Infocommunications Journal

A PUBLICATION OF THE SCIENTIFIC ASSOCIATION FOR INFOCOMMUNICATIONS (HTE)

June 2013 Volume V Number 2 ISSN 2061-2079

Technically Co-Sponsored by
IEEE COMMUNICATIONS SOCIETY

hte

IEEE HUNGARY SECTION

Editorial Board

Editor-in-Chief: CSABA A. SZABO, Budapest University of Technology and Economics (BME), Hungary

ÖZGÜR B. AKAN
Koc University, Istanbul, Turkey
JAVIER ARACIL
Universidad Autónoma de Madrid, Spain
LUIGI ATZORI
University of Cagliari, Italy
STEFANO BREGNI
Politecnico di Milano, Italy
LEVENTE BUTTYÁN
Budapest University of Technology and Economics, Hungary
TEREZA CRISTINA CARVALHO
University of Sao Paulo, Brasil
TIBOR CINKLER
Budapest University of Technology and Economics, Hungary
FRANCO DAVOLI
University of Genova, Italy
VIRGIL DOBROTA
Technical University Cluj, Romania
KÁROLY FARKAS
Budapest University of Technology and Economics, Hungary
VIKTORIA FODOR
KTH, Royal Institute of Technology, Sweden
AURA GANZ
University Massachusetts at Amherst, USA
EROL GELENBE
Imperial College London, UK
MARIO GERLA
UCLA, Los Angeles, USA
ENRICO GREGORI
CNR IIT, Italy
CHRISTIAN GUETL
Graz University of Technology, Austria
ASHWIN GUMASTE
Indian Institute of Technology Bombay, India
LAJOS HANZO
University of Southampton, UK
THOMAS HEISTRACHER
Salzburg University of Applied Sciences, Austria
JUKKA HUHTAMÄKI
Tampere University of Technology, Finland
SÁNDOR IMRE
Budapest University of Technology and Economics, Hungary
EBROUL IZGUEIRDO
Queen Mary University of London, UK
RAJ JAIN
Washington University in St. Lois, USA
ANDRZEJ JAJSZCZYK
AGH University of Science and Technology, Krakow, Poland
LÁSZLÓ T. KÓCZY
Széchenyi University, Győr, Hungary
ANDREY KOUCHERYAVY
St. Petersburg State University of Telecommunications, Russia
DÓRA MAROS
Óbuda University, Budapest, Hungary
MAJA MATIJASEVIC
University of Zagreb, Croatia
VASHEK MATYAS
Masaryk University, Brno, Czechia
OSCAR MAYORA
Create-Net, Trento, Italy
MIKLÓS MOLNÁR
Université Montpellier 2, France
JAUDELICE DE OLIVEIRA
Drexel University, USA
ALGIRDAS PAKSTAS
London Metropolitan University, UK
MICHAL PIORO
Warsaw Technical University, Poland
ROBERTO SARACCO
EIT ICT LABS, Italy
BURKHARD STILLER
University of Zürich, Switzerland
JÁNOS SZTRIK
University of Debrecen, Hungary
YUTAKA TAKAHASHI
Kyoto University, Japan
DAMLA TURGUT
University of Central Florida, USA
ESZTER UDVARY
Budapest University of Technology and Economics, Hungary
SCOTT VALCOURT
University of New Hampshire, USA
WENYE WANG
North Carolina State University, USA
ADAM WOLISZ
Technical University Berlin, Germany
JINSONG WU
Bell Laboratories, China
GERGELY ZARUBA
University of Texas at Arlington, USA
HONGGANG ZHANG
Ecole Supérieur d'Electricité, France

Indexing information

Infocommunications Journal is covered by Inspec, Compendex and Scopus.

Infocommunications Journal

Technically co-sponsored by IEEE Communications Society and IEEE Hungary Section

Editorial Office (Subscription and Advertisements)**:**
Scientific Association for Infocommunications
H-1055 Budapest, Kossuth Lajos tér 6-8, Room: 422
Mail Address: 1372 Budapest Pf. 451. Hungary
Phone: +36 1 353 1027, Fax: +36 1 353 0451
E-mail: info@hte.hu
Web: www.hte.hu

Articles can be sent also to the following address:
Budapest University of Technology and Economics
Department of Networked Systems and Services
Tel.: +36 1 463 3261, Fax: +36 1 463 3263
E-mail: szabo@hit.bme.hu

Subscription rates for foreign subscribers:
4 issues 50 USD, single copies 15 USD + postage

Publisher: PÉTER NAGY • Manager: ANDRÁS DANKÓ

HU ISSN 2061-2079 • Layout: MATT DTP Bt. • Printed by: FOM Media

www.infocommunications.hu

Multicast Routing in WDM Networks without Splitters

Dinh Danh Le, Miklós Molnár and Jérôme Palaysi

***Abstract*—Multicasting in WDM core networks is an efficient way to economize network resources for several multimedia applications. Due to their complexity and cost, multicast capable switches are rare in the proposed architectures. The paper investigates the multicast routing without splitters in directed (asymmetric) graphs. The objective is to minimize the number of used wavelengths and if there are several solutions, choose the lowest cost one. We show that the optimal solution is a set of light-trails. An efficient heuristic is proposed to minimize conflicts between the light-trails, and so to minimize the number of used wavelengths. The performance is compared to existing light-trail based heuristics. Our algorithm provides a good solution with a few wavelengths required and a low cost.**

***Keywords*: WDM network, multicast routing, light-trail, wavelength minimization, heuristic**

I. INTRODUCTION

All-optical networks are serious candidates to become high speed backbone networks with huge capacity. In optical routing, the messages are transmitted by light signal without electronic processing. Routes should satisfy the physical (optical) constraints in static connection based networks and also in the case of burst and packet switching.

Multicast communications are present in networks to efficiently perform data transmission from a source to several destinations. Usually, without physical constraints, multicast routes corresponds to trees in the topology graph. To perform multicast, there should be multicast capable nodes (splitters) at all the branching nodes of the tree. However, one of the most hard constraints for optical multicasting is the constraint on the availability of light splitters in the switches. In fact, splitters are expensive and the light power can be decreased considerably by splitting (inversely proportional with the number of outgoing ports [2]). This constraint prevents all-optical multicasting from employing splitters.

In our paper, we investigate an interesting question: how to perform multicast without splitters? Trivially, a set of light-paths from the source to the destinations can be used as a solution, but this solution is expensive in term of wavelengths. Our objective is to perform multicasting without splitters and minimizing the number of used wavelengths. Solutions in bidirectional networks (where a wavelength is available in both directions between the connected switches) are known, but we investigate the arbitrarily directed case which is very practical. Even if the network is designed to be bidirectional, when some demands hold some of the resources of the network, the resulting network graph is now arbitrarily directed, therefore the routing for subsequent demands will be calculated on a digraph.

Due to its interest, WDM multicast routing has been investigated intensively in the literature and several propositions exist to adapt multicast routing algorithms to the optical constraints (cf. [9] for some basic algorithms and [11] for a survey). The minimization of the number of used wavelengths was investigated at first in [5] where the wavelengths are supposed to be unevenly distributed in the networks. The considered network is assumed to be equipped with splitters and wavelength converters. The multicasting is based on a tree: the objective is to construct a tree T meeting optical constraints such that the number of wavelengths used to cover T is minimized. The NP-hardness of the problem is proved, and an approximation algorithm has been proposed. An improved approximation can be found in [8].

The case of switching without splitters in symmetric networks has been discussed in [1]. The problem is to find a Multiple-Destination Minimum Cost Trail (MDMCT) that starts from a source and spans all the destinations with minimizing the total cost of the edges traversed. To ensure a feasible solution, a low-cost cross-connect architecture called Tap-and-Continue (TaC) has been proposed to replace splitters. TaC cross-connects can tap a signal with small power at the local station and forward it to one of its output ports. Moreover, every link is assumed to be equipped with at least two fibers in order to support bidirectional transmission on the same link.

The authors proved that the MDMCT problem is NP-hard and then developed a heuristic (called MDT) that finds a feasible trail in polynomial time. The algorithm has two steps. The first step is computing an approximated Steiner tree for a multicast request using the Minimum Cost Path Heuristic proposed in [7]. A trail is then computed based on the backtracking method following the tree.

The advantage of MDT heuristic is that it uses only one wavelength (and one transmitter) for each multicast request (and thus, the wavelength is minimized). However, because of multitude of round-trip traversing, a large number of links is required in both directions, hence the total cost and the diameter of the light-trail is always very high. To improve the total cost, it is necessary to reduce the round-trip traversing. Moreover, it is worth noting that, the source can inject the light signal by multiple transmitters independently. By taking this feature into account, one can considerably reduce the reversal arcs (that backtrack to the source), then the total cost and

Manuscript submitted April 2013, revised May 28, 2013. This work is sponsored in part by VIED, Vietnam.

Dinh Danh Le, Miklós Molnár and Jérôme Palaysi are with the laboratory LIRMM, Université Montpellier 2, France (email: dinhdanh.le@lirmm.fr, miklos.molnar@lirmm.fr, jerome.palaysi@lirmm.fr).

the diameter can also be reduced. This is the idea to make a modified version of MDT, called MMDT that is detailed in Section V.

In [3], Der-Rong Din posed the Minimal Cost Routing Problem which minimizes the cost under WDM symmetric networks using only TaC cross-connects. Unlike the approach of [1] that based on light-trail, the approach of Der-Rong Din is based on *light-forest* (a set of the *light-trees* [11]), rooted at the source and covering all the destinations. Besides, the source can inject the signal by multiple transmitters so that each light-tree can use a single wavelength. Furthermore, to produce a trade-off between the total cost of the light-forest and the number of wavelengths used, the author developed an objective function which combines the actual total cost of the light-forest and the cost for using wavelengths: $f = cost(F) + \alpha * numWL$, in which F is the resultant light-forest, $numWL$ is the number of wavelengths used, and α is a specific coefficient.

The author proposed two heuristic algorithms, namely Farthest-Greedy (FG) and Nearest-Greedy (NG). The two algorithms are based on the shortest path tree (SPT). The idea of these algorithms is: first construct the SPT from the source to the destinations, then keep one path for each subtree of the source, and finally reroute the other destinations that have not been reached (*unreached* destinations). The difference between the two algorithms is: FG keeps the farthest (in term of cost) destination routed by the computed shortest path, and chooses the farthest destination in the unreached set to reroute, whereas NDF keeps the nearest (minimal cost) destination and chooses the nearest destination in unreached set to reroute in the rerouting phase. The rerouting phase is performed by the shortest paths from the source or from the leaves of computed trees to each unreached destination, that do not share any nodes and edges with all the computed trees (each tree is computed in a different wavelength graph that is initialized by the original graph). When there is no possible path in the computed trees or the path exists but with larger cost than the path found in the new wavelength graph, the unreached destination is routed by the shortest path found in the new tree with a new wavelength. The author also gave the comparison between FG, NG and MDT by simulations, and the results show that FG is better than NG and MDT.

Most of the solutions proposed in the literature (excluding MDT) are based on simple routes in which cycles are not allowed. However, one can operate multicasting by non-simple routes which permit nodes to be visited several times, as long as the routes using the same wavelength are arc-disjoint. MDT in [1] gives a special structure which allow cycles, but with special cycles which are 2-cycles[1]. In fact, one can construct structures that allow not only 2-cycles but also arbitrary cycles. These structures correspond to a hierarchy that was proposed in [6]. For multicast routing in WDM networks, the light-hierarchy concept has been proposed in [10]. A light-hierarchy is a hierarchy using a single wavelength.

In the paper, we show that the optimal route minimizing the number of wavelengths is a set of (non-simple) light-trails. The computation of the optimum in directed graphs is NP-hard. So, we propose some heuristic algorithms, which try to minimize the number of wavelengths, taking into account the availability of fibers in the network, with a low cost. We compare the performance of them with two previously proposed multicast routing algorithms.

The structure of the paper is the following. Section II formulates the problem. Several useful concepts and properties are given in Section III. Our heuristics are described in Section IV followed by the experimental results in Section V.

II. Problem Formulation

Let us suppose that nodes are equipped by TaC option and can be traversed by the same wavelength several times having different incoming and outgoing ports for each pass. So, light-trails from the source to the destinations can perform the multicast (cf. [1]). There is no splitter in the nodes. Moreover, the topology graph $G = (V, A)$ is an arbitrary directed graph in which each arc represents a fiber between the pair of nodes. Each arc $e \in A$ is associated with the a value $cost(e) > 0$. Let $s \in V$ the source and $D \subseteq V \setminus \{s\}$ the set of destinations. We suppose that the multicast is feasible, i.e., there is at least one directed path from the source to each destination. Different objectives for the multicast routing can be formulated as follows.

Routing using a minimum number of wavelengths (P_1): A multicast route colorable with the least number of wavelengths is required.

Minimum cost routing (P_2): The least cost multicast route is required.

In both cases (with the given constraints), the optimal solution is a set of light-trails routed at the source and covering all of the destinations. Notice that this set corresponds to a hierarchy obtained from a star (cf. [6] for the definition of a hierarchy). The solution of Problem P_1 can be composed from very long trails. The optimum of Problem P_2 can use a high number of wavelengths. Trade-off can be interesting.

Minimum cost multicast routing using a given number of wavelengths (P_3): A minimum cost multicast route colorable with the given number of wavelengths is required.

Length limited multicast route using a minimum number of wavelengths (P_4): A multicast route in which the longest trail is smaller than a given limit and colorable with the minimum number of wavelengths is required.

The solutions of the last two problems are also sets of light-trails.

The mentioned routing problems are hard optimization problems. Problem P_1 corresponds to finding the solution with minimal set of colors. Problem P_2 is equivalent the Degree Constrained Directed Minimum Spanning Tree Problem in the distance graph of the problem.

In our study, we focus on the first problem but some results can be useful to solve Problem P_3 and Problem P_4. For the limited space, the reader is recommended to refer our research paper [4] for the proof of NP-hardness of P_1.

[1] An n-cycle is a cycle with n vertices.

III. USEFUL DEFINITIONS AND PROPERTIES

In order to describe our algorithm, some concepts should be given in the following.

- **Directed shortest path tree (DSPT)**: A directed tree rooted at the multicast source covering all the destinations by shortest paths, in which each path is (one of) the shortest path(s) from the source to a destination. To compute DSPT, any shortest path algorithms, e.g., Dijkstra algorithm, can be employed.
- **Conflict graph:** A graph used to represent the conflicts among the trails. Formally, in our study, a conflict graph is $G_C = (T, E)$, in which T is a set of nodes corresponding to the trails and E is a set of edges such that $e = \{t, p\} \in E$ if and only if there is a conflict between trail t and trail p, i.e., two trails share a common arc. In this study, we just consider conflicts such that shared arcs form the prefix of the concerned trails and this property is preserved during the algorithm. For this condition, each of connected components in the conflict graph corresponds to a subtree of the DSPT (Figure 1).
  Property 1: Each connected component in the conflict graph composes a (conflict) clique[2].
  Trivially, if the shared arcs of conflicting trails are the prefix of the trails, then conflicts are transitive (if there is a conflict between T_1 and T_2 and between T_2 and T_3, then there is a conflict between T_1 and T_3). Indeed, all the trails in the same connected component share the first arc from the source. So the Property 1 follows.
  Property 2: The number of colors needed to color all the nodes of a clique is equal to the number of nodes of that clique.
  Obviously, there is a (conflict) edge between every pair of nodes in each clique. To avoid the conflict, the nodes must be colored with different colors. So the Property 2 follows.
  Property 3: The number of wavelengths needed to perform the routing respecting the *distinct wavelength constraint*[3] in network fibers is equal to the number of nodes (the size) of the maximal clique in the conflict graph in our study case.
  In deed, each clique corresponds to a subtree of the DSPT. These subtrees are arc-disjoint, so the corresponding trails in each clique do not share any arcs with the corresponding trails in the other ones. Thus, the minimal number of colors needed to color all the nodes in the conflict graph is equal to the size of the maximal clique, because we can use some colors that have been used in the maximal clique to re-color the other nodes in the other cliques. Moreover, to guarantee the distinct wavelength constraint, the number of colors needed in each clique is equal to the number of wavelengths needed to assign the corresponding trails in that clique. So this property holds.
  Thus, the problem of minimizing the number of used wavelengths reduces to the problem of minimizing the number of nodes (trails) of the maximal clique in the conflict graph.

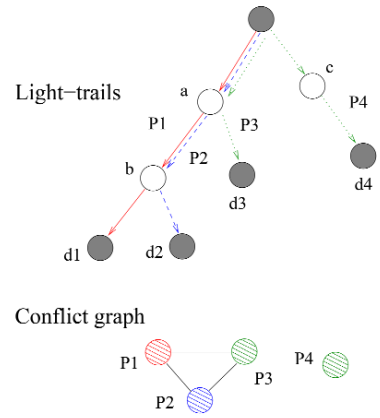


Fig. 1: Example of a DSPT and its conflict graph

Figure 1 illustrates a set of paths (trails) composing a DSPT for the multicast request r = (s, $\{d_1, d_2, d_3\}$) and the corresponding conflict graph. In Figure 1, there are two cliques corresponding to two subtrees of the DSPT. The maximal clique is composed from the paths T_1, T_2, T_3 starting from the source to the destinations d_1, d_2, d_3, respectively. It needs three wavelengths to color the three trails. The other clique composed from only the path T_4 that can re-use one wavelength that were assigned for the maximal clique.

IV. PROPOSED HEURISTICS

A. Algorithm Framework

The idea of the algorithm is to diminish the number of nodes (trails) in the maximal clique of the conflict graph until it cannot be reduced. Informally, the algorithm starts from a set of directed trails (at first, simple trails or paths in the DSPT). Then it tries to iteratively diminish the number of trails in the maximal clique, say C_{max}. At each step, it chooses one trail from the maximal clique, say T_0, that can be replaced by another one. Some mechanism can be employed to select this trail. When the trail is selected, the algorithm looks for all the other trails and choose the one, say T_k, such that the terminal of it has the arc-disjoint shortest path to the first destination of the trail T_0. Then T_0 is replaced by the corresponding trail T_k, and the cardinality of the C_{max} is reduced by 1. The algorithm iterates until the maximal clique cannot be reduced.

The framework of the algorithm consists of four main steps that can be described as follows:

THE ALGORITHM FRAMEWORK

Step 1: Compute a directed shortest path tree (DSPT) from the source s covering all destinations. If there is no branching node in the DSPT, then DONE. Otherwise, do Step 2.

[2]A clique of a graph is a complete subgraph of that graph

[3]*Distinct wavelength constraint:* Different light-paths or light-trees sharing the common link must be allocated distinct wavelengths [11].

Step 2: Compute the conflict graph from the DSPT. Each conflict clique corresponds to a sub-tree rooted at the source of the DSPT.

Step 3: Repeat Step (3.1) to Step (3.4) in the following until the cardinality of the maximal clique C_{max} cannot be reduced (is equal to 1 for the best case).

Step 3.1: Find the maximal clique C_{max}[4].

Step 3.2: Choose a trail T_0 in C_{max} with some mechanism that is mentioned below. Calculate the terminal l_0 of T_0 and the first destination f_0[5] of T_0. For example, in Figure 1 a), C_{max} is composed by the paths T_1, T_2, T_3; the selected trail T_0 is T_3, l_0 is node d_3 and f_0 is also d_3.

Step 3.3: For every terminal l_i of the remaining trails T_i in the set of trails (except T_0), compute the trail T_k such that the path (l_k, f_0) is *arc-disjoint* with all the current trails, and it is the shortest path among the paths (l_i, f_0). If there is an arc-disjoint path from the source s to f_0, then take the shorter one between (l_k, f_0) and (s, f_0). In Figure 1 a), T_k is T_4, and l_k is node d_4.

Step 3.4: Graft the path (l_k, f_0) and the path (f_0, l_0) to the trail T_k, set l_0 as the terminal of the new trail T_k, remove the trail T_0, reduce the cardinality of clique C_{max} by 1. If the path (s, f_0) exists and is selected, create a new trail (s, l_0) and remove the trail T_0.

Step 4: Record the set of trails, the cardinality of clique C_{max} as the minimum number of wavelengths required. Employ the trail-wavelength-assignment (TWA) algorithm (described below) to assign wavelengths for the set of final trails.

B. Trail-wavelength-assignment (TWA) algorithm

The TWA algorithm mentioned in Step 4 works as follows. Let k be the minimum number of wavelengths returned by the routing algorithm above, and $w_1, w_2, .., w_k$ be the k wavelengths reserved for the multicast request. With each wavelength $w_i, i = 1, .., k$, assign w_i for every clique in the set of the remaining cliques, one trail for each clique. Repeat that until there is no trail in the set of remaining cliques.

C. Two greedy heuristics

Our algorithm can be developed to result in two heuristics. In step 3.2, two greedy mechanisms to select the first trail T_0 in the maximal clique C_{max}, resulting in two heuristics of our algorithm, namely Farthest First (FF) if T_0 is the longest trail in term of cost among all the other trails in C_{max}, i.e., the terminal l_0 of T_0 is the *farthest* terminal l_0 among the others in C_{max}; and Nearest First (NF) if l_0 is the nearest terminal in C_{max}. The detail descriptions of the two heuristics and the complexity of them are given in our research paper [4].

D. Illustration of the algorithm

In order to demonstrate the algorithm, we use a network in Figure 2 in which the source is s, the destination set is $D = \{2, 6, 7, 8, 10, 11, 12, 13\}$. Moreover, due to the limited space, and because FF and NF have the same principle, we just illustrate the heuristic FF in the Figure 3 below.

After the Step 1 and Step 2, the DSPT and the initial conflict graph are shown in Figure 3 (a). The maximal clique comprises three paths T_{10}, T_{12}, T_{13}, in which T_{12} is the farthest one, so it is selected first. The first destination f_0 of T_{12} is node 8, the shortest arc-disjoint path computed is the path passing the nodes {10, 5, 8}. Thus T_{12} is replaced by T_{10}, the new trail is $T_{12}^{'}$ (Figure 3 (b)). Similarly, $T_{12}^{'}$ is then replaced by $T_{12}^{''}$ in the next run (Figure 3 (c)). The final set of trails are shown in Figure 3 (d).

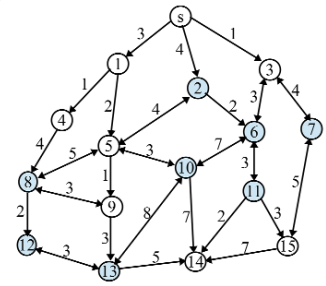
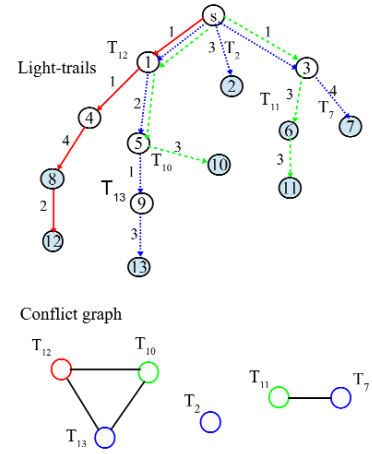


Fig. 2: The network to consider



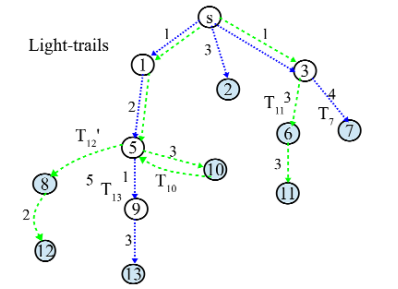
(a) The DSPT and the initial conflict graph

V. EXPERIMENTAL RESULTS

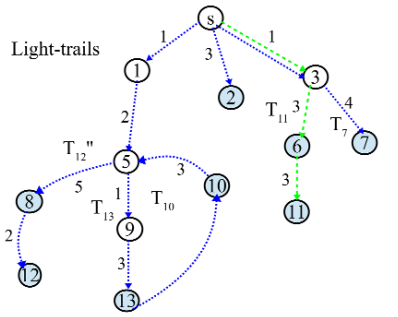
In this section, we show the performance of our algorithm and compare with the algorithms proposed in [1] (MDT) and [3] (Farthest Greedy and Nearest Greedy). In order to fairly compare with MDT in [1], the modified version of it is developed, namely Modified-MDT (MMDT for short) which is mentioned in the following.

[4]To accelerate this step, C_i is organized in a priority queue in which the priority value is the size of C_i, and only cliques C_i with the size larger than 1 are pushed into the queue

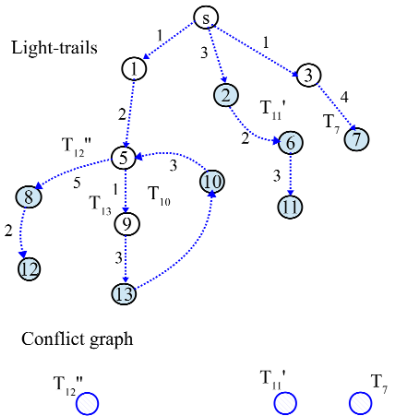
[5]f_0 is the first destination on the path from the nearest branching node of l_0 to l_0



(b) Replace T_{12} by T'_{12}



(c) Replace T'_{12} by T''_{12}



(d) Replace T_{11} by T'_{11}

Fig. 3: Illustration of the Farthest First heuristic

A. MDT and MMDT

As mentioned in Section II, the MDT algorithm has two steps. The first step is to compute an approximated Steiner tree (AST) for a multicast request using the Minimum Cost Path Heuristic (MCPH) proposed in [7]. In the second step, a trail is computed based on the backtracking method following the AST. The backtracking phase starts from the root of the tree, and recursively repeats at each non-leaf node in the tree, say, the *current* node. In the downstream direction, the algorithm tries to include all the downstream links between the *current* node and all its children destinations. Backtracking is required when a leaf node is reached and there are still some destination nodes not yet visited.

However, the total cost and the diameter of the MDT trail are high because of multitude of round-trip traversing. Moreover, it is worth noting that, the source can inject the light signal by multiple transmitters on the same wavelength independently. By taking advantage of this feature, we developed the algorithm MMDT by modifying MDT in the backtracking phase, in such a way that it can eliminate the round-trip traversing the source while using only one wavelength.

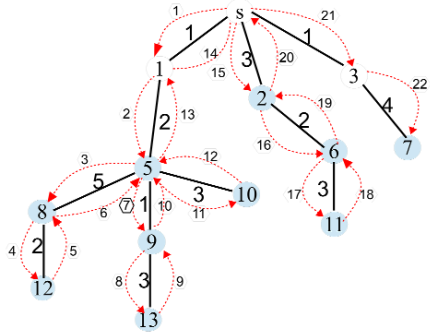
The MMDT works as follows. First, it generates an AST using the MCPH just like the way of MDT. Then the backtracking method to each subtree of the AST (the nodes 1, 2 and 3 in Figure 4) is evoked, with a greedy sequence such that the trails growing to the nearest branch first (in term of cost of the branch). Consequently, there is no reversal arcs needed in the farthest branch for each sub-tree. Accordingly, the result is the set of trails rooted at the source, covering all the destinations with only one wavelength, but with multiple transmitters, one transmitter for each trail. Obviously, the diameter and the total cost of the resultant trials are less than those resulted by MDT.

To demonstrate MDT and MMDT, we use the same topology as the one shown in Figure 2 with a few changes: all the links are now bidirectional and the destination set is $D' = \{2, 5, 6, 7, 8, 9, 10, 11, 12, 13\}$. Figure 4 (a) demonstrates the computation of the multiple-destination trail according to the MDT algorithm and Figure 4 (b) demonstrates the MMDT for the same request $r = (s, D')$. As we can see, MMDT can reduce seven arcs compared with MDT, while both use only one wavelength.

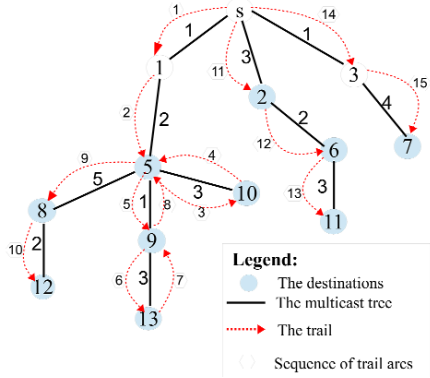
B. Two simulation settings and the performance metrics

Our algorithms can work in arbitrary directed graphs, meaning that unidirectional arcs and edges corresponding to two arcs on opposite directions can coexist in the graph, and the costs for the arcs can be given differently, even with arcs on opposite directions. However, the algorithms proposed in [1] and [3] supposed to work with bidirected graph, in which all the links are all bidirectional. Thus, for a fair comparison, we divided the simulations into two settings. In the first setting, all the algorithms are run on bidirected graphs, and in the second one, they are run on arbitrary directed ones.

Three performance metrics are taken into account in the simulations: the number of wavelengths required, the total cost and the diameter of the resultant routes (light-trails or light-forests). The diameter is defined as the number of maximal hop counts from the source to all the destinations.



(a) The MDT



(b) The MMDT

Fig. 4: Illustration of the MDT and MMDT

The reason for evaluation of this metric is that it can be represented for the end-to-end maximal delay. In fact, the delay can be combined by switching, queueing, transmission and propagation components. In all-optical networks, because of the high light speed, the propagation delay can be assumed to be the same on different links, and it is much less than the other components at the hops. Thus, the number of hops that the light signal has to pass by is usually used to represent for the delay.

C. Experimental results with bidirected graphs

In this setting, the considered algorithms have been run on several random bidirected graphs with different number of nodes $N = \{100, 200, 300\}$, the costs of arcs are randomly selected from the set of integer $\{1, 2, .., 20\}$, and the set of destinations D are also randomly selected with different size $|D| = \{10, 20, .., N/2\}$. To be sure that there is a feasible solution for all the algorithms, the selected graph must be connected and there is at least one directed path from the source to each destination for every simulation. Moreover, in order to guarantee a good confidence interval, for each size $|D|$, we run 100 simulations with different source and destination set. That means, each point in the resultant figures below is calculated on average of 100 (successful) simulations.

Besides, to respect the effect of the coefficient α on the performance of the proposed algorithms in [3] (FG and NG), we also set the coefficient α in $\{50, 100, 150\}$. The simulations showed that, in the cases of $\alpha = \{50, 100\}$ only FG and NG give slightly different results, in which the number of wavelengths is slightly higher and the total cost is slightly lower than those in the case of $\alpha = 150$. Thus we just show the results for the case of $\alpha = 150$. Likewise, we just show the results for the case of $N = \{200, 300\}$ (Figures 5 and 6). In the case of $N = 100$, the results are quantitatively the same.

In Figures 5(a) and 6(a), FG and NG result in large number of wavelengths, ranging from 1 to 5 in 200 node-networks, and from 1 to 9 in 300 node-networks when the group size varies from 10 to $N/2$, while MDT and MMDT and the two heuristics of our algorithm draw a horizontal line with just one wavelength.

In Figures 5(b) and 6(b), the two heuristics proposed in [1] appear with the highest cost, then the two heuristics proposed in [3], the two heuristics of our algorithm have the lowest cost with a small difference between them.

In Figures 5(c) and 6(c), about the diameter, unsurprisingly, the two heuristics proposed in [1] appear with highest diameter, the two heuristics proposed in [3] achieve a constant low diameter, and the two heuristics of our algorithm are in the middle with lower diameter of NF.

Generally, in bidirectional graphs, our algorithms result in the light-trail(s) in which the number of required wavelengths is close to one, low cost but quite high diameter compared with the other algorithms.

D. Experimental results with directed graphs

The configurations of this setting are similar as the ones in the first setting, except that the graphs we work with are all arbitrary directed graphs. Similarly, we just show the figures for the case of $N = \{200, 300\}$, $\alpha = 150$ (Figures 7 and 8). In the other cases ($N = 100$, $\alpha = \{50, 100\}$) the results are quantitatively the same.

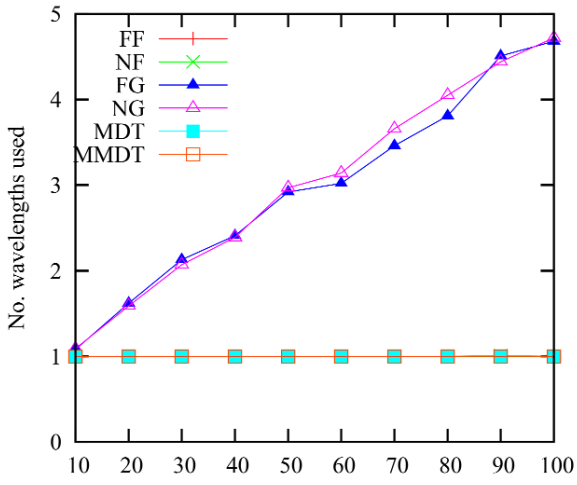
At first, in these configurations, MDT and MMDT almost have no solution so we do not show their results.

