Journal of Information & Communications

ISSN 1859 - 3534

MINISTRY OF INFORMATION AND COMMUNICATIONS

Volume 2019
Number 1
September

Research and Development on
Information & Communication Technology

1

Các công trình nghiên cứu phát triển
Công nghệ Thông tin và Truyền thông

Journal of Information & Communications
MINISTRY OF INFORMATION AND COMMUNICATIONS
ISSN: 1859-3534
VOLUME 2019, NUMBER 1, SEPTEMBER

JOURNAL ADDRESS:

115 Trần Duy Hưng street, Hanoi, Vietnam
Tel: (84)-24-37737136
Fax: (84)-24-37737130, Website http://ictmag.vn
Email: ict.research@mic.gov.vn (editorial) chuyensanbcvt@mic.gov.vn (administrative)

Journal Office Branch

27 Nguyễn Bỉnh Khiêm Street, District 1, Hồ Chí Minh City, Vietnam
Tel/Fax: (84)-28-38992898

JOURNAL MARKETING AND DISTRIBUTION

Phạm Thị Lâm
Email: ptlam@mic.gov.vn, Tel: 84-948503999

License No 365/GP-BTTTT, dated 19 December 2014; Modified license No 233/GP-BTTTT dated 23 May 2017.
Printed by Hanoi Printing Stock Company.
Printed and delivered in Sep 2019.

Price: 30,000 VND

EDITORIAL

Dear Readers,

The year 2019 marks the 20-th anniversary of our journal of Research and Development on Information and Communication Technology – ICT Research, in short. Since the last 20 years, we have kept our eyes and ears to the world scientific trends and developments in the field of ICT. Again and again, The purpose of ICT Research is to provide a forum for researchers and professionals to disseminate original and innovative ideas in the fields of information technology, communications, and electronics in Vietnam and worldwide.

Now, for a moment, imagine what has been for the past 20 years, every single day for over 7,000 days. Many people are responsible for creating and sustaining ICT Research. Without kind support and invaluable contribution of readers and authors, and hard work of the anonymous reviewers and editors under the former editorship of Prof. Nguyễn Thúc Hải – the very first Technical Editor-in-chief–, Prof. Trần Văn Lộc and Prof. Nguyễn Cảnh Tuấn, ICT Research would not be what it is today.

Since the first issue published in September 1999, 57 printed issues (41 issues in Vietnamese and 16 issues in English) have been delivered to the reader of ICT Research, nicely matched with 57 years of construction and development of its mother journal – the Journal of Information and Communication, which is a scientific entity of Vietnam Ministry of Information and Communications.

Now, it's the time for our generation to take over the duty of developing ICT Research to become a truly international journal. The current editor board, under the complementary technical editorship of Prof. Đỗ Ngọc Minh (University of Illinois at Urbana-Champaign, United States), Prof. Nguyễn Hoàng Hà (University of Saskatchewan, Canada) and Prof. Nguyễn Linh Trung (Vietnam National University, Hanoi), assume responsibility to make ICT Research indexed in SCOPUS by the end of 2020.

To contribute to the development of research in Vietnam, toward standard practices, high quality and international visibility, ICT Research has been taking measures by following current practices of prestigious international research journals. The editorial board of ICT Research is currently being extended to over 30 scientists, one quarter of them are international prominent scientists, thus forming a team of international area editors.

All manuscripts are reviewed and edited strictly according to international scientific editorial procedure. Each submission is assigned to an ICT Research area editor who then coordinates the review process and makes editorial decision. For improved paper quality in terms of organization and presentation, authors are guided to good practice of technical paper writing. In addition, accepted submissions are copy-edited by our editors and laid out using LaTeX.

Since June 2019, ICT Research has upgraded to the newest version of online journal management and publishing system – the OJS 3.1 (http://ictmag.vn).

We would like to inform you some changes in publication. An accepted article will be immediately published online as Early Access, with a DOI number, having been copy-edited and laid out in compliance with ICT Research editing rules. Also, in our printed issues, we apply a new numbering method: [VOLUME by Year, NUMBER as the running number of the year, MONTH of publishing]. For example, this very issue is numbered as "Volume 2019, Number 1, September".

Taking the opportunity of informing the above changes, we would like to, again, express our sincere gratitude and appreciation to the readers, authors, reviewers and editors of ICT Research, and to the leadership of the Ministry of Information and Communications and its predecessors – the Directorate General of Posts and Telecommunications, the Ministry of Posts and Telecommunications – for their continued support and contribution to ICT Research.

We look forward to your comments and feedback for better developing the MIC ICT Research Journal for Vietnam.

Sincerely,

Vũ Chí Kiên, Editor-in-Chief

Research and Development on
Information and Communication Technology
http://www.ictmag.vn **Volume 2019, Number 1**
September

TABLE OF CONTENTS

Diagonal Space Time Block Coded Spatial Modulation 1
.... ***Nguyen Tien Dong, Tran Xuan Nam, Le Minh Tuan***

SSG - A Solution to Prevent Saturation Attack on the Data Plane and Control Plane in SDN/Openflow Networks for Vietnam's Urban Classification .... ***Dang Van Tuyen, Truong Thu Huong*** 8

Performance Analysis of Repeated Index Modulation with Coordinate Interleaving over Nakagami-m Fading Channel .... ***Le Thi Thanh Huyen, Tran Xuan Nam*** 23

Matchmaking for Multi-Cloud Marketplace Application 31
.... ***Huynh Hoang Long, Nguyen Huu Duc, Le Trong Vinh***

Performance Comparison of Dynamic Elastic Optical Networks with Optical Regeneration 43
.... ***Le Hai Chau, Dang Hoai Bac***

Effect of Multi-site Variabilities on Electrovestibulography: Environmental and Physical Factors ... 52
.... ***Brian J. Lithgow, Zahra Moussavi***

An Ensemble Co-Evolutionary based Algorithm for Classification Problems 62
.... ***Vu Van Truong, Bui Thu Lam, Nguyen Thanh Trung***

Journal of Information & Communications
MINISTRY OF INFORMATION AND COMMUNICATIONS

Research and Development on Information and Communication Technology (ICT Research)
http://ictmag.vn

CALL FOR PAPERS

SPECIAL ISSUE ON BIOINFORMATICS AND COMPUTATIONAL BIOLOGY (BCB 2019)

The Special Issue on Bioinformatics and Computational Biology (BCB 2019) in the journal of Research and Development on Information and Communication Technology (ICT Research) aims at bringing together researchers for exchange of new developments in all areas of bioinformatics and computational biology.

We invite original full papers on any topic related to bioinformatics and computational biology. The submitted papers have not been published or under consideration for publication in any other journal or conference. Topics of interest include, but are not limited to:

- Expression Data Analysis
- Biological Sequence Analysis
- Disease Gene Prioritization
- Protein-Protein Interaction Network
- Gene Regulatory Network
- Statistical Genetics
- Systems Biology
- Computational Proteomics
- Computational Structural Biology
- Large-scale Data Management
- Systems Pharmacology
- Human genomics
- Drug discovery
- Personalized Medicine/Pharmacogenomics

Editor

Lê Đức Hậu, Thuyloi University

Paper Submission

Manuscripts should be submitted online at the website of ICT Research http://ictmag.vn.
Please see https://ictmag.vn/cntt/about/submissions#onlineSubmissions for guidelines and information.

The submissions will be peer-reviewed for originality and scientific quality.

Important Dates

- Submission: October 15, 2019
- Notification of acceptance: November 20, 2019
- Final manuscript: November 30, 2019
- Publication: December, 2019

Contact

- Lê Đức Hậu, duchaule@tlu.edu.vn
- Lê Hoàng Sơn, sonlh@vnu.edu.vn
- Nguyễn Linh Trung, linhtrung@vnu.edu.vn

Diagonal Space Time Block Coded Spatial Modulation

Invited article

Nguyen Tien Dong[1], Tran Xuan Nam[1], Le Minh Tuan[2]
[1] Le Quy Don Technical University, Hanoi, Vietnam
[2] Mobifone R&D Center, Mobifone Corporation, Hanoi, Vietnam
Correspondence: Tran Xuan Nam, namtx@mta.edu.vn
Communication: received 27 October 2018, revised 24 December 2018, accepted 26 December 2018
Online early access: 28 February 2019, Digital Object Identifier: 10.32913/mic-ict-research.v2019.n1.832
The Area Editor coordinating the review of this article and deciding to accept it was Dr. Truong Trung Kien

Abstract: In this paper, a new Spatial Modulation (SM) scheme, called Diagonal Space Time Coded Spatial Modulation (DS-SM), is designed by embedding the Diagonal Space Time Code in SM. The DS-SM scheme still inherits advantages of SM while enjoying further benefits from spatial constellation (SC) designs. Based on rank and determinant criteria, a new set of four SC codewords is proposed for the DS-SM system with four transmit antennas to achieve the fourth-order diversity. Then a general design procedure for an even number of transmit antennas, larger than four, is developed by cyclically shifting two rows of the SC codewords. Simulation results show that DS-SM surpasses several existing SM schemes at the same spectral efficiency and antenna configuration. DS-SM also exhibits better performance than the benchmark systems under spatially correlated channels. The complexity of DS-SM is also analyzed and compared to other SM schemes.

Keywords: *Multiple-input multiple-output (MIMO), space time block code, spatial modulation.*

I. Introduction

Spatial Modulation (SM), proposed by Mesleh *et al.* in [1], is a new transmission technique which can overcome many drawbacks of the conventional Multiple-Input Multiple-Output (MIMO) system. Different from the previous MIMO transmission schemes such as Vertical Bell-Labs Layered Space-Time (V-BLAST) or Space-Time Block Codes (STBC), the SM system activates only one transmit antenna during a time slot to transmit a modulated symbol. Thus, SM can avoid Inter-Channel Interference (ICI) among transmitted streams and does not require strict transmit antenna synchronization. Moreover, as only one radio frequency (RF) chain is used, SM is more advantageous than the other MIMO schemes in terms of energy saving. But the most important advantage of SM is that the spectral efficiency is increased as the antenna indices are utilized to convey information bits. However, SM lacks transmit diversity and multiple receive antennas are needed at the SM receivers to attenuate the fading effect.

Various efforts have been made to cope with the problem of channel fading and improve SM performance (see [2] and the references therein). Among these works only some solutions can help SM to increase its transmit diversity. In [3], the authors proposed the so-called Coherent Space Time Shift Keying (CSTSK) which achieves the second order transmit diversity. The Time-Orthogonal-Signal-Design Assisted Space Shift Keying (TOSD-SSK) proposed in [4] can also obtain the same diversity order of CSTSK by using shaping filters at the transmitter. In [5], Basar *et al.* proposed the so-called Space Time Block Coded Spatial Modulation (STBC-SM) by combining STBC and SM. By exploiting the orthogonal structure of the Alamouti STBC the STBC-SM scheme also achieve the second-order transmit diversity with low-complexity maximum-likelihood detection.

In [6], Le *et al.* introduced the concept of Spatial Constellation (SC) and proposed a high-rate Space-Time Block Coded Spatial Modulation (STBC-SM) scheme for four and six transmit antennas. This STBC-SM scheme has higher spectral efficiency than STBC-SM in [5] thanks to the increased number of spatial constellation matrices. In [7], based on cyclic structure and complex constellation rotation another SM scheme, abbreviated as STBC-CSM, was proposed to further improve the spectral efficiency of STBC-SM while still maintaining the second-order transmit diversity. In [8], an improved SM scheme called Spatially Modulated Orthogonal Space Time Block Coding (SM-OSTBC) was proposed. This scheme attains the maximum spectral efficiency of $(n_T - 2 + \log_2 M)$ bit per channel

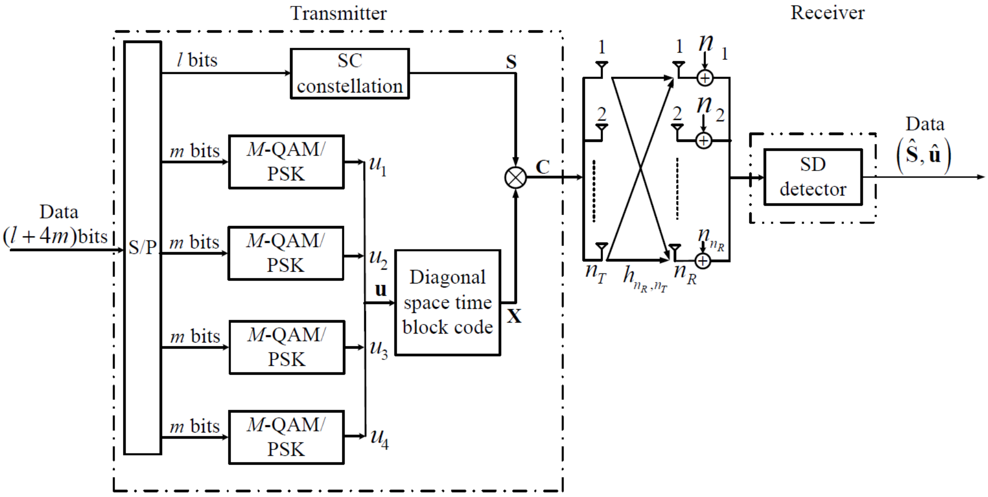


Figure 1. Block diagram of the DS-SM scheme.

use (bpcu) when the number of active antennas is equal the number of transmit antennas, i.e., $n_A = n_T$, where M is the modulation order. However, SM-OSTBC has a limitation that it is only applicable to the systems equipped with an even number of transmit antennas greater than or equal to four. To overcome this drawback, Wang *et al.* [9] proposed the so-called Spatially Modulated Diagonal Space Time Code (SM-DC) that can apply to the systems with the number of transmit antennas equal to or less than four. The SM-DC scheme also achieves the second-order transmit diversity.

The objective of the current paper is to improve the SM performance by increasing its transmit diversity order. Inspired by the concept of the SC matrices in [6] and the Diagonal STBC in [9], we propose an enhanced SM scheme by designing a new set of SC matrices and incorporating them with a Diagonal Space Time Block Code. The proposed scheme is referred to as DS-SM. Compared to SM-DC, the DS-SM scheme has the following advantages. First, our proposed DS-SM can apply to MIMO systems with an even number of transmit antennas greater than or equal to four. Second, we propose to use an optimal linear matrix to maximize the minimum product distance between any two points of the signal constellation. Finally, our scheme achieves the fourth-order transmit diversity in contrast to the second order by SM-DC.

Our contributions in this paper are as follows.

1) A new set of four SC codewords is proposed for the DS-SM system equipped with four transmit antennas to achieve the fourth-order transmit diversity.
2) A general procedure to design extended SC codewords is formulated for the DS-SM systems with an even number of transmit antennas greater than four.
3) The proposed scheme requires only one RF transmit chain, therefore eliminating the ICI effect and facilitating the IAS requirement.
4) Theoretical upper bound of the bit error probability (BEP) of the proposed scheme is derived to verify simulation results. The proposed scheme is demonstrated to surpass the related SM-based MIMO ones including SM-DC, STBC-SM, and STBC-CSM in both uncorrelated and correlated fading environments for the same antenna configuration and spectral efficiency.

The remainder of this paper is organized as follows. Section II presents the system model of the proposed DS-SM scheme. Section III describes the SC codeword design followed by the signal detection introduced in Section IV. Performance evaluation is presented in Section V, and finally conclusions are drawn in Section VI.

Notation: The following mathematical notations are used throughout the paper. $(\cdot)^T$ and $(\cdot)^H$ denote vector/matrix transpose and conjugate transpose, respectively. $\Re(\cdot)$ and $\Im(\cdot)$ denote the real and the imaginary part of a complex number, respectively. $\mathrm{vec}(\mathbf{A})$ denotes the column-vectorial stacking operation of matrix $\mathbf{A}$. $\mathrm{diag}(\mathbf{x})$ denotes a diagonal matrix built from vector $\mathbf{x}$.

II. System Model

Figure 1 illustrates the block diagram of the proposed SM scheme with $n_T = 4$ transmit antennas and n_R receive antennas. It is assumed that data bits arrive at the transmitter in blocks each of which consists of $(l + 4m)$ bits. The first l bits are mapped into a 4×4 SC matrix out of $K = 2^l$ SC matrices in the spatial constellation Ω_S. The remaining $4m$ bits are modulated by M-QAM/PSK modulators, where $M = 2^m$, to make a 4×1 modulated symbol vector

$\mathbf{u} = \begin{bmatrix} u_1 & u_2 & u_3 & u_4 \end{bmatrix}^T$. Using DSTBC, $\mathbf{u}$ is linearly combined with a rotation matrix $\mathbf{W}$ as follows:

$$\tilde{\mathbf{u}} = \mathbf{W}[\Re(\mathbf{u}), \Im(\mathbf{u})]^T. \tag{1}$$

The multidimensional rotation matrix $\mathbf{W}$ [10] for maximizing the minimum product distance between any two points of the signal constellation is given as

$$\mathbf{W} = \sqrt{\frac{2}{n}} \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix}, \tag{2}$$

where $w_{tk} = \cos\left(\frac{\pi}{4n}(4t-1)(2k-1)\right)$, for $1 \le t,\ k \le n$, $n = 2n_T = 8$, and $\sqrt{2/n}$ is used to normalize $\mathbf{W}$.

Then, the elements of the rotated signal vector $\tilde{\mathbf{u}}$ are rearranged to obtain the following vector

$$\tilde{\mathbf{x}} = \begin{bmatrix} \tilde{u}_1 + j\tilde{u}_5 & \tilde{u}_2 + j\tilde{u}_6 & \tilde{u}_3 + j\tilde{u}_7 & \tilde{u}_4 + j\tilde{u}_8 \end{bmatrix}^T. \tag{3}$$

The 4×4 diagonal STBC matrix is then obtained as $\mathbf{X} = \mathrm{diag}(\tilde{\mathbf{x}})$. Finally, the $4{\times}4$ transmitted codeword $\mathbf{C}$ is created simply by multiplying $\mathbf{S}$ by $\mathbf{X}$, i.e., $\mathbf{C} = \mathbf{SX}$. This resulted codeword $\mathbf{C}$ will be transmitted from four transmit antennas within $T = 4$ periods.

Under the assumption that the channel is quasi-static and flat fading, the $n_R \times 4$ received signal matrix $\mathbf{Y}$ at the receiver is given by

$$\mathbf{Y} = \sqrt{\frac{\gamma}{E_s}}\mathbf{HC} + \mathbf{N} = \sqrt{\frac{\gamma}{E_s}}\mathbf{HSX} + \mathbf{N}, \tag{4}$$

where γ is the average signal-to-noise ratio (SNR) at each receive antenna. E_s is the average energy of the M-QAM/PSK modulated symbols, $\mathbf{H}$ and $\mathbf{N}$ respectively denote an $n_R \times 4$ channel matrix and an $n_R \times 4$ noise matrix whose entries are assumed to be independent and identically distributed (i.i.d.) random variables with zero mean and unit variance.

III. SC Codeword Design

1. Basic SC Codewords for Four Transmit Antennas

As seen in Figure 1, a part of data bits are conveyed by SC codewords, so these codewords should be carefully designed. Based on the SC concept [6], a basic set of four SC codewords is proposed as follows:

$$\mathbf{S}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{5a}$$

$$\mathbf{S}_2 = \begin{bmatrix} 0 & e^{j\theta} & 0 & 0 \\ 0 & 0 & e^{j\theta} & 0 \\ 0 & 0 & 0 & e^{j\theta} \\ j & 0 & 0 & 0 \end{bmatrix}, \tag{5b}$$

$$\mathbf{S}_3 = \mathbf{S}_2^2, \tag{5c}$$

$$\mathbf{S_4} = \mathbf{S}_2^3. \tag{5d}$$

Using the rank and determinant criteria [11], the rotation angle θ is optimized to attain full diversity and maximum coding gain. Particularly, an exhaustive search is applied to find the optimal value of the angle $\theta \in [0, \pi/2]$ that maximizes the coding gain distance (CGD) $\delta_{\min}(\theta)$ as follows:

$$\delta_{\min} = \min_{\mathbf{C} \neq \mathbf{C}'} \det(\mathbf{C} - \mathbf{C}')^H (\mathbf{C} - \mathbf{C}'), \tag{6}$$

$$\theta_o = \arg\max_{\theta} \delta_{\min}(\theta). \tag{7}$$

Table I presents the resulted optimal angle θ_o and CGD for different modulation techniques. The spectral efficiency of DS-SM equipped with four transmit antennas is given by

$$C_{\text{DS-SM}} = \frac{1}{4}\left(\log_2 4 + \log_2 M^4\right) = \frac{1}{2} + \log_2 M \,(\text{bpcu}). \tag{8}$$

2. SC Codeword Design for an Even Number of Transmit Antennas

Based on the above basic set of SC codewords for four antennas, an extended set of SC codewords are constructed for DS-SM with even number of transmit antennas. For illustration purpose, a DS-SM system for six transmit antennas is considered. From the first SC codeword $\mathbf{S}^1 = \begin{bmatrix} \mathbf{s}_{1,1} & \mathbf{s}_{1,2} \end{bmatrix}^T$, adding a $2{\times}4$ zero matrix below the columns of the $\mathbf{S}^1$ matrix, three SC codewords are generated by cyclically shifting two rows of the new matrix as follows:

$$\mathbf{S}_1^1 = \begin{bmatrix} \mathbf{s}_{1,1} \\ \mathbf{s}_{1,2} \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{S}_2^1 = \begin{bmatrix} \mathbf{0} \\ \mathbf{s}_{1,1} \\ \mathbf{s}_{1,2} \end{bmatrix}, \quad \mathbf{S}_3^1 = \begin{bmatrix} \mathbf{s}_{1,2} \\ \mathbf{0} \\ \mathbf{s}_{1,1} \end{bmatrix}. \tag{9}$$

TABLE I
Optimal values of θ and corresponding CGDs for the basic SC codewords

Modulation	BPSK	4QAM	8QAM	16QAM
θ	0.52	1.36	0.2	0.4
$\delta_{\min}$	0.11	0.037	3.6×10^{-3}	7×10^{-4}

When the number of transmit antennas is even and greater than four, the number of SC matrices increase to $2n_T$. As a result, the number of data bits conveyed by the SC matrices are $\lfloor \log_2(2n_T) \rfloor$ bits while the number of data bits carried by modulated signals are $4m$ bits ($M = 2^m$). Therefore, the spectral efficiency of DS-SM is given by

$$C_{\text{DS-SM}} = \frac{4m + \lfloor \log_2(2n_T) \rfloor}{T} = m + \frac{\lfloor \log_2(2n_T) \rfloor}{4} \text{ (bpcu)}. \quad (10)$$

Comparing to SM-DC in [9] which has the spectral efficiency of $\log_2 M_1 + 2\log_2 M$ over $T = 2$ symbol periods resulting in the actual spectral efficiency of $\frac{\log_2 M_1}{2} + m$, DS-SM can become competitive in the systems with large number of transmit antennas and small modulation size. Note that the spectral efficiency of SM-DC does not depend on the number of transmit antennas n_T but the modulation order M_1 of the symbols in the spatial codewords. Thus, SM-DC can easily achieve higher spectral efficiency than DS-SM when the number of transmit antennas is small, e.g. $n_T < 4$, by increasing M_1. However, this increase is more vulnerable to spatial codeword errors, which sacrifices BER performance of SM-DC as demonstrated by simulation results in the later section.

In summary, a general procedure to design the SC codeword for a DS-SM system with an even number of transmit antennas is given as follows:

1) For a given number of even transmit antennas n_T and an arbitrary level of modulation M, using Table I to choose the suitable value of θ and generate the basic set of four SC codewords.
2) Adding an $(n_T - 4) \times 4$ zero matrix under the columns of the basic SC codewords and cyclically shifting two rows of these matrices to increase the number of SC codewords to $2n_T$.
3) Generating a new set of the SC codewords as $\mathbf{S}_k$, for all $k = 1, 2, \ldots, \lfloor \log_2(2n_T) \rfloor$.

IV. Signal Detection

For a given matrix $\mathbf{S}_k$, $k = 1, 2, \ldots, K$, we can construct the $n_R \times 4$ equivalent matrix $\tilde{\mathbf{H}}_k = \sqrt{\gamma/E_s}\mathbf{H}\mathbf{S}_k$. Therefore, the system equation in (4) can be re-written as

$$\mathbf{Y} = \tilde{\mathbf{H}}_k \mathbf{X} + \mathbf{N}. \quad (11)$$

Based on the diagonal structure of $\mathbf{X}$, Equation (11) can be represented as follows:

$$\mathbf{y} = \mathbf{H}_{e,k}\tilde{\mathbf{x}} + \mathbf{n}, \quad (12)$$

where $\mathbf{H}_{e,k} = \left[\text{diag}\left(\tilde{\mathbf{h}}_1\right), \ldots, \text{diag}\left(\tilde{\mathbf{h}}_{n_R}\right)\right]^T$, $\tilde{\mathbf{h}}_k$ is the k-row of $\tilde{\mathbf{H}}_k$, $k = 1, 2, \ldots, n_R$, $\mathbf{y} = \text{vec}\left(\mathbf{Y}^T\right)$, and $\mathbf{n} = \text{vec}\left(\mathbf{N}^T\right)$. Converting equation (12) into the equivalent real system-equation and using (1) and (4), we have

$$\mathbf{v} = \mathbf{M}_k \mathbf{s} + \mathbf{w}, \quad (13)$$

where

$$\mathbf{s} = \left[\Re(\mathbf{u}) \quad \Im(\mathbf{u})\right]^T,$$
$$\mathbf{w} = \left[\Re(\mathbf{n}) \quad \Im(\mathbf{n})\right]^T,$$
$$\mathbf{v} = \left[\Re(\mathbf{y}) \quad \Im(\mathbf{y})\right]^T,$$
$$\mathbf{M}_k = \begin{bmatrix} \Re\left(\mathbf{H}_{e,k}\right) & -\Im\left(\mathbf{H}_{e,k}\right) \\ \Im\left(\mathbf{H}_{e,k}\right) & \Re\left(\mathbf{H}_{e,k}\right) \end{bmatrix} \mathbf{W}.$$

The system equation in (13) is similar to that of a conventional spatial multiplexing scheme. Therefore, a Sphere Decoder (SD) in [12, 13] can be used to detect $\mathbf{s}$, as follows:

$$(\hat{\mathbf{s}})_k = \arg\min_{\mathbf{s}} \|\mathbf{t}_k - \mathbf{R}_k \mathbf{s}\|^2, \quad (14)$$

where $\mathbf{t}_q = \mathbf{Q}_k^H \mathbf{v}$, $\mathbf{Q}_k$ and $\mathbf{R}_k$ are the resulting matrices from the QR decomposition of $\mathbf{M}_k$, i.e., $\mathbf{M}_k = \mathbf{Q}_k \mathbf{R}_k$.

The index k of the transmitted SC codeword is then determined as [8]

$$\hat{k} = \arg\min_k \|\mathbf{t}_k - \mathbf{R}_k(\hat{\mathbf{s}})_k\|^2 + \mathbf{v}^H\mathbf{v} - \mathbf{t}_k^H \mathbf{t}_k. \quad (15)$$

Finally, the transmitted information bits are recovered from a pair of the detected SC codeword and the detected signal vector $(\hat{\mathbf{S}}_\mathbf{k}, \hat{\mathbf{u}}_\mathbf{k})$ at the receiver.

1. Complexity Analysis

In this section, computational complexity for signal processing at the DS-SM receivers is analyzed and compared with related SM-based MIMO schemes using SD [12, 13]. It is assumed that each real arithmetic calculation accounts for a floating point operation (flop). Therefore, a complex addition or subtraction requires two flops while a complex multiplication requires six operations including four real multiplications and two real additions. We also assume that the channel remains unchanged within T-symbol periods.

In the pre-processing state, the complexity of computing $\tilde{\mathbf{H}}_k$ in (11), $\mathbf{M}_k$ in (13), QR decomposition of $\mathbf{M}_k$ in (13) and a signal vector $\mathbf{t}_k, k = 1, 2, \ldots, K$, in (15) is given as

$$\Delta_{\text{pre}} = \frac{4}{T}\left(2048 n_R + 3n_T^2 n_R + 6 n_R n_T - 34\right) K + (143 n_R + 7) K. \quad (16)$$

Therefore, the complexity of DS-SM is given by

$$\Delta = \frac{\Delta_{\text{pre}} + \Delta_S}{4m + 2}, \quad (17)$$

where Δ_S is the average number of operations used in the SD searching stage.

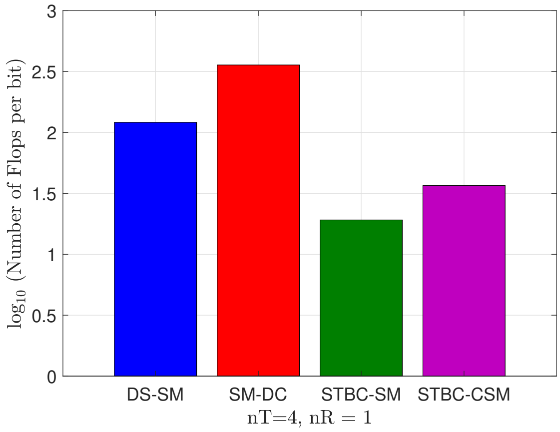


Figure 2. Complexity comparison of DS-SM, SM-DC, STBC-SM, STBC-CSM at the spectral efficiency 3.5 bpcu, SNR 9 dB, 4 transmit antennas, 1 receive antenna, $T = 80$ symbol periods.

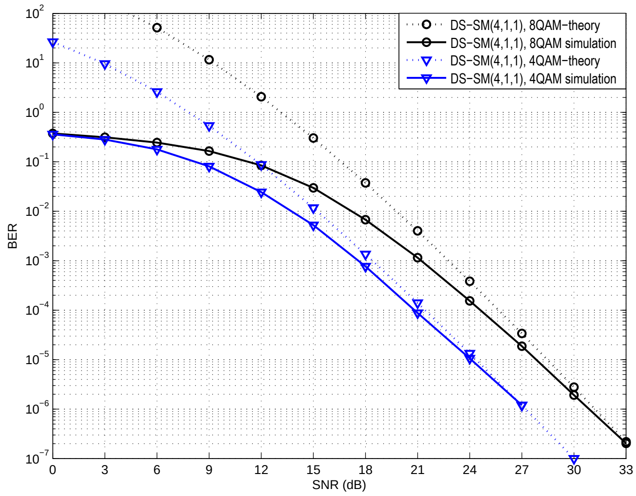


Figure 3. Simulated average BER and theoretical BEP of DS-SM.

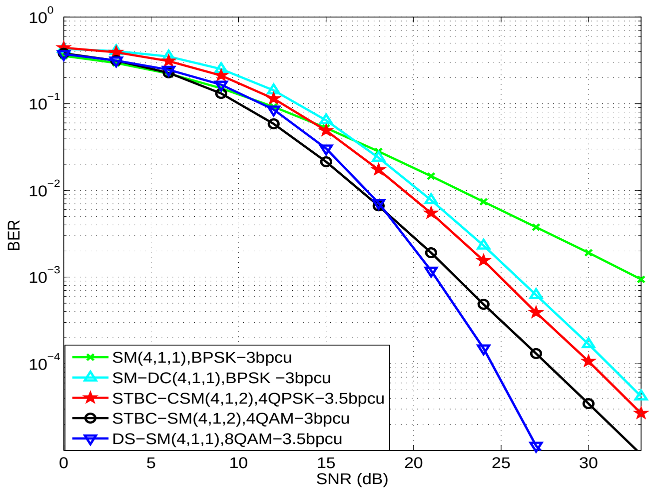


Figure 4. BER performance of DS-SM, SM, STC-SM, SM-DC, and STBC-CSM when $n_R = 1$ and spectral efficiency is 3 bpcu.

Figure 2 compares the detection complexity of the DS-SM to the related SM schemes such as SM-DC, STBC-SM, and STBC-CSM, in a MIMO scheme with $n_T = 4$, $n_R = 1$, and at a spectral efficiency of 3.5 bpcu. It can be seen that the complexity of DS-SM is larger than those of STBC-SM and STBC-CSM but lower than that of SM-DC.

2. Theoretical Upper Bound for BEP

Based on the pairwise error probability (PEP), the upper bound of BEP for the DS-SM scheme can be derived. By definition, PEP, denoted by $P(\mathbf{C}_i \to \mathbf{C}_j)$, is the probability that a matrix $\mathbf{C}_i$ is transmitted while the receiver mistakenly decides it by another matrix $\mathbf{C}_j$. The upper bound of PEP is given by [14]

$$P_b \leq \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{P(\mathbf{C}_i \to \mathbf{C}_j)w_{i,j}}{\log_2 N}, \tag{18}$$

where $N = KM^4$ and $w_{i,j}$ is the number of bits in error between the matrices $\mathbf{C}_i$ and $\mathbf{C}_j$.

The conditional PEP of the DS-SM system is given by

$$P\left(\mathbf{C}_i \to \mathbf{C}_j|\mathbf{H}\right) = Q\left(\sqrt{\frac{\gamma}{2}d^2\left(\mathbf{C}_i, \mathbf{C}_j\right)}\right), \tag{19}$$

where $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^{\infty} e^{-y^2/2}dy$. From [15], PEP is given by

$$P(\mathbf{C}_i \to \mathbf{C}_j) = \frac{1}{\pi}\int_0^{\pi}\left[\left(\frac{1}{1+\frac{\gamma\lambda_{i,j,1}}{4\sin^2\phi}}\right)^{n_R}\left(\frac{1}{1+\frac{\gamma\lambda_{i,j,2}}{4\sin^2\phi}}\right)^{n_R} \times\left(\frac{1}{1+\frac{\gamma\lambda_{i,j,3}}{4\sin^2\phi}}\right)^{n_R}\left(\frac{1}{1+\frac{\gamma\lambda_{i,j,4}}{4\sin^2\phi}}\right)^{n_R}\right]d\phi. \tag{20}$$

V. Simulation Results

In this section, performance of the proposed SM is evaluated and compared with several existing SM systems such as SM, SM-DC, STBC-SM and STBC-CSM for the case using different modulation techniques. The MIMO configuration with number of transmit antennas n_T, receive antennas n_R, and active antennas n_A in each scheme is denoted by (n_T, n_R, n_A). Furthermore, it is assumed that all schemes employ SD for signal detection.

1. Performance Evaluation Under Uncorrelated Channel

Figure 3 illustrates the BER obtained using the theoretical analysis and simulations for DS-SM in a DS-SM(4, 1, 1) system. Two modulation schemes, i.e. 4-QAM and 8-QAM,

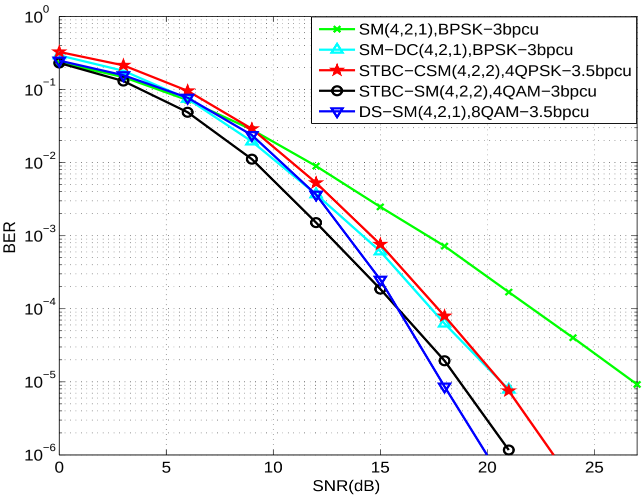


Figure 5. BER performance of DS-SM, SM, STC-SM, SM-DC, and STBC-CSM when $n_R = 2$ and spectral efficiency is 3 bpcu.

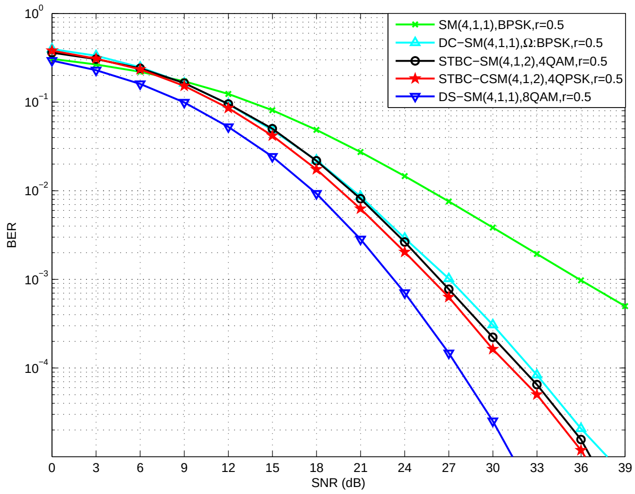


Figure 6. BER performance of DS-SM, SM, SM-DC, STBC-SM, STBC-CSM (4,1,1) at spectral efficiency of 3 bpcu and $r = 0.5$.

are used for evaluation. Note from the figure that simulation and analytical results are coincident at the high SNR region, validating the tightness of our analytical results.

Figures 4 and 5 compare performance of DS-SM$(4, n_R, 1)$ to SM$(4, n_R, 1)$, SM-DC$(4, n_R, 1)$, STBC-SM$(4, n_R, 2)$, and STBC-CSM$(4, n_R, 2)$, when being equipped with $n_R = 1$ and $n_R = 2$ receive antennas. It can be seen that DS-SM outperforms the existing schemes at high SNR region. Specifically, in Figure 4 at BER $= 10^{-3}$, DS-SM achieves SNR gains of 1.1 dB, 2.7 dB, 4.7 dB, and 11.5 dB compared with STBC-SM, STBC-CSM, SM-DC, and SM, respectively. These gaps become smaller in Figure 5 as all the curves have deeper slopes. At the BER of 10^{-5}, DS-SM provides SNR gains of 0.8 dB, 2.5 dB, and 18 dB over STBC-SM, STBC-CSM, and SM-DC, and SM, respectively. At low SNR region, DS-SM exhibits small degradation in SNR gain compared with STBC-SM. However, it is worth mentioning that DS-SM still has 0.5 bpcu higher spectral efficiency and requires one less RF chain compared with STBC-SM.

2. Performance Evaluation Under Spatially Correlated Channel

In order to evaluate performance of DS-SM under spatially correlated channels we use a modified channel matrix with spatial correlation effect, which is given by [16]

$$\bar{\mathbf{H}} = \mathbf{R}_R^{1/2}\mathbf{H}\mathbf{R}_T^{1/2}. \tag{21}$$

where $(n_T \times n_T)\,\mathbf{R}_T$ and $(n_R \times n_R)\,\mathbf{R}_R$ are the transmit and receive spatial correlation matrix, respectively. Each element of these matrices is generated using the exponential correlation matrix model [17], $r_{ij} = r_{ji}^*$ for $i \le j$ where r is the correlation coefficient of the neighboring transmit and receive antennas. This is an appropriate and common model to evaluate performance of MIMO systems under the effect of spatial correlation at both transmitter and receiver.

Figure 6 illustrates the BER performance of DS-SM versus those of SM, SM-DC, STBC-SM, and STBC-CSM, all equipped with 1 transmit and 1 receive antenna. The spectral efficiency of all schemes is 3 bpcu and the correlation coefficient is $r = 0.5$. It can be seen from the figure that DS-SM is more robust than all other schemes under spatial correlation effect. For example, at BER $= 10^{-3}$, DS-SM offers about 2.5 dB, 3 dB, 3.7 dB, and 12 dB SNR gain over STBC-CSM, STBC-SM, SM-DC, and SM, respectively.

VI. Conclusions

In this paper, we have proposed a new MIMO scheme, called DS-SM, by embedding the Diagonal STBC into the SM system and using the rank and determinant criteria to optimize its spatial codewords. The proposed DS-SM scheme still inherits promising benefits of SM including ICI avoidance and elimination of IAS while enjoying full diversity provided by STBC. It achieves significant performance improvement over the existing SM schemes including SM, STBC-SM, SM-DC, and STBC-CSM, especially at high SNR region due to higher diversity gain, at the cost of small additional complexity. Particularly, the proposed DS-SM scheme is more robust than the benchmark schemes under the spatially correlated fading channels.

Acknowledgement

This work is sponsored by National Foundation for Science and Technology Development (NAFOSTED) under project number 102.02-2015.23.

REFERENCES

[1] R. Mesleh, H. Haas, C. W. Ahn, and S. Yun, "Spatial modulation - A new low complexity spectral efficiency enhancing technique," in *Proceedings of the First International Conference on Communications and Networking in China*. IEEE, 2006, pp. 1–5.

[2] M. D. Renzo, H. Haas, A. Ghrayeb, S. Sugiura, and L. Hanzo, "Spatial modulation for generalized MIMO: challenges, opportunities and implementation," *Proceedings of the IEEE*, vol. 102, no. 1, pp. 56–103, 2014.

[3] S. Sugiura, S. Chen, and L. Hanzo, "Coherent and differential space-time shift keying: A dispersion matrix approach," *IEEE Transactions on Communications*, vol. 58, no. 11, pp. 3219–3230, 2010.

[4] M. Di Renzo and H. Haas, "Space shift keying (SSK) MIMO over correlated Rician fading channels: Performance analysis and a new method for transmit-diversity," *IEEE Transactions on Communications*, vol. 59, no. 1, pp. 116–129, 2011.

[5] E. Başar, Ü. Aygölü, E. Panayırcı, and H. V. Poor, "Space-time block coding for spatial modulation," in *Proceedings of the 21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*. IEEE, 2010, pp. 803–808.

[6] M.-T. Le, V.-D. Ngo, H.-A. Mai, and X. N. Tran, "High-rate space-time block coded spatial modulation," in *Proc. 2012 Int'l Conf. Advanced Technol. Commun.,*, 2012, pp. 278–282.

[7] X. Li and L. Wang, "High rate space-time block coded spatial modulation with cyclic structure," *IEEE Communications Letters*, vol. 18, no. 4, pp. 532–535, 2014.

[8] M.-T. Le, V.-D. Ngo, H.-A. Mai, X. N. Tran, and M. Di Renzo, "Spatially modulated orthogonal space-time block codes with non-vanishing determinants," *IEEE Transactions on Communications*, vol. 62, no. 1, pp. 85–99, 2014.

[9] L. Wang and Z. Chen, "Spatially modulated diagonal space time codes," *IEEE Communications Letters*, vol. 19, no. 7, pp. 1245–1248, 2015.

[10] J.-C. Belfiore, X. Giraud, and J. Rodriguez-Guisantes, "Optimal linear labelling for the minimization of both source and channel distortion," in *Proceedings of the IEEE International Symposium on Information Theory*. IEEE, 2000, p. 404.

[11] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information theory*, vol. 45, no. 5, pp. 1456–1467, 1999.

[12] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Transactions on Information theory*, vol. 45, no. 5, pp. 1639–1642, 1999.

[13] M.-T. Le, V.-S. Pham, L. Mai, and G. Yoon, "Rate-one full-diversity quasi-orthogonal STBCs with low decoding complexity," *IEICE transactions on communications*, vol. 89, no. 12, pp. 3376–3385, 2006.

[14] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE transactions on information theory*, vol. 44, no. 2, pp. 744–765, 1998.

[15] M. S. Alaouni and M. K. Simon, *Digital Communications over Fading Channels*. John Wiley & Sons, 2002.

[16] A. Paulraj, R. Nabar, and D. Gore, *Introduction to space-time wireless communications*. Cambridge Univ. Press, 2003.

[17] S. L. Loyka, "Channel capacity of mimo architecture using the exponential correlation matrix," *IEEE Communications letters*, vol. 5, no. 9, pp. 369–371, 2001.

Nguyen Tien Dong was born in Quang Tri, Vietnam in 1982. He received his B.Eng. and M.Sc. degrees both in electronic engineering from Le Quy Don Technical University, Vietnam, in 2006 and 2012, respectively. From 2006 to 2010 he worked as a lecturer at the Telecommunication University, Vietnam. He is currently working toward his Ph.D. degree in electronic engineering at Le Quy Don Technical University.

Tran Xuan Nam is currently an associate professor and head of the research group in advanced wireless communications in Le Quy Don Technical University, Vietnam. He received his M.Eng. in Telecommunications Engineering from University of Technology Sydney, Australia in 1998, and Dr.Eng. in Electronic Engineering from The University of Electro-Communications, Japan in 2003. From November 2003 to March 2006 he was a research associate at the University of Electro-Communications, Tokyo, Japan. Dr. Tran research interests are in the areas of space-time signal processing for communications such as adaptive antennas, space-time coding, MIMO, spatial modulation and cooperative communications. Dr. Tran is a recipient of the 2003 IEEE AP-S Japan Chapter Young Engineer Award, and of two best papers from the 2012 International Conference on Advanced Technologies for Communications and the 2014 National Conference on Electronics, Communications and Information Technology. He is a member of IEEE, IEICE and the Radio-Electronics Association of Vietnam.

