JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY

1/2018

(*Contents Continued on Back Cover*)

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

Circulation: 300 copies Sowa – Druk na życzenie, www.sowadruk.pl, tel. 22 431-81-40

Ministry of Science and Higher Education
Republic of Poland

Improvement of language quality; Assigning DOIs; Subscription to the plagiarism detection system – tasks financed under 556/P-DUN/2017 agreement from the budget of the Ministry of Science and Higher Education under the science dissemination fund.

JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY

Preface

The current issue of the *Journal of Telecommunications and Information Technology* contains twelve papers. In their articles, the authors consider important problems related to modern telecommunications.

The first paper entitled *Call-level Analysis of a Two-Link Multirate Loss Model based on a Convolution Algorithm* by Sagkriotis *et al*. presents a multirate teletraffic loss model of a two-link system that accommodates Poisson arriving calls from different service-classes. The model was worked out under assumption that each link has two thresholds and based on a convolution algorithm.

The next two papers deal with Elastic Optical Networks. The paper by Kabaciński, Rajewski, and Al-Tameemi *Rearrangeability of* 2×2 *W-S-W Elastic Switching Fabrics with Two Connection Rates*, considers rearreangeable conditions for the 2×2 three-stage switching fabric of a W-S-W (Wavelength-Space-Wavelength) architecture for elastic optical switches. The authors present rearrangeable conditions and an appropriate control algorithm.

The next paper by Rajewski, titled *Defragmentation in W-S-W Elastic Optical Networks*, discusses defragmentation in an elastic optical network's node. The author bases his considerations on the assumption that the W-S-W (Wavelength-Space-Wavelength) switching architecture has been used as a node.

The paper *SMM Clos-Network Switches under SD Algorithm* by Kleban and Warczyński is devoted to evaluating the performance of Space-Memory-Memory (SMM) Clos-network switches under a packet dispatching scheme employing static connection patterns, referred to as Static Dispatching (SD).

The paper entitled *Ganging of Resources via Fuzzy Manhattan Distance Similarity with Priority Tasks Scheduling in Cloud Computing* by Priya, Mehata, and Banu, proposes a fuzzy Manhattan distance-based similarity for gang formation of resources (FMDSGR) method with priority task scheduling in cloud computing.

Other papers deal with issues related to various problems occurring in wireless communication. The article titled *Observation of WiMAX Radio Parameters to Enhance Spectrum Utilization in Mixed Environment* by Kowalik *et al*. presents statistical characteristics of

actual, IEEE 802.11e compliant WiMAX signals, as seen from the point of view of improving spectrum utilization by means of simultaneous use of given frequency bands by two wireless systems.

In the paper titled *Performance of Hybrid Sensing Method in Environment with Noise Uncertainty*, Kustra, Kosmowski, and Suchański present a novel hybrid spectrum sensing method used in cognitive radio and present a hybrid detector (HD), which improves the sensing performance.

In *Performance Analysis of SPSK with Dual Polarized Transmit Antennas over Rayleigh Fading Channel*, Subramani, Neduncheran and Ponnusamy study the Space Polarization Shift Keying (SPSK) system, which is an extended version of Space Shift Keying (SSK) and includes both space and polarization dimensions with dual polarized antennas.

The article *Outage Performance of Bidirectional Full-Duplex Amplify-and-Forward Relay Network with Transmit Antenna Selection and Maximal Ratio Combining* by Rajesh *et al*. proposes a bidirectional full-duplex amplify-and-forward (AF) relay network with multiple antennas at source nodes.

Another paper entitled *Miniaturized Spectacles Shaped Tapered Slotted Patch Antenna for UWB Applications* by Tarikul Islam *et al*. presents a compact planner patch ultra-wideband (UWB) antenna.

In *Protocols for Wireless Sensor Networks: A Survey*, Kochhar, Kaur, Preeti and Sharma present a review of the MAC and network layer of Wireless Sensor Networks. Performance requirements of the MAC layer are explored too.

In *Underwater Acoustic Sensor Node Scheduling using an Evolutionary Memetic Algorithm*, Sivakumar and Rekha show how to optimize the utilization of acoustic sensor node bandwidth by maximizing the possible node transmissions in the TDMA frame and by minimizing the node's turnaround wait time for its subsequent transmissions by using an evolutionary memetic algorithm (MA).

I would like to thank all authors and reviewers for the effort they have put into preparing this issue of *Journal of Telecommunications and Information Technology*.

Sławomir Hanczewski, Ph.D.
Guest Editor

https://doi.org/10.26636/jtit.2018.preface1

Paper

Call-level Analysis of a Two-Link Multirate Loss Model based on a Convolution Algorithm

Stefanos G. Sagkriotis[1], Spyros K. Pantelis[2], Ioannis D. Moscholios[2], and Vassilios G. Vassilakis[3]

[1] *School of Computer Science, University of Glasgow, Glasgow, United Kingdom*
[2] *University of Peloponnese, Tripolis, Greece*
[3] *University of York, York, United Kingdom*

https://doi.org/10.26636/jtit.2018.123217

Abstract—We consider a two-link system that accommodates Poisson arriving calls from different service-classes and propose a multirate teletraffic loss model for its analysis. Each link has two thresholds, which refer to the number of in-service calls in the link. The lowest threshold, named support threshold, defines up to which point the link can support calls offloaded from the other link. The highest threshold, named offloading threshold, defines the point where the link starts offloading calls to the other link. The adopted bandwidth sharing policy is the complete sharing policy, in which a call can be accepted in a link if there exist enough available bandwidth units. The model does not have a product form solution for the steady state probabilities. However, we propose approximate formulas, based on a convolution algorithm, for the calculation of call blocking probabilities. The accuracy of the formulas is verified through simulation and found to be quite satisfactory.

Keywords—call blocking, convolution, loss model, offloading, product form.

1. Introduction

Quality of service (QoS) mechanisms are necessary in contemporary communication networks in order to provide the required bandwidth needed by calls. In the case of call-level traffic in a single link, modeled as a loss system, such a QoS mechanism is a bandwidth sharing policy [1]. The simplest bandwidth sharing policy is the complete sharing (CS) policy, where a new call is accepted in the system if there exist enough available bandwidth units (b.u.). Otherwise, call blocking occurs. The simplest teletraffic loss model that adopts the CS policy is the classic Erlang model [1]. In this model, the call arrival process is Poisson, while each call requires one b.u. to be accepted in the system. An accepted call has a generally distributed service time. The fact that call blocking probabilities (CBP) are calculated via the classic Erlang B formula has led to numerous extensions of Erlang's model for the call-level analysis of wired (e.g. [2]–[16]), wireless (e.g. [17]–[28]), satellite (e.g. [29]–[31]) and optical networks (e.g. [32]–[37]).

In the work of [25], the Erlang B formula has been adopted for the determination of CBP in a two access link system that accommodates Poisson arriving calls of a single service-class. Each access link is modelled as a loss system (i.e. no queueing is permitted) and has two thresholds, which refer to the number of in-service calls in the link. The lowest threshold, named support threshold, defines up to which point the access link can support calls offloaded from the other access link. The highest threshold, named offloading threshold, defines the point where the access link starts offloading calls to the other access link. By the term offloaded call, we refer to a call that initially arrived in a link, but is served by the other link, if there exist available b.u. The model of [25] does not have a product form solution (PFS) for the steady state probabilities. This is due to the fact that the offloading mechanism destroys local balance (LB) between adjacent states (states that differ only by one call) of the system. To calculate the various performance measures of the system, e.g. CBP or link utilization, either a linear system of global balance (GB) equations should be solved or an approximate method that relies on the independence between the links and the classic Erlang B formula can be adopted. The system of GB equations leads to an accurate calculation of the performance measures but it requires the knowledge of the state space of the two-link system. Such a state space may consist of millions of states if the capacity of the links is high. Thus, the method of solving the GB equations can only be applied in small (tutorial) systems [38]. On the other hand, the link independence assumption and the Erlang B formula facilitate the necessary calculations.

A potential application of the offloading scheme of [25] is in the area of mobile/Wi-Fi networks. To manage the increasing traffic in mobile networks, traffic can be offloaded to Wi-Fi networks [39], [40]. To further increase the available bandwidth of Wi-Fi access links, recent research focuses on the aggregation of backhaul access link capacities and on the bandwidth sharing policies that should be adopted (see e.g. the BeWi-Fi concept that enables users in proximity to share their Internet access if their link utilization is below a threshold) [41]. The impact of such an aggregation to CBP in the case of a single service-class can be well studied by the offloading scheme of [25].

In this paper, we extend the model of [25] to include the important case of multirate traffic, i.e. we consider a two access link system that accommodates Poisson arriving calls of different service-classes and different bandwidth-per-call requirements. The CBP calculation in the proposed two-link model under the CS policy is based on the classic Erlang Multirate Loss Model (EMLM) [42], [43], which refers to a single link. To differentiate, we name the proposed model 2EMLM. In the 2EMLM model, the determination of CBP can be done via a 3-step convolution algorithm. The latter exploits the PFS of the EMLM and the principle of independency among service-classes and, therefore, the link occupancy distribution can be determined by successively convolving the link occupancy distributions obtained for each service-class. Contrary to macro-state recursive formulas (such as the classic Kaufman-Roberts formula used for CBP calculation in EMLM [42], [43]), a convolution algorithm keeps the micro-state information of the number of in-service calls in a link. Such information is necessary when studying more complicated (than the CS policy) call admission policies (e.g. [44]–[53]).

The remainder of this paper is organized as follows: in Section 2, we review the system of [25]. In Section 3, we propose the 2EMLM and provide a convolution algorithm for CBP determination. In Section 4, we provide analytical and simulation CBP results for the proposed model. We conclude in Section 5. In the Appendix, we provide a tutorial example of the system of [25].

2. The Two-Link System with Single-Rate Traffic

We consider a system of two links with capacities C_1 and C_2 b.u., respectively. Each link accommodates Poisson arriving calls of a single service-class which require one b.u. in order to be connected in a link. Let λ_1 and λ_2 be the arrival rates in the 1st and the 2nd link, respectively. We also denote by j_1 and j_2 the occupied b.u. in the 1st and the 2nd link, respectively. Then, $0 \le j_1 \le C_1$ and $0 \le j_2 \le C_2$. Since calls require one b.u., the values of j_1, j_2 also represent the number of in-service calls in the 1st and the 2nd link, respectively.

Each link l ($l = 1,2$) has two different thresholds: the support threshold th_{1l} and the offloading threshold th_{2l}, with $th_{1l} < th_{2l}$ and $0 \le th_{1l}, th_{2l} \le 1$. Assuming that $\lfloor x \rfloor$ is the largest integer not exceeding x, the role of these thresholds, in the l-th link, is the following (see Fig. 1):

- If $0 \le j_l < \lfloor th_{1l} C_l \rfloor$ then the l-th link is in a *support mode* of operation, i.e. it accepts and serves not only new calls that initially arrive in the l-th link, but also new calls offloaded from the m-th link ($m = 1,2$, $m \ne l$).

- If $\lfloor th_{1l} C_l \rfloor \le j_l < \lfloor th_{2l} C_l \rfloor$ then the l-th link is in a *normal mode* of operation, i.e. it does not accept calls offloaded from the m-th link. It only accepts calls that initially arrive in the l-th link.

- If $\lfloor th_{2l} C_l \rfloor \le j_l$ then the l-th link is in an *offloading mode* of operation, i.e. a new call that initially arrives in the l-th link will be offloaded to the m-th link. If the m-th link is in *support mode* (i.e. $0 \le j_m < \lfloor th_{1m} C_m \rfloor$) then the call will be accepted in the m-th link. If the m-th link is not in *support mode* and $j_l \le C_l - 1$, the call will be accepted in the l-th link. Otherwise the call will be blocked and lost.

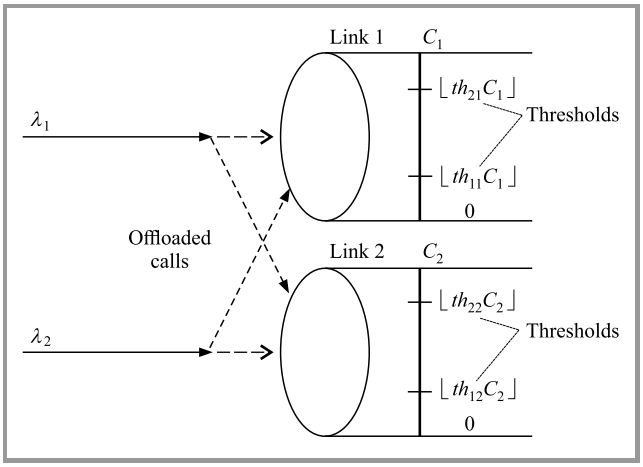


Fig. 1. The system of the two links.

Based on the above, the admission of a new call that initially arrives in the l-th link ($l = 1,2$) is summarized in the following steps:

1) If $\left(0 \le j_l < \lfloor th_{2l} C_l \rfloor\right)$ then the call is accepted by the l-th link and remains for a generally distributed service-time with mean μ^{-1}.

2) If $\lfloor th_{2l} C_l \rfloor \le j_l$ then:

   2a) if $0 \le j_m < \lfloor th_{1m} C_m \rfloor$ the call is offloaded to the m-th link and remains for a generally distributed service-time with mean μ^{-1};

   2b) if $\lfloor th_{1m} C_m \rfloor \le j_m$, the m-th link is in a *normal mode* of operation and does not support offloaded calls from the l-th link. In that case, the call will try to be accepted in the l-th link. If $j_l \le C_l - 1$, then the call is accepted in the l-th link and remains for a generally distributed service-time. Otherwise, the call is blocked and lost without further affecting the system of the two links.

A tutorial example in the Appendix, presents in detail the call admission mechanism and the required calculations for CBP determination.

Due to the *support* and *offloading modes* of operation of the two links, the 2-D Markov chain of the system is not reversible and, therefore, LB between adjacent states (states that differ only by one call) is destroyed. Thus, the steady

state distribution, $P(j)=P(j_1,j_2)$, of this system cannot be described by a PFS. To determine the values of $P(j_1,j_2)$ (and consequently CBP) there exist two different methods. The first method provides accurate results (compared to simulation) but requires the knowledge of the state space of the system and the solution of the set of linear GB equations for each state $j=(j_1,j_2)$ expressed as *rate into state j = rate out of state j*:

$$\begin{aligned}&\lambda_1(j_1-1,j_2)P(j_1-1,j_2)+\lambda_2(j_1,j_2-1)P(j_1,j_2-1)+\\&+(j_1+1)\mu P(j_1+1,j_2)+(j_2+1)\mu P(j_1,j_2+1)=\\&=\lambda_1(j_1,j_2)P(j_1,j_2)+\lambda_2(j_1,j_2)P(j_1,j_2)+\\&+(j_1\mu+j_2\mu)P(j_1,j_2)\ ,\end{aligned}\tag{1}$$

where:

$$\lambda_1(j_1,j_2)\overset{l=1,2,m\neq l}{=}\begin{cases}\lambda_l+\lambda_m, & \text{if } (j_l<\lfloor th_{1l}C_l\rfloor)\\ & \cap(j_m\geq\lfloor th_{2m}C_m\rfloor)\\ 0, & \text{if } (j_l\geq\lfloor th_{2l}C_l\rfloor)\\ & \cap(j_m<\lfloor th_{1m}C_m\rfloor)\ .\\ 0, & \text{if } (j_1,j_2) \text{ is}\\ & \text{a boundary state}\\ \lambda_l, & \text{otherwise}\end{cases}\tag{2}$$

Having obtained the values of $P(j_1,j_2)$, we can determine the CBP in the 1st and the 2nd link, P'_{b_1} and P'_{b_2} via Eqs. (3) and (4), respectively [25]:

$$P'_{b_1}=\sum_{j_2=\lfloor th_{12}C_2\rfloor}^{C_2}P(C_1,j_2)\,,\tag{3}$$

$$P'_{b_2}=\sum_{j_1=\lfloor th_{11}C_1\rfloor}^{C_2}P(j_1,C_2)\,.\tag{4}$$

In addition, we can calculate the total blocking probability in the system via the following weighted summation:

$$P'_b=\frac{\lambda_1}{\lambda_1+\lambda_2}P'_{b_1}+\frac{\lambda_2}{\lambda_1+\lambda_2}P'_{b_2}\,.\tag{5}$$

Before we proceed with the second method, we emphasize that the state space determination and the solution of the set of GB equations can be quite complex even for systems of moderate size and, therefore, is only practically used for small tutorial examples (see Appendix).

The second method provides approximate CBP results by assuming that the two links operate independently from one another. Such an assumption simplifies the necessary CBP calculations. Since each independent link behaves as an Erlang loss system, the CBP in the 1st and the 2nd link can be approximated by Eqs. (6) and (7), respectively:

$$P_{b_1}=P_1(C_1)P_2(j_2\geq\lfloor th_{12}C_2\rfloor)\,,\tag{6}$$

$$P_{b_2}=P_2(C_2)P_1(j_1\geq\lfloor th_{11}C_1\rfloor)\,,\tag{7}$$

where $P_l(C_l)$ refers to the CBP in the l-th link ($l=1,2$) which can be determined by the Erlang B formula:

$$P_l(C_l)=\frac{\dfrac{\alpha_l^{C_l}}{C_l!}}{\displaystyle\sum_{i=0}^{C_l}\frac{\alpha_l^i}{i!}},\qquad\alpha_l=\frac{\lambda_l}{\mu}\,.\tag{8}$$

As far as the values of $P_l(j_l\geq\lfloor th_{1l}C_l\rfloor)$ are concerned they are given by:

$$P_l(j_l\geq\lfloor th_{1l}C_l\rfloor)=\sum_{j_l=\lfloor th_{1l}C_l\rfloor}^{C_l}P_l(j_l)\,,\tag{9}$$

where $P_l(j_l)$ is determined by the truncated Poisson distribution:

$$P_l(j_l)=\frac{\dfrac{\alpha_l^{j_l}}{j_l!}}{\displaystyle\sum_{i=0}^{C_l}\frac{\alpha_l^i}{i!}},\qquad\alpha_l=\frac{\lambda_l}{\mu}\,.\tag{10}$$

The rationale behind Eqs. (6) and (7) is that a call that initially arrives in the l-th link will be blocked if there are no available b.u. in that link and the m-th link is not in *support mode* of operation.

Finally, the total blocking probability can be determined via the following formula:

$$P_b=\frac{\lambda_1}{\lambda_1+\lambda_2}P_{b_1}+\frac{\lambda_2}{\lambda_1+\lambda_2}P_{b_2}\,.\tag{11}$$

3. The Proposed 2EMLM

In the proposed 2EMLM, we consider again the system of the two links. Each link accommodates Poisson arriving calls of K service-classes. Calls of service-class k ($k=1,...,K$) require b_k b.u. in order to be connected in a link. Let λ_{1k} and λ_{2k} be the arrival rates in the 1st and the 2nd link of service-class k calls, respectively. We also denote by j_1 and j_2 the occupied b.u. in the 1st and the 2nd, respectively. Then, $0\leq j_1\leq C_1$ and $0\leq j_2\leq C_2$. Similar to Section 2, each link l ($l=1,2$) has a support threshold th_{1l} and an offloading threshold th_{2l}, with $th_{1l}<th_{2l}$ and $0\leq th_{1l},th_{2l}\leq 1$.

The call admission of a new service-class k call that initially arrives in the l-th link ($l=1,2$) is summarized in the following steps:

1) If $\left(0\leq j_l<\lfloor th_{2l}C_l\rfloor\right)\cap(j_l+b_k\leq C_l)$ then the call is accepted by the l-th link and remains for a generally distributed service-time with mean μ_k^{-1}.

2) If $\lfloor th_{2l}C_l\rfloor\leq j_l$ then:

   2a) if $\left(0\leq j_m<\lfloor th_{1m}C_m\rfloor\right)\cap(j_m+b_k\leq C_m)$ the call is offloaded to the m-th link and remains for a generally distributed service-time with mean μ_k^{-1};

   2b) if $\lfloor th_{1m}C_m\rfloor\leq j_m$, the m-th link is in *normal mode* of operation and does not support offloaded calls from the l-th link. In that case, the call will try to

be accepted in the l-th link. If $j_l + b_k \leq C_l$, then the call is accepted in the l-th link and remains for a generally distributed service-time with mean μ_k^{-1}. Otherwise, the call is blocked and lost.

To determine in an approximate but efficient way the CBP of service-class k calls we assume that the two links operate independently from one another. In that case, each independent link behaves as an EMLM system, and therefore the CBP of service-class k calls in the 1st and the 2nd link can be approximated by Eqs. (12) and (13), respectively:

$$P_{b_{1k}} = P_{1k}(C_1)P_2(j_2 \geq \lfloor th_{12}C_2 \rfloor), \tag{12}$$

$$P_{b_{2k}} = P_{2k}(C_2)P_1(j_1 \geq \lfloor th_{11}C_1 \rfloor), \tag{13}$$

where $P_{lk}(C_l)$ refers to the CBP of service-class k calls in the l-th link ($l = 1,2$).
The values of $P_{lk}(C_l)$ in Eqs. (12) and (13) are determined by:

$$P_{lk}(C_l) = \sum_{j_l=C_l-b_k+1}^{C_l} G_l^{-1} q(j_l), \tag{14}$$

where $q(j_l)$ refers to the unnormalized values of the link occupancy distribution of link l ($l = 1,2$) while $G_l = \sum_{j_l=0}^{C_l} q(j_l)$ is the normalization constant.
In Eq. (14), the values of $q(j_l)$ can be recursively determined via a 3-step convolution algorithm. To describe it, let $q_{l,k}(j)$ ($k = 1,\ldots,K$) be the link occupancy distribution assuming that only service-class k exists in the link l. Then, the 2EMLM convolution algorithm is as follows:
Step 1. Determine $q_{l,k}(j)$ of each service-class k via:

$$q_{l,k}(j_l) = q_{l,k}(0)\frac{\alpha_{lk}^i}{i!}, \text{ for } 1 \leq i \leq \left\lfloor \frac{C_l}{b_k} \right\rfloor \text{ and } j_l = i \times b_k, \tag{15}$$

where $\alpha_{lk} = \lambda_{lk}/\mu_k$ is the offered traffic-load (in Erl) of service-class k calls in link l.
Step 2. Determine the aggregated occupancy distribution $Q_{(-k)}$ based on the successive convolution of all service-classes (in link l) apart from service-class k:

$$Q_{(-k)} = q_{l,1} \cdot \ldots \cdot q_{l,k-1} \cdot q_{l,k+1} \cdot \ldots \cdot q_{l,K}. \tag{16}$$

The term "successive" means that initially we convolve $q_{l,1}$ and $q_{l,2}$ to obtain $q_{l,12}$. Then we convolve $q_{l,1}$ with $q_{l,3}$ to obtain $q_{l,123}$ etc. The convolution operation between service-classes k and r is as follows:

$$q_{l,k} \cdot q_{l,r} = \left\{ q_{l,k}(0)q_{l,r}(0), \sum_{x=0}^{1} q_{l,k}(x)q_{l,r}(1-x), \ldots, \right.$$
$$\left. \ldots, \sum_{x=0}^{C} q_{l,k}(x)q_{l,r}(C_l - x) \right\}. \tag{17}$$

Step 3. Determine the values of $q(j_l)$ based on the convolution operation of $Q_{l,(-k)}$ (step 2) and $q_{l,k}$ as follows:

$$Q_{l,(-k)} \cdot q_{l,k} = \left\{ Q_{l,(-k)}(0)q_{l,k}(0), \sum_{x=0}^{1} Q_{l,(-k)}(x) \right.$$
$$\left. q_{l,k}(1-x), \ldots, \sum_{x=0}^{C_l} Q_{l,(-k)}(x)q_{l,k}(C_l - x) \right\}. \tag{18}$$

Normalizing the values of (18), we obtain the occupancy distribution $q(j_l)$, $j = 0,1,\ldots,C_l$ via the formulas:

$$\begin{aligned} q(0) &= \frac{Q_{l,(-k)}(0)q_{l,k}(0)}{G_l} \\ q(j) &= \frac{\sum_{x=0}^{j} Q_{l,(-k)}(x)q_{l,k}(j-x)}{G_l},\ j = 1,\ldots,C_l \end{aligned}. \tag{19}$$

As far as the values of $P_l(j_l \geq \lfloor th_{1l}C_l \rfloor)$, in Eqs. (12) and (13), are concerned, they are calculated by:

$$P_l(j_l \geq \lfloor th_{1l}C_l \rfloor) = \sum_{j_l=\lfloor th_{1l}C_l \rfloor}^{C_l} G_l^{-1} q(j_l), \tag{20}$$

where $q(j_l)$ is determined via (19).
Finally, we propose the following formula for the total blocking probability of service-class k calls in the system of the two links:

$$P_{b_k} = \frac{\lambda_{1k}}{\lambda_{1k}+\lambda_{2k}} P_{b_{1k}} + \frac{\lambda_{2k}}{\lambda_{1k}+\lambda_{2k}} P_{b_{2k}}. \tag{21}$$

4. Numerical Examples – Evaluation

In this section, we present an application example and provide analytical and simulation results of the total CBP of the proposed model. Simulation results are derived via the Simscript III simulation language [54] and are mean values of 7 runs. As far as the reliability ranges are concerned, they are less than two orders of magnitude, and therefore are not presented in the following figures. All simulation runs are based on the generation of eight million calls per

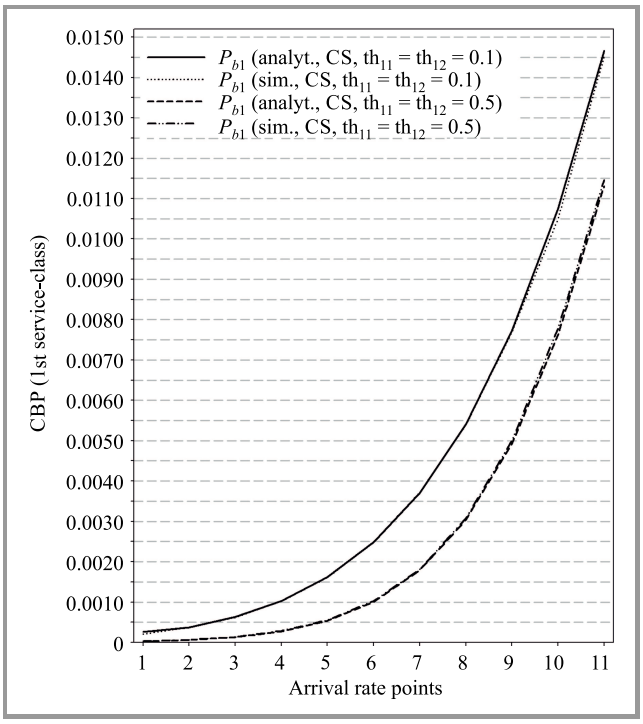


Fig. 2. CBP under the CS policy – 1st service-class.

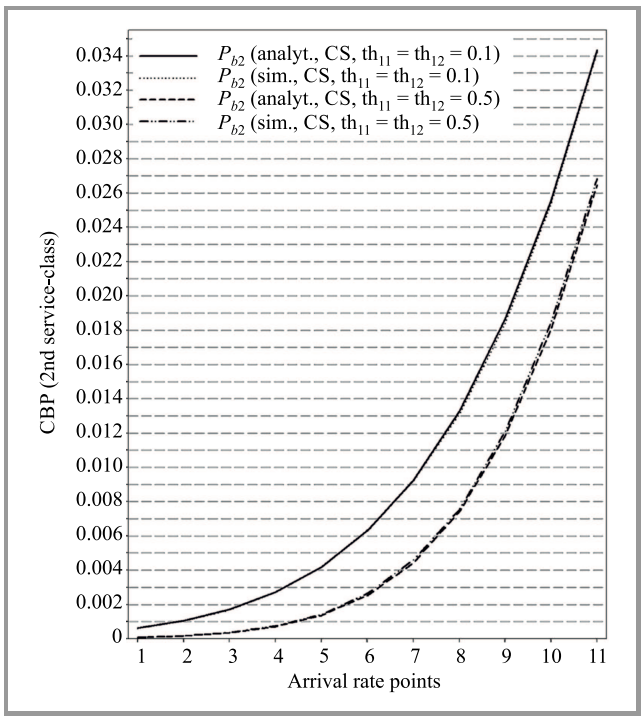


Fig. 3. CBP under the CS policy – 2nd service-class.

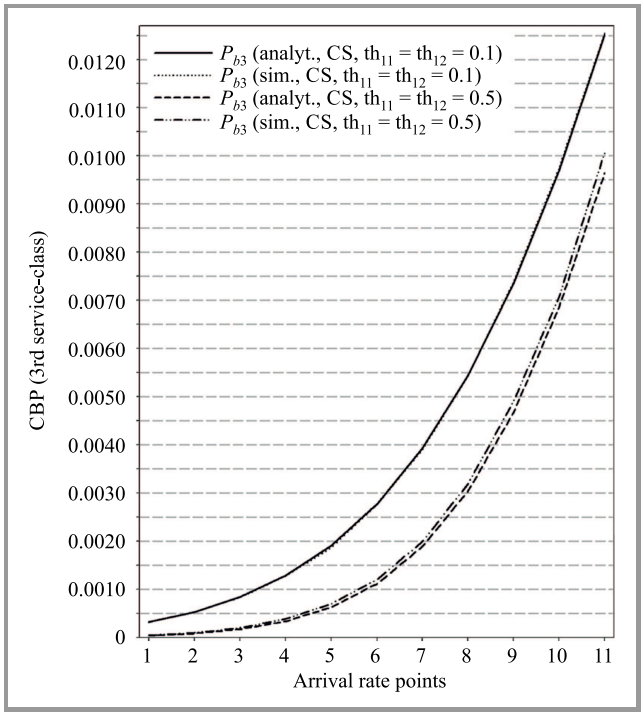


Fig. 4. CBP under the CS policy – 3rd service-class.

run. To account for a warm-up period, the first 5% of these generated calls are not considered in the CBP results.

As an application example, consider a system of two links of capacities $C_1 = 30$ and $C_2 = 25$ b.u., that accommodates $K = 3$ service-classes whose calls require $b_1 = 1$, $b_2 = 2$ and $b_3 = 5$ b.u., respectively. For the 1st link, let: $\lambda_{11} = 4.0$, $\lambda_{12} = 1.0$, $\lambda_{13} = 0.5$. Similarly, for the 2nd link, let: $\lambda_{21} = 2.0$, $\lambda_{22} = 1.0$, $\lambda_{23} = 0.5$. We also assume that $\mu_1^{-1} = \mu_2^{-1} = \mu_3^{-1} = 1.0$.

We consider two different support thresholds: 1) $th_{11} = th_{12} = 0.1$ and 2) $th_{11} = th_{12} = 0.5$. In both cases, we assume that the offloading thresholds do not alter and are equal to: $th_{21} = th_{22} = 0.7$.

In the x-axis of Figs. 2–4, λ_{11} and λ_{21} increase in steps of 1.0 and 1.0, respectively. So, point 1 is: ($\lambda_{11} = 4.0$, $\lambda_{12} = 1.0$, $\lambda_{13} = 0.5$, $\lambda_{21} = 2.0$, $\lambda_{22} = 1.0$, $\lambda_{23} = 0.5$) while point 11 is: ($\lambda_{11} = 14.0$, $\lambda_{12} = 1.0$, $\lambda_{13} = 0.5$, $\lambda_{21} = 12.0$, $\lambda_{22} = 1.0$, $\lambda_{23} = 0.5$).

In Figs. 2–4, we present CBP in the 2EMLM for the three service-classes, respectively. Figures 2–4 show that the analytical CBP results: a) are close to the simulation results and b) decrease as the support thresholds increase, an intuitively expected fact since both links cooperate with each other. Similar conclusions have been observed for systems of more than three service-classes but are not presented herein.

5. Conclusion

In this paper we propose a multirate loss model for a two-link loss system that accommodates Poisson arriving calls. A link can share a part of its capacity in order to support calls from the other link and vice versa. The proposed model does not have a PFS for the steady state distribution due to the existence of the offloading mechanism. However, we show that an approximate method does exist (based on a convolution algorithm) that provides quite satisfactory CBP results compared to simulation. As a future work, we intend to study this two-link system under the assumption that it serves different service-classes whose calls follow a quasi-random process, i.e. calls that are generated by a finite number of sources.

Appendix – Tutorial Example

Consider a system of two links with $C_1 = 6$ and $C_2 = 5$ b.u., that accommodates calls of a single service-class. Let $\lambda_1 = 4$ calls/min, $\lambda_2 = 2$ calls/min and $\mu^{-1} = 1$ min. The thresholds for this system are the following:

1st link ($l = 1$): $th_{11} = 0.2, th_{21} = 0.7$,

2nd link ($l = 2$): $th_{12} = 0.2, th_{22} = 0.7$.

Based on the thresholds' values we have:

First link

a) If $0 \leq j_1 < \lfloor th_{11}C_1 \rfloor \Rightarrow 0 \leq j_1 < 1$ then the 1st link is in a *support mode* of operation.

b) If $\lfloor th_{11}C_1 \rfloor \leq j_1 < \lfloor th_{12}C_1 \rfloor \Rightarrow 1 \leq j_1 < 4$ then the 1st link is in a *normal mode* of operation.

c) If $\lfloor th_{21}C_1 \rfloor \leq j_1 \Rightarrow 4 \leq j_1$ then the 1st link is in an *offloading mode* of operation.

Second link

a) If $0 \leq j_2 < \lfloor th_{12}C_2 \rfloor \Rightarrow 0 \leq j_2 < 1$ then the 2nd link is in a *support mode* of operation.

b) If $\lfloor th_{12}C_2 \rfloor \leq j_2 < \lfloor th_{22}C_2 \rfloor \Rightarrow 1 \leq j_2 < 3$ then the 2nd link is in a *normal mode* of operation.

c) If $\lfloor th_{22}C_2 \rfloor \leq j_2 \Rightarrow 3 \leq j_2$ then the 2nd link is in an *offloading mode* of operation.

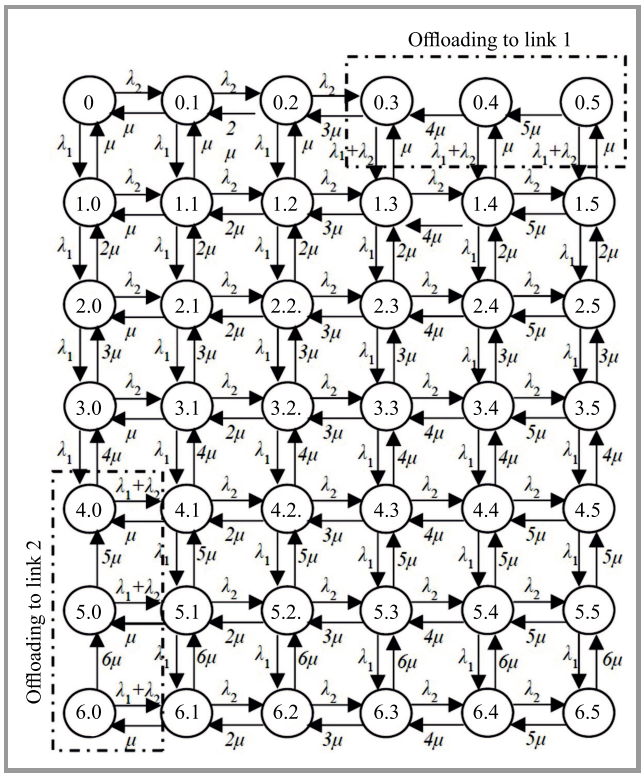


Fig. 5. State transition diagram of the tutorial example.

The state space of the system consists of 42 states of the form (j_1, j_2), depicted in Fig. 5 together with the corresponding transition rates. To help a reader understand the state transition diagram of Fig. 5 and the offloading mechanism, assume that the system is in state (0, 2) when a new call arrives in the 2nd link. Then, the call will be accepted in the 2nd link and the new state will be (0, 3). If another new call arrives in the 2nd link then the call will be offloaded to the 1st link (and served by that link) and the new state will be (1, 3). If now, another call arrives in the 2nd link, then this call cannot be offloaded to the 1st link (since $j_1 = 1$) but it can be served by the 2nd link due to bandwidth availability. In that case the new state will be (1, 4). A similar rationale exists when we consider call arrivals in the 1st link and the states (3, 0), (4,0), (4,1) and (5,1).
Based on the solution of the 42 GB equations of Fig. 5, the CBP in the 1st and 2nd link is given by:

$$P'_{b_1} = \sum_{j_2=\lfloor th_{12}C_2 \rfloor} P(C_1, j_2) = \sum_{j_2=1}^{5} P(6, j_2) = 0.10370,$$

$$P'_{b_2} = \sum_{j_1=\lfloor th_{11}C_1 \rfloor} P(j_1, C_2) = \sum_{j_1=1}^{6} P(j_1, 5) = 0.03758.$$

On the same hand, the total blocking probability in the two-link system is determined by:

$$P'_b = \frac{\lambda_1}{\lambda_1+\lambda_2} P'_{b_1} + \frac{\lambda_2}{\lambda_1+\lambda_2} P'_{b_2} \overset{\lambda_1=4,\lambda_2=2}{=} 0.08166.$$

Based on the approximate method of link independence and Eqs. (6), (7), we have:
$P_{b_1} = P_1(C_1)P_2(j_2 \geq \lfloor th_{12}C_2 \rfloor) = P_1(6)P_2(j_2 \geq 1) = 0.11716 \times 0.862386 \Rightarrow P'_b = 0.10104.$
$P_{b_2} = P_2(C_2)P_1(j_1 \geq \lfloor th_{11}C_1 \rfloor) = P_2(5)P_1(j_1 \geq 1) = 0.03670 \times 0.979405 \Rightarrow P'_b = 0.03594.$
The total blocking probability in the two-link system is determined by:

$$P_b = \frac{\lambda_1}{\lambda_1+\lambda_2} P_{b_1} + \frac{\lambda_2}{\lambda_1+\lambda_2} P_{b_2} \overset{\lambda_1=4,\lambda_2=2}{=} 0.07934.$$

The previous results reveal that the approximate method provides quite satisfactory results compared to the exact values, even in small tutorial examples.

References

[1] M. Stasiak, M. Głąbowski, A. Wiśniewski, and P. Zwierzykowski, *Modeling, and Dimensioning of Mobile Networks: From GSM to LTE*, Chichester, UK: Wiley & Sons, 2010 (doi: 10.1002/9780470976036.ch3).

[2] M. Stasiak and M. Głąbowski, "A simple approximation of the link model with reservation by a one–dimensional Markov chain", *Performance Evaluation*, vol. 41, no. 2–3, pp. 195–208, 2000.

[3] I. Moscholios, M. Logothetis, and G. Kokkinakis, "Connection Dependent Threshold Model: A Generalization of the Erlang Multiple Rate Loss Model", *Performance Evaluation*, vol. 48, no. 1–4, pp. 177–200, 2002.

[4] M. Głąbowski and M. Stasiak, "Point-to-point blocking probability in switching networks with reservation", *Annals of Telecommunications*, vol. 57, no. 7–8, pp. 798–831, 2002.

[5] S. Rácz, B. Gerő, and G. Fodor, "Flow level performance analysis of a multi–service system supporting elastic, and adaptive services", *Performance Evaluation*, vol. 49, no. 1–4, pp. 451–469, 2002.

[6] M. Głąbowski and M. Stasiak, "Multi-rate model of the limited availability group with finite source population", in *Proc. 10th Asia–Pacific Conf. on Commun.*, Beijing, China, 2004.

[7] I. Moscholios and M. Logothetis, "Engset Multirate State-Dependent Loss Models with QoS Guarantee", *Int. J. of Commun. Systems*, vol. 19, no. 1, pp. 67–93, 2006.

[8] V. Vassilakis, I. Moscholios, and M. Logothetis, "Call-level performance modelling of elastic, and adaptive service-classes with finite population", *IEICE Transact. on Commun.*, vol. E91–B, no. 1, pp. 151–163, 2008.

[9] Q. Huang, King-Tim Ko, and V. Iversen, "Approximation of loss calculation for hierarchical networks with multiservice overflows", in *IEEE Transact. on Communic.*, vol. 56, no. 3, pp. 466–473, 2008.

[10] M. Stasiak, M. Sobieraj, J. Weissenberg, and P. Zwierzykowski, "Analytical Model of the Single Threshold Mechanism with Hysteresis for Multi-service Networks", *IEICE Transact. on Commun.*, vol. E95–B, no. 1, pp. 120–132, 2012.

[11] I. Moscholios, J. Vardakas, M. Logothetis, and M. Koukias, "A Quasi-random Multirate Loss Model supporting Elastic, and Adaptive Traffic under the Bandwidth Reservation Policy", *Int. J. on Advances in Networks and Services*, vol. 6, no. 3–4, pp. 163–174, 2013.

[12] S. Yan, M. Razo, M. Tacca, and A. Fumagalli, "A Blocking Probability Estimator for the Multi–Application, and Multi-Resource Constraint Problem", in *Proc. Int. Conf. Comp., Network., and Communic.*, Honolulu, HI, USA, 2014.

[13] I. Moscholios, M. Logothetis, J. Vardakas, and A. Boucouvalas, "Performance Metrics of a Multirate Resource Sharing Teletraffic Model with Finite Sources under both the Threshold and Bandwidth Reservation Policies", *IET Networks*, vol. 4, no. 3, pp. 195–208, 2015.

[14] Y. Huang, Z. Rosberg, K. Ko, and M. Zukerman, "Blocking Probability Approximations, and Bounds for Best-effort Calls in an Integrated Service System", *IEEE Transact. on Commun.*, vol. 63, no. 12, pp. 5014–5026, 2015.

[15] I. Moscholios, M. Logothetis, J. Vardakas, and A. Boucouvalas, "Congestion Probabilities of Elastic and Adaptive Calls in Erlang-Engset Multirate Loss Models under the Threshold and Bandwidth Reservation Policies", *Comput. Networks*, vol. 92, no. 1, pp. 1–23, 2015.

[16] M. Głąbowski and M. Sobieraj, "Analytical modelling of multiservice switching networks with multiservice sources, and resource management mechanisms", accepted for publication in *Telecommunication Systems*, 2017 (doi:10.1007/s11235-017-0305-4).

[17] I. Widjaja and H. Roche, "Sizing X2 bandwidth for Inter-connected eNBs", in *Proc. IEEE VTC Fall*, Anchorage, AK, USA, 2009.

[18] M. Stasiak, P. Zwierzykowski, and D. Parniewicz, "Modelling of the WCDMA interface in the UMTS network with Soft Handoff Mechanism", in *Proc. IEEE GLOBECOM*, Honolulu, HI, USA, 2009.

[19] B. Renard, S. Elayoubi, and A. Simonian, "A dimensioning method for the LTE X2 interface", in *Proc. IEEE WCNC*, Shanghai, China, 2012.

[20] M. Stasiak, D. Parniewicz, and P. Zwierzykowski, "Traffic Engineering for Multicast Connections in Multiservice Cellular Network", *IEEE Transactions on Industrial Informatics*, vol. 9, no. 1, pp. 262–270, 2013.

[21] I. Moscholios, G. Kallos, M. Katsiva, V. Vassilakis, and M. Logothetis, "Call Blocking Probabilities in a W–CDMA cell with interference cancellation, and bandwidth reservation", in *Proc. IEICE ICTF*, Poznań, Poland, 2014.

[22] I. Moscholios, G. Kallos, V. Vassilakis, and M. Logothetis, "Congestion Probabilities in CDMA–based networks supporting batched Poisson input traffic", *Wireless Person. Commun.*, vol. 79, no. 2, pp. 1163–1186, 2014.

[23] M. Khedr, and R. Makki Hassan, "Opportunistic call admission control for wireless broadband cognitive networks", *Wireless Networks*, vol. 20, no. 1, pp. 105–114, 2014.

[24] A. Machado de Medeiros, and M. Yacoub, "BlockOut: Blocking, and Outage in a Single Performance Measure", *IEEE Transact. on Vehic. Technol.*, vol. 63, no. 7, pp. 3451–3456, 2014.

[25] V. Burger, M. Seufert, T. Hossfeld, and P. Tran-Gia, "Performance evaluation of backhaul bandwidth aggregation using a partial sharing scheme", *Physic. Communic.*, vol. 19, pp. 135–144, 2016.

[26] I. Moscholios, V. Vassilakis, M. Logothetis, and A. Boucouvalas, "A Probabilistic Threshold–based Bandwidth Sharing Policy for Wireless Multirate Loss Networks", *IEEE Wireless Commun. Let.*, vol. 5, no. 3, pp. 304–307, 2016.

[27] V. Vassilakis, I. Moscholios, and M. Logothetis, "Uplink Blocking Probabilities in Priority–Based Cellular CDMA Networks with Finite Source Population", *IEICE Transact. on Commun.*, vol. E99–B, no. 6, pp. 1302–1309, 2016.

[28] V. Vassilakis, I. Moscholios, and M. Logothetis, "Quality of Service Differentiation of Elastic, and Adaptive Services in CDMA Networks: A Mathematical Modelling Approach", accepted for publication in *Wireless Networks*, 2017 (doi:10.1007/s11276-016-1411-z).

[29] Z. Wang, P. Mathiopoulos, and R. Schober, "Performance Analysis and Improvement Methods for Channel Resource Management Strategies of LEO-MSS with Multiparty Traffic", *IEEE Transact. on Vehic. Technol.*, vol. 57, no. 6, pp. 3832–3842, 2008.

[30] D. Yiltas and A. Zaim, "Evaluation of call blocking probabilities in LEO satellite networks", *Int. J. of Satel. Communic.*, vol. 27, no. 2, pp. 103–115, 2009.

[31] Z. Wang, P. Mathiopoulos, and R. Schober, "Channel partitioning policies for multi-class traffic in LEO-MSS", *IEEE Trans. on Aerosp., and Electr. Sys.*, vol. 45, no. 4, pp. 1320–1334, 2009.

[32] J. Vardakas, I. Moscholios, M. Logothetis, and V. Stylianakis, "An Analytical Approach for Dynamic Wavelength Allocation in WDM-TDMA PONs Servicing ON-OFF Traffic", *IEEE/OSA J. of Optic. Communic. and Network.*, vol. 3, no. 4, pp. 347–358, 2011.

[33] Y. Deng and P. Prucnal, "Performance analysis of heterogeneous optical CDMA networks with bursty traffic, and variable power control", *IEEE/OSA J. of Optic. Commun. and Network.*, vol. 3, no. 6, pp. 487–492, 2011.

[34] J. Vardakas, I. Moscholios, M. Logothetis, and V. Stylianakis, "On Code reservation in Multi-rate OCDMA Passive Optical Networks", in *Proc. 8th IEEE Int. Symp. on Commun. Sys., Networks & Digit. Signal Process. CSNDSP*, Poznań, Poland, 2012.

[35] J. Vardakas, I. Moscholios, M. Logothetis, and V. Stylianakis, "Performance Analysis of OCDMA PONs Supporting Multi–Rate Bursty Traffic", *IEEE Transact. on Commun.*, vol. 61, no. 8, pp. 3374–3384, 2013.

[36] V. Casares-Giner, "Some teletraffic issues in optical burst switching with burst segmentation", *Electr. Let.*, vol. 52, no. 11, pp. 941–943, 2016.

[37] Y. Guan, H. Jiang, M. Gao, S. Bose, and G. Shen, "Migrating Elastic Optical Networks from Standard Single-Mode Fibers to Ultra-Low Loss Fibers: Strategies, and Benefits", in *Proc. Optic. Fiber Communic. Conf.*, Los Angeles, CA, USA, 2017.

[38] S. Pantelis, I. Moscholios, and S. Papadopoulos, "Call-level evaluation of a two-link single rate loss model for Poisson traffic", in *Proc. IEICE ICTF*, Poznań, Poland, 2017.

[39] L. Mamatas, I. Psaras, and G. Pavlou, "Incentives, and Algorithms for Broadband Access Sharing", in *Proc. ACM SIGCOMM Workshop on Home Networks*, New Delhi, India, 2010.

[40] I. Psaras and L. Mamatas, "On demand connectivity sharing: Queuing management, and load balancing for user-provided networks", *Comput. Networks*, vol. 55, no. 2, pp. 399–414, 2011.

[41] BeWi-Fi, October 2017 [Online]. Available: http://www.tid.es/research/areas/beWi-Fi

[42] J. Kaufman, "Blocking in a shared resource environment", *IEEE Transact. on Communic.*, vol. 29, no. 10, pp. 1474–1481, 1981.

[43] J. Roberts, "A service system with heterogeneous user requirements", in: G. Pujolle (Ed.), *Performance of Data Communications systems, and their applications*, Amsterdam, North Holland, the Netherlands, pp. 423–431, 1981.

[44] V. Iversen, "The exact evaluation of multi-service loss system with access control", *Teleteknik*, vol. 31, no. 2, pp. 56–61, 1987.

[45] D. Tsang and K. Ross, "Algorithms to determine exact blocking probabilities for multirate tree networks", *IEEE Transact. on Commun.*, vol. 38, no. 8, pp. 1266–1271, 1990.

[46] Q. Huang, K. Ko, and V. Iversen, "Approximation of loss calculation for hierarchical networks with multiservice overflows", *IEEE Transact. on Commun.*, vol. 56, no. 3, pp. 466–473, 2008.

[47] M. Głąbowski, A. Kaliszan, and M. Stasiak, "Asymmetric convolution algorithm for full–availability group with bandwidth reservation", in *Proc. Asia–Pacific Conf. on Commun.*, Busan, South Korea, 2006.

[48] M. Głąbowski, A. Kaliszan, and M. Stasiak, "Asymmetric convolution algorithm for blocking probability calculation in full–availability group with bandwidth reservation", *IET Circuits, Devices & Systems*, vol. 2, no. 1, pp. 87–94, 2008.

[49] M. Głąbowski, A. Kaliszan, and M. Stasiak, "Two-dimensional convolution algorithm for modelling multiservice networks with overflow traffic", *Mathemat. Problems in Engin.*, vol. 2013, Article ID 852082, 18 pages.

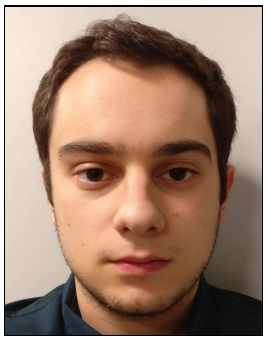
[50] I. Moscholios, "Call blocking probabilities in an Erlang Multirate Loss Model under a State–dependent Threshold Policy", in *Proc. of IEICE Infor. and Communic. Technol. Forum ICTF*, Patras, Greece, 2016.

[51] S. Hanczewski, A. Kaliszan, and M. Stasiak, "Convolution model of a queueing system with the cFIFO service discipline", *Mob. Infor. Sys.*, vol. 2016, Article ID 2185714, 15 pages.

[52] S. Sagkriotis and I. Moscholios, "Evaluation of convolution algorithms in the Erlang Multirate Loss Model under the bandwidth reservation policy", in *Proc. IEICE ICTF*, Poznań, Poland, 2017.

[53] I. Moscholios, V. Vassilakis, M. Logothetis, and A. Boucouvalas, "State–dependent Bandwidth Sharing Policies for Wireless Multirate Loss Networks", *IEEE Transact. on Wireless Commun.*, vol. 16, no. 8, pp. 5481–5497, 2017.

[54] Simscript III, October 2017 [Online]. Available: http://www.simscript.com



Stefanos G. Sagkriotis received his B.Sc. in Informatics and Telecommunications from the University of Peloponnese, Tripolis, Greece. Currently he is an M.Sc. student at the School of Computing Science, University of Glasgow, Glasgow, United Kingdom. His research interests include teletraffic engineering and network performance evaluation.

E-mail: : 2353747s@student.gla.ac.uk
School of Computing Science
University of Glasgow
Glasgow, United Kingdom

Spiridon K. Pantelis is an undergraduate student in the Department of Informatics and Telecommunications, University of Peloponnese, Greece. His research interests include teletraffic engineering and network performance evaluation.

E-mail: tst12081@uop.gr
Department of Informatics and Telecommunications
University of Peloponnese
Tripolis, Greece

Ioannis D. Moscholios received his D.Eng. degree in Electrical and Computer Engineering from the University of Patras, Patras, Greece, in 1999, M.Sc. degree in Spacecraft Technology and Satellite Communications from the University College London, UK, in 2000 and Ph.D. degree in Electrical and Computer Engineering from the University of Patras, in 2005. From 2005 to 2009 he was a Research Associate at the Wire Communications Laboratory, Department of Electrical and Computer Engineering, University of Patras. From 2009 to 2013 he was a Lecturer in the Department of Telecommunications Science and Technology, University of Peloponnese, Tripolis, Greece. Currently, he is an Assistant Professor in the Department of Informatics and Telecommunications, University of Peloponnese. His research interests include teletraffic engineering, simulation and performance analysis of communication networks. He is an IARIA Fellow and a member of the Technical Chamber of Greece.

E-mail: idm@uop.gr
Department of Informatics and Telecommunications
University of Peloponnese
Tripolis, Greece

Vassilios G. Vassilakis is a lecturer in the Department of Computer Science at the University of York, UK. He received his Ph.D. degree in Electrical and Computer Engineering from the University of Patras, Greece, in 2011. From 2011 to 2013, he was with the Network Convergence Laboratory, University of Essex, where he conducted research on information-centric networking and network security, and contributed to the EC FP7 PURSUIT project. In 2013, he joined the Institute for Communication Systems, University of Surrey, and conducted research on 5G wireless networks. After that, he was with the Computer Laboratory, University of Cambridge, where he conducted research on future Internet technologies and contributed to the EC H2020 RIFE project. In 2015, he joined the Secure and Dependable Software Systems Research Cluster at the University of Brighton, where he conducted research on 5G network security and contributed to the EC H2020 SESAME project. His main research interests are in the areas of 5G wireless and mobile networks, future Internet technologies, network security, software-defined networks, and IoT. He has published over 80 journal/conference papers.

E-mail: vv274@cl.cam.ac.uk
Department of Computer Science
University of York
York, United Kingdom

Paper

Rearrangeability of 2×2 W-S-W Elastic Switching Fabrics with Two Connection Rates

Wojciech Kabaciński, Remigiusz Rajewski, and Atyaf Al-Tameemi

Faculty of Electronics and Telecommunications, Poznan University of Technology, Poznań, Poland

https://doi.org/10.26636/jtit.2018.123417

Abstract—The rearreangeable conditions for the 2×2 three-stage switching fabric of a W-S-W architecture for elastic optical switches are considered in this paper. Analogies between the switching fabric considered and the three-stage Clos network are shown. On the other hand, differences are also shown, which presented the modifications required in the control algorithm used in rearrangeable networks. The rearrangeable conditions and the control algorithm are presented and proved. Operation of the proposed control algorithm is shown based on a few examples. The required number of frequency slot units in interstage links of rearrangeable switching fabrics is much lower than in the strict-sense non-blocking switching fabrics characterized by the same parameters.

Keywords—elastic optical networks, elastic optical switching nodes, interconnection networks, rearrangeable non-blocking conditions

1. Introduction

The Elastic Optical Network (EON) architecture has been proposed to utilize the bandwidth available in optical fiber more efficiently. By breaking the fixed-grid spectrum allocation limit of conventional Wavelength Division Multiplexing (WDM) networks, EONs increase flexibility in connection provisioning [1], [2]. To do so, depending on the traffic volume, an appropriately sized optical spectrum is allocated to the connections in EON. This optical spectrum is referred to as the Frequency Slot Unit (FSU).

Furthermore, unlike the rigid optical channels of conventional WDM networks [3], a light-path can expand or contract elastically to meet different bandwidth demands in EON. In this way, the incoming connection request can be served in a spectrum-efficient manner. This technological advance poses additional challenges on the networking level, especially in terms of efficient establishment of the connection.

Similarly to WDM networks, an elastic optical connection must occupy the same spectrum portion between its end nodes, that is, ensuring the so-called spectrum continuity constraint. However, when wavelength conversion (or spectrum conversion) is introduced in WDM (EON) networks, blocking probability is significantly reduced. In addition, in EONs, the entire bandwidth of each connection must be contiguously allocated. Bandwidth assigned to an optical channel depends on the required transmission data rate, distance to be covered, path-quality, wavelength spacing between channels, and/or the modulation scheme used [2], [4]–[6].

Several architectures of elastic optical switching nodes were proposed in literature [7]–[10]. In this paper, we deal with one of these switching fabric architectures, i.e. the W-S-W (wavelength-space-wavelength) switching fabric, called the WSW1 [11]. Strict-sense non-blocking (SSNB) conditions for the WSW1 architecture have been proved in [11] as well. We proposed rearrangeable non-blocking (RNB) conditions for this architecture in [12] for simultaneous routing of connections with a limited number of connection rates. The term *simultaneous connections* means that all connections arrive at the same time at all inputs, and must be served simultaneously.

Simultaneous connections can be routed using the modified matrix decomposition algorithm. Several such algorithms were proposed in literature, for instance the Neiman's algorithm [13], which consists of a relatively simple iteration phase followed by a relatively complex iterative phase. The latter is necessary only if the matrix cannot be decomposed completely after using phase one. One of the modifications to phase one of Neiman's algorithm was proposed for instance in [14]. Neiman's algorithm is used to route connections simultaneously in the three-stage Clos switching fabric [15]. The three-stage Clos network consists of two outer stages of rectangular switches, and of an inner stage of square switches. The WSW1 switching fabric can be modeled by the Clos network, as it will be shown later in this paper. However, we cannot use the same routing algorithms which are used for the three-stage Clos switching networks directly in the WSW1 switching fabric, for reasons mentioned in [12].

In our model, the number of simultaneous connection rates that can be served is limited to z. The upper bound for RNB connections when $r > 2$ was derived in [12]. The aim for using the RNB switching fabric is to reduce the required number of FSUs in the interstage links, i.e. to reduce the cost of this switching fabric. In this paper, we improve the result presented in [16]. The necessary and sufficient RNB conditions have been derived in [12] for the special case

when $r = z = 2$ and $\frac{n}{m_1}, \frac{n}{m_2}$, and $\frac{m_2}{m_1}$ are integers. In [16], we generalized these conditions to the general case, when $r = z = 2$ and for any values of n, m_1, and m_2. We also showed that after applying the decomposition algorithm mentioned in [12], we will get a set of connections that can be set up through the interstage links.

The main idea of this paper is to propose merge operation for matrices, and to show how to calculate the required number of FSUs in the interstage links. We also aim to present how to determine the sequence of matrices that can be merged together, satisfying the condition that each connection must use adjacent FSUs.

The remaining portions of the paper are organized as follows. In the next section, the WSW1 switching fabric is presented and the problem is described in a more detailed way. The connection model and its representation are presented as well. In Section 3, the RNB results for the proposed model are derived and proved. In Section 4 examples of the algorithm's operation are presented. The paper ends with conclusions.

2. Switching Fabrics and the Model Used

The WSW1 switching fabric considered in this paper was described in more detail in [11]. Here, we will only provide a short description which will make the paper easier to follow. This architecture is presented in Fig. 1. In the first and third stages, there are r Bandwidth-Variable Wavelength converting Switches (BV-WSs), and one Bandwidth-Variable wavelength selective Space Switch (BV-SS) of capacity $r \times r$ is in the second stage. Each BV-WS in the first stage has one input fiber with n FSUs and one output fiber with k FSUs, while each BV-WS in the third stage has one input fiber with k FSUs and one output fiber with n FSUs. The internal architecture of BV-WSs and BV-SS can be found in [11]. The switching fabric serves m-slot connections, FSUs in input/output fibers are numbered from 1 to n, BV-WSs in both input and output stages are numbered from 1 to r, and FSUs in interstage fibers are numbered from 1 to k (see Fig. 1).

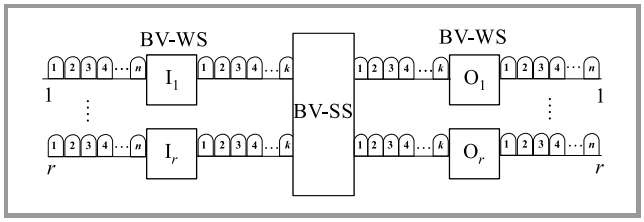


Fig. 1. The WSW1 switching fabric architecture.

In the presented considerations we assumed that BV-WSs have full range conversion capability, i.e. an m-slot connection which uses a set of m adjacent FSUs in the input fiber can be switched to a set of any other m adjacent FSUs in the output fiber. A new m-slot connection from input switch I_i to output switch O_j will be denoted by (I_i, O_j, m). When the numbers of FSUs occupied by this connection are important, the number of the first FSU will be also provided. Thus, $(I_i[x], O_j[y], m)$ denotes the m-slot connection in the input fiber of switch I_i which occupies FSUs from x to $x+m-1$, and FSUs from y to $y+m-1$ of output fiber of switch O_j. In the switching fabric, when a new connection (I_i, O_j, m) arrives, a control algorithm must find a set of m adjacent FSUs in interstage links, which can be used for this connection, and these must be FSUs with the same numbers in the interstage links from I_i and to O_j, since BV-SS has no spectrum conversion capability. In the case of the simultaneous connection model, we have a set of compatible connection requests which occupy most of FSUs in the input and output fibers, i.e. the number of free FSUs in each input/output fiber is less than m_1. This set of connections is denoted by $\mathbb{C}$ and is divided into two different types of connections: m_1 and m_2.