In Figures 7(a) and 8(a), all the algorithms result in the increasing number of wavelengths when the group size increases. FG and NG result in larger number of wavelengths with the highest of NG, while FF and NF slowly increase, with the lowest of FF. In short, FF and NF provide a lowest number of wavelengths with a slight difference between them.

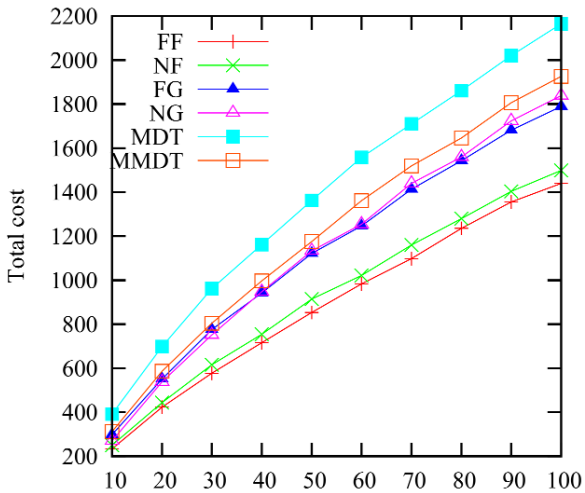
In Figures 7(b) and 8(b), all the algorithms appear with the same cost but FF being a slightly better.

In Figures 7(c) and 8(c), again, FG and NG achieve a constant low diameter with a slight difference between them (lower than 25% of the number of nodes). Between the two our heuristics, NF results in lower diameter (lower than 35% of the number of nodes). Especially, it gets closer to the FG and NG, and almost the same when the group size gets closer to 50%. In contrast, FF results in high diameter (around 50% of the number of nodes).

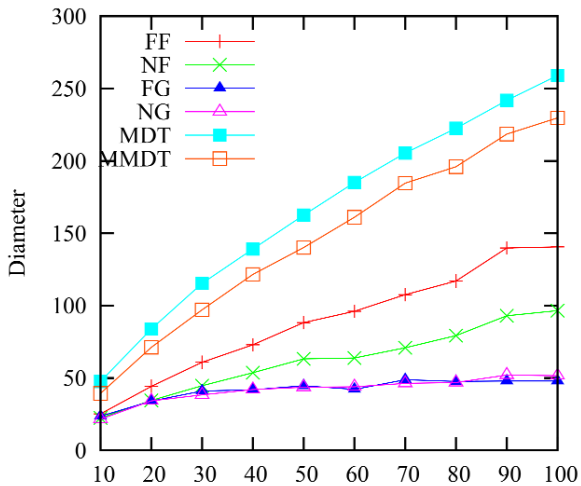
In short, in arbitrary directed graphs, our algorithms produce the light-trails with low number of required wavelengths, a low cost but quite high diameter compared with the other algorithms. Between the two heuristics, NF provides a good trade-off among the performance metrics.



(a) No. Wavelengths vs. Group size

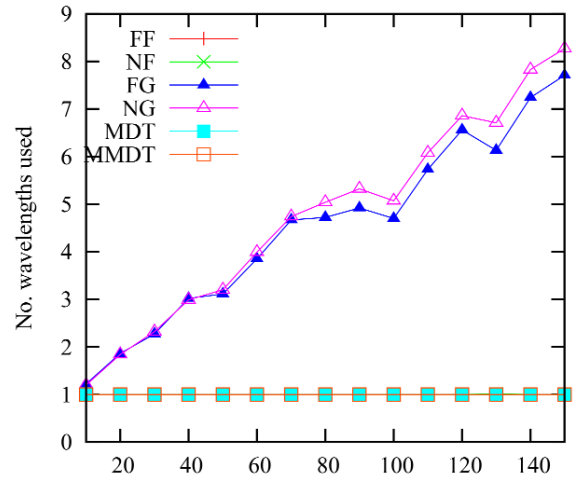


(b) Total Cost vs. Group size

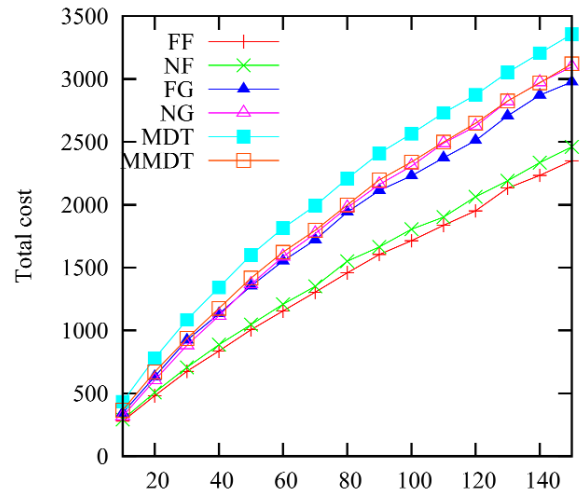


(c) Diameter vs. Group size

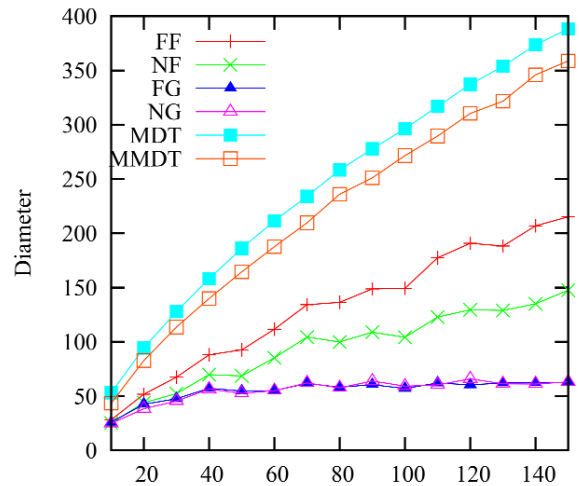
Fig. 5: Performance of algorithms on 200-node bigraphs



(a) No. Wavelengths vs. Group size

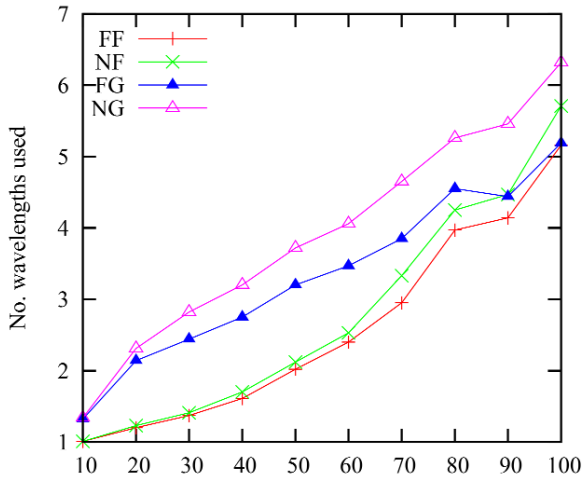


(b) Total Cost vs. Group size

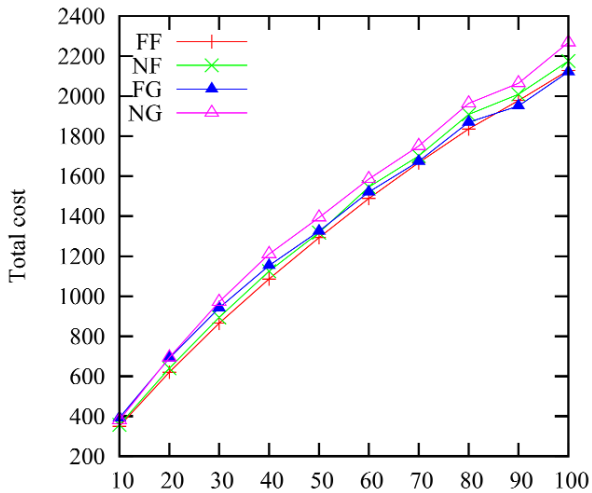


(c) Diameter vs. Group size

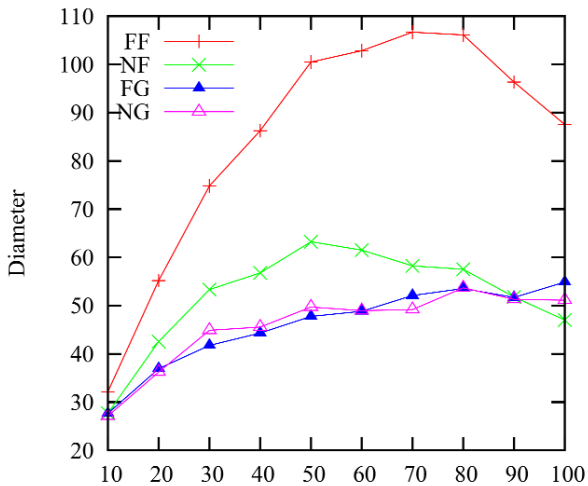
Fig. 6: Performance of algorithms on 300-node bigraphs



(a) No. Wavelengths vs. Group size

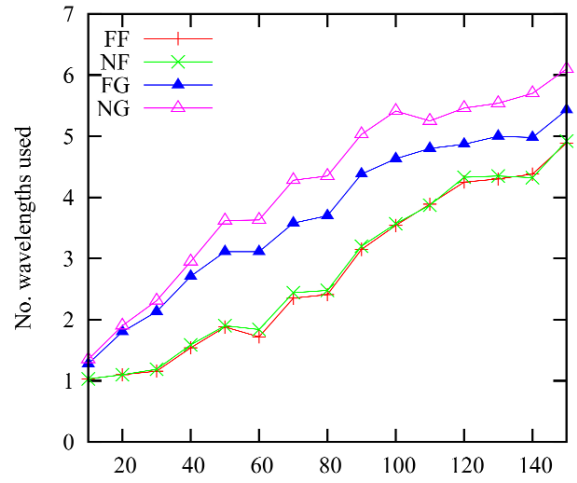


(b) Total Cost vs. Group size

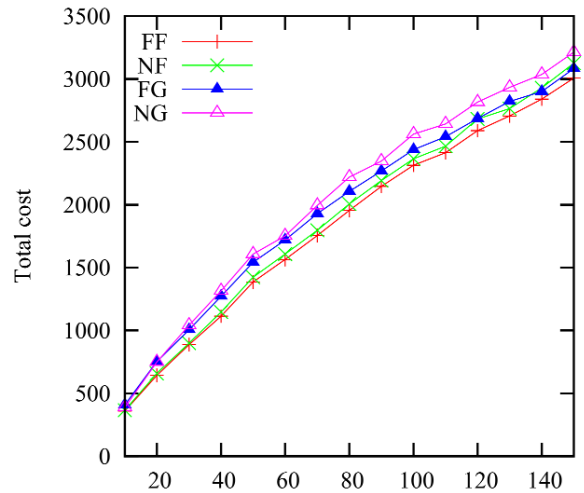


(c) Diameter vs. Group size

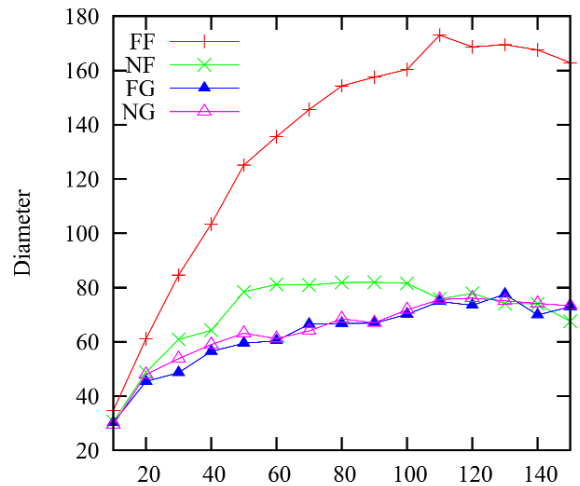
Fig. 7: Performance of algorithms on 200-node digraphs



(a) No. Wavelengths vs. Group size



(b) Total Cost vs. Group size



(c) Diameter vs. Group size

Fig. 8: Performance of algorithms on 300-node digraphs

E. Experimental result analysis

MDT versus MMDT

MDT and MMDT are just suitable in symmetric networks in which they always result in optimal wavelengths but high cost and diameter. By taking advantage of multiple transmitters, MMDT also needs just one wavelength (Figures 5(a) and 6(a)) and is even better than MDT in term of cost (Figures 5(b) and 6(b)) and diameter (Figures 5(c) and 6(c)). However, both have very poor performances in arbitrary directed graphs, even have no solution in most of the cases. This is because only one arc missing on the computing trail can make MDT and MMDT fail to get a solution.

Light-hierarchy based solutions versus light-tree based solutions

As seen in the experimental results above, our light-hierarchy based solutions (FF and NF) outperform light-tree based solutions (FG and NG) in term of the number of wavelengths and also the total cost, especially in bidirectional networks, with the expense of the diameter. This can be explained as follows.

The two approaches start with the same DSPT tree. In the rerouting phase, FG and NG try to extend the tree but always keep the tree structure which does not allow multiple visits at nodes. Moreover, the nodes used for the tree extension are always restricted by the source or the leaves which are either the farthest (FG) or nearest (NG) destinations in each subtree of the computed trees. These two properties limit the number of destinations to be covered in one tree, inducing larger number of wavelengths needed and higher cost. Besides, when the destinations cannot be routed in the current tree, they are routed in a new tree by the shortest paths, so the diameter is short.

In contrast, our approach is more flexible. After the first step, the route structure is no longer a tree, but a set of trails (composing a hierarchy) which allows to return some vertices more than once without conflicts. Hence, more arcs can be used for the trail extension. In addition, since all the terminals of the existing trails (and the source) can be considered, more nodes can be used for the trail extension. In short, more nodes and arcs can be used for the trail extension. This property helps to increase the number of destinations that can be covered in one trail, resulting in a fewer number of wavelengths required, but, of course, with a longer diameter. Besides, because more nodes are considered for the lowest cost, a lower total cost of the final trails can be achieved.

Farthest First versus Nearest First

As shown in the experimental results above, FF results in higher diameter but a slightly fewer wavelengths and a quasi-similar cost compared with NF. This can be explained as follows.

First, the diameter is the number of hops (the length) of the longest trail. When the selected trail (the *routed trail*) is replaced by an other one (the *routing trail*), the new routing trail must be longer (in term of cost) than the routed trail. Thus, the longer the routed trail is, the longer the new length can be. FF chooses the longest trail in the maximal clique, hence it makes the new trail longer than NF does. Furthermore, because the new trail is usually longer than routed trail in the old maximal clique, and it will probably become the farthest one in the new maximal clique and will be first considered next time. Hence it becomes longer and longer, and finally it can correspond to the diameter of the final trails. That is the reason for the fact that FF results in a longer diameter than NF.

Similarly, since FF tends to include more destinations in a long trail, the probability that the number of wavelengths that can be reduced by FF is higher than by NF. So FF results in a fewer wavelengths than NF does.

Finally, when the routed trail is replaced, the reduced cost is calculated by the cost of the routed trail minus the cost of the extended path of the routing trail (or the total cost of the routing trail if the source is selected). Thus, the longer the selected trail is, probably that the more the reduced cost can be. Since FF chooses the longest trail and NF chooses the shortest one in the maximal clique to diminish first, FF can reduce more cost than NF.

VI. CONCLUSION

In this paper we address the multicasting problem in all-optical networks without splitters. The problem is to find a set of light-trails which minimizes the number of required wavelengths with a low cost. The problem is proved to be NP-hard, and two heuristics based on Shortest Path Tree are proposed to make a feasible solution, Farthest First and Nearest First. The idea of our algorithm is to diminish the conflict between the light-trails until it cannot be reduced. Especially, unlike the common approaches which assume to work on networks with bidirectional links, our algorithm can work well in arbitrary networks.

The two heuristics of our algorithm are compared with the proposed algorithms in the literature, and the simulation results showed that our algorithm achieves low number of used wavelengths, low cost but quite high diameter. Between the two, although the Farthest First can result in smaller number of wavelengths and lower cost, the Nearest First provides a better trade-off among the three performance metrics.

REFERENCES

[1] M. Ali and J. S Deogun. Cost-effective implementation of multicasting in wavelength-routed networks. *IEEE/OSA Journal of Lightwave Technology*, 18:1628–1638, 2000.

[2] M. Ali and J. S Deogun. Power-efficient design of multicast wavelength routed networks. *IEEE Journal on Selected Areas in Communications*, 18:1852–1862, 2000.

[3] Der-Rong Din. Heuristic Algorithms for Finding Light-Forest of Multicast Routing on WDM Network. *Information Science and Engineering*, 25:83–103, 2009.

[4] Dinh Danh Le, Miklos Molnar, and Jerome Palaysi. All-Optical Multicast Routing Algorithms without Splitters. Technical report, May 2013.

[5] Deying Li, Xiufeng Du, Xiaodong Hu, Lu Ruan, and Xiaohua Jia. Minimizing number of wavelengths in multicast routing trees in WDM networks. *Networks*, 35(4):260–265, 2000.

[6] Miklos Molnar. Hierarchies to Solve Constrained Connected Spanning Problems. Technical Report 11029, LIRMM, September 2011.

[7] H. Takahashi and A. Matsuyama. An approximate solution for the Steiner problem in graphs. *Mathematica Japonica*, 24:573–577, 1980.

[8] Yingyu Wan and Weifa Liang. On the minimum number of wavelengths in multicast trees in WDM networks. *Networks*, 45(1):42–48, 2005.

[9] Xijun Zhang, John Wei, and Chunming Qiao. Constrained multicast routing in WDM networks with sparse light splitting. *IEEE/OSA Journal of Lightwave Technology*, 18:1917–1927, 2000.

[10] Fen Zhou, Miklos Molnar, and Bernard Cousin. Light-Hierarchy: The Optimal Structure for Multicast Routing in WDM Mesh Networks. In *Computers and Communications (ISCC), 2010 IEEE Symposium on The 15th IEEE Symposium on Computers and Communications (ISCC2010), 2010*, pages 611 – 616, Riccione Italie, June 2010.

[11] Yinzhu Zhou and Gee-Swee Poo. Optical multicast over wavelength-routed WDM networks: A survey. *Optical Switching and Networking*, 2(3):176 – 197, 2005.

Dinh Danh Le is now a PhD student in Informatics Department at laboratory LIRMM, University Montpellier 2, France. He was graduated at the Information Technology Faculty of Hong Duc University, Vietnam in 2004 and received a master degree in Computer Science at Information Technology Faculty of Vietnam National University, Hanoi in 2007. His main research topics and interests include network evaluation, multicast routing, optical networks, and combinatorial optimization.

Miklós Molnár is with the University Montpellier 2, laboratory LIRMM and he is a full professor in Computer Science. He was graduated at the Faculty of Electrical Engineering, University BME of Budapest in 1976 and received the Ph.D. degree in Computer Science from the University of Rennes 1 in 1992 and the French HDR degree in 2008. His main research activity is in combinatorial optimization to solve network related problems. His results are related to NP-hard optimization problems, constrained spanning and Steiner problems, routing algorithms for unicast, in-cast and multicast communications in optical and in wireless networks, dependable communications, energy aware protocols and optimization.

Jérôme Palaysi is an associate professor at University Montpellier 2 since September 2003. He is currently a member of the MAORE research team at LIRMM and teacher at the department of Computer Science of the IUT. He obtained his Ph.D in computer science (”Algorithmic Problems in all-optical networks”) at Montpellier 2 University. During past years, he focused his research on algorithmic studies of routing problem in networks (all-optical network, wireless networks and sensor networks). He is also applying well-known methods of scheduling theory and is working with researchers in this field.

The effect of RF unit breakdowns in sensor communication networks

Tamás Bérczes, Béla Almási, János Sztrik, Attila Kuki

Abstract—**In this paper the wireless transmission problem in sensor networks is investigated. The server (RF unit) is assumed to be subject to random breakdowns both in busy and idle states. The sensors of the networks are grouped in two classes. The first one is the "Emergency" class, which performs the notification of special emergency situations (eg. fire alarms). The second one is the "Normal" class, which measures and transmits environmental data (eg. temperature). The novelty of investigations is the inclusion of the non-reliability of the server. Our main interest is to give the main steady-state performance measures of the system computed by the help of the MOSEL tool. Several Figures illustrate the effect of the failure and repair rates of the server on the mean queue lengths and on the probability of server's breakdowns.**

I. INTRODUCTION

Wireless sensor networks are widely used to implement low cost non-attend monitoring of different environments. Baronti et al. [1] showed that the technology limits are far beyond the current usage. Chiany [2] represented the wireless sensor networks as a system containing three main components see Figure 1. Buchmann [3] showed that the operation mechanisms depending on the vendor implementations can be totally different, but also common features are observable. For example, power saving is a standard requirement to achieve long time operation of the wireless nodes. Similarly, a common feature that the wireless data transmission can appear as a bottleneck in the operation.

Retrial queues have been widely investigated and used to model many problems arising in telephone switching systems, telecommunication networks, computer networks, optical networks and most recently sensor networks, etc. The main characteristic of a retrial queue is that a customer who finds the service facility busy upon arrival is obliged to leave the service area, but some time later he comes back to re-initiate his demand. Between trials a customer is said to be in orbit. The literature on retrial queueing systems is very extensive. For a recent account, readers may refer to the recent books of Falin and Templeton [4], Artalejo and Gomez-Corral [5] that summarize the main models and methods. For some recent results on retrial queues with applications the interested reader is referred to, for example see papers of Tien Van Do [6] and references therein.

Manuscript submitted April 2013, revised May 18, 2013.

Tamás Bérczes, Béla Almási, János Sztrik and Attila Kuki is with the Department of Informatics Systems and Networks, Faculty of Informatics, University of Debrecene, Hungary, e-mail: {berczes.tamas,almasi.bela,sztrik.janos,kuki}@inf.unideb.hu

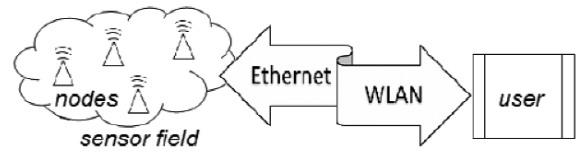


Fig. 1. Wireless sensor network.

Using wireless sensor networks, see [7], [8] one of the biggest problem is the lifetime of the sensor. Most of the time it is very hard to change or repair the RF Unit of the sensors. Because these facts, the reliability of the RF Unit is very important. The lifetime of the sensor determine the lifetime of the network too.

In this paper we introduce a finite-source retrial queueing model to investigate the performance characteristics of the wireless transmission problem in the sensor networks. We divide the sensors into two classes. The first one is the "Emergency" class, which performs the notification of special emergency situations (eg. fire alarms). The second one is the "Normal" class, which measures and transmits environmental data (eg. temperature).

The emergency class has priority over the normal class in the operation. For the performance evaluation of the wireless transmission we study and compare two cases: In the first model the RF transmission possibility will be available randomly for the sensor nodes (Non-Controlled case). In the second model the RF transmission requests coming from the emergency class will access the wireless channel immediately (Controlled case).

The main purpose of the present paper is to generalize the sensor network model (see. Bérczes et al. [9], [10]) using a more realistic case when the RF unit is subject to breakdowns during its operations. Our aim is to illustrate graphically the effect of the non-reliability of the RF unit on the steady-state system measures.

Because of the fact, that the state space of the describing Markov chain is very large, it is rather difficult to calculate the system measures in the traditional way of writing down and solving the underlying steady-state equations. To simplify this procedure we used the software tool MOSEL (Modeling, Specification and Evaluation Language), see Begail et al. [11], to formulate the model and to obtain the performance measures.

The rest of this paper is organized as follows. In Section 2 we present the corresponding queueing model. Numerical results and their discussion are provided in Section 3. Finally,

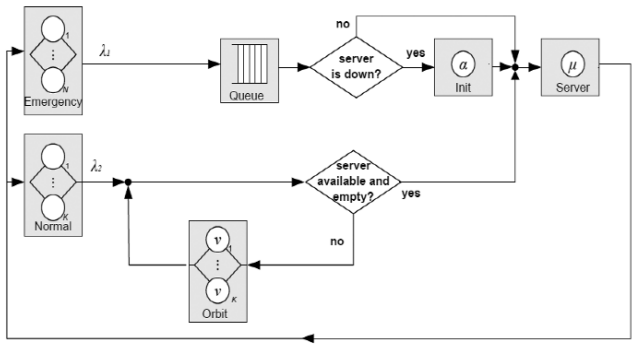


Fig. 2. A retrial queue with components

Section 4 concludes the paper.

II. System Model

Let us consider a single server queue with two classes of finite-sources which represent the sensors. The first class of sensors correspond to the emergency case (eg. fire alarms), the second class refers to the normal case (eg. temperature, humidity measurement). The number of sensors of the first class is denoted by N, and the number of sensors of the second class is denoted by K. Each sensor generates a new service request (ie. to send the measured value through the radio interface), according to an exponentially distributed time with parameter λ_1 for the emergency sensors and with parameter λ_2 for the normal class, respectively.

The server, which refers to the radio transmission in the model, can be in three states:

- *available(idle) state:* If the server is available it can start serving the arriving requests.
- *sleeping state:* The server can be in sleeping state (for power saving purposes).
- *failed state:* If the server is in failed state, it can not start serving any arriving requests until it is repaired.

The server is busy, when the server is in available state and at least one requests are in the service area. The server is idle, when the server is in available state and there is no requests in the service area.

The server can fail during the interval $(t, t + dt)$ with probability $\delta dt + o(dt)$ if it is in idle or in busy state. If the server is in sleeping state it can not be failed. If the server fails in a busy state, the interrupted request returns to the sources. The repair time is assumed to be exponentially distributed with a finite mean $1/\tau$. If the server is failed, two different cases can be treated. The first one is blocked sources case when all the operations are stopped, that is no new calls are generated. The second one is the unblocked sources case, when only service is interrupted but all the other operations are continued. In this paper we investigate only the unblocked sources case.

The server starts with a listening period. The time of this listening period is assumed to be exponentially distributed with parameter α. If no customer arrives during this period, the server will enter into the sleeping state. The time of the sleeping period is supposed to exponentially distributed with parameter β.

When the sleeping period is terminated, then the server wakes up. If there are emergency requests waiting in the queue the server begins to serve them. In the opposite case, when there is no emergency request waiting in the queue, the server remains in the available state, it will start a listening period.

Until the listening period finished, the requests arriving to the server can access to its service. If the listening period expires without any arrivals the server will enter into the sleeping mode.

A request of the emergency class goes directly to a FIFO queue waiting to be served (ie. transmitted through the radio interface).

If an emergency request arrives to the server we consider four operation possibilities:

- If the server is available, it starts the service of the emergency request.
- If the server is busy, the emergency request goes to the FIFO queue, waiting to be served.
- If the server is in sleeping state the request wakes up the server, which will start the service after an exponentially distributed initialization time with parameter γ.
- If the server is failed, the emergency request goes to the FIFO queue.

If a request from the second class finds the server busy or in sleeping state or the server is failed then the requests goes to the orbit. These requests waiting in the orbit retry to find the server idle according to a Poisson flow with retrial rate ν. We assume that emergency requests have non-preemptive priority over normal requests.

The service times for each request coming from both classes are assumed to be exponentially distributed with parameter μ.

The operational dynamics of the system can be seen in the corresponding queueing model, see Fig. 2.

We introduce the following notations (see the summary of the model parameters in Table I):

- $k_1(t)$ is the number of active sensors in the emergency source at time t,
- $k_2(t)$ is the number of active sensors in the normal source at time t,
- $q(t)$ denotes the number of emergency requests in the queue at time t,
- $o(t)$ is the number of jobs in the orbit at time t.
- $y(t) = 0$ if there is no job in the server and the server is available, $y(t) = 1$ if the server is busy with a job coming from the emergency class, $y(t) = 2$ when the server is busy with a job coming from the normal sensor class, $y(t) = 3$ if the server is in sleeping state at time t and $y(t) = 4$ if the server is failed at time t
- $c(t) = 1$ when the server is in sleeping state at time t and one emergency request has started the initialization procedure, $c(t) = 0$ in the other cases.

It is ease to see that:

$$k_1(t)+k_2(t) = \begin{cases} K+N-q(t)-o(t), & y(t)=0,4 \\ K+N-q(t)-o(t)-1, & y(t)=1,2 \\ K+N-q(t)-o(t)-c(t), & y(t)=3 \end{cases}$$

.

TABLE I
OVERVIEW OF MODEL PARAMETERS

Parameter	Maximum	Value at t
Active imergency sensors	N (population size)	$k_1(t)$
Active normal sensors	K (population size)	$k_2(t)$
Emergency generation rate		λ_1
Normal generation rate		λ_2
Total gen. rate	$\lambda_1 N + \lambda_2 K$	$\lambda_1 k_1(t) + \lambda_2 k_2(t)$
Requests in queue	N	$q(t)$
Service rate		μ
Busy servers	1 (number of servers)	$c(t)$
Cust. in service area	$N+1$	$c(t)+q(t)$
Requests in Orbit	K (orbit size)	$o(t)$
Retrial rate		ν
Server's failure rate		δ
Server's repair rate		τ

To maintain theoretical manageability, the distributions of inter-event times (i.e., request generation time, service time, retrial time, available state time, sleeping state time, failed state time) presented in the network are by assumption exponential and totally independent. The state of the network at a time t corresponds to a Continuous Time Markov Chain (CTMC) with 4 dimensions:

$$X(t) = (y(t); c(t); q(t); o(t))$$

The steady-state distributions are denoted by

$$P(y,c,q,o) = \lim_{t\to\infty} P(y(t)=y, c(t)=c, q(t)=q, o(t)=o)$$

Note, that the state space of this Continuous Time Markovian Chain is finite, so the steady-state probabilities surely exist. For computing the steady-state probabilities and the system characteristics, we use the MOSEL software tool in this paper. These computations are described in papers of Bolch and Wüchner et al. [12], [13].

As soon as we have calculated the distributions defined above, the most important steady-state system characteristics can be obtained in the following way:

- *Utilization of the server*

$$U_S = \sum_{y=1}^{2}\sum_{q=0}^{N}\sum_{o=0}^{K} P(y,0,q,o)$$

- *Availability of the server*

$$A_S = \sum_{y=0}^{2}\sum_{q=0}^{N}\sum_{o=0}^{K} P(y,0,q,o)$$

- *Average number of jobs in the orbit*

$$\overline{O} = E(o(t)) = \\ = \sum_{y=0}^{2}\sum_{q=0}^{N}\sum_{o=0}^{K} oP(y,0,q,o) \\ + \sum_{y=3}^{4}\sum_{c=0}^{1}\sum_{q=0}^{N}\sum_{o=0}^{K} oP(y,c,q,o)$$

- *Average number of jobs in FIFO*

$$\overline{Q} = E(q(t)) = \\ = \sum_{y=0}^{2}\sum_{q=0}^{N}\sum_{o=0}^{K} qP(y,0,q,o) \\ + \sum_{y=3}^{4}\sum_{c=0}^{1}\sum_{q=0}^{N}\sum_{o=0}^{K} qP(y,c,q,o)$$

- *Average number of jobs in the network*

$$\overline{M} = \overline{O} + \overline{Q} + \\ + \sum_{y=1}^{2}\sum_{q=0}^{N}\sum_{o=0}^{K} P(y,0,q,o) + \\ + \sum_{q=0}^{N}\sum_{o=0}^{K} P(3,1,q,o)$$

- *Average number of active emergency sensors*

$$\overline{\Lambda_1} = N - \overline{Q} - \sum_{q=0}^{N-1}\sum_{o=0}^{K} P(1,0,q,o)$$

- *Average number of active normal sensors*

$$\overline{\Lambda_2} = K - \overline{O} - \sum_{q=0}^{N}\sum_{o=0}^{K-1} P(2,0,q,o)$$

- *Average generation rate of emergency sensors:*

$$\overline{\lambda_1} = \lambda_1 \overline{\Lambda_1}$$

- *Average generation rate of normal sensors:*

$$\overline{\lambda_2} = \lambda_2 \overline{\Lambda_2}$$

III. NUMERICAL RESULTS

In this section, we present some numerical results in order to illustrate graphically the effect of the server's breakdown on some of the most important measures in sensor networks. The corresponding parameters are summarized in Table II. numerous interactions of parameters were investigated. The most interesting results are displayed in the following figures. In each Figure the blue lines (dotted with circles) represent the case when $\lambda = 0.5$ and the red lines (dotted with triangles) represent the case when $\lambda = 2.5$.