Le Minh Tuan received his B.E. degree in electronic engineering from Hanoi University of Science and Technology, Vietnam in 1999, M.Sc. and Ph.D. degrees in electrical engineering from the Korean Advanced Institute of Science and Technology (KAIST), Daejon, Korea, in 2003 and 2007, respectively. From 1999 to 2001 and from 2007 to 2008 he worked as a lecturer at Posts and Telecommunication Institute of Technology (PTIT), Vietnam. From November 2012 to 2015, he worked at Hanoi Department of Science and Technology, Vietnam. He is currently working at MobiFone Reasearch and Development Center, MobiFone Corporation, Vietnam. His research interests include space-time coding, space-time processing, and MIMO systems. Dr. Le is the recipient of the 2012 ATC Best Paper Award from the Radio Electronics Association of Vietnam (REV) and the IEEE Communications Society. He is a member of IEEE.

SSG - A Solution to Prevent Saturation Attack on the Data Plane and Control Plane in SDN/Openflow Networks

Dang Van Tuyen[1,2], Truong Thu Huong[1]
[1] Hanoi University of Science and Technology, Hanoi, Vietnam
[2] People's Police University of Technology and Logistics, Ministry of Public Security, Bac Ninh, Vietnam
Correspondence: Truong Thu Huong, huong.truongthu@hust.edu.vn
Communication: received 3 November 2018, revised 25 January 2019, accepted 31 January 2019
Online early access: 28 February 2019, Digital Object Identifier: 10.32913/mic-ict-research.v2019.n1.833
The Area Editor coordinating the review of this article and deciding to accept it was Dr. Truong Trung Kien

Abstract: The SDN/Openflow architecture opens new opportunities for effective solutions to address network security problems; however, it also brings new security challenges compared to the traditional network architectures. One of these challenges is that the mechanism of reactive installation for new flow entries can make the data plane and control plane easily become a target for resource saturation attacks with spoofing techniques such as SYN flood. There are a number of solutions to this problem, e.g., the Connection Migration (CM) mechanism in Avant-Guard solution. Nevertheless, most of the solutions increase the load at commodity switches and/or split benign TCP connections, which can increase the packet latency and disable some features of the TCP protocol. This paper presents a solution, referred to as SDN-based SYN Flood Guard (SSG), which takes advantages of the OpenFlow's abilities to match TCP Flags fields and the RST Cookie technique to authenticate the three-way handshake process of TCP connections in a separated device from SDN/Openflow switches. Experiment results reveal that SSG solves the aforementioned problems and improves the SYN Flood attack tolerance compared to the existing solutions.

Keywords: *OpenFlow, SDN, DDoS attack, saturation attack, TCP SYN flood, SYN cookie, RST cookie.*

I. Introduction

Recently, Software Defined Networking (SDN) [1] has attracted interest in academia and industry as a promising network technology. The philosophy of separating control plane from data plane allows SDN researchers and operators to easily add new, creative, powerful network functions and protocols in order to optimize performance as well as reduce operational cost.

In an SDN network, the switches only handle the packet forwarding function based on the configuration by a controller. The network-control policies such as routing calculation, load balancing, and security rules are implemented by the controller and configuration information in the forwarding tables of the switches via a security interface.

OpenFlow [2, 3] is generally regarded as an SDN reference implementation that defines the structure of forwarding tables at the SDN switches and the messages exchanged between the switches and controllers. At an OpenFlow switch (OFS), network policies are specified into flow entries arranged in flow tables. In a flow entry (FE), there are criterion fields which will be matched with incoming packets for finding the corresponding appropriate policies. If an incoming packet is not matched with any existing flow rules, a *table-miss* event will be issued with a *packet_in* message sent to the controller to install a new FE for processing the packet. The FE will be kept in flow tables of the OFS for a period of time based on *idle-timeout* and *hard-timeout* settings or eviction commands by application running on controllers. With this mechanism of creating and maintaining flow rules, a *table-miss* consumes a specific amount of resources (includes of CPU, memory and bandwidth) in both data plane and control plane.

Exploiting this characteristic, attackers can implement Distributed Denial of Service (DDoS) attacks in order to flood, saturate OFSes and controllers by making a massive series of *table-miss* events and useless FEs. This may be easily carried out by sending packets with randomly forging some or all packet-header fields making them hard to match any existing FEs at a victim switch. If there are enough new flows, creating a flood of table-miss events in a short time, the resource on data plane and/or control plane will be quickly exhausted.

SYN Flood attacks have been known early in the Internet development and they remain a common form of DoS and DDoS attack [4]. Attackers often spoof the source IP addresses in SYN packets, making it difficult to detect and mitigate this kind of attacks. In the SDN/OpenFlow context, by forging randomly source IP addresses to create flood of *table-miss* events and useless FEs, SYN floods can be exploited to attack OFSes (data plane) and controllers (control plane) [5–7].

Plenty of research has been performed to address the saturation of the data and control plane. However, most focus on collecting traffic information in the data plane and then applying analysis algorithms at the control plane either to produce packet policies or to enhance security processing ability in the data plane. One drawback of these techniques is that they increase in OpenFlow traffic. To address SYN Flood attacks, in Avant-Guard solution [8], S. Shin *et al.* integrate into OFSes the Connection Migration (CM) mechanism which applies SYN Cookie technique [9] to prevent Source-IP-spoofed SYN packets from requesting controller to install FEs.

CM mechanism makes OFS work as an SYN proxy. Upon receiving an SYN request from an Internet client, instead of forwarding the packet to the internal server, OFS handshakes with the client. After ensuring the connection to the client is successfully based on authenticating the three-way handshake (3HS) process, the OFS requests controller for a new FE to set up and maintain a TCP connection to the real internal server.

Although the CM mechanism prohibits attacked SYN packets from entering servers and decreases useless flow rules, it causes some problems [10] including: (1) splitting benign connections between Internet clients and internal servers, which consumes more time to process packets, disables some end-to-end functionalities of TCP protocol, and also requires the switch to maintain state information for each TCP connection; (2) an increase the processing load at the OFS, which makes the switch more vulnerable to the attack.

In this paper, we make two contributions. In the first one, we propose an SSG solution as a substitution for the CM mechanism in Avant-Guard. To protect the OFS from overloading, the SSG monitors the 3HS process of TCP connections on a separate security device. The OFS recognizes and transfers 3HS packets to the security device based on the ability to match TCP flag fields defined in OpenFlow 1.5 [3]. Moreover, to overcome the problem of splitting legitimate TCP connections, the SSG uses the RST Cookie technique to authenticate and verify the source IP address in SYN packets instead of SYN Cookie as in CM. These enhancements help the system solve the shortcomings of the CM mechanism.

In the second contribution, we build a test-bed and run a number of experiments to compare our solution against OpenFlow and the CM mechanism under SYN flood attack at different speeds. We also make an analysis on CAIDA dataset [11] to measure the increased traffic at OFS interfaces when a server operating with a real traffic is under SYN flood attacks. The analytical and experimental results show that under SYN flood attacks, the SSG improves the rate of successful connections, and decreases the average retrieve time with a lower resource consumption at both the OFS and the controller. The OpenFlow traffic between the OFS and the controller is also reduced with a small increase in the rate of traffic at the OFS interfaces. These improvements make the system more resistant under saturation attacks by SYN flooding.

The rest of this paper is organized as follows. Related works are introduced in Section II. Section III presents the principle of the CM mechanism of the Avant-Guard solution. Section IV describes our proposed network security architecture and the operating principles. The analysis of the proposed solution and experiment results are elaborated in Section V. Finally, conclusions are presented in Section VI.

II. Related Work

The SDN/OpenFlow architecture has attracted many research works including security issues [5–7, 12–18]. Some works analyze the SDN architecture and OpenFlow protocols in terms of security and identification of the threats of saturation attacks on data plane and control plane such as the study of Kreutz *et al.* [19], Kloti *et al.* [20] and Shin [16]. Solutions to mitigate this type of saturation attacks can be divided into three groups.

In the first solution group, processing functionalities are added to the controller. Traffic characteristics of flows passed across the system are collected and sent to the controller. The application running on the controller analyzes traffic characteristics and network topology to detect anomalies and through installing FEs into the OFS produce appropriate policies such as load balancing, dropping packets to mitigate attacks. The common drawback of these solutions is that the controller must collect sufficient information on each new flow on the entire network using statistical query messages and therefore the network traffic increases dramatically when an attack occurs. Some proposed solutions such as ONIX [21], HyperFlow [22], and ONOS [23] use logically centralized but physically distributed controllers to increase the responsiveness for requests from the OFSs to the controllers. In [24], Anilkumar proposed SmartTime, an OpenFlow controller system that combines an adaptive timeout heuristic to compute efficient idle timeouts with the proactive eviction of flow rules, to utilize effectively TCAM space in the OFSs. The SDN-Guard solution proposed by Lobna *et al.* [25] combines dynamically rerouting potential malicious traffic, adjusting flow timeouts and aggregating flow rules while in FloodGuard [26], Wang *et al.* introduced two techniques: proactive flow rule analyzer and packet migration to reduce both the number of FEs on the OFSs and communication between the data and control planes.

The second solution group deals with the problem by enhancing the functionality of switches. A typical function of OpenFlow 1.3 [2] is Rate Limit that restricts the rate packets processed at the switches. The DEFANE solution [27] proposed by Minlan et. al. adds topology discovery and FE distribution that allow the OFSs to easily enforce network access policies. Cutis et. al. proposed DevoFlow [28] that allows to detect elephant flows and to create and install a new FE by copying the same actions without refering to the controller. In [29], Mekky *et al.* proposed a solution to limit the communication between the data plane and the control plane by allowing packet forwarding based on application level information. By the same way, in [30], Kotani and Okabe reduced the data-to-control plane communication by adding two data structures for low-priority packets, Pending Rule Table (PRT) and Pending Flow Table (PFT). The packet are matched with FEs in the flow table and in the PRT and PFT. Only packets that are not matched with any FEs in the flow table can create *packet_in* events and be sent to the controller. However, this defense mechanism can be avoided by flooding attack packets with header modifications such as spoofed source IP fields. In such cases, the PRT and PFT filters will be ineffective and all the packets will be forwarded to the controller. The common downside of these solutions is that they require structure modification and higher capacity at the OFSs for storing additional detailed information and processing actions. With these modifications, OFSs must also have additional packet processing functionalities. Attackers can exploit this feature to increase the processing task at the switches.

The third solution group implements the packet handling rules to eliminate and prevent resource starvation attacks based on communication protocols. In Avant-Guard [8], S. Shin uses the SYN Cookie technique [9] to build the CM mechanism that prevents and minimizes impacts of SYN Flood attacks. As a result, the switch will censor all TCP connections and use SYN Cookies to authenticate the 3HS process. The OFS requires the controller to install FEs only for those connections that complete the 3HS process. Ambrosin introduced an improvement to the LineSwitch solution [10] in order to reduce the load of the OFS by ignoring the authentication of the 3HS process for SYN packets coming from trusted sources then sending a packet-processing request to the controller immediately at a given probability p. LineSwitch also immediately removes the incoming SYN packets if their source addresses are from untrusted sources (in the blacklist). Although LineSwitch reduces the load at the OFS and the connection processing time between the client and the server, the solution requires the switch to manage a large number of IP addresses for both trusted and untrusted source lists, so the search time in these lists will also increase. When a system is attacked by an SYN Flood with IP address spoofing, the number of IP addresses in the untrusted source list increases dramatically. Solutions in this group only address one type of attack. Thus far, the proposed solutions have only focused on SYN Flood attacks. Due to the popularity and influence of this type of attack [4, 31, 32], we study and propose the SSG solution based on the idea of Avant-Guard's CM mechanism. LineSwitch techniques can be fully applied to the SSG without compromising the performance of the OFS. To easily compare the performance and improvement, in Section III, the working mechanism and the shortcomings of the CM mechanism in Avant-Guard are summarized.

III. Introduction of Connection Migration Mechanism

The main idea of Avant-Guard is to introduce the CM mechanism into the OpenFlow architecture to ensure the legitimacy of an SYN-packet source by monitoring the 3HS process of TCP connections. For the TCP connections that complete the 3HS process, the OFS sends a *packet_in* message to the controller to request a new FE. The CM mechanism is integrated into the OFSs and using the SYN Cookie technique [9] with four phases as shown in Figure 1.

a) Classification Phase:

When the OFS receives an SYN packet (step 1 in Figure 1), it directly replies the client an SYN-ACK packet using the SYN Cookie [9] technique (step 2). Depending on whether an ACK packet is received from the client (CliACK) (step 3), the SYN Cookie technique checks the 3HS process and determines whether the SYN packet has a forged Source IP Address or not.

The SYN Cookie technique [9] was invented by Bernstein to combat SYN Flood attacks, where a server can verify a client SYN request without storing and managing connection status information in Transmission Control Blocks (TCB) while waiting for a CliACK packet replied from the client. When an SYN packet is received, the server generates a 32-bit cookie value, which is encrypted

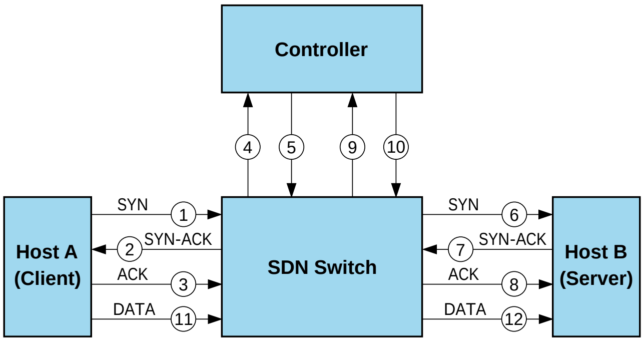


Figure 1. Avant-Guard connection migration mechanism.

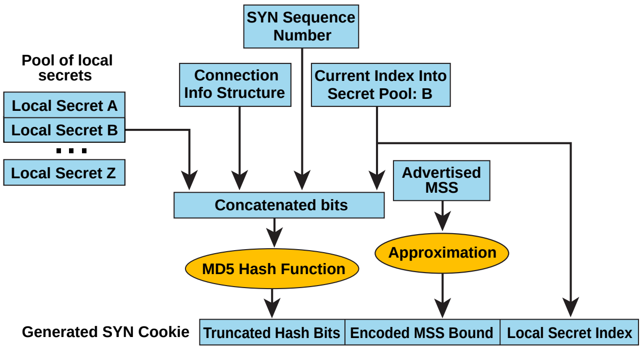


Figure 2. Cookie encryption process in SYN Cookie technique.

by the TCP connection parameters (including the source and destination IP addresses, source and destination ports, the initial *Sequence Number* of the SYN packet) and the server's secret parameters. This cookie value is used as the *Sequence Number* of the replied SYN-ACK packet (Figure 2). Upon receiving the acknowledgement packet CliACK from the client, the server encrypts cookie again and compares its value to the *Acknowledgment Number* in the received CliACK. If these two values match, the CliACK packet is validated and the 3HS process is assumed to be successful.

In Avant Guard, the CM applies the SYN Cookie technique directly into an OFS to authenticate the source IP address of an SYN packet by directly handshaking with the client. This allows the system to completely prevent SYN attack packets that spoof IP addresses from being forwarded to internal application servers without having to store the states of TCP connections in the OFS. Only authenticated SYN packets that complete the 3HS process will be passed to the Report phase.

b) Report Phase:

If the 3HS process of the SYN packet is completed, the OFS sends a *packet_in* message to the controller (step 4, Figure 1) to request to install a new FE that will be in charge of processing subsequent packets from the authenticated TCP connection from the client (step 5).

c) Migration Phase:

Upon receiving the controller's consent, the OFS establishes a TCP connection to the server (steps 6, 7 and 8 in Figure 1). Similar to the connection to the client, after the 3HS process with the server is successfully completed, the OFS also sends a *packet_in* message to the controller to request an FE for processing incoming packets of the connection from the server side (steps 9 and 10).

d) Relay Phase:

When both handshakes with the two sides are successful, the OFS maintains a communication session for exchanging packets between the client and the server by two FEs that have been created during the report phase and the migration phase.

A prominent advantage of the CM mechanism is that for only TCP connections that have completed the 3HS process, the OFS sends *packet_in* messages to the controller and the corresponding FEs are installed. SYN packets with forged source IP address will not complete the 3HS, hence their impact on the switch (data plane) and also on the controller (control planes) is mitigated. Moreover, the SYN Flood attack traffic is also prevented going to servers inside. Further more, applying the SYN Cookie technique helps the OFS reduce memory usage for TCBs especially when an SYN Flooding attack occurs.

However, the CM mechanism has three main disadvantages, as described next. First, CM uses the SYN Cookie, which is a host-based solution against SYN flood attacks applied directly into a server [33], which is a terminal entity of TCP connections. This approach is not suitable for deployment on intermediate devices such as OFSs because each benign TCP connection will then be split into two parts: one between the switch and the client while the other between the switch and the server. To link these two partial connections, the switch has to spend memory and computing resources for managing and maintaining state information as well as exchanging the *SEQ_Num/ACK_Num* value pair in packets for each TCP connection. After these value fields change, the packet's checksum must be calculated and updated from Layer 4 down to Layer 1 [34]. This reduces the system performance, increasing packet processing latency for benign TCP connections.

Second, the splitting of TCP connections between clients and servers due to the SYN Cookie also disables some optional features of the TCP protocol such as Maximum Segment Size (MSS), Selective ACKs, or TCP Window Scaling [34], etc during the 3HS process.

Third, the process of generating SYN-ACK packets, validating CliACK packets in the SYN-Cookie technique requires a number of certain computational resources including creating a packet, encrypting and generating cookies, calculating checksums, and so on. The CM mechanism performs the 3HS process by the SYN Cookie in the OFS for all incoming SYN packets even if the destination port on the server is closed. This wastes computing resources on the switch, making the switch vulnerable and easily become a target of SYN flooding and scan port attacks. In fact, integrating the additional hardware - CM module into OFS loses the original nature of an SDN switch that was designed to function as forwarding plane only. Therefore, it will be difficult to apply this type of solution to commercial OFSs that are designed as the original OFS (i.e. working in the data plane only).

As an enhanced solution of Avant-Guard, Lineswitch [10] supplements algorithms in order to reduce the usage of connection proxy at the OFS, through managing lists of trusted and untrusted source IP addresses and comparing them with the addresses of incoming packets. However, Lineswitch still uses the CM mechanism in a similar manner to Avant-Guard so the above shortcomings still prevail.

IV. SSG - Proposed Solution to Prevent Saturation Attack using SYN Flood on the Data Plane and Control Plane

1. Research Scope and Methodology

SDN-based SYN Flood Guard (SSG) is proposed to protect the OFSs (data plane) and controllers (control planes) of a small to medium size data centers from saturation attacks using TCP SYN Flood. The internal application servers are supposed to be trusted, not infected or under control of malwares. The techniques discussed in this paper only work for TCP SYN flood, not for other attack types.

With the aim of overcoming the disadvantages of CM mentioned in Section III, the SSG solution, first, minimizes resource and memory consumption at the OFSs by moving the function of monitoring the 3HS process from the switches in CM to a separated security device that is connected reliably with the controller. This helps SSG increase attack tolerance, eliminate the risk of becoming a target for saturation SYN Flood attacks, and make the solution feasible for all OpenFlow-supported commercial switches.

Second, SSG does not change OpenFlow packet-processing procedures when the system is free of SYN Flood attacks, in order not to affect the connection time or the delay of TCP connections. To do this, SSG must be equipped with a mechanism to detect the attack for

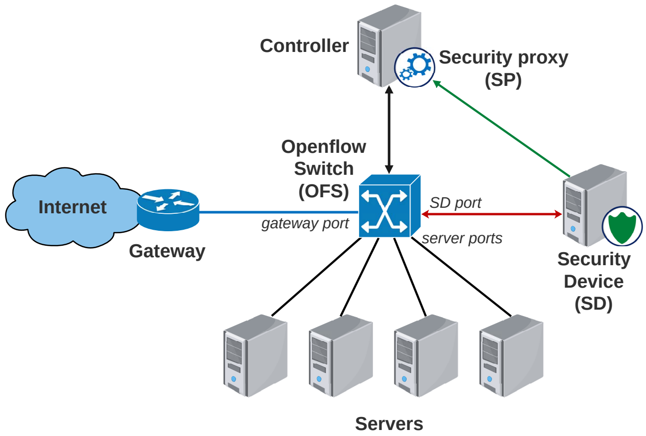


Figure 3. SSG system architecture.

each internal server and change the corresponding packet handling policies when server's state changes.

Finally, SSG uses the RST Cookie technique to verify the source IP address of untrusted SYN packets instead of using the SYN Cookie in Avant-Guard. This allows direct TCP connections from the client end to the server end without being split, and therefore it is easy to discard SYN requests to closed ports on internal application servers. Not using the SYN Cookie technique at the intermediate device also enables optional features in the TCP protocol during the 3HS process.

2. System Architecture and Operation Principle

With the outlined methodology above, Figure 3 shows the SSG system architecture. Figure 4 describes the detailed structure of main components of the entities and the interactions between them. It is proposed to implement the SSG at the edge network of a data center. This architecture is inherited from our previous work [35] in which the Internet traffic is passed through a gateway to servers via an OFS.

The OFS takes charge of (i) capturing packets during the 3HS process of TCP connections and forward them to a Security Device (SD) for monitoring and (ii) maintaining FEs for exchanging data of each TCP connection after it accomplishes the 3HS process. Making use of the ability to match TCP Flags fields [3] and the packet processing pipeline, SSG arranges proactive and reactive FEs with suitable actions into flow tables in the OFS to handle these tasks. The OFS distinguishes three separate port types: gateway port to receive and process incoming and outgoing traffic from the Internet, SD port to connect to SD, and other normal ports (server ports) to connect to servers of the data center. This distinction is based on the information of the *in-port* matching field specified by the OpenFlow protocol and allows the system to distinguish the packets

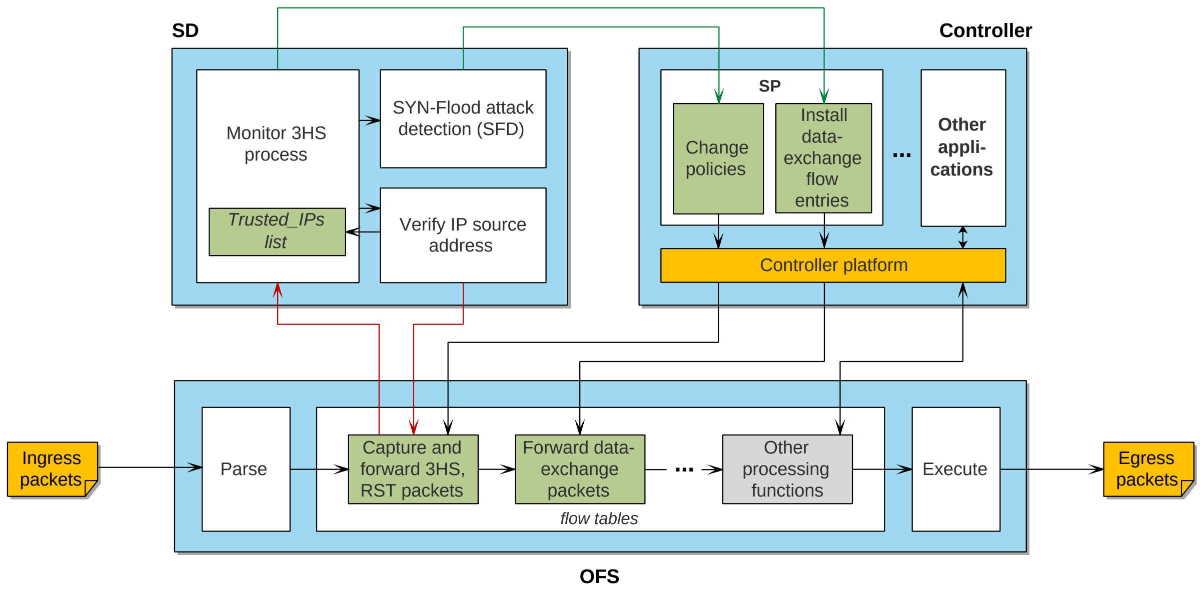


Figure 4. Detailed structure and interactions between main system components.

from clients, servers or SD and to have appropriate packet processing policies.

The SD is a stand-alone device which analyzes traffic forwarded from the OFS in order to: (i) detect SYN Flood attacks, (ii) monitor and authenticate the 3HS process of TCP connections between the outside clients and the inside servers, (iii) verify the source IP address of incoming SYN packets by using the RST Cookie technique when a server is under an SYN Flood attack, and (iv) request the controller to modify, install FEs for related packet processing rules. The SD uses packet processing tools such as DPDK [36] instead of relying on common network protocols at general terminals.

The Security Proxy (SP) is an application running on the controller, which receives the requests from the SD and interacts with the controller platform to (i) modify FEs for changing 3HS packet-forwarding policies when the attack state of a server changes and (ii) set up FEs for exchanging data between two ends of a TCP connection which is accomplished the 3HS process.

The SSG operation is based on the following principles. First, alll TCP connections from external clients must be controlled to ensure that their original source IP addresses are real and not spoofed by checking the 3HS process at the SD. The OFS just installs FEs for subsequent data-exchange packets of TCP connections of which the 3HS processes are accomplished.

Second, as an alternative to the CM mechanism, a 3HS process is carried out from the client end to the server end without being split. If there is no SYN Flood attack occurred, 3HS packets are sent to the SD for monitoring. Depending on the number of Half Open Connections (HOCs), the SD detects the presence of SYN Flood attack for each server and changes its state to either *"under SYN flood attack"* or *"attack free"*.

Third, when the system is under an SYN Flood attack, the SD verifies the source IP address of SYN packets before monitoring the 3HS process by using the RST Cookie technique. The monitoring process is only applied to the SYN packets with verified source IP addresses.

Finally, the change of processing rules for the 3HS packets depends on the state of the servers regarding to an attack. Whenever the state alters, the SD module will request the SP to modify the corresponding FEs for changing rules.

Details on packet processing at the OFS, the SD and the mechanism of the attack detection are described in the following subsections.

3. Forwarding TCP Packets at OFS

Using the ability to match packet headers with TCP flags specified in OpenFlow 1.5 [3], the ability to arrange flow entries with different priorities in multiple flow tables [2], in SSG, the OFS is set up and configured to match and capture the 3HS packets of TCP connections and route them to the servers or the SD based on the state of the internal destination server. If the state is "attack free", the packets are forwarded directly to the client and the server, and the copies are sent to the SD for monitoring the 3HS process. In case the destination server is under an SYN Flood attack, the SYN packet is sent to the SD for validating its source

TABLE I
ORGANIZATION OF FLOW TABLES AT THE OFS

Flow Table	Function	Flow entry type	When table miss occurs
FT1	Capture SYN packets from clients, SYN-ACK packets from servers and SD for directing and monitoring 3HS process, detecting SYN Flood attacks	Proactive (FEs exist permanently)	Direct to the flow table FT2
FT2	Direct data packets between clients and servers of TCP connections after 3HS completion	Reactive (FE exists until the TCP connection ends. The lifetime is based on idle-timeout and hard-timeout settings)	Direct to the flow table FT3
FT3	Capture CliACK, RST packets from clients, RST packet from servers for authenticating 3HS of a Half-Open Connection and processing the TCP connections	Proactive (FEs exist permanently)	Direct to the flow table FT4
FT4 and subsequent FTs	Contain FEs for other applications		

TABLE II
CHARACTERISTICS OF FLOW ENTRIES IN FLOW TABLE FT1 AND FT3 TO CAPTURE AND FORWARD 3HS PACKETS

Flow entry	In Flow table	Matched with packet	Packet appears on port	Actions
FE1x	FT1	SYN	gateway	– *If server is attack free:* pass to the corresponding server and a copy to SD – *If server is under SYN Flood attack:* pass to SD
FE2x	FT1	SYN	SD	– Pass to the corresponding server
FE3x	FT1	SYN-ACK	server	– Pass to the corresponding client through the gateway port – Create a copy and send to SD
FE4x	FT1	SYN-ACK	SD	– Pass to the client via gateway port
FE5	FT3	RST	gateway or server	– Pass to SD
FE6x	FT3	CliACK	gateway	– Pass to SD – *If server is attack free:* pass to the server

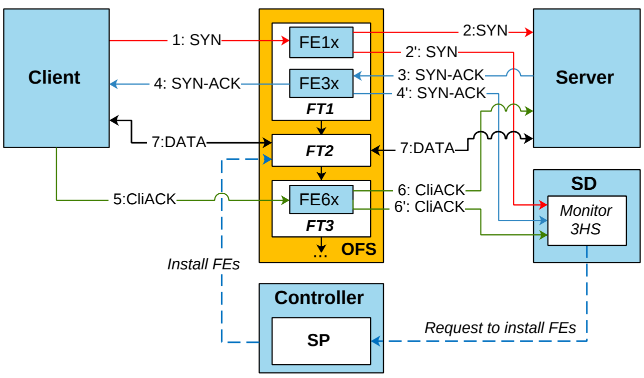


Figure 5. The process of capturing and directing packets of a legitimate TCP connection via FEs in OFS when the destination server is attack free.

IP address. Flow entries at the OFS are organized into flow tables which operate in the pipeline mechanism. The functions of the flow tables are shown in Table I. Table II describes the properties and actions of flow entries that capture and direct packets of the 3HS process.

Figure 5 is an example of the process of capturing and directing packets of a legitimate TCP connection via FEs at the OFS when the destination server is in the state of no SYN Flood attack.

4. Monitoring Three-way Handshake Process at SD

A normal 3HS process of a benign TCP connection must meet two requirements: (i) the client acknowledges the server by a CliACK packet upon receiving an SYN-ACK packet, and (ii) the *Acknowledgement Number (ACK_Num)* of the CliACK packet must be equal to the consecutive value of the *Sequence Number (SEQ_Num)* provided in the SYN-ACK packet. Monitoring the 3HS process is in order to identify and discard the SYN packets of which the source IP address is spoofed. The process of monitoring 3HS in SD is described in the flowchart of Figure 6. If there is no SYN Flood attack on the server, the monitoring process is applied to all incoming SYN packets. When the server is under a SYN Flood attack, the 3HS process of SYN packets from only trusted IP addresses are monitored. The trusted IP addressed are managed in a list named *Trusted_IPs* and

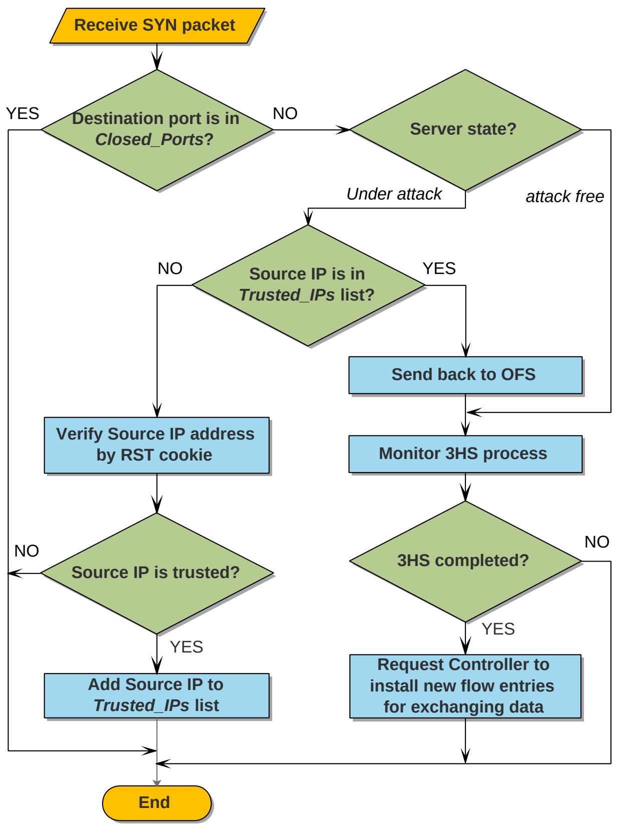


Figure 6. The flowchart of processing an incoming SYN packet in SD.

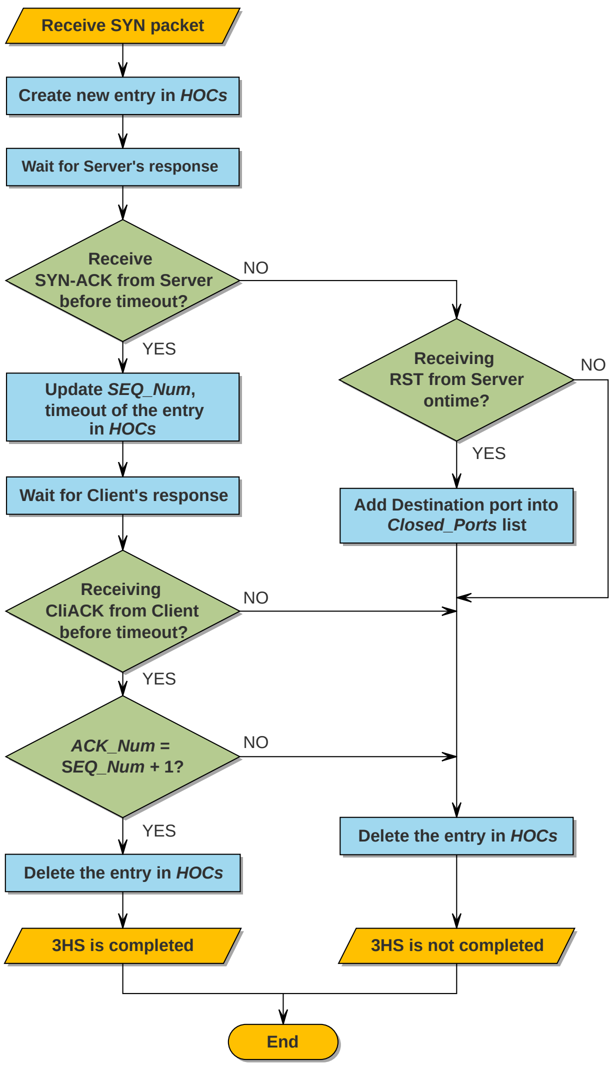


Figure 7. The flowchart of monitoring 3HS in SD.

Source IP address	Destination IP address	Source port	Destination port	SEQ_num (of SYN-ACK packet)	timeout

Figure 8. Structure of an item in *HOCs* list.

kept in the list for a period $T_{\text{IP-timeout}}$. SYN packets with the source address not in *Trusted_IPs* will be checked the source-address legitimacy by the RST cookie technique (the legitimacy check up will be described in subsection IV-E). If the source IP address is validated, it is added to the *Trusted_IPs* and the 3HS of the SYN request packet in the second time from the source will be monitored.

To limit port scanning attacks on the system, the SD also maintains a list *Closed_Ports* containing ports that have been recently closed in the servers. After receiving an SYN packet from a client, if the server returns a corresponding passive RST packet, the system understands that the port is currently closed. At this point, the port number is updated to the list and maintained for a period $T_{\text{port-timeout}}$. The SD discards the SYN packets that request to connect to destination ports on the server in the *Closed_Ports* list.

TCP connections are monitored for the 3HS process as described by the flowchart of Figure 7. To manage and monitor the 3HS, the SD maintains a list named *HOCs* containing half-open TCP connections. Each item in *HOCs* corresponds to a TCP connection and contains the information as shown in Figure 8. When the 3HS of a TCP connection is completed or the lifetime exceeds $T_{\text{HOC-timeout}}$, the corresponding item will be deleted.

5. Verifying Legitimacy of SYN Packet Source IP Address

Verifying the source IP address of an SYN packet is the process to check if the address provided in the *Source IP Address* field is the real source which generates the packet. The RST Cookie technique is based on the principle of TCP reliable exchange between two entities (client and server) that each party initializes a 32-bit value for *SEQ_Num* and send to the other. A receiving packet from the opposite side is verified by the value of *ACK_Num*, which must be equal

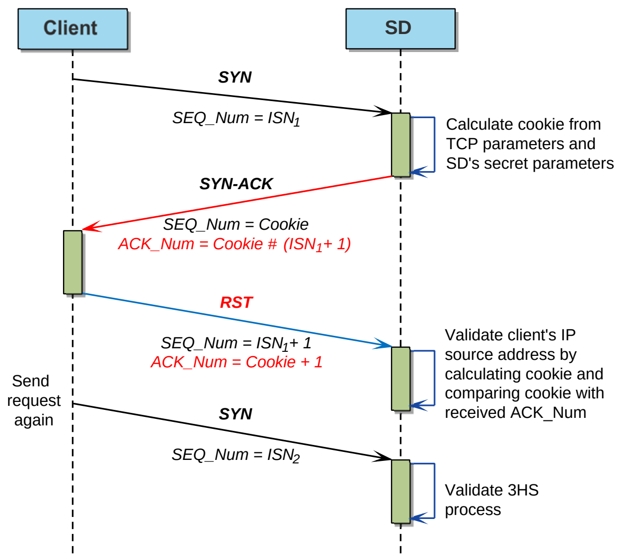


Figure 9. Sequence diagram of verifying the source IP address of an SYN packet using the RST cookie.

to the consecutive value of the sent *SEQ_Num*, that is,

$$ACK\_Num_{\text{receiving}} = SEQ\_Num_{\text{sent}} + 1. \qquad (1)$$

If this rule is violated, the detecting side understands that an error has occurred and resets the connection. If this error is detected by the client, it will send an active RST packet and resend another SYN request.

Upon receiving an SYN request of which the *Source IP Address* is not in the *Trusted_IPs* list, the SD generates and replies the client with an SYN-ACK packet with the *ACK_Num* different from the consecutive value of the received *SEQ_Num* (Figure 9). At the same time, a cookie value generated by encrypting the TCP connection information and secret parameters is used as the initiated *SEQ_Num* for the direction from SD to the client. If the SYN packet is not spoofed, upon receiving this SYN-ACK, the client sends an RST with *ACK_Num* equal to the consecutive value of the cookie. The SD checks the validity of the RST packet by re-encrypting the cookie value and comparing with the received *ACK_Num*. If the values match, the SD understands that the SYN packet is generated from a real source and not being spoofed. In the case that SD does not receive a corresponding RST in time or the received *ACK_Num* is not correlated with the encrypted cookie, the field *Source IP Address* of the SYN packet has been forged for an SYN Flood attack.

6. Monitoring and Detecting SYN Flood Attacks at SD

To minimize the impact on the system performance and packet latency, SSG has different packet-handling policies for the cases of SYN Flood attack and attack free. The

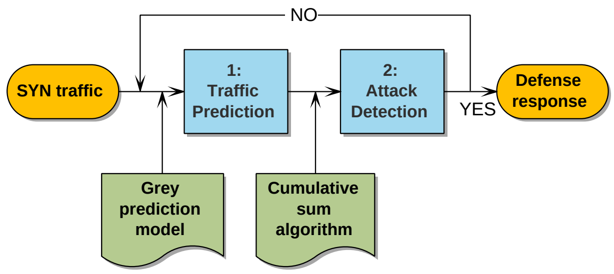


Figure 10. The model of SYN Flood detection used in SFD module.

SYN Flood attack Detection (SFD) module developed in SD detects the attack and asks the controller to distinguish these two states by using the corresponding packet handling policies. For the detection of the attack in SFD, the solution based on Grey prediction model [37–39] and Cumulative sum algorithm [40, 41] introduced by Wang *et al.* [42] is applied, Figure 10. When the system is under SYN Flood attack, the number of HOCs (k_{HOC}) increases quickly so this figure in each monitoring period T_{SFD} is chosen as the input parameter:

$$k_{\text{HOC}} = k_{\text{SYN}} - k_{\text{3HS-completed}}, \qquad (2)$$

where k_{SYN} is the number of incoming SYN packets and $k_{\text{3HS-completed}}$ is the number of connections completing the 3HS process of a server in a monitoring period.

Depending on the classification result after each monitoring period T_{SFD}, when the attack-state of a server changes (from normal to attacked or vice versa), the SFD sends a message via the SP module to request the controller to modify corresponding flow entries FE1x, FE2x and FE6x for changing actions as described in Table II. With these modifications, subsequence SYN packets are processed by desired policies which are suitable for the current server state as described in subsection IV-3.

V. System Performance Evaluation

To compare the performance of the SSG with the CM mechanism of the Avant-Guard solution, we built a test-bed and implemented different SYN Flood attack scenarios by varying attack rates and analyzed the measured results. The test-bed is described in Figure 11.

The OFS is an OpenVswitch [43] version 2.7.90 soft switch built on a computer with five 1-Gbps ethernet cards, Intel Core$^{\text{TM}}$i3-2330M CPU @ 2.2 GHz, 500GB HDD, 2GB RAM.

The SD is a stand-alone PC running Ubuntu 14.04 with Intel Core$^{\text{TM}}$i3-2330M @ 2.2 GHz CPU, 500GB HDD, 2GB RAM. To adapt to SD functionalities, packets on the interface between the SD and the OFS are captured and

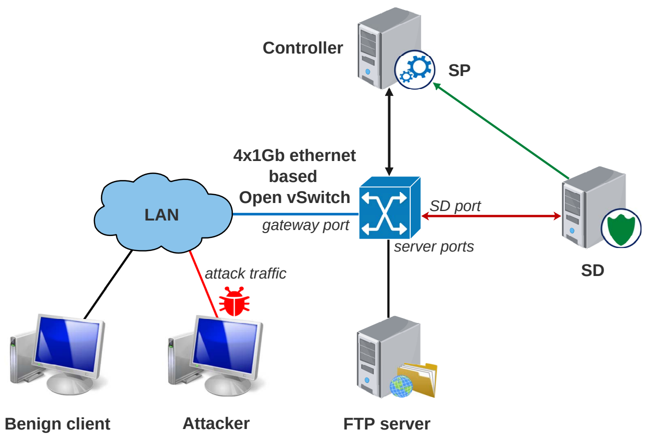


Figure 11. Structure of the testbed.

processed not according to the usual TCP protocol but a module developed by the DPDK tool [36].

The controller is based on the Floodlight scheme [44] installed on the Ubuntu 14.04 PC with the following hardware configurations: Intel Core^{TM}i3-2330M @ 2.2GHz CPU, 500GB HDD, 2GB RAM.

SYN Flood attack traffic is generated from the BONESI [45] tool with attack rates varying from 100 pps, 200 pps to 5500 pps. To ensure the homogeneity for different configurations, the generated attack traffic was logged using the WireShark [46] tool and regenerated with the TCPReplay [47] tool. The emulated attack time is 500 s for each attack rate. Legitimate traffic is generated from a software running on a computer that continually requests connections and downloads files of the FTP service on the server at a rate of 30 connections per second. The traffic between the benign client and the FTP server is recorded using the WireShark tool for statistical analysis.

Attack scenarios are performed in the same way for three schemes: the standard OpenFlow, the CM mechanism of Avant-Guard, and the SSG solution. Measured results and analysis are presented in the following subsections.

1. Successful Connection Rate and Retrieval Time when System is under SYN Flood Attacks

The goal of the SYN Flood TCP attack is to make the system less capable of serving benign TCP connections, in other words, to reduce the successful connection rate *(SCR)* and increase the retrieval time of benign TCP connections. *SCR* is measured by the number of TCP sockets that download a desired file successfully over the total TCP sockets sent from the benign clients. The average retrieval time *ART* is calculated as the average time of the successful TCP sockets from the time a benign client sends an SYN packet until it receives the first data packet from the FTP server.

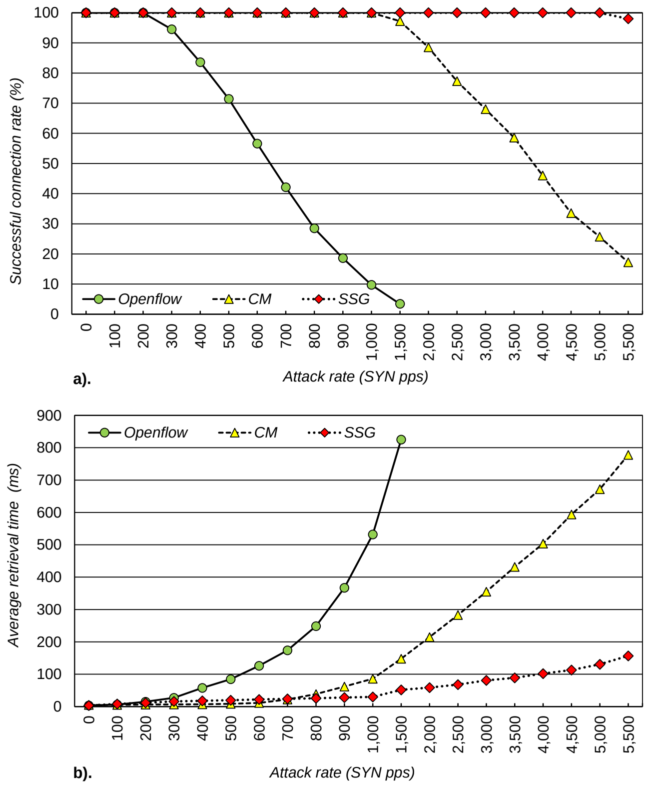


Figure 12. The *SCR* (a) and *ART* (b) of OpenFlow, CM and SSG schemes.