Example 1. Let us introduce a simple example. The set of connection requests $\mathbb{C}$ for the switching fabric of capacity 2×2 with $n = 13$ consists of eight connections (see Fig. 2). These connections are divided into two types: with $m_1 = 2$ and $m_2 = 5$. There are three m_2-slot connections and five m_1-slot connections. Additionally, one FSU remains free in input fiber no. 1.

The exact mechanism of routing these connections in the WSW1 switching fabric will be explained in detail later, in Section 3. The problem now is which FSUs in interstage

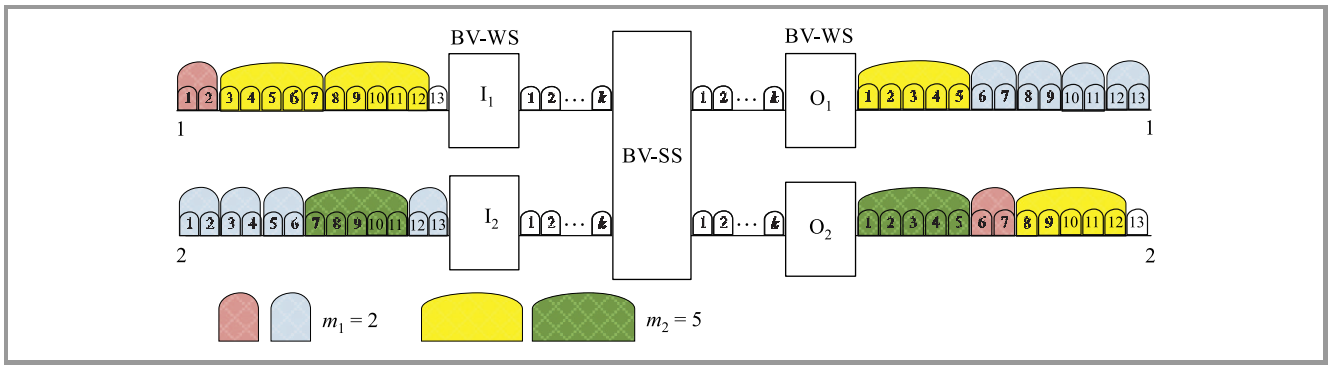


Fig. 2. The 2×2 WSW1 switching fabric with $\mathbb{C} = \{(I_1[1], O_2[6], 2); (I_1[3], O_1[1], 5); (I_1[8], O_2[8], 5); (I_2[1], O_1[6], 2); (I_2[3], O_1[8], 2); (I_2[5], O_1[10], 2); (I_2[7], O_1[1], 5); (I_2[12], O_1[12], 2)\}$.

links should be used by these connections, and how many FSUs are needed to set up all these connections, i.e. when the switching fabric is RNB. In [12] we proposed a control algorithm to assign FSUs to particular connection requests using the matrix decomposition algorithm, and showed the RNB conditions when $\frac{n}{m_1}$, $\frac{n}{m_2}$, and $\frac{m_2}{m_1}$ are integers. The case with any number of m_1, m_2, and n is considered in Section 3.

The WSW1 switching fabric could be represented by the three-stage Clos network shown in Fig. 3. The Clos equivalent of the WSW1 switching fabric shown in Fig. 4 is presented in Fig. 5. The space switches in the first stage of the Clos network correspond to the first stage switches in the WSW1 switching fabric. Similarly, the space switches

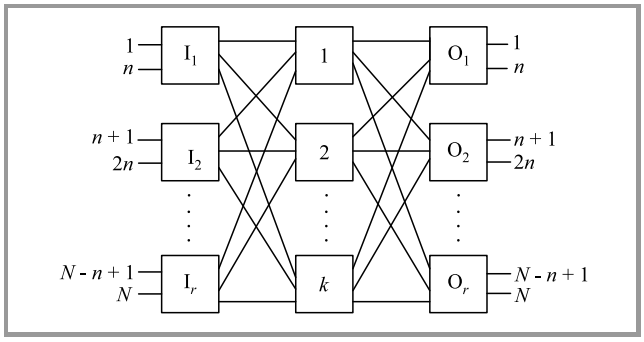


Fig. 3. Three-stage Clos network architecture.

in the third stage of the Clos network correspond to the third stage switches in the WSW1 switching fabric. Each FSU in the input fiber of the WSW1's switch I_i is represented by one input of switch I_i in the Clos network. Similarly, each FSU in the output fiber of the WSW1's switch O_j is represented by one output of switch O_j in the Clos network. In interstage links of the WSW1 fabric, each FSU corresponds to one center stage switch in the Clos network. Therefore, we have k switches in the center stage. The Clos network with these parameters can be as $C(k,n,r)$. It is known that if $k \geq n$ the Clos network is rearrangeable and if $k \geq 2n-1$ — it is strictly non-blocking [15]. The number of inputs and outputs to the Clos network is $N = nr$.

A matrix decomposition algorithm starts by deriving the $\mathbf{H}_n$ matrix of size $r \times r$, where each element $\mathbf{H}_n[i,j]$ denotes the number of connection requests at input switch I_i which are directed to output switch O_j. Because each first stage switch has n inputs, the sum of the entries in each row is n,

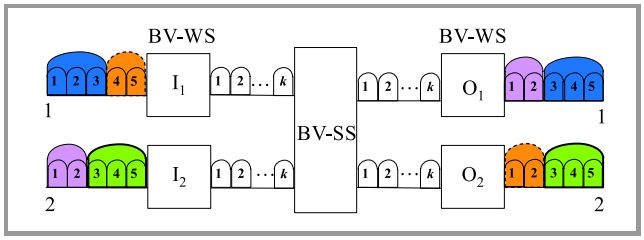


Fig. 4. 2 × 2 WSW1 switching fabric with $\mathbb{C} = \{(I_1[1], O_1[3], 3); (I_1[5], O_2[1], 2); (I_2[1], O_1[1], 2); (I_2[3], O_2[3], 3)\}$.

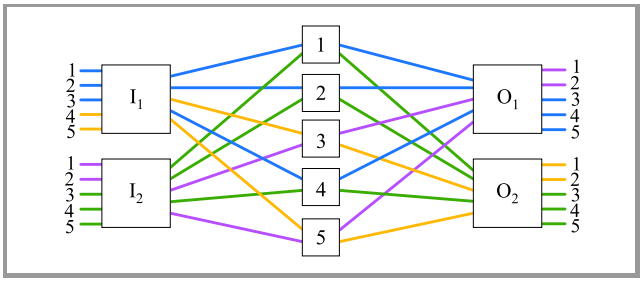


Fig. 5. Three-stage Clos network architecture with $\mathbb{C} = \{(I_1[1], O_1[3], 3); (I_1[5], O_2[1], 2); (I_2[1], O_1[1], 2); (I_2[3], O_2[3], 3)\}$. (See color pictures online at www.nit.eu/publications/journal-jtit)

and since each last stage switch has n outputs, the sum of the entries in each column is also n.

Let us consider the WSW1 switching fabric in Fig. 4. This switching fabric serves 4 connections which occupy 2 or 3 FSUs. In Fig. 5, this WSW1 is modeled as a three-stage Clos network. In Fig. 4, we used different colors to recognize these connections. The connection marked with a solid line (blue connection) occupies 3 adjacent FSUs from I_1 to O_1, and it is represented in Fig. 5 by solid lines. Similarly, the connection marked with a dashed line (orange connection), which occupies 2 adjacent FSUs from I_1 to O_2, is marked orange in Fig. 5. Other connections from I_2 are represented in the same way. These connections can be represented by the connection matrix $\mathbf{H}_5 = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$.

According to Neman's algorithm [13], this matrix can be decomposed into 5 permutation matrices: $\mathbf{P}_1$, $\mathbf{P}_2$, $\mathbf{P}_3$, $\mathbf{P}_4$, and $\mathbf{P}_5$. Each permutation matrix represents one switch from the middle stage of the Clos network (see Fig. 5). FSUs belonging to connections represented by $\mathbf{P}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ are set up through the first switch from the middle stage (or first FSUs in interstage links in the WSW1 switching fabric), FSUs belonging to $\mathbf{P}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ are set up through the second switch, and so on for $\mathbf{P}_3 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $\mathbf{P}_4 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, and $\mathbf{P}_5 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. As a result, the blue connection which occupies 3 adjacent FSUs in the input link, is set up through 1st, 2nd, and 4th FSUs in the WSW1's interstage link. However, this is not correct, since these FSUs are not adjacent to each other. In general, the problem of routing connections in the WSW1 structure looks similar to routing connections in the three-stage Clos network. However, some important differences include the following:

- instead of finding connections which can be set up though one center stage switch, we have to find connections which can be set up using the same set of FSUs in the interstage link,
- connections which occupy several FSUs must use adjacent FSUs.

3. Rearrangeability Conditions

We consider 2×2 WSW1 switching fabric with the number of connection rates limited to 2, i.e. there are only m_x-slot connections, where $x = 1,2$. A set of compatible connections in $\mathbb{C}$ is represented by $\mathbf{H}^{m_x}$ matrices:

$$\mathbf{H}^{m_x} = \begin{bmatrix} h_{11}^{m_x} & h_{12}^{m_x} \\ h_{21}^{m_x} & h_{22}^{m_x} \end{bmatrix}, \tag{1}$$

where $h_{ij}^{m_x}$ is equal to the number of m_x-slot connection requests from switch I_i to switch O_j. According to Algorithm 1 given in [12], the $\mathbf{H}^{m_x}$ matrix can be decomposed into $c_{\max}^{m_x}$ permutation matrices $\mathbf{P}_i^{m_x}$, where $c_{\max}^{m_x}$ represents the maximum number of m_x-slot connections in one input or output, while $c_{\min}^{m_x}$ represents the minimum number of such connections. We can use this algorithm to set up the set of connection requests given in Example 1 (see Fig. 2). In Table 1, steps of decomposition of the given set of connection requests are presented one by one. In the first row, connection matrices $\mathbf{H}^{m_1}$ and $\mathbf{H}^{m_2}$ are given. In the next rows, matrices that result from decomposition of $\mathbf{H}^{m_1}$ and $\mathbf{H}^{m_2}$ matrices are presented. For each decomposed matrix, the number of assigned FSUs are provided in second and fourth columns. As can be noticed, the number of occupied FSUs in interstage links is 18. After applying the decomposition algorithm, it is time to commence the merging operation which can reduce the required number of FSUs in the intestage links to 14. The final arrangement for the 8 connections mentioned is shown in Fig. 6.

Table 1
Assignment of FSUs to connections used in Example 1

Matrices representing m_1-slot connections	FSUs in interstage links	Matrices representing m_2-slot connections	FSUs interstage links
$\mathbf{H}^{m_1} = \begin{bmatrix} 0 & 4 \\ 1 & 0 \end{bmatrix}$	—	$\mathbf{H}^{m_2} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$	—
$\mathbf{P}_1^{m_1} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$	1–2	$\mathbf{P}_1^{m_2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	9–13
$\mathbf{P}_2^{m_1} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$	3–4	$\mathbf{P}_2^{m_2} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$	14–18
$\mathbf{P}_3^{m_1} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$	5–6	—	—
$\mathbf{P}_4^{m_1} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$	7–8	—	—

In [12], we proved that the WSW1 switching fabric presented in Fig. 1 with $r = 2$ is rearrangeably non-blocking when $m \in \{m_1; m_2\}$ if and only if:

$$k \geq n, \tag{2}$$

where $m_1 < m_2$, and $\frac{n}{m_1}$, $\frac{n}{m_2}$, and $\frac{m_2}{m_1}$ are integers. In this article we propose a new theorem to find value k that makes the WSW1 switching fabric RNB for a scenario of a more general nature.

Theorem 1: The WSW1 switching fabric presented in Fig. 1 with $r = 2$ is rearrangeably non-blocking for m-slot connections, where $m \in \{m_1; m_2\}$, if:

$$k \geq \left\lfloor \frac{n}{m_2} \right\rfloor \cdot m_2 + \left(\left\lfloor \frac{n}{m_1} \right\rfloor - \left\lfloor \frac{n}{m_2} \right\rfloor \cdot \left\lfloor \frac{m_2}{m_1} \right\rfloor \right) \cdot m_1. \tag{3}$$

Proof: Let $\mathbb{C}$ denote a set of compatible connections. We have two connection rates, m_1 and m_2, and all connections can be represented by $\mathbf{H}^{m_1}$ and $\mathbf{H}^{m_2}$ matrices. According to the decomposition algorithm given in [12], $\mathbf{H}^{m_1}$ and $\mathbf{H}^{m_2}$ can be decomposed into $c_{\max}^{m_1}$ and $c_{\max}^{m_2}$ permutation matrices $\mathbf{P}^{m_x}$, respectively. Each $\mathbf{P}^{m_x}$ matrix represents a set of m_x-slot connections which can be set up using the same m_x FSUs in interstage links. From these $\mathbf{P}^{m_x}$ matrices, only $c_{\min}^{m_1}$ and $c_{\min}^{m_2}$ matrices contain exactly one value 1 per each row and each column. Other $\left(c_{\max}^{m_1} - c_{\min}^{m_1}\right)$ matrices $\mathbf{P}^{m_1}$ and $\left(c_{\max}^{m_2} - c_{\min}^{m_2}\right)$ matrices $\mathbf{P}^{m_2}$ contain some rows and/or columns with 0 values only. Permutation matrices $\mathbf{P}^{m_1}$ with 0s only in certain row(s) or column(s) can be merged with matrices $\mathbf{P}^{m_2}$ with 0s only, but in other row(s) or column(s). In this case, at most $\left(c_{\max}^{m_2} - c_{\min}^{m_2}\right)$ matrices $\mathbf{P}^{m_2}$ can be merged with at most $\left(c_{\max}^{m_1} - c_{\min}^{m_1}\right)$ matrices $\mathbf{P}^{m_1}$. The required number of FSUs in interstage links k, which allows to set up all connections simultaneously, is given by the flowing formula:

$$\begin{aligned} k \geq & c_{\min}^{m_1} \cdot m_1 + c_{\min}^{m_2} \cdot m_2 + \left(c_{\max}^{m_2} - c_{\min}^{m_2}\right) \cdot m_2 \\ & + \left(\left(c_{\max}^{m_1} - c_{\min}^{m_1}\right) - \left(c_{\max}^{m_2} - c_{\min}^{m_2}\right) \cdot \left\lfloor \frac{m_2}{m_1} \right\rfloor \right) \cdot m_1 \end{aligned}. \tag{4}$$

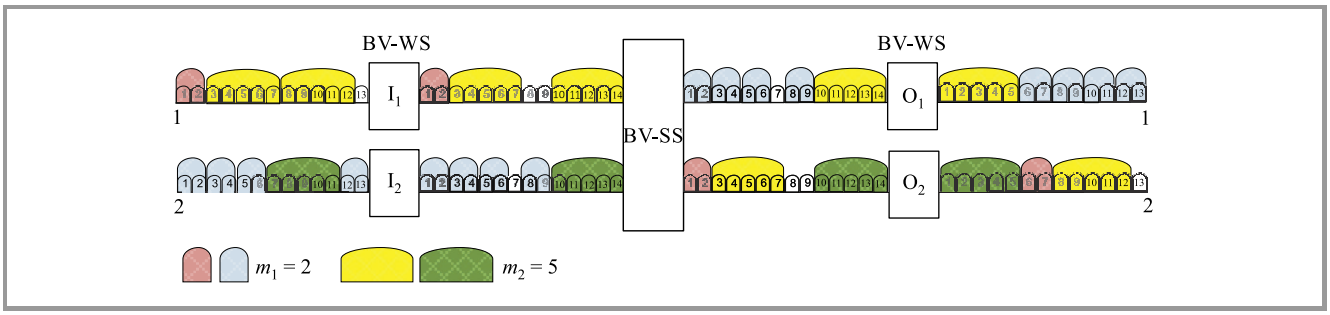


Fig. 6. 2×2 WSW1 switching fabric with set $\mathbb{C}$ of requested connections (see Fig. 1) set up with the number of FSUs decreased due to the merging operation.

Equation (4) can be simplified into the following:

$$k \geq c_{\max}^{m_2} \cdot m_2 + \left(c_{\max}^{m_1} - \left(c_{\max}^{m_2} - c_{\min}^{m_2} \right) \cdot \left\lfloor \frac{m_2}{m_1} \right\rfloor \right) \cdot m_1 . \quad (5)$$

Equation (5) must be maximized through all possible sets $\mathbb{C}$. Since $c_{\max}^{m_x}$ represents the maximum number of m_x-slot connections in one of the inputs or outputs, the number of such connections in one input/output will never be greater than $\left\lfloor \frac{n}{m_x} \right\rfloor$. When $c_{\max}^{m_x}$ value is maximized, the value of $c_{\min}^{m_x}$ is minimized. When we put $c_{\max}^{m_x} = \left\lfloor \frac{n}{m_x} \right\rfloor$ and $c_{\min}^{m_x} = 0$ to Eq. (5) we get:

$$k \geq \left\lfloor \frac{n}{m_2} \right\rfloor \cdot m_2 + \left(\left\lfloor \frac{n}{m_1} \right\rfloor - \left\lfloor \frac{n}{m_2} \right\rfloor \cdot \left\lfloor \frac{m_2}{m_1} \right\rfloor \right) \cdot m_1 \quad (6)$$

which gives number of FSUs in each interstage links. ■

4. Examples of Algorithm's Operation

Let us present a few examples rendering the idea behind the proof presented clearer. The first example for the 2×2 WSW1 switching fabric was already presented in Section 2.

Example 2. In this example the WSW1 switching fabric has the following parameters: $r = 2$, $n = 12$, $z = 2$, $m_1 = 3$, $m_2 = 5$, and the set of connection requests $\mathbb{C}$ is shown in Fig. 7. All connections from the set of requested connections $\mathbb{C}$ can be represented by matrices $\mathbf{H}^{m_1} = \begin{bmatrix} 0 & 0 \\ 2 & 2 \end{bmatrix}$ and $\mathbf{H}^{m_2} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$. There are $c_{\max}^{m_1} = 4$ and $c_{\max}^{m_2} = 2$ permutation matrices for $\mathbf{H}^{m_1}$ and $\mathbf{H}^{m_2}$, respectively. The matrices that do not contain one element 1 in each row and each column can be merged together because it means that they represent connections which are at different inputs and are directed to different outputs. After decomposition, the number of permutation matrices received from $\mathbf{H}^{m_1}$ and $\mathbf{H}^{m_2}$ that cannot be merged with other matrices, is equal to $c_{\min}^{m_1} = 0$ and $c_{\min}^{m_2} = 0$, respectively. But this does not mean that all of the permutation matrices can be merged together, and this is because values of $\frac{n}{m_1}$, $\frac{n}{m_2}$, and $\frac{m_2}{m_1}$ (or at least the third value) are not integers. Generally, we can merge only $\left\lfloor \frac{m_2}{m_1} \right\rfloor$ $\mathbf{P}^{m_1}$ matrices with one matrix $\mathbf{P}^{m_2}$, since connections in $\mathbf{P}^{m_2}$ occupy m_2 FSUs, while connections in $\mathbf{P}^{m_1}$ – only m_1 FSUs. In the presented example, we have $m_2 = 5$ and $m_1 = 3$, so $\left\lfloor \frac{m_2}{m_1} \right\rfloor = 1$ and only one $\mathbf{P}^{m_1}$ can be merged with one $\mathbf{P}^{m_2}$.

In the first step of $\mathbf{H}^{m_1}$ decomposition, we get $\mathbf{P}_1^{m_1} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$, and $\mathbf{H}_1^{m_1} = \mathbf{H}^{m_1} - \mathbf{P}_1^{m_1} = \begin{bmatrix} 0 & 0 \\ 1 & 2 \end{bmatrix}$. The next permutation matrix is $\mathbf{P}_2^{m_1} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$, and $\mathbf{H}_2^{m_1} = \mathbf{H}^{m_1} - \mathbf{P}_2^{m_1} = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$. Finally, $\mathbf{H}_2^{m_1}$ can be decomposed into two equal permutation matrices $\mathbf{P}_3^{m_1} = \mathbf{P}_4^{m_1} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$.

For $\mathbf{H}^{m_2}$, the first permutation matrix is $\mathbf{P}_1^{m_2} = \mathbf{P}_2^{m_2} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, and the second permutation matrix is obtained from $\mathbf{H}_1^{m_2} = \mathbf{H}^{m_2} - \mathbf{P}_1^{m_2} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \mathbf{P}_2^{m_2}$. When no merging operation is performed, we need 22 FSUs in interstage links, i.e. four $\mathbf{P}^{m_1}$ matrices, each uses three FSUs, and two $\mathbf{P}^{m_2}$ matrices occupying five FSUs each. When two $\mathbf{P}^{m_2}$ matrices are merged with two $\mathbf{P}^{m_1}$ matrices, the number of required FSUs is reduced to 16, as shown in Fig. 7.

Example 3. In this example, most connections in I_1 are m_2-slot connections, and the rest of FSUs are used by one m_1-slot connection. In I_2, all FSUs are occupied by m_1-slot connections This switching fabric with $n = 12$, $z = 2$, $m_1 = 2$, and $m_2 = 5$, as well as the set of connection requests $\mathbb{C}$ are shown in Fig. 8. Set $\mathbb{C}$ is represented by matrices $\mathbf{H}^{m_1} = \begin{bmatrix} 1 & 0 \\ 5 & 1 \end{bmatrix}$ and $\mathbf{H}^{m_2} = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}$. The number of permutation matrices for $\mathbf{H}^{m_1}$ is $c_{\max}^{m_1} = 6$, and for $\mathbf{H}^{m_2}$ is $c_{\max}^{m_2} = 2$. After decomposition, for $\mathbf{H}^{m_1}$ we get the following set of permutation matrices:$\mathbf{P}_1^{m_1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, and $\mathbf{P}_2^{m_1} = \mathbf{P}_3^{m_1} = \mathbf{P}_4^{m_1} = \mathbf{P}_5^{m_1} = \mathbf{P}_6^{m_1} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$, while for $\mathbf{H}^{m_2}$ we get two permutation matrices: $\mathbf{P}_1^{m_2} = \mathbf{P}_2^{m_2} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. We can merge $\mathbf{P}_1^{m_2}$ with $\mathbf{P}_2^{m_1}$ and $\mathbf{P}_3^{m_1}$, and $\mathbf{P}_2^{m_2}$ with $\mathbf{P}_4^{m_1}$ and $\mathbf{P}_5^{m_1}$, to get number of FSUs in the interstage links $k = 14$ instead of $k = 22$.

Example 4. In the fourth example, all connections in I_1 are m_2-slot connections, and in I_2 we have one m_2-slot connection and the rest of FSUs are occupied by m_1-slot connections. This switching fabric with $n = 11$, $z = 2$, $m_1 = 2$, $m_2 = 5$, and the set $\mathbb{C}$ are shown in Fig. 9. It can be represented by matrices $\mathbf{H}^{m_1} = \begin{bmatrix} 0 & 0 \\ 3 & 0 \end{bmatrix}$ and $\mathbf{H}^{m_2} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. The

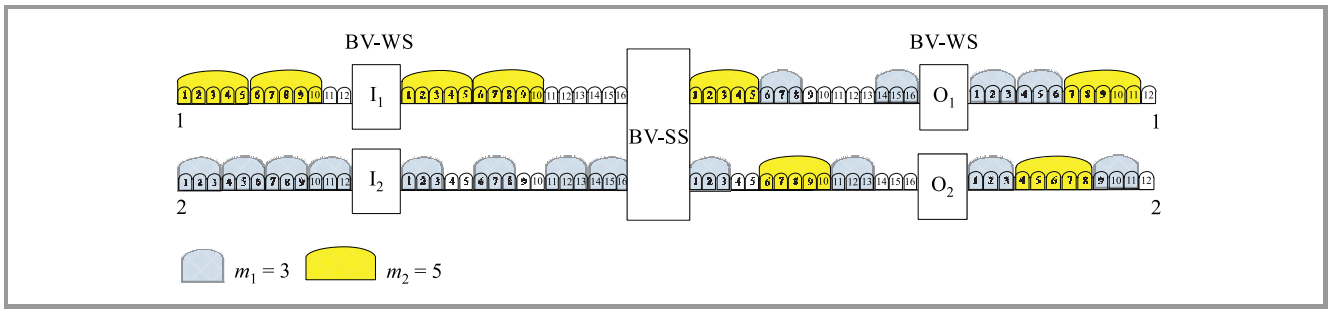


Fig. 7. 2×2 WSW1 switching fabric with $\mathbb{C} = \{(I_1[1], O_1[7], 5);\ (I_1[6], O_2[4], 5);\ (I_2[1], O_1[1], 3);\ (I_2[4], O_1[4], 3);\ (I_2[7], O_2[1], 3);\ (I_2[10], O_2[9], 3)\}$, where all connections from input 1 are of size m_2 and from input 2 are of size m_1.

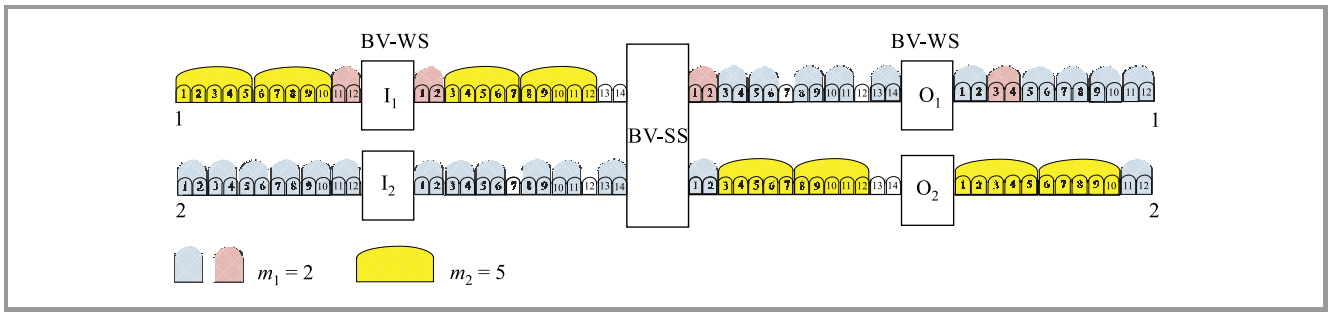


Fig. 8. 2×2 WSW1 switching fabric with $\mathbb{C}=\{(I_1[1],O_2[6],5);\ (I_1[6],O_2[1],5);\ (I_1[1],O_1[3],2);\ (I_2[1],O_1[1],2);\ (I_2[3],O_2[11],2);\ (I_2[5],O_1[1],2);\ (I_2[7],O_1[5],2);\ (I_2[9],O_1[7],2);\ (I_2[11],O_1[11],2)\}$, where all connections from input 1 are of size m_2 or m_1, and from input 2 are of size m_1.

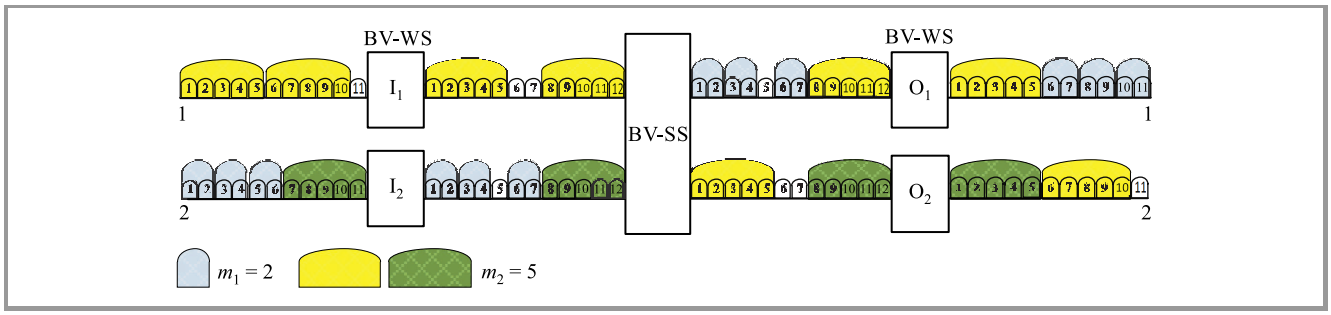


Fig. 9. 2×2 WSW1 switching fabric with $\mathbb{C}=\{(I_1[1],O_1[1],5);\ (I_1[6],O_2[6],5);\ (I_2[1],O_1[6],2);\ (I_2[3],O_1[8],2);\ (I_2[5],O_1[10],2);\ (I_2[7],O_2[1],5)\}$, where all connections from input 1 are of size m_2 and from input 2 are of size m_1 or m_2.

number of permutation matrices for $\mathbf{H}^{m_1}$ is $c_{\max}^{m_1}=3$, and for $\mathbf{H}^{m_2}$ is $c_{\max}^{m_2}=2$. After decomposition, the set of permutation matrices obtained from $\mathbf{H}^{m_1}$ is $\mathbf{P}_1^{m_1}=\mathbf{P}_2^{m_1}=\mathbf{P}_3^{m_1}=\begin{bmatrix}0&0\\1&0\end{bmatrix}$, while form $\mathbf{H}^{m_2}$: $\mathbf{P}_1^{m_2}=\begin{bmatrix}1&0\\0&1\end{bmatrix}$ and $\mathbf{P}_2^{m_2}=\begin{bmatrix}0&1\\0&0\end{bmatrix}$. In this example, we can merge matrix $\mathbf{P}_2^{m_2}$ with matrices $\mathbf{P}_1^{m_1}$ and $\mathbf{P}_2^{m_1}$. The number of FSUs in the interstage links is $k=12$ instead of $k=16$ without merging.

5. Conclusions

In this article, we considered the rearrangeability of WSW1 switching fabrics for elastic optical network nodes. Up till now, rearrangeable conditions for switching fabrics with two inputs, two outputs, two connection rates, and when $\frac{n}{m_1}$, $\frac{n}{m_2}$, and $\frac{m_2}{m_1}$ are integers, have been given. We extended these conditions to the case with any relations between values of n, m_1, and m_2. We described also a few examples which show the operation of the proposed algorithms and how merging operation results in reducing the required number of FSUs in interstage links.

Acknowledgements

The work of Wojciech Kabaciński and Remigiusz Rajewski was supported by the National Science Centre, Poland (NCN) under grant UMO-2016/21/B/ST7/02257 (ERP: 08/84/PNCN/2257), and the work of Atyaf Al-Tameemi was supported with funds of the Ministry of Science and Higher Education for the year 2017 under grant 08/82/DSMK/8222.

References

[1] V. López and L. Velasco, *Elastic Optical Networks. Architectures, Technologies, and Control*. Switzerland: Springer Int. Publishing, 2016.

[2] M. Jinno *et al.*, "Spectrum-Efficient and Scalable Elastic Optical Path Network: Architecture, Benefits, and Enabling Technologies", *IEEE Commun. Mag.*, vol. 47, no. 11, pp. 66–73, 2009.

[3] ITU-T Recommendation G.694.1. Spectral Grids for WDM Applications: DWDM Frequency Grid, 2012.

[4] O. Gerstel *et al.*, "Elastic Optical Networking: A New Dawn for the Optical Layer?", *IEEE Commun. Mag.*, vol. 50, no. 2, pp. S12–S20, 2012.

[5] I. Tomkos *et al.*, "A Tutorial on the Flexible Optical Networking Paradigm: State of the Art, Trends, and Research Challenges", *Proc. IEEE*, vol. 102, no. 9, pp. 1317–1337, 2014.

[6] R. Proietti *et al.*, "3D Elastic Optical Networking in the Temporal, Spectral, and Spatial Domains", *IEEE Commun. Mag.*, vol. 53, no. 2, pp. 79–87, 2015.

[7] P. Zhang *et al.*, "Comparison of Node Architectures for Elastic Optical Networks with Waveband Conversion", *China Commun.*, vol. 10, no. 8, pp. 77–87, 2013.

[8] Y. Chen *et al.*, "Demonstration of Petabit scalable optical switching with subband-accessibility for elastic optical networks", in *Proc. The OptoElectr. and Commun. Conf. and Austral. Conf. on Optical Fibre Tech., OECC/ACOFT 2014*, Melbourne, VIC, Australia, 2014, pp. 350–351.

[9] G. Danilewicz, W. Kabaciński, and R. Rajewski, "Strict-Sense Nonblocking Space-Wavelength-Space Switching Fabrics for Elastic Optical Network Nodes", *J. of Opt. Commun. and Netw.*, vol. 8, no. 10, pp. 745–756, 2016.

[10] W. Kabaciński, M. Michalski, and R. Rajewski, "Strict-Sense Nonblocking W-S-W Node Architectures for Elastic Optical Networks", *J. of Lightwave Tech.*, vol. 34, no. 13, pp. 3155–3162, 2016.

[11] W. Kabaciński, M. Michalski, and M. Abdulsahib, "The Strict-Sense Nonblocking Elastic Optical Switch", in *Proc. IEEE 15th Int. Conf. High Perform. Switching and Routing HPSR*, Budapest, Hungary, 2015 (doi: 10.1109/HPSR.2015.7483108).

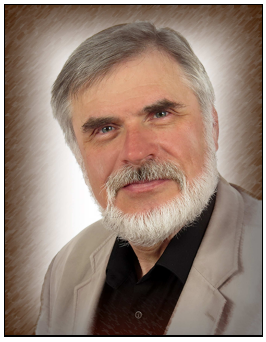
[12] W. Kabaciński, R. Rajewski, and A. Al-Tameemi, "Simultaneous Connections Routing in W-S-W Elastic Optical Switches with Limited Number of Connection Rates", in Proc. *21st Int. Conf. on Opt. Net. Design and Models ONDM*, Budapest, Hungary, 2017.

[13] V. I. Neiman, "Structure et commande optimales des réseaux de connexion sans biocage", *Annales des Telecommun.*, vol. 24, no. 7–8, pp. 232–238, 1969 (doi: 10.1007/BF02997720) [in French].

[14] N. T. Tsao, "On Neiman's Algorithm for the Control of Rearrangeable Switching Networks", in *IEEE Trans. on Commun.*, vol. 3, no. 6, 1947.

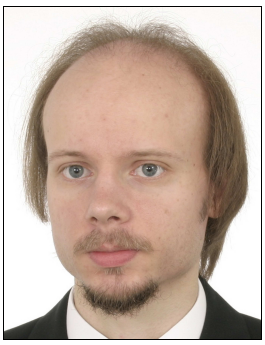
[15] C. Clos, "A study of non-blocking switching networks", in *Bell Syst. Tech. J.*, vol. 32, no. 2, pp. 406–424, 1953.

[16] W. Kabaciński, R. Rajewski, and A. Al-Tameemi, "Rearrangeability of 2x2 W-S-W Elastic Switching Fabric with Two Connection Rates", in *Proc. IEICE Infor. and Commun. Technol. Forum ICTF 2017*, Poznań, Poland, 2017.



Wojciech Kabaciński is a Professor at the Poznan University of Technology, Poland. He received his degree in Telecommunications in 1983 (with honors) from the Poznan University of Technology (PUT). In 1988 he received his Ph.D. degree (with the thesis receiving an award granted by the Ministry of National Education). In 1999 he received his D.Sc. degree, both from PUT, and in 2006 he became a Full Professor. Since 1983 he has been working at the Institute of Electronics and Telecommunications, Poznan University of Technology, where he currently is a Full Professor. He worked also as a consultant for the telecom industry. He is also a professor of College of Communications and Management. He is the author of the book titled "Non-blocking Electronic and Photonic Switching Fabrics", Springer, 2005, four other books published in Polish, over 140 papers and has 10 patents. His main research interests include: digital switching systems, photonic switching networks and systems, switching network architectures. He served as a reviewer for 5 IEEE magazines, and was also one of Guest Editors of the Feature Topic in IEEE Communications Magazine devoted to Clos switching networks. He was or currently is a member of technical program committees of many international and national conferences, symposia and workshops. Professor Kabaciński is a senior member of the IEEE Communications Society and the Association of Polish Electrical Engineers (SEP). Between 2001–2009 he served as the secretary, vice-chair and then the chair of the Communications Switching and Routing Technical Committee of the Communications Society. In 2001–2007 he was also the chair of the IEEE Communications Society Chapter in Poznań.

E-mail: wojciech.kabacinski@put.poznan.pl
Faculty of Electronics and Telecommunications
Poznan University of Technology
Poznań, Poland



Remigiusz Rajewski received his M.Sc. in Telecommunications from Poznan University of Technology (PUT), Poland, in 2006. In 2015 he received the Ph.D. degree, with honors. Since 2010 he has been with the Chair of Communication and Computer Networks, Faculty of Electronics and Telecommunications, PUT. His main research interests include: switching fabric architectures, multirate connection in switching fabrics, digital switching systems, photonic switching networks and systems, elastic optical networks, network security. He is also interested in: Linux systems (he is a Linux Academy instructor) and software (C, C++, C# and others). Rajewski was involved in several ICT projects, such as: COST Actions, BONE, Future Internet Engineering, and ALIEN. He served as a reviewer for journals: *Journal of Lightwave Technology (IEEE/OSA)*, *Journal of Optical Communications and Networking (IEEE/OSA)*, *Optical Switching and Networking (Elsevier)*, and *Computer Networks (Elsevier)*. He also served as a reviewer for many conferences. Rajewski was or currently is a member of technical program committees of numerous international conferences.

E-mail: remigiusz.rajewski@put.poznan.pl
Faculty of Electronics and Telecommunications
Poznan University of Technology
Poznań, Poland

Atyaf Al-Tameemi received her M.Sc. from the Computer Sciences and Communication Department at Arts, Sciences and Technology University (AUL) in Lebanon in 2012. From 2008 to 2014 she worked in Iraq at the University of Diyala, at the College of Sciences, Computer Science Department. Since 2014 she is a Ph.D. student at Poznan University of Technology (PUT), Faculty of Electronics and Telecommunications, Chair of Communication and Computer Networks. Her main area of interest covers elastic optical networks.

E-mail: atyaf.al-tameemi@doctorate.put.poznan.pl
Faculty of Electronics and Telecommunications
Poznan University of Technology
Poznań, Poland

Paper

Defragmentation in W-S-W Elastic Optical Networks

Remigiusz Rajewski

Faculty of Electronics and Telecommunications, Poznan University of Technology, Poznań, Poland

https://doi.org/10.26636/jtit.2018.123317

Abstract— In most cases defragmentation occurs, in elastic optical networks, in the links between the network's nodes. In this article, defragmentation in an elastic optical network's node is investigated. The W-S-W switching architecture has been used as a node. A short description of a purpose-built simulator is introduced. Several methods of defragmentation which are implemented in this simulator are described as well.

Keywords—W-S-W switching fabric, defragmentation.

1. Introduction

Nowadays, a typical optical WDM network offers enough sufficient bandwidth. However, it is highly probable that in the nearest future it will not be sufficient to handle a quickly increasing online traffic. Of course, a higher transmission speed could be used to solve that problem, but an optical path with the speed of 100 Gb/s, 400 Gb/s or even 1 Tb/s is not needed by all users. Such speeds will be used mostly by network operators inside the core network. Hence, some cost-effective and scalable solutions to convey such diverse traffic will be required. Therefore, the use of Elastic Optical Networks (EONs) has been proposed [1], enabling flexible assignment of optical bandwidth. The total optical bandwidth is divided into a lot of frequency slots, where one such frequency slot constitutes the smallest amount of optical bandwidth which can be assigned to an optical path. Therefore, any connection could demand a different number of such slots. In general, one connection demands m such slots. Currently, the slot width granularity equals 12.5 GHz, and it is referred to as a Frequency Slots Unit (FSU) [2].

EONs make bandwidth management easier. However, they offer also new challenges, such as, for instance, spectrum fragmentation. A sequence of connection and disconnection operations caused by the dynamic nature of the network's operation sooner or later results in the existence of non-aligned, isolated, and small-size blocks of spectrum segments. These segments can seldom be used for future connections. In most cases this results in a low spectrum utilization rate and a high probability of blocking. Therefore, the use of different defragmentation techniques allows to set up some, or sometimes all connections which normally will not be set up due to improper utilization of the spectrum.

Table 1
Abbreviations used in the paper

Abbreviation	Description
BV-WCS	Bandwidth-Variable Wavelength Converting Switch
BV-WSSS	Bandwidth-Variable Wavelength Selective Space Switch
BV-WSS	Bandwidth-Variable Wavelength Selective Switch
EON	Elastic Optical Network
FSU	Frequency Slot Unit
NED	Network Elements Description
PC	Passive Coupler
S-W-S	Space-Wavelength-Space switching fabric
TWBC	Tunable Waveband Bandwidth Converters
W-S-W	Wavelength-Space-Wavelength switching fabric

Table 2
Symbols used in the paper

Symbol	Description
c	Number of TWBCs
k	Number of FSUs in each interstage fiber
m	Number of FSUs occupied by one connection
$m_{\max}$	Maximum number of FSUs occupied by one connection
n	Number of FSUs in each input/output fiber
p	Number of switches in the center stage
q	Number of input/output fibers in each input/output switching element
r	Number of switches in the input/output stage

Several architectures of elastic optical switching nodes are known [3]–[5]. Recently, new architectures of EONs, referred to as Wavelength-Space-Wavelength (W-S-W) [6] and Space-Wavelength-Space (S-W-S) [7] were proposed. In this paper, two instances of the W-S-W architecture are considered, called WSW1 and WSW2, respectively [8]. Some abbreviations and symbols used in this paper have already been introduced, and some will be defined later. For the reader's convenience, they are presented in Tables 1 and 2.

The remaining portion of the paper is organized as follows. In Section 2 a short description of the EON architectures used is provided. In Section 3 problem statement, and in Section 4 defragmentation methods are described. Section 5 introduces the simulator which allows to simulate W-S-W EONs. The last Section presents conclusions and the future work.

2. EON's Architectures

As mentioned before, two W-S-W switching architectures are considered in this paper: WSW1 and WSW2. In paper [9], only the WSW1 structure was implemented in the simulator proposed in Section 5. As the simulator in question still remains in the development phase, both WSW1 and WSW2 switching fabrics are already implemented.

The WSW1 switching fabric (see Fig. 1a) consists of r Bandwidth-Variable Wavelength Converting Switches (BV-WCSs) with the capacity of 1×1 in the first and third stages, and of one Bandwidth-Variable Wavelength Selective Space Switch (BV-WSSS) with the capacity of $r \times r$ in the center stage. Each BV-WCS contains one Bandwidth-Variable Wavelength Selective Switch (BV-WSS), one Passive Coupler (PC), and c Tunable Waveband Bandwidth Converters (TWBCs). The role of BV-WSS is to direct connections from the input fiber to different TWBCs, one connection to one TWBC. In the TWBC, the connection is moved from one set of FSUs (one frequency slot) to another set of FSUs (another frequency slot). After conversion in TWBCs, all connections are combined to the output fiber by the PC. In turn, one BV-WSSS has r BV-WSSs and r PCs. For a detailed description of the WSW1 switching fabric see [8].

The WSW2 switching fabric (see Fig. 1b) consists of r BV-WCSs with the capacity of $q \times p$ in the first stage, r BV-WCSs with the capacity of $p \times q$ in the third stage, and p BV-WSSSs with the capacity of $r \times r$ in the center stage. Each BV-WCS of the first stage contains q BV-WSS, p PCs and c TWBCs. Each BV-WCS of the third stage contains p BV-WSS, q PCs and c TWBCs. For WSW2, $c = qp$. The role of BV-WSS is, similarly as in the WSW1 switching fabric, to direct connections from the input fiber to different TWBCs, one connection to one TWBC. Then, in TWBC, the connection is moved from one set to another set of FSUs. After spectrum conversion, all connections are combined by PC to the output fiber. In turn, each BV-WSSS has r BV-WSSs and r PCs. For a detailed description of the WSW2 switching fabric see [8].

Each input and each output fiber in the W-S-W switching fabric has n FSUs and each interstage fiber has k FSUs (see Fig. 1). As mentioned before, a new connection could require m frequency slots, where m is typically limited by $1 \leq m \leq m_{\max} \leq n$. Of course, the following is always true: $n \leq k$.

3. Defragmentation

Defragmentation of EONs very often occurs at the network level [10]–[12]. This means that input and output node's

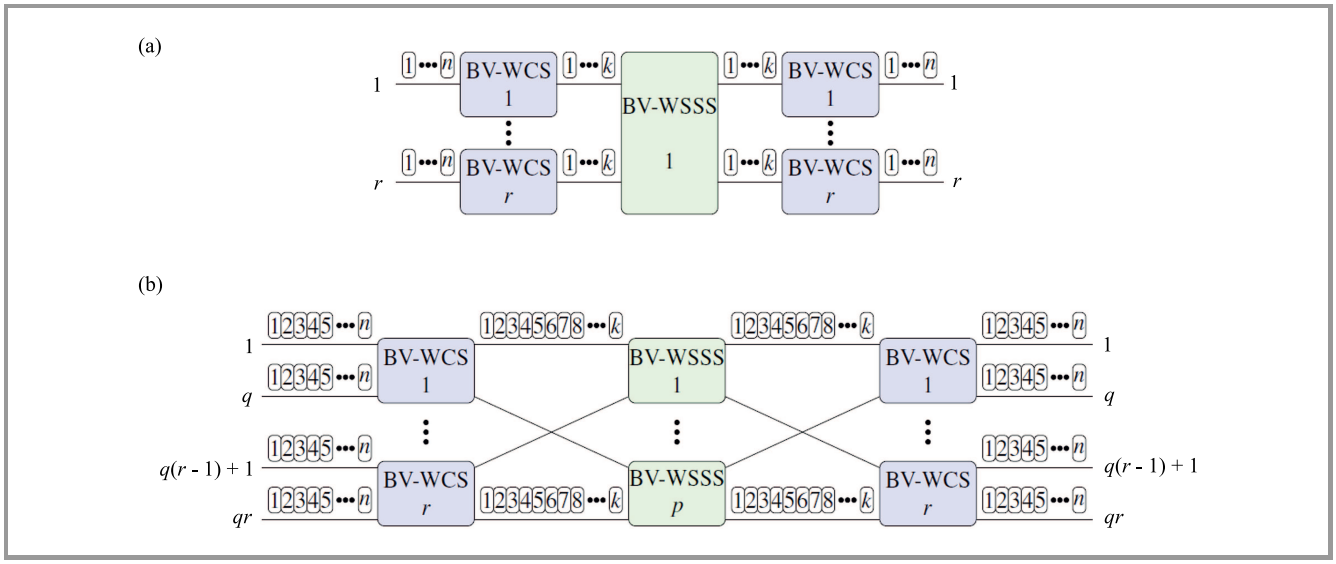


Fig. 1. W-S-W elastic optical networks: (a) WSW1 structure, (b) WSW2 structure.

FSUs are defragmented according to the network state, and for such a defragmentation it is not important which structure of the network node is used. However, from the node point of view, defragmentation could be performed inside the node as well. Defragmentation in an EON node occurs inside interstage links only. This means that the order of FSUs at each input, as well as at each output link of the W-S-W switching fabric will not be changed. The only place where the order of FSUs can be changed is at the input and output link of the BV-WSSS switch.
The frequency slot unit defragmentation algorithm needs to address the following questions: *when to defragment*, *what to defragment*, and *how to defragment?*

3.1. When to Defragment?

Defragmentation in the WSW1 or in the WSW2 switching nodes could be performed at a different moments:

- The first moment of defragmentation is when an *m*-slot connection has just been disconnected and *m* free FSUs have appeared in the switching node for new, future connection(s);
- The second moment of defragmentation is when a new *m*-slot connection appears in a node and there are enough FSUs to establish this connection, with those FSUs not being adjacent, however.

3.2. What for Defragment?

Defragmentation allows to establish a new connection when there are enough FSUs in a switching node, with the said FSUs not being adjacent to each other, however. There are several methods of defragmentation and they differ in the in which the unused FSUs are ordered. Therefore, choosing the right defragmentation method enables to reorder FSUs in the optical spectrum, which means that after the defragmentation process there will be enough adjacent FSUs to establish a new connection.

3.3. How to Defragment?

Several methods used to defragment FSUs may be distinguished: re-optimization [13], make-before-break [14], push-and-pull [15], and hop-tuning [16]. All of them are commonly used for defragmentation of an available spectrum used in links between EON nodes. In this article they were used to defragment an optical spectrum inside a W-S-W node.

4. Defragmentation Methods

In a special simulator, the four methods referred to above are used to defragment an optical bandwidth inside an EON node There are, of course, more known methods of defragmentation. However, they are have not yet been implemented in the simulator available. The simulator is described shortly in Section 5 and it is still under development.

4.1. Re-optimization

In the re-optimization method, all existing connections have to be disconnected and set up once again. The advantage of this method is that no additional transmitters are required. The time of defragmentation is a disadvantage. Sometimes, this lead time is very long. A simple example showing how this method workings is presented in Fig. 2.

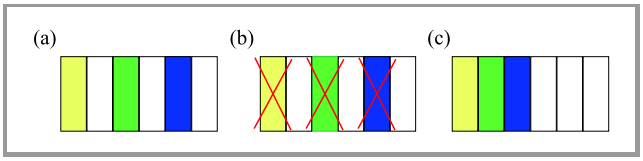


Fig. 2. Re-optimization defragmentation method: (a) step 0 – state before defragmentation, (b) step 1 – disconnecting all existing connections, (c) step 2 – setting up all connections once again (state after defragmentation). (See color pictures online at www.nit.eu/publications/journal-jtit)

4.2. Make-before-break

In the make-before-break method, a copy of some existing connection is created in free FSUs and two identical connections exist simultaneously at a certain time. To handle this, an additional transmitter is needed. Moreover, during the defragmentation process, a new connection cannot be established due to fact that more FSUs are occupied compared to the "before defragmentation state" in the EON node. It is obvious that in order to perform defragmentation using this method, an additional number of free slots is required (in fact twice as many new FSUs are required by the new connection) and it will not always be possible to defragment an optical bandwidth inside the EON's node. A simple example showing how this method works is presented in Fig. 3.

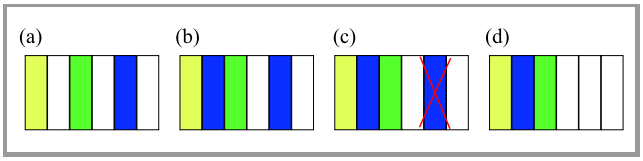


Fig. 3. Make-before-break defragmentation method: (a) step 0 – state before defragmentation, (b) step 1 – creating a copy of the existing (blue) connection, (c) step 2 – disconnecting original connection, (d) step 3 – state after defragmentation.

4.3. Push-and-pull

In the push-and-pull method, some existing connections are moved within the optical bandwidth. Any connection may

be moved inside the free adjacent FSUs, until the connection in question becomes adjacent to the other connection. In the case in which there are no free FSUs adjacent to the connection under consideration, this connection cannot be moved to other FSUs. A simple example showing how this method works is presented in Fig. 4.

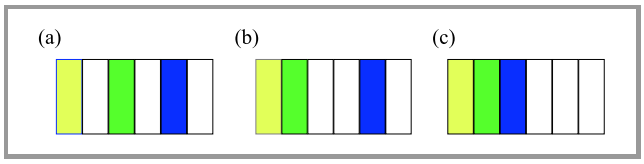


Fig. 4. Push-and-pull defragmentation method: (a) step 0 – state before defragmentation, (b) step 1 – moving the second (green) connection to the left side, (c) step 2 – moving the third (blue) connection to the left side (state after defragmentation).

4.4. Hop-tuning

In the hop-tuning method, an existing connection is moved to any free FSUs that are not necessarily adjacent to the connection under consideration. Unlike with the make-before-break and push-and-pull methods, this method allows to move several connections at the same time. This is a big advantage, as the time needed for such a defragmentation is very often shorter than 1 μs. In this method, there is no need to use additional transmitters, which is another advantage. A simple example showing how this method works is presented in Fig. 5.

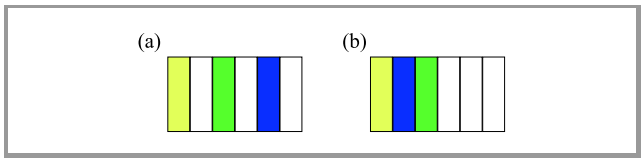


Fig. 5. Hop-tuning defragmentation method: (a) step 0 – state before defragmentation, (b) step 1 – moving (black) connection (state after defragmentation).

5. Simulator

In order to simulate the EON WSW1 architecture, a special software simulator has been developed, based on the OMNeT++ discrete event simulator, version 5.0 [17]. The same tool has been expanded to enable the simulation of WSW2 switching fabrics. At present, the simulator is operating based on OMNeT++ version 5.2. All functionalities (such as the behavior of all elements) and algorithms (such as defragmentation methods) have been developed with the use of C++. All graphical representations of each element, in turn, have been prepared with the use of Network Elements Description (NED) – a special language used in the OMNeT++ environment. OMNeT++ ensures also a variety of generators which can be used as traffic generators. The traffic is very efficient in simulating all connections appearing in the elastic optical node under consideration. It also allows to randomly choose a different size of the new connection – in the case of W-S-W, it is simply the m value.

The first version of this tool was used to simulate strict-sense non-blocking conditions for the WSW1 switching fabric. The results obtained were compared with these achieved in paper [8]. Then, the second version of the simulator was expanded to provide for rearrangeable non-blocking conditions in the WSW1 architecture. The results obtained were compared with these presented in paper [18]. However, in both simulator versions mentioned above, graphical representation of the WSW1 switching fabric was not possible – it merely had the shape of one or several gray elements. Therefore, a third version of the simulator was developed, where the graphical representation of a WSW1 EON was possible (see Fig. 6). Four defragmentation algorithms were added to it as well. In the current version of the simulator, a new structure of W-S-W was added, namely the WSW2 switching fabric. In the future, more defragmentation algorithms will be added and more different EON structures will be supported (like, for instance, S-W-S-type switching fabrics).

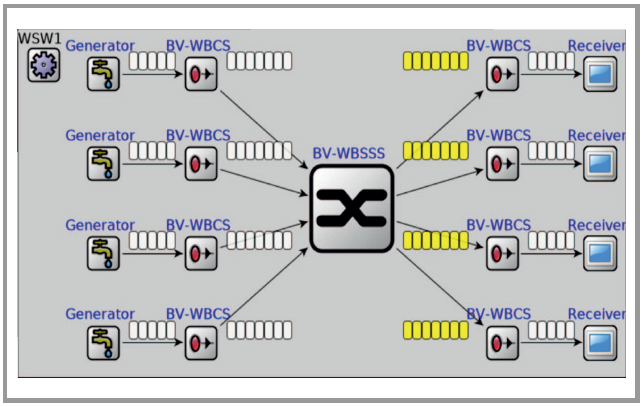


Fig. 6. WSW1 node with $n = 5$, $r = 4$, and $k = 7$ operating in the OMNeT++ environment.

The current software version uses $qr+1$ *cMersenneTwister* random number generators that are already available in the OMNeT++ simulation environment. There are qr different generators (one per input) which draw the number of the output to which the new connection is directed. The size of the new connection m, in turn, is drawn by another generator, assigning an integer from the range of 1 and $m_{\max}$. In the future, a generator responsible for the Poisson type of traffic will be added as well.

The WSW1 switching node simulator asks the user, immediately after its start-up, which parameters (like n, r, and k) should be used during the simulation. By specifying different values of n, r and k, different structures of WSW1 will be obtained. For example, when $n = 5$, $r = 4$ and $k = 7$, the WSW1 EON looks as presented in Fig. 6. Meanwhile, with $n = 10$, $r = 2$ and $k = 15$, the WSW1 structure looks as shown in Fig. 7.

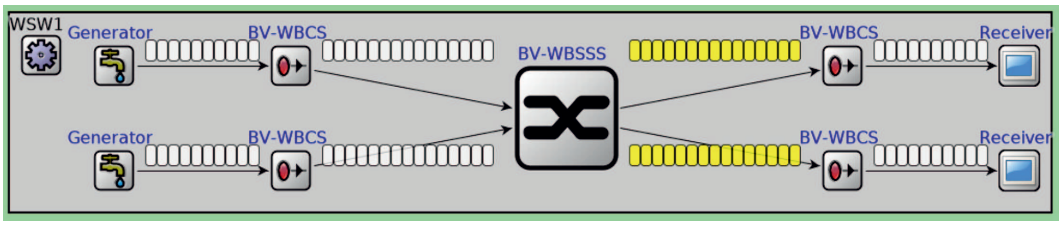


Fig. 7. WSW1 node with $n = 10$, $r = 2$, and $k = 15$ operating in the OMNeT++ environment.

In the simulator, each BV-WBCS, as well as BV-WBSSS, are represented by a proper module which consists of smaller pieces, such as PCs, BV-WSSs and TWBCs. For example, for $n = 5$, each BV-WBCS looks as shown in Fig. 8.

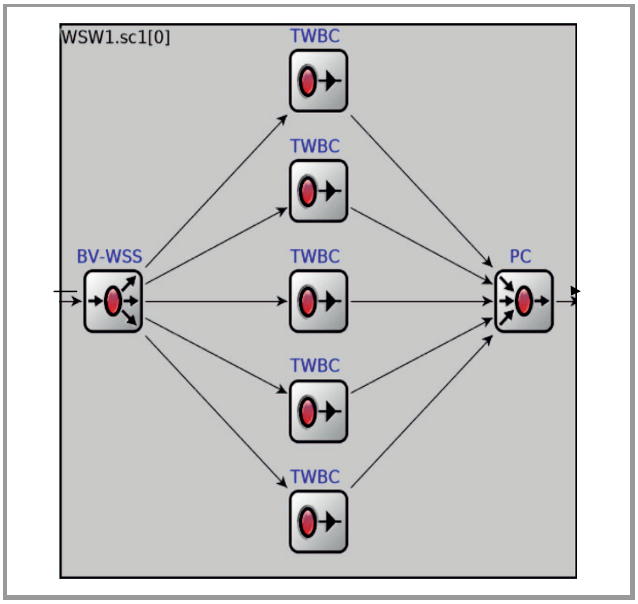


Fig. 8. The BV-WBCS switch implemented in OMNeT++.

For $r = 2$, the BV-WBSSS switch looks as presented in Fig. 9, and for $r = 7$, as shown in Fig. 10.

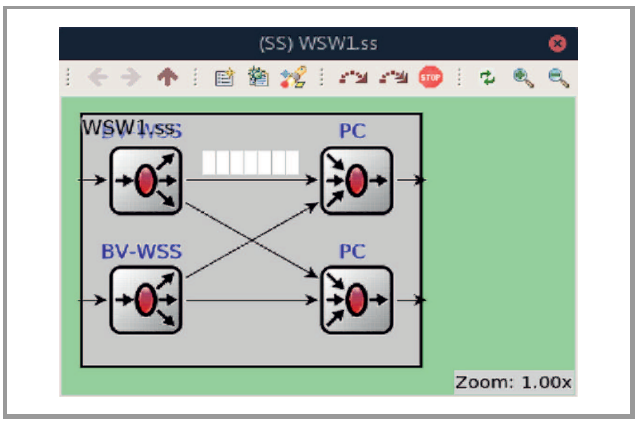


Fig. 9. The BV-WSSS switch for $r = 2$.

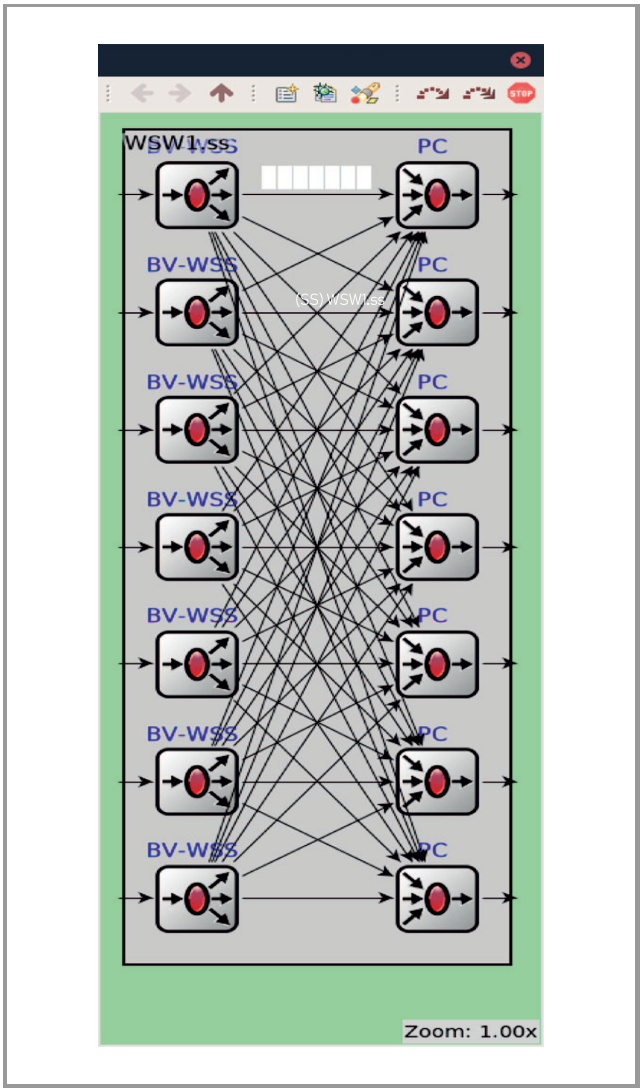


Fig. 10. The BV-WSSS switch for $r = 7$.