In Figure 3 one can see the effect of the server's failure rate on the mean queue length of emergency request in the FIFO. The length is increasing in both cases, but at higher generation rate this increase is faster.

Figure 4 shows how orbit fills up at different failure rates and request generation rates. It can be seen, that at high value of the generation rates the server's failure rate has almost no effect on the orbit size. At lower generation rate the orbit size significantly increases when the failure rate has large values.

Figure 5 illustrates the probability of the idle state of the server. In natural way, this idle state probability is much higher at large generation rates than at low ones. If we investigate the effect of the failure rate of the radio unit, a decrease of this probability can be observed. The speed of this decrease is much faster at the low generation rates.

In Figure 6 the probability of the failed state of the server is calculated, where the server initially was in idle state. It means, that there is no loss of requests here. Requests arriving during this failed period will be forwarded to queue or to orbit, depending on their priorities.

At low request generation rates the idle periods are longer than at higher rates, so the failed periods are more likely. With higher failure rates the probability of down periods increases, faster at low load and lower at high load of requests.

Figure 7 shows the same probability, but in this case the server breakdown occurs in busy state, ie. there was a request under servicing. The service terminates and the request is transmitted to the source. In the physical environment it means, that the signal of the sensor has been lost. Important result, because at relatively high failure rate the probability of lost request is significantly high, and it hardly depends on the generation rate.

It should be underlined that because the terminated requests under servicing return to the source without service, there is no sense to investigate response and waiting times.

Figure 8 illustrates the probability that the server fails in an idle state as a function of repair rates. It means, that there is no loss of requests in this case. As one can see increasing the repair rate, the probability of the failure decreases. For all repair rates we get smaller probabilities by using higher generation rate.

Figure 9 shows the probability that the server's breakdown occurs in busy state, ie. there is a request under servicing. The service terminates and the request returns to the source without service. In this case using higher generation rate we get higher probabilities because the server's busy periods last longer than using smaller generation rate.

IV. Conclusion

In this paper we have investigated a finite-source retrial queueing model with non-preemptive priorities and repeated vacations with non-reliable server. The MOSEL tool was used to formulate and solve the problem, and the main performance and reliability measures were derived and illustrated graphically. The main goal of the evaluation of the proposed system was to show how the of server's failure rate influences the performance of the system. To the best knowledge of the

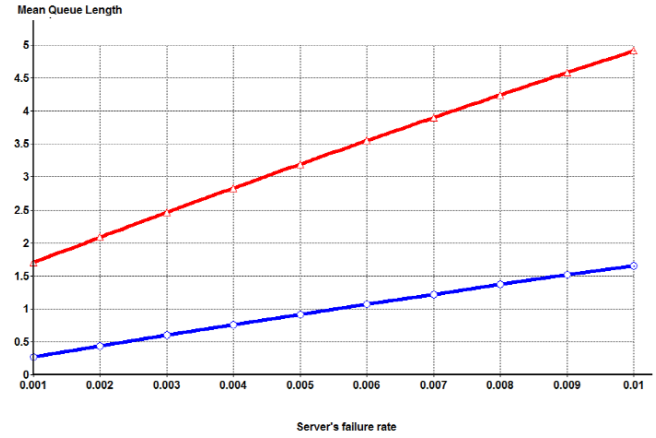


Fig. 3. Mean queue length vs Server's failure rate, repair rate = 0.1

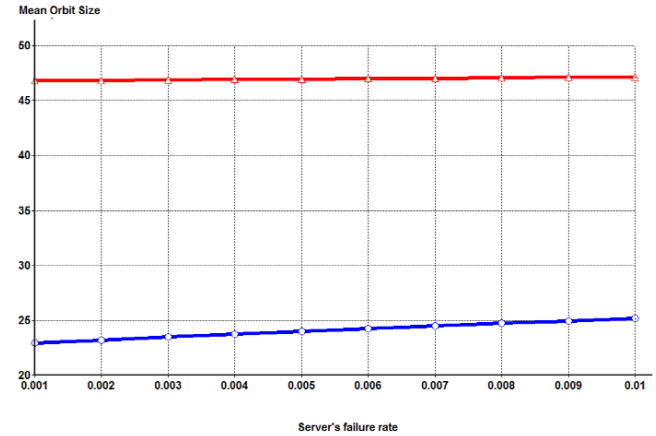


Fig. 4. Mean orbit size vs Server's failure rate, repair rate = 0.1

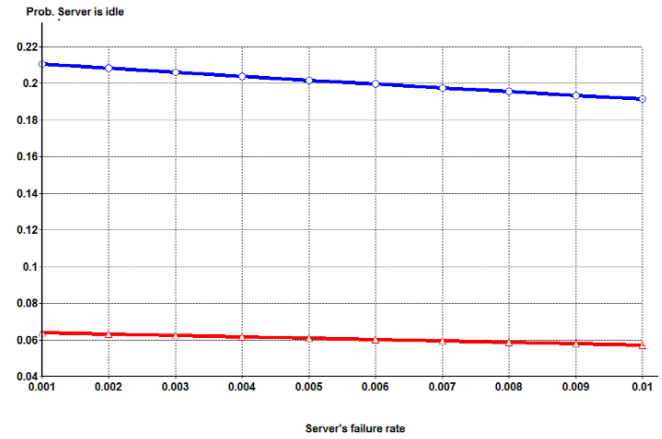


Fig. 5. Probability that the server is in idle state vs Server's failure rate, repair rate = 0.1

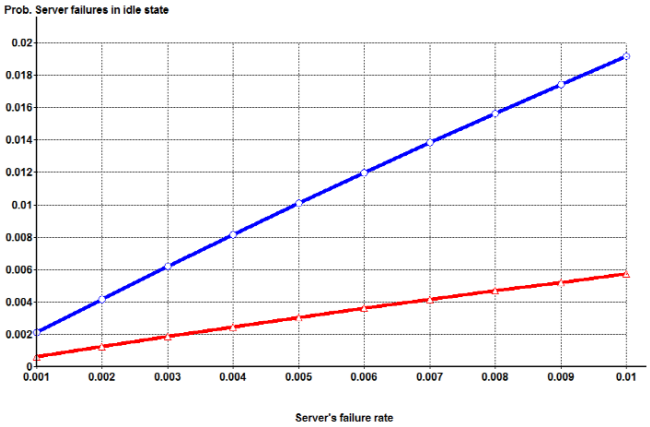


Fig. 6. Probability that the server failures in idle state vs Server's failure rate, repair rate = 0.1

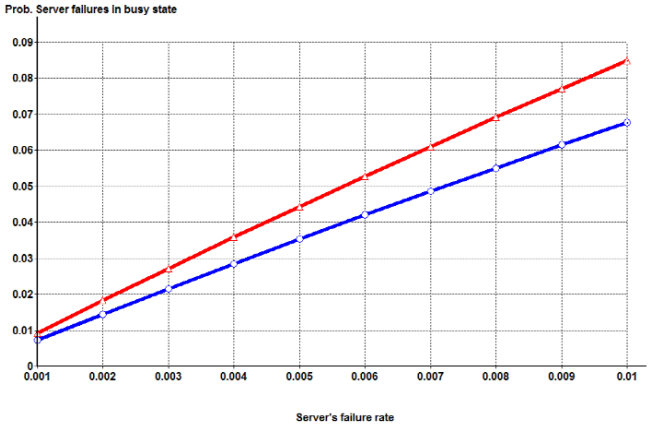


Fig. 7. Probability that the server failures in busy state vs Server's failure rate, repair rate = 0.1

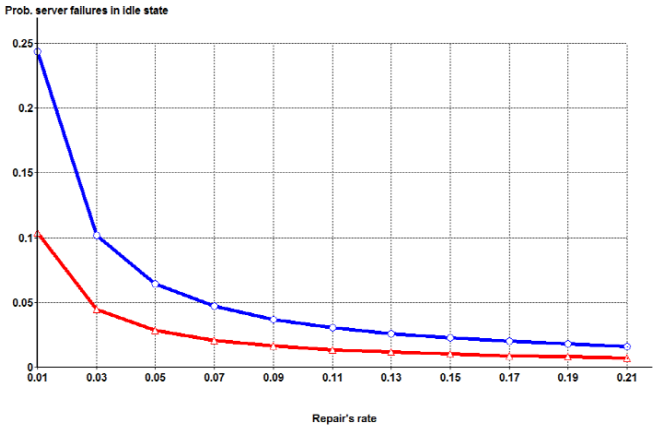


Fig. 8. Probability that the server failures in idle state vs Server's repair rate, failure rate = 0.005

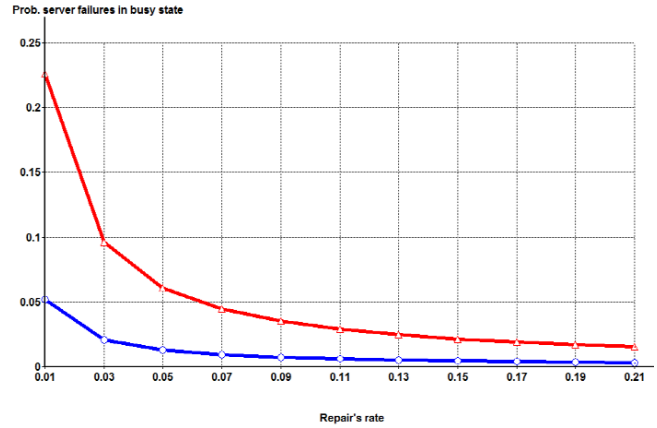


Fig. 9. Probability that the server failures in busy state vs Server's repair rate, failure rate = 0.005

TABLE II
NUMERICAL VALUES OF MODEL PARAMETERS

Parameter	Symbol	Value
Overall generation rate	λ	0.5,2.5
Emergency generation rate	$\lambda_1 = \frac{\lambda}{10}$	[0.05,0.25]
Normal generation rate	$\lambda_2 = \frac{9}{10}\lambda$	[0.45,2.25]
Number of Emergency sensors	N	50
Number of Normal sensors	K	50
Retrial rate	ν	2
Service rate	μ	20
Server's failure rate	δ	[0.001..0.01]
Server's repair rate	τ	[0.01..0.21]
Initialization rate	γ	10
Mean time of sleeping period	$\frac{1}{\beta}$	2.5
Mean time of listening period	$\frac{1}{\alpha}$	1.5

authors, this is the first proposal for the use of the theory of retrial queues to model sensor networks with priority finite-source with orbit, state dependent vacation times and a non-reliable server. In the future work we would like to investigate more complex sensor models by using finite-source queueing models.

ACKNOWLEDGMENT

The publication was supported by the TÁMOP-4.2.2.C-11/1/KONV-2012-0001 project. The project has been supported by the European Union, co-financed by the European Social Fund. The research was also supported by the European Union with the help of the GOP-1.3.2-08-2009-0007 program. This research was realized in the frames of TÁMOP 4.2.4.A/2-11-1-2012-0001 National Excellence Program Elaborating and operating an inland student and researcher personal support system The project was subsidized by the European Union and co-financed by the European Social Fund.

REFERENCES

[1] P. Baronti, "Wireless sensor networks: A survey on the state of the art and the 802.15.4 and zigbee standards," *Computer Communications*, vol. 30, pp. 1655–1695, 2007.

[2] M. W. Chiang, "Architectures of increased availability wireless sensor network nodes," *ITC International Test Conference*, 2004.

[3] A. Buchman, "An overview of hardware platforms used in wireless sensor nodes," *Carpathian Journal of Electronic and Computer Engineering*, vol. 4, pp. 19–22, 2011.

[4] G. Falin and J. Templeton, *Retrial Queues*. Chapman & Hall, 1997.

[5] J. Artalejo and A. Gomez-Corral, *Retrial Queueing Systems: A Computational Approach*. Springer Verlag, 2008.

[6] T.V.Do, "An efficient solution to a retrial queue for the performability evaluation of dhcp," *Computers & OR*, vol. 37, no. 7, pp. 1191–1198, 2010.

[7] I. F. Akyildiz and M. C. Vuran, *Wireless Sensor Networks*. John Wiley & Sons, July 2010.

[8] F. Dressler, *Self-Organization in Sensor and Actor Networks*. John Wiley & Sons, 2007.

[9] T. Bérczes, J. Sztrik, P. Orosz, P. Moyal, N. Limnios, and S. Georgiadis, "Tool supported modeling of sensor communication networks by using finite-source priority retrial queues," *Carpathian Journal of Electronic and Computer Engineering*, vol. 5, pp. 13–18, 2012.

[10] T. Bérczes, B. Almási, A. Kuki, and J. Sztrik, "A contribution to modeling sensor communication networks by using finite-source queueing systems," *Proceedings of the 8th IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI 2013).Timisoara, Romania, To appear*.

[11] K. Begain, G. Bolch, and H. Herold, *Practical performance modeling, application of the MOSEL language*. Boston: Kluwer Academic Publisher, 2001.

[12] G. Bolch, S. Greiner, H. de Meer, and K. Trivedi, *Queueing Networks and Markov Chains*, 2nd ed. New York: John Wiley & Sons, 2006.

[13] P. Wüchner, J. Sztrik, and H. de Meer, "Modeling wireless sensor networks using finite-source retrial queues with unreliable orbit," *Springer Lecture Notes in Computer Science*, vol. 6821, pp. 275–285, 2011.

Tamás Bérczes received his MSc Degree in Mathematics in 2000. at the University of Debrecen, Hungary. He is currently assistant professor at the Department of Informatics Systems and Networks of the same university. He received his Ph.D. in 2011. His primary research interests are performance analysis of re- trial queues and their application.

Béla Almási is a associate professor at the Department of Informatics Systems and Networks at the Faculty of Informatics, University of Debrecen, Hungary. He received his Ph.D. in 1998. at the University of Debrecen, Hungary and Habilitation from University of Debrecen in 2006. His primary research interests are network systems, performance analysis of retrial queues and their application.

Attila Kuki He is currently an assistant professor at the Department of Informatics Systems and Networks at the Faculty of Informatics, Universy of Debrecen. H e received his Ph.D. in 1997 in Mathematics. His primary research interests are performance analysis of retrial queues and their application.

János Sztrik is a Full Professor at the Faculty of Informatics, University of Debrecen, Hun- gary. He obtained the Candidate of Mathe- matical Sciences Degree in Probability Theory and Mathematical Statistics in 1989 from the Kiev State University, Kiev, USSR, Habilitation from University of Debrecen in 1999, Doctor of the Hungarian Academy of Sciences, Buda- pest, 2002. His research interests are in the field of production systems modeling and analysis, queueing theory, reliability theory, and computer science.

FPGA Implementation of Pipelined CORDIC for Digital Demodulation in FMCW Radar

Amritakar Mandal, Rajesh Mishra

***Abstract*— Now-a-days Radar Signal Processing system is gaining a great deal of attention for realization of on-chip programmable signal processor for its real time applications. Application specific systems are being implemented using wide spectrum of Digital Signal Processing (DSP) algorithms. Such is the case for COordinate Rotation DIgital Computer (CORDIC) algorithm which is turned out to be widely researched topic in the field of vector rotated DSP applications. In this paper we have designed an application specific pipelined CORDIC architecture for digital demodulation in low power, high performance FMCW Radar. A complex Digital Phase Locked Loop (DPLL) has been used for digital demodulation. The FPGA implementation of CORDIC based design is suitable because of its inherent high system throughput due to its pipelined architecture where latency is reduced in each of the pipelined stage. Substantial amount of resource utilization has been reduced in proposed design. For better loop performance of first order complex DPLL during demodulation, the convergence of the CORDIC architecture is also optimized. Hardware synthesized result using Cadence design tools are presented.**

***Index Terms*— FMCW Radar, CORDIC, FPGA, DSP, DPLL, Loop performance.**

I. INTRODUCTION

PHASE DETECTION is a widely researched topic in radar demodulation, especially in low power Doppler radar where accurate detection is needed [1]. In general, radar system uses coherent oscillator as a reference frequency for Doppler detection. In analog demodulation, Voltage Control oscillator (VCO) is used. But perfect demodulation is not possible with the use of VCO as it suffers to maintain linearity over the desired frequency range [2]. In present decade, digital or mixed signal design based demodulator is widely used for superior performance. To ensure linearity over the range of frequency, numerically controlled oscillator is being used. Phase detection in communication receiver is very much sensitive to quantization noise. This kind of distortion is basically due to bit resolution. Efforts have been made to design a quantization error free pipelined CORDIC architecture based digital demodulator on FPGA platform. We have proposed a first order complex Digital Phase Locked Loop (DPLL) for efficient demodulation of radar signals.

Manuscript received March 31, 2013; revised May 22, 2013.

Amritakar Mandal and Rajesh Mishra are with the School of Information and Communication Technology, Gautam Buddha University, Greater Noida, U.P.- 201308 , India (e-mail: amritkar2k@gmail.com; rmishra@gbu.ac.in).

The iterative formulation of CORDIC algorithm was first developed by Jack E. Volder in 1959 [3] for the multiplication, division and computation of trigonometric functions like sine, cosine, magnitude and phase with great precision. The key concept of CORDIC algorithm is simply shift and adds. Although the same functions can be implemented using multipliers, variable shift registers or Multiply Accumulator (MAC) units, but CORDIC can implement these functions efficiently while saving enough silicon area which is considered to be a primary design criteria in application specific on chip implementation where high performance and low cost hardware solutions for DSP are required [4]. This kind of demodulator can be used in wide variety of receivers where real time signal processing is important [5].

This paper designs first order complex DPLL for I/Q channel Radar demodulator using pipelined CORDIC architecture. In digital PLL, an adjustable local sine wave generator and phase detector is required. The CORDIC offers the opportunity to calculate the desired trigonometric computation in a simple and efficient way. Due to the simplicity of the involved operations, the CORDIC realization of complex DPLL is very well suited for on-chip hardware design and its implementation. The analysis of various error sources is necessary for optimal design of system using the CORDIC processor. In DSP systems, signals are required to be quantized and represented in fixed word-length. A limited word-length results in the round-off noise and degradation of Signal-to-Quantization Noise Ratio (SQNR) performance [6,7]. In general, larger the dynamic range of the signals, more severe is the round-off noise. To reduce the computation error, a processor designer might simply increase the number of iterations and that will be a huge wastage of processing time and power. Therefore, exact computation of word-length is necessary for designing an architecture for CORDIC. If word-length is larger, the computational speed of CORDIC reduces significantly [8,9]. On the other hand, if we implement with smaller word-length, the design will suffer from danger of overflow. To design an optimal application specific CORDIC processor for a high performance signal processing, the choice of word-length and number of iterations in error analysis are needed to be taken into consideration. In this paper, both the problems of overflow and quantization noise have been addressed adequately for the design process.

The remainder of this paper proceeds as follows. In section II, the conventional CORDIC algorithm is briefly reviewed.

Design of pipelined CORDIC and design related issues have been discussed in section III. Evaluation and analysis of computational error has been discussed in section IV. In section V and VI, complex signal processing through CORDIC has been explained for demodulation purpose. In section VII, Hardware synthesis result has been discussed and conclusion may be found in section VIII.

II. REVIEW OF CORDIC ALGORITHM

The theory of CORDIC computation is to decompose the desired rotation angle into the weighted sum of a set of predefined elementary rotation angles. Each of them can be accomplished with simple shift-add operation for a desired rotational angle θ. It can be represented for M iterations of an input vector $(x,y)^T$ setting initial conditions: $x_0=x$, $y_0=y$, and $z_0=\theta$ as $z_f = \theta - \sum_{i=0}^{M-1} \delta_i \alpha_i$. If $z_f=0$ holds, then $\theta = \sum_{i=0}^{M-1} \delta_i \alpha_i$, *i.e.* the total accumulated rotation angle is equal to θ. δ_i, $0 \le i \le M-1$, denote a sequence of ± 1s that determine the direction of each elementary rotation. When M is the total number of elementary rotation angles, i-th angle α_i is given by:

$$\alpha_{m,i} = \frac{1}{\sqrt{m}} \tan^{-1}[\sqrt{m}\, 2^{-s(m,i)}] = \begin{cases} 2^{-s(0,i)} \\ \tan^{-1} 2^{-s(1,i)} \\ \tanh^{-1} 2^{-s(-1,i)} \end{cases} \quad (1)$$

where m=0, 1 and -1 correspond to the rotation operation in linear, circular, and hyperbolic coordinate system respectively. For a given value of θ, the CORDIC iteration is given by:

$$\begin{bmatrix} x_{i+1} \\ y_{i+1} \end{bmatrix} = \begin{bmatrix} 1 & -\delta_i 2^{-i} \\ \delta_i 2^{-i} & 1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix} \quad (2)$$

and $z_{i+1} = z_i - \delta_i \alpha_i$

where $\alpha_i = \tan^{-1} 2^{-i}$. To bring a unit vector to desired angle θ, the CORDIC algorithm gives known recursive rotations to the vector. The known rotational values are shown in **Table. I** as a Pre-Computed angle. Once the vector is at desired angle, the outcome of the X and Y coordinates of the vector are equal to Cosθ and Sinθ respectively. Let a unit vector in iteration 'i' is rotated by some angle θ_i , then the recursively updated equations are generated in the following form:

$$\begin{aligned} x_{i+1} &= x_i \cos\delta_i\alpha_i - y_i \sin\delta_i\alpha_i \\ y_{i+1} &= y_i \cos\delta_i\alpha_i + x_i \sin\delta_i\alpha_i \end{aligned} \quad (3)$$

The above equation can be simplified and written as

$$\begin{aligned} x_{i+1} &= \cos\delta_i\alpha_i (x_i - y_i \tan\delta_i\alpha_i) \\ y_{i+1} &= \cos\delta_i\alpha_i (y_i + x_i \tan\delta_i\alpha_i) \end{aligned} \quad (4)$$

Here, $\tan\alpha_i$ is restricted to $\pm 2^{-i}$. So multiplication is converted in an arithmetic right shift. Since cosine is an even function, therefore $\cos(\alpha) = \cos(-\alpha)$. The iterative equation can be reduced to-

$$\begin{aligned} x_{i+1} &= K_i (x_i - y_i \delta_i 2^{-i}) \\ y_{i+1} &= K_i (y_i + x_i \delta_i 2^{-i}) \end{aligned} \quad (5)$$

Where $K_i = \cos\left(\arctan 2^{-i}\right) = \sqrt{(1+2^{-2i})}$ is known as gain factor for each iteration. If M iterations are performed, then scale factor, K, is defined as the multiplication of every K_i .

$$K = \prod_{i=0}^{M-1} K_i = \prod_{i=0}^{M-1} \sqrt{(1+2^{-2i})} . \quad (6)$$

The elementary functions sine and cosine can be computed using the rotation mode of the CORDIC algorithm if the initial vector starts at $(|K|,0)$ with unit length. The final outputs of the CORDIC for the given input values $x_0 = 1, y_0 = 0$ and $z_0 = \theta$ are as follows:

$x_f = K\cos\theta$, $y_f = K\sin\theta$ and $z_f = 0$.

Since the scale factor is constant for a given number of rotations, $x_0 = 1/K$ can be set to get purely $\sin\theta$ and $\cos\theta$ values.

Table I
Pre-Computed Angles for Pipelined CORDIC

i	$2^{-i} = \tan\alpha_i$	$\alpha_i = \arctan(2^{-i})$	α_i in radian
0	1	45°	0.7854
1	0.5	26.565°	0.4636
2	0.25	14.063°	0.2450
3	0.125	7.125°	0.1244
4	0.0625	3.576°	0.0624
5	0.03125	1.7876°	0.0312
6	0.015625	0.8938°	0.0156
7	0.0078125	0.4469°	0.0078
..	…..	……..	…….

III. CORDIC ARCHITECTURE

In this CORDIC architecture, a number of identical rotational modules have been incorporated and each module is responsible for one elementary rotation. Because of identical CORDIC iterations, it is convenient to map them into pipelined architecture [9]. The purpose of pipelined implementation is to device a minimum critical path. Therefore, this kind of architecture provides better throughput and lesser latency compared to other designs. It is associated with a number of stages of CORDIC Units where each of the pipelined stages consists of a basic CORDIC engine. The CORDIC engines are cascaded through intermediate latches as shown in **Fig. 1**. The shift operations are hardwired using permanent oblique bus connections to perform multiplications by 2^{-i} reducing a large silicon area as required by barrel shifters.The critical-path of the pipelined CORDIC is the time required by the Add/Substract operations in each of the stages. Every stage contributes critical path delay amounts to $T_{Path}=T_{Add} + T_{MUX} + T_{2C}$,where T_{Add}, T_{MUX} and T_{2C} are the time required for addition , 2:1 Multiplexing and 2's Complement operations, respectively. The pre-computed angles (**Table 1**)

of i-th iteration angle α_i required at each CORDIC engine can be stored at a ROM memory location, are known. Therefore, the need of multiplexing and sign detection is avoided to reduce critical path. The latency of computation is thus depends primarily on the adder used. Since no sign detection is needed to force $z_f = 0$, the carry save adders are well suited in this architecture. The use of these adders reduces the stage delay significantly. The delay can be adjusted by using proper bit-length in the shift register. With the pipelining architecture, the propagation delay of the multiplier is the total delay of a single adder. So ultimately the throughput of the architecture is increased to a many fold as the throughput is given by: "1/delay due to a single adder". It implies that speed up factor becomes more than M and latency of the design is M times of the delay of a single adder. It is obvious that if we increase the number of iterations then the latency of the design also will increase significantly. If an iterative implementation of the CORDIC were used, the processor would take several clock cycles to give output for a given input. But in the pipelined architecture, it converts iterations into pipeline phases. Therefore, an output is

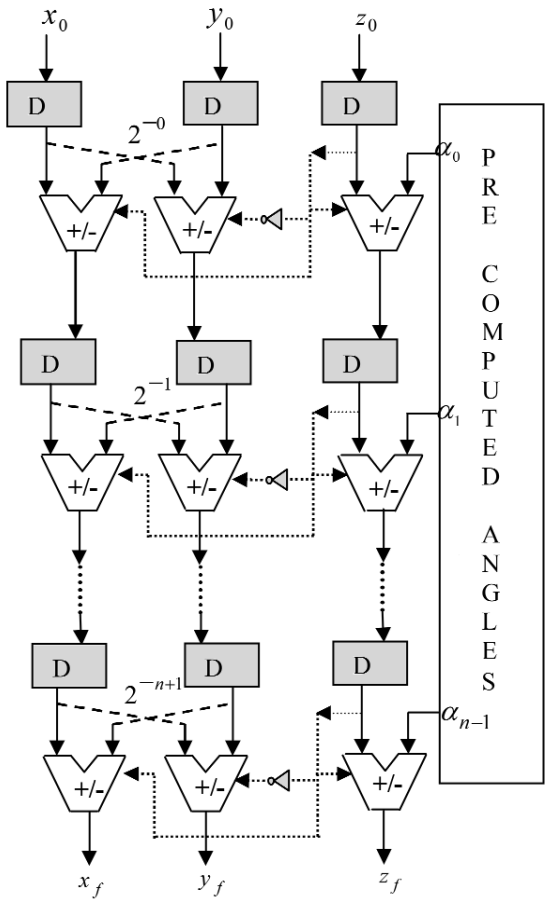


Fig. 1. Pipelined CORDIC Architecture

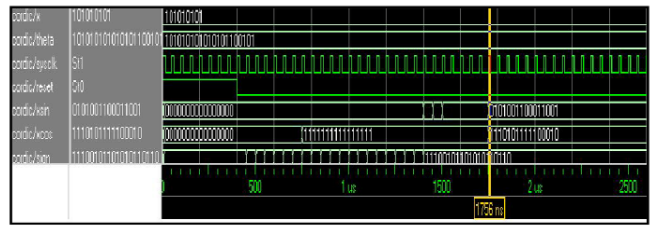


Fig. 2. Simulation Result of CORDIC

obtained at every clock cycle after pipeline stage propagation. Each pipeline stage takes exactly one clock cycle to pass one output (Simulated output shown in **Fig. 2**).

The most recurrent problems for a CORDIC implementation are overflow. Since the first tangent value is $2^0 = 1$, then rotation range will be $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. The difference in binary representation between these two angles is one bit. Overflow arises when a rotational angle crosses a positive right angle to a negative one. To avoid overflow, an overflow control is added. It checks for the sign of the operands involve in addition or subtraction and the result of the operation. If overflow is produced, the result keeps its last sign without affecting the final result. In the overflow control, the sign of z_i determines whether addition or subtraction is to be performed.

IV. NUMERICAL ERROR ANALYSIS

Theoretically, CORDIC realization has infinite number of iterations and that leads to accurate result. But practically CORDIC realization uses finite number of iterations resulting in approximation error. This kind of error arises due to approximations in angle as well as finite word length [8]. To get total approximation error, the error due to the angle approximation process will be derived and followed by the error due to the truncation of word length will be derived. The total error is taken as the summation of the two.

In the angle approximation process the angle θ is approximated as the algebraic sum of predefined elementary angles.

$$\theta = \sum_{i=0}^{M-1} \sigma_i \alpha_i + \delta \tag{7}$$

δ = Angle yet to be rotated after completion of the CORDIC iterations.

Due to convergence relationship of the CORDIC algorithm

$$\delta_{\max} = \tan^{-1} 2^{-(M-1)} \tag{8}$$

Let v^* be the ideal result obtained by the rotation of the vector $v(0) = [x(0) \quad y(0)]^t$ by an angle of θ. Let $\tilde{v}(M)$ be the output of the CORDIC block after scaling operation and assuming that there is infinite precision in the CORDIC operation module. Then,

$$v^* = \begin{bmatrix} \cos\delta & -\sin\delta \\ \sin\delta & \cos\delta \end{bmatrix} \tilde{v}(M) \tag{9}$$

Let $D=\begin{bmatrix}\cos\delta & -\sin\delta \\ \sin\delta & \cos\delta\end{bmatrix}$

The error in the output due to the process of angle approximation is $v^* - \tilde{v}(M)$.

$$= v^* - \tilde{v}(M) = [D-I]*\tilde{v}(M) \quad (10)$$

It can be easily shown that

$$\| D-I \| = \sqrt{(\cos\delta - 1)^2 + (\sin\delta)^2}$$

$$= 2*\sin|\delta/2|$$

$$\| D-I \| = 2*\sin|\delta/2| \le 2*|\delta/2| = |\delta| \quad (11)$$

From equation 9,

$$\| D-I \| \le |\delta| \le \tan^{-1}(2^{-(M-1)}) < 1/2^{M-1} \quad (12)$$

we can get a consolidated truncation error due to finite wordlength using scale factor K and number of finite iterations (M).

$$K*\sqrt{2}*2^{-b}(1+\sum_{j=0}^{M-1}(\prod_{i=j}^{M-1}\sqrt{(1+2^{-2i})})) \quad (13)$$

The scaling operation also introduces some error which amounts to maximum of 2^{-b}. So the final expression for the total quantization error is addition of all previously mentioned errors.

|Total error| ≤

$$\frac{1}{2^{M-1}}*|v^*|+K*\sqrt{2}*2^{-b}(1+\sum_{j=0}^{M-1}\prod_{i=j}^{M-1}\sqrt{(1+2^{-2i})})+2^{-b} \quad (14)$$

Let the output of the CORDIC block has 12 bits in its fractional part. Therefore, the upper limit of the total quantization error can be taken as 2^{-12}.

|Total error|

$$\le \frac{1}{2^{M-1}}*|v^*|+K*\sqrt{2}*2^{-b}(1+\sum_{j=0}^{M-1}\prod_{i=j}^{M-1}\sqrt{(1+2^{-2i})})+2^{-b} \le 2^{-12} \quad (15)$$

The above inequality is simulated in MATLAB to find out fractional bits of the internal word length of the CORDIC.

