it's necessary to derive the relationship between the load and network capacity (number of IN equipment) in order to provide the specified QoS. The IN information streams can be transmitted with the usage of all possible data transfer modes (DTM): channel, packet, frame and cell. The transfer mode defines a combination of the methods for switching, multiplexing and packing the information while transmitting across the network. All the DTM are standardized by the ITU and correspond to the first three levels of the open systems interconnection model [Gayvoronska, 2000], and differ in the used multiplexing method: positional (PM) or label (LM), as well as in algorithms for establishing, maintaining and disengaging the connections (physical or logical) [Gayvoronska, 2007 (1)]. The modern IN use all possible DTM simultaneously, so while calculating the network equipment it's necessary to take into account not only the switching method, but a whole totality of parameters, united by the concept of the information transfer mode. Such situation demands a correction of used method for calculating the IN equipment in such a way that the calculations become adequate to the IS model. It means that a selected mathematical model matches a processes in real systems and networks of the information distribution.

Analysis of Existing Methods for Calculating the Information Network Equipment

As it's known from the classical teletraffic theory [Stormer et al., 1971], [Eldin, Lind, 1972], method for calculating the network equipment depends on a structure of service nodes, service discipline and model of information streams, circulating in this network.

Basic works in the queueing theory [Kendall, Stuart, 1961], [Kleinrock, 1975], [Kleinrock, 1964] involve an adopted the five-part descriptor A/B/n/M/N that denotes indicated relationship as follows:

A – the interarrival time distribution;

B – the service time distribution;

n – the number of servers of the queueing system;

M – the system's storage capacity (number of positions for waiting the service start);

N – the size of the customer population (number of load sources).

For a complete specification of a queueing system more information is required and in these cases ITU adopted the six-part notation [ITU, 2005]

A/B/n/M/N/X

where A and B describe arrival process and service time distribution, respectively, of n-server queueing system, M – the total capacity of the system, N – the population size of customers, X – the queueing discipline.

From the first works of the TT founder Erlang A.K. his first and second formulae describing determination of the losses probability in a full-accessible trunk of loss system and the waiting probability (implicit loss) of queuing system, respectively, traditionally involve well known expressions representing a little unity inherent the corresponding mathematical models [Gayvoronska, 2007 (2)]. Works [Sergeev, 1983], [Sergeev, Chekmareva, 1988] show that such homological unity exists and propose a number of formulae with following notations:

E – the threshold value of the service losses;

λ – the rate of the requests (corresponding stream parameter);

$'$ – the loss system indication;

$''$ – the queuing system indication;

ν – the SN number;

^ – indication of the system with re-calls.

Traditional Erlang's first and second formulae

$$P(e<E)=\frac{\lambda^{\nu}/\nu!}{\sum_{i=0}^{\nu}\frac{\lambda_i}{i!}},\ P(e<E)=\frac{1\Big/\left(\frac{\lambda}{\nu}\right)}{1\Big/E+\frac{\lambda}{\nu}\Big/\left(1-\frac{\lambda}{\nu}\right)} \tag{1, 2}$$

can be presented in a generalized form:

for the loss systems

$$E'=\left[1+\sum_{i=0}^{\nu-1}\prod_{j=0}^{i}\frac{\nu-j}{\lambda'}\right]^{-1}, \tag{3}$$

for the queuing systems

$$E''=\left[1+\left(1-\frac{\lambda''}{\nu}\right)\sum_{i=0}^{\nu-1}\prod_{j=0}^{i}\frac{\nu-j}{\lambda''}\right]^{-1}. \tag{4}$$

In so doing we suppose that in case of generally distributed IS the service rate is equal to one.

Queuing system with the limited number M of positions for waiting

$$E''_M = \left[1 + \frac{1 - \frac{\lambda^M}{\nu}}{1 - \left(\frac{\lambda}{\nu} \right)^{M+1}} \sum_{i=0}^{\nu-j} \prod_{j=0}^{i} \frac{\nu - j}{\lambda''} \right]^{-1} . \tag{5}$$

Joint (loss and queuing) service system

$$E''' = \left[1 + \left(1 - \frac{\lambda'''}{\nu} \right) \sum_{i=0}^{\nu-1} \prod_{j=0}^{i} \frac{\nu - j}{\lambda' + \lambda''} \right]^{-1} , \; E'''_M = \left[1 + \frac{1 - \frac{\lambda''}{\nu}}{1 - \left(\frac{\lambda''}{\nu} \right)^{M+1}} \sum_{i=0}^{\nu-1} \prod_{j=0}^{i} \frac{\nu - j}{\lambda' + \lambda''} \right]^{-1} . \tag{6}$$

Loss system, subject to a busyness of the users (degenerate case – only a part of the requests occupy free SN)

$$E'_p = \left[1 + \sum_{i=0}^{\nu-j} \prod_{j=0}^{i} \frac{\nu - j}{\lambda'(1 - p_b)} \right]^{-1} , \tag{7}$$

where p_b – probability of the user's busyness.

System with repeated requests at the isolated trunk which consists of SN

$$E^{\wedge} = \left[1 + \sum_{i=0}^{\nu-j} \prod_{j=0}^{i} \frac{(\nu - j)/\lambda^{\wedge}}{\lambda'(1 - \alpha_d s)} \right]^{-1} , \tag{8}$$

where α_d – user's insistence rate;

s – number of the repeated requests' sources.

System with repeated requests, subject to connection time τ_b, insistence α_p and probability of the user's busyness (the pre-loading system)

$$E^{\wedge} = \left[1 + \sum_{i=0}^{\nu-j} \prod_{j=0}^{i} \lambda^{\wedge} \frac{\frac{\nu - j}{\tau_b + (1 - p_b)}}{1 - \alpha_d s - (1 - s)\alpha_p p_b} \right]^{-1} . \tag{9}$$

System with arbitrary number of arbitrary variants of repeated requests' generating

$$E^{\wedge} = \left[1 + \sum_{i=0}^{\nu-j} \prod_{j=0}^{i} \frac{\dfrac{\nu - j}{\lambda(1-p)\ldots}}{1-(x+y+z+\ldots)} \right]^{-1}, \tag{10}$$

$$x = \alpha_z E^{\wedge} \alpha_Q Q;\ \ y = \alpha_d E^{\wedge}(1-Q)p;\ \ z = (1-E^{\wedge}) \cdot \alpha_p p.$$

Joint service system with repeated requests

$$E^{\wedge} = \left[1 + \left(1 - \frac{\lambda''}{\nu}\right) \sum_{i=0}^{\nu-1} \prod_{j=0}^{i} \frac{\nu - j}{\lambda' + \lambda'' / \left(1 - \alpha E^{\wedge}\right)} \right]^{-1}, \tag{11}$$

where α – the rate of the primary requests' source.

Modification of Method for Calculating the Network Equipment, Subject to Probabilistic-temporal Structure of Call Streams in Modern Information Networks

Analysis of the given expressions made in [Gayvoronska, Somsikov, 2008] revealed a unity in the calculation formulae. According to this author offered unified expression generalizing all variants of researched service systems and proved its adequacy for determination of losses probability in the service system

$$E = \left(1 + K \sum_{i=0}^{\nu-1} \prod_{j=0}^{i} \frac{\nu - j}{L} \right)^{-1}. \tag{12}$$

That allowed a unification of a method for calculating the network equipment. Depending on the queueing system and discipline only K and L coefficients in (12) should be modified. However the expression doesn't affect third component of the methods for calculating the network resources – a probabilistic-temporal structure of the streams, circulating in the network and arriving for the service. Parameters of the requirements stream have recently changed, so they should be clarified. Application of the calculation method which doesn't match real network streams (their structure) reduces the accuracy of its result [Gayvoronska, 1997]. Underestimated number of SN makes worse the QoS and causes the losses increase, while its overestimated value reduces the network usage efficiency. Therefore further generalization was a unification of the method by introducing the coefficient reflecting a probabilistic-temporal structure of the stream.

The TT existing methods allow calculating only the particular network fragments while for the network as a whole these methods can be applied only like an approximate. Basic formulae for calculating the network equipment, colligating message losses, load and trunks capacity, were obtained under the assumption of Poisson nature of arriving call stream. Analysis of the results of real call streams' measurements in the networks showed that their characteristics deviated significantly from the Poisson distribution model (so-called simplest stream with an infinite number of load sources). Significance of these deviations can be observed by the value of divergence between the moments of the second and higher degrees or the value of divergence between the coefficients of variation, asymmetry and excess [Gayvoronska, 1999]. Due to the fact that the Poisson stream doesn't always match real streams, in the case of the existing information networks it's necessary to select some other distributions allowing modify a shape of the probability density curve loosely enough and provide therefore safe matching the measurements. There are quite a large number of mathematical models describing the characteristics of call streams. However, there is a matter how much do real characteristics of modern networks' call streams match the mathematical models developed up to 90s. And in case of presence of the differences it's necessary to estimate their impact on the modern IN functioning.

In order to solve this problem goodness of fit of the arrival process in real networks at arbitrary time to normal, exponential, log-normal, Poisson, Pareto, Rayleigh, Weibull and other distributions was verified by the Pearson criterion, cf. [Gayvoronska, Gannitsky, 2007], [Gannitsky, 2009]. Real measurements of the researched random variable were obtained by a processing the statistics of call streams of “Ukrtelecom” LLC (Odessa branch) for the 3,5 years (2003, 2004, 2005 and 2006). Total amount of processed data is 100 GB (parent population consists of 711608646 elements in the database). The real arrival process doesn’t match any of previously used models, the call streams are unordinary with aftereffect and can be represented by empirical parameter characterizing the real stream’s nature on the basis of the log-normal distribution, cf. [Gannitsky, 2012]

$$\lambda = \frac{de^{\mu+\delta^2/2}}{3600}, \tag{13}$$

where d – the predicted value of the call stream’s parameter;

μ – the scale parameter of the log-normal distribution;

δ – the shape parameter of the log-normal distribution.

Subject to this, calculation coefficients in (12) for the loss system

$$K = 1,\ L = \frac{de^{\mu+\delta^2/2}}{3600}, \tag{14}$$

for the queuing system with the limited number M of positions for waiting

$$K = \frac{1 - \dfrac{de^{\mu+\delta^2/2}}{3600} \Big/ \nu}{1 - \left(\dfrac{de^{\mu+\delta^2/2}}{3600} \Big/ \nu \right)^{M+1}}, \; L = \frac{de^{\mu+\delta^2/2}}{3600}. \tag{15}$$

For the other service systems the K and L coefficients are being modified similarly. According to this the unified method for calculating the network equipment, adaptable to the service system and discipline, was further improved by an implementation of the coefficients reflecting the real call stream's nature. So the method's usage can be expanded to the data communication network with the label multiplexing and packet switching. That makes it an adaptable to the switching method and information transfer mode, cf. [Gannitsky, 2012].

Justification of the Message Stream Model Usage for Calculating the Equipment of Information Network

While researching the modern information networks it makes sense to consider the message streams instead the call streams, since a call is only a requirement for the network service in order to transmit the message, and a message – it is an information, transformed into the electromagnetic signals, that should be transferred across the network [Gayvoronska, 2000]. In this connection it isn't enough to consider only the need of the data transfer: it's necessary to know what kind of the information (both from the quantitative and qualitative point of view), what amount of information should be transferred and how, as well as to take into account a number of other parameters of the transmitted information. This is important in terms of the information transformation in time and space. The state-of-the-art analysis reveals that there is still no general theory for the distribution and calculation the quantitative and qualitative indicators of the information streams in the IN [Davydov et al., 1977], [Gayvoronska, 1998], [Gayvoronska, Kalnev, 2001], [Gayvoronska, 2007 (3)]. Thus, in order to represent the IS it's expedient to use the model of message stream, circulating in the network, instead of the call stream model, cf. [Gayvoronska, Solomitsky, 2012 (1)], [Solomitsky, 2013], [Gayvoronska, Solomitsky, 2012 (2)]. Further we denote the message stream as MS to avoid confusion with the call stream (requirements for the service).

Each information message is characterized by a parameters reflecting the required quality of their service by the network e.g. maximum permissible: probability of symbol corruption, mean delay and its variance.

Model of the information message transmitted in the streams

$$\overline{\mathbf{u}_{\vartheta r}^{\Theta}} = (\vartheta, r, \Theta, t),$$

where the information types ϑ depend on the information class ς and together with the network functioning technology define different requirements to the indicators of the quality of messages transfer, priority levels r determine the information urgency, amount of information Θ is determined by the ϑ and ς.

The message stream definition is related to the determination of: number of the messages and amount of the information in the MS, time while the messages are in the network and value of the information in the MS.

Accordingly, the MS analytical model in a generalized form is the set of the following functions.

1. Distribution function of the messages number in the stream at any time.
2. Distribution function of the information content in the stream as a whole.
3. Distribution function of the time while each message of the common sequence (the stream) are in the network.
4. Function of the value of the transmitted information.

Moreover the MS characteristics are affected by the external and intranetwork hampering factors (disturbances, glitches, etc.) arising during the MS transferring across the network in the form of errors and equipment failures which are stochastic too. Thus the MS mathematical model should include models or distribution functions specifying the probabilities of the corresponding stochastic variables E^t.

In general terms, form of the MS model and corresponding certain functions is determined at a concrete time moment at the IN concrete point, which is defined by the vector $x(v, a_k, a_l)$ [Gayvoronska, Solomitsky, 2014],

where v – the information transfer channel or the SN;

a_k, a_l – pair of the nodes between which the researched stream is circulating.

The messages number in the MS in addition to a time depends on the number of the information sources n, the channel bandwidth υ and parameter Ξ characterizing the information receiver and specifying according to the Erlang classification the system structure and the service discipline of the node-information receiver. Form of the distribution function $\Lambda^t(n, \upsilon, \Xi)$ of the messages number in the stream depends on the place in the network, where it's analyzed, which is defined by the vector x.

In order to define the MS types and amount of the information, we used a concept of the user information message (UIM) representing a finite data sequence, formed for the network transmission and having a complete sense content. The UIM is transferred across the network in the form of the data transfer units (DTU): packets, datagrams, frames, cells. There are several types of the MS, which differ in the ratio ξ of the number of elements contained in the information part of the DTU (m_e) and UIM (m_c):

- simple stream, where $m_e = m_c$ and $\xi = m_e / m_c = 1$;
- thinned stream, where $m_e < m_c$ and $\xi = m_e / m_c < 1$;
- complex stream, where $m_e > m_c$ and $\xi = m_e / m_c > 1$.

The distribution function $\mathrm{T}_\rho(c, \Theta, r, q)$ of the time while messages are in the network depends on the number of the messages c, the amount of the information in one message Θ, the integral quality parameter q, determining the delay of the signal transfer, the error rate, the rate of failures and the preset priority r.

The priority is set by the network, so it's considered as the objective characteristic of the information importance/value for the network operator. The information value, as distinct from the priority, defines the subjective information importance to the end users.

The function of the information value for the user is defined as $Q(A, B, W_A, W_B, \varepsilon, \chi, \psi)$,

where A, B – a pair of the corresponding users;

W_A, W_B – the information importance for the user *A* and *B*, respectively;

ε – the timeliness of the information transfer;

χ – preservation of the emotional nuance of the transmitted information;

ψ – preservation of the intonational nuance of the transmitted information.

Generalized model of the message stream is a set of the above expressions

$$MS = \{\, x(v, a_k, a_l), \Lambda^t(n, \upsilon, \Xi), \Delta^t(c, \Theta), \mathrm{T}_\rho(c, \Theta, r, q), \mathrm{E}^t, Q(A, B, W_A, W_B, \varepsilon, \chi, \psi) \,\},$$
$$v \in V, V = \{v_1, \ldots, v_b\}, a_k, a_l, A, B \in N, r \in R, R = \{r_1, \ldots, r_h\}. \tag{16}$$

At the same time in the most general terms the information transfer process in the IN can be represented as a streams with the intensity $\lambda, \lambda = \lambda(\hbar)$, where in the general case $\hbar$ is a vector whose components are defined by the control algorithms of exchange, multiplexing and switching in the IN as well as by the network structure and its components' structure.

In order to estimate the network capacity for the equipment-assisted QoS it's necessary to develop a model of the network load of the message streams.

Model Development of the Load of the Message Streams in Information Network

The load model of the IN message streams is based on the proposed MS model and takes into account all the possible data transfer modes. The IN load rate is estimated as the sum total of the information content of the stream (label multiplexing) and the time while messages are in the network (positional multiplexing).

The time while message is in the network at the PM for the each type and priority messages consists of the waiting time and service time [Solomitsky, 2014 (1)]

$$\rho_\wp = \omega_\wp + \bar{\tau}_\wp , \tag{17}$$

where $\omega_\wp$ – the mean value of the waiting time;

$\bar{\tau}_\wp$ – the mean value of the service time.