6. Conclusions

Four defragmentation methods for the W-S-W architecture have been described in this paper. All of these methods were already implemented in a purpose-developed simulator. The current version of the simulator offers a graphical interface as well. However, it is still under development. The future step is to represent the S-W-S switching fabric

(another example of an elastic optical network) in the simulator as well. Both SWS1 and SWS2 architectures will be implemented. Such a future solution will make it possible to compare both types of structures with one another. Other defragmentation methods will be deployed as well.

Acknowledgments

The work described in this paper was financed with the funds of the Ministry of Science and Higher Education for 2017, under grant 08/82/DSMK/8222.

References

[1] M. Jinno *et al.*, "Spectrum-efficient and scalable elastic optical path network: Architecture, benefits, and enabling technologies", *IEEE Communications Magazine*, vol. 47, no. 11, pp. 66–73, 2009.

[2] ITU-T Recommendation G.694.1. Spectral Grids for WDM Applications: DWDM Frequency Grid, 2012.

[3] P. Zhang *et al.*, "Comparison of node architectures for elastic optical networks with waveband conversion", *China Commun.*, vol. 10, no. 8, pp. 77–87, 2013.

[4] Y. Chen *et al.*, "Demonstration of petabit scalable optical switching with subband-accessibility for elastic optical networks", in *Proc. The OptoElectr. and Commun. Conf. and Austr. Conf. on Opt. Fibre Technol. OECC/ACOFT*, Melbourne, VIC, Australia, July 2014, pp. 350–351.

[5] B. C. Chatterjee, N. Sarma, and E. Oki, "Routing and spectrum allocation in elastic optical networks: A tutorial", *IEEE Commun. Surveys and Tutor.*, vol. 17, no. 3, pp. 1776–1800, 2015.

[6] W. Kabaciński, M. Michalski, and M. Abdulsahib, "The strict-sense nonblocking elastic optical switch", in *Proc. IEEE 16th Int. Conf. on High Perfor. Switching and Routing HPSR*, Budapest, Hungary, July 2015 (doi: 10.1109/HPSR.2015.7483108).

[7] G. Danilewicz, W. Kabaciński, and R. Rajewski, "Strict-sense nonblocking space-wavelength-space switching fabrics for elastic optical network nodes", *J. of Optic. Commun. and Network.*, vol. 8, no. 10, pp. 745–756, 2016.

[8] W. Kabaciński, M. Michalski, and R. Rajewski, "Strict-sense nonblocking W-S-W node architectures for elastic optical networks", *J. of Lightwave Technol.*, vol. 34, no. 13, pp. 3155–3162, 2016.

[9] R. Rajewski, "Defragmentation in the W-S-W Elastic Optical Network," in *Proc. IEICE Infor. and Commun. Technol. Forum ICTF 2017*, Poznań, Poland, 2017.

[10] M. Zhang, W. Shi, L. Gong, W. Lu, and Z. Zhu, "Bandwidth defragmentation in dynamic elastic optical networks with minimum traffic disruptions", in *Proc. IEEE Int. Conf. on Commun. ICC*, Budapest, Hungary, 2013, pp. 3894–3898.

[11] C. You, M. Zhang, and Z. Zhu, "Reduce Spectrum Defragmentation Latency in EONs with Effective Parallelization of Connection Reconfigurations", in *Proc. Opt. Fiber Commun. Conf. and Exh. OFC*, San Francisco, CA, USA, 2014, pp. 1–3.

[12] S. Ba, B. C. Chatterjee, and E. Oki, "Defragmentation scheme based on exchanging primary and backup paths in 1+1 path protected elastic optical networks", *IEEE/ACM Transact. on Network.*, vol. 25, no. 3, pp. 1717–1731, 2017.

[13] I. Tomkos, S. Azodolmolky, J. Solé-Pareta, D. Careglio, and E. Palkopoulou, "A tutorial on the flexible optical networking paradigm: State of the art, trends, and research challenges", in *Proceedings of the IEEE*, vol. 102, no. 9, pp. 1317–1337, 2014.

[14] A. Pagès, J. Perelló, and S. Spadaro, "Lightpath fragmentation for efficient spectrum utilization in dynamic elastic optical networks", in *Proc. 2012 16th Int. Conf. on Opt. Network Design and Modelling ONDM*, Colchester, United Kingdom, 2012, pp. 1–6.

[15] F. Cugini *et al.*, "Push-pull defragmentation without traffic disruption in flexible grid optical networks", *J. of Lightwave Technol.*, vol. 31, no. 1, pp. 125–133, 2013.

[16] M. Zhang, Y. Yin, R. Proietti, Z. Zhu, and S. J. B. Yoo, "Spectrum defragmentation algorithms for elastic optical networks using hitless spectrum retuning techniques", *Proc. 2013 Opt. Fiber Commun. Conf. and Exp. and the Nation. Fiber Optic Engineers Conf. (OFC/NFOEC)*, Anaheim, CA, USA, 2013, pp. 1–3.

[17] The OMNeT++ official website, 20 Oct 2017 [Online]. Available: https://omnetpp.org

[18] W. Kabaciński, R. Rajewski, and A. Al-Tameemi, "Simultaneous connections routing in elastic optical switches with limited number of connection rates", in *Proc. 21st Int. Conf. on Opt. Net. Design and Models ONDM*, Budapest, Hungary, 2017.

Remigiusz Rajewski – for biography, see this issue, p. 17.

Paper

SMM Clos-Network Switches under SD Algorithm

Janusz Kleban[1] and Jarosław Warczyński[2]

[1] *Faculty of Electronics and Telecommunications, Poznan University of Technology, Poznań, Poland*
[2] *Faculty of Electrical Engineering, Poznan University of Technology, Poznań, Poland*

https://doi.org/10.26636/jtit.2018.123017

Abstract—This paper is devoted to evaluating the performance of Space-Memory-Memory (SMM) Clos-network switches under a packet dispatching scheme employing static connection patterns, referred to as Static Dispatching (SD). The control algorithm with static connection patterns can be easily implemented in the SMM fabric due to bufferless switches in the first stage. Stability is one of the very important performance factors of packet switching nodes. In general, a switch is stable for a particular arrival process if the expected length of the packet queues does not increase without limitation. To prove the stability of the SMM Clos-network switches considered under the SD packet dispatching scheme the discrete Markov chain model of the switch is used and Foster's criteria to extend Lyapunov's second (direct) method of stability investigation of discrete time stochastic systems are used. The results of simulation experiments, in terms of average cell delay and packet queue lengths, are shown as well.

Keywords— Clos-network switch, packet dispatching algorithms, packet switching network, stability of switching network

1. Introduction

Connecting paths between input and output ports in switches/routers are provided by switching fabrics, which are the main part of every packet switching node. The switching fabrics replace buses which are too slow, mainly in medium-sized and high-end routers and switches. They can establish connections between input ports and requested output ports, while simultaneously transmitting packets. Single-stage switching fabrics known as crossbar switches are used mainly in medium-sized routers/switches [1].

Basically, an $N \times N$ crossbar switch consists of a square array of N^2 individually operated crosspoints *(*N represents the number of inputs and outputs). Each crosspoint has two possible states: cross (default) and bar, and corresponds to the input-output pair. A connection between input port i and output port j is established by setting the (i, j)-th crosspoint to the bar state, while letting other crosspoints along the connecting paths remain in the cross state. The crossbar switch can transfer up to N cells from different input ports to different output destinations within the same time slot. The control algorithm for the crossbar fabric is very simple, as the bar state of the crosspoint can be triggered individually by each incoming packet when its destination matches the output address. Crossbar fabrics are complex in terms of the number of crosspoints, which grows as N^2. The arbitration process that has to choose packets to be sent from inputs to outputs within each time slot can also become the system's bottleneck, as the switch size increases.

In high-end routers, multi-stage or even multi-stage and multi-plane switching fabrics are used. These types of switching fabrics are currently used by network equipment vendors in core routers, e.g. Cisco's CRS series, Juniper's T series, and Brocade's BigIron RX Series. For example, in Cisco's new router called Carrier Routing System-X (CRS-X), a multi-stage and multi-plane switching fabric is used. This family of routers focuses on the extreme scale. One standard deployment of a 7-ft rack chassis of CRS-X routers can deliver up to 12.8 terabits per second. The system can be clustered together in a massive configuration of up to 72 chassis, which would deliver up to 922 Tb/s of throughput [2].

Clos-network switches are a very attractive solution for core routers because of their modular and scalable architecture. The Clos-network fabric is composed of crossbar switches arranged in stages [3]. According to the required combinatorial properties, it is possible to build [4]:

- strict-sense nonblocking (SSNB),
- wide-sense nonblocking (WSNB),
- rearrangeable (RRNB),
- repackable (RPNB) non-blocking networks.

In SSNB [3] networks, no call is blocked at any time. WSNB [5], [6] networks are able to connect any idle input and any idle output, but a special path-searching algorithm must be used. RRNB [5] networks can also establish the required connections between any idle input-output pair, but a rearrangement of some existing connections to other connecting paths may be needed to change the network state in order to unblock a blocked call. RPNB [7], [8] networks employ rearrangements after call termination to prevent the switching fabric from entering blocking states. The presented classes of switching networks were proposed in the past, when circuit-switching telephone exchanges supported voice traffic.

Currently, telecommunication networks focusing on packet services and high-speed switching fabrics adopt the use

of fixed-length packets called cells. All incoming variable-length packets (e.g. IP packets) are segmented at ingress line cards into fixed-size cells. Next, they are transmitted within time slots through the switching fabric, and reassembled into packets at egress line cards, before they depart [1]. In high-speed routers it is not necessary to use SSNB switching fabrics, because a new set of connecting paths may be set up for each time slot. RRNB fabrics are sufficient to satisfy all requirements related to one-to-one connections between all sources and destinations. They can establish connecting paths for all possible permutations of input-output pairs. These connections can be established on a call-by-call basis or simultaneously for a given set of inputs and outputs. The former technique employs rearrangements of the existing paths when a new call cannot be set up. In the latter method, parallel processing of all required input-output connections is carried out and, next, the connecting paths are set up simultaneously in the switching fabric.

The main difference between circuit switching and packet switching fabrics is that in circuit switching systems, when the output port is busy, the call is lost. In packet switching fabrics, when outputs are busy, cells are buffered and wait in queues. This means that queues of cells destined for, let's say, a very popular output port, may grow, because many cells should be sent to the same output port, but only one cell can be sent out within one time slot.

While a cell is being routed in a packet switching fabric, it can face a contention problem resulting from the fact that two or more cells compete for a single resource. Algorithms that can solve contentions, are usually called packet dispatching schemes. Cells that have lost contention must be either discarded or buffered. Buffers are also used to alleviate the complexity of packet dispatching algorithms and to absorb possible contentions. According to buffer allocation schemes, Clos-network packet switches are classified as: Space-Space-Space (SSS or S^3), Memory-Memory-Memory (MMM), Memory-Space-Memory (MSM), and Space-Memory-Memory (SMM) switches. MSM Clos-network switch seems to be the best architecture investigated. The basic packet dispatching algorithms for this kind of switching fabrics were proposed in [9], [10]. A modified MSM Clos-network switch was proposed and investigated in [11].

In this paper, we analyze the SMM Clos-network switch, where bufferless modules are used in the first stage and buffered crossbars in the second and third stages. Due to bufferless modules in the first stage, a very simple control algorithm may be implemented to distribute cells to the central modules, e.g. static dispatching (SD). The SMM architecture was proposed in [12], where an analytical analysis for admissible traffic was performed. In [13], different kinds of backpressure schemes between central modules and input modules are evaluated, in terms of maximum buffer usage in central modules. The packet dispatching scheme proposed in [14] uses static dispatching patterns and internal backpressure signals. It is dedicated for SMM Clos-network switches, where the second and third stages are made of crosspoint queued (CQ) switches [15]. In [16], a fault-tolerant desynchronized static round-robin (FT-DSRR) cell dispatching algorithm was proposed. The FT-DSRR algorithm is an adaptation of the DSRR algorithm to SMM Clos-network switches, where serious crosspoint faults induced by harsh space radiation environment may take place. It may be used to control onboard switches.

This paper deals with an SMM Clos-network switch, where output queued switches are used in the second and third stages. The main contribution of this paper is the proof of stability of the SMM Clos-network switch using the Discrete Time Markov Chain (DTMC) model and an analytical approach based on Foster's stochastic criteria, analogous to the direct method of Lyapunov which was aimed at inferring about the stability of deterministic dynamic systems. The theoretical results were verified by simulation investigations of RRNB and SSNB architectures of a network under uniform and non-uniform traffic distribution patterns.

The remainder of this paper is organized as follows. Section 2 introduces some background knowledge concerning the SMM Clos-network switch and the SD algorithm. In Section 3 input traffic analysis is performed. Using a stochastic Lyapunov-like analytical method, we prove that the investigated switching fabric is stable under the SD packet dispatching scheme in Section 4. Section 5 presents simulation results obtained for the SD scheme. We conclude this paper in Section 6.

2. SMM Clos-Switching Fabric and SD Scheme

The three-stage Clos switching fabric architecture is denoted by $C(m, n, r)$, where the parameters m, n, and r entirely determine the structure of the network. There are r input modules (IM) of capacity $n \times m$ in the first stage, m central modules (CM) of capacity $r \times r$, and r output modules (OM) of capacity $m \times n$ in the third stage. The capacity of this switching system is $N \times N$, where $N = nr$. The three-stage Clos-network switch is strictly non-blocking if $m \geq 2n - 1$ and rearrangeable non-blocking if $m \geq n$.

In the basic SMM Clos-network switch (shown in Fig. 1), the first stage consists of r bufferless IMs with n input ports (IPs) each. The second stage consists of m CMs, and each of them has r FIFO buffers (COQs), one per output. A maximum of r cells from r IMs may arrive at one COQ buffer, so it must work r times faster than the line rate. The third stage consists of r OMs, where each output port $OP(j,h)$ has FIFO output buffer (OQ). A maximum of m cells from m CMs may arrive at one OQ, so to store all cells during one time slot it must work m times faster than the line rate. The interstage links between IMs and CMs are denoted by $L_I(i,k)$, where i represents the number of IMs, and k – the number of CMs, whereas $L_C(k,j)$ denotes interstage links between $CM(k)$, and $OM(j)$. In-

stead of using shared-memory CM and OM modules, it is possible to employ CQ switches, where speed-up is not necessary [15].

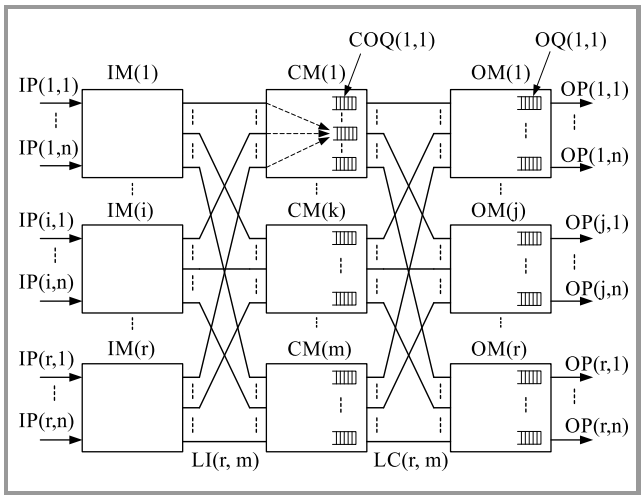


Fig. 1. An SMM Clos-network switch.

The SD scheme investigated in this paper seems to be the simplest packet dispatching algorithm that can be implemented in the SMM Clos-network switch. It is an adaptation of the Static Round-Robin Dispatching (SRRD) to the SMM Clos-network switch, and is less demanding in terms of hardware, in comparison with other proposed schemes (e.g. [14]). The SD scheme does not need any special arbitration, e.g. the handshaking processes, to distribute cells to the CMs. The key idea of the scheme is based around static connection patterns which are used in each IM. The consecutive static connection patterns used in IMs are shown in Fig. 2.

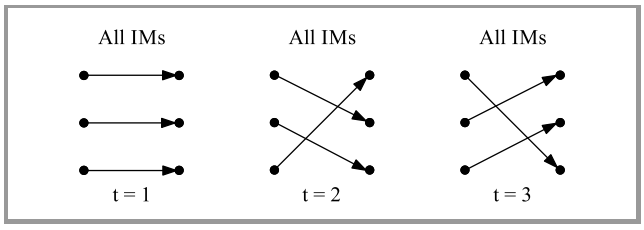


Fig. 2. A sequence in which the static connection patterns should be changed in each IM of capacity 3×3.

The connection patterns are the same in all IMs and are shifted to the next one in consecutive time slots. Cells arriving at each input are at once distributed to the CMs, and are stored in COQs related to the destined OMs. In the first time slot, cells from $IP(x,1)$ are sent to $CM(1)$, from $IP(x,2)$ to $CM(2)$, from $IP(x,3)$ to $CM(3)$; in the second time slot, cells from $IP(x,1)$ are sent to $CM(2)$, from $IP(x,2)$ to $CM(3)$, from $IP(x,3)$ to $CM(1)$, and so on. Arriving cells are evenly distributed to CMs, to decrease cell delay within the SMM Clos-network switch. The SD scheme may be also adopted in the MSM Clos-network switch [7].

3. Input Traffic Analysis

We assume that the traffic directed to each input port IP(*i, h*) can be modeled by an i.i.d. Bernoulli process, where the number of successes – which means the number of cells arriving in t time slots (in t trails) is tp_B with p_B denoting the probability of success in one trial. In such a case, the ports' arrival rate is expressed by the expected value:

$$\lambda_{IP} = \lim_{t\to\infty} \frac{t\,p_B}{t} = p_B\,. \tag{1}$$

Therefore, the input traffic arriving at one input module is equal to $\lambda_{IM}{=}np_B$, and at the whole switching fabric, all input modules – rnp_B. The SD algorithm balances this input load on CMs and after m time slots the central modules' arrival rate can be expressed in the following way:

$$\lambda_{CM} = \frac{n\,p_B\,r}{m}\,. \tag{2}$$

There are output queues (COQs) in each central module which stores cells destined for the predetermined OMs. When analyzing the input rate of these queues, it is easy to see that this rate can be assessed as:

$$\lambda_{COQ(i,j)} = \frac{np_B\,r}{m}\,p_{ij}\,, \tag{3}$$

where p_{ij} represents the probability of a cell arriving from the i-th input module being destined for the j-th output module. For example, with traffic uniformly distributed to the output ports and, in consequence, to the output modules OMs, $p_{ij} = 1/r$. This means that even for the maximum input port load, i.e. for $p_B = 1$, the rate $\lambda_{COQ(i,j)}$ is less than or equal to 1, if the number of OMs is $m \geq n$.
In the investigated SMM Clos-network architecture, each central module CM has one link to each OM. This ensures that in each time slot from any non-empty COQ(i, j), one cell will be sent to the appropriate OM(j), which can be described by the COQ(i, j) queue's service rate $\mu = 1$.

4. Stability Proof

The theory of stability for deterministic dynamic systems was founded by Lyapunov [18] (see also [19] for survey of stability ideas) who invented two methods of stability investigation. His second method, known as *Lyapunov's second method* or *indirect method*, turned out to be very effective in proving the stability of a very wide spectrum of deterministic systems – linear, non-linear, continuous and discrete. Later, Lyapunov's ideas were extended to stochastic systems, mainly by Foster [20]. The application of this theory to Markov chains was due to Meyn and Tweedie [21].
The term *stability*, in the context of dynamic systems described by ordinary differential equations, is commonly used to mean asymptotic stability, i.e. convergence of a system's state paths to a fixed, stable, point.
With Markovian systems, convergence must be understood in a distributional sense and, therefore, is called stochastic

stability. It considers stochastic convergence in the time of a Markov chain $X = (X_n;\ n = 0,1,\dots)$, as $n \to \infty$.
In general terms, the Markov chain is topologically stable if there is a positive probability that it does not leave the compact center of the state space (which is called *non-evanescence* [21]), or, using a stronger condition, if the distributions of the chain as time evolves are tight (bounded in probability [21]). Meyn and Tweedie say the chain is probabilistically stable if it returns to sets of positive measure (Harris recurrence), or if there is a unique invariant probability measure for it (positive Harris recurrence) [21]. According to [20] and [21], the stability proof of stochastic systems modeled by Markov chains must show:

1. the irreducibility of the chain, which means that starting from any initial state, it is possible to arrive in subsequent transitions on any other state of the chain. Figs. 3 and 4 show examples of irreducible and non-irreducible Markov chains;
2. the positive recurrence of the chain, which can be done by demonstrating the negative drift of the Lyapunov function.

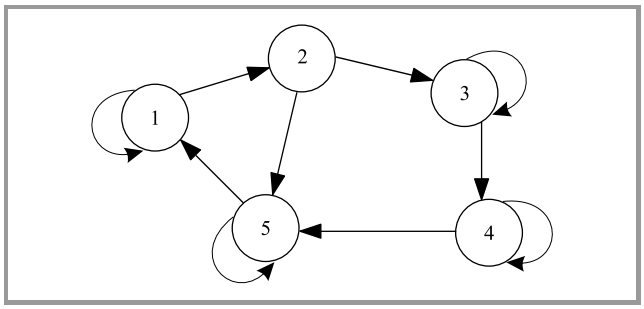


Fig. 3. Example of an irreducible Markov chain.

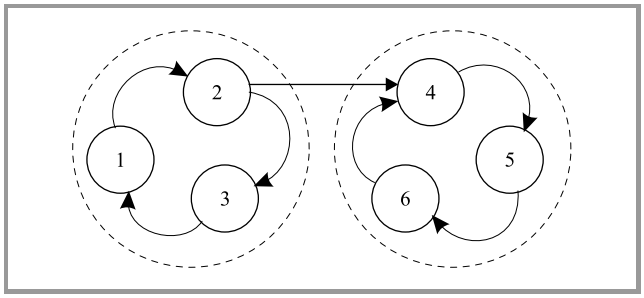


Fig. 4. Example of a non-irreducible Markov chain.

The positive recurrence of the irreducible chain's state means (see for example [21]) that

$$E(\tau_i | X_0 = i) < \infty. \tag{4}$$

That is, if state i is positive recurrent, then the chain comes back infinitely often to state i and the time τ_i between two consecutive visits is finite.
Denoting by P_i the conditional probability of the process started at state i, we say, by definition (see for example [21]), that a state i is:

1. transient if $P_i(\tau_i = \infty) > 0$,
2. null recurrent if $P_i(\tau_i < \infty) = 1$,
3. positive recurrent if it is recurrent and $E_i(\tau_i) < \infty$.

For irreducible Markov chains, condition (4) implies a positive recurrence of state i and, hence, a positive recurrence of the whole chain, i.e. in a given class, all states are either positive recurrent, null recurrent or transient.

Lemma 1: If X is irreducible, then all states are of the same type.

Proof: The proof can be based on the following fact: If X is irreducible and $j \neq k$ are any two states, then $P_j(\tau_k < \tau_j) > 0$. Now, let us assume the opposite – that this probability equals 0. Then, by the strong Markov property, the process starting from j would never visit state k. This is, however, in contradiction with the irreducibility of X.
Let S be the state space of a given DTMC and let $P \subset S$ be a finite subset of S. Denoting by τ_P the time of the first visit to set P, one can state the following generalization of condition (4) (according to the guidelines in [21]):

Lemma 2: Let $X = (X_n;\ n = 0,1,\dots)$ denote an irreducible DTMC with state space S and let $P \subset S$ be a finite subset of S. Chain X is positive recurrent if and only if:

$$E(\tau_P | X_0 = i) < \infty \text{ for all } i \in P. \tag{5}$$

However, it is rather difficult to determine with Eq. (5) whether a given Markov chain is positive recurrent or not. Here, the Lyapunov-Foster criteria can be used [20]:
Let $X = (X_n;\ n = 0,1,\dots)$ be an irreducible Markov chain defined on some countable space S with transition probabilities p_{ij}, $i, j \in S$. On the basis of [20], we can state:

Theorem 1: The Markov chain X is positive recurrent if and only if there exists a finite set $S_0 \in S$ and a function $V\colon S \to R^+$ with $\inf\{f(i) :\ i \in S\} > -\infty$ and a constant $\varepsilon > 0$ such that:

$$\sum_{j \in S} p_{ij} V(j) < \infty \quad \text{for all } i \in S_0, \tag{6}$$

and

$$\sum_{j \in S} p_{ij} V(j) \leq V(i) - \varepsilon \quad \text{for all } i \notin S_0. \tag{7}$$

The function $V\colon S \to R^+$ is commonly referred as the Lyapunov-Foster function.
Equations (6)–(7) can be rewritten in the equivalent form:

$$E[V(X_{n+1}) | X_n = i] < \infty \text{ for } i \in S_0, \tag{8}$$

and

$$E[V(X_{n+1}) - V(X_n) | X_n = i] \leq -\varepsilon \text{ for } i \notin S_0. \tag{9}$$

Looking at Eq. (8)–(9), it is easy to notice that Foster's criteria can be interpreted as conditions for the Lyapunov's function drift, which is in analogy with Lyapunov's stability direct method for dynamical systems described by ordinary differential equations.
The function fulfilling Lyapunov's conditions can be regarded as a Lyapunov candidate function (only a candidate function which allows stability proving is called a Lyapunov function). There are requirements imposed on Lyapunov candidate function $V(x)$ [18]:

1. $V(x)$ is scalar on the investigated system's state vector **x**; switching networks' states are determined by queue lengths,
2. is positive definite, i.e.: $\underset{x\neq 0}{\forall} V(x) > 0;\ V(0) = 0$,
3. $V(x)$ grows with the state growth of the investigated system which, in our case, means that it grows with the length of switching network queues,
4. for continuous systems: $V(x) \in C_1$.

Speaking generally, there are two levels of stability [18]–[21] – the so-called *weak stability* and the *asymptotic stability*. A proof of weak stability for a given switch network guarantees its full, 100% throughput, but does not predetermine the maximum delay of cells, which in general may be unlimited. The asymptotic stability is a more demanding level of stability, which guarantees not only full throughput of the network, but also a finite value of the maximum cell delay.
Formally, the switching system in which the packet (cell) arrival is an independent random process is characterized by the weak (in Lyapunov sense) stochastic stability if for every $\varepsilon > 0$ there exists $\delta > 0$ that:

$$\underset{\varepsilon>0}{\forall}\ \exists \delta > 0\ \lim_{t\to\infty} P\{\|q_t\| > \delta\} < \varepsilon\,, \tag{10}$$

or

$$\underset{\varepsilon>0}{\forall}\ \exists \delta > 0\ \lim_{t\to\infty} P\{\|q_t\| < \delta\} < 1-\varepsilon\,, \tag{11}$$

where $P\{Z\}$ denotes the probability of event Z, and $||q_t||$ is any norm of q_t – the measure of queues in the system.
The asymptotic stochastic stability is defined as follows: a switching fabric in which the packet (cell) arrival is an independent random stationary process characterized by asymptotic stochastic stability if:

$$\limsup_{t\to\infty} E\{\|q_t\|\} < \infty\,. \tag{12}$$

Inequality (12) means that the maximum expected value of $||q_t||$ is finite. Asymptotic stochastic stability guarantees limited average queue lengths and limited cell delay times. As shown above, the dynamics of the SMM switching fabric is determined by COQ queues (due to static connections of the central stage with the first and third stages, contentions are possible only in the COQ queues).
Let us note that the dynamics of the COQ(*i, j*) queue can be represented by the Markov chain's state diagram

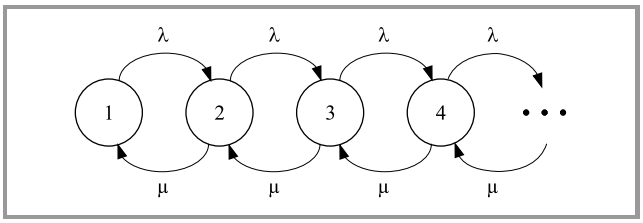


Fig. 5. State graph of a COQ(i, j) queue.

depicted in Fig. 5, where λ represents the queue arrival rate – $\lambda_{COQ(i,j)}$, and μ – is the queue service rate.
The proof of stability of this queue can be based on the Foster-Lyapunov criterion [19]–[21]. It requires that the Lyapunov candidate function $V(q_t)$, defined on the queue length, has a negative drift, strictly that:

$$\underset{\|q_t\|>\varepsilon}{\forall} E\left[V\left(q_{t+1}\right) - V(q_t) | q_t\right] < -\delta\,. \tag{13}$$

In the following proof of stability, Lyapunov candidate function is chosen as the simplest possible one:

$$V(q_t) = q_t\,. \tag{14}$$

The selected function $V(q_t)$ satisfies the Lyapunov candidate function requirements specified above. After substituting the selected function $V(q_t)$ into the left-hand side of inequality (13) and taking into account the graph in Fig. 5:

$$\begin{aligned} E\left[V\left(q_{t+1}\right) - V(q_t) | q_t\right] &= E[q_{t+1}|q_t] - E[q_t|q_t] = \\ &= E[q_{t+1}|q_t] - q_t = \left[\tfrac{\lambda}{\lambda+\mu}(q_t+1) + \right. \\ &\left. + \tfrac{\mu}{\lambda+\mu}(q_t-1)\right] - q_t = \tfrac{\lambda-\mu}{\lambda+\mu}\,, \end{aligned} \tag{15}$$

eventually, the stability condition is:

$$\frac{\lambda-\mu}{\lambda+\mu} < 0\,. \tag{16}$$

The drift is negative when $\lambda < \mu$. For $\mu = 1$, the system will be weakly stable (stable in Lyapunov sense) for $\lambda < 1$. It is worth noting that it does not follow that for $\lambda = 1$ the system will not be stable. The Lyapunov method proves only the stability, and if that fails, the instability of the studied system does not follow from it.
In order to prove the asymptotic stochastic stability, it should be shown that:

$$\underset{\|q_t\|>\varepsilon}{\forall} E\left[V\left(q_{t+1}\right) - V(q_t) | q_t\right] < -\delta \|q_t\|\,. \tag{17}$$

For this purpose, we need another Lyapunov candidate function $V(q_t)$ – we choose it as:

$$V(q_t) = q_t^2\,. \tag{18}$$

The drift of this function is:

$$\begin{aligned} E\left[V\left(q_{t+1}\right)-V\left(q_t\right)\middle|q_t\right] &= E\left[q_{t+1}^2\middle|q_t\right]-E\left[q_t^2\middle|q_t\right] = \\ &= E\left[q_{t+1}^2\middle|q_t\right]-q_t^2 = \left[\tfrac{\lambda}{\lambda+\mu}\left(q_t+1\right)^2+\right. \\ &\left.+\tfrac{\mu}{\lambda+\mu}\left(q_t-1\right)^2\right]-q_t^2 = \tfrac{\lambda}{\lambda+\mu}\left(q_t^2+2q_t+1\right)+ \\ &+\tfrac{\mu}{\lambda+\mu}\left(q_t^2-2q_t+1\right)-q_t^2 = q_t^2\left(\tfrac{\lambda}{\lambda+\mu}+\tfrac{\mu}{\lambda+\mu}-1\right)+ \\ &+2q_t\left(\tfrac{\lambda}{\lambda+\mu}-\tfrac{\mu}{\lambda+\mu}\right)+\tfrac{\lambda+\mu}{\lambda+\mu} = \\ &= 2q_t\tfrac{\lambda-\mu}{\lambda+\mu}+1 = 2q_t\tfrac{-(\mu-\lambda)}{\lambda+\mu}+1\,. \end{aligned} \tag{19}$$

Solving the inequality:

$$2q_t\frac{-(\mu-\lambda)}{\lambda+\mu}+1<0\,, \tag{20}$$

the conditions for asymptotic stability can be determined. For $\mu=1$ we obtain:

$$q_t>\frac{\lambda+1}{2(1-\lambda)} \text{ and } \lambda<1\,. \tag{21}$$

This means that the asymptotic stability will only occur for a sufficiently large q_t, for example, assuming $\lambda=0.9$, this will be an average of 10 cells, that is, when the value is reached, the cell delay will be limited and stabilized.

5. Simulation Experiments

The experiments have been carried out mainly for the RRNB Clos-network switch $C(8,8,8)$ of size 64 × 64 (8 switches in each stage) under the SD algorithm. The SSNB $C(8,16,8)$ architecture, with 8 switches in the first and last stages, and 15 switches in the second stage was also investigated. A wide range of traffic loads per input port, from $p_B=0.05$ to $p_B=1$, with the step of 0.05, was considered in each simulation experiment. 95% confidence intervals that have been calculated after t-student distribution for ten series with 250,000 time slots (after the starting phase comprising 50,000 time slots, which enables reaching the stable state of the SMM Clos-network switch) are at least one order lower than the mean value of the simulation results, therefore they are not shown in the figures. It is assumed that in the second and third stages the switches with output buffers are used, and the size of buffers is not limited. Three main performance measures have been evaluated: average cell delay in time slots, maximum size of OQs, and throughput. A switch can achieve 100% throughput under uniform or non-uniform traffic, if the switch is stable, as it was defined in [22]. It means that the cell queues do not grow without limitation.

Two packet arrival models are considered in simulation experiments: the Bernoulli arrival model, and the bursty traffic model, where the average burst length is set to 16 cells. Several traffic distribution models (the most popular one in this area of research) have been considered, which determine probability p_{ij} that a cell, which arrives at input i, will be directed to output j. The considered cell distribution models are: uniform – $p_{ij}= p_B/N$, diagonal – $p_{ij}= 2p_B/3$ for $i = j$ and $p_{ij}= p_B/3$ for $j = (i+1) \bmod N$, and 0 otherwise, and Hot-spot: $p_{ij}= p_B/2$ for $i = j$, and $p_B/2(N-1)$ for $i \neq j$. Selected simulation results are shown in Figs. 6, 7, and 8. Figure 6 shows the average cell delay, in time slots, obtained for Bernoulli and bursty arrival models, and different kinds of cell distribution models. The SD algorithm provides 100% throughput for the investigated switching fabric only for uniform traffic and the Bernoulli arrival model. Under Bernoulli arrivals, the throughput is limited to 90% for non-uniform traffic, such as diagonal and Hot-spot. It is possible to say that the SD scheme, for uniform and non-uniform traffic distribution patterns under Bernoulli arrivals, performs quite well when the input load is lower than 0.85. In this case, the average cell delay is not greater than 10 time slots. For the bursty arrival model, the SMM Clos-network switch controlled by the SD algorithm is not

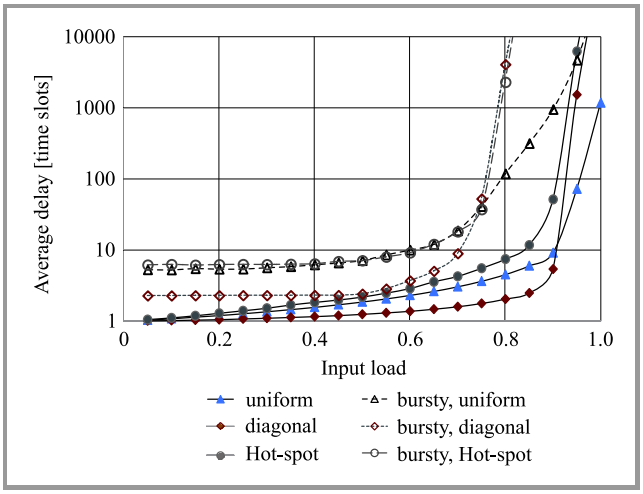


Fig. 6. Average cell delay at egress side of the SMM Clos-network switch under the SD scheme.

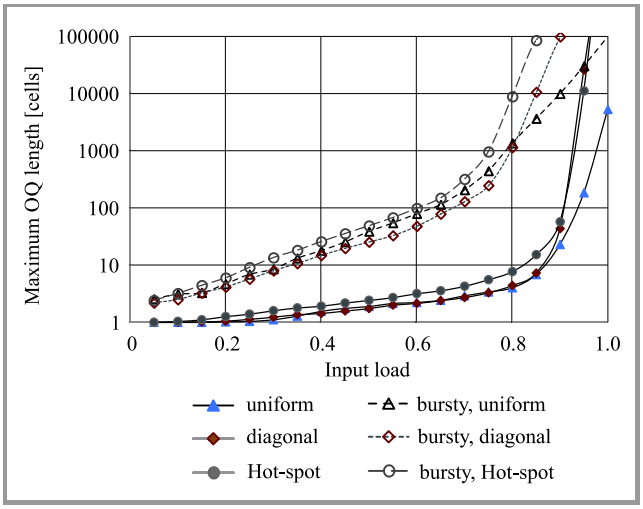


Fig. 7. Maximum OQ length in OMs under the SD scheme.

able to achieve 100% throughput for both the uniform and non-uniform traffic distribution patterns. For the uniform traffic, the throughput is close to 98%, but for the non-uniform traffic, the throughput is limited to 80%.

Figure 7 shows the maximum OQ length obtained during simulation experiments. These results are consistent with the charts presented in Fig. 6. It can be seen that for Bernoulli arrivals the OQ length rapidly grows for a heavy input load and non-uniform traffic (p_B>0.9). In bursty traffic, the OQ length increases very fast for p_B>0.75, especially for non-uniform cell distribution patterns.

Speaking generally, the SD algorithm is very simple to implement within the SMM Clos-network switch and can produce good results for the input load of p_B<0.7, both for uniform and non-uniform traffic distribution patterns. The results related to throughput are not impressive, but the complexity of this algorithm is very low.

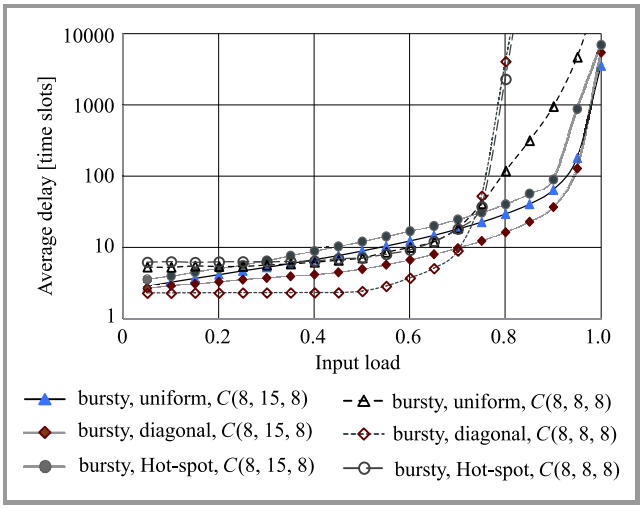


Fig. 8. Average cell delay at egress side of the SMM Clos-network switch for $C(8,8,8)$ and $C(8,15,8)$ architectures under bursty traffic.

Figure 8 shows a comparison of the average cell delay under bursty traffic for RRNB $C(8,8,8)$ and SSNB $C(8,15,8)$ architectures. As it was stated above, in packet switching nodes, connection patterns may be changed in every time slot, so the RRNB switching fabric can set up all connections possible between inputs and outputs. The SSNB architecture contains more switches in the second stage than the RRNB architecture. In this case, cells are distributed more evenly to a higher number of buffers located in the second stage, and can thus reach the destined output with a lower delay. As it is shown in Fig. 8, the SSNB fabric can offer 100% throughput and produces better results than the RRNB fabric under the SD scheme and bursty traffic. For input loads lower than 0.9, the average delay is lower than 100 time slots for all traffic distribution patterns investigated. The delay grows very fast for input load greater than 0.9, but the average delay is very high only for input loads equal to 1, and equals about 5000 time slots.

6. Conclusions

This paper aims to evaluate performance of the SMM Clos-network under the packet dispatching scheme employing static connection patterns, called SD. The system was evaluated in terms of stability and basic performance measures, such as average cell delay and packet queue lengths. In Section 4 we showed how to use the DTMC model and an analytical approach based on Foster's stochastic criteria, analogous to the direct Lyapunov's method, to prove the stability of the SMM Clos-network switch under the SD algorithm. Taking into account that the stability is proven for ideal, theoretical traffic, in Section 5 we showed simulation results obtained for uniform and non-uniform traffic distribution patterns, and for Bernoulli and bursty arrival models. Two architectures of the SMM Clos-network switch were taken into account: RRNB $C(8,8,8)$ and SSNB $C(8,15,8)$. The investigated cell dispatching scheme is very simple, but it is not able to provide satisfactory performance of the RRNB SMM Clos-network switch for very high input load, greater than 0.7, especially for bursty traffic The results are better for the SSNB architecture, but in this case more switches in the second stage must be used, and the cost of such a network will be higher. It is also impossible to provide in-sequence service under this algorithm, which results in special resequencing buffers at outputs. Furthermore, this re-sequence function makes a switch more difficult to implement, especially as the port speed and switch size increase.

Acknowledgements

The work described in this paper was financed with the use of the funds of the Ministry of Science and Higher Education for the year 2017 under Grants 08/82/DSPB/8221 and 04/45/DSPB/0162.

References

[1] H. J. Chao and B. Liu, *High Performance Switches and Routers. Wiley-Interscience*, New Jersey, USA: Wiley, 2007.

[2] Router-Switch.com, Cisco CRS-X Core Router to Offer 10 Times Capacity of Original [Online]. Available: http://blog.router-switch.com/2013/06/cisco-crs-x-core-router-to-offer-10-times-capacity-of-original/ (accessed on February 14 , 2018).

[3] C. Clos, "A Study of Non-Blocking Switching Networks", *Bell Sys. Tech. Jour.*, vol. 32, no. 2 pp. 406–424, 1953.

[4] W. Kabaciński, *Nonblocking Electronic and Photonic Switching Fabrics,* Berlin: Springer, 2005.

[5] V. E. Beneš, *Mathematical Theory of Connecting Networks and Telephone Traffic,* New York: Academic Press, 1965.

[6] V. E. Beneš, "Semilattice characterization of nonblocking networks". *The Bell System Techn. J.*, vol. 52, no. 5, pp. 697–706, 1973.

[7] H. M. Ackroyd, "Call repacking in connecting networks", *IEEE Transact. on Commun.*, vol. 27, no. 3, pp. 589–591, 1979.

[8] A. Jajszczyk and G. Jekel, "A New Concept – Repackable Networks", *IEEE Transact. on Commun.*, vol. 41, no. 8, pp. 1232–1237, 1993.

[9] E. Oki, Z. Jing, R. Rojas-Cessa, and H. J. Chao, "Concurrent round-robin-based dispatching schemes for Clos-network switches", *IEEE/ACM Transact. on Network.*, vol. 10, no. 6, pp. 830–844, 2002.

[10] J. Kleban and A. Wieczorek, "CRRD-OG: A Packet Dispatching Algorithm with Open Grants for Three-Stage Buffered Clos-Network Switches", in *Proc. 2006 Workshop on High Performance Switching and Routing HPSR2006*, Poznań, Poland, 2006, pp. 315–320.

[11] J. Kleban, "Packet dispatching using module matching in the modified MSM Clos-network switch", *Telecommun. Sys,*, vol. 66, no. 3, pp 505–513, 2017.

[12] X. Li., Z. Zhou, and M. Hamdi, "Space-Memory-Memory architecture for Clos-network packet switches". in *Proc. IEEE Int. Conf. on Commun. – ICC 2005*, Seoul, Korea (South), 2005, vol. 2, pp. 1031–1035.

[13] A. V. Manolova, S. Ruepp, A. Rytlig, M. Berger, H. Wessing, and L. Dittmann, "Internal backpressure for terabit switch fabrics", *IEEE Commun. Let.*, vol. 16, no. 2, pp. 265–267, 2012.

[14] J. Kleban and U. Suszyńska, "Static Dispatching with Internal Backpressure Scheme for SMM Clos-Network Switches", in *Proc. The Eighteenth IEEE Symp. on Computers and Commun., ISCC'13*, Split, Croatia, 2013, pp. 654–658 (doi:10.1109/iscc.2013.6755022).

[15] K. Yoshigoe, "The Crosspoint-Queued Switches with Virtual Crosspoint Queueing". in *Proc. 5th Int. Conf. on Signal Proces. and Commun. Sys. ICSPCS 2011)*, Honolulu, HI, USA, 2012, pp. 277–281.

[16] K. Liu, J. Yan, and J. Lu, "Fault-tolerant Cell Dispatching for Onboard Space-Memory-Memory Clos-Network Packet Switches", in *Proc. 16th Int. Conf. on High Performance Switching and Routing HPSR*, Budapest, Hungary, 2015 (doi:10.1109/HPSR.2015.7483090).

[17] J. Kleban and H. Santos, "Packet Dispatching Algorithms with the Static Connection Patterns Scheme for Three-Stage Buffered Clos-Network Switches", in *Proc. IEEE Int. Conf. on Commun. 2007 ICC-2007*, Glasgow, United Kingdom, 2007 (doi:10.1109/ICC.2007.1046).

[18] A. M. Lyapunov, "The General Problem of the Stability of Motion", *Int. J. of Control*, vol. 55, no. 3, pp. 531–773, 1992 (doi: 10.1080/00207179208934253).

[19] J. Kleban and J. Warczyński, "Stabilność buforowanych pól komutacyjnych Closa", *Przegląd Telekomunikacyjny i Wiadomości Telekomunikacyjne*, no. 8–9, pp. 976–981, 2016 (in Polish).

[20] F. G. Foster, "On the stochastic matrices associated with certain queuing processes". *Ann. Math. Statistics.* vol. 24, np. 3, pp. 355–360, 1953.

[21] S. Meyn and R. Tweedie, *Markov Chains and Stochastic Stability*, New York, USA: Springer, 1993.

[22] N. McKeown, A. Mekkittikul, V. Anantharam, and J. Walrand, "Achieving 100% Throughput in an Input-queued Switch", *IEEE Transact. on Commun.,* vol. 47, no. 8, pp. 1260–1267, 1999 (doi: 10.1109/26.780463).

Janusz Kleban is an Assistant Professor at the Poznan University of Technology (PUT), the Chair of Communication and Computer Networks of the Faculty of Electronics and Telecommunications. He received his M.Sc. and Ph.D. degrees in telecommunications from PUT in 1982 and 1990, respectively. His scientific interests include packet dispatching algorithms for single and multistage switching fabrics, photonic broadband switch architectures, optical switching systems, control algorithms for lightwave networks, and Network on Chip (NoC).
E-mail: janusz.kleban@put.poznan.pl
Faculty of Electronics and Telecommunications
Poznan University of Technology
Poznań, Poland

Jarosław Warczyński received his M.Sc. degree in Electrical Engineering (Control Engineering) in 1978 and the Ph.D. degree in Computer Since (Operations research) in 1983 from Poznan University of Technology, Poland. He is a faculty member at the University's Institute of Control, Robotics and Information Engineering. His general areas of research are in operations research, robotics, control engineering and currently in high speed packet switching and routing in Clos-network switches.
E-mail: jaroslaw.warczynski@put.poznan.pl
Faculty of Electrical Engineering
Poznan University of Technology
Poznań, Poland

Paper

Ganging of Resources via Fuzzy Manhattan Distance Similarity with Priority Task Scheduling in Cloud Computing

S. Sharon Priya[1], K. M. Mehata[2], and W. Aisha Banu[1]

[1] *Department of Computer Science and Engineering, B. S. Abdur Rahman University, Chennai, Tamil Nadu, India*
[2] *School of Computer, Information and Mathematical Sciences at B. S. Abdur Rahman University, Chennai, Tamil Nadu, India*

https://doi.org/10.26636/jtit.2018.108916

Abstract—This paper proposes a fuzzy Manhattan distance-based similarity for gang formation of resources (FMDSGR) method with priority task scheduling in cloud computing. The proposed work decides which processor is to execute the current task in order to achieve efficient resource utilization and effective task scheduling. FMDSGR groups the resources into gangs which rely upon the similarity of resource characteristics in order to use the resources effectively. Then, the tasks are scheduled based on the priority in the gang of processors using gang-based priority scheduling (GPS). This reduces mainly the cost of deciding which processor is to execute the current task. Performance has been evaluated in terms of makespan, scheduling length ratio, speedup, efficiency and load balancing. CloudSim simulator is the toolkit used for simulation and for demonstrating experimental results in cloud computing environments.

Keywords—clustering, pre-processing, quality of service, resource allocation.

1. Introduction

In recent years, cloud computing has been used for developing a manner for providing assorted services and computational resources to the customers [1]. The Internet and central remote servers were used in cloud computing to produce scalable services for its customers. There is a requirement to establish an autonomic resource management approach that enhances both QoS targets: resource centric (reliability, availability, utilization) and user centric (budget, implementation time) [2]. Task scheduling is performed by the cloud computing platform. Firm resources, computational power and allocated tasks are used to restrict the task node. The job management node has various subtasks when a cloud computing task is assigned and stored in the task pool [3].

The process of resource allocation in cloud computing may be divided into two categories. The first one is static allocation. In this type, the cloud customer needs to create a prior demand for the resources. In this scenario the customer recognizes which resources are essential and in how many cases the resources are required. All that is performed prior to employing the system. Such an approach leads to overutilization or underutilization of resources, depending on the time of application. This is one of the disadvantages of using static allocation [4].

The other type is dynamic allocation. Here, cloud resources are required by the cloud customer when there is a demand for application. At this point, overutilization and underutilization may be prevented [5]. The service provider has to allocate resources from an alternative cloud information center [6], [7]. Cloud computing performs most important tasks between virtual machine (VM) placements, which is considered to be one of the most severe problems. This method will select the perfectly suited physical hosts with respect to energy efficiency and resource consumption – from the point of view of the cloud provider.

Calculations concerned with the capacity of each resource node are commonly unstable in heterogeneous or homogeneous multi-cluster and multi-resource environments [8]. Therefore, cloud computing scenarios need a consistent job scheduling scheme and a suitable resource management model [9], [10].

Scheduling algorithms which are commonly run on dedicated and homogeneous resources, in parallel, and on distributed systems, cannot perform well in the new cloud computing scenarios [11]. Transmission delay is a significant issue that disturbs the scheduling algorithm in cloud task scheduling [12], [13]. The creation and full usage of resources are also significant factors during scheduling [14]–[17].

Gang scheduling is used in distributed systems, and it is a very useful function [18]. The scheduling algorithm is necessary for assigning processors to the current activities in a distributed system. A scheduling algorithm assigns the functions that need to be performed simultaneously in the case of similar events, meaning mainly in communication operations. It is a powerful method for scheduling such activities, and it depends on time-space sharing [19]. The major role of the scheduling policy of gang scheduling algorithms is to collect the functions of a parallel job, and to perform them instantaneously on various proces-

sors. So, the threat of having wait for a function that is not currently working on any processor is eliminated [20]. In light of the above, one may notice that the quantity of functions within a gang cannot surpass the number of accessible processors.

The main objective of this paper is to address an efficient resource allocation aspect of cloud computing for selecting proper resources from a specific gang. It supports task scheduling with the aim of reducing the selection time. In this paper, a novel fuzzy Manhattan distance similarity-based ganging of resources (FMDSGR) technique, developed for gang formation of resources, with its characteristics aiming to attain better resource scheduling results in cloud computing, is discussed. Then, gang priority scheduling (GPS) is proposed to schedule tasks within a gang of processors.

The subsequent parts of the paper are ordered as follows. Section 2 reviews the various related works. Section 3 presents proposals for effective pre-processing of resources and task scheduling. The results are presented and discussed in Section 4. Lastly, the relevant research work is described in Section 5.

2. Related Work

In manufacturing relying on cloud systems, there is a considerable amount of assets which have comparable useful qualities, as indicated by the requirements of the manufacturing task at hand. Step-by-step instructions to choose the ideal assets series or assets combinations to finish the manufacturing task with a number of distributed subtasks was a key consideration in the investigation of cloud manufacturing. Cao *et al*. [21] had developed a multivariate process capability indicator to assess the manufacturing process with the unwavering quality of information on cloud scheduling. As indicated by the fuzzy quality theory, the strength level of intuitionistic fuzzy esteem was considered in decision making while selecting the ideal asset chain. Another technique was proposed for ideal determination of assets based on the multivariate process indicator and on the predominance level of intuitionistic fuzzy esteem. A practical contextual investigation was relied upon to outline the proposed strategy and technique.

Due to the burstiness of VM in computing clouds, real application workloads are characterized by spikes that are more often characterized by periodic occurrence, low frequency and a brief span. Zang *et al*. [22] explored the burstiness-mindful server solidification issue from the viewpoint of asset reservation, i.e. holding an accurate measure of additional assets on every physical machine (PM) to maintain a strategic distance from live movements. The creator first modeled the asset necessity pattern of each VM as a two-state Markov chain to catch burstiness, then an asset reservation system was intended for every PM, based on the stationary distribution of the Markov chain.

With the promotion and advancement of cloud computing, loads of logical computing applications have been developed in cloud scenarios. A general elastic resource provisioning and task scheduling mechanism performing the logical work in the cloud was used to address this issue. Shi *et al*. [23] had proposed elastic resource provisioning to perform logical work process employments in the cloud environment. The issue of elastic resource provisioning was displayed as a problem that consisted of elastic resource provisioning algorithms related to VM occupancy.

Saraswathi *et al*. [24] had proposed a technique that concentrated on the assignment of VM to the customer. This work sets a fundamental standard for low priority occupations (due date of the employment was high), without postponing the execution of high priority occupations (due date of the employment was low) and progressively allocated VM assets to the customer who was able to complete the work within the deadline.

Cloud computing offers the ability to share resources over various geographical destinations. Cloud resource scheduling has been a task that needs to be performed on a repetitive basis, as the issue of finding the best match for the workload needs to be tackled. Proficient management of the dynamic nature of resources should be possible with the assistance of cloud workloads. Very few effective resource scheduling strategies for energy, cost and time-imperative cloud workloads have been accounted for in writing. Singh *et al*. [25] had proposed an effective cloud workload administration framework, in which cloud workloads have been recognized, dissected and clustered through K-means on the premise of weights relegated and their QoS necessities. The execution of the proposed algorithm had been assessed with the existing scheduling strategies through the CloudSim toolkit. While clustering the cloud workloads utilizing K-means, diverse preparatory allotments can bring various final clusters, and it does not work well with clusters of different sizes and clusters prone to minimization.

Sfrent and Pop [26] dealt with the problem of scheduling a set of jobs across a set of machines and specifically analyzed the behavior of the system at very high loads, which was specific to big data processing. Under certain conditions, they could easily discover the best scheduling algorithm, prove its optimality and compute its asymptotic throughput. Vasile *et al*. [27] proposed a scheduling algorithm for different types of computation requests: independent tasks, like a bag of tasks model, or tasks with dependencies modeled as directed acyclic graphs, and they will be scheduled for execution in a cloud data center. The tasks in the requests are scheduled based on the available resources using the scheduling algorithm suitable for each request.

Vasile *et al*. in [28] proposed a resource-aware hybrid scheduling algorithm for different types of applications, such as batch jobs and workflows. The proposed algorithm considers hierarchical clustering of the available resources into groups in the allocation phase. Task execution was performed in two steps. In the first phase, tasks were assigned to groups of resources and in the second phase, a classical scheduling algorithm was used for each group

of resources. The proposed algorithm was suitable for heterogeneous distributed computing, especially for modern high-performance computing systems in which applications were modeled with various requirements (both IO and computationally intensive), with a particular emphasis on data from multimedia applications.

Li [29] approach was to design and analyze the performance of heuristic algorithms based on the equal-speed method. Pre-power and post-power determination algorithms were developed for both energy and time-constrained scheduling of precedence-constrained parallel tasks on multiple-core processors with continuous or discrete speed levels.

Rimal and Maier in [30] proposed a new cloud-based workflow scheduling policy for compute-intensive workflow applications in multi-tenant cloud computing environments, which helps minimize the overall workflow completion time, tardiness, cost of execution of the workflows, and utilize idle cloud resources in an effective manner. The proposed algorithm was compared with the state-of-the-art algorithms, i.e. first come first serve (FCFS), easy backfilling, and minimum completion time (MCT) scheduling policies to evaluate the performance levels.

Keshanchi *et al*. [31] proposed a powerful and improved genetic algorithm to optimize task scheduling solutions. The proposed algorithm uses the advantages of evolutionary genetic algorithm and heuristic approaches.

Taking into consideration the work mentioned above, we can say that there is a need to develop a resource management technique that optimizes both QoS targets: user-centric (cost and execution time) and resource-centric (reliability, availability, and utilization). Due to unavailability of the required resources at a point in time, some sub-tasks are stuck in a deadlock condition and continue to struggle for resources, which further leads to customer dissatisfaction (increasing execution time and cost). The QoS-aware resource management technique needs to be decentralized. In centralized distributed systems, it is tough to manage large numbers of user requests in multiple service queues, which further leads to performance degradation (decreased reliability and scalability). Due to the difficulty in predicting the behavior (regarding QoS requirements) and demand (in terms of resources) of the workload/application, there is a need for an effective QoS-aware resource management technique that can easily make the right decision concerning the dynamic scaling of resources.

3. Proposed Resource Pre-processing and Task Scheduling

The proposed work is divided into in two stages: pre-processing and task scheduling. Fuzzy clustering is done in the first stage, where resource characteristics, such as processing performance, average communication ability, maximum transmission capacity, network position, and many links are grouped into a gang of resources. Resources with a similar computing capability are grouped into one gang

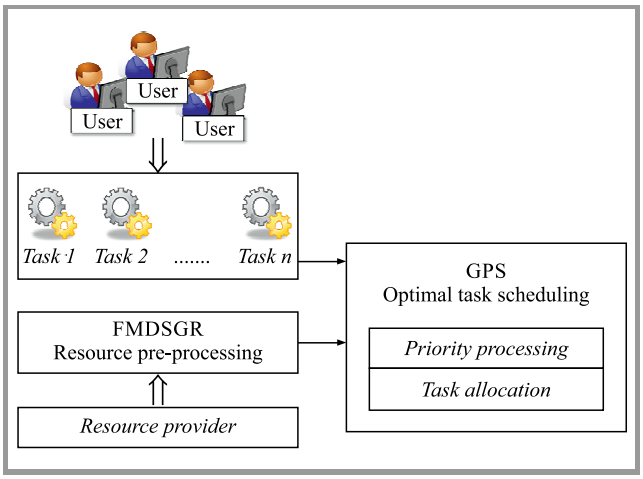


Fig. 1. Architecture of the proposed scheme.

by separating all resources into many gangs based on the Manhattan distance similarity. The resources in one gang have a similar data transfer rate, as they share a similar network for communicating with each other. In the second stage, the tasks from the users are scheduled based on gang priority scheduling in order to minimize the scheduling length or makespan. The architecture of the proposed work is given in Fig. 1.

3.1. Resource Pre-processing using FMDSGR algorithm

There are many typical workflow task scheduling algorithms which are deployed in a heterogeneous environment. Those methods are based on quantitative characteristics of the task and do not consider the service-oriented resources. Moreover, it is hard to describe exactly the task's demand for resources. Meanwhile, resource attributes cannot be described accurately. The fuzzy theory provides various effective means to solve the uncertain problems in the real world. Therefore, fuzzy clustering is used to divide the resources, and the Manhattan distance is used to group the resources into gangs based on similarity to improve the efficiency of task scheduling.

Also, computation and communication ability of resources allocated by the task affects the completion of the subsequent task. Therefore, distinguishing the performance of resources is conductive to choosing the appropriate resources for workflow task scheduling. However, it is difficult to describe the attributes of resources accurately, and there are no attributes to distinguish the processing units strictly. Fuzzy clustering is an effective method to divide resources.

Design objective: The purpose of the resource pre-processing method is to minimize the distance vector of the processor's characteristics to achieve effective utilization of resources. We can dynamically group the resources as gangs during scheduling in cloud systems, partitioning the given data by minimizing the distance objective function:

$$f(x) = \min(M_t) \in \varepsilon\,. \tag{1}$$

We have to gang the processing unit with minimum Manhattan distance M_t of the n-th characteristic feature t dependent upon the threshold value ε.

Initialization of resources: Consider the original dataset U with the processing unit $\{p_1, p_2, \ldots, p_M\}$ in which each processor has its separate characteristics $u = \{t_1, t_2, \ldots, t_N\}$. In which, p_M is the x-th processor of the M-th process unit and t_N is the y-th characteristic value in the N-th characteristic elements. The performance of the scheduling process is improved by forming the resources in the cloud system as a gang. Qualitative characteristics refer to qualities or properties of cloud computing resources.