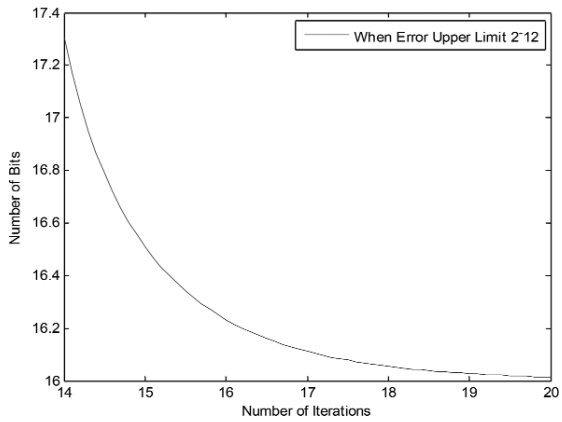


Fig. 3. The required Bits Vs. Iterations for CORDIC Internal Design

V. BASICS OF I/Q DEMODULATION

As per the Euler's theorem, vector sum of cosine component is completely real whereas the spectrum of sine component is totally imaginary. If the cosine and sine components are combined, the resultant spectrum becomes one sided with direction of rotation (positive or negative frequency) and with known real (cosine) and imaginary (sine) components [11].

$$\cos\omega_c t + j\sin\omega_c t = e^{j\omega_c t} \quad (16)$$

The process of recovering both real and imaginary signal component is known as *I/Q* demodulation. *I* stand for in-phase channel which processes cosine (real) components. *Q* stands for in-quadrature channel which processes sine (imaginary) component. The input of *I/Q* channel is Intermediate Frequency (IF) as shown in **Fig. 4**. If the carrier frequency of IF is f_c with a time varying amplitude $a(t)$ and time varying phase $\phi(t)$, then input signal $s(t)$ will be:

$$S(t) = a(t)\cos[(2\pi f_c t + \phi(t)]. \quad (17)$$

In *I* channel, the IF signal is multiplied by reference carrier frequency produced by crystal oscillator at zero phase reference. The output of the *I* channel mixer is, *I(t)* , given by :

$$I(t) = a(t)\cos[(2\pi f_c t + \phi(t)].\cos(2\pi f_c t)$$

$$= a(t)\cos[\phi(t)] + a(t)\cos[4\pi f_c t + \phi(t)] \quad (18)$$

The first term is the average value (DC) of the product and represents cosine of the signal phase and amplitude. The second term with high frequency component is suppressed by Low Pass Filter (LPF). So the output of the *I* channel is

$$I(t) = a(t)\cos[\phi(t)]. \quad (19)$$

Similarly, the *Q* channel output can be derived. The LPF output at *Q* channel is:

$$Q(t) = a(t)\sin[\phi(t)]. \quad (20)$$

Thus, *I* and *Q* channel together provide the amplitude and phase modulation.

VI. PHASE DETECTION TECHNIQUE IN FMCW RADAR

The down converted and filtered baseband signal has two components: real and imaginary parts. Therefore, the baseband signal can hold both the amplitude and phase of the sinusoidal signal at the same time. It does not hold any image frequency. So only loop filter is sufficient for the digital demodulation using Digital PLL [10]. Using the vector rotation operator $[x, y]^T \angle\theta$, the complex first-order DPLL demodulator equations for a given input signal can be stated as:

a). The real part of the output in phase comparator equation : $\varepsilon_n = \Re\{v(n)\angle -\theta_n\}$

b). The loop filter equation: $c_n = 2\pi K_l \varepsilon_n$,where K_l is the loop filter coefficient. The loop filter coefficient K_l depends on the sampling frequency and number of

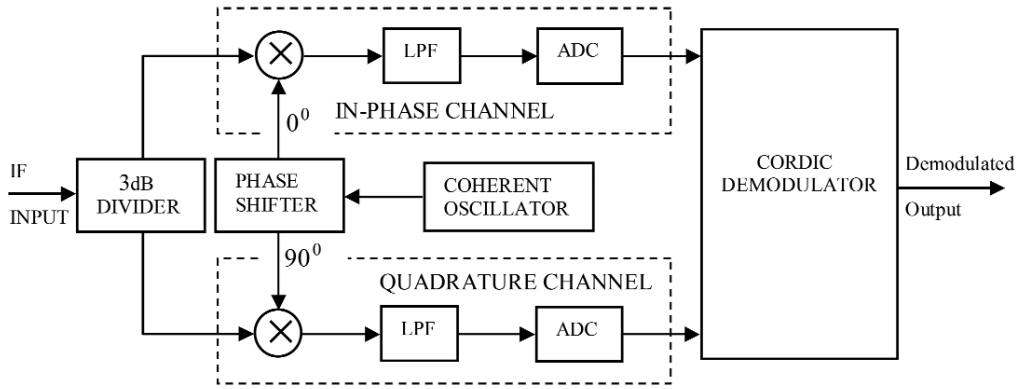


Fig. 4. CORDIC Based DPLL in I/Q Channel

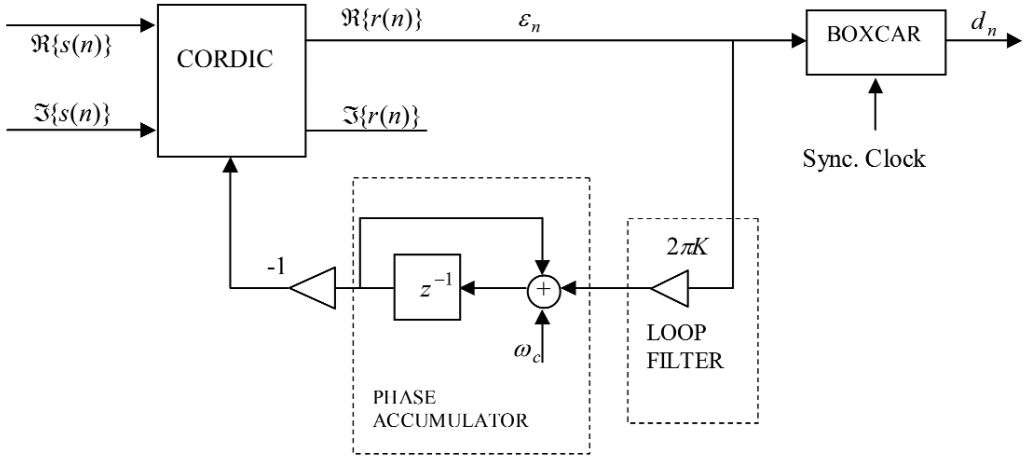


Fig. 5. Complex DPLL using CORDIC

iterations of CORDIC algorithm. For M number of iterations, the loop filter coefficient K_l can be given by:

$$K_{CORDIC} = \frac{K_l}{\prod_{i=0}^{M-1} \sqrt{1+2^{-2i}}} \tag{21}$$

c). The phase accumulator equation:

$$\theta_{n+1} = (\theta_n + 2\pi K_l \varepsilon_n + \omega_c) \bmod 2\pi \,, \tag{22}$$

where $\omega_c = 2\pi f_c$ is the center frequency.

The CORDIC based DPLL (as shown in **Fig. 5**) tries to adjust the continuous phase rotation in such a way that the complex component of the rotated vector will always be zero.

The post detection filter after the DPLL can be a Boxcar filter instead of low-pass filter. It rejects the unwanted noise Low-pass filter averages the signal and produces an output from the demodulated signal efficiently. BOXCAR filter has added advantage over the LPF as far as information recovery from narrow samples with little energy signal is concerned. Low-pass filter averages the signal and produces an output with weak amplitude signals. To avoid this problem, the samples can be stretched for entire inter sample period by increasing their sample amplitude at the filter output. The VLSI implementation of Boxcar generator is very easy as it performs only addition operation.

The simulation study has been carried out in MATLAB simulator. The down converted IF (with Doppler shift) is taken as input signal for the demodulator. The coherent oscillator frequency has been taken as a reference frequency to the demodulator. Any phase and amplitude variation (as shown in equations 18 and 19) in radar echo can be retrieved in the phase detector stage. Initial phase shift of 30^0 is introduced for the simulation. Simulated radar detected signal with reference to coherent reference signal and the final phase-amplitude variation at the demodulator output have been shown in **Fig. 6**, **Fig. 7** and **Fig. 8** respectively.

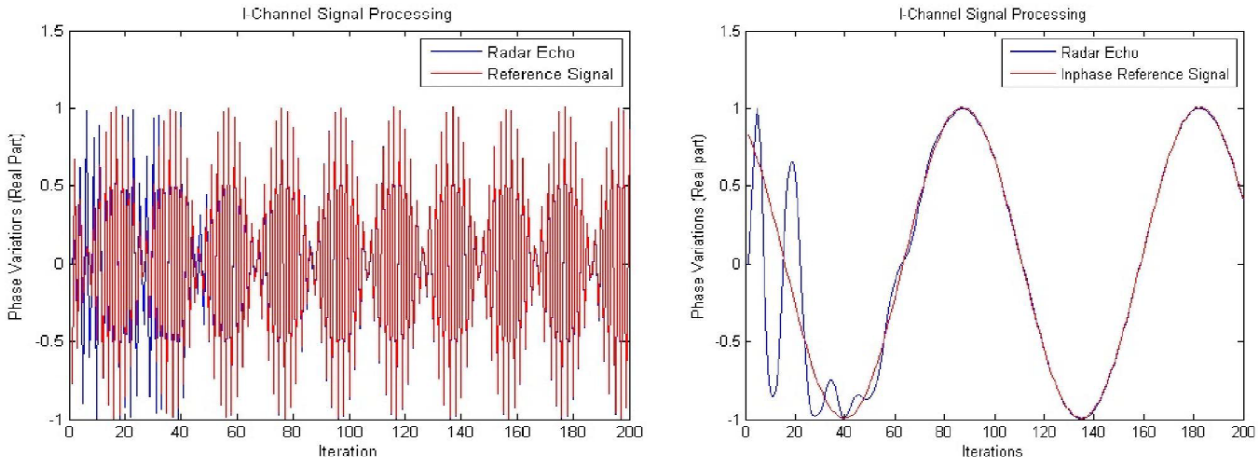


Fig. 6. Phase variation in In-phase Channel Radar Receiver

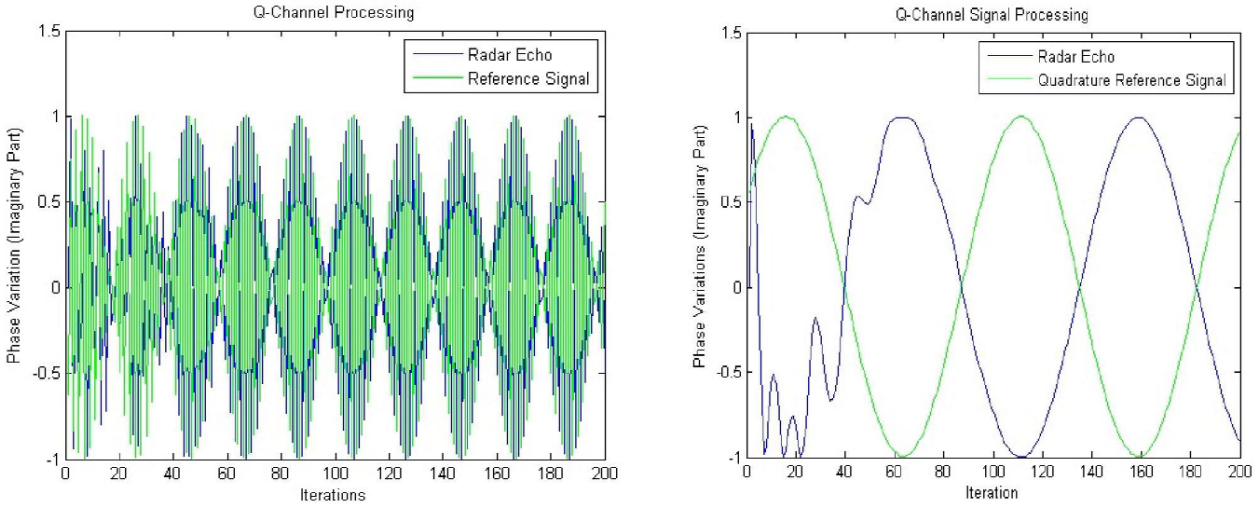


Fig. 7. Phase variation in Quadrature Channel Radar receiver

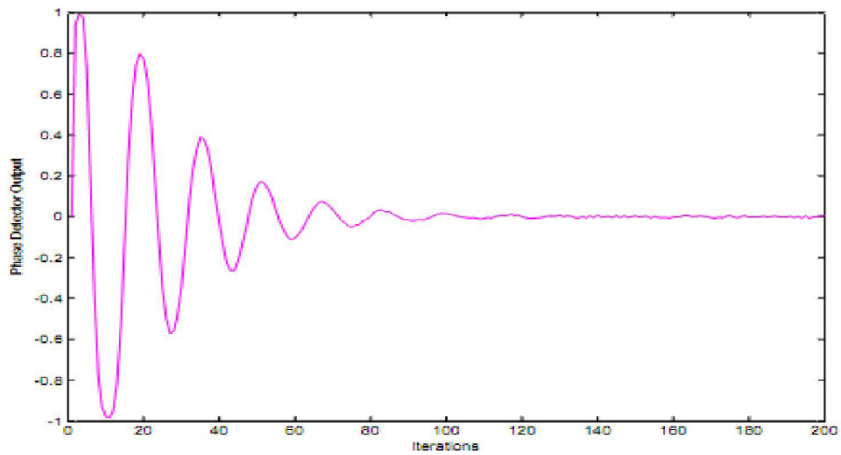


Fig. 8. Phase Detected Output at Radar Receiver

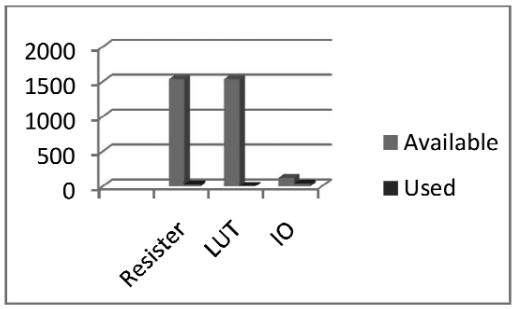


Fig. 9. Resource Utilization

VII. HARDWARE SYNTHESIS RESULT

The main part of VLSI design is optimization of design in terms of speed, power, resource utilization and delay etc. The proposed architecture design was synthesized on Spartan-3 based xc3s50pq208-5 FPGA device using XILINX ISE 10.1 and simulated on ModelSim. The area utilization of proposed design is implemented on above said FPGA kit in terms of Register, LUT and IOs. The area consumed by register, LUTs and IOs are 3%, 2% and 35% respectively of available resources as shown in **Fig. 9**.The result of power measurement for proposed architecture also has been measured using Cadence custom IC design tool and shown in **Table II**.

Table II.
Power Measurement

Cells	Leakage Power (nW)	Dynamic Power (µW)	Total Power (µW)
95	112.132	112.64	112.752

VIII. CONCLUSION

The paper presents the demodulation technique in a high speed FMCW Radar receiver using complex Digital Phase Locked Loop. To maintain high degree of accuracy during radar signal processing, quantization error is needed to be minimized in design level itself. For that purpose, numerical error analysis has been done to minimize angle approximation as well as truncation errors. With using reduced number of micro-rotation and adequate optimized convergence property of CORDIC design, implementation of this kind of demodulator becomes easier. Inherent issue of overflow is quite appropriately resolved in the proposed design. Numbers of micro-rotations have been adjusted so as to achieve better loop performance and speed of operation while minimizing quantization error. The phase and amplitude variations in the received echo signal are aptly demodulated. The BOXCAR filter gives additional facility in reconstructing output signal efficiently.

REFERENCES

[1] W. Xu, C. Gu, C. Li, M. Sarrafzadeh, "Robust doppler radar demodulation via compressed sensing", Electronics Letters, vol. 48, no. 22, pp. 1428-1430, Oct 2012.

[2] N. Da Dalt, "Linearized Analysis of a Digital Bang-Bang PLL and its Validity Limits Applied to Jitter Transfer and Jitter Generation", *IEEE Transactions on Circuits and Systems I:Regular paper*, vol. 55, no. 11, pp. 3663-3675, Dec. 2008.

[3] J. E. Volder, "The CORDIC Trigonometric Computing Technique", *IRE Transactions on Electronic Computing*, vol. EC-8, pp. 330-334, Sept, 1959.

[4] Y. H. Hu, "CORDIC-Based VLSI Architectures for Digital Signal Processing", *IEEE Signal Processing Magazine*, vol.9, no.3, pp.16-35, 1992.

[5] S. Ugazio, " Design of real time adaptive DPLLs for generic and variable Doppler frequency", *In proc. Of International Conference on Localization and GNNS (ICL-GNSS)*", pp.169-174, Tampere, Jun 2011.

[6] K. Kota and J.R. Cavallaro, "Numerical accuracy and hardware tradeoffs for CORDIC arithmetic for special-purpose processors," IEEE Trans. Comput., vol.42, no.7, pp.769–779, Jul 93.

[7] Y. H. Hu, "The Quantization Effects of the CORDIC Algorithm", *IEEE Transactions on signal processing,* vol.40, no.4, pp.834-844, Apr 1992.

[8] S.Y. Park and N.I. Cho, "Fixed-point error analysis of CORDIC processor based on the variance propagation formula," *IEEE Trans. Circuits Syst. I*, vol.51, no.3, pp.573–584, March 2004.

[9] T. –Y. SUNG, H. –C. HSIN, "Fixed-Point Error Analysis of CORDIC Arithmetic for Special-Purpose Signal Processors", *IEICE Trans. Fundamentals*, vol. E90-A, no. 9, Sept 2007.

[10] J. Vuori, "Implementation of a Digital Phase-Locked Loop Using CORDIC Algorithm", *IEEE International Symposium on Circuits and Systems,* vol. 4, pp. 164-167, Atlanta, 1996.

[11] B. R. Mahafza, A. Elsherbeni, "MATLAB Simulations for Radar Systems Design", CRC, Florida, USA, 2003.

Amritakar Mandal received his Bachelor of Engineering Degree in Electronics and Communication Engineering from The Institution of Engineers (India) in 2006. He completed M.Tech with specialization in VLSI from Shobhit University, India in 2009. He is currently working toward the Ph.D Degree in Electronics and Communication Engineering in the Department of Information and Communication Technology, Gautam Buddha University, India. He has vast experience in Electronic Warfare and Surveillance Radar systems. His research interests include Radar Signal Processing and High Speed VLSI Design for Communication Systems.

Rajesh Mishra is currently working as Assistant Professor in School of Information and Communication Technology, Gautam Buddha University Greater Noida, Delhi NCR (India). He received his BE (Electronics Eng.), M. Tech and Ph. D. degree (Reliability Eng.) from Reliability Engineering Centre, IIT Kharagpur (India) in year 2000, 2004, and 2009 respectively. He has research interest in the area of reliability engineering, layout design for capacitated networks, and network optimization. He has published papers in several international journals such as IJPE, IEEE, RESS, QTQM etc.

Movement modeling in cellular networks - Markov Movement Model Creator Framework (MMCF)

Péter Fülöp, Dr. Sándor Szabó

Abstract- **Nowadays, in the wireless networks the number of users and the transferred packet switched data speed are increasing dramatically. Due to the demands and the market competition the services are becoming more and more complex.**
The efficient network dimensioning and configuration highly depends on the underlying mathematical model of user distribution and expected data transfer level. In this paper we propose a Markov Movement-Model Creator Framework (MMCF) for setting up a model based on the network parameters and requirements with optimal number of states.
Firstly we describe a method that yields an abstract model of the mobile network and the node, and we introduce a simple classifying method that defines the necessary parameters of the exact Markov movement model. The mathematical solutions for determining these parameters are also presented in the paper. Finally we analyze the accuracy, complexity and usability of the proposed MMCF and an analytical comparison is made with other mobility models, the comparison is proved by simulations. The movement model created with the framework helps the network operators in setting up an effective authorization, fraud detection system or solving self-configuration issues.

Index Terms-Movement model; Framework; Markov; Accuracy; Complexity; User movements;

I. INTRODUCTION

The number of users and the amount of transferred data is increasing dynamically and substantially in the mobile networks. There are also more and more new technologies, standards (for example HSXPA-High Speed Packet Data, LTE-Long Term Evolution [1]), and future solutions [2][11] to support efficient mobility. Hence the network providers and operators face more and more complex management systems and operation tasks. Wireless multimedia and other services have many requirements and the resources in the network are often expensive and limited. Nowadays the operation tasks have some critical parts, i.e. guaranteeing the security of user-related information and data, providing QoS (Quality of Service), location management and maintaining the service levels in the network.[1]

In recent years people increasingly rely on wireless devices in their daily life for very sensitive tasks such as shopping and bank transactions. Although many authentication protocols are used in wireless, mobile networks, it is still a challenging task to design a fully secure mobile environment because of the open radio transmission environment and the vulnerability of mobile devices. Anomaly-based detection as part of the detection-based techniques creates normal profiles of system states or user behaviours, stores and periodically compares them with the current activities. If significant deviation is detected, the network system raises an alarm. A user profile is very difficult to build up, but it could largely increase the security of a wireless system [4][5]. Movement mobility models and location prediction take significant parts in creating user profiles as well.

The previously mentioned fast evolution of network applications and services require skilled individuals to install, configure and maintain these systems. Other possibility is to introduce mechanisms and procedures, which enable a system to reconfigure, heal or install itself [6]. These systems shall be capable of modifying their own behaviour and adapting environmental changes based on performance measures. A well developed movement mobility model can be used as a proper trigger to rebuild a cell boundary in 4G LTE (Long Term Evolution) network, recognize radio interface problem of mobile access points [1].

Different areas have been introduced above, where mobility models should be used, but the scope of this research area is far more wider nowadays. The discussion of individual or group mobility modelling have been addressed in many papers in the literature [5][7][8][9][10][12][23]. Beside other approaches a few propositions are using Markov model as a sophisticated mathematical solution [8][9][10][13]

As highlighted above an efficient mobility movement model is necessary for network providers nowadays. In this paper our aim is to give general design guidelines to create Markov movement mobility models with optimal number of states and proper accuracy according to the network and user movement parameters.

The structure of the paper is the following. In Section II we give a brief on the mobility models. The model for the network and the mobile node with its mobility parameters are introduced in Section III. This is followed with the definitions of classifying system for Markov movement models. In Section V the Markov Movement model Creator Frameworks is introduced and analysed, while Markov model examples are derived in Section VI. Finally simulation and numerical results are given in Section VII.

II. RELATED WORK

Different mobility models have been proposed in the literature to cope with user mobility in different wireless and mobile networks (e.g. cellular networks, ad hoc networks etc.). In this section we give a short overview of mobility models.

In the traditional Random Walk Mobility Model the node moves from its current location to a new contiguous location by randomly choosing a direction and a speed. The Random Walk Model defines user movement from one position to the next with randomly selected speed and direction. Many derivatives of the Random Walk mobility model have been

Manuscript submitted April 2013, revised May 22, 2013.
Péter Fülöp, Department of Telecommunication, Budapest University of Technology and Economics Magyar tudósok körútja, Budapest 1117, Hungary, fpeti@mcl.hu
Dr. Sándor Szabó, Department of Telecommunication, Budapest University of Technology and Economics Magyar tudósok körútja, Budapest 1117, Hungary, szabos@hit.bme.hu

developed including one, two, three - and d-dimensional walks [20].

A flexible mobility framework for hybrid motion patterns is the Mobility Vector model [7]. A mobility vector expresses the mobility of a node as the sum of two sub-vectors: the Base Vector (BV) $\vec{B} = (bx_v, by_v)$ and the Deviation Vector (DV) $\vec{V} = (vx_v, vy_v)$. The BV defines the major direction and velocity of the node while the DV stores the mobility deviation from the base vector. The mobility vector $\vec{M}$ is expressed as $\vec{M} = B + \alpha V$ where α is an acceleration factor.

The location history of a mobile user is exploited in High-Order Markov Model that is described in [5][8]. The model focuses on the identification of a group of especially harmful internal attackers. The order-o Markov predictor assumes that the location can be predicted from the current context, which is the sequence of the previous o most recent characters in the location history.

F. Lassabe et al. [9] present a mobility model adapted to the logging of mobile positioning or to the tracking of mobiles. This model is based on the All-Kth Markov Model. They present two predictive models from the AKMM: the K-to-1 past Model and its improvement, the K-to-1 past* Model. The model defines a Markov state-space constructed of the possible user trajectories. Each state describes a trajectory section of 1 to K previous locations. The model predicts future locations based on the possibilities of each transition between states. A threshold value is used to select a group of locations which are likely to be visited in the next step, so a handoff procedure can be prepared for each one.

Shiang-Chun Liou et al present a mobility model with two-tier cell structure in [14]. The user trajectory is defined based on the logical function of velocity, direction, acceleration and position. This logical function is converted to a model that uses three preceding geographical locations to estimate the fourth parameter. The location prediction with this estimation enables the network operator to make preparations for a future handoff in the group of cells that are likely to be crossed. Two-tier cell structure is used to decrease the waste of bandwidth due to reserved resources of a future handoff. The two tiers can be described in a mobile cell as a function of distance from the base station (first tier). While the mobile node is close to the base station, it is unlikely that even with a sudden trajectory modification the mobile node steps into another cell. On the other hand, if the mobile node is more close to the cell boundary (second tier), the possibility of a handoff is increasing.

A ring-based mobility prediction and resource reservation algorithm is proposed in [10]. A cell cluster is divided into three cell groups, where the first group is equivalent to the central cell of the cluster, the second and third groups consist of the cells that are located in the first and second cell ring around the central cell, respectively. The pre-handoff resource reservation is derived from the possibilities of the event that the mobile node steps from the central cell into a cell of the second or third cell-group. This approach can be considered as the generalization of the two-tier cell structure described in [14] to an inter-cell level.

W. Ma et al propose a user mobility pattern (UMP) based model (Mobility Pattern-Based Scheme – MPBS) in [16]. The MPBS is a general method to follow users in the network without expensive paging operations if the user meets some requirements. The model defines a personal mobility pattern list which consists of a sequence of register areas (RA, i.e. mobile cells), and a time-sequence of the trajectory on the RA sequence. The time-sequence is built up by the timestamps of handoffs between RAs, and the dwell times for each RA. Based on the time- and RA-sequence, an exact timeline can be defined which is followed by the user. The operator does not need to page the user in different RAs because the timeline shows which RA is the user located in at the actual timestamp. Naturally, the ideal user who always follows the timeline does not exist, but the time-sequence and RA-sequence provide information even if the actual timeline differs from the pre-recorded one. A categorization is presented with four categories where the first category is the ideal user with a timeline-compatible trajectory. The second category involves users who are following the RA-sequence but with time delays or hurries, that is the network operator can find the user in the remaining RA set after the last paging or location update. Users who are located in the appropriate RA set, but are not following the sequence are in the third category. The fourth category is for the users who are located out of their UMPs, that is their actual trajectories are not close the pre-recorded ones.

III. MODELLING THE NETWORK

In this section we collect the most significant properties and parameters of the mobile network that can describe an abstract network model.

A. Basic notations and descriptions

We define the basic notations that we use is the article. The basic model will resemble to the abstract one in our previous work [16][17].

- The specific network with all its parameters is denoted with N.
- The Mobile Nodes (MN, alias mobiles, moving entities, users) are the mobile equipments that want to communicate with other mobile nodes or fixed partners and move between the radio access points. The number of users (number of MNs) in the model is denoted by n_u.
- There are Mobility Access Points (MAP, alias cells), these are the only entities that are capable of communicating with the mobile nodes via radio interface. All mobility access points have their own geographical areas. While the MN moves in an area, it is always connected to the owner of the area. The number of mobility access points in the model is denoted by n_m.
- The user can connect to MAPs with handovers from the neighbouring MAPs, each user is connected to only one MAP at a time. The neighbour MAPs could use even different access technologies than the current MAP and they could be located in the very same geographical place as well, the model does not require single access technology in the whole network. The number of neighbour MAPs of MAP_i is denoted by n^i_{nm}.
- There are other network elements which provide the communication in the core network behind the MAPs. We denote these as Network Elements (NE).

- Network trace is an abstraction of the network operation log, it contains 4-tuples of a timestamp, user ID, MAP ID and network event. A trace entry could mean for example that the selected user connected with handover to the MAP at the given timestamp. The network trace contains all information of the mobility of the users in the network.

B. *Deriving parameters of a given network*

A way to describe a network is to observe the network trace. We introduce a method to process the network traces to calculate typical parameters of the mobility. The trace entry describes the events in a cellular network. An event might be a state change of the given user (e.g. mobile node is in idle status, voice call or data transfer is set up, cell boundary crossing). The logical location of the event is determined by the MAP ID where the user is located at the timestamp of the event. (Table 1). The events are recorded in the network management system's logs, thus the information can be extracted from the management system of cellular mobile networks.

TABLE 1. AN EXAMPLE OF WIRELESS NETWORK MANAGEMENT SYSTEM'S LOG

Timestamp	User ID	MAP id	State or Event
...			
09:21:43:12	41	4951	Idle
09:21:43:12	41	4957	Idle
09:21:43:12	41	4957	Voice call
09:21:43:12	19	5341	Data Call/Traffic class2
09:21:43:12	84	7120	Idle
09:21:43:12	19	5348	Data Call/Traffic class2
09:21:43:12	19	5348	Idle
...			

The aim is to derive the parameters of the user mobility, therefore we should pick the relevant entries from the network trace. In our work we focused on location changes of users, handovers, and initializing or receiving calls. These events are observed during a time interval that is considered to be the reference interval for deriving model parameters.

We assumed that the user distribution in the network is given at the first moment of the reference interval. We created a discrete sample series where samples are taken at Δt time intervals, that is a location state is assigned to every users per Δt time. Δt is defined system-wide as the minimum of the time intervals elapsed between two events registered to the same user. That is the sample frequency is set to the "fastest" user in the network. This ensures that every user event and all state reports are processed. With this sampling a MAP ID and a state can be determined to every user in every timeslot. The sampling results a n_u x n_T sized P matrix, where n_T denotes the number of timeslots, and n_u the number of users in the model. P matrix stores the MAP IDs and state of each user in each timeslot.

The relative frequency of any state can be determined based on P matrix, for instance the relative frequency of receiving voice call in a MAP, or even the handover rate between two different MAPs. We defined the S set, which contains all possible states and events appearing in the logs. The important ones are the following:

-receiving voice call -initialling data call
-receiving data call -fall back into idle status
-initialling voice call

Depending on the detail of the logs and on the requirements more or different states, events could be investigated as well. For example if more data traffic classes are determined based on the logs, then they could also be differentiated. But for us the above mentioned states are sufficient in the model.

To determine the relative frequencies of states in a MAP, the state must occur frequently enough, otherwise it is neglected. Let us define the n_u x n_m $C_s=[c^s_{i,j}]$ matrix, where $c^s_{i,j}$, $s \in S$, is the occurrence of s state with user i, in MAP j from the P matrix. The average occurrence of s state in the network is:

$$c^s = \frac{\sum_{\forall i}\sum_{\forall j} c_{i,j}}{n_u \cdot n_m} \qquad (1)$$

The parameter c_s can be used as a main requirement in order to create a valid model based on the network description. So the network must be monitored for sufficient time before we create a model from it.

Parameter ε_c denotes the minimal occurrence of a state for acceptance, if the occurrence is smaller than ε_c, the rate will be 0. Based on this the relative frequency matrix of a state can be determined as the following:

$$D^s = [d^s_{i,j}] = \begin{cases} 0, & c^s_{i,j} < \varepsilon_C \\ \dfrac{c^s_{i,j}}{c^s_j} \end{cases} \qquad c^s_j = \sum_{\forall i} c^s_{i,j} \; s \in S \quad (2)$$

In fact $d^s_{i,j}$ is the probability of getting into the state s happens to user i in MAP j.

We determine the rate of receiving a call with μ. It can be determined for every MN in every MAP from the $D^\mu=[\mu_{i,j}]$, but the average value is also calculable in similar manner described above:

$$\mu = \frac{\sum_{\forall i}\sum_{\forall j} \mu_{i,j}}{n_u \cdot n_m} \qquad (3)$$

Let us have the corresponding network graph given with its weighted adjacency matrix: A.

Let us assume that the aggregated behaviour of the Mobile Nodes can be modelled with a finite state continuous Markov chain (the handover or call arrival rate is a Poisson process with various intensity parameters as in many works, e.g. [18]). The chain is given by a rate matrix $B_Q = [b_{ij}]$. In this matrix, all the possible MAP-s are listed so the matrix will be a n_m x n_m matrix where each element b_{ij} denotes how frequent the movement of the mobile is from $MAP_i \rightarrow MAP_j$. If an MA is not a MAP then there are 0 values in its row and column (i.e. we treat it the same way that the MN cannot or never attaches to it). From the rate matrix the transition matrix B_Π can be determined easily. We assume that the matrix B_Π, without the non-MAP nodes, is practically irreducible and aperiodic that implies that the chain is stable and there exists a stationary distribution. This will be denoted by a density vector $\underline{b}$. Other B matrices can be determined for a single user, user group or all the users as well and they can be assigned to a state in the network also. According these assumptions, for example $B^{i,s}_\Pi$ is the transition probability matrix of user i, when it is in state s. Another network describing parameter, which is useful during

the modeling, is the number of visited MAPs by the user or users. That is calculated as follows:

$$n_{vm} = \underline{1} \cdot sign(B_{\Pi} \cdot \underline{1}) \qquad (4)$$

A general network describing parameter, the weighted average of visited MAPs is:

$$w_{vm} = \frac{\sum_{\forall i} \underline{b} \cdot sign(B^{i}{}_{\Pi} \cdot \underline{1})}{n_u} \qquad (5)$$

The other parameter is the average number of neighbouring MAs that can be accessed via a wire from a given node: w_{nm}. It should be also weighted with the probability density of the MN.

$$w_{nm} = \frac{(\underline{b} * sign(B_{\Pi})) \cdot \underline{1}}{n_a} \qquad (6)$$

When talking about an existing network, the parameters described in this section can be calculated easily, producing the base of the model.