For the estimation of the mean value of the waiting time and the probability of waiting at the mean service time $\bar{\tau}_\wp$ it's expedient to use approximate formulae, cf. [Kramer, 1975] and [Langenbach-Belz, 1972]

$$\omega_\wp = \overline{\tau_\wp} \frac{A_\wp}{2(1-A_\wp)}\left(\kappa_{A\wp}^2 + \kappa_{H\wp}^2\right) g\left(A_\wp, \kappa_{A\wp}^2, \kappa_{H\wp}^2\right),$$

where $$g\left(A_\wp, \kappa_{A\wp}^2, \kappa_{H\wp}^2\right) = \begin{cases} \exp\left\{-\dfrac{2(1-A_\wp)}{3A_\wp}\dfrac{\left(1-\kappa_{A\wp}^2\right)^2}{\kappa_{A\wp}^2+\kappa_{H\wp}^2}\right\}, & \kappa_{A\wp} < 1 \\ \exp\left\{-(1-A_\wp)\dfrac{\kappa_{A\wp}^2-1}{\kappa_{A\wp}^2+4\kappa_{H\wp}^2}\right\}, & \kappa_{A\wp} \ge 1; \end{cases}$$

$$A_\wp = \frac{\lambda_\wp}{\mu_\wp},\ \lambda_\wp = \lambda_{0\wp} + \sum_{j=1}^{N} \lambda_{\wp j} q_{ji},\ i = 1,2,\ldots,N,\ \mu_\wp = \bar{\tau}_\wp^{-1};$$

q_{ji} – transition probability for messages circulating from node *j* to a node *i*;

node 0 represents the outside world of the researched IN;

$\kappa_{A\wp}$ – variation coefficient of the arrival process;

$\kappa_{H\wp}$ – variation coefficient of the service process,

$$W_\wp = A_\wp + \left(\kappa_{A\wp}^2 - 1\right) A_\wp (1 - A_\wp) \begin{cases} \dfrac{1 + \kappa_{A\wp}^2 + A_\wp \kappa_{H\wp}^2}{1 + A_\wp(\kappa_{H\wp}^2 - 1) + A_\wp^2(4\kappa_{A\wp}^2 + \kappa_{H\wp}^2)}, & \kappa_{A\wp}^2 < 1 \\ \dfrac{4A_\wp}{\kappa_{A\wp}^2 + A_\wp^2(4\kappa_{A\wp}^2 + \kappa_{H\wp}^2)}, & \kappa_{A\wp}^2 \ge 1. \end{cases}$$

The information content $\Delta^t(c,\Theta)$ in the MS depends on the messages number c and amount of information in the message.

Formulae of the information content for the three above MS modifications, cf. [Solomitsky, 2014 (2)]

$$P(U_e^1 = c) = \frac{\lambda_{\Sigma\ell}^c}{e^{\lambda_{\Sigma\ell}} c!},\ c = 0, 1, 2, \ldots,\ M[U_e^1] = D[U_e^1] = \lambda_{\Sigma\ell};$$

$$P(U_e^2 = c) = 1 - \frac{1}{e^{\lambda_{\Sigma\ell} c}} \sum_{i=0}^{\xi-1} \frac{(\lambda_{\Sigma\ell} c)^i}{i!},\ M[U_e^2] = \xi / \lambda_{\Sigma\ell},\ D[U_e^2] = \xi / \lambda_{\Sigma\ell}^{\ 2};$$

$$P(U_e^3 = c) = \begin{cases} \dfrac{1}{e^{\lambda_{\Sigma\ell}}},\ c = 0 & M[U_e^3] = \lambda_{\Sigma\ell}/\xi\ , \\ \dfrac{1}{e^{\lambda_{\Sigma\ell}}} \displaystyle\sum_{i=1}^{c} C_{c-1}^{i-1} \frac{(\lambda_{\Sigma\ell}\xi)^i (1-\xi)^{c-i}}{i!},\ c = 1, 2, \ldots, & D[U_e^3] = \lambda_{\Sigma\ell}(2-\xi)/\xi^2\ , \end{cases}$$

where U_e – the arrival process of DTU, it depends on a change of amount of information in the UIM and DTU and is determined by a rate of the m_e and m_c values;

$$\lambda_{\Sigma\ell} = \lambda_\ell + s\beta,\ \lambda_\ell = \alpha n,\ s = \frac{\dfrac{1}{\nu}\displaystyle\sum_{i=0}^{\nu-1}\prod_{j=0}^{i}\frac{\nu-j}{\lambda_\ell} + \left(1 - \frac{\lambda_\ell}{\nu}\right)\frac{d}{d\lambda_\ell}\left(\sum_{i=0}^{\nu-1}\prod_{j=0}^{i}\frac{\nu-j}{\lambda_\ell}\right)}{\left(1 + \left(1 - \dfrac{\lambda_\ell}{\nu}\right)\displaystyle\sum_{i=0}^{\nu-1}\prod_{j=0}^{i}\frac{\nu-j}{\lambda_\ell}\right)^2},$$

α, β – rate of the primary and repeated requests' sources, respectively.

The mean value of amount of information in the messages of the stream at the LM

$$\pi_\ell = \begin{cases} M[U_e^1] = \lambda_{\Sigma\ell},\ \xi = 1 \\ M[U_e^2] = \xi/\lambda_{\Sigma\ell},\ \xi < 1 \\ M[U_e^3] = \lambda_{\Sigma\ell}/\xi,\ \xi > 1 \end{cases} . \quad (18)$$

We determine the load intensities at the PM and LM for the ϑ^{th} type and r^{th} priority information between nodes k and l by a matrixes $[\lambda_{\vartheta r}^{\ell kl}]$ and $[\lambda_{\vartheta r}^{\wp kl}]$, respectively. So we derived simplified expression for estimating the information type- and priority-averaged time while messages are in the network

$$\rho^{\wp} = \lambda^{-\wp} \sum_{\substack{i \in I_\wp \\ j \in J_\wp}} \lambda_{\vartheta r}^{\wp ij} \rho_{\wp ij},\ \lambda^{\wp} = \sum_{\substack{i \in I_\wp \\ j \in J_\wp}} \lambda_{\vartheta r}^{\wp ij}\ ,$$

and amount of the information in them

$$\pi^{\ell} = \lambda^{-\ell} \sum_{\substack{i \in I_\ell \\ j \in J_\ell}} \lambda_{\vartheta r}^{\ell ij} \pi_{\ell ij} \Theta_{\vartheta}^{\ell ij}, \; \lambda^{\ell} = \sum_{\substack{i \in I_\ell \\ j \in J_\ell}} \lambda_{\vartheta r}^{\ell ij},$$

where $I_\wp, I_\ell, J_\wp, J_\ell$ – set of the information types and priorities at the PM and LM usage, respectively.

The resultant IN load for the concrete DTM, respectively

$$\Lambda^{\wp} = \rho^{\wp} \sum_{r \in R} \sum_{\vartheta \in \Upsilon} MS, \; \Lambda^{\ell} = \pi^{\ell} \sum_{r \in R} \sum_{\vartheta \in \Upsilon} MS. \quad (19), (20)$$

As a result, the unified representation of the load

$$\Lambda^{\Phi} = \lambda^{-\Phi} \sum_{\substack{i \in I_\Phi \\ j \in J_\Phi}} \lambda_{\vartheta r}^{\Phi ij} \rho_{\Phi ij} \pi_{\Phi ij} \Theta_{\vartheta}^{\Phi ij}. \quad (21)$$

This authors load model is determined subject to: a sum of the information messages $\overline{\mathbf{u}_{\vartheta r}^{\Theta}}$ transferring in the streams, the time $\rho^{\wp}$ while the messages are in the network, amount of the information π^{ℓ} in these messages, distribution law of the DTU number in the UIM, priority class at the information transfer, the information value and other parameters which can be specified by the designer.

Modification of the Method for Calculating the Equipment of Information Network Based on the Message Stream Model

The MS load model of the information network is used for evaluation of the parameter λ determining values of the coefficients K and L from (12). So we provide the usage adequacy of the systematized unified method for calculating the network equipment for the IN. Possible variants of service systems used in the researched IN and corresponding modifications of the unified method for calculating the network equipment are given below.

For the loss system (cf. Fig. 1) $K = 1; L = \Lambda^{\wp}$, where the load intensity for the PM is $\Lambda^{\wp} = \rho^{\wp} \sum_{r \in R} \sum_{\vartheta \in \Upsilon} MS$

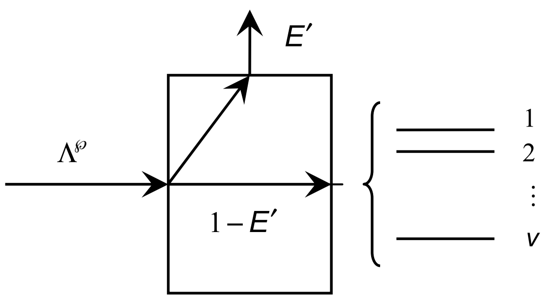

Fig. 1. The loss system

An example of such a system is a service node for voice messages hardwired e.g. by a public switched telephone network (PSTN) or a mobile communication network. The SN is a digital switching system (DSS) of the arbitrary network fragment (cf. Fig. 2) serving the MS load at the channel data transfer mode.

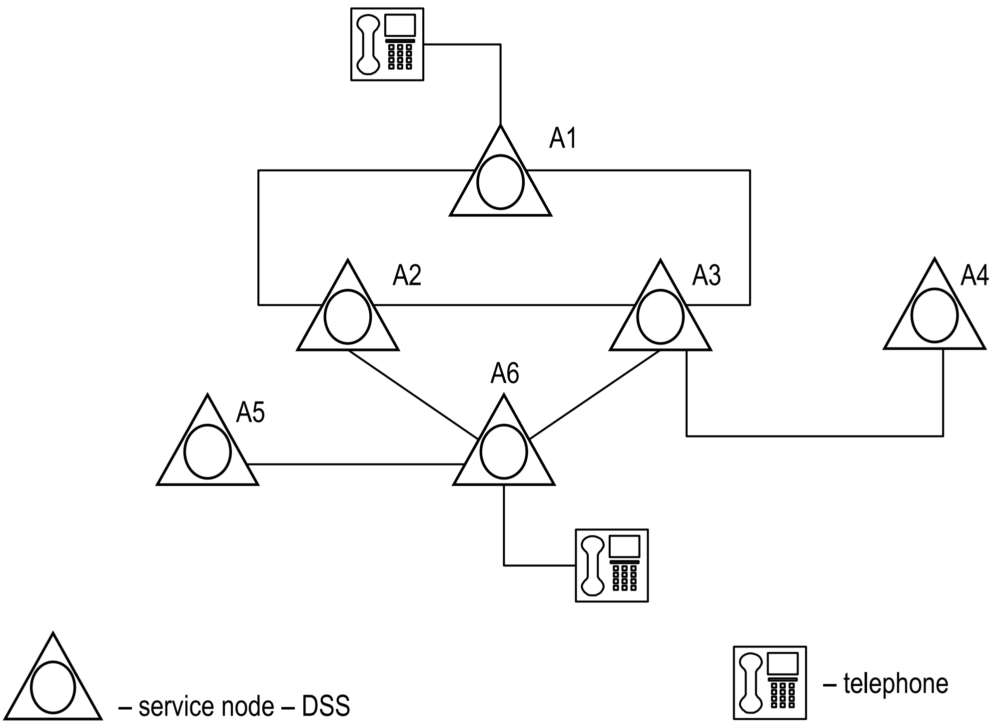

Fig. 2. The PSTN fragment (channel DTM)

For the queuing system with the unrestricted queue (cf. Fig. 3a) $K = 1 - \Lambda^{\ell}/v\,; L = \Lambda^{\ell}$, where the resultant load for the LM is $\Lambda^{\ell} = \pi^{\ell}\sum_{r\in R}\sum_{\vartheta\in \Upsilon} MS$.

For the queuing system with the restricted queue (M of positions for waiting) $K = \dfrac{1-\Lambda^{\ell}/v}{1-\left(\Lambda^{\ell}/v\right)^{M+1}}; L = \Lambda^{\ell}$ (cf. Fig. 3b). An example of such a system is a SN of any data communication network (DCN), cf. Fig. 4.

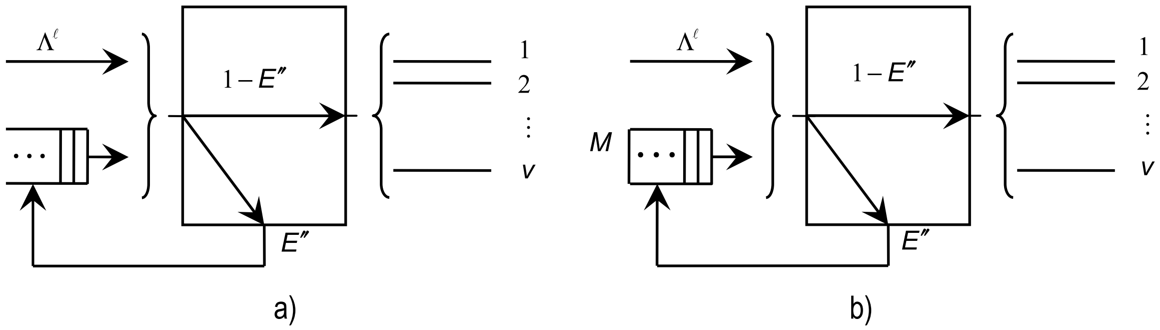

Fig. 3. The queuing system

a) with the unrestricted queue

b) with the restricted queue

The SN is a central switch of the example network fragment (cf. Fig. 4) serving the MS load at the packet DTM.

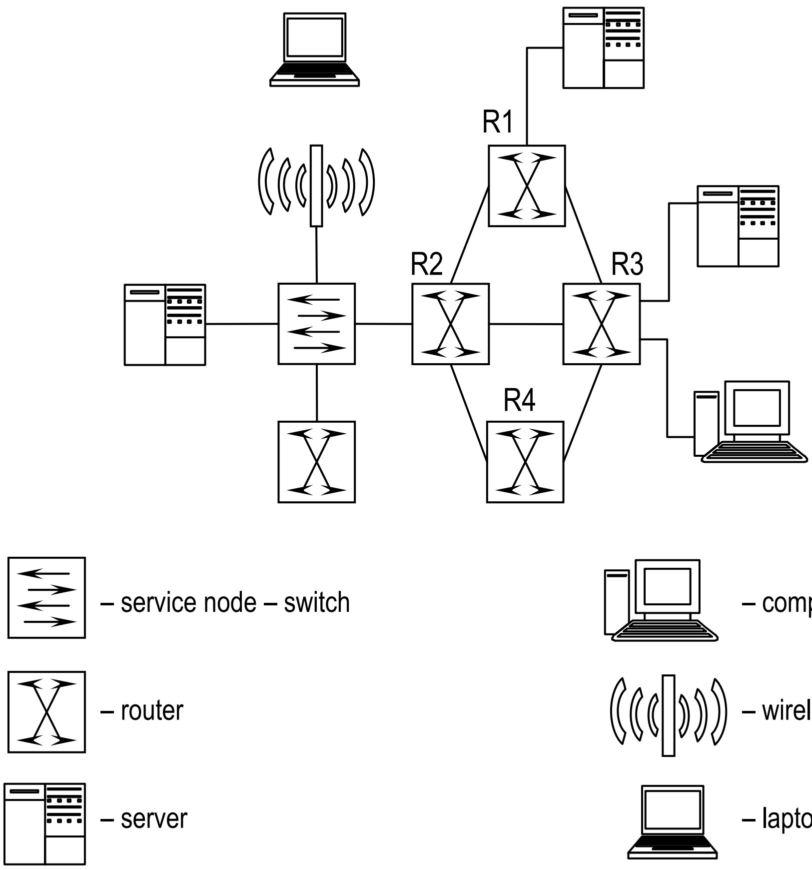

Fig. 4. The DCN fragment

For the joint system $K = \dfrac{1-\Lambda^{\ell}/v}{1-\left(\Lambda^{\ell}/v\right)^{M+1}}$; $L = \Lambda^{\Phi}$, where the resultant load $\Lambda^{\Phi} = \lambda^{-\Phi} \sum\limits_{\substack{i \in I_{\Phi} \\ j \in J_{\Phi}}} \lambda_{\vartheta r}^{\Phi ij} \rho_{\Phi ij} \pi_{\Phi ij} \Theta_{\vartheta}^{\Phi ij}$ includes simultaneous usage of all the DTM (cf. Fig. 5). An example of such a system is a node for servicing the information streams both with losses and queuing (waiting probability – implicit loss).

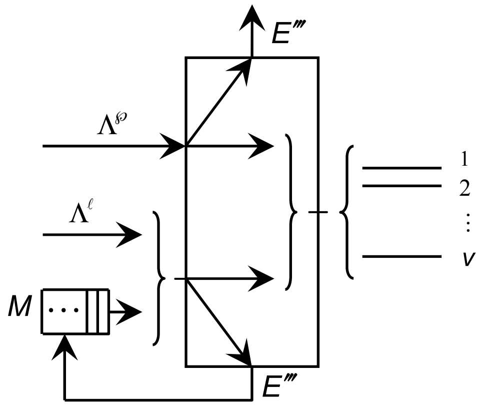

Fig. 5. The joint (loss and queuing)

The SN is an intelligent multifunctional switch of the example network fragment (cf. Fig. 6) serving the MS load of both label and positional multiplexing (packet (frame, cell) and channel DTM, respectively).