Data standardization and extreme standardization: We obtain the standardization data U' to deal with the data U in the target system using mean and standard deviation. The standardized value u_{ik} of each data is equal to:

$$u'_{xy} = \frac{u_{xy} - \mu_{xy}}{\sigma_{xy}}, \tag{2}$$

where u_{ik} denotes the k-th eigenvector of original data. In the Eq. 2, $\mu_{xy} = \frac{1}{MN}\sum_{x=1}^{M}\sum_{y=1}^{N} u_{xy}$ is the mean value of u_{xy} and $\sigma_{xy} = \sqrt{\frac{1}{MN}\sum_{x=1}^{M}\sum_{y=1}^{N}(u_{xy} - \mu_{xy})^2}$ is the standard deviation of u_{xy}. The final value of the standardized data $u'_{xy}$is not in the range $[0, 1]$. Therefore, in order to normalize the U' to U'' an extremely standardized method is presented. The extreme standardized method is defined as:

$$u''_{xy} = \frac{u'_{xy} - u'_{y\min}}{u_{y\max'} - u'_{y\min}}, \tag{3}$$

where $u'_{y\min}$ is the minimum value in $u'_{1y}, u'_{2y}, \ldots, u'_{Ny}$ and $u'_{y\max}$ is the maximum value in $u'_{1k}, u'_{2k}, \ldots, u'_{Nk}$.

Fuzzy similarity matrix: The correlation coefficient between the elements in the fuzzy matrix includes the similar index coefficient method. The similar index coefficient can be calculated as:

$$I_{xy} = \frac{1}{N}\sum \mathrm{e}^{\frac{3(u_{xk}-u_{yk})^2}{4\sigma''^2_{xk}}}, \tag{4}$$

where n denotes the number of characteristic features. The values of x and y belongs to the processing units. The value of k fits with the characteristic feature.

Consequently, a fuzzy similarity matrix is obtained as a result of the correlation coefficient between the elements and is expressed as:

$$F = [I_{xy}]_{2(n+1)} = \begin{bmatrix} I_{11} & I_{12} & \cdots & I_{1(n+1)} \\ I_{21} & I_{22} & \cdots & I_{2(n+1)} \\ \vdots & \vdots & \vdots & \vdots \\ I_{(n+1)1} & I_{(n+1)2} & \cdots & I_{2(n+1)} \end{bmatrix}_{2(n+1)}. \tag{5}$$

Then, fuzzy relation F among x and y is established as a fuzzy subset of $x \times y$, in which I_{xy} is the membership function in the interval $[0, 1]$. The fuzzy similarity between the processing units is referred to as I_{xy} in the formula:

$$I_{xy} = \begin{cases} 1 & x = y \\ [0, 1] & x \neq y \end{cases}. \tag{6}$$

The fuzzy similarity is based upon the fuzzy relations, such as equivalence relation, fuzzy relation and similarity relation, which are necessary for reflexivity and symmetry.

Manhattan distance based gang formation: The process of grouping the elements or objects into clusters for a problem with a certain objective is called gang formation. It can expose formerly hidden relations in a multifaceted data set. Finding the similarity between two objects is the main issue in gang formation, so that the gang can be formed with a high similarity between the objects. Manhattan distance computes the absolute differences between the coordinates of a pair of objects. This means that the distance between two items is the total sum of the differences of their parallel elements. Figure 2 presents the process flow of the proposed FMDSGR.

The similarity between the data items of a gang is measured using the Manhattan distance M. The formula for this distance between a standardized data item and a center point is:

$$M_t = \sum_{x=1}^{M}\sum_{y=1}^{N} \left\| I'_{xy} - c_t \right\|. \tag{7}$$

In Eq. (7) $c_t = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N} u''_{xy}}{MN}$ represents the center point of the t characteristic value calculated after the extremely standardized output.

The maximum number of center points depends upon the minimum value of the objective function. Then, the gangs are formed using those center points represented as a gang. Members other than the center point could not be assigned to any gang. Therefore, they belong to two or more gangs.

After forming the gang, we have to decide, which gang executes the task. Hence, we have to add the separate characteristics of each gang and sort it in the descending order. We can decide, which gang processes the task based on the arrangement.

The objective of this algorithm is to reduce the completion time of every task in the workflow, whereas the start time and the finish time are the two important factors affecting the completion time. If the task is allocated to the different processing units, the execution implementation is also different. We schedule each task based on its suitable processing unit, which minimizes the finishing time of the task.

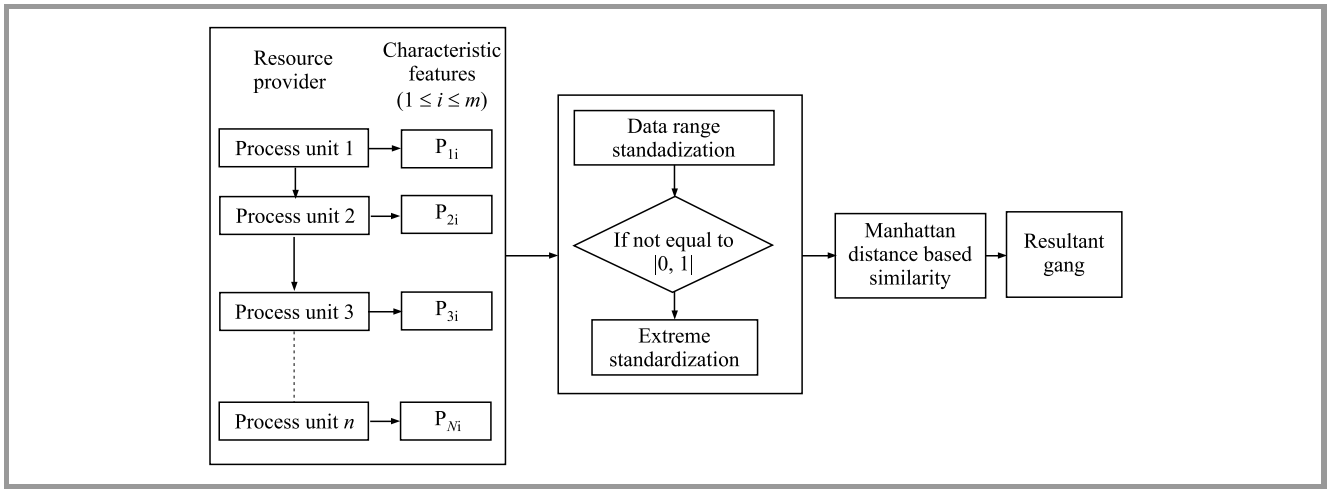


Fig. 2. Process flow of proposed FMDSGR.

Algorithm 1: FMDSGR flowchart

Input: N number of processors with characteristics
Output: gang
The algorithm comprises the following steps:

1. Random selection of the gang center
2. Initialization of resources $U = [u_{ij}]$ matrix, $U^{(0)}$
3. Data standardization
$$u'_{xy} = \frac{u_{xy} - u_{xy}}{\delta_{xy}}$$
4. Extreme standardization
$$u''_{xy} = \frac{u'_{xy} - u'_{y\min}}{u'_{y\max} - u'_{y\min}}$$
5. Fuzzy similarity matrix
$$F^e = [I'_{xy}]_{(n+1)\times(n+1)} = \begin{bmatrix} I'_{11} & I'_{12} & \cdots & I'_{1(n+1)} \\ I'_{21} & I'_{22} & \cdots & I'_{(n+1)(n+1)} \\ \vdots & \vdots & \vdots & \vdots \\ I'_{(n+1)1} & I'_{(n+1)2} & \cdots & I'_{2(n+1)} \end{bmatrix}_{(n+1)\times(n+1)}$$
6. At k-step: calculate the centers vectors $C^{(k)} = [c_t]$ with $U^{(k)}$
$$c_t = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N} u''_{xy}}{MN}$$
7. Update $U^{(k)}$, $U^{(k+1)}$
$$M_t = \sum_{x=1}^{M}\sum_{y=1}^{N} \|I'_{xy} - c_t\|$$

If $\|U^{(k+1)} - U^{(k)}\| < \varepsilon$ or the minimum objective is achieved, then Stop. Otherwise return to step 2.

3.2. Task Scheduling Using GPS Algorithm

Gang based priority scheduling is used to reduce the scheduling length (makespan) of the task scheduling workflow. The scheduling length gets reduced every time the task is performed and a ready task list (RTL) is established for priority allocation. If there is any task in the RTL that denotes the scheduled tasks of the parent node, the priority of the task is based up on the current-to-exit length (CEL) value. It represents the largest distance from the present node to the existing node:

$$CEL(t_x) = \frac{W(t_x)}{M_p} + \max_{t_y \in succ(t_x)} \left\{ \frac{T_{xy}}{M_c} + CEL(t_y) \right\}, \qquad (8)$$

where M_p is the median of the computing ability of the processing units, M_c is the median of the transmission ability between processing units, T_{xy} is the inter-task communication between t_x, and $t_y \in succ(t_y)$ is the closest successor or child of the task t_x. The larger the CEL, the greater its priority. If greater the attribute priority value of a task means priority is also greater. After that, group the workflow task into gangs or jobs based on their priority.

Let us consider all the tasks within the same job being allocated at the same time. The job consists of a set of tasks called gangs and each task in the gang starts at the same time. Assume that every gang G consists of x tasks and $1 \le x \le \frac{P}{s}$, in which $s = 2$ and P denotes the processor. Gang thus formed by cluster the task of parallel jobs and allocate the tasks simultaneously to different processors. The degree of parallelism of a gang is the number of the task in a job. The size of the gang is equal to the number of tasks in a gang G. If p is the number of processors required by gang G, then $1 \le x \le p \le \frac{P}{s}$. Before scheduling, we should know that a small gang requires a limited number of the processing units, and a large gang requires a large processing unit.

3.2.1. Gang Size Dissemination

The number of tasks of the gang (job) is evenly disseminated with in the range $\lfloor 1, \ldots, \frac{P}{s} \rfloor$. The mean of the gang size is represented as:

$$\gamma = \left\lceil \frac{1 + \frac{P}{s}}{2} \right\rceil. \qquad (9)$$

The number of jobs that can be processed in parallel depends on the gang size and the scheduling policy. The

scheduling algorithm selects the task in the gang which is on the top of RTL. The tasks will be released after executing each task from the first gang, so the RTL is updated after each task allocation process. The task from the gang with the highest priority completes the task with a minimum completion time during scheduling. The processor with the nearest center unit is selected when two processors have the same completion time.

3.2.2. Computing the Primitive Begin Time (PBT)

While allocating tasks to the suitable processor, in order to achieve the smallest PBT value, calculate all the processing units' PBT assigned by t_i to decide about the processing unit p_1.

$$PBT(t_x, p_y) = \max\left\{RFT(t_x, P_y), \max_{t_y \in PRED(t_x)}(AT(t_x, t_y, p_y)\right\}, \tag{10}$$

where $pred(t_x)$ is the nearest predecessor of t_x, $RFT(t_x)$ is the real finish time of the task t_x.

$$RFT(t_x, p_y) = RST(t_x, p_y) + \frac{w(t_x)}{w(p_y)}. \tag{11}$$

When $t_y \in PRED(t_x)$ and tasks t_x and t_y on p_x and p_y. $w(t_x)$ is the computation time of the task t_x, the processing unit p_y computation time is represented as $w(p_y)$. The obtained time on t_x and t_y on p_x and as p_y is:

$$AT(t_x, p_y) = RST(t_x, p_y) + \frac{C(t_x, t_y)}{C(p_x, p_y)}. \tag{12}$$

In Eq. (12) $C(t_x, t_y)$ is the inter-task communication time between t_x and t_y, $C(p_x, p_y)$ is the communication time of the processing unit between p_x and p_y.

3.2.3. Computing the Primitive Completion Time

In the same way, the processing unit p_2 is selected and provides the minimum PCT value:

$$PCT(t_x, p_y) = PBT(t_x, p_y) + w_{xy}, \tag{13}$$

where w_{xy} is the computing cost of the processing unit.

If the selected processing units are the same, then each task from the high priority gang is allocated to the first processor. Then, simultaneously update the LRT. If the selected processing unit is different, then select the nest gang of the task in the RTL with a low priority. Continue the step until the processors p_1 and p_2 become equal, and simultaneously update RTL.

The task with the highest priority is scheduled on the processing unit that can complete the task with the minimum completion time. When there is no vacancy in the optional set, we choose the processing unit of the cluster with the highest comprehensive performance in the backup set and calculate the completion time of the current task on this processing unit. If the completion time of this processing unit in the backup set is lower than the smallest completion time of the processing unit in an optional set, we put the current task schedule on the one in the backup set. If there are two processing units that have the same completion time, we can choose the processing unit whose network location is nearer to the center unit. Because of the consideration of communication cost, whenever possible, we place the tasks on the critical workflow path of the same processing unit.

Algorithm 2: Gang based priority scheduling

Input: Current LRT[LRT= $G_1, G_2, \ldots, G_n$];
G_1 – largest gang, G_n – smallest gang;
$G = p_1, \ldots, p_n$
Task: Still exist tasks that haven't been allocated,
$T = t_1, t_2, \ldots, t_n$
LRT [head]: Head of the LRT
LRC [tail]: Tail of the LRT

Output: Refresh LRT to improve the resources scheduling performance

```
1: Cluster the processing unit with fuzzy theory
2: Compute each node's CEL
3: Establish the LRT
4: Choose the first LRT
5: While (Gang) do {
6: If (LRC→LRT [head]) Then {
        Compute the EST and choose the minimum
        value processing unit p1.
        Compute the EFT and choose the minimum
        value processing unit p2.
7: If (p1 = p2) Then {
                    Allocate each task of ti to p1
                    Remove task of ti from LRT,
                    update the LRT
              Else
                    To the next task in LRT
                  }
                    Else
                    //LRC→LRT[tail]
                    Go back to the first task
                    in LRT,
                    Compute the difference, and
                    Choose the processing unit.
8: Allocate ti to pi and delete ti from LRT
                  }          }
```

4. Results and Discussion

The experimental results were achieved by using the CloudSim toolkit as a simulation platform to simulate heterogeneous cloud environments. CloudSim [26] was also proposed by R. Buyya, extending the range of features of GridSim end enabling, modeling and simulating cloud environments, data centers, virtual machines, cloudlets, etc.

Five characteristic features have been employed in this paper to improve scheduling performance in the resource network:

- processing performance t_1: the average computing ability among processing units in the resource system, it represents the average time per task executed by a processing unit. If the resultant value is small, this means that the processing unit costs are also lower during computation;
- average communication ability t_2: the average communication ability value of the processor-connected links that can be calculated from the average weight of edges connected to the processing elements;
- maximum transmission capacity t_3: the maximum number of edges connected to a processing unit with the transmission capacity;
- network position t_4: position of the processing element in the network, defined as the network position. If the processing unit is far away from the center of the network, then the value is greater;
- number of links t_5: the number of links that are connected to a processing element.

Table 1 shows the characteristics of the original data U for the processing units. The cloud resources should be formed as gangs to use the resources effectively and to divide the processing units.

Table 1
Original dataset

Processing unit	Characteristic features				
	t_1	t_2	t_3	t_4	t_5
P_0	20	0.05	0.05	6	1
P_1	30	0.08	0.05	9	2
P_2	35	0.1	0.1	10	2
P_3	50	0.13	0.1	14	4
P_4	45	0.13	0.1	13	4
P_5	50	0.13	0.1	14	3
P_6	30	0.1	0.1	9	1
P_7	28	0.1	0.1	8	1
P_8	48	0.11	0.03	13	5
P_9	33	0.1	0.1	9	1
P_{10}	30	0.08	0.08	8	1
P_{11}	35	0.1	0.1	10	1
P_{12}	40	0.12	0.1	11	1

In the proposed work, first the original data U can be standardized. Then, the similarity is calculated based on the Manhattan distance, and we obtain the overall gang formation results as gang 1: $\{P_0, P_2, P_6, P_7, P_7, P_9, P_{10}, P_{11}, P_{12}\}$ and gang 2: $\{P_1, P_3, P_4, P_5, P_8\}$. The first gang result forms gang 1, and the next gang result forms gang 2.

The proposed work ensures better performance when the data sets are different or detached from one another. Manhattan distance-based similarity of the presented method is compared with the arithmetic mean method in Table 2. The successful result has not been achieved due to the random selection of gang center in the arithmetic mean method, and this method does not work well for categorical data, i.e. it is applicable only when the mean is defined.

Table 2
Comparison of gangs in Manhattan distance and arithmetic mean methods

Manhattan distance (MFC)		Arithmetic mean	
Gang 1	Gang 2	Gang 1	Gang 2
P_0	P_1	P_0	P_3
P_2	P_3	P_1	P_4
P_6	P_4	P_2	P_5
P_7	P_5	P_6	P_6
P_9	P_8	P_7	–
P_{10}	–	P_9	–
P_{11}	–	P_{10}	–
P_{12}	–	P_{11}	–
–	–	P_{12}	–

In order to decide which gang to choose to execute the task at hand, we have to add any of the characteristics of the processor in the gang and sort the results in the descending order. In this paper, we have used the average communication ability t_2, and the maximum transmission capacity t_3 as our characteristics, and added all t_2 and t_3 features of the gang 1 and gang 2 separately.
The total average communication ability equals 0.74 for gang 1 and 0.59 for gang 2. The total maximum transmission capacity amounts to 0.75 for gang 1 and 0.38 for gang 2.
After calculating the characteristics, gangs with high average communication and transmission capacity values have been sorted in the descending order as [0.75, 0.59] and [0.75, 0.38]. From this, we can decide that gang 1 is to to execute the task. Thus, the Manhattan distance based fuzzy clustering (MFC) is an effective technique for gang formation enabling a considerable reduction in the cost of deciding which processor is to execute the task and the selection of the gang depends on the performance characteristics.
The various performance metrics to compute the results are:

Makespan: makespan (M) is referred to as the scheduling length. It is measured by calculating the ending time of the final task using the proposed algorithm:

$$M = \max\left\{PCT(t)\right\}, \tag{14}$$

in which $PCT(t)$ is the primitive completion time of the task.

Scheduling length ratio: the scheduling length ratio (SLR) in the lower bound is the time taken to execute the task on a serious path. SLR is defined to normalize the schedule length:

$$SLR = \frac{M}{\sum\limits_{t_i \in CP_{\min}}^{n} \min\limits_{p_j \in P}\{w_{i,j}\}}. \tag{15}$$

SLR is defined as the ratio of makespan to the total sum of minimum computation cost $CP_{\min}$ of all tasks. The resultant graph value of SLR is not less than one. When the SLR output is low, then the scheduling algorithm provides better performance.

Speed up: the ratio of parallel execution time and the sequential execution time is the speedup (S). The scheduling length on a limited number of processors is the parallel execution time and the sum of entire computation time of every task is the sequential execution time:

$$S = \min_{p_j \in P} \left\{ \frac{\sum\limits_{i=1}^{n} w_{i,j}}{M} \right\}, \tag{16}$$

where the total sum of the computational time of tasks is defined as $\sum\limits_{i=1}^{n} w_{i,j}$, $i = 1, 2, 3, \ldots, n$.

Efficiency: the efficiency is the measure of the utilization of the processor of a parallel program:

$$\eta = \frac{S}{NP}, \tag{17}$$

where S stands for speed up and NP is the number of processors.

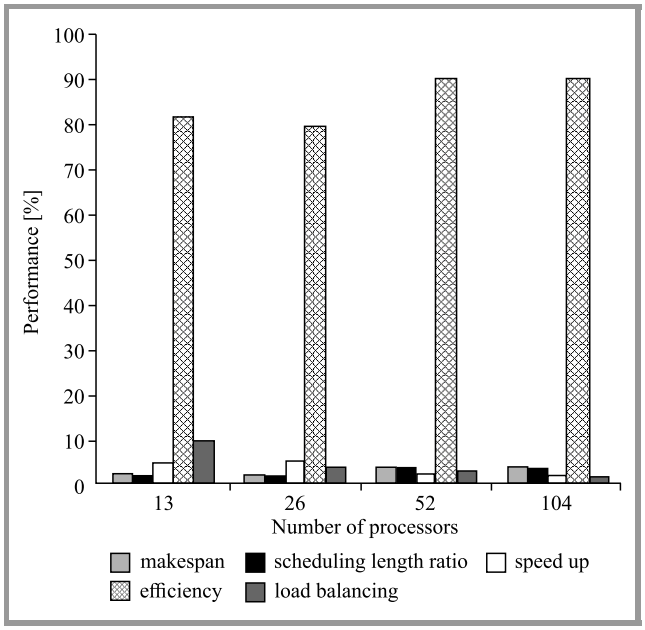


Fig. 3. Performance comparison of the proposed GPS method for different number of processors.

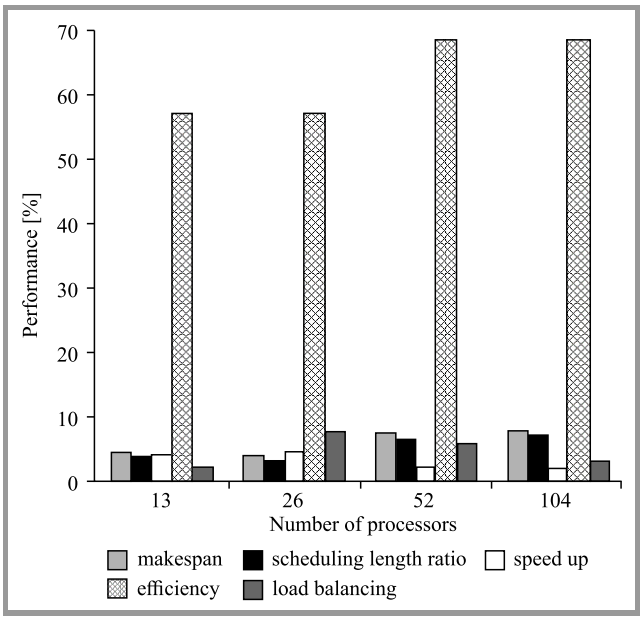


Fig. 4. Comparison of the existing LJFS scheduling.

Load Balancing: the ratio of the overall makespan of processors and the average execution time is the measure of load balancing (LB):

$$LB = \frac{M}{A}, \tag{18}$$

where M is scheduling length. The average A is taken from the ratio of a sum of the processing time of each processor and the number of processors used.

The proposed fuzzy-based gang priority scheduling method is compared with the largest come first serve (LJFS), Yarn, and Mesos scheduling method.

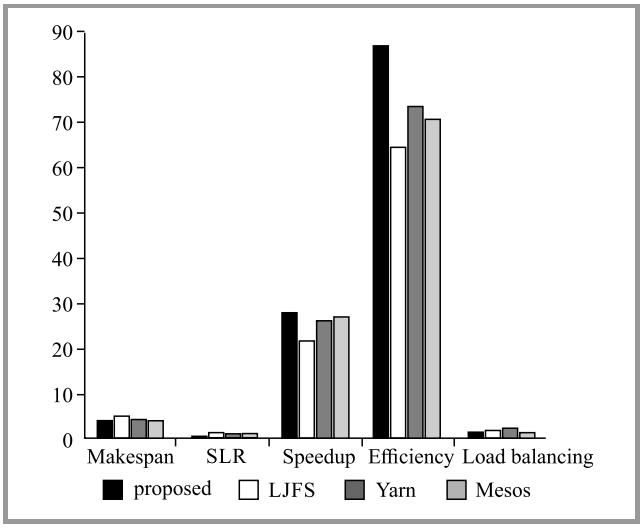


Fig. 5. Comparison of the proposed GPS method with the LJFS scheduling.

The comparison of the various performance metrics shows that the proposed method provides better performance when

compared with the LJFS method, as presented in Figs. 3 and 4.

The overall comparison of the proposed method with the LJFS, Yarn, and Mesos scheduling is given in Fig. 5. As a result, the proposed method minimizes the waiting time between tasks and reduces the idle time interval among the processing elements.

5. Conclusion

The work proposed in this paper has led to developing efficient pre-processing (gang forming) and task scheduling of cloud resources. Similar resources are ganged based on the FMDSGR approach to reduce the scheduled time while selecting the processor. The proposed pre-processing algorithm is fast, robust and easier to understand. When data sets are different or separated from each other, improved results are achieved. Thus, the resultant method enables to form gangs more effectively when compared with the existing arithmetic mean method. The resources can be utilized effectively with the resultant gang. The scheduling length of the workflow is prioritized using the GPS algorithm. Finally, the QoS performance metrics have been evaluated in terms of makespan, SLR, speedup, efficiency and load balancing. The proposed method provides improved performance when compared with the existing LJFS, Yarn, and Mesos method.

References

[1] M. Armbrust *et al.*, "A view of cloud computing", *Commun. of the ACM*, vol. 53, no. 4, pp. 50–58, 2010 (doi: 10.1145/1721654.1721672).

[2] M. Masdari, S. ValiKardan, Z. Shahi, and S. I. Azar, "Towards workflow scheduling in cloud computing: A comprehensive analysis", *J. of Network and Comp. Appl.*, vol. 66, pp. 64–82, 2016 (doi: 10.1016/j.jnca.2016.01.018).

[3] J. Ma, W. Li, T. Fu, L. Yan, and G. Hu, "A novel dynamic task scheduling algorithm based on improved genetic algorithm in cloud computing", in *Wireless Communications, Networking and Applications*, Q.-A. Zeng, Ed. Springer, 2016, pp. 829–835.

[4] Q. Zhang, L. Cheng, and R. Boutaba, "Cloud computing: state-of-the-art and research challenges", *J. of Internet Serv. and Appl.*", vol. 1, no. 1, pp. 7–18, 2010.

[5] L. F. Bittencourt, E. R. M. Madeira, and N. L. S. Da Fonseca, "Scheduling in hybrid clouds", *IEEE Commun. Mag.*, vol. 50, no. 9, pp. 42–47, 2012.

[6] B. Saovapakhiran, M. Devetsikiotis, G. Michailidis, and Y. Viniotis, "Average delay SLAs in cloud computing", in *Proc. IEEE Int. Conf. Commun. ICC 2012*, Ottawa, ON, Canada, 2012, pp. 1302–1308 (doi: 10.1109/ICC.2012.6364548).

[7] S. Saha, S. Pal, and P. K. Pattnaik, "A novel scheduling algorithm for cloud computing environment", in *Computational Intelligence in Data Mining – Volume 1*, H. S. Behera and D. P. Mohapatra, Eds. Springer, 2015, pp. 387–398.

[8] K. Chandran, V. Shanmugasudaram, and K. Subramani, "Designing a fuzzy-logic based trust and reputation model for secure resource allocation in cloud computing", *Int. Arab J. of Inform. Technol.* (*IAJIT*), vol. 13, no. 1, pp. 30–37, 2016.

[9] T. Luis, C. A. Caminero, B. Caminero, and C. Carrion, "A strategy to improve resource utilization in Grids based on network-aware meta-scheduling in advance", in *Proc. 12th IEEE/ACM Int. Conf. on Grid Comput. GRID 2011*, Lyon, France, 2011, pp. 50–57, 2011.

[10] X. Li *et al.*, "Cloud tasks scheduling meeting with QoS", in *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, Y. Qin, L. Jia, J. Feng, M. An, and L. Diao, Eds. Springer, 2016, pp. 289–297.

[11] N. Moganarangan, R. G. Babukarthik, S. Bhuvaneswari, M. S. Saleem Basha, and P. Dhavachelvan, "A novel algorithm for reducing energy-consumption in cloud computing environment: Web service computing approach", *J. of King Saud Univ. – Comp. and Inform. Sci.*, vol. 28, no. 1, pp. 55–67, 2016.

[12] J. Dümmler, R. Kunis, and G. Rünger, "SEParAT: scheduling support environment for parallel application task graphs", *Cluster Comput.*, vol. 15, no. 3, pp. 223–238, 2012.

[13] J. J. Durillo, H. M. Fard, and R. Prodan, "MOHEFT: A multi-objective list-based method for workflow scheduling", in *Proc. IEEE 4th Int. Conf. Cloud Comput. Technol. and Sci. CloudCom 2012*, Taipei, Taiwan, China, 2012, pp. 185–192 (doi: 10.1109/CloudCom.2012.6427573).

[14] K. R. Remesh Babu and P. Samuel, "Enhanced bee colony algorithm for efficient load balancing and scheduling in cloud", in *Innovations in Bio-Inspired Computing and Applications*, V. Snášel *et al.*, Eds. Springer, 2016, pp. 67–78.

[15] G. Peng, H. Wang, J. Dong, and H. Zhang, "Knowledge-based resource allocation for collaborative simulation development in a multi-tenant cloud computing environment", *IEEE Trans. on Services Comput.*, 2016 (doi: 10.1109/TSC.2016.2518161).

[16] F. Koch, D. M. Assunção, C. Cardonha, and A. S. M. Netto, "Optimising resource costs of cloud computing for education", *Future Gener. Comp. Systems*, vol. 55, pp. 473–479, 2016 (doi: 10.1016/j.future.2015.03.013).

[17] D. Saxena, R. K. Chauhan, and R. Kait, "Dynamic fair priority optimization task scheduling algorithm in cloud computing: concepts and implementations", *Int. J. of Comp. Netw. and Inform. Secur.*, vol. 8, no. 2, pp. 41–48, 2016.

[18] J. Choi, T. Adufu, Y. Kim, S. Kim, and S. Hwang, "A job dispatch optimization method on cluster and cloud for large-scale high-throughput computing service", in *Proc. Int. Conf. on Cloud and Autonom. Comput. ICCAC 2015*, Cambridge, MA, USA, pp. 283–290.

[19] I. A. Moschakis and H. D. Karatza, "Evaluation of gang scheduling performance and cost in a cloud computing system", *The J. of Supercomput.*, vol. 59, no. 2, pp. 975–992, 2012.

[20] I. A. Moschakis and H. D. Karatza, "Performance and cost evaluation of Gang Scheduling in a Cloud Computing system with job migrations and starvation handling", in *Proc. 16th IEEE Symp. on Comp. and Commun. ISCC 2011*, Kerkyra, Corfu, Greece, 2011, pp. 418–423. IEEE, 2011.

[21] Y. Cao, Z. Wu, T. Liu, Z. Gao, and J. Yang, "Multivariate process capability evaluation of cloud manufacturing resource based on intuitionistic fuzzy set", *The Int. J. of Adv. Manufactur. Technol.*, vol. 84, no. 1–4, pp. 227–37, 2016.

[22] S. Zhang, Z. Qian, Z. Luo, J. Wu, and S. Lu, "Burstiness-aware resource reservation for server consolidation in computing clouds", *IEEE Trans. on Parallel and Distrib. Syst.*, vol. 27, no.4, pp. 964–77, 2016.

[23] J. Shi, J. Luo, F. Dong, J. Zhang, and J. Zhang, "Elastic resource provisioning for scientific workflow scheduling in cloud under budget and deadline constraints", *Cluster Comput.*, vol. 19, no. 1, pp. 167–182, 2016 (doi: 10.1007/s10586-015-0530-0).

[24] A. T. Saraswathi, Y. R. Kalaashri, and S. Padmavathi, "Dynamic resource allocation scheme in cloud computing", *Procedia Comp. Science*, vol. 47, pp. 30–36, 2015 (doi: 10.1016/j.procs.2015.03.180).

[25] S. Singh and I. Chana, "A survey on resource scheduling in cloud computing: issues and challenges", *J. of Grid Comput.*, vol. 14, no. 2, pp. 217–264, 2016.

[26] A. Sfrent and F. Pop, "Asymptotic scheduling for many task computing in big data platforms", *Inform. Sciences*, vol. 319, pp. 71–91, 2015 (doi: 10.1016/j.ins.2015.03.053).

[27] M. A. Vasile, F. Pop, M. C. Nita, and V. Cristea, "MLBox: Machine learning box for asymptotic scheduling", *Inform. Sciences*, 2017 (doi: 10.1016/j.ins.2017.01.005).

[28] M. A. Vasile, F. Pop, R. I. Tutueanu, V. Cristea, and J. Kołodziej, "Resource-aware hybrid scheduling algorithm in heterogeneous distributed computing", *Future Gener. Comp. Systems*, vol. 51, no. C, pp. 61–71, 2015.

[29] K. Li, "Scheduling parallel tasks with energy and time constraints on multiple manycore processors in a cloud computing environment", *Future Gener. Comp. Systems*, 2017 (doi: 10.1016//j.future.2017.01.010).

[30] B. P. Rimal and M. Maier, "Workflow scheduling in multi-tenant cloud computing environments", *IEEE Trans. on Parallel and Distrib. Syst.*, vol. 28, no. 1, pp. 290–304, 2017.

[31] R. Calheiros, R. Ranjan, A. Beloglazov, C. DeRose, and R. Buyya, "CloudSim: A toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms", *Software: Pract. and Exper.*, vol. 41, no. 1, pp. 23–50, 2011.

S. Sharon Priya completed her B.Tech. in Information Technology and M.Tech. in Computer Science Engineering from Anna University in 2008. She is presently working as an Assistant Professor at the Department of Computer Science and Engineering at B. S. Abdur Rahman University, Chennai, India. Her areas of interest include computing and image processing. At present, she is pursuing Ph.D. from B.S. Abdur Rahman University. She has published/presented several research papers in international journals and at conferences.
E-mail: sharonpriyaphd@gmail.com
sharonpriya@bsauniv.ac.in
Department of Computer Science and Engineering
B. S. Abdur Rahman University
Chennai, Tamil Nadu, India

K. M. Mehata obtained his Ph.D. from the Indian Institute of Technology, Madras, and has been, since 2009, Professor & Dean at the School of Computer, Information and Mathematical Sciences at B. S. Abdur Rahman University since 2009. He guided more than fifteen doctoral candidates and published about 70 research papers concerned with image processing, computer networks, software engineering, web mining and medical informatics. He is currently guiding 10 doctoral students. He is an expert member of AICTE, National Board of Accreditation, DRDO, MHRD and other universities' academic councils study boards.
E-mail: mehatakmphd@gmail.com
School of Computer
Information and Mathematical Sciences
B. S. Abdur Rahman University
Chennai, Tamil Nadu, India

W. Aisha Banu has a Ph.D. degree and focuses on information retrieval and natural language processing. She has a rich teaching experience of seventeen years and has published research papers in peer-reviewed journal and has presented them at conferences. She is a life member of ISTE and a member of ACM.
E-mail: aisha@bsauniv.ac.in
Department of Computer Science and Engineering
B. S. Abdur Rahman University
Chennai, Tamil Nadu, India

Paper

Observation of WiMAX Radio Parameters to Enhance Spectrum Utilization in Mixed Environment

Karol Kowalik[1], Adrian Kliks[2], Bartosz Musznicki[1], Michał Kołodziejski[1], and Paweł Kryszkiewicz[2]

[1] *INEA, Poznań, Poland*

[2] *Chair of Wireless Communications, Poznan University of Technology, Poznań, Poland*

https://doi.org/10.26636/jtit.2018.123917

Abstract—It is believed that 5G networks will provide 1000 times more capacity than current solutions. One of the keys to achieve that goal is not only the utilization of additional radio bands, but also and foremost, the dynamic and efficient spectrum sharing. To successfully implement it such feature statistical observation and analysis of currently operational legacy systems are required. Comprehensive data on the signal parameters will allow then to determine and tune the approach to simultaneous bandwidth usage by existing and new systems. Therefore, to define and introduce the problem this paper presents a conceptual analysis of IEEE 802.16e based WiMAX network operating in the 3.6–3.8 GHz band on the eve of spectrum sharing introduction.

Keywords—5G, bandwidth sharing, spectrum utilization, statistical characteristics, WiMAX.

1. Introduction

The 5G wireless networks have to face different rigorous requirements and expectations, broadly defined as 5G Key Performance Indicators (KPIs). Based on [1] and [2], the following KPIs can be identified:

- 1000-fold increase in mobile data volume in certain geographical areas,
- 10 to 100 times higher typical user data rate,
- up to 10 times reduced energy consumption,
- less than 1 ms end-to-end latency,
- 10 to 100 times increased number of simultaneously connected devices,
- it is assumed that the ubiquitous access will include also low-density areas.

Such a wide set of requirements entails that the delivery of various services to the end-users will be possible by the means of dynamic spectrum aggregation under the umbrella of 5G networks [3]–[6]. Therefore, the on-going improvement of existing radio access methods and new ideas development is required. Prospective enablers include new waveforms [7], massive Multiple Input Multiple Output (MIMO) [8], [9], network visualization, and Software Defined Radio (SDR) [10]–[13]. Some approaches aim to utilize new frequency bands exceeding 6 GHz, such as centimeter and millimeter waves [14], [15], as they enable allocation of a wide spectrum for short-range technologies. Alternatively, flexible utilization of sub-6 GHz frequencies is envisaged as a key enabler for guaranteeing long-distance service delivery and more efficient spectrum utilization.

Given that the spectrum is allocated to different operators, classic coexistence solutions within the licensed and unlicensed portions of the spectrum may not provide the isolation required. Moreover, typical approaches based on orthogonal spectrum sharing between operators within licensed bands may prove inefficient in future scenarios. Hence, a new vision for spectrum utilization is required, and it is highly expected that in the context of future wireless communications systems two traditional models of spectrum management and licensing schemes, i.e. exclusive use and license-exempt, will be complemented by more flexible versions.

One may notice that numerous solutions to advanced resource sharing have been foreseen so far. The examples include infrastructure sharing, as in the Multi-Operator Core Network (MOCN) approach, Licensed Shared Access (LSA), with European standards for the 2.3–2.4 GHz band established by the European Telecommunications Standards Institute (ETSI) [16], or 3-tier sharing model promoted in the USA for the 3.55 –3.7 GHz band and known as Citizen Broadband Radio Service with Spectrum Access System (CBRS/SAS) [17]. However, implementation of these standards is still in the trial phase and numerous field-tests and experiments have to be conducted to verify the applicability of the proposed solutions.

At the same time, network operators aim to maximize their revenues from existing infrastructure. One particular example is the utilization of the 3.5 GHz band for new wireless communications systems, as this is the band already allocated to systems such the IEEE 802.16e-2005 [18], based on WiMAX networks. As Long Term Evolution (LTE) and LTE-Advanced (LTE-A) networks develop, the number of customer connected to WiMAX networks declines on

a continuous basis. So, even though the family of WiMAX standards is not part of 5G and is considered to be of the legacy variety and nearly extinct, these networks still operate and occupy radio spectrum resources.

However, radio resources associated with the WiMAX technology may be also shared with other technologies. In general, future radio systems may need to share the radio spectrum with legacy communications systems. It is important to analyze the real characteristics describing the existing systems, and to draw conclusions regarding prospective utilization of WiMAX frequency bands for 5G purposes. Therefore, we have performed an initial field analysis concerning key parameters of the WiMAX system which have to be considered while deploying new radio systems. Moreover, considerations have been complemented with relevant observations related to 2.4 and 5 GHz IEEE 802.11 access networks [19]. In that context, the observations and conclusions presented further in this paper may be useful for operators who need to keep their WiMAX networks running in the future, but intend to utilize, in the meantime, the spectrum allocated to them in a more efficient manner, by sharing it with other radio systems.

The remainder of the paper is organized as follows: Section 2 briefly presents deployment of WiMAX within an INEA network and discusses its key system parameters, as well as the spectrum sharing scheme considered. Next two sections present the results of long-term (two years) and short-term (two weeks) WiMAX signal observations. Then, Section 5 introduces general similarities with parameters of 2.4 and 5 GHz IEEE 802.11 networks. Section 6 concludes the paper.

2. Spectrum Sharing Scenario

2.1. INEA's WiMAX Deployment

INEA is the largest regional fixed-access telecommunications operator in the Greater Poland region and provides various multimedia and connectivity services to over 250,000 of private and business users, relying on different backbone and access technologies, i.e. Gigabit Passive Optical Network (GPON), point-to-point carrier Ethernet optical fibers, IEEE 802.16e WiMAX, IEEE 802.11 Wi-Fi, as well as Hybrid Fiber-Coaxial (HFC), twisted pair-based xDSL and IEEE 802.3 Ethernet [20].

Since 2010, INEA has been rolling-out and maintaining WiMAX-based services aimed to meet the needs of home users scattered across 30,000 km^2 within the region [21]. Selection of the IEEE 802.16e-2005 standard and the Time Division Duplexing (TDD) mode made it possible to adjust the downlink/uplink ratio to suit the needs of end-users' data transmissions. Following comprehensive testing of infrastructure and equipment supplied by various vendors, INEA has chosen the Motorola (currently Cambium Networks) PMP320 solution due to its compact form and energy-efficient components. The choice was also influenced by its uncomplicated management and installation, which ensured low cost of ownership. So far, this deployment has provided Internet access and telephony services to almost 6000 households.

2.2. Key WiMAX Signal Parameters

As far as WiMAX signal parameters are concerned, particular emphasis should be placed on those which affect system capacity the most. Knowledge of their statistical characteristics may then aid and allow coexistence of two or more systems in the same spectrum, enabling the radio resources available to be shared.

First and foremost, the capacity of a WiMAX network is not fixed [22]. Each Customer-Premises Equipment (CPE) operates with spectral efficiency that changes in time and is influenced by three parameters: modulation, Forward Error Correction (FEC) coding and Multiple Input Multiple Output (MIMO) mode. Therefore, the network capacity of a given access point (AP) is a function of the number of CPEs and their spectral efficiency. For example, if one CPE operates with the modulation of 64-QAM, then its spectral efficiency is 6 bits/s/Hz. If it is using 5/6 FEC, then the spectral efficiency is reduced to 5 bits/s/Hz, whereas for MIMO-B its spectral efficiency is doubled and reaches about 10 bits/s/Hz. The three parameters: modulation, FEC coding and MIMO mode are mostly affected by Received Signal Strength Indication (RSSI) and Carrier to Interference and Noise Ratio (CINR) values. Therefore, the following analyses are focused on RSSI and CINR statistical changes.

2.3. Spectrum Sharing of 3.5 GHz Radio Resources

As it has been pointed out, it is usually economically justified for the network operator to maximize their revenues from the existing infrastructure and from the spectrum licenses held. If it was possible to either offer new services by relying on the existing infrastructure or utilize the available spectrum resources in a more efficient way, such an approach would be highly beneficial for the network operator.

In INEA's case, the second option is being considered, as the WiMAX network has to be kept operational (as the highest priority, incumbent network). Therefore, the following question arises: can other parallel services be deployed, in particular microwave lines or other point-to-point transmissions, that will seamlessly operate in the 3.5 GHz band?

To make such a solution possible, stable methods for the protection of incumbent (WiMAX) network from harmful interference have to be implemented. However, this will be only possible if the key operating parameters that describe the behavior of the WiMAX network are precisely identified. For example, in order to calculate the impact of the induced interference originating from the new network on

the WiMAX network, one needs to know what the required average values observed in the downlink (DL) and uplink (UL) transmission directions in the WiMAX system are. It is also possible however, to analyze the potential changes of these parameters over time. Thus, keeping in mind the main objective of this research project, i.e. is future co-existence of two wireless systems operating in the same frequency band, the present study focuses on the analysis of the changes in RSSI and CINR values measured, as a function of time. Two approaches have been tested, i.e. 2-year long observations to detect long-term trends, further discussed in the context of short, 2-week long measurement campaigns.

3. Long-term Signal Observations

At INEA, every CPE is queried once every hour with the use of the Network Diagnostics System (NDS) to collect statistical data about the quality of the connection. These queries are typically performed using the Simple Network Management Protocol (SNMP), and the results are stored for at least one year. In the case of the WiMAX network, data for last two years has been obtained, covering the period from April 2015 to April 2017. This data is based on information collected from about 6000 CPEs and includes various network parameters (such as MAC address, IP address, base station ID), and two PHY layer signal parameters:

- RSSI for both downlink and uplink,
- CINR for both downlink and uplink.

First of all, the distribution of RSSI and CINR values should be analyzed. Figure 1 presents the distribution of average RSSI values in downlink and uplink directions.

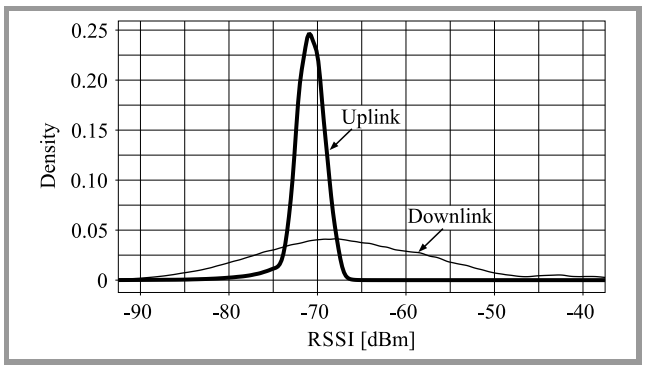


Fig. 1. Distribution of RSSI.

It can be observed that DL RSSI values are spread across a wide range of values, from –50 to –90 dBm, with the mean value at –66 dBm, while UL RSSI values oscillate closely around the mean at –71 dBm. This can be explained by the transmit power control mechanism implemented for uplink, as the WiMAX access point implements a power control mechanism to mitigate path loss, shadowing, etc.

The access point is then configured with a target RSSI value, which is set, in INEA's network, at –71 dBm. Therefore, WiMAX access points instruct all the connected CPEs to adjust their transmit (TX) power to meet the target RSSI value. Hence, the RSSI values are distributed closely to the mean value which, in turn, is equal to the configured target RSSI value. In the downlink direction, each access point transmits at maximum power, and, therefore, all DL RSSI values are proportional to the path loss, shadowing, etc. Thus, the range of values for DL RSSI is much wider than for UL RSSI.

As it is visible in Fig. 2, the CINR density function follows trends that are similar to RSSI, and DL values occupy a much wider range than UL. It can be observed that DL CINR values are ranging from 15 to 40 dB, with the mean at 30 dB, while the most of UL CINR values are located within the range of 25 to 33 dB, with the mean at 28.5 dB.

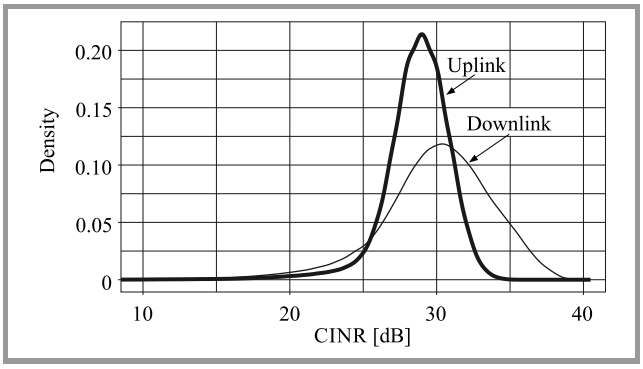


Fig. 2. Distribution of CINR.

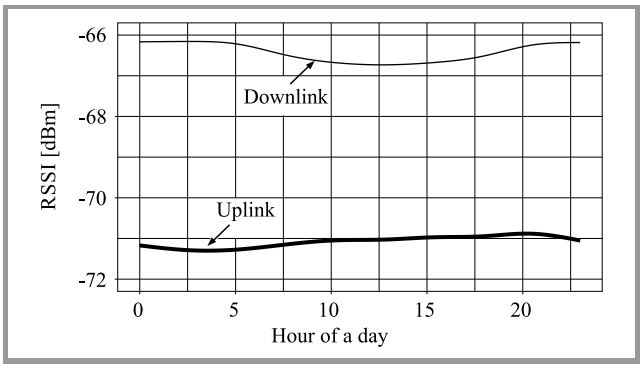


Fig. 3. Average RSSI.

Moreover, signal parameter variations during an average day have been examined. They are calculated as an average for all of CPEs during the two-year observation period. Figures 3 and 4 depict the average RSSI and CINR values as observed round-the-clock. It can be noticed that the values oscillate around the mean values, and that the variation is much smaller for UL parameters than for DL. Moreover, one can observe that DL RSSI values are inversely proportional to the expected seasonal temperature trends.

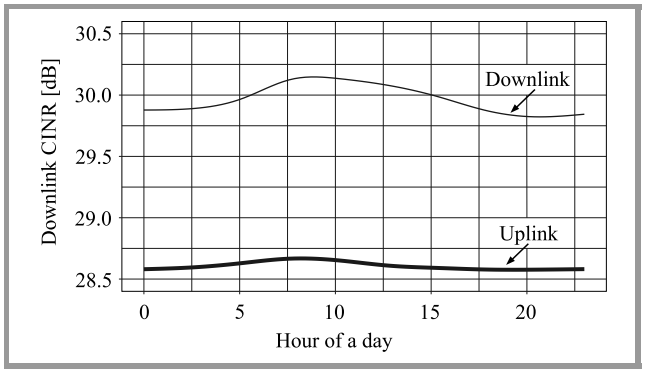


Fig. 4. Average CINR.

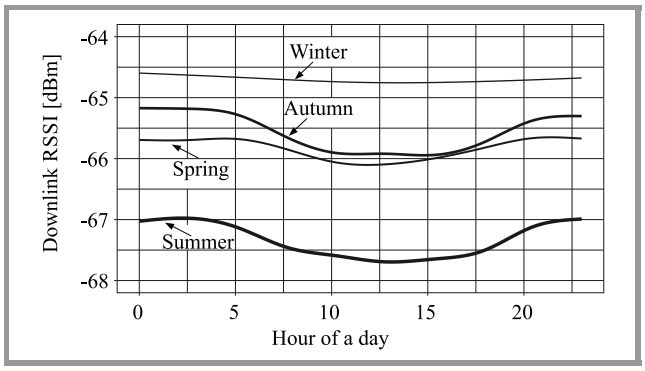


Fig. 5. Average RSSI for each season.

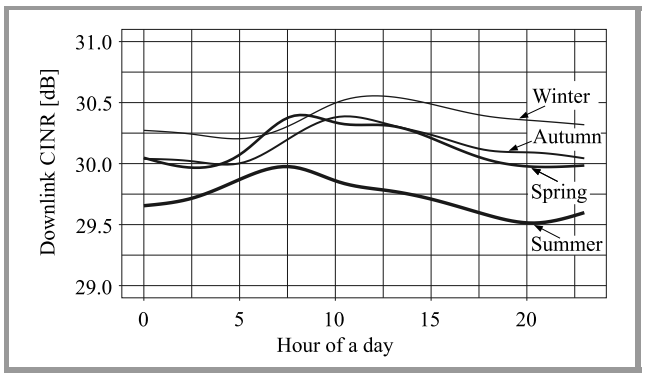


Fig. 6. Average CINR for each season.

To verify if there is any correlation between the RSSI values and temperature, downlink RSSI and CINR for different seasons of the year (calculated as an average for all CPEs during a given season or the two-year observation period) have been plotted in Figs. 5 and 6 respectively. Indeed, Fig. 5 suggests that there is an inverse correlation between RSSI and temperature, because the highest RSSI values can be observed during winter, while in spring and autumn RSSI values are lower and also exhibit a significant decrease during daylight hours. Accordingly, in summer, RSSI values are the lowest. Figure 6 shows some ambivalent trends though. Apparently, in each season a mid-day CINR increase is observed. Still, in summer the CINR values are the lowest, while their highest levels are recorded in winter.

4. Short-term Signal Observations

Since two years of observations presented in the previous section did not exhibit a clear correlation between CINR values and temperature, we have decided to capture more dense data from the WiMAX network, and additionally combine it with humidity and temperature values from the Industrial Institute of Agricultural Engineering (PIMR)

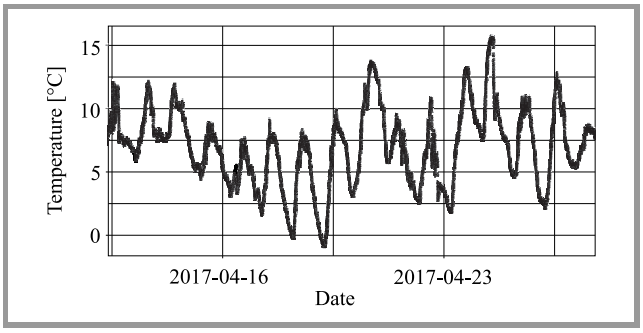


Fig. 7. Temperature for the two weeks period.

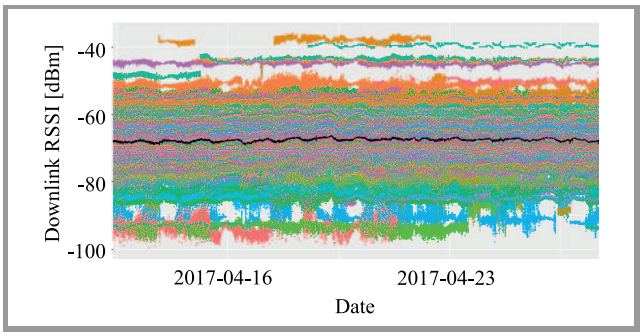


Fig. 8. Download RSSI for the two weeks period. (See color pictures online at www.nit.eu/publications/journal-jtit)

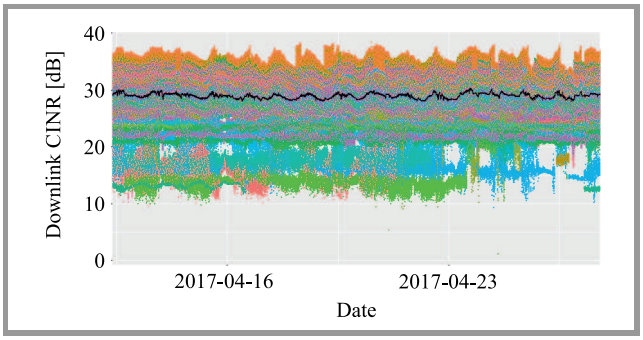


Fig. 9. Download CINR for the two weeks period.

weather station located in Poznań, Poland[1]. In this experiment, 200 WiMAX CPEs located in close proximity to the PIMR weather station have been used. Each CPE has been queried using SNMP every minute during a two-week period between 14 and 28 April 2017. In this way, the following parameters have been collected:

[1] Data are available at http://www.pimr.poznan.pl/bup/gethd2003.php

- RSSI for both downlink and uplink,
- CINR for both downlink and uplink,
- average throughput (expressed in kbit/s) for both downlink and uplink,
- Modulation and Coding Scheme (MCS) index for both DL and UL,
- TX power for UL,
- temperature,
- humidity.

As a reference, Fig. 7 demonstrates varying temperature values during the two-week period of interest.
Due to frequent and transient changes of RSSI and CINR, RSSI and CINR mean values for all 200 CPEs have been calculated. These mean trend lines, as well as data points for each minute within the two weeks for all CPEs, are shown in Fig. 8 and 9, respectively. Each of the 200 CPEs was assigned with a different color and the black line in the middle represents the mean value for all CPEs. The mean trajectory expresses a trend which is not specific to any particular CPE, but to the whole network under consideration.

4.1. Correlation between Parameters

To analyze the correlation between CINR, RSSI and other network and environmental parameters, a correlation matrix has been plotted, as shown in Fig. 10. This matrix presents the Pearson correlation between all pairs of parameters.

An in-depth analysis of the parameter correlation matrix offers different interesting insights, such as:

- mean DL CINR exhibits positive correlation with temperature and negative correlation with humidity and mean DL RSSI,
- mean DL RSSI exhibits negative correlation with temperature and mean DL CINR,
- mean DL RSSI seams to affect DL CINR and DL MCS index,
- mean DL CINR also seems to affect DL MCS index,
- DL throughput is correlated with UL throughput (since Transmission Control Protocol (TCP) based data transfer required communication in both directions),

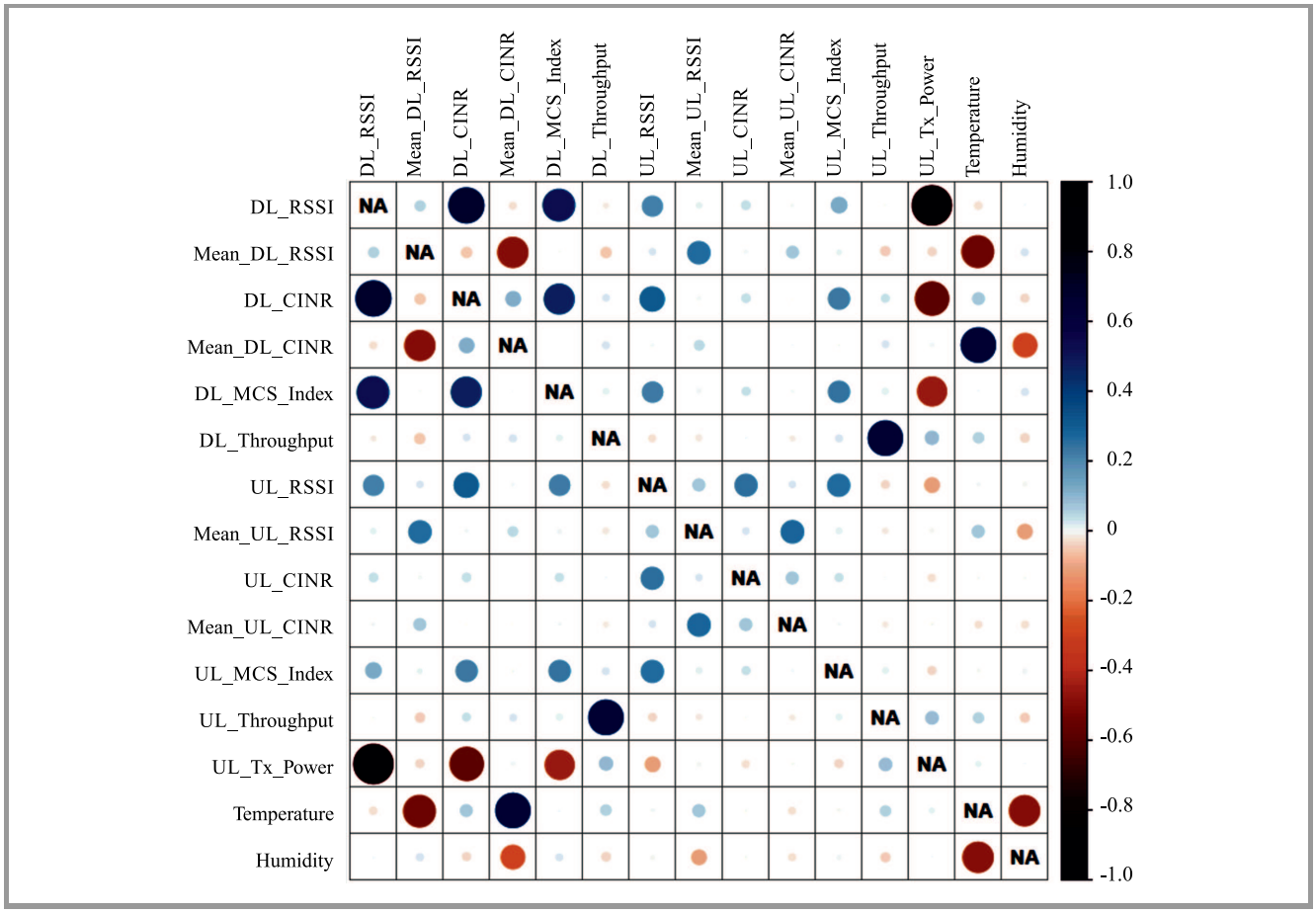


Fig. 10. Parameters correlation matrix.

- most of uplink parameters do not show any strong correlation with any other parameter, except of uplink TX power which affects DL RSSI, DL CINR and DL MCS index. We consider this finding as the most intriguing, since we cannot identify any reason for such a behavior. TX power on the UL is controlled by adaptive power control, and it should mitigate path loss and shadowing on the UL without affecting DL parameters.

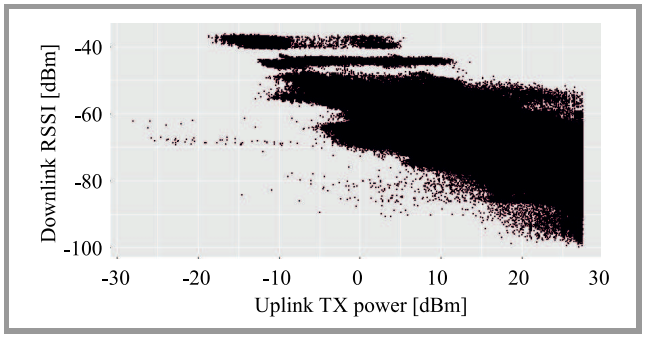


Fig. 11. UL TX power vs. DL RSSI.

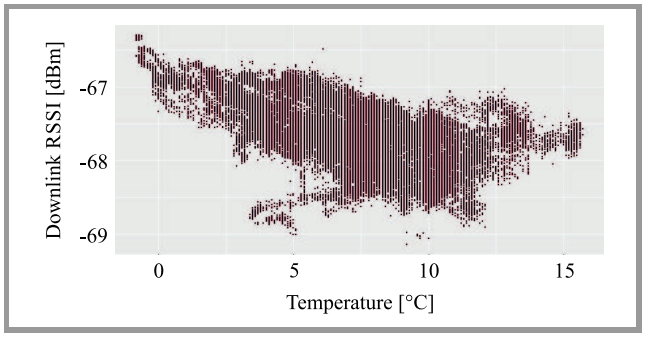


Fig. 12. Temperature vs. DL RSSI.