IV. CLASSIFYING MARKOV MOVEMENT MODELS

In this section, a simple classification of motion models is presented. The aim is to compare different models easily and to analyse them in different network environments. The purpose is to determine the attributes of the individual models, the level, the depth and the resolution and hence the models can be rated in a general marking system.

We used a discrete-time, finite-state and infinite-state Markov-chain to model the mobile movement.

The classification system handles simple, general Markov-chain based mobility models, in that a user or users can be located in different Markov states. MAP or group of MAPs (or merged MAPs based on a special relationship) is mapped to a state or more states of the Markov-chain model. Let us define *X(t)* as a random variable, which represents the movement state of a mobile terminal during timeslot t. The transition probabilities of the Markov model can be determined from the describing parameters of the network. Let us assume that the Markov chain is always irreducible and aperiodic, so the stationary user distribution is determinable.

Two main types of these Markov mobility models are distinguished, the *User-Centralized* and the *Access Point Centralized* model. The latter one is further separated into two subtypes. Figure 1 depicts this main classification.

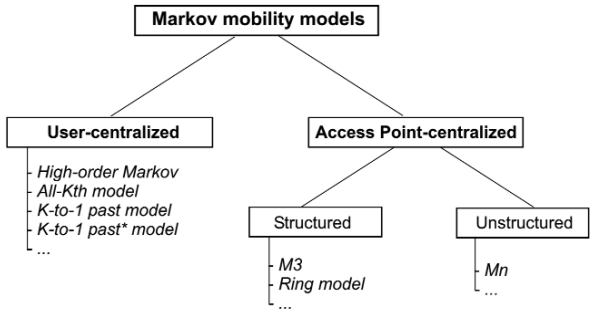


Fig. 1 The general classifying of the Markov mobility models

In the next subsection these groups of models are explained.

A. User-centralized Markov models (UCM)

A user or a group of users from the network is selected for observation in user-centralized Markov models. The users' movement behaviour is modelled with a Markov model. Only the MAPs which are visited by the selected user(s) are taken into account, other MAPs and other users do not affect the structure of the movement model.

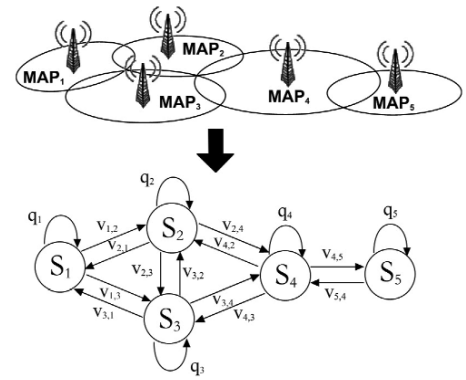


Fig. 2 The creation of user-centralized models

Figure 2 shows how to represent a user centralized model. The chosen user in the example visits only the MAPs between ID 1 and ID 5. Each MAP is mapped into one standalone Markov state. This is a very simple model, where the stationary distribution of the Markov chain is equal to density vector *b* as the stationary distribution of B_{Π} transition matrix.

This usage of this model is reasonable, if the behaviour of the user is to be investigated, or a user profile is needed for example for fraud detection.

Most of the Markov mobility models in the literature belong to this class, for example [5], [9].

B. Access Point Centralized Markov models (ACM)

The access point-centralized Markov models can be used when the user distribution in a selected MAP or group of MAPs must be determined. Instead of modelling the behaviour of an individual user, a MAP and its environment is to be observed. In these cases a MAP or more MAPs and their defined neighbours are selected according to a requirement. The users who stepped into the area of the observed MAPs are investigated and their distribution is used to build a model for prediction.

Two guidelines exist:

-In the structured model, for a predefined reason certain MAPs are grouped together, this creating a regular structure in the model.

-a MAP or MAPs are simply mapped into a state of Markov model. This method is called unstructured model.

Details and examples are presented in the next sub-sections.

Structured Markov Models (ACSM)

In the structured Markov model groups of MAPs are defined. The grouping can be derived from user behaviour, geographical specialty or even from network requirements. Figure 3 shows examples for structured solutions, the Ring Model (*RM*) [10] and the *M3* model [13].

In the *RM* (Figure 3.b) the ring consists of cells surrounding a central cell. The concept is to simplify the calculations, if we are interested only in the number of users arriving to a given ring, or leaving a given ring during a time period. Internal movements are disregarded [10].

The *M3* model handles users in four different Markov states, the right-area, left-area, stay and outside state. More details about the M3 model are given in Section VI.A [13].

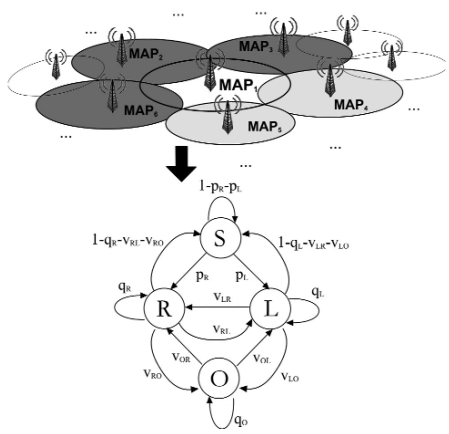


Fig. 3.a The representation of access point-centralized/structured models, M3 model [13]

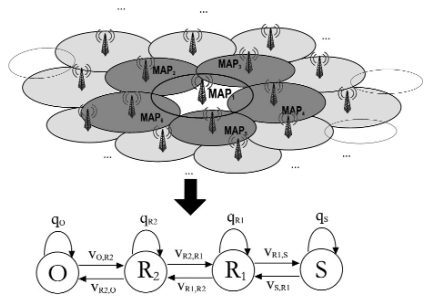


Fig. 3.b The representation of access point-centralized/structured models, Ring model [10]

Let us introduce the theoretical error E_T, which is the sum of error percentages per MAP in prediction. In detail during the prediction, if we want to determine the number of users in the MAPs of the group (for example the right-area state in *M3*, or the first ring in *RM*), then the predicted number of the users for the group is distributed uniformly between the MAPs in the group. Obviously this step brings a theoretical error (E_T) into the prediction process. For example in the left area state of the *M3* model there are MAP_1, MAP_2 handled together as a group. We know that in the left area state 100 users move. For lack of further information 50-50 users are predicted in MAP_1, MAP_2. Actually there is 25 users in MAP_1 and 75 in MAP_2. In this case the E_T is 50% in MAP_1, MAP_2 as well.

Unstructured Markov Models (ACSM)

If we try to predict the user's distribution in a city having irregular, dense road system, or in a big park where people are able to move around then, the handover intensities could differ thus the calculations above could produce errors. From this point of view the best way is if we represent all of neighbour MAPs as a separated Markov state, so this results the Unstructured Markov Model. The results from determined stationary distribution are easy to map back, into the MAPs.

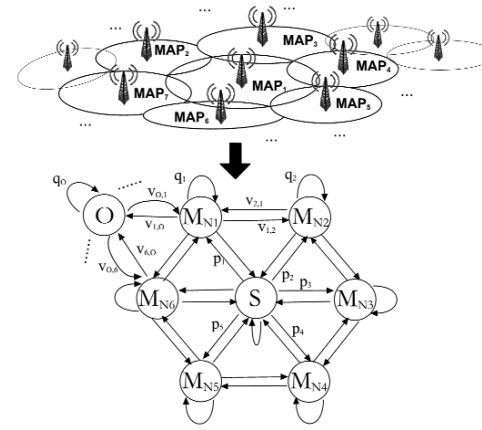


Fig. 4 Access point-centralized/Unstructured/M7 model [13]

Figure 4 depicts the methodology of unstructured model representation. The Markov chain in Figure 4 is similar to *M3* model. In this *M7* model all of neighbour MAPs are mapped into Markov-chain states. The *M7* and generalized *Mn* model described later, in 'Markov model examples' section.

C. *Attributes*

The example models introduced above are the simplest ones in their class. In this section we determine attributes to the classifying system which describe important parameters of the Markov models. Supported by these attributes, more complex, more sophisticated models could be classified or constructed for solving more difficult problems.

The examples mentioned in the introduction of this section used present only one attribute at a time to keep the simplicity and distinctness. Of course the attributes could be used together in any number and combination.

Level of the model

As mentioned in Section III.B, B_{II} could be determined from the *P* matrix for every state as well. There are two main reasons to handle the states differently:

-the users behave differently in certain states,

-the users in distinct states must be modelled in a different way (for example different CAC is used for the users in voice call, than the users downloading data from the internet).

In these cases the B_{II} must be calculated for different states. This diversity in the model is represented by 'levels' or 'dimension' (Figure 5). The transition rates between the levels show the intensity state changes in the current MAPs. This model is similar to the one described in [19].

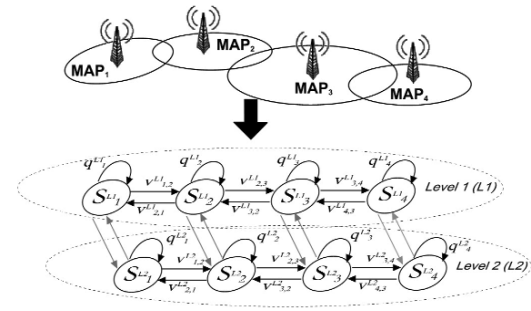


Fig. 5 Attributes in Markov mobility model classifying: Example for meaning of 'level'

The number of levels in the model is determined with n_L. A level is denoted with L, the levels in the model are marked with L vector, where $\underline{L} = [L_1,...,L_{n_L}]$. A specific L is based on its B^s_{II} matrix.

Resolution of the model

There is a possibility to merge adjacent MAPs together, if those MAPs are not needed to be handled separately. If outgoing predictions of users in two adjacent MAPs match within a certain limit, the two adjacent MAPs could be merged together and handled henceforward as a new major MAP. By this the complexity of the model can be decreased. Figure 6 shows an example, in which the 6 neighbour MAPs of an access point-centralized, unstructured model, are merged into 3 new major MAPs.

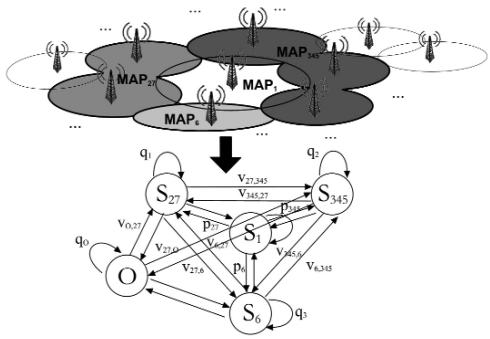


Fig. 6 Attributes in Markov mobility model classifying: Example for meaning of 'resolution'

'Grouping' explained in Section IV.B.1 (structured models) is not equal to 'merging' mentioned here. As the result of 'merging' new, major MAPs are created instead of the initial ones. A 'grouping' organizes the MAPs into a structure.

Every level could have its own resolution. The resolution is denoted with R, $R = \{G_1,...,G_{n_M}\}$ where G is a set of merged MAPs, and n_M the number of new MAPs after merging. The R is described with a general rate, n_m:n_M. The vector $\underline{R} = [R_1,...,R_{n_L}]$ contains the resolution rules to every level.

Memory of the model

The application of the recent user locations has a crucial importance in a variable, directional user motion. Neglecting the preceding transition series of a user in the MAP results that the estimation could work with a theoretical error (E_T, like in M3 and RM model) [13].

It is very important that the usage of this type of 'memory' does not violate the Markovian property, the memorylessness is still true for the Markov model created by the MMCF.

We present a simple example which shows the effect of depth or memory in the model in our previous work [13]. If we consider the two roads shown is Figure 7.b, the accuracy of the transition probability estimations is higher when the model knows where the users come from than an estimation which cannot distinguish the users on the two roads (Figure 7.a.).

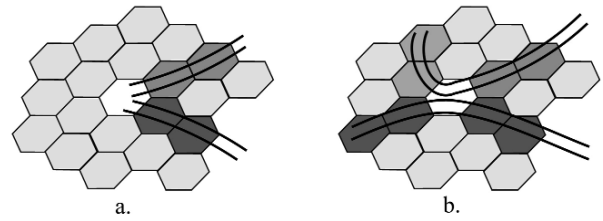


Fig. 7 User prediction methods
a. model without using *memory*, unknown where is the users come from
b. model with *memory*, the previous steps of the users taken in account

The results show that our proposition of using memory in a mobility model significantly increases the accuracy of the model in cases when the ID distribution in an arbitrary cell has high variance, or has periodicity without stationary distribution.

Therefore, an o-depth could be determined for our Markov models similar like in [5]. In our model, sequence of MAP IDs can be assigned to every MAP not to a user; ID_1, ID_2, . . ., ID_i, . . . , where ID_i denotes the identity of the MAP visited by the mobile before it stepped into the current MAP. The last element of the sequence is always the current MAP. The future locations of the mobile in most of the cases are correlated with its movement history. The probability that the user moves to a particular MAP depends on the location of the current cell and a list of cells recently visited. If only the current cell is taken into account, like in previous examples, the depth is 1.

For every MAP different depth could be assigned, which determines the length of the recently visited MAP ID list before the current MAP. Since a MAP could be reached on different paths by the users, therefore a more specific MAP ID list could belong to a MAP, and for this reason more Markov state assigned to a MAP, see Figure 9.

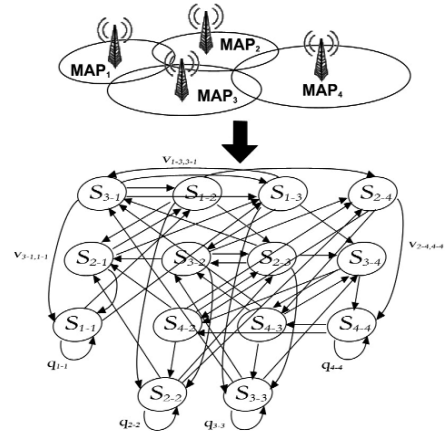


Fig. 9 Example for meaning 'memory', 2nd model [9]

Thus $\underline{O} = [o_1,..., o_{n_m}]$ matrix denotes the depth of the model for a level, where o_i is the applied sequence length of previous visited MAPs to MAP_i. Generally the n_m x n_L O matrix ($\underline{\underline{O}} = [\underline{O}_1,...,\underline{O}_{n_L}]$) belongs to a Markov model.

The weighted average depth for a model is the following:

$$w_O = \frac{\sum_{i=1}^{n_L} \underline{b}^i \cdot \underline{O}^i}{n_L} \qquad (7)$$

Complexity

The complexity of the model could be denoted by the number of states. Following the determining of attributes the number of states is:

$$n_{states} = \sum_{\forall l \in L} {n^l}_m / {n^l}_M \cdot {w^l}_{vnm}{}^{w^l O} \qquad (8)$$

D. *Performance analysis*

In this section we present a simple performance analysis between the Markov models introduced in the previous sections. The models are examined in different network environments. In every scenario the analysis is performed on a cluster of 7 hexagonal radio cells (Figure 10). In our interpretation the performance of the mentioned Markov models depends on the theoretical error (E_T) and the number of states (n_{states}), in this case a special theoretical cost $C_{MM}=f(n_{states},E_T)$. The lower this cost is, the better the performance. The computational capacity increases, hence the steady state probabilities determination of Markov models with more states are less difficult. But the theoretical error in prediction is more important. For these reasons we calculate the theoretical cost this way:

$$C_{MM} = \log n_{states}(1+E_T)^2 \qquad (9)$$

where n_{states} is the number of states and E_T is the theoretical error. In previous sections, in case of *RM*, *M3* and *o*-th models the E_T was introduced, which is the sum of error percentages per MAP in prediction.

The following models were examined in the performance analysis:

-RM: The model introduced in Section IV.B.1 (Figure 3.b) with the difference that only one ring is around the central MAP, so only states *S*, *R1*, and *O* exist. The predicted number of the users in *R1* state is distributed uniformly between the MAP_1-MAP_6.

-M3: Also introduced in Section IV.B.1 (Figure 3.a.). MAP_2, MAP_3 and MAP_7 belong to the left-area state (L). The right-area state (R) includes MAP_4, MAP_5, and MAP_6.

-M7: The model presented in Section IV.B.2 (Figure 4).

-2nd: The extension of model M7, the memory is increased to 2. It is similar to 2nd model in Figure 9.

-3D: This is a two-level model resulted by duplicating the Markov chains of the M7 model. The first level represents the users who establish data connections, while the second level represents the mobile users initiating and receiving voice calls.

We investigated special network environments (scenarios) to highlight the advantages and disadvantages of each model. The different network scenarios are shown in Figure 10. These are the following cases:

- Scenario a.: 'A park, uniform user distribution'
- Scenario b.: 'Simple road'
- Scenario c.: 'Highway to city'
- Scenario d.: 'Directional motion'
- Scenario e.: 'Differentiated users'

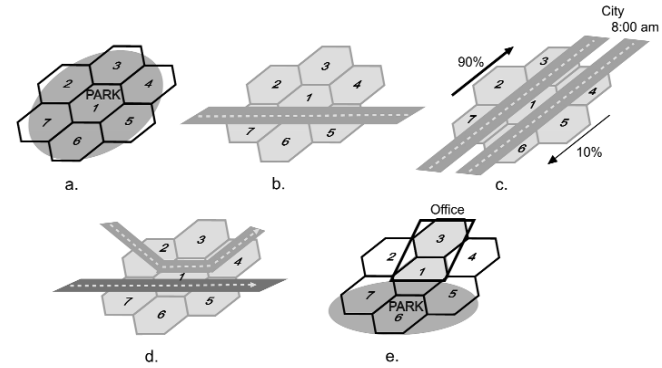


Fig. 10 The network scenarios for performance analysis

TABLE 2. THE RESULT OF PERFORMANCE ANALYSIS

C_{MM}	a.	b.	c.	d.	e.
RM (n_{states}=3)	*0,47*	*2,55*	*1,52*	*2,55*	*2,55*
M3 (n_{states}=4)	*0,6*	***3,2***	*0,6*	*3,2*	*3,2*
M7 (n_{states}=8)	*0,9*	*0,9*	*0,9*	*3,6*	*3,6*
2nd (n_{states}=49)	***1,69***	*1,69*	***1,69***	*1,69*	***6,76***
3D (n_{states}=16)	*1,2*	*1,2*	*1,2*	***4,8***	*1,2*

Table 2 shows the result of the analysis. C_{MM} was calculated in every scenarios for every model. The highlighted values are the best, the boldfaced the worst results for the current scenario.

Scenario a. represents a simple park, where the distribution of the users is uniform so thus the E_T is 0 in each model. In this case the best performance belongs to RM, because it has the minimum number of states.

In scenario b. a road crosses the examined area. The users are distributed uniformly on the road. In the left state of the *M3* model, the predicted distribution of the users in 33,3% in each MAP group (MAP_7, MAP_2, MAP_3), while all of them stay in *M7*. Therefore the model has 133,3% theoretical error. The theoretical error of the RM model can be determined the same way. *M7*, 2nd and 3D make no error. The optimal choice is *M7* in this situation.

The scenario c. presents a morning, rush-hour traffic situation towards the city. On the road the users are distributed uniformly. The road to the city located fully in the left-area state of the *M3*, and the road from the city is in the right-area state. Because these states are handled separately there is no theoretical error. However the *RM* based prediction distribute the users uniformly between the 6 neighbor MAPs.

Scenario d. is very similar to the example presented in Section IV.B.3.3. The users from MAP_2 move to MAP_4 via MAP_1. The other path is MAP_7-MAP_1-MAP_5. 50% of users step out from MAP_1 to MAP_2, the other 50% to MAP_5. In the measured time-slot all of the users are in MAP_1, on the road below, so the next step is MAP_5. Because only the 2nd model

takes into account the previous step, all other model have the theoretical error.

Scenario e. is a special case. The mobile users in the office generate data traffic, in the park there are rather voice calls. At current t timeslot in MAP_1 the park is empty, users moves only in the building. In this case 3D model makes no error as opposed to other models.

One can see that each model performs best in specific situations, accordingly it is important to choose the proper model in every network scenario.

V. MARKOV MOVEMENT-MODEL CREATOR FRAMEWORK - MMCF

In the previous section we showed that every model works well in a given network environment. Therefore the optimal model has to be chosen, or constructed for a specific network environment.

If every MAP is handled separately and all previous steps, user groups are taken into account then the modelling Markov chain will contain a high number of states. Therefore a minimal error rate should be allowed and the appropriate Markov motion model with the minimal number of states should be found.

In this section we give guidelines to construct Markov movement mobility model for a network, and we present the determination of the classification system attributes.

We proposed a general Markov Movement model Creator System (MMCS). For every $N<A,D,B>$ network, where A is the adjacent matrix of the network, D is the occurrence matrix and B_{II} is the handover matrix and for ε error vector, an optimal number of states $M<L,R,O>$ Markov model can be determined, where L is the level-, R is the resolution-, and O is the depth of the model.

A. Error vector

The error vector is an input for the MMCF. The elements of the vector are as follows:

- ε_{ds} the limit of mean difference between the transition matrixes for different user states.
- ε_{rd} the acceptable error rate of outgoing prediction probability, when MAPs are merging together.
- ε_{op} the limit of the difference of the outgoing probabilities of the sequences from current MAP.
- ε_{uv} the limit of fluctuation of the number of users arriving from a certain sequence directions is investigated .

More details about the error rates are presented in V.C, V.D and V.E Sections.

B. Main type of the model

The proper main type of the model is determined by the goals and requirements, not by mathematical computation. Beside the mentioned examples there are some guidelines for selecting the best model type according terms and conditions:

- User- Centralized
  - Modelling from user point of view
  - User profile creation
  - Fraud detection
- Access Point-Centralized
  - Modelling from cell point of view
  - CAC in a MAP
  - Movement modelling of a geographical area

C. Determining the level

A new level should be applied in the model, if the mean difference between transition matrices for different user states is greater than a predefined limit. Of course if there is a requirement to use levels, then it must applied independently from the calculation.

Let us define ε_{ds} as the limit of the mean difference between the B_{II} and transition matrix for different s user states (B^s_{II}). The average weighted deviation can be calculated by the following form:

$$w^s{}_{ds} = \frac{\underline{b} \cdot abs(B_{\Pi} - B^s{}_{\Pi}) \cdot \underline{1}}{n_u{}^2} \quad s \in S \qquad (10)$$

If $w^s{}_{ds} > \varepsilon_{ds}$ is true for a state s, then a new level must be introduced into the model for state s. This inequality must be analysed for all $s \in S$.

The number of levels can be calculated as follows:

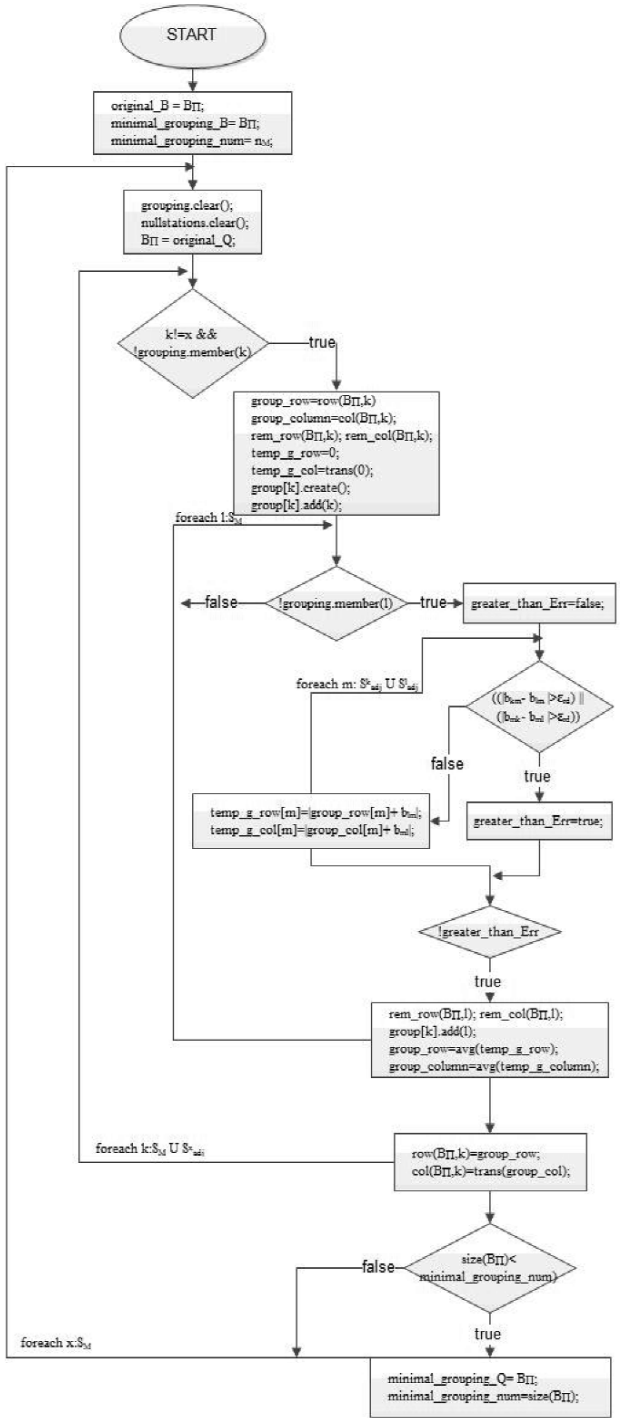
$$n_L = \sum_{\forall s \in S} sign(v^s{}_{ds} + abs(v^s{}_{ds})), \qquad v^s{}_{ds} = w^s{}_{ds} - \varepsilon_{ds} \quad (11)$$

D. Determining the resolution

As we explained earlier if outgoing predictions of users in more adjacent MAPs match within a certain limit, then the MAPs could be merged together to create a new, major MAP. With this step the complexity of the model is decreasing.

We present a simple algorithm to determine the resolution. The input parameters of the algorithms are:

- ε_{rd}, the acceptable error rate of outgoing prediction probability, when MAPs are merging together,
- transition matrix, B_{Π} , as it was defined earlier,



where S_M the set of MAPs in the model, S^a_{adj} the set of neighbouring MAPs of MAP_a , *row(A,a)* the *a*-th row of matrix *A*, *col(A,a)* the *a*-th column of matrix *A*. This results in a smaller transition matrix, B_{II}*, which leads to a simpler model.

E. Determining the memory

The future movement of the users is highly influenced by the path they have taken in the past to reach the investigated point. Leaving this out of consideration would introduce large errors into the mobility model. However it is not always useful to look back into each direction or to look back in equal depth into each direction from every MAP.

The determination of *memory* or *depth* needs proper precaution. The depth exponentially increases the number of states in the model. This can be seen in Figure 9, where the depth is generally 2, for all MAPs ($\underline{Q} = [2,2,2,2]$).

The main idea is to analyse each MAP sequence, visited by the users and decide its importance for consideration. The analysis starts with a sequence of length 2 (length 1 means that only the current MAP is observed) and it is increased one by one. If a sequence of length *i* belongs to a MAP that is not important, then it will decreased, and *i-1* depth will denoted for the MAP. The importance of *k* depth is decided based on the following basic criteria:

- Take the MAP ID sequences for k length, which differ in the first MAP ID and belong to a current MAP. The difference of the outgoing probabilities of the sequences from current MAP must be investigated. Let us define ε_{op} as a limit for this difference. The difference for a MAP_i and k-depth is determined the following way:

$$Df[op]_i^k = \frac{\sum_{b\in Q_i^k}\sum_{a\in Q_i^k}\frac{\sum_{l\forall NB_i}\left|b_{q(a),l}-b_{q(b),l}\right|}{n^i_{nm}}}{(w_{vnm}{}^k)^2} \qquad (12)$$

where Q_{ik} is the set of existing *k* length sequences from MAP_i, NB_i is the set of neighbour MAP IDs of MAP_i and *q* denotes a sequence from the set.

The first criterion of importance is: $Df[op]_i^k > \varepsilon_{op}$.

- Take the MAP ID sequences for *m* length, which differ in the first MAP ID and belong to a current MAP. The fluctuation of the number of users arriving from a certain sequence directions is investigated. Let us define ε_{uv} as a limit for this variance. The variance of number of incoming users from a sequence into MAP_i:

$$V_i^a = E\left(\left|E\left(\frac{n^{i,a}{}_u}{n^i{}_u}\right)-\left(\frac{n^{i,a}{}_u}{n^i{}_u}\right)\right|\right)=\sigma\left(\frac{n^{i,a}{}_u}{n^i{}_u}\right) \qquad (13)$$

where $n^{i,a}{}_u$ is the number of users in MAP_i arrived from a sequence (path), $n^i{}_u$ is the number of users in MAP_i.

This must be examined for all of the incoming sequences:

$$Df\,[uv]_i^k = \frac{\sum_{a\in Q_i^k} V_i^{q(a)}}{w_{vnm}{}^k} \qquad (14)$$

The second criterion of importance is: $Df[uv]_i^k > \varepsilon_{uv}$.

This two criteria, $Df[op]_i^m > \varepsilon op$ and $Df[uv]_i^m > \varepsilon_{uv}$ must be applied for all MAP in order to determine $\underline{Q}$.

VI. MARKOV MODEL EXAMPLES

In this section we describe some of the previous introduced, classified models.

A. The M3 model

In the *M3*, Markov-chain based model: a user can be located in four different states during each time slot, the stay state (S), the left-area state (L), the right-area state (R) and outside-area (O) (Figure 3.a).

The grouping can be derived from the user behaviour. If the users in right-hand side cells behave similarly from the current cell's point of view, the neighbouring cells will be merged into a common cell group, which represents a state in the Markov model (R state). Other grouping methods can be used as well, i.e. a standalone cell can constitutes a group also. In this example model each of the two groups (R and L) contains three cells. The state O represents the outside area, where users can come in to the L, and R state from, and where users can go from the L, and R state to [13].

This model performs well when the user's distribution in the left- or right-area state is uniform.

The *M3* model could be determined as an MMCF model, with the following parameters:

n_L=1, $\underline{O}$=[1,1,1], $\underline{R}$ =[1,{2,3,4},{5,6,7}].

B. Generalized Mn model

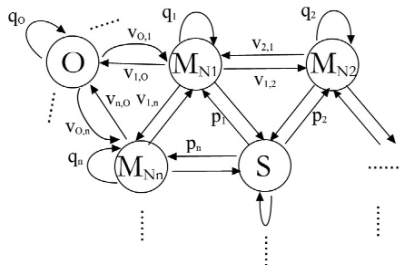
We have enhanced the access point-centralized Markov model, *M3* to generalized *n+2* state Markov model (*Mn*).

If we try to predict the user's distribution in a city having irregular, dense road system, or in a big park where people are able to move around then the handover intensities could differ thus the *M3* model could produce errors. From this point of view the best way is if we represent all of the neighbour cells as a separated Markov state.

As we described above, in the unstructured models a MAP is simply mapped into a Markov-chain state, so 8 states are created, because 7 elements are assumed in a theoretical cluster (direction dimension is 6) and another one to the outside world. This results the M7 model [13].

It is to be taken into account that in the real networks a cell does not always have six neighbours depending on the coverage. This model has to be generalized for a common case when a cell has n neighbour cells. We expanded our previously mentioned model to n-neighbour case (Figure 11), when all the n neighbours are represented with a Markov state:

- stationary state (*S*)
- neighbour 1...n state (M_{N1}... M_{Nn})
- outside area state (*O*)



$$\Pi = \begin{bmatrix} 1-\sum_{m=1}^{n} p_m & p_1 & p_2 & \dots & \dots & \dots & p_n & 0 \\ 1-q_1-v_{1,2}-v_{1,n}-v_{1,O} & q_1 & v_{1,2} & 0 & \dots & \dots & v_{1,n} & v_{1,O} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1-q_i-v_{i,i-1}-v_{i,i+1}-v_{i,O} & 0 & \dots & v_{i,i-1} & q_i & v_{i,i+1} & \dots & v_{i,O} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1-q_n-v_{n,1}-v_{n,n-1}-v_{n,O} & v_{n,1} & \dots & \dots & 0 & v_{n,n-1} & q_n & v_{n,O} \\ 0 & v_{O,1} & v_{O,2} & \dots & \dots & v_{O,n-1} & v_{O,n} & q_O \end{bmatrix}$$

Fig. 11 State diagram and Π matrix of n+2-state Markov model (*Mn*)

The steady state probabilities (*Ps*, P_{N1}...P_{Nn}, P_{NO}) can be calculated.