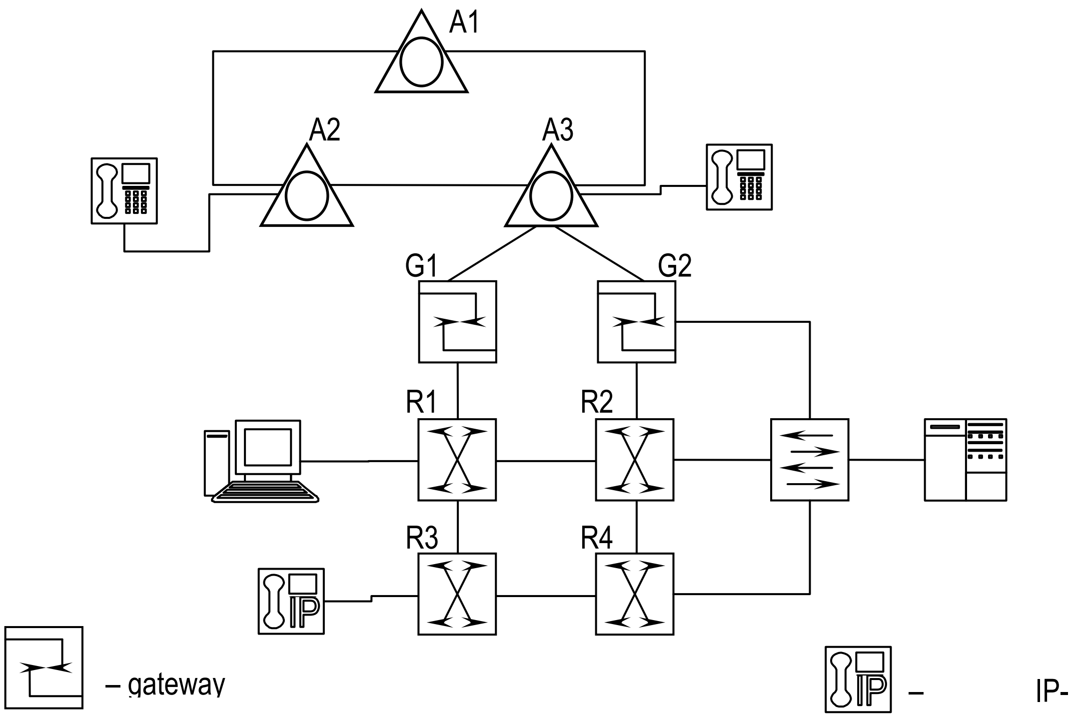

Fig. 6. Convergent network fragment aggregating the PSTN, DCN and VoIP segments

For the loss system, subject to a busyness of the users $K = 1; L = \Lambda^{\wp}(1 - p_b)$, cf. Fig. 7.

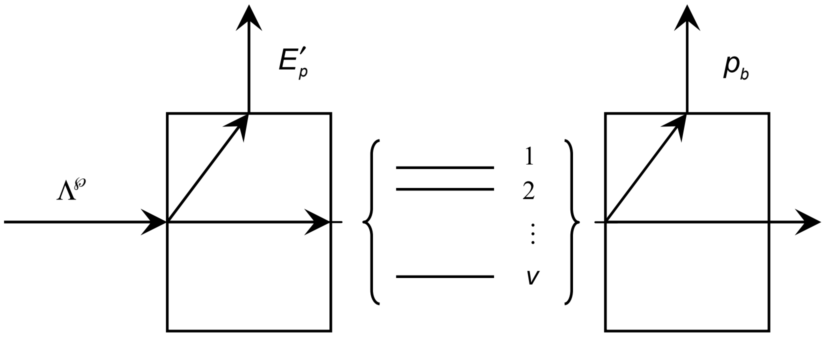

Fig. 7. The loss system subject to a busyness of the users

Model at Fig. 7 can be applied for calculating the SN of the network fragment at Fig. 2 at the channel DTU usage. Expression for the system at Fig. 7 allows consideration of the SN user's busyness probability p_b in comparison with the simplest variant at Fig. 1. For the system with repeated requests at the isolated trunk which consists of SN (cf. Fig. 8) $K = 1$; $L = \Lambda^{\wp} + \Lambda^{\wp\Re}\left(1 - \alpha_d \hat{E}'\right)$, where $\Lambda^{\wp\Re} = \dfrac{d\hat{E}'}{d\Lambda^{\wp}}\beta$.

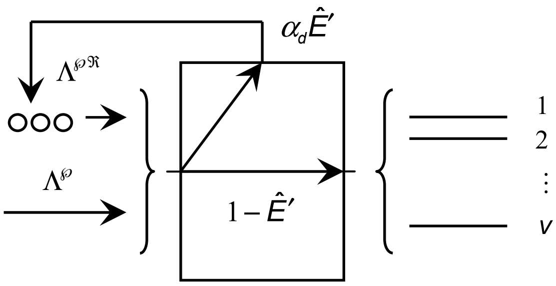

Fig. 8. The system with repeated requests at
the isolated trunk which consists of SN

Strictly speaking, all the SN of the IN should be considered as the systems with repeated requests since while real networks are functioning losses due to the network failures and timeouts leading to the repeated messages' initiation are unavoidable. The total absence of losses (both implicit and explicit) is typical only for ideal networks which are only theoretically valuable. For the system with repeated requests, subject to connection time, insistence and probability of the user's busyness (the preloading system) $K = 1$; $L = \Lambda^{\wp} + \Lambda^{\wp\Re}\dfrac{\tau_b + (1 - p_b)}{1 - \alpha_d \hat{E}' - (1 - \hat{E}')\alpha_p p_b}$, where $\beta \to 0$, cf. Fig. 9.

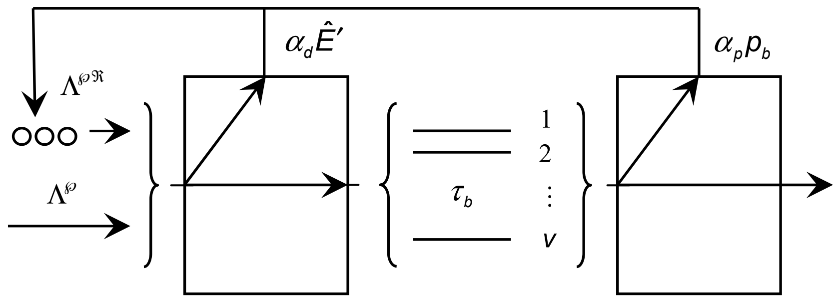

Fig. 9. The system with repeated requests subject to connection time, insistence and probability of the user's busyness (the preloading system)

For the system (cf. Fig. 10) with arbitrary value of arbitrary variants of repeated requests' generating $K = 1$; $L = \Lambda^{\wp} + \Lambda^{\wp\Re} \dfrac{(1-p)\ldots}{1-(x+y+z+\ldots)}$, $x = \alpha_z E^{\wedge} \alpha_Q Q$, $y = \alpha_d E^{\wedge}(1-Q)p$, $z = (1-E^{\wedge}) \cdot \alpha_p p$, $\beta \to 0$.

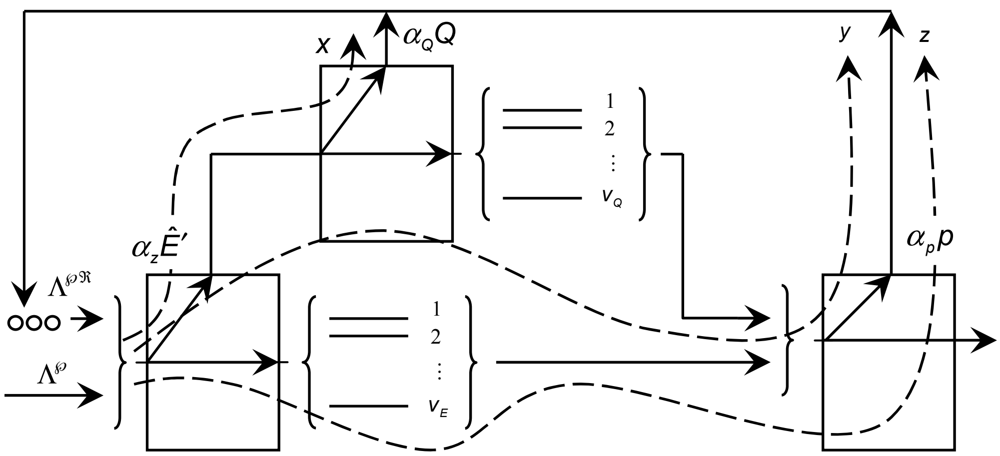

Fig. 10. The system with arbitrary value of arbitrary variants of repeated requests'

For the queuing system (cf. Fig. 11) with the restricted queue and repeated messages $K = \frac{1 - \Lambda^{\ell}/v}{1 - \left(\Lambda^{\ell}/v\right)^{M+1}}$; $L = \Lambda^{\ell} + \Lambda^{\ell\Re}\left(1 - \hat{E}''\right)$, where $\Lambda^{\ell\Re} = \frac{d\hat{E}''}{d\Lambda^{\ell}}\beta$.

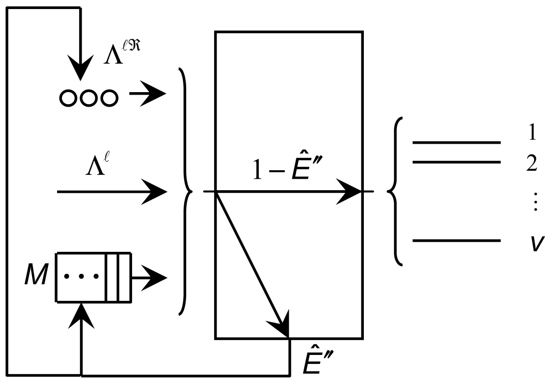

Fig. 11. The queuing system with the restricted queue and repeated messages

For the joint system with the repeated messages $K = \frac{1 - \Lambda^{\ell}/v}{1 - \left(\Lambda^{\ell}/v\right)^{M+1}}$; $L = \Lambda^{\Phi} + \frac{\Lambda^{\Phi\Re}}{1 - \alpha\hat{E}'''}$, where $\Lambda^{\Phi\Re} = \frac{d\hat{E}'''}{d\Lambda^{\Phi}}\beta$, $\beta \to 0$.

On the basis of the actual situation at the modern IN it's expedient to research the SN supporting simultaneously the packet, channel, frame and/or cell DTM of the MS load as the joint service system with the messages requiring retransmission.

Proposed representation of the unified method for calculating the network equipment can be used for estimation of any system with any service discipline of the message streams. At the same time proposed approach causes an adaptability of the method for calculating the network equipment to the DTM in the researched IN.

Results of the Developed Method Application for the Network Equipment Calculating at the Information Network Design

In order to evaluate an efficiency of the proposed method for calculating the network equipment we compared several possible methods for calculating the information network's fragment (cf. Fig. 12).

This is a converged fragment; it is aggregate of segments of the several networks: PSTN, DCN, the CATV and VoIP network. Additionally the fragment aggregates the PSTN segments with the ADSL/VDSL technologies' usage and CATV based on the DOCSIS standard. The analyzed network fragment consists of 46 transit nodes, of which 12 are hardwired as the digital switching system, 21 – as the router, 8 are functioning as the gateway between the network segments, 5 are implementing the Softswitch system.

All the calculations are based on the assumption of a stationary mode of the network functioning. The initial data are based on the real Ukrainian networks' values of users number, intensity of information sources, information types and classes, used at the design by the "Ukrainian State Institute for the Design and Development of the Information-communication Infrastructure "Diprozvyazok". Mentioned parameters' values correspond to the aggregated segments' features which are typical specially for such networks.

We compared the usage results of four methods (M1…M4) for calculating the network SN. **M1** is based on the Erlang's first formula which was base until the last time for all the Ukrainian design organizations (including the "Diprozvyazok"). **M2** takes into account a possibility of the requirements' repeated service due to the rejects (explicit loss) or waiting in excess of the allowed time (implicit loss). **M3** is theoretically possible calculation for the joint service system, subject to repeated messages and the MS aggregated nature, based on the Erlang's first formula. **M4** is based on the authors' unified calculation method for the joint system with the messages requiring retransmission.

These methods were applied subject to the features of the equipment functioning at the researched network fragment's nodes. We give analysis of the methods usage results for calculating the transit node (S3) hardwired by the Softswitch equipment (it's marked out by bold lines at the Fig. 12). This node services the MS load of the packet DTM as well as realizes the joint service of aggregated load of the cell, channel and frame DTM. This load arrives from the seven network segments (PSTN, DCN, VoIP, CATV, PSTN with ADSL and VDSL, CATV with DOCSIS, the telephone and data transfer networks' hybrid segment) by the five routes via routers R6, R7, R12, R13, R17. The DCN segments load arrives to the Softswitch (S3) via routers (R6) and (R13), the PSTN voice load – via gateway (G3) and router (R7), the load from the DCN, CATV (including the DOCSIS CATV) and the telephone and data transfer networks' hybrid segments – via router (R12). Generally the node, servicing the corresponding MS, is the queuing system with the restricted queue.

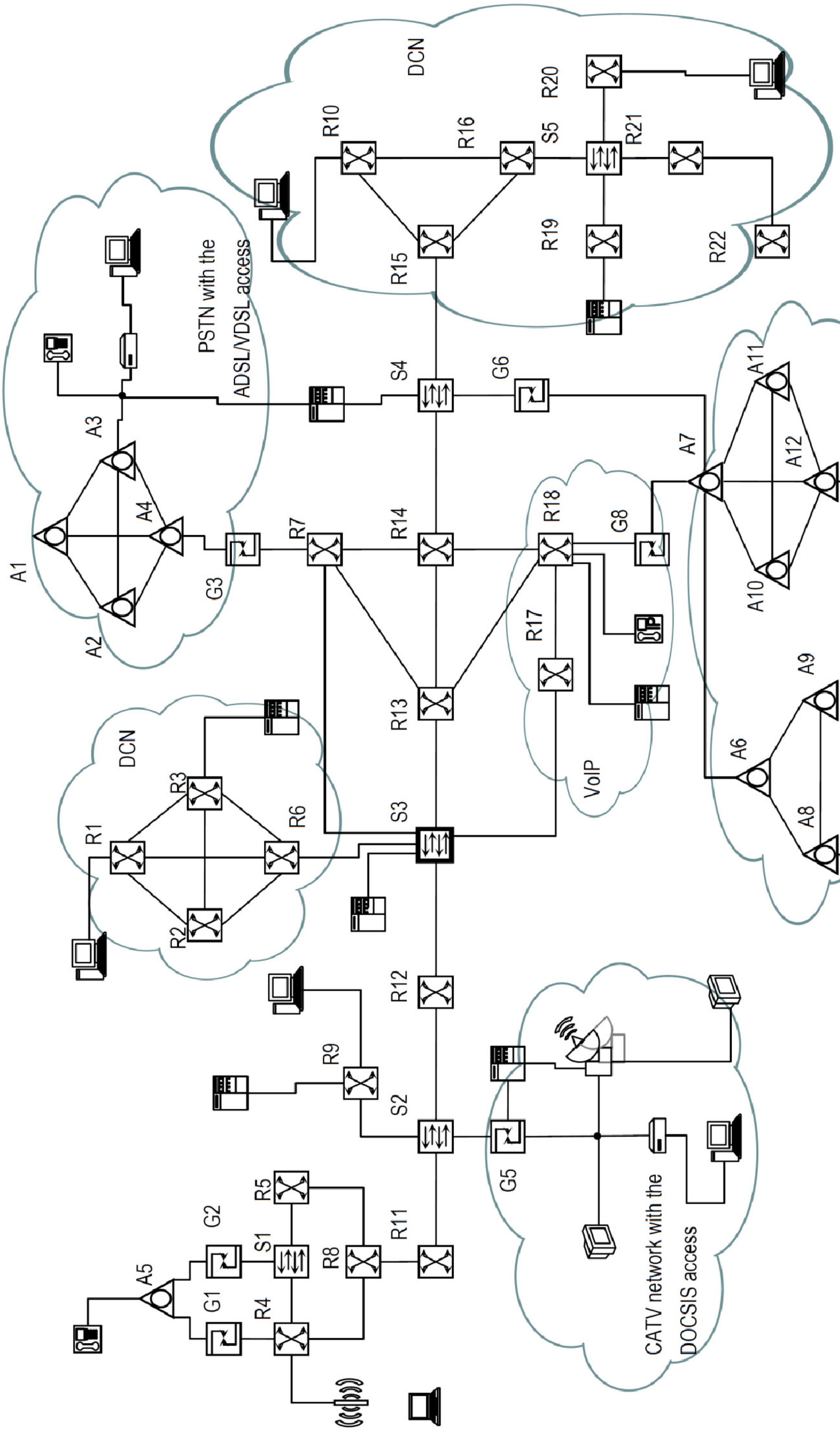

Fig. 12. The convergent network fragment (aggregation of the PSTN, DCN, the CATV and VoIP network)

The calculation results of the S3 service node for the transit load of aggregated MS via the five mentioned above routes (routers) are shown in Fig. 13. The histogram's first row represents the result of calculating the SN number at the M1 usage. The value is 60474 SNs. Deficiency of this method is caused by a non consideration of the messages re-arrival, characteristically for the real service systems' functioning. This leads to increase of the losses and the QoS degradation correspondingly. So the histogram's second row represents the calculation result according to the M2. The result of consideration of the messages requiring retransmission is 67731 SNs. This is more than the M1 result for 12%. The histogram's third row represents the result of calculating the SN as the joint service system, subject to the repeated messages and aggregated nature of the MS. The value of usage of the M3, based on the Erlang's first formula, is 47412 SNs.

Usage of the existing methods for a designing the IN converged segments leads to the significant overstatement of the SNs number, cf. Fig. 13. It's caused by the fact that these methods are based on the load which has additive nature in terms of the arrival information streams. I.e. the SN number is being calculated separately for each IS generating the load. But the load is being jointly served by the researched node. However derived results are adding and their sum determines necessary number of the SN. Actually the SN is the joint service system of a pre-aggregated MS. This fact should be considered in the calculation method. That's why the M4 result (cf. Fig. 13 fourth row) is the most accurate among the compared methods. The value is 37929 SNs (in spite the fact that it joint to the messages requiring retransmission, cf. Fig. 14). It's lower by 20% than the best theoretically possible result (i.e. the M3, subject to the load aggregation, which isn't being considered at the real design) and by 44% than the M2 result subject to the re-service. The 20% accuracy enhancement of the result, provided by the authors' method usage, effects a saving of 9483 SNs, the 44% accuracy enhancement of the result – 29802 SNs.