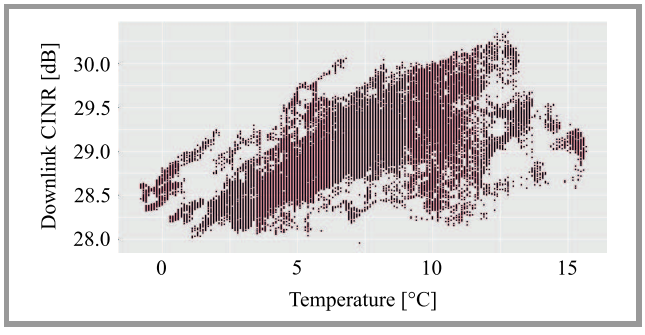


Fig. 13. Temperature vs. DL CINR.

To investigate the correlation between parameters further, we have plotted three pairs exhibiting the highest and almost linear correlation with some limited variations. In Fig. 11 negative correlation between DL RSSI and UL TX power is visible. Figure 12 presents negative DL RSSI correlation with temperature. Conversely, in Fig. 13 positive DL CINR correlation with the temperature can be observed.

5. Neighboring Radio Bands

In order to perceive changes in WiMAX signal characteristics in a broader sense, it is worthwhile to refer to similar parameters of widely utilized neighboring Internet-bearing unlicensed radio bands. As a matter of fact, the present research has been partially inspired by an earlier study conducted at INEA between 10 June and 10 July 2016 [23], where 330 mobile 2.4 GHz and 10 access points were used. Mobile access points were mounted on-board public buses and trams in Poznań and Konin. Their main everyday task is to provide passengers (and other nearby users) with a Wi-Fi service. Each device is a RouterBoard RB751U equipped with a 2.5 dBi omnidirectional antenna and a 4G network modem to provide Internet connection. Each of the selected stationary access points is based on Router Board RB433 connected to a 16 or 19 dBi external sector antennas.

Data has been obtained using SNMP queries issued every 15 minutes to collect information on the noise floor level perceived by each AP, and the RSSI value for each connected client. The sampling interval was chosen to guarantee the frequency required to ensure a complete picture of round-the clock variations, without affecting operation of the access point and the end-user experience.

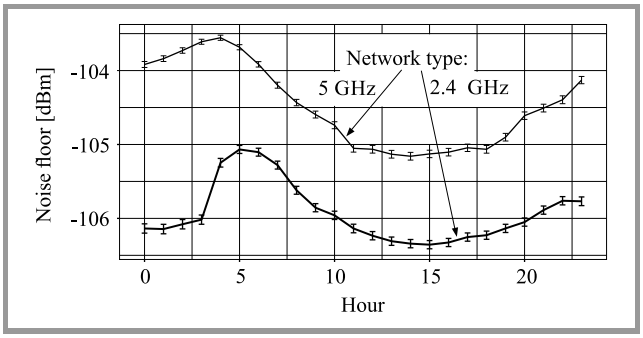


Fig. 14. Average noise floor in 2.4 and 5 GHz networks.

As a result, RSSI for 5 GHz network and noise floor characteristics presented in Fig. 14 have been obtained. It is clearly visible that their shapes resemble the curvature of average RSSI in a 3.5 GHz WiMAX network in summer, as depicted in Fig. 5. Most interestingly, 2.4 and 5 GHz networks operated in seemingly different environmental con-

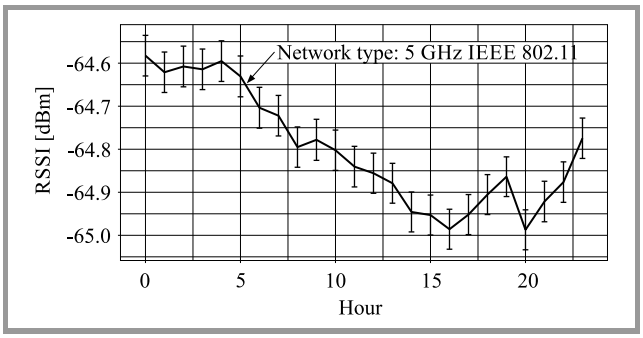


Fig. 15. Average RSSI in a 5 GHz network.

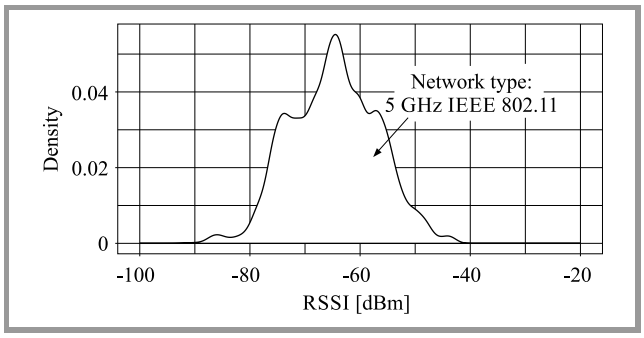


Fig. 16. Distribution of RSSI in a 5 GHz network.

ditions, and yet, the similarities to WiMAX are striking. Access points were not only operating in separated radio bands, but were also equipped with antennas of significantly different gains, and thus, exposed to different angles of noise sources. Mobile APs were traversing busy metropolitan streets, while stationary access points were mounted on aerial masts in more scattered locations. Due to their fixed position, it is also beneficial to refer to the comparable shape of average RSSI in the 5 GHz network presented in Fig. 15. The similarities with WiMAX parameters can be further illustrated by 5 GHz RSSI distribution, as depicted in Fig. 16.

6. Conclusions

In this paper, we have presented statistical characteristics of actual, IEEE 802.11e compliant WiMAX signals, as seen from the perspective of improving spectrum utilization by means of simultaneous use of given frequency bands by two wireless systems. As interference will play a key role in such a scenario, in our discussion we have revealed the changes of RSSI and CINR parameters as functions of time. Based on the highlighted results, the following conclusion can be drawn.

There is a direct correlation between, firstly, the mean RSSI and CINR values observed, and, secondly, the season and time of the day. However, in the context of season changes, these variations are rather negligible (around 1 dB), and in practice, an appropriate interference margin can be included to reflect these changes.

On the other hand, there are high variations of RSSI and CINR changes observed, noticeable in a shorter time scale. When analyzed per CPE, not per network, as the average RSSI and CINR, they are again more or less stable. These may reach even up to a few dB, and such a change has to be taken into consideration while deploying a new and parallel wireless network. Finally, the specificity of the WiMAX system (i.e. power steering in upload direction) entails the need for the system to adjust to the changes in ambient environment in UL, probably without any loss in average RSSI or CINR. The key problem will affect DL, and due for that circumstance, application of careful network planning algorithms may be needed. The analysis presented has demonstrated that some changes in transmission parameters are natural even in the fixed (stationary) WiMAX network that INEA operates. It is also worth pointing out that spectrum sharing requires long-term measurements of the parameters of the entire radio system (WiMAX in the presented case) in order to obtain key characteristics of the signal and to identify bottlenecks. For example, deterioration in CINR values must not necessarily be due to interference from the secondary spectrum user.

It may be due to ambient temperature (or other external factors) as shown in the article, both for 3.5 GHz WiMAX and for 2.4 and 5 GHz networks. In order to explain it with a higher level of certainty, additional investigation of a broader scope and nature seems to be required. Moreover, the presented study suggests that various not obvious relations between observed parameters may emerge, which indicates that further research is required in order to explain the causes of and the practical significance of the findings. As it has been stressed, even though the WiMAX technology is currently considered outdated, networks based on this technology still operate and occupy radio spectrum resources. Moreover, they often have to be maintained due to the legal commitments towards customers, and this will still be the case, at least in the nearest future. Therefore, the findings presented may be of key importance for operators who need to continue maintaining WiMAX networks and use the spectrum more efficiently in the presence of other (new) systems coexisting in the same band. Further studies and results of similar nature may be also valuable to equipment vendors and regulatory offices to aid their efforts aiming to optimize utilization of the spectrum.

Acknowledgement

The work has been funded by the EU H2020 project COHERENT (contract no. 671639).

References

[1] "What Will The 5G-Infrastructure-PPP Deliver?", 5G PPP: The 5G Infrastructure Public Private Partnership [Online]. Available: https://5g-ppp.eu/kpis/

[2] "5G Vision – The 5G Infrastructure Public Private Partnership: the next generation of communication networks and services", 5G PPP: The 5G Infrastructure Public Private Partnership, 2015 [Online]. Available: https://5g-ppp.eu/wp-content/uploads/2015/02/5G-Vision-Brochure-v1.pdf

[3] M. Mustonen, M. Matinmikko, M. Palola, S. Yrjölä, and K. Horneman, "An evolution toward cognitive cellular systems: licensed shared access for network optimization", *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 68–74, 2015.

[4] A. Kliks, O. Holland, A. Basaure, and M. Matinmikko, "Spectrum and license flexibility for 5G networks", *IEEE Commun. Mag.*, vol. 53, no. 7, pp. 42–49, 2015.

[5] H. Bogucka, P. Kryszkiewicz, and A. Kliks, "Dynamic spectrum aggregation for future 5G communications", *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 35–43, 2015.

[6] R. H. Tehrani, S. Vahid, D. Triantafyllopoulou, H. Lee, and K. Moessner, "Licensed spectrum sharing schemes for mobile operators: A survey and outlook", *IEEE Commun. Surv. and Tutorials*, vol. 18, no. 4, pp. 2591–2623, 2016.

[7] P. Banelli, S. Buzzi, G. Colavolpe, A. Modenini, F. Rusek, and A. Ugolini, "Modulation formats and waveforms for 5G networks: Who will be the heir of OFDM?: An overview of alternative modulation schemes for improved spectral efficiency", *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 80–93, 2014.

[8] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey", *IEEE Commun. Surv. and Tutorials*, vol. 18, no. 3, pp. 1617–1655, 2016.

[9] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas", *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, 2010.

[10] C. Liang and F. R. Yu, "Wireless Network Virtualization: A Survey, Some Research Issues and Challenges", *IEEE Commun. Surv. and Tutorials*, vol. 17, no. 1, pp. 358–380, 2015.

[11] I. Khan, F. Belqasmi, R. H. Glitho, N. Crespi, M. Morrow, and P. Polakos, "Wireless sensor network virtualization: A survey", *IEEE Commun. Surv. and Tutorials*, vol. 18, no. 1, pp. 553–576, 2015 (doi: 10.1109/COMST.2015.2412971).

[12] A. Blenk, A. Basta, M. Reisslein, and W. Kellerer, "Survey on network virtualization hypervisors for software defined networking", *IEEE Commun. Surv. and Tutorials*, vol. 18, no. 1, pp. 655–685, 2015 (doi: 10.1109/COMST.2015.2489183).

[13] Y. Li and M. Chen, "Software-defined network function virtualization: A survey", *IEEE Access*, vol. 3, pp. 2542–2553, 2015.

[14] T. S. Rappaport, R. W. Heath, R. C. Daniels, and J. N. Murdock, *Millimeter Wave Wireless Communications*. Prentice Hall, 2015.

[15] S. Kutty and D. Sen, "Beamforming for millimeter wave communications: An inclusive survey", *IEEE Commun. Surv. and Tutorials*, vol. 18, no. 2, pp. 949–973, 2016.

[16] ETSI TS 103 154 V1.1.1. – "Reconfigurable Radio Systems (RRS); System requirements for operation of Mobile Broadband Systems in the 2300 MHz – 2400 MHz band under Licensed Shared Access (LSA)" [Online]. Available: http://www.etsi.org/deliver/etsi_ts/103100_103199/103154/01.01.01_60/ts_103154v010101p.pdf

[17] Federal Communications Commission FCC 15-55, Order on reconsideration and second report and order in the matter of amendment of the commission's rules with regard to commercial operations in the 3550-3650 MHz band, May 2016 [Online]. Available: http://transition.fcc.gov/Daily_Releases/Daily_Business/2016/db0502/FCC-16-55A1.pdf

[18] IEEE Std 802.16e-2005 – IEEE Standard for Local and Metropolitan Area Networks – Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems – Amendment for Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands, 2006.

[19] IEEE Std 802.11-2012 – IEEE Standard for Information technology, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, 29 March 2012 [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6178212

[20] P. Walkowiak, R. Szalski, B. Musznicki, D. Dudek, K. Kowalik, and P. Zwierzykowski, "Evaluation of CARMNET System in INEA HOTSPOT Network", in *Proc. of IEICE Inform. and Commun. Technol. Forum IEICE ICTF*, Poznań, Poland, 2014 [Online]. Available: http://i-scover.ieice.org/proceedings/ICTF/2014/pdf/ALIEN-4.pdf

[21] K. Kowalik, D. Dudek, M. Kołodziejski, B. Musznicki, E. Grzybek, and J. Jarzina, "Lessons Learned from WiMAX Deployment at INEA", *J. of Telecommun. and Inform. Technol.*, no. 3, pp. 34–41, 2014. [Online]. Available: http://www.itl.waw.pl/czasopisma/JTIT/2014/3/34.pdf

[22] C. So-In, R. Jain, and A.-K. Tamimi, "Capacity evaluation for IEEE 802.16e mobile WiMAX", *J. of Comp. Syst., Netw., and Commun.*, Special issue on WiMAX, LTE, and WiFi Interworking, vol. 2010, Article ID 279807, 2010 (doi: 10.1155/2010/279807).

[23] B. Musznicki, K. Kowalik, P. Kłodziejski, and E. Grzybek, "Mobile and residential INEA Wi-Fi hotspot network", in *Proc. 13th Int. Symp. on Wireless Commun. Syst. ISWCS 2016*, Poznań, Poland, 2016.

Karol Kowalik received his M.Sc. from Poznan University of Technology in 2000 and Ph.D. from Dublin City University in 2004. He is currently the Technical Development Manager at INEA and is responsible for technical innovation and validation on new ideas. His research interests include networking, switching, routing.

E-mail: karol.kowalik@inea.com.pl
INEA S.A.
Wysogotowo, Wierzbowa st 84
62-081 Przeźmierowo

Adrian Kliks received his M.Sc. and Ph.D. degrees in telecommunication (with a distinction) from the Poznan University of Technology, in 2005 and 2011, respectively. Since 2011, he has been working at the chair of wireless communications, holding the position of an Assistant Professor. His research interests cover a wide spectrum of wireless communications. Kliks is particularly interested in new waveforms for future wireless systems, including (non-)orthogonal and noncontiguous multicarrier schemes, in application of cognitive radio technology, in advanced spectrum management, but also in deployment and resource management in small-cells, and network virtualization. He is the author of approximately 100 scientific publications or presentations for technical conferences and journals.

E-mail: adrian.kliks@put.poznan.pl
Chair of Wireless Communications
Faculty of Electronics and Telecommunications
Poznan University of Technology
M. Skłodowskiej-Curie sq. 5
60-965 Poznań, Poland

Bartosz Musznicki is pursuing Ph.D. at Poznan University of Technology. His main research interests include topology control and routing in wireless networks. He gained network engineering and managerial experience at INEA. He is an author of four book chapters, seven journal articles, and six conference papers.

E-mail: bartosz.musznicki@inea.com.pl
INEA S.A.
Wysogotowo, Wierzbowa st 84
62-081 Przeźmierowo

Michał Kołodziejski received M.Sc. in Electronics and Telecommunications from Szczecin University of Technology in 2008. In the meantime, he spent six months at Technical Faculty CAU Kiel in Germany. After graduation, he worked at Atos IT Services Poland in Bydgoszcz as a Network Operator (WAN, LAN monitoring and management). In 2010 he joined INEA in Poznań. Currently he is a radio communications engineer and senior network administrator. He deals with various technologies, especially Wi-Fi, WiMAX, microwave links and customer distribution cable networks.

E-mail: michal.kolodziejski@inea.com.pl
INEA S.A.
Wysogotowo, Wierzbowa st 84
62-081 Przeźmierowo

Paweł Kryszkiewicz received his M.Sc. degree in Telecommunications from Poznan University of Technology in 2010. Since 2010 he has been employed at the Chair of Wireless Communications in the Faculty of Electronics and Telecommunications of Poznan University of Technology, as a senior researcher. His research interests focus mostly on physical layer for the future wireless communications systems. In particular, he is interested in the following: multicarrier modulation schemes, e.g., NC-OFDM, their application to cognitive radio systems, problems of spectrum shaping for protection of primary users and mitigation of interference caused by nonlinear effects in radio front-end. Since 2010 he has been a member of IEEE. He was involved in international projects (like NEWCOM#, COGEU, ACROPOLIS, COST Action IC-0902, COST-Terra).

E-mail: pawel.kryszkiewicz@put.poznan.pl
Chair of Wireless Communications
Faculty of Electronics and Telecommunications
Poznan University of Technology
M. Skłodowskiej-Curie sq. 5
60-965 Poznań, Poland

Paper

Performance of Hybrid Sensing Method in Environment with Noise Uncertainty

Mateusz Kustra, Krzysztof Kosmowski, and Marek Suchański

Military Communication Institute, Zegrze, Poland

https://doi.org/10.26636/jtit.2018.123117

Abstract—The paper presents a novel hybrid spectrum sensing method used in cognitive radio and presents a hybrid detector (HD) which improves the sensing performance. The proposed HD takes advantage of the energy detection (ED) principle and a method based on Covariance Absolute Value (CAV), as well as on Cyclic Autocorrelation Function (CAF). The paper shows the limitations of using ED, resulting from the uncertainty of spectral density of noise power estimation, known as the SNR wall. The paper describes a system model and presents simulation results for the OFDM signal of a WiMAX-based communications system. The simulation results refer to an ideal environment with well-known parameters, and to an environment with uncertain spectral density of noise power estimation.

Keywords—Covariance Absolute Value, Cyclic Autocorrelation Function, hybrid detector, noise uncertainty, OFDM, SNR wall, WiMAX.

1. Introduction

Cognitive radio systems [1], [2] are an effective solution to the problem of spectrum scarcity, providing dynamic spectrum access to frequencies that are temporarily not used by primary users (PU). Spectrum sensing is one of the basic tasks of cognitive radio which must be carried out to enable communications. It relies on monitoring wide-band spectrum and finding the channels not occupied by PU (licensed) users, which can be used by secondary users (SU).

There is a lot of research dealing with optimization of spectrum sensing. A common approach is to increase efficiency of hybrid architecture detectors, based on a combination of various detection methods [3], [4]. The structure of a hybrid sensing model depends on the spectrum recognition scenario used. A two-phase system which uses energy detections (ED) in the first phase could be an example of the simplest and fastest method of sensing. It enables reliable detection of strong signals, using a relatively small number of samples. In other cases, if the detected energy level does not allow for accurate ED estimation, another, more accurate method can be used.

ED is characterized by low computational complexity and simple implementation [5]. Unfortunately, it is sensitive to the uncertainty of spectral density of noise power estimation [6], [7]. Therefore, the second phase of the hybrid detector (HD) uses a method that does not require this parameter. These methods most often use distinctive features which let us distinguish noise from modulated signals. However, they are usually complex or require many samples to ensure high detection reliability. Examples of methods that can be used in the second HD phase include the following: matched filter, cyclostationary features detector, eigenvalue-based sensing detector, wavelet-based sensing detector or covariance-based detector.

The results of HD research show, inter alia, superiority of the hybrid method [8], [9]. However, these papers refer to an ideal scenario in which the uncertainty of spectral density of noise power estimation is considered. In real systems it is not possible to accurately estimate noise variance, which results in restrictions affecting the use of ED. Any measurements are characterized by finite accuracy and, thus, uncertainty. In the case of ED, this uncertainty in relation to the measurement of the spectral density of noise power is revealed as the so-called SNR wall [10].

When noise is affected by uncertainty, the existing approach turns out to be too idealistic. For this reason, the paper shows an analysis of HD efficiency in an environment with uncertainty associated with spectral density of noise power estimation.

The remaining parts of this paper present two hybrid sensing methods (HD_{CAV} and HD_{CAF}) using ED and CAV or ED and CAF, respectively. A system model for which simulations have been carried out is characterized. The results of the study for the WiMAX system are presented for two cases: the ideal case of an environment with well-known conditions, as considered in the literature so far, and for an environment with uncertainty related to spectral density of noise power estimation.

2. Hybrid Detector

A two-phase hybrid detector is proposed, combining the advantages of ED and CAV or CAF sensing approaches (Fig. 1).

For each channel, first the presence of PU is determined in the ED detection phase. Although this method is sensitive to good noise uncertainty, its undoubted advantage is the

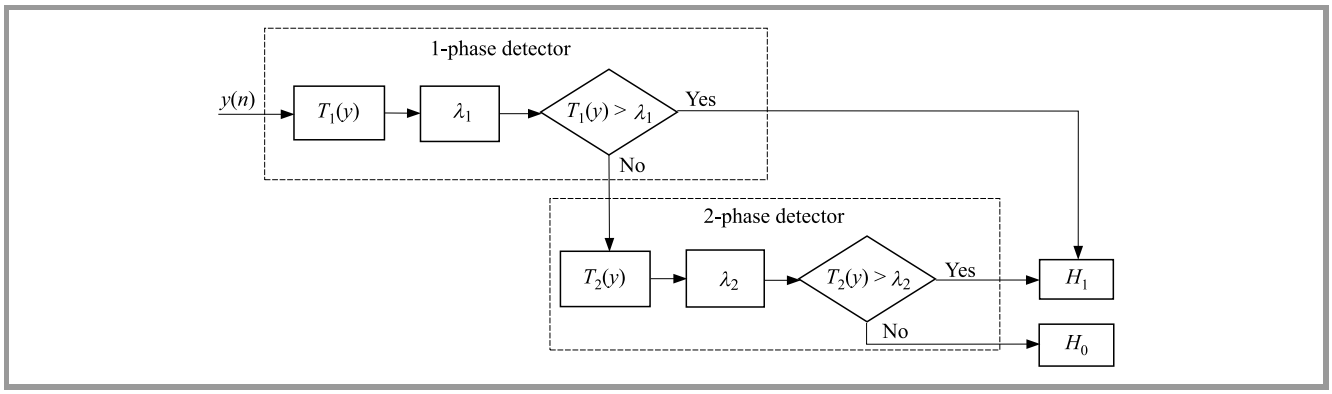


Fig. 1. Hybrid detector block diagram.

high speed of detection and accuracy at high SNR values. The decision about PU signal presence H_1 is taken only when the energy of the received signal ($T_1 = T_{ED}$) is higher than the first phase detection threshold ($\lambda_1 = \lambda_{ED}$) calculated for the assumed probability of a false alarm (P_{fa}).
If the decision cannot be made using ED, CAV or CAF, the second phase of hybrid detection is used as a more accurate approach. The decision about PU signal presence is taken when decision statistic T_2 is greater than the second phase threshold λ_2. Otherwise, a decision about PU signal absence H_2 is made. Depending on the detector used in the second phase (CAV or CAF), here $T_2 = T_{CAV}$ and $\lambda_2 = \lambda_{CAV}$, or $T_2 = T_{CAF}$ and $\lambda_2 = \lambda_{CAF}$.

2.1. ED Method

The decision rule for the energy detector can be expressed by [5], [11]:

$$T_{ED} = \frac{1}{N_S} \sum_{n=0}^{N_S-1} |y(n)|^2 , \tag{1}$$

where: $y(n)$ is the received signal, N_S is the number of signal samples.
The detection threshold for the assumed P_{fa} value is expressed as:

$$\lambda_{ED} = \sigma_\eta^2 \left(Q^{-1}(P_{fa}) \sqrt{2N_S} + N_S \right), \tag{2}$$

where: σ_η^2 is noise variance, $Q(t)$ is the Q function given by:

$$Q(t) = \frac{1}{\sqrt{2\pi}} \int_t^{+\infty} \mathrm{e}^{-\frac{u^2}{2}} \mathrm{d}u . \tag{3}$$

Equation 2 can be used in an ideal environment, for which it is possible to estimate the noise variance with a high level of accuracy. Under real conditions, the uncertainty of measurement needs to be taken into consideration [10], assuming that the actual variance of noise is within the uncertainty interval such as:

$$\sigma^2 = \left\langle \left(\frac{1}{\rho} \right) \sigma_\eta^2 ; \rho \sigma_\eta^2 \right\rangle , \qquad \rho > 1 , \tag{4}$$

where ρ is parameter that quantifies the uncertainty degree.

Considering the uncertainty associated with spectral density of noise power measurements, the detection threshold is:

$$\lambda_{ED} = \rho \sigma_\eta^2 \left(Q^{-1}(P_{fa}) \sqrt{2N_S} + N_S \right). \tag{5}$$

The time (represented by number of samples N_S) required to the channel state corresponds to the probability values assumed and is expressed as [10]:

$$N \approx \frac{2 \left(Q^{-1}(P_{fa}) - Q^{-1}(P_d) \right)^2}{\left(SNR - \left(\rho - \frac{1}{\rho} \right) \right)^2} . \tag{6}$$

Equation 6 shows that the required number of samples reaches infinity when the decreasing SNR reaches a value comparable to the area of approximated spectral density of noise power uncertainty. Figure 2 shows the number of samples needed to obtain the assumed probabilities in the SNR function [10]. Depending on the accuracy of the spectral density of noise power estimation expressed as uncertainty ($x = 10 \log \rho$), the SNR wall level is achieved at lower SNRs, but as the limit approaches, the number of samples necessary to maintain the required credibility increases rapidly.

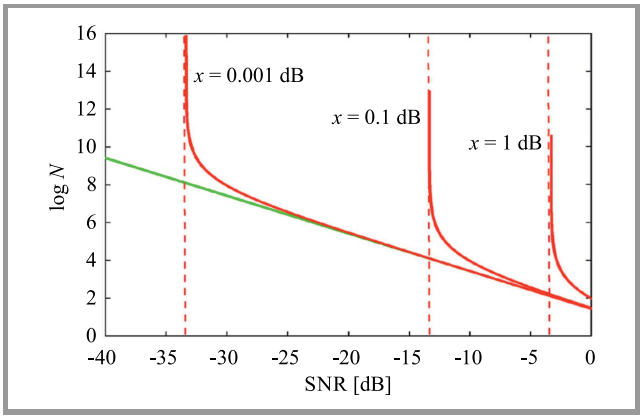


Fig. 2. Number of samples as a function of SNR, depending on the uncertainty of spectral density of noise power estimation.

The detector cannot provide a reliable decision if the signal power level is lower than the uncertainty associated with the spectral density of noise power measurement. SNR wall as

function of uncertainty is expressed by Eq. 7 and shown in Fig. 3:

$$SNR_{Wall} = \frac{\rho^2 - 1}{\rho}. \tag{7}$$

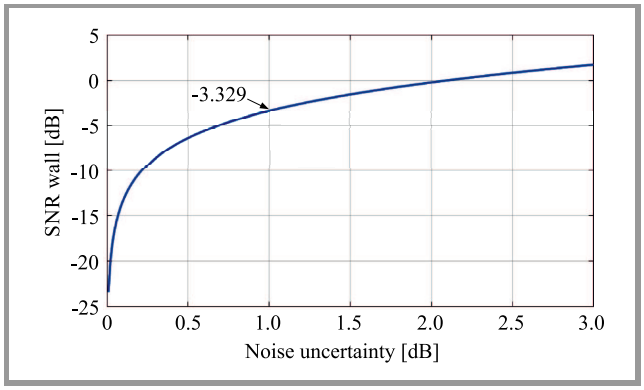


Fig. 3. SNR wall as a function of noise uncertainty.

2.2. CAV Method

CAV is based on differences between noise and signal autocorrelation. The autocorrelation of received signal is [12]:

$$\phi(l) = \frac{1}{N_S} \sum_{n=0}^{N_S-1} y(n) \cdot y(n-l), \quad l = 0, 1, \ldots, L-1, \tag{8}$$

where N_S is number of signal samples, L is the smoothing factor.

Statistical covariance matrices R_x of the entire signal and noise can be estimated using an $\hat{R}_x$ matrix symmetric and Toeplitz formed for L consecutive signal samples:

$$\hat{R}_x(N_S) = \begin{bmatrix} \phi(0) & \phi(1) & \ldots & \phi(L-1) \\ \phi(1) & \phi(1) & \ldots & \phi(L-2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi(L-1) & \phi(L-2) & \ldots & \phi(0) \end{bmatrix}. \tag{9}$$

Based on the symmetric property of the autocorrelation matrix, T_1 and T_2 ratios are expressed as follows:

$$T_1 = \frac{1}{L} \sum_{n=1}^{L} \sum_{m=1}^{L} |r_{nm}|, \tag{10}$$

$$T_2 = \frac{1}{L} \sum_{n=1}^{L} |r_{nn}|, \tag{11}$$

where r_{nm} and r_{nn} are $\hat{R}_x$ matrix elements.

The decision statistic for CAV is:

$$T_{CAV} = \frac{T_1}{T_2}, \tag{12}$$

and detection threshold λ_{CAV} is calculated as:

$$\lambda_{CAV} = \left(1 + (L-1)\sqrt{\frac{2}{N_S \pi}}\right)\left(1 - \frac{P_{fa}}{Q}\sqrt{\frac{2}{N_S}}\right)^{-1}. \tag{13}$$

2.3. CAF Method

According to [13], the complex $x(t)$ process with the average zero value is cyclostationary in a wide sense, if its autocorrelation function (varying in time domain) is periodic with the repetition period T_f and can be represented as a Fourier series:

$$R_{xx}(t, \tau) = \sum_{\alpha} R_{xx}^{\alpha}(\tau) \mathrm{e}^{\mathrm{j}2\pi\alpha t}, \tag{14}$$

where values are added by integral multiplies of the basic frequency $\alpha = \frac{k}{T_f}$, $k = 1, 2, 3 \ldots$. The Fourier series coefficients depending on the time lag have the following form:

$$R_{xx}^{\alpha}(\tau) = \lim_{T \to \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} R_{xx}(\tau) \mathrm{e}^{-\mathrm{j}2\pi\alpha t} \mathrm{d}t \tag{15}$$

The $R_{xx}^{\alpha}(\tau)$ function is called the cyclic autocorrelation function (CAF) [14], and the CAF Fourier transform:

$$S_{xx}^{\alpha}(f) = \int_{-\infty}^{\infty} R_{xx}^{\alpha}(\tau) e^{-\mathrm{j}2\pi f\tau} \mathrm{d}t \tag{16}$$

is called the spectral correlation density function.

One can see that CAF is discrete in terms of frequency and continuous in terms of time lag.

For non-cyclostationary CAFs processes, $R_{xx}^{\alpha}(\tau) = 0$, $\forall \alpha \neq 0$. Each non-zero value of the α parameter, where $R_{xx}^{\alpha}(\tau) = 0$, is called the cyclic frequency.

CAF for the OFDM signal has the following form [15]:

$$R_{xx}^{\alpha} = \frac{A}{T_S} \frac{\sin(\pi N_S \Delta f \tau)}{\sin(\pi \Delta f \tau)} \mathrm{e}^{\mathrm{j}2\pi\left(f_0 + \Delta f \frac{N_S - 1}{2}\right)} \times \\ \times \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{j}2\pi(\alpha_n - f)} G(f) G(\alpha_n - f) \mathrm{d}f, \tag{17}$$

where $G(f)$ is the Fourier transform of a rectangular pulse shape, A is the variance of symbol sequence, $T_s = T_u + T_g$ is the symbol duration, $T_u = \frac{1}{\Delta f}$ is the useful symbol duration, Δf is the subcarrier spacing, and T_g is the guard interval duration. The detection threshold λ_{CAF} is:

$$\lambda_{CAF} = tg \cdot \frac{1}{2} \pi \left(1 - P_{fa\_CAF}\right). \tag{18}$$

3. System Model

In cognitive radio, the sensing of the primary user's signals is directly connected with the cognitive system scenario. In this paper the WiMAX (IEEE 802.16-2004 [16]) was assumed as the licensed system with its parameters specified in Table 1. The following detection parameters were also assumed:

- probability of a detection $P_d = 0.9$,
- probability of a false alarm $P_{fa} = 0.1$,
- uncertainty associated with spectral density of noise power estimation $x = \pm 1$ dB.

For the HD second phase using the CAF, detection of a single CAF peak is used ($\alpha = 0$ and $\tau = T_u$). This case is similar to [17], and the difference lies in other decision statistics.

Table 1
Parameters of the licensed system used

Parameter	Value
Bandwidth	3.5 MHz
OFDM symbol duration	80 μs
OFDM useful symbol duration	64 μs
Cyclic prefix ratio	1/4
FFT size	256

The decision statistics for proposed CAF is:

$$T_{CAF} = \left| \frac{R_x^\alpha}{R_y^\alpha} \right|, \tag{19}$$

where: R_x^α is the empirical CAF of the OFDM signal, R_y^α is the empirical noise CAF. T_{CAF} test is a simple ratio test between R_x^α and R_y^α evaluated for $\alpha = 0$ and $\tau = T_u$. The test compares characteristic points of CAF for OFDM signals and noise.

The question that remains open is how to acquire noise samples for the test. One of the solutions proposed in literature is to take data from a rarely used channel. American channel 37 reserved for radio astronomy is a good example here. Another proposal is to use samples from the tested channel, provided that a previous decision has been made that there is no emission in the PU channel.

4. Simulation Results

The aim of the simulations was to check the efficiency of HD_{CAV} and HD_{CAF} methods in comparison to other available techniques, i.e. ED, CAV, CAF. Three metrics were used to evaluate the sensing efficiency:

- sensitivity of the sensing P_d,
- reliability of the sensing P_{fa},
- sensing time.

HD sensing should significantly increase efficiency. However, insertion of the uncertainty of noise variance into the scenario may significantly worsen the results. For this reason, the proposed hybrid detectors were first tested for the ideal case, i.e. in an environment that did not take into account the uncertainty of spectral density of noise power estimation. Then, the tests were repeated for an environment with such uncertainty.

To determine the dependence of P_d on SNR with the assumed number of samples, the probability of a false alarm was set at 10% ($P_{fa} = 0.1$).

Figure 4 shows a comparison of HD_{CAV} performance with ED and CAV sensing techniques for N OFDM signal symbols versus SNR for the ideal case. For 10 OFDM symbols, HD_{CAV} reaches $P_d = 90\%$ for SNR lower by at least 0.8 dB, and for 50 symbols, it is 2 dB referring to the best of of the two single methods (ED). The hybrid detection scheme considered achieves better results than detectors based on exclusively on ED or CAV.

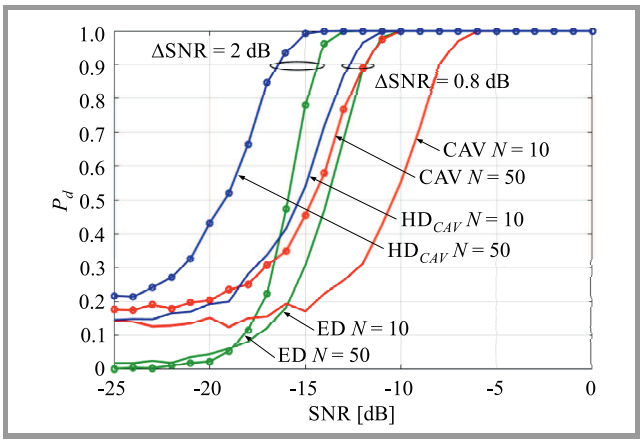


Fig. 4. Probability of detection vs. SNR for HD_{CAV} without the influence of uncertainty of spectral density of noise power estimation.

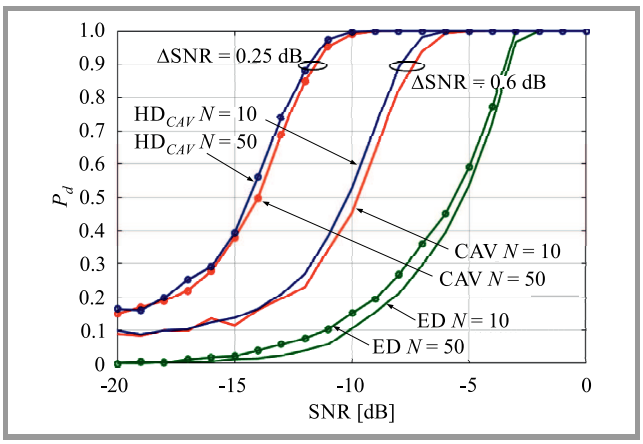


Fig. 5. Probability of detection vs. SNR for HD_{CAV} with the influence of uncertainty of spectral density of noise power estimation.

Figure 5 shows the same comparison as presented in Fig. 4, but with the uncertainty of noise variance. In this situation the results are considerably worse. For 10 OFDM symbols, HD_{CAV} reaches $P_d = 90\%$ for SNR lower by almost 0.6 dB, and for 50 symbols, it is 0.25 dB referring to the best of the two single methods (CAV). The uncertainty of noise variance leads to significant deterioration of the HD detection performance. One can see that the biggest gain from the use of HD is achieved for short signals. So, the longer the signal, the more dependent HD performance becomes on the method used in the second phase of detection.

Figure 6 shows the comparison of HD_{CAF} performance with the ED and CAF sensing techniques for N OFDM signal symbols versus SNR for the ideal case. For 10 OFDM sym-

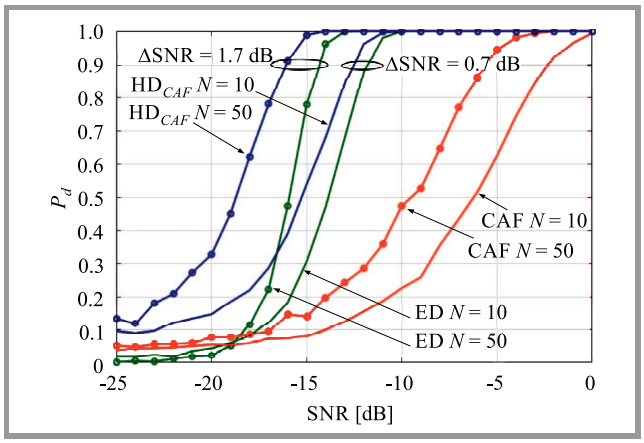


Fig. 6. Probability of detection vs. SNR for HD$_{CAF}$ (without the influence of uncertainty of spectral density of noise power estimation).

bols, HD$_{CAF}$ reaches $P_d = 90\%$ for SNR lower by 0.7 dB, and for 50 symbols, it is 1.7 dB referring to the best of the two single methods (ED). Also, in this case, HD shows better detection parameters than other methods. For HD, the assumed $P_d = 0.9$ is reached at lower SNR values than for the other methods.

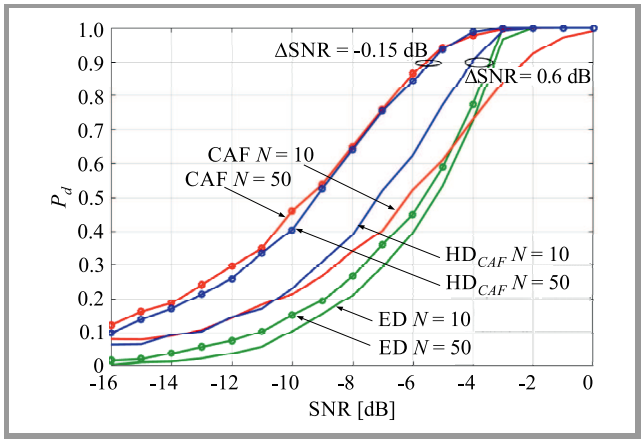


Fig. 7. Probability of detection vs. SNR for HD$_{CAF}$ with the influence of uncertainty of spectral density of noise power estimation.

Similarly, Fig. 7 shows the performance of the same detectors as in Fig. 6, with the uncertainty of noise variance. In this scenario, the results are much worse. For 10 OFDM symbols, HD$_{CAF}$ reaches $P_d = 90\%$ for SNR lower by almost 0.6 dB referring to the best of the two single methods (ED). However, for 50 symbols, HD$_{CAF}$ is worse than the best of the two single methods (CAF) by 0.15 dB. It can be seen that for the environment with the uncertainty of spectral density of noise power estimation, the gain from the use of HD$_{CAF}$ is achieved just for a short signal observation time.

In order to compare the presented detectors, the receiver operating characteristic (ROC) curves were determined (for HD$_{CAV}$ – Fig. 8, Fig. 9, and for HD$_{CAF}$ – Fig. 10, Fig. 11).

It can be noticed that for the ideal case (Fig. 8), HD$_{CAV}$ is significantly better than the other single detectors. HD$_{CAV}$ reaches $P_d = 90\%$ for P_{fa} lower than 6.5%, compared to the better of the single methods (ED). According to the theoretical assumptions, introduction of HD increases reliability sensing due to minimizing P_{fa}.

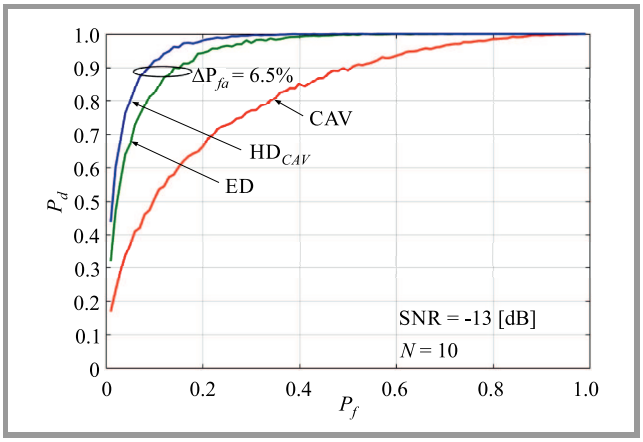


Fig. 8. ROC curves for HD$_{CAV}$ (without the influence of uncertainty of spectral density of noise power estimation).

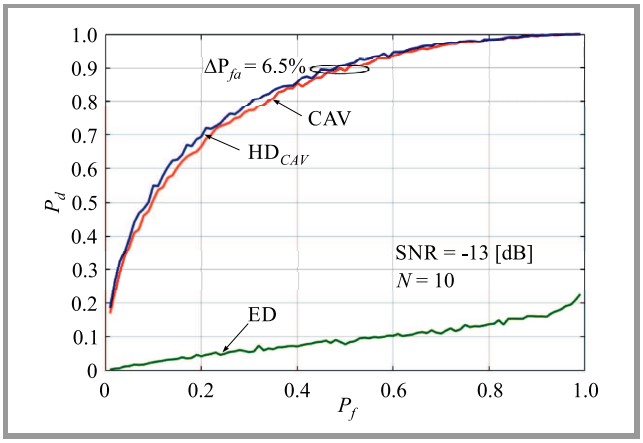


Fig. 9. ROC curves for HD$_{CAV}$ (with the influence of uncertainty of spectral density of noise power estimation).

Figure 9 shows the ROC curves taking into account the uncertainty of the noise variance effect. In this case, the results are much worse. HD$_{CAV}$ reaches $P_d = 90\%$ for P_{fa} lower by at least 2.5%, compared to the better of the single methods (CAV).

In the ideal case (Fig. 8), the detection threshold for the first phase based on ED ($\lambda_1 = \lambda_{ED}$) was calculated from Eq. 2, which did not account for the uncertainty of noise variance. That is why the results show HD superiority compare to other methods. However, by analyzing the ROC curves after taking into account the uncertainty (Fig. 9), one may notice that ED and SNR wall have a great impact on the reliability of HD.

Considering hybrid detection based on CAF in the second phase, one can see that for the ideal case (Fig. 10), HD$_{CAF}$ is also better than other single detectors. HD$_{CAF}$ reaches $P_d = 90\%$ for P_{fa} lower than 6%, compared to the better

of the single methods (ED). This time, the introduction of HD (by minimizing P_{fa}) also increases the reliability of sensing.

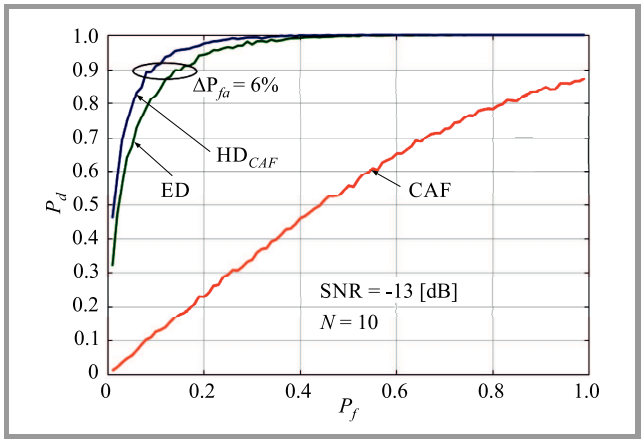


Fig. 10. ROC curves for HD_{CAF} (without the influence of uncertainty of spectral density of noise power estimation).

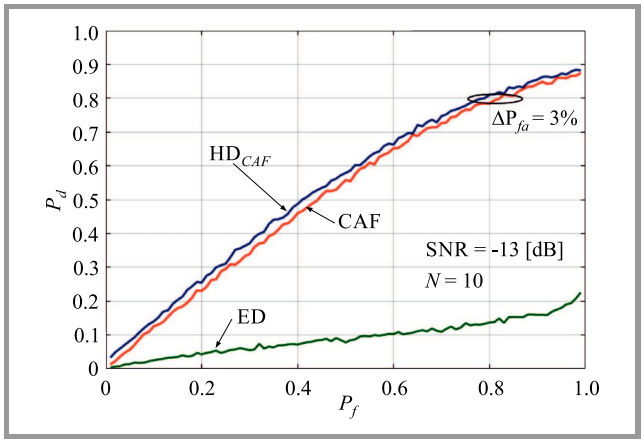


Fig. 11. ROC curves for HD_{CAF} (with the influence of uncertainty of spectral density of noise power estimation).

Figure 11 shows the ROC curves taking into account the uncertainty of the noise variance effect. In this case the results are much worse. For the conditions under consideration, HD_{CAF} does not reach $P_d = 90\%$. But generally, HD_{CAF} allows for decreasing P_{fa} by 3% comparing to the better of the single methods (CAF).

This time, the weak performance of ED in an environment with the uncertainty of spectral density of noise power estimation results in the fact that HD_{CAF} is useless and the SNR wall has too big an impact on the reliability of HD.

To compare both HD solutions in terms of detection time, the results achieved were presented and compared with the number of samples.

The simulation results show P_d vs. sensing time, expressed in the number of samples for HD_{CAV} and HD_{CAF} in Figs. 12 and 13, respectively. The results have been presented just for an environment with the uncertainty of spectral density of noise power estimation, in order to show how considerable a reduction of sensing time is possible with the HD method.

Figure 12 shows that for –5 dB SNR, $P_d = 90\%$ can be achieved for a number of samples lower by at least 400, which represents a reduction of sensing time by 26%. For –10 dB SNR, $P_d = 90\%$ can be achieved 1600 signal samples faster (17% less time).

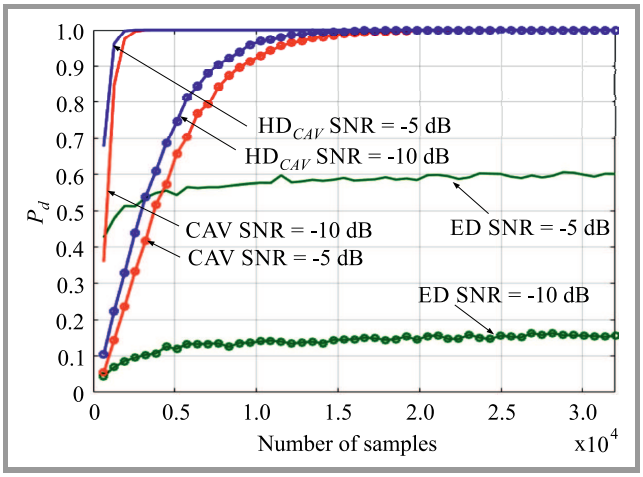


Fig. 12. Probability of detection vs. sample numbers function for HD_{CAV} (with the influence of uncertainty of spectral density of noise power estimation).

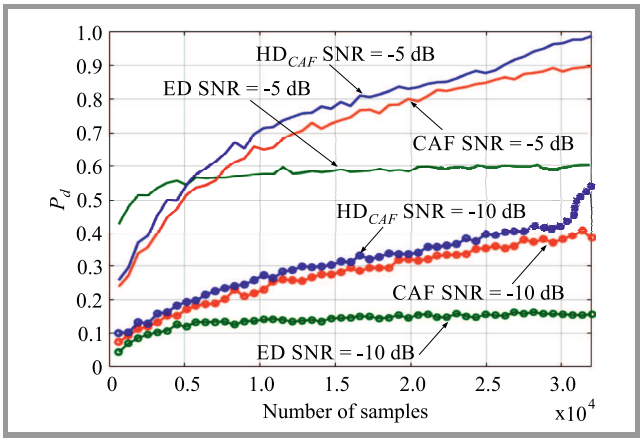


Fig. 13. Probability of detection vs. sample numbers function for HD_{CAF} (with the influence of uncertainty of spectral density of noise power estimation).

In Fig. 13, simulation results for HD_{CAF} show that for –5 dB SNR, $P_d = 90\%$ can be achieved for a number of samples lower by at least 4000 (13% reduction of sensing time). For –10 dB SNR, HD_{CAF} does not reach the required level of P_d for the taken number of samples considered in the simulations. It can be concluded that HD_{CAF} is slower than HD_{CAV} – it requires more samples.

5. Conclusions

The HD allows for the increase of sensing efficiency in cognitive radio, especially in comparison to individual methods, i.e. ED, CAV or CAF. By taking into account two

extreme cases: the ideal and the worst ones (with 1 dB uncertainty of spectral density of noise power estimation), it is possible to conclude that the more accurate the estimation of SNR, the higher the HD gain. And even in the worst scenario, HD makes it possible to detect the signal quicker (even by 26%), at the same time lowering P_{fa} and increasing P_d.

References

[1] Federal Communications Commission: Notice of proposed rulemaking and order, Facilitating opportunities for flexible, efficient and reliable spectrum use employing cognitive radio technologies. FCC 03-322, 2003.

[2] S. Kapoor and G. Singh, "Non-cooperative spectrum sensing: A hybrid model approach", in *2011 Int. Conf. on Dev. and Commun. ICDeCom 2011*, Mesra, India, 2011.

[3] H. Bogucka, *Technologie Radia Kognitywnego*. Warsaw: PWN, 2013 (in Polish).

[4] S. Haykin, D. J. Thomson, and J. Reed, "Spectrum sensing for cognitive radio", *Proc. IEEE*, vol. 97, no. 5, pp. 849–877, 2009.

[5] H. Urkowitz, "Energy detection of unknown deterministic signals", *Proc. IEEE*, vol. 55, no. 4, pp. 523–531, 1967 (doi: 10.1109/PROC.1967.5573).

[6] D. Cabric, A. Tkachenko, and R. Brodersen, "Experimental study of spectrum sensing based on energy detection and network cooperation", in *Proc. of the 1st ACM Int. Worksh. on Technol. and Policy for Access. Spectrum TAPAS'06*, Boston, USA 2006, Article no. 12 (doi: 10.1145/1234388.1234400).

[7] A. Sahai and D. Cabric, "Spectrum sensing: Fundamental limits and practical challenges", in *Proc. of the IEEE Int. Symp. on New Frontiers in Dynam. Spec. Access Netw. DySPAN 2005*, Baltimore, MD, USA, 2005.

[8] S. Geethu and G. Lakshminarayanan, "A novel selection based hybrid spectrum sensing technique for cognitive radios", in *2013 IEEE Int. Conf. on Emerging Trends in Comput., Commun. and Nanotechnol. ICE-CCN 2013*, Tirunelveli, India, 2013 (doi: 10.1109/ICE-CCN.2013.6528546).

[9] J. Nikonowicz, P. Kubczak, and L. Matuszewski, "Hybrid detection based on energy and entropy analysis as a novel approach for spectrum sensing", in *Proc. Int. Conf. on Sig. and Elec. Syst. ICSES 2016*, Cracow, Poland, 2016 (doi: 10.1109/ICSES.2016.7593852).

[10] R. Tandra and A. Sahai, "SNR Wall for signal detection", *IEEE J. of Selec. Topics In Sig. Process.*, vol. 2, no. 1, pp. 4–17, 2008.

[11] Y. Zeng, Y.-Ch. Liang, A. T. Hoang, and R. Zhang, "A review on spectrum sensing for cognitive radio: Challenges and solutions", *EURASIP J. on Adv. in Sig. Process.*, vol. 2010, Article no. 2, 2010 (doi: 10.1155/2010/381465).

[12] Y. Zeng and Y.-Ch. Liang, "Covariance based signal detections for cognitive radio", in *Proc. 2nd IEEE Int. Symp. on New Front. in Dynam. Spectrum Access Netw. DySPAN 2007*, Dublin, Ireland, 2007, pp. 202–207.

[13] M. Öner and F. Jondral, "Air interface identification for software radio systems", *Int. J. Electron. Commun. (AEU)*, vol. 61, no. 2, pp. 104–117, 2007.

[14] W. Gardner, *Statistical Spectral Analysis: A Nonprobabilistic Theory*. NJ: Prentice Hall, 1987.

[15] L. P. Goo, "Detectors for Cognitive Radio. Master Thesis", National University of Singapore, 2007.

[16] IEEE Standard 802.16-2004 [Online]. Available: http://ieee802.org/16/pubs/80216-2004.html

[17] Z. Ye, G. Memik, and J. Grosspietsch, "Energy detection using estimated noise variance for spectrum sensing in cognitive radio networks", in *Proc. IEEE Wirel. Commun. and Netw. Conf. WCNC 2008*, Las Vegas, NV, USA, 2008, pp. 711–716.

Mateusz Kustra received his M.Sc. degree from the Faculty of Electronics of the Military University of Technology, Warsaw, Poland, in 2012. He has been working at the Military Communications Institute since 2012. He works in a team of Radio Communications and Electronic Warfare Department. He is a co-author of several scientific publications and a co-designer of a spectrum management application. His research interests includes spectrum management, dynamic spectrum access, radio communications systems and signal processing.
E-mail: m.kustra@wil.waw.pl
Military Communication Institute
Warszawska 22A
05-130 Zegrze, Poland

Krzysztof Kosmowski received his Ph.D. degree in Wireless Communications from the Military University of Technology, Warsaw. He has been with the Military Communications Institute in Zegrze since 2001. He works as the head of a team and an assistant at the Radio Communications and Electronic Warfare Department. His research focus includes signal processing, radio propagation and physical layers of radio communications systems.
E-mail: k.kosmowski@wil.waw.pl
Military Communication Institute
Warszawska 22A
05-130 Zegrze, Poland

Marek Suchański received his Ph.D. and D.Sc. degrees from the Military University of Technology, Faculty of Electronics, Warsaw, Poland in 1978 and 2013, respectively. His main areas of interests are: wireless communications, propagation of radio waves and spectrum management. He currently works as a Professor at the Military Communications Institute. He has published papers in many communications magazines, and has contributed to numerous domestic and international conferences.
E-mail: m.suchanski@wil.waw.pl
Military Communication Institute
Warszawska 22A
05-130 Zegrze, Poland

Paper

Performance Analysis of SPSK with Dual Polarized Transmit Antennas over Rayleigh Fading Channel

Malarvizhi Subramani, Arumbu Vanmathi Neduncheran, and Vijayakumar Ponnusamy

Department of Electronics and Communications Engineering (ECE), SRM IST-Kattankulathur, Chennai, India

https://doi.org/10.26636/jtit.2018.118717

Abstract—In this paper, the Space Polarization Shift Keying (SPSK) system, which is an extended version of Space Shift Keying (SSK) and includes both space and polarization dimensions with dual polarized antennas, is studied. The capacity and error probability analysis of SPSK for different polarization configurations are dealt with over the Rayleigh fading channel with a rich scattering environment. The analysis conducted shows that the SPSK with a dual polarized antenna (DP) offers better results than SPSK with a single orthogonal polarized antenna SPSK(SP), under non-LOS conditions. SPSK(DP) outperforms SSK by 4.2 dB over the correlated Rayleigh fading channel at 3 bits/s/Hz.

Keywords—ABER, dual polarized antenna, Rayleigh fading, spatial modulation, SPSK, SSK.

1. Introduction

In wireless communications, the need for high data exchange rates and spectral efficiency is increasing on a continuous basis. Demand for spectral efficiency means that the power required for transmission is growing as well. As the power increases, more radiation will be emitted from the antenna which, in turn, increases the pollution. So, reduction of the transmit power, achieved with a simultaneous data rate increase, is the major concern. High data exchange and spectral efficiency rates can be achieved by using more antennas, both in the transmitter and the receiver, and this means that the Multiple Input Multiple Output (MIMO) scheme is employed. The higher number of transmit antennas increases the data exchange rate and information is transmitted in a parallel way.

One such case is the Vertical-Bell Laboratories Layered Space- Time system (V-BLAST), where information is spilt into many parallel sub-streams which are allowed to transmit from all transmit antennas at the same time. This increases both the data rate and the spectral efficiency, but synchronization is required for all the transmit antennas to send information at the same time. V-BLAST employs spatial multiplexing here. The major disadvantage is linked to Inter Channel Interference (ICI) and Inter Antenna Interference (IAI). As more antennas are active at the same time, many RF channels are required to transmit the information, which increases the level of radiation.

In the development of green wireless communication spatial modulation (SM), as introduced by Mesleh *et al.*, where multiplexing was used to increase the data rate of a single antenna system, it is proposed that the complexity of the receiver design should remain low [1]. In SM, only one antenna is active at a given time, hence one RF chain is required which, in turn, reduces radiation. IAI and ICI can be avoided. In SM information is conveyed through both modulation index M and antenna index N_t. As the number of transmit antennas increases, the number of information bits increases as well. The total number of information bits transmitted is $\log_2(M \times N_t)$. In Space Shift Keying (SSK), however, the information is conveyed through the antenna index only [2].

Polarization Shift Keying Modulation (PSKM) is presented in [3], where binary bits are encoded as orthogonal states of polarization. In contrast to binary PSKM scheme, the multi-level PSKM scheme supports higher data rates, as multiple bits are encoded in one symbol. An additional polarization index is also added to the antenna index to relay the information, known as Space Polarization Shift Keying (SPSK) [4]. In SPSK, instead of a modulation index, a polarization index P is used along with dual polarized antennas at both input and output points. So, the total number of bits transmitted is $\log_2(2N_t \times P)$. For example, instead of using four single polarized transmit antennas in SM, only two dual polarized antennas are used here, which reduces the installation cost and also improves the spectral efficiency compared to single polarized antennas. Applications like Internet of Things and mobile devices with receive antenna orientation may opt for SPSK(DP) to improve spectral efficiency.

This paper is organized as follows: Section 2 introduces the system model. Section 3 deals with the performance analysis in detail. Section 4 discusses the results achieved by the system proposed in this paper.

2. System Model

A general system model is shown in Fig. 1, consisting of N_t dual polarized transmit antennas and N_r dual polarized receive antennas. A random sequence of information bits

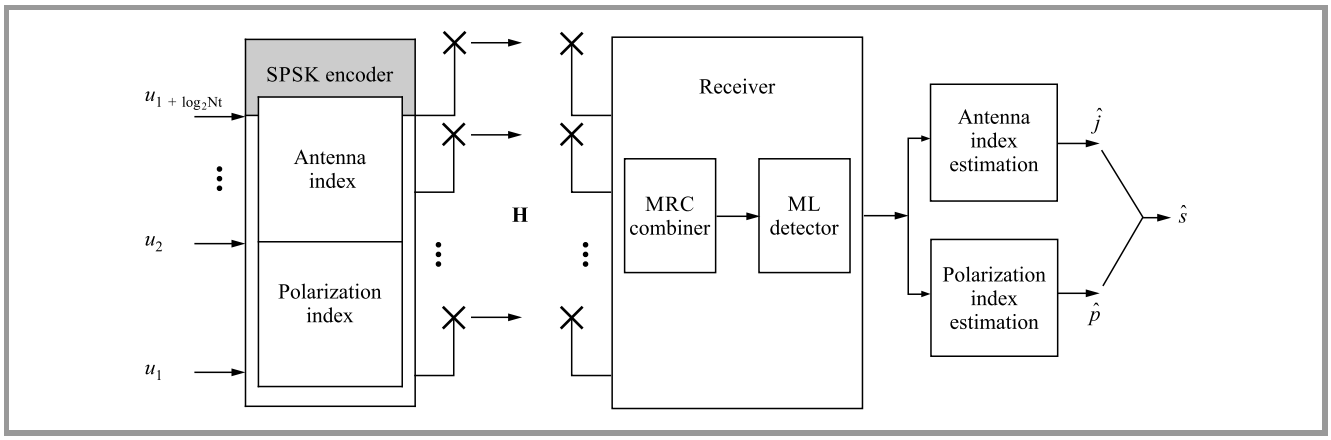


Fig. 1. SPSK system model with dual polarized antennas.

enters the encoder which groups the information sequence of size $m = 2\log_2 N_t + 1$ to a constellation vector:

$$s_{jp} = \left\{ \begin{array}{cccccccccc} 0 & 0 & \ldots & 0 & & 1^p & & 0 & \ldots & 0 & 0 \\ & & & & & \uparrow & & & & & \\ & & & & & j\text{-th position} & & & & & \end{array} \right\},$$

where j is the active antenna index and p is the polarization index (x denotes horizontal and y denotes vertical polarization).