Using the result the distribution of the mobile users is determinable in the steady state. The predicted number of users in the next time slot is given in Eq. 15.

$$N_i^{t+1} = N_i^t \cdot (1-\sum_{m=1}^{n} p_m(i)) + \\ + \sum_{j,MAP_j \in S_{adj}^i} N_j^t \cdot (1-q_j(i)-v_{j,j+1 \bmod n}(i)-v_{j,j-1 \bmod n}(i)-v_{j,O}(i)) \qquad (a) \\ \dots \\ N_k^{t+1} = N_k^t \cdot q_k(i) + N_i^t \cdot p_k(i) + \sum_{j,MAP_j \in (S_{adj}^k \cap S_{adj}^i)} N_j \cdot v_{j,k}(i) + \\ + N_O^t \cdot v_{O,k}(i) \qquad k \in S_{adj}^i \qquad (b) \qquad (15)$$

Because all of neighbour MAPs is represented as separated states, the prediction based *Mn* model is more accurate than the *M3* model, or at least as accurate as model *M3*.

The inaccuracy of prediction N_i^{t+1} from MAP_j is the following:

$$\left(\frac{1}{3}-\frac{N_j^t}{N_K^t}\right) N_K^t \cdot b_{j,i} \qquad (16)$$

where N_K is the number of users in the group (left-area or right-area in *M3* model) where MAP_j belongs.

This is calculated for all MAP and weighted with the number of users in the group:

$$\sum_{K=L,R} \frac{\sum_{j,MAP_j \in S_k} \left\{ \left| \frac{1}{3} - \left(\frac{N_j^t}{N_K^t} \right) \right| N_K^t \cdot b_{i,j} \right\}}{N_K^t} \qquad (17)$$

Let us determine the inaccuracy in a different way. The difference between the two predictions, *M3* and *Mn*, weighted with the real number of users:

$$\frac{N_{i\ M3}^{t+1} - N_{i\ Mn}^{t+1}}{N_i^{t+1}} \qquad (18)$$

Instead of transition probabilities *p*, *q*, *v* the elements of the rate matrix $B_Q = [b_{ij}]$ are used. The Eq. 19 shows the proof, which follows the same result as in Eq. 17:

$$\left(N_i^t \cdot b_{i,i} + N_L^t \cdot b_{L,i} + N_R^t \cdot b_{R,i}\right) - \left(N_i^t \cdot b_{i,i} + \sum_{j,MAP_j \in S_{adj}^i} N_j^t \cdot b_{j,i}\right) \\ \left(\sum_{K=L,R} \sum_{j,S_{adj(K)}^i} N_j \cdot b_{j,i}\right) - \left(\sum_{j,S_{adj}^i} N_j^t \cdot b_{j,i}\right) = \\ \sum_{K=L,R} \left(\sum_{j,S_{adj}^i} \left(\frac{1}{3} - \frac{N_j}{N_k} \right) N_k \cdot b_{j,i} \right) = \qquad (19)$$

To generalize the equation, instead of $b_{j,i}$ the average handover rate, λ is used:

$$\sum_{K=L,R} \frac{\sum_{j,MAP_j \in S_k} \left\{ \left[E\left| \frac{1}{3} - \left(\frac{N_j}{N_K} \right) \right| \right] N_K \cdot \lambda \right\}}{N_K} \qquad (20)$$

In the M3 model the neighbouring MAPs are grouped into two areas. But it is possible any other grouping as well. For this case equation 20 transformed into fully general form:

$$\sum_{K,\forall G} \frac{\sum_{j,MAP_j \in S_k} \left\{ \left[E\left| \frac{1}{n_G+1} - \left(\frac{N_j}{N_K} \right) \right| \right] N_K \cdot \lambda \right\}}{N_K} \qquad (21)$$

where G is the set of areas, and n_G is the number of areas in the model.

VII. SIMULATION RESULTS

In this section we compare the accuracy of our Markov movement models to other models found in the literature. The estimation procedure was validated by a simulation environment of a cell cluster shown in Figure 12.a.

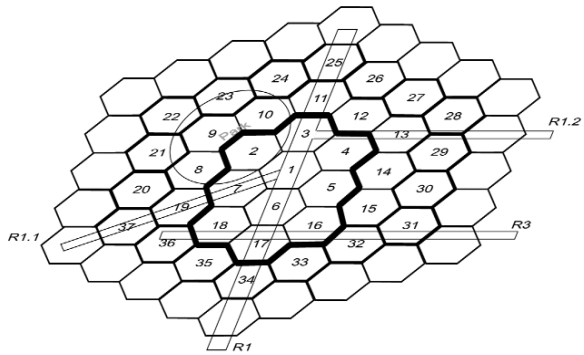


Fig. 12.a The logical cell-cluster in the simulation environment

The simulation was written in the open source OMNet++ using C++ language. The simulation environment consisted of a cluster with 61 named cells and it also included geographical data that is interpreted as streets and a park on the cluster area. The drift of the movement is heading to the streets from neutral areas.

The simulation used 610 mobile terminals (10 for each cell), in the initial state uniformly distributed in the cluster. The average motion velocity of the users is parameterized with a simple phase-type (PH) cell dwell time simulator (reciprocal of exponentially distributed values). In the simulation time mobile terminals appear and disappear, in order to simulate the active and inactive states.

The simulation consists of two parts. The trace simulation is the series of cell-transitions that the mobiles have initiated. It produces a time-trace that contains the actual location data for each mobile terminal in the network (reference interval). We have used this trace simulation as if it was a provider's real network trace.

The second part is the estimation procedure that uses the past and the current reference simulation results to estimate future number of users in each cell. The estimation error is interpreted as the measure of accuracy of each mobility model in this paper.

The prediction starts 100 timeslots after the reference simulation initiation. During the warm-up process the reference simulation produces enough sample data for the correct estimation, which uses the previous reference results as an input to estimate the future user distribution. Each user-transition in the 100-timeslot reference period is used to derive transition probabilities, motion speed and patterns in the simulation cell-space. These patterns serve as an input for the simulation threads of each mobility model. The models have the same input throughout the simulation process so that the results are comparable.

A widely used modified Random Walk estimation, *M3* and *M7* models were used in the simulation as references.

We used the MMCF parameter calculation algorithms, introduced above in Section V for the simulation environment. The smaller examined area contains the cells in the bold circle (cell 1-7, cell 16-18) in Figure 13. In MMCF generated optimal Markov movement model estimation compared to the fix *M3*, *M7* models. The input parameters of MMCF for this simulation environment are:

S={handover during voice call}, ε_c=5, ε_{uv}=0.4, ε_{op}=0.2,

We examined only the handover event, so the D matrix is empty and because of the limits of this paper the C and B_Q matrices are not presented. Structured, access point centralized model with one level was chosen. The result of the algorithms is:

$\underline{\underline{Q}}$ = [[2,1,1,1,2,1,1,2,1]] - which means for MAPs 1,6,17 the depth is 2, for the others it is 1,

$\underline{R}$ = [{1,2,3,{4,5},6,7,16,17,18}] - which means that MAP 4 and 5 is merged together.

The Markov-chain is (for clear interpretation not all of the edges depicted):

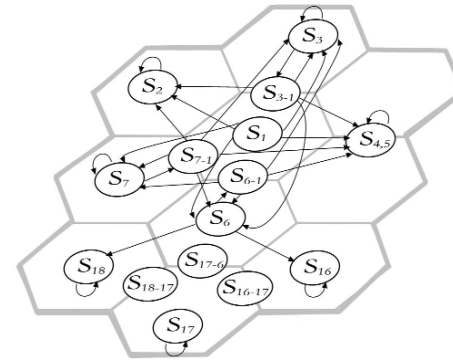


Fig. 13 The TAEV values in RW, ExtRW, M3 and M7 models with direction=(1,4)

The following plots (Figure 14.) show the average error of the estimations in every t timeslot. Random Walk model performed worst, it cannot follow the patterns in user fluctuation as it was expected. The *M3* and *M7* models work with significantly lower error rate, but in t=105,t=120 and t=135 timeslots the average error rate increased suddenly. This is caused by the change of distribution of the directional moving users (suddenly increased the number of active mobile users), what the simple Markov models cannot follow. The MMCF generated Markov approach holds the average error rate, it followed the changes in user motion appropriately, and it is able to learn the directional motion patterns and the fluctuation of user distribution, which proves the strength of the Markov Model Creator Framework.

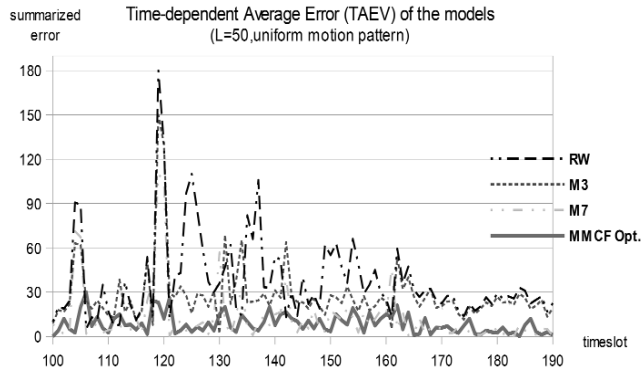


Fig. 14 The TAEV values in RW, M3, M7 and optimal MMCF model

CONCLUSIONS

In this paper we selected some significant parameters of mobility and proposed a method to model the mobile node and the network independently of the technology used. We proposed a simple classification for Markov mobility models, and we have shown examples for the most important types. We showed the attributes of a general Markov model, and we prepared processes for definition. Obviously these algorithms could be further refined.

By using the MMCF it is not necessary to create a new Markov model, only the description of the network, parameters and the requirement of the accuracy must be given and a Markov movement model is generated with minimal number of states. The network operator may use this Markov model to make predictions on the future distribution and location of users among radio cells. It is able to support self-configuring system in 4G mobile networks as well.

REFERENCES

[1] Harri Holma (Editor), Antti Toskala (Co-Editor), "WCDMA for UMTS: HSPA Evolution and LTE, 4th Edition", ISBN: 978-0-470-31933-8, Wiley and Sons, 2007.

[2] László Bokor, Zoltán Faigl, Sándor Imre, "Flat Architectures: Towards Scalable Future Internet Mobility", Lecture Notes in Computer Sciences; LNCS 6656, In: John Domingue, et al. (ed.) Future Internet Assembly 2011: Achievements and Technological Promises. 2011. pp. 35-50.

[3] F. Yu and V. C. M. Leung, "Mobility-based Predictive Call Admission Control and Bandwidth Reservation in Wireless Cellular Networks," Elsevier Computer Networks, vol. 38, no. 5, Apr. 2002, pp. 577-589.

[4] H. Debar, M. Dacier, and A. Wespi, "A Revised Taxonomy for Intrusion-Detection Systems", Annales des Telecommunications, vol. 55, 2000, pp. 361-378.

[5] B. Sun, F. Yu, K. Wu and VCM Leung, "Mobility-Based Anomaly. Detection in Cellular Mobile Networks," ACM WiSe'04, Philadelphia, PA,. 2004, pp. 61-69.

[6] M. Dottling and I. Viering, "Challenges in Mobile Network Operation: Towards Self-Optimizing Networks," Proc. IEEE Int'l. Conf. Acoustics, Speech, and Sig. Processing, Apr. 2009, pp. 3609–12.

[7] X. Hong, T. Kwon, M. Gerla, D. Gu and G. Pei, "A Mobility Framework for Ad Hoc Wireless Networks", In Proceedings of ACM 2nd International Conference on Mobile Data Management, (MDM 2001), Hong Kong, (2001, January).

[8] Bo Sun, Fei Yu, Kui Wu, Yang Xiao, Victor C. M. Leung, "Enhancing Security using Mobility-Based Anomaly Detection in Cellular Mobile Networks", Vehicular Technology, IEEE Transactions on (Volume:55 , Issue: 4) July 2006.

[9] Frédéric Lassabe, Damien Charlet, Philippe Canalda, Pascal Chatonnay, and François Spics, "Predictive Mobility Models based on Kth Markov Models" In IEEE Int. Conf. on Pervasive Services 2006 (ICPS'06), Lyon, France, pp 303-306, June 2006.

[10] Sándor Szabó, "The effects of user mobility on the performance of wireless networks", Ph.D dissertation, 2010.

[11] Hanzo, L.; Haas, H.; Imre, S.; O'Brien, D.; Rupp, M.; Gyongyosi, L.: "Wireless Myths, Realities, and Futures: From 3G/4G to Optical and Quantum Wireless", Proceedings of the IEEE, Volume: 100, Issue: Special Centennial Issue, pp. 1853-1888, May 2012.

[12] Wencho Ma, Yuguang Fang, Phone Lin, "Mobility Management Strategy Based on User Mobility Patterns in Wireless Networks", IEEE Transactions on Vehicular Technology, vol. 56. no. 1., pp 322-330, January 2007.

[13] P. Fülöp, T. Szalka, S. Szabó and S. Imre, "Accurate Mobility Modelling and Location Prediction based on Pattern Analysis of Handover Series in Mobile Networks", Mobile Information Systems (ISSN: 1574-017X), pp. 1-35, 2009.

[14] Shiang-Chun Lion, Queh-Min Huang, "Trajectory Predictions in Mobile Networks", International Journal of Information Technology, vol. 11 no. 11 2005.

[15] Wencho Ma, Yuguang Fang, Phone Lin, "Mobility Management Strategy Based on User Mobility Patterns in Wireless Networks", IEEE Transactions on Vehicular Technology, vol. 56. no. 1., pp 322-330, January 2007.

[16] P. Fülöp, B. Kovács, and S. Imre, "Mobility Management Algorithms for the Client-driven Mobility Frame System – Mobility from a Brand new point of view", Mobile Information Systems (ISSN: 1574-017X), pp. 313-337, 2009.

[17] P. Fülöp, B. Kovács and S. Imre, "Enhanced Mobility Management Modelling Framework", Proc. 6th Computer Information Systems and Industrial Management Applications (CISIM 2007), Elk, Poland, June 28-30, 2007. pp. 53-58.

[18] Fang, Y., & Lin, Y., Portable Movement Modelling for PCS Networks, , IEEE Transactions on Vehicular Technology, 1356-1362, 2000.

[19] Ahed Alshanyour and Anjali Agarwal, „Three-Dimensional Markov Chain Model for Performance Analysis of the IEEE 802.11 Distributed Coordination Function" Global Telecommunications Conference, GLOBECOM 2009.

[20] M. M. Zonoozi and P. Dassanayake, "User mobility modeling and characterization of mobility patterns", IEEE Journal on Selected Areas in Communications, Volume 15, No.7, September 1997.

[21] Y.-B. Lin, M. Chen, and H. Rao, "Potential Fraudulent Usage in Mobile Telecommunications Networks", IEEE Transactions on Mobile Computing, vol. 1, no. 2, 2002, pp. 123-131.

[22] P. Fazekas, S. Imre, M. Telek: "Modeling and Analysis of Broadband Cellular Networks with Multimedia Connections", Telecommunication Systems (Kluwer Academic Publishers) Vol. 19, No. 3-4, pp. 263-288, March-April 2002 .

[23] Shiang-Chun Lion, Queh-Min Huang, "Trajectory Predictions in Mobile Networks", International Journal of Information Technology, vol. 11 no. 11 2005.

Péter Fülöp received his M.Sc. degree in Technical Informatics from the Budapest University of Technology and Economics in 2005. He is a former PhD student on the faculty of Informatics on the Department of Telecommunication and fellow researcher in the Mobile Innovation Centre of BME. Currently he is working as a system test engineer at Ericsson Telecommunications Hungary at Research and Development Department. His research interests are IP mobility, interworking of heterogeneous mobile networks and movement modeling in cellular, mobile networks. The PhD thesis is going to be written on Complex Mobility Management Systems and Applications.

Sándor Szabó received the M.Sc. degree in Electronic Engineering from the Budapest University of Technology (BME) in 2000. Next he started his Ph. D. studies at BME. He received the Ph.D. degree in 2011. Currently he is carrying his teaching activities as an assistant professor at the Dept. of Telecommunications of BME. He leads research projects, and also a project leader at the Mobile Innovation Centre of BME. His research interests include mobile and wireless systems, mobility modeling, mobility management and the integration of heterogeneous wired and wireless networks.

Applying Opportunistic Spectrum Access in Mobile Cellular Networks

Ádám Horváth

***Abstract*—Nowadays, mobile service providers (SPs) have the exclusive access of certain spectrum for the mobile telecommunication services. This approach may lead to the inefficient use of the spectrum, which can be alleviated by opportunistic spectrum access.**

In this paper, we propose a cooperation scheme for mobile SPs, in which the SPs can use each other's unutilized frequency bands, when their own speech channels are occupied. Our simulation results show that significant gain in the blocking ratio and in the utilization can be achieved with the proposed cooperation scheme. Moreover, we calculate the average profit rate, and show that our scheme is beneficial for both cooperating party.

***Index Terms*—opportunistic spectrum access, mobile cellular networks, spectrum sharing policy;**

I. Introduction

PRESENTLY, mobile network operators have exclusive right for the use of given frequency bands, which were assigned to them by the governments on spectrum auctions. Each mobile operator is restricted to its dedicated channels when allocating an incoming call, and the requests for speech channels will be rejected if the dedicated speech channels are occupied. However, the exclusive frequency usage may lead to an inefficient use of the spectrum [1], [2], [3].

To handle the inefficiency, several researchers proposed spectrum sharing techniques such as spectrum renting or opportunistic spectrum access (OSA). Bernal-Mor et al. analyzed three different models for renting individual speech channels [4]. Many papers dealt with the technical aspects of OSA [1], [2], [5], [6] and some proposed spectrum auctions for achieving the dynamic spectrum access [7], [8]. Nevertheless, only few made an attempt to really model the operation of a network assuming cooperation between service providers [9], [10], [3].

Moreover, the authors in [9] and [10] tried to model the spectrum renting. However, they ignored that the current specification of mobile cellular networks does not allow the renting of a single speech channel [3]. Therefore, these models are not realistic in the current technical environment. In [3], the authors proposed a Markovian model for modeling spectrum renting and investigated some renting policies. They showed that their model using exponentially distributed holding times is a good approximation of the operation of a mobile cellular network, where the holding times follow the log-normal distribution [11]. In their model [3], only one service provider is present and the environment is modeled as a stochastic process, which makes their model mathematically tractable.

Manuscript submitted March 28, 2013, revised May 14, 2013.
Á. Horváth is with the University of West Hungary, Institute of Informatics and Economics, Sopron, H-9400, Hungary (email: horvath@inf.nyme.hu)

In this paper, we re-positioned the model of [3] as follows. First, we assume that two competing SPs are present, by which we can investigate the effects of the cooperation at both SPs at the same time. Second, we propose a new cooperation scheme using a spectrum sharing policy for mobile SPs, in which the SPs minimize the renting time by the opportunistic use of external frequency bands. Our simulation results show that the high level of cooperation (OSA) results a significant gain for the SPs. Namely, the blocking ratio decreases extremely and the utilization is higher. We also show that achieving a significant gain in the blocking ratio does not need the frequent use of external frequency bands. Moreover, we calculate the average profit rate (APR) to demonstrate that our cooperation scheme is financially advantageous for both SPs. Since this work focuses on the operation of the model, we do not deal with the technical details like signaling between the SPs.

The rest of the paper is organized as follows. In Section II, we describe a cooperation policy between the SPs in detail. In Section III, we describe the proposed operational practice, while we present some simulation results in Section IV. We conclude the paper in Section V.

II. The Proposed Spectrum Sharing Policy

At present, there are competitive service providers delivering mobile communication service in each country. The frequency bands which they use to serve their clients are owned by the local government. The mobile communication companies compete for the license of the frequency usage on spectrum auctions. On each frequency band, the SPs can allocate 8 full-rate or 16 half-rate speech channels, on which they can allocate the incoming calls.

Each service provider divides its network into disjoint cells and tries to cover the whole area of the given country. In most cases, the mobile cells of the different providers cover each other, so more than one SPs are available on a given geographical place. Presently, if all speech channels of one service provider are occupied new incoming calls are blocked while there may be available channels at another operator.

However, if two arbitrary SPs (SP_A and SP_B) would cooperate according to the scheme presented below, more clients could be served. To facilitate it, we propose the following cooperation scheme. The SPs continuously monitor each other's spectrum to obtain which frequency bands are unused (a frequency band is unused, if there is no allocated call on it). If all speech channels of SP_A are occupied, SP_A may opportunistically allocate the new incoming calls on SP_B's unused frequency band f. Due to the current technical constraints, the SPs cannot share a frequency band

for simultaneous use [3]. Therefore, while SP_A is using f, SP_B cannot use it, even if only one speech channel is occupied on f.

In our model, a SP allows to use its unused frequency bands if the other SP needs it (of course, a renting fee must be paid for the frequency renting). Moreover, SP_B uses as few frequency bands as possible to increase the probability of successful frequency renting for SP_A. To achieve this, SP_B consolidates its network every time when an ongoing call is terminated, i.e. re-allocates the rest of ongoing calls. If SP_A is already using at least one frequency band of SP_B, the consolidation aims to vacate the rented frequency bands first in order to minimize the renting fee. The consolidation process ensures that a SP uses always the "required number" of frequency bands simultaneously, i.e., at most one used frequency band exists in a given moment on which there is at least one free channel. Formally, if the number of channels per frequency band is c, the number of ongoing calls is n, the number of used frequency bands can be computed as $\lceil \frac{n}{c} \rceil$.

If SP_B's speech channels are also occupied, and needs the channels which are currently used by SP_A, SP_A must vacate the rented frequency band and terminate the ongoing calls allocated on that frequency band. These ongoing calls suffer the forced termination, since the consolidation rule ensures that all speech channels in SP_A's network are occupied. Therefore, there are two ways of blocking in OSA: the normal blocking (NB), when the home network is saturated and there is no other available frequency band; and the forced blocking (FB), when an ongoing call is broken due to frequency withdrawal. In case of forced blocking, a user has got the service for a given time, however, it is annoying for him to suffer the break of the connection.

From SP_B's point of view, the frequency renting results financial benefits for him, since the renting fee must be paid for it. When SP_B needs the rented frequency band, it can be withdrawn without any disadvantageous consequence. The only disadvantage for SP_B is that the frequency renting helps the concurrent provider, too. From SP_A's viewpoint, the frequency renting is an additional option to avoid the blocking of the incoming calls. Without renting, the incoming calls would be blocked, however, the forced termination problem still holds. The break of the connection is annoying for the users, but it is still possible that most users are served without forced termination and without knowing that their call was allocated on an external frequency band.

In our investigations, we assume that two SPs (SP_A and SP_B) are present, and we compare the cases when both SPs use OSA or they do not cooperate.

III. System Model

In this section, we describe our system model, in which SPs use OSA and follow the policy that was detailed in Section II.

A. Notations

Henceforth, we use some notations that we collected in Table I.

TABLE I
NOTATIONS

Notation	Description
N_x	The number of speech channels of $SP_x, x \in A, B$
n_x	The number of frequency bands of $SP_x, x \in A, B$
c	The number of channels per frequency band ($N_x = n_x \cdot c$) $x \in A, B$
b_x	The number of occupied channels at $SP_x, x \in A, B$ (including the occupied rented channels)
f_x	The number of own frequency bands used at $SP_x, x \in A, B$
g_x	The number of rented frequency bands by $SP_x, x \in A, B$

B. The operation of the model

Using the variables of Table I, we can unambiguously determine whether an incoming call at a given SP can be served or will be blocked. Moreover, we can determine if an incoming call at a given SP causes forced blocking at the other SP.

In our model, the system is controlled by two main events: a *call arrival* (CA), when a SP should serve an incoming call, and a *channel release* (CR), when a call is terminated by the user. The following flowcharts present how our model operates when these events occur (Fig. 1 and Fig. 2).

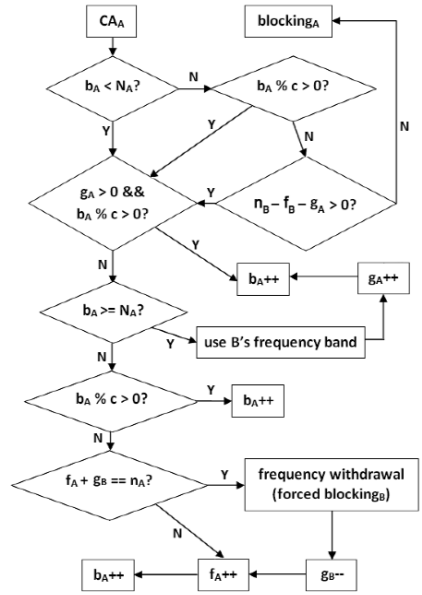


Fig. 1. System operation when a CA event occurs at SP_A

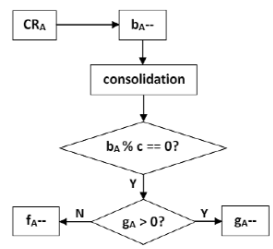


Fig. 2. System operation when a CR event occurs at SP_A

Fig. 1 represents the CA event. If an incoming call arrives e.g. at SP_A, there are three possible ways of serving it. First, if SP_A has got more own channels than the number of calls under service (even if SP_A must withdraw a channel from SP_B). Then, if the home network is saturated, but SP_A is currently using a frequency of SP_B and there is at least one free channel on that frequency band (namely, $b_A \% c > 0$). Finally, if the previous two conditions are not met, SP_A can use another frequency band of SP_B, if at least one of them is free. If none of the above mentioned conditions are fulfilled, the incoming call will be blocked. Note that without OSA, SP_A must have blocked the incoming call after the $first$ condition is unsatisfied.

If SP_A can serve the incoming call, a channel must be allocated for it, so it must be checked which condition of the above mentioned ones is met. First, if SP_A already uses at least one frequency band of SP_B ($g_A > 0$), and there is no need to use another frequency band of SP_B ($b_A \% c > 0$), SP_A serves the call on a frequency band of SP_B that SP_A has already been using (b_A++). If the first condition is not satisfied, SP_A checks whether the number of users under service is greater or equal to the number of its own channels ($b_A \geq N_A$). If so, SP_A must use another free frequency band of SP_B and serve the incoming call on that frequency band. If neither of the first two conditions are satisfied, SP_A will serve the incoming call on its own frequency band, on which there is no ongoing call of its own customers. However, SP_A must check if it must withdraw an own frequency band from SP_B to serve the incoming call ($f_A + g_B \stackrel{?}{=} N_A$). If so, SP_A will withdraw its own frequency band from SP_B (forced blocking at SP_B), else SP_A will use its own free frequency band.

Fig. 2 represents the CR event. If an ongoing call is terminated by the user e.g. in SP_A's system, SP_A's network must be consolidated to ensure that SP_A always uses only the minimum number of frequency bands on which its customers can be served. If one frequency band will be free after the consolidation ($b_A \% c = 0$), SP_A must check if that frequency band is SP_B's band ($g_A > 0$). If so, SP_A stops using SP_B's frequency band.

IV. Simulation Results

In this section, we present some numerical results obtained with the use of SimPack [12] that was proved as an effective event-driven simulation tool in many performance evaluation studies [3], [13], [14], [15].

A. Definitions

In the following scenarios, we present our results as a function of the normalized load $load_{norm}$, which can be computed as follows.

$$load_{norm} = \frac{\lambda_A}{N_A}\frac{1}{\mu_A}, \tag{1}$$

where $1/\lambda_A$ is the average time interval between the arrivals at SP_A, $1/\mu_A$ is the average holding time, and N_A is the number of SP_A's speech channels.

If we really want to demonstrate the financial benefits of our cooperation scheme, we must explicitly present financial parameters like APR, which can be computed as follows.

$$APR = \alpha \times \bar{S} - \beta \times (\bar{R} - \bar{L}), \tag{2}$$

where α and β are cost coefficients measured in cost units per time unit for the calls and the renting, respectively [3]; $\bar{S}$ is the total duration of the calls, $\bar{R}$ is the total duration of the renting, and $\bar{L}$ is the duration when the other SP was using the current SP's frequency bands. When we determine the renting fee, we must take into account that a SP will have c additional channels when renting a frequency band. In our model, we assume that at least one ongoing call is using the rented frequency band. On the other hand, the rented channel is usually not fully utilized, since there are expectedly less than 8 ongoing calls on it. Therefore, the renting fee should be less than the total income from a fully utilized frequency band (i.e. $\alpha \times c$). So, we define a discount factor and calculate the renting fee as follows.

$$\beta = \alpha \times \frac{c}{d}, \tag{3}$$

where $c > d \geq 1$ is the discount factor (note that $d = 1$ means no discount).

B. Simulation settings

In the case study, the channel holding times are log-normally distributed with the mean of 4.0 seconds and the standard deviation of 1.17 seconds. These values are taken from [11] and derived from observing real data traffic. This means that the expected channel holding time is 108.25 seconds in our model.

Moreover, Pattavina and Parini showed that the interarrival time distribution also follows the log-normal distribution [16]. Based on the parameter settings of [16], the mean interarrival time runs from 7.83 to 7.21 seconds to achieve a blocking ratio of 1-2% when the SPs did not cooperate. In the simulations, each SP has got 3 frequency bands in the investigated mobile cell, which means 24 full-rate speech channels for each SP if they do not cooperate.

Besides, we applied a policy, according to which the SPs do not pay any renting fee to each other in the cases when the given frequency band was withdrawn. This rule can be a compensation for breaking the ongoing connections (of course, $\bar{R}$ will be not increased with the time interval of the broken renting when computing the APR).

We investigated some properties at a given SP assuming the use of OSA, then without cooperation, as well. These properties are the blocking ratio; the average number of occupied channels; the ratio of external frequency bands used ($g_A/(f_A + g_A)$ or $g_B/(f_B + g_B)$); and the APR.

The simulation results are obtained with the confidence level of 99.9% and the relative precision (i.e. the ratio of the half-width of the confidence interval and the mean of collected observations) of 0.8%.

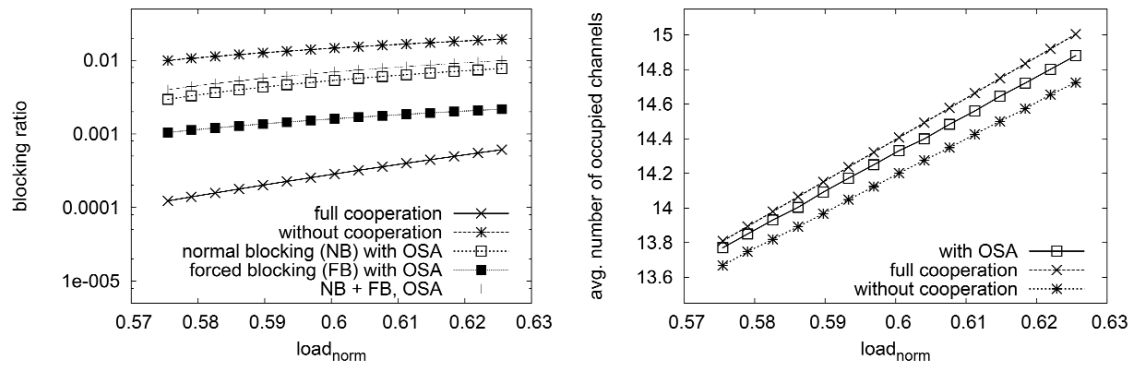


Fig. 3. Scenario 1 – the ratio and the average number of occupied channels

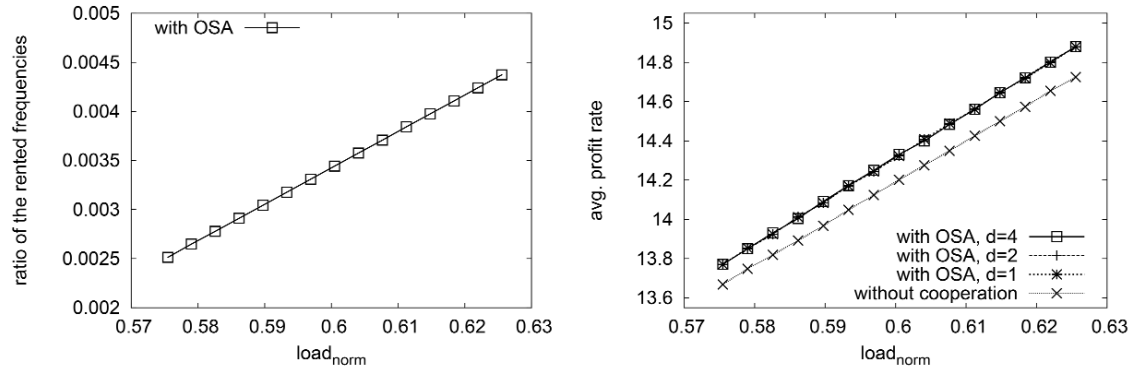


Fig. 4. Scenario 1 – the ratio of the rented frequency bands and the average profit rate

C. Scenario 1

In the first scenario, the normalized load is increasing at both SPs. Fig. 3 and Fig. 4 show the simulation results.