The authors' method usage specially for the joint service systems of the aggregated load generated by the heterogeneous information streams both at the channel and packet DTM is the best evidence of its adequacy and expedience. All the results given in this paper are documented by National Commission for the State Regulation of Communications and Informatization, “Diprozvyazok” PJSC, “Ukrtelecom” LLC (Odessa branch).

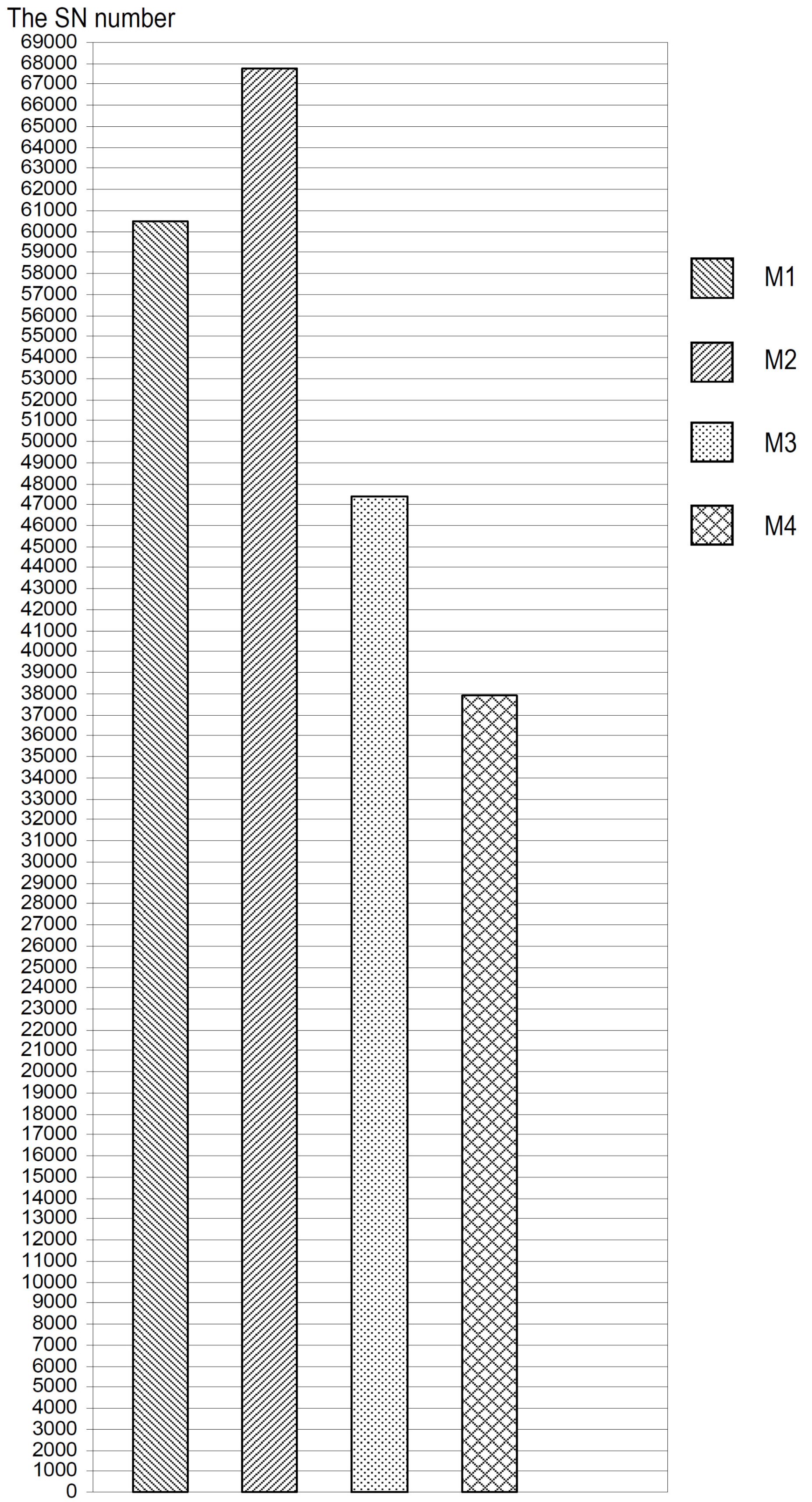

Fig. 13. The calculation results of the Softswitch (S3) service system

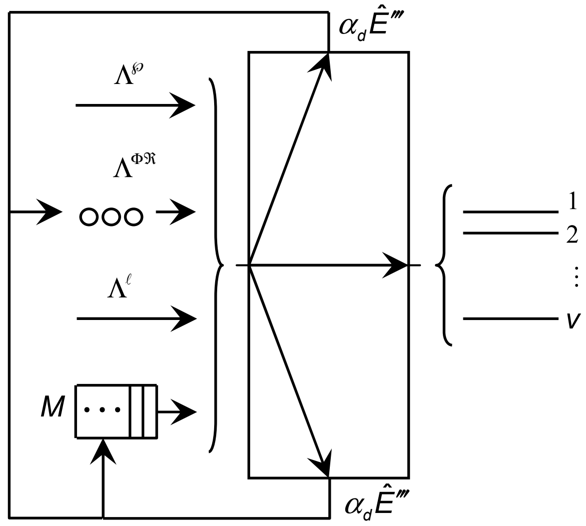

Fig. 14. The joint service system of the messages requiring retransmission

Conclusion

The developed unified method for calculating the network equipment capacity is an adaptable to the service system structure and discipline as well as to the data transfer mode which is being used in the network.

The method is improved with introduction of the coefficients, reproducing the real information stream's nature and in this way expanding its adequacy and sphere of application by means of specification of a used stream model.

An usage necessity of the message stream model instead of the call stream model is justified for the modern convergent information network. Such a model allows adequate consideration of features of the probabilistic-temporal structure of the information streams at the load intensity estimation and the QoS determination.

The developed load model is used for modification of the unified method for calculating the network equipment to any service system and discipline at all the data transfer modes, subject to parameters of the message streams circulating in the network. These parameters consist of the distribution functions of: number of the messages both in time and space, amount of the information in the message stream, time while the messages are in the network and value of the information, as well as the QoS parameters (a maximum permissible probability of symbol corruption, mean delay and its variance, etc.).

The developed method for calculating the equipment of information networks based on the message stream model assures the 18…22% accuracy enhancement of the results and so extends accuracy of the network design on the whole. The implementation results of the developed methods and models to the real design of Ukrainian information networks are verification which confirms their reliability.

Bibliography

[Leland et al., 1994] Will E. Leland, Murad S. Taqqu, Walter Willinger, Daniel V. Wilson. On the Self-Similar Nature of Ethernet Traffic (Extended version) // IEEE/ACM Transactions on Networking. – 1994.

[Duffy et al., 1994] Diane E. Duffy, Allen A. McIntosh, Mark Rosenstein, Walter Willinger. Statistical Analysis of CCSN/SS7 Traffic Data from Working CCS Subnetworks // IEEE Journal on Selected Areas in Communications. – 1994.

[Crovella, Bestravros, 1996] Crovella M., Bestravros A. Self-Similarity in World-Wide Web Traffic: Evidence and Possible Cause // Proceedings, ACM Sigmetrics Conference on Measurements and Modeling of Computer Systems. – 1996.

[Giroux, Ganti, 1999] Giroux N., Ganti S. Quality of Service in ATM Networks: State-of-the-Art Traffic Management. – NJ, USA: Prentice Hall, 1999.

[Borella, Brewster, 1998] Borella M., Brewster G. Measurement and Analyses of Long-Range Packet Dependent Behavior of Internet Packet Delay // 17th Annual Joint Conference of the IEEE Computer and Communications Societies. – 1998.

[Oliveira et al., 2003] Carlos Oliveira, Jaime Bae Kim, Tatsuya Suda. Long-Range Dependence in IEEE 802.11b Wireless LAN Traffic: An Empirical Study // IEEE 18th Annual Workshop on Computer Communications. – 2003.

[Ilnickis, 2004] Sergejs Ilnickis. M/M/1 and G/M/1 Systems with a Self-Similar Input Traffic. – 2004.

[Min, Ould-Khaoua, 2004] Geyong Min, Mohamed Ould-Khaoua. A Performance Model for Wormhole-Switched Interconnection Networks under Self-Similar Traffic // IEEE Tansactions on Computers. – Glasgow, 2004.

[Stallings, 2002] William Stallings. High-Speed Networks and Internets: Performance and Quality of Service (2nd Edition). – Prentice Hall, 2002.

[Gayvoronska, 2013 (1)] Gayvoronska G.S. NGN Access Networks' Features Demanding a New Approach for Their Synthesis // PIC&T. – Kharkov, 2013. (in Russian) = Гайворонская Г.С. Особенности сетей доступа к NGN, требующие нового подхода к их синтезу // I

Международная научно-практическая конференция “Проблемы инфокоммуникаций. Наука и Технологии”. – Харьков, 2013.

[Gayvoronska, Britsky, 2016] Galina Gayvoronska, Sergey Britsky. Research of Features and Demand of Modern Infocommunication Services in Ukraine // Information Content and Processing. – Bulgaria: ITHEA, 2016. (in Russian) = Галина С. Гайворонская, Сергей В. Брицкий. Исследование особенностей и востребованности современных инфокоммуникационных услуг в Украине // Information Content and Processing. – Bulgaria: ITHEA, 2016.

[Gayvoronska, 2015] Galina Gayvoronska. Designing Features of Access Networks // Second International Scientific-Practical Conference “Problems of Infocommunications. Science and Technology” – Kharkov, 2015.

[Gayvoronska et al., 2014] Galyna Gayvoronska, Oleg Domaskin, Svitlana Sakharova. Development of Expert System for Forecasting the Users Demands to Telecommunication Network // First International Conference “Information Technologies and Interactions”. – Kyiv, 2014. (in Ukrainian) = Галина Гайворонська, Олег Домаскін, Світлана Сахарова. Розробка експертної системи прогнозування вимог користувачів до телекомунікаційної мережі // IT&I 2014 – International Journal "Information Technologies & Knowledge". – Bulgaria: ITHEA, 2014.

[Gayvoronska, 2013 (2)] Galina Gayvoronska. Infocommunications // Science Intensive Infocommunication Techologies: Collective Monograph edited by Bezruk V.M., Barannyk V.V. – Kharkov, 2013. (in Russian) = Гайворонская Г.С. Инфокоммуникации // Наукоемкие технологии в инфокоммуникациях: коллективная монография под ред. В.М. Безрука, В.В. Баранника. – Харьков: Компания СМИТ, 2013.

[Gayvoronska et al., 2012] Gayvoronska G.S., Sakharova S.V., Kryzanovska A.A. Determination of Quality of Service at Multiservice Access Networks // Donetsk Railway Transport Institute. – 2012. (in Russian) = Гайворонская Г.С., Сахарова С.В., Крыжановская А.А. Определение качества обслуживания на сетях доступа к мультисервисным сетям // Наукові праці Донецького інституту залізничного транспорту. – Донецьк, 2012.

[ITU-T, 2008] ITU-T Recommendation E.800: Definitions of terms related to quality of service.

[Gannitsky, 2012] Gannitsky I.V. Telecommunication Networks' Functioning Efficiency's Increase by Unification of Method for Calculating a Network Equipment: PhD thesis. – Odessa, 2012. (in Russian) = Ганницкий И.В. Повышение эффективности функционирования телекоммуникационных сетей за счет унификации метода определения сетевого оборудования: дис. канд. техн. наук: 05.12.12 – телекоммуникационные системы и сети. – Одесса, 2012.

[Gayvoronska, 2000] Gayvoronska G.S. Main ITU information networks' recommendations. – Odessa: UNCA, 2000. (in Russian) = Гайворонская Г.С. Основные рекомендации МСЭ по информационным сетям. – Одесса: УГАС, 2000.

[Gayvoronska, 2007 (1)] Gayvoronska G.S. Level model of infocommunication services representation // Communications. – 2007. (in Russian) = Гайворонская Г.С. Уровневая модель представления инфокоммуникационных услуг // Зв'язок. – 2007.

[Stormer et al., 1971] Stormer H. et al. Teletraffic Theory. – M.: Communications, 1971. (in Russian) = Хоранд Штермер, Эрика Белендорф, Норберт Бининда, Г. Бретшнайдер, Эрик Хоффман, Ганс Зухландт. Теория телетрафика. – М.: Связь, 1971.

[Eldin, Lind, 1972] Eldin A., Lind G. Teletraffic Theory. – M.: Communications, 1972 (in Russian) = Элдин А., Линд Г. Основы теории телетрафика. – М.: Связь, 1972.

[Kendall, Stuart, 1961] Kendall M.G., Stuart A. The Advanced Theory of Statistics. Inference and Relationship. – London: Charles Griffin & Co., Ltd., 1961.

[Kleinrock, 1975] Leonard Kleinrock. Queueing Systems: Volume I – Theory. – New York: Wiley Interscience, 1975.

[Kleinrock, 1964] Leonard Kleinrock. Communication Nets: Stochastic Message Flow and Design. – McGraw-Hill, 1964.

[ITU, 2005] ITU-D Teletraffic Engineering Handbook. – Geneva, 2005.

[Gayvoronska, 2007 (2)] Gayvoronska G.S. Multiple-factor Optimization of Spatio-temporal Structure of Telecommunication Network: Doctor of Engineering thesis. – Kiev, 2007. (in Russian) = Гайворонская Г.С. Многофакторная оптимизация пространственно-временной структуры телекоммуникационных сетей: дисс. доктора техн. наук: 05.12.12. – телекоммуникационные системы и сети. – Киев, 2007.

[Sergeev, 1983] Sergeev V.O. Polynomial formula for calculating the Erlang losses at fully accessible loss system and its usage // Quasi-electronic and electronic switching system. – M., 1983. (in Russian) = Сергеев В.О. Полиноминальная формула для вычисления эрланговских потерь на полнодоступном пучке с отказами и её применение // Квазиэлектронные и электронные системы коммутации: сб. науч. тр. центр. науч.-исслед. ин-та связи. – М., 1983.

[Sergeev, Chekmareva, 1988] Sergeev V.O. About a upper and lower bounds of common probability of losses at fully accessible system with re-calls // Models of information distribution and methods of their analysis. – M., 1988. (in Russian) = Сергеев В.О., Чекмарёва Е. О верхней и нижней оценке общей вероятности потерь в системе с повторными вызовами на полнодоступном пучке // Модели распределения информации и методы их анализа. – М.: ИППИ АН СССР, 1988.

[Gayvoronska, Somsikov, 2008] Gayvoronska G.S., Somsikov D.A. Unification of method for calculating the telecommunication networks' equipment // Refrigeration Engineering and Technology. – Odessa, 2008. (in Russian) = Гайворонская Г.С., Сомсиков Д.А. Унификация метода расчёта оборудования телекоммуникационных сетей // Холодильная техника и технология. – Одесса, 2008.

[Gayvoronska, 1997] Gayvoronska G.S. Research of aggregation parameters of different services' information streams into integrated call stream // Third international conference on telecommunication, television and sound broadcasting. – Odessa, 1997. (in Russian) = Гайворонская Г.С. Исследование параметров объединения иформационных потоков различных видов связи в единый интегральный поток вызовов // Тр. III междунар. науч.-тех. конф. по электросвязи, телевизионному и звуковому вещанию. – Одесса, 1997.

[Gayvoronska, 1999] Gayvoronska G.S. Research of Aggregated Call Streams in Integrated Services Digital Network: PhD thesis. – Kiev, 1999. (in Russian) = Гайворонская Г.С. Исследование параметров объединенных потоков вызовов интегральных цифровых сетей связи: дисс. кандидата техн. наук: 05.12.12. – телекоммуникационные системы и управление ими. – Киев, 1999.

[Gayvoronska, Gannitsky, 2007] Gayvoronska G.S., Gannitsky I.V. Evaluation of impact of probabilistic-temporal structure of call stream on telecommunication network capacity // Telecommunication Problems. – K., 2007. (in Russian) = Гайворонская Г.С., Ганницкий И.В. Оценка влияния вероятностно-временной структуры потоков вызовов на объем оборудования телекоммуникационных сетей // Проблеми телекомунікацій. – К.: НТУУ «КПІ», 2007.

[Gannitsky, 2009] Gannitsky I.V. Usage of a normalized swing method for estimating the telecommunication network's call stream's characteristics // Digital Technologies. – Odessa, 2009. (in Russian) = Ганницкий И.В. Применение метода нормированного размаха для оценки характеристик потока вызовов на телекоммуникационной сети // Цифрові технології. – Одеса: ОНАЗ, 2009.

[Davydov et al., 1977] Davydov G.B., Roginsky V.N., Tolchan A.Y. Telecommunication Networks. – M.: Communications, 1977. (in Russian) = Г.Б. Давыдов, В.Н. Рогинский, А.Я. Толчан. Сети электросвязи. – М.: Связь, 1977.

[Gayvoronska, 1998] Gayvoronska G.S. Aggregation of load streams in telecommunication networks with bypass routes // Informatics and Communications. – 1998. (in Russian) = Гайворонская Г.С. Объединение потоков нагрузки на сетях связи с обходными направлениями // Труды УГАС "Информатика и связь". – 1998.