Each group is divided into two blocks, the first bit chooses the polarization index and the next $2\log_2 N_t$ bits choose the antenna index. In the SPSK system, only one antenna is active during the transmission. All others are idle. Compared to the SPSK system with a single polarization transmit antenna, the SPSK(DP) system proposed, with dual polarized transmit antennas, increases the spectral efficiency twice compared to the earlier system. The receiver combines the signal received from N_r receive antennas. ML detector estimates the polarization index, the antenna index and the estimated symbol $\hat{s}$ is obtained.

3. SPSK(DP) Performance Analysis

The horizontal polarized component E_x and vertical polarized component E_y are represented as:

$$\begin{aligned} E_x &= a_1 \cos(\omega t + \theta_1) \\ E_y &= a_2 \cos(\omega t + \theta_2) \end{aligned}, \tag{1}$$

where a_1 and a_2 are Rayleigh distributed random variables and θ_1 and θ_2 are uniformly distributed random variables. The signal received at the two antenna elements is given by:

$$\begin{aligned} V_1 &= E_y \cos\alpha + E_x \sin\alpha \\ V_2 &= -E_x \cos\alpha + E_y \sin\alpha \end{aligned}, \tag{2}$$

where only the vertical component is present for $\alpha = 0^\circ$ and only the horizontal component is present for for $\alpha = 90^\circ$.

The dual polarized antenna at the receiver side, receives the transmitted signal with two branches. Due to the multipath effect, the signal with either horizontal or vertical polarization may get depolarized. So, the channel is assumed to be constant during the symbol time. The perfect knowledge of the channel state at the receiver side is assumed. Maximum Likelihood (ML) detection is performed at the receiver side. From the combined receive signal Y, first antenna index $\hat{j}$ is estimated, then the polarization index $\hat{p}$ of the active antenna is estimated based on cross-correlation comparing the signal strength from co-polar and cross-polar branches. The two estimation processes are interdependent. The average received power on the co-polar branches is greater than on the cross-polar branches. Therefore, polarization index estimation does not lead to errors. Hence, in the system with N_r receive and N_t transmit antennas the overall probability of error is reduced compared to SSK.

The signal at the receiver is:

$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_{N_r} \end{bmatrix} = \sqrt{E_S} \begin{bmatrix} a_{k,1} e^{-j\theta_{k,1}} \\ a_{k,2} e^{-j\theta_{k,2}} \\ . \\ . \\ . \\ a_{k,N_r} e^{-j\theta_{k,N_r}} \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \\ . \\ . \\ . \\ n_{N_r} \end{bmatrix}. \tag{3}$$

The received signal in vector form is,

$$Y = S_k + N. \tag{4}$$

In Eq. (4), the received signal is represented as $[y_1, y_2, \ldots, y_{N_r}]^T$, $\sqrt{E_s}$ is the signal power, $a_{k,j}$ is amplitude and $\theta_{k,j}$ is angle of signal s_k multiplied with channel coefficients h_j. $[n_1, n_2, \ldots, n_{N_r}]^T$ is the noise vector with AWGN symmetric parameters.

The decision metric D_k is given by:

$$\begin{aligned} D_{i|u_k} &= < y, \hat{s}_i > - \frac{1}{2} < \hat{s}_i, \hat{s}_i > \\ D_{i|u_k} &= < \hat{s}_k, \hat{s}_i > + < w, \hat{s}_i > - \frac{1}{2} \left\| \hat{s}_i \right\|^2 \end{aligned}. \tag{5}$$

The ML decision rule of SPSK(DP) for BPSK modulation is written as:

$$\hat{u} = \left\{ \begin{array}{ll} u_1 & \text{if } D_1 > D_2 \\ u_2 & \text{otherwise} \end{array} \right., \tag{6}$$

where $<\hat{s}_i, \hat{s}_i>$ is the inner product of vectors $\hat{s}_i$ and $\hat{s}_i$. u_1, u_2 are binary symbols transmitted by signals $s_1(t)$ and $s_2(t)$.

The conditional error probability is given by [4]:

$$p(e|H) = \frac{1}{2} p_{u_1}(e|H) + \frac{1}{2} p_{u_2}(e|H), \tag{7}$$

$$p(e|H) = p(||\hat{y} - \hat{s_1}||^2 - ||\hat{y} - \hat{s_2}||^2), \tag{8}$$

$$p(e|H) = p(X > ||\hat{s_1} - \hat{s_2}||^2), \tag{9}$$

$$p(e|H) = Q\left(\sqrt{\frac{E_s||\hat{s_1} - \hat{s_2}||^2}{4N_0}}\right). \tag{10}$$

By the explanation of Q function

$$p(e|H) = \frac{1}{\pi}\int_0^{\infty} \exp\left(\frac{\rho}{4\sin^2\theta}\right)^{N_r} f_\rho(\rho)\mathrm{d}\rho\mathrm{d}\theta, \tag{11}$$

$$p(e|H) = \mathscr{M}_\rho(s)^{N_r}\Big|_{s=\frac{\rho}{4\sin^2\theta}} \mathrm{d}\theta. \tag{12}$$

The Eq. 12 is reduced as closed form expression using [5]:

$$p(e|H) = \left(\frac{1-\mu(c)}{2}\right)^{N_r} \sum_{k=0}^{m-1} \binom{m-1+k}{k} \left(\frac{1+\mu(c)}{2}\right)^k, \tag{13}$$

where $c = \frac{\rho}{4}$ and $\mu(c) = \sqrt{\frac{c}{1+c}}$.

The average error probability of N_t transmit antennas is [4]:

$$P_b = \frac{\frac{N_t}{2}}{N_t - 1}\left[\frac{1}{N_t}\sum_{m=1}^{N_t}\sum_{m\neq n=1}^{N_t} p(e|H)\right]. \tag{14}$$

4. Results and Discussion

For $N_t = 4$ and $N_r = 4$ transmit antennas SPSK symbol with the speed of 3 bits/s/Hz is analyzed in the SPSK(DP) modulation scheme. The dual polarized antenna at the receiver side, receives the transmitted signal with two branches. The theoretical expressions are validated by Monte Carlo simulation. From Fig. 2, the SPSK system with dual polarized antennas SPSK(DP) outperforms both SSK and SPSK with a single polarized antenna. The power allocated to SPSK(SP) and SPSK(DP) is the same in the rich scattering Rayleigh environment. For the SNR of 14 dB, SSK, SPSK(SP) and SPSK(DP) achieve the error performance of $1.48 \cdot 10^{-5}$, $4.9 \cdot 10^{-6}$ and $3.57 \cdot 10^{-7}$, respectively. In the case of LOS, single polarized antennas will outperform dual polarized transmit antennas [4]. For the fixed average bit error rate (ABER) value, SPSK(DP) outperforms SSK and SPSK(SP) by 4.2 dB and 3 dB, respectively.

From Fig. 3, the capacity of SPSK schemes is increased compared to SSK. In the SSK system, the information bits are transmitted only through the antenna index, and in SPSK, orthogonal polarization is also indexed for the transmission of information.

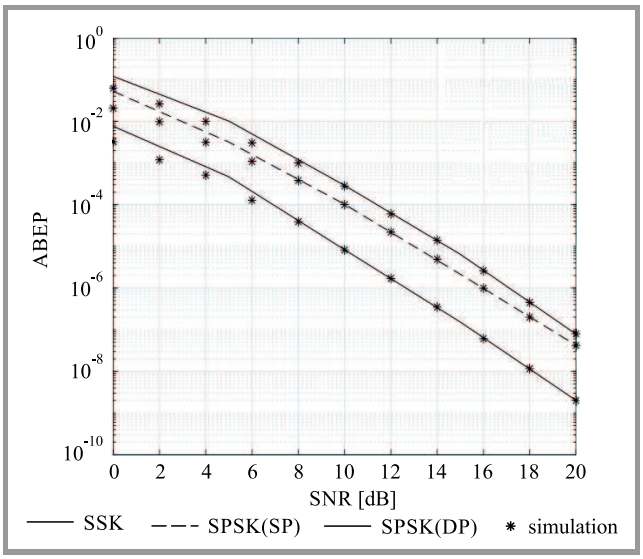


Fig. 2. Average bit error probability of different configurations compared to SSK.

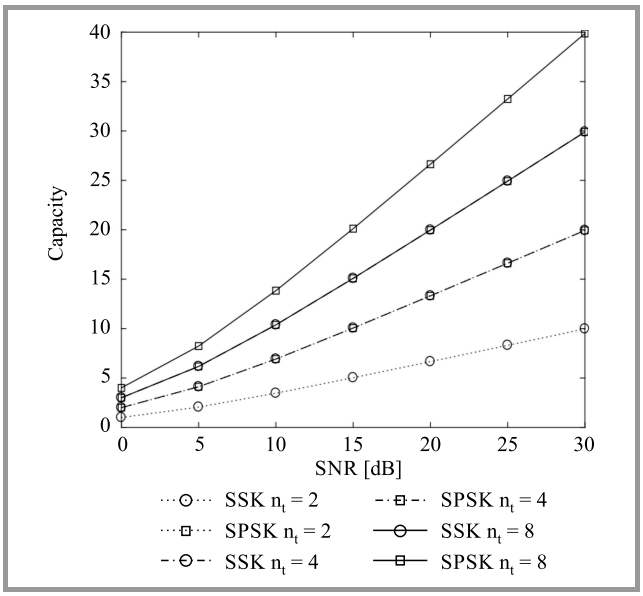


Fig. 3. Channel capacity of different configurations compared to SSK.

5. Conclusions

This paper compares the performance of SSK, SPSK(SP), SPSK(DP) systems. Dual polarized antennas at the transmitter and receiver of SPSK increase channel capacity and reduce error probability compared to SSK systems in the rich scattering environment. The increase in size of spatial constellation with dual polarized antennas results in better performance.

System performance studied over generalized fading channels is the subject of future work. The SPSK(DP) system can be experimentally evaluated using the NI LabVIEW software and the 18-slot PXIe chassis hardware platform. USRP RIO 2943R with the 2×2 MIMO configuration can be used as transceivers.

References

[1] R. Mesleh and H. Haas, "Spatial modulation", *IEEE Trans. on Veh. Technol.*, vol. 57, no. 4, pp. 2228–2241, 2008.

[2] M. Di Renzo and H. Haas, "A general framework for performance analysis of space shift keying (SSK) modulation for MISO correlated Nakagami-m fading channels", *IEEE Trans. on Commun.*, vol. 58, no. 4, pp. 2590–2603, 2010.

[3] A. Kaur and G. Kaur, "Polarization shift keying modulation: A novel modulation technique for FSO systems", *Int. J. of Engin. Develop. and Res.*, vol. 4, no. 2, pp. 1454–1459, 2016.

[4] S. Dhanasekaran, "Space-polarization shift keying modulation for MIMO channels", *Wirel. Personal Commun.*, vol. 86, no. 3, pp. 1509–1539, 2016.

[5] M. K. Simon and M. S. Alouni, *Digital Communication over Fading Channels: A Unified Approach to Performance Analysis*, 1st ed. Wiley-Interscience, 2000.

Malarvizhi Subramani is working as a Professor in the ECE Department, SRM University, Kattankulathur, Chennai, India. She finished her Ph.D. at the Wireless Communication College of Engineering, Guindy, Chennai. Her research interests are in the area of wireless communications, wireless sensor networks, MIMO and massive MIMO communications, as well as implementation of communications algorithms in FPGA.
E-mail: malarvizhi.g@ktr.srmuniv.ac.in
Department of Electronics and Communications Engineering (ECE)
SRM IST-Kattankulathur
Chennai, India

Arumbu Vanmathi Neduncheran is currently a research scholar in the ECE Department, SRM University, Kattankulathur, Chennai, India. She completed her M.Eng. in Communication Systems at the College of Engineering, Guindy, Chennai. She is a life member of ISTE. Her research interests are spatial modulation, cooperative communication, and massive MIMO.
E-mail: arumbu.n@ktr.srmuniv.ac.in
Department of Electronics and Communications Engineering (ECE)
SRM IST-Kattankulathur
Chennai, India

Vijayakumar Ponnusamy is working as an Assistant Professor in the ECE Department, SRM University, Chennai, India. He has completed his M.Eng. in Applied Electronics at the College of Engineering, Guindy, Chennai. He has completed his B.Eng. in Electronics and Communication Engineering from the Madras University. He is a member of the Communication Society of IEEE. His current research interests are in the area of MIMO wireless communications, cognitive radio networks, software defined radio and intelligent systems.
E-mail: vijayakumar.p@ktr.srmuniv.ac.in
Department of Electronics and Communications Engineering (ECE)
SRM IST-Kattankulathur
Chennai, India

Paper

Outage Performance of Bidirectional Full-Duplex Amplify-and-Forward Relay Network with Transmit Antenna Selection and Maximal Ratio Combining

R. Rajesh[1], P. G. S. Velmurugan[2], S. J. Thiruvengadam[2], and P. S. Mallick[3]

[1] *School of Electronics Engineering, VIT University, Vellore, India*
[2] *Thiagarajar College of Engineering, Madurai, India*
[3] *School of Electrical Engineering, VIT University, Vellore, India*

https://doi.org/10.26636/jtit.2018.116417

Abstract—In this paper, a bidirectional full-duplex amplify-and-forward (AF) relay network with multiple antennas at source nodes is proposed. Assuming that the channel state information is known at the source nodes, transmit antenna selection and maximal ratio combining (MRC) are employed when source nodes transmit information to the relay node and receive information from the relay node respectively, in order to improve the overall signal-to-interference plus noise ratio (SINR). Analytical expressions are derived for tight upper bound SINR at the relay node and source nodes upon reception. Further, closed form expressions are also derived for end-to-end outage probability of the proposed bidirectional full-duplex AF relay network in the Nakagami-*m* fading channel environment. Although self-interference at the relay node limits the performance of the full-duplex network, the outage performance of the proposed network is better than that of conventional bidirectional full-duplex and half-duplex AF relay networks, due to the selection diversity gain in TAS and diversity and array gain in MRC.

Keywords—channel state information, Nakagami-m fading channel, self interference, SINR.

1. Introduction

Bidirectional full-duplex relay assisted wireless networks have attracted considerable interest in the field of wireless communications due to their capability to improve network coverage, capacity, data rate and spectral efficiency [1]–[4]. Several IEEE protocol standards, such as IEEE 802.16j and LTE-A, adopted the relay technology to improve coverage and capacity [5], [6]. In a bidirectional full-duplex amplify-and-forward (AF) relay network, source nodes transmit data to the relay node and, at the same time, the relay node forwards the amplified signal back to the source nodes to utilize the spectral resources more efficiently [7]. Therefore, the data rate is doubled compared to the conventional half-duplex decode-and-forward (DF) relay network with physical-layer network coding (PLNC) for mutual data exchange [8]. However, its performance is limited mainly by self-interference due to the leakage from the full-duplex relay node transmission to its own reception.

Bidirectional relaying has become a potential candidate for sustaining the evolution of fifth generation (5G) technologies towards denser heterogeneous networks, with flexible relaying modes being studied recently [9]. There are many issues, such as synchronization and channel estimation errors, which are considered in [10]–[13]. It has been reported that the increased degree of freedom (DoF) offered by spatial domain antenna arrays of multiple input multiple output (MIMO) networks may be utilized to provide a range of new solutions for self-interference cancellation/suppression [14]. Wireless transmission and reception using multiple antennas also improves the overall signal-to-noise ratio (SNR) of the network through diversity and array gains, and, hence, provides high capacity and reliability [15], [16].

However, as the number of antennas increases, the requirement to provide expensive radio frequency chains also grows [17]. Further, designing power amplifiers with the required large linearity region is a major challenge. Transmit antenna selection (TAS) is one of the alternatives enabling to solve this problem. In TAS, the transmitter selects the best antenna based on the channel state information from the receiver, obtained through the feedback channel. Thus, it requires only one RF chain, although many cheap antenna elements are used in the network. In [18]–[21] different algorithms have been proposed for antenna selection in a one way AF relay network.

In [22] antenna selection is done at both transmitter and receiver ends of the bidirectional half-duplex MIMO AF relay network. In [23], [24], the combination of TAS and maximal ratio combining (MRC) at the receiver is used in a one-way AF relay network. In [25], the performance of

a dual hop AF MIMO multi-antenna relay network with the best end-to-end antenna selection was considered and compared with the transmit beamforming MRC in [26]. In [27], a wide variety of antenna selection schemes based on low complexities are proposed for one way full-duplex relaying. In [28], source transmit antenna selection for MIMO DF relay networks is proposed based on the channel state information to maximize the diversity and joint coding for multiple relay networks.

The focus of earlier research work was on transmit antenna selection in half-duplex bidirectional wireless relay networks, self-interference is the major limiting factor in full-duplex bidirectional relay network. In this paper, a bidirectional full-duplex AF relay network with multiple antennas at source nodes is proposed. Assuming that the channel state information is known at the source nodes, TAS and MRC are employed when source nodes transmit information to the relay node and receive information from the relay node respectively, in order to improve the overall SINR.

The major contributions of this paper are:

- TAS and MRC are employed at source nodes in the proposed bidirectional full-duplex AF relay network, which maximizes the overall SINR.
- A closed form analytical expression is derived for end-to-end outage probability of the proposed network in the Nakagami-*m* fading channel environment. The Nakagami-*m* fading model represents a wide variety of realistic line of sight (LoS) and non-LoS fading channels encountered in practice. Hence, the derived analytical expression of outage probability can be used to investigate the outage characteristics under different fading severity conditions of the proposed bidirectional AF full-duplex relay network.
- The outage performance of the proposed network is also compared with a half-duplex bidirectional AF relay network with TAS and MRC in the Nakagami-*m* fading environment.

The remaining part of this paper is organized as follows. The system model of the proposed bidirectional full-duplex AF relay network is presented in Section 2. Section 3 provides the outage performance analysis of the proposed bidirectional AF relay network with TAS and MRC. Numerical results are presented in Section 4, and concluding remarks are given in Section 5.

2. System Model

A bidirectional FD AF relay network is shown in Fig. 1. The source nodes S_j, $j = 1,2$ are equipped with N multiple antenna terminals and an intermediate FD AF relay node terminal R is equipped with two antennas, one for transmitting and the other for receiving.

Let P_s and P_r be the transmit power at source nodes S_1, S_2 and relay node R respectively.

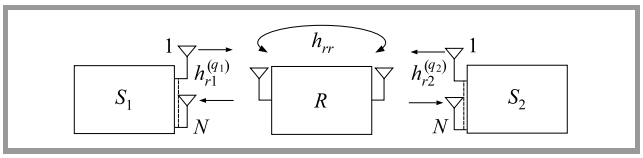


Fig. 1. System model of a bidirectional FD AF relay network.

The direct link between the source nodes does not exist due to the shadowing effect. The channel impulse response vector between source nodes S_1, S_2 and relay node R is written as $\mathbf{h}_{jr} = \left[h_{rj}^{(1)}, h_{rj}^{(2)}, \ldots, h_{rj}^{(N)}\right]^T$, $j = 1,2$, where T denotes the transpose of channel vector. Each element of the channel impulse response vector is denoted by $h_{rj}^{(l)} = \left|h_{rj}^{(l)}\right| e^{j\theta}$, $l = 1,2,\ldots,N$, where $\left|h_{rj}^{(l)}\right|$ is Nakagami-m distributed with shape and scale parameters m, Ω and θ being uniformly distributed over $[0, 2\pi]$. The probability density function (PDF) of $\left|h_{rj}^{(l)}\right|$ is given by:

$$f_{\left|h_{rj}^{l}\right|}(x) = \frac{2}{\Gamma(m_l)}\left(\frac{m_l}{\Omega_l}\right)^{m_l} x^{2m_l-1}\, \mathrm{e}^{-\frac{m_l}{\Omega_l}x^2}, \quad x \geq 0, \quad l = 1,2,\ldots,N, \tag{1}$$

where $\Gamma(.)$ is the Gamma function, defined as $\Gamma(\Phi) = \int_0^\infty x^{\Phi-1}e^{-\Phi}\mathrm{d}x$. The shape parameters $m_1, m_2, \ldots, m_N$ and scale parameters $\Omega_1, \Omega_2, \ldots, \Omega_N$ are independent. Nakagami-m fading represents a wide range of fading scenarios through its shape parameter m, which ranges from 0.5 to ∞. The $m = 0.5$ represents worst-case fading and $m = \infty$, no fading. $m = 1$ corresponds to Rayleigh fading. Hence, Rayleigh fading becomes a special case of the Nakagami-m fading channel. As m increases further, the severity of fading reduces [16]. It is assumed that channel reciprocity exists between source nodes and the relay node; then $h_{rj}^{(l)} = h_{jr}^{(l)}$ $l = 1,2,\ldots,N$ holds.

The transmit antenna selection q_j at the source nodes $S_j, j = 1,2$ is based on the following criteria:

$$q_j = \underset{q\in\{1,2,\ldots,N\}}{\arg\max}\left[\left|h_{jr}^{q}\right|^2\right], \quad j = 1,2. \tag{2}$$

The antenna selection information is sent to the relay node R through a dedicated control channel. In k-th time slot, the receive signal at relay node R is given by:

$$y_r(k) = \sqrt{P_s}h_{r1}^{q_1}x_1(k) + \sqrt{P_s}h_{r2}^{q_2}x_2(k) + h_{rr}t_r(k) + n_r(k). \tag{3}$$

The symbols $x_j(k)$, $j = 1,2$ transmitted by both source nodes S_j, $j = 1,2$ belong to the unit symbol energies $E\left\{\left|x_1\right|^2\right\} = 1$ and $E\left\{\left|x_2\right|^2\right\} = 1$. The $n_r(k)$ is an additive white Gaussian noise at relay node R with zero mean and variance of σ_n^2. As the relay node R operates in a full-duplex mode, it broadcasts the receive signal at the $(k-1)$-th time slot to both the source nodes S_j, $j = 1,2$ with an

amplification factor β_r. The transmit signal $t_r(k)$ at relay node R at k-th time slot is given by:

$$t_r(k) = \sqrt{P_r}\beta_r y_r(k-1). \tag{4}$$

The self-interference h_{rr} at relay node R is also modeled as Gaussian distributed with zero mean and variance $\sigma_{e,r}^2$. Substituting Eq. (3) with Eq. (4), $t_r(k)$ can be expressed as:

$$t_r(k) = \beta_r\sqrt{P_r}\sum_{n=1}^{\infty}\left(h_{rr}\beta_r\sqrt{P_r}\right)^{n-1} g(k-n), \tag{5}$$

where $g(k) = \sqrt{P_s}h_{r1}^{q_1}x_1(k-n) + \sqrt{P_s}h_{r2}^{q_2}x_2(k-n) + n_r(k-n)$.

The variance of $t_r(k)$ is derived as:

$$\mathrm{E}\left[\left|t_r(k)\right|^2\right] =$$
$$= \left(P_s\left|h_{r1}^{q_1}\right|^2 + P_s\left|h_{r2}^{q_2}\right|^2 + \sigma_n^2\right)\frac{\beta_r^2 P_r}{1-\left|h_{rr}\right|^2\beta_r^2 P_r} = P_r. \tag{6}$$

The variance of $t_r(k)$ should be equal to P_r. Equating the right-hand side expression of Eq. (6) with P_r to prevent the oscillation at the relay node R, the amplification factor β_r of the relay node is calculated as:

$$\beta_r = \frac{1}{\sqrt{P_s\left|h_{r1}^{q_1}\right|^2 + P_s\left|h_{r2}^{q_2}\right|^2 + P_r\left|h_{rr}\right|^2 + \sigma_n^2}}. \tag{7}$$

The receive signals at l-th antenna of source nodes S_j, $j = 1, 2$ are:

$$y_j^l(k) = h_{jr}^l t_r(k) + \sqrt{P_s}h_{jj}^{q_j}\delta(l-q_j)x_j(k) + n_j^l(k),$$
$$l = 1, 2, \ldots, N, \quad j = 1, 2 \tag{8}$$

where $\delta(l-q_j) = \begin{cases} 1 & \text{if } l = q_j \\ 0 & \text{if } l \neq q_j \end{cases}$.

By substituting Eq. (4) and Eq. (7) in Eq. (8), the receive signal at $S_j, j = 1, 2$ by l-th antenna is expressed as:

$$\begin{aligned}\mathbf{y}_j^l(k) = {} & \sqrt{P_rP_s}\beta_r h_{jr}^l h_{jr}^{q_j}x_j(k-1) + \\ & + \sqrt{P_rP_s}\beta_r h_{jr}^l h_{\bar{j}r}^{q_{\bar{j}}}x_{\bar{j}}(k-1) + \\ & + \sqrt{P_r}\beta_r h_{jr}^l h_{rr}t_r(k-1) + \\ & + \sqrt{P_s}h_{jj}^{q_j}\delta(l-q_j)x_j(k) + \sqrt{P_r}\beta_r h_{jr}^l n_r(k-1) + \\ & + n_j^l(k), \\ & l = 1, 2, \ldots, N, \ j = 1, 2,\end{aligned} \tag{9}$$

where $\{j, \bar{j}\} = \{1, 2\}$ or $\{2, 1\}$. In vector form, it is written as:

$$\begin{aligned}\mathbf{y}_j(k) = {} & \sqrt{P_rP_s}\beta_r\mathbf{h}_{jr}h_{jr}^{q_j}x_j(k-1) + \\ & + \sqrt{P_rP_s}\beta_r\mathbf{h}_{jr}h_{\bar{j}r}^{q_{\bar{j}}}x_{\bar{j}}(k-1) + \\ & + \sqrt{P_r}\beta_r\mathbf{h}_{jr}h_{rr}t_r(k-1) + \\ & + \sqrt{P_s}\mathbf{D}_{qj}\mathbf{h}_{jj}x_j(k) + \sqrt{P_r}\beta_r\mathbf{h}_{jr}n_r(k-1) + \\ & + \mathbf{n}_j(k), \\ & j = 1, 2.\end{aligned} \tag{10}$$

The first term in Eq. (10) is the self-information signal at the source node S_j, $j = 1, 2$ and it can be subtracted. The second term is the useful signal from $S_{\bar{j}}$. The third and fourth terms are the self-interference at relay node R and source S_j respectively. The last two terms are the noise at relay node R and source node S_j. The term $\mathbf{D}_{qj}$ is a diagonal matrix in which the q_j-th diagonal element is unity. After subtracting, the self-information signal term is:

$$\begin{aligned}\mathbf{y}_j(k) = {} & \sqrt{P_rP_s}\beta_r\mathbf{h}_{jr}h_{\bar{j}r}^{q_{\bar{j}}}x_{\bar{j}}(k-1) + \\ & + \sqrt{P_r}\beta_r\mathbf{h}_{jr}h_{rr}t_r(k-1) + \\ & + \sqrt{P_s}\mathbf{D}_{qj}\mathbf{h}_{jj}x_j(k) + \sqrt{P_r}\beta_r\mathbf{h}_{jr}n_r(k-1) + \\ & + \mathbf{n}_j(k), \\ & j = 1, 2.\end{aligned} \tag{11}$$

Since each node is equipped with multiple antennas, MRC is applied at each source node to obtain diversity gain. Now the MRC output at the source nodes S_j is expressed as:

$$z_j(k) = \mathbf{h}_{jr}^*\mathbf{y}_j(k). \tag{12}$$

Substituting Eq. (11) in Eq. (12), the MRC output is:

$$\begin{aligned}z_j(k) = {} & \sqrt{P_rP_s}\beta_r\left\|\mathbf{h}_{jr}\right\|^2 h_{\bar{j}r}^{q_{\bar{j}}}x_{\bar{j}}(k-1) + \\ & + \sqrt{P_r}\beta_r\left\|\mathbf{h}_{jr}\right\|^2 h_{rr}t_r(k-1) + \sqrt{P_s}\mathbf{D}_{qj}\mathbf{h}_{jj}x_j(k) + \\ & + \sqrt{P_r}\beta_r\left\|\mathbf{h}_{jr}\right\|^2 n_r(k-1) + \mathbf{h}_{jr}^*\mathbf{n}_j(k).\end{aligned} \tag{13}$$

Since source nodes S_j, $j = 1, 2$ are ideal and non-iterative, the effect of self-interference term $\sqrt{P_s}\mathbf{D}_{qj}\mathbf{h}_{jj}x_j(k)$ is neglected. Now, the MRC output is rewritten as:

$$\begin{aligned}z_j(k) = {} & \sqrt{P_rP_s}\beta_r\left\|\mathbf{h}_{jr}\right\|^2 h_{\bar{j}r}^{q_{\bar{j}}}x_{\bar{j}}(k-1) + \\ & + \sqrt{P_r}\beta_r\left\|\ \mathbf{h}_{jr}\right\|^2 h_{rr}t_r(k-1) + \\ & + \sqrt{P_r}\beta_r\left\|\mathbf{h}_{jr}\right\|^2 n_r(k-1) + \mathbf{h}_{jr}^*\mathbf{n}_j(k).\end{aligned} \tag{14}$$

Given h_{jr}, the end-to-end receive SINR at each source node is:

$$\gamma_j = \frac{P_rP_s\beta_r^2\left|h_{\bar{j}r}^{q_{\bar{j}}}\right|^2\left\|\mathbf{h}_{jr}\right\|^4}{P_r^2\beta_r^2\left|h_{rr}\right|^2\left\|\mathbf{h}_{jr}\right\|^4 + P_r\beta_r^2\left\|\mathbf{h}_{jr}\right\|^4\sigma_n^2 + \left\|\mathbf{h}_{jr}\right\|^2\sigma_n^2},$$
$$j = 1, 2. \tag{15}$$

Substituting for β_r from Eq. (7) in Eq. (15), the SINR expression is:

$$\gamma_j = \frac{P_rP_s\left|h_{\bar{j}r}^{q_{\bar{j}}}\right|^2\left|\mathbf{h}_{jr}\right|^2}{\left[\begin{array}{c}P_r^2|h_{rr}|^2\left\|\mathbf{h}_{jr}\right\|^2 + P_r\left\|\mathbf{h}_{jr}\right\|^2\sigma_n^2 + P_s\left|h_{jr}^{q_j}\right|^2\sigma_n^2 \\ + P_s\left|h_{\bar{j}r}^{q_{\bar{j}}}\right|^2\sigma_n^2 + P_r|h_{rr}|^2\sigma_n^2 + \sigma_n^4\end{array}\right]},$$
$$j = 1, 2. \tag{16}$$

Let $\xi_r = \frac{P_r}{\sigma_n^2}$, $\xi_s = \frac{P_s}{\sigma_n^2}$ and $\sigma_{e,r}^2 = |h_{rr}|^2$. The term ξ_s denotes the SNR at source nodes, ξ_r denotes the SNR at relay and

$\sigma_{e,r}^2$ denote the variance of the self-interference at relay node R. After simple manipulations, the end-to-end receive SINR at source nodes S_j, $j=1,2$ is:

$$\gamma_j = \\ = \frac{\xi_s \left|h_{\bar{j}r}^{q_{\bar{j}}}\right|^2 \xi_r \left\|\mathbf{h}_{jr}\right\|^2}{\xi_r(\sigma_{e,r}^2+1)\left\|\mathbf{h}_{jr}\right\|^2+\xi_s\left(\left|h_{\bar{j}r}^{q_{\bar{j}}}\right|^2+\left|h_{\bar{j}r}^{q_j}\right|^2\right)+\sigma_{e,r}^2+1}. \quad (17)$$

The variance of the self-interference $|h_{rr}|^2 \xi_r$ at relay node is defined as $\sigma_{e,r}^2$. As the variances of the self-interference terms are too small, the upper bound of instantaneous SINR at the source nodes is:

$$\gamma_j < \frac{\xi_r \xi_s \left|h_{\bar{j}r}^{q_{\bar{j}}}\right|^2 \left\|\mathbf{h}_{jr}\right\|^2}{\xi_r(\sigma_{e,r}^2+1)\left\|\mathbf{h}_{jr}\right\|^2+\xi_s\left|h_{\bar{j}r}^{q_{\bar{j}}}\right|^2}. \quad (18)$$

3. Performance Analysis

The overall outage probability of the proposed system occurs when γ_j falls below the threshold $\gamma\text{-th} = 2^{R_t}-1$, where R_t is the required target data rate. To derive the closed form expression for overall outage probability, SINR expression is to be rewritten in a suitable form.

The upper bound of instantaneous end-to-end SINR at source nodes S_j, $j=1,2$:

$$\frac{1}{\gamma_j} > \frac{(\sigma_{e,r}^2+1)}{\xi_s\left|h_{\bar{j}r}^{q_{\bar{j}}}\right|^2}+\frac{1}{\xi_r\left\|h_{jr}\right\|^2}. \quad (19)$$

Further, it can be approximated as:

$$\frac{1}{\gamma_j} > \max\left\{\frac{(\sigma_{e,r}^2+1)}{\xi_s\left|h_{\bar{j}r}^{q_{\bar{j}}}\right|^2}, \frac{1}{\xi_r\left\|h_{jr}\right\|^2}\right\}. \quad (20)$$

Hence, the tight upper bound of end-to-end SINR using [29] at the source nodes is:

$$\gamma_j \le \min\left\{\frac{\xi_s\left|h_{\bar{j}r}^{q_{\bar{j}}}\right|^2}{(\sigma_{e,r}^2+1)}, \xi_r\left\|\mathbf{h}_{jr}\right\|^2\right\}, \\ j=1,2. \quad (21)$$

Then, the outage probability for the proposed bidirectional AF relay network is expressed as

$$P_{out}^{s_j} = \Pr(\gamma_j < \gamma\text{-th}) \quad (22)$$

Substituting Eq. (21) in Eq. (22), the outage probability is:

$$P_{out}^{s_j} = \Pr\left[\min\left\{\frac{\xi_s\left|h_{\bar{j}r}^{q_{\bar{j}}}\right|^2}{(\sigma_{e,r}^2+1)}, \xi_r\left\|\mathbf{h}_{jr}\right\|^2\right\} \le \gamma\text{-th}\right], \\ j=1,2. \quad (23)$$

Let $P_x = \frac{\xi_s}{(\sigma_{e,r}^2+1)}$ and $P_y = \xi_r$. Equation (23) can be rewritten as:

$$P_{out}^{s_j} = \Pr\left\{\min\left\{P_x\left|h_{\bar{j}r}^{q_{\bar{j}}}\right|^2, P_y\left\|h_{jr}\right\|^2\right\} \le \gamma\text{-th}\right\}, \\ j=1,2, \quad (24)$$

$$P_{out}^{s_j} = 1-\left(1-\Pr\left\{\left|h_{\bar{j}r}^{q_{\bar{j}}}\right|^2 < \frac{\gamma\text{-th}}{P_x}\right\}\right)\times \\ \times\left(1-\Pr\left\{\left\|h_{jr}\right\|^2 < \frac{\gamma\text{-th}}{P_y}\right\}\right), \\ j=1,2. \quad (25)$$

Since $\left|h_{\bar{j}r}^{q_{\bar{j}}}\right|$ is Nakagami-m, $\left|h_{\bar{j}r}^{q_{\bar{j}}}\right|^2$ follows the Gamma distribution with shape parameter $m_{q_{\bar{j}}}$ and scale parameter $\beta_{q_{\bar{j}}} = \frac{\Phi_{q_{\bar{j}}}}{m_{q_{\bar{j}}}}$, the PDF of $X = \left|h_{\bar{j}r}^{q_{\bar{j}}}\right|$ is given by:

$$f_X(x) = \frac{1}{\Gamma(m_{q_{\bar{j}}})\beta_{q_{\bar{j}}}} x^{m_{q_{\bar{j}}}-1} \mathrm{e}^{-\frac{x}{\beta_{q_{\bar{j}}}}}. \quad (26)$$

Similarly, $\left\|\mathbf{h}_{jr}\right\|^2 = \sum_{l=1}^{N}\left|h_{jr}^l\right|^2$ follows the Gamma distribution with shape parameter, $l=1,2,\ldots,N$ and shape parameter $\beta_{q_{\bar{j}}} = \frac{\Phi_{jr}^l}{m_{jr}^l}$, $l=1,2,\ldots,N$, the PDF of $Y = \left\|\mathbf{h}_{jr}\right\|^2$ is given by:

$$f_Y(y) = \frac{1}{\Gamma\left(Nm_{jr}^l\right)\beta_{jr}^l} y^{Nm_{jr}^l-1} \mathrm{e}^{-\frac{y}{\beta_{jr}^l}}. \quad (27)$$

Substituting the Eq. (27) in Eq. (25), $\Pr\left\{\left\|h_{jr}\right\|^2 < \frac{\gamma\text{-th}}{P_y}\right\}$ is computed as:

$$\Pr\left\{\left\|h_{jr}\right\|^2 < \frac{\gamma\text{-th}}{P_y}\right\} = \\ = \int_0^{\frac{\gamma\text{-th}}{P_y}} \frac{1}{\Gamma\left(Nm_{jr}^l\right)\beta_{jr}^l} y^{Nm_{jr}^l-1} \mathrm{e}^{\frac{-y}{\beta_{jr}^l}} \mathrm{d}y = \quad (28)$$

$$= \frac{1}{\Gamma\left(Nm_{jr}^l\right)\beta_{jr}^l} \int_0^{\frac{\gamma\text{-th}}{P_y}} y^{Nm_{jr}^l-1} \mathrm{e}^{\frac{-y}{\beta_{jr}^l}} \mathrm{d}y. \quad (29)$$

The outage probability at the source nodes due to MRC is computed as [30]:

$$\Pr\left\{\left\|h_{jr}\right\|^2 < \frac{\gamma\text{-th}}{P_y}\right\} = \frac{1}{\Gamma\left(Nm_{jr}^l\right)}\Gamma\left(\frac{\gamma\text{-th}}{P_y\beta_{jr}^l}, Nm_{jr}^l\right). \quad (30)$$

It is worth noting that, even though a single active antenna is selected at both the source nodes, it is possible to achieve the N diversity order between the source nodes and the relay node. Similarly by using the PDF given in Eq. (26), the outage probability at the relay node is:

$$\Pr\left\{\left|h_{\bar{j}r}^{q_{\bar{j}}}\right|^2 < \frac{\gamma\text{-th}}{P_x}\right\} = \left[\frac{1}{\Gamma\left(m_{q_{\bar{j}}}\right)}\Gamma\left(\frac{\gamma\text{-th}}{P_x\beta_{q_{\bar{j}}}}, m_{q_{\bar{j}}}\right)\right]^N. \quad (31)$$

Substituting Eq. (30) and Eq. (31) in Eq. (25), the outage probability of the proposed system is:

$$\Pr(\gamma_j < \gamma\text{-th}) = 1 - \left\{1 - \left[\frac{1}{\Gamma\left(m_{q_{\tilde{j}}}\right)}\Gamma\left(\frac{\gamma\text{-th}}{P_x\beta_{q_{\tilde{j}}}}, m_{q_{\tilde{j}}}\right)\right]^N\right\} \times$$
$$\times \left[1 - \frac{1}{\Gamma\left(Nm^l_{jr}\right)}\Gamma\left(\frac{\gamma\text{-th}}{P_y\beta^l_{jr}}, Nm^l_{jr}\right)\right], \qquad (32)$$

where $\Gamma(...)$ is the incomplete Gamma function and it is given by $\Gamma(z,\Phi) = \int_0^z x^{\Phi-1}e^{-x}\mathrm{d}x$.

4. Numerical Results and Discussions

In this section, the outage performance of the proposed bidirectional full-duplex AF relay assisted network with TAS and MRC is analyzed using the analytical expressions derived in Section 3. The parameter values for the analytical expression and simulations are based on [7] for outage analysis and are listed in Table 1.

Table 1
List of parameters

Parameters	Values
Target data rate R_t	1, 2, 4 bps/Hz
Number of antennas at source nodes N	1, 2, 4
Shape parameter m	1, 1.5, 2
Variance of channel β	2
Variance of self-interference $\sigma^2_{\tilde{e},r}$	–20 to 30 dB
SNR at source nodes	0–30 dB
Relay power P_r	30 dB

The outage performance of the proposed bidirectional full-duplex AF relay network is shown in Fig. 2 at a target rate of $R_t = 4$ bps in the Nakagami-m fading environment with shape parameter $m = 1$ and for various numbers of antennas N at source nodes. At the outage probability of 10^{-1} it is observed that the minimum SNR requirement is approximately 20 dB when a single antenna $N = 1$ is employed at the source nodes. In the case of a bidirectional full-duplex (BFD) AF relay network without TAS and MRC [7], the SNR requirement is approximately 25 dB at $N = 1$. As the number of antennas N at the source nodes increases to 2, the minimum SNR requirement decreases to 14 dB. This improvement is due to the transmit diversity at the source nodes and the receive diversity and array gains at source nodes from MRC. It is observed that the analytical results precisely match exactly with the simulation results for $N \geq 2$. But, in a single antenna environment, there is a small deviation between analytical data and simulations at high SNR, due to the approximations that have been made to obtain the tight upper bound.

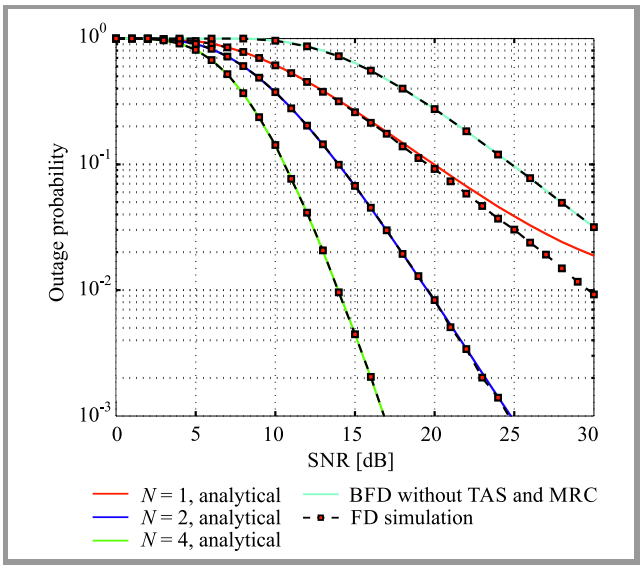


Fig. 2. Outage performance of the proposed bidirectional AF relay network in Rayleigh fading channel.

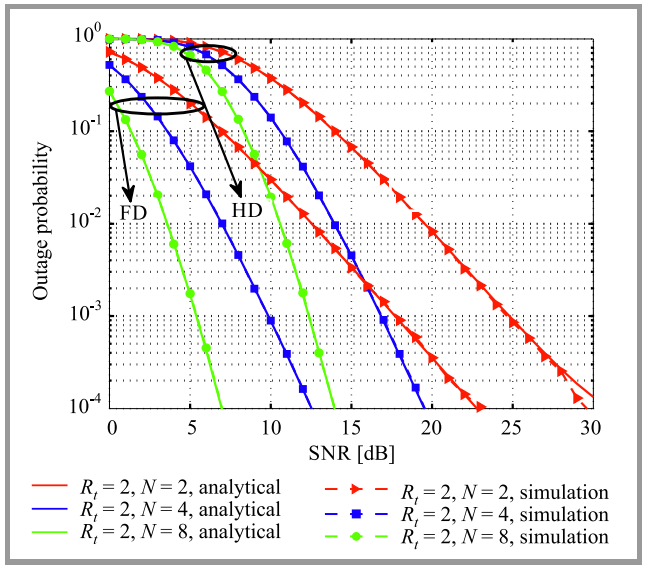


Fig. 3. Outage performance of the proposed bidirectional full-duplex AF relay network in Rayleigh fading channel and bidirectional half-duplex.

The outage performance of the proposed bidirectional full duplex AF relay network is compared with the bidirectional half-duplex AF relay network at a target rate of $R_t = 2$ bps/Hz using Fig. 3 for various numbers of antennas N at source nodes. For bidirectional half-duplex AF relay network, the outage probability expression is:

$$\Pr\left(\gamma_j < \gamma^{th}_{HD}\right) =$$
$$= 1 - \left\{1 - \left[\frac{1}{\Gamma\left(m_{q_{\tilde{j}}}\right)}\Gamma\left(\frac{\gamma^{th}_{HD}}{P^{HD}_x\beta_{q_{\tilde{j}}}}, m_{q_{\tilde{j}}}\right)\right]^N\right\} \times$$
$$\times \left[1 - \frac{1}{\Gamma\left(Nm^l_{jr}\right)}\Gamma\left(\frac{\gamma^{th}_{HD}}{P_y\beta^l_{jr}}, Nm^l_{jr}\right)\right], \qquad (33)$$

where $\gamma^{th}_{HD} = 2^{2R_t} - 1$ and $P^{HD}_x = \xi_s$.

It is observed that the proposed bidirectional full-duplex AF relay network has better outage performance when compared with a half-duplex relay network. At the outage probability of 10^{-2}, the SNR requirement in a half-duplex relay network is 20 dB when $N=2$, whereas in a full duplex network it is 13 dB. With a further increase in N value to 4, the SNR requirement in half duplex decreases to 14 dB, whereas in full-duplex it falls to 7 dB.

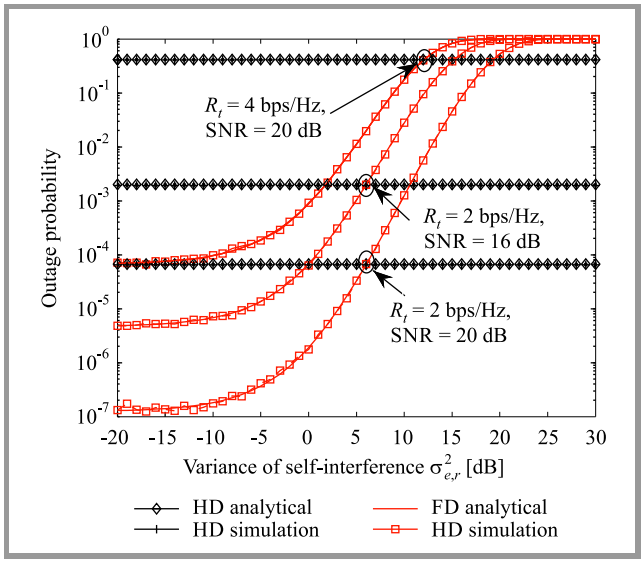


Fig. 4. Outage performance of the proposed bidirectional full-duplex AF relay network by varying the variance of the self-interference.

In Fig. 4, the effect of self-interference on the outage performance of the proposed full duplex AF relay network is examined under various target rates R_t in bps/Hz and SNR with the number of antennas at the source nodes $N = 4$. The outage probability of an FD network increases as the variance of the self interference increases. As the HD network has no effect of self-interference, its outage performance does not vary. However, the outage performance of an FD network is better than that of an HD network when the variance of self-interference $\sigma^2_{e,r}$ is very small. It is observed that the bidirectional AF FD with $R_t = 4$ bps/Hz, SNR = 20 dB and HD with $R_t = 2$ bps/Hz, SNR = 20 dB almost have the same outage performance in the case of small interference, which proves that FD is better than HD, and it doubles the rate when the self-interference is small enough.

In Fig. 5, the outage performance of the proposed bidirectional full-duplex relay network is shown for various shape parameters in the Nakagami-m fading channel. The number of antennas at the source nodes is fixed at $N = 4$ and the target rate at $R_t = 4$ bps/Hz. At an outage probability of 10^{-2}, the minimum SNR requirement of the proposed network is 14 dB when $m = 1$ (Rayleigh fading environment). When $m = 1.5$ and $m = 2$, the SNR requirement decreases to 11 dB and 9 dB respectively, as the effect of fading severity decreases.

In Fig. 6, the outage performance of the proposed bidirectional full-duplex relay network is shown for a target rate

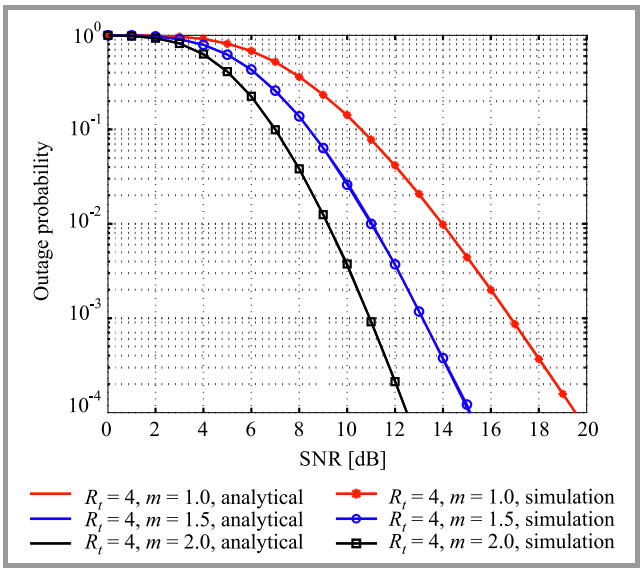


Fig. 5. Outage performance of the proposed bidirectional full-duplex relay network for different shape parameters.

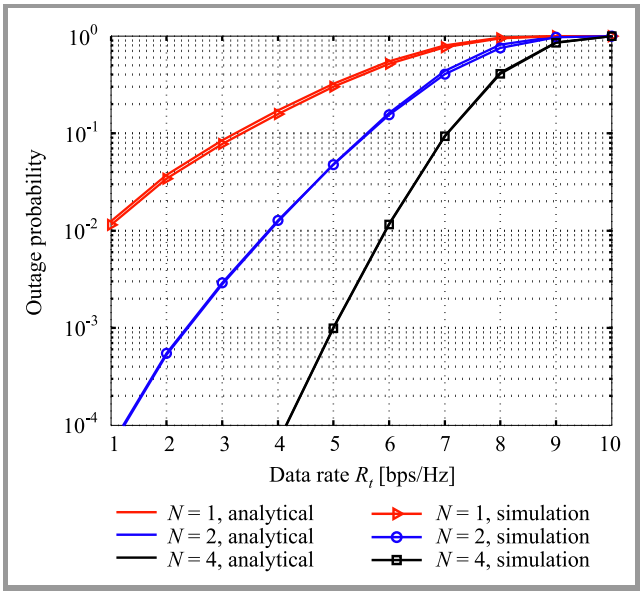


Fig. 6. Outage performance of the proposed bidirectional full-duplex relay at various data rates in Rayleigh fading channels ($m = 1$).

varying from 1 bps/Hz to and an outage probability of 10^{-3}, the data rate 10 bps/Hz. The number of antennas at source nodes is assumed to be $N=1,2,4$. For $N=2$ and an outage probability of 10^{-3}, the data rate for the proposed network is 2.5 bps/Hz, when the number of antennas N increases to 4, the same outage probability is maintained at a data rate of 5 bps/Hz. It is observed that increasing the number of antennas at the source nodes improves the data rate.

5. Conclusion

In this paper, a TAS- and MRC-based bidirectional full-duplex AF relay network is proposed to overcome the ef-

fect of self-interference at full duplex nodes and to improve the overall SINR. A closed form analytical expression is derived for the end-to-end outage probability of the proposed network in the Nakagami-*m* fading channel environment. The proposed network has better outage performance when compared with a full-duplex bidirectional AF relay network with a single antenna at the source nodes. Further, it provides better performance than a half-duplex bidirectional AF really network.

References

[1] P. Popovski and H. Yomo, "Bi-directional amplification of throughput in a wireless multi-hop network", in *Proc. IEEE 63rd Veh. Technol. Conf. VTC2006-Spring*, Melbourne, Australia, 2006, vol. 2, pp. 588–593 (doi: 10.1109/VETECS.2006.1682892).

[2] R. Vaze and R. W. Heath, "On the capacity and diversity-multiplexing tradeoff of the two-way relay channel", *IEEE Trans. on Inform. Theory*, vol. 57, no. 7, pp. 4219–4234, 2011.

[3] X. Cheng *et al.*, "Communicating in the real world: 3D MIMO", *IEEE Wirel. Commun.*, vol. 21, no. 4, pp. 136–144, 2014.

[4] D. Bharadia, E. McMilin, and S. Katti, "Full-duplex radios", *ACM SIGCOMM Comp. Commun. Rev.*, vol. 43, no. 4, pp. 375–386, 2013.

[5] V. Genc, S. Murphy, Y. Yu, and J. Murphy, "IEEE 802.16j relay-based wireless access networks: an overview", *IEEE Wirel. Commun.*, vol. 15, no. 5, pp. 56–63, 2008.

[6] Y. Yuan, *LTE-Advanced Relay Technology and Standardization*. Springer, 2012.

[7] H. Cui, M. Ma, L. Song, and B. Jiao "Relay selection for two-way full-duplex relay networks with amplify-and-forward protocol", *IEEE Trans. on Wirel. Commun.*, vol. 13, no. 7, pp. 3768–3777, 2014.

[8] S. C. Liew, S. Zhang, and L. Lu, "Physical-layer network coding: Tutorial, survey, and beyond", *Phys. Commun.*, vol. 6, pp. 4–42, 2013 (doi: 10.1016/j.phycom.2012.05.002).

[9] S. Hong *et al.*, "Applications of self-interference cancellation in 5G and beyond", *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 114–121, 2014.

[10] D. S. Michalopoulos, H. A. Suraweera, G. K. Karagiannidis, and R. Schober, "Amplify-and-forward relay selection with outdated channel estimates", *IEEE Trans. on Commun.*, vol. 60, no. 5, pp. 1278–1290, 2012.

[11] S. Won and L. Hanzo, "Synchronization issues in relay-aided cooperative MIMO networks", *IEEE Wirel. Commun.*, vol. 21, no. 5, pp. 41–55, 2014.

[12] C.-L. Wang, P.-C. Chiu, and H.-C. Wang, "Joint time synchronization and channel estimation for two-way amplify-and-forward relay systems", in *Proc. IEEE Global Commun. Conf. GLOBECOM 2014*, Austin, TX, USA, 2014, pp. 3543–3548, 2014.

[13] J. M. Moualeu, W. Hamouda, and F. Takawira, "Relay selection for coded cooperative networks with outdated CSI over Nakagami-*m* fading channels", *IEEE Trans. on Wirel. Commun.*, vol. 13, no. 5, pp. 2362–2373, 2014.

[14] Z. Zhang, X. Chai, K. Long, A. V. Vasilakos, and L. Hanzo, "Full duplex techniques for 5G networks: Self-interference cancellation, protocol design, and relay selection", *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 128–137, 2015.

[15] K. Song, B. Ji, Y. Huang, M. Xiao, and L. Yang, "Performance analysis of antenna selection in two-way relay networks", *IEEE Trans. on Sig. Process.*, vol. 62, no. 10, pp. 2520–2532, 2015.

[16] J. Yang, P. Fan, T. Q. Duong, and X. Lei" "Exact performance of two-way af relaying in Nakagami-*m* fading environment", *IEEE Trans. on Wirel. Commun.*, vol. 10, no. 3, pp. 980–987, 2011.

[17] S. Sanayei and A. Nosratinia, "Antenna selection in MIMO systems", *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 68–73, 2004.

[18] S. W. Peters and R. W. Heath Jr, "Nonregenerative MIMO relaying with optimal transmit antenna selection", *IEEE Sig. Process. Lett.*, vol. 15, pp. 421–424, 2008.

[19] H. A. Suraweera, P. J. Smith, A. Nallanathan, and J. S. Thompson, "Amplify-and-forward relaying with optimal and suboptimal transmit antenna selection", *IEEE Trans. on Wirel. Commun.*, vol. 10, no. 6, pp. 1874–1885, 2011.

[20] G. Zhang, W. Zhan, and J. Qin, "Transmit antenna selection in the alamouti-coded MIMO relay systems", *Wirel. Personal Commun.*, vol. 62, no. 4, pp. 879–891, 2012.

[21] M. Ding, S. Liu, H. Luo, and W. Chen, "MMSE based greedy antenna selection scheme for AF MIMO relay systems", *IEEE Sig. Process. Lett.*, vol. 17, no. 5, pp. 433–436, 2010.

[22] K. Yang, N. Yang, C. Xing, and J. Wu, "Relay antenna selection in MIMO two-way relay networks over Nakagami-*m* fading channels", *IEEE Trans. on Vehicular Technol.*, vol. 63, no. 5, pp. 2349–2362, 2014.

[23] S. Chen, W. Wang, X. Zhang, and D. Zhao, "Performance of amplify-and-forward MIMO relay channels with transmit antenna selection and maximal-ratio combining", in *IEEE Wirel. Commun. & Netw. Conf. WCNC 2009*, Budapest, Hungary, 2009, pp. 808–813.

[24] A. Yılmaz and O. Kucur, "Performance of transmit antenna selection and maximal-ratio combining in dual hop amplify-and-forward relay network over Nakagami-*m* fading channels", *Wirel. Personal Commun.*, vol. 67, no. 3, pp. 485–503, 2012.

[25] H. A. Suraweera, G. K. Karagiannidis, Y. Li, H. K. Garg, A. Nallanathan, and B. Vucetic, "Amplify-and-forward relay transmission with end-to-end antenna selection", in *IEEE Wirel. Commun. & Netw. Conf. WCNC 2010*, Sydney, Australia, 2010, pp. 1–6.

[26] R. H. Louie, Y. Li, and B. Vucetic, "Performance analysis of beamforming in two hop amplify and forward relay networks", in *Proc. IEEE Int. Conf. on Commun. ICC 2008*, Beijing, China, 2008, pp. 4311–4315.

[27] H. A. Suraweera, I. Krikidis, and C. Yuen, "Antenna selection in the full-duplex multi-antenna relay channel", in *Proc. IEEE Int. Conf. on Commun. ICC 2013*, Budapest, Hungary, 2013, pp. 4823–4828.

[28] X. Jin, J.-S. No, and D.-J. Shin, "Source transmit antenna selection for MIMO decode-and-forward relay networks", *IEEE Trans. on Sig. Process.*, vol. 61, no. 7, pp. 1657–1662, 2013.

[29] E. Li, S. Yang, and H. Wu, "A source-relay selection scheme in two-way amplify-and-forward relaying networks", *IEEE Commun. Lett.*, vol. 16, no. 10, pp. 1564–1567, 2012.

[30] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. *Academic Press*, 2014.

R. Rajesh received his B.Eng. degree in Electronics and Communication Engineering from Mohammad Sathak College of Engineering, Kilakarai, India in 2002, and the M.Tech. degree in Communication Engineering from VIT University, Vellore, India in 2005. He is now an Assistant Professor in the School of Electronics Engineering, VIT University, Vellore. He is currently pursuing Ph.D. in wireless communications. His research interests includes signal processing in full-duplex relay assisted communication.

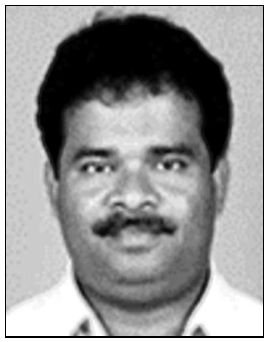
E-mail: rajesh@vit.ac.in
School of Electronics Engineering
VIT University
Vellore, India

P. G. S. Velmurugan received his B.Eng. degree in Electronics and Communication Engineering (ECE) from Tamilnadu College of Engineering, Coimbatore, India in 2002, the M.Eng. degree in Applied Electronics from Kumaraguru College of Technology, Coimbatore in 2007, and the Ph.D. degree from Anna University, Chennai, India in 2015. He is now an Assistant Professor in the ECE Department, Thiagarajar College of Engineering, Madurai, India and pursues a Ph.D. degree at the Information and Communication Engineering department. His research interests include signal processing in cognitive radio networks.

E-mail: pgsvels@tce.edu
Thiagarajar College of Engineering
Madurai, India

S. J. Thiruvengadam received his B.Eng. degree in Electronics and Communication Engineering from Thiagarajar College of Engineering, Madurai, India in 1991, the M.Eng. degree in Applied Electronics from Anna University, Chennai, India, in 1994, and the Ph.D. degree from Madurai Kamaraj University, Madurai, in 2005. In 2008, he was a visiting Associate Professor with the Department of Electrical Engineering, Stanford University, under a postdoctoral fellowship, by the Department of Science and Technology, Government of India. He is currently a Professor with the Department of Electronics and Communication Engineering, Thiagarajar College of Engineering. His areas of research interest include statistical signal processing and multi-input-multi-output wireless communication.

E-mail: sjtece@tce.edu
Thiagarajar College of Engineering
Madurai, India



P. S. Mallick received his M.Sc. degree from the University of Chittagong, Bangladesh, and his Ph.D. from Jadavpur University, India. He worked for 4.5 years in a Sweden-based electronics industry company named *IAAB Electronics*, as a Technical Head. His current area of research interest includes nanoscale CMOS, nanoelectronics and VLSI engineering. He has published 38 research papers in different journals and conferences of international repute and authored a book on Matlab and Simulink. At present he is working for the School of Electrical Engineering, VIT University, Tamilnadu, India, holding the position of a Professor. In 1998, he received the prestigious Jawaharlal Nehru Scholarship for his doctoral research work. He has been one of the enlisted technical innovators of India since 2007.