The left graph in Fig. 3 represents the blocking ratio of SP_A as a function of the normalized load. As Fig. 3 illustrates, the blocking ratio by using OSA is significantly lower than if the SPs would not cooperate (note that we use logarithmic scale when depicting the blocking ratio). In the left graph of Fig. 3, we also represent the blocking ratio assuming a full cooperation, which is equivalent with the fusion of the two SPs ensuring double capacity and double customer number. We also would get the full cooperation case if the SPs could rent single speech channels from each other. This graph also shows the forced blocking when using OSA, which is significantly lower than the normal blocking. Moreover, we represented the sum of the normal blocking and the forced blocking when we used OSA.

In the right graph of Fig. 3, the average number of occupied channels is shown. Using OSA, the average number of occupied channels is around 14.9 when the load is 0.625, while this value is only 14.7 without cooperation. Therefore, beside the decrease of the blocking ratio, the utilization increases using OSA.

The left graph of Fig. 4 depicts the ratio of external frequency bands used to the total number of frequency bands used by SP_A when OSA was applied. Even when the load is 0.625, this value is below 0.45%, which means that SP_A could serve its customers on its own frequency bands in most cases. When not, SP_A had the possibility of frequency renting, which resulted a notable gain in the blocking ratio, as we can see in the left graph of Fig. 3.

In the left graph of 4, the APR value is illustrated. When the SPs cooperate according to our scheme, a significant gain can be achieved in the APR. We investigated the APR with different values of d, however, the results show that there was no impact of d on the APR in this case. The explanation of this is that when the traffic is equal and equally changes at both SPs, the probability that they will use each other's channel is also equal. Henceforth, we will demonstrate the impact of d in Section IV-D.

D. Scenario 2

In the second scenario, we investigated the APR when the load of SP_A is increasing as in Scenario 1, while SP_B's load is constant with around 1% expected blocking ratio.

The graphs of 5 represent the APR values of SP_A (left side) and SP_B (right side). As the traffic increases, the APR of SP_A increases, as well. Of course, the higher the discount factor is, the higher APR can be realized for the renter. Even if the discount factor is one, the APR is higher than without cooperation, which can be explained as follows. First, SP_A also has income from the renting fee, since SP_B also

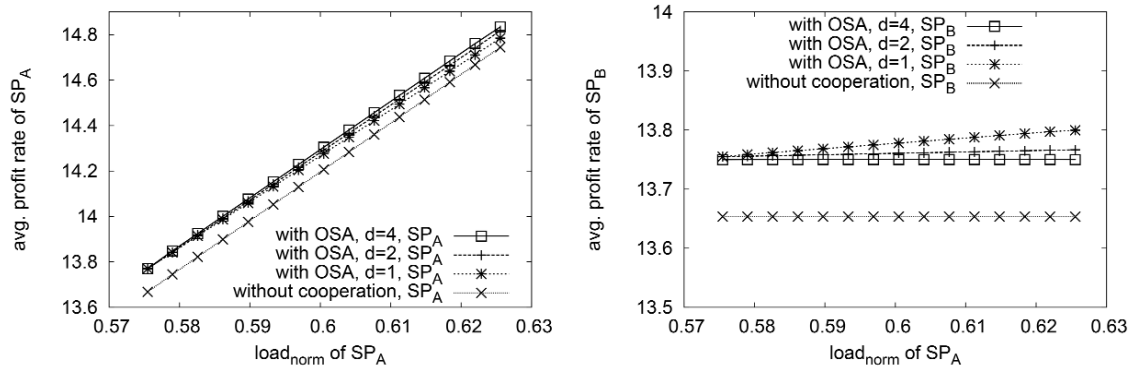


Fig. 5. Scenario 2 – the average profit rate of SP_A and SP_B

uses SP_A's frequency bands when necessary. Second, the compensation rule allows SP_A not to pay any renting fee if the given frequency band was withdrawn. On the other hand, SP_B can increase its profit only due to the increase of SP_A's traffic, when the SPs cooperate. Even if SP_B offers a discount of $d = 2$ or $d = 4$, the gain is significant.

V. Conclusions

In this paper, we presented a realistic model for investigating the effects of opportunistic spectrum access in mobile cellular networks. In our simulations, two competing SPs cooperate and allow each other to opportunistically use their spectrum. The results showed that both SPs have significant gain in the main system parameters like the blocking ratio and the utilization, while the forced blocking was marginal in each investigated case. Besides, our investigations showed that both SPs can increase their average profit rate when they cooperate: one from the renting fee, the other due to the increased capacity. Since our model involves a new signaling method, developing the signaling algorithm would be an interesting future work. Moreover, a work on a mathematical model is under progress.

Acknowledgement

The author thanks his supervisors: Prof. Tien Van Do, Department of Telecommunications, Budapest University of Technology and Economics and Prof. László Jereb, University of West Hungary, for the professional support and guidance during the preparation of this paper.

Ádám Horváth received the M.Sc. degree in Computer Science from the Budapest University of Technology and Economics (BME) in 2007. After his M.Sc. he joined the Institute of Informatics and Economics at the University of West Hungary as a Ph.D. student under the supervision of Prof. Károly Farkas. His research interests cover security of wireless networks and mathematical modeling, mainly focused on Markovian processes.

References

[1] J. M. Peha, "Sharing spectrum through spectrum policy reform and congitive radio," in *in Proceedings of the IEEE*, vol. 97, pp. 708–719, 2009.

[2] B. Jabbari, R. Pickholtz, and M. Norton, "Dynamic Spectrum Access and Management (Dynamic Spectrum Management)," *IEEE Wireless Communications*, vol. 17, no. 4, pp. 6–15, 2010.

[3] T. V. Do, N. H. Do, and R. Chakka, "A new queueing model for spectrum renting in mobile cellular networks," *Computer Communications*, vol. 35, no. 10, pp. 1165–1171, 2012.

[4] E. Bernal-Mor, V. Pla, and J. Martinez-Bauset, "Analysis of different channel sharing strategies in cognitive radio networks," in *Proceedings of the Third international conference on Multiple access communications*, MACOM'10, (Berlin, Heidelberg), pp. 70–73, Springer-Verlag, 2010.

[5] A. Ghasemi and E. Sousa, "Optimization of spectrum sensing for opportunistic spectrum access in cognitive radio networks," in *in Proceedings of the IEEE Consumer Communications and Networking Conference, Las Vegas, Nevada, USA*, pp. 1022–1026, 2007.

[6] S. Öner and F. Jondral, "On the Extraction of the Channel Allocation Information in Spectrum Pooling Systems," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, pp. 558–565, 2007.

[7] S. Gandhi, C. Buragohain, L. Cao, H. Zheng, and S. Suri, "Towards Real-Time Dynamic Spectrum Auctions," *Computer Networks*, vol. 52, no. 4, pp. 879–897, 2008.

[8] E. M. Noam, "Taking the Next Step Beyond Spectrum Auctions: Open Spectrum Access," *Telecommunications Policy*, vol. 21, pp. 461–475, 1997.

[9] S.-S. Tzeng and C.-W. Huang, "Threshold based call admission control for qos provisioning in cellular wireless networks with spectrum renting," in *Novel Algorithms and Techniques in Telecommunications and Networking* (T. Sobh, K. Elleithy, and A. Mahmood, eds.), pp. 17–22, Springer Netherlands, 2010.

[10] S.-S. Tzeng, "Call Admission Control Policies in Cellular Wireless Networks with Spectrum Renting," *Computer Communications*, vol. 32, pp. 1905–1913, 2009.

[11] C. Jedrzycki and V. C. M. Leung, "Probability Distribution of Channel Holding Time in Cellular Telephony Systems," in *in Proceedings of the IEEE VTC'96, Atlanta, Georgia, USA*, pp. 247–251, 1996.

[12] P. Fishwick, "SimPack: Getting Started With Simulation Programming In C and C++," in *in Proceedings of the Winter Simulation Conference*, pp. 154–162, 2002.

[13] T. V. Do, "An Efficient Solution to a Retrial Queue for the Performability Evaluation of DHCP," *Computers & OR*, vol. 37, no. 7, pp. 1191–1198, 2010.

[14] T. V. Do, "A new computational algorithm for retrial queues to cellular mobile systems with guard channels," *Computers & Industrial Engineering*, vol. 59, no. 4, pp. 865 – 872, 2010.

[15] T. V. Do, "Solution for a retrial queueing problem in cellular networks with the fractional guard channel policy," *Mathematical and Computer Modelling*, vol. 53, no. 11–12, pp. 2058–2065, 2011.

[16] A. Pattavina and A. Parini, "Modelling voice call inter-arrival and holding time distributions in mobile networks," *Performan. Challenges Efficient Next Generat. Networks, Proc. 19th Int. Teletraf. Congr*, pp. 729–738, 2005.

Video Services in Information Centric Networks: Technologies and Business Models

A. Difino, R. Chiariglione and G. Tropea

***Abstract*—Information Centric Networking (ICN) is an emerging communication paradigm in which information is stored and retrieved based on its "name". It can be seen as moving away from a host-based to a content-based communication model. The CONVERGENCE project has developed a comprehensive ICN solution integrating network and middleware, glued together by a unified, interoperable container called Versatile Digital Item (VDI). The paper describes the implementation called Video Centric Networking – Publish/Subscribe (VCN-PS) and shows that ICN can be realistically implemented and made accessible to applications in popular devices, and can thus be considered as a viable medium-term alternative to other multimedia delivery technologies. The paper also checks the ability of VCN-PS to support business models for video streaming services.**

***Index Terms*—Dynamic Adaptive Streaming Information Centric Networking, Internet Protocol, Middleware, Publish/Subscribe**

I. INTRODUCTION

CONSIDERING the premises – a network born for connecting computers and later to transport hypertext information using the HTTP protocol – Internet has been quite successful in responding to the expanding needs of its users, as demonstrated by the wide range of different formats used today in an ever growing number of multimedia data. Video data, e.g. MPEG standards like AVC [13] and, shortly, HEVC [28] take an increasingly lion's share of the global traffic and the amount of streamed bits is growing by the day. Among the different protocols used to transport data formats, of particular interest for the short term is Dynamic Adaptive Streaming over HTTP (DASH) [29] that uses HTTP to transport on demand segments of audio and video at different levels of quality – and hence of bitrate – to cope with unpredictable bandwidth variations from source to destination.

Additionally, content replication at appropriate network locations is an effective strategy to cope with the unpredictable mismatch between source and destination in terms of bandwidth availability over long distances. The relevant logic is currently implemented in such overlay networks as Peer-to-Peer (P2P) Networks [1] and Content Distribution Networks (CDN) [2]: P2P leverages self-organised, adaptive, and fault-tolerant distribution among peers by retrieving fragments of information from selected available peers, while CDNs redirect requests for web resources to appropriately located caches using mechanisms such as DNS-based redirection.

The paradigm shift represented by both solutions is to move away from a host-based to a content-based communication model: URIs and DNS names are interpreted in a way that allows accessing cached copies of information units in the network.

Information Centric Networking (ICN) is an emerging communication paradigm in which information is stored and retrieved based on its "name" or identifier, without relying on point-to-point (client-server) communication primitives. It can be seen as a move towards incorporation of the above issues into the network layer, with the promise to overcome some of their existing shortcomings. In ICN each information unit, when stored, is individually identified at an appropriate granularity level so that it can be retrieved by simply using the identifier, as opposed to retrieving from a logical location whose physical addresses are provided by the Domain Name System (DNS). In other words ICN does not require an additional level of indirection, from identifiers to logical or physical content locations, being able to directly "route-by-name" requests to the "closest" copy of the content.

Since the Xerox PARC CCN [3] and the NSF funded Named Data Networking (NDN) [4] projects, ICN has benefitted from European Commission funded Framework Program 7 R&D investments in different projects such as 4WARD [5], PSIRP/PURSUIT [6], SAIL [7], and COMET [8].

Section 2 lists the main requirements for generic ICN Publish/Subscribe. Section 3 details the main technology choices adopted by CONVERGENCE, an FP7-funded ICN project [9]. Section 4 describes the peculiarities of video pub/sub in ICN based on the comprehensive ICN solution based on the combination of CCN for the network part and a range of technologies specified by such MPEG standards as MPEG-7, MPEG-21 and MPEG-M for the middleware part. The middleware is glued together by a unified, interoperable container named Versatile Digital Item (VDI), a specialisation of the MPEG-21

Paper first submitted on 2013/03/08, revised 2013/04/18

This work was partially supported by the EU FP7-257123 CONVERGENCE project.

Angelo Difino was with CEDEO.net, Via Borgionera 103, I-10040 Villar Dora (TO), Italy. Currently he works as an independent consultant (email: angelo.difino@gmail.com).

Riccardo Chiariglione is with CEDEO.net (email:riccardo@cedeo.net).

Giuseppe Tropea is with CNIT, Italy, (email: giuseppe.tropea@cnit.it)

Digital Item [16]. This paper also provides novel contributions toward a user-friendly publish/subscribe of audio-visual content streamed using the DASH format natively embedded in the CONVERGENCE ICN, implemented as a running platform accessible from the web called Video Centric Networking – Publish/Subscribe (VCN-PS). VCN-PS proves that ICN features can be realistically implemented and made accessible to applications in such popular devices as the Mac PC, and can thus be considered as a viable medium-term alternative to CDNs for the delivery of multimedia content. Section 5 analyses business models for video services in ICN using the operational video platform WimTV [10] mapped to VCN-PS. Section 6 draws conclusions and identifies future work areas.

II. Requirements for Generic Information-Centric Publish/Subscribe

In this section we will make an analysis of requirements that a generic (content-agnostic) ICN PS system shall support, by using results published in the literature.

A. Network Layer Requirements

- **Caching** is the functionality whereby the ICN knows where to find the information element that has been requested. As a subscriber's request can be satisfied with data coming from any source holding a copy of the object, ICN requires that publisher- and subscriber-neutral caching be natively part of ICN. Objects can be replicated, distributed without owner control.
- **Security and privacy** is the functionality whereby the source may sign and protect an information element so that other entities, e.g. network elements and consumer devices, can verify its signature and hence its integrity and authenticity, because publishers and subscribers are decoupled in space and time, and the source can control access to it, because the network has visibility of the information elements.
- **Information transport** is the functionality providing reliable, and congestion- and flow-controlled transport of information elements from one or more sources to a receiver, including support for caching, multi-path and disruption tolerance.
- **Routing** is the functionality of
  - *Name-based forwarding*: forward on names, using a specific routing protocol;
  - *Name resolution*: resolve names to data objects, leveraging underlying forwarding and routing infrastructure;
  - *Congestion control and QoS*.

B. Middleware Layer Requirements

Publish/Subscribe is the functionality, dealing with descriptions (Publish) and queries (Subscribe) of information elements, whereby an ICN user can

- Request content that satisfies certain criteria. In a publication the following data should be provided: the issuer, the descriptions, the names of those to be notified when a match occurs etc.
- Make known that content with certain features has been stored on the ICN. In a subscription the following data should be provided: the issuer, the query, the names of those to be notified when a match occurs etc.

Please note that both subscribe-first/publish-later and publish-first/subscribe-later models have to be supported.

C. Joint Requirements

Content Identification is the functionality whereby each information element, down to the appropriate granularity level, must be identified uniquely and unambiguously. This involves two main sub-requirements:

- Information elements should be identified at the application level;
- The syntax of the information element identifier should support scalability since ICN should be able to cope with an extremely large number of information objects.

Therefore the naming/addressing architecture should be designed to allow an information element identification that is global, provides efficient lookup, is scalable to objects possibly numbering quadrillions, and is able to deal with node and information object mobility.

D. Meta-Requirements

The following requirements are of general "trans-layer" interest:

- The availability of metrics that make it possible to evaluate implementations in a consistent manner;
- The existence of business, legal and regulatory frameworks, including incentives and novel business models to engage relevant stakeholders;
- The ability to deploy ICN incrementally, e.g. to gradually migrate without obliterating existing IPv4/v6 infrastructure.

To fully satisfy the above (and broader) requirements, the CONVERGENCE project ended-up developing a reference layered ICN architecture where layers are operated independently and glued by the notion of VDI and its identification scheme. The outline is provided by *Figure 1* where three different kinds of ICN device, namely CoNet, a network node, CoSec, a security node, and Peer, or user device, are required.

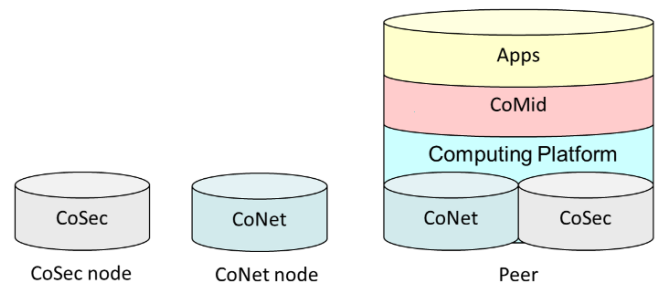


Figure 1: CONVERGENCE devices

CoNet node is a device running the ICN stack of functionalities only, including caching, routing-by-name and

securing of network data objects, while CoSec is a device responsible for handling the majority of cryptographic protocols and security related tasks.

Peer is a layered device whose layers are:

1. Computing Platform that includes computing resources and operating system interfaces able to run concurrently CoNet and CoSec implementations;
2. CoMid a middleware with API interfaces to the underlying ICN functionalities;
3. Apps that includes applications offering VCN-PS to humans or higher-level machine users.

Figure 2 depicts the architecture from a distributed point-of-view. The functionalities at the computing platform level of each device are collectively distributed across an Infrastructure level. Special devices that only run CoNet (core network routers) or only run CoSec (specialised security servers) are referred to as nodes. Devices that additionally run the CoMid are referred to as peers. The middleware is **distributed across all peers**: however, some of them can be more specialised in providing certain services, some of them (typically devices for end-user but also for service providers) will additionally run CONVERGENCE-compliant applications.

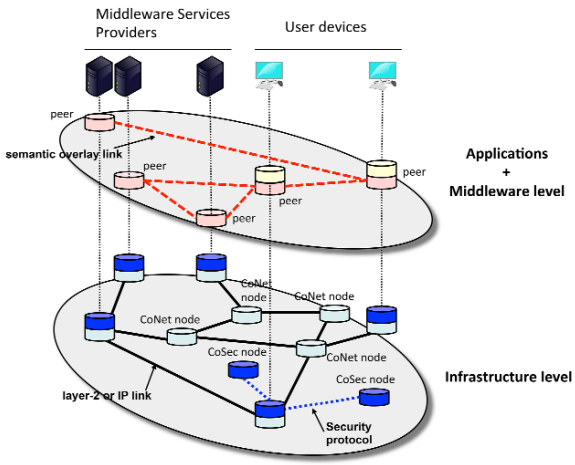


Figure 2: Distributed view of the CONVERGENCE architecture

CoMid peers exchange messages with peer entities of a semantic overlay. CoNet nodes are connected via physical or traditional IP links and communicate over the ICN. CoSec nodes talk using distributed, specialised security protocols.

III. TECHNOLOGIES AT THE HEART OF CONVERGENCE

A priori there is no obligation to rely on standards but the specification of a complete information-centric system is such a huge undertaking that it is preferable to be able to tick out some problem elements by relying on the technical consensus achieved by standards groups in selected technology areas, particularly when these are tightly connected with the special information type called media.

A suite of International Standards developed by MPEG, namely the Multimedia Service Platform Technologies (MPEG-M), provides technology support to CONVERGENCE architecture.

The CONVERGENCE extension of the MPEG-M architecture [23] represented in *Figure 3* includes:

- An intermediate Tool layer between the MXM Layer and the Application Layer. Tools include reusable code for applications;
- New southward APIs for ICN (CoNet) and Security (CoSec).

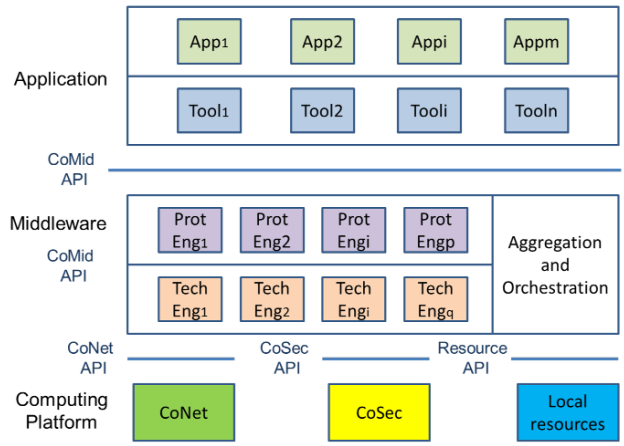


Figure 3: MPEG-M based CONVERGENCE ICN Architecture

When an Application calls the API defined in part 2 of MPEG-M [24], to access the MPEG Extensible Middleware (MXM), different possibilities exist:

1. The App calls just one local Technology Engines (TE). A TE is a module providing defined functionalities, such as a Media Framework to play a video. Part 2 of MPEG-M defines some high level APIs and provides placeholders to define new ones. Part 3 of MPEG-M [25] provides software implementations of a range of TEs released with a Berkeley Software Distribution (BSD) licence;
2. The App calls a chain of local TEs. This TE serialization is called "Technology Orchestration";
3. The App calls just one Protocol Engine (PE). A PE is an implementation of an Elementary Service (ES) such as Create Licence which in turn calls just one or a sequence of local or remote TEs. Part 4 of MPEG-M [26] defines a set of ESs and the corresponding PEs;
4. The App calls a chain of PEs. Part 5 of MPEG-M [27] defines a machine-readable representation of the PE workflow that represents the "Service Aggregation" implied by the sequence of PEs.

Figure 4 shows an Application calling the PEa→PEb→PEc Aggregated Service and where PEa calls just one TE, PEb calls 3 Orchestrated TEs and PEc calls 2 Orchestrated TEs. Typically a special TE called Orchestrator drives the sequence of TEs to accomplish the goal.

A. Resource Identification

In the following the term Resource indicates information primarily intended for human-consumption. Versatile Digital Item (VDI) is the CONVERGENCE-defined data structure that includes:

- VDI Identifier
- Resource Identifier
- Metadata describing Resources or query on Resources
- Licence expressing what rights are given to a User to act on Resources on what conditions
- Event Report Requests (ERR) to instruct a Peer to issue an Event Report (ER) to specific users/peers in the event certain actions (e.g. play or store) are performed on a resource etc.

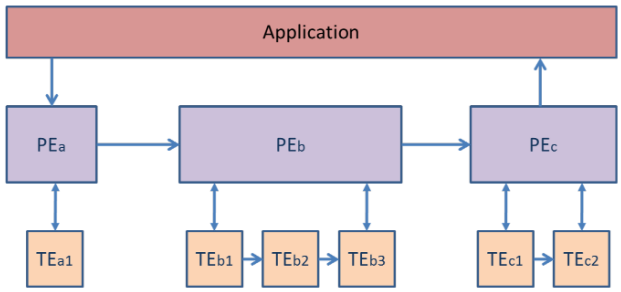


Figure 4: MPEG-M Aggregation and Orchestration

CONVERGENCE has further defined 3 types of VDI

1. R-VDI (Resource VDI) carrying resource (e.g. a video) identifier and other data
2. S-VDI (Subscription VDI) carrying the query and other data
3. P-VDI (Publication VDI) carrying the resource description and other data.

A VDI created by the CONVERGENCE Middleware receives a unique and persistent Identifier. Please note that it is useful to be able to place a semantic link between a VDI and one or more than one VDI, e.g. in case a VDI is "superseded" by a new version of it or a new independent VDI appears that has some connection with the existing VDI. CONVERGENCE Ontology Services lets users establish such links.

B. Resource Classification

The CONVERGENCE Core Ontology (CCO) allows for a semantic organisation of peers in a virtual overlay network of "topics" built on the basis of users' interests in the different types of published content [11]. Note that

1. *Figure 5*, an example of fractal organisation of topics, shows that a peer typically belongs to more than one topic, depending on which resources users are interested in;
2. All P-VDIs and S-VDIs related to a certain Metadata are stored in some Peers of that fractal;
3. In general more than one peer populates a topic, to accommodate the fact that some peers may be off when access to them is needed.

C. Resource Publication and Subscription

Peers are endowed with the embedded ability to

- Perform matches between P-VDIs and S-VDIs that are stored on them;
- Communicate any match found to the specified users/peers depending on licences and ERRs;
- Remove S-VDIs and P-VDIs whose validity date has expired from the tables from the topic.

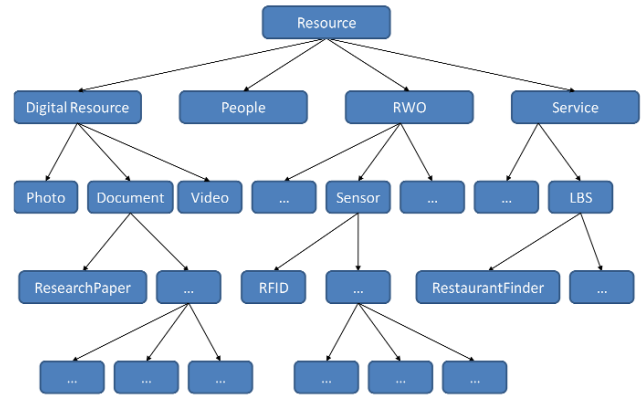


Figure 5: Topics

Figure 6 depicts how Publish/Subscribe works in CONVERGENCE.

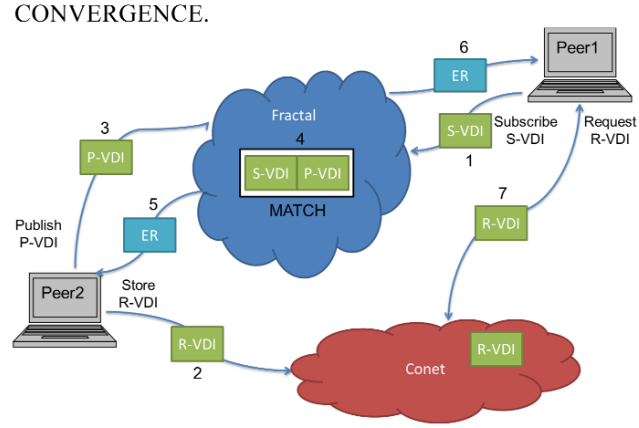


Figure 6: Publish/Subscribe in CONVERGENCE ICN

This is the sequence of steps:

1. Whenever a user (on Peer1) subscribes to receive a Resource with particular features, a description of its interest will be created as an S-VDI and will be injected in the Overlay (step 1);
2. When another user (on Peer2) wants inform a particular class of Users that a Resource has been created, an R-VDI will be created that includes Resource description and will be Stored on CoNet (step 2);
3. A description of the features of the media will be created as a P-VDI and injected in the same Overlay topic (step 3);
4. Behind the scene, a Peer dedicated to the particular class of Resource will eventually match the publication with the subscriptions previously injected (step 4);
5. If a match is detected, Peer2 and the Peer1 are notified back about the match (steps 5 and 6);
6. Peer1 User can now retrieve the R-VDI that includes the Resource previously stored on CoNet (step 7).

With reference to Figure 6 this is the complete walkthrough:

1. Peer1
   a. Creates S-VDI
   b. Subscribes S-VDI

2. Peer2
    a. Creates R-VDI
    b. Stores R-VDI
    c. Stores Resource
    d. Creates P-VDI
    e. Publishes P-VDI
3. PeerN (the peer hosting the relevant topic)
    a. Finds Match between Peer1's S-VDI and Peer2's P-VDI
    b. Notifies Peer1 and Peer2
4. Peer1
    a. Retrieves R-VDI
    b. Retrieves Resource

IV. PECULIARITIES OF VIDEO PUB/SUB IN ICN

This section focuses on the peculiarities of the Middleware and Application layers for video streaming in a DASH-based environment. The focus is on a complete implementation of a specialised ICN for handling and distributing video content in a consumer-friendly manner. The last section will discuss the described VCN-PS can be used to support the WimTV video service.

The VCN-PS solution has been designed to enable a user of an ICN application to perform the functions listed in Table 1, which represent a subset of the functionality requirements capturing the fundamental need of a Publish/Subscribe ICN application.

Please note that in the following

1. Words beginning with a capital letter will refer to terms with defined semantics;
2. The term data structure will refer to the specifically defined XML structure conveying information related to a Resource.

Table 1: VCN-PS requirements

1	Create a data structure containing
1.a	Identifier of the data structure
1.b	Identifier of the Multimedia Resource it refers to
1.c	Metadata describing the Multimedia Resource it refers to
1.d	Rights available to a User and conditions to be satisfied for using the Multimedia Resource
1.e	Requests to report Resource-related Events
2	Encrypt parts of the data structure
3	Sign the data structure
4	Verify the authenticity of a data structure or Multimedia Resource
5	Establish links of a data structure with other data structures (e.g. a previous version)
6	Optionally make a package of the data structure and the Multimedia Resource it refers to
7	Subscribe to a set of Multimedia Resource attributes by providing a query and a request to report to identified Users any match of the Subscription with a Publication
8	Publish information about a Resource Stored on the ICN by providing appropriate Metadata and a request to
	report to identified Users any match of the Publication with a Subscription
9	Interface to Store (put) and Retrieve (get) data from ICN indicating the data Identifier

These requirements are largely met by two suites of International Standards developed by MPEG group and listed in Table 2, namely: Multimedia Content Description Interface (MPEG-7) and Multimedia Framework (MPEG-21).

Table 2: MPEG-7 and MPEG-21 standards for VCN-PS

Req.	Title	Std	Acronym	Ref.
1	Digital Item Declaration	21	DID	[16]
1.a	Digital Item Identification	21	DII	[17]
1.b	Digital Item Identification	21	DII	[17]
1.c	Simple Metadata Profile	7	SMP	[14]
1.d	Rights Expression Language	21	REL	[20]
1.e	Event Report	21	ER	[22]
2	Intellectual Property Management and Protection Components	21	IPMP	[19]
5	Related identifier types	21	DII	[18]
6	MPEG-21 File Format	21	MP21FF	[21]
7	MPEG Query Format	7	MPQF	[15]

A. Implementation

The scenario in *Figure 6* has been implemented as depicted in *Figure 7*.

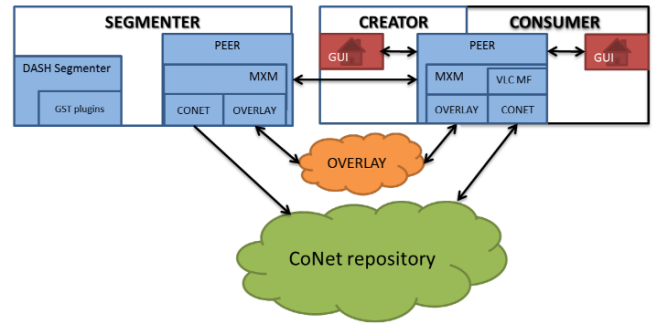


Figure 7: An implementation of VCN-PS

In *Figure 7* two Peers are depicted:

1. A Peer with the Creator and Consumer functionalities. Both share CoMid with 3 Engines highlighted because of their relevance to this paper: CoNet TE to Store/Retrieve data from ICN, Overlay TE to subscribe S-VDIs and publish P-VDIs and VideoLAN (VLC) [30] based Media Framework TE to play DASH-streamed videos from CoNet;
2. A Peer with the functionality of DASH segmenter. This has a CoMid instance that does not include the Media Framework TE and has additionally an internal engine (DASH Segmenter) that chops a video into DASH segments.