[Gayvoronska, Kalnev, 2001] Gayvoronska G.S., Kalnev A.N. Packet service for public data transfer in GSM network // Ukrainian Scientific Research Institute of Radio and TV. – 2001. (in Russian) =

Гайворонская Г.С., Кальнев А.Н. Служба пакетной передачи данных общего пользования сети стандарта GSM // Труды УНИИРТ. – 2001.

[Gayvoronska, 2007 (3)] Gayvoronska G.S. Synthesis problem of spatio-temporal structure of telecommunication network // State University of Information and Communication Technologies. – 2007. (in Russian) = Гайворонская Г.С. Проблема синтеза пространственно-временной структуры телекоммуникационной сети // Вісник ДУІКТ. – 2007.

[Gayvoronska, Solomitsky, 2012 (1)] Gayvoronska G.S., Solomitsky M.Y. Justification of the message stream research in convergent telecommunication network // Modern Information and Communication Technologies. – Kyiv, 2012. (in Ukrainian) = Гайворонська Г.С., Соломицький М.Ю. Обґрунтування необхідності дослідження моделі потоку повідомлень в конвергентній телекомунікаційній мережі // Науково-технічна конференція «Сучасні інформаційно-комунікаційні технології». – Київ: ДУІКТ, 2012.

[Solomitsky, 2013] Solomitsky M.Y. Research aspects of semistructured infocommunication mediums // Science Intensive Infocommunication Techologies: Collective Monograph edited by Bezruk V.M., Barannyk V.V. – Kharkov, 2013. (in Russian) = Соломицкий М.Ю. Аспекты исследования слабоструктурированных сред инфокоммуникаций // Наукоемкие технологии в инфокоммуникациях: коллективная монография под ред. В.М. Безрука, В.В. Баранника. – Харьков: Компания СМИТ, 2013.

[Gayvoronska, Solomitsky, 2012 (2)] Galyna Gayvoronska, Maxim Solomitsky. About the problem of development of messages stream model in the convergent telecommunication network // Problems of Computer Intellectualization. – Kyiv-Sofia: National Academy of Sciences of Ukraine V.M.Glushkov Institute of Cybernetics, ITHEA, 2012.

[Gayvoronska, Solomitsky, 2014] Gayvoronska G.S., Solomitsky M.Y. Formalization of message streams at technologies convergence in telecommunication networks // Refrigeration Engineering and Technology. – Odessa, 2014. (in Russian) = Г.С. Гайворонская, М.Ю. Соломицкий. Формализация потоков сообщений при конвергенции технологий в телекоммуникационной сети // Холодильна техніка і технологія. – Одеса: ОНАХТ, 2014.

[Solomitsky, 2014 (1)] Solomitsky M.Y. Research of time characteristics of message streams in convergent telecommunication network // Donetsk Railway Transport Institute. – 2014. (in Russian) = Соломицкий М.Ю. Исследование временных характеристик потоков сообщений конвергентной телекоммуникационной сети // Збірник наукових праць ДонІЗТ. – Донецьк: ДонІЗТ, 2014.

[Kramer, 1975] Kramer W. Investigations of Systems with Queues in Series (Untersuchung von Systemen mit seriellem Warten) // Dissertation University of Stuttgart 1975 and 22nd Report on Studies in Congestion Theory – Institute of Switching and Data Technics, 1975.

[Langenbach-Belz, 1972] Langenbach-Belz M. Sampled Queueing Systems Proc.Symp. on Computer Communications Networks & Teletraffic // Polytechnic Press – Brooklyn, New York, 1972.

[Solomitsky, 2014 (2)] Solomitsky M.Y. Model of Message Streams of Convergent Telecommunication Network: PhD thesis. – Odessa, 2014. (in Ukrainian) = Соломицький М.Ю. Модель потоків повідомлень конвергентної телекомунікаційної мережі: дис. канд. техн. наук: 05.12.12. – телекомунікаційні системи та мережі. – Одеса, 2014.

Authors' Information

Galina Gayvoronska *– adviser to the chairman of the supervisory board «Ante Mediam» LLC, expert of UCCI committee on electronic communications, head of ITHEA infocommunication section, Odessa, Ukraine; e-mail: gsgayvoronska@gmail.com*

Major Fields of Scientific Research: optimization of telecommunication networks' evolution, information streams, load, perspective access networks, all-optical switching systems

Maxim Solomitsky *– School of information technologies and cybersecurity of V.S. Martinovsky Institute of Refrigeration Cryotechnologies and Ecoenergetics ONAFT, senior lecturer of the computer engineering department, Odessa, Ukraine; e-mail: sage89@mail.ru*

Major Fields of Scientific Research: traffic engineering, convergent telecommunication networks

ROAD TRACKING FROM UAV IMAGERY USING GRADIENT INFORMATION [1],[2]

David Asatryan, Samvel Hovsepyan, Vardan Kurkchiyan

Abstract: *Nowadays an unmanned aerial vehicle (UAV) has many applications in a variety of fields. Detection and tracking of a road in UAV videos play an important role in traffic monitoring, ground–vehicle tracking, automatic UAV navigation etc. In this paper, an efficient road tracking framework in UAV videos is proposed. It includes the procedures of images similarity assessment and dominant direction determination. The methodology is based on using the gradient information of the sequentially captured images and corresponding correction of the motion direction of UAV.*

Keywords: *UAV, road tracking, image similarity, dominant direction, gradient*

ACM Classification Keywords: *ACM Classification Keywords: Image Processing and Computer Vision*

Introduction

There is currently a rapid growth in unmanned aerial vehicles application in various fields of human activity. The use of drones as part of modern navigation, video and communication equipment makes UAVs an indispensable tool for many tasks.

One of the main problems with the use of UAVs is the use of UAVs for road monitoring, traffic monitoring, tracking are objects on it, automatic navigation and others.

Road tracking is the automated or semi-automated process which uses video information captured by a camera, and special image processing methods.

The complexity of this problem is associated with the presence of the numerous factors that influence the quality and content of the images on the video shooting frames. The main factors that are constantly operating in the process of a video shooting are the evolution of the system, the turbulence

[1] *This work was supported by the RA MES State Committee of Science, in the frames of the research project № 15T-1B274.*

[2] *This work was supported by Russian-Armenian (Slavonic) University, in the frames of the research project, financed by the RF State Committee of Science.*

of the atmosphere, the presence of wind, vibration side of the device, as well as external intentional and unintentional interference and distortion of the real scene.

In addition, images of the same objects in the adjacent frames of a video sequence differ both in size and perspective, and the images themselves are fluctuating due to the vibration of UAV board and other mentioned factors. Therefore, for the efficient processing of information received UAV, you must either create special approaches and methods or improve existing algorithms. At the same time it should be noted that the effectiveness of digital image analysis depends largely on the available prior information about the properties of the image characterizing features of a scene and numerical values of its major parameters. On the other hand, since the analysis results ultimately are estimated by a person, the used methods should take into account the properties and capabilities of human visual system (HVS).

It should be noted that the use of global positioning GPS systems or GLONASS and other types of navigation technology significantly simplifies the solution of the problem, but there are actual problems in recent years, in which the use of these means is inappropriate (for stealth) or impossible (in case of failure for a variety of reasons of related equipment). In such cases, the important and almost the only information to carry out the problem being solved is the video shooting of the area and online processing of video sequences. In the present study we attempted to develop methods for road tracking by control of similarity of images of various parts of the road, received by UAV.

In the literature suggested a number of approaches and methods related to the problem of search, detection, and tracking of a road and / or objects on it [1, 2 and 3]. They use a variety of features that characterize the presence of the road sections on the image or of objects on it, such as the movement [4], the background [5-6] and others [7-8]. In the paper [9] was used the fact of the similarity of neighbor video frames and corresponding similarity measure. We confine ourselves to a situation where the initial part of the road has already been found or defined, so in this study the issue of search and detection of the road in general is not considered. Instead, it is assumed that preliminarily is available a template, showing some characteristic of the road and its surroundings, and in each moment for the UAV video camera is accessible only a small part of the road contained in the current frame. Consequently, the road tracking procedure is reduced only to the use of information on road features, contained in the above-mentioned template, additionally using only electronic compass to determine the course of the flight. It is necessary to use the structural characteristics that are stable under various transformations and distortions of the image.

Description of Mathematical Models

The proposed procedure is based on repeated application of two algorithms - of the algorithms for estimating the similarity of the images and for evaluation of the dominant orientation of the images of discussed scenes. These algorithms previously proposed in [10-13], and showed the effectiveness of their using in various applications. Both algorithms are based on mathematical models, using a variety of characteristics of the image gradient field and lead to results corresponding to the perception of HVS.

Here is a brief description of these algorithms.

The basis for usage of gradient field is well known fact that HVS confidently recognizes the content of an image by the aggregate of available edges. Methods of edge detection in an image are basically based on gradient field analyze using different mathematical methods [14]. One of approaches to gradient field processing is its presentation as two-dimensional random variable and creating certain statistical models for image structure investigation [10].

Let $G_H = \|G_H(m,n)\|$ and $G_V = \|G_V(m,n)\|$ be the horizontal and vertical gradients of an image I correspondingly, and $M = \|M(m,n)\|$ be the gradient magnitude, where

$$M(m,n) = \sqrt{G_H^2(m,n) + G_V^2(m,n)} \ . \tag{1}$$

a. *Model for dominant orientation* of an image is based on usage the term of orthogonal regression, the equation of which is as follows

$$\frac{1}{1-\rho_{HV}^2}\left[\frac{(g_H-\mu_H)^2}{\sigma_H^2} - \frac{2\rho_{HV}(g_H-\mu_H)(g_V-\mu_V)}{\sigma_H\sigma_V} + \frac{(g_V-\mu_V)^2}{\sigma_V^2}\right] = C^2 , \tag{2}$$

where $\mu_H, \mu_V, \sigma_H, \sigma_V$ mathematical expectations and MSE of random variables G_H and G_V, ρ_{HV} is the coefficient of correlation between them, C – is a constant. Dominant orientation α of an image is determined as follows

$$tg\alpha = \frac{2 * \rho_{HV}}{\sigma_H^2 - \sigma_V^2 + \sqrt{\left(\sigma_H^2 - \sigma_V^2\right)^2 + 4\rho_{HV}^2}} \ . \tag{3}$$

Models (1)-(3) were applied in different tasks for analyzing of images with different sizes and orientations [10].

b. *Model for distribution of magnitude gradient.* Let assume that the gradient magnitude (1) is a random variable with Weibull distribution density

$$f(x;\eta,\sigma) = \frac{\eta}{\sigma}\left(\frac{x}{\sigma}\right)^{\eta-1} \exp\left[-\left(\frac{x}{\sigma}\right)^{\eta}\right], x \geq 0,$$

where $\eta > 0$ - shape parameter, $\sigma > 0$ - scale parameter.

It must be noted that the fact of using only two parameters in this model has significant advantage in certain tasks of image processing, connected with searching in big data bases. Particularly, it is proposed images similarity assessment measure in [10], which shows the proximity of images which have gradient magnitudes with Weibull distribution densities $f_1(x;\eta_1,\sigma_1)$ and $f_2(x;\eta_2,\sigma_2)$. The measure is determined by formula as follows

$$W^2 = \frac{\min(\eta_1,\eta_2)\min(\sigma_1,\sigma_2)}{\max(\eta_1,\eta_2)\max(\sigma_1,\sigma_2)},\ 0 < W^2 \leq 1,$$

where the gradients are calculated using Sobel operator and the parameters in (5) are estimated by method of moments.

Algorithm for Updating of UAV flight Course

Description of the algorithm we give on a concrete example. Fig. 1 shows an example of an image of the road, located in the forest, on which can be seen some moving vehicles on it. It is assumed that the initial small land area, containing a road, is known, hence we consider that the flight UAV begins with the video shooting of this place. A part of the road is chosen as shown in Fig. 1. The template will be compared to the current image area on similarity with him during the process of the road tracking. It is also assumed that the direction of flight is measured by an electronic compass and may be updated according to the control program, managing the flight of the UAV.

Fig. 1. Choosing the initial part of the road to be a template.

To illustrate the possibility of tracking the road on the basis of this procedure let's conduct the following experiment. We compute the similarity measure of the template to all areas of the same image size Fig. 1 which are obtained by the sliding window. Then we the resulting map of similarity binarized and compare to the binarized original image (Fig. 2).

a b

Fig. 2. Similarity map for chosen template (a) and binary pattern of the road (b).

Analysis and comparison of the images of Figures 2a and 2b indicates that the road surface is distinguished from the background significantly sharper in the first case, which illustrates the advantage of the proposed approach.

It is expected that in the course of the flight on board of the UAV is recorded a sequence of frames of video and for each frame it is making a decision about changing the direction of the UAV. This algorithm analyzes the video frames and changes direction of flight according to the following steps.

Step 1. Choosing the initial part of the road image (see. Fig. 1)

Step 2. Select ordinary frame of the video sequence, filmed in the course of the UAV flight. The image is divided into sections with approximately the same size, the amount of which is determined in advance, based on a preliminary study of the situation, flight conditions etc. (in this paper, these issues are not addressed). For definiteness, we assume it is equal to the number of 3x3 = 9 (see Fig.3a). The arrow on the image Fig.3a shows the direction of movement of UAV. In Fig. 3b marked selected template, which will be compared with the received image parts. Given the direction of motion shall be considered only six sites located in the upper part of Fig. 3a. It is assumed that the central portion corresponds to the UAV position at the moment of the frame shooting.

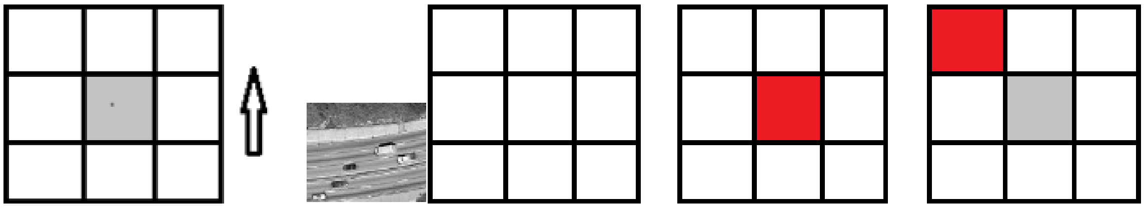

Fig. 3. Description of the scheme of comparison the image parts.

Step 3. Each part of the resulting partition of the image is compared by similarity with selected one, and find out the part with a maximum value of the similarity measure. Two situations are possible: a site with a maximum value of the similarity measure is coincide (Situation 1, Fig 3c.) or not (Situation 2, Fig 3d) with a central part. Case 1 indicates that the UAV is located above the road; therefore the direction of motion may be determined according to the dominant direction of the central part. In Case 2 the direction of movement must be corrected by the direction of the site with the maximum value of the similarity measure.

Step 4. Go to Step 2 for the processing of the next video frame.

Fig. 4 illustrates the procedure described above on the same example. The top three parts which lie in the frame captured by the UAV are highlighted in red. The second row shows the images of these areas (with the numbers 1, 2 and 3), and the third row shows the values of the measures of similarity of these images with the template (located on the lower left corner of the image). As can be seen, the maximum similarity is observed with a central portion (W02 = 0.71), hence we observe the above-described Situation 1. Therefore, the direction of the central portion of the image is equal to 1700, and the command may generate compensation of the motion direction of flight of the UAV.

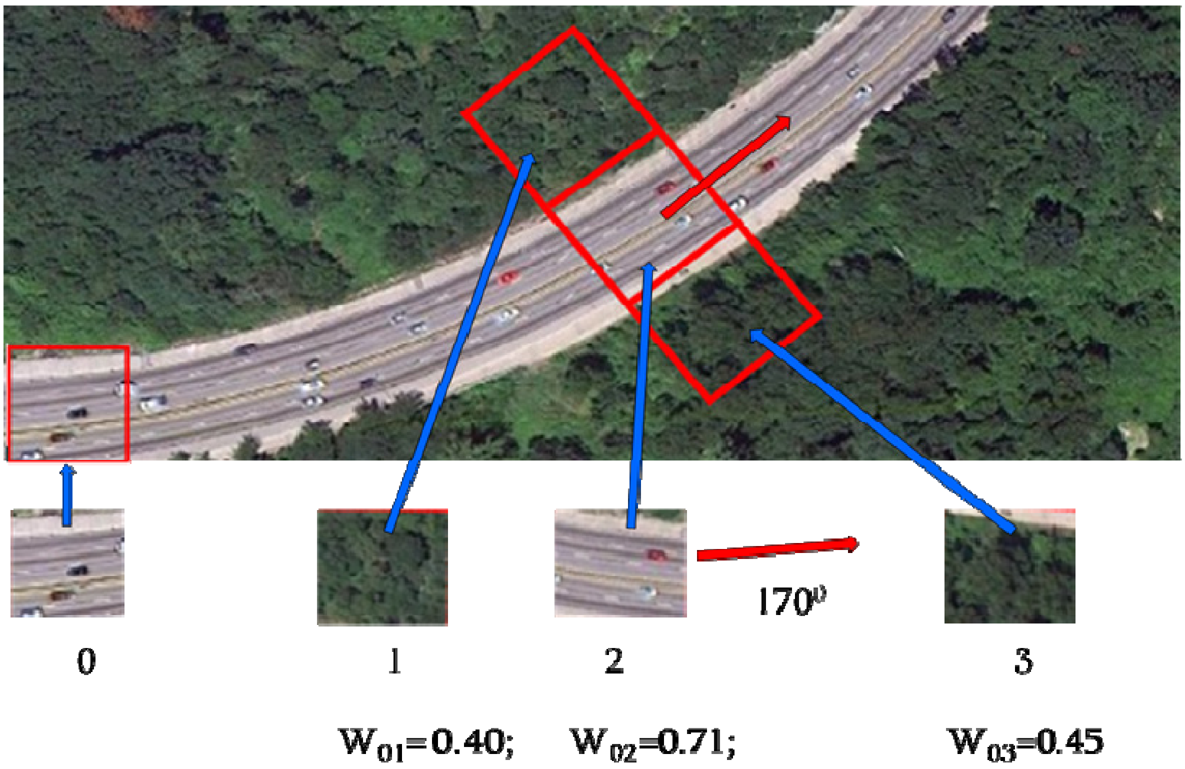

Fig. 4. Illustration of application of proposed procedure for road tracking.