The measured parameters of the three schemes at different attack rates are presented in Figure 12. When the test-bed is attack free or under attack with low rate (below 200 pps), there is no considerable difference between the three schemes. In addition, under the increased attack rates, *SCR* of the OpenFlow scheme is quickly decreased to 3% at the rate of 1500 pps and the benign clients can not connect to the internal FTP server during the attack of 2000 SYN pps. The tolerance of CM mechanism is better as the *SCR* is maintained at 97% at a rate of 1500 pps and diminishes down to under 20% when the rate is 5500 pps.

Moreover, compared to the two other schemes, the connection ability to the server of the SSG solution is superior as the *SCR* remains stable at 100% and when the attack rate increases to 5500 pps, the rate just starts slightly decreasing to 98%, much higher than the figure of CM mechanism.

Finally, the results also point out that there is a correlation between the decline of *SCR* and the *ART* of TCP connections from benign clients to an internal application server for all of the three schemes. However, at low attack rates (from 100 pps to 700 pps), *ART* in SSG solution is a bit higher than that in the CM mechanism. The reason is that in SSG, a client must resend SYN requests when its IP address is not in the *Trusted IPs* list.

The testbed results show that the SSG solution significantly improves the system's attack tolerance over the OpenFlow and CM mechanism in Avant-Guard.

2. Resource Usage on the Switch and SD

Figure 13 presents the average amount of resource consumption in OFS and SD of the three schemes during states of attack free and under SYN flood attack. The resource consumption is defined using two parameters: the memory usage and the CPU utilization of OFS and SD processes on corresponding systems.

It can be seen that, when there is no attack, the memory occupancy and the CPU resource utilization of the SSG solution are approximately the same as the OpenFlow scheme with the respective values of over 150 MB and about 8%. Meanwhile, in the CM mechanism of the Avant-Guard solution, both these figures are higher, with values near 200 MB and 13%.

When the system is under SYN Flood attack, the resource consumption in the OFS of the OpenFlow scheme increases fastest and reaches nearly 1600 MB memory and over 50% CPU resource at an attack rate of 1500 pps. This is a demonstration of a saturation attack on the data plane, which occupies most of the switch's resource and causes the decrease in *SCR* as described in subsection V-A. The CM mechanism and the SSG solution help the OFS to reduce the impact of the attack and the resource occupancy rates just are under 300 MB of memory and about 30% of CPU at the same attack rate.

In comparison between the CM mechanism and the SSG solution, it can be seen that the resource usage in OFS of the SSG solution is always lower than of the CM mechanism. Especially, in the case of high attack rates (from 2500 pps to 5500 pps), the CM mechanism needs much more CPU resource than SSG which is the main reason for the *SCR* reduction presented in Section V-1.

The improvement in reducing resource usage, and packet processing capabilities of OFS in the SSG solution are due to transferring the function of 3HS monitoring from OFS to SD. The results in Figure 13 show that the resource consumption on the SD is not significantly affected by the increase in attack rate. Compared to the case of no attack traffic, at the highest attack rate in the testbed (5500 pps), the memory usage increases from 180 MB to 300 MB over 2 GB total memory and the CPU resource is increased from 12% to 23%. Choosing the right configuration for OFS processing capability, this combination helps the system increase its SYN Flood tolerance.

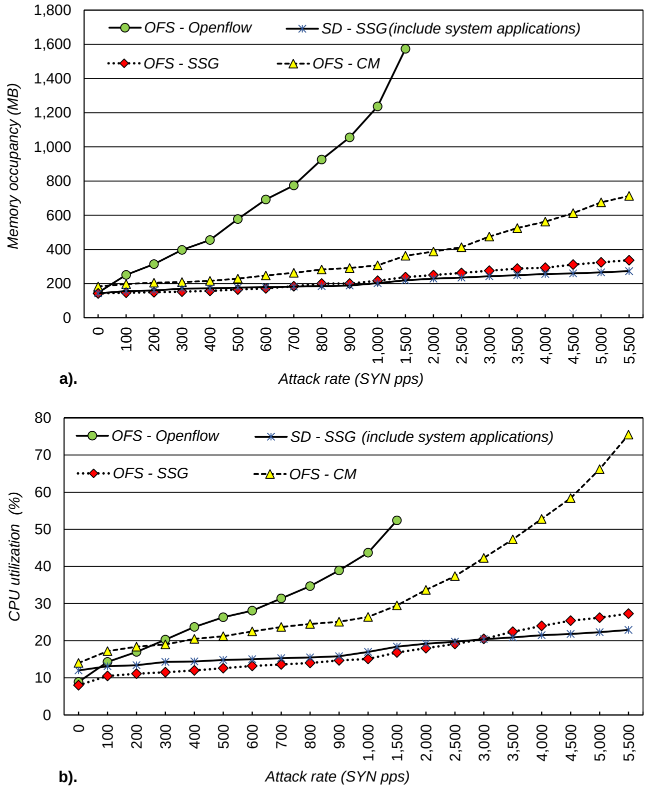


Figure 13. The memory occupancy (a) and CPU utilization (b) in OFS and SD.

3. Increase in Traffic Volume at OFS Interface

Compared to the CM scheme, moving the function of monitoring the 3HS process to SD increases the processed traffic at OFS interfaces. Here, the change in traffic at OFS interface is analyzed in both server states: free and under TCP SYN Flood.

a) When a server is not under SYN Flood attack:

By comparing the packet exchange between OFS and the other entities of SSG solution (Figure 5) and the CM mechanism (Figure 1), it can be seen that OFS in the CM scheme carries out 10 actions while in the SSG, the switch takes only 9 steps for the 3HS process of a benign TCP connection. Of these actions, 4 of CM and 1 of SSG are OpenFlow messages, the others are 3HS TCP packets. As defined in TCP protocols [34], 3HS packets usually are the smallest size TCP packets with the main information containing flag settings. Meanwhile, OpenFlow messages are TCP packets containing different information, so their size is much larger than 3HS packets. Therefore, taking account of the size of interaction packets, during the state of no attack, the traffic at the OFS interfaces of SSG solution is smaller than in the CM scheme.

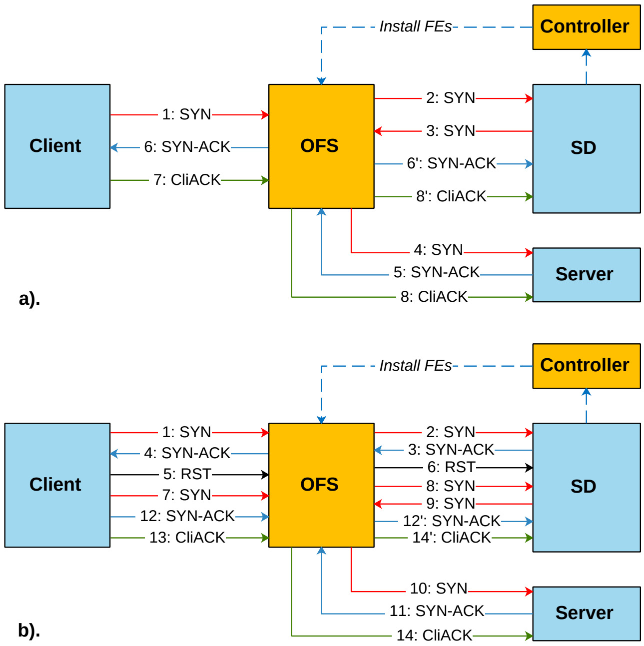


Figure 14. The interaction in SSG during 3HS process of a benign SYN packet with *Source IP address*: (a) in *Trusted_IPs* and (b) not in *Trusted_IPs* list

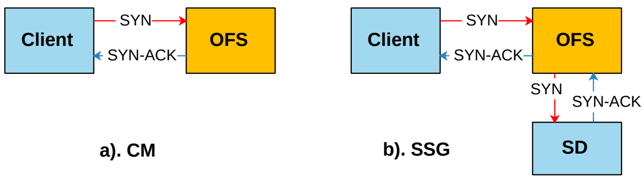


Figure 15. The interaction for processing an Source-IP spoofed SYN packet in: (a) CM and (b) SSG schemes.

b) When a server is under SYN Flood attack:

Incoming SYN packets belong to one of three cases: (i) a benign packet with source IP address in the *Trusted_IPs* list, (ii) a benign packet with source IP address not in the *Trusted_IPs* list, and (iii) an attack packet with a spoofed source IP address. With packet processing mechanism in the SSG solution, OFS has to handle 9 actions in the 3HS process of a benign TCP connection for the first case (Figure 14.a) and 15 steps for the second case (Figure 14.b). In the CM mechanism, the interactions for 3HS process of SYN packets in the both cases are implemented in the same way through 10 steps as described in Figure 1. Concerning the third case, the number of OFS interaction steps in CM and in SSG are respectively 2 and 4 (Figure 15).

To evaluate the increase in traffic at OFS interfaces, the attack traffic is measured and analyzed in combination with attack-free traffic. For the attack-free traffic, logged packets from 100 randomly-selected application servers in the CAIDA 2013 dataset [11] are processed and analyzed. The statistical characteristics results of the traffic are described

TABLE III
STATISTICAL CHARACTERISTICS OF ATTACK-FREE TRAFFIC

Number of analyzed servers	100 servers
Analysis time	45 minutes
Number of TCP flows	3,947,883 flows
Number of new flows perserver per second	14.62 flows/server/s
Average total throughput from clients to each server	11,658,649.46 bytes/server
Average data rate from client to each server	5,181.62 bits/s/server
Average data rate in both ways per server	72,404,791.51 bits/s/server
Rate of SYN packets having the same source IP address with the previous packets ($T_{\text{IP-timeout}}$ = 5 minutes)	44,7 %

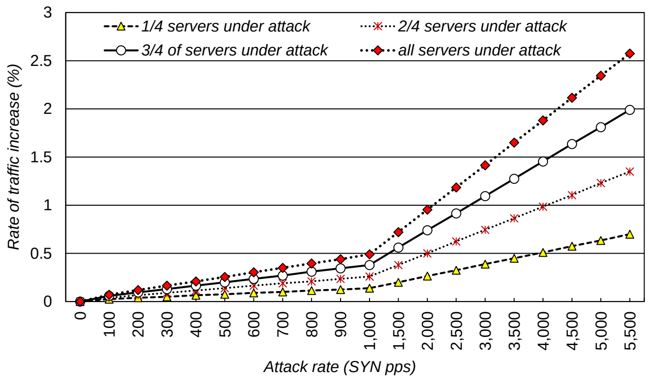


Figure 16. Traffic increase rate of SSG in comparison with the CM mechanism.

in Table III. The combined traffic is analyzed for four cases: 1/4, 1/2, 3/4 and all of internal application servers being attacked by SYN Flood with the source IP spoofing technique. The analyzed result in Figure 16 shows that when the system is attacked by SYN Flood, in comparison with the CM mechanism, OFS traffic increases with a small rate, just about 1.8% in case half of internal servers being attacked and over 2.5% when all the servers in the system are attacked with the rate of 5500 SYN pps each server.

4. Reduction in OpenFlow Traffic and Load on the Controller

Both the Avant-Guard's CM mechanism and the SSG solution operate based on monitoring the 3HS process of a TCP connection before requesting the controller to install FEs on the OFS. This not only prevents the resource consumption in the OFS by useless attack SYN packets but also protects the controller from being overloaded by messages of those attack TCP connections and reduces the OpenFlow traffic between the OFS and the controller.

Comparing the interaction between the controller and system entities in CM shown as Figure 1 and in SSG described as in Figures 5 and 14, it can be seen that for each TCP connection, SSG needs only one request to the controller but this figure is two in CM. With such a difference, the total number of messages exchanged with the controller to install FEs for legitimate TCP connections in SSG solution would be a half of the number in CM mechanism. This enhancement diminishes the load on the controller and so makes SSG more resistant to SYN Flood attacks than the Avant-Guard CM mechanism.

VI. Conclusion

Inspired by the CM mechanism in Avant-Guard scheme, the proposed SSG solution moves the SYN proxy, which monitors TCP 3HS connection processes, inside the OFS to locate it in a separated device, termed the SD. The ability to match the TCP Flag fields specified in OpenFlow 1.5 is applied to filter related packets in the OFS and forward them to the SD for monitoring the 3HS process. SSG uses the RST Cookie technique to authenticate Source IP address instead of SYN Cookie as in CM. Besides, by integrating SYN-Flood attack detection module, SSG processes incoming SYN packets depending on the attack state of the destination application server. Thanks to these improvements, SSG overcomes the shortcomings of the CM mechanism and can be used as an alternative solution that can be applied to all OpenFlow 1.5 supported switches without any modification. The experiment results show that during attack-free state, SSG does not affect packet exchange between the external clients and the internal application servers. When an internal server is under SYN Flood attack, SSG consumes less resources than CM with a negligible total traffic increase at the OFS interfaces. This shows that SSG is more resistant than CM under saturation attack by SYN Flooding on the data and control planes.

References

[1] Open Networking Foundation, "SDN architecture overview Version 1.0," 2013. [Online]. Available: https://www.opennetworking.org/images/stories/downloads/sdn-resources/technical-reports/SDN-architecture-overview-1.0.pdf

[2] ——, "OpenFlow switch specification version 1.3.0 (Wire protocol 0x04)," 2012. [Online]. Available: https://www.opennetworking.org/wp-content/uploads/2014/10/openflow-spec-v1.3.0.pdf

[3] ——, "OpenFlow switch specification version 1.5.1 (Protocol Version 0x06)," 2015. [Online]. Available: https://www.opennetworking.org/wp-content/uploads/2014/10/openflow-switch-v1.5.1.pdf

[4] T. Peng, C. Leckie, and K. Ramamohanarao, "Survey of network-based defense mechanisms countering the DoS and DDoS problems," *ACM Computer Survey*, vol. 39, no. 1, Article 3, 2007.

[5] S. Scott-Hayward, G. O'Callaghan, and S. Sezer, "SDN security: A survey," in *Proceedings of the IEEE SDN for Future Networks and Services (SDN 4FNS)*, Trento, Italy, Nov. 11-13 2013, pp. 1–7.

[6] S. Shin and G. Gu, "Attacking software-defined networks: A first feasibility study," in *Proceedings of the Second ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking (HotSDN'13)*, Hong Kong, China, Aug. 16 2013, pp. 165–166.

[7] R. Kandoi and M. Antikainen, "Denial-of-service attacks in OpenFlow SDN networks," in *Proceedings of the IFIP/IEEE International Symposium on Integrated Network Management (IM)*, Ottawa, ON, Canada, May 11-15 2015, pp. 1322–1326.

[8] S. Shin, V. Yegneswaran, P. Porras, and G. Gu, "AVANT-GUARD: scalable and vigilant switch flow management in software-defined networks," in *Proceedings of the ACM SIGSAC Conference on Computer & Communications security (CCS'13)*, Berlin, Germany, Nov. 04 - 08, 2013, pp. 413–424.

[9] Daniel J. Bernstein, "SYN cookies." [Online]. Available: https://cr.yp.to/syncookies.html

[10] M. Ambrosin, M. Conti, F. De Gaspari, and R. Poovendran, "LineSwitch: Tackling control plane saturation attacks in software-defined networking," *IEEE/ACM Transactions on Networking*, vol. 25, no. 2, pp. 1206 – 1219, 2017.

[11] CAIDA: Center for Applied Internet Data Analysis, "The CAIDA anonymized internet traces 2013 dataset." [Online]. Available: http://www.caida.org/data/passive/passive_2013_dataset.xml

[12] P. Porras, S. Shin, V. Yegneswaran, M. Fong, M. Tyson, and G. Gu, "A security enforcement kernel for OpenFlow networks," in *Proceedings of the first workshop on Hot topics in software defined networks (HotSDN '12)*, Helsinki, Finland, Aug. 13 2012, pp. 121–126.

[13] L. Wei and C. Fung, "FlowRanger: A request prioritizing algorithm for controller DoS attacks in software defined networks," in *Proceedings of the IEEE International Conference on Communications (ICC)*, London, UK, Jun. 8-12, 2015, pp. 5254–5259.

[14] N.-N. Dao, J. Park, M. Park, and S. Cho, "A feasible method to combat against DDoS attack in SDN network," in *Proceedings of the International Conference on Information Networking (ICOIN)*, Siem Reap, Cambodia, Jan. 12-14 2015, pp. 309–311.

[15] R. Sahay, G. Blanc, Z. Zhang, and H. Debar, "Towards autonomic DDoS mitigation using software defined networking," in *Proceedings of the Workshop on Security of Emerging Networking Technologies*, San Diego, CA, Feb. 8 2015, p. 7 pages.

[16] S. Shin, Y. Song, T. Lee, S. Lee, J. Chung, P. Porras, V. Yegneswaran, J. Noh, and B. B. Kang, "Rosemary: A robust, secure, and high-performance network operating system," in *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS'14)*, Scottsdale, Arizona, USA, Nov. 03 - 07 2014, pp. 78–89.

[17] S. Hong, L. Xu, H. Wang, and G. Gu, "Poisoning network visibility in software-defined networks: New attacks and countermeasures," in *Proceedings of the Network and Distributed System Security Symposium*, San Diego, CA, Feb. 8-11 2015, p. 15 pages.

[18] M. Dhawan, R. Poddar, K. Mahajan, and V. Mann, "SPHINX: Detecting security attacks in software-defined networks," in *Proceedings of the Network and Distributed System Security Symposium*, San Diego, CA, Feb. 8-11 2015, p. 15 pages.

[19] D. Kreutz, F. M. Ramos, and P. Verissimo, "Towards secure

and dependable software-defined networks," in *Proceedings of the second ACM SIGCOMM workshop on Hot topics in software defined networking (HotSDN '13)*, Hong Kong, China, Aug. 16 2013, pp. 55–60.
[20] R. Kloti, V. Kotronis, and P. Smith, "OpenFlow: A security analysis," in *Proceedings of the 21st IEEE International Conference on Network Protocols (ICNP)*, Goettingen, Germany, Oct. 07-10 2013, pp. 1–6.
[21] T. Koponen, M. Casado, N. Gude, J. Stribling, L. Poutievski, M. Zhu, R. Ramanathan, Y. Iwata, H. Inoue, T. Hama, and S. Shenker, "Onix: A distributed control platform for large-scale production networks," in *Proceedings of the 9th USENIX Conference on Operating Systems Design and Implementation (OSDI'10)*, Vancouver, BC, Canada, Oct. 4–6 2010, pp. 351–364.
[22] A. Tootoonchian and Y. Ganjali, "HyperFlow: A distributed control plane for OpenFlow," in *Proceedings of the Internet Network Management Conference on Research on Enterprise Networking (INM/WREN'10)*, San Jose, CA, Apr. 28-30 2010, p. 6 pages.
[23] P. Berde, W. Snow, G. Parulkar, M. Gerola, J. Hart, Y. Higuchi, M. Kobayashi, T. Koide, B. Lantz, B. O'Connor, and P. Radoslavov, "ONOS: towards an open, distributed SDN OS," in *Proceedings of the Third Workshop on Hot Topics in Software Defined Networking (HotSDN'14)*, Chicago, Illinois, USA, Aug. 22 2014, pp. 1–6.
[24] A. Vishnoi, R. Poddar, V. Mann, and S. Bhattacharya, "Effective switch memory management in OpenFlow networks," in *Proceedings of the 8th ACM International Conference on Distributed Event-Based Systems (DEBS'14)*, Mumbai, India, May 26-29 2014, pp. 177–188.
[25] L. Dridi and M. F. Zhani, "SDN-guard: DoS attacks mitigation in SDN networks," in *Proceedings of the 5th IEEE International Conference on Cloud Networking (Cloudnet)*, Pisa, Italy, Aug. 17 2016, pp. 212–217.
[26] H. Wang, L. Xu, and G. Gu, "FloodGuard: A DoS attack prevention extension in software-defined networks," in *Proceedings of the 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks*, Rio de Janeiro, Brazil, Jun. 22-25 2015, pp. 239–250.
[27] M. Yu, J. Rexford, M. J. Freedman, and J. Wang, "Scalable flow-based networking with DIFANE," in *Proceedings of the ACM SIGCOMM 2010 Conference*, New Delhi, India, Aug. 30 - Sep. 03 2010, pp. 351–362.
[28] A. R. Curtis, J. C. Mogul, J. Tourrilhes, P. Yalagandula, P. Sharma, and S. Banerjee, "DevoFlow: Scaling flow management for high-performance networks," in *Proceedings of the ACM SIGCOMM conference*, Toronto, ON, Canada, Aug. 15 - 19 2011, pp. 254–265.
[29] H. Mekky, F. Hao, S. Mukherjee, Z.-L. Zhang, and T. Lakshman, "Application-aware data plane processing in SDN," in *Proceedings of the Third Workshop on Hot Topics in Software Defined Networking (HotSDN'14)*, Chicago, Illinois, USA, Aug 2014, pp. 13–18.
[30] D. Kotani and Y. Okabe, "A packet-in message filtering mechanism for protection of control plane in OpenFlow switches," *IEICE Transactions on Information and Systems*, vol. E99.D, no. 3, pp. 695–707, 2016.
[31] C. Schuba, I. Krsul, M. Kuhn, E. Spafford, A. Sundaram, and D. Zamboni, "Analysis of a denial of service attack on TCP," in *Proceedings of the IEEE Symposium on Security and Privacy*, Oakland, CA, USA, May 04-07 1997, pp. 208–223.
[32] C. Douligeris and A. Mitrokotsa, "DDoS attacks and defense mechanisms: Classification and state-of-the-art," *Computer Networks*, vol. 44, pp. 643–666, 2004.
[33] Eddy Wesley M., "Defenses against TCP SYN flooding attacks," *The Internet Protocol Journal*, vol. 9, no. 4, pp. 2–16, 2006.
[34] Postel, J., "Transmission control protocol, DAPRA internet program - protocol specification, rfc 793," 1981. [Online]. Available: https://tools.ietf.org/html/rfc793
[35] D. V. Tuyen and T.T. Huong, "A multi-criteria based software defined networking system architecture for DDoS-attack mitigation," *REV Journal on Electronics and Communications*, vol. 6, no. 3, pp. 50–60, 2017.
[36] The Linux Foundation Projects, "Data plane development kit (DPDK)." [Online]. Available: https://www.dpdk.org
[37] J. L. Deng, "Introduction to grey system theory," *The Journal of Grey Systems*, vol. 1, no. 1, pp. 1–24, 1989.
[38] T. Bohlin, *Practical Grey-box Process Identification: Theory and Applications (Advances in Industrial Control)*. Berlin, Heidelberg, Germany: Springer-Verlag, 2006.
[39] E. Kayacan, B. Ulutas, and O. Kaynak, "Grey system theory-based models in time series prediction," *Expert Systems with Applications*, vol. 37, no. 2, pp. 1784–1789, 2010.
[40] D. Zhang, H. Wang, and K. G. Shin, "Change-point monitoring for the detection of DoS attacks," *IEEE Transactions on Dependable and Secure Computing*, vol. 1, no. 4, pp. 193–208, 2004.
[41] J. L. Hellerstein, F. Zhang, and P. Shahabuddin, "A statistical approach to predictive detection," *Computer Networks*, vol. 35, no. 1, pp. 77–95, 2001.
[42] S. Wang, Q. Sun, H. Zou, and F. Yang, "Detecting SYN flooding attacks based on traffic prediction: A demonstration of the security communication networks class file," *Security and Communication Networks*, vol. 5, no. 10, pp. 1131–1140, 2012.
[43] The Linux Foundation Collaborative Project, "Open vSwitch." [Online]. Available: https://www.openvswitch.org
[44] Project Floodlight, "Floodlight OpenFlow controller." [Online]. Available: http://www.projectfloodlight.org/floodlight
[45] M. Goldstein, "BoNeSi: the DDoS botnet simulator." [Online]. Available: https://github.com/Markus-Go/bonesi
[46] W. Foundation, "Wireshark." [Online]. Available: https://www.wireshark.org
[47] Fred Klassen, "TCPreplay-PCAP editing and replaying utilities." [Online]. Available: https://tcpreplay.appneta.com

Dang Van Tuyen obtained his Bachelor degree in Electronics and Telecommunications in 1999 and Master degree in Telecommunication Engineering in 2008 from Hanoi University of Science and Technology (HUST), Vietnam. He is a lecturer at the Faculty of Electronics and Telecommunications, People's Police University of Technology and Logistics, Ministry of Public Security, Vietnam. Currently, he is pursuing the PhD program in the field of Telecommunication Engineering of HUST. His research interest includes: network security, SDN network, quality of service, wireless sensor network.

Truong Thu Huong is Associate Professor of the School of Electronics and Telecommunication, Hanoi University of Science and Technology (HUST). She is also Vice Director of Elitech, a project to promote all elite education programs of HUST. Truong's educational, research, and development work is oriented toward next generation networks, protocols and mechanism, traffic analyses, QoE/QoS measuring, green networking and deployment of new integrated multimedia services into fixed and mobile networks, network security and applications in the Internet of Things. She has been serving various international research conferences in different roles such as TPC member, publicity chair, organizer, track chair and serving international journals as reviewer. She is also active in capacity building; research and education development.

Performance Analysis of Repeated Index Modulation with Coordinate Interleaving over Nakagami-m Fading Channel

Le Thi Thanh Huyen, Tran Xuan Nam

Le Quy Don Technical University, Hanoi, Vietnam

Correspondence: Le Thi Thanh Huyen, huyen.ltt@mta.edu.vn

Communication: received 17 May 2019, revised 21 June 2019, accepted 22 June 2019

Online early access: 25 June 2019, Digital Object Identifier: 10.32913/mic-ict-research.v2019.n1.863

The Area Editor coordinating the review of this article and deciding to accept it was Dr. Truong Trung Kien

Abstract: In this paper, a new Space-Time Block Coded Spatial Modulation (SM) scheme based on the Golden Code, called the In this paper, we evaluate the symbol error performance of an extended Index Modulation for Orthogonal Frequency Division Multiplexing (IM-OFDM), namely repeated index modulation-OFDM with coordinated interleaving (abbreviated as ReCI), over the Nakagami-m fading channel. The ReCI system attains higher error performance than the conventional IM-OFDM with coordinate interleaving (IM-OFDM-CI). In order to investigate the system performance over the Nakagami-m fading channel, we derive the closed-form expressions for the symbol error probability (SEP) and the bit error probability (BEP). The analytical results give interesting insights into the dependence of SEP on system parameters. Their tightness is also validated by numerical results, which show that our proposed scheme can provide considerably better error performance than the conventional IM-OFDM and IM-OFDM-CI at the same spectral efficiency.

Keywords: *Index modulation, IM-OFDM, coordinate interleaving, pair-wise error probability, symbol error probability, maximum likelihood detection, Nakagami-m fading.*

I. Introduction

Index modulation for orthogonal frequency division multiplexing (IM-OFDM) has emerged as an effective index modulation scheme in the frequency domain as it achieves not only higher energy efficiency but also better error performance than the conventional OFDM. The IM-OFDM system activates only a subset of sub-carriers for transmission. In comparison with the classical OFDM, the IM-OFDM system uses both the M-ary modulated symbols and indices of active sub-carriers to convey information bits.

In the early IM-OFDM system [1], a fixed number of information bits were employed to activate the sub-carriers, thus its spectral efficiency and error performance were limited. In order to overcome this disadvantage, an IM-OFDM scheme with adjustable active sub-carriers according to incoming bits was proposed in [2]. In another effort, the sub-carriers were interleaved in [3] to extend the Euclidean distance among the complex data symbols to deal with the error performance issue. Moreover, the study in [4] proposed the interleaved sub-carrier grouping and investigated the achievable rate of the system.

Regarding the spectral efficiency issue, the IM-OFDM-I/Q scheme in [5] performed joint index modulation over both the in-phase and quadrature components of the M-ary modulated symbols. In another solution, the dual-mode OFDM (DM-OFDM) was introduced in [6]. This model utilized inactive sub-carriers to carry more data bits together with active sub-carriers. Different signal constellations were used for complex data symbols over the active and inactive sub-carriers. The work in [7] introduced a multi-mode IM-OFDM (MM-IM-OFDM) scheme which used all sub-carriers and extra information bits were conveyed by permutation of transmission modes, thus attaining further increase in spectral efficiency. In order to achieve not only spectral efficiency but also diversity gain, the authors in [8] successfully designed a linear constellation precoder (LPC) for IM-OFDM. The recent study in [9] attained an improvement for both diversity gain and energy efficiency by applying compressed sensing to IM-OFDM.

Meanwhile, there were also a variety of researches focused on analyzing the performance of IM-OFDM. The authors in [10] provided a frame work for deriving a tight bound for BER of IM-OFDM. The work [11] evaluated the outage probability of the IM-OFDM system over two-way diffused-power fading channels. The transmission reliability of IM-OFDM and IM-OFDM employing greedy detection under imperfect channel state information (CSI) were investigated through SEP in [12] and [13], respectively. Besides, an opportunistic scheduling IM-OFDM (OS-IM-OFDM) scheme which exploits multi-user diversity gain to improve the system performance of IM-OFDM is reported in [14]. Diversity and coding gains of OS-IM-OFDM and OS-OFDM were also compared in this study.

Focusing on diversity gain issue, the coordinate interleaved IM-OFDM scheme in [15] distributed the real and imaginary components of the M-ary modulated data symbols over distinctive sub-carriers. The paper [16] introduced an IM-OFDM scheme with transmit diversity, which utilized multiple signal constellations to carry the same data bits over the active sub-carriers. In the recent work [17], the coded IM-OFDM with transmit diversity (TD-IM-OFDM) was proposed to increase the reliability for index detection. The authors of [18] introduced a novel spread IM-OFDM scheme to improve the transmit diversity and low-complexity detectors were also designed in this work. To further improve the diversity gain and error performance, the IM-OFDM concept was extended to MIMO systems in [19, 20]. In order to reduce complexity while still improving the diversity gain of IM-OFDM, the study in [21] introduced the IM-OFDM with greedy detection and diversity reception. Its BER performance under imperfect CSI was analyzed in [22]. The repeated IM-OFDM with transmit diversity (ReMO) and the closed-form expressions for its SEP and BER were presented in [23].

Recently, there was a proposal to apply deep learning to detect data bits of the IM-OFDM systems in [24]. The proposed deep-learning assisted detector could provide a near optimal performance as the ML and considerably reduces the runtime of existing detectors.

Aiming at improving the error performance of the conventional IM-OFDM-CI system, we proposed a repeated index modulation-OFDM with coordinated interleaving scheme in [25], which is abrreviated as ReCI. In this scheme, coordinate interleaving is simultaneously applied to the M-ary modulated symbols in two different clusters. Additionally, these distinct clusters repeatedly utilize the same set of active sub-carrier indices. Combining coordinate interleaving and the index repetition allows ReCI to attain better index detection and the M-ary modulated symbol error performance over the conventional IM-OFDM and IM-OFDM-CI systems, even at the same spectral efficiency. Furthermore, while IM-OFDM-CI requires an even number of active sub-carriers, our proposed scheme can work well with an arbitrary number of sub-carriers and thus is more flexible in terms of achieving error performance and spectral efficiency. However, this work lacks theoretical analysis on SEP and BEP performance. Additionally, the impacts of system parameters on the transmission reliability were also not properly analyzed. Besides, most previous IM-OFDM studies only investigated the system performance over the Rayleigh fading channel. For realistic conditions, received signal intensity follows the Nakagami-m distribution, which is more suitable for multi-path channels than the Rayleigh. The Nakagami-m fading channel is most suitable for the urban, suburban, as well as both the indoor and outdoor propagation model [26].

In this paper, we analyze performance of ReCI in terms of SEP and BEP. The analytical results allow us to obtain an helpful insight into the impact of system parameters on the error performance. Thus, it is possible for us to design a system configuration with the best error performance.

The remainder of this paper is organized as follows. Section II describes the system model of the proposed ReCI. The error performance is analyzed in Section III. Simulation results and discussions are given in Section IV. Finally, the conclusion is given in Section V.

Notation: Vectors and matrices are denoted by bold letters. $C(.,.)$ and $\lfloor \cdot \rfloor$ present the binomial coefficient and the floor function, respectively. $\mathbb{E}\{\cdot\}$ is the expectation operation. $(\cdot)^R$ and $(\cdot)^I$ denote the real and imaginary parts of a complex number, respectively. The moment generating function is denoted as $\mathcal{M}(\cdot)$.

II. System Model

The block diagram of one ReCI sub-block is depicted in Figure 1. Assume that the system with N_t sub-carriers is split into G sub-blocks of N_G sub-carriers, i.e., $N_t = GN_G$. Then, each sub-block is partitioned into two clusters of N sub-carriers, i.e., $N_G = 2N$. Since the operation in each sub-block is independent, without loss of generality, we consider only one sub-block.

In each sub-block transmission, p incoming bits are divided into two bit sequences. The first $p_1 = \lfloor \log_2 (C(N,K)) \rfloor$ bits are sent to an index selector to determine K out of N sub-carriers for each cluster, using either a look-up table (LUT) or the combinatorial method [2]. Similar to the IM-OFDM system, an additional number of information bits is transfered through the indices of active sub-carriers. The remaining $N - K$ sub-carriers are set to zero. The output of the index selector is a set of activated sub-carrier indices θ, i.e., $\theta = \{\alpha_1, \ldots, \alpha_K\}$, where $\alpha_k = \{1, \ldots, N\}$ and $k = \{1, \ldots, K\}$. Let $\mathcal{I}$ denote the set of possible active indices. For a given N and K, there are a total of $c = 2^{\lfloor \log_2 C(N,K) \rfloor}$ combinations of active indices. Different from the conventional IM-OFDM-CI, ReCI employs the same set of active sub-carrier indices θ for two clusters in one sub-block as illustrated in Figure 1. It is noteworthy that such index repetition can improve the

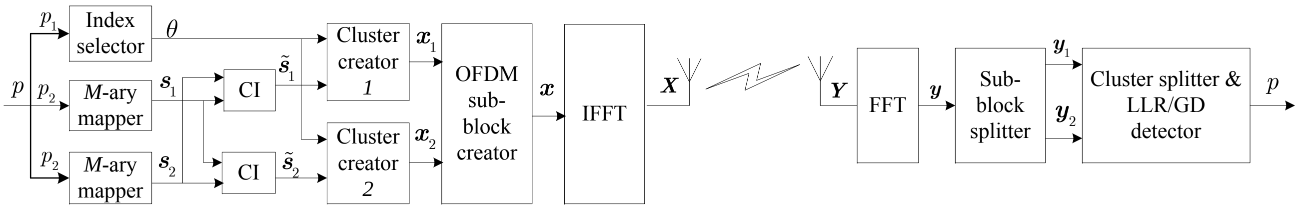


Figure 1. Block diagram of one ReCI sub-block.

index error probability (IEP) over the conventional scheme at the cost of spectral efficiency.

The second bit sequence is equally separated into two clusters of $p_2 = K\log_2 M$ bits. By repeating the indices of sub-carriers between two clusters, we can jointly apply the coordinate interleaving technique to the M-ary modulated symbols from two distinguishable clusters. For cluster i, $p_2 = K\log_2 M$ bits are mapped into a vector of K M-ary modulated symbols $\mathbf{s}_i \in \mathbb{C}^{K\times 1}$, for $i = 1, 2$. Let $\mathbf{s}_1 = [a_1\ a_2\ \dots\ a_K]^T$ and $\mathbf{s}_2 = [b_1\ b_2\ \dots\ b_K]^T$. Prior to the coordinate interleaving, all M-ary modulated symbols in $\mathbf{s}_i$ are rotated by an angle ϕ which is defined based on the modulation type. Let $\mathcal{S}^\phi$ denote the rotated constellation. The values of ϕ can be selected by the computer search method [15]. For example, the quadrature amplitude modulation (QAM) with constellation size $M = \{2, 4, 16, 64\}$, rotation angles are respectively given by $\phi = \{45°, 15°, 8.5°, 4.5°\}$. After coordinate interleaving between the symbols in $\mathbf{s}_1$ and $\mathbf{s}_2$ from different clusters, the symbol vectors in each cluster, $\tilde{\mathbf{s}}_i \in \mathbb{C}^{K\times 1}$, are obtained as follows:

$$\tilde{\mathbf{s}}_1 = \begin{bmatrix} c_{1,1} \\ c_{1,2} \\ \vdots \\ c_{1,K-1} \\ c_{1,K} \end{bmatrix} = \begin{bmatrix} a_1^R + jb_1^I \\ a_2^R + jb_2^I \\ \vdots \\ a_{K-1}^R + jb_{K-1}^I \\ a_K^R + jb_K^I \end{bmatrix}, \tag{1}$$

$$\tilde{\mathbf{s}}_2 = \begin{bmatrix} c_{2,1} \\ c_{2,2} \\ \vdots \\ c_{2,K-1} \\ c_{2,K} \end{bmatrix} = \begin{bmatrix} b_1^R + ja_1^I \\ b_2^R + ja_2^I \\ \vdots \\ b_{K-1}^R + ja_{K-1}^I \\ b_K^R + ja_K^I \end{bmatrix}. \tag{2}$$

where $j = \sqrt{-1}$. Using $\tilde{\mathbf{s}}_i$ and θ, the transmitted codeword over N sub-carriers $\mathbf{x}_i = [x_{i1}, \dots, x_{iN}]^T$ is generated. Since only K out of N sub-carriers are activated, K symbols corresponding to K active sub-carriers are nonzero, i.e., $x_{i\alpha_k} = c_{i,k}$ when $\alpha_k \in \theta$, and $x_{i\alpha} = 0$ when $\alpha \notin \theta$, where $\alpha_k = \{1, \dots, N\}$, $k = \{1, \dots, K\}$, $i = 1, 2$. An example of a lookup table to determine the transmitted codewords in each cluster when $N = 4$, $K = 2$, $p_1 = 2$ is presented in Table I. It can be seen that for each data symbol in a cluster, its real and imaginary components are transferred over different sub-carriers. In addition, combining the index repetition and the joint coordinate interleaving allows ReCI to activate an arbitrary number of sub-carriers, leading to higher flexibility in terms of the transmission reliability and spectral efficiency than the conventional IM-OFDM-CI.

The OFDM sub-block creator takes into account $\mathbf{x}_1$ and $\mathbf{x}_2$ to generate the transmitted signal per sub-block $\mathbf{x} = \left[\mathbf{x}_1^T\ \mathbf{x}_2^T\right]^T$. For each ReCI sub-block, the received signal in the frequency domain can be expressed as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{3}$$

where $\mathbf{y} = [\mathbf{y}_1\ \mathbf{y}_2]^T$, $\mathbf{n} = \left[\mathbf{n}_1^T\ \mathbf{n}_2^T\right]^T$, and

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2 \end{bmatrix}.$$

The components $\mathbf{y}_i = [y_{i1}, \dots, y_{iN}]$ and $\mathbf{H}_i = \mathrm{diag}\{h_{i1}, \dots, h_{iN}\}$, for $i = 1, 2$, are the received signal and channel matrix of the cluster i, respectively. The noise component is given by $\mathbf{n}_i = [n_{i1}, \dots, n_{iN}]^T$. The channel coefficient over each sub-carrier is represented by $h_{i\alpha}$. Each element $|h_{i\alpha}|^2$ is a Gamma distributed random variable with $|h_{i\alpha}|^2 \sim \mathcal{G}(m, \beta)$, where m is the distribution parameter and β is defined as

$$\beta = \frac{m}{\mathbb{E}\{|h_{i\alpha}|^2\}}. \tag{4}$$

Noise on each sub-carrier follows the distribution $n_{i\alpha} \sim \mathcal{CN}(0, N_0)$, where N_0 represents the noise variance. The average signal to noise ratio (SNR) per sub-carrier is determined by $\bar{\gamma} = \omega E_s / N_0$, where ωE_s denotes the average transmit power per M-ary modulated symbol, i.e., $\mathbb{E}\{|s|^2\} = \omega E_s$. The power allocation factor and the average power per sub-carrier are respectively denoted by $\omega = N/K$ and E_s.

Consequently, the total number of information bits transmitted per sub-block is $p = p_1 + 2p_2$ bits. As a result, the spectral efficiency of the ReCI system is given by

$$\eta = \frac{\lfloor \log_2 (C(N, K)) \rfloor + 2K\log_2 M}{2N}\ \ [\text{b/s/Hz}]. \tag{5}$$

In order to estimate the transmitted signal, the receiver uses an ML detector to jointly detect the indices of active sub-carriers and the corresponding M-ary modulated

TABLE I
EXAMPLE OF LUT WITH $N = 4$, $K = 2$, $p_1 = 2$

p_1	θ	$\mathbf{x}_1^T$	$\mathbf{x}_2^T$
00	[1,2]	$\left[a_1^R + jb_1^I \quad a_2^R + jb_2^I \quad 0 \quad 0\right]$	$\left[b_1^R + ja_1^I \quad b_2^R + ja_2^I \quad 0 \quad 0\right]$
01	[2,3]	$\left[0 \quad a_1^R + jb_1^I \quad a_2^R + jb_2^I \quad 0\right]$	$\left[0 \quad b_1^R + ja_1^I \quad b_2^R + ja_2^I \quad 0\right]$
10	[2,4]	$\left[0 \quad a_1^R + jb_1^I \quad 0 \quad a_2^R + jb_2^I\right]$	$\left[0 \quad b_1^R + ja_1^I \quad 0 \quad b_2^R + ja_2^I\right]$
11	[1,3]	$\left[a_1^R + jb_1^I \quad 0 \quad a_2^R + jb_2^I \quad 0\right]$	$\left[b_1^R + ja_1^I \quad 0 \quad b_2^R + ja_2^I \quad 0\right]$

symbols for both clusters. The ML detector selects the codeword that minimizes the following decision metric:

$$\left(\hat{\theta}, \hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2\right) = \arg\min_{\theta, \mathbf{s}_1, \mathbf{s}_2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2. \tag{6}$$

III. PERFORMANCE ANALYSIS

1. SEP Derivation

In this section, we derive the closed-form expression for SEP of the ReCI system using ML detection. SEP is defined by the ratio of the number of error symbols to the total number of transmitted index and data symbols. Following the frame work in [10], SEP of the ReCI system is given by

$$P_s = \frac{P_I + KP_M}{K+1}, \tag{7}$$

where P_I and P_M denotes the index and the M-ary modulated symbol error probability, respectively. We calculate SEP of the M-ary modulated data symbols in ReCI by using the pair-wise error probability (PEP) of modulated symbols. PEP is determined by the probability that a transmitted symbol a_1 is mistakenly detected by symbol $\hat{a}_1$. Particularly, the received signals per sub-carrier of two clusters are respectively rewritten as

$$y_{1\alpha} = h_{1\alpha} x_{1\alpha} + n_{1\alpha}, \tag{8}$$

$$y_{2\alpha} = h_{2\alpha} x_{2\alpha} + n_{2\alpha}, \tag{9}$$

where $h_{i\alpha} \sim \mathcal{CN}(0,1)$, $n_{i\alpha} \sim \mathcal{CN}(0, N_0)$, $\alpha = 1, \ldots, N$, and $i = 1, 2$. For presentation convenience, we will ignore the sub-carrier index α in the following derivations. Then, we can rewrite $y_{1\alpha}$, $y_{2\alpha}$ as follows:

$$\begin{aligned} y_1 &= \left(h_1^R a_1^R - h_1^I b_1^I + n_1^R\right) + j\left(h_1^I a_1^R + h_1^R b_1^I + n_1^I\right) \\ &= y_1^R + j y_1^I, \end{aligned} \tag{10}$$

$$\begin{aligned} y_2 &= \left(h_2^R b_1^R - h_2^I a_1^I + n_2^R\right) + j\left(h_2^I b_1^R + h_2^R a_1^I + n_2^I\right) \\ &= y_2^R + j y_2^I. \end{aligned} \tag{11}$$

As a result, we can deduce

$$\begin{aligned} h_1^R y_1^R + h_1^I y_1^I &= \left[\left(h_1^R\right)^2 + \left(h_1^I\right)^2\right] a_1^R + h_1^R n_1^R + h_1^I n_1^I \\ &= |h_1|^2 a_1^R + \overline{n}_1^R, \end{aligned} \tag{12}$$

$$\begin{aligned} h_2^R y_2^I - h_2^I y_2^R &= \left[\left(h_2^R\right)^2 + \left(h_2^I\right)^2\right] a_1^I + h_2^R n_2^I - h_2^I n_2^R \\ &= |h_2|^2 a_1^I + \overline{n}_1^I. \end{aligned} \tag{13}$$

Then, we obtain

$$\begin{aligned} \tilde{y}_1 &= \left(\frac{h_1^R y_1^R + h_1^I y_1^I}{|h_1|}\right) + j\left(\frac{h_2^R y_2^I - h_2^I y_2^R}{|h_2|}\right) \\ &= |h_1| a_1^R + j|h_2| a_1^I + \tilde{n}_1^R + j\tilde{n}_1^I \\ &= |h_1| a_1^R + j|h_2| a_1^I + \tilde{n}_1, \end{aligned} \tag{14}$$

where $\tilde{n}_1^R = \overline{n}_1^R / |h_1|$, $\tilde{n}_1^I = \overline{n}_1^I / |h_2|$, and both $\tilde{n}_1^R$ and $\tilde{n}_1^I$ have the same distribution $\mathcal{N}\left(0, \frac{N_0}{2}\right)$. Thus, the distribution of $\tilde{n}_1 = \tilde{n}_1^R + j\tilde{n}_1^I$ is given by $\tilde{n}_1 \sim \mathcal{CN}(0, N_0)$. Following Equation (5) of [27], the conditional PEP of the ReCI system can be computed as follows:

$$\begin{aligned} P(a_1 \to \hat{a}_1 \mid h_1, h_2) &= \Pr\left[\left|\tilde{y}_1 - |h_1|\hat{a}_1^R - j|h_2|\hat{a}_1^I\right|^2 \right. \\ &\qquad \left. < \left|\tilde{y}_1 - |h_1| a_1^R - j|h_2| a_1^I\right|^2 = |\tilde{n}_1|^2\right] \\ &= Q\left(\sqrt{\frac{|h_1|^2 \Delta_R^2 + |h_2|^2 \Delta_I^2}{2N_0}}\right) \\ &= Q\left(\sqrt{\frac{\gamma_1 \Delta_R^2 + \gamma_2 \Delta_I^2}{2}}\right), \end{aligned} \tag{15}$$

where $\Delta_R^2 = |a_1^R - \hat{a}_1^R|^2$ and $\Delta_I^2 = |a_1^I - \hat{a}_1^I|^2$. Assume that the average SNR is given by $\overline{\gamma} = 1/N_0$, let us denote $\gamma_1 = \overline{\gamma}|h_1|^2$ and $\gamma_2 = \overline{\gamma}|h_2|^2$.