E-mail: psmallick@vit.ac.in
School of Electrical Engineering
VIT University
Vellore, India

Paper

Miniaturized Spectacles Shaped Tapered Slotted Patch Antenna for UWB Applications

M. Tarikul Islam, M. Samsuzzaman, M. Z. Mahmud, and M. T. Islam

Centre of Advanced Electronic and Communication Engineering, The National University of Malaysia, Malaysia

https://doi.org/10.26636/jtit.2018.115717

Abstract—A compact planner patch ultra-wideband (UWB) antenna is presented in this paper. The antenna configuration consists of a spectacles-shaped patch and a slotted ground plane. Different parameters are investigated for improving the antenna's properties and for achieving the preferred UWB band (3.1–10.6 GHz). The experimental and simulated results demonstrate that the proposed antenna acquires an operating bandwidth of 117% (3–11.5 GHz) with a stable omnidirectional radiation pattern, about 89% of average radiation efficiency and 4.2 dBi of average gain with the maximum of 5.7 dBi at 10.2 GHz.

Keywords— patch antenna, tapered ground, UWB, wireless communication

1. Introduction

The main problem with most antennas is that the size and the bandwidth of an antenna are contradictory features. The wider the bandwidth, the bigger the antenna. Therefore, a lot of research has been conducted to overcome this problem. Recently, adding a finite metal plane [1], inserting an additional stub to one side of the circular patch [2], adding steps to the lower edge of the patch, adding a slot to one side of the radiating element [3] have been described as methods used for increasing the operating bandwidth of elliptically and circularly planar monopole antennas.

The properties of the ultra-wideband (UWB) antennas can be improved also by changing the patch shape (rectangular, elliptical, circular, heart-shaped). Wider bandwidth can be also achieved by reducing dimensions of the ground plane [4]–[7]. Over the past decade, numerous antennas have been proposed for UWB applications. Unfortunately, a large ground plane increases bandwidth, but also increases the dimensions. Hence, the antenna is no longer compact and is hard to integrate with microwave technology [8]. To overcome this problem for UWB applications, various line-feeding and waveguide-feeding antennas are investigated.

Ray *et al*. reported a small elliptical ring antenna for UWB applications, where the antenna having a bandwidth of 4.6–10.3 GHz does not meet the requirements of UWB systems [2]. Despite its compact size, the antenna cannot offer a full range of UWB features [9], but its dimensions are smaller than those of the antennas reported in [10], [11].

In this paper, an antenna with a spectacles-shaped radiator and a tapered slot ground plane, with the operating bandwidth of 3–11.5 GHz and meeting all UWB criteria is presented. Its dimensions are reduced, but performance is increased due to the modified design structures. Combination of the spectacles-shaped patch and the tapered slot ground plane make the design more efficient and suitable for use in UWB applications.

2. Design

Figure 1 shows the proposed antenna layout to be fabricated with the use of typical FR4 1.6 mm thick PCB laminate. The front side, having the form of a spectacles-shaped radiator, is shown in Fig. 1a. Figure 1b shows the ground plane,

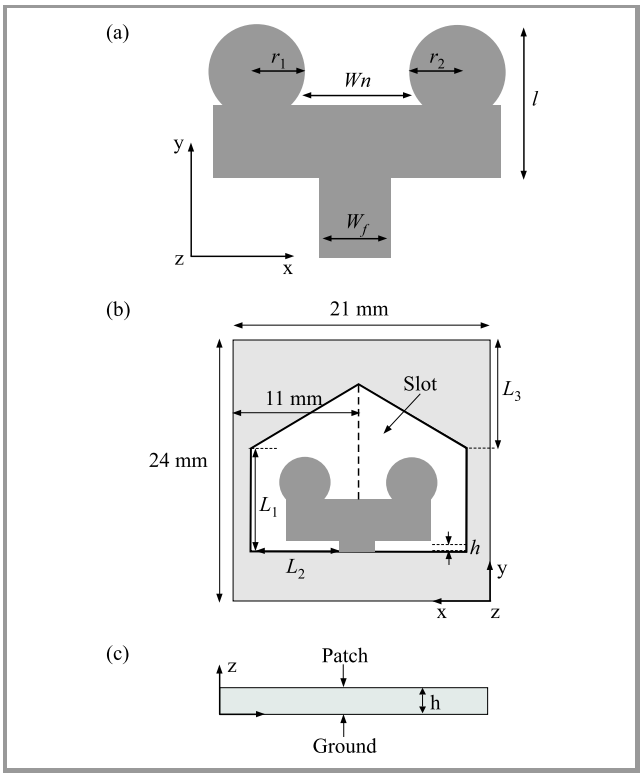


Fig. 1. Antenna geometry: (a) top view, (b) bottom view, (c) side view.

which is calculated on the other side of FR4. Two circular slots r_1 and r_2 form the spectacles-shaped modified patch, ensuring better performance. A 50 Ω microstrip transmission line is connected to the bottom of the feed line. The overall dimensions of the design are (W × L) 21 × 24 mm. The specific parameters of the modified ground plane and patch are presented in Table 1.

Table 1
Parameter of the presented design

Parameter	Value [mm]	Parameter	Value [mm]
W	21	w_1	6.5
L	24	L_1	8.66
l	7.30	L_2	7.52
W_n	5.56	L_3	10.12
W_f	3	h	1.6
r_1, r_2	2.5	h_1	5.564

Different shapes tested in order to achieve a wide UWB bandwidth are shown in Fig. 2a and simulated results of S_{11} are presented in Fig. 2b. It is observed that the tapered design offers a wider bandwidth compared to circle, elliptical or square slot ground planes. For the circle slot ground

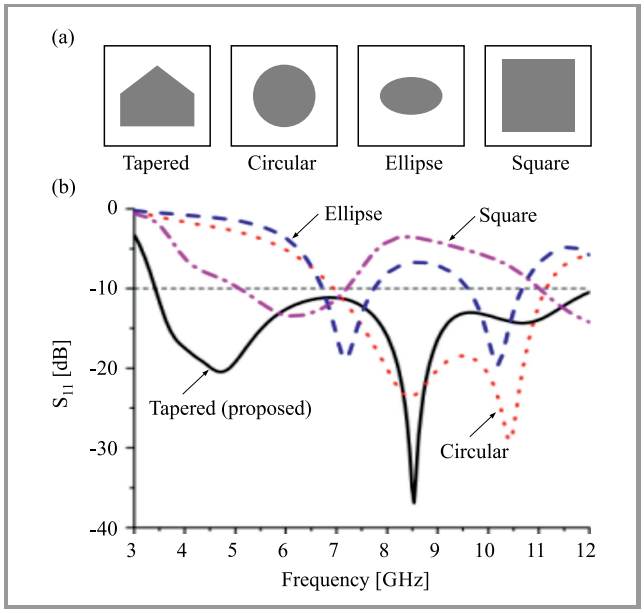


Fig. 2. Effect of ground plane shape on S_{11}: (a) different ground plane shapes and (b) S_{11} parameter.

plane, the operating bandwidth achieved is 7.3–11 GHz and fails to meet UWB requirement. For elliptical and square slot ground planes, the resonant frequency is not satisfactory to meet the desired specification. Finally, after applying the tapered slot ground plane, the operating bandwidth attained (< 10 dB) meets the requirements set. Figure 3 demonstrates the effect of patch shape on bandwidth.

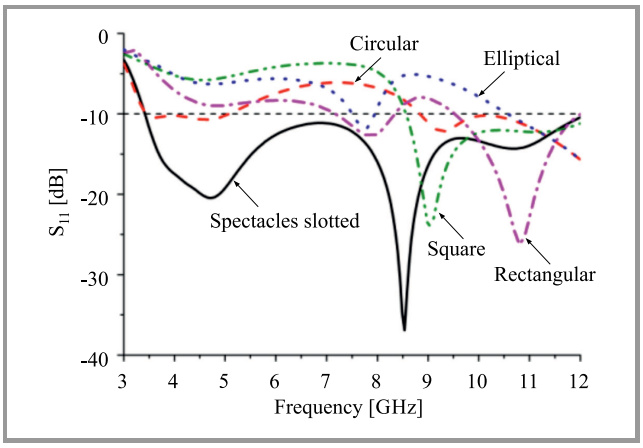


Fig. 3. Effects of patch shape on S_{11} reflection coefficient.

The simulated results show that only the spectacles-shaped patch proposed attains a full bandwidth of 3–11.5 GHz, covering the entire UWB spectrum. The lower frequency bandwidth is meaningfully affected by the modified patch shape.

3. Results

3.1. Frequency-Domain Performance

A prototype of the proposed antenna is presented in Fig. 4. The optimization measurements were performed with 3D HF Ansoft's HFSS EM software simulator, based on the finite element method and the CST Studio solver. The results

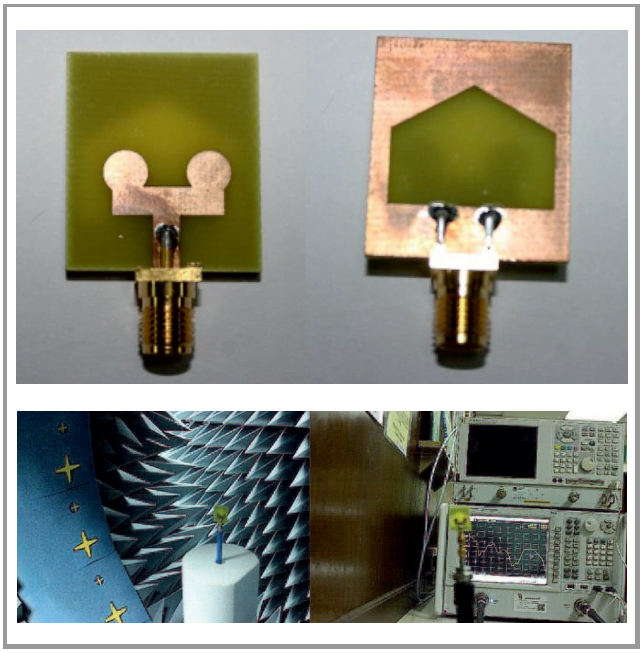


Fig. 4. Prototype of the proposed antenna (top and bottom view) difference and measurment setup.

measured were attained from the Agilent E8362C vector network analyzer in a Satimo near field anechoic chamber (UKM StarLab).

The reflection coefficient S_{11}, measured and simulated vs. frequency is shown in Fig. 5. It is observed that the operating bandwidth ranges from 3 to 11.5 GHz. There is a slight distortion between the measured and simulated results, because of faulty soldering and the coaxial cable that was used for the measurements. The design covers the full UWB band (3.1–10.6 GHz).

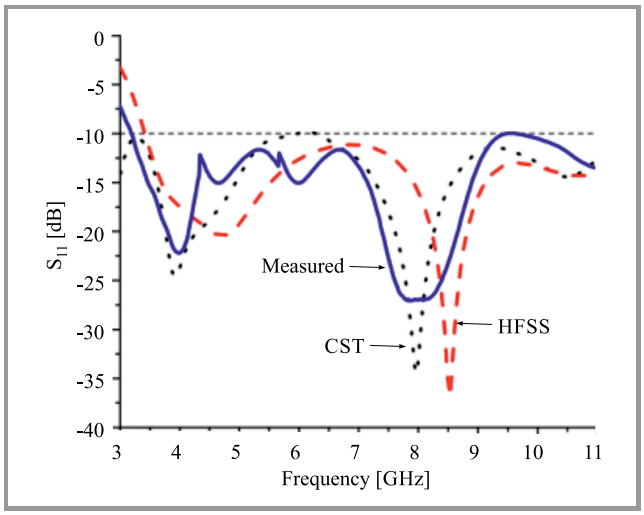


Fig. 5. Return loss vs. frequency.

The simulated and experimental gain across the operating bandwidth is presented in Fig. 6. The maximum gain of 5.7 dBi is recorded at 10.2 GHz, and the average gain across UWB equals 4.2 dBi. The radiation efficiency measured and simulated for the antenna is presented in Fig. 7. Radiation varies between lower and higher bands from 81% to 9%. Efficiency and gain are both affected by the use of low-cost FR4 PCB laminate as the substrate. Gain and efficiency can be improved by using microwave substrate materials.

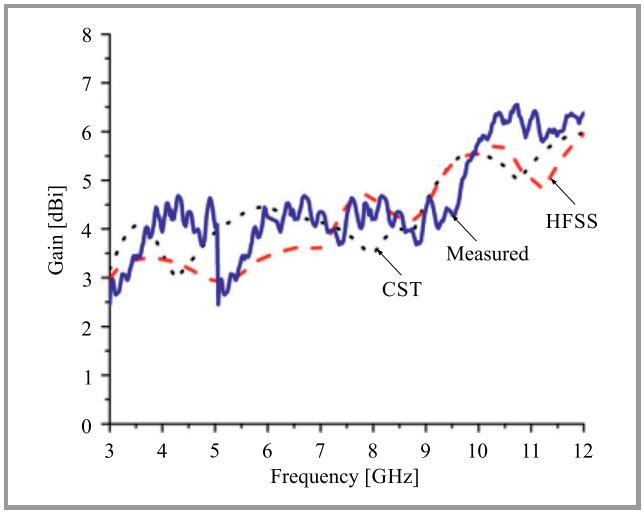


Fig. 6. Measured and simulated gain.

The surface current distribution at 4.5 GHz and 8.5 GHz is shown in Figs. 8 and 9, respectively. The antenna's

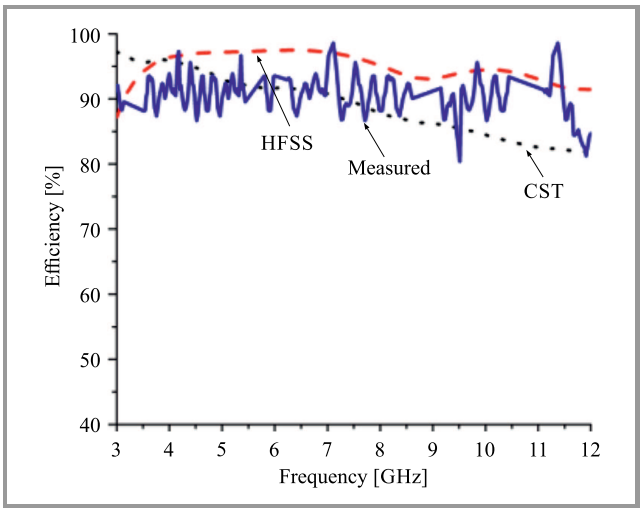


Fig. 7. Measured and simulated efficency.

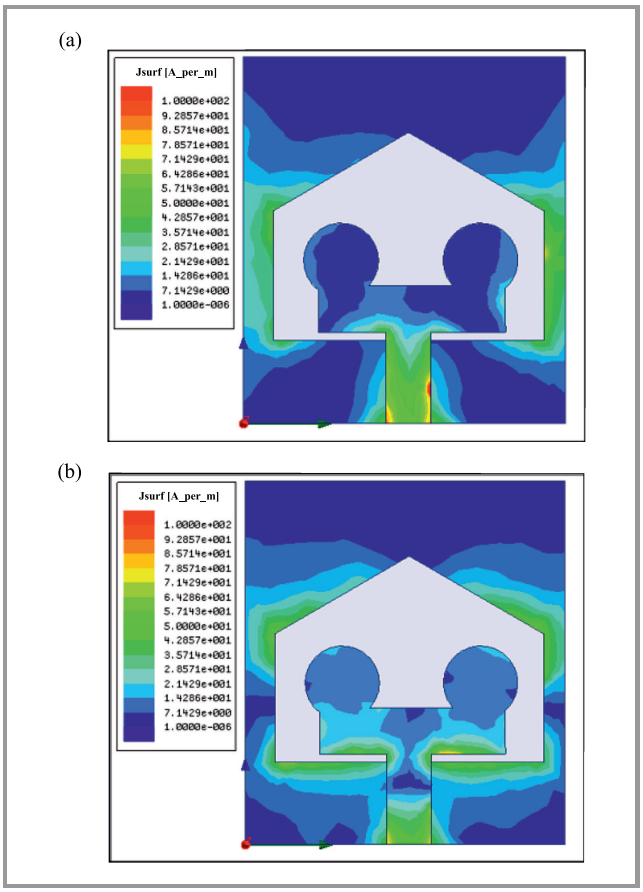


Fig. 8. Surface current distribution at: (a) 4.5 GHz and (b) 8.5 GHz. (See color pictures online at www.nit.eu/publications/journal-jtit)

measured and simulated radiation pattern at 3.5 GHz, 6.5 GHz and 9.5 GHz with both cross-polarization and co-polarization of two major planes, is shown in Fig. 9. The surface current is evenly distributed over the lower frequency of 4.5 GHz. At higher frequencies, the antenna shows the development of higher order current modes and the density of current is lower on the patch. It is observed that the antenna has, over the UWB band, a stable radiation

pattern that is almost omnidirectional at lower frequencies. Unwanted cross-polarization occurs with the increase of frequency when changing current distribution. As a result, the radiation pattern is slightly more directional at higher frequencies. Several null points are also observed in current distribution at higher frequencies. With higher order modes, the radiating element is excited and causes the directional radiation pattern.

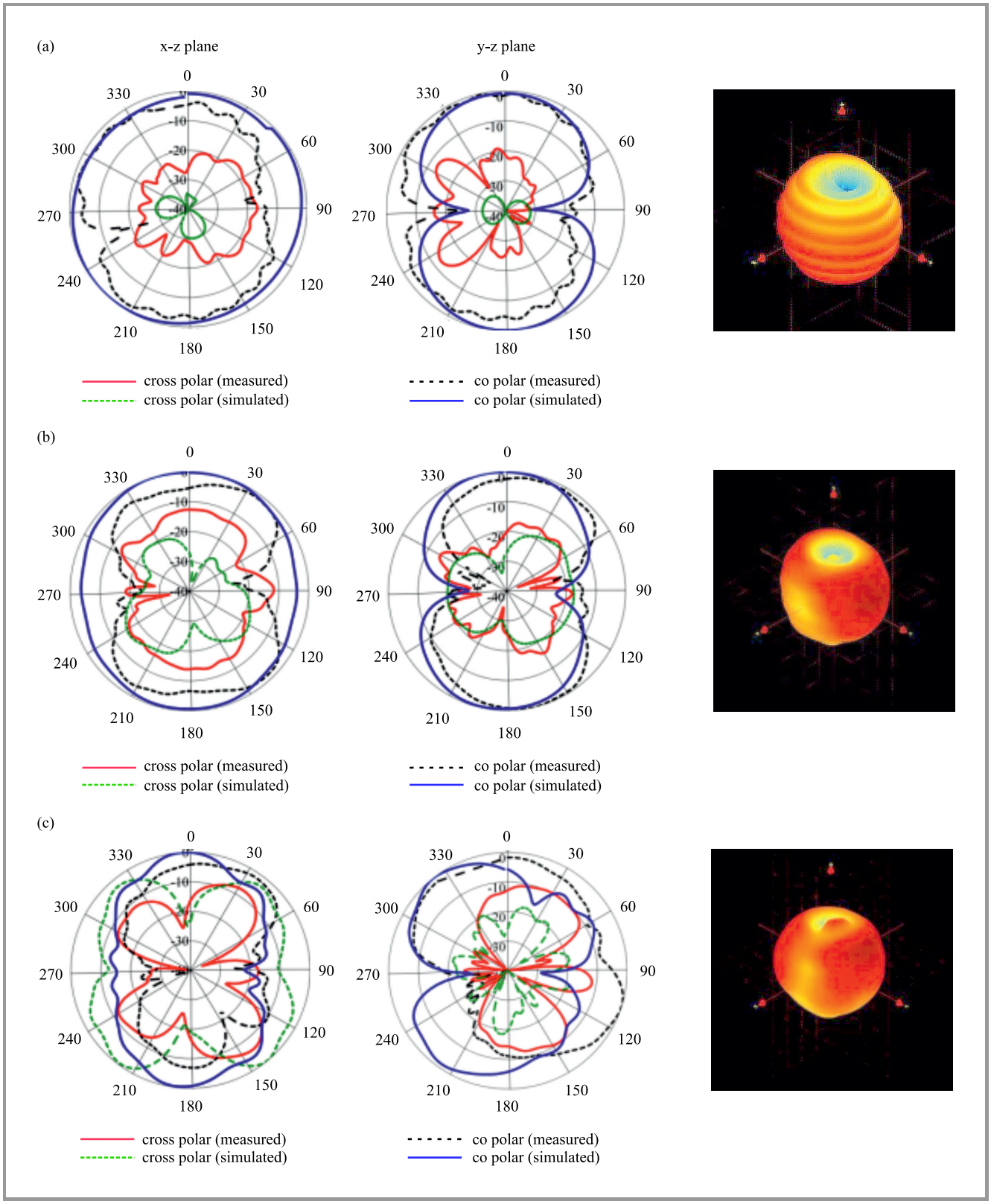


Fig. 9. 2D and 3D radiation pattern at: (a) 3.5 GHz, (b) 6.5 GHz, and (c) 9.5 GHz.

3.2. Time Domain Performance

It is clear, from frequency domain observations, that the proposed antenna demonstrates good frequency domain performance. However, decent frequency domain characteristics cannot guarantee the antenna's equal behavior in the time domain. Hence, in order to validate the design, its time-domain behavior needs to be examined, including the transmission coefficient, input-output pulse waveform, and group delay investigation. Figure 10 shows the transmission coefficient $|S_{21}|$ with two matching proposed antennas used, placed in front of each other and in a side-by-side configuration, at a distance of 300 mm, considering far field environments across the whole UWB range. Figure 10 shows flat magnitude of the transmission coefficient line over the operating band. A slight decrease is observed in the side-by-side scenario at 10.1 GHz, with the magnitude remaining flat over the rest of the bandwidth, which indicates a stable UWB transmission capability in both face-to-face and side-by-side configurations.

The group delay is defined as the first derivative of the far field phase of the transmission response with respect to radial frequency ω [12]. Figure 11 presents the measured group delay in the face-to-face and side-by-side scenarios. One can see two sharp variations of the group delay at 4.5 and 8.5 GHz, which specifies a slight non-conformity with the linear phase response. Group delay remains almost constant at other frequencies outside the affected bands that show good phase linearity. The input and received signals in the face-to-face and side-by-side scenarios (with the distance of 300 mm in the case of the latter) are shown in Fig. 12. The received signals in both orientations have parallel waveforms showing a small disparity. The highest value of cross-correlation between the transmitting and receiving pulse estimates the signal distortion, known as fidelity factor (FF). Typically, the pulse becomes almost unrecognizable if the alteration is higher than 50% (FF < 0.5). For face-to-face and side-by-side setups, FF equals 83% and 90%, which indicates that the proposed system offers a lower distortion of the signal while transmitting UWB impulse signals, simultaneously displaying a low-variation transmission coefficient, a constant group delay and a decent FF, demonstrating good phase linearity.

A comparison of the proposed antenna and the ones described in literature is presented in Table 2. The parameters

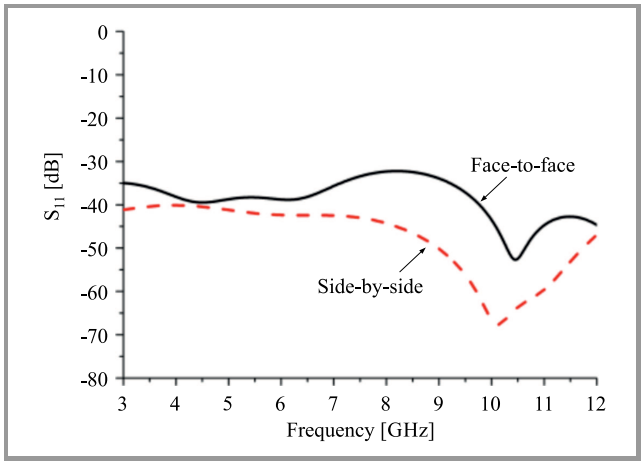


Fig. 10. Transmission coefficient in face-to-face and side-by-side scenarios.

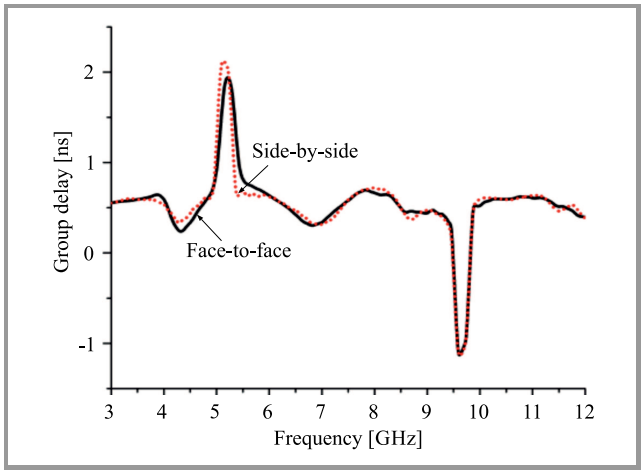


Fig. 11. Group delay.

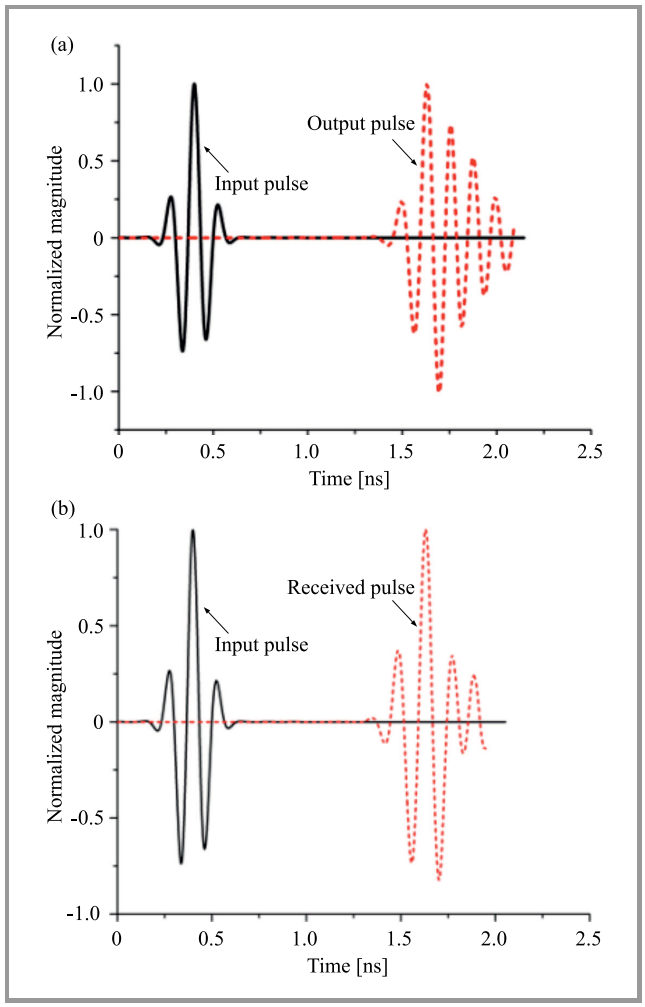


Fig. 12. Input and received pulse waveforms in: (a) face-to-face and (b) side-by-side scenario.

considered include bandwidth (BW), fractional bandwidth (FBW), gain, and applications.

Table 2
Comparison of bandwidth, fractional bandwidth, gain, and applications

Reference	BW ($S_{11} <$ 10 dB) [GHz]	FBW [%]	Gain [dBi]	Applications
[8]	3.4–9.9	97	4.8	Near field imaging
[5]	4–14	111	2.32–4.4	UWB
[6]	3.7–18	132	3.97	UWB
[7]	3.1–15.2	132	2.5	UWB
[9]	8.39–9.7	84	4.25	UWB
Proposed design	3–11.5	117	5.7	UWB

4. Conclusions

The spectacles-shaped patch antenna is a miniature 21×24 mm design with an operating bandwidth of 117% (3–11.5 GHz) and a return loss of < 10 dB. The antenna shows an average peak gain of 4.2 dBi across the UWB band, with high efficiency levels of over 81% and a stable omnidirectional radiation pattern. Both the frequency and time domain characteristics of the proposed design are properly analyzed. The antenna is very simple in design, low-cost and highly efficient, which makes it suitable for integration with portable devices.

References

[1] H. Nazli, E. Bicak, B. Turetken, and M. Sezgin, "An improved design of planar elliptical dipole antenna for UWB applications", *IEEE Antennas and Wireless Propag. Let.*, vol. 9, pp. 264–267, 2010.

[2] K. Ray and Y. Ranga,"Ultrawideband printed elliptical monopole antennas", *IEEE Transact. on Antennas and Propag.*, vol. 55, no. 4, pp. 1189–1192, 2007.

[3] K.-H. Kim and S.-O. Park, "Analysis of the small band-rejected antenna with the parasitic strip for UW", *IEEE Transact. on Antennas and Propag.*, vol. 54, no. 6, pp. 1688–1692, 2006.

[4] M. Sharma and V. Shrivastava, "Printed fractal elliptical monopole antenna for UWB application", in *Proc. 2008 Int. Conf. of Recent Advances in Microwave Theory and Appl. Microwave-08*, Jaipur, Rajasthan, 2008, pp. 374–376.

[5] A. A. Shaalan and M. Ramadan,"Design of a compact hexagonal monopole antenna for ultra-wideband applications", *J. of Infrared, Millimeter, and Terahertz Waves*, vol. 31, no. 8, pp. 958–968, 2010.

[6] L. Liu, S. Cheung, R. Azim, and M. T. Islam, "A compact circular-ring antenna for ultra-wideband applications", *Microwave and Optic. Technol. Let.*, vol. 53, no. 10, pp. 2283–2288, 2011.

[7] Y. B. Yang, F. S. Zhang, F. Zhang, L. Zhang, and Y. C. Jiao, "A novel compact CPW-fed planar monopole antenna with modified stair-style ground for ultra-wideband application", *Microwave and Optic. Technol. Let.*, vol. 52, no. 9, pp. 2100–2104, 2010.

[8] M. N. Shakib, M. T. Islam, and N. Misran, "Stacked patch antenna with folded patch feed for ultra-wideband application", *IET Microwaves, Antennas & Propagation*, vol. 4, no. 10, pp. 1456–1461, 2010.

[9] M. Islam, M. T. Islam, and M. R. I. Faruque, "Design of an UWB patch antenna for dual frequency operations", *Research J. of Applied Sc., Engineer. and Technol.*, vol. 7, no. 4, pp. 822–825, 2014.

[10] M. Hossain, M. R. I. Faruque, and M. T. Islam, "Design of a patch antenna for ultra wide band applications", *Microwave and Optic. Technol. Let.*, vol. 58, no. 9, pp. 2152–2156, 2016.

[11] K. Kikuta and A. Hirose,"Compact folded-fin tapered slot antenna for UWB applications", *IEEE Antennas and Wireless Propag. Let.*, vol. 14, pp. 1192–1195, 2015.

[12] Y. J. Cho, K. H. Kim, D. H. Choi, S. S. Lee, and S.-O. Park, "A miniature UWB planar monopole antenna with 5-GHz band-rejection filter and the time-domain characteristics", *IEEE Transact. on Antennas and Propag.*, vol. 54, no. 5, pp. 1453–1460, 2006.

[13] G. Quintero, J.-F. Zurcher, and A. K. Skrivervik, "System fidelity factor: A new method for comparing UWB antennas", *IEEE Transact. on Antennas and Propag.*, vol. 59, no. 7, pp. 2502–2512, 2011.

M. Tarikul Islam received his B.Sc. in Computer Science and Engineering from Patuakhali Science and Technology University (PSTU) in 2016. Currently he is working as an M.Sc. student at the National University of Malaysia (UKM), Malaysia. He has authored or co-authored a number of referred journals and conference papers. He is currently a Graduate Research Assistant at the Department of Electrical, Electronic and Systems Engineering, UKM, Malaysia. His research interests include communication antenna design, wireless communication, RF engineering and microwave imaging.
E-mail: p94299@siswa.ukm.edu.my
Centre of Advanced Electronic and Communication Engineering
The National University of Malaysia
Malaysia

M. Samsuzzaman received his B.Sc. and M.Sc. degrees in Computer Science and Engineering from the Islamic University Kushtia, Bangladesh in 2005 and 2007, respectively, and the Ph.D. degree from the National University of Malaysia, Malaysia in 2015. Between 2008 and 2011, he worked as a Lecturer at the Patuakhali Science and Technology University (PSTU), Bangladesh. From February 2011 to 2015, he worked as an Assistant Professor at the same university. He is also an Associate Professor at the same university and is currently working as a post-doctoral fellow at the National University of Malaysia, Malaysia. He has authored or co-authored

70 referred journals and conference papers. His research interests include communication antenna design, satellite antennas and satellite communication.
E-mail: samsuzzaman@siswa.ukm.edu.my
Centre of Advanced Electronic and Communication Engineering
The National University of Malaysia
Malaysia

M. Zulfiker Mahmud is an Assistant Professor at the AIS Department of Jagannath University Bangladesh. He received his B.Sc. and M.Sc. degrees in Computer Science and Engineering from the Islamic University Kushtia, Bangladesh. Currently, he works as a Ph.D. student at the National University of Malaysia (UKM), Malaysia. He is currently a Graduate Research Assistant at the Department of Electrical, Electronic and Systems Engineering, UKM, Malaysia. He has authored or co-authored 25 referred journals and conference papers. His research interests include microwave imaging, antenna design, satellite antennas, satellite communication, and wireless communication.
E-mail: zulfikerm@siswa.ukm.edu.my
Centre of Advanced Electronic and Communication Engineering
The National University of Malaysia
Malaysia

M. Tariqul Islam is a Professor at the Department of Electrical, Electronic and Systems Engineering of the National University of Malaysia (UKM) and a visiting Professor of Kyushu Institute of Technology, Japan. He has authored about 350 research journal articles, nearly 165 conference articles, 4 research level books and a few book chapters on various topics related to antennas, microwaves and electromagnetic radiation analysis with 13 inventory patents filed. He is a Senior Member of IEEE, a Chartered Professional Engineer (CEng), a Member of IET (UK) and a member of IEICE (Japan).
E-mail: tariqul@ukm.edu.my
Centre of Advanced Electronic and Communication Engineering
The National University of Malaysia
Malaysia

Paper

Protocols for Wireless Sensor Networks: A Survey

Aarti Kochhar[1,2], Pardeep Kaur[1], Preeti Singh[1], and Sukesha Sharma[1]

[1] *University Institute of Engineering and Technology, Panjab University, Chandigarh, India*
[2] *Lovely Professional University, Phagwara, India*

https://doi.org/10.26636/jtit.2018.117417

Abstract—This paper presents a survey on the MAC and network layer of Wireless Sensor Networks. Performance requirements of the MAC layer are explored. MAC layer protocols for battery-powered networks and energy harvesting-based networks are discussed and compared. A detailed discussion on design constraints and classification of routing protocols is presented. Several routing protocols are compared in terms of such parameters as: energy consumption, scalability, network lifetime and mobility. Problems that require future research are presented. The cross-layer approach for WSNs is also surveyed.

Keywords—cross layer, Medium Access Control, protocols, Wireless Sensor Networks.

1. Introduction

Any Wireless Sensor Network (WSN) application requires the physical environment to be sensed for data transmitted over a channel to a base station. Power is required in order to sense data and send it to the base station. It can be obtained from a battery or may be harvested from a natural source. One of the basic architectures of a sensor node is shown in Fig. 1 [1]. It comprises 4 units responsible for power, processing and communications. Most energy is consumed by processing and communications.

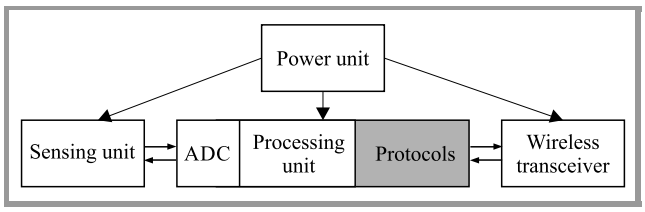


Fig. 1. Architecture of a typical sensor node.

In order to overcome energy, lifetime, traffic and mobility constraints, the communication protocol stack needs to be carefully designed.

The basic structure of a WSN protocol stack is discussed in Section 2. Transport and upper layers add reliability to the transmission of data only, which is not a key concern in the majority of WSN applications. Hence, only the data link layer (DLL) and the network layer are discussed in this paper. Energy consumption sources, classification, design constraints, respective protocols and outstanding research problems are discussed for the MAC and network layers. In Section 3, a sub-layer of DLL – Medium Access Control (MAC) layer – is surveyed. In Section 4, the network layer is examined. A comparison of both MAC and routing protocols has been tabulated in the respective sections. In Section 5, the cross layer approach, a technological advancement enhancing efficiency, is discussed.

2. Protocol Stack in WSNs

Proper design of the protocol stack is important for the overall efficiency of a WSN. WSN differs from conventional computer communication networks in the following ways:

- Contrary to computer network's well planned physical topology, the nodes are densely and randomly deployed in WSNs.
- Once designed, computer networks remain static, whereas WSNs are dynamic in nature. Failure of one node can change the entire topology. So, WSNs need to be self-configurable.
- Computer networks have IP addresses for their global identification. WSN nodes have no global identification because it creates a large overhead.
- Computer networks have a continuous supply of energy, whereas WSNs have limited resources. So, the WSN protocol stack needs to be energy-aware.

Protocol stacks in WSNs comprise five horizontal and five vertical levels. They have five layers and five management planes, as shown in Fig. 2 [2].

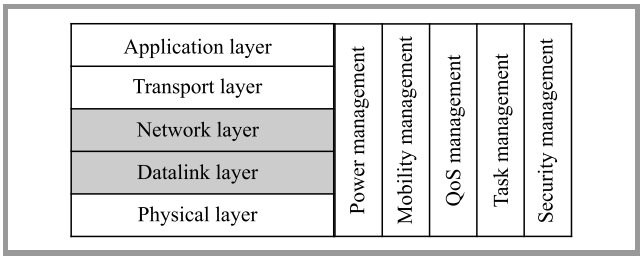


Fig. 2. Protocol stack of WSNs.

3. Data Link Layer

DLL has two sub-layers: MAC and Logical Link Control (LLC). LLC is used for link management, flow and error control. MAC is responsible for assembling data into frames and for disassembling frames to retrieve information. Nodes may be sharing a single channel for sending data over to the sink or to another node. Simultaneous transmission of data on a single channel will lead to a collision, causing loss of data and energy. To avoid this, nodes should agree on a time slot at which a particular node would be sending. To agree on timeslots, nodes need to communicate, which requires a channel too. Considering the propagation delay, it is difficult for a node to know the instantaneous status of another node. The transceiver also consumes a large amount of energy while accessing the media. MAC controls activity of the transceiver to conserve energy [3].

3.1. Energy Consumption Sources

There are a few energy consumption sources at the MAC layer [4]:

Collision – when two or more nodes try to send information on a single channel at the same time, the packets collide. Collided packets need to be discarded and retransmitted.

Overhearing – when a node receives a packet destined for another node, it consumes unnecessary energy.

Overhead – sending and receiving control information also requires energy, causing an additional overhead.

Idle listening – idle listening is listening to an idle channel on which traffic is expected.

Over-emitting – sending information to a node which is not ready to receive. Hence, packets are discarded and need to be retransmitted.

3.2. Performance Requirements for the MAC Layer

While designing MAC layer protocols, one needs to consider the following requirements [5]:

Throughput: Protocol efficiency is measured by its throughput. In the case of a wireless link, it may be related to capacity.

Scalability: Scalability refers to the protocol's adaptation to an increase in network size, traffic, overhead and load. One way to deal with this is to localize the interactions so that nodes need less global knowledge to operate.

Latency: Latency can be referred as the time delay between message transmission and message arrival. Latency is an important constraint for time-critical applications, and needs to be minimized.

Number of hops: It is the number of hops taken by packets to reach the sink. Operation of the MAC protocol varies between single-hop and multi-hop scenarios. In the case of multiple hops taken to reach the sink, data needs to be aggregated before sending it to the sink.

3.3. Classification of MAC Protocols

MAC protocols can be categorized into two types [6]:

- **schedule-based MAC** protocol in which nodes agree upon a fixed schedule to access the channel. So, each node has a fixed slot for communication. Outside their slots, nodes move into sleep mode, avoiding collision and overhearing. The lifetime of nodes is enhanced, as they do not communicate over the complete duty cycle;
- **random access-based protocol** in which nodes need to compete to reserve access to a channel. After collision, each node waits for a random time before accessing the channel again. Energy efficiency of random access-based protocol is low.

3.4. MAC Layer Protocols

A protocol for an application can be chosen based on performance and specific requirements. In the battery-powered area Sensor-MAC, a T-MAC is presented. Then, MAC protocols based on energy harvesting are presented (Fig. 3).

Sensor-MAC (S-MAC): In general, nodes are synchronized locally, to operate a periodic sleep-and-listen schedule. Each node belongs to a virtual cluster and each cluster has a common listen-and-sleep schedule, as shown in Fig. 4. This represents the basic idea of S-MAC [7]. Each node discovers its neighbors regularly and establishes a link with them. Then, it assigns a distinct frequency, time or

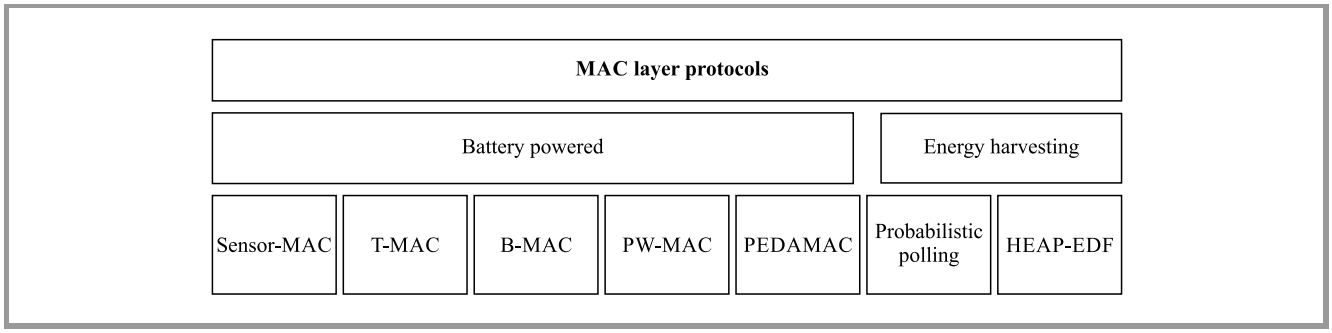


Fig. 3. MAC layer protocols.

code to each link. Long messages are divided before sending. Such a solution offer various advantages, as it self-organizes the network to a variation in topology. This change in topology can be the consequence of deaths or movement of a node. It also operates a lower duty cycle, so the consumption of power used for overhearing and idle listening is reduced. Network latency increases as nodes alternate between active and sleep mode. It can be avoided altogether if a node wakes up after sensing the wake-up of its neighbor. Since sleep-and-listen periods are predefined, efficiency of the protocol may decrease under variable traffic, as traffic may be forwarded to a sleeping node.

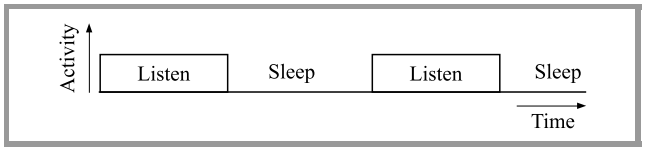


Fig. 4. Sleep-and-listen periods.

Timeout MAC (T-MAC): S-MAC has fixed listen and sleep periods, but applications with variable loads need dynamic listen and sleep periods. In T-MAC, the listen period ends when no event, such as receipt of data or sensing of activity has taken place for a threshold period (TP), as shown in Fig. 5 [8]. The listen period depends on current load. Transmission is based on Request-To-Send (RTS), Clear-To-Send (CTS) and acknowledgment (ACK) packets. Nodes close to the sink may have more data to send, so their listen periods are longer. The advantages are: RTS, CTS and ACK packets reduce collision rates and increase reliability. If listen periods are fixed, then nodes with less data will waste energy by idle listening. Energy consumption and idle listening are reduced as data can be sent in variable bursts. T-MAC has low sensitivity to latency, but it has a few drawbacks, such as it cannot support high data rate applications. Also, it has to trade-off throughput to maintain low energy consumption.

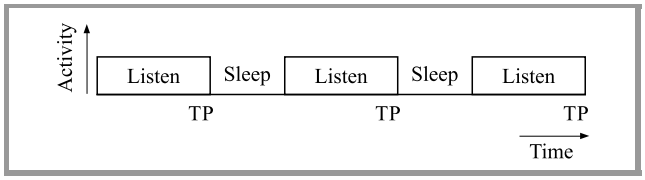


Fig. 5. Adaptive listen and sleep periods.

Berkeley MAC (B-MAC): B-MAC uses preamble sampling [9], [10]. Each time a node wakes-up, it checks for any activity before sending. The node is also waiting only for a certain period of time to receive data. After time-out, the node returns to sleep mode. B-MAC uses clear channel assignment, and makes local policy decisions to optimize network performance. Owing to preamble sampling, the duty cycle is reduced, which increases efficiency and throughput. Energy consumption is lower because of low-power listening. Also, it supports reconfiguration to improve latency. B-MAC has a few drawbacks, such as it has no ability to handle multi-packet environments and suffers form a hidden terminal problem. Also, overhead of the protocol increases. The protocol can be enhanced further using adaptive preamble sampling.

Predictive Wake-up MAC (PW-MAC): In PW-MAC, the wake-up schedule of nodes can be randomized [11], [12]. To inform the intended transmitters, the node will send a signal upon waking up. A sender can predict the receiver's wake-up time and can wake-up simultaneously to save energy. To address timing challenges, PW-MAC has an on-demand prediction-based error correction mechanism. PW-MAC has a reduced duty cycle, as it has a random node wake-up schedule. It has improved performance compared to S-MAC and B- MAC, as collisions can be avoided. Latency is less than 5% of that typical of other MAC protocols. A node needs only 10 bytes of memory to store the prediction state of other nodes. Each node has to send a signal on waking-up, so the overhead of the protocol is increased, although it is low compared to other protocols. Also, hardware can induce errors in predicting wake-up times of the receiver.

Power Efficient and Delay Aware MAC (PEDAMAC): To minimize energy consumption due to overhearing, PEDAMAC transmits data at more than one power level. The access points (also called sinks) coordinate sensor nodes. Access points are assumed to have no power constraints, while sensor nodes have limited power. PEDAMAC assumes that each node can reach the sink in one hop. It has four phases: topology learning, topology collection, scheduling and adjustment. The protocol allows the nodes to operate at different power levels, as per the requirement of the task being processed by the node. It has three power levels: maximum power P_m, medium P_x, and minimum P_s. Synchronization is done at P_m. The sink can broadcast topology-related information at P_x. Data is transmitted at P_s. Low transmission power saves energy and it is used in delay-bound applications, but it has a few drawbacks, such as the fact that protocol assumes a one hop distance to the sink, which may not always be the case. Distinct power levels increase the protocol overhead. Also, data may be dropped before delivered, if transmission power is too low, i.e. the range of radio is decreased because of power limitation. PEDAMAC can be enhanced by increasing the number of media or channels to further reduce the delay [12]–[14].

Energy harvesting is considered as the only energy source by Eu *et al.* [15]. It is not easy to predict the wake-up schedule of nodes powered by energy harvesters. Authors exploited the uncertain nature of energy harvesting sources to increase the performance of MAC protocols. MAC protocols based on battery-powered WSNs have different goals, such as increased lifetime compared to energy harvesting based WSN (EH-WSN). So, there is a need to have protocols designed specifically for EH-WSN.

Probabilistic polling: In probabilistic polling, the sink sets contention probability P_c in each node through a polling

Table 1
Performance evaluation of MAC protocols

Protocol	Throughput	Energy conservation	Maximum % of energy saved vs. S-MAC	Latency	Overhead	Scalability
S-MAC [7]	Low	Low	0	High	Low	High
T-MAC [9]	Low	High	85	N/A	Moderate	Low
B-MAC [9], [10]	High	Moderate	57	Moderate	High	Low
PW-MAC [11], [12]	High	High	80	Low	Moderate	High
PEDAMAC [13], [14]	Moderate	Moderate	38	Low	High	Low
Probabilistic polling [15]	High	N/A	N/A	Depends on energy harvesting rate	Low	High
HEAP-EDF [16]	Moderate	N/A	N/A	Depends on energy harvesting rate	Moderate	Low

packet [15]. Each node generates a random number ν, and when it is less than contention probability ($\nu < P_c$), the node is allowed to send. Otherwise, the node can go to the charging state. The sink keeps on changing contention probability depending on network response. If no sensor responds, the sink increases P_c. Also, when a node leaves the network, P_c is increased. In the case of collision and joining of new node, P_c is decreased by a larger amount. This approach is known as additive-increase and multiplicative-decrease. Contention probability P_c offers maximum throughput when it is equal to the inverse of the number of nodes receiving polling packets:

$$P_{opt} = \frac{1}{N_r}, \tag{1}$$

where P_{opt} is the optimal probability that maximizes throughput. N_r is the number of nodes receiving polling packets ($N_r \geq 1$).

This protocol can adapt to varying energy harvesting rates to ensure high throughput and the sink can also adjust P_c in the case of a collision. Hence, the protocol increases scalability of the network. Since P_c keeps changing due to collisions or when a node joins or leaves a network, it takes quite some time for the network to stabilize. This leads to increased network latency. Also, bandwidth is wasted until the network stabilizes at an appropriate P_c. Another drawback is that the protocol assumes a single hop distance to the sink, limiting protocol scalability.

HEAP-EDF: Power generated by ambient energy harvesting sources (HEAP), may vary, i.e. solar energy has different rates in the morning and in the afternoon. To overcome this, Earliest Deadline First (HEAP-EDF) uses a predict-and-update algorithm to reduce the temporal variations [16]. In HEAP-EDF, the sink polls the node with the minimum or the earliest wake-up time. The sensor will not poll the node whose energy has decreased below the transmission level because of previous polling. At the power-balance ratio of 0.5, HEAP-EDF offers the best fairness. The power-balance ratio is given as:

$$\emptyset = \sum_{n=1}^{N} \frac{T_c}{T_n}. \tag{2}$$

In Eq. (2), T_c is the duration of polling cycle, T_n is energy harvesting delay for n-th node and N is the number of sensor nodes. Simulations in [16] show that channel utilization reduces as the link error probability increases. HEAP-EDF performs worse in the case of large networks. Also, the single-hop approach is assumed, which limits application of the protocol to small networks.

3.5. Comparison of Protocols

Table 1 shows the performance of MAC protocols reviewed. B-MAC has a high throughput owing to preamble sampling, but this increases the overhead too. Since probabilistic polling and HEAP-EDF are based on an ambient energy harvesting source, energy consumption is not a relevant factor to be compared. In this case of HEAP-EDF, overhead can be decreased if energy harvesting rates are correlated. Protocols with high overheads cannot be scaled to a large network due to the increase in the number of control

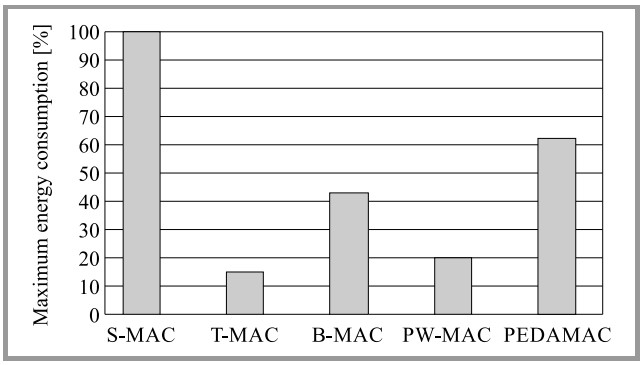


Fig. 6. Maximum energy consumption of protocols (considering S-MAC as a full-scale benchmark).

packets. In PEDAMAC, as transmission power decreases, the range of radio also decreases, which affects the network scalability. Column 4 represents the relative percentage of energy saved. In the reviewed papers, simulations are performed under different scenarios and with different considerations, so it is difficult to directly compare these protocols. Hence, the comparison values are presented as percentages of S-MAC serving as a benchmark. Figure 6 shows the energy consumption analysis. S-MAC consumes 2.8 mA/node and T-MAC consumes 0.4 mA/node [8]. B-MAC saves 57% more energy than S-MAC for a throughput of 240 b/s, because synchronization overhead increases in S-MAC [10]. PW-MAC protocol's duty cycle is only 11%, while duty-cycle of S-MAC is 50% [11]. Decreased duty-cycle leads to decreased energy consumption. Also, owing to operation at distinct power levels, PEDAMAC saves 38% more energy than S-MAC [14].

Table 2
Comparison of MAC protocols

Protocol	Type	Cross layer optimization	Energy conservation factors
S-MAC [7]	Single hop	No	Overhearing, idle listening
T-MAC [8]	Single hop		Idle listening, collision
B-MAC [9], [10]	Single hop		Overhearing, collision
PW-MAC [11], [12]	Multi hop		Idle listening, collision, retransmission
PEDAMAC [13], [14]	Single hop		N/A
Probabilistic polling [15]	Single hop		N/A
HEAP-EDF [16]	Single hop		N/A

Table 2 shows the comparison of MAC protocols. Column 4 represents the factors that were focused on while designing the respective protocols, in order to reduce energy consumption. The key consideration of probabilistic polling and HEAP-EDF is the optimal use of harvested energy rather than conservation of energy.

3.6. Open Research Problems

A number of MAC protocols have been proposed and designed for WSN, but there are still many open issues that need to be addressed. Cross-layer interaction and optimization are potential areas of research which can enhance the performance of MAC protocols. The MAC protocols available can be analyzed for various traffic generation and node distribution models. Development of multi-hop MAC protocols, in order to extend range and scalability, is another task to be considered in the future.

4. Network Layer

The main task of WSN is to sense and transmit data while using minimum resources. An efficient routing protocol is required at the network layer to choose a path with the minimum cost of delay, lifetime, energy or any other parameter that is more relevant to the application.

4.1. Energy Consumption Sources

Routing overhead is the main source of energy consumption at this particular level. Wang *et al*. presents one of the criteria to design the routing protocol with a minimum overhead to minimize energy consumption [17]. The overhead of a routing protocol varies with hop count and hop distance. In the case of small distances, single hop routing has less overhead. However, if the distance is long and cannot be covered with the available transmitted power, multi-hop routing is more efficient.

4.2. Design Constraints of a Routing Protocol

As compared to routing protocols designed for computer networks, WSN routing protocols need some distinctive features to handle a unique set of challenges [18], [19]:

Network scale: Node density may vary from hundreds to thousands, depending upon application. It is difficult to supervise such large, distributed structures. So, sensor nodes should be able to self-organize. The routing protocol should also deal with maintenance of global knowledge of such a large deployment.

Dynamics of node: WSNs are highly dynamic in nature. Owing to movement, power depletion and addition of new nodes, the topology of a WSN keeps changing. The routing protocol should be capable of adapting to frequent changes.

Resource constraints: A WSN need to operate on limited battery resources. Hence, the routing protocol should be able to transmit information over less than half a duty cycle. Some information-possessive applications need accuracy in data transmission. Therefore, the protocol should be able to trade-off energy consumption for accuracy.

Nature of node: Nodes operating over a certain coverage area may be homogenous or heterogeneous. Hence, the routing protocol needs to support nodes with unlike parameters and capabilities.

QoS: In a few applications delayed transmission of the sensed data may result in the loss of its significance. For such applications, delay is a critical parameter. Similarly, for a few other applications, other parameters – such as accuracy – may play a critical role. To maintain the quality of response of the application, these parameters need to be carefully traded-off for energy.

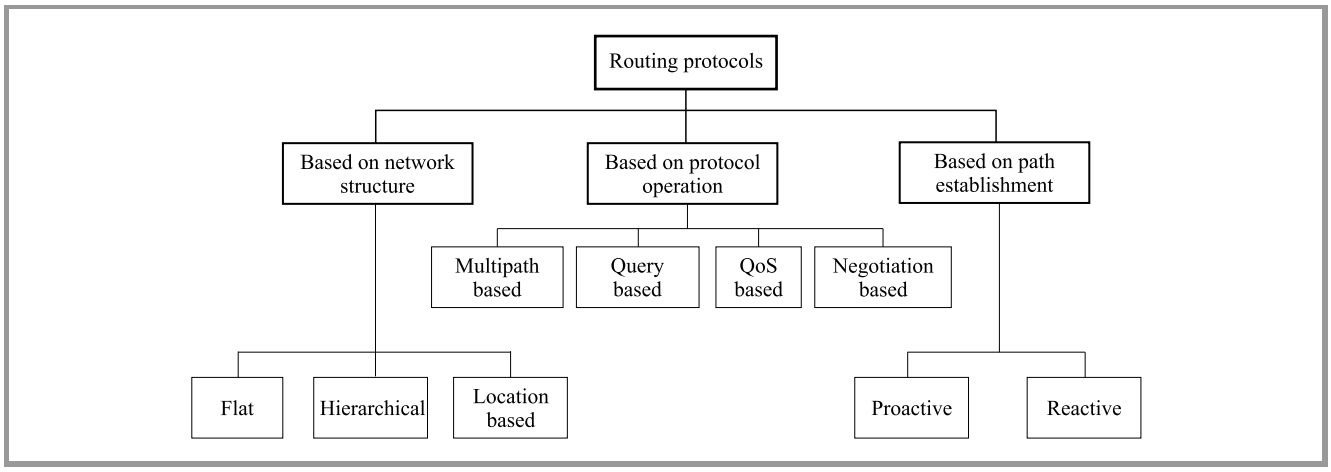


Fig. 7. Taxonomy of routing protocols.

4.3. Classification of Routing Protocols

Nodes can select from a number of available paths to transmit data to the sink or the base station. Routing protocols can be classified based on different criteria establishing the path to the sink, as shown in Fig. 7 [18], [20].

Routing protocols based on network structure:

Flat structure routing: All the nodes play the same role, i.e. each node is considered to be a base station and each node is provided with all information, so that the user can send a query to any node to get information.

Hierarchical structure routing: Not all nodes have the same capability. Higher capability nodes perform critical tasks, whereas less critical tasks are assigned to nodes with low capability. It is a two-level or multi-level structure.

Location-based routing: Nodes can be addressed based on their locations, whereas the location of a sensor can be detected using a satellite, provided the system is equipped with a low power GPS receiver. Another way is to measure the relative distance of the node from its neighbors based on strength of the signal received.

Routing protocols based on protocol operation:

Multipath routing protocol: In order to deliver data from source to destination, the protocol may rely on multiple paths. Multiple paths increase fault tolerance of the network, but also increase energy consumption and protocol overhead. An extension of the algorithm considers only the path with nodes having the highest energy. The path keeps changing whenever the protocol discovers a better path. By using the multipath routing protocol, reliability of the network can be increased in highly unreliable environments. A large packet can be divided into sub-packets and sent over different paths. A message can be reconstructed even if one of the sub-packets is lost due to link errors. Such an approach is known as multipath routing.

Query based routing: In query based routing, a node initiates a query and propagates it through the network. Each node receives the query and only the node having data that matches responds. Instead of propagating the query throughout the entire network, the node may send it in a random direction and wait for the response. If none of the nodes respond, then the node can propagate it through the whole network.

QoS based routing: In applications where parameters like delay, resources and bandwidth are critical, the routing protocol needs to maintain the quality and specifications of the critical parameter while delivering data. The routing protocol is responsible for maintaining a trade-off between energy and other metrics.

Negotiation based routing: Flooding and gossiping produce implosion and a single node may receive multiple data copies. The basic concept of the negotiation based protocol is to avoid propagation of duplicated packets. A sequence of negotiation messages is shared among the nodes to transmit redundant data to the next node. It reduces energy consumption and network congestion. The SPIN protocol discussed later is an example of the negotiation based protocol.

To deliver data from source to destination, the node initiating communication should know the path to the destination, i.e. path-based routing protocols are established in two ways:

Reactive path establishment: such protocols are event-driven. After a data packet has reached a node, the protocol decides the next node to be taken towards the destination. The decision about the next node may depend upon cache history, but in most cases the nodes have limited memory and low computational capability, hence no cache history is maintained. Another metric to decide the next hop can include distance, cost, bandwidth and energy of the node.