Conclusion

The problem of road tracking by UAV imagery, using only the information delivered with video sequence and by electronic compass to determine the flight course, is considered. The possibility of reducing the problem to multiple applications of procedures of estimating similarity of images of different sizes and of determining the dominant direction of an image in each frame of a video sequence is analyzed. A corresponding motion control algorithm for UAV is proposed.

Bibliography

[1] Kalyani Chopade, Mrs. Shanthi K. Guru. A Survey: Unmanned Aerial Vehicle for Road Detection and Tracking. International Journal of Innovative Research in Computer and Communication Engineering Vol. 3, Issue 11, p. 10761-10766, 2015.

[2] Joshan.J Athanesious and Suresh .P. Systematic survey on Object Tracking Methods in Video. International Journal of Advanced Research in Computer Engineering and Technology. Vol. 1, Issue 8, p. 242-247, 2012.

[3] Yilmaz, A., Javed, O., and Shah, M. Object tracking: A survey. ACM Comput. Surv. 38, 4, Article 13 (Dec. 2006), 45 pages.

[4] Abhishek Kumar Chauhan and Prashant Krishan. Moving Object Tracking Using Gaussian Mixture model and Optical flow. International Journal of Advanced Research in Computer science and Research Engineering. Vol. 3, Issue 4, p. 243-246, 2013.

[5] Sen Ching S. and Chandrika Kamath.2011.Robust Techniques for background subtraction in urban Traffic Video. EURASIP Journal on Applied Signal Processing archive. Vol. 2005, p. 2330-2340, 2005.

[6] Ruolin Zhang and Jian Ding.2012.Object Tracking and Detecting based on Adaptive Background Subtraction. International Workshop on Information and Electronics Engineering. Vol. 29, p. 1351-1355, 2012.

[7] Hailing Zhou, Hui Kong, Lei Wei, Douglas Creighton, and Saeid Nahavandi. Efficient Road Detection and Tracking for Unmanned Aerial Vehicle. IEEE Transactions on Intelligent Transportation Systems, Vol. 16, No. 1, 2015, p. 297-309.

[8] J. Candamo, R. Kasturi, D. Goldgof. Using Color Profiles for Street Detection in Low-Altitude UAV Video. Proc. SPIE 7307, Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications VI, 73070O (April 28, 2009);

[9] Łoza, Artur and Mihaylova, Lyudmila and Bull, David and Canagarajah, Nishan (2009) Structural similarity-based object tracking in multimodality surveillance videos. Machine Vision and Applications, 20 (2). pp. 71-83. ISSN 1432-1769 (SSIM)

[10] Asatryan D., Egiazarian K. Quality Assessment Measure Based on Image Structural Properties. Proc. of the International Workshop on Local and Non-Local Approximation in Image Processing, Finland, Helsinki, pp. 70-73, 2009.

[11] David G. Asatryan. Structure-Based Technique for Object Detecting in UAV Imagery. Mathematical Problems of Computer Science 44, 51-58, 2015.

[12] David Asatryan, Karen Egiazarian, Vardan Kurkchiyan. Orientation Estimation with Applications to Image Analysis and Registration. International Journal “Information Theories and Applications”, Vol. 17, Number 4, pp. 303-311, 2010.

[13] Samvel Hovsepyan. Technique for Road Automated Tracking with Unmanned Aerial Vehicles. International Journal "Information Technologies & Knowledge" Volume 9, Number 4, p. 376-388, 2015.

[14] R.C.Gonzales, R.E. Woods. Digital Image Processing. 2-nd Edition, Prentice Hall, 2002.

Authors' Information

David Asatryan *– Professor, Head of Center for Critical Technologies of Russian-Armenian (Slavonic) University, Head of group of the Institute for Informatics and Automation Problems of NAS Armenia, 1, P.Sevaki Str., 0014, Yerevan, Armenia; e-mail: dasat@ipia.sci.am*

Major Fields of Scientific Research: Digital signal and image processing.

Samvel Hovsepyan *– Candidate of Technical Sciences, Software developer at RAU. Hovsep Emin Str., 0051, Yerevan, Armenia;*

e-mail: samvel1207@gmail.com

Major Fields of Scientific Research: Digital signal and image processing

Vardan Kurkchiyan *– Candidate of Technical Sciences, Software developer at RAU. Hovsep Emin Str., 0051, Yerevan, Armenia;*

e-mail: vakur@gmail.com.

Major Fields of Scientific Research: Digital signal and image processing.

A SHORT SURVEY ON CAR ALGORITHMS

Ilia Mitov, Stefan Karastanev, Krassimir Markov

Abstract: *In this paper we discuss in detail CAR algorithms. We present the main types of algorithms for association rule mining, pruning techniques, quality rule measures and rule ordering strategies. We also describe a number of specific CAR algorithms.*

Keywords: *CAR algorithms, Survey*

1. Introduction

Association rule mining quickly became a popular instrument to model relationships between class labels and features from a training set [Bayardo, 1998]. It appeared initially within the field of market basket analysis for discovering interesting rules from large data collections [Agrawal et al, 1993]. Since then, many associative classifiers were proposed, mainly differing in the strategies used to select rules for classification and in the heuristics used for pruning rules. "Class association rules" (CAR) algorithms have its important place in the family of classification algorithms.

Zaïane and Antonie suggested that the five major advantages of associative classifiers are the following [Zaïane and Antonie, 2005]:

- the training is very efficient regardless of the size of the training set;
- training sets with high dimensionality can be handled with ease and no assumptions are made on dependence or independence of attributes;
- the classification is very fast;
- classification based on association methods present higher accuracy than traditional classification methods [Liu et al, 1998] [Li et al, 2001] [Thabtah et al, 2005] [Yin and Han, 2003];
- the classification model is a set of rules easily interpreted by human beings and can be edited [Sarwar et al, 2001].

Within the data mining community, research on classification techniques has a long and fruitful history. However, classification techniques based on association rules, are relatively new. The first associative classifier CBA was introduced by [Liu et al, 1998]. During the last decade, various other associative classifiers were introduced, such as CMAR [Li et al, 2001], ARC-AC and ARC-BC [Zaiane and Antonie, 2002], CPAR [Yin and Han, 2003], CorClass [Zimmermann and De Raedt, 2004], ACRI [Rak et al, 2005], TFPC [Coenen and Leng, 2005], HARMONY [Wang and Karypis, 2005], MCAR [Thabtah et al, 2005], CACA [Tang and Liao, 2007], ARUBAS [Depaire et al, 2008], etc.

CAR-algorithms are based on a relatively simple idea. Given a training set with transactions where each transaction contains all features of an object in addition to the class label of the object, the association rules are constructed, which have as consequent a class label. Such association rules are named "class association rules" (CARs).

Generally the structure of CAR-algorithms consists of three major data mining steps:

1. Association rule mining.
2. Pruning (optional).
3. Recognition.

The mining of association rules is a typical data mining task that works in an unsupervised manner. A major advantage of association rules is that they are theoretically capable of revealing all interesting relationships in a database. But for practical applications the number of mined rules is usually too large to be exploited entirely. This is why the pruning phase is stringent in order to build accurate and compact classifiers. The smaller the number of rules a classifier needs to approximate the target concept satisfactorily, the more human-interpretable is the result.

2. Association Rule Mining

Association rule mining was first introduced in [Agrawal et al, 1993]. It aims to extract interesting correlations, frequent patterns, associations, or casual structures among sets of instances in the transaction databases or other data repositories.

Association rule mining itself has a wide range of application domains such as market basket analysis, medical diagnosis/research, Website navigation analysis, homeland security and so on. In parallel, it participates as a step in the training process of CAR classifiers.

The datasets can be represented in two forms:

- transactional datasets;
- rectangular datasets.

In transactional datasets each record (transaction) can contain different number of items and order of the items can be arbitrary.

In rectangular datasets each record has the same number of attributes and position of the attribute value is fixed and corresponds to the attribute.

These differences are not particularly difficult to address since there is an easy way of converting transactional to binary rectangular dataset by ordering all possible items and pointing the presence of a concrete item with 1 (true) and respectively the absence with 0 (false).

The rectangular dataset also become transactional representation using attribute-value pairs in description of each record.

The description of the problem of association rule mining is firstly presented in [Agrawal et al, 1993]. The description of the problem provided below follows the one given in [Goethals, 2002].

Let $\mathbf{D}$ be a set of items.

A set $X = \{i_1, ..., i_k\} \subseteq \mathbf{D}$ is called an itemset or a k-itemset if it contains k items.

A transaction over $\mathbf{D}$ is a couple $T = (tid, I)$ where tid is the transaction identifier and I is an itemset. A transaction $T = (tid, I)$ is said to support an itemset $X \subseteq \mathbf{D}$ if $X \subseteq I$.

A transaction database D over $\mathbf{D}$ is a set of transactions over $\mathbf{D}$.

The cover of an itemset X in D consists of the set of transaction identifiers of transactions in D that support X: $cover(X, D) := \{tid \mid (tid, I) \in D, X \subseteq I\}$.

The support of an itemset X in D is the number of transactions in the cover of X in D: $support(X, D) := |cover(X, D)|$. Note that $|D| = support(\{\}, D)$.

An itemset is called frequent if its support is no less than a given absolute minimal support threshold $MinSup$, with $0 \le MinSup \le |D|$.

Let D be a transaction database over a set of items $\mathbf{D}$, and $MinSup$ a minimal support threshold. The collection of frequent itemsets in D with respect to $MinSup$ is denoted by $F(D, MinSup) := \{X \subseteq \mathbf{D} \mid support(X, D) \ge MinSup\}$.

An association rule is an expression of the form $X \Rightarrow Y$, where X and Y are itemsets, and $X \cap Y = \{\}$. Such a rule expresses the association that if a transaction contains all items in X, then that transaction also contains all items in Y. X is called the body or antecedent, and Y is called the head or consequent of the rule.

The support of an association rule $X \Rightarrow Y$ in D, is the support of $X \cup Y$ in D. An association rule is called frequent if its support exceeds a given minimal support threshold $MinSup$.

The confidence or accuracy of an association rule $X \Rightarrow Y$ in D is the conditional probability of having Y contained in a transaction, given that X is contained in that transaction:

$$confidence(X \Rightarrow Y, D) := P(Y \mid X) = \frac{support(X \cup Y, D)}{support(X, D)}.$$

The rule is called confident if $P(Y \mid X)$ exceeds a given minimal confidence threshold $MinConf$, with $0 \le MinConf \le 1$.

Especially in the case of classification association rules the head consists of only one attribute-value pair. In the case of rectangular data one of the columns contains class labels that divide the dataset into separate extensional parts.

Generally, an association rules mining algorithm consists of the following steps:

1. The set of candidate k-item-sets is generated by 1-extensions of the large (k-1)-item-sets generated in the previous iteration.
2. Supports for the candidate k-item-sets are generated by a pass over the database.
3. Item-sets that do not have the minimum support are discarded and the remaining item-sets are called large (frequent) k-item-sets.
4. This process is repeated until no more large item-sets are found to generate association rules from those large item-sets with the constraints of minimal confidence.

In many cases, the algorithms generate an extremely large number of association rules, often in thousands or even millions; in addition to this the association rules are sometimes very large. It is nearly impossible for the end-users to comprehend or validate such large number of complex association rules, thereby limiting the usefulness of the data mining results. Several researchers suggested strategies aimed at reducing the number of association rules:

— extracting of rules based on user-defined templates or instance constraints [Baralis and Psaila, 1997] [Ashrafi et al, 2004];

— developing interestingness measures to select only interesting rules [Hilderman and Hamilton, 2002]. For instance [Jaroszewicz and Simovici, 2002] proposed a solution to the problem using the Maximum Entropy approach;

— proposing inference rules or inference systems to prune redundant rules and thus present smaller, and usually more understandable sets of association rules to the user [Cristofor and Simovici, 2002];

- creating new frameworks for mining association rule to find association rules with different formats or properties [Brin et al, 1997].

Depending of the specificity of the observed problem many additional question arise. For instance [Liu et al, 1999] present an approach to the rare instance problem. The dilemma that arises in the rare instance problem is that searching for rules that involve infrequent (i.e., rare) instances requires a low support but using a low support will typically generate many rules that are of no interest. Using a high support typically reduces the number of rules mined but will eliminate the rules with rare instances. The authors attack this problem by allowing users to specify different minimum supports for the various instances in their mining algorithm.

The computational cost of association rules mining can be reduced by sampling the database, by adding extra constraints on the structure of patterns, or through parallelization.

Techniques for association rule discovery have gradually been adapted to parallel systems in order to take advantage of the higher speed and greater storage capacity that they offer. The transition to a distributed memory system requires the partitioning of the database among the processors, a procedure that is generally carried out indiscriminately. [Parthasarathy et al, 2001] wrote an excellent survey on parallel association rule mining with shared memory architecture covering most trends, challenges, and approaches adopted for parallel data mining.

3. Creating Association Rules

During the first stage, several techniques for creating association rules are used, which mainly are based on:

- Apriori algorithm [Agrawal and Srikant, 1994] (CBA, ARC-AC, ARC-BC, ACRI, ARUBAS);
- FP-tree algorithm [Han and Pei, 2000] (CMAR);
- FOIL algorithm [Quinlan and Cameron-Jones, 1993] (CPAR);
- Morishita & Sese Framework [Morishita and Sese, 2000] (CorClass).

Generating association rules can be made from all training transactions together (such it is in ARC-AC, CMAR, CBA) or can be made for transactions grouped by class label (as it is in ARC-BC), which offers small classes a chance to have representative classification rules.

We provide a brief overview of some distinctive algorithms created during the recent years, which are used or can be implemented at the step of creating the pattern set of CAR algorithms.

3.1. AIS

The AIS algorithm [Agrawal et al, 1993] was the first algorithm proposed for mining association rule in the early 90s, when a task for emulating the biological immune system in the real world scenarios became actual. AIS algorithm uses candidate generation to detect the frequent item-sets. The candidates are generated on the fly and are compared with previously found frequent item-sets. In this algorithm only one instance of consequent association rules are generated, which means that the consequent of those rules only contain one instance, for example we only generate rules like $X \cap Y \Rightarrow Z$ but not those rules as $X \Rightarrow Y \cap Z$. The main drawbacks of the AIS algorithm are too many passes over the whole database and too many candidate item-sets that finally turned out to be small are generated, which requires considerable memory and involves significant effort that turned out to be useless.

3.2. Apriori

The Apriori [Agrawal and Srikant, 1994] is the most popular algorithm for producing association rules. It created new opportunities to mine the data. Since its inception, many scholars have improved and optimized the Apriori algorithm and have presented new Apriori-like algorithms. Apriori uses pruning techniques to avoid measuring certain item-sets, while guaranteeing completeness. These are the item-sets that the algorithm can prove will not turn out to be large.

However, there are two bottlenecks of the Apriori algorithm. One is the complex candidate generation process that uses most of the time and memory because of the multiple scans of the database. Based on the Apriori algorithm, many new algorithms were designed with some modifications or improvements.

The Apriori algorithm for finding frequent item-sets makes multiple passes over the data. In the k-th pass it finds all item-sets having k instances called the k-item-sets. Each pass consists of two phases. Let F_k represent the set of frequent k-item-sets, and C_k the set of candidate k-item-sets (potentially frequent item-sets). The candidate generation phase where the set of all frequent (k-1)-item-sets, F_{k-1}, found in the ($k-1$)-th pass is applied first and it is used to generate the candidate item-sets C_k. The candidate generation procedure ensures that C_k is a superset of the set of all frequent k-item-sets. A specialized hash-tree data structure is used to store C_k. Then, data is scanned in the support counting phase. For each transaction, the candidates in C_k contained in the transaction are determined using the hash-tree data structure and their support count is incremented. At the end of the pass, C_k is examined to determine which of the candidates are frequent, yielding F_k. The algorithm terminates when F_k or C_{k+1} becomes empty.

Several optimizations of Apriori algorithm are available, such as:

- PASCAL [Bastide et al, 2000], which introduces the notions of key patterns and use inference of other frequent patterns from the key patterns without access to the database;
- Category-based Apriori algorithm [Do et al, 2003], which reduces the computational complexity of the mining process by bypassing most of the subsets of the final item-sets;
- Apriori-T [Coenen et al, 2004], which makes use of a "reverse" set enumeration tree where each level of the tree is defined in terms of an array (i.e. the T-tree data structure is a form of Trie);
- FDM [Cheung et al, 1996], which is a parallelization of Apriori for shared machines, each with its own partition of the database. At every level and on each machine, the database scan is performed independently on the local partition. Then a distributed pruning technique is employed.