Applying the approximation of the Q-function [28] as $Q(t) \approx \frac{1}{12} e^{-\frac{t^2}{2}} + \frac{1}{4} e^{-\frac{2}{3}t^2}$, the average PEP of the M-ary modulated data symbol is obtained by

$$P(a_1 \to \hat{a}_1) = \mathbb{E}\left\{Q\left(\sqrt{\Omega}\right)\right\} = \mathbb{E}\left\{\frac{1}{12} e^{-\frac{\Omega}{2}} + \frac{1}{4} e^{-\frac{2\Omega}{3}}\right\}, \tag{16}$$

where $\Omega = \frac{\gamma_1 \Delta_R^2 + \gamma_2 \Delta_I^2}{2}$.

Using the moment generating function (MGF) for the Nakagami-m fading channel [29]

$$\mathcal{M}_\gamma(z) = \mathbb{E}_\gamma\left\{e^{-z\gamma}\right\} = \left(1 - \frac{z\gamma}{m}\right)^{-m}, \tag{17}$$

the MGF of Ω is given by

$$\mathcal{M}_\Omega(z) = \frac{1}{\left(1 - \frac{\bar{\gamma}\Delta_R^2 z}{2m}\right)^m \left(1 - \frac{\bar{\gamma}\Delta_I^2 z}{2m}\right)^m}. \tag{18}$$

Following the MGF approach, the average PEP of the M-ary modulated data symbol is given by

$$\begin{aligned} P\left(a_1 \to \hat{a}_1\right) &= \frac{\mathcal{M}_\Omega\left(-\frac{1}{2}\right)}{12} + \frac{\mathcal{M}_\Omega\left(-\frac{2}{3}\right)}{4} \\ &= \frac{1}{12}\frac{1}{\left(1 + \frac{\bar{\gamma}\Delta_R^2}{4m}\right)^m}\frac{1}{\left(1 + \frac{\bar{\gamma}\Delta_I^2}{4m}\right)^m} \\ &\quad + \frac{1}{4}\frac{1}{\left(1 + \frac{\bar{\gamma}\Delta_R^2}{3m}\right)^m}\frac{1}{\left(1 + \frac{\bar{\gamma}\Delta_I^2}{3m}\right)^m}. \end{aligned} \tag{19}$$

Using the union bound, the M-ary modulated symbol error probability is determined by

$$P_M = \frac{1}{M}\sum_{a_1 \in \mathcal{S}^\phi}\sum_{\hat{a}_1 \neq a_1} P\left(a_1 \to \hat{a}_1\right). \tag{20}$$

Generally, the erroneous symbol occurs when index and/or data symbols are erroneously detected at the receiver. An error index symbol causes the data symbol corresponding to the index error symbol to become erroneous. Since different clusters in the ReCI system employ a same set of active indices, this index repetition allows P_I to achieve higher diversity gain than P_M. Therefore, SEP of the ReCI system can be approximated by

$$P_s \approx \frac{KP_M}{K+1}. \tag{21}$$

2. Asymptotic Analysis

In the high SNR region, the average SEP of the ReCI system can be approximated by

$$P_s \approx \frac{K\left((4m)^{2m} + 3\,(3m)^{2m}\right)}{12\,(K+1)}\frac{1}{\left(\Delta_I^2\Delta_R^2\right)^m}\bar{\gamma}^{-2m}. \tag{22}$$

Remark 1: It can be seen from (22) that the ReCI system achieves the diversity order of $2m$. Besides, the error performance depends on K, the reliability of the ReCI system deteriorates when K increases and choosing a small K allows ReCI to improve its error performance. This statement will be validated using simulation results in Section IV.

Remark 2: As can be seen from (22) that at a given K, SEP of ReCI is strongly influenced by the product $\Delta_I^2\Delta_R^2$. The best SEP is achieved by maximizing $\Delta = \Delta_I^2\Delta_R^2$.

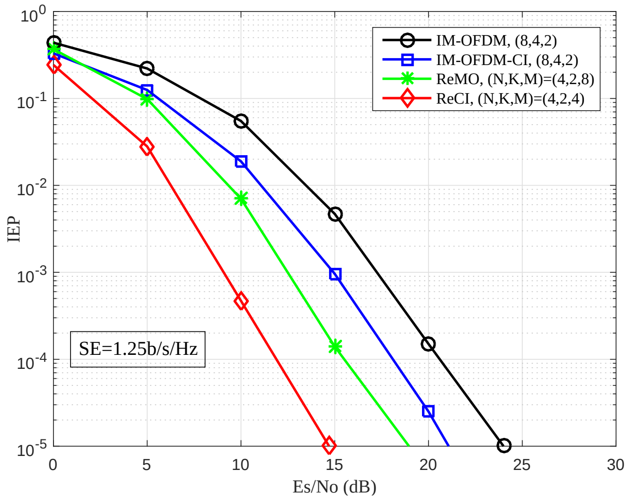


Figure 2. Index error performance of ReCI in comparison with the conventional IM-OFDM, IM-OFDM-CI and ReMO systems at the spectral efficiency (SE) of 1.25 b/s/Hz when $m = 2$, $M = \{2, 4, 8\}$, $N = \{4, 8\}$, $K = \{2, 4\}$.

Remark 3: Based on the framework to derive BER of IM-OFDM in [12], the approximated average BER of ReCI can be obtained by

$$P_b \approx \frac{KP_M}{p_I + p_c}. \tag{23}$$

IV. Simulation Results and Discussions

This section presents analytical and simulation results of the ReCI scheme in comparison with IM-OFDM [2], IM-OFDM-CI [15], and ReMO [23]. The ML detector is employed in all the addressed schemes. Assume that the channel over each sub-carrier is flat Rayleigh fading channel. For simplicity, a system configuration with N sub-carriers, K active sub-carriers and modulation order M is referred to as (N, K, M).

Figure 2 compares the index error probability (IEP) of the proposed ReCI, the conventional IM-OFDM, IM-OFDM-CI and ReMO system at the same spectral efficiency of 1.25 b/s/Hz when $m = 2$, $M = \{2,4,8\}$, $N = \{4,8\}$, $K = \{2,4\}$. It can be seen that the proposed scheme significantly improves the IEP performance of the reference systems. Particularly, at IEP of 10^{-4}, the proposed scheme attains SNR gains of 9 dB, 5.5 dB, 3.5 dB over the IM-OFDM, IM-OFDM-CI, and ReMO systems, respectively. Since the proposed scheme jointly uses index repetition and coordinate interleaving, it can achieve a better diversity gain in the index domain than IM-OFDM, IM-OFDM-CI, and ReMO.

Figure 3 reports the SEP performance comparison between the ReCI, IM-OFDM, IM-OFDM-CI, and ReMO systems at the same spectral efficiency of 1.25 b/s/Hz

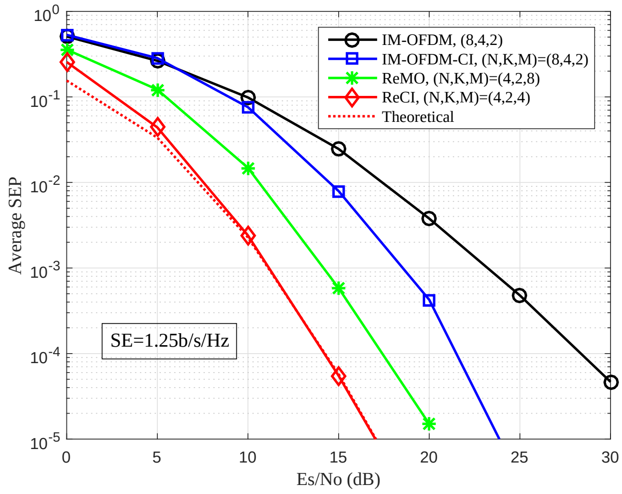


Figure 3. Comparison between the SEP performance of ReCI, IM-OFDM and IM-OFDM-CI at the spectral efficiency of 1.25 b/s/Hz when $m = 2$, $M = \{2, 4, 8\}$, $N = \{4, 8\}$, $K = \{2, 4\}$.

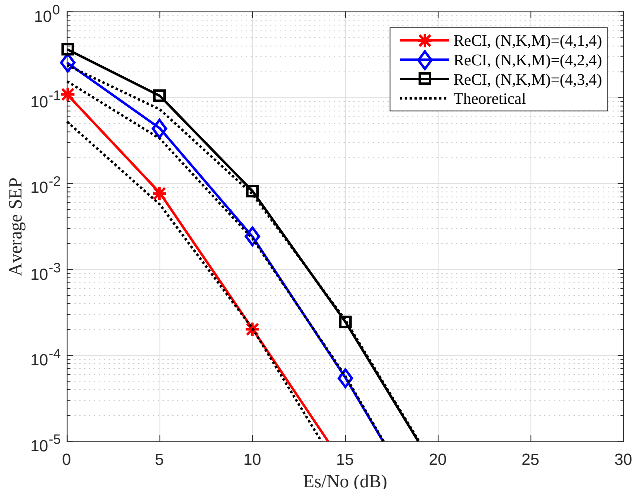


Figure 5. Impact of K on the performance of ReCI when $m = 2$, $M = 4$, $N = 4$, $K = \{1, 2, 3\}$.

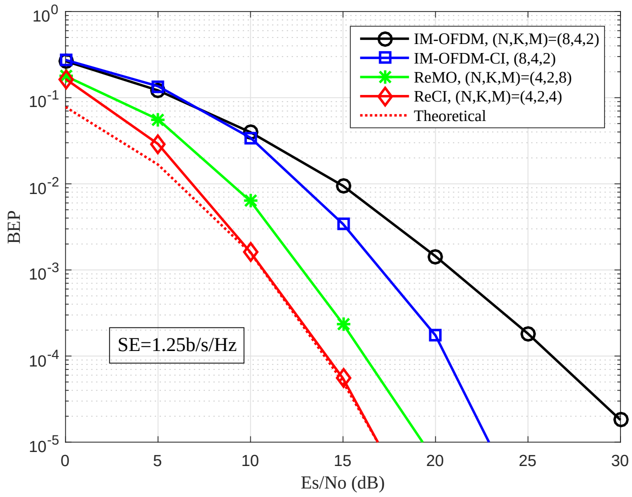


Figure 4. BEP comparison between the ReCI scheme and benchmarks at the spectral efficiency of 1.25 b/s/Hz when $m = 2$, $M = \{2, 4, 8\}$, $N = \{4, 8\}$, $K = \{2, 4\}$.

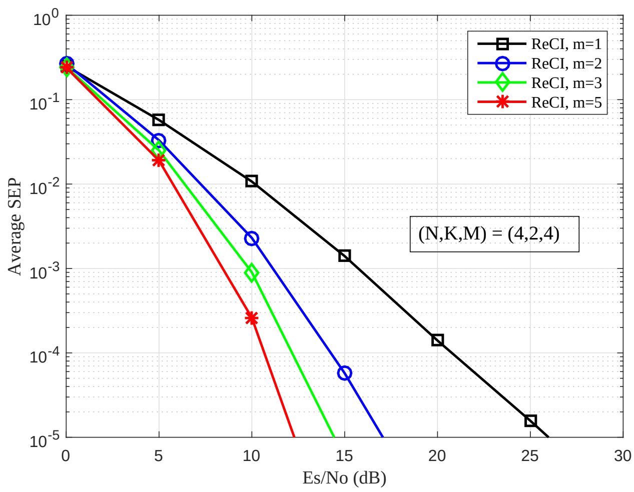


Figure 6. Impact of m on the SEP performance of ReCI when $m = \{1, 2, 3, 5\}$, $M = 4$, $N = 4$, $K = 2$.

when $m = 2$, $M = \{2, 4, 8\}$, $N = \{4, 8\}$, $K = \{2, 4\}$. As seen from the figure, at the same spectral efficiency and SEP of 10^{-4}, the ReCI scheme provides an SNR gain of about 14.5 dB, 6.5 dB, and 2.5 dB over IM-OFDM, IM-OFDM-CI, and ReMO, respectively. This achievable improvement is due to the IEP improvement, as shown in Figure 2, which can reduce errors of the M-ary modulated symbol detection, leading to improvement in the overall error probability compared with the benchmark schemes.

Figure 4 depicts the BEP comparison between ReCI, IM-OFDM, IM-OFDM-CI and ReMO when $m = 2$, $M = \{2, 4, 8\}$, $N = \{4, 8\}$, $K = \{2, 4\}$. All considered schemes use the ML detection. As can be seen from this figure, at the same spectral efficiency of 1.25 b/s/Hz, the proposed ReCI scheme achieves better BEP performance than the reference systems. For example, at BEP of 10^{-4}, the proposed scheme can obtain SNR gains of about 13 dB, 5.5 dB and 1.5 dB over IM-OFDM, IM-ODFM-CI and ReMO, respectively. Moreover, the tightness of theoretical and simulation curves validates the accuracy of the closed-form expression for BEP derived for the proposed scheme in (23).

The accuracy of the approximate SEP expressions in (21) and (22) is validated by using the simulation results for various system parameters in Figure 5. As shown in the figure, the gap between the theoretical and simulation SEP curves are very tight in the large SNR regime for all system configurations. Therefore, the derived closed-form expression of SEP in (21) can be used to evaluate performance of ReCI working at the high SNR regime. Moreover, it can be observed from Figure 5 that the SEP performance deteriorates as K increases. Clearly, choosing $K = 1$ allows ReCI to achieve the highest transmission reliability. Thus, it is recommended to select a small number of active sub-carriers K for better system performance. This validates Remark 1.

The impact of the channel parameter m on the SEP performance of ReCI is represented in Figure 6. It can be observed that the diversity gain of ReCI is strongly depended on m. As expected in Remark 1, when the value of m increases, the system attains better SEP performance. The worst case occurs with $m = 1$, which is equivalent to the Rayleigh fading channel.

V. Conclusions

This paper has analyzed the error performance of the repeated index modulation-OFDM with coordinated interleaving in terms of both SEP and BEP performances. The proposed scheme provides higher reliability and flexibility than the conventional IM-OFDM-CI system. By applying the same set of active sub-carrier indices in two distinguishable clusters and jointly employing the coordinate interleaving to the M-ary modulated symbols in them, the proposed ReCI scheme can improve both the index and symbol error performances while having higher flexibility between the transmission reliability and the spectral efficiency. The derived closed-form expressions gives an insight into the influence of the system parameters on the error performance. It allows to select the system configuration with better system performance. The analytical and simulation results clearly confirmed the advantages of the proposed scheme over the reference systems.

References

[1] R. Abu-Alhiga and H. Haas, "Subcarrier-index modulation OFDM," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*. IEEE, Sep. 2009, pp. 177–181.

[2] E. Başar, Ü. Aygölü, E. Panayırcı, and H. V. Poor, "Orthogonal frequency division multiplexing with index modulation," *IEEE Transactions on Signal Processing*, vol. 61, no. 22, pp. 5536–5549, Aug. 2013.

[3] Y. Xiao, S. Wang, L. Dan, X. Lei, P. Yang, and W. Xiang, "OFDM with interleaved subcarrier-index modulation," *IEEE Communications Letters*, vol. 18, no. 8, pp. 1447–1450, Jun. 2014.

[4] M. Wen, X. Cheng, M. Ma, B. Jiao, and H. V. Poor, "On the achievable rate of OFDM with index modulation," *IEEE Transactions on Signal Processing*, vol. 64, no. 8, pp. 1919–1932, Apr. 2016.

[5] B. Zheng, F. Chen, M. Wen, F. Ji, H. Yu, and Y. Liu, "Low-complexity ML detector and performance analysis for OFDM with in-phase/quadrature index modulation," *IEEE Communications Letters*, vol. 19, no. 11, pp. 1893–1896, Nov. 2015.

[6] T. Mao, Z. Wang, Q. Wang, S. Chen, and L. Hanzo, "Dual-mode index modulation aided OFDM," *IEEE Access*, vol. 5, pp. 50–60, Feb. 2017.

[7] M. Wen, E. Basar, Q. Li, B. Zheng, and M. Zhang, "Multiple-mode orthogonal frequency division multiplexing with index modulation," *IEEE Transactions on Communications*, vol. 65, no. 9, pp. 3892–3906, May. 2017.

[8] M. Wen, B. Ye, E. Basar, Q. Li, and F. Ji, "Enhanced orthogonal frequency division multiplexing with index modulation," *IEEE Transactions on Wireless Communications*, vol. 16, no. 7, pp. 4786–4801, May. 2017.

[9] H. Zhang, L.-L. Yang, and L. Hanzo, "Compressed sensing improves the performance of subcarrier index-modulation-assisted OFDM," *IEEE Access*, vol. 4, pp. 7859–7873, 2016.

[10] Y. Ko, "A tight upper bound on bit error rate of joint OFDM and multi-carrier index keying," *IEEE Commun. Lett.*, vol. 18, no. 10, pp. 1763–1766, Oct. 2014.

[11] T. V. Luong and Y. Ko, "Symbol Error Outage Performance Analysis of MCIK-OFDM over Complex TWDP Fading," in *2017 European Wireless Conference*. VDE, May 2017, pp. 1–5.

[12] ——, "A Tight Bound on BER of MCIK-OFDM With Greedy Detection and Imperfect CSI," *IEEE Communications Letters*, vol. 21, no. 12, pp. 2594–2597, Aug. 2017.

[13] ——, "Impact of CSI uncertainty on MCIK-OFDM: Tight closed-form symbol error probability analysis," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 2, pp. 1272–1279, Feb. 2018.

[14] ——, "Impact of Opportunistic Transmission on MCIK-OFDM: Diversity and Coding Gains," in *Proc. 4th EAI INISCOM*. Springer, Aug. 2018, pp. 221–231.

[15] E. Basar, "OFDM with index modulation using coordinate interleaving," *IEEE Wireless Communications Letters*, vol. 4, no. 4, pp. 381–384, Aug. 2015.

[16] J. Zheng and R. Chen, "Achieving transmit diversity in OFDM-IM by utilizing multiple signal constellations," *IEEE Access*, vol. 5, pp. 8978–8988, Aug. 2017.

[17] J. Choi, "Coded OFDM-IM with transmit diversity," *IEEE Transactions on Communications*, vol. 65, no. 7, pp. 3164–3171, Jul. 2017.

[18] T. V. Luong and Y. Ko, "Spread OFDM-IM with precoding matrix and low-complexity detection designs," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 12, pp. 11 619–11 626, Dec. 2018.

[19] E. Başar, "Multiple-input multiple-output OFDM with index modulation," *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2259–2263, Dec. 2015.

[20] B. Zheng, M. Wen, E. Basar, and F. Chen, "Multiple-input multiple-output OFDM with index modulation: Low-complexity detector design," *IEEE Transactions on Signal Processing*, vol. 65, no. 11, pp. 2758–2772, Jun. 2017.

[21] J. Crawford, E. Chatziantoniou, and Y. Ko, "On the SEP analysis of OFDM index modulation with hybrid low complexity greedy detection and diversity reception," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 9, pp. 8103–8118, Apr. 2017.

[22] T. V. Luong and Y. Ko, "The BER analysis of MRC-aided greedy detection for OFDM-IM in presence of uncertain CSI," *IEEE Wireless Communications Letters*, vol. 7, no. 4, pp. 566–569, Aug. 2018.

[23] T. V. Luong, Y. Ko, and J. Choi, "Repeated MCIK-OFDM With Enhanced Transmit Diversity Under CSI Uncertainty," *IEEE Transactions on Wireless Communications*, vol. 17, no. 6, pp. 4079–4088, Jun. 2018.

[24] T. V. Luong, Y. Ko, N. A. Vien, D. H. Nguyen, and M. Matthaiou, "Deep Learning-Based Detector for OFDM-IM," *IEEE Wireless Communications Letters*, 2019.

[25] H. L. T. Thanh, V.-D. Ngo, M.-T. Le, and X. N. Tran, "Repeated index modulation with coordinate interleaved ofdm," in *2018 5th NAFOSTED Conference on Information and Computer Science (NICS)*. IEEE, Nov. 2018, pp. 114–118.

[26] M. Nakagami, "The m-distribution– A general formula of intensity distribution of rapid fading," in *Statistical Methods in Radio Wave Propagation*. Elsevier, 1960, pp. 3–36.

[27] H. Lee, J. G. Andrews, R. W. Heath Jr, and E. J. Powers, "The performance of space-time block codes from coordinate interleaved orthogonal designs over Nakagami-m fading channels," *IEEE Transactions on Communications*, vol. 57, no. 3, pp. 653–664, Mar. 2009.

[28] M. Chiani, D. Dardari, and M. K. Simon, "New exponential bounds and approximations for the computation of error probability in fading channels," *IEEE Transactions on Wireless Communications*, vol. 2, no. 4, pp. 840–845, Jul. 2003.

[29] M. K. Simon and M.-S. Alouini, *Digital communication over fading channels*. John Wiley & Sons, 2005, vol. 95.

Le Thi Thanh Huyen was born in Hanoi, Vietnam in 1986. She received her B.Eng. and M.Sc. degrees both in Electronic Engineering from Le Quy Don Technical University, Hanoi, Vietnam, in 2010 and 2014, respectively. From 2010 to 2016 she worked as a lecturer at Le Quy Don Technical University. She is currently working toward her Ph.D. degree in Electronic Engineering at Le Quy Don Technical University.

Tran Xuan Nam is currently an associate professor and head of the research group in advanced wireless communications in Le Quy Don Technical University, Hanoi, Vietnam. He received his M.Eng. in Telecommunications Engineering from University of Technology Sydney, Australia in 1998, and Dr.Eng. in Electronic Engineering from The University of Electro-Communications, Japan in 2003. From November 2003 to March 2006 he was a research associate at the University of Electro-Communications, Tokyo, Japan. Dr. Tran research interests are in the areas of space-time signal processing for communications such as adaptive antennas, space-time coding, MIMO, spatial modulation and cooperative communications. Dr. Tran is a recipient of the 2003 IEEE AP-S Japan Chapter Young Engineer Award, and of two best papers from the 2012 International Conference on Advanced Technologies for Communications and the 2014 National Conference on Electronics, Communications and Information Technology. He is a member of IEEE, IEICE and the Radio and Electronics Association of Vietnam.

Matchmaking for Multi-Cloud Marketplace Application

Huynh Hoang Long[1], Nguyen Huu Duc[1], Le Trong Vinh[2]
[1] Hanoi University of Science and Technology, Hanoi, Vietnam
[2] University of Science, Vietnam National University, Hanoi, Vietnam
Correspondence: Nguyen Huu Duc, ducnh@soict.hust.edu.vn
Communication: received 19/05/2019, revised 07/07/2019, accepted 14/07/2019
Online early access: 14/07/2019, Digital Object Identifier: 10.32913/mic-ict-research.v2019.n1.854
The Area Editor coordinating the review of this article and deciding to accept it was Prof. Le Hoang Son

Abstract: Software as a Service has been developing according to the trend that leverages the advantage of multi-cloud environment to avoid vendor lock-in problem. To consume cloud resource provided by various providers, cloud software should be decomposed into components which can be deployed across different clouds. Ideally, components of a component-based cloud software are independently developed and offered through multi-cloud marketplace. To bring the highest benefit to consumers, there should be an efficient cost strategy for finding a group of compatible cloud platforms for their cloud softwares. In this paper, after redefining and developing simple formal definition for Composable Application Model (CAM), we present a matchmaking method that could be useful for verifying the correctness of software composition as well as for checking the correct deployment of a software composition on specified cloud platforms. As an illustration, we experimentally transform the cloud application model achieved by our matchmaking method into TOSCA-based specification template, which is known as a standard representation for multi-cloud applications.

Keywords: *Cloud computing, multi-cloud, multi-cloud marketplace, composable application model (CAM), software as a service, matchmaking.*

I. Introduction

In recent years, the Software as a Service delivery model (SaaS) is increasingly used and it has become a reasonable choice to consumers [1]. A simple way for a consumer to find a suitable SaaS for his needs is through a cloud marketplace where cloud softwares are developed and provided by vendors themselves and/or individual developers.

Usually, a cloud provider publishes a set of application programming interfaces (APIs) and tools to developers for implementing applications on the top of their cloud resources. Thus, cloud applications are often tightly coupled to the cloud services at different levels and restrictions. Developers are totally bound to the ecosystem of technology from the cloud provider where they want to host their on-developing application. As a result, the vendor lock-in problem arises, SaaS development is single, isolated and discursive in narrow ranges. There are some limitations of that approach which are identified as follows: (i) the lack of integration mechanisms among cloud providers making the development of a cloud application whose components are designated to be hosted on different cloud infrastructures become infeasible; (ii) the capability limit of cloud providers preventing the development of large-scale software system which requires a tremendous amount of resources; (iii) lack of standards supporting a uniform development of cloud applications.

To overcome these limitations, cloud application is designed in a fashion that they consume various types of resource offered by cloud providers. A good idea is that a cloud application in a multi-cloud environment should be decomposed into different components which can be developed and distributed across heterogeneous cloud platforms, it was mentioned in some papers such as Guillén *et al.* [2] and Baryannis *et al.* [3]. This view becomes clearer through Copil *et al.* [4] proposed a generic composition model of cloud service to represent the entire cloud application or system which can be further decomposed into service topologies and service units. Service units represent individual software or cloud offering services, and can be grouped into a cloud service topology for establishing semantically connections. Also from this point of view and combine with the component-based approach presented in [5–8], in our previous studies [9] and [10], we proposed a component-based cloud application model in which the cloud software of multi-cloud marketplace is composed from software components, each of them can be independently developed by different developers and can reside in different clouds.

Developers could be free to develop cloud software components as they wish without being dependent on any cloud API of a certain cloud provider, and cloud software development could be not tightly coupled to any cloud provider. This brings some benefits as follows: (i) this enables the distribution of software components on heterogeneous multi-cloud platforms; (ii) this separates the cloud software development from cloud providers and so minimizes vendor lock-in problem; (iii) this helps consumers exploit the most advanced features from a cloud provider by only choosing its best platform services to host some appropriate components instead of a full SaaS solution; (iv) developers do not need to pay too much attention to cloud infrastructure information runtime parameters such as CPU, RAM, Storage, middleware, network, etc.

In this approach, cloud software development is separated from cloud providers. Thus, we need an efficient way to make sure that a cloud software and its underlying execution platform are compatible. There are some studies related to this issue. Guillén *et al.* [2] presented a cloud development framework for developing managing cloud application that is separated from the source code and managed, source code of application could be deployed on multiple cloud platforms. The framework is intended to be applied upon applications targeted towards IaaS and PaaS clouds. Baryannis *et al.* [3] presented a cloud service composition approach for multi-cloud applications so that would be able to find the most optimal resource by different cloud providers for satisfying all end-user requirements. Kolb [11] introduced an approach for matching web application among PaaS providers based on their profiles. The matchmaking is successful if and only if all web application properties exactly match with a compared PaaS profile. Zeng *et al.* [12] proposed a Wordnet-based matching algorithm that considers the semantic similarity of the concepts mapping to the I/O parameters of the services. QoS information is utilized to rank the search. Zilci [13] introduced an idea to compare services based on their quality of service (QoS) requirements in cloud service marketplaces by using constraint programming to solve the service matchmaking problem.

Garg *et al.* [14] proposed a framework and a mechanism for ranking Cloud services based on their performance on QoS properties and the weights given to these properties, by exploiting an Analytical Hierarchy Process (AHP)-based algorithm. The matchmaking solution of Cloud4SOA proposed by D'Andria *et al.* [15], allows searching among the existing PaaS offerings those that best match the user requirements and ranks them based on the number of satisfied user preferences. The matchmaking mechanism uses semantic technologies to align the user requirements and the compatible PaaS offerings. García-Gómez *et al.* [16] introduced an ideal for matchmaking via the use of blueprints. A set of alternative Abstract Resolved Blueprints (ARBs) are created in design process of software developer based on the offerings of other available third-party source blueprints that can be queried and purchased from the marketplace. Each ARB is a possible combination of blueprints constituting a Cloud application. Elshareef [17] propose a matchmaking strategy between the incoming requests and various resources in the cloud environment to satisfy the requirements of users and to load balance the workload on resources.

An interesting approach of incorporating Domain Specific Languages (DSL) which facilitate to model cloud application and support between matchmaking deployment requirements and infrastructure descriptions were presented by Sledziewski *et al.* [18] and Brandtzaeg *et al.* [19]. Baryannis [3] introduced an ideal to matchmaking application with cloud infrastructure. Constraint satisfaction rules are employed in order to match requirements with existing infrastructure descriptions/capabilities, resulting in one or more proposed plans for deployment. The matchmaking process is conducted through the matchmaker engine of which inputs are requirement specifications provided from application developer, cloud infrastructure descriptions are derived from knowledge base, and constraint satisfaction rules are derived from a rule base. In addition, there are some open source solutions such as Heat [20] and Juju [21] which describes and model composite cloud applications and support deploying them on several cloud providers. However, these proposals could only be done on a single cloud. An effort for distributing cloud application on many clouds is showed in a recent study which was presented by Saatkamp [22], he introduced the Split and Match Method for TOSCA specification. His method splits TOSCA topology according to the specified targets providers and matches the resulting topology fragments with the cloud provider services to support an automated deployment of the application to multiple clouds. The goal of the Split and Match Method is to enable a customized distribution of the components of an application to different cloud providers.

In general, these matching solutions have limitations as follows: (i) the compatibility and the dependencies among components in a multi-component cloud application is not considered; (ii) these matching solutions have not been fully defined in a specific multi-cloud application pattern; (iii) a cloud software and its runtime systems were not described in a standard specification model so that cloud software components can be distributed across various clouds.

In this paper, we present a novel matchmaking method for cloud application of multi-cloud marketplace as the next step of our previous works in [9] and [10]. Our contributions are summarized as follows:

- We improve Composable Application Model (CAM) proposed in [9] that organizes multi-component cloud software in a nested structure. Then, we define a simple formal definition for abstract model of CAM.
- We propose an approach to explicitly specify the technological constraints among software components and between a component and its expected underlying runtime platforms by matching rules. We use these constraints to verify the correctness of software composition and deployment.
- We develop a matchmaking algorithm to retrieve a group of compatible cloud platforms for a component-based cloud software that suits with a criteria set ordered by cloud consumer.

Multi-cloud marketplace also plays a role of a service broker. Therefore, our matchmaking method aims to create an effective broker mechanism that support consumers to find suitable cloud platforms that satisfy their demands for each cloud software that independently developed by developers and released through multi-cloud marketplace. The output of this work is a group of compatible cloud platforms for a component-based cloud software; on this basis, a multi-cloud application defined by CAM is made up. Relying on a case study a case study presented in Section IV, we validate our proposal by experimenting a transformation from a specification of CAM into TOSCA-based specification template which is used for representing multi-cloud application structure.

The rest of the paper is organized as follows. Section II introduces our work for improving Composable Application Model. Matchmaking algorithm is presented in Section III. The illustration of transformation from CAM specification to TOSCA application template is shown in Section IV. Finally, we conclude and summarize the contributions of the paper in Section V.

II. Composable Application Model

In our approach, the development of cloud software should not be bound to any specific cloud provider. A cloud software should be able to deploy on compatible cloud platforms to form a cloud application system without having to re-engineer or to re-develop. Thus, we divide a cloud application into two separated parts: a cloud software and an underlying runtime system which is provided by specific cloud providers. We also adopt the concept of component-based application model in which a cloud software is decomposed into software components, each of them can be hosted on separated cloud platforms. This

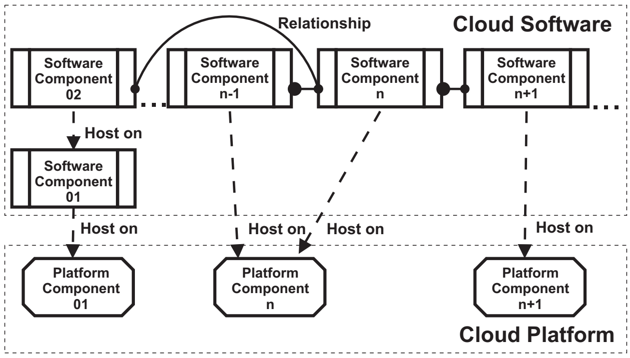


Figure 1. Cloud Application.

component-based approach requires the dependency among software components to be explicitly clarified for the sake of separate development and testing of the components. We call this cloud application model is *Composable Application Model* (CAM) and are going to redefine the model in following sub-sections.

1. General Concepts

Before going to further details of CAM, we need to clarify several related concepts as follows:

1) Cloud application: A cloud application is a runtime service that offers specific business functionalities. To customers, this refers to a similar concept to Software as a Service (SaaS). Internally, a cloud application may be constructed as a distributed system whose elements are compute servers running specific softwares and located at specific cloud providers.

We divide a cloud application into two separated parts: a cloud software and an underlying runtime system (which is a single or a group of cloud platforms) as illustrated in Figure 1. This separation allows us avoiding vendor lock-in problem. The software part can be developed separately from the underlying cloud platforms. All dependencies between the cloud software and the cloud platforms should be explicitly described as they will be used to check the compatibility at the time of deployment. Depends on the number of cloud platforms are used, we classify cloud applications into two categories:

- *Single-platform application:* A single platform application is a cloud application operating on a single cloud platform.
- *Multi-platform application:* a multi-platform application is a cloud application operating on a group of cloud platforms which may be distributed among different cloud providers.

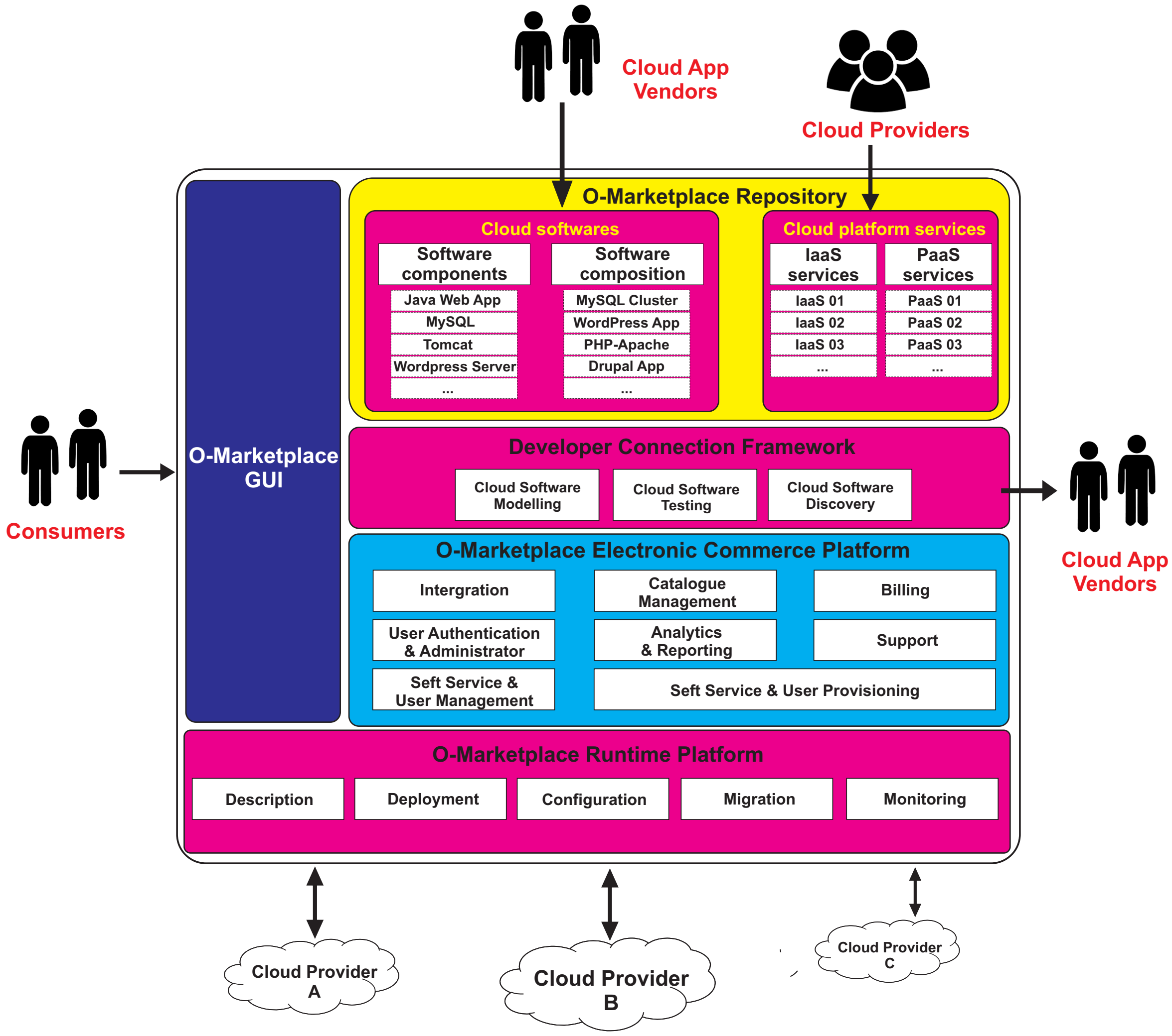


Figure 2. O-marketplace.

2) Cloud software: A cloud software is a collection of artifacts (i.e., code and data) grouped as a software bundle in a standard format. There are several types of cloud software. In the simplest form, cloud software is just a single component which can resides in a single cloud platform. In more complex situations, a cloud software is a composition of software components that could be distributed across many cloud platforms. With this component-based approach, developers can build up an application by just incorporating existing components into their own software solution. This would help us to increase re-usability of the software components and reduces unnecessary efforts in the software development process.

Cloud Software Component, or *component* for short, is a basic element of CAM. Components are used to build up more complex cloud softwares. Beside the content of code and data, the definition of a component should specify necessary conditions for the component to be integrated with others and to be deployed on a specific platform and/or a platform group. A component may be an atomic entity or a combination of other components. We classify cloud software components into three following types:

- *Simple component* type stands for a set of atomic entities, the building blocks for cloud softwares. A simple component packs its code and data together with the necessary requirements for running properly. It also explicitly defines the capabilities which other components may need when combining to form a cloud software.
- *Cloud Software Stack*, or *stack* for short, is a special kind of multi-component cloud composition. It defines a sequence of cloud software components in which a component consumes the software services provided by the following components in the sequence, and in the same time it sets up necessary environment for the previous components in the sequence. Dependencies

among layers of components should be satisfied when developing a software stack. A stack can only be deployed on a single platform.
- *Cloud Software Composition*, or *composition* for short, is a more generic multi-component cloud composition. Different from cloud software stack, which can only be deployed on a single platform, a cloud software composition can host its components on multiple platforms (may be from different cloud providers). Dependencies among software components should be satisfied according to the composition specification.

3) Cloud platform: Cloud platform is a kind of runtime system provided by cloud providers for hosting cloud software components. We also refer the term *cloud platform* as a model of delivery IaaS (Infrastructure as a Service) and/or PaaS (Platform as a Service). Various cloud providers can develop and provide the same kind of cloud platform service, but with different price, QoS, resource capacity, policies, etc. In this paper, we only consider the technical capabilities of cloud platforms for matching them with the proposed software model.

4) Cloud marketplace: Cloud marketplace is known as a kind of marketplace for selling or leasing cloud softwares which can be able to deploy on and to deliver from existing cloud providers. Most current cloud marketplaces are operated by a single cloud provider. They normally provide a complete application solution and/or a set of proprietary tools and APIs for developers to develop softwares on the top of their cloud environment. That obviously leads to the vendor lock-in problem.

To avoid this problem, in [10], we proposed a model of multi-cloud marketplace, which was called *O-Marketplace* (Figure 2). O-Marketplace targets to an open environment for developers who do not own cloud infrastructure, especially to startups whose resources are limited but their creativity is very plentiful. Developers are free to evolve their cloud software without any technology restriction from cloud providers. In this model, cloud software components are developed by individual developers, are published and sold on the marketplace, and can be able to deploy on compatible cloud platforms suggested by the marketplace.

To facilitate the suggestion feature of the marketplace, we develop a matchmaking algorithm. This algorithm would help us to find suitable platforms to host a specific cloud software by matching the platform requirements of the cloud software to the capabilities of the cloud platforms registered in their profiles. More details of the algorithm will be given in the Section III.

2. Matching Definition

In CAM, a cloud software is composed from independent software components in a composition, a software component is deployed on a platform component which is a cloud resource service bundle such as IaaS or PaaS offered from different cloud providers. To depict the relationships of components within a cloud application modeled by CAM, we specify two types of dependencies in a cloud application model: *software dependence* and *platform dependence*. Software dependence denotes the interconnection between two software components. The platform dependence denote deployment capability of a pair of software components or a pair of a software component and a cloud platform. We define two elements to make up a dependence: *Requirement* and *Capability*. Relying on these elements, matching rules is constructed to denote the dependencies within a cloud application. Requirement specification, capability specification, and matching condition is defined as follows:

- Requirement specification: *requirement* denotes the dependency of a component on other another component. It poses the necessary conditions for component to perform its functionality and behaviors. Requested interfaces and properties of a component are encapsulated in requirement. We specify two types of requirement: *software requirement* and *platform requirement*. Software requirement *(sreq)* is a constraint on the functionality and behavior of another component. Platform requirement *(preq)* is a constraint on the environment and technology of another component.
- Capability specification: *capability* denotes the available capacity that can meet the request from outside. its functionality and behavior are performed when it satisfies the condition from another external component. Responded interfaces and properties of a component are encapsulated in capability. We specify two types of capability: *software capability* and *platform capability*. Software capability *(scap)* is a supply of the functionality and behavior of a component. Platform capability *(pcap)* is a supply of the environment and technology of a component.
- Matching condition: we define a dependence between two components is valid if *requirement* of a component fully match with *capability* of the other.

3. An Abstract Model of CAM

Our ultimate goal is to develop a full specification for the proposed composable cloud application. This development requires a great effort on various aspects including: (1) defining standards and frameworks for coding cloud software components; (2) specifying the deployment and management operations of composable cloud application; (3)

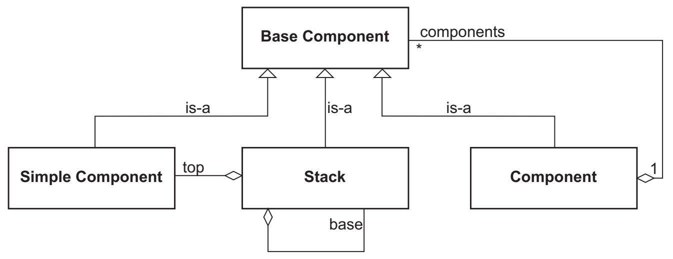


Figure 3. Overall structure of CAM.