The walkthrough described in this section covers the same scenario presented above, adding more details about the video streaming capabilities of the system. Resource retrieval the is achieved using a similar approach as in [12]. Since video is delivered in DASH format, it is simply identified by an HTTP URI inserted in the R-VDI (*Figure 8*). The DASH segments and Media Presentation Descrip-tion (MPD) are stored in

CoNet. The HTTP URI contains a dummy common prefix followed by the real name of the video to enable the consumer to forward the HTTP request to Conet (*Figure 9*).

```
<mpegm-didl:DIDL>
…
  <mpegm-didl:Component>
    <mpegm-didl:Resource mimeType="video/mp4"
      ref="http://www.ict-
convergence.eu:9080/convergence/FerraraBalloonFestiv
al_1350451738512.mpd"/>
  </mpegm-didl:Component>
…
</mpegm-didl:DIDL>
```

Figure 8: R-VDI fragment with MPD URI

```
<MPD
  xmlns:xsi="http://www.w3.org/2001/XMLSchema"
  xmlns="urn:mpeg:DASH:schema:MPD:2011"
  xsi:schemaLocation="urn:mpeg:DASH:schema:MPD:2011"
  profiles="urn:mpeg:dash:profile:isoff-main:2011"
  type="static"
  minBufferTime="PT3.0S">

  <BaseURL>http://www.ict-
convergence.eu:9080/convergence/</BaseURL>
  <Period start="PT0S">
    <AdaptationSet bitstreamSwitching="true">
      <Representation codecs="avc1"
        mimeType="video/mp4"
        width="1024" height="576" startWithSAP="1"
        bandwidth="438272" minBufferTime="3000">

        <SegmentBase>
          <Initialization
sourceURL="FerraraBalloonFestival_1350451738512_4382
72.mp4"/>
        </SegmentBase>

        <SegmentList duration="3">
…
</MPD>
```

Figure 9: MPD fragment containing the dummy common prefix

The DASH segmentation related operations are covered by an ad-hoc remote service invoked by Creator to chop the video in DASH segments and to store the segments and its MPD in CoNet. As soon as this set of operations is concluded, Creator proceeds as usual: store the R-VDI (with the MPD URI inside) on CoNet and inject P-VDI in Overlay.

As highlighted above, whenever a match occurs, the consumer peer receives an ER that includes the R-VDI identifier. Consumer can now retrieve that R-VDI from CoNet using the Resource URI inside the ER. Therefore the MF engine can use the MPD URI to render the video, with the help of an HTTP proxy that forwards the HTTP request to ICN get operation. When the MPD and/or its segments are retrieved, they are forwarded back to the MF engine that presents the information.

This is the detailed walkthrough:

1. Consumer
    a. Creates S-VDI
    b. Subscribes S-VDI to Overlay
2. Creator sends video to Segmenter Service
3. Segmenter Service
    a. Segments video
    b. Stores video segments and MPD to CoNet
4. Creator
    a. Polls Segmenter until it is notified that video segments and MPD have been stored to CoNet
    b. Creates R-VDI
    c. Stores R-VDI
    d. Creates P-VDI
    e. Publishes P-VDI
5. PeerN (hosting the relevant topic)
    a. Detects match between S-VDI and P-VDI
    b. Notifies Consumer
6. Consumer
    a. Retrieves R-VDI
    b. Extracts the MPD URI
    c. Via MXM MF VLC Engine, request the MPD
    d. HTTP PROXY intercepts the request and forwards it to ICN
    e. The MPD is returned to MXM MF VLC Engine, which starts to request DASH segments
    f. HTTP PROXY intercepts the requests and forwards them to ICN

Currently VCN-PS includes the MXM engines listed in the following *Table 3*.

Table 3: MXM engines in VCN-PS

Engine name	Function	Engine type
convergence-conetTE	CoNet	TE
convergence-eventReportTE	Event Report	TE
convergence-matchTE	Match	TE
convergence-metadataTE	Metadata	TE
convergence-overlayTE	Overlay	TE
convergence-vdiTE	VDI	TE
mxm-digitalitemTE	Digital Item	TE
mxm-mediaframeworkTE	Mediaframework	TE
mxm-metadataTE	Metadata	TE
mxm-createcontentPE	Create Content	PE
mxm-identifycontentPE	Identify Content	PE

V. BUSINESS MODELS FOR VIDEO SERVICES IN ICN

In this section we explore the suitability of ICN to support the video streaming business by taking as WimTV, a platform supporting the activity of a plurality of interacting businesses, a case study, The platform is conceptually represented in *Figure 10.*

In WimTV a User can

1. Import Resources and relevant Metadata (including Rights and payment terms) to his digital locker (MyMedia)
2. Export Resources out of MyMedia (if export rights are available)

3. Post videos to his WebTV (MyStreams) for End Users to stream
4. Post videos to the marketplace (MyShop) for Buyers to assess and possibly acquire
5. Issue calls where User declares his availability to acquire specific content at specific conditions (MyCalls) that he can then, depending on Licence terms, post on MyShop or MyStreams, or export
6. Upload content with the specific value of "advertisement"
7. Get rights to content from another User's MyShop and post it on his Web TV.

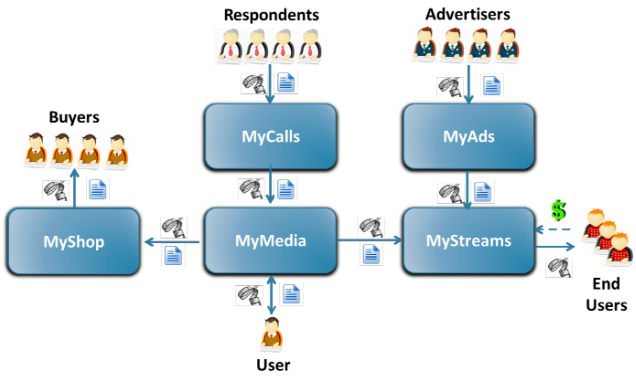


Figure 10: The WimTV model

All videos are stored on the WimTV server and can be referenced (e.g. streamed). Every time a User gets rights to a content item this is recorded by the WimTV server application.

In today's internet, video services are regularly based on a large servers managed by service providers. End users access the server, make their selection and the selected video is streamed to them.

VCN-PS offers new ways to look at networked video services but their precise mapping is a matter for discussion. How this can be done in ICN is the purpose of an analysis that is based on the following assumptions:

1. Videos are stored in ICN as clear-text or encrypted form;
2. R-VDIs are stored in ICN along with video Identifier, Metadata, Licence describing the rights granted to the principal (i.e. the licensee) and ERRs containing the names of the entities to be notified when video is played;
3. P-VDIs are posted on the overlay with search-friendly Metadata, Licence describing the match rights granted to the overlay and ERRs containing the names of the entities to be notified when match is found;
4. R-VDIs are injected on the overlay with Queries, Licence describing the match rights granted to the overlay and ERRs containing the names of the entities to be notified when match is found.

Table 4 shows how an operation (of the Function column) of a WimTV User (in the User type column) can be mapped to ICN steps (the Action column).

Function	User type	Action
Import to MyMedia	GU	Create and Store R-VDI
Export from MyMedia	GU	Create and Store R-VDI
Post to MyStreams	SP	Create and Store P-VDI
Search MyStream	GU	Create and Store S-VDI
View MyStream	GU	Get R-VDI
Post to MyShop	CP	Create and Store P-VDI
Search MyShop	SP	Create and Store S-VDI
Get Resource	SP	Get R-VDI
Post to MyCall	SP	Create and Store S-VDI
Respond to MyCall	CP	Create and Store P-VDI
Provide Ads	AD	Create and Store P-VDI
Look for Ads	SP	Create and Store S-VDI
Get Ad	SP	Get R-VDI

Table 4: MPEG-7 and MPEG-21 standards for ICN

Legend: GU=Generic User CP=Content Provider SP=Service Provider AD=Advertiser

NB1 "Create and Store R-VDI" may be executed separately on the VDI and the Resource or on a file that includes both VDI and the Resource. Obviously P-VDI and S-VDI do not involve Resources.

NB2 The Middleware provides support to a wide range of business relationships, e.g.
   1. Metadata to describe videos
   2. Licence to describe the rights to videos, but also what the Overlay can actually do
   3. Event Report Requests to notify which Users

NB3 The VDI can be extended to convey additional information regarding payments, a key enabler of WimTV value chains.

It is clear that a blind use of ICN features would probably make business differentiation rather difficult. Fortunately some options exist to achieve differentiation, e.g.

1. **Use of two-layered Metadata**. The first is common to all businesses and the second is proprietary. This would make matching between P-VDI and S-VDI work well only for a particular Service Provider
2. **Use of Licences** to provide clear benefits to only to subscribers to a particular Service Provider. Of course this requires that Peers have a sufficiently high level of security
3. **Use of proprietary ontologies**. A Service Provider bases its offer on proprietary ontologies that allow effective tagging and searches
4. **Use of VDI creation services**. A service provider offers a VDI creation service. The structure of the VDIs is such that R-VDI licences allow video retrieval only by principals recognised by the SP.
5. **Use of private overlay**. The Service Provider manages a private overlay that performs matches between Publishers and Subscribers in case Peers do not offer sufficiently high security.

VI. CONCLUSIONS AND FUTURE WORK

Introduction of ICN offers clear benefits to network operators because they are able to recover certain functions (P2P and CDN) that are currently implemented as overlay by other operators. The current competitive environment and the abundance of network infrastructure may also offer P2P and CDN operators the opportunity to become ICN operators themselves.

This paper has presented the standards-based CONVERGENCE Middleware that provides a rich set of functionalities for application developers and a complete implementation of the key video-oriented ICN Publish/Subscribe functionality called VCN-PS. The MXM Middleware is available as BSD OSS from http://mxm.wg11.sc29.org/download/check-out-the-source-code/java/.

A summary analysis of how VCN-PS can be applied to the existing WimTV video platform has shown that exploitation of VCN-PS is possible and that there is ample room to specific optimisations for the provisioning of meaningful personalisation of video service business.

The next steps in this research will include 1) making VCN-PS robust, 2) adding new functionalities, 3) deploying an experimental ICN, 4) implementing parts of the WimTV platform for actual operation on ICN and 5) furthering the business model analysis.

VII. ACKNOWLEDFGEMENTS

The Authors would specially like to acknowledge the work carried out by Andrea Detti and Salvatore Signorello (CNIT), Fernando Almeida (INESC), Andreas Khols (Morpho), Angelos-Christos Anadiotis and Aziz Mousas (NTUA), Bogdan Ardeleanu (UTI) from the CONVERGENCE project, and Jonathan Florido (DMAG-UPC) for their contribution to the development of MXM Engines.

VIII. REFERENCES

[1] Jiangchuan Liu; Rao, S.G.; Bo Li; Hui Zhang; "Opportunities and Challenges of Peer-to-Peer Internet Video Broadcast," Proceedings of the IEEE, vol.96, no.1, pp.11-24, Jan. 2008

[2] CDN Business Models - Wetzel Consulting, LLC, www.wetzelconsultingllc.com/CDNArticle.pdf

[3] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. Briggs, R. Braynard, ”Networking named content”, ACM CoNEXT 2009

[4] NDN, http://named-data.net/

[5] P. A. Aranda et al. Final Architectural Framework, June 2010, http://www.4ward-project.eu/

[6] D. Trossen et al. Conceptual Architecture: Principles, Patterns and Sub-components Descriptions, May 2011, http://www.fp7-pursuit.eu/PursuitWeb/

[7] Scalable and Adaptive Internet Solutions (SAIL), http://www.sail-project.eu/

[8] COntent Mediator architecture for content-aware nETworks (COMET), http://www.comet-project.org/

[9] The CONVERGENCE FP7 project: http://www.ict-convergence.eu/

[10] WimTV, http://wim.tv/

[11] Lioudakis, Georgies V.; Anadiotis, Angelos-Christos G.; Mousas, Aziz S.; Patrikakis, Charalampos Z.; Kaklamani, Dimitra I.; Venieris, Iakovos S., Routing in Content-Centric Networks: From Names to Concepts, 5th International Conference on New Technologies, Mobility and Security (NTMS), 2012

[12] Offloading cellular networks with Information-Centric Networking: the case of video streaming; A. Detti, M. Pomposini, N. Blefari-Melazzi, S. Salsano, A. Bragagnini, IEEE International Symposium on a World of Wireless Mobile and Multimedia Networks (WoWMoM 2012), S. Francisco, California, June 25-28, 2012

[13] ISO/IEC 14496-10:2012 Information technology – Coding of audio-visual objects – Part 10: Advanced Video Coding

[14] ISO/IEC 15938-9:2005 Information technology – Multimedia content description interface – Part 9: Profiles and levels

[15] ISO/IEC 15938-12:2012 Information technology – Multimedia content description interface – Part 12: Query format

[16] ISO/IEC 21000-2:2005 Information technology – Multimedia framework (MPEG-21) – Part 2: Digital Item Declaration

[17] ISO/IEC 21000-3:2003 Information technology – Multimedia framework (MPEG-21) – Part 3: Digital Item Identification

[18] ISO/IEC 21000-3:2003/Amd 2 Digital item semantic relationships

[19] ISO/IEC 21000-4:2006 Information technology – Multimedia framework (MPEG-21) – Part 4: Intellectual Property Management and Protection Components

[20] ISO/IEC 21000-5:2004 Information technology – Multimedia framework (MPEG-21) – Part 5: Rights Expression Language

[21] ISO/IEC 21000-9:2005 Information technology – Multimedia framework (MPEG-21) – Part 9: File Format

[22] ISO/IEC 21000-15:2006 Information technology – Multimedia framework (MPEG-21) – Part 15: Event Reporting

[23] ISO/IEC 23006-1:2013 Information technology – Multimedia Service Platform Technologies – Part 1: Architecture

[24] ISO/IEC 23006-2:2013 Information technology – Multimedia Service Platform Technologies – Part 2: MPEG Extensible Middleware API

[25] ISO/IEC 23006-3:2013 Information technology – Multimedia Service Platform Technologies – Part 3: Reference software and conformance

[26] ISO/IEC 23006-4:2013 Information technology – Multimedia Service Platform Technologies – Part 4: Elementary Services

[27] ISO/IEC 23006-5:2013 Information technology – Multimedia Service Platform Technologies – Part 5: Service Aggregation

[28] ISO/IEC 23008-2:2013 Information technology – High Efficiency Coding and Media Delivery in Heterogeneous Environments – Part 2: High Efficiency Video Coding

[29] ISO/IEC 23009-1:2012 Information technology – Dynamic adaptive streaming over HTTP (DASH) – Part 1: Media presentation description and segment formats

[30] VideoLAN, http://www.videolan.org/vlc/index.html

IX. BIOGRAPHIES

Angelo Difino
Angelo received his Computer Science degree from Università di Torino, Italy, in December 2000 and in 2005 he also got the degree in Computer Science Engineering from Politecnico di Torino. He started to work in the research lab of the telecommunication Italian company, Telecom Italia Lab, onpersonalization and user profiling activities. He contributed to several international standards: MPEG, SIP and FIPA. He spent last 6 years working on a startup and on a consulting agency about Multimedia technologies and Metadata descriptions, having the lead on the CEDEO.net contributions on Italian and European research projects.

Riccardo Chiariglione
Riccardo graduated in Science of Communication from Turin University. He worked on his final year project at Hill&Knowlton, London, UK, and sDae, Madrid, Spain (interships), where he started to investigate the Digital Media domain. Since June 2005 Riccardo has been working at CEDEO.net as Digital Media Analyst, developing a series of studies and analysis in the field of Digital Media. Riccardo focuses on the evolution of Digital Media environment and is currently involved in the User-Generated-Content and Social Networking areas.

Giuseppe Tropea
Giuseppe received his Laurea Degree in Informatics Engineering (spec. Computer Systems) from University of Catania, Italy, in October 2002. Since 2006 he is with Consorzio Nazionale Inter-universitario per le Telecomunicazioni (CNIT), inside the DISCREET, MOMENT and CONVERGENCE European projects, dealing with privacy protection in digital services and obfuscation of sensible IP traffic data by means of semantic, ontology-bases approaches. He is active in the MOI (Monitoring Ontology for IP traffic) ETSI Industry Specification Group and in MPEG-21 and MPEG-M. His research interests include routing protocols, neural networks, software modeling, semantic ontologies and natural language parsers, as well as data mining and GIS applications.

CALL FOR PAPERS
Special Issue on Autonomic Communications

The explosion in the size of communication networks, including the Internet, and their increasing diversity, and the ever increasing burden imposed on communication networks by pervasive computing and the Internet of Things, means that networks that are statically organized are unable to cope with ever changing conditions. The speed at which conditions change are also making it very difficult, and extremely expensive, to rely on human intervention to provide the adaptation that is constantly needed. Autonomic Communications, and the underlying Autonomic Computing, with their requirements for system Self-Awareness, Learning and Adaptation, Self-Healing, and Adaptive QoS, offer a way forward to address this dilemma. Thus since the early 2000's significant work has been carried out both in research environments and in industry to define the principles and main component technologies for Autonomic Communications. Worldwide, numerous research projects have been conducted in this area, including in a series of successful EU FP6 and FP7 projects.

Thus our journal is calling for original and unpublished contributions to this important area that will be refereed. Selected papers will appear in a Special Issue to be published in September of 2013. Original and unpublished papers should be submitted by 1st of May, 2013 in the form of pdf files in IEEE format according to the formatting instructions available at *http://www.ieee.org/publications_standards/publications/authors/authors_journals.html#sect2.*

Submissions should be sent in the form of an email attachment to the Special Issue Editor Prof Erol Gelenbe at the following address: *e.gelenbe@imperial.ac.uk.* Authors will be informed of acceptance or rejection before the 1st of June 2013, and papers in final form will be due on 1st July 2013. Papers rejected for the Special Issue, may later be revised and resubmitted for consideration as regular papers in the Infocommunications Journal.

Guest Editor:

EROL GELENBE is the Professor in the Dennis Gabor Chair at Imperial College, London, and has actively developed the field of Autonomic Communications. His other areas of interest include neuronal networks and bioinformatics, performance engineering, energy modeling for computer systems and networks, and network security. He currently leads the EU FP7 Project NEMESYS on mobile network security. He participates in the ECROPS project on energy savings in computing and communications, and in the European Institute of Technology Project on Smart Networks at the Edge. A Fellow of IEEE and ACM, he was elected to Foreign Membership of the Hungarian Academy of Sciences, and is a member of the French National Academy of Engineering and of the Science Academy of Turkey.

CALL FOR PAPERS
Special Issue on Cryptology

This special issue will focus on the area of cryptology and will include selected papers from the 2013 Central European Conference on Cryptology. CECC 2013 covers various aspects of cryptology, including but not limited to:

- cryptanalysis,
- design of cryptographic systems,
- general cryptographic protocols,
- pseudorandomness,
- cryptographic applications in information security,
- encryption schemes,
- post-quantum cryptography,
- signature schemes and steganography.

Detailed information on submissions to CECC 2013 and other information is provided at *http:// ww.fi.muni.cz/cecc/*, with April 15, 2013 being the deadline for the submission of abstracts that will be reviewed by the program committee and authors will be informed about acceptance or rejection by April 29, 2013. The conference registration deadline will be May 22, 2013, and the conference dates are June 26-28.

Submissions and presentations at the conference will be evaluated by the program committee and authors will be informed about the evaluation results no later than June 30.

No more than 5 papers from the workshop shall be selected for the special issue of the Infocommunications Journal, and authors of these papers will have the opportunity to revise their papers (including typesetting in the IEEE format) after the conference - final versions for the special issue will be due July 22, 2013.

Papers from the general public are also welcome. The deadline for submission of such papers is May 15. Papers will be peer-reviewed as usual.

Guest Editors:

VÁCLAV (VASHEK) MATYÁS is a Professor at the Masaryk University, Brno (CZ), and serves as a Vice-Dean for Foreign Affairs and External Relations, Faculty of Informatics. His research interests relate to applied cryptography and security, publishing over a hundred peer-reviewed papers and articles, and co-authoring six books. He was a Fulbright Visiting Scholar with Harvard University, Center for Research on Computation and Society, and also worked with Microsoft Research Cambridge, University College Dublin, Ubilab at UBS AG, and was a Royal Society Postdoctoral Fellow with the Cambridge University Computer Lab. Vashek was one of the Editors-in-Chief of the Identity in the Information Society journal, and he also edited the Computer and Communications Security Reviews, and worked on the development of Common Criteria and with ISO/IEC JTC1 SC27. Vashek is a member of the Editorial Board of the Infocommunications Journal. He received his PhD degree from Masaryk University, Brno and can be contacted at *matyas@fi.muni.cz*.

ZDENEK RÍHA is an Assistant Professor at the Masaryk University, Faculty of Informatics, in Brno (CZ). He received his PhD degree from the Faculty of Informatics, Masaryk University. In 1999 he spent 6 months on an internship at Ubilab, the research lab of the bank UBS, focusing on security and usability aspects of biometric authentication systems. Between 2005 and 2008 he was seconded as a Detached National Expert to the European Commission's Joint Research Centre in Italy, where he worked on various projects related to privacy protection and electronic passports. He was involved in the ePassport interoperability group known as the Brussels Interoperability Group. Zdenek has been working with the WG 5 (Identity management and privacy technologies) of ISO/IEC JTC 1/SC 27. Zdenek's research interests include smartcard security, PKI, security of biometric systems and machine readable travel documents. Zdenek can be contacted at *zriha@fi.muni.cz*.

MAREK KUMPOST is a Research Assistant at the Masaryk University, Faculty of Informatics, in Brno (CZ). He received his PhD in 2009 from the Faculty of Informatics, Masaryk University. The primary area of his doctoral research was oriented on privacy protection, anonymity and user profiling. He was involved in two European-wide project on privacy protection and identity management – FIDIS (Future of Identity in the Information Society) and PICOS (Privacy and Identity in Management for Community Services). He spent 3 months with the LIACC (Laboratory of Artificial Intelligence and Computer Science) group working on user profiling based on information from NetFlow. He is also interested in network security, web application security and cloud security. Marek can be contacted at *kumpost@fi.muni.cz*.

Guidelines for our Authors

Format of the manuscripts

Original manuscripts and final versions of papers should be submitted in IEEE format according to the formatting instructions available on
http://www.ieee.org/publications_standards/publications/authors/authors_journals.html#sect2,
"Template and Instructions on How to Create Your Paper".

Length of the manuscripts

The length of papers in the aforementioned format should be 6-8 journal pages.
Wherever appropriate, include 1-2 figures or tables per journal page.

Paper structure

Papers should follow the standard structure, consisting of *Introduction* (the part of paper numbered by "1"), and *Conclusion* (the last numbered part) and several *Sections* in between.
The Introduction should introduce the topic, tell why the subject of the paper is important, summarize the state of the art with references to existing works and underline the main innovative results of the paper. The Introduction should conclude with outlining the structure of the paper.

Accompanying parts

Papers should be accompanied by an *Abstract* and a few *index terms (Keywords)*. For the final version of accepted papers, please send the *short cvs* and *photos* of the authors as well.

Authors

In the title of the paper, authors are listed in the order given in the submitted manuscript. Their full affiliations and e-mail addresses will be given in a footnote on the first page as shown in the template. No degrees or other titles of the authors are given. Memberships of IEEE, HTE and other professional societies will be indicated so please supply this information.
When submitting the manuscript, one of the authors should be indicated as corresponding author providing his/her postal address, fax number and telephone number for eventual correspondence and communication with the Editorial Board.

References

References should be listed at the end of the paper in the IEEE format, see below:

a) Last name of author or authors and first name or initials, or name of organization
b) Title of article in quotation marks
c) Title of periodical in full and set in italics
d) Volume, number, and, if available, part
e) First and last pages of article
f) Date of issue

[11] Boggs, S.A. and Fujimoto, N., "Techniques and instrumentation for measurement of transients in gas-insulated switchgear," IEEE Transactions on Electrical Installation, vol. ET-19, no. 2, pp.87–92, April 1984.

Format of a book reference:

[26] Peck, R.B., Hanson, W.E., and Thornburn, T.H., Foundation Engineering, 2nd ed. New York: McGraw-Hill, 1972, pp.230–292.

All references should be referred by the corresponding numbers in the text.

Figures

Figures should be black-and-white, clear, and drawn by the authors. Do not use figures or pictures downloaded from the Internet. Figures and pictures should be submitted also as separate files. Captions are obligatory. Within the text, references should be made by figure numbers, e.g. "see Fig. 2."
When using figures from other printed materials, exact references and note on copyright should be included. Obtaining the copyright is the responsibility of authors.

Contact address

Authors are requested to send their manuscripts via electronic mail or on an electronic medium such as a CD by mail to the Editor-in-Chief:

Csaba A. Szabo
Department of Networked Systems and Services
Budapest University of Technology and Economics
2 Magyar Tudosok krt.
Budapest, 1117 Hungary
szabo@hit.bme.hu

IEEE INTERNATIONAL CONFERENCE ON COMMUNICATIONS

Sydney, Australia
10-14 JUNE 2014

IEEE INTERNATIONAL CONFERENCE ON COMMUNICATIONS
INDUSTRY FORUM & EXHIBITION

IEEE
IEEE COMMUNICATIONS SOCIETY

www.ieee-icc.org/2014

CALL FOR PAPERS AND PROPOSALS

The 2014 IEEE International Conference on Communications (ICC) will be held in the beautiful city of Sydney, Australia from 10 – 14 June 2014. Themed "Communications: The Centrepoint of Digital Economy," this flagship conference of IEEE Communications Society will feature a comprehensive technical program including twelve Symposia and a number of Tutorials and Workshops. IEEE ICC 2014 will also include an exceptional expo program including keynote speakers and Industry Forum & Exhibition.

TECHNICAL SYMPOSIA: We invite you to submit original technical papers in the following areas:

Selected Areas in Communications Symposium

Data Storage Track
Brian M. Kurkoski, JAIST, JP

e-Health Track
Nazim Agoulmine, University of Evry, FR

Internet of Things Track
Khaled Boussetta, University Paris 13, FR

Communications for the Smart Grid Track
Vincent Guillet, Landis+Gyr, FR

Satellite & Space Communication Track
Igor Bisio, University of Genoa, IT

Green Communications and Computing Track
John S. Thompson, University of Edinburgh, UK

Cloud Computing Track
Yonggang Wen, Nanyang Technical University, SG

Access Networks and Systems Track
Tarek S. El-Bawab, Jackson State University, USA

Nanoscale, Molecular, and Quantum Network Track
Tadashi Nakano, Osaka University, JP

Social Networking Track
Neeli Prasad, Aalborg University, DK

Wireless Communications Symposium
Yahong Rosa Zheng, Missouri University of S&T, USA
Yiqing Zhou, Chinese Academy of Sciences, CN
Cheng Li, Memorial University of Newfound, CA
Peter M. R. Rost, NEC Labs Europe, DE
Jinhong Yuan, University of NSW, AU

Mobile and Wireless Networking Symposium
Weihua Zhuang, University of Waterloo, CA
Pascal Lorenz, University of Haute-Alsace, FR
Jian Tang, Syracuse University, USA
Nurul Sarkar, Auckland University of Technology, NZ

Communication Theory Symposium
Fulvio Babich, University of Trieste, IT
Bechir Hamdaoui, Oregon State University, USA
Huaiyu Dai, North Carolina State University, USA

Signal Processing for Communications Symposium
Tomohiko Taniguchi, Fujitsu Labs, JP
Lingyang Song, Beijing Univesity, CN
Rose Qingyang Hu, Utah State University, USA

Optical Networks and Systems Symposium
Arun Somani, Iowa State University, USA
Philippe Perrier, Xtera Communications, USA
Nathan Gomes, University of Kent, UK

Next-Generation Networking Symposium
Mohammed Atiquzzaman, University of Oklahoma, USA
Konstantinos Samdanis, NEC Europe, DE
Antonio Pescapè, University of Napoli
Federico II, IT

Communication QoS, Reliability & Modeling Symposium
Kohei Shiomoto, NTT, JP
Stefano Giordano, University of Pisa, IT
Wei Song, University of New Brunswick, CA

Ad Hoc and Sensor Networking Symposium
Jalel Ben-Othman , University of Paris 13, FR
Jiming Chen, Zhejiang University, CN
Somaya Charkaoui, University of Sherbrooke, CA
Zubair Md. Fadlullah, Tohoku University, JP

Communication Software, Services and Multimedia Application Symposium
Abdelhamid Mellouk, University of Paris 12, FR
Lingfen Sun, University of Plymouth, UK

Communication and Information Systems Security Symposium
Peter Mueller, IBM Zurich Research, CH
Shui Yu, Deakin University, AU
Thorsten Strufe, Technische Universität Darmstadt, DE

Cognitive Radio and Networks Symposium
Jacques Palicot, Supelec, FR
Jaime L Mauri, Polytechnic University of Valencia, ES
Lin Cai, Huawei Technologies, USA

TUTORIALS: Proposals should provide a focused lecture on new and emerging topics within the scope of communications.

WORKSHOPS: Proposals should emphasize current topics of particular interest, and should include a mix of regular papers, invited presentations and panels that encourage the participation of attendees in active discussion.

Accepted and presented technical and workshop papers will be published in the IEEE ICC 2014 Conference Proceedings and in IEEE Xplore®. See the website for requirements of accepted authors.

IMPORTANT DATES	Paper Submission	Acceptance Notification	Camera-Ready	Tutorial Proposal	Workshop Proposal
	15 September 2013	**12 January 2014**	**13 February 2014**	**31 October 2013**	**31 March 2013**

General Chair:
Farzad Safai, University of Wollongong, AU

Vice General Chair:
Leith Campbell, Ovum, AU

Technical Program Chair:
Abbas Jamalipour, University of Sydney, AU

Technical Program Vice Co-Chairs:
Sanjay Jha, University of New South Wales, AU
Grenville Armitage, Swinburne University of Technology, AU

Symposia Chair:
Nei Kato, Tohoku University, JP

Workshops Co-Chairs:
Hsiao-Hwa Chen, National Cheng Kung University, TW
Jean Armstrong, Monash University, AU

Tutorials Co-Chairs:
Sherman Shen, University of Waterloo, CA
Chun Tung Chou, University of New South Wales, AU

Full details of submission procedures are available at
www.ieee-icc.org/2014

SCIENTIFIC ASSOCIATION FOR INFOCOMMUNICATIONS

Who we are

Founded in 1949, the Scientific Association for Infocommunications (formerly known as Scientific Society for Telecommunications) is a voluntary and autonomous professional society of engineers and economists, researchers and businessmen, managers and educational, regulatory and other professionals working in the fields of telecommunications, broadcasting, electronics, information and media technologies in Hungary.

Besides its more than 1300 individual members, the Scientific Association for Infocommunications (in Hungarian: Hírközlési és Informatikai Tudományos Egyesület, HTE) has more than 60 corporate members as well. Among them there are large companies and small-and-medium enterprises with industrial, trade, service-providing, research and development activities, as well as educational institutions and research centers.

HTE is a Sister Society of the Institute of Electrical and Electronics Engineers, Inc. (IEEE) and the IEEE Communications Society. HTE is corporate member of International Telecommunications Society (ITS).

What we do

HTE has a broad range of activities that aim to promote the convergence of information and communication technologies and the deployment of synergic applications and services, to broaden the knowledge and skills of our members, to facilitate the exchange of ideas and experiences, as well as to integrate and harmonize the professional opinions and standpoints derived from various group interests and market dynamics.

To achieve these goals, we…

- contribute to the analysis of technical, economic, and social questions related to our field of competence, and forward the synthesized opinion of our experts to scientific, legislative, industrial and educational organizations and institutions;
- follow the national and international trends and results related to our field of competence, foster the professional and business relations between foreign and Hungarian companies and institutes;
- organize an extensive range of lectures, seminars, debates, conferences, exhibitions, company presentations, and club events in order to transfer and deploy scientific, technical and economic knowledge and skills;
- promote professional secondary and higher education and take active part in the development of professional education, teaching and training;
- establish and maintain relations with other domestic and foreign fellow associations, IEEE sister societies;
- award prizes for outstanding scientific, educational, managerial, commercial and/or societal activities and achievements in the fields of infocommunication.

Contact information

President: **DR. GÁBOR HUSZTY** • *ghuszty@entel.hu*

Secretary-General: **DR. ISTVÁN BARTOLITS** • *bartolits@nmhh.hu*

Managing Director, Deputy Secretary-General: **PÉTER NAGY** • *nagy.peter@hte.hu*

International Affairs: **ROLLAND VIDA, PhD** • *vida@tmit.bme.hu*

Addresses

Office: H-1055 Budapest, V. Kossuth Lajos square 6-8, Room: 422.

Mail Address: 1372 Budapest, Pf. 451., Hungary

Phone: +36 1 353 1027, Fax: +36 1 353 0451

E-mail: *info@hte.hu*, Web: *www.hte.hu*