3.3. FP-Tree

FP-Tree [Han and Pei, 2000] is another milestone in the development of association rule mining, which breaks the main bottlenecks of Apriori [Kotsiantis and Kanellopoulos, 2006]. The frequent item-sets are generated with only two passes over the database and without any candidate generation process. FP-tree is an extended prefix-tree structure storing crucial, quantitative information about frequent patterns. Only frequent length-1 instances will have nodes in the tree, and the tree nodes are arranged in such a way that more frequently occurring nodes will have better chances of sharing nodes than less frequently occurring ones. FP-Tree scales much better than Apriori because as the support threshold goes down, the number as well as the length of frequent item-sets increase dramatically. The frequent patterns generation process includes two sub processes: constructing the FT-Tree, and generating frequent patterns from the FP-Tree. The mining result is the same with Apriori series algorithms.

To sum up, the efficiency of FP-Tree algorithm accounts for three reasons:

- The FP-Tree is a compressed representation of the original database because only those frequent instances are used to construct the tree, other irrelevant information are pruned.
- This algorithm only scans the database twice.
- FP-Tree uses a divide and conquers method that considerably reduced the size of the subsequent conditional FP-Tree.

Every algorithm has his limitations, for FP-Tree it is difficult to be used in an interactive mining system. Another limitation is that FP-Tree is that it is not suitable for incremental mining.

3.4. TreeProjection

The innovation brought by TreeProjection [Agarwal et al, 2000] is the use of a lexicographic tree, which requires substantially less memory than a hash tree. The number of nodes in its lexicographic tree is exactly that of the frequent item-sets. The support of the frequent item-sets is counted by projecting the transactions onto the nodes of this tree. This improves the performance of counting the number of transactions that have frequent item-sets.

The lexicographic tree is traversed in a top-down fashion. The efficiency of TreeProjection can be explained by two main factors:

- the transaction projection limits the support counting in a relatively small space.
- the lexicographical tree facilitates the management and counting of candidates and provides the flexibility of picking efficient strategy during the tree generation and transaction projection phrases.

3.5. Matrix Algorithm

The Matrix Algorithm [Yuan and Huang, 2005] generates a matrix, which entries 1 or 0 by passing over the database only once, and then the frequent candidate sets are obtained from the resulting matrix. Finally, association rules are mined from the frequent candidate sets. Experimental results confirm that the proposed algorithm is more effective than Apriori Algorithm.

3.6. Sampling Algorithms

For obtaining associations, several algorithms use sampling. Some examples are provided below:

- Toivonen's sampling algorithm [Toivonen, 1996]. This approach is a combination of two phases. During phase 1 a sample of the database is obtained and all associations in the sample are found. These results are then validated against the entire database. To maximize the effectiveness of the overall approach, the author makes use of lowered minimum support on the sample. Since the approach is probabilistic (i.e. dependent on the sample containing all the relevant associations) not all the rules may be found in this first pass. Those associations that were deemed not frequent in the sample but were actually frequent in the entire dataset are used to construct the complete set of associations in phase 2;

- Progressive sampling [Parthasarathy, 2002] is yet another approach; it relies on a novel measure of model accuracy (self-similarity of associations across progressive samples), the identification of a representative class of frequent item-sets that mimic (extremely accurately) the self-similarity values across the entire set of associations, and an efficient sampling methodology that hides the overhead of obtaining progressive samples by overlapping it with useful computation;
- Sampling Error Estimation algorithm [Chuang et al, 2005] aims to identify an appropriate sample size for mining association rules. It has two advantages. First, it is highly efficient because an appropriate sample size can be determined without the need of executing association rules. Second, the identified sample size is very accurate, meaning that association rules can be highly efficiently executed on a sample of this size to obtain a sufficiently accurate result;
- Sampling large datasets with replacement [Li and Gopalan, 2004] is used when data comes as a stream flowing at a faster rate than can be processed. Li and Gopalan derive the sufficient sample size based on central limit theorem for sampling large datasets with replacement.

3.7. Partition

Partition [Savasere et al, 1995] is fundamentally different from other algorithms because it reads the database at most two times to generate all significant association rules. In the first scan of the database, it generates a set of all potentially large item-sets by scanning the database once and dividing it in a number of non-overlapping partitions. This set is a superset of all frequent item-sets so it may contain item-sets that are not frequent. During the second scan, counters for each of these item-sets are set up and their actual support is measured.

3.8. FOIL

FOIL (First Order Inductive Learner) is an inductive learning algorithm for generating classification association rules (CARs) developed by Quinlan and Cameron-Jones in 1993 [Quinlan and Cameron-Jones, 1993] and further developed by Yin and Han to produce the PRM (Predictive Rule Mining) CAR generation algorithm [Yin and Han, 2003]. PRM was then further developed, by Yin and Han, to produce CPAR (Classification based on Predictive Association Rules).

FOIL is a sequential covering algorithm that learns first-order logic rules. It learns new rules one at a time, removing the positive examples covered by the latest rule before attempting to learn the next rule.

The hypothesis space search performed by FOIL is best understood by viewing it hierarchically. Each iteration through FOIL'S outer loop adds a new rule to its disjunctive hypothesis. The effect of each new rule is to generalize the current disjunctive hypothesis (i.e., to increase the number of instances it classifies as positive), by adding a new disjunct. Viewed at this level, the search is a specific-to-general search through the space of hypotheses, beginning with the most specific empty disjunction and terminating when the hypothesis is sufficiently general to cover all positive training examples. The inner loop of FOIL performs a finer-grained search to determine the exact definition of each new rule. This inner loop searches a second hypothesis space, consisting of conjunctions of literals, to find a conjunction that will form the preconditions for the new rule. FOIL employs a specific performance FOIL Gain that differs from the entropy measure. This difference follows from the need to distinguish between different bindings of the rule variables and from the fact that FOIL seeks only rules that cover positive examples [Mitchell, 1997].

3.9. Morishita & Sese Framework

This framework [Morishita and Sese, 2000] efficiently computes significant association rules according to common statistical measures such as a chi-squared value or correlation coefficient. Because of anti-monotonicity of these statistical metrics, Apriori algorithm is not suitable for associative rule generation. Morishita and Sese present a method of estimating a tight upper bound on the statistical metric associated with any superset of an item-set, as well as the novel use of the resulting information of upper bounds to prune unproductive supersets while traversing item-set lattices.

4. Rule Quality Measures

The process of generating association rules usually creates an extremely big number of patterns. This bottleneck imposes the necessity of measuring the significance, respectively redundancy of the generated rules and ordering using different criteria.

Here we will mention some examples of used ranking of association rules.

For a rule P and a class-labeled data set $D = \{R_i \mid i = 1, \ldots, n\}$ several kinds of rule quality measures and combinations of them are used:

- The time of generation of the rule. This is a weak restriction used when all constrains before order two rules in equal places;
- $ncovers(P)$: the number of instances covered by P (i.e. $R_i : body(P) \subseteq body(R_i)$);

— $pos(P)$: the number of instances correctly classified by P (i.e. $R_i : body(P) \subseteq body(R_i)$ and $head(P) = head(R_i)$);

— $neg(P)$: the number of negative instances covered by P (i.e. $R_i : body(P) \subseteq body(R_i)$ and $head(P) \neq head(R_i)$);

— $|D|$: the number of instances in D;

— Coverage: $coverage(P) = \frac{ncovers(P)}{|D|}$;

— Accuracy: $accuracy(P) = \frac{pos(P)}{ncovers}$;

— Cardinality: $card(P) = |body(P)|$;

— Pessimistic error rate: $PER(P) = \frac{neg(P)+1}{neg(P)+pos(P)+2}$

— p_i is the probability of class c_i in D;

— Expected information: $Info(D) = -\sum_{i=1}^{m} p_i * \log_2(p_i)$;

— Information gain: $InfoGain(D) = Info(D) - \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times Info(D_j)$;

— FOIL gain (it favors rules that have high accuracy and cover many positive instances): $FOILGain(P,P') = pos(P') \times \left(\log_2 \frac{pos(P')}{pos(P')+neg(P')} - \log_2 \frac{pos(P)}{pos(P)+neg(P)} \right)$;

Further measures can be defined but those listed above are the most basic ones.

5. Pruning

In order to reduce the produced association rules, pruning in parallel with (pre-pruning) or after (post-pruning) creating association rules is performed. Different heuristics for pruning during rule generation are used, mainly based on minimum support, minimum confidence and different kinds of error pruning [Kuncheva, 2004]. In post-pruning phase, criteria such as data coverage (ACRI) or correlation between consequent and antecedent (CMAR) are also used.

During the pruning phase or in classification stage, different ranking criteria for ordering the rules are used. The most common ranking mechanisms are based on the support, confidence and cardinality of the rules, but other techniques such as the cosine measure and coverage measure (ACRI) also exist; we can mention amongst them:

- Pruning by confidence: retain more general rules with higher accuracy: $|R_1| < |R_2|$ and $conf(R_1) < conf(R_2)$, than R_1 is pruned (used in ARC-AC, ARC-BC);
- Pruning by precedence: special kind of ordering using "precedence" (CBA and MCAR);
- Correlation pruning: statistical measuring of the rule significance using weighted χ^2 (CMAR).

6. Recognition

In the recognition stage, three different approaches can be discerned [Depaire et al, 2008]:

1. using a single rule.
2. using a subset of rules.
3. using all rules.

An example which uses a single rule is CBA. It classifies an instance by using the single best rule covering the instance.

CPAR uses a subset of rules. It first gathers all rules covering the new instance and selects the best n rules per class. Next, it calculates the average Laplace accuracy per class and predicts the class with the highest average accuracy.

Additionally to support, coverage and confidence, ACRI uses also the cosine measure.

CMAR uses all rules covering a class to calculate an average score per class.

CMAR selects the rule with the highest χ^2 measure from the candidate set.

ARC-AC and ARC-BC use the sum of confidences as score statistics.

A different approach is proposed in TFPC, which suggests to consider the size of the antecedent and to favor long rules before making an allowance for confidence and support.

When a subset or all rules is being used, several order-based combined measures can be applied:

- Select all matching rules;
- Group rules per class value;
- Order rules per class value according to criterion;
- Calculate combined measure for best Z rules;

— Laplace Accuracy (CPAR): if k is the number of class values then $LA = \frac{support(R)+1}{support(body(R))+k}$

7. Some Representatives of CAR Algorithms

In this subsection we present briefly several representatives of the CAR Algorithm.

7.1. CBA

In CBA [Liu et al, 1998], Apriori is applied to create the association rules.

For measuring the significance of the rules a special "precedence" definition is given:

$P^1 \succ P^2$ (rule P^1 precedes rule P^2) if:

1. $confidence(P^1) > confidence(P^2)$;
2. $confidence(P^1) = confidence(P^2)$ but $support(P^1) > support(P^2)$;
3. $confidence(P^1) = confidence(P^2)$, $support(P^1) = support(P^2)$, but P^1 is generated earlier than P^2.

Pruning is based on the pessimistic error rate based pruning method in C4.5.

Condition 1. Each training case is covered by the rule with the highest precedence among the rules that can cover the case.

Condition 2. Every rule correctly classifies at least one remaining training case when it is chosen.

The key point is instead of making one pass over the remaining data for each rule, the algorithm to find the best rule to cover each case.

During the recognition CBA just searches in the pruned and ordered list for the first rule that covers the instance to be classified. The prediction is the class label of that classification rule. If no rule covers the instance, CBA uses the default class calculated during pruning. If the decision list is empty, the majority class of the training instance will be assigned to each test instance as default.

7.2. CMAR

CMAR [Li et al, 2001] employs a novel data structure, CR-tree, to compactly store and efficiently retrieve a large number of rules for classification. CR-tree is a prefix tree structure to explore the sharing among rules, which achieves substantial compactness.

In the phase of rule generation, CMAR computes the complete set of rules. CMAR prunes some rule and only selects a subset of high quality rules for classification. CMAR adopts a variant of FP-growth method, which is much faster than Apriori-like methods, especially in the situations where large data sets, low support threshold, and long patterns exist. The specificity of CMAR is also that it finds frequent pattern and generates rules in one step.

For every pattern, CMAR maintains the distribution of various class labels among data objects matching the pattern. This is done without any overhead in the procedure of counting (conditional) databases. On the other hand, CMAR uses class label distribution to prune. Once a rule is generated, it is stored in a CR-tree.

The number of rules generated by class-association rule mining can be huge. To make the classification effective and also efficient, we need to prune rules to delete redundant and noisy information. According to the facility of rules on classification, a global order of rules is composed.

CMAR employs the following methods for rule pruning:

1. Using general and high-confidence rule to prune more specific and lower confidence ones.
2. Selecting only positively correlated rules.
3. Pruning rules based on database cover.

In the phase of classification, for a given data object, CMAR selects a small set of high confidence matching the object, highly related rules and analyzes the correlation among those rules.

7.3. ARC-AC and ARC-BC

In 2002, Zaïane and Antonie offered new associative-based classifiers for text categorization – ARC-AC and ARC-BC [Zaïane and Antonie, 2002]. For building association rules they used Apriori-like algorithm. They have considered two different approaches for extracting term-category association rules and for combining those rules to generate a text classifier.

In the first approach ARC-BC (Association Rule-based Categorizer by Category), each category is considered as a separate collection and the association rule mining applied to it. Once the frequent item-sets are discovered, the rules are simply generated by making each frequent item-set the antecedent of the rule and the current category the consequent.

The ARC-AC (Association Rule-based Categorizer for All Categories) considers all categories at whole. In this case one antecedent can be found with different consequents. During the recognition they introduce "dominant factor", which is the proportion of rules of the most dominant category in the applicable rules to the query.

7.4. CPAR

A greedy associative classification algorithm called CPAR was proposed in [Yin and Han, 2003]. CPAR adopts FOIL [Quinlan and Cameron-Jones, 1993] strategy in generating the rules from data sets. It seeks for the best rule condition that brings the most gain among the available ones in the data set. Once the condition is identified, the weights of the positive examples associated with it will be deteriorated by a multiplying factor, and the process will be repeated until all positive examples in the training data set are covered.

The searching process for the best rule condition is time consuming process for CPAR since the gain for every possible item needs to be calculated in order to determine the best item gain. Thus, CPAR uses an efficient data structure, i.e. PN Array, to store all the necessary information for calculation of the items gain. In the rules generation process, CPAR derives not only the best condition but all close similar ones since there are often more than one attribute items with similar gain.

7.5. CorClass

CorClass [Zimmermann and De Raedt, 2004] directly finds the best correlated associations rules for classification by employing a branch-and-bound algorithm, using so called Morishita & Sese Framework [Morishita and Sese, 2000]. It follows the strategy in which calculating the upper bounds on the values attainable by specializations of the rule currently considered. The upper bound finally allows dynamic rising of the pruning threshold, differing from the fixed minimal support used in existing techniques. This will result in earlier termination of the mining process. Since the quality criterion for rules is used directly for pruning, no post-processing of the discovered rule set is necessary.

The algorithm uses two strategies for classifying a new object

1. Decision List: Rank all the rules (rules are ranked by quality according to some criterion) and use the first rule satisfied by an example for classification.
2. Weighted Combination: The general way to do this is to collect all such rules, assign each one a specific weight and for each class predicted by at least one rule sum up the weights of corresponding rules. The class value having the highest value is returned.

7.6. ACRI

The task of ACRI (Associative Classifier with Reoccurring Items) [Rak et al, 2005] is to combine the associative classification with the problem of recurrent items.

A delicate issue with associative classifiers is the use of a subtle parameter: support. Support is a difficult threshold to set, inherited from association rule mining. It is known in the association rule mining field that the support threshold is not obvious to tune in practice. The accuracy of the classifier can be very sensitive to this parameter.

The algorithm for mining associations in ACRI is based on earlier work of the authors Apriori-based MaxOccur [Zaiane et al, 2000]. The building of the classification model follows their previous ARC-BC approach. The rational is based on the efficiency of this method in the case of non-evenly distributed class labels. MaxOccur run on transactions from each known class separately makes the core of the rule generator module. It mines the set of rules with reoccurring items from the training set.

These rules associate a condition set with a class label such that the condition set may contain items preceded by a repetition counter. The classification process might be considered as plain matching of the rules in the model to the features of an object to classify. Different classification rules may match, thus the classifier module applies diverse strategies to select the appropriate rules to use.

In addition, simple matching is sometimes not possible because there is no rule that has the antecedent contained in the feature set extracted from the object to classify. With other associative classifiers, a default rule is applied, either the rule with the highest confidence in the model or simply assigning the label of the dominant class. The ACRI approach has a different strategy allowing partial matching or closest matching by modeling antecedents of rules and new objects in a vector space.

7.7. TFPC

TFPC (Total From Partial Classification) [Coenen and Leng, 2005] is a classification association rule mining algorithm founded on the TFP (Total From Partial) association rule mining algorithm; which, in turn, is an extension of the Apriori-T (Apriori Total).

TFP (Total From Partial) algorithm builds a set enumeration tree structure, the P-tree, that contains an incomplete summation of support-counts for relevant sets. Using the P-tree, the algorithm uses an Apriori-like procedure to build a second set enumeration tree, the T-tree, that finally contains all the frequent sets (i.e. those that meet the required threshold of support), with their support-counts. The T-tree is built level by level, the first level comprising all the single items (attribute-values) under consideration. In the first pass, the support of these items is counted, and any that fail to meet the required support threshold are removed from the tree. Candidate-pairs are then generated from remaining items, and appended as child nodes. The process continues, as with Apriori, until no more candidate sets can be generated.

The class-competition is solved by using support and confidence measures.