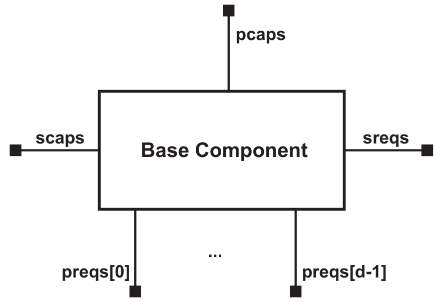


Figure 4. Base component.

verifying the correctness of cloud software composition; (4) matching composable cloud software to existing cloud platforms. In this paper, we limit our study to the latter two problems. We develop an abstract model of CAM in which we ignore the implementation details of the components but focus on the relationship among components inside a software composition and the relationship between a component and its underlying platforms.

1) Overall structure: We define the abstract model of CAM as a nested structure where each component may be a composition of other components. Some of them may again be compositions of other smaller components. This approach helps increase the degree of re-usability since the implementation of a complex component may be reused in other compositions. This also motivates the cloud software marketplace where developers can develop and sell their small components instead of complete software solutions.

According to the concepts presented in section II-C, we classify components in the model into the following three types: (1) *Simple Component*, (2) *Stack*, and (3) *Composition*. Figure 3 shows the overall structure of CAM and the relationship among these three types of the components. Details for each type of component will be explained in the following sections.

2) Base component: As shown in Figure 3, three types of software components are modeled as sub-types of the Base Component. This formalism allows us to treat components of different types uniformly as specializations of the Base Component. Each component has its own properties. We use the dot-notation to refer to a property of a component. For example, $X.y$ refers to the property y of the component X. Since we focus on the relationship among components inside a composition and the relationship between a component and its underlying platforms, the properties of a component should explicitly describe its requirements and capabilities. More specifically, as shown in Figure 4, a component should have at least the following properties: (1) description of software services provided by the component (software capabilities - $scaps$), (2) the establishment of runtime environment providing to upper components in a stack manner (platform capabilities - $pcaps$), (3) descriptions of extra software services required for running the component properly (software requirements - $sreqs$), (4) and the necessary requirements for underlying platforms to host the component (platform requirements - $preqs$). The specifications of components, i.e,. the description of component's properties, are used for verifying the correctness of software composition and for checking the compatibility between a component and its underlying platforms. These specifications are written by developers, and they are not necessary to collate the actual implementation of the components. Instead, developers can select the requirements and capabilities of a component from a predefined set of terms.

Note that a component may be hosted on multiple cloud platforms from different cloud providers. The platform requirements should be specified for each platform. We call $degree$ of a component to be the total number of underlying platforms for hosting the component. Thus, the platform requirements for the platform i is defined as $preqs[i]$.

For a convenient development, we also use the following derived properties:

- $preqs$: set of all platform requirements,

$$preqs = \bigcup_{i \in [0 \ldots d-1]} preqs[i].$$

- $reqs$: set of all software requirements and platform requirement,

$$reqs = preqs \bigcup sreqs.$$

- $caps$: set of all software requirements and platform requirement,

$$caps = pcaps \bigcup scaps.$$

3) Simple Component: Simple component is an atomic type of components in CAM. In a full definition, a simple component should specify its code and data as well as the necessary specification of requirements and capabilities. A simple component is hosted on a single platform, i.e., ($degree = 1$). If the set of requirements of a simple

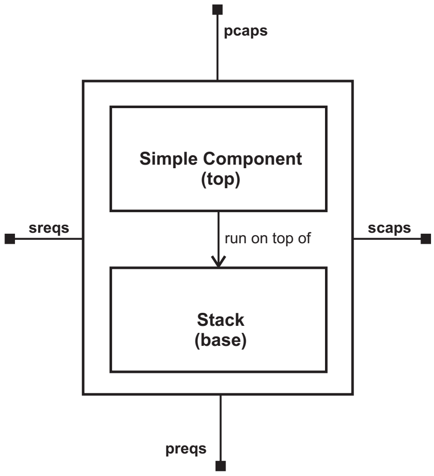


Figure 5. Cloud software stack.

component is empty ($regs = \emptyset$), the component can represent a complete cloud software. In general cases, a simple component can combine with other components to form a more complex component (i.e., a stack or a composition).

4) Stack: A cloud software stack, or stack for short, denotes a sequence of simple components deployed on top of each other vertically. For simplicity, we define a Stack in nested manner. Each stack has two elements: (1) *top* element is a simple component lay on top of the stack, (2) and *base* which is either a simple component or another stack representing the remain components in the sequence. Similar simple component, a software stack can only be deployed on a single platform ($degree = 1$). Beside the advantage of re-usability, the combination of multiple components into a single form of stack would help us reducing the cost of deployment and management by combining corresponding operations from the stack's elements. We will discuss this interesting problem in another study from our on-going research.

When creating a stack, a developer must specify the *top* element and the *base* element (Figure 5). The developer must also specify the stack requirements and capabilities (i.e., *sreqs*, *preqs*, *scaps*, and *pcaps*), like those mentioned in the Base Component. A correct combination of a stack S should satisfy the following validation rules:

$$\begin{aligned}
S.top.preqs &\subseteq S.base.pcaps \\
S.pcaps &\subseteq S.top.pcaps \bigcup S.base.pcaps \\
S.preqs &\supseteq S.base.preqs \\
S.sreqs &\supseteq S.top.sreqs \bigcup S.base.sreqs \\
S.scaps &\subseteq S.top.scaps \bigcup S.base.scaps
\end{aligned}$$

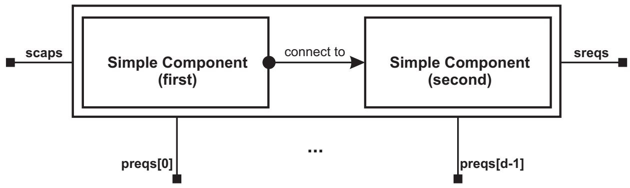


Figure 6. Cloud software composition.

5) Composition: Cloud software composition, or composition for short, denotes a set of software components combined in a single form of component to hide the complexity of its own inter-dependency. A software composition is considered as a directed graph whose vertex are either a simple component or a stack. An edge from a component A to a component B in a composition specifies that the component A consumes services provided by the component B at runtime.

Similar with stack, we apply the nested structure for defining compositions. Each composition consists of a set of components and some of them may be other compositions. For simplicity, we define a composition with two elements (Figure 6): the $first$, and the $second$. The interdependence of a composition is restricted to the software dependence of the $first$ to the $second$. Degree of the composition is calculated by the sum of degrees of the $first$ and the $second$. When creating a composition, developers need to specify its external requirements and capabilities. At the moment, we do not allow extra components to lay on top of a composition. So, the property *pcaps* should be empty. The following validation rules should be applied for verifying the correctness of a composition C:

$$\begin{aligned}
C.scaps &\subseteq C.first.scaps \bigcup C.second.scaps \\
C.sreqs &\supseteq C.second.sreqs \bigcup \\
&\quad (C.first.sreqs \backslash C.second.scaps) \\
C.preqs[i] &\supseteq C.first.preqs[i] \\
&\quad \forall i \in [0 \ldots (C.first.degree - 1)] \\
C.preqs[i] &\supseteq C.second.pregs[i - d_1] \\
&\quad \forall i \in [d_1 \ldots (d_1 + d_2 - 1)] \\
&\quad d_1 = C.first.degree \\
&\quad d_2 = C.second.degree
\end{aligned}$$

6) Cloud platform: Cloud platform, or platform for short, is used for modeling the cloud platform profile in a cloud marketplace. A specification of a single platform P should describe its capabilities (*pcaps*) in order to match with the platform requirements (*preqs*) of cloud software components which are aimed to deploy on it. Practically, the specification of a platform is provided by the cloud provider or a service broker such as cloud marketplace. In this case, developers and cloud providers should agree on the same set of predefined terms of the platform requirements and capabilities.

A cloud software composition may require more than one platform. We define *platform group* as an order set of cloud platforms, and *compatible platform group* as a platform group that satisfies the platform requirements of a composition. The following predicates are used for checking this property.

- $compatibleWithReqs(P, preqs)$ if $P.pcaps \supseteq preqs$
- $compatibleWith(PG = [P_0, P_{d-1}], C)$
  if $d = C.degree$ and
  $\forall i \in [0 \ldots d-1].compatibleWithReqs(P_i, C.preqs[i])$

At the time of software deployment, platform compatibility should be checked to ensure that the software component will work fine on the underlying platform. The platform compatibility check is also a basis for the matchmaking algorithm presented in the next section.

III. Matchmaking Method for Cloud Application

In multi-cloud marketplace, besides providing a place for publishing and selling cloud softwares, the marketplace may also play a role of a service broker. Based on the specification of platform requirements of cloud software component, the marketplace would be able to suggest optimal solutions for renting compatible cloud platforms. The term "optimal" could be defined according to the user needs. Some user wants to find a platform solution with smallest price. Others may need a solution with the best quality of service. In the context of this paper, we do not dig deeper in this aspect.

We assume that a platform solution for hosting a cloud software component C is a compatible platform group PG to the component C, i.e. $compatibleWith(PG, C)$ is satisfied. There may be more than one solution existed since several cloud providers may offer a same kind of platform (but with different price and quality of service). Choosing an optimal solution means that we need to make a comparison among the solutions based on a designated cost function. We call $Cost(PG)$ for the cost function of a platform solution, i.e., the platform group PG. Some reasonable examples for the cost function would be:

- $Cost([P_0, \ldots, P_{d-1}]) = \sum_{i=0}^{d-1} Price(P_i)$. This cost function helps customers select a platform solution with optimal price.
- $Cost([P_0, \ldots, P_{d-1}]) = |\{P_0, \ldots, P_{d-1}\}|$. This cost function helps customers select a platform solution with a smallest number of platforms will be used.

Algorithm 1: Matchmaking(C, PS)

```
1  Inputs:
2      - A cloud software component C.
3      - A set of cloud platforms PS.
4  Output:
5      - A compatible platform group PG to the component
6        C whose members are selected from PS.
7  begin
8      d = C.degree ;
9      for i = 1 to d − 1 do
10         CPL[i] = ∅;
11         foreach P in PS do
12             if compatibleWithReqs(P, C.preqs[i]) then
13                 CPL[i] = CPL[i] ∪ {P};
14             end
15         end
16     end
17     min_cost = INFINITY;
18     best_solution = NULL;
19     Let S = (CPL[0] × CPL[1] × . . . × CPL[d − 1]);
20     foreach PG in S do
21         if Cost(PG) < min_cost then
22             min_cost = Cost(PG);
23             best_solution = PG;
24         end
25     end
26     return best_solution;
27 end
```

Using such a cost function, we develop a matchmaking algorithm for suggesting compatible platform solution as presented in Algorithm 1.

The first part of the algorithm (lines 8-16) calculates a list of compatible cloud platforms $CPL[i]$ for each set platform requirements $preqs[i]$ of the component C. In the second part (lines 17-25), the algorithm select the best platform solution among all possible solutions, the Cartesian product $\mathcal{S} = CPL[0] \times CPL[1] \times \cdots \times CPL[d-1]$.

Correctness: Correctness of the algorithm is obvious since the optimal solution $best\_solution$ is selected from a set of all compatible platform groups to the component C, i.e., $\mathcal{S} = CPL[0] \times CPL[1] \times \cdots \times CPL[d-1]$.

Performance analysis: Let n be the number of cloud platforms in PS. The first part of the algorithm (lines 8-16) requires $d \times n$ compatibility checks. Thus the complexity of this part is $O(d \times n)$. The second part of the algorithm (lines 17-25) requires n^d calculation for the $Cost$ function. Assume that $O(f(n))$ is the complexity of the $Cost$ function. The complexity of the second part would be $O(n^d \times f(n))$. In summary, the complexity of the algorithm is $O(n^d \times f(n))$.

The algorithm is rather simple but flexible since we always have chance to change the cost function according to the consumer needs.

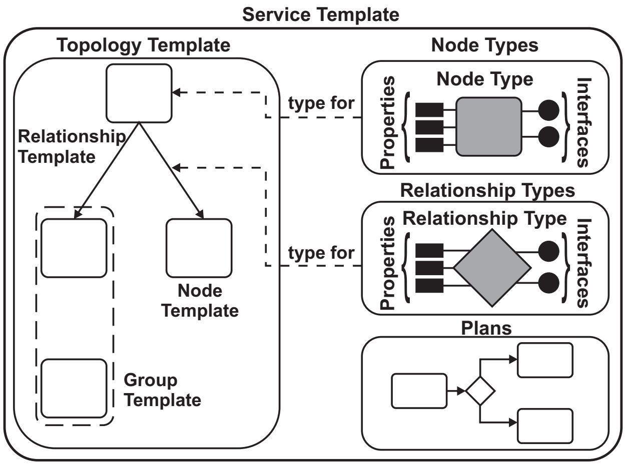


Figure 7. Structural elements of a service template and their relations.

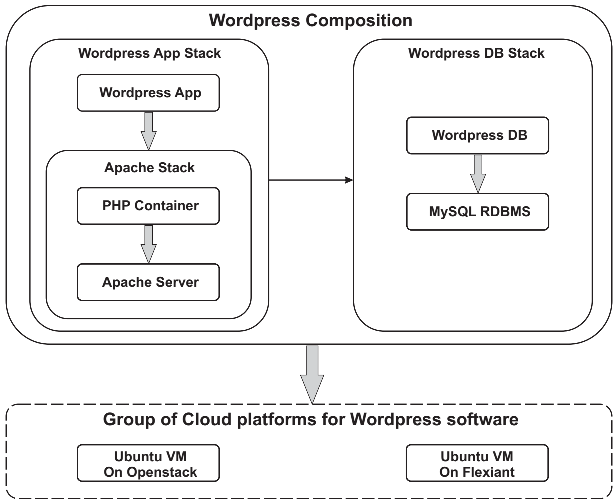


Figure 8. Wordpress Application represented in CAM.

IV. Mapping to TOSCA

In this section, we demonstrate the feasibility of our proposed composable application model (CAM) by translating a specification of CAM into TOSCA topology template, a standard specification for creating multi-cloud application.

1. Topology and Orchestration Specification for Cloud Application

Topology and Orchestration Specification for Cloud Applications (TOSCA) [23] is an open standard built by OASIS that defines the inter-operable description of cloud application hosted on the cloud; including its components, relationships, dependencies, requirements, and capabilities. TOSCA enables portability and automated management across cloud providers which have different underlying platform or infrastructure. Thus, the portable deployment of cloud applications could be done on any compliant cloud because the relationships among parts of the service and the operational behaviours of these services are independently described with any cloud provider.

TOSCA provides a meta model and packaging format to realize the automated implementation, deployment, configuration, management, and orchestration of cloud applications in an automated manner through a combination of two core concepts: Topology Templates and Management Plans (Figure 7). In order to support such packaging format. Target clouds have to have runtime environments that support TOSCA standard. Both IaaS and PaaS are feasible with this approach.

TOSCA is a middle-level language for the specification of the topology and orchestration of an IT service in the form of a service template. It is very efficient in software provisioning, deployment and management of cloud application. It enhances the portability of cloud application and has been supported by many partners like IBM, Reb Hat, Cisco, Citrix, EMC, etc.

Since TOSCA is currently a mature standard with a lot of supporting tools for designing and orchestration. In this paper, we select this standard as a target of the translation to demonstrate the feasibility of our proposed composable application model.

1) Method Overview: The major differences between CAM specification and TOSCA specification are:

- TOSCA specification represents a complete blueprint of a cloud application system including the software and the cloud infrastructure/platform parts while CAM separates these two parts into independent entities.
- TOSCA expresses the internal structure of a cloud service by a topology template which is a graph representing software/hardware components (as nodes), and relationship among these components (as edges). CAM represents a cloud application in a nested manner.

Thus, we design the translation from a CAM specification to TOSCA specification in a process of the following steps:

- First, we combine the cloud software and its compatible platform group into a single form of CAM specification.
- Second, we flatten the nested structure of the CAM specification into a graph representation. Nodes of the graph correspond to simple components and platform, and edges of the graph represent the relationships among simple components and platforms. This translation is done recursively based on the nested structure

```
Simple components:
   - Apache Server:
        type: software
        preqs: [Linux]
        pcaps: [Apache]
   - PhP Container:
        type: software
        preqs: [Apache]
        pcaps: [PhP, Apache]
   - WordpressApp:
        type: software
        preqs: [Apache, PhP]
        sreqs: [WordpressDB]
   - MySQLDB:
        type: software
        preqs: [Linux]
        pcaps: [MySQL]
   - WordpressDB:
        type: software
        preqs: [MySQL]
        scaps: [WordpressDB]
```

```
Stacks:
   - PhP-Apache Stack:
        type: stack
        top: PhP Container
        base: Apache Server
        preqs: [Linux]
        pcaps: [Apache, PhP]
   - WordpressApp Stack:
        type: stack
        top: WordpressApp
        base: Php-Apache Stack
        preqs: [Linux]
        sreqs: [WordpressDB]
   - WordpressDB Stack:
        type: stack
        top: WordpressDB
        base: MySQLDB
        preqs: [Linux]
        scaps: [WordpressDB]
```

```
Compositions:
   - Wordpress Composition:
        type: composition
        first: WordpressApp Stack
        second: WordpressDB Stack
        preqs:
            - [Linux]
            - [Linux]
```

```
Platforms:
    - UbuntuVM on Openstack:
        Type: os
        provider: OpenStack
        pcaps: [Linux]
    - UbuntuVM on Flexiant:
        Type: os
        provider: Flexiant
        pcaps: [Linux]
```

```
Wordpress Application:
   software: Wordpress Composition
   platforms: [UbuntuVM, UbuntuVM]
```

Figure 9. Wordpress application specification.

of CAM specification. The relationships are derived from the specifications of stacks and compositions.

- Finally, we map the resulting graph into TOSCA topology template. Here, the most cryptic part is that we have to map each node of the graph to a node template of TOSCA, and each edge of the graph to a relationship template of TOSCA.

The following sub-section demonstrating this process in a particular example – the Wordpress application.

2. Experiment with Wordpress Application

A typical Wordpress application is represented in CAM as shown in Figure 8. The specification of the Wordpress application is written in YAML as shown in Figure 9. We temporary ignore all implementation details of the cloud software components but only focus on the specification of their requirements and capabilities. If a property of a component is missed from the specification, we consider the value of this property is null or empty.

In the second step of the translation process, we flatten the Wordpress Application into a graph representation which is depicted in Figure 10. This translation ignores the specifications of requirements and capabilities of software stacks and software compositions. They are only used for verifying the correctness of the combination and for checking the compatibility between the software composition and its underlying platforms.

Finally, we map the graph representation to TOSCA topology template as shown in the following XML code:

```
Nodes:
   - Apache Server:
        type: SimpleComponent
        preqs: [Linux]
        pcaps: [Apache]
   - PhP Container:
        type: SimpleComponent
        preqs: [Apache]
        pcaps: [PhP, Apache]
   - WordpressApp:
        type: SimpleComponent
        preqs: [Apache, PhP]
        sreqs: [WordpressDB]
   - MySQLDB:
        type: SimpleComponent
        preqs: [Linux]
        pcaps: [MySQL]
   - WordpressDB:
        type: SimpleComponent
        preqs: [MySQL]
        scaps: [WordpressDB]
   - UbuntuVMonOpenStack:
        provider: OpenStack
        pcaps: [Linux]
   - UbuntuVMonFlexiant:
        provider: Flexiant
        pcaps: [Linux]
```

```
Relationships:
   - from: WordpressApp
     to: PhP Container
     type: host-on
   - from: PhP Container
     to: Apache Server
     type: host-on
   - from: Apache Server
     to: UbuntuVMonOpenStack
     type: host-on
   - from: WordpressDB
     to: MySQLRDBMS
     type: host-on
   - from: MySQLRDBMS
     to: UbuntuVMonFlexiant
     type: host-on
   - from: WordpressApp
     to: WordpressDB
     type: connect-to
```

Figure 10. Wordpress application graph.

```
<?xml version="1.0"?>
<ns2:Definitions id="Wordpress"
      xmlns:ns2="http://docs.oasis-open.org/tosca/ns/2011/12"
      name="Wordpress">
<ns2:ServiceTemplate id="WordpressTopology">
<ns2:TopologyTemplate>
<ns2:RelationshipTemplate
      id="WordpressApp_HostOn_PHPContainer"
      type="HOSTON">
<ns2:SourceElement ref="PhP Container"/>
<ns2:TargetElement ref="WordpressApp"/>
</ns2:RelationshipTemplate>
<ns2:RelationshipTemplate
      id="PHPContainer_HostOn_ApacheServer"
      type="HOSTON">
<ns2:SourceElement ref="Apache Server"/>
<ns2:TargetElement ref="PhP Container"/>
</ns2:RelationshipTemplate>
<ns2:RelationshipTemplate
      id="ApacheServer_HostOn_UbuntuVM"
      type="HOSTON">
<ns2:SourceElement ref="UbuntuVM"/>
<ns2:TargetElement ref="Apache Server"/>
</ns2:RelationshipTemplate>
<ns2:RelationshipTemplate
      id="WordpressDB_HostOn_MySQLDB"
      type="HOSTON">
<ns2:SourceElement ref="MySQLDB"/>
<ns2:TargetElement ref="WordpressDB"/>
</ns2:RelationshipTemplate>
<ns2:RelationshipTemplate
      id="MySQLDB_HostOn_UbuntuVM" type="HOSTON">
<ns2:SourceElement ref="UbuntuVM"/>
<ns2:TargetElement ref="MySQLDB"/>
</ns2:RelationshipTemplate>
<ns2:RelationshipTemplate
      id="WordpressApp_ConnectTo_WordpressDB"
      type="CONNECTTO">
<ns2:SourceElement ref="MySQLDB"/>
<ns2:TargetElement ref="WordpressApp"/>
</ns2:RelationshipTemplate>
<ns2:NodeTemplate id="UbuntuVM" type="os" >
```

```
 ...
</ns2:NodeTemplate>
<ns2:NodeTemplate id="Apache Server" type="software">
 ...
</ns2:NodeTemplate>
<ns2:NodeTemplate id="PhP Container" type="software">
 ...
</ns2:NodeTemplate>
<ns2:NodeTemplate id="WordpressApp" type="software">
 ...
</ns2:NodeTemplate>
<ns2:NodeTemplate id="WordpressDB" type="software">
 ...
</ns2:NodeTemplate>
<ns2:NodeTemplate id="MySQLDB" type="software">
 ...
</ns2:NodeTemplate>
</ns2:TopologyTemplate>
</ns2:ServiceTemplate>
<ns2:ArtifactTemplate
      id="Artifact_93f8753b-17ab-43c5-8f11-e3ec98fe3224"
      type="sh">
 ...
</ns2: ArtifactTemplate >
 ...
</ns2: Definitions >
```

As seen in the code above, all nodes of the graph are mapped to Node Template of TOSCA, implementation details of the Node Template are omitted from the specification for brevity. Edges of the graph are mapped to Relationship Template. According to the original of the edges, i.e., from a stack or from a composition, the type of corresponding Relationship template will be given as HOSTON or CONNECTTO.

Although CAM has not been fully developed, this demonstration has proved the feasibility of the proposal with distinct advantages.

V. Conclusion

In this study, we present a matchmaking method that supports to find suitable cloud platforms, whose profiles registered in a multi-cloud marketplace, to multi-component cloud software. The goal of this work is to verify compatibility among the components within a component-based cloud application, as well as between the components and their underlying runtime platforms. To implement this idea, firstly, we redefine *Composable Application Model* (CAM) that organizes multi-component cloud software in a nested structure. Secondly, we develop an abstract model of CAM which only covers the inter-dependency of the components. We proposed validation rules for the CAM components based on that type of dependency. Thirdly, We proposed matchmaking algorithm which is the core of brokerage mechanism of multi-cloud marketplace. Finally, our proposal is validated by experimenting a transformation from a specification of CAM into TOSCA specification.

To sum up, our work creates an effective brokerage mechanism to retrieve optimal compatible platform solutions for a specific multi-cloud software according to consumer's demand in the context of O-Marketplace. This result is a premise for our future research related to Service Level Agreement (SLA) in multi-cloud environment [24].

Acknowledgement

This work is supported by Hanoi University of Science and Technology under Project "Energy-aware workflow service in cloud computing environment", No. T2017-PC-077. Simulations and technical realization were achieved on the hardware equipment at the Center for Data and Computation Technology, Hanoi University of Science and Technology. We would like to thank all colleagues and domain experts for collaborations and consultations.

References

[1] A. Nayyar, *Handbook of Cloud Computing*, 04 2019.

[2] J. Guillén, J. Miranda, J. M. Murillo, and C. Canal, "A service-oriented framework for developing cross cloud migratable software," *Journal of Systems and Software*, pp. 2294–2308, 2013.

[3] G. Baryannis, P. Garefalakis, K. Kritikos, K. Magoutis, A. Papaioannou, D. Plexousakis, and C. Zeginis, "Lifecycle management of service-based applications on multi-clouds," in *Proceedings of the 2013 International workshop on Multi-cloud applications and federated clouds (Multicloud'13)*, 2013, pp. 13–20.

[4] G. Copil, D. Moldovan, H.-L. Truong, and S. Dustdar, "Multi-level elasticity control of cloud services," in *Service-Oriented Computing*, S. Basu, C. Pautasso, L. Zhang, and X. Fu, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 429–436.

[5] D. Garlan, R. T. Monroe, and D. Wile, "Acme: Architectural description of component-based systems," in *Foundations of Component-Based Systems*, G. T. Leavens and M. Sitaraman, Eds. Cambridge University Press, 2000, pp. 47–68.

[6] B. Wallace, "A hole for every component, and every component in its hole," *Existential Programming*, 2010.

[7] Mcllroy and M. Douglas, "Mass produced software components," *Scientific Affairs Division, NATO*, p. 79, 1969.

[8] R. Niekamp, "Software component architecture," *Gestión de Congresos - CIMNE/Institute for Scientific Computing*, p. 4, 2011.

[9] H.-L. Huynh, H.-D. Nguyen, V.-T. Le, and T.-T. Nguyen, "A composable application model for cloud marketplace," *Journal of Vietnam Science and Technology*, vol. 16, no. 5, pp. 40–45, 2017.

[10] H.-L. Huynh, H.-D. Nguyen, V.-T. Le, and D.-H. Le, "Towards the cloud marketplace for multi-cloud infrastructures," in *18th Vietnam National Conference: Selected issues of information technology and communication*, 2015.

[11] S. Kolb and G. Wirtz, "Towards application portability in platform as a service," in *Proceedings - IEEE 8th International Symposium on Service Oriented System Engineering, SOSE 2014*, 2014.

[12] C. Zeng, X. Guo, W. Ou, and D. Han, "Cloud computing service composition and search based on semantic," in *Cloud Computing*, M. G. Jaatun, G. Zhao, and C. Rong, Eds.

Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 290–300.
[13] B. I. Zilci, M. Slawik, and A. Küpper, "Cloud service matchmaking using constraint programming," in *2015 IEEE 24th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises*, June 2015, pp. 63–68.
[14] S. K. Garg, S. Versteeg, and R. Buyya, "SMICloud: A framework for comparing and ranking cloud services," in *2011 Fourth IEEE International Conference on Utility and Cloud Computing*, Dec 2011, pp. 210–218.
[15] F. DAndria, S. Bocconi, J. G. Cruz, J. Ahtes, and D. Zeginis, "Cloud4SOA: Multi-cloud application management across PaaS offerings," in *2012 14th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing*, Sep. 2012, pp. 407–414.
[16] S. García-Gómez, M. Jiménez-Gañán, Y. Taher, C. Momm, F. Junker, J. Bíró, A. Menychtas, V. Andrikopoulos, and S. Strauch, "Challenges for the comprehensive management of cloud services in a PaaS framework," *Scalable Computing: Practice and Experience*, vol. 13, 2012.
[17] W. Elshareef, H. A. Ali, and A. Y. Haikal, "A matchmaking strategy of mixed resource on cloud computing environment," *International Journal OF Scientific and Technology Research*, vol. 4, 2015.
[18] K. Sledziewski, B. Bordbar, and R. Anane, "A DSL-based approach to software development and deployment on cloud," in *2010 24th IEEE International Conference on Advanced Information Networking and Applications*, April 2010, pp. 414–421.
[19] E. Brandtzaeg, S. Mosser, and P. Mohagheghi, "Towards CloudML, a model-based approach to provision resources in the clouds," in *8th European Conference on Modelling Foundations and Applications (ECMFA)*, 2012, p. 18–27.
[20] "OpenStack HEAT URL," accessed: 2019-04-26. [Online]. Available: https://docs.openstack.org/heat/latest
[21] "Juju Charms URL," accessed: 2019-04-26. [Online]. Available: https://jujucharms.com
[22] K. Saatkamp, U. Breitenbücher, O. Kopp, and F. Leymann, "Topology splitting and matching for multi-cloud deployments," in *Service-Oriented Computing-ICSOC 2017 Workshops*, 2017, pp. 379–383.
[23] OASIS, "Topology and orchestration specification for cloud applications version 1.0," *Organization for the Advacement of Structured Information Standards*, 2013.
[24] D. Kourtesis, K. Bratanis, A. Friesen, Y. Verginadis, A. J. H. Simons, A. Rossini, A. Schwichtenberg, and P. Gouvas, "Brokerage for quality assurance and optimisation of cloud services: An analysis of key requirements," in *Service-Oriented Computing – ICSOC 2013 Workshops*, A. R. Lomuscio, S. Nepal, F. Patrizi, B. Benatallah, and I. Brandić, Eds. Cham: Springer International Publishing, 2014, pp. 150–162.

Huynh Hoang Long received B.Sc. degree from Nhatrang University in 2008 and M.Sc. degree from Hanoi University of Science and Technology in 2012. His research interest includes cloud computing.

Nguyen Huu Duc received Ph.D. degree in Computer Science from Japan Advanced Institute of Science and Technology (JAIST), Japan, in 2006. He is currently the director of the Center for Data and Computation Technologies, Hanoi University of Science and Technology.

His main research topics include compiler construction, high performance computing, distributed systems and big data.

Le Trong Vinh received Ph.D. degree in Computer Science from Japan Advanced Institute of Science and Technology (JAIST), Japan, in 2006. He is currently Associate Professor and the director of the Center for Information Technology and Communication, University of Science, Vietnam National University, Hanoi.

His main research topics include theory of algorithms, computer networks.

Performance Comparison of Dynamic Elastic Optical Networks with Optical Regeneration

Le Hai Chau, Dang Hoai Bac

Posts and Telecommunications Institute of Technology, Hanoi, Vietnam

Correspondence: Le Hai Chau, chaulh@ptit.edu.vn

Communication: received 23/04/2019, revised: 25/07/2019, accepted: 26/07/2019

Online early access: 26/07/2019, Digital Object Identifier: 10.32913/mic-ict-research.v2019.n1.853

The Area Editor coordinating the review of this article and deciding to accept it was: Dr. Nguyen Tan Hung

Abstract: We have investigated optical regeneration issues and application in elastic optical networks that are capable of providing dynamically optical paths with flexible bandwidths. We have analyzed the impact of optical regeneration in elastic optical networks and clarified various usage scenarios. We have then evaluated and compared the performance, in terms of the overall blocking probability and the total accommodated traffic volume, of three possible network scenarios with regeneration capability including (i) no regeneration, (ii) 3R regeneration, and (iii) 4R regeneration for practical network topologies. Numerical simulation proved that deployment of optical regeneration devices can exploit elastic optical networking to enhance the network performance for provisioning dynamically bandwidth-flexible lightpath services. It is also demonstrated that using re-modulation function while regenerating optical signals (4R regeneration) can further improve the network performance. However, due to the high cost of optical regeneration devices, especially all-optical ones, and more functional regenerators, the trade-off between the performance enhancement and the necessary number of regenerating devices needs to be carefully considered.

Keywords: *Optical regenerator, routing, modulation format and spectrum assignment, network control algorithm, elastic optical network.*

I. Introduction

Internet traffic is growing at incredible rates, driven by high-performance applications including video on demand, and cloud and grid computing [1, 2] which demand even more ubiquity, mobility, and heterogeneous bandwidths [3, 4], in recent decades. This traffic growth places an importance of extremely large data capacity yet flexible and efficient optical network technologies to support broadband services up to Terabit/s in the near future. To scale to Terabit/s, optical networks based on current WDM technology using fixed ITU-T frequency grid will face serious problems due to the stranded bandwidth provisioning, inefficient spectral utilization, and high cost [5]. Present researches on optical transmission and networking technologies are oriented forward more efficient, flexible, and scalable network solutions. Recently, elastic optical network (EON) utilizing a flexible frequency grid has been proposed as a promising candidate for future ultra-high capacity optical networks [6, 7]. Elastic optical networking technology helps greatly improve the spectral efficiency and flexibility of the network by eliminating stranded spectrum between channels, supporting both sub-channel and super-channel traffic, and therefore allowing flexible bandwidth connections of multiple data rates and modulation formats [6–8]. However, EON is currently facing challenges owing to the lack of architectures and technologies to efficiently support bursty traffic on flexible spectrum.

Elastic optical networks are able to allocate spectrum resources flexibly for handling not only legacy low-bitrate services but also new super-channel services [6]. Although elastic optical networks are also capable of provisioning dynamic bandwidth-flexible and spectrum-efficient end-to-end optical paths while enable an economical scalability of networks adapting to the growing trend and the heterogeneity of bandwidth requirements for Telcos/Internet service providers [7, 8], more sophisticated network design and provision control strategies are required for realizing efficient and robust network operations [7]. As a result, routing and wavelength assignment (RWA) problem of elastic optical networks becomes more complicated and is known as routing and spectrum assignment problem which includes three sub-problems that are routing, modulation format assignment and spectrum allocation [7–9].

Moreover, optical signals suffer from the physical impairments, such as dispersion and noise, etc. which are being accumulated along the optical paths and consequently, cause a limitation on the optical transmission reach [10–12]. Particularly in large optical networks in which optical paths may travel ultra-long distances, the optical impairment impact becomes very critical. Hence, optical regenerators need to be deployed to clear the impairments and improve the optical transmission reach [12, 13].

Among the regenerator realization technologies, all-optical regeneration can potentially offer significant cost savings thanks to less power and size requirement, especially at high-data rates. Employing advanced all-optical regeneration that applies multi-channel packaging can be a disruptive technology to reduce the power and size footprint.

Besides conventional optical regeneration techniques including the signal re-amplifying along with re-shaping (2R), and even/or re-timing (3R), a new elastic optical regenerator, which is Virtualized Elastic Regenerator (VER), has been lately proposed for elastic optical networks [14]. Different from traditional optical regenerators, the developed VER not only offers 3R regeneration functions (re-amplifying, re-shaping and re-timing), but also naturally supports re-modulating function (so called 4R) [15, 16]. In addition, all 3R/4R optical regenerators can also provide the spectrum conversion capability and therefore, the use of optical regenerators can help to avoid spectrum collision efficiently and significantly enhance spectrum resource utilization in elastic optical networks.

In this paper, we investigate optical regeneration issue and its impact on dynamic elastic optical networks with bandwidth flexible optical paths provisioning capability. We also analyze the performance limits of elastic optical transmission system. The network performance, in term of the overall blocking probability and the accommodated traffic volume, has been then evaluated and compared among three typical optical regeneration deploying scenarios that are

(i) no (without) regeneration,
(ii) 3R regeneration, and
(iii) 4R (3R plus Re-modulation) regeneration in two practical network topologies.

Numerical simulation is used to verified the efficiency of using optical regenerators to exploit elastic optical networking and enhance the network performance. It is proved that at least 29.8% (or 45.6%) more traffic volume for NSF (USNET) network can be obtained. Re-modulation feature of elastic regenerators can also further improve the network performance. However, in order to build cost-effective, bandwidth-abundant and flexible optical networks, the balance between the network performance and the required regeneration resource cost should be carefully considered due to costly optical regeneration devices.

II. Optical Regeneration and Its Application in Dynamic Elastic Optical Networks

1. Performance Limits of Long-Haul Elastic Optical Transmission Systems

Unlike traditional WDM networks, elastic optical networks can support one or several modulation formats. Even more, distance-adaptive EONs are able to dynamically assign the modulation format to each lightpath according to its length [9]. Backbone elastic optical networks also consist of long-haul transmission systems which are set up by many spans. The performance of systems is mainly limited by Optical Signal to Noise Ratio (OSNR) parameter.

Optical Signal to Noise Ratio is defined so as to include the nonlinear interference noise as follows [17],

$$\mathrm{OSNR} = \frac{P_{tx}}{P_{\mathrm{ASE}} + P_{\mathrm{NLI}}}, \tag{1}$$

where P_{tx} is the average transmitted power per channel, P_{ASE} is the amplified spontaneous emission (ASE) noise power and P_{NLI} is the nonlinear interference noise power which accounts for the nonlinear effects.

In fact, P_{ASE} is calculated by [18],

$$P_{\mathrm{ASE}} = S_{\mathrm{ASE}} \times B_0, \tag{2}$$

where S_{ASE} is the ASE noise power spectral density (PSD), B_0 is the noise bandwidth. Depending on the amplifiers applied (i.e. Erbium-Doped Fiber Amplifier (EDFA) or distributed amplification like Raman), S_{ASE} is calculated differently.

If EDFA is used, S_{ASE} can be calculated as the following formulation:

$$S_{\mathrm{ASE}} = N_s h\nu (G-1) N_F, \tag{3}$$

where G is the EDFA gain that is equal to the span loss, N_F is the EDFA noise figure, N_s is the number of EDFAs that is equal to the number of spans, h is Planck's constant and, ν is the optical carrier frequency of the signal being amplified.

On the other hand, P_{NLI} is approximated as [19]

$$P_{\mathrm{NLI}} = \left(\frac{2}{3}\right)^3 N_s \gamma^2 P_{tx}^3 L_{\mathrm{eff}} \frac{\log\left(\pi^2 |\beta_2| N_{ch}^2 R_s^2 L_{\mathrm{eff}}\right)}{\pi |\beta_2| R_s^3} B_0, \tag{4}$$

where P_{tx} is the transmitted power, N_{ch} is the number of channels, γ is the fiber nonlinear parameter, β_2 is the fiber dispersion parameter, R_s is the baudrate ($R_s = R_b/\log_2(M)$ with R_b bitrate and M is the modulation order) and L_{eff} is the effective length is given by

$$L_{\mathrm{eff}} = \frac{1 - e^{-\alpha L}}{\alpha}, \tag{5}$$

where L is the span length, α is the fiber loss parameter.

TABLE I
SUMMARY OF TRANSMISSION PARAMETERS

Parameter	Value
Baud rate	12.5 Gbaud
Channel spacing	12.5 GHz
Centered wavelength	1550 nm
Number of channels	80
Modulation	M-QAM
Fiber loss	0.2 dB/km
GVD coefficient	-21.7 ps^2/km
Nonlinear coefficient	1.4 W^{-1}/km
Amplifier noise figure	5 dB

For distributed amplification, the S_{ASE} and P_{NLI} are calculated as follows:

$$S_{\text{ASE}} = 4\alpha h\nu K_T L_{\text{tot}}, \tag{6}$$

$$P_{\text{NLI}} = \left(\frac{2}{3}\right)^3 \gamma^2 P_{tx}^3 L_{\text{tot}} \frac{\log\left(\pi^2|\beta_2|N_{ch}^2 R_s^2 L_{\text{tot}}\right)}{\pi|\beta_2|R_s^3} B_0, \tag{7}$$

where L_{tot} is the total fiber length, K_T is a constant ($K_T \backsimeq 1.13$ for Raman amplification).

Furthermore, system capacity at Nyquist limit is given in the following equation [17, 20]:

$$\frac{C}{B} = \log_2(1 + \text{SNR}) \text{ (bits/s/Hz/pol)}, \tag{8}$$

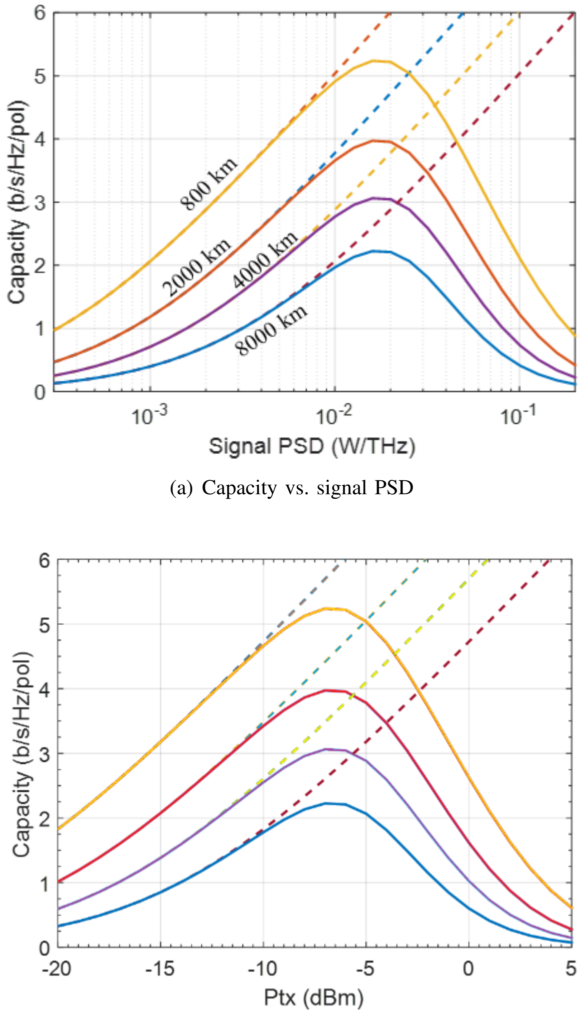
$$\text{SNR} = \frac{2B_0}{R_b}\text{OSNR} = \frac{2B_0}{R_s \log_2 M}\text{OSNR}. \tag{9}$$

And, the bit error rate (BER) for the systems with QPSK and M-QAM (Quadrature Amplitude Modulation) can be estimated respectively as in formulas (10) and (11) [20]:

$$\text{BER} = \frac{1}{2}\text{erfc}(\sqrt{\text{SNR}}), \tag{10}$$

$$\text{BER} = \frac{2\left(1 - \frac{1}{\sqrt{M}}\right)}{\log_2 M}\text{erfc}\left(\sqrt{\frac{3\log_2 M}{2(M-1)}\text{SNR}}\right). \tag{11}$$

In order to estimate the performance limits of an elastic optical transmission system, we assumed values of key parameters as given in Table I. Figure 1 demonstrates the capacity limits per polarization of 16-QAM modulation format (M=16) with respect to the signal PSD and the given transmitted power when the transmission distance is various [20]. Here, the capacity is estimated without and with optical fiber nonlinear effects (shown as dashed and solid graphs). The results confirm that the system capacity is limited due to the nonlinear effects of optical fibers. Consequently, the system performance, in terms of BER, is also strongly depended on the fiber distance and the transmitted power (as shown in Figure 2).



(a) Capacity vs. signal PSD

(b) Capacity vs. signal transmission power

Fig. 1. Information capacity limits per polarization of 16-QAM for linear (dashed line) and nonlinear (solid line) transmission at different transmission distances (L_{span} = 80 km).

On the other hand, the dependence of the system performance, in terms of capacity, BER and the maximum optical reach at BER of 10^{-3}, on the transmission distance and the applied modulation format is shown in Figures 3 and 4 correspondingly. Applying higher-order modulation formats which support higher capacity per spectrum slot and consequently, require less number of spectrum slots. In other words, employing higher-order modulation format can attain higher spectrum efficiency but it suffers from shorter optical transparent reach and as a result, more frequent regeneration and/or more regeneration resources are required. On the other hand, using lower-order modulation formats may cause less spectrum slot capacity and thus, may result in an increment in the required number of spectrum slots. Hence, there exists a trade-off between the system performance and the transmission length/modulation format in elastic optical networks.