Proactive path establishment: under this scenario, the protocol decides the path to destination when the data packet is at the first hop or at the node where communication initiates. The path can be established based on

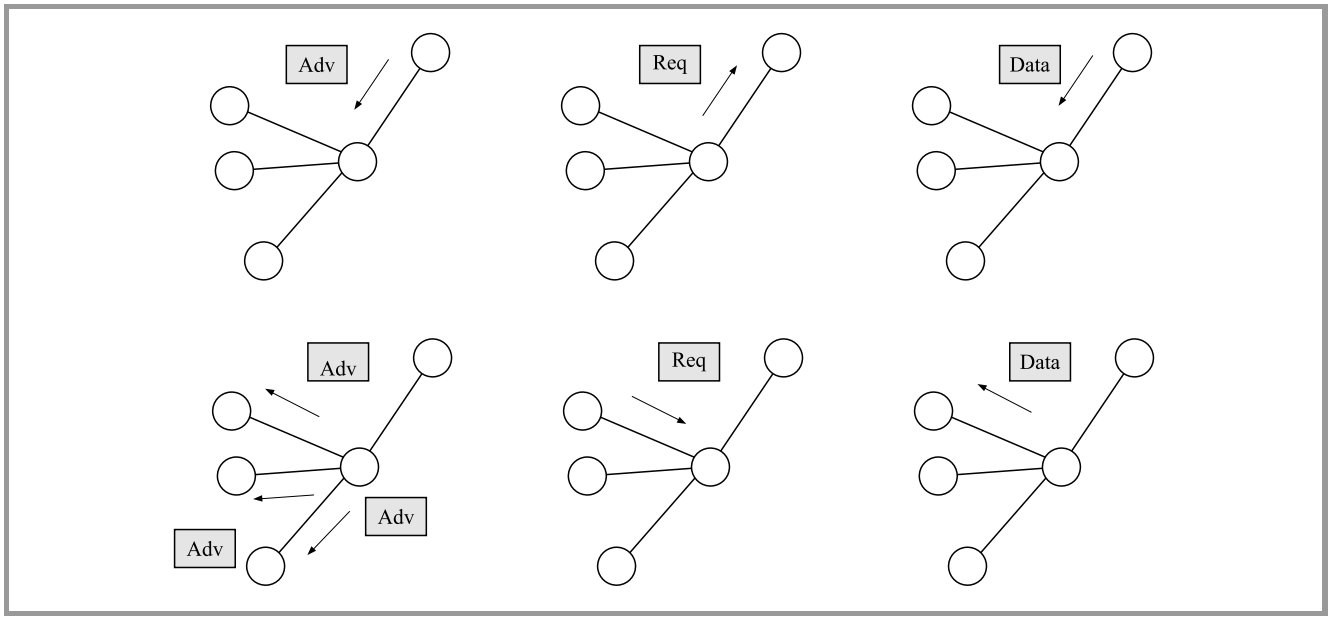


Fig. 8. SPIN protocol flowchart.

the minimum cost, maximum bandwidth or nodes with the highest energy levels. Once the path is established, all data packets propagate through the path selected. These protocols are not fault-tolerant, as data will be lost if the established link fails.

4.4. Network Layer Protocols

Flooding and gossiping: Flooding allows a node to send packets via all links. To avoid a packet looping indefinitely, the hop count or time-to-live is included in the packet. Another approach is gossiping, in which the packet is not sent via all outward links. The packet is transmitted only to a randomly selected neighbor. This saves bandwidth of the network but increases the delay from source to destination [21].

Sensor Protocols for Information via Negotiation (SPIN): SPIN overcomes the drawbacks of conventional dissemination protocols. SPIN is based on metadata [22]. Transmitter broadcasts the metadata of data. The receiver checks the information about data and sends the request to the transmitter if interested. Finally, the transmitter transmits the information to the interested receiver.

The SPIN protocol is presented in Fig. 9, where Adv are advertisement packets advertising metadata, Req are request for data packets from interested nodes to transmitter and Data are packets carrying data.

Directed diffusion: In directed diffusion, data packets are propagated through the network as interests, whereas the reverse reply link towards the transmitter is known as gradient [23]. Each node maintains a cache. When an event occurs, the node searches its cache. If the entry is not in the cache, it is added for future use. Caching increases efficiency and decreases energy consumption. Using the sequence of interests and gradients, the best path is reinforced between the transmitter and the receiver. Directed diffusion is based on the localized demand-driven query model. Receiver queries the sender node through interests for data and gets the response. The query-driven model increases the overhead.

Low-Energy Adaptive Clustering Hierarchy (LEACH): LEACH is a hierarchical cluster-based protocol. Nodes with higher energy are cluster heads [24] collecting information from all nodes in the cluster. Aggregated data is compressed and sent to the sink. LEACH reduces energy consumption, because cluster heads can be selected efficiently to increase network lifetime. The node generates a random number between 0 and 1, if the number generated is less than $T(n)$, the node can become a cluster head. The threshold ensures that the node has not become the cluster head in last $\frac{1}{p}$ rounds:

$$T(n)=\begin{cases} 0 & \text{if } n \notin G \\ \dfrac{p}{1-p\left(r \bmod \left(\frac{1}{p}\right)\right)} & \forall\, n \in G \end{cases}, \qquad (3)$$

where $T(n)$ is the threshold to choose the cluster head, G is the set of all nodes eligible for cluster head role, p is probability of being the cluster head and r is the current round number.

LEACH protocol is demonstrated in Fig. 9, where sensor nodes send data to cluster heads and cluster heads send aggregated data to the base station.

LEACH with Spare Management (LEACH-SM): It is a modification of the LEACH protocol. LEACH-SM has spare nodes which are normally in the sleep mode [25].

Table 3
Performance evaluation of MAC protocols

Protocol	Type	Energy consumption	Network lifetime	Mobility	Scalability
Flooding and gossiping [21]	Flat	High	Small	Yes	Low
SPIN [22], [28]	Negotiation based	Low	Small	Yes	Low
Directed diffusion [23]	Multipath	Moderate	Small	Limited	Low
LEACH [24], [29]	Hierarchical	High	Medium	No	Moderate
LEACH-SM [25]	Hierarchical	Moderate	Long	No	Moderate
DEEC [26]	Hierarchical, multilevel, heterogeneous	Low	Long	No	High
BLR [27], [28]	Location based	Low	Moderate	Limited	High

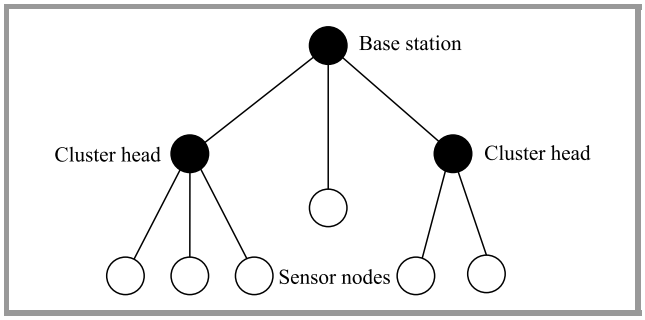


Fig. 9. Structure of a LEACH protocol network.

When the network is out of energy, spare nodes provide redundancy and increase network lifetime. LEACH-SM also has the capability to avoid deadlocks that may occur due to redundancy of nodes, and thus offers extended lifetime.

Distributed Energy Efficient Clustering (DEEC): DEEC was proposed for heterogeneous WSNs [26]. It considers multi-level heterogeneity. Other clustering protocols did not consider energy while choosing cluster heads. DEEC uses the knowledge about initial and residual energy of nodes while choosing cluster heads. DEEC used the same threshold in Eq. (3) to determine the cluster head, but threshold probability to select the cluster head depends on the heterogeneity of nodes.

$$p = \begin{cases} \dfrac{p'E(r)}{(1+am)E'(r)} & \text{if node is normal} \\ \dfrac{p'(1+a)E(r)}{(1+am)E'(r)} & \text{if node is advanced} \end{cases}, \qquad (4)$$

where p' is reference probability, $E(r)$ is residual energy and $E'(r)$ is average energy of the network, m is fraction of advanced nodes whose energy is a times higher than that of normal nodes. Normally, a cluster dies as its cluster head is out of energy. DEEC keeps on reassigning the role of the cluster head depending upon energy, to extend the lifetime of network. The DEEC stability period is 15% longer than in the stable election protocol. Also, it does not require any global knowledge to select the cluster head. Hence, it is more efficient than other clustering protocols.

Beacon-less Routing (BLR): In location based routing, nodes exchange a few messages called beacons to know the position of each other. These beacons create a large overhead and work inefficiently in erroneous wireless links. Therefore, BLR was proposed. BLR selects the next hop by computing the dynamic forwarding delay. A node broadcasts a data packet to all its neighbors but only the receiving node which is best positioned towards the destination, will forward the packet. Nodes within a certain area take part in forwarding. These areas are called forwarding areas and can be of any shape. The receiving node sends a passive acknowledgment back to the sending node [27], [28].

4.5. Comparison Table

Table 3 shows the comparison of the protocols' performance. Since the SPIN protocol is based on metadata, its energy consumption is low. The BLR protocol has a lower overhead. Hence, it is highly scalable and can be used for large networks. Figure 10 represents a lifetime analysis of the hierarchical protocols reviewed. The lifetime of LEACH-SM equals 183% of the lifetime of LEACH [25]. Also the lifetime of DEEC is 130% of the lifetime of LEACH [26].

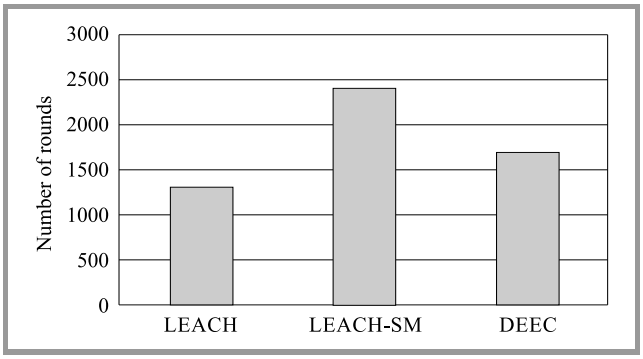


Fig. 10. Lifetime of hierarchical protocols.

4.6. Open Research Problems

Most routing protocols assume a channel to be loss-less. So, future work includes development of routing protocols for lossy wireless channels. Also, existing routing protocols can be evaluated for lossy channels. QoS can be enhanced to ensure latency-free routing. Security management can be explored to avoid such threats as sleep deprivation attacks, packet dropping attacks and collecting sensitive information [30]. The routing protocols considered can be upgraded to handle various types of traffic, and the effect of traffic on the lifetime of a network can be minimized. Different traffic profiles should be modeled and analyzed to design efficient routing protocols [31]. New routing protocols are required for mobile networks. Also, most routing protocols available assume that an ideal MAC protocol exists. Cross-layer optimization can be used to improve the performance of a network as a whole.

5. Cross Layer Approach

To improve network performance, interaction of parameters across the protocol stack is necessary. Energy is a parameter of the physical layer and routing is considered at the network layer. Layers need to interact to obtain the value of energy in a routing packet. This helps the routing protocol to choose an energy efficient path. Route energy packets which are used to exchange energy values among nodes are generated using the cross-layer design. Hoesel *et al*. presents a cross-layer approach in which the routing protocol uses topology and infrastructure information available at the MAC layer [32]. It reestablishes the route utilizing information at the MAC layer and outperforms S-MAC and Dynamic Source Routing (DSR) in mobile sensor networks [33]. Cross Layer MAC (CL-MAC) makes and optimizes scheduling decisions based on cross layer information [34]. Path-Loss Ordered Slotted Aloha (PLOSA) protocol is designed using cross-layer design for wireless data collection networks [35]. It helps in observing physical signals and orders the access of nodes accordingly. Nodes at a greater distance from the collector get an earlier chance to access the slot of the transmission channel. PLOSA has a high delivery rate and low latency.

A cross-layer approach has been presented by Catarinucci *et al*., in which protocol solutions are integrated with hardware [36]. A new wake-up system consisting of a sensor node and a power meter circuit was suggested in the paper, where the suggested protocol exploits the hardware to reduce power consumption. It is indicated by Alrajeh *et al*. that to design a secure routing protocol, the cross-layer approach was necessary [37]. Most security attacks are multilayered. The sleep deprivation attack, for instance, occurs at the physical layer, whereas the packet dropping attack occurs at the network layer. The author proposed to keep an eye on the packet count, in order to prevent malicious nodes from sending unnecessary packets and creating congestion. Hence, in order to select an energy efficient and secure path, cross-layer communication is necessary.

6. Conclusions and Future Works

In this paper existing MAC protocols and routing protocols have been surveyed. MAC protocols have been reviewed for both type of nodes followed by their advantages and disadvantages. This paper suggests that because of its low latency, PEDAMAC can be used for delay sensitive applications. Owing to the random wake-up schedule, PW-MAC offers high throughput. It sends one update in 1400 s, so overhead is moderate.

Diversified routing protocols ranging from flat to multilevel have been discussed in this paper. This paper analyses that DEEC has 30% more rounds than LEACH, because of low energy consumption. LEACH-SM has 83% more rounds than LEACH. Then, the cross-layer approach has been presented that improves performance of the protocol stack as a system. Although the protocols discussed may seem promising, there are still many challenges that need to be faced in WSNs. The cross-layer approach is a research area that needs to be studied and analyzed widely. Traffic modeling is another prospective area which can be analyzed and studied for improving performance or security of networks. Energy harvesting algorithms and models for WSNs also are subject to great advancements in the future.

References

[1] M. Abozahhad, M. Farrag, and A. Ali, "A comparative study of energy consumption sources for wireless sensor networks", *Int. J. of Grid Distrib. Comput.*, vol. 8, no. 3, pp. 65–76, 2015.

[2] J. Zheng and A. Jamalipour, *Wireless Sensor Networks: A Networking Perspective*. Wiley, 2014.

[3] A. S. Althobaiti and M. Abdullah, "Medium access control protocols for wireless sensor networks classifications and cross-layering", *Procedia Comp. Sci.*, vol. 65, pp. 4–16, 2015.

[4] W. Ye, J. Heidemann, and D. Estrin, "An energy-efficient MAC protocol for wireless sensor networks", in *Proc. of 21st Ann. Joint Conf. of the IEEE Comp. and Commun. Soc. IEEE INFOCOM 2002*, New York, NY, USA, 2002, pp. 1567–1576.

[5] K. Sohraby, D. Minoli, and T. Znati, *Wireless Sensor Networks Technology, Protocols and Applications*, 2nd ed. Wiley, 2013.

[6] U. Roedig, "f-MAC: A deterministic media access control protocol without time synchronization", in *Wireless Sensor Networks – Third European Workshop, EWSN 2006 Zurich, Switzerland, February 13-15, 2006 Proceedings*, K. Römer, H. Karl, and F. Mattern, Eds., *LNCS* 3868. Zurich: Springer, 2006, pp. 276–291.

[7] M. A. Ameen, S. M. R. Islam, and K. Kwak, "Energy saving mechanisms for MAC protocols in wireless sensor networks", *Int. J. of Distrib. Sensor Netw.*, vol. 6, no. 1, Article ID 163413, 2010 (doi: 101155/2010/163413).

[8] L. Wang and K. Liu, "An adaptive energy-efficient and low-latency MAC protocol for wireless sensor networks", in *Proc. Int. Conf. on Wirel. Commun., Netw. and Mob. Comput. WiCOM 2007*, Shanghai, China, 2007, pp. 2440–2443.

[9] C. Cano, B. Bellalta, A. Sfairopoulou, and M. Oliver, "Low energy operation in WSNs: A survey of preamble sampling MAC protocols", *Comp. Netw.*, vol. 55, no. 15, pp. 3351–3363, 2011.

[10] J. Polastre, J. Hill, and D. Culler, "Versatile low power media access for wireless sensor networks", in *Proc. 2nd Int. Conf. on Embed. Netw. Sensor Syst. SenSys 2004*, Baltimore, MD, USA, 2004.

[11] D. B. Johnson, "PW-MAC: An energy-efficient predictive-wakeup MAC protocol for wireless sensor networks", in *Proc. 30th IEEE Int. Conf. on Comp. Commun. IEEE Infocom 2011*, Shanghai, China, 2011, pp. 1305–1313.

[12] J. Kabara and M. Calle, "MAC protocols used by wireless sensor networks and a general method of performance evaluation", *Int. J. of Distrib. Sensor Netw.*, vol. 8, no. 1, 2012 (doi: 10.1155/2012/834784).

[13] S. Coleri and P. Varaiya, "PEDAMACS: Power efficient and delay aware medium access protocol for sensor networks", *IEEE Trans. on Mob. Comput.*, vol. 5, no. 7, pp. 920–930, 2006.

[14] M. G. C. Torres, "Energy consumption in wireless sensor networks using GSP", Master's Thesis, University of Pittsburgh, 2006.

[15] Z. A. Eu, H. P. Tan, and W. K. G. Seah, "Design and performance analysis of MAC schemes for Wireless Sensor Networks Powered by Ambient Energy Harvesting", *Ad Hoc Netw.*, vol. 9, no. 3, pp. 300–323, 2011.

[16] Y. Jin and H. P. Tan, "Optimal performance trade-offs in MAC for wireless sensor networks powered by heterogeneous ambient energy harvesting", in *Proc. IEEE IFIP Networking Conf. IFIP Networking 2014*, Trondheim, Norway, 2014 (doi: 10.1109/IFIPNetworking.2014.6857125).

[17] Q. Wang, M. Hempstead, and W. Yang, "A realistic power consumption model for wireless sensor network devices", in *Proc. 3rd Ann. IEEE Commun. Soc. on Sensor and Ad Hoc Commun. and Netw. SECON 2006*, Reston, VA, USA, 2006, pp. 286–295 (doi: 10.1109/SAHCN.2006.288433).

[18] J. N. Al-Karaki and A. E. Kamal, "Routing techniques in wireless sensor networks: A survey", *IEEE Wirel. Commun.*, vol. 11, no. 6, pp. 6–28, 2004 (doi: 10.1109/MWC.2004.1368893).

[19] K. Akkaya and M. Younis, "A survey on routing protocols for wireless sensor networks", *Ad Hoc Netw.*, vol. 3, no. 3, pp. 325–349, 2005.

[20] S. P. Singh and S. C. Sharma, "A survey on cluster based routing protocols in wireless sensor networks", *Procedia Comp. Sci.*, vol. 45, no. C, pp. 687–695, 2015.

[21] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey", *Comp. Netw.*, vol. 38, no. 4, pp. 393–422, 2002.

[22] W. R. Heinzelman, J. Kulik, and H. Balakrishnan, "Adaptive protocols for information dissemination in wireless sensor networks", in *Proc. 5th Ann. ACM/IEEE Conf. on Mob. Comput. and Netw. MobiCom'99*, Seattle, WA, USA, 1999, pp. 174–85.

[23] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed diffusion: A scalable and robust communication", in *Proc. 6th Ann. Int. Conf. on Mob. Comput. and Netw. MobiCom'00*, Boston, MA, USA, 2000, pp. 56–67.

[24] M. M. Afsar and M. H. Tayarani-N, "Clustering in sensor networks: A literature survey", *J. of Netw. and Comp. Appl.*, vol. 46, pp. 198–226, 2014.

[25] B. A. Bakr and L. T. Lilien, "Comparison by simulation of energy consumption and WSN lifetime for LEACH and LEACH-SM", *Procedia Comp. Sci.*, vol. 34, pp. 180–187, 2014.

[26] L. Qing, Q. Zhu, and M. Wang, "Design of a distributed energy-efficient clustering algorithm for heterogeneous wireless sensor networks", *Comp. Commun.*, vol. 29, no. 12, pp. 2230–2237, 2006.

[27] J. A. Sanchez, P. M. Ruiz, and R. Marin-Perez, "Beacon-less geographic routing made practical: Challenges, design guidelines, and protocols", *IEEE Commun. Mag.*, vol. 47, no. 8, pp. 85–91, 2009.

[28] M. Heissenbüttel, T. Braun, T. Bernoulli, and M. Wälchli, "BLR: Beacon-less routing algorithm for mobile ad hoc networks", *Comp. Commun.*, vol. 27, no. 11, pp. 1076–1086, 2004.

[29] L. J. G. Villalba, A. L. S. Orozco, A. T. Cabrera, and C. J. B. Abbas, "Routing protocols in wireless sensor networks", *Sensors*, vol. 9, no. 11, pp. 8399–8421, 2009.

[30] K. Chelli, "Security issues in wireless sensor networks: Attacks and countermeasures", in *Proc. of the World Congr. on Engin. WCE 2015*, London, United Kingdom, 2015, vol. I, pp. 519–524.

[31] Q. Wang, "Traffic analysis and modeling in wireless sensor networks and their applications on network optimization and anomaly detection", *Netw. Prot. and Algorithms*, vol. 2, no. 1, pp. 74–92, 2010 (doi: 10.5296/npa.v2i1.328).

[32] L. Van Hoesel, T. Nieberg, J. Wu, and P. J. M. Havinga, "Prolonging the lifetime of wireless sensor networks by cross-layer interaction", *IEEE Wirel. Commun.*, vol. 11, no. 6, pp. 78–86, 2004.

[33] D. B. Johnson and D. A. Maltz, "Dynamic source routing in ad hoc wireless networks", *Mob. Comput.*, vol. 353, pp. 153–181, 1996.

[34] M. S. Hefeida, T. Canli, and A. Khokhar, "CL-MAC: A cross-layer MAC protocol for heterogeneous wireless sensor networks", *Ad Hoc Netw.*, vol. 11, no. 1, pp. 213–225, 2013.

[35] D. Espes, X. Lagrange, and L. Suárez, "A cross-layer MAC and routing protocol based on slotted aloha for wireless sensor networks", *Annales des Telecommun./Annals of Telecommun.*, vol. 70, no. 3–4, pp. 159–169, 2015.

[36] L. Catarinucci *et al.*, "A cross-layer approach to minimize the energy consumption in wireless sensor networks", *Int. J. of Distrib. Sensor Netw.*, vol. 10, no. 1, 2014 (doi: 10.1155/2014/268284).

[37] N. A. Alrajeh, J. Lloret, and J. Loo, "Secure routing protocol using cross-layer design and energy harvesting in wireless sensor networks", *Int. J. of Distrib. Sensor Netw.*, vol. 9, no. 1, 2013 (doi: 10.1155/2013/374796).

Aarti Kochhar is working as an Assistant Professor at Lovely Professional University, Phagwara, India. She has completed her M.E. degree in Electronics and Communication Engineering. She conducts research in the areas of wireless sensor networks.

E-mail: aarti.kochhar92@gmail.com
University Institute of Engineering and Technology
Panjab University
Chandigarh, India
Lovely Professional University
Phagwara, India

Pardeep Kaur is working as an Assistant Professor at the Electronics and Communication Engineering Department at U.I.E.T., Panjab University, Chandigarh, India. She received her B.Tech. and M.E. degrees in Electronics and Communication Engineering. She is pursuing her Ph.D. in wireless sensor networks. Her areas of interest include optical communication and wireless communication.

E-mail: pardeep.tur@gmail.com
University Institute of Engineering and Technology
Panjab University
Chandigarh, India

Sukesha Sharma is working as an Assistant Professor at the Electronics & Communication Engineering Department at U.I.E.T., Panjab University, Chandigarh, India. She has completed her B.Tech. and M.E. degrees in Electronics and Communication Engineering. Her research interests include embedded systems, automation and control, active vibration control and energy harvesting.
E-mail: er_sukesha@yahoo.com
University Institute of Engineering and Technology
Panjab University
Chandigarh, India

Preeti Singh is working as an Assistant Professor at the Electronics & Communication Engineering Department at U.I.E.T., Panjab University, Chandigarh, India. She has completed her B.Tech. and M.E. degrees in Electronics and Communication Engineering. She obtained her Ph.D. degree in 2013. Her areas of interest include optical communication (wired and wireless), optical biosensors and cognitive neuroscience.
E-mail: preets.singh.82@gmail.com
University Institute of Engineering and Technology
Panjab University
Chandigarh, India

Paper

Underwater Acoustic Sensor Node Scheduling using an Evolutionary Memetic Algorithm

V. Sivakumar and D. Rekha

School of Computing Science and Engineering, VIT, Chennai, Tamil Nadu, India

https://doi.org/10.26636/jtit.2018.116217

Abstract—Underwater Acoustic Sensor Networks (UWASNs) play an important role in monitoring the aqueous environment which has created a lot of interest for researchers and scientists. Utilization of underwater acoustic sensor node (UASN) scheduling for transmission remains, due to the limited acoustic bandwidth available, a challenge in such an environment. One of the methods to overcome this problem is to efficiently schedule UASN data using time division multiple access (TDMA) protocols the parallel transmissions, simultaneously avoiding interference. The paper shows how to optimize the utilization of acoustic sensor node bandwidth by maximizing the possible node transmissions in the TDMA frame and also by minimizing the node's turnaround wait time for its subsequent transmissions by using an evolutionary memetic algorithm (MA). The simulation of MA-TDMA proves that as the size of the network increases, every node in UWASN transmits with an average minimal turnaround transmission time. It also proves that as the TDMA cycle repeats, the overall network throughput gets maximized by increasing the possible node transmissions in the MA-TDMA frame.

Keywords—broadcast UASN scheduling, memetic algorithm, time division multiple access, underwater acoustic sensor network.

1. Introduction

Underwater Acoustic Sensor Networks (UWASNs) play an important role in weather monitoring and in the aqueous environment. Over a past few decades, underwater sensor nodes have been deployed for data collection performed manually, by recording [1]. This method of gathering information evoked much enthusiasm for the advancement of underwater sensor networks enabling the sensors to be connected. Acoustic communication is the most reliable and adaptable method in the case of the time-varying underwater channel [2]. The acoustic signal can be sent over longer distances (many kilometers), while electromagnetic waves are highly attenuated even over short distances, and they require using large aerials with high transmission power for communication [3]. On the other hand, optical links work well in underwater environments for short distance communication, but in large networks the data gets quickly absorbed and scattered [4].

Although acoustic communication eliminates the disadvantages of optical and electromagnetic signals in underwater environments, it suffers from limited bandwidth and large propagation delays due to the low speed of sound in water (approximately 1500 m/s) [5]. In general, UWASNs are one-hop and multi-hop networks that are directly based on hardware or are software-defined [6]. Every node within the network can communicate with its neighbor node directly, in the one-hop fashion. In a multi-hop network, all one-hop nodes collect and forward the data via an acoustic link.

In UWASNs, one of the important research area involves medium access control (MAC), which can provide efficient access to the shared underwater acoustic communication medium. Due to the large propagation delay in water, and also being half-duplex in nature for communication, terrestrial MAC protocols do not work effectively underwater [7]. There are two types of MAC protocols that are used for underwater communication. These are schedule based and non-schedule based protocols [8]. The existing non-schedule based protocols are like ALOHA [9], [10] and CSMA/CA [11]. Their implementation is simple, but they suffer from increased collisions at low data rates.

The schedule based protocols for UWASNs provide a high data rate with fewer collisions, and a good network throughput. The terrestrial sensor network protocols cannot be applied directly in UWASNs, due to the continuous change in the underwater environment [12]. In UWASNs, Time Division Multiple Access (TDMA) can work well for a prolonged period of time over which collision is been avoided efficiently [13]. The exact constraints in the underwater acoustic link can be solved using TDMA protocols [14]. TDMA is more appropriate for the underwater acoustic channel than FDMA and CDMA. FDMA and CDMA require more transmission capacity, which is not available in the underwater acoustic channel. In this paper we have considered fixed or anchored UWASNs for surveillance purposes, where each node has certain neighbors for communication. By using TDMA for scheduling, each node is assigned a different time slot for its transmission. It is to be noted that each node within the network takes a very long turnaround schedule for its next transmission. As the size of the network increases, time slots increase as well, which leads to a longer turnaround wait time for the trans-

mission performed by individual nodes. This results in less effective utilization of acoustic bandwidth and low network throughput in big networks.

Underwater TDMA-based MAC protocols are introduced in [13], [15], [16]. All proposals have major issues in broadcast scheduling. In [17] hybrid spatial reuse TDMA (HSR-TDMA), the problem of broadcast scheduling has been solved using the hybrid spread spectrum method, but the hidden and exposed terminal issues have not been addressed effectively. The other drawback in HSR-TDMA is that few nodes in the network suffer from very long schedules for its subsequent transmissions, which directly affects the overall throughput. This broadcast scheduling problem can be addressed with evolutionary algorithms [18] to reach an optimal solution.

In this paper, the UASN broadcast scheduling problem experienced in applications that require frequent and periodic transmissions is solved by using the memetic algorithm. The aim is to minimize the node's turnaround transmission wait time and maximize the number of the node's possible transmissions in the TDMA frame that do not interfere within the same time slot, which results in full utilization of the acoustic channel limit available.

The remaining parts of this paper are arranged as follows. Section 2 explores underwater acoustic sensor node (UASN) for scheduling conflicts. Section 3 investigates the formation of the UASN broadcast scheduling problem. Section 4 describes the memetic algorithm used for solving the underwater broadcast scheduling problem. Section 5 reports the experimental simulation results and Section 6 draws the conclusions.

2. Underwater Acoustic Sensor Node Scheduling Conflicts

In multi-hop UWASNs, there are two types of major packet collisions that can occur in broadcast scheduling: primary collision and secondary collisions [19]. If node i and j start transmitting in the same time slot, then it will end up into a primary collision occurs. The secondary collision can occur, when node i in the network intends to receive two or more number of packets from the its directly linked acoustic nodes within the same time slot.

In underwater scheduling, in addition to primary and secondary conflicts, it is the non-trifling propagation delay that poses another serious problem. Because of the continuous change in water temperature, salinity, pressure, etc., the propagation delay may vary from time to time. In TDMA-based scheduling, the propagation delay can be avoided by considering guard interval time and the maximum expected propagation delay [20]. The primary and secondary collision can avoided in TDMA by considering node transmission performed in two or more hops [21].

Explanations are presented in [22] and [20], where guard time is added after every TDMA packet. MAC avoids packet collision early and delayed reception of the packet. The largest size of the expected propagation delay experienced underwater is reduced by using propagation estimation to stagger transmission [23]. HSR-TDMA presented in [17] overcomes the above conflicts and increases the number of node transmissions, but a few nodes within the network suffered from a long turnaround wait time for their subsequent transmissions. As the size of UWASN increases, the HSR-TDMA frame length also increases, which shows that a few nodes in the network suffer from very long turnaround times for their subsequent transmissions also, which means they will be the cause of the lowest successful packet transmission rate (STR) for those few nodes, which affects the overall network throughput [17]. Therefore, the main aim of this work is to develop an optimized TDMA schedule with an average minimum turnaround time for subsequent node transmissions, and to succeed in maximizing the overall UWASN throughput.

3. Formation of UASN Broadcast Scheduling Problem

UWASN is a special kind of an ad-hoc network that can be represented as an undirected graph $U(S,E)$ [17], where S holds the number of nodes comprising the network and E accounts for the link between nodes. Acoustic link $E(i,j)$ between nodes i and j shows that the two nodes are connected within the transmission range. Figure 1 presents a simple multi-hop acoustic sensor network referred to in [17]. Every node within the network is connected to its neighbor node to form a link.

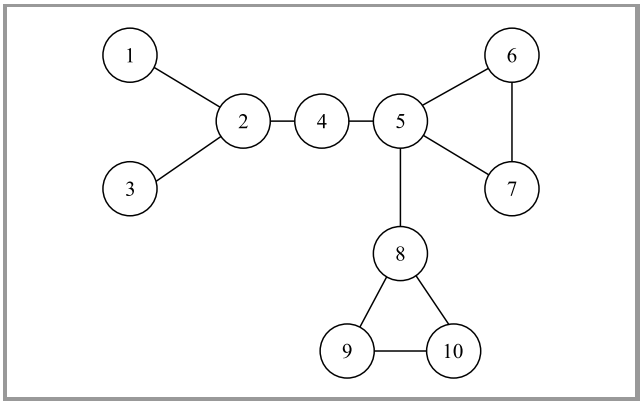


Fig. 1. A simple multi-hop UWASN.

The scheme from Fig. 1 has $|S| = 10$ network nodes in broadcasting. Primary conflicts are avoided by identifying adjacent node connections and secondary conflicts are eliminated by recognizing the nodes of two-hop connectivity. The identified connectivity matrix Con_M for the 10-node network is shown in Fig. 2. Rows in Con_M form the association between nodes. Columns represent sensor nodes. Zeros in Con_M indicate that there is no connection between nodes, while ones indicate that links exist. Two-hop connectivity for a given UWASN is considered to consti-

Con_M =

	1	2	3	4	5	6	7	8	9	10
1	0	1	0	0	0	0	0	0	0	0
2	1	0	1	1	0	0	0	0	0	0
3	0	1	0	0	0	0	0	0	0	0
4	0	1	0	0	1	0	0	0	0	0
5	0	0	0	1	0	1	1	1	0	0
6	0	0	0	0	1	1	0	1	0	0
7	0	0	0	0	1	1	0	0	0	0
8	0	0	0	0	1	0	0	0	1	1
9	0	0	0	0	0	0	0	1	0	1
10	0	0	0	0	0	0	0	1	1	0

Fig. 2. UWASN connectivity matrix.

Hop_M =

	1	2	3	4	5	6	7	8	9	10
1	0	1	1	1	0	0	0	0	0	0
2	1	0	1	1	1	0	0	0	0	0
3	1	1	0	1	0	0	0	0	0	0
4	1	1	1	0	1	1	1	1	0	0
5	0	1	0	1	0	1	1	1	1	1
6	0	0	0	1	1	0	1	1	0	0
7	0	0	0	1	1	1	0	1	0	0
8	0	0	0	1	1	1	1	0	1	1
9	0	0	0	0	1	0	0	1	0	1
10	0	0	0	0	1	0	0	1	1	0

Fig. 3. UWASN two-hop matrix.

tute a hop matrix Hop_M (Fig. 3). The rows of the two-hop matrix represent one- or two-hop connections between the nodes, while columns represent sensor nodes. The ones in Hop_M show that it might be one or two hops apart from the selected node.

The TDMA frame TDMA_M is shown as the $S \times T$ matrix, where T is the number of time slots in the TDMA frame and $S = \{S_1, S_2, \ldots, S_n\}$ represents the number of UASNs involved in the network shown in Figs. 4 and 5. The rows represent the number of time slots and all columns represent the transmitting nodes (i.e. adjacent nodes) in the network. The ones in the TDMA_M show that the nodes to transmit within that particular time slot without any conflicts and zeros are considered as receiving nodes in the network. Figure 4 represents a conventional TDMA frame for a 10-node network. Figure 5 shows an optimum MA-TDMA frame with minimum time slots and maximum possible transmission available for the same network. Nodes 1 and 5 are transmitted in the first slot, without any interference.

TDMA_M =

	1	2	3	4	5	6	7	8	9	10
1	1	0	0	0	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0	0	0
3	0	0	1	0	0	0	0	0	0	0
4	0	0	0	1	0	0	0	0	0	0
5	0	0	0	0	1	0	0	0	0	0
6	0	0	0	0	0	1	0	0	0	0
7	0	0	0	0	0	0	1	0	0	0
8	0	0	0	0	0	0	0	1	0	0
9	0	0	0	0	0	0	0	0	1	0
10	0	0	0	0	0	0	0	0	0	1

Fig. 4. Conventional TDMA frame.

TDMA_M =

	1	2	3	4	5	6	7	8	9	10
1	1	0	0	0	1	0	0	0	0	0
2	0	1	0	0	0	1	0	0	1	0
3	0	0	1	0	0	0	1	0	0	1
4	0	0	0	1	0	0	0	0	1	0
5	1	0	0	0	0	0	0	1	0	0

Fig. 5. Optimal MA-TDMA frame.

The underwater TDMA scheduling issue has proven to be of the non-deterministic polynomial (NP) variety as presented in [19]. It is formulated as a special vertex coloring problem, whose solution for a given graph is NP-complete [24]. In this article, the problem is approached by adopting an evolutionary memetic algorithm. The optimum TDMA solution is determined based on the fitness criteria given by Eq. (1).

The available underwater bandwidth utilization is evaluated by Eqs. (2)–(3), and the average time delay for UWASN is calculated by Eq. (5). The tight lower bound $\Delta = 1$ terminates the algorithm. Let $Dg(s)$ be a set of UASN network connectivity, and let Max_D represent the highest degree of connectivity.

$$Max\_D = \max_{s \in S} |Dg(s)| . \tag{1}$$

Then the tight lower bound for MA-TDMA frame is created [21] as:

$$\Delta = |T| - Max\_D \geq 1. \tag{2}$$

If $\Delta = 1$, the solution is optimal.

The acoustic bandwidth utilization is calculated for the entire UWASN:

$$\alpha = \frac{1}{|S| \, . \, |T|} \left[\sum_{i=1}^{|S|} \sum_{j=1}^{|T|} \text{TDMA\_M}_{\text{ij}} \right]. \tag{3}$$

For each node:

$$\alpha_s = \left[\sum_{i=1}^{|S|} \text{TDMA\_M}_{\text{is}} \right]. \tag{4}$$

The average time delay is calculated as:

$$\eta = \frac{|S|}{|T|} \sum_{i=1}^{T} \left[\frac{1}{\sum_{j=1}^{S} \text{TDMA\_M}_{\text{ij}}} \right], \tag{5}$$

where η is the average node availability of UASN in the network. By minimizing the η value, the optimal optimum network design can be achieved [25].

4. Memetic Algorithm

Memetic algorithms (MAs) were developed from evolutionary algorithms that apply local search processes in the agents to improve their fitness [26]. MA belongs to the family of meta-heuristic methods and used the hybrid population approach which combines genetic algorithm and local search methods [27]. Adoption of MA successfully solves any difficult optimization problem. MA is used in computing to find the actual or nearby optimal solution. Algorithm 1 establishes the structure of MA to solve the problem of optimal TDMA node scheduling in underwater environments.

Algorithm 1: Memetic algorithm

Memetic algorithm (Mempop, maxgen, Mem_s, Mem_c, Mem_m)

Initialization

Generate initial population (Mempop)

MA operations

Mem_s = select(Mempop)

while *condition not terminated* **do**

Mem_m = mutation(Mem_c)

Mem_{opt} = optimizer(α_s, Mem_m)

Mem_{imp} = improver((Hop_M), Mem_{opt})

Mem_{new} = evaluate(Mem_{imp})

Mempop = survival(Mempop, Mem_{new})

end while

MA is based on the following inputs: the population for MA, maximum number of generations for the algorithm to terminate, probability measurement for the selection, crossover and mutation operations (Table 1). MA will iterate its process until any one of the following conditions is satisfied, either by achieving the tight lower bound $\Delta = 1$ or by reaching the maximum number of generations.

Table 1
Simulation parameters

Parameters	Values
Population size	60
Crossover rate	0.32
Mutation rate	0.01
Maximum generation	300

4.1. Selection and Survival Phase

The different probable conventional TDMA frames are chosen as MA population. The iteration process begins with the selection phase. The reproduction process is done by way of the k-tournament selection pressure. Depending upon the selection pressure, two parents are picked up from the mating pool for evolution. As per the *survival of the fittest* rule, only the fittest population would be retained for next generations by removing the unwanted population.

4.2. Crossover and Mutation Phase

The selected parents will go through the crossover and mutation process to produce a new generation. In this study a single point crossover is performed for each row, with random bit strings selected from the TDMA frame and with information interchanged between them. For example, consider two parents with single row P1 = 1111 and P2 = 0000, with a random crossover point as two. The produced children are C1 = 0011 and C2 = 1100. This child replaces the parents and checks for the constraints by referring to the Hop_M. If they violate the constraints, then it gets dropped from the process. The mutation phase is performed only at the rows of the TDMA frame obtained. It is done only by flipping the zeros to ones based on their non-violating condition. The mutation may increase utilization of the acoustic channel.

4.3. Optimizer Phase

In this phase MA tries to minimize the number of time slots in the TDMA frame. This process is performed by identifying the utilization factor α_s of each channel, based on Eq. (4). For example, if a node transmits more than once in the TDMA frame $(\alpha_s > 1)$, then this row is removed from the frame.

4.4. Improver Phase

This phase improves channel utilization by increasing the number of node transmissions by referring to the Hop_M. Since the optimizer and improver operation is carried out in each iteration of MA, the optimum TDMA frame is obtained in a lower number of generations, within an acceptable computation time. Fitness for the new population is evaluated based on two criteria: overall UWASN channel utilization factor and tight lower bound $\Delta = 1$. If both are satisfied, the algorithm gets terminated. In the worst case scenario, the nearby optimum solution is obtained by reaching the maximum number of generations specified in the algorithm.

5. Simulation Results for Underwater Broadcast Scheduling

Typical simulation results were obtained in Matlab using parameters shown in Table 1. Fitness is evaluated after every generation in the memetic algorithm. The UWASN is configured with 10, 50, 80, 100, 200, 300, 400 and 500 randomly located nodes, respectively. For a 10-node UWASN it is observed that existing HSR-TDMA node transmission

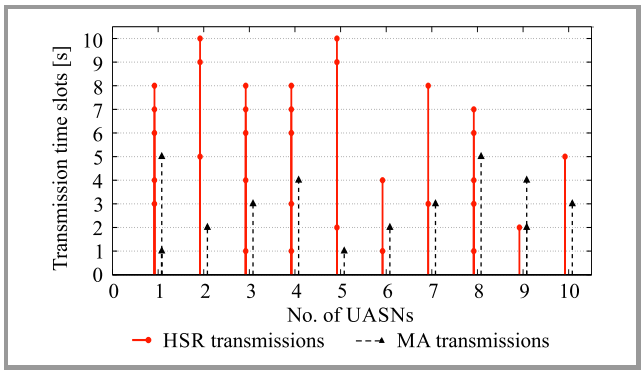


Fig. 6. Comparison between HSR and MA TDMA frame for 10 node UWASN.

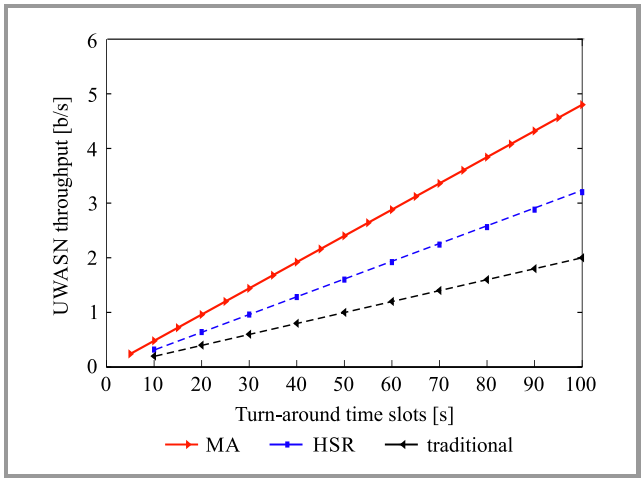


Fig. 7. Comparison of 10-node UWASN throughput.

Table 2
Comparison of MA-TDMA and HSR-TDMA frame length

No. of UASNs	No. of acoustic links	MA-TDMA frame length [bytes]	HSR-TDMA frame length [bytes]
10	22	5	10
25	43	7	25
50	98	9	50
100	200	11	100
250	430	12	250
500	1000	15	500

takes place within 10 slots, for whereas MA-TDMA employs only five slots as shown in Fig. 6. Also, it can be seen in Fig. 6 that nodes 9 and 10 in HSR-TDMA have transmitted only once in time slot 2 and 5 respectively, which shows that it has to wait nine slots for their next transmission. This long turnaround wait time leads to the low successful packet transmission rate. Whereas in MA-TDMA, every node in the UWASN schedules with an average minimal turnaround wait time of five slots. This shows

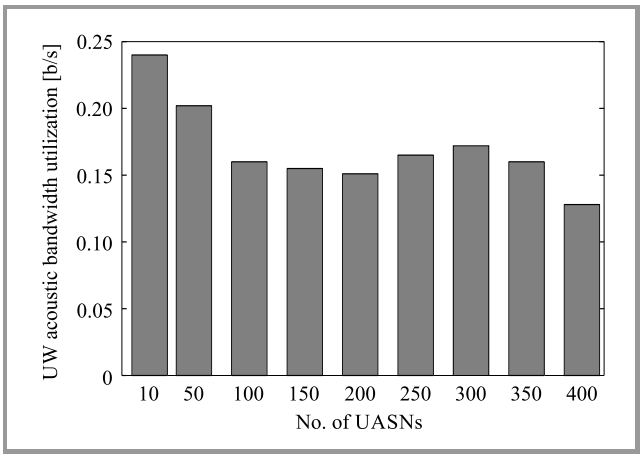


Fig. 8. Underwater acoustic bandwidth utilization for various network sizes.

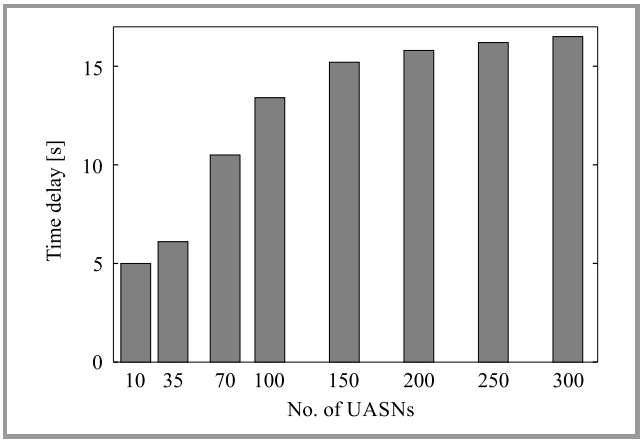


Fig. 9. Average time delay for different network sizes.

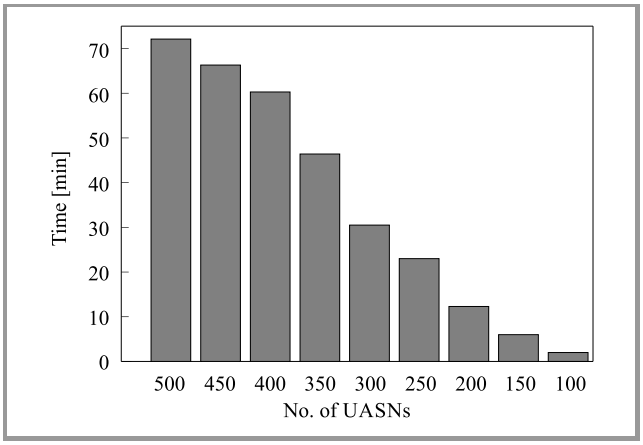


Fig. 10. Computation time take by MA for various network sizes.

that the network throughput has increased along with the increase of the TDMA cycle, as shown in Fig. 7. Table 2 shows that as the number of nodes increases in UWASN, the HSR-TDMA frame length increases as well. Meanwhile, in MA-TDMA, the optimum frame length is maintained for various network sizes, which shows that MA-TDMA performs best as the network size increases.

Table 3
Simulation results of memetic algorithm

No. of UASNs	No. of acoustic links	Avg. ND	Max. ND	Optimal TDMA frame length	α	Avg. no. of generations	Computation time
10	22	4	4	5	0.245	2.3	1.3 s
50	84	4	6	7	0.202	5.10	7 s
80	150	7	9	10	0.175	7.48	11 s
100	200	7.5	10	11	0.160	16.5	2.0 min
200	400	8	10	11	0.151	30	12.3 min
300	600	7	10	11	0.172	54.3	30.5 min
400	800	12	16	17	0.160	60.28	65.3 min
500	1000	9	14	15	0.128	89.03	72.11 min

In the second stage, the MA process is performed 50 times. The average value of simulated results is shown in Table 3. The average node degree and maximum ND for various network sizes has been evaluated and the optimum TDMA frame was obtained. This optimum TDMA frame length is obtained by satisfying the tight lower bound $\Delta = 1$ of the MA. This proves that for various sizes of the network, the average turnaround time is maintained. It is also observed that nodes 1 and 9 in MA-TDMA, as shown in Fig. 6, have transmitted twice within 5 slots, which is a sign of better utilization of the acoustic bandwidth available. While comparing MA with traditional and HSR approaches, as seen in Fig. 7, one may observe that throughput for MA-TDMA was increased along with the increase of the TDMA frame cycle. Figures 8 and 9 show the utilization of bandwidth and the average time delay of UWASNs of various sizes, with the said parameters calculated from Eqs. (3) and (5). It can be observed that the average time delay increases as the size of network gets bigger. The channel utilization result obtained for various sizes of acoustic networks depends on the connectivity between nodes within the network. Figure 10 shows the computation time taken by MA for various sizes of UWASNs. As the number of nodes increases, the computation time taken by MA increases as well. Therefore, the simulation results for various simulation scenarios prove that MA-TDMA outperforms traditional and HSR-TDMA in turnaround transmission wait time and also produces high network throughput by utilizing the acoustic bandwidth available.

6. Conclusion

In this study, the UWASN TDMA frame has been optimized by using the memetic algorithm. The optimizer reduces the number of time slots in the TDMA frame which, in turn, reduces the turnaround transmission wait time in the network. The algorithm uses an improver that increases the number of possible transmissions in the frame, thereby increasing the effectiveness of utilization of the limited acoustic bandwidth available. MA-TDMA proves, even for large networks, that the average turnaround wait time for node transmission is very low when compared with traditional and HSR-TDMA. This proves that MA-TDMA increases network throughput, as the TDMA cycle gets increased. Compared to other evolutionary algorithms, MA provides an improved solution with an acceptable computation time. The simulation results prove that MA works well for UWASN broadcast node scheduling.

References

[1] I. Vasilescu, K. Kotay, D. Rus, M. Dunbabin, and P. Corke, "Data Collection, Storage, and Retrieval with an Underwater Sensor Network", in *Proc. 3rd Int. Conf on Embed. Network. Sens. Sys. ACM SenSys*, San Diego, CA, USA, November 2005 (doi: 10.1145/1098918-1098936).

[2] X. Han, J. Yin, G. Yu, and X. Zhang, "Underwater acoustic communication in time-varying channel environment based on passive time reversal", *J. Acoust. Soc. Am.*, vol. 139, no. 4, pp. 3326–3334, 2015.

[3] I. F. Akildiz, D. Pompili, and T. Melodia, "Underwater acoustic sensor networks: Research Challenges", *J. of Ad Hoc Networks*, vol. 3, no. 3, pp. 257–279, 2005.

[4] H. Kaushal and Georges Kaddoum, "Underwater Optical Wireless Communication", *IEEE Access*, vol. 4, pp. 1518–1547, 2016.

[5] G. E. Burrowes and J. Y. Khan, "Investigation of a short-range underwater communication channel for MAC protocol design", in *Proc. 4th Conf. on Signal Proces. and Commun. Sys. ICSPS*, Gold Coast, QLD, Australia, December 2010, pp. 1–8 (doi: 10.1109/ICSPCS.2010.5709665).

[6] I. F. Akyildiz, P. Wang, and S.-C. Lin, "SoftWater: Software-defined networking for next-generation underwater communication systems", *J. of Ad Hoc Networks*, vol. 46, pp. 1–11, 2016 (doi: 10.1016/j.adhoc.2016.02.016).

[7] I. F. Akildiz, P. Wang, and Z. Sun, "Realizing underwater communication through magnetic induction", *IEEE Commun. Magaz.*, vol. 53, no. 11, pp. 42–48, 2015.

[8] S. Climent, A. Sánchez, J. V. Capella, N. Meratnia, and J. J. Serrano, "Underwater Acoustic Wireless Sensor Networks: Advances and Future Trends in Physical, MAC and Routing Layers", *Sensors*, vol. 14, no. 1, pp. 795–833, 2014.

[9] K. S. Geethu and A. V. Babu, "Improving energy efficiency performance of ALOHA based underwater acoustic sensor networks", in *Proc. IEEE Distributed Computing, VLSI, Electrical Circuits and Robotics DISCOVER*, Mangalore, Karnataka, India, 2016 (doi:10.1109/DISCOVER.2016.7806247).

[10] A. Syed, W. Ye, and J. Heidemann, "Comparison and evaluation of the T-Lohi MAC for underwater acoustic sensor networks", *IEEE J. on Select. Areas in Commun.*, vol. 26, no. 9, pp. 1731–1743, 2008 (doi: 10.1109/JSAC.2008.081212).

[11] X. Guo, M. R. Frater, and M. J. Ryan, "Design of a propagation-delay-tolerant MAC protocol for underwater acoustic sensor networks", *IEEE J. of Ocean. Engineer.*, vol. 34, no. 2, pp. 170–180, 2009.

[12] N. Li, J.-F. Martínez, J. M., Meneses Chaus, and M. Eckert, "A survey on underwater routing protocols", *Sensor*, vol. 16, no. 3, 2016 (doi: 10.3390/s16030414).

[13] L. Hong, F. Hong, Z. Guo, and Z. Li, "ECS: Efficient communication scheduling for underwater sensor networks", *Sensors* vol 11, no. 3, pp. 2920-2938, 2011.

[14] M. Chitre, S. Shahabodeen, and M. Stojanovic, "Underwater Acoustic Communications and Networking: recent advances and future challenges", *Marine Technol. Soc. J.*, vol. 42, pp. 103–116, 2008.

[15] P.-H. Huang, Y. Chen, B. Krishnamachari, and A. Kumar, "Link scheduling in a single broadcast domain underwater network", in *IEEE Int. Conf. on Sensor Networks, Ubiquitous and Trustworthy Comput. SUTC* , Newport Beach, CA, USA, June 2010, pp. 205–212.

[16] J. W. Lee and H. S. Cho, "Cascading Multi-Hop Reservation and Transmission in underwater acoustic sensor networks", *Sensors*, vol. 14, no. 10, pp. 18390–18409, 2014.

[17] R. Diamant and L. Lampe, "Spatial Reuse TDMA for broadcast Ad-Hoc Underwater Acoustic Communication Networks", *IEEE J. of Ocean. Engineer.*, vol. 36, no. 2, pp. 172–185, 2011.

[18] D. Arivudainambi and D. Rekha, "Memetic algorithm for minimum energy broadcast problem in wireless ad hoc networks", *Swarm and Evolut. Computat., Elsevier*, vol. 12, pp. 57-64, 2013 (doi: 10/1016/j.swevo.2013.04.001).

[19] C.-C. Hsu, M.-S. Kuo, C.-F. Chou, K. C.-J. Lin, "The elimination of spatial-temporal uncertainty in underwater sensor networks", *IEEE/ACM Transact. on Network.*, vol. 21, no. 4, pp. 1229-1249, 2013.

[20] P. Anjangi and M. Chitre, "Design and implementation of super-TDMA: A MAC protocol exploiting large propagation delays for underwater acoustic networks", in *ACM Int. Conf. on Underwater Networks and System WUWNET*, Washington DC, USA, October 2015, Article no. 1, pp. 1–8.

[21] G. Chakraborty, "Genetic algorithm to solve optimum TDMA transmission schedule in broadcast packet radio networks", *IEEE Transact. on Commun.*, vol. 52, no. 5, pp. 765–777, 2004.

[22] L. Hong, F. Hong, Z. Guo, and X. Yang, "A TDMA-based MAC protocol in Underwater Sensor Networks", in *Proc. IEEE 4th Conf. on Wireless Communic., Network. and Mobile Comput. WiCOM*, Dalian, Liaoning, China, October 2008, pp. 1–4 (doi: 10.1109/WiCOM.2008.838).

[23] K. Kredo, P. Djukic, and P. Mohuputra, "STUMP: Exploiting position diversity in the staggered TDMA underwater MAC protocol", in *Proc. IEEE Conf. on Comput. Communic. Infocom*, Rio de Janeiro, Brazil, April 2009, pp. 2961–2965.

[24] I. Holyer, "The NP-completeness of edge-colouring", *SIAM J. Comput.*, vol. 10, no. 4, pp. 718–720, 1981.

[25] M. Sun *et al.*, "Novel hysteretic noisy chaotic neural network for broadcast scheduling problems in packet radio networks", *IEEE Transact. on Neural Networks*, vol. 21, no. 9, pp. 1422–1433, 2010.

[26] P. Moscato, "Memetic Algorithms: A Short Introduction New Ideas in Optimization", UK Maidenhead: MCGraw-Hill, 1999, pp. 219–234.

[27] R. Dawkins, "The Selfish Gene", Oxford: Clarendon Press, 1976.

[28] Y.-D. Chen, C.-Y. Lien, S.-W. Chuang, and K.-P. Shih, "DSSS: A TDMA-based MAC protocol with dynamic slot scheduling strategy for underwater acoustic sensor networks", in *Proc. IEEE OCEANS*, Santander, Cantabria, Spain, June 2011, pp. 1–6 (doi: 10.1109/Oceans-Spain.2011.6003632).

V. Sivakumar received his B.Eng. and M.Eng. degrees from Anna University, Chennai. Now he is pursuing Ph.D. from VIT, Chennai. He has nearly 5 years of experience in teaching. He is currently working as an Assistant Professor in SRM Institute of Science and Technology (Deemed to be University), Chennai. His research interests include underwater wireless sensor networks and underwater Internet of things.
E-mail: sivakumarvit2013@gmail.com
School of Computing Science and Engineering
VIT
Chennai, Tamil Nadu, India

D. Rekha is currently working as Senior Assistant Professor in VIT (Deemed to be University), Chennai. She has more than 14 years of experience which includes 10 years of teaching and 4 years of Research. She is also the Division Chair of the Cyber Physical System Research Group, her area of research includes IoT, smart grid, wireless sensor networks, multi-hop networks, evolutionary algorithms, mobile cloud computing and Internet of things. She has published many papers in reputed journals. She has served as General and Technical Program Chair of numerous conferences.
E-mail: rekha.d@vit.ac.in
School of Computing Science and Engineering
VIT
Chennai, Tamil Nadu, India

Information for Authors

Journal of Telecommunications and Information Technology **(JTIT)** is published quarterly. It comprises original contributions, dealing with a wide range of topics related to telecommunications and information technology. **All papers are subject to peer review**. Topics presented in the JTIT report primary and/or experimental research results, which advance the base of scientific and technological knowledge about telecommunications and information technology.

JTIT is dedicated to publishing research results which advance the level of current research or add to the understanding of problems related to modulation and signal design, wireless communications, optical communications and photonic systems, voice communications devices, image and signal processing, transmission systems, network architecture, coding and communication theory, as well as information technology.

Suitable research-related papers should hold the potential to advance the technological base of telecommunications and information technology. Tutorial and review papers are published only by invitation.

Manuscript. TEX and LATEX are preferable, standard Microsoft Word format (.doc) is acceptable. The authors JTIT LATEX style file is available:
http://www.nit.eu/for-authors

Papers published should contain up to 10 printed pages in LATEX authors style (Word processor one printed page corresponds approximately to 6000 characters).

The manuscript should include an abstract about 150200 words long and the relevant keywords. The abstract should contain statement of the problem, assumptions and methodology, results and conclusion or discussion on the importance of the results. Abstracts must not include mathematical expressions or bibliographic references.

Keywords should not repeat the title of the manuscript. About four keywords or phrases in alphabetical order should be used, separated by commas.

The original files accompanied with pdf file should be submitted by e-mail: redakcja@itl.waw.pl

Figures, tables and photographs. Original figures should be submitted. Drawings in Corel Draw and PostScript formats are preferred. Figure captions should be placed below the figures and can not be included as a part of the figure. Each figure should be submitted as a separated graphic file, in .cdr, .eps, .ps, .png or .tif format. Tables and figures should be numbered consecutively with Arabic numerals.

Each photograph with minimum 300 dpi resolution should be delivered in electronic formats (TIFF, JPG or PNG) as a separated file.

References. All references should be marked in the text by Arabic numerals in square brackets and listed at the end of the paper in order of their appearance in the text, including exclusively publications cited inside. Samples of correct formats for various types of references are presented below:

[1] Y. Namihira, Relationship between nonlinear effective area and mode field diameter for dispersion shifted fibres, *Electron. Lett*., vol. 30, no. 3, pp. 262264, 1994.

[2] C. Kittel, *Introduction to Solid State Physics*. New York: Wiley, 1986.

[3] S. Demri and E. Orłowska, Informational representability: Abstract models versus concrete models, in *Fuzzy Sets, Logics and Knowledge-Based Reasoning*, D. Dubois and H. Prade, Eds. Dordrecht: Kluwer, 1999, pp. 301314.

Biographies and photographs of authors. A brief professional authors biography of up to 200 words and a photo of each author should be included with the manuscript.

Galley proofs. Authors should return proofs as a list of corrections as soon as possible. In other cases, the article will be proof-read against manuscript by the editor and printed without the author's corrections. Remarks to the errata should be provided within one week after receiving the offprint.

A copy of the JTIT is provided to each author of paper published.

Journal of Telecommunications and Information Technology has entered into an electronic licencing relationship with EBSCO Publishing, the worlds most prolific aggregator of full text journals, magazines and other sources. The text of *Journal of Telecommunications and Information Technology* can be found on EBSCO Publishings databases. For more information on EBSCO Publishing, please visit www.epnet.com.

(*Contents Continued from Front Cover*)

INSTYTUT ŁĄCZNOŚCI
PAŃSTWOWY INSTYTUT BADAWCZY

Editorial Office

National Institute
of Telecommunications
Szachowa st 1
04-894 Warsaw, Poland

tel. +48 22 512 81 83
fax: +48 22 512 84 00
e-mail: redakcja@itl.waw.pl
http://www.nit.eu

JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY 1/2018