7.8. HARMONY

HARMONY [Wang and Karypis, 2005] directly mines for each training instance one of the highest confidence classification rules that it supports and satisfies a user specified minimum support constraint, and builds the classification model from the union of these rules over the entire set of instances. Thus HARMONY employs an instance-centric rule generation framework and mines the covering rules with the highest confidence for each instance, which can achieve better accuracy. Moreover, since each training instance usually supports many of the discovered rules, the overall classifier can better generalize to new instances and thus achieve better classification performance.

To achieve high computational efficiency, HARMONY mines the classification rules for all the classes simultaneously and directly mines the final set of classification rules by pushing deeply some effective pruning methods into the projection-based frequent item-set mining framework. All these pruning methods preserve the completeness of the resulting rule-set in the sense that they only remove from consideration rules that are guaranteed not to be of high quality.

7.9. MCAR

MCAR (Multi-class Classification based on Association Rule) [Thabtah et al, 2005] uses an efficient technique for discovering frequent items and employs a rule ranking method which ensures detailed rules with high confidence.

During the rules generation MCAR scans the training data set to discover frequent single items, and then recursively combines the items generated to produce items involving more attributes. After that the rules are used to generate a classifier by considering their effectiveness on the training data set, using expanded definition of "precedence".

$P^1 \succ P^2$ (rule P^1 precedes rule P^2) if:

$confidence(P^1) > confidence(P^2)$;

$confidence(P^1) = confidence(P^2)$ but $support(P^1) > support(P^2)$;

$confidence(P^1) = confidence(P^2)$, $support(P^1) = support(P^2)$,

but $ActAcc(P^1) = ActAcc(P^2)$;

All conditions before are the same, but $card(P^1) < card(P^2)$;

Last condition: P^1 is generated earlier than P^2.

7.10. CACA

The following innovations are integrated in CACA [Tang and Liao, 2007]:

- use the class-based strategy to cut down the searching space of frequent pattern;
- design a structure call Ordered Rule Tree (OR-Tree) to store the rules and their information which may also prepare for the synchronization of the two steps;
- redefine the compact set so that the compact classifier is unique and not sensitive to the rule reduction;
- synchronize the rule generation and building classifier phases.

Class-based strategy: Given a training data set D with k classes, the principle idea of class based rule mining is to divide the single attribute value set C_{all} for all classes into k smaller ones for every class, that is, to limit the searching in k low dimensional spaces other than a high dimensional one.

OR-Tree: To facilitate they design a structure, called Ordered-Rule-Tree (OR-Tree), under the inspiration of CR-Tree used in CMAR to store and rank rules. It is composed with a tree structure and an ordered list. When a rule $P^i = (c^i \mid a_1^i, ..., a_n^i)$ satisfying the support and confidence thresholds is generated, attribute values $a_1^i, ..., a_n^i$ are stored as nodes in this tree according to their frequency in D in descending order. The last node points to an information node storing the rule's information such as class label, support and confidence. Each rule can and only can have one information node. The ordered list is designed to organize all rules in the tree. Each node in the chain points to a certain rule. Nodes pointing to the rules with higher priority are closer to the head node, while those pointing to the rules with lower priority are farther from the head node.

The ranking rule criteria are as follows:

$P^1 \succ P^2$ (P^1 precedes P^2) if:

$confidence(P^1) > confidence(P^2)$;

$confidence(P^1) = confidence(P^2)$ but $support(P^1) > support(P^2)$;

$confidence(P^1) = confidence(P^2)$ and $support(P^1) = support(P^2)$ but $card(P^1) < card(P^2)$ (P^1 is more general than P^2);

Equal previous conditions, but P^1 is generated earlier then P^2.

To ensure compact classifier to be unique and not sensitive to the rule reduction, the redundant rules are defined as follows:

Definition of redundant rule: Given P^1, P^2 and P^3, that belong to rule set R, P^2 is redundant if:

– $P^1 = (c^1 \mid a_1^1,...,a_{k1}^1)$, $P^2 = (c^2 \mid a_1^2,...,a_{k2}^2)$:

$c^1 \neq c^2$, $(a_1^1,...,a_{k1}^1) \subseteq (a_1^2,...,a_{k2}^2)$, $P^1 \succ P^2$;

– $P^1 = (c^1 \mid a_1^1,...,a_{k1}^1)$, $P^2 = (c^2 \mid a_1^2,...,a_{k2}^2)$:

$c^1 = c^2$, $(a_1^1,...,a_{k1}^1) \subset (a_1^2,...,a_{k2}^2)$, $P^1 \succ P^2$;

– $P^1 = (c^1 \mid a_1^1,...,a_{k1}^1)$, $P^2 = (c^2 \mid a_1^2,...,a_{k2}^2)$, $P^3 = (c^3 \mid a_1^3,...,a_{k3}^3)$:

$c^1 = c^2 \neq c^3$, $(a_1^1,...,a_{k1}^1) \subset (a_1^2,...,a_{k2}^2)$, $(a_1^1,...,a_{k1}^1) \subset (a_1^3,...,a_{k3}^3)$, $P^1 \succ P^2 \succ P^3$.

Definition of compact rule set: For rule set R, if $\mathrm{R}' \subset \mathrm{R}$, any redundant rule $P \notin \mathrm{R}'$ and R' is unique, then R' is the compact set of R.

CACA technically combined the rule generation and the building classifier phases together. Once a new rule is generated, the algorithm visits the OR-Tree partially to recognize its redundancy, stores it in the OR-Tree and ranks it in the rule set. Not only can the synchronization simplify the procedure of associative classification but also apply the pruning skill to shrink the rule mining space and raise the efficiency.

7.11. ARUBAS

In contrast with many existing associative classifiers, ARUBAS [Depaire et al, 2008] uses class association rules to transform the feature space and uses instance-based reasoning to classify new instances. The framework allows the researcher to use any association rule mining algorithm to produce the class association rules. Five different fitness measures are used for classification purposes.

The main idea behind the ARUBAS framework, is transformation of the original feature space into a more powerful feature space. The original feature space is called the attribute space, where each record $R^i = (c^i \mid a_1^i,...,a_n^i)$ is coded as a set of attribute values and a class value.

In attribute space, each dimension consists of a single attribute. In the new feature space, which we will call pattern space, each dimension will consist of a combination of attributes, also called a pattern, which is denoted as $P_p = \langle (A_{i1}, a_{i1}),...,(A_{ik}, a_{ik}) \rangle$. For achieving more power for the feature space, only combinations of attributes (or patterns) which are strongly associated with a single class value is given.

The first step in the ARUBAS framework is to use any CAR mining technique to find a set of CARs, which is used to transform the feature space. The antecedent of each CAR, which represents an itemset, will become a pattern P_p and hence a dimension in the new feature space. The value of an instance R^i for a pattern dimension P_p is 1 (if the instance contains the pattern) or 0 (if it doesn't).

The instance similarity is used for classifying new instances. To measure the similarity between a new instance R^i and a known training instance R^t ARUBAS focuses on the patterns contained by both instances and how many patterns both instances have in common, but on those patterns coming from the CARs which predicted the class value of the training instance R^t.

The main idea behind the association rule based similarity framework is that classification is based on similarity between a new instance and an entire class. This similarity is not measured in the original attribute space, but in the pattern space, which is constructed by means of CARs.

8. Conclusion

This chapter provided an overview of the area of CAR-classifiers. CAR algorithms have its important place in the family of classification algorithms with several advantages, such as: efficiency of the training regardless of the training set; easy handling with high dimensionality; very fast classification; high accuracy; human comprehensible classification model.

We observed all typical steps in the whole classification process of CAR algorithms: generating the rules, pruning, and recognition.

In the phase of generating the rules several techniques are observed: the pioneer AIS, most used Apriori, alternative FP-Tree, TreeProjection, Matrix Algorithm, Sampling Algorithms, Partition, FOIL and Morishita & Sese Framework.

The pruning is important step in the learning process of CAR algorithms, applied as preprocessing step, in parallel of association rule mining or after it. Here we made a brief observation of several rule quality measures and rule ordering schemes, used in CAR algorithms.

In the recognition phase we also observed different types of choosing final decision – using simple rule or set of rules with different types of ordering schemas.

Finally, using the proposed framework, typical for CAR algorithms, we analyze the some representatives of CAR algorithms: CBA, CMAR, ARC-AC and ARC-BC, CPAR, CorClass, ACRI, TFPC, HARMONY, MCAR, CACA, ARUBAS, showing wide variety of proposed techniques.

Bibliography

[Agarwal et al, 2000] Agarwal, R. Aggarwal, C., Prasad V.: A tree projection algorithm for generation of frequent item-sets. In J. of Parallel and Distributed Computing, 61/3, 2001, pp.350-371.

[Agrawal and Srikant, 1994] Agrawal, R., Srikant, R.: Fast algorithms for mining association rules. In Proc. 20th Int. Conf. Very Large Data Bases, 1994, pp.487-499.

[Agrawal et al, 1993] Agrawal, R., Imieliński, T., Swami, A.: Mining association rules between sets of items in large databases. Proc. of the ACM SIGMOD Int. Conf. on Management of Data, Washington, DC, 1993, pp. 207-216.

[Ashrafi et al, 2004] Ashrafi, M., Taniar, D. Smith, K.: A new approach of eliminating redundant association rules, LNCS, Vol.3180, 2004, pp.465-474.

[Baralis and Psaila, 1997] Baralis, E., Psaila, G.: Designing templates for mining association rules. Journal of Intelligent Information Systems, 9/1, 1997, pp.7-32.

[Bastide et al, 2000] Bastide, Y., Taouil, R., Pasquier, N., Stumme, G., Lakhal, L.: Mining frequent patterns with counting inference. ACM SIGKDD Explorations Newsletter, 2000

[Bay, 2000] Bay, S.: Multivariate discretization of continuous variables for set mining. In Proc. of the 6th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, 2000, pp.315-319.

[Brin et al, 1997] Brin, S., Motwani, R., Ullman, J., Tsur, S.: Dynamic itemset counting and implication rules for market basket data. In Proc. ACM SIGMOD Int. Conf. on Management of Data, 1997, pp.255-264.

[Cheung et al, 1996] Cheung, D., Han, J., Ng, V., Fu, A., Fu, Y.: A fast distributed algorithm for mining association rules. In Proc. of 1996 Int. Conf. on Parallel and Distributed Information Systems, Miami Beach, Florida, 1996, pp.31-44.

[Chuang et al, 2005] Chuang, K., Chen, M., Yang, W.: Progressive sampling for association rules based on sampling error estimation, LNCS, Vol.3518, 2005, pp.505-515.

[Coenen and Leng, 2005] Coenen, F., Leng, P.: Obtaining Best Parameter Values for Accurate Classification. Proc. ICDM'2005, IEEE, pp.597-600.

[Coenen et al, 2004] Coenen, F., Goulbourne, G., Leng, P.: Tree structures for mining association rules. Data Mining and Knowledge Discovery, Kluwer Academic Publishers, 8, 2004, pp.25-51.

[Cristofor and Simovici, 2002] Cristofor, L., Simovici, D.: Generating an informative cover for association rules. In Proc. of the IEEE Int. Conf. on Data Mining, 2002.

[Depaire et al, 2008] Depaire, B., Vanhoof, K., Wets, G.: ARUBAS: an association rule based similarity framework for associative classifiers. IEEE Int. Conf. on Data Mining Workshops, 2008, pp.692-699.

[Do et al, 2003] Do, T.D., Hui, S.-C., Fong, A.: Mining frequent item-sets with category-based constraints, LNCS, Vol. 2843, 2003, pp.76-86.

[Goethals, 2002] Goethals, B.: Efficient Frequent Pattern Mining. PhD Thesis, Transnationale Univeriteit Limburg, 2002.

[Han and Pei, 2000] Han, J., Pei, J.: Mining frequent patterns by pattern-growth: methodology and implications. ACM SIGKDD Explorations Newsletter 2/2, 2000, pp.14-20.

[Hilderman and Hamilton, 2002] Hilderman, R., Hamilton, H.: Knowledge Discovery and Interest Measures. Kluwer Academic, Boston, 2002.

[Jaroszewicz and Simovici, 2002] Jaroszewicz, S., Simovici, D.: Pruning redundant association rules using maximum entropy principle, LNCS, Vol. 2336, 2002, pp.135-142.

[Kotsiantis and Kanellopoulos, 2006] Kotsiantis, S., Kanellopoulos, D.: Association rules mining: a recent overview. GESTS International Transactions on Computer Science and Engineering, 32/1, 2006, pp. 71-82

[Kuncheva, 2004] Kuncheva, L.: Combining Pattern Classifiers: Methods and Algorithms. Willey, 2004.

[Li and Gopalan, 2004] Li, Y., Gopalan, R.: Effective sampling for mining association rules. LNCS, Vol. 3339, 2004, pp.391-401.

[Li et al, 2001] Li, W., Han, J., Pei, J.: CMAR: Accurate and efficient classification based on multiple class-association rules. In: Proc. of the IEEE Int. Conf. on Data Mining ICDM, 2001, pp.369-376.

[Liu et al, 1998] Liu, B., Hsu, W., Ma, Y.: Integrating classification and association rule mining. In Knowledge Discovery and Data Mining, 1998, pp.80-86.

[Liu et al, 1999] Liu, B. Hsu, W., Ma, Y.: Mining association rules with multiple minimum supports. Proc. Knowledge Discovery and Data Mining Conf., 1999, pp.337-341.

[Mitchell, 1997] Mitchell, T.: Machine Learning, McGraw-Hill, 1997.

[Morishita and Sese, 2000] Morishita, S., Sese, J.: Transversing itemset lattices with statistical metric pruning. In Proc. of the 19th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, 2000, pp.226-236.

[Parthasarathy et al, 2001] Parthasarathy, S., Zaki, M., Ogihara, M., Li, W.: Parallel data mining for association rules on shared memory systems. Journal Knowledge and Information Systems, 3/1, 2001, pp. 1-29.

[Parthasarathy, 2002] Parthasarathy, S.: Efficient progressive sampling for association rules. Proc. of Int. Conf. on Data Mining, 2002, pp.354-361.

[Quinlan and Cameron-Jones, 1993] Quinlan, J., Cameron-Jones, R.: FOIL: A midterm report. In Proc. of European Conf. On Machine Learning, Vienna, Austria, 1993, pp.3-20.

[Rak et al, 2005] Rak, R., Stach, W., Zaiane, O., Antonie M.-L.: Considering re-occurring features in associative classifiers. In Advances in Knowledge Discovery and Data Mining, LNCS, Vol. 3518, 2005, pp.240-248.

[Sarwar et al, 2001] Sarwar, B., Karypis, G., Konstan, J., Reidl, J.: Item-based collaborative filtering recommendation algorithms. In: World Wide Web, 2001, pp.285-295.

[Savasere et al, 1995] Savasere, A., Omiecinski, E., Navathe, S.: An efficient algorithm for mining association rules in large databases. The 21st VLDB Conference, 1995, pp.

[Tang and Liao, 2007] Tang, Z., Liao, Q.: A new class based associative classification algorithm. IAENG International Journal of Applied Mathematics, 2007, 36/2, pp.15-19.

[Thabtah et al, 2005] Thabtah, F., Cowling, P., Peng, Y.: MCAR: multi-class classification based on association rule. In Proc. of the ACS/IEEE 2005 Int. Conf. on Computer Systems and Applications, Washington, DC, 2005, p.33.

[Toivonen, 1996] Toivonen, H.: Sampling large databases for association rules. In The VLDB Journal, 1996, pp.134-145.

[Wang and Karypis, 2005] Wang, J., Karypis, G.: HARMONY: Eficiently Mining the Best Rules for Classification. In Proc. of SDM, 2005, pp.205-216.

[Yin and Han, 2003] Yin, X., Han, J.: CPAR: Classification based on predictive association rules. In SIAM Int. Conf. on Data Mining (SDM'03), 2003, pp.331-335.

[Yuan and Huang, 2005] Yuan, Y., Huang, T.: A matrix algorithm for mining association rules. LNCS, Vol. 3644, 2005, pp.370-379.

[Zaiane and Antonie, 2002] Zaiane, O., Antonie, M.-L.: Classifying text documents by associating terms with text categories. J. Australian Computer Science Communications, 24/2, 2002, pp.215-222.

[Zaiane and Antonie, 2005] Zaiane, O., Antonie, M.-L.: On pruning and tuning rules for associative classifiers. In Proc. of Int. Conf. on Knowledge-Based Intelligence Information & Engineering Systems, LNCS, Vol. 3683, 2005, pp.966-973.

[Zaiane et al, 2000] Zaiane, O., Han, J., Zhu, H.: Mining recurrent items in multimedia with progressive resolution refinement. In Int. Conf. on Data Engineering, 2000, pp.461-470.

[Zimmermann and De Raedt, 2004] Zimmermann, A., De Raedt, L.: CorClass: Correlated association rule mining for classification. In Discovery Science, LNCS, Vol. 3245, 2004, pp.60-72.

Authors' Information

Ilia Mitov *– Institute of Mathematics and Informatics, BAS, Bulgaria; e-mail: mitov@foibg.com*

Major Fields of Scientific Research: *Multi-dimensional information systems; Data Processing and Mining*

Stefan Karastanev *– Institute of Mechanics, BAS, Bulgaria;*

e-mail: stefan@imbm.bas.bg

Major Fields of Scientific Research: *Software Engineering, Data Processing and Mining, Data structures in information systems.*

Krassimir Markov *– Institute of Information Theories and Applications, Bulgaria e-mail: markov@foibg.com*

Major Fields of Scientific Research: *General theoretical information research, Multi-dimensional information systems; Data Processing and Mining*

TABLE OF CONTENTS