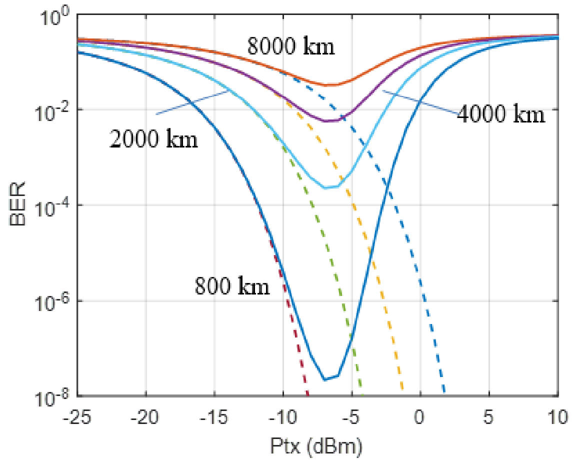
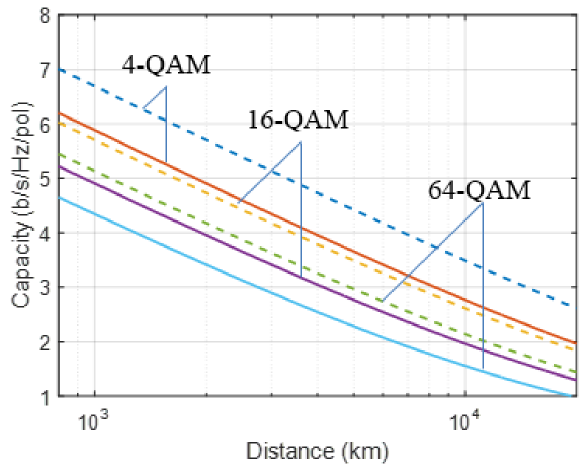
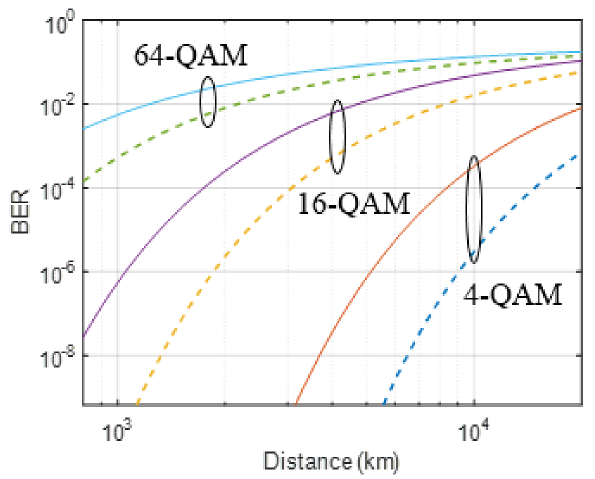


Fig. 2. Performance limits of 16-QAM format for linear (dashed line) and nonlinear (solid line) transmission.



(a) Capacity



(b) BER

Fig. 3. Performance limits vs transmission distance for linear (dashed line) and nonlinear (solid line) transmission schemes at different M-ary (L_{span} = 80 km).

2. Optical Regeneration Issue and Its Application

In general, optical regeneration can be deployed in optical networks to mitigate the accumulation of noises to limit signal degradation and extend the transmission length of optical paths. Optical regenerators are capable of re-amplifying (1R), re-shaping (2R), and even re-timing (3R) the signals and so, called 1R, 2R or 3R regenerators respectively. Among those regenerators, 3R regenerators are required for high-speed and long reach optical paths. Optical regenerator technology for WDM networks is quite mature and it strongly relies on the provided data rate; semiconductor optical amplifier (SOA)-based Mach-Zehnder interferometers have been widely utilized for amplitude modulation to support data rates up to 40 Gbps while ultrafast mechanisms in optical fibers or polarization rotation in SOAs are popularly employed to deal with higher rates such as beyond 40 Gbps [10, 14]. Recently, all-optical 3R (reamplifying, reshaping, and re-timing) regeneration devices can directly handle the optical signal degradation caused by fiber loss, dispersion and the amplified spontaneous emission noise of EDFAs or fiber nonlinearity, and overcome the bandwidth bottleneck of optical–electrical–optical (O/E/O) processing systems.

In addition to conventional regenerators, a new regeneration technology, named virtualized elastic regenerator (VER), that is able to regenerate flexibly and cost-effectively multiple-data-rate optical signals, has been proposed to realize emerging elastic optical networks [14]. VERs are constructed based on spectrum-selective sub-channel regenerators (SSRs) where each SSR is capable of regenerating a sub-channel, i.e., 100 Gbps. A VER includes a set of SSRs and all the incoming optical signals, that need to be regenerated, share the regeneration resource, says SSR pool. By sharing SSRs in the regeneration pool of VER, optical signals with various data rates can be cost-effectively regenerated. Based on the data rate of each incoming optical signal, a number of necessary SSRs is allocated to regenerate the optical signal. In order to provide the spectrum selectivity in VER, optical coherent detection with a wavelength tunable local oscillator (LO) is employed. Along with the multiple-channel and various data rate regeneration capability, a VER is also able to convert modulation format distance-adaptively (also known as re-modulation capability) and hence, it enables saving the spectrum utilization when the transmission distance is sufficiently short enough to use more spectrum-efficient (higher-order) modulation formats. In other words, VERs can provide 4R regeneration function. However, note that VER cost seems to be expensive and relies strongly on the SSR pool size, the number of SSRs [15, 16].

From network point of view, elastic optical networks that are partially with VER-equipped nodes (i.e. recon-

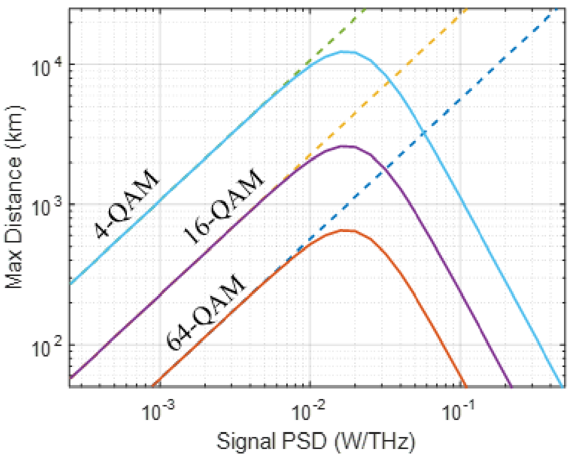


Fig. 4. Transmission distance limits vs signal PSD per channel at BER = 10^{-3} with L_{span} = 80 km for linear (dashed line) and nonlinear (solid line) transmission schemes at different M-ary QAM formats.

figurable optical add-drop multiplexer (ROADMs)/optical cross-connects (OXCs)) are capable of not only providing distance-adaptive and multiple modulation format optical paths, but also offer the optical path regeneration (including 3R and 4R regeneration). Lightpaths can be regenerated with/without being re-modulated only at the nodes that consist of VERs. Furthermore, to establish lightpaths dynamically and effectively in elastic optical networks, routing, modulation and spectrum assignment (RMSA) problem must be solved efficiently. The applied RMSA scheme must be able to overcome the optical reach limit constraint while exploiting distance-adaptive modulation capability of the networks. As discussed in [18], with a given capacity of traffic demand, higher-order modulation formats may offer a spectrum reduction while suffer from an increase in the regeneration utilization frequency and/or necessary regenerator devices. Contrarily, a lower-order modulation format can help to extend the optical reach and so, may reduce the number of necessary regenerator devices but, it may result in less spectrum utilization efficiency. Therefore, employing 4R regeneration with an effective RMSA algorithm enables improving the spectrum usage efficiency while dealing with the spectrum continuity and spectrum consecutiveness constraints in elastic optical networks.

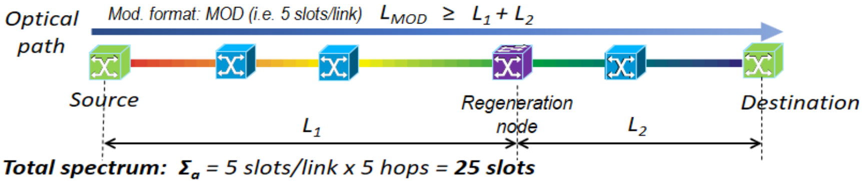
In fact, elastic optical networks can use various types of regenerators (3R or 4R) and low-order modulation formats can be employed to extend the optical reach and avoid the use of regenerators. Figure 5 illustrates the differences among three regeneration applicable elastic optical network scenarios that are a network without using regenerators, 3R and 4R regeneration capable ones. Here, L_{MOD} is the optical reach required if the modulation format, MOD, is used while L_1 and L_2 are the sub-path lengths. The distance adaptive modulation capability can be exploited by using appropriate regenerators to enhance the spectrum utilization efficiency of the networks (the required L_{MOD} is shorter). Obviously, selection of regeneration nodes and modulation format and spectrum assignment strategy are the key for minimizing the network performance.

III. Network Performance Evaluation

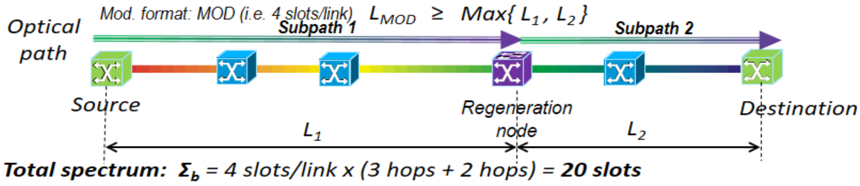
In this section, we evaluate the performances of dynamic elastic optical networks using different optical regeneration techniques including 3R- and 4R-capable regeneration, in term of the overall blocking probability and the accommodated traffic volume. In order to clarify the impact of optical regeneration in dynamic elastic optical networks, the network performances of three regeneration applicable network scenarios including (i) no regeneration, (ii) only 3R regeneration and (iii) fully-capable elastic regeneration (4R) will be tested and compared. For fair comparison, we employ similar RMSA algorithms for those cases. We also use two different RMSA algorithms which are a simple shortest path and first fit algorithm (called First Fit) [21] and the spectrum-least RSA algorithm which has been developed in [22] (denoted as Least Spectrum).

Moreover, two typical network topologies, that are (i) National Science Foundation network (NSF) consisting of 14 nodes and 22 links, and (ii) US backbone network (USNET) including 24 nodes and 43 links (shown in Figure 6) are used for numerical experiments. The regeneration capable node number is assumed to be limited at 4 and 7 for NSF and USNET respectively and each case will be tested with 20 random scenarios. We also use following parameters for the numerical simulation. Each fiber link can carry up to W spectrum slots (W is fixed at 128) and the slot bandwidth is assumed to be 12.5 GHz. The networks can flexibly and dynamically set up and release optical paths (also called lightpaths). Modulation format of each lightpath is distance-adaptively assigned. Lightpath requests arrive sequentially and follow Poisson distribution. Average arrival rate of ligthpaths is λ (requests per time unit). Distribution of lightpath holding time is assumed to be a negative exponential one with the mean hold time of $1/\mu$ (time units). Consequently, the given network traffic load in Erlangs is λ/μ. Here, the capacity, C, of each requested lightpath between node pairs is also randomly assigned between 50 and 100 Gbps following a uniform distribution. The assumed modulation formats are BPSK, QPSK, 8-QAM and 16-QAM. The slot bandwidth and the corresponding transparent reach of BPSK, QPSK, 8-QAM, and 16-QAM optical signals are given in Table II [21, 22].

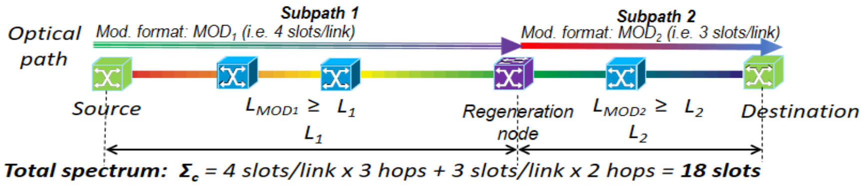
As mentioned above, we evaluate the network performance in three comparative regeneration scenarios of elastic optical networks: (i) without regeneration capability (named



(a) No regeneration scenario

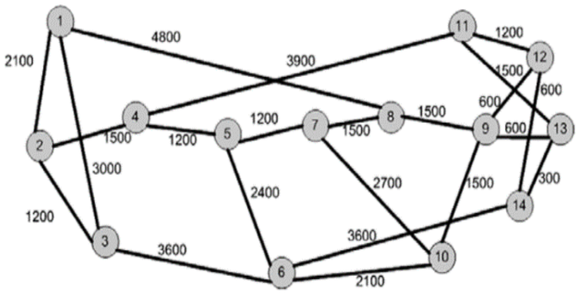


(b) 3R regeneration scenario

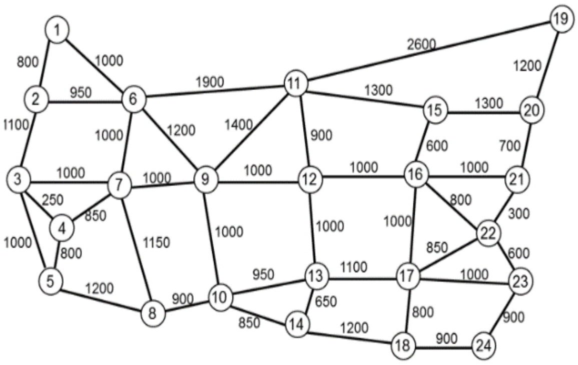


(c) 4R regeneration scenario

Fig. 5. Regeneration network scenario comparison.



(a) NSF



(b) USNET

Fig. 6. Experimental network topologies (NSF and USNET).

TABLE II
SUMMARY OF KEY PARAMETERS

Parameter			Value
Spectrum slot number per link			128
Slot bandwidth			12.5 GHz
Modulation format	BPSK	Slot capacity	12.5 Gbps
		Transparent reach	9600 km
	QPSK	Slot capacity	25 Gbps
		Transparent reach	4800 km
	8-QAM	Slot capacity	37.5 Gbps
		Transparent reach	2400 km
	16-QAM	Slot capacity	50 Gbps
		Transparent reach	1200 km
Capacity of requested lightpaths			50-100 Gbps
Connection mean hold time			1000

No regeneration), (ii) 3R regeneration with two different RSA algorithms including First-fit and Least-spectrum algorithms (so called First fit w/ 3R and Least spectrum w/ 3R correspondingly), and (iii) 4R regeneration also with First-fit and Least-spectrum RSA algorithms (denoted by First fit w/ 4R and Least spectrum w/ 4R respectively). The performance of the comparative network scenarios, in terms of the blocking probability, is shown severally in Figures 7 and 8 for NSF and USNET network topologies when the traffic load ranges from 100 to 1500 Erlangs. These figures verify that the network scenarios using optical regeneration, i.e. 3R or 4R regeneration, can dramatically decrease the blocking probability in comparison with that without regeneration. That is because implementing optical regeneration can help to resolve the spectrum collision, save the spectrum resources with higher modulation formats and enhance the optical reach to improve the network utilization. Moreover, the results also show that applying

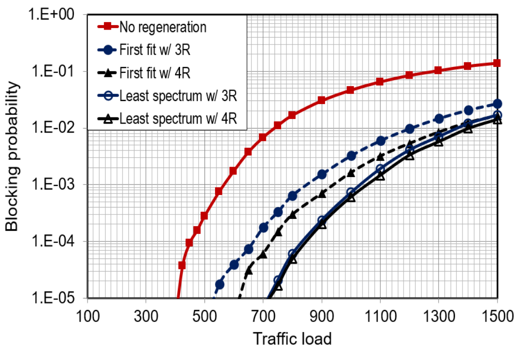


Fig. 7. Blocking probability for NSF network.

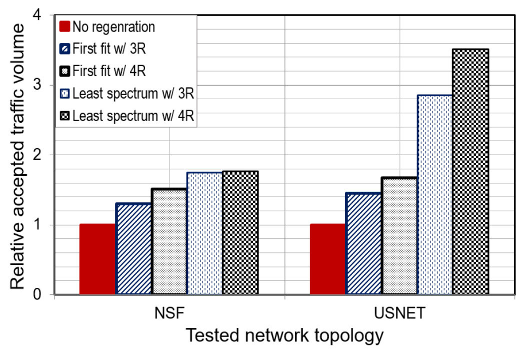


Fig. 9. Accommodated traffic volume comparison.

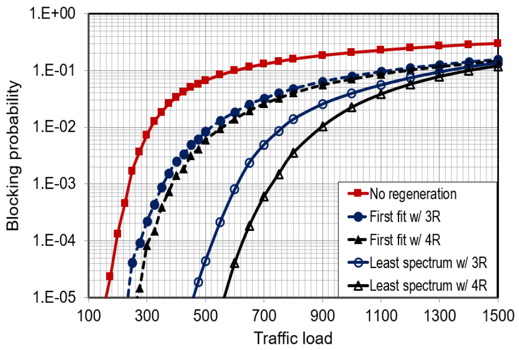


Fig. 8. Blocking probability for USNET network.

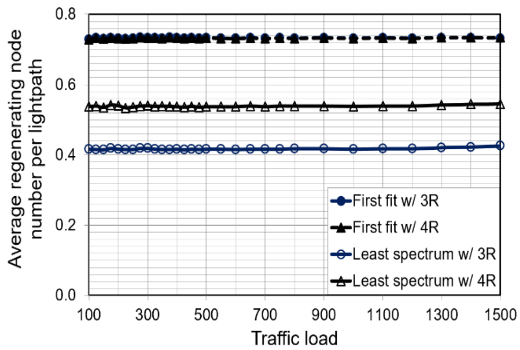


Fig. 10. Average number of regenerating nodes per lightpath for NSF network.

4R regeneration offers better performance than that of 3R regeneration. Such performance improvement is obtained by re-modulating the optical signal into the optimal modulation format for exploiting the distance-adaptive feature of elastic optical networks.

Furthermore, as being verified in Figures 7 and 8, the performance of the networks also strongly depends on the applied RMSA algorithm due to the effect of selecting the regeneration nodes and the modulation format. Optimizing the regenerating node position and assigning suitable modulation formats play an important role to enhance the overall spectrum utilization efficiency. It is confirmed that the effect of optical regeneration is enhanced with larger network. The main reason is that high-order modulation levels can be assigned to shorter lightpaths and consequently, help to lessen the number of spectrum slots required. On the other hand, larger network which contains longer average length of lightpaths requires more regeneration resources, especially for the lightpaths with high-order modulation format due to the transparent reach limitation.

Moreover, Figure 9 demonstrates the comparison of relative accepted traffic volumes among the experimented scenarios, that are No regeneration, First fit w/3R, First fit w/ 4R, Least spectrum w/ 3R and 4R, with the given blocking probability of 10^{-5}. The results obtained by No regeneration is used as the benchmark, so its graph is 1. The figure describes that, with the same network conditions, using optical regeneration provides higher accepted traffic volume thanks to the signal regeneration. We can attain at least 29.8% (45.6%) higher traffic volume with NSF (USNET) network. Implementing more-efficient RSA algorithm (i.e. spectrum least) even can further improve the network performance. However, note that it may results in higher network CAPEX. In fact, we must consider the cost of regeneration devices carefully to satisfy the tradeoff between the CAPEX and the network performance.

We have also estimated the average number of regenerating nodes and that of spectrum slots per lightpath in the experimental network topologies to determine the requirement of necessary regeneration resources. Figures 10 and 11 illustrate the obtained numerical results for NSF topology and Figures 12 and 13 show that of USNET respectively. Because of the use of same regenerating node selection strategy, both First fist w/ 3R and First fit w/ 4R scenarios need almost the same average number of regenerating nodes for each lightpath. However, the number of regenerated spectrum slots is decreased with the use of 4R regeneration thanks to modulate the signals distance-adaptively. The graphs also verify that, with the Least spectrum algorithm, the improvement of the network performance costs more regenerating resources.

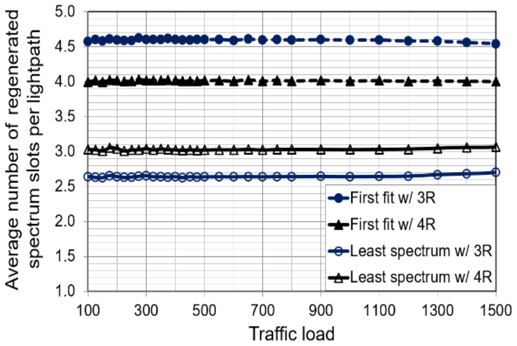


Fig. 11. Average number of regenerated spectrum slots per lightpath for NSF network.

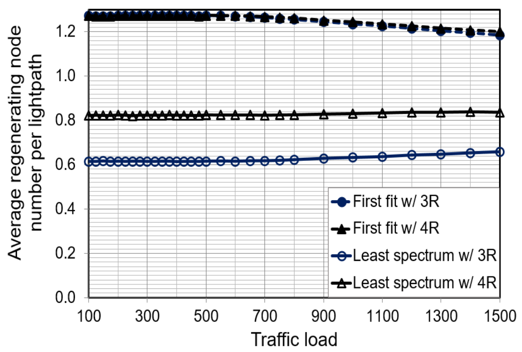


Fig. 12. Average number of regenerating nodes per lightpath for USNET network.

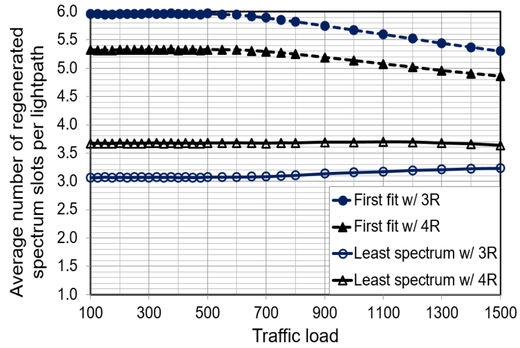


Fig. 13. Average number of regenerated spectrum slots per lightpath for USNET network.

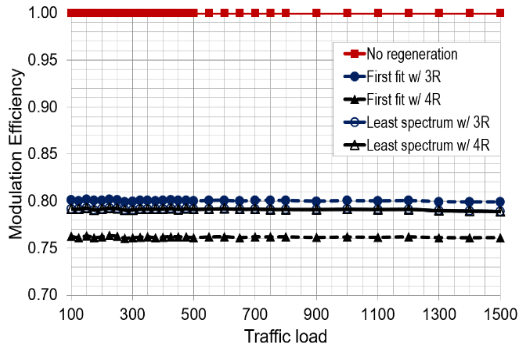


Fig. 14. Modulation efficiency of NSF network.

Finally, Figures 14 and 15 illustrate the modulation efficiency, which is the ratio of the average spectrum slot number per established lightpath of each compared scenario to that of No regeneration one, of both NSF and USNET. The spectrum slot number of No regeneration scenario is used as a benchmark and consequently, its normalized value is one. It demonstrates that although First fit strategy offers better modulation efficiency (less ratio) for established lightpaths, it does not consider the spectrum collision an as a result, its performance is poorer than that of Least spectrum strategy.

IV. Conclusion

In this paper, we have investigated the impact of various optical regeneration techniques including 3R and 4R on the performance of dynamic elastic optical networks. We firstly discussed and clarified the differences among optical regenerating usage scenarios without regeneration and with regeneration (both 3R and 4R regeneration). We have, then, evaluated and compared the dynamic elastic optical network performance, in term of the blocking probability and the total accepted traffic volume, etc. for three comparative network scenarios including (i) no regeneration, (ii) with 3R regeneration and (iii) with 4R regeneration capabilities. Numerical simulation results prove that using optical regenerators, especially 4R regenerators, can help to exploit elastic optical networking and enhance the dynamic network performance for provisioning bandwidth-flexible lightpath services. It was demonstrated that at least 29.8% (or 45.6%) more traffic volume can be attained for NSF (USNET) network. However, optical regenerating devices are expensive and hence, the trade-off between the required network performance and the necessary regenerating resource cost should be considered carefully in order to create future cost-effective, spectrum-efficient and bandwidth-flexible optical networks.

Acknowledgment

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.02-2015.39.

References

[1] Cisco, "Cisco visual networking index: Forecast and methodology, 2014–2019." [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.pdf

[2] K. Sato and H. Hasegawa, "Optical networking technologies that will create future bandwidth-abundant networks," *J. Optical Com. Net.*, vol. 1, pp. A81–A93, 2009.

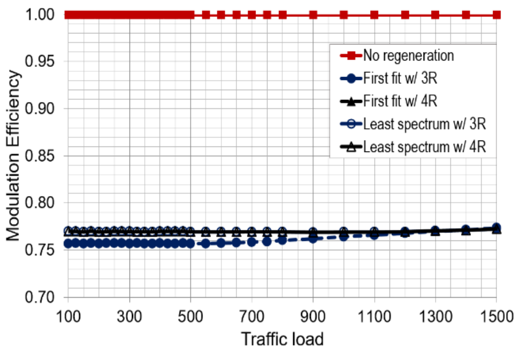


Fig. 15. Modulation efficiency of USNET network.

[3] G. Zhang, M. De Leenheer, A. Morea, and B. Mukherjee, "A survey on OFDM-based elastic core optical networking," *IEEE Com. Surveys & Tutorials*, vol. 15, pp. 65–87, 2012.

[4] A. Jukan and J. Mambretti, "Evolution of optical networking toward rich digital media services," *Proceedings of the IEEE*, vol. 100, no. 4, pp. 855–871, 2012.

[5] M. Jinno, H. Takara, B. Kozicki, Y. Tsukishima, Y. Sone, and S. Matsuoka, "Spectrum-efficient and scalable elastic optical path network: architecture, benefits, and enabling technologies," *IEEE Com. Mag.*, vol. 47, pp. 66–73, 2009.

[6] O. Gerstel, M. Jinno, A. Lord, and S. B. Yoo, "Elastic optical networking: A new dawn for the optical layer?" *IEEE Com. Mag.*, vol. 50, 2012.

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

[8] A. Lord, P. Wright, and A. Mitra, "Core networks in the flexgrid era," *Journal of Lightwave Technology*, vol. 33, no. 5, pp. 1126–1135, 2015.

[9] M. Jinno, B. Kozicki, H. Takara, A. Watanabe, Y. Sone, T. Tanaka, and A. Hirano, "Distance-adaptive spectrum resource allocation in spectrum-sliced elastic optical path network," *IEEE Com. Mag.*, vol. 48, pp. 138–145, 2010.

[10] A. Klekamp, R. Dischler, and F. Buchali, "Limits of spectral efficiency and transmission reach of optical-OFDM superchannels for adaptive networks," *IEEE Photonics Technology Letters*, vol. 23, no. 20, pp. 1526–1528, 2011.

[11] I. Cerutti, F. Martinelli, N. Sambo, F. Cugini, and P. Castoldi, "Trading regeneration and spectrum utilization in code-rate adaptive flexi-grid networks," *Journal of Lightwave Technology*, vol. 32, no. 23, pp. 3894–3901, 2014.

[12] A. Bocoi, M. Schuster, F. Rambach, M. Kiese, C.-A. Bunge, and B. Spinnler, "Reach-dependent capacity in optical networks enabled by OFDM," in *Conf. Optical Fiber Communication*, 2009, pp. 1–3.

[13] H.-C. Le, H. Hasegawa, and K.-i. Sato, "Impact of electrical grooming and regeneration of wavelength paths in creating hierarchical optical path networks," *Optics express*, vol. 19, no. 26, pp. B882–B894, 2011.

[14] M. Jinno, K. Yonenaga, H. Takara, K. Shibahara, S. Yamanaka, T. Ono, T. Kawai, M. Tomizawa, and Y. Miyamoto, "Demonstration of translucent elastic optical network based on virtualized elastic regenerator," in *National Fiber Optic Engineers Conf.* Optical Soc. America, 2012, pp. PDP5B–6.

[15] M. Jinno and H. Takara, "Elastic optical transponder and regenerator: Toward energy and spectrum efficient optical transport networks," in *2012 International Conference on Photonics in Switching (PS)*, 2012, pp. 1–3.

[16] H.-C. Le, N. T. Dang, and S. Vu-Van, "Dynamic multi-domain elastic optical networks with 4R regeneration capable border nodes," in *National Foundation for Science and Technology Development Conference on Information and Computer Science (NICS)*, 2015, pp. 169–173.

[17] P. Poggiolini, A. Carena, V. Curri, G. Bosco, and F. Forghieri, "Analytical modeling of nonlinear propagation in uncompensated optical transmission links," *IEEE Photonics technology letters*, vol. 23, no. 11, pp. 742–744, 2011.

[18] G. Bosco, P. Poggiolini, A. Carena, V. Curri, and F. Forghieri, "Analytical results on channel capacity in uncompensated optical links with coherent detection," *Optics express*, vol. 19, no. 26, pp. B440–B451, 2011.

[19] R.-J. Essiambre, G. Foschini, P. Winzer, and G. Kramer, "Capacity limits of fiber-optic communication systems," in *Optical Fiber Com. Conf.* Optical Soc. America, 2009.

[20] J. G. Proakis and M. Salehi, *Digital Communications*, 4th ed. McGraw-hill New York, 2000.

[21] Z. Zhu, W. Lu, L. Zhang, and N. Ansari, "Dynamic service provisioning in elastic optical networks with hybrid single-/multi-path routing," *Journal of Lightwave Technology*, vol. 31, no. 1, pp. 15–22, 2012.

[22] H.-C. Le, N. T. Dang, and N. D. Nguyen, "Dynamic RMSA algorithm for multi-domain EONS with limited virtualized elastic regenerators," in *International Conference on Communications and Electronics (ICCE)*, 2016, pp. 46–51.

Le Hai Chau received the B.E. degree in Electronics and Telecommunications Engineering from Posts and Telecommunications Institute of Technology (PTIT) of Vietnam in 2003, and the M.Eng. and D.Eng. degrees in Electrical Engineering and Computer Science from Nagoya University of Japan in 2009 and 2012, respectively. From 2003 to 2006, he worked at PTIT as a lecturer. He was a researcher in Nagoya University from 2012 to 2014. He then became a research scientist at University of California, Davis. Now, he is a lecturer at PTIT. He is an IEEE member.

Dang Hoai Bac received the B.E. degree from the Hanoi University of Technology ,Hanoi, Vietnam, in 1997, and both the M.E. degree and the Ph.D. degree in Electronics and Telecommunications from the Posts and Telecommunications Institute of Technology (PTIT), Hanoi, Vietnam in 2004 and 2010, respectively. He is currently an Associate Professor/Vice Director at PTIT. His current research interests include the area of communication theory with a particular emphasis on modeling, design, and performance evaluation.

Effect of Multi-site Variabilities on Electrovestibulography: Environmental and Physical Factors

Invited article

Brian J. Lithgow[1,2], Zahra Moussavi[1]
[1] University of Manitoba, Winnipeg, MB, Canada
[2] Monash University, Melbourne, Victoria, Australia
Correspondence: Brian J. Lithgow, brian.lithgow@umanitoba.ca

Communication: received 11 August 2019, revised 28 August 2019, accepted 30 August 2019
Online early access: 30 August 2019, Digital Object Identifier: 10.32913/mic-ict-research.v2019.n1.885
The Editor coordinating the review of this article and deciding to accept it was Prof. Nguyen Linh Trung

Abstract:

Background: **There are physiological changes in pathologies such as Alzheimer's Disease (AD) within the lower vestibulo-acoustic system, which may be potentially useful when used as neurodegeneration features. We hypothesize two Electrovestibulography feature types (Field Potential (FP) shape and the Firing pattern of detected FP's) may have utility as Neurodegeneration features. Our long term objective is to use a population of Parkinson's Disease (PD), AD, Post-Concussion Syndrome (PCS), Bipolar Disorder (BD) and Major Depressive Disorder (MDD) patients together with individual pathology-wise age and gender matched control cohorts to determine the degree to which each of these pathologies varies from controls and in proportion to the level of Neurodegeneration often associated (either temporarily or permanently) with each pathology. However, before such a comparison can be made it is necessary to ensure the various populations recorded across different countries are comparable. This paper determines which populations are comparable.**

Methods: **An initial comparison of AD (with N = 16) and a best matched healthy control population (with specific age/gender/recording site/electrode matched controls) from a pool of 112 controls produced two EVestG features (FP shape and FP firing pattern). These features were examined for their variability with respect to electrode type, age, gender, powerline frequency and environmental factors.**

Results: **Age and gender did not have a significant impact on the features. Powerline and environmental artefacts could be accounted for by filtering; thus, they did not significantly affect the features measured. However, electrode type had a significant effect on the extracted features.**

Conclusions: **For the two EVestG features tested only electrode type had a significant effect on the recordings, and hence the extracted features. Thus, only populations with the same electrode type can be compared.**

Keywords: *Multi-site recording variability, electrodes, powerline frequency, gender, age, Electrovestibulography, biological markers, vestibular.*

Glossary

AD	Alzheimer's Disease
Background phase	1.5sec EVestG recording immediately prior to motion
EVestG	Electrovestibulography
EVS	Efferent Vestibular System
FP	Field Potential
$IH_{33}1$	EVestG long window interval firing pattern feature recorded during static (no motion) time
LDA	Linear Discriminant Analysis
N1, N2	Negative peaks of compound Action potential
NEER	Neural Event Extraction Routine
PCS	Post-Concussion Syndrome
ROC	Region of Convergence
Sh1	Background (no motion) shape feature
VN	Vestibular Nucleus

I. Introduction

In studies care is taken to match age, gender, recording methodology and environmental factors (e.g. powerline frequency). This is even more important when comparing populations recorded in multi-site studies, wherein there are differences that can be potentially very difficult to control for such as powerline frequency (artefacts). Electrovestibulography (EVestG) is a non-invasive measure from the vestibulo-acoustic system in both background (no motion)

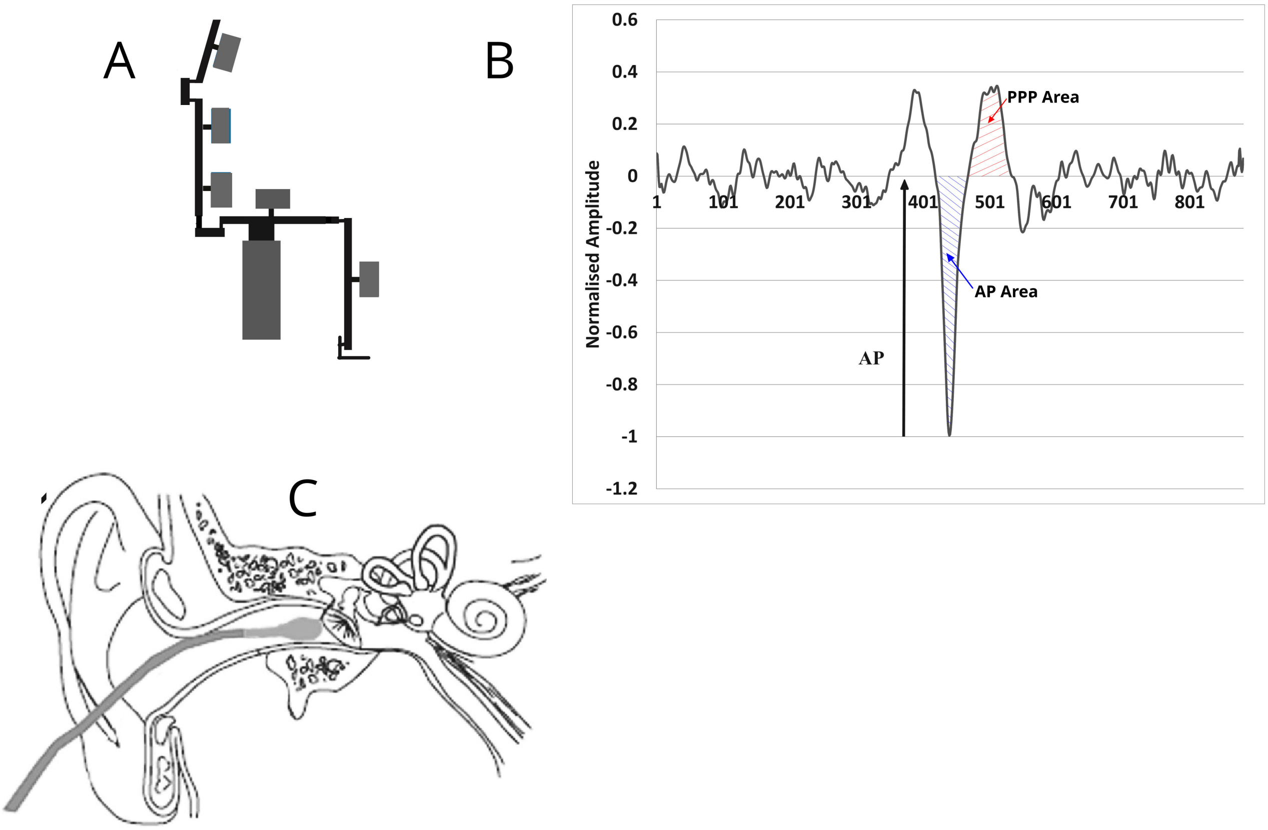


Figure 1: EVestG Methodology: (A) EVestG chair; (B) Example averaged field potential (FP) showing the AP and PPP areas (horizontal scale is in samples (41.67 kHz)); (C) Ear canal electrode placement, a reference is placed on the earlobe or outer ear canal, a common ground silver/silver chloride electrode is placed on the forehead.

and/or in response to passive whole-body tilt [1–3] stimuli. There is evidence that the EVestG responses are vestibulo-acoustic and predominantly vestibular [1–5]. EVestG assessment has been applied to the detection of a number of pathologies including Alzheimer's Disease (AD) [5], Parkinson's Disease (PD) [6], Post-Concussion Syndrome (PCS [7–9], Vertiginous disorders [10–12], Major Depressive Disorder (MDD) [2], non-manic Bipolar Disorder [1] and the separation of Bipolar and MDD [13]. In previous studies we have collected Electrovestibulography recording data from Parkinson's Disease (PD), Alzheimer Disease (AD), Post-Concussion Syndrome (PCS), Major Depression and Bipolar Depression populations as well as a pool of healthy controls (N = 117). These were recorded across two sites (Australia and Canada), each with different powerline frequencies, slightly different equipment setups, different age groups and gender ratios, and also quite different recording electrodes (BiopacTM versus soft cotton-wool electrodes). In this study, we aim to determine which populations are or can be made directly comparable for the two EVestG main feature types that we commonly measure and apply in neurodegeneration studies.

In our recent neurodegeneration studies (i.e. Parkinsons, PCS, AD) we have commonly used two feature types. Those features types are the average extracted field potential shape and the pattern of occurrence of the detected field potentials. The results have been quite promising with accuracies above 80%. These feature types have been hypothesized to be representatives of channelopathies and neurotransmitter changes.

Alzheimer's Disease (AD) and EVestG: In a recent study [7], one feature of the EVestG response used to identify post-concussion syndrome (PCS) individuals from age-and-gender-matched healthy individuals with an 84% leave-one-out cross-validated test accuracy (ROC = 84%) was the shape of the average Field Potential (FP) signal. This feature was hypothesized to be linked to channelopathies. A recent study suggests the existence of a time- and dose-dependent risk of developing dementia more than 30 years after Traumatic Brain Injury (TBI); the risk being higher for more severe and multiple TBIs than for single mild TBI [14]. These ideas were carried forward in a pilot AD study using EVestG [5]. In that AD study the EVestG features used were developed based on:

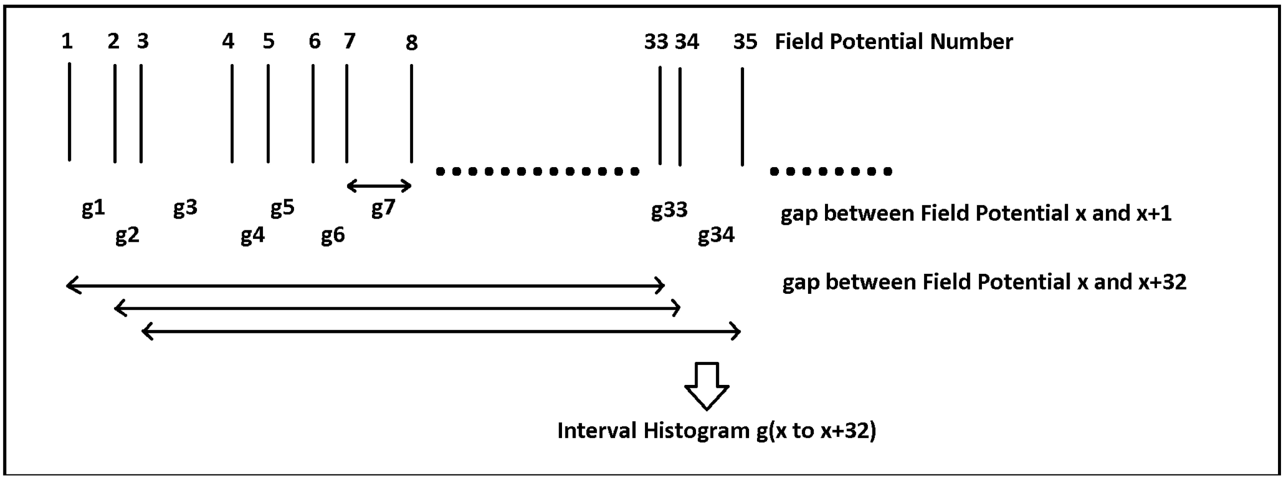


Figure 2: IH_{33} formation.

1) A link between the amyloid precursor protein and an increase in the surface expression of some voltage dependent Na+ channels [15]. Thus, one AD classification/detection feature should search for FP shape changes relevant to an increase in the number of Na+ channels; this was hypothesized to be and then shown to appear as a steeper descending arm in the average EVestG AP part (Figure 1B) of the recorded FP in AD patients.
2) The contribution of GABAergic dysfunction to AD is known or suspected [16, 17]. Cortes, 2013 suggests GABA could act as a facilitator in the spontaneous discharge of the vestibular afferents by altering intracellular Ca++ concentration [18, 19]. Reduced GABA then may lead to a de-facilitation in spontaneous (and driven) discharge. Thus, a feature sensitive to firing pattern changes was hypothesized and then shown to be a sensitive second AD biomarker [5].

Thus, there are physiological changes in AD within the lower vestibulo-acoustic system [5], which might be potentially useful as Neurodegeneration features. We have hypothesized the two EVestG feature types (FP shape and FP firing pattern) may have utility as neurodegeneration features. Our focus herein is to use these two features initially derived from a comparison of an AD patient population with its best matched healthy control population and examine those features for age/gender/recording site/electrode type variability using best matched healthy subgroups from our pool of 112 controls. This will allow us to determine which previously recorded neurodegenerative populations are directly comparable for the two EVestG feature types that have been commonly measured and applied to neurodegeneration detection studies.

II. METHODS

The detailed methodology for EVestG recordings is given in [1–3, 9]. Briefly, EVestG recordings are made inside an electromagnetically shielded and sound attenuated (greater than 30dB) chamber. To minimize movement artefacts subjects are positioned in a sitting position on a hydraulic chair. During EVestG recordings participants have their eyes closed and are in a relaxed state with their neck supported. A gelled wick electrode is positioned close to the tympanic membrane (Figure 1C). Recordings analyzed herein were made with each subject stationary. Each trial is repeated twice, the first trial is for subject familiarization and data from the second trial recorded. The subject remains passive with closed eyes, and is not required to interact cognitively during the measurement. A background recording of spontaneous activity 1.5sec immediately prior to any potential chair's movement is considered as the static phase response. Only static recordings were used in this study.

The Neural Event Extraction Routine (NEER-V4) [3] averages the detected spontaneous field potentials (FPs) to produce an average FP plot (e.g. Fig. 1B) and generate a firing pattern interval histogram (IH_{33}) of the FP occurrence times. These intervals were calculated as shown in Figure 2. The IH_{33} plot represents low frequency (10Hz) modulations of detected spontaneous FP interval activity, as hypothesized [2] to occur in response to efferent and or EEG (potentially α band) activity. Spontaneous vestibular efferent activity is seen at 10 – 50 spikes/sec [20] and the α band is 8 – 13 Hz. Average interval histograms were generated based on every 33^{rd} FP gap (IH_{33}). A 33 FP gap corresponds to about 100ms (10Hz) and the lower range of any hypothesized potential efferent/α band modulatory

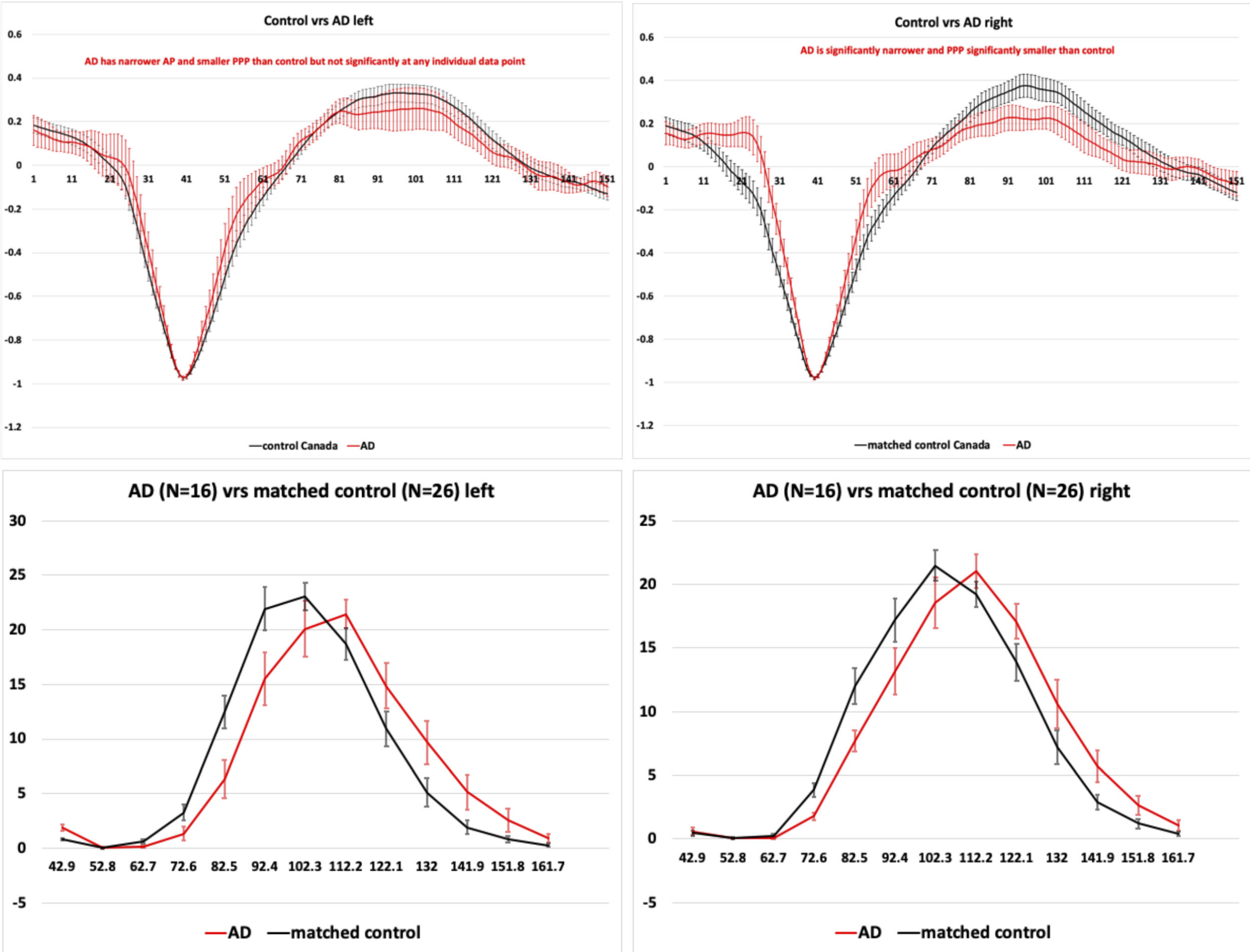


Figure 3: AD versus matched control data (95% error bars). Upper plots are average FP responses for left and right sides. Lower plots are $IH_{33}1$ responses for left and right sides. Best match means matched electrodes, age, gender, site frequency etc.

burst firing effect as the average experimentally detected gap measured with NEER was 3.3ms. Figure 3 shows an example IH_{33} plot. SPSS Version 24 software (IBM, New York, NY, USA) was used for statistical analysis. Statistical significance (p-value) was set at 0.05 unless specified.

Study cross-compatibility

The data recorded in Australia was with a powerline frequency of 50Hz and hydraulic jitter frequency (HJF) of 940 or 1024 Hz. HJF is the frequency at which the hydraulic controller used in controlled tilts of the chair is applied to prevent "sticking" of the piston between and during motion. Comparatively, the data recorded in Canada was with a powerline frequency of 60Hz and HJF of 980Hz. Within NEER-V4 notch filtering was applied at all odd powerline harmonics and at multiples of the HJF up to 4500Hz. This ensured best match of responses across sites. The AD population were all recorded in Canada, and included 16 individuals (with 9 females, age: 66.7 ± 10.1 (SD), MOCA = 13.4 ± 5.7). The control population demographics were: Australia, N = 57, Age = 46.7 ± 15.5yrs (SD), 29 Females; Canada, N = 54, Age = 53.0 ± 18.2yrs (SD), 29 Females. The Australian site used Sanibel^{TM} in the ear recording electrodes (Figure 4), had multiple of 50Hz powerline notch filters and a hydraulic artefact frequency of 1024 or 940 Hz (selected automatically). The Canada site used both Sanibel^{TM} and cotton wool (in house) in the ear recording electrodes (Figure 4), had multiple of 60Hz powerline notch filters and a hydraulic artefact frequency of 980 Hz.

III. Results

1. AD versus Control Feature formation

It was hypothesized that the AD response relative to the other neurodegenerative pathologies would be most different from best matched controls if the features used

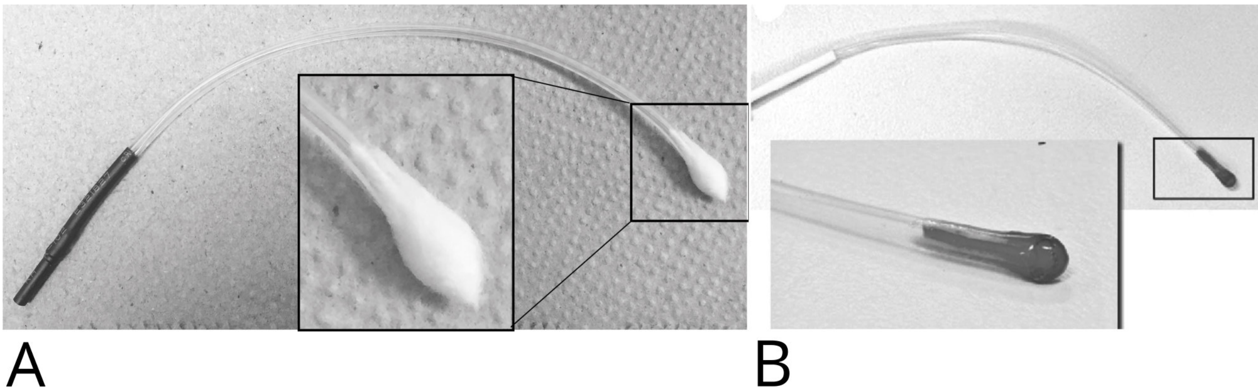


Figure 4: Electrode Types: (left) In-house Cotton-wool tipped soaked in saline and conductive Gel; (right) TM ECochG Electrode, Sanibel.

were neurodegenerative in nature. Examination of the FP response for AD (N = 16) and best matched controls (N = 26) (Figure 3A) showed not only the Action potential (AP) area as significantly smaller (narrower) but also the Post-Potential Peak (PPP see Figure 1B) area. Accordingly, these combined areas were chosen as a feature Sh1. Arguing similarly, the AD IH_{33} should be most different from matched controls. Examination of the AD IH_{33} plot shows a marked shift to the right (Figure 3B) relative to matched controls. This shift was chosen as feature $IH_{33}1$.

Shape (Sh1)- Average Sh1 data for AD versus control groups were found to have significantly ($p < 0.05$) smaller AP and PPP areas on both the left and right sides useful in characterizing populations (Figure 3A). The area of the AP plus PPP was calculated as the area between the baseline ($y = 0$) and the AP plus the area between the baseline and PPP (Figure 1B). For AD the Figure 3 upper plots show a narrowing in the AP and PPP areas on both the left and right sides with the right side displaying significant pointwise differences. The Figure 3 lower plots show a significant rightward shift for the right and left sides for AD relative to matched controls.

For AD versus matched control classification accuracy using the 2 right side features Sh1 and $IH_{33}1$, the one-way MANOVA revealed a significant multivariate main effect (Wilkes $\lambda = 0.687$, $F(2,39) = 8.869$, $p = 0.001$, partial $\eta^2 = 0.313$, power= 0.961). The hypothesis that the interaction of features Sh1 and $IH_{33}1$ has a significant effect on the classification of AD from matched controls is confirmed. The Boxes M test for homogeneity of covariances ($p = 0.206$) and the Levene's test for equality of variances ($p = 0.601$_Sh1, $p = 0.526$_$IH_{33}1$) were not significant. Given the significance of the overall test, the univariate main effects were examined. Significant univariate main effects were obtained supporting features Sh1 and $IH_{33}1$ being applied to AD versus matched control classification: feature Sh1 $F(1,40) = 13.393$, $p = 0.001$, $\eta^2 = 0.259$, power = 0.954; and feature IH_{33} $F(1,40) = 16.853$, $p < 0.001$, $\eta^2 = 0.296$, power = 0.980.

2. Site Comparability Using Best Matched Controls

1) Comparison of electrode types: At the Canada site, it was determined the signal to noise ratio of the EVestG recording was markedly improved by using impedance matched electrodes [21]. For this reason, in-house made cotton wool (active in the ear and reference earlobe/outer ear canal) electrodes (Figure 4) replaced the previously used higher impedance and impedance mismatched Sanibel/Biopac electrode combinations for recording. The in-house ear electrodes were silastic wrapped silver wire with the tip covered in cotton wool soaked in a mixture of saline and conductive gel to reduce interface impedance. Biopac electrodes (in the ear-TM ECochG Electrode, Sanibel (Figure 4) earlobe reference Biopac silver-silver chloride 4mm, EL254) were exclusively used at the Australian site as these, date wise, were the first EVestG recordings. At the Canada site we could compare the responses of the two electrode types.

Figure 5 clearly shows the response of the cottonwool electrodes (N = 17 left, N = 17 right) is narrower in the AP and PPP than that of the Sanibel electrodes (N = 17 left, N = 17 right) for the Canada site. The narrowing was significant ($p < 0.05$) on the right in both the AP and PPP areas. The yellow ellipses mark regions of significant difference. The populations were age (Sanibel, 58.7 ± 4.3 (SD), cotton wool, 54.7 ± 12.6yrs) and gender (11F, 10F) matched. Accordingly, comparisons should be made only between common electrode types.

Figure 6 shows the Cottonwool (54.7 ± 4.3yrs, N = 17, and 10F) compared to the Sanibel (58.7 ± 12.7yrs, N = 17, and 11F) age and gender matched responses from the same

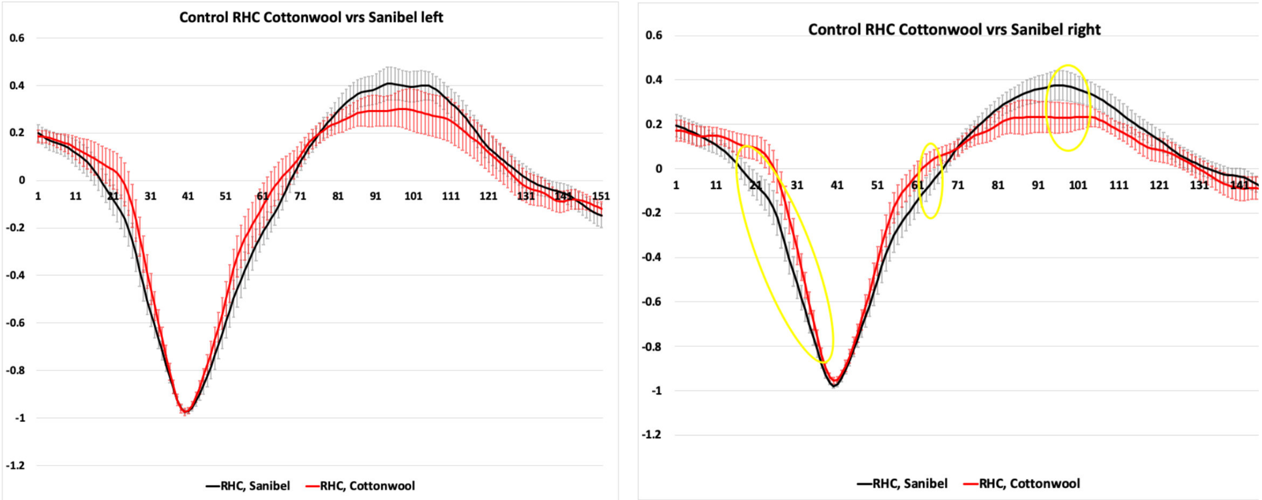


Figure 5: FP response: AP plus PPP area shape response at the Canada site comparing the Sanibel and in-house cottonwool electrodes. 95% Confidence bars are overlaid. Sample number is on the horizontal axis (44.1 samples/ms). Vertical axis is the normalised magnitude.

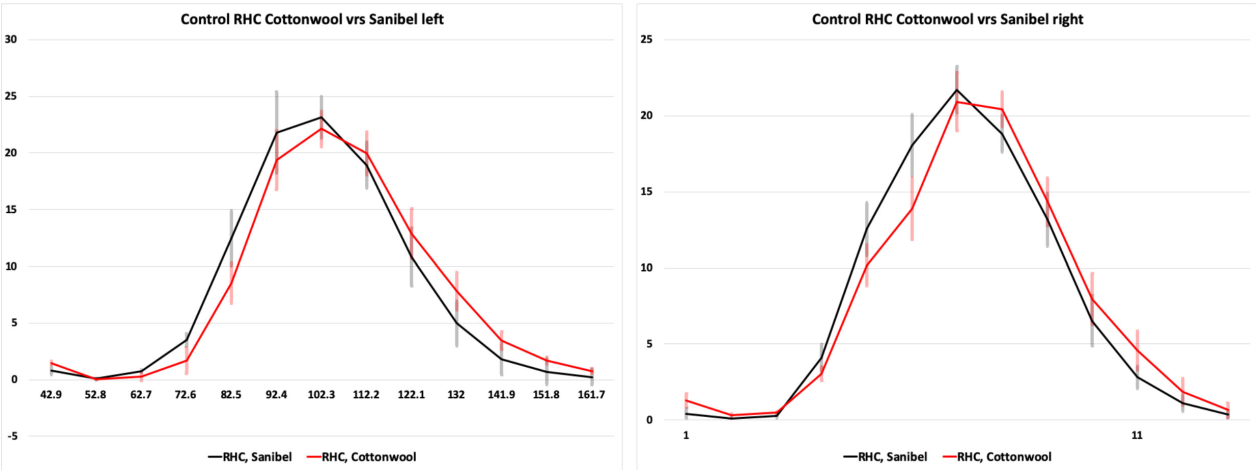


Figure 6: IH33 response at the Canada site comparing Cottonwool (N = 17) and Biopac (N = 17) age and gender matched responses. 95% Confidence bars are overlaid. The horizontal axis is in ms. Vertical axis is population percentage.

(Canada) site on the left or right side are right shifted. A yellow ellipse marks a bin of significant difference. Accordingly, comparisons should be made only between populations with matched electrode types.

2) Gender matching: Figure 7 shows the female response (N = 24 female left and right) is marginally but not significantly narrower than that of the comparable male (N = 24 male left and right) response at the same (Australian) site. The populations were age and electrode type matched. Ideally, comparisons should be made only between male or female types but as the difference is not significant gender is considered not a major factor for AP area herein. However, gender matching for combined male/female populations should be applied when possible.

Figure 8 shows the age matched male and female responses at the same Australian site and with same electrode type on the left side are very similar (N = 24). The right response shows the female population with a slightly broader response, a small right shift and lower peak value however, these differences are not significant. Ideally, comparisons should be made only between gender matched types but as the difference is not significant gender is considered not a major factor for the IH_{33} plots herein. However, gender matching should be applied when possible.

3) Age matching: Figure 9 shows the older (greater than 51 years old) and younger responses at the same site with the same electrode type on the left and right side are very similar (N = 23, 11F). The age division was determined

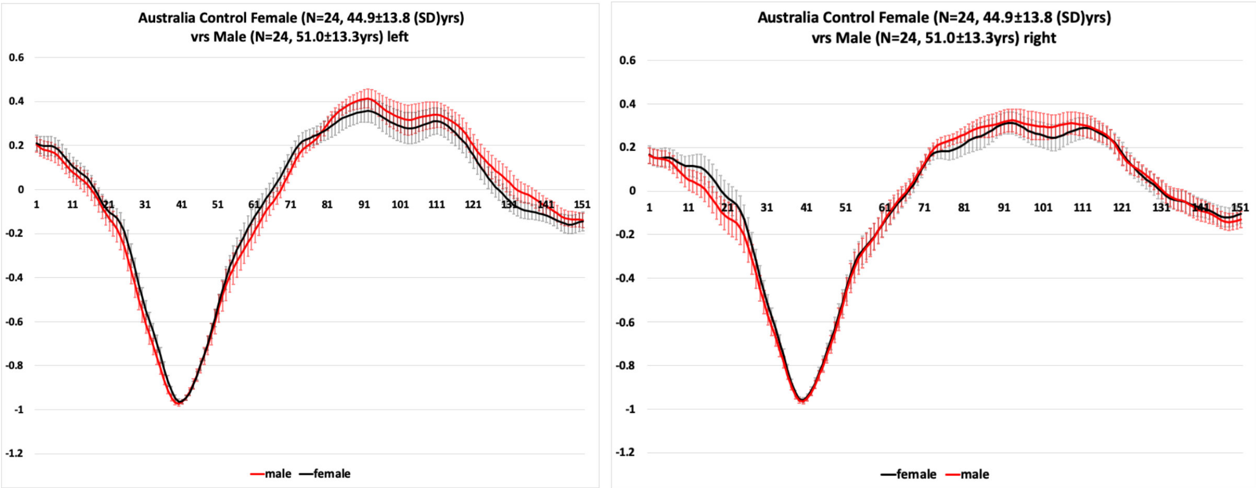


Figure 7: AP plus PPP shape response at the Australia site comparing male and female responses. 95% Confidence bars are overlaid. Sample number is on the horizontal axis (44.1 samples/ms). Vertical axis is normalised magnitude.

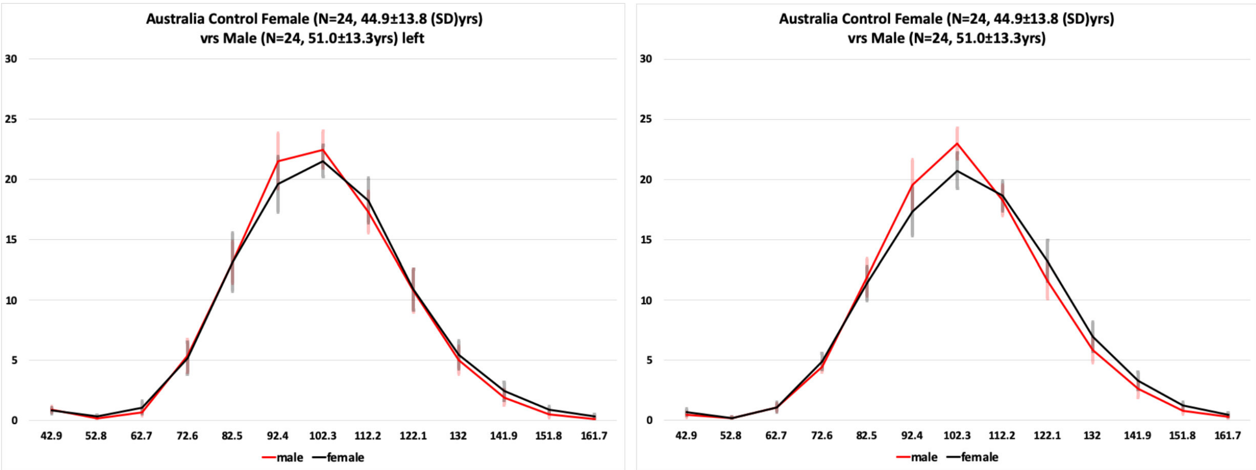


Figure 8: IH33 response at the Australia site comparing male (N = 24) and female (N = 24) responses. 95% Confidence bars are overlaid. The horizontal axis is in ms. Vertical axis is population percentage.

by the available population and approximate halving of that population. The right response shows the older population with a slightly narrower (smaller AP and PPP area) response but this is not significant. Ideally, comparisons should be made only between age matched types but as the difference is not significant age is considered not a major factor for AP area herein. However, age matching should be applied when possible.

Figure 10 shows the older and younger responses at the same site and with same electrode type on the right side are very similar (N = 24, 11F). The left response shows the older population with a slightly broader response and lower peak value however, these differences are not significant. The older responses are very marginally right shifted. Ideally, comparisons should be made only between age matched types but as the difference is not significant age is considered not a major factor for the IH33 plots herein. However, age matching should be applied when possible.

*4) Site comparison (*50 *versus* 60 *Hz and* 940/1024 *versus* 940*Hz hydraulic artefact removal):* Figure 11 shows the Australian (N = 36, 22F) and Canadian (N = 20, 13F) site responses for a control population that is age, electrode type (Sanibel) and gender matched are, on the left side and right side similar. The right response shows the Australian site with a marginally narrower (smaller AP and PPP area) response but this is not significant. Ideally, comparisons should be made only between populations from the same site but when comparing studies this is not possible. Site differences for age and gender matched populations recorded with the same electrode type appear minor.

Figure 12 shows the Australian and Canadian age and gender matched responses with same electrode type (Sani-

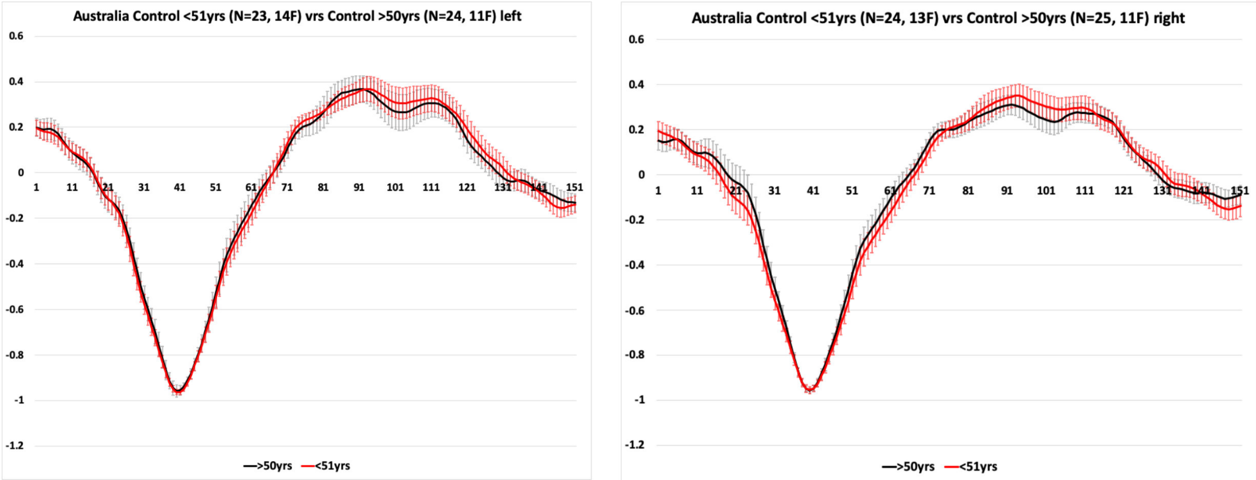


Figure 9: AP plus PPP shape response at the Australia site comparing older (greater than 50yrs) and younger (less than 51yrs) responses. 95% Confidence bars are overlaid. Sample number is on the horizontal axis (44.1 samples/ms). Vertical axis is normalised magnitude.

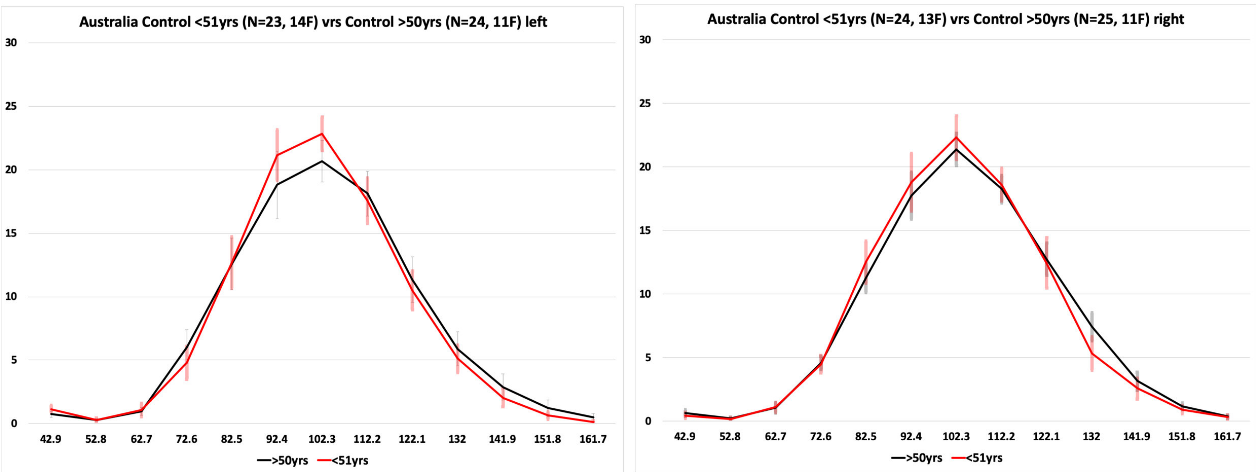


Figure 10: IH33 response at the Australia site comparing older (greater than 50yrs) and younger (less than 51yrs) responses. 95% Confidence bars are overlaid. The horizontal axis is in ms. Vertical axis is population percentage.

bel) on the right side are very similar. The left response shows the Australian population with a slightly broader response and lower peak value however, these differences are not significant. There is no rightward shift of populations. Ideally, comparisons should be made only between age and gender matched types from the same site but as the difference is not significant site is considered not a major factor for the IH_{33} plots herein. However, within site comparisons should be applied when possible.

IV. Discussion

Figure 3 shows significant features can be derived; thus, it is possible to distinguish AD population from best matched healthy controls. On the other hand, Figure 5 and Figure 6 show the comparative effects on healthy controls of using different electrode types at the same site (Canada) can be significant; hence, the need for any subsequent comparisons only to be between the recordings with the same electrode types. The Sanibel™ FP response (Figure 5) has more low frequency content relative to that of the in-house cotton-wool electrode on both right and left sides. The Sanibel™ IH_{33} response (Figure 6) is left shifted relative to that of the in-house cotton wool electrode on both right (significant different bin marked with a yellow ellipse) and left sides.

It is interesting to ask why the electrode type should affect IH detection. Perhaps there is more noise detected with the Sanibel electrode configuration; that is likely as the SNR was improved significantly with matched cotton wool electrode configuration [21]. Figures 7-12 show the effect

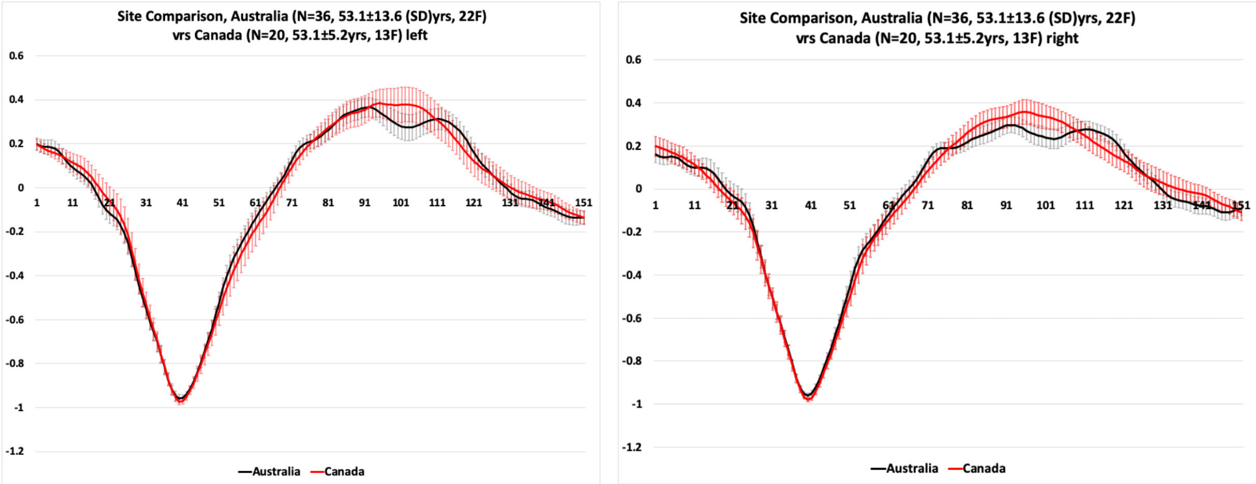


Figure 11: Left and Rightside AP plus PPP shape response at the Australia site compared with the Canada site. 95% Confidence bars are overlaid. Sample number on the horizontal axis (44.1 samples/ms). Vertical axis is normalised magnitude.

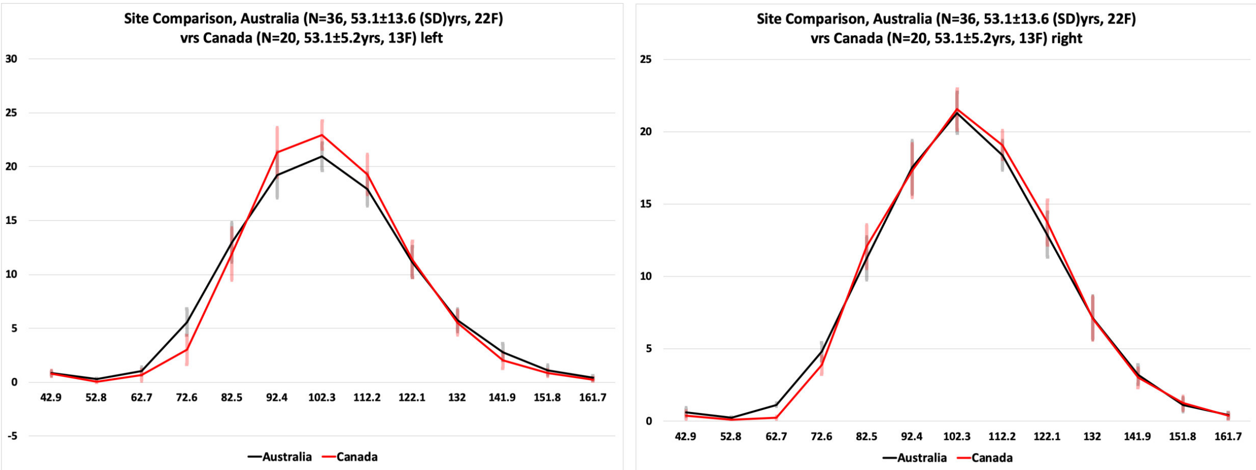


Figure 12: IH33 response comparing the age and gender matched Australian ($N = 36$) and Canadian ($N = 20$) responses. 95% Confidence bars are overlaid. The horizontal axis is in ms. Vertical axis is population percentage.

of gender, age and site (powerline frequency/hydraulic artefact) on FP shape and IH_{33} plot response were not significant; however, gender, age and site matching for all populations should best applied where possible. However, there is a need for comparisons only to be made for comparable electrode types and where possible only within sites.

V. Conclusion

It is possible to discern AD from best matched healthy controls using EVestG given the comparison recordings are matched in terms of electrode type. Age and gender did not have a significant factor on the main EVestG feature types. Powerline and environmental artefacts could be accounted for adequately; implying they did not significantly affect the features measured. For the two EVestG features tested only electrode type had a significant effect on recordings. Therefore, only populations with the same electrode type can be compared.

References

[1] B. J. Lithgow, Z. Moussavi, C. Gurvich, J. Kulkarni, J. J. Maller, and P. B. Fitzgerald, "Bipolar disorder in the balance," *European archives of psychiatry and clinical neuroscience*, pp. 1–15, 2018.

[2] B. J. Lithgow, A. L. Garrett, Z. M. Moussavi, C. Gurvich, J. Kulkarni, J. J. Maller, and P. B. Fitzgerald, "Major depression and electrovestibulography," *The world journal of biological psychiatry*, vol. 16, no. 5, pp. 334–350, 2015.

[3] B. Lithgow, "A methodology for detecting field potentials from the external ear canal: NEER and EVestG," *Annals of biomedical engineering*, vol. 40, no. 8, pp. 1835–1850, 2012.

[4] B. Blakley, A. Suleiman, G. Rutherford, Z. Moussavi, and B. Lithgow, "EVestG recordings are vestibuloacoustic

signals," *Journal of Medical and Biological Engineering*, vol. 39, no. 2, pp. 213–217, 2019.

[5] B. J. Lithgow and et al., "Electrovestibulography identifies Alzheimer's Disease from Mixed Dementia and from Healthy Controls," *Journal of Alzheimer's Disease*, 2019, submitted.

[6] B. J. Lithgow and M. Shoushtarian, "Parkinson's disease: disturbed vestibular function and levodopa," *Journal of the neurological sciences*, vol. 353, no. 1-2, pp. 49–58, 2015.

[7] A. Suleiman, B. Lithgow, Z. Dastghieb, B. Mansouri, and Z. Moussavi, "Quantitative measurement of post-concussion syndrome using Electrovestibulography," *Scientific reports*, vol. 7, no. 1, p. 16371, 2017.

[8] A. Suleiman, B. Lithgow, B. Mansouri, and Z. Moussavi, "Investigating the validity and reliability of Electrovestibulography (EVestG) for detecting post-concussion syndrome (PCS) with and without comorbid depression," *Scientific reports*, vol. 8, no. 1, p. 14495, 2018.

[9] Z. Moussavi, A. Suleiman, G. Rutherford, O. R. Pouya, Z. Dastgheib, W. Zhang, J. Salter, X. Wang, B. Mansouri, and B. Lithgow, "A pilot randomised double-blind study of the tolerability and efficacy of repetitive transcranial magnetic stimulation on persistent post-concussion syndrome," *Scientific reports*, vol. 9, no. 1, p. 5498, 2019.

[10] Z. A. Dastgheib, B. Lithgow, B. Blakely, and Z. Moussavi, "Application of vestibular spontaneous response as a diagnostic aid for Meniere's disease," *Annals of biomedical engineering*, vol. 44, no. 5, pp. 1672–1684, 2016.

[11] Z. A. Dastgheib, B. Lithgow, B. Blakley, and Z. Moussavi, "A new diagnostic vestibular evoked response," *Journal of Otolaryngology-Head and Neck Surgery*, vol. 44, no. 1, p. 14, Apr. 2015.

[12] B. Blakley, Z. A. Dastgheib, B. Lithgow, and Z. Moussavi, "Preliminary report: neural firing patterns specific for Meniere's disease," *Journal of Otolaryngology-Head & Neck Surgery*, vol. 43, no. 1, p. 52, 2014.

[13] B. J. Lithgow, Z. Moussavi, and P. B. Fitzgerald, "Quantitative separation of the depressive phase of bipolar disorder and major depressive disorder using electrovestibulography," *The World Journal of Biological Psychiatry*, pp. 1–14, 2019.

[14] A. Nordström and P. Nordström, "Traumatic brain injury and the risk of dementia diagnosis: A nationwide cohort study," *PLoS medicine*, vol. 15, no. 1, p. e1002496, 2018.

[15] C. Liu, F. C. K. Tan, Z.-C. Xiao, and G. S. Dawe, "Amyloid precursor protein enhances Nav1. 6 sodium channel cell surface expression," *Journal of Biological Chemistry*, vol. 290, no. 19, pp. 12 048–12 057, 2015.

[16] R. A. Rissman, A. L. De Blas, and D. M. Armstrong, "GABAA receptors in aging and Alzheimer's disease," *Journal of neurochemistry*, vol. 103, no. 4, pp. 1285–1292, 2007.

[17] T. J. Marczynski, "GABAergic deafferentation hypothesis of brain aging and Alzheimer's disease revisited," *Brain research bulletin*, vol. 45, no. 4, pp. 341–379, 1998.

[18] P. Theofilas, A. J. Ehrenberg, S. Dunlop, A. T. D. L. Alho, A. Nguy, R. E. P. Leite, R. D. Rodriguez, M. B. Mejia, C. K. Suemoto, R. E. D. L. Ferretti-Rebustini *et al.*, "Locus coeruleus volume and cell population changes during Alzheimer's disease progression: A stereological study in human postmortem brains with potential implication for early-stage biomarker discovery," *Alzheimer's & Dementia*, vol. 13, no. 3, pp. 236–246, 2017.

[19] C. Cortes, F. Galindo, S. Galicia, J. Cebada, and A. Flores, "Excitatory actions of GABA in developing chick vestibular afferents: effects on resting electrical activity," *Synapse*, vol. 67, no. 7, pp. 374–381, 2013.

[20] V. Marlinsky, "The effect of somatosensory stimulation on second-order and efferent vestibular neurons in the decerebrate decerebellate guinea-pig," *Neuroscience*, vol. 69, no. 2, pp. 661–669, 1995.

[21] C. Kumaragamage, B. Lithgow, and Z. Moussavi, "A new low-noise signal acquisition protocol and electrode placement for electrocochleography (ECOG) recordings," *Medical & biological engineering & computing*, vol. 53, no. 6, pp. 499–509, 2015.

Brian J. Lithgow is currently the Leader of the Diagnostic and Neurosignal Processing Research Groups at the Alfred Hospital (Melbourne, Australia) and Riverview Health Centre (Winnipeg, Canada). He is also a Senior Research Fellow at the Alfred Hospital and Research Affiliate at Riverview Health Center. His appointments include Adjunct Professor at the University of Manitoba, Canada and Adjunct Assoc. Professor at Monash Alfred Psychiatry Research Center. He founded the Monash University Centre for Biomedical Engineering (MUCBE) and was the Director of Teaching for MUCBE from 2001-2010.

His recent research is diagnostics development for neurological and neurodegenerative disorders, including Dementia, Parkinson's Disease, Vertiginous Disorders, Post-concussion syndrome, Bipolar disorder and Major Depressive disorder (4 patents). The successful quantitative separation of unipolar and bipolar depression is a world first. Current research aims at separating Dementia types and predicting treatment efficacy. An animal study is looking at modelling of vestibular electro-neurophysiology. He has published more than 130 refereed publications, 6 books, 2 book chapters, 5 patents.

Zahra Moussavi Zahra Moussavi is a professor, a Canada Research Chair, and the founder and director of Biomedical Engineering Graduate Program at University of Manitoba. Her current research focuses are on medical devices instrumentation and signal analysis for sleep apnea management and Alzheimer's diagnosis and treatment using virtual reality, rTMS and EVestG technologies. She is the recipient of several awards including the "Canada's Most Powerful Women (Top 100)" and "Manitoba Distinguished Women" in 2014 and IEEE EMBS Distinguished Lecturer, 2014 and 2019. She has published more than 259 peer-reviewed papers in journals and conferences, and has given 94 invited talks/seminars including 2 Tedx Talks and 9 keynote speaker seminars at national and international conferences. Aside from academic work, on her spare time, she writes science articles for public; also has developed and offered memory fitness programs for aging population.

An Ensemble Co-Evolutionary based Algorithm for Classification Problems

Vu Van Truong[1], Bui Thu Lam[1], Nguyen Thanh Trung[2]
[1] Le Quy Don Technical University, Hanoi, Vietnam
[2] Liverpool Logistics Offshore and Marine Research Institute, Liverpool, United Kingdom
Correspondence: Vu Van Truong, truongvv@mta.edu.vn
Communication: received 14 April 2019, revised 22 August 2019, accepted 28 August 2019
Online early access: 28 August 2019, Digital Object Identifier: 10.32913/mic-ict-research.v2019.n1.852
The Editor coordinating the review of this article and deciding to accept it was Prof. Le Hoang Son

Abstract: In this paper, the authors propose a co-evolutionary algorithm using an ensemble learning approach (E-SOCA) to simultaneously solve both feature subset selection and optimal classifier design. Unlike previous studies where each population retains only one best individual (Elite) after co-evolution, in this study, an elite community will be stored and calculated together through an ensemble learning algorithm to produce the finally classified result. Experimental results on the University of California, Irvine (UCI) problems with a variety of input features ranging from small to large sizes show that the proposed algorithm results in more accuracy and stability than traditional algorithms.

Keywords: *Ensemble learning, co-evolution algorithm, feature selection, genetic algorithm.*

I. Introduction

For the classification domain there are basically two main problems to be solved. To be specific, one problem is to select the classification algorithm and the other is to select the input feature sets. Normally for each object to be classified, there will be many features. However, it is necessary to identify important features having the greatest impact on the classification results, and at the same time, with this selected feature set, what is the best configuration for the classification algorithm. Typically, the more features there are, the longer the training and classification process will take. In addition, it also makes the program take up more memory and hard disk space. Therefore, for machine learning methods, it is necessary to select a smaller subset that still guarantees the acceptable accuracy of the classification process. That is called a feature selection (or other names like variable selection, feature reduction, attribute selection or variable subset selection). In contrast to other dimensionality reduction techniques like PCA (Principal Component Analysis) or compression, feature selection techniques only select a subset of them rather than alter the original representation of the variables.

So far there have been many methods used to solve the feature selection problem (more details are presented in Section II.4). In general, these methods can be solved by using only a single algorithm or combining multiple algorithms. The co-evolutionary algorithm is one of the second methods. These algorithms use more than two populations to solve the problem. Depending on how these populations interact with each other, they can be divided into two main categories: co-operative and competitive. In co-operative co-evolution, a problem is broken down into many sub-problems and each population is designed to solve a sub-problem. The final solution will be the combination of those populations' outputs. Meanwhile, in competitive co-evolution, during evolution, individuals in this population will have to fight against a group of individuals in the remaining population and vice versa, this leads to an arms race between these two populations.

The number of competitive co-evolution studies is mainly focused on games and robots [1, 2]. Studies related to classification problems are still relatively small and have started to be more interested in recent years [3–5]. In [3], the authors have proposed a competitive co-evolutionary model using a double-elimination tournament (DET) mechanism. This research employs two populations which are the population of the multilayer perceptron artificial neural network (MLPANN) and the population of basis radial artificial neural network (RBFANN). During evolutionary process, individuals between these two populations interact with each other through the DET mechanism. The proposed method achieved high accuracy rates in the comparison with other versions of co-evolutionary ANNs, co-evolutionary RBFNNs and other classification algorithms on 15 datasets. In [4], a Predator and Prey model was used to train multi-layer perceptron (MLP) classifiers. The Predators are MLPs and the Preys are training samples. In the evolutionary process, the Predator tries to classify as correct as possible, while the Prey tries to offer the most difficult training data sets to make the Predator misclassify as much as possible. Thanks to this process, the algorithm will find a strong classifier that has the ability to classify difficult data sets. It is noting that the idea of the Predator-Prey model was also used in the Generative Adversarial Networks (GANs) model [5], which is considered a revolution in deep learning. In this model, the Prey acts as a neural network Generator, it always tries to generate fake data that is as similar as possible with real data. Meanwhile, the Predator acts as a Discriminator network, always trying to identify the fake data with the real data.

In this study, we focus primarily on the co-operative co-evolution. Up to now, there have been many publications related to the classification problem. In [6], the authors presented a hybrid approach based on a co-operative co-evolutionary algorithm with dual populations for designing the RBFNN (Radial Basis Function Neural Network) models with feature selection problem (named DC-RBFNN). In this research, each feature is encoded as a binary string and an RBFNN structure is encoded in a real-encoded matrix-form. The authors used five objectives to evaluate the fitness of individuals. The proposed method was tested on 26 datasets from the UCI Repository. Experiment results showed that the proposed algorithm was able to obtain better results on complicated classification tasks than other training algorithms did. In [7], these authors proposed an ensemble coevolutionary algorithm (named Co-NNE). Unlike the DC-RBFNN, in Co-NNE, the authors used multiple subpopulations instead of dual populations, and utilized an Elite Pool to store a community of Elites. Each Elite is the best individual of a subpopulation. Experimental results showed that the Co-NNE got better results on complicated classification tasks in comparison with other ensemble algorithms. Following the success of this research direction, in [8] the authors continued to improve the DC-RBFNN with a new algorithm called subspace-based RBFNN model (SBRBFNN). This algorithm has two versions which are hard CEA-SBRBFNN (CEA-SBRBFNNH - using binary encoding) and soft CEA-SBRBFNN (CEA-SBRBFNNS – using real encoding). Experimental results on 22 UCI datasets show that the proposed method outperforms the DC-RBFNN and other feature selection algorithms. In [9], the authors proposed a multi-population co-evolutionary algorithm using genetic programming (MOGP) for software defect prediction (SDP) problem. In this study, the authors utilized three ensemble selection strategies to select a subset of solutions at the end of the coevolution process. Their results show that the MOGP was a promising solution for solving the SDP with unbalanced data. In [10] and [11], the authors used a co-operative multi-objective evolutionary approach with the support vector machines (SVMs) for solving multiclass classification problems. The idea of the algorithm is to break down the multiclass problem into multiple binary classification problems and utilize a subpopulation to solve each binary problem. From results obtained from 25 datasets, the proposed method outperformed some SVMs Multiclass Extensions.

In [12], the authors proposed a single objective coevolutionary algorithm (named COCEA) for the classification problems. In COCEA, the authors use two populations with different individual coding, one population using binary encoding (we call it as binary or feature population) and the other using real number encoding (real number population). The authors used the Differential evolution (DE) algorithm to evolve the real population and the GA algorithm to evolve the binary one. The results were tested on both oil spill dataset and three other UCI datasets. A notable point in the COCEA algorithm is that both populations share a fitness function which is the MSE error. This means both populations try to evolve to achieve the best possible classification value. The results show that this idea is appropriate because the classification result of the COCEA algorithm obtains the best one on all test data sets. However, the authors did not focus on further analysis of the aspect of the final selected features.

Through the literature review, we found that studies of co-evolution still have great potential and are still being studied extensively to solve problems related to the classification problem. In particular, after studying carefully the dual-population co-evolutionary models in previous studies, we have found that these studies can still be further expanded, and this is the motivation for us to continue studying in this direction. In this paper, we propose a wrapper method using an ensemble co-operative co-evolution approach with a dual population to optimize the ANN classifier and deal with the feature selection problem.

The novelty and main contributions of the article are as follows. First, in terms of elite retention mechanism, we utilize an Elite Pool to store the best-known individuals from two populations during the coevolutionary process (each the best individual is called an elite). However, unlike previous studies when only two elites were stored from beginning to end of evolution. In this study, all intermediate elites found during co-evolution are stored for further processing. This may increase the chances of finding a good classifier.

Second, in terms of ensemble learning step, in previous studies, the combination of two individuals in the Elite pool was often considered as the final solution. Some studies using a multi-objective co-evolution solution, after the end of co-evolution, a community of individuals on the first Pareto front was chosen as an ensemble of classifiers. This selection method may encounter problems with the diversity of classifiers. The voting technique was commonly used in this case to make the final decision. Our approach is completely different, the mechanism of storing elites found through each generation may help us maintain the diversity. A boosting algorithm is then used to combine these classifiers to create a more powerful classifier.

This paper extends preliminary results in the conference paper [12]. In general, compared to the algorithm in the conference paper (COCEA), the proposed algorithm has three main differences. First, the COCEA algorithm works only with binary classification problems. In this study, the proposed method works with multiclass classification problems.

Second, in the COCEA, the real population was evolved by using the DE algorithm, meanwhile, the binary population used the GA algorithm to evolve. It is noted that the DE algorithm has proven to be able to converge faster than the conventional GA algorithm. This may lead to an imbalance between the two populations. In this proposed algorithm we use the conventional GA instead of the DE algorithms with some changes in evolutionary steps corresponding to two different populations.

Third, in the COCEA algorithm, after the end of the co-evolution process, only the best individual (elite) of each population is retained for calculation. In the proposed algorithm, a community of elite individuals in the evolutionary process is archived. These elites play as weak classifiers then an ensemble learning algorithm (the AdaBoost) is used to create a strong classifier from these weak learners.

The rest of the paper is organized as follows: Section II describes the fundamentals related to the proposed method. In Section III, the proposed algorithm is presented in detail. Section IV reports the experiments of the proposed method on six UCI datasets together with other algorithms. Finally, Section V concludes with remarks for future research.

II. Background

1. The Artificial Neural Network (ANN)

For many years, the ANN has been considered a powerful tool to solve machine learning problems, especially classification problems. When working with the ANN, there are two main issues need to be addressed: one is to select the network architecture and the other is to select the initial initialization value for the Weight set. The ANN randomly generates initially values for the weight set, then a learning algorithm is used to train the ANN; the most common training algorithm is Backpropagation (BP). Because the BP algorithm is based on the gradient descending mechanism, it can quickly converge to the global extreme. However, it also can be easily stuck at local extremes in the weight-space. This depends greatly on initiating the ANN's weight. There have been many studies addressing this problem [13, 14]. In this study, we use one of two populations of co-evolution approach to solve this problem.

2. Genetic Algorithm (GA)

Genetic algorithm [15] is an algorithm based on Darwin's theory to simulate the process of natural evolution. The GA applies the genetic principles: natural selection, mutation, and cross-exchange. In the GA, a population is a collection of chromosomes, each chromosome represents a solution of the problem (in this study each chromosome will be the weight matrix of the ANN or a binary string present for a feature set). The GA is often used widely for feature selection problems. There are many types of encoding strategies for an individual in the GA. The most common strategy is to treat each individual as a binary string in which value "1" means the corresponding feature is selected and vice versa with the value "0". Another encoding type is to use a real number sequence in the range $[0, 1]$ to represent the probability of being selected for a feature. By comparing a given threshold it is possible to convert this real number encoding to conventional binary encoding. Another strategy is a bio-encoding, each chromosome is encoded as a pair of strings. The first one is a binary string to represent the selection of features and the second one is encoded as real numbers to show the weights of features. In this study, we use the binary encoding to present the first population and the real number to encode the second one.

3. Ensemble Learning

In machine learning, ensemble learning is a method using multiple learning algorithms to obtain a better performance than any single learning algorithm [16]. In ensemble learning, Bagging and Boosting are two common techniques. By far, the most common algorithm of boosting is AdaBoost. The AdaBoost (standing for "Adaptive Boosting") was first introduced by Freund and Schapire [17]. After that, there were several versions of this algorithm such as AdaBoost.M1 and AdaBoost.M2 [18]. The AdaBoost combines multiple "weak classifiers" into a single "strong classifier". A weak classifier is simply a classifier like an ANN or a KMeans or any classification algorithm. The AdaBoost determines weights of each classifier. These weights are then used to combine weak learners to form a stronger classifier. The larger weight value, the higher the level of importance of that weak learner. In this study, we will use the AdaBoost algorithm to combine ANN classifiers to create a more powerful classifier.

4. Feature Selection Algorithms

In general, depending on the combination of the feature selection search and the classification model, the Feature subset selection techniques can be organized into filter approach and wrapper approach [19]. In the Filter approach [20], a feature relevance score is first calculated and features having low-scoring are removed. After that, the selected features becoming the input of the classification algorithm, they are done independently of the classification algorithm. Two steps are done independently. Although they are computationally simple, they are assessed by another criterion rather than classifiers, which leads to worse classification performance. Conversely, in the wrapper approach, the feature subset selection algorithm is wrapped with the classification function. Various subsets of features are generated, and the predictor plays as a black box to evaluate these subsets. In general, the wrapper approach outperforms the filter approach. There are several search algorithms used to find the optimal subsets. In [21–23], the authors used exhausting strategies to solve a feature selection problem. Normally, these algorithms only work effectively in relatively small dimensional data problems, and vice versa, because the computational space is too large these algorithms are not effective. In recent years, evolutionary techniques (EC) have been used extensively for solving this problem. Readers can be found more details of studies using the EC techniques for feature selection problems in a review paper [24].

III. Proposed Method

The general diagram of the E-SOCA is given in Algorithm 1 and Figure 1. In this study, the authors use two populations. The first population (called the feature population or binary population) and the second one (called the populations of ANNs or real population). At each generation, these two populations will interact with each other. This population will use the best individual (i.e. Elite) of the other population. $Elite1$ will determine the number of inputs of ANNs in the population 2 and vice versa, $Elite2$ will be used to calculate the adaptive value of each individual in population 1. Undergoing evolution, the overall solution will be a combination of the Elites from two populations. As can be seen that the biggest difference between the proposed solution and the conventional feature selection methods is that the classifiers (i.e. ANNs) change constantly during the evolution process instead of fixed configuration.

In traditional wrapper methods, a classifier is fixed to evaluate the performance of each selected feature set. This can reduce the performance of the classifier, since each ANN is usually appropriate with a set of inputs. It cannot fit all types of inputs. Therefore, the use of the co-evolutionary solution here will hopefully solve this problem. Another interesting point is that unlike previous co-evolutionary studies, only a pair of Elites in each population is selected at the last evolutionary step. In this study, a list of Elite pairs is stored so that we can continue to use an ensemble learning algorithm to combine them together.

In Figure 1, the E-SOCA consists of four main steps: population initialization, co-evolution, ensemble learning, and fine-tune. A more detailed explanation of the E-SOCA will be presented below.

1. Encoding

Two populations use two different coding strategies. In the feature population, suppose that we have m individuals $(F_1, F_2, \dots, F_m)$, each individual F_i is encoded as a binary string. In particular, the value 1 corresponds to the feature being selected and vice versa (Figure 2).

In the second population, suppose that this population has n individuals $(\mathrm{ANN}_1, \mathrm{ANN}_2, \dots, \mathrm{ANN}_n)$, each individual ANN_i is a weight and bias of an ANN. Therefore, we use a real-encoding to represent an ANN (Figure 3).

2. Initialization

In general, at the initialization step, individuals will be generated randomly. However, if individuals in the real population receive any random value, random in the binary population initialization must ensure the following two constraints: (1) this binary string must contain at least one bit of "1" to ensure there must be at least one selected feature. (2) There is only one string containing all bits of "1" (line 3) to ensure that all features must be considered.

At first, an $Elite1$ with all the bits assigned to "1" will be used as the input of the individuals in the population 2. Thus, it is possible to assess the adaptability of each individual in the second population. After that we make a ranking based on the fitness value and choose the best individual as an elite (i.e. $Elite2$) and put it into the Elite Pool. The $Elite2$ is then used to evaluate fitness values of the individuals in population 1, after ranking, the best one (named $Elite1$) is selected and put into the Elite Pool. At this time, $\langle Elite1, Elite2 \rangle$ will be stored in the Elite Pool, at the same time $Elite2$ is also included in a list (named $BestList$) to serve the ensemble learning step later.

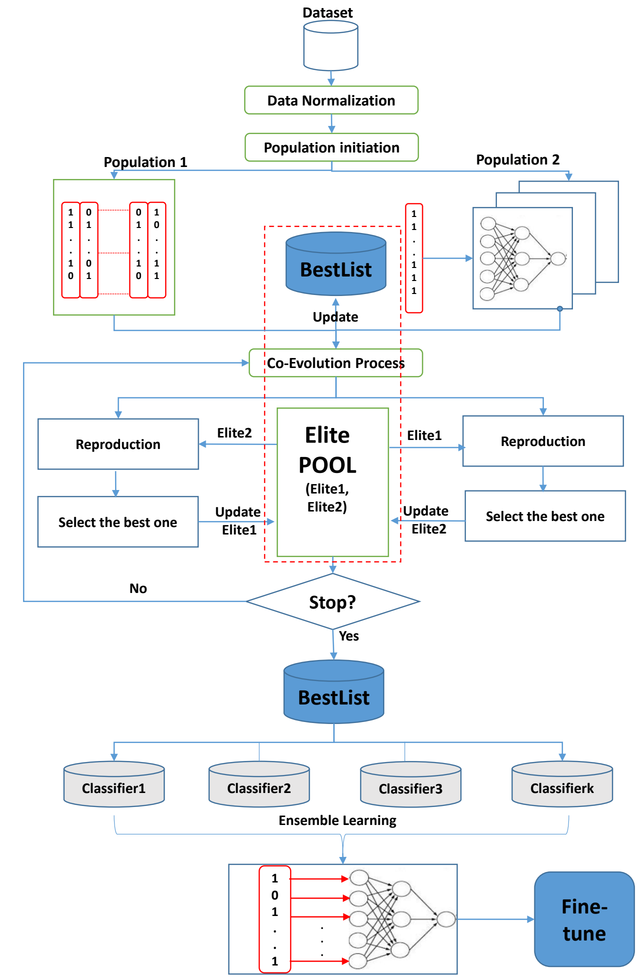


Figure 1. Diagram of the proposed method.

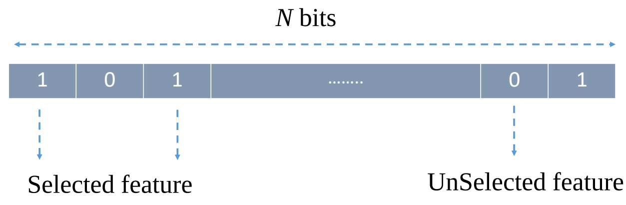


Figure 2. Encoding a feature individual.

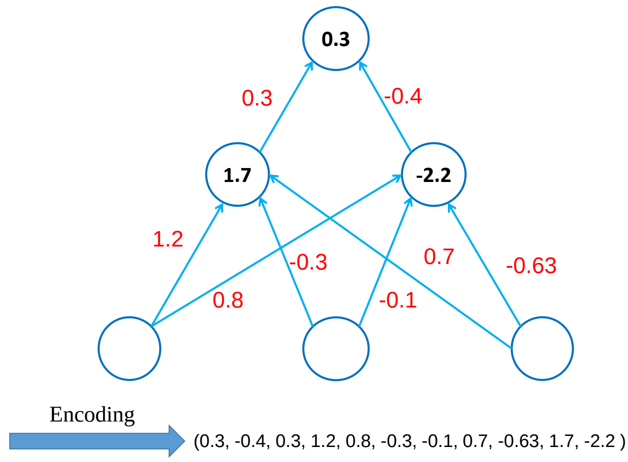


Figure 3. Encoding an ANN individual.

3. Co-Evolution Process

After the population initialization is the process of co-evolution. This is an iterative process; the step basically is quite similar to the initialization step in which each sub-population will try to evolve and find the elites to update to the pool. The elite in this population will be used for the other population in the next iteration.

However, there are two differences in this step compared to the initialization step. Suppose the Elite Pool consists of two elements $\langle Elite1, Elite2\rangle$. The first difference is the $Elite1$ update mechanism. If there is anyone (denoted by $Elite1_t$ at the loop t^{th}) dominating the corresponding elite in the pool (i.e. $Elite1$), the $Elite1$ will be substituted by $Elite1_t$ and the fitness of the new $Elite1$ is updated as well. However, we do not compare $Elite1_t$ with $Elite1$, instead we replace $Elite1$ by $Elite1_t$ immediately. This is also a different point from the COCEA algorithm. The reason for this change is that, through experimenting with the COCEA algorithm, we realize that this update mechanism is the main reason that the number of updated Elite Pool times is quite small. This leads to the result of co-evolution being not significant. Therefore, to increase the probability of updating the Elite Pool, we make a change in this update mechanism for $Elite1$. Meanwhile, with the $Elite2$, the update mechanism still happens normally, which means we choose the best individual from the population 2 (denoted by $Elite2_t$ at the loop t^{th}) compared to the $Elite2$ in the Elite Pool, if the MSE value of the $Elite2_t$ is smaller than the $Elite2$ then we replace the $Elite2$ by the $Elite2_t$. The second difference here is that, whenever the $Elite2$ is replaced by the $Elite2_t$, instead of taking new better values (i.e. $Elite2_t$), we store the old values ($Elite2$) and put them into $BestList$. These elements in $BestList$ continue to be combined through an ensemble learning algorithm.

The main difference between the two populations is the evolutionary operation (the Reproduction step in the Figure 1). Because each population uses different encoding types (binary and real encoding), therefore evolutionary operation applied to them are varied. Specifically, in population 1 we use *Two points crossover* and *Bit flip mutation* operations. Meanwhile, the *SBX Crossover* and *Polynomial Mutation* are applied to the population 2.

Algorithm 1: Ensemble Single Objective Coevolutionary Algorithm (E-SOCA).

1 **Input**: DataSet.
2 **Output**: BestClassifier.
3 $P1 \leftarrow$ InitializePopulation1(N);
4 $P2 \leftarrow$ InitializePopulation2(N);
5 $Elite1 \leftarrow [11\ldots1]$;
6 $Pool \leftarrow \Theta$; $BestList \leftarrow \Theta$;
7 Pool.Add(Elite1);
8 **CalculateFitness**(P2, Elite1, DataSet);
9 Eltite2 $\leftarrow Sort(P2)$;
10 Pool.Add(Elite2);
11 BestList.Add(Elite2);
12 **while** *stop condition false* **do**
13   **Reproduction**(P1, Eltite2);
14   Eltite1 $\leftarrow Sort(P1)$;
15   **Pool.Update**(Elite1);
16   **Reproduction**(P2, Eltite1);
17   Eltite2 $\leftarrow Sort(P2)$;
18   **Pool.Update**(Elite2);
19   **if** *update = true* **then**
20   BestList.Add(Elite2);
21   **end**
22 **end**
23 WeakLearners = BestList;
24 BestClassifier = **EnsembleLearning**(WeakLearners);
25 **FineTune**(BestClassifier);
26 **Return** BestClassifier;

4. Ensemble Learning

After the coevolutionary process, a list of Elites is stored in *BestList*. These Elites play as *weakLearners* in the AdaBoost algorithm. Ideas and ways to implement this algorithm are shown in [13]. The biggest difference here is that the AdaBoost version is used for the classification problem instead of the forecasting problem. In the AdaBoost algorithm, weak learners may be different models or the same models with different configurations. In this paper, we use an ensemble of ANNs (i.e. *BestList*) with the same topology but different weights. Besides, each weak learner may be trained on a random subset of the training set and the subsets can overlap. However, in this study, instead of dividing into subsets we use the entire training data set for all weak learners. By using the AdaBoost, each Elite will have different influence coefficients, the better the result of the classification, the greater the effect factor and vice versa. These coefficients will be used to combine Elite sets to create a stronger classifier (named *BestClassifier*).

TABLE I
THE PROPERTIES OF THE BENCHMARK DATASETS

Dataset	Train	Val	Test	Class	Features
Sonar	104	-	104	2	60
Satimage	4435	-	2000	7	36
Segment	1162	574	574	7	19
Ionosphere	175	88	88	2	34
Pima	384	192	192	2	8
Heart	134	68	68	2	13

5. Fine-Tune

A population-based evolutionary mechanism helps the E-SOCA approach find out the regions that contain the global maximum points, but it is very hard to find exactly these extreme points. From this characteristic, we use the GA algorithm to optimize the ANNs first (i.e. using the GA to take the ANNs to escape from the local extremes and approach the global ones), then use the Back-Propagation algorithm to put them closer to the global extreme points. This is the reason that we continue using this step for *BestClassifier* (line 24 of Algorithm 1) after going through the ensemble learning step.

6. Fitness function

In this paper, we choose the Mean Squared Error (MSE) which is calculated as the fitness function as follows:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(x_{\text{real}} - x_{\text{output}})^2, \quad (1)$$

where n is the number of data value, x_{real} is a real value, and x_{output} is an output value.

IV. EXPERIMENTS

1. Data Description

In order to evaluate the performance of the proposed method, the authors conduct experiments on 6 different UCI datasets (Sonar, Satimage, Segment, Ionosphere, Pima, and Heart). Most datasets are divided into three subsets: 50% of the dataset are utilized for training, 25% for validating, and the remaining 25% for testing, except the Sonar and the Satimage. For each dataset, the final result is averaged over 20 runs of the algorithms. These datasets are briefly characterized in Table I.

TABLE II
PARAMETER SETTING

Method	Parameters	Value
ANN	Learning rate	0.3
	Number hidden nodes	20
	The iterations of Fine-tune	1000
	Alpha value	2
GA	The population size	20
	The probability of mutation	$1/\lambda$
E-SOCA	The iterations of E-SOCA	150
	The iterations of reproduction	50
	The Population size	100

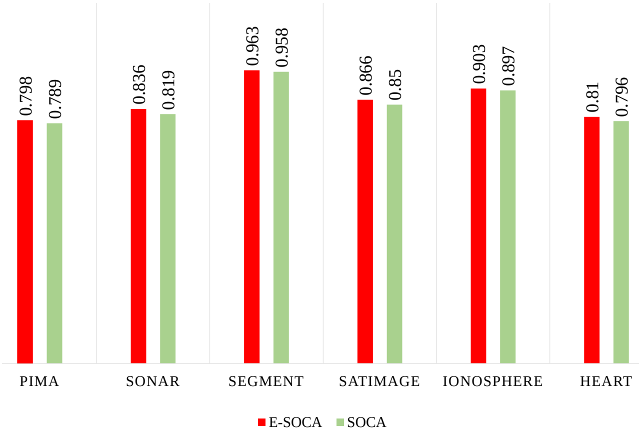


Figure 4. Testing accuracies of the proposed method E-SOCA comparing with the non-ensemble version SOCA.

2. Parameter Setting

The experiment parameters used in the proposed algorithm are in Table II, where λ is the chromosome length.

3. Results and Analysis

To evaluate the potential of the proposed algorithm we conduct some experiments as follows.

Experiment 1. Verify effect of the ensemble learning algorithm:

In order to measure the effect of the ensemble learning method, we compare the proposed method E-SOCA with a version which does not utilize the ensemble learning algorithm (named SOCA - Single Objective Co-Evolutionary Algorithm). As can be seen from Figure 4, the proposed method E-SOCA is better than the SOCA in all test cases. One common point among all 6 experimental data sets is that both methods give quite high results, especially in the Segment dataset, the correct classification results can reach 96.32% and 95.84% by the E-SOCA and the SOCA, respectively. Since both these algorithms differ only in ensemble learning step, only the classification results are different. However, in terms of criteria for the number of features selected, both methods are the same because they use Elite1 at the last step of the co-evolutionary process. Through this experimental result, the effect of community learning can be clearly seen. Especially with the Sonar dataset, thanks to this step, the classification accuracy has increased from 81.9% to 83.6%.

Experiment 2. Verify proposed algorithm with other conventional machine learning algorithms:

There are five machine learning algorithms (including the Naive Bayes - Bayes, the C4.5, the K-means, the K-neighbour algorithm (KNN), the Multilayer Perceptron (MLP). Two other algorithms are also used to compare with the proposed method including the probabilistic neural network (PNN) [25] — a 4-layers MLP widely used in classification problems, and the sequential minimal optimization (SMO) algorithm [26] — an algorithm used for training a support vector classifier. The empirical results with these algorithms are derived from [6].

Table III shows the results of proposed algorithms E-SOCA in comparison with six machine learning algorithms on six data sets. The bold value is the best one in each row. It can be easily seen that the E-SOCA algorithm achieves the best results in four out six test cases except for the Ionosphere and the Heart datasets. This shows the stability and power of the proposed algorithm in terms of classification accuracy criteria. Specifically, the proposed algorithm E-SCOCA outperforms the other ones with Pima dataset. It can reach 79.8%. Meanwhile SMO ranks the second, but its level of accuracy just reaches 77.26%. The remaining algorithms are lower, even the PNN can only reach 69.36%. In the Sonar dataset, the proposed E-SOCA algorithm continues to give the best classification results (83.6%), outperforming other methods. Meanwhile, the figures for the MLP is slightly slower, at 82.01%. In this dataset, PNN results in very low classification, only 53.2%. With the Segment dataset, the proposed algorithm E-SCOCA gives the best classification results compared to other methods. To be specific, it can achieve accuracy of 96.3%, while the C4.5 and the MLP are slightly lower, at 96.26% and 96.01% respectively. the KMeans produces the worst results, only 76%. With the Satimage dataset, the E-SOCA, PNN and SMO algorithms show quite similar results, the figures are 86.6%, 86.42%, 86.42%, and 86.58% respectively. Meanwhile, Bayes gives the worst classification results, reaching only 79.67%. The E- ranks the third compared to other algorithms on the Ionosphere dataset. In this case, the PNN gives the best results, at 94.24% while the KNN produces a very poor result, at only 79.13%. The result with the Heart dataset is similar to the Ionosphere dataset. The E-SOCA does not get the best results, just reaching 81%

TABLE III
TESTING ACCURACIES OF THE PROPOSED METHOD E-SOCA COMPARING WITH SOME OTHER MACHINE LEARNING ALGORITHMS.

Dataset	E-SOCA	Bayes	C4.5	K-means	KNN	MLP	PNN	SMO
Pima	**0.798**	0.7542	0.7345	0.7323	0.7149	0.7476	0.6936	0.7726
Sonar	**0.836**	0.6564	0.7251	0.7087	0.758	0.8201	0.5321	0.7733
Segment	**0.963**	0.8019	0.9626	0.76	0.9383	0.9601	0.9561	0.9229
Satimage	**0.866**	0.7967	0.8608	0.8282	0.8101	0.8614	0.8642	0.8658
Ionosphere	0.903	0.8307	0.8983	0.8848	0.7913	0.896	**0.9424**	0.885
Heart	0.81	**0.8505**	0.7636	0.8064	0.827	0.7922	0.7225	0.8406

while with this data the Bayes gives the highest value with 85.05%. Meanwhile, the PNN is the worst one with the figure is 72.25%.

Experiment 3. Verify proposed algorithm with other ensemble learning algorithms:

To verify the performance of the proposed method, we compare the proposed algorithm with some other ensemble learning algorithms such as the CO-NNE using the majority voting rule (CO-NNE (VOT)), AdaBoost (AB) [18], Bagging (BA) [27], Decision Forests (DF) [28] and Random Subspace method (RS) [29]. The experimental results of these algorithms are shown in [7]. In particular, the CO-NNE (VOT) (published in 2012) is the main algorithm proposed in this study while the other algorithms are renowned ensemble learning machine algorithms. We focus more on comparing the proposed method with the CO-NNE (VOT) algorithm, which is also an ensemble co-evolutionary approach, but the implementation method is completely different. As can be seen in Table IV (the symbol "–" indicates unavailable result), the E-SOCA significantly outperforms the CO-NNE (VOT) on three the datasets Pima, Sonar and Segment. There are improvements of 4.2%, 6.7% and 5.4% on testing accuracy respectively. Meanwhile, with the Ionosphere and Heart datasets, the CO-NNE (VOT) is slightly better than the E-SOCA. The figures for the difference are 1.6% and 1.1% correspondingly. In comparison with other ensemble learnings, the E-SOCA gets the best results on two data sets the Pima and the Sonar. Meanwhile, the DF reaches the best values on the Segment and the Ionosphere datasets with 97.4% and 93.1%, respectively. For the Heart dataset, the BA shows the best testing accuracy at 84.4%, outperforming other algorithms.

Experiment 4. Verify proposed algorithm with other state-of-the-art co-evolutionary algorithms:

To further verify the power of the proposed algorithm, we compare the E-SOCA to the latest recent dual population co-evolutionary algorithms. There are four algorithms used in this research in which two algorithms (the CEA-SBRBFNNH and the CEA-SBRBFNNS algorithms) were proposed in 2016 [8]. Two remaining algorithms (the DC-RBFNN, and the GA-RBFNN algorithm) were proposed in 2010 [6]. As can be seen in Table V, the E-SOCA significantly outperforms the other coevolutionary algorithms on most of the datasets except the Inono and Heart. Especially with the Sonar dataset, the proposed algorithm gives much better results than the remaining algorithms do. The accuracy reaches 83.6%. Meanwhile, the corresponding figures for CEA-SBRBFNNH, CEA-SBRBFNNS, DC-RBFNN and GA-RBFNN algorithms are 77.1%, 75.6%, 77% and 73.4%, respectively. With the Pima and Satimage datasets, the results of E-SOCA are slightly better than other algorithms. In contrast, with the Ionosphere dataset, the algorithm proposes the worst result, however, the difference between the algorithms is not much. The DC-RBFNN algorithm gives the highest outcome at 92.20% while the figure for the E-SOCA is 90.30%. While the two CEA-SBRBFNN versions produced higher results than others with 84.3% and 83.4%, the E-SOCA algorithm produced similar results with DC-RBFNN, reaching 81% and 81.6%, respectively. The GA-RBFNN algorithm gives the lowest result, at 78.2%. In terms of the number of selected features, this is also the biggest difference between co-evolutionary algorithms and other single algorithms (Bayes, C4.5, . . .). With the single algorithms, all features are selected to participate in the classification. Meanwhile, with the co-evolutionary algorithms only a certain number of features are selected to classify. If the number of features is large which can be up to hundreds, choosing the most important features will not only help improve classification results but also reduce time and space complexity. To assess the effectiveness of the proposed algorithm in reducing the number of features, we conduct a comparison with another co-evolutionary feature selection algorithm (the DC-RBFNN).

Figure 5 visually displays the results of the algorithms on these six sample datasets in terms of the feature selected of two algorithms. It can be easily seen that, in the DC-RBFNN algorithm, the number of selected features in the DC-RBFNN algorithm is much lower than the figure in the proposed algorithm, 4.167/8 compared to 6.45/8

TABLE IV
TESTING ACCURACIES OF THE PROPOSED METHOD E-SOCA COMPARING WITH SOME OTHER ENSEMBLE LEARNING METHODS

Dataset	E-SOCA	CO-NNE (VOT)	AB	BA	DF	RS
Pima	**0.798**	0.756	0.743	0.758	0.742	0.753
Sonar	**0.836**	0.769	0.825	0.762	0.799	0.720
Segment	0.963	0.909	0.922	0.961	**0.974**	0.903
Satimage	0.866	–	–	–	–	–
Ionosphere	0.903	0.919	0.933	0.914	**0.931**	0.916
Heart	0.81	0.821	0.806	**0.844**	0.801	0.827

TABLE V
TESTING ACCURACY OF THE PROPOSED METHOD E-SOCA COMPARING WITH STATE-OF-THE-ART COEVOLUTIONARY ALGORITHMS

Dataset	E-SOCA	CEA-SBRBFNNH	CEA-SBRBFNNS	DC-RBFNN	GA-RBFNN
Pima	**0.798**	0.77 0	0.769	0.758	0.750
Sonar	**0.836**	0.771	0.756	0.770	0.734
Segment	**0.963**	–	–	0.951	0.9448
Satimage	**0.866**	–	–	0.8648	0.861
Ionosphere	0.903	0.921	0.914	**0.922**	0.912
Heart	0.810	0.834	**0.843**	0.816	0.782

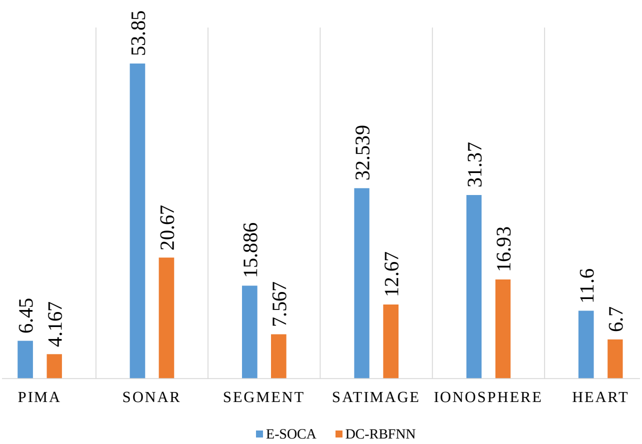


Figure 5. The number of features selected by the proposed method comparing with the DC-RBFNN.

in the Pima dataset; 6.7/13 compared to 11.6/13 in the Heart dataset; 7.567/19 compared to 15.886/19 in the Segment dataset; 16.93/34 compared to 31.37/34 in the Ionosphere dataset; 32.539/36 compared with 12.67/36 in the Satimage dataset and 53.85/60 compared with 20.67/60 in the Sonar dataset. It can be clearly seen that the larger the number of input features, the number of characteristics chosen from the DC-RBFNN algorithm remains stable (approximately 50% of the features are retained), while the figure for the proposed algorithm is variable and unstable. This is the weakness of the E-SOCA. The cause of this phenomenon will be explained more thoroughly in the Discussion section.

Experiment 5. Verify running times of proposed algorithm:

In this experiment, we will check that in the 3 main steps of the proposed algorithm including the co-evolution step, the community learning step (AdaBoost) and the Fine-tune step (BP), which step takes up computational time the most. The running time of each step with the data sets is shown in Figure 6. The unit of measurement in seconds. The computer configuration used here is core I7, 8Gb RAM. One thing that can be seen is that the co-evolution is the most time-consuming step. The remaining two steps take a very small time. The two columns represent the total computational time and the running time of the co-evolutionary step is always approximately the same on all data sets. As can be seen that the calculation time depends mainly on three factors: the number of co-evolution steps, the size of the training set and the size of the input feature set. The Heart data set (with a feature number of 13 and the number of training samples of 134) occupy the smallest running time, reaching 288.27 (s). Meanwhile, the Segment data set, with 19 input features, 1162 training samples, takes more time consuming than the Heart, occupying 1992.58 (s).

4. Discussion

In short, from the proposed algorithm model along with experimental results and comparisons with other algorithms. We can draw some points. First, the strength of the E-SOCA algorithm is that it is possible to give good accuracy classification results with high stability. This is partly because the E-SOCA uses a population of ANNs

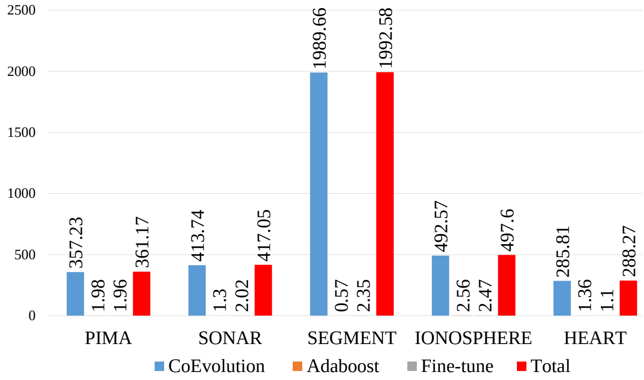


Figure 6. The computational time for each steps in the proposed method.

to evolve and the accuracy of ANNs in next-generations will be better. Another reason is the effect of the ensemble learning algorithm, by combining the weak learners into a strong learner, this algorithm contributes to the stability and also improves the classification accuracy.

Second, the weakness of the E-SOCA lies in the single-objective function. Although the E-SOCA uses two different populations, one for evolving the classifier and the other one for evolving the feature set. However, both populations use the same fitness function. This is the classification accuracy (MSE). This means both populations only focus on optimizing this goal, and the criteria for feature selection are still secondary. This cannot guarantee that the number of features selected will be gradually reduced. This is the main reason why the number of features retained by the E-SOCA is often large and uncontrollable. This problem can be handled by applying a multi-objective optimization algorithm (like the DC-RBFNN) instead of the single-objective algorithm being used in this version.

Third, time-consuming is also a drawback of the proposed method. Because the proposed method is currently still running sequentially, therefore the running time of this method will be greater when the size of the features as well as the training sample number increases. This problem can be solved by using parallel, multi-thread, and distributed computing solutions.

Finally, the E-SOCA algorithm provides a common model that can be applied to many different evolutionary algorithms for each population. In this study we are using the GA algorithms for evolving both populations, other multi-objective evolution algorithms can be applied to this model (such as NSGA-II). Besides, other ensemble learning algorithms can also be applied to the final set of Elites.

V. Conclusion

In this paper, an ensemble dual-population based co-operative co-evolutionary approach (E-SOCA) is presented. In the E-SOCA, two sub-populations are simultaneously maintained to achieve both good classification accuracy and prominent input features. These sub-populations interact with each other via an elitist selection mechanism then are combined by an ensemble learning method. The performance of the proposed algorithm is compared with the conventional machine learning algorithms and other state-of-art ensemble learning and coevolutionary algorithms on the six datasets. The empirical results on the test instances demonstrated the effectiveness of the new co-operative co-evolutionary approach for designing the ANN to solve the classification problems. This is our preliminary study, in the follow-up study, we will use parallel computing solutions to speed up the proposed algorithm as well as utilize the multi-objective optimization solution to further optimize both the number of selected features and classification results and in further research, we could apply this model to hyper-parameter searching in Deep Learning models.

References

[1] J. B. Pollack and A. D. Blair, "Co-evolution in the successful learning of backgammon strategy," *Machine Learning*, vol. 32, no. 3, pp. 225–240, Sep 1998.

[2] E. Uchibe and M. Asada, "Incremental coevolution with competitive and cooperative tasks in a multirobot environment," *Proceedings of the IEEE*, vol. 94, no. 7, pp. 1412–1424, July 2006.

[3] B. Y. Hiew, S. C. Tan, and W. S. Lim, "A double-elimination-tournament-based competitive co-evolutionary artificial neural network classifier," *Neurocomputing*, vol. 249, pp. 345 – 356, 2017.

[4] M. Castellani, "Competitive co-evolution of multi-layer perceptron classifiers," *Soft Computing*, vol. 22, no. 10, pp. 3417–3432, May 2018.

[5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems 27*. Curran Associates, Inc., 2014, pp. 2672–2680.

[6] J. Tian, M. Li, and F. Chen, "Dual-population based coevolutionary algorithm for designing rbfnn with feature selection," *Expert Systems with Applications*, vol. 37, no. 10, pp. 6904 – 6918, 2010.

[7] J. Tian, M. Li, F. Chen, and J. Kou, "Coevolutionary learning of neural network ensemble for complex classification tasks," *Pattern Recognition*, vol. 45, no. 4, pp. 1373 – 1385, 2012.

[8] J. Tian, M. Li, F. Chen, and N. Feng, "Learning subspace-based rbfnn using coevolutionary algorithm for complex classification tasks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 1, pp. 47–61, Jan 2016.

[9] G. Mauša and T. G. Grbac, "Co-evolutionary multi-population genetic programming for classification in software defect prediction: An empirical case study," *Applied Soft Computing*, vol. 55, pp. 331 – 351, 2017.

[10] A. Rosales-Pérez, A. E. Gutierrez-Rodríguez, S. García, H. Terashima-Marín, C. A. C. Coello, and F. Herrera, "Cooperative multi-objective evolutionary support vector machines

for multiclass problems," in *Proceedings of the Genetic and Evolutionary Computation Conference*, ser. GECCO '18. New York, NY, USA: ACM, 2018, pp. 513–520.
[11] A. Rosales-Perez, S. Garcia, H. Terashima-Marin, C. A. Coello Coello, and F. Herrera, "Mc2esvm: Multiclass classification based on cooperative evolution of support vector machines," *IEEE Computational Intelligence Magazine*, vol. 13, no. 2, pp. 18–29, May 2018.
[12] V. Van Truong, B. Lam Thu, and N. Trung Thanh, "A coevolutionary approach for classification problems: Preliminary results," in *2018 5th NAFOSTED Conference on Information and Computer Science (NICS)*, Nov 2018, pp. 81–86.
[13] L. T. Bui, V. T. Vu, and T. T. H. Dinh, "A novel evolutionary multi-objective ensemble learning approach for forecasting currency exchange rates," *Data & Knowledge Engineering*, vol. 114, pp. 40 – 66, 2018, special Issue on Knowledge and Systems Engineering (KSE 2016).
[14] T. T. H. Dinh, V. T. Vu, and T. L. Bui, "A multi-objective ensemble learning approach based on the non-dominated sorting differential evolution for forecasting currency exchange rates," in *2016 Eighth International Conference on Knowledge and Systems Engineering (KSE)*, Oct 2016, pp. 96–102.
[15] M. Mitchell, *An Introduction to Genetic Algorithms*. Cambridge, MA, USA: MIT Press, 1998.
[16] L. Rokach, "Ensemble-based classifiers," *Artificial Intelligence Review*, vol. 33, no. 1, pp. 1–39, Feb 2010.
[17] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119 – 139, 1997.
[18] ——, "Experiments with a new boosting algorithm," in *Proceedings of the Thirteenth International Conference on International Conference on Machine Learning*, ser. ICML'96. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1996, pp. 148–156.
[19] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16 – 28, 2014.
[20] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, 2007.
[21] M. Dash and H. Liu, "Feature selection for classification," *Intelligent Data Analysis*, vol. 1, no. 1, pp. 131 – 156, 1997.
[22] H. Liu and Z. Zhao, *Manipulating Data and Dimension Reduction Methods: Feature Selection*. New York, NY: Springer New York, 2012, pp. 1790–1800.
[23] H. Liu, H. Motoda, R. Setiono, and Z. Zhao, "Feature selection: An ever evolving frontier in data mining," in *Proceedings of the Fourth International Workshop on Feature Selection in Data Mining*, ser. Proceedings of Machine Learning Research, H. Liu, H. Motoda, R. Setiono, and Z. Zhao, Eds., vol. 10. Hyderabad, India: PMLR, 21 Jun 2010, pp. 4–13.
[24] B. Xue, M. Zhang, W. N. Browne, and X. Yao, "A survey on evolutionary computation approaches to feature selection," *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 4, pp. 606–626, Aug 2016.
[25] D. F. Specht, "Probabilistic neural networks," *Neural Networks*, vol. 3, no. 1, pp. 109 – 118, 1990.
[26] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya, and K. R. K. Murthy, "Improvements to platt's smo algorithm for svm classifier design," *Neural Computation*, vol. 13, no. 3, pp. 637–649, March 2001.
[27] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, Aug 1996.
[28] Tin Kam Ho, "Random decision forests," in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, vol. 1, Aug 1995, pp. 278–282 vol.1.
[29] ——, "The random subspace method for constructing decision forests," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832–844, Aug 1998.

Vu Van Truong received the B.S. degree in Geomatics from Le Quy Don Technical University, Ha Noi, Viet Nam in 2010 and the M.S. degree in Information systems from Le Quy Don Technical University, in 2014. He is currently pursuing the Ph.D. degree in Information technology at Le Quy Don Technical university. His research interest includes the Evolutionary Computations specialized with evolutionary multi-objective optimization, Machine learning, GIS and Remote sensing.

Bui Thu Lam received the Ph.D. degree in computer science from the University of New South Wales (UNSW), Australia, in 2007. He did postdoctoral training at UNSW from 2007 until 2009. He has been involved with academics including teaching and research since 1998. Currently, he is an Associate Professor at Le Quy Don Technical University, Hanoi, Vietnam. He is doing research in the field of evolutionary computation, specialized with evolutionary multi-objective optimization.

Nguyen Thanh Trung is a Reader in Operational Research and a co-director of the Liverpool Logistics, Offshore and Marine (LOOM) Research Institute. He is leading a research group on optimization, simulation and operational research, with focuses on maritime, transport (rail and road), logistics and IoT problems. He is working on solving various combinatorial problems such as vehicle scheduling, vehicle fleet sizing, vehicle routing, bin packing, berth planning, seat allocation for trains, minimizing train delays, crane trajectory optimizing, container stacking, vessel stowage planning, container loading, yard layout design problems. He is also leading research on using data analytics to study transportation systems and improve transportation operations.

INFORMATION FOR AUTHORS

Research and Development on Information and Communication Technology – *The MIC Journal of ICT Research*, or *ICT Research* in short – is a scientific publication of the Journal of Information and Communication, published by Vietnam Ministry of Information and Communications (MIC) since 1999. ICT Research is peer-reviewed, open-access, and published four issues a year: two in its *English Edition* (ISSN: 1859-3534), and two in its *Vietnamese Edition* (ISSN: 1859-3526) (Chuyên san Các công trình nghiên cứu phát triển Công nghệ Thông tin và Truyền thông).

The purpose of ICT Research is to provide a forum for researchers and professionals to disseminate original and innovative ideas in the fields of information technology, communications, and electronics in Vietnam and worldwide. Its Editorial Board is composed of renowned scientists from research institutions throughout Vietnam and other countries.

AUTHOR GUIDELINES

Authors are encouraged to follow good practice of technical paper writing, such as the guidelines on "How to write for technical periodicals & conferences" of the Institute of Electrical and Electronics Engineers.

Submissions must be made online on the website of ICT Research [https://ictmag.vn]. Authors should format the manuscript as closely as possible to the ICT Research manuscript template.

By submitting a manuscript, the authors are required to ensure that the submission has not been previously published, nor is currently submitted for publication elsewhere. Submission also implies that the corresponding author has consent of all authors. The submission file must be in PDF format. To faciliate double-blind review, authors are requested to submit two versions of the manuscript, one with full information of all authors, the other without.

Submissions of revised manuscripts should follow the same guidelines as for the initial manuscript, plus with a document entitled "Response to reviewers' comments" which explains the modifications according to the comments of the anonymous reviewers.

Upon formal acceptance of a manuscript for publication,the authors are required to prepare the final manuscript in LaTeX, using the IEEE Transactions style.

PUBLICATION ETHICS

ICT Research is committed to following the rules, standards, and guidelines set forth by the Committee on Publication Ethics (COPE) in ensuring the publication integrity and in handling cases of misconduct. All involved parties (Editors, Authors, Reviewers, and Publisher) are expected to be aware of and to conform to the respective rules, standards, and guidelines, available on COPE website.

The Editorial Board of ICT Research will immediately screen all articles submitted for publication. All submissions we receive are checked for plagiarism by using online available tools. Any type of plagiarism is unacceptable and is considered unethical publishing behavior. Such manuscripts will be rejected.

PEER REVIEW PROCESS AND PUBLICATION

An area editor of ICT Research (to be considered as the Editor from the point of view of the Authors) is assigned, by the Technical Editor-in-Chief, to each submission for coordinating the review process. The Editor pre-screens the submission and determines whether it is appropriate to the interests of readers of ICT Research. If appropriate, the Editor then initiates the review process. Each submission is then reviewed by at least two Reviewers. The review process is doubly blind, that is, identities of authors and reviewers are not revealed to each other.

Review decisions are: *(A) Accept submission* (i.e., no changes required), *(B1) Revision required* (i.e., minor revision), *B2) Resubmit for review* (i.e., major revision, the submission needs to be re-worked with major changes and is then subject to a second round of review), and *(C) Decline submission* (i.e., reject). Each major round of review may take approximately 2 months. Should authors be requested by the Editor to revise the manuscript, the revised version should be submitted within 6 weeks for a major revision or 2 weeks for a minor revision. No more than 1 major revision and 2 minor revisions are allowed.

An accepted article will be immediately published online as Early Access, with a DOI number, having been edited in compliance with ICT Research editing rules. ICT Research articles are published collectively as part of an issue with its own table of contents.

COPYRIGHT NOTICE